



Spectroscopic and dynamic properties of electronically excited retinal in C1C2 channelrhodopsin

INAUGURAL-DISSERTATION

zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

IRINA DOKUKINA

AUS ODESSA

Düsseldorf, September 2018

aus dem Institut für Theoretische Chemie und Computerchemie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. PD Dr. Oliver Weingart
2. Univ.-Prof. Dr. Christel M. Marian

Tag der mündlichen Prüfung: 26.10.2018

Eidesstattliche Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf erstellt worden ist. Die Dissertation wurde vorher bei keiner anderen Fakultät vorgelegt und es hat keine erfolglosen bzw. erfolgreichen Promotionsversuche gegeben.

Unterschrift:

Datum:

*The only way of discovering the limits of the possible is to venture
a little way past them into the impossible.*

Clarke's second law

Meiner Familie gewidmet

Danksagung

An dieser Stelle möchte ich mich bei den Menschen bedanken, die mich während meiner Doktorzeit begleitet haben. Ohne sie wäre die vorliegende Doktorarbeit nicht realisierbar.

In erster Linie möchte ich mich bei meinen Doktoreltern Professor Christel M. Marian und PD Dr. Oliver Weingart herzlich bedanken.

Professor Marian bin ich zunächst für die Aufnahme in den Bereich der Theoretischen Chemie dankbar und für die Möglichkeit das fehlende Hintergrundwissen nachzuholen, um den Aufgaben im Rahmen der Doktorarbeit gewachsen zu sein. Ich bin Ihr für die Übernahme der Erstbetreuung beim größten Teil meiner Promotion und für Ihr Engagement dankbar. Ich möchte mich bei Ihr auch für bereichernde wissenschaftliche Diskussionen bedanken. Diese gemeinsamen Gespräche hatten stets eine motivierende Wirkung und halfen mir herausfordernde Fragestellungen in einem anderen Licht zu betrachten. Nicht zuletzt bin ich Professor Marian auch für die Ermunterung dankbar, mir meiner Fähigkeiten und Kompetenzen bewusst zu werden.

Ich möchte mich bei PD Dr. Oliver Weingart für ein spannendes und vielseitiges Projekt bedanken, bei dem sich die meisten meiner wissenschaftlichen Interessen vereint haben. Zudem bin ich Ihm für die kontinuierliche fachliche Begleitung meiner Arbeit und für die Übernahme der Erstbetreuung am Ende meiner Promotion dankbar. Seine Hilfsbereitschaft (auch zur späten Stunde) und optimistische Art haben zum Fortschritt meiner Dissertation beigetragen. Er hat immer eine Lösung für manch ein fachliches oder technisches Problem gesehen, bei dem ich des öfteren anfang zu verzweifeln. Ich bin Ihm außerdem für die Möglichkeit dankbar, während der Zusammenarbeit die zahlreichen „Werkzeuge“ und Tricks kennengelernt zu haben, die einem nicht nur das Arbeiten vereinfachen, sondern auch für weitere berufliche Perspektiven von Bedeutung sein könnten.

Sowohl Professor Christel M. Marian als auch PD Dr. Oliver Weingart bin ich dafür dankbar, dass ich die Möglichkeit hatte an wissenschaftlichen Konferenzen teilzunehmen. Dadurch konnte ich mein Wissen erweitern, für mich interessante Bereiche kennenlernen und andere Wissenschaftler mit meiner Arbeit in Berührung bringen.

Herrn Professor Wolfgang Gärtner danke ich für das Einverständnis mein Mentor im Rahmen der Promotionsvereinbarung zu sein.

Bei der Arbeitsgruppe von Professor Marco Garavelli (Bologna) und insbesondere bei Dr. Artur Nenov möchte ich mich für Ihre Kooperation bedanken.

Mein herzlicher Dank gilt Jelena Föller, Oliver Schillinger und Fabian Dinkelbach für die entgegengebrachte Geduld bei meinen Fragen (fachlicher und nicht immer fachlicher Natur) und für Teerunden bei denen Angelegenheiten unterschiedlicher Wichtigkeit besprochen wurden. Ich bin froh nicht nur von netten Kollegen umgeben gewesen zu sein, sondern auch Freundschaften geknüpft zu haben. An dieser Stelle möchte ich mich auch

bei Gudrun Brauwers vielfach bedanken, die mich immer mit einem guten Rat und einem aufmunternden Wort unterstützt hat.

Ich möchte mich auch bei der gesamten Arbeitsgruppe der Theoretischen Chemie (sowohl bei den ehemaligen als auch bei den aktuellen Mitgliedern) bedanken, die unterschiedlich zu meinem Promotionsdasein beigetragen haben, aber insgesamt für ein angenehmes Arbeitsklima sorgten und immer bereit waren zu helfen.

Rita Sevcuk und Jelena Föller bin ich zudem für das Korrekturlesen meiner Arbeit und für fachliche und stilistische Anmerkungen dankbar.

Ein besonderer Dank gilt meiner Familie und meinem Freund Jakob Schlegel. Meinen Eltern (Elena und Gennadij Dokukin) und Großeltern (Johanna und Yevgen Gorbulski) bin ich für eine bedingungslose Unterstützung während meines Werdegangs noch lange vor der Promotion dankbar. Sie haben zu meiner Affinität für die Naturwissenschaften beigetragen und mit Ihrer Hilfe konnte ich meinen Interessen nachgehen. Jakob Schlegel bin ich dafür dankbar, dass er schon nicht die erste Abschlussarbeit mit mir durchgestanden hat und mich während dieser Zeit aufgebaut und mir den Rücken freigehalten hat. Auch für seine Geduld beim Zuhören und jeden praktischen Ratschlag bin ich ihm dankbar.

Bei meinen Freunden möchte ich mich für die emotionale Unterstützung bedanken und auch dafür, dass sie geduldig darauf warten bis ich wieder mehr Zeit für sie habe.

Für die Finanzierung meiner Arbeit möchte ich mich bei mehreren Menschen bedanken. Zum einen danke ich PD Dr. Oliver Weingart für die Gründung des Projekts und für das Erwerben der Förderung durch den Strategischen Forschungsfonds der Heinrich-Heine-Universität Düsseldorf. Bei Professor Christel M. Marian und bei Professor Lutz Schmitt möchte ich mich für das Ermöglichen der Finanzierung in der fortgeschrittenen Phase meiner Promotion bedanken. Anschließend gilt mein Dank Professor Christel M. Marian für die Finanzierung des letzten Abschnitts meiner Doktorarbeit.

Es war eine nicht immer einfache, aber eine für mich wichtige und lehrreiche Zeit, in der ich mich persönlich und wissenschaftlich weiterentwickelt habe.

Contents

List of publications included in the thesis	xiii
Abbreviations	xiv
Summary	xvii
Zusammenfassung	xviii
1 Introduction	1
1.1 Opsin proteins	1
1.2 Optogenetics	2
1.3 Channelrhodopsins	3
1.3.1 General description	3
1.3.2 Crystal structures	4
1.4 Retinal chromophore	5
1.4.1 Polyenes	5
1.4.2 Spectral properties of retinal in the protein environment	8
1.4.3 Quantum yield and reaction time of retinal isomerization	9
1.5 State of the art	11
2 Motivation and objectives	15
3 Theory and methods	17
3.1 Photochemical reactions	17
3.1.1 Relaxation processes after photoexcitation	17
3.1.2 Conical intersections	17
3.2 Quantum mechanical approach	20
3.2.1 Description of potential energy surfaces	20
3.2.2 Configuration interaction	22
3.2.3 CASSCF	23
3.2.4 CASPT2	23
3.2.5 DFT and DFT/MRCI	26
3.2.6 Semiempirical methods	28
3.3 Resonance Raman spectra	31
3.4 Molecular mechanical approach	32

3.5	QM/MM methods	34
3.5.1	The QM/MM energy	34
3.5.2	Handling of the QM/MM boundary	36
3.6	Nonadiabatic dynamics	37
3.6.1	Initial conditions	37
3.6.2	Trajectory surface hopping	38
3.6.3	Transition probabilities	40
3.6.4	Kinetic energy redistribution	41
3.6.5	Decoherence correction	41
3.6.6	Integration of Newton's equations with the velocity Verlet algorithm	42
4	Implementations (Paper III)	45
4.1	General description	45
4.2	Active space related problems	49
4.3	Adaptive time step algorithm by Spörkel and Thiel	49
4.4	Corrections introduced into the MNDO interface	49
5	Results	51
5.1	<i>Ab initio</i> and DFT QM/MM study (Paper I)	51
5.1.1	Ground state properties of retinal in C1C2	51
5.1.2	Absorption properties of retinal in C1C2	52
5.1.3	Geometry optimizations in the excited state	53
5.1.4	Resonance Raman spectra	54
5.2	Distinct excited state properties of linear polyenes and PSBs (Paper II)	55
5.2.1	The two S_1 minima of PSBs	56
5.2.2	OMx/MRCI at a test	58
5.3	Excited state dynamics simulations (Paper IV)	58
5.3.1	Static calculations	58
5.3.2	Photodynamics simulations	60
6	Conclusions and Outlook	67
6.1	Concluding remarks	67
6.2	Outlook	68
	Bibliography	71
	Publications	85

List of publications included in the thesis

• PAPER I

Spectral properties and isomerisation path of retinal in C1C2 channelrhodopsin

I. Dokukina, O. Weingart, *Phys. Chem. Chem. Phys.* **2015**, *17*, 25142-25150.

Own contribution: Execution and analysis of all calculations, preparation of supplementary material, generation of all figures, co-writing and proofreading of the manuscript.

• PAPER II

New perspectives on an old issue—a comparative MS-CASPT2 and OM2-MRCI study of polyenes and protonated Schiff bases

I. Dokukina, C. M. Marian, O. Weingart, *Photochem. Photobiol.* **2017**, *93*, 1345-1355.

Own contribution: Execution and analysis of all calculations, generation of all figures, preparation of the manuscript and supplementary material.

• PAPER III

COBRAMM 2.0 – A software interface for tailoring molecular electronic structure calculations and running nano-scale (QM/MM) simulations

O. Weingart, A. Nenov, P. Altoè, I. Rivalta, J. Segarra-Martí, **I. Dokukina**, M. Garavelli, *J. Mol. Model.* **2018**, *24*, 271-301.

Own contribution: Implementation of an interface for the MNDO program package, proofreading of the manuscript.

• PAPER IV

QM/MM photodynamics of retinal in the channelrhodopsin chimera C1C2 with OM3/MRCI

I. Dokukina, A. Nenov, M. Garavelli, C. M. Marian, O. Weingart, *manuscript submitted to: ChemPhotoChem* **2018**.

Own contribution: Execution and analysis of all calculations except pump-probe spectra, generation of all figures in the main paper and most of the figures for the supplementary section, providing data for calculation of pump-probe spectra, preparation of the manuscript and supplementary information.

Abbreviations

BO	Born-Oppenheimer
CASPT2	Complete Active Space Second-Order Perturbation Theory
CASSCF	Complete Active Space Self Consistent Field
CT	Charge Transfer
ChR	Channelrhodopsin
CoIn	Conical Intersection
DFT	Density Functional Theory
FC	Franck-Condon
GS	Ground State
HF	Hartree-Fock
HOMO	Highest Occupied Molecular Orbital
HOOP	Hydrogen Out-Of-Plane
LUMO	Lowest Unoccupied Molecular Orbital
MD	Molecular Dynamics
MM	Molecular Mechanics
MO	Molecular Orbital
MRCI	Multi Reference Configuration Interaction
MS	Multi State
OMx	Orthogonalization-corrected Methods
PSB	Protonated Schiff Base
QM	Quantum Mechanics
RPSB	Retinal Protonated Schiff Base
RR	Resonance Raman
SA	State Averaged
SE	Schrödinger Equation
TFS	Tully's Fewest Switches
TSH	Trajectory Surface Hopping

One letter amino acid code

<u>One letter code</u>	<u>Amino acid</u>
A	Alanine
C	Cysteine
D	Aspartic acid
E	Glutamic acid
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
W	Tryptophan
Y	Tyrosine

Summary

Channelrhodopsins (ChRs) are light-gated ion channels that play an important role in optogenetics. Their application ranges from basic neurobiological research to the development of therapies against severe degenerative diseases. The opening of an ion channel is triggered by all-*trans* to 13-*cis* isomerization of a co-factor, retinal, covalently bound to the protein. A detailed understanding of photoisomerization is particularly important for optimization of absorption and kinetic properties in ChRs. The research conducted within my thesis focused on the investigation of retinal isomerization in the chimeric channelrhodopsin C1C2. For this purpose, a combined quantum mechanical molecular mechanical (QM/MM) approach was applied. The current work comprises static calculations in ground and excited states of retinal as well as nonadiabatic excited state surface hopping dynamics. For the QM part, *ab initio* (HF, CASSCF, MS-CASPT2) and semiempirical strategies (DFT, DFT/MRCI, OMx/MRCI) were used. The MM part was described by the Amber force field. All QM/MM calculations were performed by using the in-house software package COBRAMM, which is developed in Bologna, Düsseldorf and Lyon. Static calculations provided first insight into the photoreaction and document the influence of the environment on the absorption properties of retinal. These calculations showed that highly demanding multireference strategies are necessary to properly describe the electronic properties of retinal in C1C2. I decided to use the semiempirical OMx/MRCI approach for my calculations because it provides qualitatively accurate results at low computational cost. To use OMx/MRCI with COBRAMM, I implemented an interface to the semiempirical program package MNDO developed by the group of Prof. Walter Thiel. Validation of the new implementation was performed on small linear polyenes and corresponding protonated Schiff base models. This study did not only confirm the applicability of the OMx/MRCI methodology for dynamics studies, but also provided new insights into the excited state properties of π -conjugated systems. As a final step, I performed OM3/MRCI dynamics of retinal within C1C2, which delivered quantum yields and photoproduct distributions along with time constants of the excited state processes. These calculations also provided details on the possible isomerization mechanisms. An intriguing finding concerns the relationship between the structure of retinal and the speed and efficiency of the photoisomerization. An additional twist in the chromophore structure leads to an increase in photoreaction speed. Accelerated photoreaction in turn resulted in a highly selective photoproduct formation and high quantum yield of the 13-*cis* retinal. This relationship might be used as a design strategy for novel ChR variants.

Zusammenfassung

Channelrhodopsine (ChR) sind lichtgesteuerte Ionenkanäle, die eine besondere Rolle in der Optogenetik spielen. Deren Anwendung reicht von neurobiologischer Grundlagenforschung bis zur Entwicklung von Therapiemethoden gegen degenerative Erkrankungen. Die Öffnung des Ionenkanals wird durch lichtinduzierte all-*trans* zu 13-*cis* Isomerisierung des kovalent an das Protein gebundenen Co-Faktors Retinal ausgelöst. Ein grundlegendes Verständnis der Photoreaktion ist besonders für die Optimierung der Absorptions- und der Kinetikeigenschaften der ChR von Bedeutung. Im Fokus meiner Dissertation lag die Untersuchung der Retinalisomerisierung im chimären ChR C1C2 mit kombinierten quantenmechanischen und molekülmechanischen Methoden (QM/MM). Die von mir durchgeführte Arbeit beschäftigt sich sowohl mit der statischen Beschreibung des Grundzustands und der angeregten Zustände des Retinals als auch mit der Simulation der nicht-adiabatischen *Surface-Hopping* Dynamik im angeregten Zustand. Das Spektrum der QM-Methoden umfasste *ab initio* (HF, CASSCF, MSCASPT2) und semiempirische (DFT, DFT/MRCI, OMx/MRCI) Ansätze. Die MM-Beschreibung erfolgte mit dem Amber-Kraftfeld. Für die Ausführung der QM/MM Berechnungen wurde das Programmpaket COBRAMM verwendet, welches in Bologna, Düsseldorf und Lyon entwickelt wurde. Statische Berechnungen lieferten einen ersten Einblick in den möglichen Verlauf der Isomerisierung. Außerdem konnte der Einfluss der Umgebung auf die Absorptionseigenschaften des Retinals erfasst werden. Die während der statischen Studie durchgeführte Methodenvalidierung hat die Notwendigkeit der Multireferenzbeschreibung für das Retinal in C1C2 aufgezeigt. Im Hinblick auf die anspruchsvolle Modellierung der Dynamik bot sich der Einsatz semiempirischer OMx/MRCI Methoden an. Diese liefern qualitativ gute Ergebnisse bei geringem Rechenaufwand. Um den OMx/MRCI Ansatz verwenden zu können, habe ich das Programmpaket COBRAMM um eine Schnittstelle zum semiempirischen Programmpaket MNDO (entwickelt in der Gruppe von Prof. Walter Thiel) erweitert. Die Validierung der OMx-Methoden vor der Dynamik-Studie erfolgte an kleinen symmetrischen Polyenen und entsprechenden Modellen für protonierte Schiff Basen. In dieser Studie konnte nicht nur die Eignung der OMx-Methoden für die Dynamik gezeigt werden, sondern es wurden auch neue Einblicke in die Eigenschaften von π -konjugierten Systemen gewonnen. Die Untersuchung der Isomerisierung wurde durch OM3/MRCI Dynamikberechnungen vervollständigt. Diese lieferten die Quantenausbeuten der Photoprodukte sowie deren Verteilung und entsprechende Reaktionskonstanten. Außerdem konnten mögliche Reaktionsmechanismen beschrieben werden. Zusätzlich wurde ein Zusammenhang zwischen der Retinalstruktur und der Geschwindigkeit und Effizienz der Photoreaktion festgestellt. Daraus ergibt sich, dass eine Vorverdrillung der Chromophorstruktur zur Beschleunigung der Photoreaktion führen kann. Eine beschleunigte Photoreaktion wiederum führte in meinen Berechnungen zu einer erhöhten Photoproduktselektivität und

einer hohen Ausbeute des 13-*cis* Isomers. Dieser Zusammenhang bietet neue Ansätze für die weitere Optimierung der Channelrhodopsine.

Chapter 1

Introduction

1.1 Opsin proteins

The ability to capture and respond to light has manifested itself in all domains of life. For this purpose, numerous organisms use organic chromophores as cofactors, which are covalently or non-covalently bound to a protein. [1, 2] In photosystems, pigments like chlorophyll, pheophytin and carotenoids are non-covalently bound to protein subunits enabling photosynthesis. [3] Retinal (vitamin A aldehyde) triggers the photoactivity of retinylidene proteins e.g. visual rhodopsins, channelrhodopsins or bacteriorhodopsin, all being members of the largest photosensitive protein family of opsin proteins, also generally referred to as rhodopsins. [4]

Opsins can be categorized into two subfamilies: microbial (type I) opsins and animal (type II) opsins. Both families originate from distinct genetic sequences. They, however, share a common scaffold: 7 transmembrane helices embedding a retinal cofactor. Retinal is covalently bound via a protonated Schiff base linkage (retinal protonated Schiff base, RPSB¹, Fig. 1.1) to a lysine side chain in one of the helices. [1, 4, 5] The initial retinal conformation is *all-trans* in microbial and *11-cis* in animal rhodopsins.

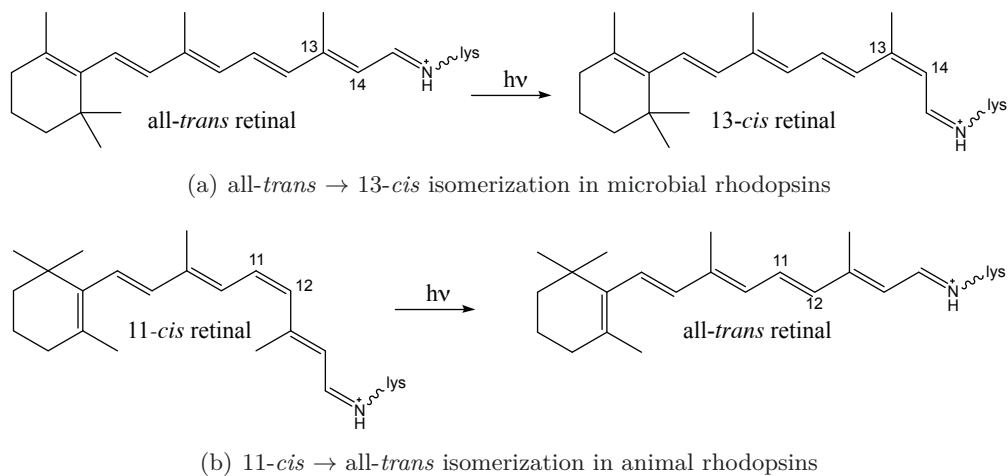


FIGURE 1.1: Retinal isomerization in type I (microbial) and type II (animal) opsins.

¹Within the current work RPSB is also termed just retinal.

Light induced isomerization of retinal (all-*trans*/13-*cis* in microbial [1] and 11-*cis*/all-*trans* in animal rhodopsins [6, 7] as shown in Figure 1.1) subsequently leads to protein activation. Animal rhodopsins are part of a G protein-coupled receptor superfamily. Thus, *cis-trans* isomerization is coupled to a G protein mediated signalling cascade. After isomerization, all-*trans* retinal disconnects from the protein and has to be replaced by a new 11-*cis* molecule. In contrast to animal rhodopsins, the initial all-*trans* retinal conformation of type I opsins is restored after isomerization during a cycle of thermal events. [4]

Considering their function, type I opsins can be subdivided into ion pumps, ion channels and transducer coupled light sensors (Fig. 1.2). [1, 5] Subsequent to retinal iso-

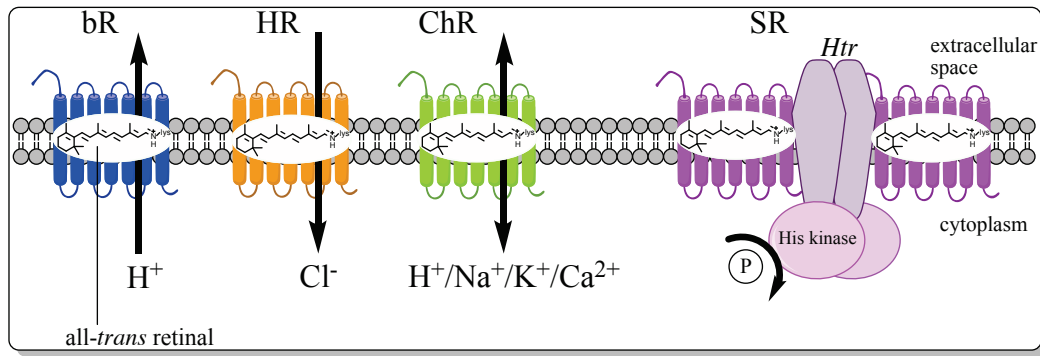


FIGURE 1.2: Proteins of type I opsin family: bacteriorhodopsin (bR), halorhodopsin (HR), channelrhodopsin (ChR) and sensory rhodopsin (SR) with indicated all-*trans* retinal chromophore. Image reproduced close to Ref. [5].

merization, ion pumps relocate ions across the membrane against the electrochemical gradient, while no intermediary proteins (like in animal rhodopsins) are involved. This way, they build up a membrane potential, which is used for ATP-synthesis. Thus, ion pumps serve as light-energy converters. The first discovered microbial opsins were the ion pumps bacteriorhodopsin (bR) [8] and halorhodopsin (HR) [9] from *Halobacterium salinarium*. Two more photosensitive proteins from *Halobacterium salinarium* are the sensory rhodopsins SRI and SRII. [10] Photoisomerization in SRs leads to activation of an associated transducer molecule *Htr*. Activated *Htr* initiates a phosphorylation cascade, which subsequently controls flagellar motion of the bacterium and affects its phototaxis (light-dependent motion). [5, 10] Light-gated ion channels termed channelrhodopsins (ChRs) form a pore in the cell membrane and conduct ions according to the electrochemical gradient. [11, 12] The first discovered ChRs were responsible for cation influx, thus leading to cell membrane depolarization.

1.2 Optogenetics

Ion conducting properties of light-sensitive pumps and channels lead to a change of membrane polarization and thus are closely related to neuronal signalling processes.

A logical consequence of such similarity was to introduce microbial rhodopsins into neurons enabling light mediated control of neural firing. Remote control of neurons by light was already attempted earlier by application of spatially guided lasers and multicomponent biological effectors. [13, 14] Nevertheless, a real success was achieved with microbial rhodopsins. It resulted in the development of optogenetics, a technology that genetically introduces photosensitive proteins into cells enabling their control by light. [15–17] The advantages of microbial rhodopsins in optogenetics are their one-gene-one-protein encoding (compared to multicomponent systems which are difficult to express in cells) and fast response to light. Before the discovery of ChRs, bR and HR were applied for selective silencing of neuronal signals due to their hyperpolarizing properties. [15, 16, 18] The discovery of ChRs not only extended the optogenetical toolbox, but also revolutionized the entire field by providing an activation tool due to their depolarization properties. Neurons equipped with ChRs can be fast and precisely switched on by light, enabling neuroscientists to study neuron communication and, e.g., understand brain activity. [19–22] A highly promising application of ChRs is the development of new therapies against, e.g. Parkinsons disease, heart diseases or even blindness. [18, 22–25] Application of ChRs is also highly desirable for neuron deactivation as an alternative to the ion pumps. The biggest advantage of ChRs compared to ion pumps is the stoichiometry of the ion conductance. While ion pumps transport less than one ion per activating photon, channels let a swarm of ions flow into the cell providing faster response to the light stimulus. [26] In recent years, several anion ChRs were discovered (see next section for details). A more detailed description of ChR properties will follow below.

1.3 Channelrhodopsins

1.3.1 General description

The roots of ChRs lead to the green alga *Chlamydomonas reinhardtii*, a unicellular organism inhabiting freshwater. Already in the early 20th century a connection between light perception of the eyespot and flagellar movement of algae was mentioned. [27] Several decades later (2002-2003), two proteins, Channelrhodopsin-1 (ChR1) [12] and Channelrhodopsin-2 (ChR2) [11], responsible for this behavior, were discovered by Nagel and co-workers. They revealed a new class of rhodopsins – light gated ion channels. ChR1 and ChR2 are permeable for positive ions (H^+ , Na^+ , K^+ , Ca^{2+}). Light-induced all-*trans* to 13-*cis* retinal isomerization (Fig. 1.1(a)) leads to structural changes within the protein resulting in channel opening. Subsequently, cations flow into the cell and depolarize it. Due to its high expression level, ChR2 became one of the most used ChRs in optogenetics. Later on (2009), two further ChR species, *Volvox* channelrhodopsin-1 (VChR1) and *Volvox* channelrhodopsin-2 VChR2 [28, 29] from the multicellular green alga *Volvox carteri* were described. While VChR2 shares similarity

with ChR2, VChR1 shows a more red-shifted absorption. Meanwhile, numerous ChR variants were discovered or identified via targeted genetic screening. Moreover, new variants with desired properties are steadily designed. Until recently, all discovered ChRs were conducting cations. In 2014, C1C2, a chimera protein of ChR1 and ChR2, was modified to form a Cl^- selective channel (iC1C2), providing a tool for inhibition of neuronal signals. [30] Further mutagenesis engineering of C1C2 and ChR2 resulted in new highly selective Cl^- channels (iChloC [31] and iC++ [32]). Later on, a naturally occurring anion conducting ChR (AChR) was reported. [13, 33] So far, light absorption of ChRs was mostly restricted to wavelengths in the blue-green spectral range. Nowadays, many red-shifted opsins exist including C1V1 (a chimera of ChR1 and VChR1), ReaChR (engineered red-activatable channelrhodopsin with VChR1 as a template) as well as MChR1 (*Mesostigma viride* ChR1), Chrimson (nicknamed due to its spectral sensitivity) and bReaCHES (ReaChR with mutations based on an accelerated ChR2 variant ChETA). (Ref. [13] and citations therein)

1.3.2 Crystal structures

C1C2

One of the biggest challenges in studying proteins is to obtain a crystal structure in high resolution. Structural information provides insight into the functionality of the particular protein and species related to it (homologues). In 2012, Kato *et al.* presented a crystal structure of a chimeric channelrhodopsin termed C1C2 at 2.3 Å resolution. [34] Until lately, it was the first and the only known crystal structure for channelrhodopsins. C1C2 comprises five helices of ChR1 and two helices of ChR2. In the membrane it forms a homo-dimer (Fig. 1.3(a)). As in all rhodopsin proteins, retinal is covalently bound to a lysine (K296), forming a protonated positively charged Schiff base. The protonated Schiff base terminus (NH^+) is exposed to a hydrophilic environment, while the polyene chain and the β -ionone ring experience influence of the hydrophobic residues. In the vicinity of the NH^+ -terminus two negatively charged counterions D292, E162, a positively charged lysine residue (K132) and a water molecule are located. The distance between the RPSB nitrogen and the counterions E162 and D292 is 3.4 Å and 3.0 Å, respectively (Fig. 1.3(b)). A similar arrangement is known from bacteriorhodopsin. In bR, a water molecule acts as a proton shuttle during deprotonation. It receives a proton from the Schiff base terminus and passes it to the counterion D85. However, in C1C2 the counterions are closer to the NH^+ -terminus than the water molecule, suggesting one of the counterions as a proton acceptor rather than the water molecule. Furthermore, Kato *et al.* conclude that D292 has the highest probability to accept the proton, while E162 is protonated from the beginning. They justify their hypothesis by the closer distance of D292 compared to E162 and also by results of a mutation study. While D292A mutation causes lack of photocurrent, substitution of E162 by alanine leads to a moderately decreased current. The arrangement of the negatively charged

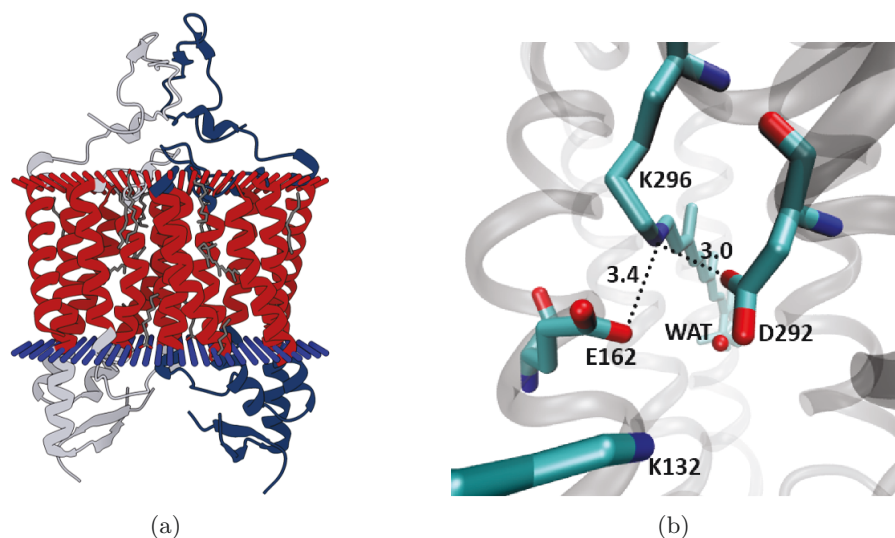


FIGURE 1.3: (a) C1C2 dimer in a membrane. Figure adopted from pdb.org (b) Residues close to the RPSB terminus in the C1C2 crystal structure.

residues close to the positive charge of the RPSB is suggested to be responsible for the blue-shift of the maximum absorption in ChRs compared to bR. In bR, negatively charged amino acids are located further away. [34]

ChR2

The investigation of ChRs is a fast developing research area. Recently, a crystal structure of ChR2 was obtained at a resolution of 2.39 Å. [35] Its general structure is similar to that of C1C2 and comprises a dimer consisting of two protomers. However, the authors note differences in length and helix orientation and in the number of water molecules compared to the C1C2 crystal structure. Furthermore, the presence of two retinal conformations, *all-trans*,15-*anti* and 13-*cis*,15-*syn* in the ground state (GS) was suggested. The distance between the RPSB nitrogen and the counterions E123 and D253 is 2.8 Å and 3.2 Å respectively, which is closer than in C1C2. Another difference refers to the number of gates and the gating mechanism in both proteins, which is not discussed in the current work. The differences between the two crystal structures are supported by recent electrophysiological and Fourier-transform infrared spectroscopy (FTIR) measurements, where C1C2 exhibited light-induced responses different from ChR2. [36]

1.4 Retinal chromophore

1.4.1 Polyenes

Linear π -conjugated systems (polyenes) have been a matter of research through decades involving theoretical and experimental studies. Naturally, they occur as cofactors in

photosensitive biological systems: β -carotene in light harvesting complexes of photosystems or retinal in animal and microbial rhodopsins. Moreover, polyene macrolide antibiotics are known for their antifungal properties. [37]. Figure 1.4 illustrates some naturally occurring polyenes. Within the scope of the current work, photochemical prop-

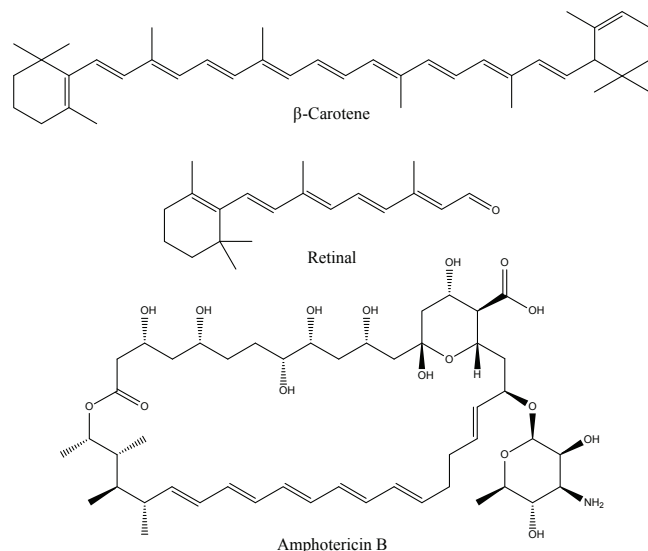


FIGURE 1.4: Different polyenic systems. β -carotene, retinal and the polyenic antibiotic amphotericin B.

erties of the retinal protonated Schiff base are of particular interest. Protonated Schiff bases (PSBs) are structurally related to symmetric linear polyenes (e.g. β -carotene or simple polyenes like butadiene, hexatriene etc.). However, their electronic properties differ in some aspects due to alteration of the π -conjugated system by an NH^+ terminal group. One of my studies included in the present work [38] compares symmetric linear polyenes (termed linear polyenes hereafter) and corresponding PSB models in vacuum (Fig. 5.5). The aim of the current section is to provide a general description of linear polyenes and PSBs relevant for understanding of the properties of retinal in a protein environment.

In the ground state (GS), the conjugated π -system leads to a geometric structure characterized by alternating double and single bonds. Primarily, the excitation wavelength depends on the length of the conjugated chain: The longer the chain the less energy is required for the excitation. With regard to retinal isomerization, the character of low-lying excited states is of particular interest. In all linear π -conjugated systems, one-photon absorption results from π - π^* excitation, which is mainly governed by the transition of one electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO). [39, 40]

In linear polyenes, the HOMO–LUMO state (also referred to as symmetry allowed) has $^1\text{B}_u$ symmetry. According to valence bond theory (VB), the nature of the $^1\text{B}_u$ state is ionic. [41] This gives rise to a large transition dipole moment responsible for

strong photoabsorbance of this state. The inner bond lengths of the 1^1B_u geometry are equalized compared to an alternating character of the GS structure (1^1A_g) as shown in Figure 1.5. [39, 42, 43] Close to the singly excited HOMO–LUMO state lies

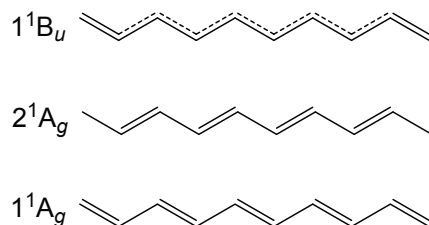


FIGURE 1.5: Schematic representation of bond length patterns in decapentaene. Relative to the ground state (1^1A_g), the 2^1A_g state is characterized by inverted bonds and the 1^1B_u state by an equalization of inner bonds.

an optically forbidden 2^1A_g state termed covalent by VB definition. [41] It entails a $\text{HOMO}^2\text{--LUMO}^2$ double excitation augmented by HOMO-1 to LUMO and HOMO to LUMO+1 single excitations. The geometry of the 2^1A_g state is characterized by reversing of the bond length pattern compared to the 1^1A_g state, thus double bonds gain single bond character and vice versa. [39, 44]

In Schiff bases, introduction of a positive charge through protonation leads to a red-shift of the HOMO–LUMO excitation compared to linear polyenes of the same length or equivalent unprotonated Schiff bases. [45] Due to this shift, absorption of RPSB emerges in the biologically relevant spectral range. Another consequence of the positive charge is a localization of the π -orbitals. In the HOMO of PSBs, more electron

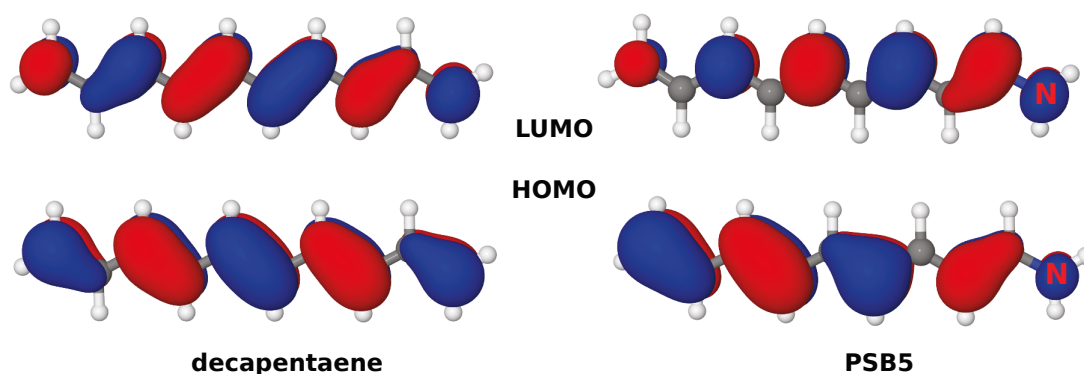


FIGURE 1.6: Highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals in decapentaene and the corresponding protonated Schiff base model PSB5.

density is distributed at the opposite side of the positively charged N-terminus and the opposite trend is found in the LUMO (Fig. 1.6). Thus, HOMO–LUMO excitation involves relocation of the positive charge from the nitrogen to the opposite end of the π -system, a so called charge transfer (CT) process. In contrast to linear polyenes, the ionic HOMO–LUMO state of PSBs is usually characterized by bond length inversion [46–48]

more similar to the 2^1A_g state of linear polyenes. I described this property in my second publication. [38] On the other hand, the doubly excited covalent HOMO²–LUMO² state is characterized by an equalization of the bond lengths similar to the 1^1B_u state in linear polyenes, a finding also documented in my second publication. [38] Moreover, in PSBs the interaction of HOMO–LUMO singly and doubly excited states is symmetry allowed, resulting in a non-zero absorbance of the latter.

1.4.2 Spectral properties of retinal in the protein environment

As already mentioned before, the spectrally visible absorption of retinal comes from HOMO–LUMO single excitation, which usually occurs in the first excited S_1 state. Primary factors determining the absorption of the RPSB are the length of the polyene chain and the protonation state of the Schiff base terminus (Sec. 1.4.1). The absorption spectra of rhodopsin proteins cover the entire visible spectral region. For example, the maximum absorption wavelength of C1C2 is ca. 470 nm [34, 49], while bR absorbs at ca. 570 nm [8]. This effect emerges from the influence of the protein environment, which can be electrostatic or steric in nature.

Electrostatic interaction of RPSB and the environment arises from the CT character of the HOMO–LUMO state (Sec. 1.4.1). In the ground state, the positive charge is located at the NH^+ -terminus. Upon excitation, it moves toward the β -ionone ring (Fig. 1.7). Thus, charges placed in the vicinity of the NH^+ -terminus (e.g. counterions)

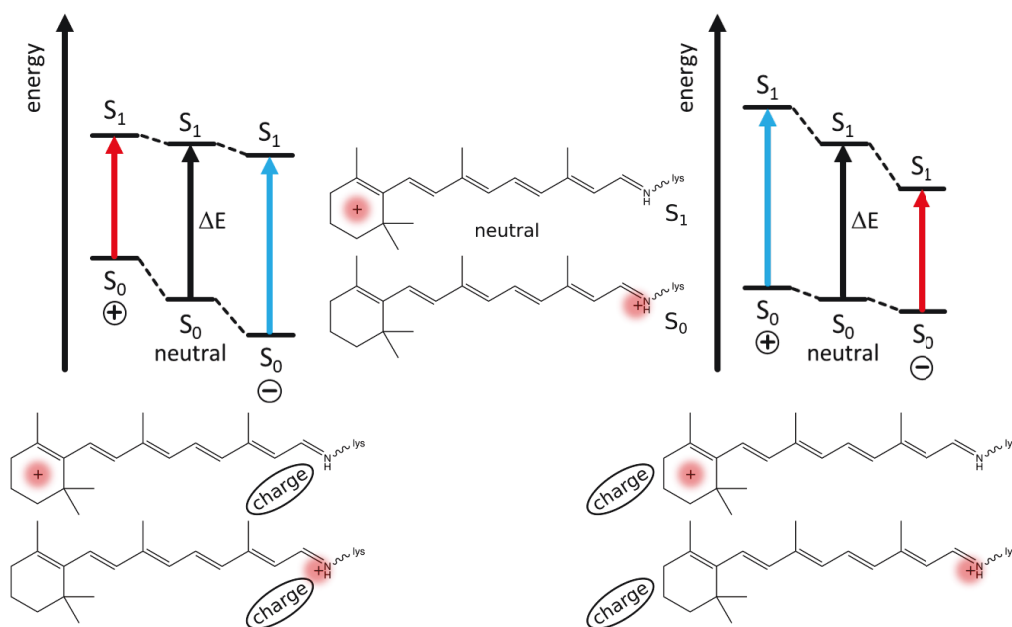


FIGURE 1.7: Influence of external charges on vertical excitation energies of retinal caused by a CT process assigned to the S_1 state. Left: Charges placed at NH^+ -terminus. Right: Charges placed at β -ionone ring. Blue arrows denote an increase of the excitation energies (blue-shift) and red arrows denote a decrease (red-shift) of the excitation energies relative to the neutral environment.

or near the β -ionone ring affect the CT process and retinal absorption compared to a neutral environment. [1, 50–52] Negative charges placed in the vicinity of the RPSB terminus stabilize the ground and the first excited state. In the GS, the distance between the positive NH^+ -terminus and surrounding negative charges is shorter than in the S_1 state. Consequently, stabilization of this state is stronger compared to the S_1 state. As a result, the energy amount required for the absorption of a photon increases compared to the non-charged surrounding. This effect is known as hypsochromic- or blue-shift. Red- or bathochromic-shift occurs when positive charge is located near the NH^+ -terminus. In this arrangement, stronger destabilization of the GS with regard to the S_1 state leads to a decrease of the absorption energy relative to a neutral environment. [53, 54] I noted these effects in my first publication. [55] If charges are placed near the β -ionone ring, the S_1 energy gets more affected relative to the ground state. Consequently, negative charges lower the S_1 energy stronger than the S_0 energy, causing a bathochromic-shift. In turn, positive external charges near the β -ionone ring cause the opposite effect, resulting in a hypsochromic-shift. [1]

Besides electrostatic interactions, the protein pocket can influence the absorption by affecting the geometry of the chromophore. Planar retinal achieves maximum conjugation and thus maximum length of the π -system. Such a structure would exhibit the maximum possible absorption wavelength. A twisted geometry may shorten the π -system, which would cause a blue-shift compared to the fully conjugated chromophore. This effect was already exploited for the design of blue-shifted ChR species. [56]

Electrostatic and steric interactions between retinal and its protein environment open numerous possibilities for color tuning of the chromophore. However, these effects occur simultaneously and their mutual influence has to be taken into account. For example, a change in the electrostatic environment in combination with some unfavourable geometry may not lead to the desirable result. In this case, detailed knowledge of the electronic and geometric structure is required, which must be provided through computational studies.

1.4.3 Quantum yield and reaction time of retinal isomerization

Retinal isomerization occurs in the excited state and is one of the fastest processes in photobiology (fs-timescale). In microbial rhodopsins it involves torsion about $\text{C13}=\text{C14}$ double bond. In the ground state, this reaction coordinate is characterized by a barrier too high for a thermal process. An ultrafast reaction is possible due to a conical intersection (CoIn) connecting the excited and ground states. Similar to the absorption efficiency, the speed of the reaction is influenced by the interaction of retinal with the protein environment. A pre-twist of the chromophore in the direction of the photoproduct may speed up the reaction. [57] Charges in the vicinity of retinal could influence the location of the CoIn or even cause it to vanish. [50]

Immediately after excitation, retinal leaves the Franck-Condon (FC) region following

two reaction coordinates: one involves bond length alternation and the second one a twist about the C13=C14 bond. Along the latter coordinate, retinal arrives at the S_1/S_0 CoIn. [51, 58, 59] From there, the reaction can proceed to the 13-*cis* photo-product or return to the starting material. The exact bond length pattern strongly depends on the electrostatic interactions of retinal with its environment (this effect is detailed in my studies [38, 55]). Irrespective of bond length inversion or bond length equalization, the C13=C14 bond would be elongated in the excited state facilitating the isomerization. Furthermore, efficient photoisomerization requires a charged counterion complex. [1, 60, 61]

Several computational studies in bovine rhodopsin and small retinal models illustrated the influence of a hydrogen out-of-plane (HOOP) motion on the outcome of the isomerization. [62–65] In particular, a correlation between the number of surface approaches prior to the hop event and photoproduct formation was reported. At the first surface approach, the HOOP-motion (Fig. 1.8(a)) and the reaction coordinate defined by skeletal carbon motion (Fig. 1.8(b)) are mostly in phase. In *cis*-PSBs, HOOP-torsion increases at this point. If the conical intersection is passed at the first surface approach,

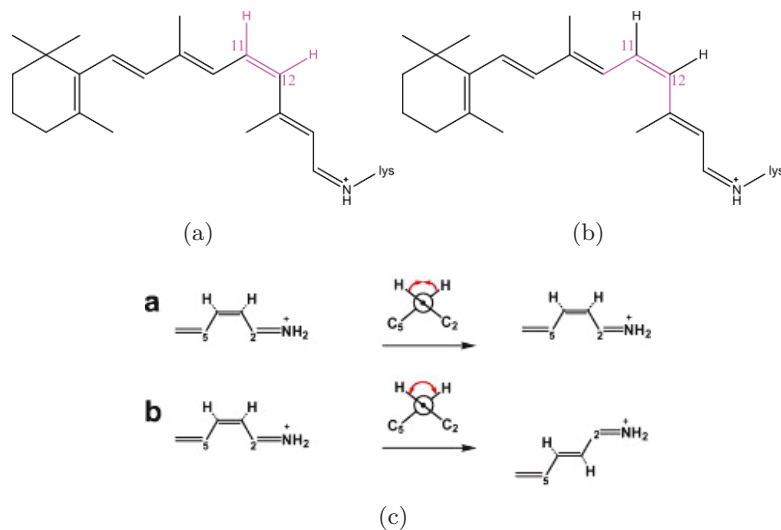


FIGURE 1.8: (a) HOOP and (b) reaction coordinates in 11-*cis* retinal denoted in magenta. (c) Decreasing HOOP torsion leading to *cis-cis* and increasing HOOP torsion leading to *cis-trans* reactions in PSB3. Image from Ref. [62].

positive HOOP-torsion is maintained in the ground state resulting in the preferred formation of the *trans*-photoproduct. At later surface approaches, HOOP and reaction coordinate torsions begin to dephase and the HOOP-torsion decreases as shown in Figure 1.8(c). Complete dephasing leads to equal probabilities for the formation of *cis*- and *trans*-products. I described this relationship for RPSB in C1C2. [66]

1.5 State of the art

The most important studies concerning the chimera protein C1C2 are chronologically summarized in this section. The interaction of retinal with the surrounding residues (E162, D292, K132) and the electronic properties of retinal are especially important in this context. Furthermore, protonation states of E162 and D292 are of particular interest with regard to electrostatic interactions between the positively charged retinal and the counterions. Moreover, recently reported findings about the isomerization of retinal in C1C2 are briefly reviewed.

2012-2013

Kamiya *et al.* [67] conducted molecular mechanical (MM) molecular dynamics (MD) simulations and computed QM/MM excitation energies using the newly published crystal structure of C1C2. In their study, the authors aimed to examine structural properties of the chimera protein and analyzed the electrostatic contributions of the surrounding residues to the spectral shift of retinal. The performed MD simulations revealed a hydrogen bond between the retinal NH^+ -terminus and the counterion E162. Analysis of electrostatic contributions has shown that major influence on the absorption comes from charged residues in the vicinity of retinal. The negatively charged E162, D292, and S295 provided a strong blue-shift and the positively charged K132 led to a strong red-shift of the maximum absorption. However, also charged residues at a larger distance to retinal influence its absorption, although at a lower rate. Based on their study, the authors suggested mutation strategies for engineering of new ChR color variants.

Sneskov and co-workers [68] performed polarizable embedding QM/MM mutation studies based on the C1C2 crystal structure. Similar to Kamiya *et al.*, the authors reported a significant blue-shift of the retinal absorption (0.3 eV) arising from the charged counterions (E162 and D292) compared to E162G and D292G neutral mutations. Furthermore, they noted that the same residues have major influence on the two-photon absorption properties of retinal, thus two-photon absorption might be an alternative to red-shifted ChR mutants. According to this study, application of polarizable embedding results in a ca. 0.21 eV lowering of the excitation energy.

2014

Ito *et al.* [69] investigated the hydrogen bonding network in the protein pocket of C1C2 by low-temperature (77K) FTIR spectroscopy. The FTIR measurements of C1C2 indicated that the counterions E162 and D292 are deprotonated. Absorption and FTIR measurements involving E162Q and D292N mutants confirmed the ionic nature of the counterions. Furthermore, FTIR results imply a direct hydrogen bond between NH^+

and E162. Another conclusion of the FTIR study is related to a water molecule in the proximity of the NH^+ -terminus. The obtained results suggest that this water molecule bridges D292 to the neighboring positively charged K132.

Pescitelli *et al.* [49] have reported experimental and calculated circular dichroism (CD) spectra based on the C1C2 monomer and dimer structures. The obtained spectra show an exciton coupling between retinal chromophores in the dimer. The authors furthermore provide time dependent density functional calculations (TDDFT) using three chromophore models. Their models refer to (1) RPSB in vacuum, (2) RPSB and counterions E162 and D292 in vacuum and (3) QM/MM calculations with RPSB and counterions in the QM-region and the MM-region containing residues within 10 Å from the atoms of the ion pair. The authors obtained an S_0/S_1 transition dominated by HOMO–LUMO π - π^* excitation for model (1), while for the models including counterions in the QM-region excitation is denoted as π - π^* like. Further observation refers to π and π^* orbitals partially localized on E162. The energy of the (crude) QM/MM model is blue-shifted compared to the experiment. Moreover, the authors mention a relation between the low polarity of the protein pocket and the vibronic fine structure of the spectrum.

2015

Takemoto and co-workers [70] investigated the channel functionality in C1C2 by electrophysiological experiments and MM/MD simulations. The authors also modelled C1C2 with a 13-*cis* retinal based on the bacteriorhodopsin 13-*cis* retinal structure to analyze conformational changes of the protein upon isomerization.

Kato *et al.* [56] designed blue-shifted variants of ChRs and archaerhodopsin-3 (AR3) employing theoretical and computational methods. The design of the ChR mutants was based on the crystal structure of C1C2. Kato and co-workers introduced mutations to enforce a twist about the C6-C7 bond in retinal. Such twist leads to a shortening of the π -system and results in a blue-shift of the absorption.

Inaguma *et al.* [36] discovered through difference FTIR measurements that C1C2 undergoes light induced conformational changes distinct from ChR2. Furthermore, the authors noted a transient protonation band of a carboxylate group, which may be assigned to either E162 or D292. This finding suggests both counterions as possible candidates for the role of a proton acceptor.

2017

VanGordon and colleagues [71] investigated intramolecular interactions in the closed state of C1C2 by MM/MD simulations. The authors paid particular attention to the gating region of the conducting pore. This study suggested D292 as a primary proton acceptor, but it also advises to consider E162 for this role.

Hontani *et al.* provided details of the retinal photoisomerization in C1C2 in their experimental study combined with QM/MM calculations (Fig. 1.9). [72] QM/MM computa-

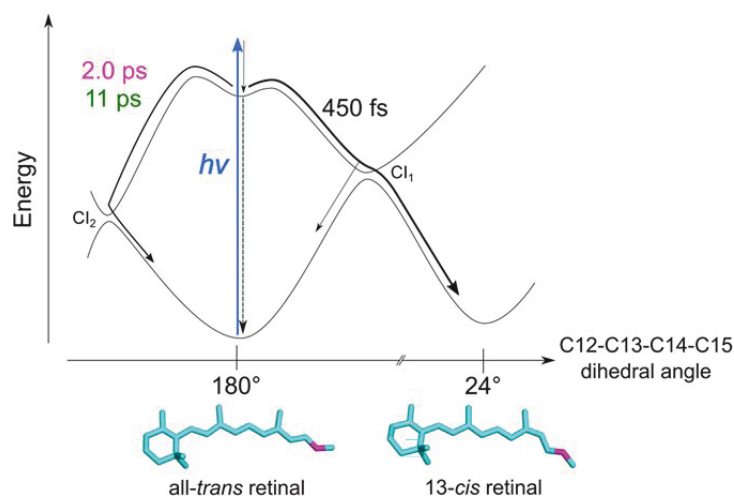


FIGURE 1.9: Model of excited state reaction dynamics of retinal in C1C2 suggested by Hontani *et al.* from Ref. [72].

tions were employed for geometry optimization, calculation of relaxed torsion scans and one trajectory without initial velocities (termed 0K trajectory). For the calculations, two initial models were used: one with NH^+ hydrogen bonding with the counterion E162 and a second one with a hydrogen bond between NH^+ and a water molecule. The authors reported three time constants (450 fs, 2 ps and 11 ps) for the photoreaction. They furthermore suggested relaxation to the ground state through two conical intersections. One conical intersection (termed CI₁) is suggested to result in the 13-*cis* retinal as photoproduct upon clockwise torsion, while the second CoIn (termed CI₂) is unproductive and leads back to all-*trans* retinal involving counter-clockwise torsion. A 450 fs time constant was assigned to the clockwise channel that yielded 50% of 13-*cis* and 50% of all-*trans* retinal. The slower 2 ps and 11 ps time components describe the return to all-*trans* retinal via CI₂. The overall quantum yield estimated for the 13-*cis* photoproduct with regard to all reaction channels comprises 30%.

Photoisomerization in microbial rhodopsins initiates a cycle of thermal reactions. During this photocycle, a series of retinal de- and reprotonation events take place, while stable photointermediates are formed. Hontani and co-authors suggested models for the C1C2 photocycle (Fig. 1.10(a)). An important aspect in the photocycle refers

to the isomeric composition of the ground state. In ChR2, the dark state is entirely composed of the all-*trans*,15-*anti* isomer. [73] However, there is evidence from Raman and nuclear magnetic resonance (NMR) measurements that upon minor illumination a second 13-*cis*,15-*syn* configured GS isomer emerges. The occurrence of two ground state isomers consequently leads to a branched photocycle. [74] The all-*trans*,15-*anti* isomer converts to 13-*cis*,15-*anti* upon excitation, while 13-*cis*,15-*syn* transforms to the all-*trans*,15-*anti* configuration (Fig. 1.10(b)). Similar findings were reported earlier for the bR photocycle. [1] For C1C2, no evidence for a second ground state isomer was

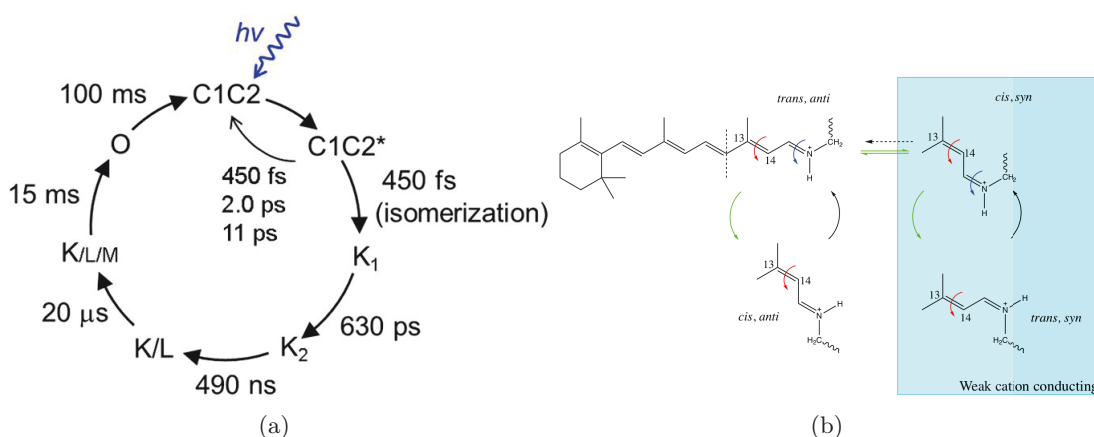


FIGURE 1.10: (a) Photocycle model proposed for the retinal in C1C2 at pH8 from Ref. [72]. (b) Configurations of RPSB in ChR2 reproduced from Ref. [13].

provided yet. Hontani *et al.* report only the all-*trans*,15-*anti* isomer as initial structure appearing in the ground state at their conditions.

2018

Recently, Adam and Bondar [75] published results of their QM/MM study based on the C1C2 crystal structure. The authors investigated proton transfer possibilities prior to photoisomerization. In particular, the main object of the study was to understand how the protein environment stabilizes the protonated state of the retinal Schiff base in the resting state and how it avoids early deprotonation by nearby charged counterions. The authors attribute an essential role to the neighboring lysine (K132). This residue showed a complex hydrogen bonding dynamics, which is reflected in a hydrogen bonding between K132 and water, K132 and E136, K132 and counterions (E162 and D292). Such interactions might have an impact on the protonation energetics resulting in unfavourable protonation of E162 and D292 prior to photoisomerization. In general, both counterions are considered for the role of the possible proton acceptor.

Chapter 2

Motivation and objectives

Channelrhodopsins are among the most important tools in optogenetics. Since their discovery, numerous studies were conducted to unravel structural-functional relationships within the channels and to apply this knowledge for efficient design of novel ChR variants. In this sense, theoretical atomistic studies play a crucial role. They provide "live" monitoring of the investigated system and also allow to introduce modifications in a time- and cost-saving manner (compared to laboratory studies). Furthermore, molecular simulations can provide predictions that may help to understand experimental results (e.g. spectra, kinetics) or to conceive new experiments.

A primary step in ChR activation and also one of the key targets for the design of new properties is retinal photoisomerization. For its control, a detailed understanding of spectral features (characterization of involved electronic states and geometries) and reaction parameters (speed, efficiency, possible photoproducts) is inevitable. A first insight into the photoreaction can be provided by static calculations. A quantitative characterization is, however, only possible through simulation of its excited-state dynamics. Such studies should preferably include a molecular ensemble, i.e. a relevant number of molecules to deliver information about photoproducts, their quantum yields, time constants, reaction channels and reaction mechanisms. Statistical analysis of such an ensemble delivers information about these quantities. Furthermore, dynamical simulations provide a basis for the investigation of factors influencing speed and efficiency of the reaction.

The major objective of my thesis was the detailed understanding of the retinal photoisomerization mechanism in C1C2. For the description of the chromophore in a protein environment, a combined QM/MM approach was chosen. QM/MM MD simulations in the microcanonical ensemble are known to be a valuable tool for understanding light-induced processes in photosensitive proteins like bR or visual rhodopsin. [76–78] Until yet, no such studies were reported for ChRs. An important step prior to dynamics studies is the choice of a QM method, which should provide a cost efficient, yet accurate description of electronic properties and potential energy surfaces. In this step, a detailed characterization of the structural and electronic properties in the ground and low-lying excited states is performed. Furthermore, the influence of the protein environment in the vicinity of retinal on these properties is investigated. The obtained

results are then compared to available experimental data (e.g. IR, Raman and absorption spectra) and other computational studies. Knowledge gained through static calculations is then applied to study retinal dynamics.

I started my thesis with static calculations on retinal within C1C2, which provided a first insight into its photoreaction and allowed for method validation with regard to further dynamical investigations. The results of this study are included in the first publication (Sec. 5.1 and Ref. [55]). The calculations therein refer to ground- and excited-state geometry optimizations, calculation of vertical excitation energies and screening of residues influencing the absorption properties. Results from the investigation of static properties and method validation have demonstrated the demand for multireference methodologies to properly describe the chromophore in C1C2. Furthermore, this study has shown the need for revisiting the protonated Schiff base properties with regard to structurally related symmetric linear polyenes. To include a multireference method feasible with regard to dynamics simulations, the COBRAMM software package was extended by a QM interface to the semiempirical methods provided in the MNDO program package. This implementation is documented in the third publication (Chap. 4 and Ref. [79]). To address the issue related to the difference between PSBs and linear polyenes and to test the semiempirical approach, a new study was conducted, which is included in the second publication (Sec. 5.2 and Ref. [38]). Finally, equipped with findings from my previous studies a simulation of retinal excited state dynamics in C1C2 was performed for the first time. The results of this study are captured in my final publication (Sec. 5.3 and Ref. [66]).

Chapter 3

Theory and methods

3.1 Photochemical reactions

In a photochemical sense, light provides energy to overcome activation barriers for which thermal energy in the ground state is not enough. Every photochemical reaction starts with the absorption of a photon and subsequent excitation to a higher lying electronic and vibronic level. Immediately after excitation, the excess energy is dissipated via different mechanisms that will be described below.

3.1.1 Relaxation processes after photoexcitation

In a static picture, the main events occurring upon excitation and relaxation to the ground state can be summarized in a Jablonski scheme. [80] The first process is vibronic relaxation (VR), which promotes the molecule rapidly and radiationlessly from the higher vibronic level to the lowest one within the same electronic state. From there, deactivation involves different electronic states and can be non-radiative or radiative. Non-radiative transitions within the same spin multiplicity proceed via internal conversion (IC), while radiative decay results in fluorescence. Non-radiative transitions between states of different multiplicities (e.g. singlet-triplet) are possible via intersystem crossing (ISC). If a triplet state is populated, it can relax non-radiatively via ISC, IC, VR or through photon emission via phosphorescence from the T_1 state.

At certain geometries potential energy surfaces may intersect, enabling radiationless transition (IC or ISC). A special case of such crossings are conical intersections (CoIns). [81] They are of immense importance in the photochemistry of ultrafast processes (e.g. photoisomerization), which are ubiquitous in photochemistry. [82]

3.1.2 Conical intersections

Usually, CoIns are pictured as funnels connecting ground and excited states that guide a molecule radiationless and fast between potential energy surfaces. In practice, the funnel is only one part of a high-dimensional picture. In diatomic molecules, states of the same symmetry are not allowed to be degenerate (noncrossing rule). [83] States of different spatial symmetry can, however, cross. In polyatomic molecules, states of same

as well as of different symmetries can intersect. [84, 85] States of different spatial symmetries cross in an $(m-1)$ -dimensional space, while states of the same symmetry lead to a CoIn in $m-2$ dimensions, where m describes the number of degrees of freedom in the particular molecule. [86, 87] In multidimensional space, CoIns can be described as points on a $(m-2)$ -dimensional hyperline, referred to as a crossing seam. To derive the crossing rules, two electronic states representing potential energy curves that may cross are considered. They are described by two arbitrary functions ϕ_1 and ϕ_2 that are orthogonal and build an orthonormal basis with already known solutions of the electronic Schrödinger equations. [84] Consequently, each of the two unknown eigenfunctions can be expressed as a linear combination of ϕ_1 and ϕ_2 :

$$\psi = c_1\phi_1 + c_2\phi_2 \quad (3.1)$$

To obtain energy eigenvalues, a secular equation is used:

$$\begin{bmatrix} H_{11} - E & H_{12} \\ H_{21} & H_{22} - E \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = 0 \quad (3.2)$$

where $H_{ij} = \langle \phi_i | H | \phi_j \rangle$. The energy eigenvalues are evaluated as following:

$$E_{1,2} = \frac{H_{11} + H_{22}}{2} \pm \sqrt{\left(\frac{H_{11} - H_{22}}{2}\right)^2 + H_{21}H_{12}} \quad (3.3)$$

Equation (3.3) allows to derive the conditions for a crossing of the two states. At the crossing point, the energies of the involved states are equal, i.e. $E_1 = E_2$, which happens when two conditions from the Equation (3.3) are simultaneously fulfilled:

$$H_{11} = H_{22} \wedge H_{12} = 0 = H_{21} \quad (3.4)$$

Equation (3.4) requires at least two independent molecular degrees of freedom x_1 and x_2 . In diatomic molecules, only one degree of freedom (the interatomic distance) exists. Consequently, states of the same symmetry would never cross leading to the noncrossing rule. This rule does not apply to the case of different symmetries, because H_{12} is already equal to zero and only one condition has to be met. In polyatomic molecules, it is generally possible to find two independent degrees of freedom which satisfy the conditions from Equation (3.4). In the case of different symmetries, the intersection space is $(m-1)$ -dimensional for the same reason as for diatomic molecules. For states of the same symmetry it leads to a crossing in the $(m-2)$ -dimensional space. In the $(m-2)$ -dimensional intersection space, degeneracy is preserved even if coordinates are varied. [85] The remaining degrees of freedom x_1 and x_2 span the so-called branching plane. Motion in this plane lifts the degeneracy. Plotting the potential energy against x_1 and x_2 leads to the well known double cone shape of the crossing potential energy

surfaces (PESs) (Fig. 3.1(a)). [81, 82] In the beginning of this section, I mentioned that CoIns are points along the crossing seam. A relaxation of a molecule through a conical intersection occurs in the x_1/x_2 plane when the reaction path of the investigated system crosses the seam as shown in Figure 3.1(b). The branching plane is associated with

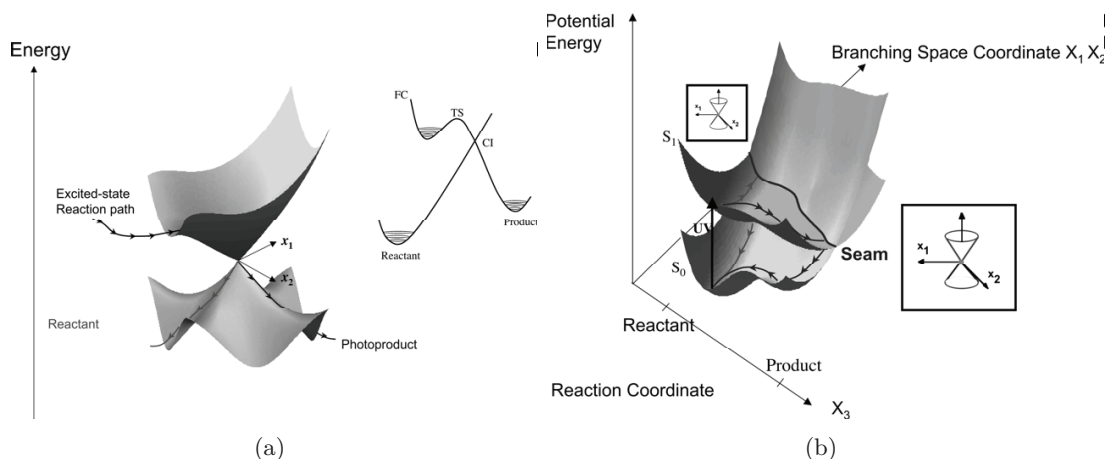


FIGURE 3.1: Different representations of conical intersections: (a) representation as a double cone; (b) conical intersection in a space spanned by the reaction coordinate X_3 and branching coordinate X_1X_2 . Figures from Ref. [82].

the molecular motion after exiting the degeneracy region. Consequently, the outcome of the photochemical reaction is influenced by the branching vectors. The branching space coordinates x_1 and x_2 can be described as gradient difference vector and gradient of the interstate coupling vector. [86, 88] The x_1 direction is referred to as the gradient difference vector defined as

$$x_1 = \frac{\delta(E_1 - E_2)}{\delta R} \quad (3.5)$$

It is also called tuning mode, since it adjusts the energy gap between the interacting states. The x_2 direction is given as the gradient of the interstate coupling vector

$$x_2 = \left\langle \psi_i \left(\frac{\delta H}{\delta R} \right) \psi_j \right\rangle \quad (3.6)$$

where ψ_i and ψ_j are electronic wave functions, H is the electronic Hamiltonian and R denotes a nuclear configuration vector. The vector x_2 is also referred to as a coupling mode, since it describes coupling of the interacting states. Furthermore, x_2 is parallel to the nonadiabatic coupling vector f_{ij} . [86] The relation between f_{ij} and x_2 is expressed as:

$$f_{ij} = \left\langle \psi_i \left| \frac{\delta \psi_j}{\delta R} \right. \right\rangle = \frac{\left\langle \psi_i \left(\frac{\delta H}{\delta R} \right) \psi_j \right\rangle}{E_1 - E_2} \quad (3.7)$$

where E_1 and E_2 are the energy eigenvalues. Equation (3.7) shows that nonadiabatic coupling (also denoted as derivative coupling) becomes important when the interacting

states become close in energy. The derivative coupling is required to describe nonadiabatic nuclear dynamics (Sec. 3.6). [89–91]

3.2 Quantum mechanical approach

3.2.1 Description of potential energy surfaces

Born-Oppenheimer Approximation

The time-independent Schrödinger equation (SE) can be written as

$$H(r, R)\Psi(r, R) = E\Psi(r, R) \quad (3.8)$$

where r corresponds to electronic and R to nuclear coordinates. H is the total nonrelativistic Hamiltonian of the system, which can be described as

$$H(r, R) = T_{nuc} + H_e(r; R) \quad (3.9)$$

where T_{nuc} denotes the nuclear kinetic energy operator and $H_e(r; R)$ corresponds to the electronic Hamiltonian that depends parametrically on R :

$$H_e(r; R) = T_e + V_{ee} + V_{en} + V_{nn} \quad (3.10)$$

The electronic Hamiltonian includes electronic kinetic energy (T_e) and Coulomb interactions (electron-electron repulsion (V_{ee}), nucleus-nucleus repulsion (V_{nn}) and electron-nucleus interaction (V_{en})). [92] Analytic solution of the SE for molecular systems with more than two particles is not feasible. The Born-Oppenheimer (BO) approximation is used to simplify this task. [93] It is based on the fact that electrons are much lighter than nuclei and thus both move on different time scales. Particularly, electrons move much faster than nuclei. The coupling between nuclear and electronic degrees of freedom is ignored. Consequently, the total wavefunction is approximated as a product of nuclear $\chi(R)$ and electronic $\psi(r; R)$ wavefunctions:

$$\Psi(r, R) = \psi(r; R)\chi(R) \quad (3.11)$$

Hence, the movement of both particles can be handled separately. Using this assumption, the total SE is separated into electronic and nuclear problems. From the electrons "point of view" nuclei move so slow that they can be assumed fixed. In conjunction with this assumption, the kinetic energy of the nuclei is neglected ($T_{nuc} = 0$) and the repulsion between the nuclei (V_{nn}) is considered constant. This approximation leads to the formulation of the electronic Schrödinger equation:

$$H_e(r; R)\psi_i(r; R) = E_i^e(R)\psi_i(r; R) \quad (3.12)$$

Within these assumptions, the electronic wave function explicitly depends on electronic coordinates, while its dependence on nuclear coordinates is parametric. This is also true for the electronic energy. The electronic SE is then solved for fixed nuclei arrangements and provides eigenfunctions and potential energies. The potential energies with corresponding nuclear coordinates yield a potential for nuclear motion known as potential energy surface. The solution of the electronic SE is a common task in quantum chemistry, which is addressed by application of various quantum mechanical methods. [92] The solution of the problem for the motion of the nuclei is based on the same assumptions as formulated for the electrons. Since electrons move much faster than nuclei, electronic coordinates are substituted by values averaged over the electronic wave function. This leads to an equation for the nuclear Hamiltonian and the nuclear SE. Solution of the nuclear SE provides information about rotation, vibration and translation of a molecule. [87, 92] The BO approximation is a key concept in quantum mechanics and delivers reliable results for the ground state and in cases, where potential energy surfaces do not cross.

Breakdown of the Born-Oppenheimer Approximation²

The BO approximation is valid as long as the described PESs are well separated. In the case of energy degeneracy (e.g. at conical intersections), the coupling between nuclear and electronic degrees of freedom becomes non-negligible. To describe these regions the BO approximation has to be extended by nonadiabatic terms. For this purpose, the molecular wave function is not represented as a simple product of electronic and nuclear wave functions, but rather as an expansion in a basis of electronic Born-Oppenheimer states:

$$\Psi(r, R) = \sum_i \chi_i(R) \cdot \psi_i(r; R) \quad (3.13)$$

where $\chi_i(R)$ corresponds to nuclear and $\psi_i(r; R)$ to electronic wavefunctions. The above equation is referred to as Born-Huang expansion [94] or Born-Huang ansatz and is exact when not truncated. [82] To obtain $\chi_i(R)$, Equation (3.13) is inserted into Equation (3.8), multiplied by $\psi_j^*(r; R)$ and integrated over the electronic coordinates. The resulting SE has the following form:

$$\left[T_{nuc} - \frac{1}{2\mu} K_{ii}(R) + E_i^e(R) \right] \chi_i(R) - \sum_{j \neq i} \frac{1}{2\mu} [K_{ij}(R) + 2f_{ij}(R) \cdot \nabla] \chi_j(R) = E \chi_i(R) \quad (3.14)$$

where $T_{nuc} = \frac{-1}{2\mu} \nabla^2$, ∇ is a gradient over the nuclear coordinates R and μ is a reduced mass. K_{ij} and f_{ij} are referred to as coupling terms, which are ignored in the BO approximation. They describe nonadiabatic transitions between states i and j and are

²Equations in this section are kept close to Ref. [87].

expressed as follows:

$$f_{ij}(R) = \langle \psi_i(r; R) | \nabla \psi_j(r; R) \rangle \quad (3.15)$$

$$K_{ij}(R) = \langle \psi_i(r; R) | \nabla^2 \psi_j(r; R) \rangle \quad (3.16)$$

where integration is done over electronic coordinates r . The derivative coupling vector f_{ij} expresses the efficiency of transitions between the states i and j . Furthermore, it depends on the energy difference between the states i and j (Eq. (3.7)). The derivative coupling increases with decreasing energy difference. The second order (scalar) coupling term K_{ij} is often neglected. K_{ii} describes the nonadiabatic correction to a single PES, which is usually also neglected. The diagonal term f_{ii} is zero for real wavefunctions. [87]

3.2.2 Configuration interaction

Consideration of excited state configurations in addition to the ground state wave function improves the ground state energy by adding correlation effects. Furthermore, it allows computation of the excited states. This approach is referred to as configuration interaction (CI) and expresses the wave function as a linear combination of ground state Hartree-Fock (HF) and excited determinants Ψ_i , weighted by expansion coefficients c_i :

$$\Psi_{CI} = c_0 \Psi_{HF} + \sum_S c_S \Psi_S + \sum_D c_D \Psi_D + \sum_T c_T \Psi_T + \dots = \sum_{i=0}^n c_i \Psi_i \quad (3.17)$$

where S, D, T etc. denote singly, doubly, triply etc. excited configurations relative to the HF ground state determinant. [92, 95] Each determinant is represented by a configuration state function (CSF), which is a linear combination of determinants describing equivalent states with the same spin multiplicity. Each total wave function represents one of the possible electronic states. Weighting coefficients determine the contribution of the CSFs to the particular total wave function, e.g. for the ground state the weight of the ground state determinant would be the largest. The energy is minimized in a variational procedure. In the case where all possible excitations are considered, the approach is called full CI. It delivers exact results within the used basis set and would deliver the exact ground state energy in the non-relativistic approximation and if an infinitely large basis set is used. [95] Full CI is computationally expensive and applicable only to small molecules. Usually, it is restricted to important terms, e.g., CI singles would consider all singly excited determinants, CI doubles all double excitations etc. In the CI-scheme, only CI coefficients are optimized, which means that HF orbitals are used throughout the calculation. This strategy may not be ideal for the description of excited states. An improvement can be achieved by application of a multiconfigurational SCF (MCSCF) approach.

3.2.3 CASSCF

MCSCF is a variant of the CI implementation, where CI coefficients and molecular orbitals are optimized to obtain the final wave function. [95] MCSCF can be referred to as truncated CI, where the restriction concerns the selection of the active orbital space. [96] A variant of the MCSCF method in the current work is the complete active space SCF (CASSCF) approach which will be described in the following. [97] A key concept of CASSCF is the partitioning of the orbital space into inactive and active spaces. The inactive space can be further subdivided into inactive occupied and inactive virtual orbitals. Orbitals in the inactive occupied space always contain two electrons during the calculation. The inactive virtual space contains orbitals that always remain unoccupied during the calculation. In the active space, a full CI calculation is performed and occupation numbers can vary between zero and two. The total CASSCF wavefunction is expressed by a linear combination of active CSFs. The usual notation is (n,m)-CASSCF denoting n electrons distributed over m orbitals. The selection of active orbitals is an individual problem and there is no unique rule to address this issue. The limiting factor is the number of CSFs, which grows factorially³ with n and m, restricting the manageable size of the active space currently to 16 orbitals with the same or smaller number of electrons. [99] Hence, a proper and sensible orbital selection is mandatory. A practical advice is to consider all orbitals that may be involved in the investigated reaction, e.g. if π to π^* transitions are studied, then, if possible, all π orbitals should be selected.

To provide a proper description for situations where root-switching occurs (e.g. at a conical intersection or unbalanced orbital description) the state-averaging CASSCF (SA-CASSCF) approach may be used. [100, 101] Its basic idea is to optimize the CASSCF wave function with regard to the energy averaged over all involved states. The energy of the SA wave function is given by

$$E_{av} = \sum_i \omega_i E_i \quad (3.18)$$

where ω_i refers to the weight of the i -th state. As a result, one set of molecular orbitals for all states is obtained. Consequently, CI coefficients are optimized for each state, using the same set of averaged orbitals. This strategy provides a balanced description for the simultaneous calculation of multiple states and also improves wave function convergence.

3.2.4 CASPT2

CASSCF accounts for static correlation, which is particularly important for the qualitative description of potential energy surfaces. However, dynamic correlation is important

³According to Weyl's formula. [98]

for a quantitative estimation of energies. Thus, to obtain satisfactory results, inclusion of both, static and dynamic correlation is necessary. The complete active space second order perturbation theory (CASPT2) method developed by Roos and co-workers includes static effects through the CAS reference and takes care of dynamical correlation through second order perturbation. [102, 103] Despite its successful application to a variety of molecular systems, the original CASPT2, also referred to as single-state (SS-CASPT2), can be inadequate in situations where the CASSCF wave function is not a good reference state. This could be the case when strong interaction between the electronic states is present, e.g. at (avoided) crossings or through Rydberg-mixing. To tackle these problems, Finley and co-workers suggested an extension of SS-CASPT2 to multistate MS-CASPT2. [104] It uses a multidimensional reference space, spanned by several SA-CASSCF states. In the following, SS- and MS-CASPT2 will be described in more detail based on the original publication of Finley *et al.*

SS-CASPT2

In perturbation theory, the Hamiltonian is expressed as following:

$$H = H_0 + \lambda V \quad (3.19)$$

where H_0 is a zeroth-order unperturbed Hamiltonian with known eigenfunctions, V is the perturbation operator and λ is a dimensionless parameter. λV describes a perturbative correction of H_0 . Furthermore, the Hilbert space is partitioned into a reference space P and a secondary space Q . The reference space P is one-dimensional and spanned by a CASSCF reference state $|\alpha\rangle$:

$$P = |\alpha\rangle \langle\alpha| \quad (3.20)$$

The reference state $|\alpha\rangle$ is an eigenfunction of the zeroth-order Hamiltonian

$$H_0 |\alpha\rangle = E_0^\alpha |\alpha\rangle \quad (3.21)$$

The second order correlation energy is given by

$$E_{2nd}^\alpha = E_{1st}^\alpha + \langle\alpha|H\Omega_1^\alpha|\alpha\rangle \quad (3.22)$$

where Ω_1^α is a first order wave operator, based on the following equation:

$$(E_0^\alpha - H_0)\Omega_1^\alpha |\alpha\rangle = QH |\alpha\rangle \quad (3.23)$$

Operating on the zeroth-order state $|\alpha\rangle$, the wave operator can create the exact state of interest $|\Psi\rangle$. Furthermore, it is completely determined by H_0 and $|\alpha\rangle$. In SS-CASPT2 $|\alpha\rangle$ is a CASSCF or CASCI state.

The SS-CASPT2 zeroth-order Hamiltonian is defined as

$$H_0^\alpha = |\alpha\rangle \langle \alpha| \hat{F}^\alpha |\alpha\rangle \langle \alpha| + \sum_k |k\rangle \langle k| \hat{F}^\alpha |k\rangle \langle k| + Q_{sd}^\alpha \hat{F}^\alpha Q_{sd}^\alpha + Q_{tq}^\alpha \dots \hat{F}^\alpha Q_{tq}^\alpha \dots \quad (3.24)$$

In Equation (3.24) \hat{F}^α is a generalized one-body Hartree-Fock operator, the sum over k includes all states within the CAS space that are orthogonal to the reference state $|\alpha\rangle$, Q_{sd}^α refers to a subspace spanned by states $|pqrs; \alpha\rangle$ corresponding to double excitations. These states are non-orthogonal and linearly dependent. Q_{tq}^α includes higher order excitations. Using special identities, Equation (3.23) can be recasted into

$$(E_0^\alpha - \hat{F}^\alpha) \Omega_1^\alpha |\alpha\rangle = Q_{sd}^\alpha H |\alpha\rangle \quad (3.25)$$

This equation is solved iteratively for the first order Ω_1^α .

MS-CASPT2

In MS-CASPT2, the reference space P is multi-dimensional and is spanned by several SA-CASSCF states (compare to Equation (3.20)):

$$P = \sum_{\alpha=1}^d |\alpha\rangle \langle \alpha| \quad (3.26)$$

where d denotes the dimension of the reference space. Similar to SS-CASPT2, the reference states $|\alpha\rangle$ are eigenfunctions of H_0^α , which is expressed as

$$H_0^\alpha |\alpha\rangle = E_0^\alpha |\alpha\rangle \quad (\alpha = 1, 2, \dots, d) \quad (3.27)$$

The multireference operator Ω^P operates on the model states $|\Psi_P^0\rangle$ to generate the states of interest $|\Psi_P\rangle$, while the model states are projections on the exact states. Ω^P is expanded in orders:

$$\Omega^P = 1 + \Omega_1^P + \Omega_2^P + \dots \quad (3.28)$$

Each Ω_n^P has n perturbative factors V . Each Hamiltonian is separately partitioned according to the reference space state $|\alpha\rangle$:

$$H = H_0^\alpha + V^\alpha \quad (\alpha = 1, 2, \dots, d) \quad (3.29)$$

The zeroth-order Hamiltonian is similar to SS-CASPT2:

$$H_0^\alpha = \sum_\beta |\beta\rangle \langle \beta| \hat{F}^\alpha |\beta\rangle \langle \beta| + \sum_k |k\rangle \langle k| \hat{F}^\alpha |k\rangle \langle k| + Q_{sd}^\alpha \hat{F}^\alpha Q_{sd}^\alpha + Q_{tq}^\alpha \dots \hat{F}^\alpha Q_{tq}^\alpha \dots + \dots \quad (\alpha = 1, 2, \dots, d) \quad (3.30)$$

where the first sum over β includes all reference space states. The equation for the first-order multireference wave operator is equivalent to Equation (3.25) in the SS-CASPT2 formalism:

$$(E_0^\alpha - \hat{F}^\alpha)\Omega_1^P|\alpha\rangle = Q_{sd}^\alpha H|\alpha\rangle \quad (\alpha = 1, 2, \dots, d) \quad (3.31)$$

Comparing Equations (3.25) and (3.31), it can be concluded that the multireference wave operator is a linear combination of SS wave operators:

$$\Omega_1^P = \sum_{\alpha} \Omega_1^\alpha |\alpha\rangle \langle \alpha| \quad (3.32)$$

SA-CASSCF states interact via an effective Hamiltonian H^{eff} , which is by definition expressed as

$$H^{eff} = PH\Omega^P P \quad (3.33)$$

The earlier mentioned model states are eigenfunctions of H^{eff} :

$$H^{eff}|\Psi_P^0\rangle = E_P|\Psi_P^0\rangle \quad (p = 1, 2, \dots, d) \quad (3.34)$$

and E_P is the exact energy of $|\Psi_P\rangle$. The second order effective Hamiltonian is given by

$$H_{2nd}^{eff} = PHP + PH\Omega_1^P P \quad (3.35)$$

The diagonal energies of H_{2nd}^{eff} correspond to SS-CASPT2 energies, while the off diagonal elements refer to the coupling up to second order of dynamical correlation. The MS-CASPT2 approach is described as not much more time consuming than the SS approach.

3.2.5 DFT and DFT/MRCI

The central idea of density functional theory (DFT) is to derive molecular properties (energy, geometries etc.) from the electron density instead of using an electronic wave function. The improvement of DFT compared to a wave function approach refers to the reduced dimensionality of the first. While the electronic wave function depends on $4N$ (N corresponds to the number of electrons) variables, the electron density is a function of only three coordinates regardless of the number of electrons. The validity of the DFT concept results from the correspondance of the ground state energy of a system and its electron density, which was proven by Hohenberg and Kohn. [105] The connection between electron density and ground state energy is given by a density functional. Obtaining a proper functional is a main challenge in the DFT approach. Similar to the wave function concept, the energy functional comprises three terms: a kinetic energy functional $T[\rho]$, the attractive interaction between electrons and nuclei $E_{ne}[\rho]$ and electron-electron repulsion $E_{ee}[\rho]$. The latter can be subdivided into Coulomb part $J[\rho]$ and exchange $K[\rho]$ part analogous to HF theory. While $E_{ne}[\rho]$ and $J[\rho]$ can be

derived classically, the estimation of the kinetic energy and exchange interaction $K[\rho]$ is less straight forward for a charge continuum that represents the electron density. At the early stages of DFT, the electrons were described as a uniform electron gas. This was a basis for Thomas-Fermi (TF) and Thomas-Fermi-Dirac (TFD) theories. However, this model was only valid for certain metallic systems and not applicable to molecules due to the lack of a proper bonding description. Despite subsequently applied corrections, results from TF and TFD were not good enough to compete with wave function methods. [95] A true improvement of DFT was achieved through re-introduction of orbitals by Kohn and Sham (KS), which changed the DFT orbital-free paradigm. [106] A basic concept of the KS formalism is to separate the kinetic energy functional into two parts: an exact expression and a correction term. This idea is realized by the introduction of a fictitious system of N non-interacting electrons in N KS-orbitals ϕ^{KS} , where the electron density is the same as the electron density of a real system with interacting particles. For a non-interacting system, an exact solution of the Schrödinger equation is given by a Slater-determinant. Consequently, an exact kinetic energy functional $T_{SD}[\rho]$ is expressed as

$$T_{SD}[\rho] = \sum_i^{N_{elec.}} \left\langle \phi_i^{KS} \left| -\frac{1}{2} \nabla^2 \right| \phi_i^{KS} \right\rangle \quad (3.36)$$

However, in a real system electrons interact and Equation (3.36) is only an approximation to the real kinetic energy. The deviation from the real energy is included in a special exchange-correlation term. The total DFT-energy can be expressed as follows:

$$E_{DFT}[\rho] = T_{SD}[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \quad (3.37)$$

The main goal and main problem of the DFT approach is to devise a good correlation-exchange functional $E_{xc}[\rho]$. If the exact functional would be known, DFT would deliver the exact energy in contrast to HF. This is, however, hardly possible and functionals are approximated using parameters based on a set of certain criteria or fitted to experimental values. [95] In the current study, B3-LYP [107–111] (Becke’s three-parameter functional) and BH-LYP [107–110, 112] (Becke’s half-and-half exchange) functionals both in a combination with the LYP correlation functional and denoted as hybrids, were used. Hybrid functionals include a certain amount of HF-exchange energy mixed with exchange-correlation energy derived from other functionals and also include empirical parameters.

DFT provides calculations at a speed of Hartree-Fock and accounts for electron correlation, in particular for dynamical effects. It is a good choice for calculating ground state properties, however, it is not well suitable for describing excited states, where static correlation is especially important. For this purpose, a well-established method is DFT augmented by a multi-reference configuration interaction expansion (MRCI).

MRCI goes beyond the CI expansion by using excited-state determinants as reference wave function in addition to the ground state determinant. [95] In principle, with this strategy both dynamic and static correlation effects can be taken into account. This, however, would require lengthy CI expansions. The **DFT/MRCI** approach developed by Grimme and Waletzke exploits the strengths of both methods: dynamic correlation from DFT and static correlation from short MRCI expansions. [113] Double counting of dynamical correlation from MRCI is avoided by a special damping function. The MRCI expansion is performed based on KS-orbitals and the BH-LYP functional. The effective DFT/MRCI Hamiltonian is an empirical operator that includes KS-orbital energies, two-electron integrals and empirically fitted scaling factors. Recently, a Hamiltonian improved for treatment of bi-chromophores was designed by Lyskov and co-authors. [114] Due to the lack of analytic gradients, the DFT/MRCI approach is only applicable to the calculation of energy corrections.

3.2.6 Semiempirical methods

In calculations where a quantum mechanical strategy is required, but *ab initio* or DFT methods are too time-consuming, semiempirical approaches may be considered due to their fast computational performance (about 100-1000 faster than *ab initio* or DFT). [96] In the current section, the basic principles of semiempirical approaches will be described with the focus on orthogonalization-corrected methods (OMx) applied within the scope of the current work. The concept of semiempirical methods is based on the *ab initio* molecular orbital approach. It includes approximations affecting the treatment of electrons, basis set, evaluation of integrals and treatment of the overlap matrix. The most significant speedup comes from integral approximations, which also define a semiempirical method. [115] Many of the integral terms vanish, because products of functions on different atoms are set to zero. Evaluation of the remaining integrals is performed by parametrization either against experimental or high-accuracy theoretical data.

For a better understanding of the semiempirical methods, the SCF-MO approach will be briefly reviewed in the following. [116, 117] Molecular orbitals are represented by a linear combination of atomic orbitals ϕ_μ weighted by expansion coefficients c_i :

$$\psi_i = \sum_{\mu} c_{\mu i} \phi_{\mu} \quad (3.38)$$

Expansion coefficients and energies can be obtained by the solving the secular equation (expressed here in a matrix representation)

$$FC = SCE \quad (3.39)$$

where F is the Fock matrix, C is a matrix with expansion coefficients, S is an overlap matrix and E is a diagonal matrix with orbital energies. Each Fock matrix element includes the one-electron core Hamiltonian $H_{\mu\nu}$, density matrix elements $P_{\lambda\sigma}$ and two-electron repulsion integrals $\langle\mu\nu|\lambda\sigma\rangle$, $\langle\mu\lambda|\nu\sigma\rangle$:

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda} \sum_{\sigma} P_{\lambda\sigma} [\langle\mu\nu|\lambda\sigma\rangle - \frac{1}{2} \langle\mu\lambda|\nu\sigma\rangle] \quad (3.40)$$

The total energy E_{tot} is a sum of the electronic energy E_{el} and the nuclear repulsion V_{nn} :

$$E_{tot} = E_{el} + V_{nn} \quad (3.41)$$

where E_{el} is represented as

$$E_{el} = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu}) \quad (3.42)$$

As stated above, integral approximations are representative for each semiempirical method. In most semiempirical methods, the so called zero differential overlap approximation (ZDO) [118] is used which neglects the differential overlap of all different orbitals μ and ν sharing the same electron coordinate:

$$dS = \phi_{\mu}(1)\phi_{\nu}(1)d\tau_1 = 0 \quad (3.43)$$

Consequently, the overlap matrix is a unit matrix and the Roothaan-Hall equations can be simplified to

$$FC = CE \quad (3.44)$$

which makes orthogonalization unnecessary. A more elegant way to express the ZDO-approximation in integral terms is by using Kronecker deltas:

$$\langle\mu\nu|\lambda\sigma\rangle = \delta_{\mu\nu}\delta_{\lambda\sigma} \langle\mu\mu|\lambda\lambda\rangle \quad (3.45)$$

Thus, all three- and four-center two-electron integrals are neglected, which means that computation of these integrals scales with N^2 and not anymore with N^4 , substantially speeding up the calculations. For the Fock-matrix calculation, only one- and two-center one-electron and two-electron integrals remain. Several other approximations originate from the ZDO-approximation, e.g. complete neglect of differential overlap (CNDO), intermediate neglect of differential overlap (INDO) or neglect of diatomic differential overlap (NDDO). [119] The latter is considered the golden standard. [96] It applies ZDO only for atomic orbitals on different atoms and provides a more accurate description of long-range electrostatic interactions. [120] The NDDO approximation is employed for well established semiempirical methods based on the MNDO [121] model (modified

neglect of diatomic differential overlap) (e.g. AM1 [122], PM3 [123]). Orthogonalization corrected methods (OMx; x=1, 2, 3) belong to a line of more recent semiempirical methods which go beyond the MNDO model.

In MNDO as well as in OMx methods, all one-center integrals are evaluated in the same way. One-center one-electron terms undergo parametrization, while one-center two-electron integrals are obtained from experimental data. The difference between MNDO-type and OMx methods lies in the treatment of two center terms included in the core-Hamiltonian and the Fock-matrix. The lack of orthogonalization in MNDO-type methods leads to some problems of conceptual and intrinsic character. A conceptual issue is the lack of the Pauli repulsion between electrons and core. It leads to incorrect description of conformational properties. The effect of Pauli repulsion can be partly compensated through parameterization. An intrinsic problem refers to the splitting between the bonding and anti-bonding orbitals, which is symmetric for homonuclear diatomics in the MNDO framework and can not be corrected within the MNDO paradigm. [124–126]. Orthogonalization corrections parametrically introduced into the core Hamiltonian make it possible to model Pauli exchange repulsion within the OMx methods. Depending on the level of orthogonalization correction, OMx methods can be subdivided into OM1 [126, 127], OM2 [128, 129] and OM3 [130]. The OMx one-electron core Hamiltonian can be expressed as

$$H_{\mu\nu} = U_{\mu\mu}\delta_{\mu\nu} + \sum_B [V_{\mu\nu,B}^s + V_{\mu\nu,B}^{ORT} + V_{\mu\nu,B}^{PI} + V_{\mu\nu,B}^{ECP}] \quad (3.46)$$

$$H_{\mu\lambda} = \beta_{\mu\lambda} + \sum_C V_{\mu\lambda,C}^{ORT} \quad (3.47)$$

In the above equations μ and ν refer to orbitals at the same atom A, while λ corresponds to a different orbital at atom B and C is the third atom. $U_{\mu\mu}$ represents one-electron energies as adjustable parameters. $V_{\mu\nu,B}^s$ is the core-electron attraction. V^{ORT} includes orthogonalization corrections, V^{PI} designates penetration integrals and reflects differences between core-electron attractions and electron-electron repulsions. V^{ECP} accounts for the influence of the inner core electrons. $\beta_{\mu\lambda}$ is a resonance integral that in OMx methods follows an empirical formula. Expressions for two-center ($V_{\mu\nu,B}^{ORT}$) and three-center ($V_{\mu\lambda,C}^{ORT}$) orthogonalization corrections are given by:

$$\begin{aligned} V_{\mu\nu,B}^{ORT} = & -\frac{1}{2}F_1^A \sum_{\rho \in B} (S_{\mu\rho}\beta_{\rho\nu} + \beta_{\mu\rho}S_{\rho\nu}) \\ & + \frac{1}{8}F_2^A \sum_{\rho \in B} S_{\mu\rho}S_{\rho\nu}(H_{\mu\mu,B}^{loc} + H_{\nu\nu,B}^{loc} - 2H_{\rho\rho,A}^{loc}) \end{aligned} \quad (3.48)$$

$$\begin{aligned}
V_{\mu\lambda,C}^{ORT} = & -\frac{1}{2}G_1^{AB} \sum_{\rho \in C} (S_{\mu\rho}\beta_{\rho\lambda} + \beta_{\mu\rho}S_{\rho\lambda}) \\
& + \frac{1}{8}G_2^{AB} \sum_{\rho \in C} S_{\mu\rho}S_{\rho\lambda} (H_{\mu\mu,C}^{loc} + H_{\lambda\lambda,C}^{loc} - H_{\rho\rho,A}^{loc} - H_{\rho\rho,B}^{loc})
\end{aligned} \tag{3.49}$$

where F_1^A , F_2^A , G_1^{AB} and G_2^{AB} are adjustable parameters that define the magnitude of the orthogonalization corrections. $H_{\mu\mu,X}^{loc}$ is a matrix-element of a local Hamiltonian for an atom pair A and X. OMx methods (OM1, OM2, OM3) vary by parameter sets and, most important, in the extent of orthogonalization. In OM2, all expressions above are included. OM1 includes only corrections in the one-center part of the Hamiltonian $V_{\mu\mu,B}^{ORT}$. OM3 is truncated at first order terms ($F_2=0$ and $G_2=0$). Combined with a multi-reference configuration interaction approach (MRCI) OMx methods are applicable for the calculation of excited states in addition to ground state computations. [131] The MRCI for OMx methods within the current work uses the graphical unitary group approach. [132–134]

3.3 Resonance Raman spectra

Raman spectroscopy is a technique based on molecular vibrations, applied to study structural properties of molecules or to identify them. During the experiment, a molecule is excited by a laser pulse. The light scattered by the molecule is captured and results in a Raman spectrum. This provides finger print vibrations, helping to identify a molecule. The light scattering process can be subdivided into three kinds of scattering: Rayleigh scattering (the energy of the scattered light does not change), Stokes Raman scattering (the energy of the scattered light is lower than the initial energy) and anti-Stokes Raman scattering (the energy of the scattered light is larger than the initial energy). [135] In resonance Raman (RR) spectroscopy, the energy of the light source is close to the electronic transition. The resulting frequency resonance enhances the usually weak Raman signal, making this method more sensitive to lower concentrations.

Computation of RR spectra provides valuable additional information for the experimentalists, since each Raman frequency can be examined and assigned to particular vibrations. For the calculation of RR spectra within the current work, I applied an approach developed by Orlandi *et al.* Further description is based on the following references. [136–138] This model assumes that the major contribution to the RR intensity of the vibrational bands originates from the allowed Franck-Condon excitation, while the impact of the vibronic coupling is minor. In this case, the excited modes are totally symmetric. The FC factors are non-vanishing Franck-Condon integrals arising from the difference between vibrational frequencies of the initial and resonant states and displacement between the corresponding equilibrium geometries. [139] However, the impact of the displacement on the FC-activity is known to be much larger than the

influence of the frequency difference. Thus, for simplicity, vibrational frequencies of the ground and excited states are assumed to be the same and only the displacement is considered. [136] The contribution of the displacement to the FC intensity is described by the shift parameter γ_i :

$$\gamma_i = 0.5B_i^2 \quad (3.50)$$

where B_i denotes the displacement between the equilibrium geometries and is expressed as:

$$B_i = 0.172\sqrt{\nu_i} \cdot (x_1 - x_0)\sqrt{M} \cdot L_i \quad (3.51)$$

where ν_i and L_i are i^{th} vibrational frequency and normal mode of the initial state (e.g. S_0). M is the matrix of atomic masses and x_0 and x_1 correspond to coordinates defining minimum geometries of the involved states. Another approach to calculate the displacement parameter is by using nuclear forces of the resonant electronic state at the equilibrium ground state geometry instead of the excited state minimum. In this case, the B_i factor is calculated as follows:

$$B_i = (2.41 \times 10^6) \cdot F \cdot \sqrt{M} \cdot L_i \cdot \nu_i^{-3/2} \quad (3.52)$$

In this equation, F is the Cartesian force vector of the resonant state in atomic units, M is the matrix of atomic masses (in amu), ν_i is the vibrational frequency and L_i are the vibrational normal mode coordinates of the studied state. For the current work, the latter option was applied. This rather simple approach has shown to yield reasonable results for modeling of RR spectra in retinal models. [46]

3.4 Molecular mechanical approach

A molecular mechanical approach is a common choice for the description of large systems consisting of hundreds or thousands of atoms. Its typical field of application is biomolecular modeling, but also organic and inorganic materials can be investigated. [140] Different from the QM paradigm, MM-particles are not explicitly subdivided into electrons and nuclei. Atoms are rather represented as soft spheres characterized by van der Waals (vdW) radii, polarizabilities, masses and net charges. The spheres are connected via harmonic potentials. These can be seen as elastic springs approximating molecular bonds. Electronic properties are contained implicitly in special parameters that will be discussed later in the context of the force field (FF) concept. The sphere-spring system obeys the laws of classical mechanics, making MM calculations faster compared to a quantum mechanical approach. However, a classical description is only applicable to the ground state and a description of chemical reactions is not possible without special assumptions (e.g. reactive force fields [141]). The objective of MM calculations is the estimation of the potential energy of a given system. A total energy expression is called a force field if the expression is continuously differentiable. For organic molecules it

is commonly represented by a sum of bonded and non-bonded terms (potentials) with force parameters characteristic for each atom type. The latter can be obtained from quantum mechanical calculations or experimental data (e.g. NMR, X-Ray, Raman or IR spectroscopy etc.). The functional form of a force field may vary depending on the desired purpose. The particular energy terms will be described below in more detail using the AMBER FF applied in the current work.

The Amber force field

The Amber force field family is designed for modeling of amino acids, DNA and organic molecules. [142] A basic Amber FF energy equation has the following functional form:

$$\begin{aligned}
 E_{MM} = & \sum_{bonds} K_l(l_i - l_{i,0})^2 + \sum_{angles} K_\theta(\theta_i - \theta_{i,0})^2 \\
 & + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\phi - \delta)) \\
 & + \sum_{non-bonded} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)
 \end{aligned} \tag{3.53}$$

The first two terms refer to bond stretching and angle bending. They are represented by harmonic potentials and control the change of the bond lengths l_i and angles θ_i relative to the reference equilibrium values $l_{i,0}$ and $\theta_{i,0}$, respectively. K_l and K_θ are the force constants. The torsion potential is described by a set of parameters included in a Fourier series, where V_n is the height of the torsion barrier, n is the dihedral periodicity and δ is the phase of the dihedral angle ϕ . The last term describes the non-bonded energy which involves vdW interactions expressed as Lennard-Jones potentials and electrostatic interactions modeled by Coulomb potentials. Force constants and reference values are parameters specific for each FF.

Within the current study, I used the Amber ff99SB-ILDN [143] force field with corrected side chain torsion parameters for the four amino acids Isoleucin (I), Leucin (L), Aspartate (D) und Asparagin (N). It refines the Amber ff99SB force field [144] which has already been improved for back bone torsions. Its energy function consists of the Amber ff99SB total energy E_{A99SB} with the torsion potential $V_{A99SB}(\theta)$ replaced by a new term:

$$E_{MM} = E_{A99SB} - V_{A99SB}(\theta) + k_0 + \sum_{m=1}^M k_m(1 + \cos(m\theta - \theta_0)) \tag{3.54}$$

In this representation, E_{A99SB} has the functional form of the basic Amber energy equation. Within the dihedral term, k_0 is a constant, k_m are fit parameters for the cosine expansion and θ_0 is set to zero.

3.5 QM/MM methods

A quantum mechanical approach is the method of choice to follow chemical transformations and is inevitable for the description of photochemical reactions like in channel-rhodopsins. However, the computational cost of QM methods restricts their application to small and medium sized systems. The investigation of an enzyme with several hundreds of atoms and a catalytic center solely by QM techniques would not be feasible within reasonable time. An elegant solution, which became a common strategy for the description of large scale systems, is a hybrid of QM and MM approaches, first proposed by Warshel and Levitt. [145] The basic idea of the QM/MM technique is to subdivide the entire system (S) into an inner region (I) treated quantum mechanically and an outer region (O) which is described by a force field (Fig. 3.2). [146] The main chal-

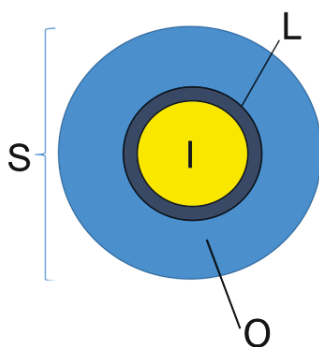


FIGURE 3.2: Partitioning of the entire molecular system (S) into outer region (O), inner region (I) and link atom (L).

lenge arising from such a hybrid approach is the evaluation of the interactions between differently treated parts. In the current section, the general QM/MM concept will be introduced. Approaches implemented in COBRAMM will be described in detail, while only a brief introduction to other possible strategies will be given.

3.5.1 The QM/MM energy

Calculation of the total QM/MM energy can follow subtractive or additive schemes. In a subtractive scheme, three energy terms are calculated: the MM energy of the entire system and QM and MM energies of the QM-region. To obtain the total QM/MM energy, the MM energy of the inner region has to be replaced by its QM energy:

$$E_{QM/MM}^{sub}(S) = E_{MM}(S) - E_{MM}(I + L) + E_{QM}(I + L) \quad (3.55)$$

The equation above refers to a calculation where a link atom (L) approach is applied to saturate dangling valencies of the QM region originating from a cut through a covalent bond across the QM/MM boundary. This approach will be discussed later. No explicit terms to describe the interaction of the QM and MM regions are required. In principle,

all QM/MM interactions can be handled at the MM level, but we will later see an improvement of this strategy. In COBRAMM an extended subtractive scheme similar to ONIOM (Our N-layered Integrated molecular Orbital and molecular Mechanics) [147, 148] scheme is implemented which includes subpartitioning of the MM-region. [79, 149] Within an additive scheme [146], the total energy is a sum of three terms: the MM energy of the outer region (O), the QM energy of the inner region (I) and a QM/MM interaction term $E_{QM/MM}(I, O)$:

$$E_{QM/MM}^{add}(S) = E_{MM}(O) + E_{QM}(I + L) + E_{QM/MM}(I + O) \quad (3.56)$$

The explicit coupling term $E_{QM/MM}(I, O)$ describes the interaction between the QM and MM regions (see next section) and includes bonded, vdW and electrostatic interactions of the QM and MM atoms. Irrespective of an additive or subtractive scheme, bonded and vdW interactions are calculated classically. VdW interactions are typically described by Lennard-Jones potentials. Bonded interactions are usually described by MM parameters for bonded interactions (bond stretchings, angle bendings and dihedral torsions).

QM/MM electrostatic interactions

Two approaches can be applied to handle electrostatic interactions: mechanical or electrostatic embedding. The concept of mechanical embedding is to treat QM/MM electrostatics solely by the MM approach. [146] Consequently, the QM density is not influenced by the MM electrostatic environment. Such a strategy is straight-forward to implement, but may be insufficient for the description of chemical reactions or excitation processes, where the electron distribution may change significantly. Also the change of the charge distribution within the MM environment (e.g. during optimization or dynamics) is not "perceived" by the wave function. This shortcoming is improved in an electrostatic embedding scheme, where MM point charges are included in the QM Hamiltonian as one-electron terms. [150] In this case, the electronic wave function interacts with its electrostatic environment. However, the MM surrounding is not polarized by the wave function. Furthermore, at the QM/MM border overpolarization effects may occur. The first problem can be tackled by application of polarizable force fields with mutual polarization, where QM and MM regions influence each other. Overpolarization may be overcome through redistribution of the MM-charges and will be discussed in the next subsection. As already mentioned before, in the subtractive scheme all QM/MM interactions can be described by a force field. However, modern subtractive schemes, including the one in COBRAMM [79, 149], explicitly incorporate electrostatic terms in QM-calculations and can profit from the accuracy of electrostatic embedding. In dynamics calculations, gradients of moving MM point charges are also considered to ensure energy conservation.

3.5.2 Handling of the QM/MM boundary

The easiest case for the QM/MM frontier, where no further assumptions have to be made, is when the boundary does not cross any covalent bonds (e.g. simulation of a solute in a solvent). However, especially in biological systems like proteins with a reactive site, the QM/MM boundary passes covalent bonds at least once. Thus, the QM region remains with an unsaturated valence, which has to be capped to avoid non-physical results during QM calculations. Some of the established strategies to achieve a well-balanced description of the QM/MM boundary are: link-atom, boundary atom or localized-orbital schemes. [146] The focus will be on the link-atom approach since it is the most commonly used strategy and is also implemented in COBRAMM. [79, 149] Within this method, the dangling QM valence is saturated with a QM atom, that is placed between the boundary of QM and MM atoms (Fig. 3.3). In COBRAMM,

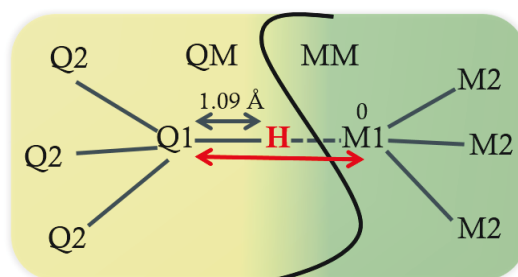


FIGURE 3.3: Description of the QM/MM boundary. The hydrogen link-atom is shown in red color. Q1 and M1 are atoms at the QM/MM border. Q2 and M2 denote neighboring atoms.

a hydrogen link-atom approach is implemented. However, the choice of the link-atom is not restricted and in principle any atom type can be considered which can saturate the free valence and has the least influence on the QM-system. Artificial insertion of atom(s) into the system can lead to some undesirable effects. The additional atom introduces extra degrees of freedom, which do not belong to the system and can cause artifacts. In COBRAMM, all additional degrees of freedom are removed by restricting the motion of the link atom along the bond axis (Fig. 3.3, grey and red arrows). Furthermore, placement of a hydrogen atom among the QM and MM atoms leads to short Q1-L and L-M1 distances. Within an electrostatic embedding scheme, overpolarization of the QM region can arise, a problem already mentioned in the previous section. To diminish the artificial interaction, the charge of the border MM atom (M1) is set to zero and its value is redistributed between the neighboring atoms (M2). Taking into account all mentioned problems that can occur, the decision where to put the boundary may be taken with caution. To avoid artifacts, the cut should be introduced as far as possible from the active region, it should not involve multiple bonds or bonds between hetero-atoms.

3.6 Nonadiabatic dynamics

The focus of the current study lies on ultrafast and radiationless photochemical reactions. These correspond to a dynamical process that usually involves passing through a conical intersection. In this case, the coupling between nuclei and electrons becomes non-negligible and the BO-approximation breaks down (Sec. 3.2.1). Such processes are called nonadiabatic and are ubiquitous in chemistry. They are also known from collision reactions and heterolytic dissociations. [151, 152]

A theoretical description of nonadiabatic dynamics is nontrivial since it requires the solution of the time-dependent molecular Schrödinger equation while considering at least two electronic states simultaneously. Generally, nonadiabatic dynamics can be addressed by fully quantum mechanical or quantum-classical approaches, while in all approaches the fast variable (electrons) is always treated quantum mechanically. The most prominent example of the full quantum mechanical approach is the multi-configuration time dependent Hartree (MC-TDH) method, where the nuclei are represented by wave packets. [153, 154] Due to its computational demand, this method is restricted to few degrees of freedom (typically 4-12). For larger systems, techniques that treat the nuclei by classical mechanics are best suited. In this regard, the most widely used methods are Ehrenfest and trajectory surface hopping (TSH) dynamics. [155, 156] The Ehrenfest method is referred to as a meanfield method, where an averaged potential energy surface drives a single trajectory. This representation is adequate for the coupling region but becomes unrealistic after leaving it, since the single trajectory will still move on the averaged PES. The TSH approach approximates the wave packet by computing a swarm of quantum-classical trajectories, which are averaged at the end of the simulation to obtain final results. [151, 157–159] The trajectories are allowed to change the surface corresponding to a determined probability distribution. Within the scope of the current project, the TSH approach in the formulation proposed by Tully and Preston [157] was applied and thus will be described in detail.

3.6.1 Initial conditions

Excitation of a molecule does not occur from one particular geometry in the ground state, but rather from one of the structures existing in a ground-state equilibrium. There are different possibilities to approximate ground state configuration distributions. An obvious strategy would be ground-state equilibration of a system within a QM/MM or pure MM framework with subsequent snapshot selection. These snapshots contain molecular geometries and corresponding velocities that can be used as starting points for the simulation of a semi-classical excited-state ensemble. [160] The major problem of this approach originates from usage of classical molecular velocities for the quantum mechanical description performed during dynamics. Within a classical system, the energy of one mode is described by $E=kT$ [64, 160, 161], which means that its energy

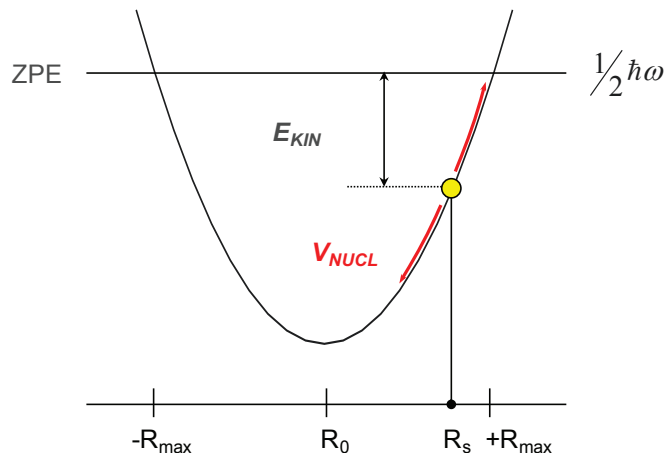


FIGURE 3.4: Zero point energy sampling scheme from Ref. [64].

is equally distributed over all degrees of freedom. However, within a QM paradigm, the energy of each vibrational mode depends on its frequency. This discrepancy leads to over-representation of the low frequency modes and under-representation of the high frequency modes in a classical ensemble, which may influence the dynamics in an unpredictable way. Furthermore, geometries originating from one MD trajectory are not independent, since the simulation time is limited and only a small part of the reaction phase space can be explored. [160] A well established method to obtain less dependent initial geometries and velocities is to sample normal modes, which originate from a vibrational frequency analysis. This approach is referred to as zero-point energy (ZPE) sampling. [162] Each normal mode is approximated by a harmonic function, which refers to its deviation from the equilibrium structure. In this representation, the zero-point energy corresponds to maximum elongation and thus maximum possible potential energy of the normal mode at zero Kelvin. During the sampling process, a list of random numbers is generated. Each random number defines a position (R_s) between the starting point (R_0) and $\pm R_{max}$ (Fig. 3.4). This position defines a distortion relative to the equilibrium geometry characterized by its potential energy E_{pot} . The difference between this potential energy and the ZPE is equivalent to the kinetic energy and delivers velocities for the geometries at R_s . Performing this procedure for all generated random numbers and normal modes results in a set of geometries and velocities which represent a starting ensemble for the dynamics calculation.

3.6.2 Trajectory surface hopping

In surface hopping, the nuclei are propagated classically using Newton's equation of motion on a single adiabatic surface. [151, 157] At the same time, the electronic degrees

of freedom are propagated using the time-dependent SE

$$i\hbar \frac{\delta \Psi(r, R, t)}{\delta t} = H_e \Psi(r, R, t) \quad (3.57)$$

The time dependent wave function defines the electronic state at time t and is expanded in an electronic basis:

$$\Psi(r, R, t) = \sum_i c_i(t) \phi_i(r; R) \quad (3.58)$$

where c_i are complex-valued expansion coefficients, r designates electronic coordinates and R corresponds to the nuclear coordinates, i.e., a classical trajectory. Insertion of this term into the time dependent Schrödinger equation (3.57), multiplication on the left by $\phi_j^*(r; R)$ and integration over r leads to a set of coupled differential equations for expansion coefficients $c_j(t)$:

$$i\hbar \frac{\delta c_j(t)}{\delta t} + \sum_i (-H_{ji} + i\hbar f_{ji} \cdot v) c_i(t) = 0 \quad (3.59)$$

In Equation (3.59), H_{ji} are matrix elements $\langle \phi_j | H_e | \phi_i \rangle$, f_{ji} is the nonadiabatic coupling vector between the states j and i , and v is the velocity vector. As already introduced in the Section 3.1.2 the nonadiabatic coupling for an atom m can be expressed as

$$f_{ji} = \left\langle \phi_j \left| \frac{\delta \phi_i}{\delta R} \right. \right\rangle \quad (3.60)$$

Propagation of the semi-classical time-dependent Schrödinger equation is usually done in the adiabatic representation, since it has been shown to yield better results. [151, 163] In the adiabatic representation, the off-diagonal matrix elements of the electronic Hamiltonian vanish and Equation (3.59) can be simplified [164]:

$$i\hbar \frac{\delta c_j(t)}{\delta t} = \epsilon_j c_j(t) - i\hbar \sum_i f_{ji} \cdot v c_i(t) \quad (3.61)$$

where ϵ_j is the energy of state j . Integration of this equation delivers state amplitudes $c_j(t)$ for each quantum state. This quantity is used to estimate hopping probabilities. COBRAMM uses Tully's fewest switches (TFS) algorithm [165] for this purpose. The TFS algorithm will be explained in the next section. New positions for a nucleus m with mass M are obtained by propagation of Newton's equations in a single electronic state:

$$\frac{\delta^2 R_m}{\delta t^2} - \frac{F_m}{M_m} = 0. \quad (3.62)$$

In this equation, F_m is the force obtained from a quantum mechanical calculation as the negative gradient of the potential energy. In COBRAMM, integration of Newton's equations is performed with a velocity Verlet scheme that will be described in the corresponding section (Sec. 3.6.6).

3.6.3 Transition probabilities

The main difference between various TSH implementations is the selection of the hopping criterion. It is also a central point of this approach. The main requirement to be fulfilled is the equality of classical occupation and quantum population during dynamics simulation (quantum-classical self-consistency). Numerous algorithms exist which satisfy this condition. One of the most popular is the fewest switches algorithm which was suggested by Tully. [165] Its idea is to avoid an averaged PES description and to ensure quantum-classical self-consistency. The latter means consistence between electronic properties of a single molecule and the entire ensemble. A general description of the algorithm for a large number of trajectories N can be described as follows:

$$N_i(t) - N_i(t + \delta t) \quad (3.63)$$

where $N_i(t)$ and $N_i(t + \delta t)$ are fractions of trajectories (ensemble occupation) residing in the state i at time steps t and $t + \delta t$ respectively. At time t , N_i is expressed as

$$N_i(t) = \rho_{ii}(t)N, \quad (3.64)$$

where ρ_{ii} are diagonal terms of the density matrix ρ_{ij} and assign electronic state populations. In terms of state coefficients, $\rho_{ii}(t)$ is given by:

$$\rho_{ii}(t) = c_i^*(t) \cdot c_i(t) \quad (3.65)$$

At time $t + \delta t$, the ensemble occupation will change to:

$$N_i(t + \delta t) = \rho_{ii}(t + \delta t)N \quad (3.66)$$

In the case $N_i(t) > N_i(t + \delta t)$, the minimum number of hops from state i corresponds to Equation (3.63), while the number of hops to state i equals zero. In the infinitesimal time interval δt the hopping probability from state i is given by

$$P_i(t)\delta t \approx \frac{-\dot{\rho}_{ii}\delta t}{\rho_{ii}} \quad (3.67)$$

where $\dot{\rho}_{ii}$ corresponds to:

$$\dot{\rho}_{ii} = 2\text{Re}(c_i^*(t)\dot{c}_i(t)) \quad (3.68)$$

Finally, using Equation (3.61), the probability for a hop from state i to state j within one MD time step Δt can be calculated as

$$P_{ij} = 2 \frac{\int_t^{t+\Delta t} \text{Re}(c_i^*(t)c_j(t)v \cdot f_{ij})\delta t}{c_i^*(t)c_i(t)} \quad (3.69)$$

The coupling vector f_{ij} is interpolated between the MD-steps with a step size usually set to $\Delta t/100$. The sign of the coupling vector is arbitrary and depends on the sign of molecular orbitals in the consecutive steps. When the signs of the current and subsequent steps differ, artificially large changes may arise from numerical calculation of the gradient. To avoid this problem, the sign of the coupling vector is monitored and adjusted to be consistent throughout the calculation. Consequently, the hopping probability is compared to a random number r_t from a range between 0 and 1. The decision to hop from surface i to surface j is made when the following condition is fulfilled:

$$\sum_{n=1}^j P_{i \rightarrow n} < r_t \leq \sum_{n=1}^{j+1} P_{i \rightarrow n} \quad (3.70)$$

In other words, if the hopping probability drops below the generated random number, the hop occurs and the molecule propagates on the new PES. Otherwise it stays on the current surface.

3.6.4 Kinetic energy redistribution

Hopping in a TSH scheme is a stochastic event, which means that the trajectories may hop at a significant energy difference. Without correction, this event would lead to loss of energy conservation during the computation within a microcanonical NVE ensemble. In an NVE ensemble, however, the total energy, which is a sum of kinetic and potential energies, must be constant. The loss of total energy can be avoided through energy redistribution along a particular coordinate. Several scaling coordinates are possible, they include the nonadiabatic coupling vector, the gradient difference vector or the velocity vector. [151, 156, 160, 165, 166] In COBRAMM, the energy is redistributed along the nonadiabatic coupling vector. However, prior to the scaling process the correct scaling direction has to be estimated. The reason is the arbitrary sign of the coupling vector. [160] If the scaling vector contains a reaction determining coordinate (e.g. HOOP modes in retinal [62, 63, 160]), scaling in a random direction might artificially influence the reaction outcome. [160] Thus, the scaling direction is estimated from the comparison of the coupling and velocity vectors by computing the angle between them. [79, 167] This approach ensures that the trajectory evolves into the velocity direction and results in a correct product distribution. In the case where velocity and coupling vectors are orthogonal, the hop is either rejected or scaling has to be performed along the velocity vector. Another case where hops are rejected concerns classically forbidden hopping to a state lying above the total energy level.

3.6.5 Decoherence correction

In a quantum mechanical description of dynamics, an initially pure electronic state evolves to a superposition of coherent states. However, in the course of dynamics a quantum decoherence between nuclear and electronic degrees of freedom arises. Thus,

at the hopping point it would be physically correct to dissipate one trajectory into multiple trajectories taking their own path and propagating along different electronic states. [168] This effect is partially taken into account e.g. in the multiple spawning approach. [169] According to the concept of a mixed quantum-classical treatment, quantum decoherence is neglected. A single trajectory either performs a hop or not, irrespective of partial hopping probabilities, leading to a divergence between the classical state occupations and the quantum state populations. To introduce decoherence effects, a nonlinear decay of mixing correction scheme suggested by Truhlar *et al.* [170] and realized by Persico *et al.* [171] is implemented in COBRAMM. At each integration step, amplitudes for the state k are obtained from Equation (3.61) and are used to calculate probabilities for estimating the current MD state m . The following corrections are applied after calculation of the probabilities:

$$\begin{aligned} c'_k &= c_k e^{-\Delta t / \tau_{km}} \quad \forall \quad k \neq m \\ c'_m &= c_m \left[\frac{1 - \sum_{k \neq m} |c'_k|^2}{|c_m|^2} \right]^{1/2} \\ \tau_{km} &= \frac{\hbar}{|E_k - E_m|} \left(1 + \frac{\alpha}{E_{kin}} \right) \end{aligned} \quad (3.71)$$

where E_k and E_m are the energies of the corresponding states, E_{kin} is the nuclear kinetic energy, Δt is an integration step and α is an empirically estimated parameter, with the value of $\alpha=0.1$ Hartree. [151, 171] The time evolution of the electronic wave function is then continued with the corrected coefficients c'_k . [171]

3.6.6 Integration of Newton's equations with the velocity Verlet algorithm

The velocity Verlet (VV) code is a self starting algorithm. [172] Integration begins with known velocity (v), force (F) and geometry (R) vectors at a starting time step t . The new geometry for the time step $t + \Delta t$ is expressed as:

$$R(t + \Delta t) = R(t) + v(t + \Delta t) \Delta t \quad (3.72)$$

and the new velocities for $t + \Delta t$ are given by:

$$v(t + \Delta t) = v(t) + \frac{1}{2M} [F(t) + F(t + \Delta t)] \Delta t \quad (3.73)$$

Since the forces for the new time step $F(t + \Delta t)$ are not known, half-step velocities are calculated:

$$v(t + \Delta t_{1/2}) = v(t) + \frac{1}{2M} F(t) \Delta t \quad (3.74)$$

These are inserted into Equation (3.72) to calculate the new full step geometry by

$$R(t + \Delta t) = R(t) + (v(t) + \frac{1}{2M}F(t)\Delta t)\Delta t \quad (3.75)$$

At the new geometry, new forces and new velocities are computed and the process continues with Equation (3.75). The VV algorithm is stable, but can exhibit fluctuations due to the half-step velocity calculation. Thus, the value for the time step Δt should be chosen small enough to preserve energy conservation. It is typically chosen in a range between 0.25-0.5 fs and depends on the highest frequency vibrations of the molecule.

Chapter 4

Implementations (Paper III)

Several studies provide evidence for the ability of computationally affordable semiempirical OMx-methods to adequately describe static and dynamic properties of organic molecules (see [131] and citations therein), in particular those of protonated Schiff bases [131, 173–178]. In the context of the current work, the implementation of an interface between COBRAMM and the MNDO program package [79] (**Paper III**) enabled fast execution of QM/MM nonadiabatic dynamics of retinal within C1C2. The implementation is documented in the following.

4.1 General description

The core idea of COBRAMM is to provide a user friendly interface that allows to perform pure QM as well as QM/MM-computations. Custom tasks are specified in the **cobram.command** file, which is created by the user. These may involve geometry optimizations, vibrational analysis, single point calculations or molecular dynamics simulations. To accomplish these tasks within a QM/MM environment, COBRAMM requires QM and MM energies and gradients that are obtained through the corresponding QM and MM interfaces. In the first step, input files for the QM and MM programs are produced and the programs are executed. In the next step, the program output files are processed by the corresponding interfaces to extract the required quantities. Finally, the selected tasks can be initialized and performed by the COBRAMM subroutines. A basic COBRAMM flowchart is depicted in Figure 4.1. The parts of the program that needed to be adapted for the COBRAMM-MNDO interface are highlighted in Figure 4.1 by dashed circles. The programmed interface is scripted in Python and is included in the file **MNDO.py**. A basic flowchart of the MNDO interface is provided in Figure 4.2. Core routines in **MNDO.py** are: **mndoinput**, **launch** and **mndoEneGradCrg**. They will be described in the following.

mndoinput The basic task of the **mndoinput** routine is to generate a suitable input file (**MNDO.input**) for the MNDO program. It furthermore initializes arrays for

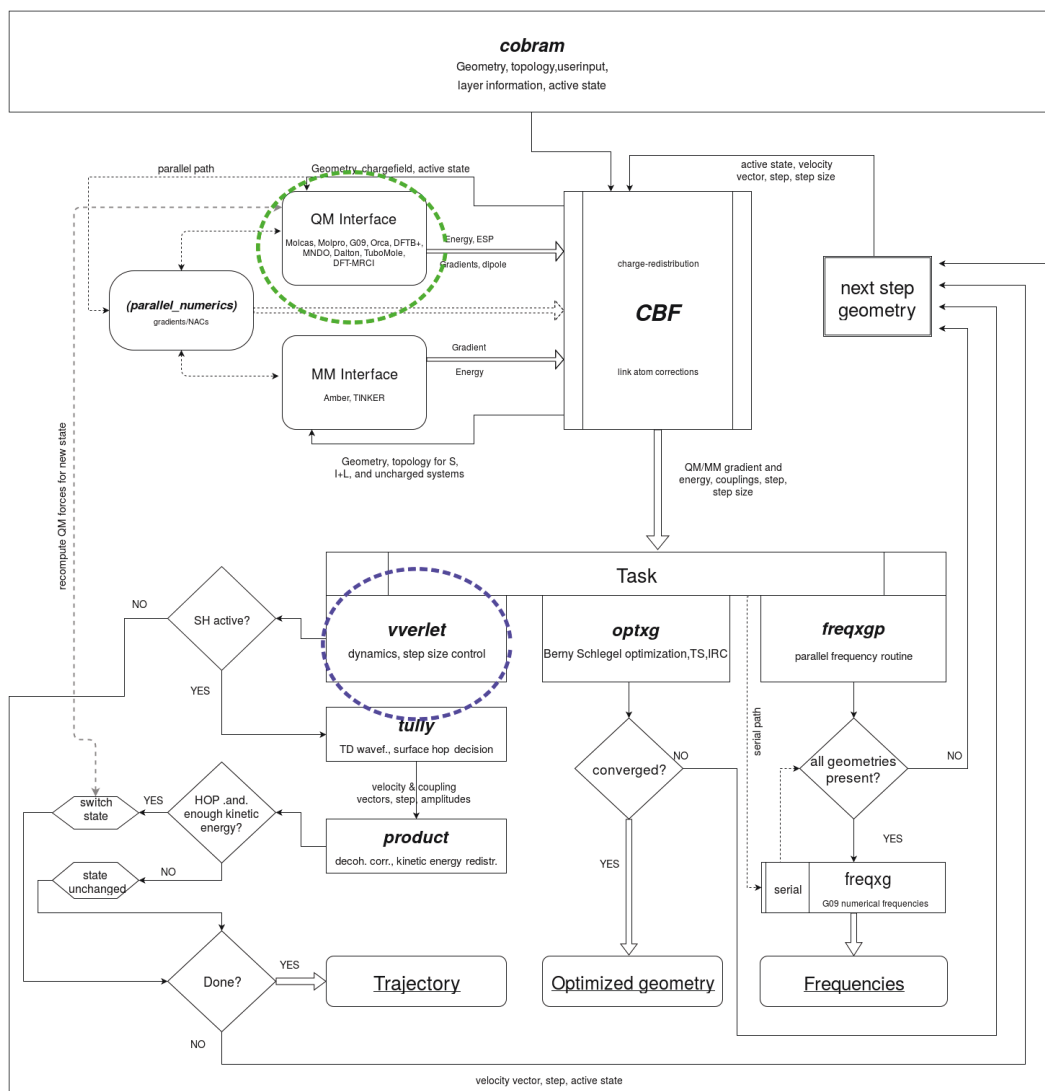


FIGURE 4.1: Basic COBRAMM flowchart. Blue and green circles denote parts of the program, which I modified within the course of my thesis. Scheme from Paper III.

nonadiabatic couplings (NAC) and energy differences (DEarray) required in dynamical calculations.

launch The second routine **launch** starts the MNDO program with the provided input file. The results of a successful calculation are provided in the output files **MNDO.log** and **fort.15**. The **MNDO.log** file holds the QM energies, gradients, dipole moments and transition dipole moments. The **fort.15** file contains QM gradients, charge gradients and nonadiabatic couplings.

mndoEneGradCrg This routine extracts energies, dipole moments, charges, gradients of the QM-region for the state of interest and gradients of the external point charges (MM gradients). It preprocesses the data from the output files and places them at disposal for the corresponding tasks i.e. **Vverlet** for dynamics, **optxg** for optimization or **freqxg** for vibrational analysis (Fig. 4.1). The extracted values are stored in the **Results** array and forwarded to the COBRAMM subroutines. For performing excited state dynamics, nonadiabatic couplings and energy differences of all involved states are required. They are stored in the **NAC** and **DEarray** arrays. Moreover, specific keywords allow to distinguish between SCF and MRCI calculations, where necessary. In the following, the extraction of quantities for static and dynamic calculations will be described separately.

- *Static calculations*

All basic quantities are extracted from the **MNDO.log** file, except for the gradients of the external point charges. If they are not contained in the **MNDO.log** file, they are taken from the **fort.15** file.

- *Dynamic calculations*

In the same manner as for static calculations, energies, dipole moments and charges are taken from the **MNDO.log** file, but NACs and gradients (QM and MM) are extracted from the **fort.15** file. In case of an MRCI calculation, the gradients of all calculated states are collected. Furthermore, energy differences between all states are calculated and stored in **DEarray**.

After each run, the initial **MNDO.input** file is deleted, while the **MNDO.log** file is stored for further analysis.

Technical problems may arise from the active-space CI treatment during OMx/MRCI dynamics, which may lead to discontinuities in the total energy in NVE MD simulations. Spörkel and Thiel suggested a solution for these problems by applying an adaptive time step scheme. [179] I implemented a similar algorithm to improve the quality of my calculations. In the following, the aforementioned difficulties, Spörkel's algorithm and the corrections implemented in the COBRAMM code will be detailed.

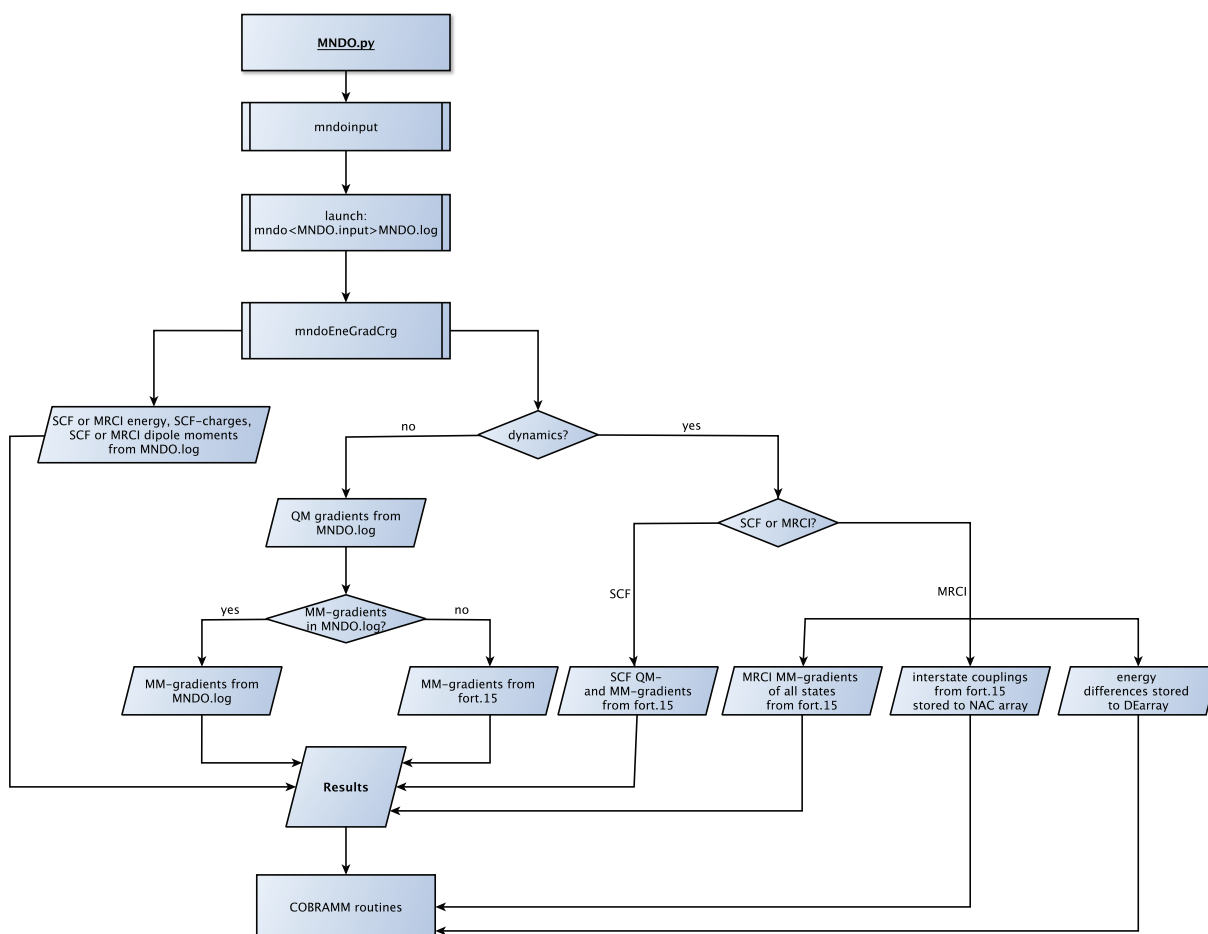


FIGURE 4.2: Basic MNDO interface flowchart.

4.2 Active space related problems

The OMx/MRCI approach requires the definition of active and inactive orbitals from the restricted closed- or open-shell HF (RHF, ROHF) reference wave function. During a dynamics simulation, the geometry changes may, however, lead to a change in the energetic ordering of the orbitals. Energetically close lying orbitals may even mix and strongly change their character from one to the next step. In turn, active orbitals may rotate out of the chosen window and are replaced by the formerly inactive orbitals, or the orbitals strongly differ from the chosen set due to mixing. Both situations will change the composition of the active space and may lead to abrupt and unphysical changes in energies and gradients. In Newtonian dynamics, such corrupted gradients and energies will influence the motion of the molecule, i.e. its velocities and thereby the kinetic energy and the potential itself, leading to loss of energy conservation in an NVE ensemble. This effect is strongly correlated with the step size used in optimization or dynamics computations. To avoid orbital rotations, MNDO includes an orbital monitoring routine that keeps track of the chosen orbitals by computing their overlaps in each step of the computation. When the overlap drops below a certain user-defined value, the computation is aborted and the user is requested to evaluate and correct the calculation. The orbital mixing problem can be partly prevented by selection of a small integration step.

4.3 Adaptive time step algorithm by Spörkel and Thiel

Spörkel and Thiel stated that the aforementioned problems are of local character and are enhanced when the chosen integration time step is too large. A strategy to overcome difficult regions on the PESs is to iteratively reduce the time step when the problem occurs. Such difficult regions are detected by monitoring energy conservation and orbital overlap. The values are compared with a user-defined threshold. When the threshold is exceeded, the integration time step is reduced by half, the data from the previous step is restored and the calculation is performed again using the new time step. This procedure runs iteratively depending on a user defined maximum number of iterations. The algorithm also ensures that after all criteria are satisfied, the initial integration time step is gradually restored.

4.4 Corrections introduced into the MNDO interface

I apply an adaptive time scheme in cases where the orbital tracking procedure fails or when the energy conservation criterium is not met. The algorithm is implemented in the velocity Verlet routine of COBRAMM (`Vverlet.py`), which is highlighted in Figure 4.1 by a blue dashed circle. In the following, implementaion related details will be explained.

- *Orbital tracking*

The threshold for monitoring of the orbital overlap is provided by the user in the `cobram.command` file. In `MNDO.py`, a special keyword is defined to communicate a failed orbital mapping to the integration routine `Vverlet.py`. In this routine, the initial time step is then reduced and subsequently the results from the previous step are restored. The calculation is restarted with the new time step. The process is repeated until the calculation is successful or the maximum number of iterations is reached.

- *Energy conservation*

To maintain energy conservation, the difference between the total energies (E_{tot}) of subsequent steps is calculated and compared to a user-defined threshold provided in the `cobram.command` file. The energy difference is controlled within the routine `Vverlet.py`. If the value exceeds the defined threshold, the time step correction algorithm is invoked and a correction is performed as described before.

In the case of a successful correction, similar to Spörkel's algorithm, the initial time step is gradually restored.

Chapter 5

Results

This chapter will provide a description of various aspects in the photoreaction of retinal in C1C2 obtained from my static and dynamic calculations. The first section will cover results of my first publication attributed to the static QM/MM description of retinal in the channelrhodopsin chimera C1C2 (**Paper I**). [55] The next section is based on a publication describing the properties of protonated Schiff base models and symmetric linear polyenes (**Paper II**). [38] Finally, the last section will give an overview of the results of QM/MM photodynamics calculations with the C1C2 chimera (**Paper IV**). [66]

5.1 *Ab initio* and DFT QM/MM study (Paper I)

The purpose of my first publication was to assess methods for a future description of retinal photodynamics in C1C2. I applied CASSCF, DFT and HF QM/MM strategies for retinal geometry optimizations in the ground state (GS) with subsequent description of its vertical excitation at CASSCF, MS-CASPT2 and DFT/MRCI levels of theory. I furthermore performed CASSCF optimizations in the HOMO–LUMO singly and doubly excited states to document the structural changes upon excitation. First insights into the isomerization process were provided by calculation of the CASSCF/MS-CASPT2 torsion path along the C13=C14 reaction coordinate. Calculation of resonance Raman spectra was also part of this study, as well as an investigation on how the retinal binding pocket influences retinal absorption.

5.1.1 Ground state properties of retinal in C1C2

Ground state geometry optimizations were performed with the retinal chromophore and ϵCH_2 group of the bound lysine (termed RET) in the QM-region and the rest of the attached lysine (termed LYP), the counterions (E162 and D292) and a nearby water molecule in the *movable* MM-region (Fig. 5.1(a)). The remaining portion of the protein was kept fixed at the crystal structure positions. All geometry optimizations started with this setup resulted in a retinal structure forming a hydrogen bond between the terminal NH^+ and D292 (Fig. 5.1(b)). When the protein pocket was allowed to relax, NH^+ is also able to form a hydrogen bond with the E162 counterion (Fig. 5.1(b)), as

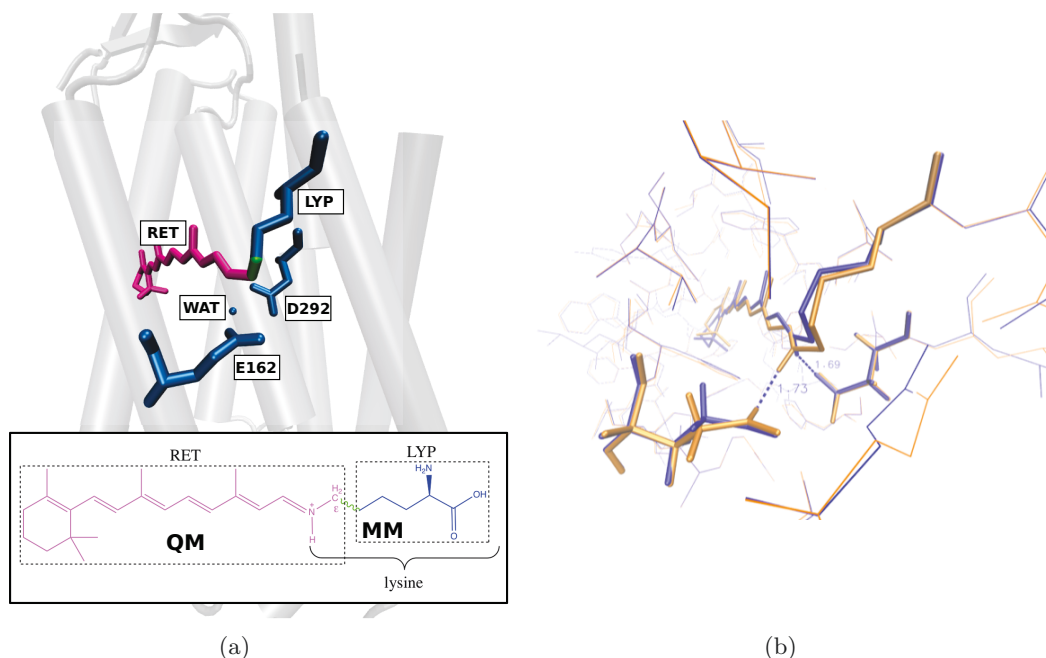


FIGURE 5.1: (a) Partitioning of the QM- and MM-regions. The QM-region is denoted in magenta. Blue color depicts the movable MM-region. Grey color corresponds to the fixed MM-part. The covalent bond between the QM- and MM-regions is highlighted by green color. (b) Models with hydrogen bonding between NH⁺ and E162 (orange) and NH⁺ and D292 (blue). Figure (b) from Paper I.

suggested by Ito *et al.* [69] For further studies, only the D292 bound configuration was considered. DFT and CASSCF ground state geometry optimizations yielded considerably different bond length patterns. While DFT tends to overestimate the conjugation of the π -system, CASSCF resulted in strongly alternating single and double bonds (Fig. 5.2(b)). The consequences of the differing bond length alternation (BLA) are discussed in the next section with regard to the excited state properties of the retinal chromophore.

5.1.2 Absorption properties of retinal in C1C2

Figure 5.2(a) gives an overview of the computed vertical excitation energy values for the different methods. The character of the S_0/S_1 absorption was assigned to HOMO–LUMO single excitation with high oscillator strength by the MS-CASPT2 as well as the DFT/MRCI approaches. The best agreement (479 nm) with the experimental absorption value (ca. 470 nm [34, 69]) was provided by MS-CASPT2 calculations on top of the CASSCF(12,12) optimized ground state retinal geometry. Also DFT/MRCI yielded reasonable results (452 nm) when applied to the CASSCF structure. Calculations at the B3LYP geometry revealed a red-shift compared to the CASSCF optimized structure (Fig. 5.2(a)), a consequence of the reduced bond length alternation described in the last section. The second excited state was dominated by the HOMO²–LUMO²

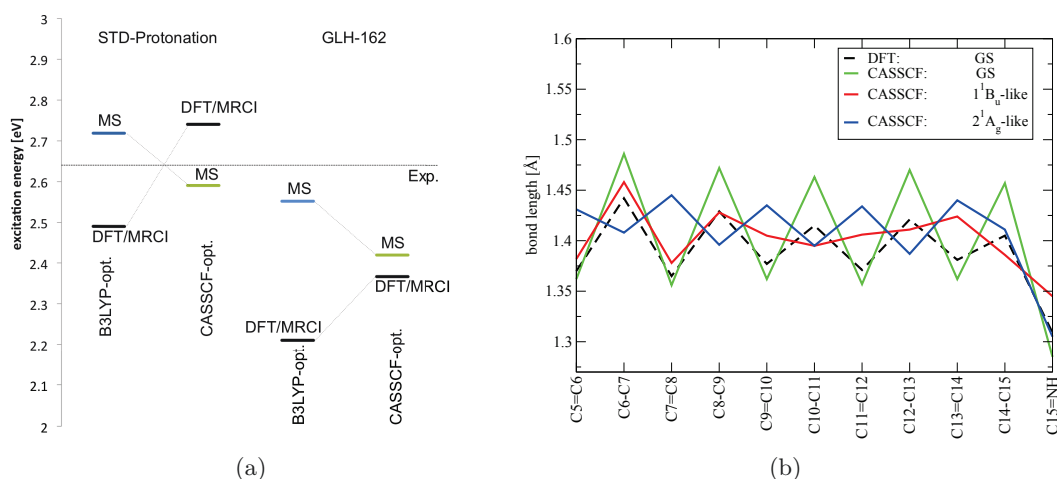


FIGURE 5.2: (a) Calculated MS-CASPT2 and DFT/MRCI vertical excitation energies for standard and alternative (GLH) protonations. Figure from Paper I. (b) Bond lengths of the singly excited HOMO–LUMO (1^1B_u -like) and doubly excited HOMO²–LUMO² (2^1A_g -like) states compared to the ground state DFT and CASSCF bond lengths.

double excitation in MS-CASPT2, while in DFT-MRCI it was characterized by several single excitations. In both descriptions it possesses a lower oscillator strength. The maximum absorption wavelengths computed only at the CASSCF level were far beyond the experimental value. Furthermore, the state ordering for the low-lying excited states was contrary to the expectations. The bright HOMO–LUMO single excitation shifts to the S_3 state, while the doubly excited HOMO²–LUMO² transition was located in the S_1 state. This finding already rules out the usage of CASSCF alone for the description of C1C2 photodynamics calculations.

The protonation state of the counterions in C1C2 is still controversial (Sec. 1.5). For this reason, I also modelled an alternative C1C2 structure with a protonated and thus neutral E162 (GLH-162 model) and repeated the optimization in S_0 . The computed absorption with this model, however, yielded a too red-shifted value relative to the experiment (Fig. 5.2(a)).

The influence of nearby residues was investigated in a series of computations which only considered the charges of one residue plus the background charge distribution of the protein. The largest electrostatic influence on the retinal absorption is provided by charged residues close to the NH^+ -terminus. The rest of the investigated residues had only a minor impact. This finding supports results obtained by Sneskov *et al.* [68] and Kamiya *et al.* [67]

5.1.3 Geometry optimizations in the excited state

To assess the geometry relaxations in the HOMO–LUMO singly and doubly excited states, the geometry was allowed to relax in the CASSCF S_3 (HOMO–LUMO) and S_1

(HOMO²–LUMO²) states. These calculations revealed, at first, an unexpected finding with respect to the bond lengths in PSBs. Bond relaxation in the HOMO–LUMO singly excited state resulted in equalization of the bond lengths, whereas optimization in the doubly excited HOMO²–LUMO² state led to a bond inverted geometry relative to the ground state (Fig. 5.2(b)). This behavior is rather typical for symmetric linear polyenes than for PSBs and indicated a major influence of negative charges in the vicinity of the retinal terminus on its electronic structure. These nearby charges hinder the relocation of positive charge occurring upon HOMO–LUMO excitation (Fig. 1.7) and are responsible for the polyenic character of retinal in C1C2.

The C13=C14 isomerization of retinal was simulated by constrained geometry optimization in the CASSCF HOMO–LUMO state along this coordinate and subsequent MS-CASPT2 energy calculation. The resulting MS-CASPT2 profile is shown in Figure 5.3(b), revealing a bi-directional pathway (arrows in Fig. 5.3(b)) with favored counter-clockwise torsion direction and S₁/S₀ crossings on either side of the potential (ca. 110 and -100 °). The CASSCF energies themselves gave no indication for a conical intersec-

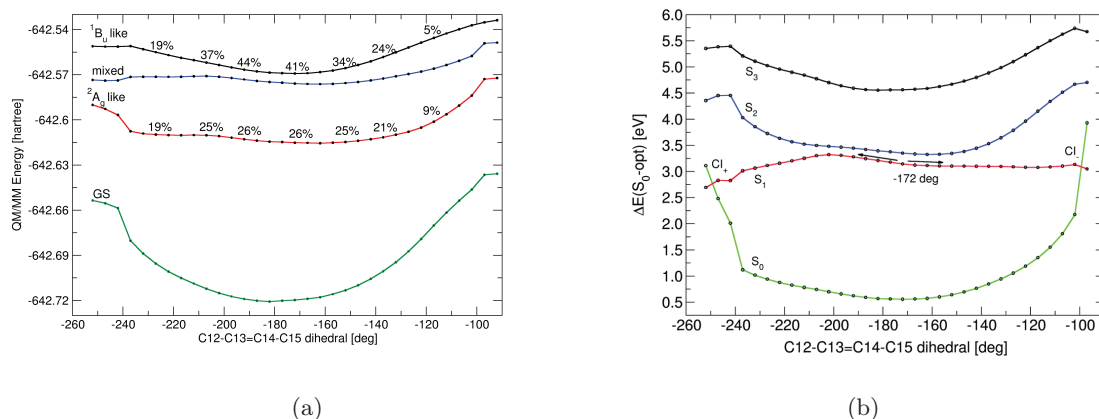


FIGURE 5.3: (a) Relaxed CASSCF C12-C13=C14-C15 torsion scan. (b) MS-CASPT2 corrected energy profile. Figures from Paper I.

tion (Fig. 5.3(a)). The geometrical changes during torsion suggested a hula-twist (HT) like mechanism for retinal isomerization, where the whole C13– ϵ CH₂ fragment moves as one rigid entity (Fig. 5.10). This mechanism results in a 13-*cis* retinal structure where the NH⁺ proton points away from the ASP counterion (13-*cis*,15-*anti*) (Fig. 7 in Paper I). A similar result was also found in the later work of Hontani and co-authors. [72]

5.1.4 Resonance Raman spectra

Although CASSCF could not provide an accurate description of the PESs and excited-state properties, it was applicable for the calculation of resonance Raman (RR) spectra using a simplified approach for spectra calculation. This approach relies only on the consideration of the excited state forces and ground state vibrational modes. [138] The

computed RR spectra (Fig. 5.4, bottom panel) for all-*trans* (black) and 13-*cis* retinal (blue) are shown below. The results were compared with ChR2 spectra measured by Nagel *et al.* (Fig. 5.4, upper panel). I noted reasonable agreement between experi-

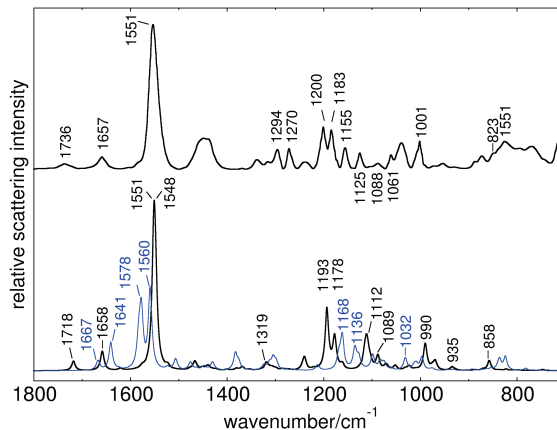


FIGURE 5.4: Calculated resonance Raman spectra of all-*trans* (black) and 13-*cis* (blue) retinal geometries (lower trace) compared to the experimental RR spectrum (upper trace) obtained by Nagel *et al.* for ChR2. Figure from Paper I.

mental ChR2 results and the computed C1C2 all-*trans* spectrum, implying similarities in the electrostatic and steric environments of retinal in ChR2 and C1C2. For C1C2, important retinal fingerprint bands were reproduced. These include C=C and C=N stretching modes, rocking of methyl-groups and hydrogen out-of-plane (HOOP) motion. The spectral signature of 13-*cis* retinal in C1C2 significantly differs from all-*trans* by frequency and intensity of characteristic modes. Hence, a mixture of all-*trans* and 13-*cis* isomers, as suggested by Nagel *et al.*, could not be confirmed computationally. Rather, my theoretical results point toward a strong predominance of all-*trans* retinal in the measured ChR2 sample, a finding that was later verified by the experiments of Baldus *et al.* [73]

5.2 Distinct excited state properties of linear polyenes and PSBs (Paper II)

The most prominent difference between symmetric linear polyenes and protonated Schiff bases is related to the geometries of the low-lying excited states. In polyenic systems, the excited state minima may exhibit a bond length pattern differing from the ground state structure (Sec. 1.4.1). In particular, for thoroughly investigated symmetric linear polyenes it is known that the bright 1^1B_u (HOMO–LUMO) minimum is associated with bond length equalization. Relaxation in the spectroscopically dark 2^1A_g (HOMO²–LUMO²) state, in turn, results in a bond inverted geometry (i.e. former double bonds become single bonds and single bonds become double bonds). In contrast, the

accepted mechanism in PSBs states that bond inversion occurs in the bright HOMO–LUMO state (Sec. 1.4.1). This empirical knowledge was never consequently probed by high-level methodologies, incorporating static and dynamical correlation effects (e.g. MS-CASPT2). This issue is of particular interest for my work, since HOMO–LUMO singly and HOMO²–LUMO² doubly excited states are involved in excited state processes of the retinal protonated Schiff base.

In Paper II, I aimed to examine the differences between electronic properties of symmetric linear polyenes (termed linear polyenes hereafter) and corresponding protonated Schiff base models (termed PSBs within the current section) in vacuum (Fig. 5.5) with MS-CASPT2 methodology. Moreover, available data for polyenic systems and accurate MS-CASPT2 calculations performed during the current study provided a good basis for validation of the newly implemented MNDO interface with regard to future dynamics simulations. In the course of my study, I performed computations using CASSCF,

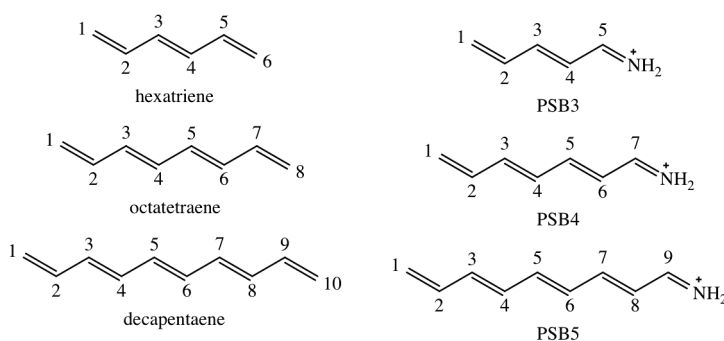


FIGURE 5.5: Symmetric linear polyenes and corresponding protonated Schiff base models. Figure from Paper II.

MS-CASPT2 and OM_x/MRCI ($x=2, 3$) methodologies. My calculations included geometry optimizations in the ground and excited states and the estimation of vertical, adiabatic and emission energies.

5.2.1 The two S_1 minima of PSBs

The S_1 state in PSBs is characterized by bright HOMO–LUMO single excitation and retains this character upon optimization. Indications for two distinct S_1 minima with inverted and unchanged bond length patterns are documented in the literature. [180–183] Mostly, they are described by CASSCF methodology, which is not always a suitable approach. Interestingly, I found that the results obtained for geometry optimization of the bright state are strongly dependent on the selected methodology and optimization strategy. While CASSCF optimization always yielded bond inverted HOMO–LUMO structures, two distinct equilibrium geometries with inverted and non-inverted

bond lengths were obtained with MS-CASPT2. Geometry optimization based on the ground state initial structure led to a ground state like bond length pattern in the HOMO–LUMO state. In turn, S_1 -geometry optimizations started with pre-inverted bonds resulted in an inverted bond length pattern. Subsequent calculation of the interpolated paths between the inverted and non-inverted minima revealed a small barrier (ca. 0.1–0.2 eV) connecting both structures. Figure 5.6(a) depicts the two minima in PSB5. Obviously, the excited state geometries of linear polyenes and corresponding PSBs possess different structures in a non-polar environment. This finding supports my conclusion from the previous study that the polyenic character of retinal in C1C2 is induced by its negatively charged environment. Another issue emerging from these results is the insufficient performance of the CASSCF approach resulting from its lack of dynamic correlation.

The existence of the two HOMO–LUMO minima raised the question about their possible impact on the photoreaction in PSBs. To address this question I computed unrelaxed PES scans about all bonds in PSB5 using inverted and non-inverted optimized geometries as a starting point. Torsion about the central bond led both structures towards a conical intersection between the ground and the first excited states. Figure 5.6(b) depicts a C4–C5=C6–C7 torsion scan based on the inverted geometry. While for this case the torsion was almost barrierless, the non-inverted minimum revealed a ca. 0.4 eV barrier (Fig. 5 in Paper II). This result is plausible considering the sin-

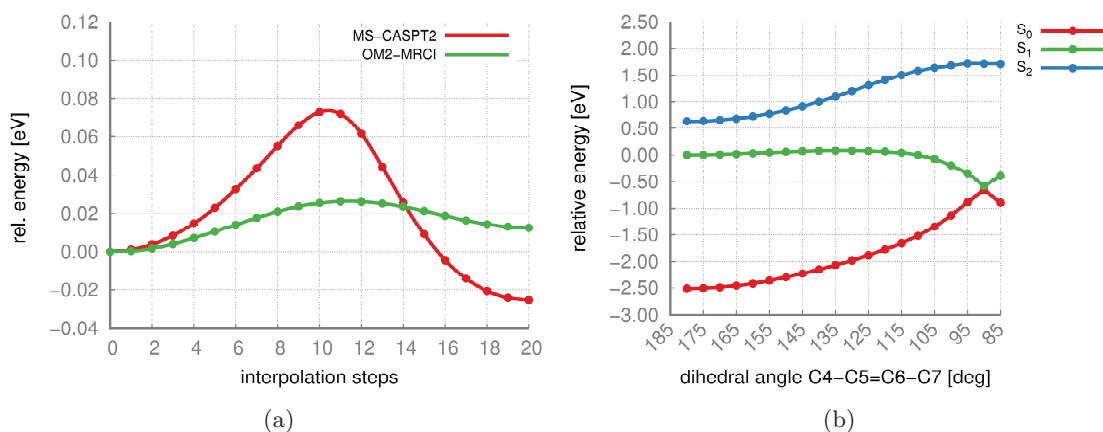


FIGURE 5.6: (a) Interpolated path for S_1 inverted (left minimum) and non-inverted (right minimum) equilibrium geometries calculated with MS-CASPT2 and OM2-MRCI for PSB5. (b) MS-CASPT2 torsion profile along the central double bond of PSB5 starting from the inverted geometry. Figures adopted from Paper II.

gle bond character of the central bond in the inverted structure and its double bond character in the non-inverted geometry.

5.2.2 OMx/MRCI at a test

The OMx/MRCI computations performed in this study confirmed the main trends for linear polyenes known from the literature. Furthermore, they were in agreement with my MS-CASPT2 results for linear polyenes as well as for longer PSBs. Two S_1 minima with HOMO–LUMO excitation character (with inverted and non-inverted bond order) were located for PSB4 and PSB5 (Fig. 5.6(a)), while PSB3 revealed only the non-inverted equilibrium structure. An OM2/MRCI unrelaxed torsion scan about the central bond of PSB5 indicated the presence of a CoIn between the ground and the first excited states also at this level (Fig. 5 in Paper II). Barrierless torsion was found for the scan starting with the inverted geometry, while a ca. 0.15 eV barrier was located on the path starting with the non-inverted structure. The remarkable agreement of the OM2/MRCI torsion scans with the MS-CASPT2 results illustrates that this method enables a reliable description of the PESs in PSBs. However, to achieve such accurate results a large active space comprising not only π -orbitals is required for the semiempirical treatment.

No significant difference was found between the OM2 and OM3/MRCI methods. Thus, the semiempirical OMx/MRCI methods provide a suitable and computationally affordable alternative to MS-CASPT2 approach.

5.3 Excited state dynamics simulations (Paper IV)

Based on the previous validation of the OMx-methodology with regard to polyenes and particularly PSBs, the OM3/MRCI approach was finally applied to perform nonadiabatic trajectory surface hopping dynamics. Prior to trajectory calculations, static properties of retinal were assessed in order to additionally validate the OM3/MRCI method within a protein environment. Static calculations comprised geometry optimizations (ground and excited states), calculations of excitation energies and computation of excited state relaxed torsion paths. Dynamics simulations were based on the ZPE sampling procedure (Sec. 3.6.1) [162], generating initial geometries and velocities.

5.3.1 Static calculations

Ground and excited state geometries

QM/MM ground state geometry optimizations were based on three models (ASP, GLU1 and GLU2) obtained through ground state geometry optimizations with different mobile regions in the MM part. The results of geometry optimizations for the different models are shown in Figure 5.7. The simplest model (ASP) contained RET in the QM-region and LYP in the movable MM-region. Optimization of this model led to formation of a hydrogen bond between the NH^+ -terminus of the chromophore and the neighboring counterion D292 (Fig. 5.7(a)). The GLU1 model is an extension of the ASP model,

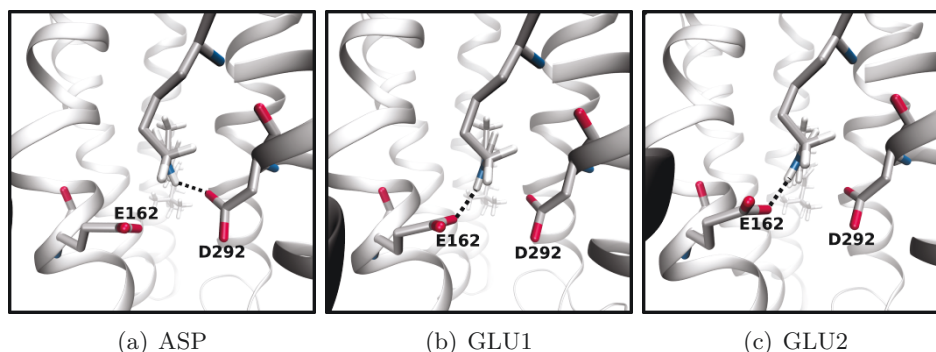


FIGURE 5.7: Orientation of the NH^+ -terminus in ground-state optimized minima of (a) ASP, (b) GLU1 and (c) GLU2 models. Hydrogen bonds with the corresponding counterion are depicted by a dashed line. Figures from Paper IV.

comprising four (K132; E162; T166; D292) additional residues in the movable MM-region. After geometry relaxation of the GLU1 model, the hydrogen bond bridges the NH^+ -terminus and the counterion E162 (Fig. 5.7(b)). Finally, GLU2 had the largest portion of mobile residues within 5 Å distance from retinal. Similar to GLU1, the hydrogen bond in the GLU2 model is formed between NH^+ and E162 (Fig. 5.7(c)). While the overall chromophore structure is rather planar and somewhat bent in ASP and GLU1, the chain in the GLU2 model is straight with additional twists along the polyene chain about the C6-C7 and C8-C9 bonds (Fig. 4 in Paper IV). Moreover, in the GLU2 model the C13- NH^+ fragment exhibits a larger twist compared to the GLU1 model. Already in the first CASSCF/CASPT2 study (Paper I), the existence of two ground state hydrogen bonding patterns between RPSB and counterions was indicated. For all further excited state calculations only RET (QM-region) and LYP (MM-region) were movable during the computations. The results for optimized ground and excited states resemble those already obtained in the first CASSCF/CASPT2 study, with pronounced bond length equalization in the HOMO-LUMO singly excited state and inverted bond lengths in the HOMO²-LUMO² doubly excited state (Fig. 5 in Paper IV). This result again confirms the polyenic nature of retinal in C1C2 due to its negatively charged environment.

Electronic structure

The electronic structure of all calculated models exhibited similar properties. At the ground state geometry, the S_1 state was characterized by the expected bright HOMO-LUMO excitation. The calculated S_0/S_1 absorption was, however, strongly blue shifted (0.64-0.84 eV) at the OM3/MRCI level. This effect was already documented in the literature, but it has no impact on the qualitative performance of the OMx/MRCI approach. [175–177] The S_2 state is governed by the almost dark HOMO²-LUMO² double excitation. Excited state character and state ordering at the S_0 geometry were in good

agreement with the results already obtained in the previous MS-CASPT2 calculations. Upon geometry relaxation of the S_1 and S_2 states, a change of the state ordering due to a crossing of these states was noted. The crossing was confirmed by a linearly interpolated S_0/S_1 path (Fig. 6 in Paper IV). Thus, after initial excitation to the bright S_1 HOMO–LUMO state the molecule relaxes toward the bond inverted $\text{HOMO}^2\text{--LUMO}^2$ doubly excited state. A similar result was obtained with MS-CASPT2 (Fig. 2 in the SI of Paper IV). This finding shows that, despite producing overestimated vertical excitation energies, the OMx/MRCI strategy provides an accurate qualitative description for the electronic structure of RPSB in C1C2.

Torsion paths

Torsion paths for all models were calculated in the S_1 state along the reactive C13=C14 coordinate (Fig. 7 in Paper IV). Similar to my previous CASSCF/MS-CASPT2 study, bi-directional torsion is in principle possible and produces two CoIns. While torsion toward one of the CoIns is almost barrierless, the second CoIn can be reached after overcoming a small barrier. In the ASP and GLU1 models, counterclockwise torsion appears favored, while the clockwise direction is the preferred direction in the GLU2 model. It is worth noting that in the GLU1 model the barriers in both torsion directions are rather small. In general, the excess energy after excitation would be sufficient to react through both possible channels. Regarding the preferable clockwise torsion direction, the GLU2 model agrees with the results of Hontani and co-authors. [72] The torsion begins in the $\text{HOMO}^2\text{--LUMO}^2$ doubly excited state which changes its character to HOMO–LUMO singly excited near the intersection point. However, the orbitals on the twisted geometry at the CoIn differ from those at the FC point and show a more localized character (Fig. 4 in the SI of Paper IV).

5.3.2 Photodynamics simulations

Photoreaction–excited state processes and time constants

The trajectory ensemble was started in the initially bright S_1 state. The subsequent de-excitation in general involves three processes reflected in the population graphs (Fig. 8 in Paper IV): a) partial depopulation of the S_1 state and simultaneous partial population of the S_2 state; b) entire depopulation of the S_2 state and almost complete repopulation of the S_1 state and c) final relaxation from the S_1 state to the ground state. In Paper IV I provide exponentially fitted time constants for the processes b) and c). The time constant of the process c) is experimentally relevant and was compared with the results of Hontani *et al.* The authors suggested three time components for the S_1 decay: 450 fs, 2 ps and 11 ps. Within my simulations, I estimated three different time constants, i.e. one for each model. The exponentially fitted reaction time of ASP (ca. 285 fs) was in reasonable agreement with the experiment. The slower de-excitation (ca.

6 ps) of GLU1 (obtained through extrapolation) is comparable with the unproductive 2 and 11 ps time constants obtained by Hontani *et al.* De-excitation of the S_1 state in the GLU2 model was, however, too fast (ca. 95 fs) and not consistent with the experiment. A possible rationale for this finding is given in one of the next sections (Sec. **Factors influencing the photoreaction**).

Photoproducts—quantum yields, deactivation channels and reaction mechanisms

Quantum yields: In total, four distinct photoproducts were obtained during the simulations (Fig. 5.8). The identified isomers include two distinct *trans* and two distinct *cis* species. The experimentalists report only the quantum yield for the 13-*cis* isomer,

	all- <i>trans</i> ,15- <i>anti</i>	13- <i>cis</i> ,15- <i>anti</i>	13- <i>cis</i> ,15- <i>syn</i>	all- <i>trans</i> ,15- <i>syn</i>
ASP	56%	29%	3%	12%
GLU1	7%	5%	-	1%
GLU2	34%	66%	-	-

FIGURE 5.8: Possible photoproducts and corresponding quantum yields.

i.e. regardless of the orientation of the NH^+ proton. My calculations based on the ASP model yielded ca. 30% of the 13-*cis* isomer (considering both *anti* and *syn* geometries). This value already perfectly reproduces the overall experimental quantum yield. Reasonable quantum yields for the GLU1 model could not be estimated, because at the end of the simulation most of the trajectories still remained in the excited state. With 66% quantum yield, GLU2 appeared as the most productive model. In addition to its high quantum yield this model is characterized by a strong product selectivity, i.e. the photoreaction produced only all-*trans*,15-*anti* and 13-*cis*,15-*anti* structures.

Reaction channels: The torsion scans discussed in the previous subsection indicated a bi-directional isomerization. This conclusion is, however, only reflected in the dynamics of the GLU1 model, where the few trajectories hopping to the ground state relax through clockwise and counterclockwise reaction channels (Fig. 5.9, lower left panel). In contrast, exclusive counterclockwise torsion was found in the ASP model (Fig. 5.9, upper panel), while GLU2 exhibited only torsion in the clockwise direction (Fig. 5.9, lower right panel). Furthermore, each of the observed reaction channels may either proceed to the 13-*cis* photoproduct or back to the all-*trans* geometry. This outcome shows that, in principle, both clockwise and counterclockwise torsions can be productive.

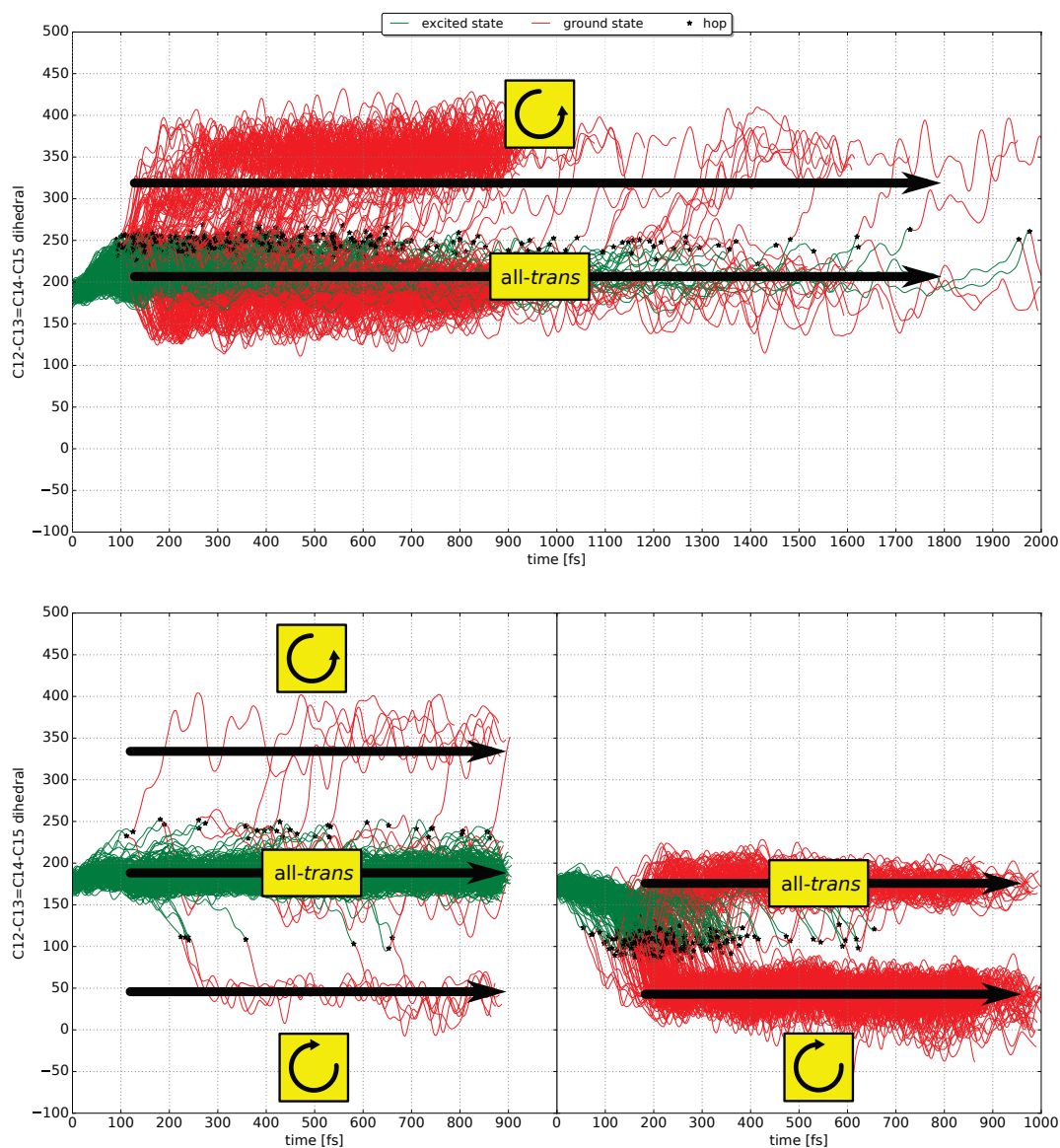


FIGURE 5.9: Evolution of the C12-C13=C14-C15 torsion in surface hopping dynamics of the ASP (upper panel), GLU1 (lower left panel) and GLU2 (lower right panel) models. Green color denotes motion in the excited state, while red color depicts relaxation to the ground state. Figure from Paper IV.

Isomerization mechanisms: Two possible 13-*cis* isomers (shown in Figure 5.8) originate from two distinct isomerization mechanisms (Fig. 5.10). The bicycle-pedal (BP)

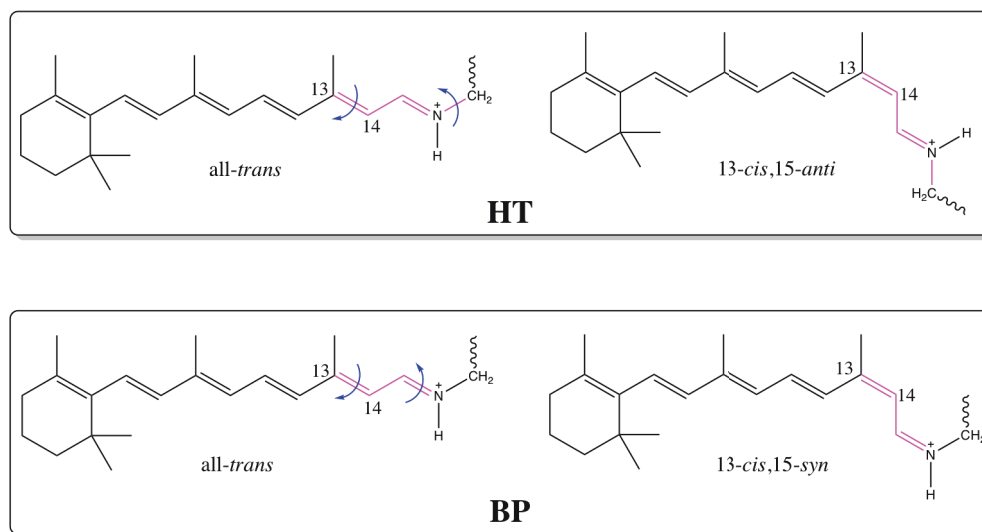


FIGURE 5.10: Hula-twist (HT) and bicycle-pedal (BP) reaction mechanisms identified in the course of the dynamics simulations. Figure from Paper IV.

mechanism results in isomerization toward the 13-*cis*,15-*syn* structure, while the space demanding hula-twist like (HT) mechanism is responsible for the 13-*cis*,15-*anti* isomer. The most frequently formed *cis*-photoproduct is 13-*cis*,15-*anti*, indicating that there is enough space in the retinal binding pocket to perform this motion.

Factors influencing the photoreaction

Successful tuning of the photoreaction requires knowledge about the factors that influence it. In the course of my simulations, I identified several factors affecting the speed and the efficiency of the photoreaction.

Influence on the reaction speed: Different reaction times observed in the investigated models may be attributed to (a) electrostatic and (b) structural effects. Compared to the ASP model, electrostatic environment near the RPSB terminus changes in GLU models. While the distance between the NH^+ and the counterions remains almost the same in all models, positively charged lysine (K132) moves closer to the negatively charged counterions (Fig. 11 in the SI of Paper IV) in the GLU models. Such arrangement may partly shield the negative charge of the counterions. As reported in the literature, the influence of negative counterions is sufficient for the photoreaction, while a neutral environment slows it down. [61] Apparently, this effect is observed in the GLU1 model and is counterbalanced in the fast GLU2 model through structural features of the chromophore. The reason for the extremely fast photoreaction of the latter model is the occurrence of strong twists at the C6-C7 and C8-C9 retinal bonds and an additional twist of the C13-NH⁺ fragment, more pronounced than in the GLU1

model (Fig. 4 and Fig. 12 in Paper IV). Such twists introduce an extra momentum during the chromophore dynamics. As a consequence, torsional motion (into the pre-twisted direction) sets in earlier and promotes faster de-excitation.

Influence on the reaction outcome: In the introductory section (Sec. 1.4.3), I explained how hydrogen out-of-plane motion and the number of PES approaches before the hopping event may influence the outcome of the photoreaction. I captured this relationship in C1C2 considering angular velocities of the two dihedrals: C17-C13=C14-H14 (out-of-plane) and C12-C13=C14-C15 (reaction defining coordinate) (Fig. 5.11(a)) at different PES approaches (Fig. 5.11(b)). The results of this analysis are shown in Figure 15 of Paper IV. Trajectories hopping at the first surface approach rather tend to

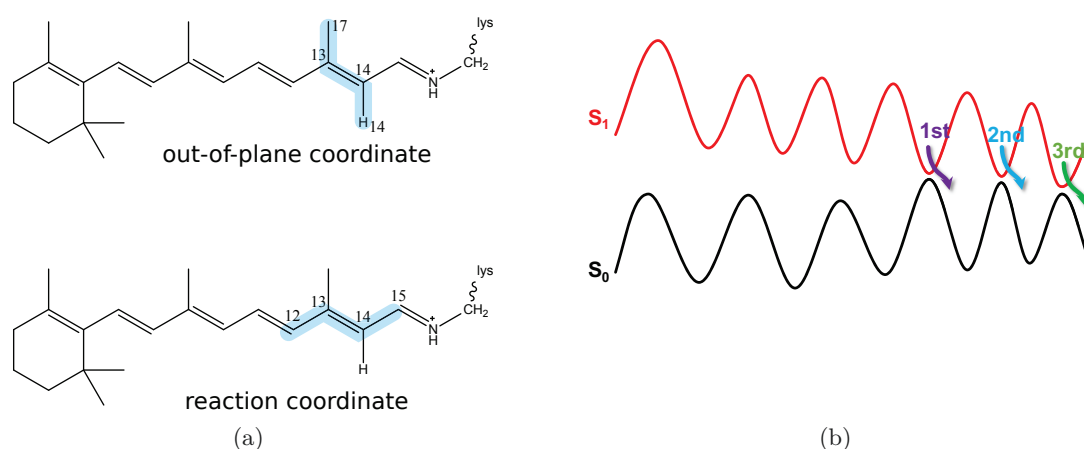


FIGURE 5.11: (a) Out-of-plane and reaction coordinates. (b) Illustration of surface approaches and corresponding hopping events at first (1st), second (2nd) and third (3rd) surface encounters indicated by arrows.

perform a successful *cis*-isomerization, whereas already at the second surface encounter a significant amount of the *trans* isomer is obtained. The fast GLU2 model revealed the largest portion of trajectories hopping at the first surface approach. This effect certainly contributed to the high 13-*cis* quantum yield estimated for this model and also indicates a relationship between the reaction time and surface hopping decision. I.e. high reaction speed leads to a higher probability of hopping at the first PES approach.

Reaction dynamics of retinal

This section aims to provide a rather simplified and general description of excited state retinal dynamics in C1C2. It is based on insights gained through static and dynamic calculations combined with simulated pump-probe spectra. The photoreaction starts in the S₁ state characterized by the bright HOMO–LUMO excitation (¹B_u-like). Immediately after excitation, C–C stretching vibrations set in and lead to transitions between the S₁ and S₂ states – a finding supported by static (linearly interpolated paths) and dynamic calculations (population plots). Subsequent geometry relaxation drives the system toward the HOMO²–LUMO² (²¹A_g-like) doubly excited S₁ state.

This part of the de-excitation process is common for all three models. The next step in the ASP and GLU2 models involves torsional motion about the C13=C14 bond (with single bond character in the excited state) in the 2^1A_g -like S_1 state, which leads the molecule toward the S_1/S_0 CoIn. Shortly before the relaxation to the ground state the state character changes to an open shell configuration. Subsequently, if the torsion is completed, *cis* photoproduct is formed and *trans* photoproduct occurs in the opposite case. In contrast to the ASP and GLU2 models, most of the trajectories from the GLU1 set lack torsional motion and stay trapped in the S_1 state during the simulation time.

Chapter 6

Conclusions and Outlook

6.1 Concluding remarks

The central result of my thesis is the realisation of nonadiabatic excited state ensemble dynamics in order to describe retinal isomerization in C1C2. For ChRs, such a study was conducted for the first time. Furthermore, the application of the OMx/MRCI strategy in this context was also unprecedented. The quality of the results obtained from OMx-calculations may encourage further application of semiempirical methods for the investigation of other retinal binding proteins (e.g. Rh, bR etc.). This strategy certainly provides a valuable alternative to the rather demanding MS-CASPT2 approach at much less computational cost. However, the resulting optimized geometries and absorption values have to be interpreted carefully. In particular, for retinal in C1C2 I could note artificial structural changes (twist in the retinal structure) and unrealistic blue-shifted absorption. For an accurate description of excited states with OMx/MRCI, a large active space, including not only π -electrons, should be envisaged. Application of CASSCF for retinal characterization in channelrhodopsins is not appropriate, since this method was not capable to correctly reproduce the state ordering of the excited states, the absorption energy and the PESs themselves. Limitations of the method originate from a lack of dynamic correlation, which is required to describe the close lying HOMO–LUMO singly and doubly excited states of retinal in C1C2. Also, in PSB models certain effects (e.g. excited-state minima) could not be captured with CASSCF. Nevertheless, resonance Raman spectra based on CASSCF results were in fair agreement with the experimental data. Best results in static calculations were achieved with MS-CASPT2 and DFT/MRCI, methods that are unfeasible due to either high computational cost and/or lack of analytical gradients.

The insights gained from static computations provide an important basis for dynamical calculations. They enable the prediction of certain trends and may, to a certain extend, explain processes occurring in photodynamics. They are, however in no case a replacement for a dynamical treatment. The most prominent evidence for this was the relation between static torsion profiles and reaction channels in the dynamical study of

the current work. For three investigated models (ASP⁴, GLU1⁵, GLU2⁶), bi-directional torsion scans were calculated, whereas during the dynamics bi-directionality was confirmed only in one of the three models. Also, the efficiency of the photoreaction and its kinetics can only be accessed through a dynamical approach. Two of the three models considered in the dynamical study offered good agreement with the recent experimental findings of Hontani *et al.* These models are ASP and GLU1, which match the fast and slow components from the experiment. Apparently, a single model is not sufficient to describe retinal dynamics in the rather flexible C1C2 binding pocket. Furthermore, this result indicates that the NH⁺-terminus can form a hydrogen bond with both counterions, E162 and D292. The fast GLU2 model was not in agreement with the experiment. It, however, offers valuable insight into the relationship between structure and photoreaction of retinal. The GLU2 model indicated that torsion introduced at specific locations of the chromophore accelerates the relaxation to the GS. This, in turn, influences the successful formation of the 13-*cis* photoproduct. This trend suggests modification strategies for more efficient chromophores.

Furthermore, my computations revealed interaction of the close-lying HOMO–LUMO singly and doubly excited states of retinal in C1C2. Consequently, both states are populated during photodynamics.

Another core aspect of the current study is related to the comparative investigation of short symmetric linear polyenes and PSB models. It not only provided a benchmark for OMx/MRCI methods, but also contributed to the basic knowledge about polyenic systems. In my work I described the difference between the excited-state behaviour of linear polyenes and PSBs applying the highly accurate MS-CASPT2 approach. For the PSBs, I located two HOMO–LUMO S₁ minima characterized by inverted and non-inverted bond lengths of the polyenic chain. In the polyenes, an inverted equilibrium structure corresponds to the nuclear arrangements in the 2¹A_g state governed by HOMO²–LUMO² double excitation. Interestingly, electrostatic interaction of RPSB with the negatively charged environment in C1C2 results in a polyenic nature of retinal. Thus, an equilibrium structure of the chromophore characterized by the HOMO–LUMO single excitation exhibited bond length equalization, while relaxation in the HOMO²–LUMO² doubly excited state yielded an inverted bond length pattern. In conclusion, it can be stated that the current project is the first comprehensive study of retinal isomerization in ChRs which ranges from static to dynamical computations.

6.2 Outlook

The calculations conducted in this thesis were based on the crystal structure of the chimera protein C1C2, which was the only one available when the PhD project started.

⁴Hydrogen bond between NH⁺ and D292.

⁵Hydrogen bond between NH⁺ and E162, planar retinal structure.

⁶Hydrogen bond between NH⁺ and E162, partially twisted retinal structure.

Recently, a crystal structure of the experimentally highly relevant ChR2 was reported. The knowledge gained in the course of my work, in particular the established protocol for nonadiabatic dynamics, can certainly be applied for simulations based on the newly discovered crystal structure.

In the course of my studies, different ground-state geometries of retinal were obtained, depending on the optimization strategy, which involved different numbers of movable residues. This effect indicates a greater flexibility of the protein pocket and suggests to investigate the influence of a movable environment during dynamics. In the course of semiempirical computations, I faced limitations of the OMx methods related to an overestimated maximum absorption of retinal and artificial twist in the retinal structure. Hence, a new parametrization of OM2 and OM3/MRCI could be helpful in this context.

Retinal isomerization is the first and central step in the activation of ChRs. Subsequent steps in the photocycle, involving de- and reprotonation of retinal, are of further importance, especially for the channel-gating mechanism. Information (absorption properties, protonation state) about the photointermediates in the photocycle is provided for C1C2 as well as for ChR2. However, only the spectral signatures are known, while the structure of retinal is unknown. Computational assessment of photointermediates was not provided yet and would also be possible to address. Especially interesting within the photocycle is the proton transfer pathway. The 13-*cis*,15-*anti* conformation of retinal, which moves the NH^+ -terminus away from the counterions, suggests a bR-like proton transfer mechanism. Furthermore, no consistent opinion on the primary proton acceptor exists until now.

In the course of my dynamical calculations, the occurrence of the 13-*cis*,15-*syn* photoproduct was detected. This structure was not reported yet for C1C2. However, in ChR2 it is suggested as a second ground state conformation leading to a branched photocycle. Dynamical calculations in C1C2 and ChR2 based on the 13-*cis*,15-*syn* isomer might provide a valuable addition to experimental findings and a basis for the description of the second part of a photocycle.

My final recommendation concerns the implemented interface. Although the implementation of the MNDO interface provided a reasonable platform for semiempirical QM/MM MD simulations, further code improvement can be suggested. I noted that MNDO calculations may stop due to failed SCF wave function convergence. With the methods applied in this study a decreasing time step correction for this problem is, however, not reasonable. This would only slow down the process of reaching the non-convergence data-point. An increase of the time step for a short period may, on the other hand, help to overcome the critical point on the PES.

Bibliography

- [1] O. P. Ernst, D. T. Lodowski, M. Elstner, P. Hegemann, L. S. Brown, H. Kandori, “Microbial and animal rhodopsins: structures, functions, and molecular mechanisms”, *Chem. Rev.* **2013**, *114*, 126–163.
- [2] L. O. Björn, *Photobiology: The science of light and life*, Springer, New York, **2015**.
- [3] N. Nelson, C. F. Yocum, “Structure and function of photosystems I and II”, *Annu. Rev. Plant Biol.* **2006**, *57*, 521–565.
- [4] J. L. Spudich, C.-S. Yang, K.-H. Jung, E. N. Spudich, “Retinylidene proteins: structures and functions from archaea to humans”, *Annu. Rev. Cell Dev. Biol.* **2000**, *16*, 365–392.
- [5] F. Zhang, J. Vierock, O. Yizhar, L. E. Fenno, S. Tsunoda, A. Kianianmomeni, M. Prigge, A. Berndt, J. Cushman, J. Polle, M. Jon, H. Peter, D. Karl, “The microbial opsin family of optogenetic tools”, *Cell* **2011**, *147*, 1446–1457.
- [6] H. Kandori, Y. Shichida, T. Yoshizawa, “Photoisomerization in rhodopsin”, *Biochem.* **2001**, *66*, 1197–1209.
- [7] R. R. Birge, “Photophysics and molecular electronic applications of the rhodopsins”, *Annu. Rev. Phys. Chem.* **1990**, *41*, 683–733.
- [8] D. Oesterhelt, W. Stoeckenius, “Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*”, *Nature New Biol.* **1971**, *233*, 149–152.
- [9] A. Matsuno-Yagi, Y. Mukohata, “Two possible roles of bacteriorhodopsin; a comparative study of strains of *Halobacterium halobium* differing in pigmentation”, *Biochem. Biophys. Res. Commun.* **1977**, *78*, 237–243.
- [10] J. L. Spudich, R. A. Bogomolni, “Mechanism of colour discrimination by a bacterial sensory rhodopsin”, *Nature* **1984**, *312*, 509–513.
- [11] G. Nagel, T. Szellas, W. Huhn, S. Kateriya, N. Adeishvili, P. Berthold, D. Ollig, P. Hegemann, E. Bamberg, “Channelrhodopsin-2, a directly light-gated cation-selective membrane channel”, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13940–13945.
- [12] G. Nagel, D. Ollig, M. Fuhrmann, S. Kateriya, A. M. Musti, E. Bamberg, P. Hegemann, “Channelrhodopsin-1: a light-gated proton channel in green algae”, *Science* **2002**, *296*, 2395–2398.

- [13] K. Deisseroth, P. Hegemann, “The form and function of channelrhodopsin”, *Science* **2017**, *357*, 5544.
- [14] I. C. Farber, A. Grinvald, “Identification of presynaptic neurons by laser photostimulation”, *Science* **1983**, *222*, 1025–1027.
- [15] A. Guru, R. J. Post, Y.-Y. Ho, M. R. Warden, “Making sense of optogenetics”, *Int. J. Neuropsychopharmacol.* **2015**, *18*, 1–8.
- [16] G. P. Dugué, W. Akemann, T. Knöpfel, “A comprehensive concept of optogenetics” in *Progress in brain research*, Vol. 196, Elsevier, **2012**, pp. 1–28.
- [17] E. S. Boyden, “Optogenetics and the future of neuroscience”, *Nat. Neurosci.* **2015**, *18*, 1200–1201.
- [18] Y. Mei, F. Zhang, “Molecular tools and approaches for optogenetics”, *Biol. Psychiatry* **2012**, *71*, 1033–1038.
- [19] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, K. Deisseroth, “Milli-second-timescale, genetically targeted optical control of neural activity”, *Nat. Neurosci.* **2005**, *8*, 1263–1268.
- [20] X. Han, E. S. Boyden, “Multiple-color optical activation, silencing, and desynchronization of neural activity, with single-spike temporal resolution”, *PLoS One* **2007**, *2*, 299.
- [21] B. R. Arenkiel, J. Peca, I. G. Davison, C. Feliciano, K. Deisseroth, G. J. Augustine, M. D. Ehlers, G. Feng, “*In vivo* light-induced activation of neural circuitry in transgenic mice expressing channelrhodopsin-2”, *Neuron* **2007**, *54*, 205–218.
- [22] F. Schneider, C. Grimm, P. Hegemann, “Biophysics of channelrhodopsin”, *Annu. Rev. Biophys.* **2015**, *44*, 167–186.
- [23] A. V. Kravitz, B. S. Freeze, P. R. Parker, K. Kay, M. T. Thwin, K. Deisseroth, A. C. Kreitzer, “Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry”, *Nature* **2010**, *466*, 622–626.
- [24] H. Tomita, E. Sugano, H. Isago, T. Hiroi, Z. Wang, E. Ohta, M. Tamai, “Channelrhodopsin-2 gene transduced into retinal ganglion cells restores functional vision in genetically blind rats”, *Exp. Eye Res.* **2010**, *90*, 429–436.
- [25] T. Bruegmann, D. Malan, M. Hesse, T. Beierr, C. J. Fügemann, B. K. Fleischmann, P. Sasse, “Optogenetic control of heart muscle *in vitro* and *in vivo*”, *Nat. Methods* **2010**, *7*, 897–900.
- [26] P. Hegemann, A. Möglich, “Channelrhodopsin engineering and exploration of new optogenetic tools”, *Nat. Methods* **2011**, *8*, 39–42.
- [27] A. Famintzin, “Die Wirkung des Lichtes auf Algen und einige andere ihnen nahe verwandte Organismen”, *Jb. Wiss. Bot.* **1867**, *6*, 1–48.

- [28] A. Kianianmomeni, K. Stehfest, G. Nematollahi, P. Hegemann, A. Hallmann, “Channelrhodopsins of *Volvox carteri* are photochromic proteins that are specifically expressed in somatic cells under control of light, temperature, and the sex inducer”, *Plant Physiol.* **2009**, *151*, 347–366.
- [29] F. Zhang, M. Prigge, F. Beyrière, S. P. Tsunoda, J. Mattis, O. Yizhar, P. Hegemann, K. Deisseroth, “Red-shifted optogenetic excitation: a tool for fast neural control derived from *Volvox carteri*”, *Nat. Neurosci.* **2008**, *11*, 631–633.
- [30] A. Berndt, S. Y. Lee, C. Ramakrishnan, K. Deisseroth, “Structure-guided transformation of channelrhodopsin into a light-activated chloride channel”, *Science* **2014**, *344*, 420–424.
- [31] J. Wietek, R. Beltramo, M. Scanziani, P. Hegemann, T. G. Oertner, J. S. Wiegert, “An improved chloride-conducting channelrhodopsin for light-induced inhibition of neuronal activity *in vivo*”, *Scientific reports* **2015**, *5*, 14807.
- [32] A. Berndt, S. Y. Lee, J. Wietek, C. Ramakrishnan, E. E. Steinberg, A. J. Rashid, H. Kim, S. Park, A. Santoro, P. W. Frankland, S. M. Iyer, S. Pak, S. Åhrlund-Richter, S. L. Delp, R. C. Malenka, S. A. Josselyn, M. Carlén, P. Hegemann, K. Deisseroth, “Structural foundations of optogenetics: Determinants of channelrhodopsin ion selectivity”, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 822–829.
- [33] E. G. Govorunova, O. A. Sineshchekov, R. Janz, X. Liu, J. L. Spudich, “Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics”, *Science* **2015**, *349*, 7484.
- [34] H. E. Kato, F. Zhang, O. Yizhar, C. Ramakrishnan, T. Nishizawa, K. Hirata, J. Ito, Y. Aita, T. Tsukazaki, S. Hayashi, P. Hegemann, A. D. Maturana, R. Ishitani, K. Deisseroth, O. Nureki, “Crystal structure of the channelrhodopsin light-gated cation channel”, *Nature* **2012**, *482*, 369–374.
- [35] O. Volkov, K. Kovalev, V. Polovinkin, V. Borshchevskiy, C. Bamann, R. Astashkin, E. Marin, A. Popov, T. Balandin, D. Willbold, G. Büldt, E. Bamberg, V. Gordeliy, “Structural insights into ion conduction by channelrhodopsin 2”, *Science* **2017**, *358*, 8862.
- [36] A. Inaguma, H. Tsukamoto, H. E. Kato, T. Kimura, T. Ishizuka, S. Oishi, H. Yawo, O. Nureki, Y. Furutani, “Chimeras of channelrhodopsin-1 and -2 from *Chlamydomonas reinhardtii* exhibit distinctive light-induced structural changes from channelrhodopsin-2”, *J. Biol. Chem.* **2015**, 115.
- [37] S. C. Kinsky, “Polyene antibiotics” in *Antibiotics*, Springer, Berlin, **1967**.
- [38] I. Dokukina, C. M. Marian, O. Weingart, “New Perspectives on an Old Issue: A Comparative MS-CASPT2 and OM2-MRCI Study of Polyenes and Protonated Schiff Bases”, *Photochem. Photobiol.* **2017**, *93*, 1345–1355.

- [39] Y. Kurashige, H. Nakano, Y. Nakao, K. Hirao, "The $\pi \rightarrow \pi^*$ excited states of long linear polyenes studied by the CASCI-MRMP method", *Chem. Phys. Lett.* **2004**, *400*, 425–429.
- [40] T. Polívka, V. Sundström, "Ultrafast dynamics of carotenoid excited states—from solution to natural and artificial systems", *Chem. Rev.* **2004**, *104*, 2021–2072.
- [41] K. Schulten, M. Karplus, "On the origin of a low-lying forbidden transition in polyenes and related molecules", *Chem. Phys. Lett.* **1972**, *14*, 305–309.
- [42] C. M. Marian, N. Gilka, "Performance of the density functional theory/multi-reference configuration interaction method on electronic excitation of extended π -systems", *J. Chem. Theory Comput.* **2008**, *4*, 1501–1515.
- [43] F. Blomgren, S. Larsson, "Bond length alternation in ground and HOMO \rightarrow LUMO excited states in polyenes. Dynamic Stokes shift?", *Theor. Chem. Acc.* **2003**, *110*, 165–169.
- [44] C. Angeli, M. Pastore, "The lowest singlet states of octatetraene revisited", *J. Chem. Phys.* **2011**, *134*, 184302.
- [45] K. Schulten, U. Dinur, B. Honig, "The spectra of carbonium ions, cyanine dyes, and protonated Schiff base polyenes", *J. Chem. Phys.* **1980**, *73*, 3927–3935.
- [46] M. Garavelli, F. Negri, M. Olivucci, "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", *J. Am. Chem. Soc.* **1999**, *121*, 1023–1029.
- [47] I. Schapiro, M. N. Ryazantsev, L. M. Frutos, N. Ferré, R. Lindh, M. Olivucci, "The ultrafast photoisomerizations of rhodopsin and batho-rhodopsin are modulated by bond length alternation and HOOP driven electronic effects", *J. Am. Chem. Soc.* **2011**, *133*, 3354–3364.
- [48] I. Schapiro, M. N. Ryazantsev, W. J. Ding, M. M. Huntress, F. Melaccio, T. Andruniow, M. Olivucci, "Computational photobiology and beyond", *Aust. J. Chem.* **2010**, *63*, 413–429.
- [49] G. Pescitelli, H. E. Kato, S. Oishi, J. Ito, A. D. Maturana, O. Nureki, R. W. Woody, "Exciton circular dichroism in channelrhodopsin", *J. Phys. Chem. B* **2014**, *118*, 11873–11885.
- [50] A. Cembran, F. Bernardi, M. Olivucci, M. Garavelli, "Counterion controlled photoisomerization of retinal chromophore models: a computational investigation", *J. Am. Chem. Soc.* **2004**, *126*, 16018–16037.
- [51] I. Rivalta, A. Nenov, M. Garavelli, "Modelling retinal chromophores photoisomerization: from minimal models *in vacuo* to ultimate bidimensional spectroscopy in rhodopsins", *Phys. Chem. Chem. Phys.* **2014**, *16*, 16865–16879.

- [52] A. Cembran, F. Bernardi, M. Olivucci, M. Garavelli, “The retinal chromophore/chloride ion pair: Structure of the photoisomerization path and interplay of charge transfer and covalent states”, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6255–6260.
- [53] P. E. Blatz, J. H. Mohler, H. V. Navangul, “Anion-induced wavelength regulation of absorption maxima of Schiff bases of retinal”, *Biochem.* **1972**, *11*, 848–855.
- [54] L. H. Andersen, I. B. Nielsen, M. B. Kristensen, M. O. El Ghazaly, S. Haacke, M. B. Nielsen, M. Å. Petersen, “Absorption of Schiff-base retinal chromophores *in vacuo*”, *J. Am. Chem. Soc.* **2005**, *127*, 12347–12350.
- [55] I. Dokukina, O. Weingart, “Spectral properties and isomerisation path of retinal in C1C2 channelrhodopsin”, *Phys. Chem. Chem. Phys.* **2015**, *17*, 25142–25150.
- [56] H. E. Kato, M. Kamiya, S. Sugo, J. Ito, R. Taniguchi, A. Orito, K. Hirata, A. Inutsuka, A. Yamanaka, A. D. Maturana, R. Ishitani, Y. Sudo, S. Hayashi, O. Nureki, “Atomistic design of microbial opsin-based blue-shifted optogenetics tools”, *Nat. Commun.* **2015**, *6*, 7177–7187.
- [57] O. Weingart, *Ab initio* Moleküldynamik der Photoisomerisierung von protonierten Polyenaldehyd Schiff Basen, PhD thesis, Universität Duisburg-Essen, Fakultät für Chemie, **2005**.
- [58] S. Gozem, H. L. Luk, I. Schapiro, M. Olivucci, “Theory and simulation of the ultrafast double-bond isomerization of biological chromophores”, *Chem. Rev.* **2017**, *117*, 13502–13565.
- [59] M. Garavelli, P. Celani, F. Bernardi, M. Robb, M. Olivucci, “The $C_5H_6NH_2^+$ protonated Schiff base: an *ab initio* minimal model for retinal photoisomerization”, *J. Am. Chem. Soc.* **1997**, *119*, 6891–6901.
- [60] L. Song, M. El-Sayed, J. Lanyi, “Protein catalysis of the retinal subpicosecond photoisomerization in the primary process of bacteriorhodopsin photosynthesis”, *Science* **1993**, *261*, 891–894.
- [61] F. Scholz, E. Bamberg, C. Bamann, J. Wachtveitl, “Tuning the primary reaction of channelrhodopsin-2 by imidazole, pH, and site-specific mutations”, *Biophys. J.* **2012**, *102*, 2649–2657.
- [62] O. Weingart, “The role of HOOP-modes in the ultrafast photoisomerization of retinal models”, *Chem. Phys.* **2008**, *349*, 348–355.
- [63] O. Weingart, P. Altoè, M. Stenta, A. Bottoni, G. Orlandi, M. Garavelli, “Product formation in rhodopsin by fast hydrogen motions”, *Phys. Chem. Chem. Phys.* **2011**, *13*, 3645–3648.
- [64] N. Klaffki, O. Weingart, M. Garavelli, E. Spohr, “Sampling excited state dynamics: influence of HOOP mode excitations in a retinal model”, *Phys. Chem. Chem. Phys.* **2012**, *14*, 14299–14305.

- [65] O. Weingart, M. Garavelli, “Modelling vibrational coherence in the primary rhodopsin photoproduct”, *J. Chem. Phys.* **2012**, *137*, 523.
- [66] I. Dokukina, A. Nenov, M. Garavelli, C. M. Marian, O. Weingart, “QM/MM photodynamics of retinal in the channelrhodopsin chimera C1C2 with OM3/MRCI”, *ChemPhotoChem*, *submitted 2018*.
- [67] M. Kamiya, H. E. Kato, R. Ishitani, O. Nureki, S. Hayashi, “Structural and spectral characterizations of C1C2 channelrhodopsin and its mutants by molecular simulations”, *Chem. Phys. Lett.* **2013**, *556*, 266–271.
- [68] K. Snedkov, J. M. H. Olsen, T. Schwabe, C. Hättig, O. Christiansen, J. Kongsted, “Computational screening of one- and two-photon spectrally tuned channelrhodopsin mutants”, *Phys. Chem. Chem. Phys.* **2013**, *15*, 7567–7576.
- [69] S. Ito, H. E. Kato, R. Taniguchi, T. Iwata, O. Nureki, H. Kandori, “Water-containing hydrogen-bonding network in the active center of channelrhodopsin”, *J. Am. Chem. Soc.* **2014**, *136*, 3475–3482.
- [70] M. Takemoto, H. E. Kato, M. Koyama, J. Ito, M. Kamiya, S. Hayashi, A. D. Maturana, K. Deisseroth, R. Ishitani, O. Nureki, “Molecular dynamics of channelrhodopsin at the early stages of channel opening”, *PLoS One* **2015**, *10*, 0131094.
- [71] M. R. VanGordon, G. Gyawali, S. W. Rick, S. B. Rempe, “Atomistic study of intramolecular interactions in the closed-state channelrhodopsin chimera, C1C2”, *Biophys. J.* **2017**, *112*, 943–952.
- [72] Y. Hontani, M. Marazzi, K. Stehfest, T. Mathes, I. H. Stokkum, M. Elstner, P. Hegemann, J. T. Kennis, “Reaction dynamics of the chimeric channelrhodopsin C1C2”, *Sci. Rep.* **2017**, *7*, 7217.
- [73] J. Becker-Baldus, C. Bamann, K. Saxena, H. Gustmann, L. J. Brown, R. C. D. Brown, C. Reiter, E. Bamberg, J. Wachtveitl, H. Schwalbe, C. Glaubitz, “Enlightening the photoactive site of channelrhodopsin-2 by DNP-enhanced solid-state NMR spectroscopy”, *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 9896–9901.
- [74] S. Bruun, D. Stoeppler, A. Keidel, U. Kuhlmann, M. Luck, A. Diehl, M.-A. Geiger, D. Woodmansee, D. Trauner, P. Hegemann, H. Oschkinat, P. Hildebrandt, K. Stehfest, “Light–Dark Adaptation of Channelrhodopsin Involves Photoconversion between the all-*trans* and 13-*cis* Retinal Isomers”, *Biochem.* **2015**, *54*, 5389–5400.
- [75] S. Adam, A.-N. Bondar, “Mechanism by which water and protein electrostatic interactions control proton transfer at the active site of channelrhodopsin”, *PLoS One* **2018**, *13*, 0201298.

- [76] D. Polli, P. Altoè, O. Weingart, K. M. Spillane, C. Manzoni, D. Brida, G. Tomasello, G. Orlandi, P. Kukura, R. A. Mathies, M. Garavelli, G. Cerullo, “Conical intersection dynamics of the primary photoisomerization event in vision”, *Nature* **2010**, *467*, 440.
- [77] S. Hayashi, E. Tajkhorshid, K. Schulten, “Photochemical reaction dynamics of the primary event of vision studied by means of a hybrid molecular simulation”, *Biophys. J.* **2009**, *96*, 403–416.
- [78] S. Hayashi, E. Tajkhorshid, K. Schulten, “Molecular dynamics simulation of bacteriorhodopsin’s photoisomerization using *ab initio* forces for the excited chromophore”, *Biophys. J.* **2003**, *85*, 1440–1449.
- [79] O. Weingart, A. Nenov, P. Altoè, I. Rivalta, J. Segarra-Martí, I. Dokukina, M. Garavelli, “COBRAMM 2.0 — A software interface for tailoring molecular electronic structure calculations and running nanoscale (QM/MM) simulations”, *J. Mol. Model.* **2018**, *24*, 271.
- [80] P. Atkins, J. De Paula, J. Keeler, *Atkins’ physical chemistry*, Oxford university press, Oxford, **2018**.
- [81] F. Bernardi, M. Olivucci, M. A. Robb, “Potential energy surface crossings in organic photochemistry”, *Chem. Soc. Rev.* **1996**, *25*, 321–328.
- [82] W. Domcke, D. Yarkony, *Conical intersections: electronic structure, dynamics & spectroscopy*, Vol. 15, World Scientific, Singapore, **2004**.
- [83] J. von Neumann, E. Wigner, “Über das Verhalten von Eigenwerten bei adiabatischen Prozessen” in *The Collected Works of Eugene Paul Wigner*, Springer, **1993**, pp. 294–297.
- [84] G. Herzberg, H. Longuet-Higgins, “Intersection of potential energy surfaces in polyatomic molecules”, *Discuss. Faraday Soc.* **1963**, *35*, 77–82.
- [85] E. Teller, “The Crossing of Potential Surfaces.”, *J. Phys. Chem.* **1937**, *41*, 109–116.
- [86] M. A. Robb, M. Garavelli, M. Olivucci, F. Bernardi, “A computational strategy for organic photochemistry”, *Rev. Comput. Chem.* **2000**, *15*, 87–146.
- [87] S. Matsika, “Conical Intersections in Molecular Systems”, *Rev. Comput. Chem.* **2007**, *30*, 83.
- [88] I. N. Ragazos, M. A. Robb, F. Bernardi, M. Olivucci, “Optimization and characterization of the lowest energy point on a conical intersection using an MC-SCF Lagrangian”, *Chem. Phys. Lett.* **1992**, *197*, 217–223.
- [89] G. A. Worth, L. S. Cederbaum, “Beyond Born-Oppenheimer: molecular dynamics through a conical intersection”, *Annu. Rev. Phys. Chem.* **2004**, *55*, 127–158.

- [90] J. P. Malhado, M. J. Bearpark, J. T. Hynes, “Non-adiabatic dynamics close to conical intersections and the surface hopping perspective”, *Front. Chem.* **2014**, *2*, 97.
- [91] X. Zhu, D. R. Yarkony, “Non-adiabaticity: the importance of conical intersections”, *Mol. Phys.* **2016**, *114*, 1983–2013.
- [92] A. Szabo, N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*, Courier Corporation, Mineola, **2012**.
- [93] M. Born, R. Oppenheimer, “Zur Quantentheorie der Molekeln”, *Ann. Phys.* **1927**, *389*, 457–484.
- [94] M. Born, K. Huang, *Dynamical theory of crystal lattices*, Clarendon press, Oxford, **1954**.
- [95] F. Jensen, *Introduction to computational chemistry*, John Wiley & Sons, Chichester, **2007**.
- [96] E. G. Lewars, *Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics*, Springer Science & Business Media, Dordrecht, **2010**.
- [97] B. O. Roos, “The complete active space self-consistent field method and its applications in electronic structure calculations”, *Advances in Chemical Physics: Ab Initio Methods in Quantum Chemistry Part 2* **1987**, *69*, 399–445.
- [98] R. Pauncz, *The Symmetric Group in Quantum Chemistry*, CRC Press, Boca Raton, **1995**.
- [99] B. O. Roos, L. Roland, P. Å. Malmqvist, V. Valera, P. Widmark, *Multiconfigurational Quantum Chemistry*, Wiley-Blackwell, Amsterdam, **2016**, Chapter 9.
- [100] S. Reine, T. Saue, *European Summerschool in Quantum Chemistry Book I*, Pirulla, Palermo, **2015**.
- [101] K. K. Docken, J. Hinze, “LiH potential curves and wavefunctions for $X\ ^1\Sigma^+$, $A\ ^1\Sigma^+$, $B\ ^1\Pi$, $^3\Sigma^+$, and $^3\Pi$ ”, *J. Chem. Phys.* **1972**, *57*, 4928–4936.
- [102] P.-Å. Malmqvist, A. Rendell, B. O. Roos, “The restricted active space self-consistent-field method, implemented with a split graph unitary group approach”, *J. Phys. Chem.* **1990**, *94*, 5477–5482.
- [103] K. Andersson, P.-Å. Malmqvist, B. O. Roos, “Second-order perturbation theory with a complete active space self-consistent field reference function”, *J. Chem. Phys.* **1992**, *96*, 1218–1226.
- [104] J. Finley, P.-Å. Malmqvist, B. O. Roos, L. Serrano-Andrés, “The multi-state CASPT2 method”, *Chem. Phys. Lett.* **1998**, *288*, 299–306.
- [105] P. Hohenberg, W. Kohn, “Inhomogeneous electron gas”, *Phys. Rev.* **1964**, *136*, 864.

- [106] W. Kohn, L. J. Sham, "Self-consistent equations including exchange and correlation effects", *Phys. Rev.* **1965**, *140*, 1133.
- [107] P. A. M. Dirac, "Quantum mechanics of many-electron systems", *Proc. R. Soc. Lond. A* **1929**, *123*, 714–733.
- [108] J. C. Slater, "A simplification of the Hartree-Fock method", *Phys. Rev.* **1951**, *81*, 385.
- [109] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior", *Phys. Rev. A* **1988**, *38*, 3098.
- [110] C. Lee, W. Yang, R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density", *Phys. Rev. B* **1988**, *37*, 785.
- [111] A. D. Becke, "A new mixing of Hartree-Fock and local density-functional theories", *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- [112] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange", *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [113] S. Grimme, M. Waletzke, "A combination of Kohn-Sham density functional theory and multi-reference configuration interaction methods", *J. Chem. Phys.* **1999**, *111*, 5645–5655.
- [114] I. Lyskov, M. Kleinschmidt, C. M. Marian, "Redesign of the DFT/MRCI Hamiltonian", *J. Chem. Phys.* **2016**, *144*, 034104.
- [115] J. Sadlej, *Semi-empirical methods of quantum chemistry*, Halsted Press, Hertfordshire, **1985**.
- [116] C. C. J. Roothaan, "New Developments in Molecular Orbital Theory", *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- [117] "The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials", *Proc. Royal Soc. Lond.* **1951**, *205*, 541–552.
- [118] R. Pariser, R. G. Parr, "A Semi-Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. I.", *J. Chem. Phys.* **1953**, *21*, 466–471.
- [119] J. A. Pople, D. P. Santry, G. A. Segal, "Approximate self-consistent molecular orbital theory. I. Invariant procedures", *J. Chem. Phys.* **1965**, *43*, 129–135.
- [120] R. A. Evarestov, *Quantum chemistry of solids: the LCAO first principles treatment of crystals*, Vol. 153, Springer Science & Business Media, Berlin, **2007**.
- [121] M. J. Dewar, W. Thiel, "Ground states of molecules. 38. The MNDO method. Approximations and parameters", *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.

- [122] M. J. Dewar, E. G. Zoebisch, E. F. Healy, J. J. Stewart, "Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model", *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- [123] J. J. Stewart, "Optimization of parameters for semiempirical methods II. Applications", *J. Comput. Chem.* **1989**, *10*, 221–264.
- [124] P. O. Dral, X. Wu, L. Spörkel, A. Koslowski, W. Weber, R. Steiger, M. Scholten, W. Thiel, "Semiempirical quantum-chemical orthogonalization-corrected methods: theory, implementation, and parameters", *J. Chem. Theory Comput.* **2016**, *12*, 1082–1096.
- [125] W. Thiel, "Semiempirical quantum-chemical methods", *WIREs Comput. Mol. Sci.* **2014**, *4*, 145–157.
- [126] M. Kolb, W. Thiel, "Beyond the MNDO model: Methodical considerations and numerical results", *J. Comput. Chem.* **1993**, *14*, 775–789.
- [127] M. Kolb, Ein neues semiempirisches Verfahren auf Grundlage der NDDO Näherung, PhD thesis, Universität Wuppertal, **1991**.
- [128] W. Weber, Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Parametrisierung und Anwendungen, PhD thesis, Universität Zürich, **1996**.
- [129] W. Weber, W. Thiel, "Orthogonalization corrections for semiempirical methods", *Theor. Chem. Acc.* **2000**, *103*, 495–506.
- [130] M. Scholten, Semiempirical Methods with Orthogonalization Corrections: The OM3 Method, PhD thesis, Heinrich-Heine-Universität Düsseldorf, **2003**.
- [131] D. Tuna, Y. Lu, A. Koslowski, W. Thiel, "Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks of electronically excited states", *J. Chem. Theory Comput.* **2016**, *12*, 4400–4422.
- [132] A. Koslowski, M. E. Beck, W. Thiel, "Implementation of a general multireference configuration interaction procedure with analytic gradients in a semiempirical context using the graphical unitary group approach", *J. Comput. Chem.* **2003**, *24*, 714–726.
- [133] S. Patchkovskii, A. Koslowski, W. Thiel, "Generic implementation of semi-analytical CI gradients for NDDO-type methods", *Theor. Chem. Acc.* **2005**, *114*, 84–89.
- [134] S. Patchkovskii, W. Thiel, "Analytical first derivatives of the energy for small CI expansions", *Theor. Chem. Acc.* **1997**, *98*, 1–4.
- [135] W. Demtröder, *Laserspektroskopie: Grundlagen und Techniken*, Springer, Berlin, **2007**.

- [136] F. Negri, G. Orlandi, A. M. Brouwer, F. W. Langkilde, R. Wilbrandt, "The lowest triplet state of 1, 3, 5-hexatrienes: Quantum chemical force field calculations and experimental resonance Raman spectra", *J. Chem. Phys.* **1989**, *90*, 5944–5963.
- [137] F. Negri, G. Orlandi, F. W. Langkilde, R. Wilbrandt, "A quantum chemical analysis of the time-resolved resonance Raman spectrum of 2,5-dimethyl-2,4-hexadiene in the T_1 state", *J. Chem. Phys.* **1990**, *92*, 4907–4919.
- [138] A. ten Wolde, H. J. Jacobs, F. W. Langkilde, K. Bajdor, R. Wilbrandt, F. Negri, F. Zerbetto, G. Orlandi, "Triplet state resonance Raman and absorption spectroscopy of a configurationally locked (Z)-hexatriene: 1,2-divinylcyclopentene", *J. Phys. Chem.* **1994**, *98*, 9437–9445.
- [139] S. A. Asher, "UV resonance Raman studies of molecular structure and dynamics: applications in physical and biophysical chemistry", *Annu. Rev. Phys. Chem.* **1988**, *39*, 537–588.
- [140] T. Schlick, *Molecular modeling and simulation: an interdisciplinary guide*, Springer Science & Business Media, New York, **2010**.
- [141] A. C. Van Duin, S. Dasgupta, F. Lorant, W. A. Goddard, "ReaxFF: a reactive force field for hydrocarbons", *J. Phys. Chem. A* **2001**, *105*, 9396–9409.
- [142] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules", *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- [143] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field", *Proteins: Struct. Funct. Bioinf.* **2010**, *78*, 1950–1958.
- [144] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters", *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 712–725.
- [145] A. Warshel, M. Levitt, "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme", *J. Mol. Biol.* **1976**, *103*, 227–249.
- [146] H. M. Senn, W. Thiel, "QM/MM methods for biomolecular systems", *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
- [147] M. Svensson, S. Humbel, R. D. Froese, T. Matsubara, S. Sieber, K. Morokuma, "ONIOM: a multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and $\text{Pt}(\text{P}(t\text{-Bu})_3)_2 + \text{H}_2$ oxidative addition", *J. Phys. Chem.* **1996**, *100*, 19357–19363.

- [148] F. Maseras, K. Morokuma, "IMOMM: A new integrated *ab initio*+molecular mechanics geometry optimization scheme of equilibrium structures and transition states", *J. Comput. Chem.* **1995**, *16*, 1170–1179.
- [149] P. Altoè, M. Stenta, A. Bottoni, M. Garavelli, "A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction", *Theor. Chem. Acc.* **2007**, *118*, 219–240.
- [150] H. Lin, D. G. Truhlar, "QM/MM: what have we learned, where are we, and where do we go from here?", *Theor. Chem. Acc.* **2007**, *117*, 185.
- [151] M. Barbatti, "Nonadiabatic dynamics with trajectory surface hopping method", *WIREs Comput. Mol. Sci.* **2011**, *1*, 620–633.
- [152] A. W. Jasper, C. Zhu, S. Nangia, D. G. Truhlar, "Introductory lecture: Nonadiabatic effects in chemical dynamics", *Faraday Discuss.* **2004**, *127*, 1–22.
- [153] H.-D. Meyer, U. Manthe, L. S. Cederbaum, "The multi-configurational time-dependent Hartree approach", *Chem. Phys. Lett.* **1990**, *165*, 73–78.
- [154] M. Ehara, H.-D. Meyer, L. Cederbaum, "Multiconfiguration time-dependent Hartree (MCTDH) study on rotational and diffractive inelastic molecule-surface scattering", *J. Chem. Phys.* **1996**, *105*, 8865–8877.
- [155] X. Li, J. C. Tully, H. B. Schlegel, M. J. Frisch, "*Ab initio* Ehrenfest dynamics", *J. Chem. Phys.* **2005**, *123*, 084106.
- [156] N. L. Doltsinis, "Nonadiabatic dynamics: mean-field and surface hopping", *Quantum simulations of complex many-body systems: from theory to algorithms* **2002**, *10*, 377–397.
- [157] J. C. Tully, R. K. Preston, "Trajectory surface hopping approach to nonadiabatic molecular collisions: the reaction of H^+ with D_2 ", *J. Chem. Phys.* **1971**, *55*, 562–572.
- [158] A. Bjerre, E. Nikitin, "Energy transfer in collisions of an excited sodium atom with a nitrogen molecule", *Chem. Phys. Lett.* **1967**, *1*, 179–181.
- [159] S. Hammes-Schiffer, J. C. Tully, "Proton transfer in solution: Molecular dynamics with quantum transitions", *J. Chem. Phys.* **1994**, *101*, 4657–4667.
- [160] O. Weingart, "Combined Quantum and Molecular Mechanics (QM/MM) Approaches to Simulate Ultrafast Photodynamics in Biological Systems", *Curr. Org. Chem.* **2017**, *21*, 586–601.
- [161] M. Barbatti, K. Sen, "Effects of different initial condition samplings on photodynamics and spectrum of pyrrole", *Int. J. Quantum Chem.* **2016**, *116*, 762–771.

- [162] C. S. Sloane, W. L. Hase, "On the dynamics of state selected unimolecular reactions: chloroacetylene dissociation and predissociation", *J. Chem. Phys.* **1977**, *66*, 1523–1533.
- [163] J. Tully, "Mixed quantum–classical dynamics", *Faraday Discuss.* **1998**, *110*, 407–419.
- [164] J. C. Tully, "Nonadiabatic molecular dynamics", *Int. J. Quantum Chem.* **1991**, *40*, 299–309.
- [165] J. C. Tully, "Molecular dynamics with electronic transitions", *J. Chem. Phys.* **1990**, *93*, 1061–1071.
- [166] E. Fabiano, T. W. Keal, W. Thiel, "Implementation of surface hopping molecular dynamics using semiempirical methods", *Chem. Phys.* **2008**, *349*, 334–347.
- [167] D. Polli, I. Rivalta, A. Nenov, O. Weingart, M. Garavelli, G. Cerullo, "Tracking the primary photoconversion events in rhodopsins by ultrafast optical spectroscopy", *Photochem. Photobiol. Sci.* **2015**, *14*, 213–228.
- [168] B. J. Schwartz, E. R. Bittner, O. V. Prezhdo, P. J. Rossky, "Quantum decoherence and the isotope effect in condensed phase nonadiabatic molecular dynamics simulations", *J. Chem. Phys.* **1996**, *104*, 5942–5955.
- [169] A. Toniolo, S. Olsen, L. Manohar, T. Martinez, "Conical intersection dynamics in solution: the chromophore of green fluorescent protein", *Faraday Discuss.* **2004**, *127*, 149–163.
- [170] C. Zhu, A. W. Jasper, D. G. Truhlar, "Non-Born-Oppenheimer Liouville-von Neumann dynamics. Evolution of a subsystem controlled by linear and population-driven decay of mixing with decoherent and coherent switching", *J. Chem. Theory Comput.* **2005**, *1*, 527–540.
- [171] G. Granucci, M. Persico, "Critical appraisal of the fewest switches algorithm for surface hopping", *J. Chem. Phys.* **2007**, *126*, 134114.
- [172] L. Verlet, "Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules", *Phys. Rev.* **1967**, *159*, 98.
- [173] T. W. Keal, M. Wanko, W. Thiel, "Assessment of semiempirical methods for the photoisomerisation of a protonated Schiff base", *Theor. Chem. Acc.* **2009**, *123*, 145–156.
- [174] T. W. Keal, A. Koslowski, W. Thiel, "Comparison of algorithms for conical intersection optimisation using semiempirical methods", *Theor. Chem. Acc.* **2007**, *118*, 837–844.
- [175] M. Wanko, M. Hoffmann, P. Strodel, A. Koslowski, W. Thiel, F. Neese, T. Frauenheim, M. Elstner, "Calculating absorption shifts for retinal proteins: computational challenges", *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

- [176] Y. Guo, F. E. Beyle, B. M. Bold, H. C. Watanabe, A. Koslowski, W. Thiel, P. Hegemann, M. Marazzi, M. Elstner, “Active site structure and absorption spectrum of channelrhodopsin-2 wild-type and C128T mutant”, *Chem. Sci.* **2016**, *7*, 3879–3891.
- [177] K. Welke, J. S. Frahmcke, H. C. Watanabe, P. Hegemann, M. Elstner, “Color Tuning in Binding Pocket Models of the Chlamydomonas-Type Channelrhodopsins”, *J. Phys. Chem. B* **2011**, *115*, 15119–15128.
- [178] M. Hoffmann, M. Wanko, P. Strodel, P. H. König, T. Frauenheim, K. Schulten, W. Thiel, E. Tajkhorshid, M. Elstner, “Color tuning in rhodopsins: The mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II”, *J. Am. Chem. Soc.* **2006**, *128*, 10808–10818.
- [179] L. Spörkel, W. Thiel, “Adaptive time steps in trajectory surface hopping simulations”, *J. Chem. Phys.* **2016**, *144*, 194108.
- [180] O. Valsson, C. Filippi, “Photoisomerization of model retinal chromophores: Insight from quantum monte carlo and multiconfigurational perturbation theory”, *J. Chem. Theory Comput.* **2010**, *6*, 1275–1292.
- [181] I. Conti, M. Garavelli, “Substituent-controlled photoisomerization in retinal chromophore models: Fluorinated and methoxy-substituted protonated Schiff bases”, *J. Photochem. Photobiol. A* **2007**, *190*, 258–273.
- [182] E. N. Laricheva, S. Gozem, S. Rinaldi, F. Melaccio, A. Valentini, M. Olivucci, “Origin of fluorescence in 11-*cis* locked bovine rhodopsin”, *J. Chem. Theory Comput.* **2012**, *8*, 2559–2563.
- [183] M. M. El-Tahawy, A. Nenov, M. Garavelli, “Photoelectrochromism in the Retinal Protonated Schiff Base Chromophore: Photoisomerization Speed and Selectivity under a Homogeneous Electric Field at Different Operational Regimes”, *J. Chem. Theory Comput.* **2016**, *12*, 4460–4475.

Publications

Paper I

Spectral properties and isomerisation path of retinal in C1C2 channelrhodopsin

I. Dokukina, O. Weingart

Phys. Chem. Chem. Phys. **2015**, *17*, 25142-25150



Cite this: *Phys. Chem. Chem. Phys.*,
2015, 17, 25142

Spectral properties and isomerisation path of retinal in C1C2 channelrhodopsin†

I. Dokukina and O. Weingart*

Structure and excited state isomerisation pathway of retinal in the channelrhodopsin chimera C1C2 have been investigated with combined quantum mechanical/molecular mechanical (QM/MM) techniques, applying CD-MS-CASPT2//CASSCF and DFT-MRCI quantum methods. The absorbing S_1 state is of 1B_u -like character, and the second excited S_2 state is dominated by HOMO–LUMO double excitation with small oscillator strength. Upon photoexcitation and torsion along the reactive C13=C14 double bond we observe bond length equalisation and a two-path deactivation mechanism in positive and negative torsion directions. The computed path is barrierless in positive direction while a small barrier exists for the opposite side. Comparative protonation studies suggest a charged glutamate E162 residue, with computed resonance Raman data in valuable agreement with experimental channelrhodopsin-2 data. The two negatively charged counter-ions and a positive lysine residue close to the retinal Schiff base terminus have the largest influence on the chromophore absorption wavelength.

Received 7th May 2015,
Accepted 2nd September 2015

DOI: 10.1039/c5cp02650d

www.rsc.org/pccp

Introduction

Channelrhodopsins (ChRs) mediate phototaxis and photophobic reactions in the green algae *Chlamydomonas reinhardtii* and *Haematococcus pluvialis*.^{1,2} Their primary function is the light induced regulation of cell conductivity through opening or closing ion channels of different size.^{3,4} Unlike G-protein coupled receptors, which involve external agonists to produce a delayed but amplified nerve stimulus,⁵ ChRs provide an immediate response after light excitation. Their fast regulation of cell polarisation and feasible expression in mammalian neurons⁶ make ChRs ideal candidates for the study of basic brain or nerve regulation functions through non-invasive external light control.^{7,8}

The initial step in the photoreaction of ChRs is the light induced isomerisation of its chromophore retinal (Fig. 1), which is bound *via* a Schiff base linkage to a lysine residue of the protein. Subsequent thermal reactions open the ion channel and lead to depolarisation of the cell.^{1–4}

The two known forms of ChR, ChR1 and ChR2, provide highest selectivity for Li^+ , Na^+ and Ca^{2+} cations (in decreasing order⁹). ChR2 is able to depolarise the cell membrane within ms and has therefore been widely used in studying neuronal activity.¹⁰ Recently, the crystal structure of dark-adapted C1C2, a chimera of ChR1 and ChR2, has been reported at a resolution of 2.3 Å, together with a model of ion vestibules which may

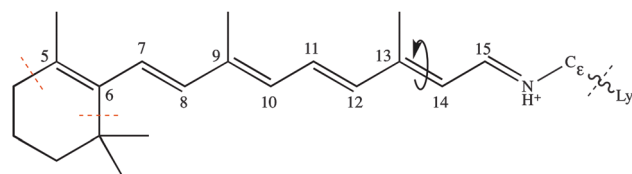


Fig. 1 All-*trans* retinal with isomerising double bond (in negative direction), numbering and QM areas indicated.

form within the protein.¹¹ The studied chimera contains mostly structural features of ChR1, the last two helices were exchanged for those of ChR2. The resulting protein has been successfully expressed in insect cells. Stable photocurrents were measured, *i.e.*, the chimera is fully functional.

The initial photoreaction in ChRs is not yet completely resolved. Resonance Raman and extraction experiments reveal two retinal conformers in dark adapted ChR2: 70% in all-*trans* and 30% in 13-*cis* configuration.¹² Femtosecond time-resolved absorption spectroscopy indicates the formation of a bacteriorhodopsin (bR) K-like ground state intermediate through torsion around the C13=C14 retinal bond.⁹ Since the absorption signal of 13-*cis* retinal is very similar to all-*trans*, participation of the 13-*cis* isomer in the initial step of the photoreaction could not be excluded.^{9,10}

The absorbing state in retinal Schiff bases can be compared to the 1B_u excited state dominated by HOMO–LUMO excitation in symmetric polyenes. Upon excitation, a charge migration from the Schiff base nitrogen terminus towards the β -ionone ring takes place.¹³ The latter process however does not occur in neutral symmetric polyenes, thus the same type of excitation

Institut für Theoretische Chemie und Computerchemie, Heinrich-Heine-Universität
Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany.

E-mail: Oliver.Weingart@hhu.de

† Electronic supplementary information (ESI) available: Excitation energies, detailed resonance Raman data and Cartesian coordinates. See DOI: 10.1039/c5cp02650d



generates different geometrical changes in the two types: while the bond length pattern is inverted in Schiff bases,¹⁴ the bond lengths equalise in polyenes.¹⁵

Currently, two models for the ChR binding pocket are discussed in the literature. Early models propose two negatively charged counter-ions in the vicinity of the retinal Schiff base linkage, E123 and D253 (E162 and D292 in C1C2), while a recent, DFTB//Charmm based computational study of Welke *et al.* gives indications for a protonated E123 residue in ChR1 and C1C2.¹⁶

Computational multi-reference MP2 QM/MM studies on the C1C2 chimera structure performed by Hayashi and coworkers (computed $\lambda_{\text{max}} = 426 \text{ nm}$)¹⁷ and polarisable embedded RI-CC2 studies by Sneskov *et al.*¹⁸ ($\lambda_{\text{max}} = 459 \text{ nm}$) reproduce with good accuracy the experimental absorption of 470 nm. Both studies were based on the double-counter-ion model, but lack information on state composition and other states beyond the spectroscopic bright state associated with ¹B_u-like HOMO–LUMO excitation. This information is of importance to correctly model and interpret transient absorption in fs-UV/Vis and 2DES spectroscopy.¹⁹

Photoreaction path analyses for retinal in C1C2 have not been pursued so far, but for similar systems as *e.g.* the anabaena sensory rhodopsin^{20,21} and bR,²² revealing unidirectional bicycle-pedal like motion²³ of two neighboring double bonds during photoisomerisation.

To investigate the nature of low lying excited states in C1C2, we applied DFT-MRCI/Amber and MS-CASPT2/Amber QM/MM strategies for both, single and double counter-ion binding pocket models. We provide excitation energies for structure models of C1C2 refined at different levels and follow the C13=C14 torsion path starting in the bright ¹B_u-like state. We investigate the spectral characteristics of 13-*cis* retinal in C1C2, and provide computed resonance Raman spectra for both retinal isomers. Finally, we examine the influence of residues on the absorption of retinal.

Computational methods

QM/MM single point computations, geometry optimisations and numerical frequency analysis have been performed using COBRAMM, a QM/MM program interfacing various QM and MM codes.²⁴ The Amber12 suite of programs²⁵ has been used to generate a molecular mechanics setup for ChR utilising the 3UG9 PDB structure, complemented by the SBildn extension to the ff99 parameter set.²⁶ Protonation states of protein residues have been computed with proPka.²⁷ The configurations of backbone histidine residues were screened by visual inspection. Missing hydrogens were added and the hydrogen positions were re-optimised with all heavy atoms fixed. Side chains with missing residues in 3UG9 were saturated with ACE or NME groups to prevent charged termini. QM/MM ground state geometries were obtained within an electrostatic embedding scheme using DFT and CASSCF routines of Molpro2010²⁸ as QM part and Amber12 ff99SBildn as MM method. MM structure parameters were generated with the Antechamber toolset of Amber.

Chromophore ground state charges were derived by fitting the electrostatic potential of the bare chromophore using the Merz–Kollmann Scheme in Gaussian09. To saturate dangling QM-bonds, a hydrogen link-atom scheme with MM charge-repartitioning has been used. The charges of the MM boundary atoms were set to zero and the charge was partitioned proportionally among the neighboring MM atoms. Forces of the link hydrogens were repartitioned among the QM/MM border atoms. For more details on the QM/MM setup, see also ref. 19.

Geometry optimisations

A full CASSCF 12/12 π -space was employed for ground state optimisation, using the 6-31G* basis set for all computations. DFT structures were obtained using the B3LYP functional. The QM layer consisted of the whole retinal chromophore and one CH₂ group of lysine. The retinal, the bound lysine residue, the two counter-ions and one close water molecule (WAT619) were free to move. The rest of the protein structure was kept fixed at the crystal structure values. Further refinement was not pursued at this stage to keep larger structural rearrangements, as *e.g.* observed in the dynamics computations by Hayashi and coworkers,¹⁷ where full equilibration of the system was pursued in solvent, or changes due to the missing loops in our model at a minimum. This strategy will allow to judge, whether the sole crystal structure can already provide valuable information on the photophysics of C1C2. A similar protocol has also been used in the work of Sneskov *et al.*¹⁸

MS-CASPT2 and DFT-MRCI²⁹ energy values for the first five electronic states and corresponding oscillator strengths were computed for the optimised structures using Molcas 8.1³⁰ and Turbomole 6.3, respectively.³¹ For MS-CASPT2 computations the IPEA shift has been set to 0.0, and a standard imaginary shift of 0.2 has been applied.

Relaxed excited state scan

Excited state torsion paths were obtained using the Molcas 8.1 CD-CASSCF routines as QM method with an active space of 12 electrons in 12 π -orbitals. A reduced QM layer, including the whole conjugated part and the C5 methyl group but neglecting the β -ionone alkyl part (C1–C4, see also Fig. 1), was used. The two counter-ions were fixed to their positions after CASSCF S₀-optimisation. The torsion profile computations were performed within a frozen environment, because the ultrafast fs-isomerisation of the chromophore prohibits an immediate energy exchange with the surrounding. Relaxation and thermalisation of the latter will only take place in a later phase, thus additional optimisation of the binding pocket residues may significantly bias the computed retinal torsion path. The wavefunction was averaged over the ground state and the three lowest excited states (state averaging, SA). SA-CASSCF gradients were computed for the third excited CASSCF state (S₃) resembling HOMO–LUMO single excitation. Gradients were corrected using the molcas mclr routine. A relaxed scan along the C12–C13=C14–C15 torsion angle was performed. The torsion angle was varied in steps of 5 degrees and kept fixed, while the rest of the movable part was fully relaxed.



CD-MS-CASPT2 energy values for the four lowest states were computed at converged geometries.

Resonance Raman data

Numerical frequencies were computed to generate resonance Raman data. To reduce the computational effort in producing numerical frequencies, the models were re-optimised at the reduced CD-CASSCF(10/10)//Amberff99SBildn level of theory, deleting the highest unoccupied and lowest occupied orbital from the former CAS(12/12) space. The β -ionone ring was deleted from the QM part and all protein residues were frozen at the crystal structure values.

Raman activities were computed using the approach formulated by Orlandi and coworkers,³² which has proven to yield reliable data for Schiff bases.³³ In the harmonic approximation, the relative displacement parameter B_i for the transition from S_0 to S_1 in mode i can be approximated as the matrix product of the S_1 force vector f^{S_1} with the mass-weighted Cartesian normal coordinates for mode i , $L_i^{S_0}$, scaled by the associated vibrational frequency $\nu_i^{-3/2}$:

$$B_i = (2.41 \times 10^6) f^{S_1} M^{1/2} L_i^{S_0} \nu_i^{-3/2}$$

M is the $3N \times 3N$ dimensional, diagonal matrix of the atomic masses. In resonance with the 0–0 transition, the activity of each totally symmetric mode is proportional to the square of the displacement parameter:

$$I_i \propto \frac{1}{2} B_i^2$$

Results and discussion

Geometry changes after optimisation

After QM/MM optimisation, we note significant geometrical changes from the crystal structure for the orientation of the retinal NH^+ moiety (Fig. 2). In all ground-state optimisations this fragment moves significantly closer towards the D292 counter-ion and away from E162. Leaving also the counter-ions D292 and E162 free to move during optimisation, leads to a small shift of the E162 position, away from the NH^+ moiety. The orientation towards D292 persists also, when the protein binding pocket is relaxed (see ESI†), suggesting D292 as the primary proton acceptor in the modelled structure. This result seems to be in contrast to UV/Vis and FTIR measurements³⁴ which support formation of a salt bridge solely between the Schiff base NH^+ and E162. The recent work of Bartl and coworkers however reports the same protonation characteristics for both counter-ions after illumination,³⁵ such that both initial orientations are plausible. This picture is further supported by reaction path computations in the next chapter. The C1C2 crystal structure consequently appears more biased towards the $\text{NH}^+ \rightarrow \text{D292}$ configuration, an effect which may also result from protonation of E162 under the slightly acidic conditions during the crystallisation procedure. This effect is obviously enhanced in our QM/MM optimisations, and can be released through further refinement and equilibration of the used model. Through optimisation of the

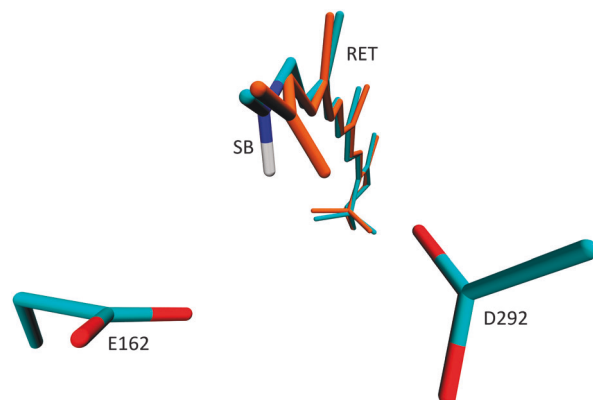


Fig. 2 Location of the two negatively charged counter-ions E162 and D292 and geometry change from crystal structure (cyan) to QM/MM optimised ground state (orange).

residues within 5 Å of retinal and using a lower level HF method for the QM region, we obtained a structure with hydrogen bonding to E162 and very similar absorption properties (see ESI†). It thus appears that both hydrogen bonded configurations may be independently stable.

The conjugated chain is significantly twisted at the C13=C14 double bond (-171°) and the C10=C11 (-169°) and C14=C15 (-163°) single bonds, *i.e.*, the chromophore has overall M-helicity. The twist persists when the geometry is optimised in the excited $^1\text{B}_u$ -like state. At the CASSCF level, this structure is a stationary point (see ESI† for an overlay of the optimised ground state and $^1\text{B}_u$ -like structures).

Fig. 3 shows the bond length alternation (BLA) pattern of single and double bonds along the conjugated chain of retinal computed at different levels of theory in the ground state and in the relaxed CASSCF $^1\text{B}_u$ -like geometry. A strong BLA (varying from *ca.* 1.35 to 1.49 Å along the carbon chain) is observed for the CASSCF-optimised structure, while DFT computes a stronger conjugation, *i.e.* a less alternating single/double bond pattern, along the chain (1.38 to 1.44 Å). No significant changes in the

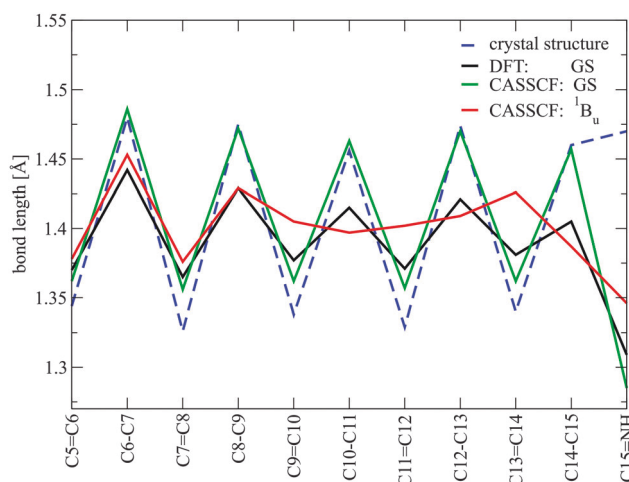


Fig. 3 Bond lengths in crystal structure and QM/MM optimised geometries.



bond pattern can be observed in the single counter-ion structure (see ESI†).

In the relaxed bright 1B_u -like excited state, the bond length pattern changes drastically in the fragment from C9 to C15. Former double bonds become shorter and single bonds longer, with values between 1.38 and 1.42 Å. Thus, the bonds become more equivalent after excitation. This observation is somewhat different from the previously computed retinal chromophore in rhodopsin (Rh), where a complete bond order inversion is observed in the excited S_1 state,¹⁴ but it is in agreement with trends in symmetric carotenoids and polyenes. These molecules show bond order inversion only in the 2A_g state due to HOMO–LUMO double excitation.¹⁵ A similar effect has also been observed in Rh with a locked retinal.³⁶ This process of bond equalisation in ChR can be attributed to the two negatively charged counter-ions in the vicinity of the Schiff base bond, which have profound influence on the nature of the excited states. In the ground state, the counter-ions strongly stabilise the positive charge at the Schiff base tail of the chromophore, leading to a large BLA. The 1B_u -like state associated with HOMO–LUMO excitation leads to a charge transfer towards the β -ionone moiety.^{13,14} A complete transfer of the positive nitrogen charge, which would invert the bond ordering, is hindered by the negatively charged environment around the Schiff base. This destabilises the energy of the excited 1B_u -like state and results in an overall stronger conjugation along the C9–C14 fragment. We suspect that the two negatively charged counter-ions lead to the more “polyene-like” character of retinal in this surrounding. The short distance of the two negatively charged counter-ions is also responsible for the significant blue shift of retinal in ChR with respect to other retinal-binding proteins.

Nature of low lying excited states for different protonation states

Vertical excitation energies computed at MS-CASPT2 and DFT-MRCI levels of theory are collected in Table 1 and Fig. 4 for the CASSCF S_0 -optimised chromophore, counter-ions and nearby water. The corresponding data for CASSCF, HF and DFT optimisations of retinal with frozen protein environment are given in the ESI†. Generally, the additional optimisation of the counter-ion positions leads to a red-shift of the computed maximum absorption wavelength (see ESI†). We note a strong influence on the CASSCF state ordering in the charged environment of the binding pocket, which is only removed at higher levels of correlation in DFT-MRCI and MS-CASPT2. *E.g.*, the bright HOMO–LUMO excitation is the first excited state in the latter two descriptions, while in CASSCF it corresponds to the third excited state. This deficiency of CASSCF, which lacks a major part of dynamic electron correlation, can only partly be corrected at the single state CASPT2 level. Too high S_1 excitation energies are obtained with a single state perturbation strategy (see ESI†). Only the multi-state CASPT2 or DFT-MRCI levels allow for a stronger stabilisation of the S_1 state, with computed excitation energies close to the experimental values.

Both methods, MS-CASPT2 and DFT-MRCI, show a large contribution of the HOMO–LUMO CSF in the first excited state and

Table 1 Excitation energies E and oscillator strengths f for retinal optimised at the CASSCF/Amber S_0 level with both counter-ions and nearby water molecule movable. The experimental value for the maximum absorption is 2.64 eV

Method	State	E (eV)	Config.	Weight	f
MS-CASPT2	S_0	0.0	GS	0.72	—
	S_1	2.59	$H \rightarrow L$	0.54	1.20
	S_2	3.95	$^2H \rightarrow ^2L$	0.23	0.13
	S_3	4.96	$H-1 \rightarrow L$	0.15	0.14
			$H-2 \rightarrow L$	0.17	
			$H-2/H-1 \rightarrow L$	0.17	
	S_4	5.39	$H-3 \rightarrow L$	0.19	0.01
			$H-1 \rightarrow L, H \rightarrow L+1$	0.08	
	S_0	0.0	GS	0.93	—
	S_1	2.74	$H \rightarrow L$	0.82	1.47
DFT-MRCI	S_2	3.52	$H-1 \rightarrow L$	0.37	0.56
	S_3	4.13	$H \rightarrow L+1$	0.10	0.25
			$H-1 \rightarrow L$	0.28	
			$H-2 \rightarrow L$	0.15	
	S_4	4.46	$H-2 \rightarrow L$	0.11	0.08
			$H-2 \rightarrow L$	0.26	
			$H \rightarrow L+1$	0.22	
			$H-1 \rightarrow L$	0.13	

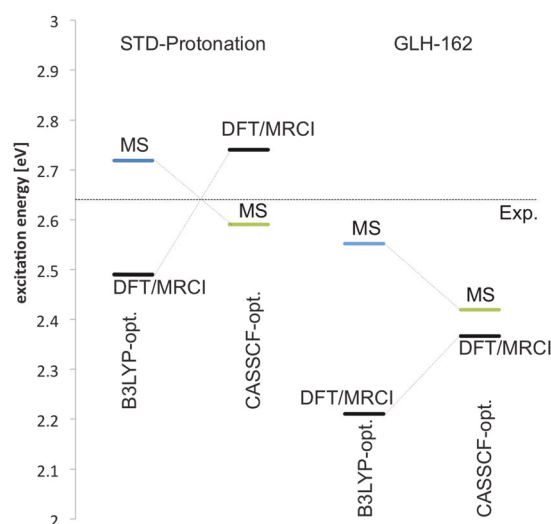


Fig. 4 Excitation energies to the 1B_u -like S_1 state of retinal in ChR with standard protonation and E162 protonated. Values are shown for B3LYP- and CASSCF optimised geometries.

a large oscillatory strength for this state. The second excited state is dominated by the HOMO–LUMO double excitation and HOMO–1 \rightarrow LUMO single excitation in MS-CASPT2. It has a small but non-negligible oscillatory strength. The same is true for the strongly blue-shifted S_3 state, which is of mixed character. The S_4 state is strongly mixed as well and essentially dark in MS-CASPT2. DFT-MRCI yields a slightly different picture for the higher excited states. S_2 , *e.g.*, is more dominated by the HOMO–1 \rightarrow LUMO excitation, a double excitation is missing. This results in a larger oscillatory strength for this state. The maximum absorption band is well reproduced at both levels (479 and 452 nm for MS-CASPT2 and DFT-MRCI, respectively). Higher excited states (S_2 – S_4) appear significantly more stabilised in the DFT-MRCI description.



To investigate the effect of the alternative E162 protonated configuration on absorption and state composition, we have re-optimised the geometry of retinal and counter-ion positions using the protonated GLH residue at the corresponding position with DFT and CASSCF as QM methods. The geometry and the composition of the excited states change only slightly with respect to the previously optimised geometry in standard protonation (see ESI†). For the CASSCF optimised geometry we note a significant red shift in the maximum absorption of *ca.* 0.17 eV (*ca.* 30 nm) in MS-CASPT2 and more than 0.37 eV (70 nm) at the DFT-MRCI level (Fig. 4). The deviation in absorption energy persists, when we use a DFT-optimised geometry to determine the excited state properties. Here, the difference between MS-CASPT2 and DFT-MRCI computed absorption values becomes even more pronounced; the 1B_u state is strongly stabilised in the latter.

Properties after HOMO–LUMO single excitation

After bond relaxation in the 1B_u -like state, the properties of S_1 and S_2 slightly change. At the optimised 1B_u -like geometry the emission wavelength from the S_1 state is 489 nm, with an oscillator strength of 1.69. The HOMO–LUMO excitation contributes *ca.* 50% to this state (see ESI†), while some fraction of the same CSF can now be found in S_2 (*ca.* 8%, osc. strength of this state 0.091) – together with the HOMO–LUMO double excitation (*ca.* 24%). At this point, the S_1 and S_2 states are very close in energy, the MS-CASPT2 energy difference is only 7.3 kcal mol^{−1}. Thus, state interaction which may lead to population of the S_2 state during dynamics, similar to the proposed three-state model by Anfinrud and coworkers,³⁷ cannot be excluded. The computed torsion path however follows the 1B_u -like excited state. The other excited states are well separated and essentially dark. Only S_3 has significant oscillator strength (305 nm, 0.12).

Torsion path in the 1B_u -like state

The torsion path was obtained by following the CASSCF S_3 state, which until *ca.* 120° along the C12–C13=C14–C15 dihedral angle keeps its HOMO–LUMO single excitation character (see ESI†). Fig. 5 shows the energy profiles of the four lowest retinal electronic states along negative and positive C13=C14 bond twists, computed at the CD-MS-CASPT2/CASSCF(12,12) level. The paths start at the optimised 1B_u -like excited state structure with a torsion angle of −172°. The positive pathway leads, nearly barrierless, towards an S_0/S_1 state crossing (CI₊) with *ca.* −100° twist. The negative twisting pathway has a *ca.* 3.8 kcal mol^{−1} barrier at 155°. This barrier is slightly higher than the available excess energy after excitation (*i.e.* the difference between the FC energy and the bond-length relaxed geometry, *ca.* 3.6 kcal mol^{−1}). Thus, the molecule will need a small amount of additional thermal energy to overcome the barrier in the negative torsion direction. After the barrier, the pathway becomes steeper and finally results, at a C13=C14 torsion value of *ca.* 110°, in an S_0/S_1 state crossing (CI_−). Both optimised pathways are extremely flat, the difference in potential energy between the −172° S_1 starting point and CI₊ is *ca.* 8.8 kcal mol^{−1}, even much smaller for the CI_− point (*ca.* 1.25 kcal mol^{−1}). During torsion,

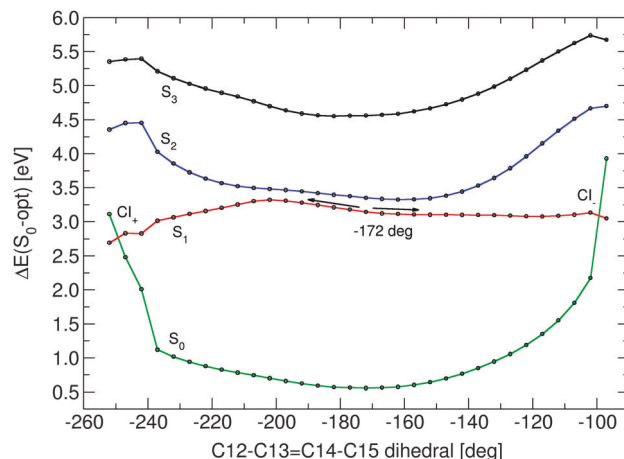


Fig. 5 State energy profiles for C13=C14 torsion in 1B_u -like excited state. The path starts at *ca.* −172° twist in negative (left side of the graph) and positive directions (right side of the graph). The energy difference to the optimised CASSCF S_0 geometry is given in eV.

the nature of the first excited state smoothly changes from 45% HOMO–LUMO single excitation character up to 55% for both directions. Along both torsion paths, the bond length alternation increases. Shortly before the crossing, the HOMO–LUMO single excitation contribution drops, and the ground state configuration slowly but steadily increases in weight. At *ca.* −100 and +110° twist, we note a faster change in state character. Now the ground state configuration dominates the upper state, and the HOMO–LUMO excitation the lower state – a clear indication that a surface crossing point has been passed between this and the last point. It must be noted that the corresponding CASSCF-profiles show a barrier of *ca.* 3 kcal mol^{−1} on both sides in the corresponding CAS- S_3 state, and all states are well separated at this level without any state crossing.

After excitation and bond relaxation, the β -ionone-part carries a charge of *ca.* +0.6 electrons, which is effectively transported to the Schiff base moiety during isomerisation (see ESI†).

During isomerisation in either direction, the whole fragment from C14 to the nitrogen Schiff base moves as one rigid part, *i.e.* the observed mechanism is somewhat different from the bicycle-pedal motion reported for other retinal proteins,^{21–23,38} where two double bonds rotate in different directions (Fig. 6). The reaction route shows much more similarities to the hula-twist mechanism suggested by Liu *et al.*,³⁹ where torsion takes place around one single and one double bond. The corresponding single bond is the N–CH₂ bond, which twists by up to 90°. The C=N double bond rotates only by *ca.* 20° in positive direction and 40° for the negative path. In the negative path, the configuration of the N–CH₂ single bond changes from *syn* towards *anti*, while the configuration is kept in the positive twisting direction.

In the negative torsion path, the hydrogen bonding to D292 is conserved at the crossing, while in the positive path the hydrogen bond is released, and the Schiff base hydrogen orients towards a nearby serine (S295, see ESI†).



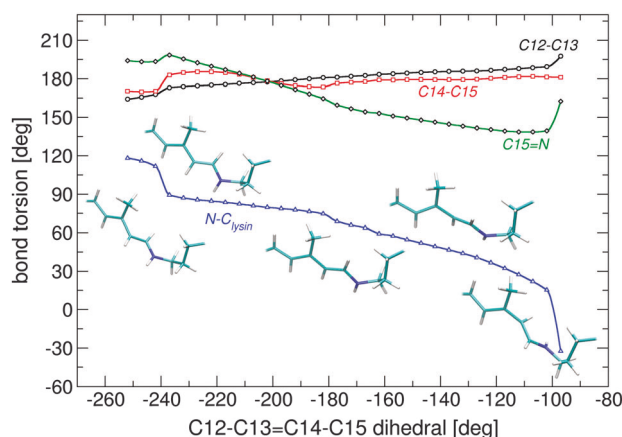


Fig. 6 Evolution of torsion angles during simulated S_1 torsion along $C13=C14$ bond.

Properties of 13-*cis* retinal in C1C2

To obtain the 13-*cis* retinal configuration in C1C2 we started a CASSCF single state optimisation in the ground state using as a starting point the *ca.* -120° twisted structure from the above mentioned torsion path. For comparability to the optimised all-*trans* structure, this optimisation was also carried out with both counter-ions movable. Fig. 7 shows an overlay of the 13-*cis* and all-*trans* optimised S_0 structures. Significant changes are apparent along the conjugated chain and the orientation of the Schiff base nitrogen, which instead of pointing towards the ASP292 counter-ion, now has a weak hydrogen bond with SER295 (2.43 Å for the latter compared to 1.68 Å in all-*trans* and ASP292). Due to the 13-*cis* configuration this chromophore is slightly shorter than all-*trans*, which leads to a small clockwise tilt in the β -ionone ring, continuing along the chain. The C13 methyl group is strongly tilted, and the chromophore occupies slightly more volume in the fragment from C12 to lysine. This leads to enhanced van der Waals and electrostatic interactions, destabilizing the ground state QM/MM energy by *ca.* 31 kcal mol $^{-1}$ for 13-*cis* at the MS-CASPT2 level. Interestingly, the chromophore changes its helicity from P to M during our simulated all-*trans* to 13-*cis* conversion. Its state compositions are very similar to the previously discussed all-*trans* retinal (see ESI †), with only 12 nm red shift. Thus, the two isomers will be difficult to distinguish in sole UV/Vis measurements. Better chances are provided by spectroscopic techniques considering the differences in vibrational modes in

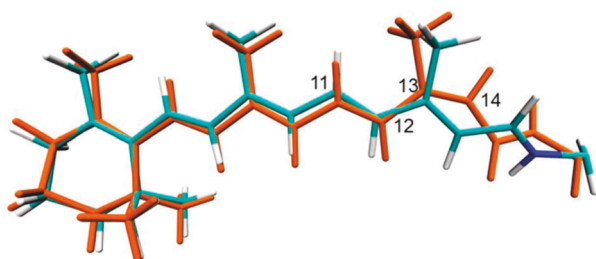


Fig. 7 Overlay of CASSCF optimised ground state structures of all-*trans* retinal (cyan) and 13-*cis* retinal (orange) in C1C2.

the two molecules in combination with UV/Vis excitation, *i.e.* resonance Raman spectroscopy, which we attempt to simulate in the following.

Resonance Raman spectra of all-*trans* and 13-*cis* C1C2

The procedure of CAS space reduction and QM/MM frontier shift (see Computational section) lead to three spurious low imaginary frequencies in both models. Further optimisation along these modes did not lead to structure relaxation and energy reduction. For generation of resonance Raman data, these modes were ignored. The obtained vibrational frequencies were scaled to match the highest intensity band in the experimental spectrum, defined by C=C in plane stretching of the polyene fragment (1551 cm $^{-1}$). The applied scaling factor was 0.877. Fig. 8 shows the computed resonance Raman spectra for all-*trans* retinal (black line) and 13-*cis* retinal (blue line) in comparison with the experimental spectrum obtained by Nack *et al.*¹²

The computed C1C2 spectrum shows good overall agreement with the experimental ChR2 spectrum, indicating distinct similarities in the geometrical and electrostatic environments of the binding pockets in C1C2 and ChR2. For both chromophores, the gradient of the 1B_u -like state (see ESI †) shows major components for bond length changes after excitation. The highest intensity band in our computed spectra is therefore produced by the double bond stretching modes of the C9–C14 fragment. In all-*trans* retinal this band consists of two separate close lying lines, describing C13=C14 stretching (1551 cm $^{-1}$) and combined C9=C10 and C11=C12 stretching (1548 cm $^{-1}$). The 13-*cis* spectrum shows slightly shifted and more distinct bands for these modes (1560 and 1578 cm $^{-1}$, respectively).

The two peaks left from the highest absorption band are coupled C15=N stretching modes, which also appear in 13-*cis* retinal but significantly shifted. The in-plane bending modes of the vinyl hydrogens, which have been assigned to 1270 cm $^{-1}$ by Nack *et al.*, appear as lower intensity band at 1319 cm $^{-1}$ in our spectrum. The 1240 cm $^{-1}$ peak is produced by H11 in-plane bending. The two bands at 1193 cm $^{-1}$ and 1178 cm $^{-1}$ which are

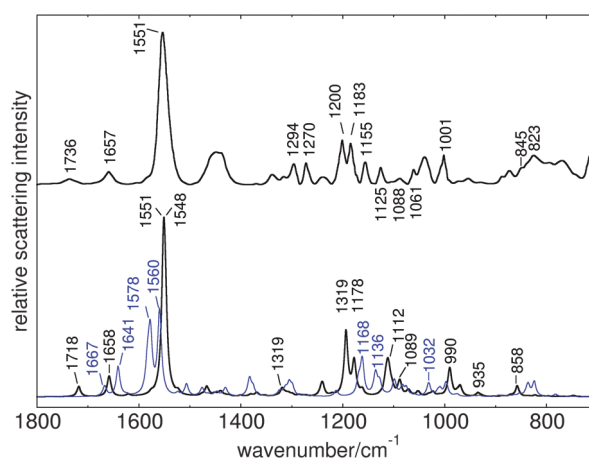


Fig. 8 Computed resonance Raman spectra of all-*trans* retinal (black line) and 13-*cis* retinal (blue line) in C1C2, in comparison to experimental ChR2 spectrum of Nack *et al.* (data reproduced from Fig. 1 of ref. 12).



very close to the characteristic all-*trans* and 13-*cis* assigned bands of Nack *et al.*, are due to single H14 in-plane bending and combined vinyl hydrogen in-plane bend including H10–H12. We suspect that in a combined 13-*cis* all-*trans* spectrum, the latter peak may be enhanced by combination with the shifted H10–H12 in-plane band at 1168 cm⁻¹, because no separate band in 13-*cis* has enough intensity to produce a single peak at 1183 cm⁻¹ when a 70:30 mixture of the two isomers is assumed.

Single bond stretchings, combined with CH₂ twisting motions, are observed at 1089 cm⁻¹ and around 1112 cm⁻¹ in all-*trans*, and 1136 cm⁻¹ in 13-*cis*. The gauche stretching bands in the latter appear at 1032 cm⁻¹ with low intensity, again very close to the experimental assignment. Also the Me rocking band at 1001 cm⁻¹ is well reproduced (990 cm⁻¹). HOOP motion is observed at 858 cm⁻¹ (mainly H14) and with very low intensity at 935 cm⁻¹ (symmetric H11–H12 HOOP). The two small peaks at the right end of the 13-*cis* spectrum are due to H12 in-plane bending.

The calculated frequencies also allow the computation of the IR difference spectrum between the two species. This spectrum, which is in valuable agreement with the recent findings of Furutani *et al.*,⁴⁰ is reported in the ESI†.

Influence of the electrostatic environment on retinal absorption

The contribution of selected protein residues to the retinal maximum absorption was investigated by computing their absorption differences to the uncharged protein pocket. In the choice of potentially contributing amino acids we follow the suggestions of Kamiya *et al.*, who provided a list of 12 charged and polarizable residues in the vicinity of the retinal chromophore as potential candidates for absorption-regulating mutations.¹⁷ These are: E122, E129, K132, C167, H173, D195, G220, F228, S295, N297 and finally the counter-ions E162 and D292 (see ESI† for the location of these amino acids in the retinal binding pocket).

The charges of all named amino acids were initially set to zero. The computed excitation energy of this configuration is used as a reference value for the uncharged environment. To study the effect of one single amino acid, we insert the charges of the specific residue, compute the absorption with MS-CASPT2 and build the difference from the reference value. As before, the SS-CASPT2 computed values largely overestimate the contribution of the charged residues.

The computed values in eV are summarised in Fig. 9. On the right-hand side the overall effect of the remaining protein backbone (compared to the retinal vacuum absorption) becomes apparent. Its contribution is a slight blue shift with a value <0.15 eV. Only K132, situated below the Schiff base terminus, contributes, in sum with the protein backbone, a significant red shift in comparison to the vacuum absorption. All other amino acids only slightly modulate the blue shift induced by the remaining protein residues. The strongest blue shift contribution results from the two counter-ions E162 (0.260 eV) and D292 (0.261 eV), followed by S295 (0.065 eV). It should be noted that the contribution of both counter-ions results in a higher value (0.728 eV) than the sum of their single contributions.

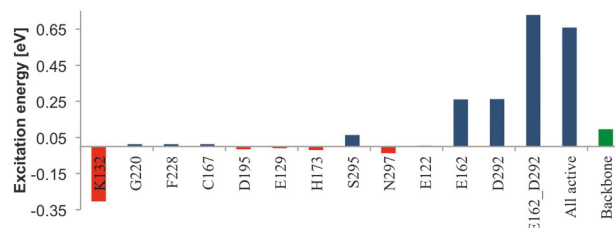


Fig. 9 Electrostatic influence of protein residues suggested by Kamiya *et al.*¹⁷ on S₀–S₁ excitation energy of retinal in C1C2. The green bar indicates the shift with respect to vacuum, resulting from the remaining protein backbone.

Conclusions

Static aspects and reaction path of the retinal photoconversion in C1C2 have been investigated with QM/MM methods and were compared to experimental findings. A crucial point in the generation of valid data is the choice of the quantum mechanical treatment to describe the chromophore in its electrostatic environment. Geometry optimisations of the chromophore in the protein binding pocket with HF, DFT and CASSCF methods lead to a slightly different orientation of the NH⁺ Schiff base moiety with respect to the crystal structure. In all models, this fragment moves towards the suggested primary counter-ion D292 and forms a strong hydrogen bond with the latter, an effect which might be balanced by the initial configuration of the crystal structure and the restrictions applied in our simulations. We could however show, that relaxation of the binding pocket does not destabilise this configuration, and we were able to obtain a stable structure with hydrogen bonding to E162 through a different optimisation strategy (see ESI†). The latter shows a similar absorption value and state composition, such that, from a computational point of view, both initial configurations appear plausible.

The optimisation with DFT and CASSCF methods furthermore leads to a strong difference in bond length alternation of the chromophore. Similar to observations made by Larsson and coworkers for retinal *in vacuo*,⁴¹ DFT tends to overestimate the conjugation along the chain, while CASSCF yields a more polyenic structure with larger BLA. The level of conjugation however strongly affects the maximum absorption of the chromophore.⁴² In any case, for an adequate description of excitation energies, especially with charged residues in the vicinity of the chromophore, a multi-state multi-reference treatment is inevitable. The best agreement with the experimental 470 nm value was obtained by using an MS-CASPT2 strategy on top of the CASSCF(12,12)//Amber optimised chromophore, employing simultaneous optimisation of both counter-ion positions (479 nm). Single state CASPT2 largely overestimates the influence of the nearby charged residues and consequently yields too high excitation energies (see ESI†). DFT-MRCI yields reasonable results with the CASSCF geometry (452 nm), but underestimates the absorption with a DFT-optimised structure (498 nm). The missing polarisation contribution in our study may slightly alter these values, a lowering of the excitation energy by *ca.* 5 kcal mol⁻¹



has *e.g.* been reported in the study of Sneskov *et al.* due to polarisation effects. Consideration of excitonic coupling effects and charge transfer, as *e.g.* applied in the works of Hasegawa *et al.*⁴³ and Wanko *et al.*,⁴⁴ may further improve the obtained values. Recently, it has been reported that MS-CASPT2 may overestimate the state energy splitting near conical intersections.⁴⁵ Application of more elaborate XMS-CASPT2 or XMCQDPT2 strategies may thus lead to more accurate PESs near the S_1/S_0 state crossing in the studied system.

The alternative protonation of the E162 residue yields in all computed models a too red-shifted absorption value. From this point of view, it does not significantly improve the initially suggested two-counter-ion model, suggesting that E162 should be deprotonated. A final conclusion on this issue can however not be drawn solely based on the comparison of the absorption values in our models.

The first excited state is associated with 1B_u -like HOMO–LUMO excitation with large oscillator strength. The higher S_2 and S_3 states also exhibit small but non-negligible contributions in the corresponding part of the absorption spectrum. In MS-CASPT2, the second excited state is similar to the polyenic 2A_g state, with significant contribution of HOMO \rightarrow LUMO double and HOMO–1 \rightarrow LUMO single excitations. Higher excited states are of strongly mixed character.

At the Franck–Condon (FC) point, the S_1 and S_2 states are well separated. This situation changes dramatically after bond length relaxation in S_1 . The energy difference drops to $7.3 \text{ kcal mol}^{-1}$, thus, further interaction of these states in the photoconversion dynamics cannot be consequently excluded. Bond relaxation in S_1 leads to equalisation of the bond lengths along the conjugated chain, very much like in the 1B_u state of symmetric carotenoids. This is a consequence of the nearby charges, which hinder the relocation of the positive charge at the Schiff base terminus upon excitation, thus resembling the situation in the neutral symmetric polyenes where no large charge migration takes place. As a consequence, the retinal chromophore has a more “polyenic” character in C1C2.

Different from reaction pathways in bRh,²¹ Rh¹⁴ and anabaena sensory Rh,²² which are strictly unidirectional, the computed CD-MS-CASPT2//CASSCF profile along the C13=C14 torsion suggests a bidirectional pathway in the conversion of C1C2. The negative torsion path in C1C2 has a small barrier of *ca.* 3 kcal mol^{-1} which, however, can easily be overcome with thermal excess energy (*i.e.* vibrational energy) after FC excitation. The location of the barrier may switch to the other side of the torsion potential, when retinal establishes a salt bridge with D292. With these properties, the reaction dynamics of retinal in C1C2 may share more similarities to the photoprocess observed in iso-rhodopsin as in bRh.⁴⁶ Reaction rates, the distribution in the proposed reaction channels and resulting photoproducts can, however, only be obtained through molecular dynamics computations, which due to the state order problems at the CASSCF level with electrostatic embedding will form a major challenge.

In the 13-*cis* isomer, which was generated from the computed positive torsion path, the hydrogen bonding to the assumed primary proton acceptor D292 is released, and the Schiff base

proton unexpectedly points toward a neighboring serine. During the ultrafast photoisomerisation, the protein environment will not participate, therefore we performed the optimisation within a frozen environment. Proton release however is reported to take place within 700 ns, which is enough time for the protein to reorient and offer a new partner for deprotonation of the Schiff base.

Computational C1C2 resonance Raman spectra indicate significant differences for all-*trans* and 13-*cis* chromophores. The computed all-*trans* retinal spectrum reproduces location and intensity of distinct bands in ChR2, indicating that, despite their differences in protein sequence, the steric and electrostatic environments which influence the excited state properties of retinal are not too different in C1C2 and ChR2.

Finally, electrostatic screening of the nearby residues reveals 3 sites with major influence on the retinal absorption: the two counter-ions E162 and D292 are mainly responsible for the strong blue shift, while a lysine K132 positioned close to the Schiff base terminus induces a major red shift. The contribution of other residues is only minor.

Acknowledgements

We thank Prof. Christel Marian (University of Düsseldorf) and Prof. Marco Garavelli (ENS Lyon) for valuable discussions. Financial support was provided through “Strategischer Forschungsfonds” of the Heinrich-Heine-University Düsseldorf, project no. F 2013 – 442-9. Computational support and infrastructure was provided by the “Centre for Information and Media Technology” (ZIM) at the University of Düsseldorf (Germany).

Notes and references

- 1 H. Harz and P. Hegemann, *Nature*, 1991, **351**, 489.
- 2 S. Ehlenbeck, G. Gradmann, F.-J. Braun and P. Hegemann, *Biophys. J.*, 2001, **82**, 740.
- 3 G. Nagel, T. Szellas, W. Huhn, S. Kateriya, N. Adeishvili, P. Berthold, D. Ollig, P. Hegemann and E. Bamberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 13940.
- 4 G. Nagel, D. Ollig, M. Fuhrmann, S. Kateriya, S. Musti, E. Bamberg and P. Hegemann, *Science*, 2002, **296**, 2395.
- 5 R. R. Birge, *Annu. Rev. Phys. Chem.*, 1990, **41**, 683.
- 6 E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel and K. Deisseroth, *Nat. Neurosci.*, 2005, **8**, 1263.
- 7 X. Li, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 17816.
- 8 A. Bi, J. Cui, Y.-P. Ma, E. Olshevskaya, M. Pu, A. M. Dizhoor and Z.-H. Pan, *Neuron*, 2006, **50**, 23.
- 9 M.-K. Verhoeven, C. Bamann, R. Blöcher, U. Förster, E. Bamberg and J. Wachtveitl, *ChemPhysChem*, 2010, **11**, 3113.
- 10 P. Hegemann, W. Gärtner and W. Uhl, *Biophys. J.*, 1991, **60**, 1477.
- 11 H. Kato, F. Zhang, O. Yizhar, C. Ramakrishnan, T. Nishizawa, K. Hirata and J. Ito, *et al.*, *Nature*, 2012, **482**, 369.
- 12 M. Nack, I. Radu, C. Bamann, E. Bamberg and J. Heberle, *FEBS Lett.*, 2009, **583**, 3676.



- 13 N. Ferré and M. Olivucci, *J. Am. Chem. Soc.*, 2008, **125**, 6868–6869.
- 14 (a) V. Bonačić-Koutecký, K. Schöffel and J. Michl, *Theor. Chim. Acta*, 1987, **72**, 459–474; (b) M. Garavelli, F. Bernardi, M. Olivucci, T. Vreven, S. Klein, P. Celani and M. A. Robb, *Faraday Discuss.*, 1998, **110**, 51–70; (c) G. Tomasello, G. Olaso-González, P. Altoè, M. Stenta, L. Serrano-Andrés, M. Merchán, G. Orlandi, A. Bottoni and M. Garavelli, *J. Am. Chem. Soc.*, 2009, **131**, 5172–5186.
- 15 (a) P. Tavan and K. Schulten, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1987, **36**, 4337–4358; (b) S. Knecht, C. M. Marian, J. Kongsted and B. Menucci, *J. Phys. Chem. B*, 2013, **117**, 13808; (c) C. M. Marian and N. Gilka, *J. Chem. Theory Comput.*, 2008, **4**, 1501.
- 16 K. Welke, H. C. Watanabe, T. Wolter, M. Gaus and M. Elstner, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6651.
- 17 M. Kamiya, H. E. Kato, R. Ishitani, O. Nureki and S. Hayashi, *Chem. Phys. Lett.*, 2013, **556**, 266.
- 18 K. Sneskov, J. M. H. Olsen, T. Schwabe, C. Hättig, O. Christiansen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2013, **15**, 7567.
- 19 D. Polli, I. Rivalta, A. Nenov, O. Weingart, M. Garavelli and G. Cerullo, *Photochem. Photobiol. Sci.*, 2015, **14**, 213.
- 20 I. Schapiro and S. Ruhman, *Biochim. Biophys. Acta, Bioenerg.*, 2014, **1837**, 589–597.
- 21 A. Strambi, B. Durbeek, N. Ferré and M. Olivucci, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 21322–21326.
- 22 P. Altoè, A. Cembran, M. Olivucci and M. Garavelli, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 20172–20177.
- 23 A. Warshel, *Nature*, 1976, **260**, 679–683.
- 24 P. Altoè, M. Stenta, A. Bottoni and M. Garavelli, *Theor. Chem. Acc.*, 2007, **118**, 219.
- 25 D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, *Amber 12*, 2012.
- 26 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950.
- 27 M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525.
- 28 H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang and A. Wolf, *Molpro 2010*, 2010.
- 29 S. Grimme and M. Waletzke, *J. Chem. Phys.*, 1999, **111**, 5645.
- 30 G. Karlström, R. Lindh, P.-Å. Malmqvist, B. O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrády and L. Seijo, *MOLCAS: a program package for computational chemistry*, *Comput. Mater. Sci.*, 2003, **28**, 222.
- 31 TURBOMOLE V6.3 2010, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.
- 32 A. ten Wolde, H. J. C. Jacobs, F. Langkilde, K. Bajdor, R. Wilbrandt, F. Negri, F. Zerbetto and G. Orlandi, *J. Phys. Chem.*, 1994, **98**, 9437.
- 33 M. Garavelli, F. Negri and M. Olivucci, *J. Am. Chem. Soc.*, 1999, **121**, 1023.
- 34 S. Ito, H. E. Kato, R. Taniguchi, T. Iwata, O. Nureki and H. Kandori, *J. Am. Chem. Soc.*, 2014, **136**, 3475.
- 35 J. Kuhne, K. Eisenhauer, E. Ritter, P. Hegemann, K. Gerwert and F. Bartl, *Angew. Chem., Int. Ed.*, 2015, **54**, 4953–4957.
- 36 E. N. Laricheva, S. Gozem, S. Rinaldi, F. Melaccio, A. Valentini and M. Olivucci, *J. Chem. Theory Comput.*, 2012, **8**(8), 2559–2563.
- 37 K. C. Hasson, F. Gai and P. A. Anfinrud, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 15124.
- 38 D. Polli, P. Altoè, O. Weingart, K. M. Spillane, C. Manzoni, D. Brida, G. Tomasello, G. Orlandi, P. Kukura, R. A. Mathies, M. Garavelli and G. Cerullo, *Nature*, 2010, **467**, 440–443.
- 39 R. Liu, *Acc. Chem. Res.*, 2001, **34**, 555.
- 40 A. Inaguma, H. Tsukamoto, H. E. Kato, T. Kimura, T. Ishizuka, S. Oisho, H. Yawo, O. Nureki and Y. Furutani, *J. Biol. Chem.*, 2015, **290**, 11623–11634.
- 41 F. Blomgren and S. Larsson, *J. Comput. Chem.*, 2005, **26**, 738–742.
- 42 P. D. Patel and A. E. Masunov, *J. Phys. Chem. A*, 2009, **113**, 8409.
- 43 J. Hasegawa, *Chem. Phys. Lett.*, 2013, **571**, 77–81.
- 44 M. Wanko, M. Hoffmann, T. Frauenheim and M. Elstner, *J. Phys. Chem. B*, 2008, **112**, 11462–11467.
- 45 S. Gozem, M. Huntress, I. Schapiro, R. Lindh, A. Granovsky, C. Angeli and M. Olivucci, *J. Chem. Theory Comput.*, 2012, **8**, 4069–4080.
- 46 D. Polli, O. Weingart, D. Brida, E. Poli, M. Maiuri, K. M. Spillane, A. Bottoni, P. Kukura, R. A. Mathies, G. Cerullo and M. Garavelli, *Angew. Chem., Int. Ed.*, 2014, **53**, 2504.



SUPPLEMENTARY INFORMATION

Spectral Properties and Isomerisation Path of Retinal in C1C2 ChannelRhodopsin

Irina Dokukina and Oliver Weingart*

Geometry changes after optimisation

Figure S1 shows an overlay of the optimised ground state geometry (cyan) and 1B_u -like structure (orange). The polyene chain twist observed in the ground state at the C13=C14 double bond persists after the optimisation in the excited 1B_u -like state.

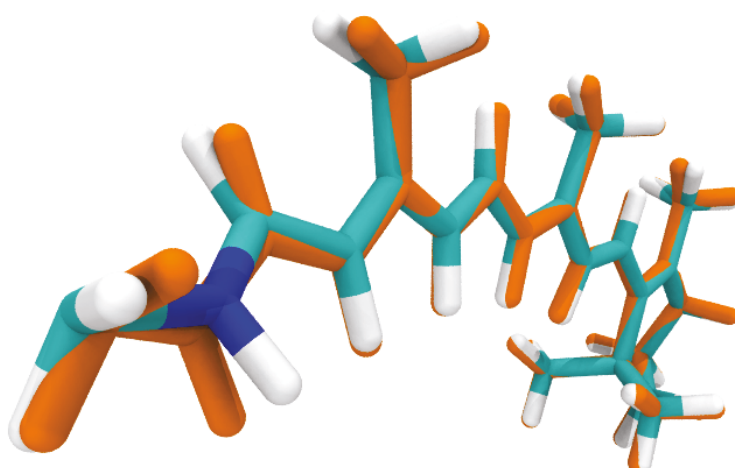


Figure S1: Overlay of S_0 -CASSCF (cyan) and 1B_u -CASSCF (orange) optimised geometries

Figure S2 shows bond lengths along polyene chain of retinal in the CASSCF optimised ground state geometries in standard (GLU) protonation with both counterions charged and alternative protonation (GLH) with protonated GLU162 residue. No significant changes can be noted for the bond pattern in alternative protonation.

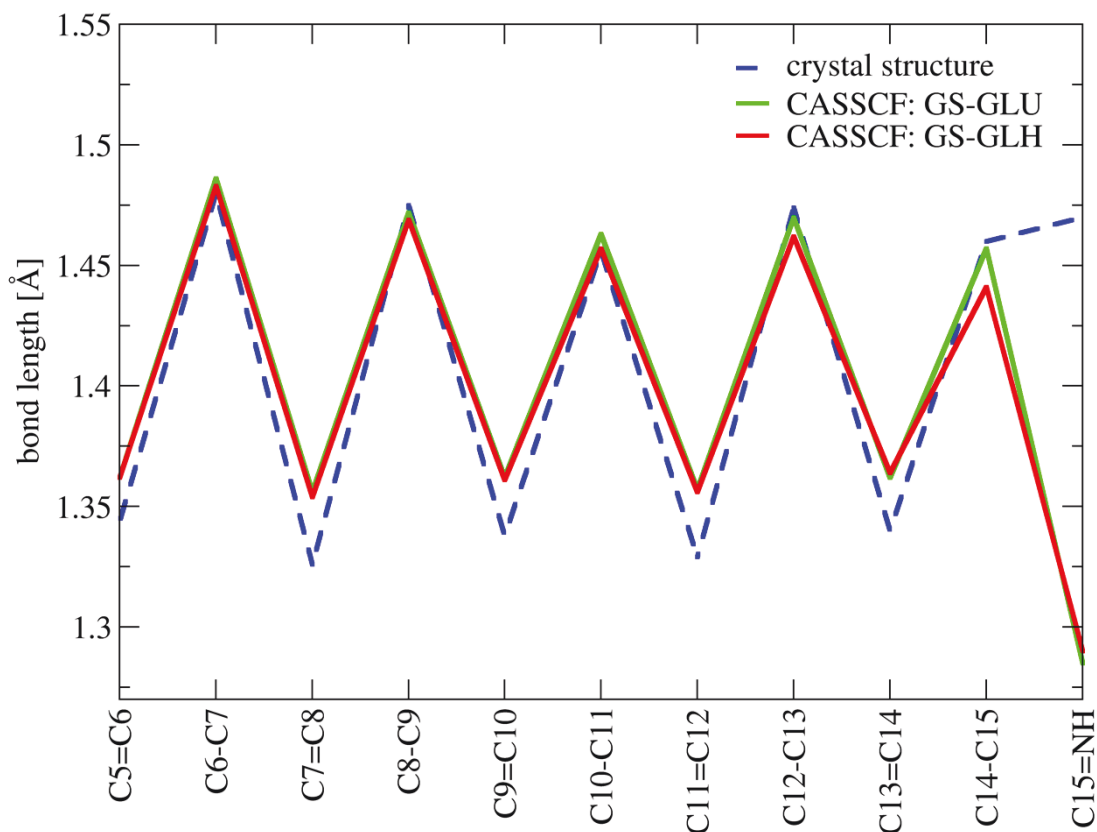


Figure S2: Bond lengths in crystal structure and CASSCF-optimised ground state geometries in standard (GLU) and alternative (GLH) protonation states of counterions.

Nature of low lying excited states for different protonation states

Standard Protonation

Tables S1-S3 show MS-CASPT2 and DFT/MRCI data for HF, DFT and CASSCF optimisations of retinal within frozen protein environment. For all geometries the MS-CASPT2 and DFT/MRCI S_0 - S_1 excitation energy is blue shifted relative to structures with additionally optimised counter ions. Only the DFT/MRCI maximum absorption wavelength for the DFT geometry (see Table S2) is red shifted.

Table S1: Excitation energies and oscillator strengths for retinal optimised at the HF/Amber S_0 level with frozen protein environment. The experimental value for the maximum absorption is 2.64 eV.

Method	State	E(eV)	Config.	Weight	f
MS-CASPT2	S_0	0	GS	0.74	-
	S_1	2.6759	$H \rightarrow L$	0.59	1.22
	S_2	4.2249	$^2H \rightarrow ^2L$	0.25	0.10
			$H-1 \rightarrow L$	0.16	
	S_3	5.2520	$H-2 \rightarrow L$	0.19	0.94×10^{-1}
			$H-1/H \rightarrow L$	0.16	
	S_4	5.6173	$H-3 \rightarrow L$	0.20	0.39×10^{-3}
DFT/MRCI	S_0	0	GS	0.94	-
	S_1	2.9415	$H \rightarrow L$	0.83	1.54
	S_2	3.7681	$H-1 \rightarrow L$	0.39	0.54
			$H \rightarrow L+1$	0.11	
	S_3	4.3511	$H \rightarrow L+1$	0.31	0.26
			$H-1 \rightarrow L$	0.18	
			$H-2 \rightarrow L$	0.10	
	S_4	4.6581	$H-2 \rightarrow L$	0.29	0.067
			$H \rightarrow L+1$	0.22	

Table S2: Excitation energies and oscillator strengths for retinal optimised at the DFT-B3LYP/Amber S_0 level with frozen protein environment. The experimental value for the maximum absorption is 2.64 eV.

Method	State	E(eV)	Config.	Weight	f
MS-CASPT2	S_0	0	GS	0.72	-
	S_1	2.7866	$H \rightarrow L$	0.37	1.12
			$H-1 \rightarrow L$	0.12	
	S_2	3.2989	$H \rightarrow L$	0.23	0.55
			$^2H \rightarrow ^2L$	0.17	
			$H-1 \rightarrow L$	0.12	
	S_3	4.1606	$H-2 \rightarrow L$	0.18	0.11
			$H-1/H \rightarrow L$	0.17	
	S_4	5.0274	$H-3 \rightarrow L$	0.20	0.59×10^{-2}
DFT/MRCI	S_0	0	GS	0.91	-
	S_1	2.5496	$H \rightarrow L$	0.62	1.21
	S_2	2.8883	$H \rightarrow L$	0.24	1.06
			$H-1 \rightarrow L$	0.19	
			$^2H \rightarrow ^2L$	0.15	
	S_3	3.6634	$H-2 \rightarrow L$	0.25	0.023
			$H-1, H \rightarrow L$	0.16	
			$H \rightarrow L+2$	0.11	
	S_4	4.1987	$H-1 \rightarrow L$	0.38	0.12
			$H \rightarrow L+1$	0.36	

Table S3: Excitation energies and oscillator strengths for retinal optimised at the CASSCF/Amber S_0 level with frozen protein environment. The experimental value for the maximum absorption is 2.64 eV.

Method	State	E(eV)	Config.	Weight	F
MS-CASPT2	S_0	0	GS	0.72	-
	S_1	2.7467	$H \rightarrow L$	0.54	1.16
	S_2	3.9286	$^2H \rightarrow ^2L$	0.23	0.14
			$H-1 \rightarrow L$	0.15	
			$H-2 \rightarrow L$	0.17	
	S_3	4.8719	$H-1/H \rightarrow L$	0.17	0.17
DFT/MRCI	S_4	5.3833	$H-3 \rightarrow L$	0.19	0.33×10^{-2}
	S_0	0	GS	0.93	-
	S_1	2.8271	$H \rightarrow L$	0.81	1.47
	S_2	3.5398	$H-1 \rightarrow L$	0.34	0.60
			$H \rightarrow L+1$	0.12	
			$H \rightarrow L+1$	0.25	
			$H-1 \rightarrow L$	0.15	
	S_4	4.4816	$H-2 \rightarrow L$	0.11	0.075
			$H \rightarrow L+1$	0.25	
			$H-2 \rightarrow L$	0.24	
			$H-1 \rightarrow L$	0.14	

Alternative Protonation

Additionally to standard protonation, a model with a protonated counter ion (E162) was created. After ground state optimisation the retinal geometry (Figure S3, orange) does not change with respect to the optimised geometry in standard protonation (Figure S3, cyan).

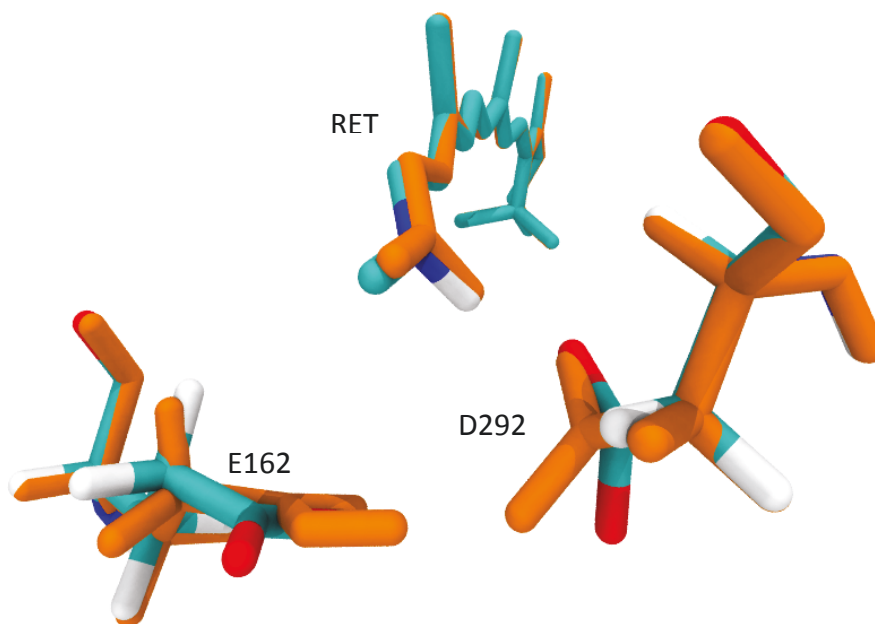


Figure S3: Overlay of S_0 /CASSCF optimised retinal structure in standard protonation (cyan) and S_0 /CASSCF optimised retinal structure in alternative protonation (orange).

The composition of the excited states (Tables S4-S5) changes only slightly with respect to the optimised geometry in standard protonation. The S_0 - S_1 excitation energy is strongly red shifted for both DFT and CASSCF optimised geometries.

Table S4: Excitation energies and oscillator strengths for retinal in alternative protonation optimised at the DFT-B3LYP/Amber S_0 level with counter ions movable. The experimental value for the maximum absorption is 2.64 eV.

Method	State	E(eV)	Config.	Weight	f
MS-CASPT2	S_0	0	GS	0.75	-
	S_1	2.5517	$H \rightarrow L$	0.39	1.65
			$H-1 \rightarrow L+1$	0.14	
	S_2	3.2347	${}^2H \rightarrow {}^2L$	0.21	0.62
			$1H \rightarrow L$	0.21	
			$H-1 \rightarrow L$	0.14	
	S_3	4.0339	$H-2 \rightarrow L$	0.21	0.65×10^{-2}
			$H-1/H \rightarrow L$	0.16	
			$H \rightarrow L/L+1$	0.11	
			$H \rightarrow L+2$	0.10	
	S_4	4.9848	$H-3 \rightarrow L$	0.24	0.44×10^{-2}
			$H-2/H \rightarrow L$	0.11	
DFT/MRCI	S_0	0	GS	0.91	-
	S_1	2.2105	$H \rightarrow L$	0.82	1.95
	S_2	2.6893	$H-1 \rightarrow L$	0.27	0.33
			${}^2H \rightarrow {}^2L$	0.24	
			$H \rightarrow L+1$	0.11	
	S_3	3.5372	$H-2 \rightarrow L$	0.27	0.0035
			$H-1, H \rightarrow L$	0.17	
	S_4	3.9227	$H-1 \rightarrow L$	0.39	0.078
			$H \rightarrow L+1$	0.18	

Table S5: Excitation energies and oscillator strengths for retinal in alternative protonation optimised at the CASSCF/Amber S_0 level with counter ions movable. The experimental value for the maximum absorption is 2.64 eV.

Method	State	E(eV)	Config.	Weight	f
MS-CASPT2	S_0	0	GS	0.75	-
	S_1	2.4197	$1H \rightarrow L$	0.53	1.54
			$H-1 \rightarrow L+1$	0.11	
	S_2	4.0700	${}^2H \rightarrow {}^2L$	0.28	0.22
			$H-1 \rightarrow L$	0.16	
	S_3	4.6387	$H-2 \rightarrow L$	0.18	0.17×10^{-2}
			$H-1/H \rightarrow L$	0.14	
			$H \rightarrow L/L+1$	0.11	
			$H \rightarrow L+2$	0.10	
	S_4	5.3165	$H-3 \rightarrow L$	0.22	0.2×10^{-2}
DFT/MRCI	S_0	0	GS	0.94	-
	S_1	2.3668	$H \rightarrow L$	0.84	1.54
	S_2	3.3872	$H-1 \rightarrow L$	0.52	0.46
	S_3		$H \rightarrow L+1$	0.31	0.25
		3.9782	$H-1 \rightarrow L$	0.14	
			$H-2 \rightarrow L$	0.13	
	S_4	4.3835	$H-2 \rightarrow L$	0.25	0.055
			$H \rightarrow L+1$	0.11	

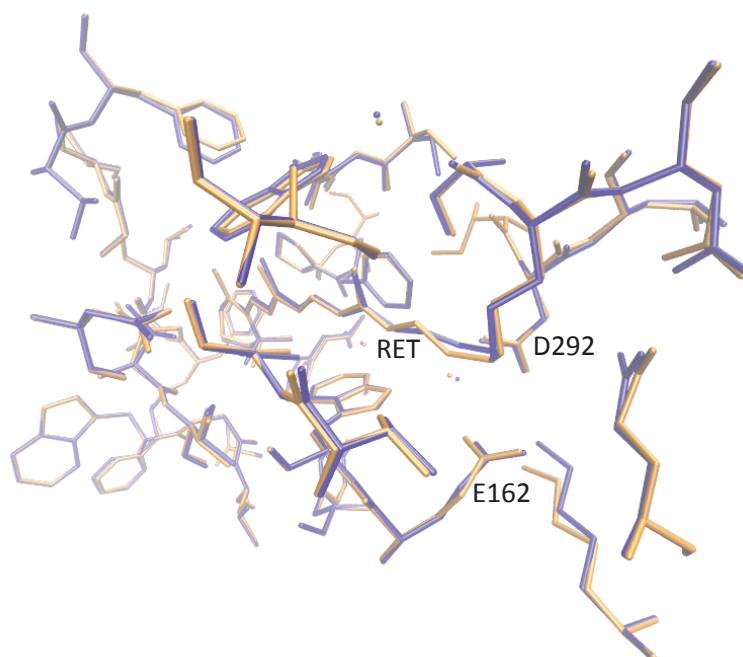


Figure S4: Crystal structure (orange) and structure after optimisation with movable pocket residues (blue)

Optimisation with movable binding pocket

The CASSCF optimized C1C2 geometry was re-optimized, but now with all residues movable within a distance of 5 Angstrom of the retinal chromophore. Figure S4 shows an overlay of the optimised structure with the crystal structure. We note only minor changes for the positions of the pocket residues. The orientation of the NH^+ group of retinal still points towards D292 after optimisation. Excitation energies computed at the MS-CASPT2 level of theory are reported in Table S6.

Table S6: MS-CASPT2 excitation energies and oscillator strengths for retinal optimised at the CASSCF/Amber S_0 level with pocket residues movable.

State	E(eV)	Config.	Weight	f
S_0	0	GS	0.73	-
S_1	2.5696	$\text{H} \rightarrow \text{L}$	0.55	1.18
S_2	3.9468	${}^2\text{H} \rightarrow {}^2\text{L}$	0.24	0.11
		$\text{H-1} \rightarrow \text{L}$	0.16	
		$\text{H} \rightarrow \text{L+1}$	0.09	
		$\text{H} \rightarrow \text{L}$	0.08	
		$\text{H-2} \rightarrow \text{L}$	0.18	
S_3	4.9890	$\text{H-1}/\text{H} \rightarrow \text{L}$	0.16	0.12
		$\text{H} \rightarrow \text{L}/\text{L+1}$	0.08	
		$\text{H} \rightarrow \text{L+2}$	0.08	
		$\text{H-3} \rightarrow \text{L}$	0.19	
S_4	5.4072	$\text{H-3} \rightarrow \text{L}$	0.19	0.75×10^{-3}

H \rightarrow L+3 0.09**Salt bridge to E162**

A structure with the retinal NH⁺ proton forming a salt bridge with the E162 counter ion was obtained by the following strategy: the missing LOOPS in the crystal structure were modelled with the program Modeller. The resulting structure was then optimised with all residues within a 5 Angstrom distance of retinal movable, using HF/6-31G* as QM method. Figure S5 shows an overlay of this structure and the corresponding CASSCF-optimised structure with mobile binding pocket (without LOOP refinement, see Fig. S4). Larger re-arrangements of the mobile residues with respect to the crystal structure are apparent in this structure. MS-CASPT2 computed absorption data is shown in Table S7. A similar absorption wavelength is obtained as for the D292 bound geometry.

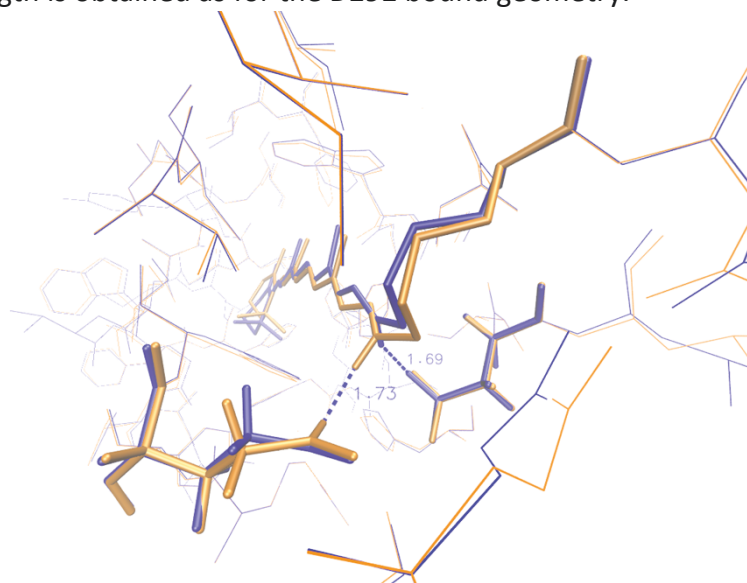


Figure S5: Overlay of CASSCF optimized structure forming a salt bridge to D292 (blue) and HF-optimized structure with salt bridge to E162 (orange). NH-O distances in Å.

Table S7: MS-CASPT2 excitation energies and oscillator strengths for C1C2 optimised at the HF/Amber S_0 level with pocket residues movable, with retinal forming a salt bridge to E162.

State	E(eV)	Config.	Weight	f
S_0	0	GS	0.75	-
S_1	2.6212	H \rightarrow L	0.60	1.27
S_2	4.2362	2 H \rightarrow 2 L	0.26	0.10
		H-1 \rightarrow L	0.16	
		H \rightarrow L+1	0.09	
S_3	5.2799	H-2 \rightarrow L	0.18	0.02
		H-1/H \rightarrow L	0.14	
		H \rightarrow L/L+1	0.10	
		H \rightarrow L+2	0.09	
S_4	5.6301	H-3 \rightarrow L	0.20	0.14×10^{-2}

H→ L+3 0.09

Properties after 1B_u excitation

Table S8 shows MS-CASPT2 data for 1B_u CASSCF optimisation of retinal with counter ions and water molecule movable. The bright ($f=1.69$) S_1 state is defined by ca. 50% HOMO-LUMO excitation with an emission wavelength of 489 nm. The dark S_2 state is mostly defined by HOMO-LUMO double excitation and lies close to S_1 state. Other excited states are well separated and only S_3 has significant oscillator strength.

Table S8: MS-CASPT2 excitation energies and oscillator strengths for retinal optimised at the CASSCF/Amber S_1 level with counter ions and water molecule movable.

State	E(eV)	Config.	Weight	f
S_0	0	GS	0.73	-
S_1	2.5347	H→L	0.50	1.69
S_2	2.8498	$^2H\rightarrow^2L$	0.24	0.91×10^{-1}
		H-1→L	0.16	
		H→L+1	0.11	
		H→L	0.08	
S_3	4.0638	H-2→L	0.16	0.12
		H→L/l+1	0.12	
		H-1/H→L	0.11	
		H→L+2	0.11	
S_4	4.8236	H-3→L	0.13	0.72×10^{-2}
		H-2→H/L	0.11	

Properties of 13-*cis* retinal in C1C2

MS-CASPT2 and DFT/MRCI calculations were performed (Table S9) to compare 13-*cis* and all-*trans* retinal ground state structures. Their state compositions are very similar. S_0 - S_1 excitation energy is red shifted by 12 nm at the MS-CASPT2 level and 17 nm at DFT/MRCI level. After optimisation, the salt bridge to D292 is released, and the retinal Schiff base NH^+ proton points towards the neighboring serine 295 residue (see Figure S6).

Table S9: Excitation energies and oscillator strengths for 13-*cis* retinal optimised at the CASSCF/Amber S_0 level with counter ions movable.

Method	State	E(eV)	Config.	Weight	f
MS-CASPT2	S_0	0	GS	0.74	-
	S_1	2.5348	$H \rightarrow L$	0.53	1.38
			$H \rightarrow L+1$	0.11	
	S_2	3.9097	$^2H \rightarrow ^2L$	0.26	0.13
			$H-1 \rightarrow L$	0.16	
	S_3	4.9140	$H-2 \rightarrow L$	0.18	0.17×10^{-4}
			$H/H-1 \rightarrow L$	0.14	
			$H \rightarrow L/L+1$	0.11	
			$H \rightarrow L+2$	0.11	
	S_4	5.3784	$H-3 \rightarrow L$	0.21	0.36×10^{-2}
DFT/MRCI	S_0	0	GS	0.93	0
	S_1	2.6414	$H \rightarrow L$	0.81	1.48
	S_2	3.4106	$H-1 \rightarrow L$	0.37	0.47
			$^2H \rightarrow ^2L$	0.11	
			$H \rightarrow L+1$	0.10	
	S_3	3.9743	$H \rightarrow L+1$	0.29	0.27
			$H-1 \rightarrow L$	0.17	
	S_4	4.3553	$H-2 \rightarrow L$	0.28	0.05
			$H \rightarrow L+1$	0.19	
			$H-1 \rightarrow L$	0.11	

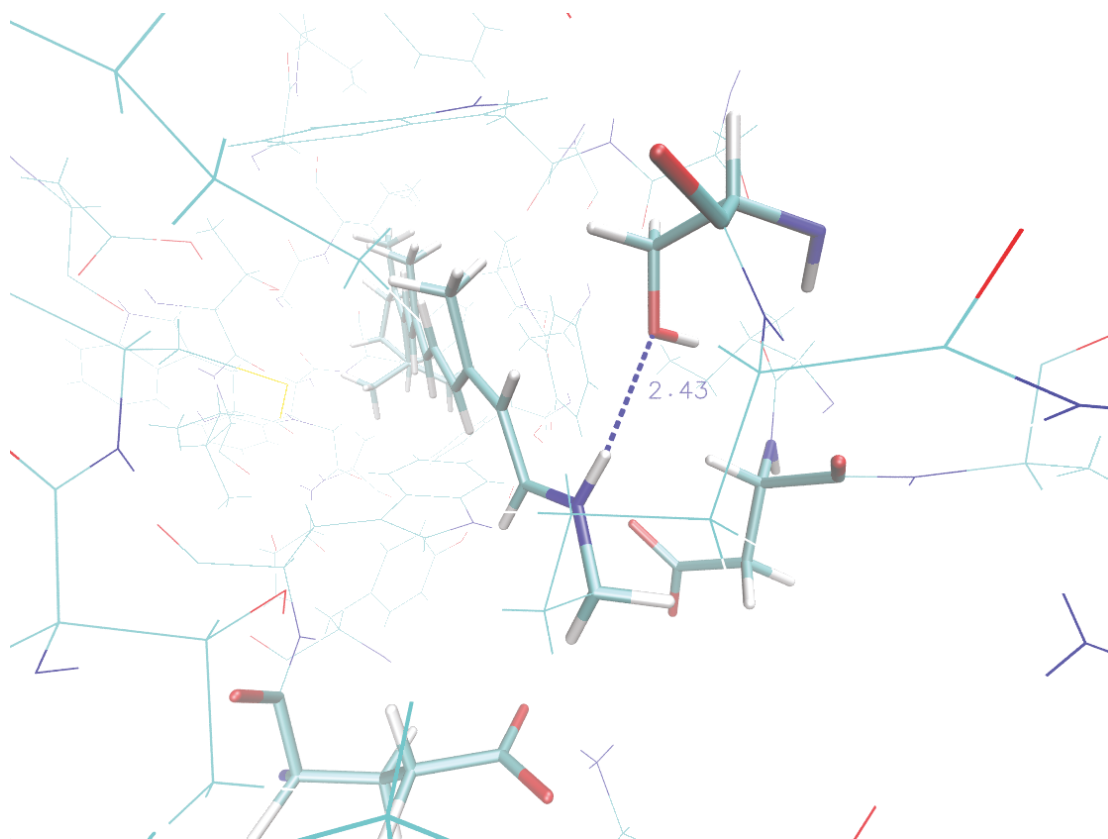


Figure S6: Structure of CASSCF optimised 13-*cis* retinal in C1C2 binding pocket, with the Schiff base NH⁺ moiety oriented towards S295. The NH-O distance is given in Å.

Effect of bond length alternation on retinal absorption in vacuo and in C1C2

The charged protein environment in C1C2 leads to a different effect of bond length alternation on the excited state absorption. This is also documented in Figure 5 of the main text, where the MS-CASPT2 computed absorption of a DFT optimised structure with less bond alternation shifts to higher energies compared to the corresponding CASSCF structure. Usually, reduced bond length alternation will lead to a red shift in absorption. The underlying process is discussed in the following.

Different from uncharged polyenes, HOMO-LUMO excitation in protonated Schiff bases relocates part of the positive charge situated at the N-terminus towards the β -ionone ring. This becomes apparent in Figure S7, where a charge of ca. +0.4 electrons migrates from the right to the left half of the chromophore after excitation. In vacuum, the β -ionone part carries a positive charge after excitation to the absorbing S_1 state. The charge transport is of similar magnitude in DFT and CASSCF computed structures. The influence of bond length alternation on the MS-CASPT2 vacuum state energies is shown in Figure S8. In the DFT optimized structure both states decrease in energy, but the excited state somewhat more than the ground state, as the electronic structure of a strongly conjugated system is

more similar to the relaxed S_1 geometry with inverted bond length alternation. This leads to the known red-shift in absorption wavelength when conjugation is enhanced.

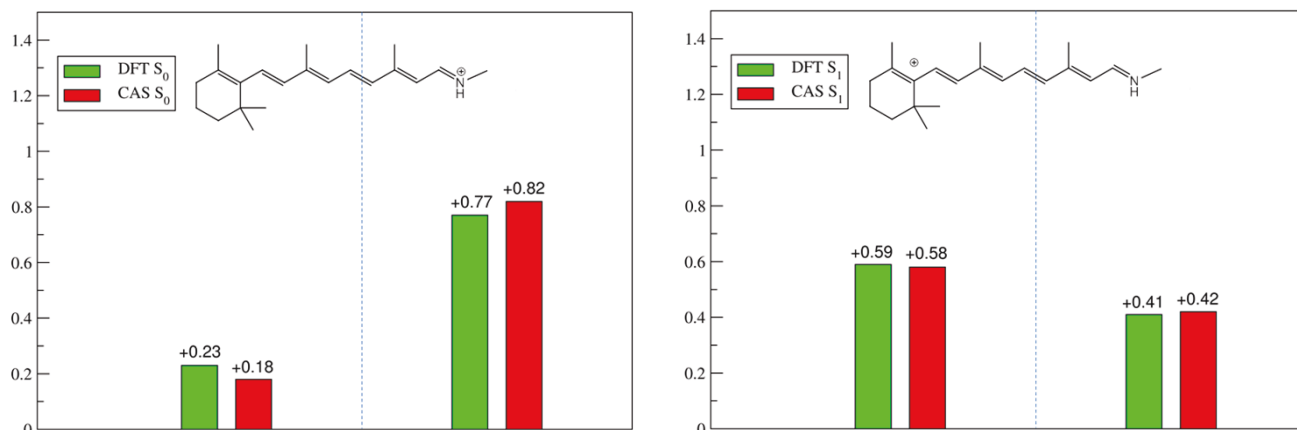


Figure S7: Charges of chromophore halves in ground and HOMO-LUMO excited state for DFT and CASSCF optimized C1C2 retinal structures in vacuo.

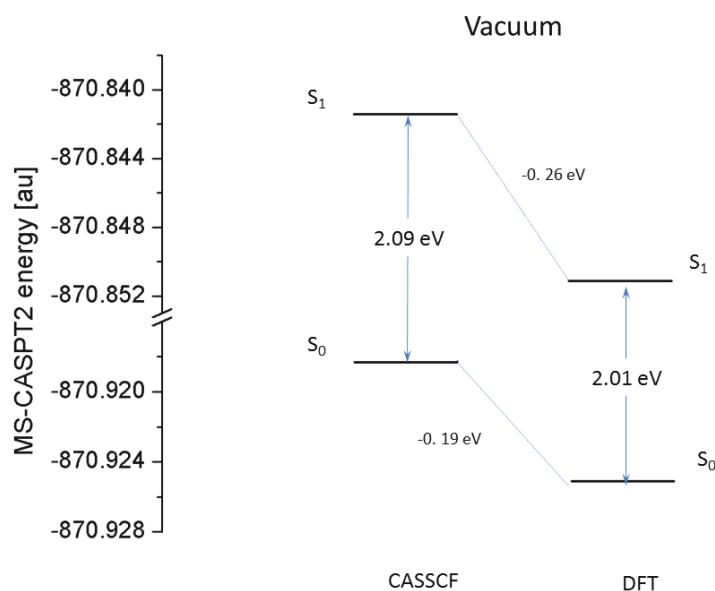


Figure S8: MS-CASPT2 energies of retinal ground state and first excited state in CASSCF and DFT-optimised C1C2 structures without surrounding.

In the protein environment, relocation of the positive charge is hindered by the negatively charged counterions near the N-terminus of retinal (Figure S9). In the ground state, the positive charge is almost entirely located near the N-terminus. An alternating structure is preferred within the charged environment, the conjugated DFT structure is slightly destabilised (Figure S10). Upon HOMO-LUMO excitation, only a small amount of charge shifts towards the β -ionone ring compared to vacuum. In the conjugated DFT structure, we note that slightly more charge (ca +0.1 electron) is transported against the electrostatic field as in the alternating CASSCF structure. Together with energy splitting due to state mixing (see next chapter) this effect may contribute to the destabilisation of the excited state energy in the DFT structure (Figure S10) and to the observed blue shift in absorption wavelength.

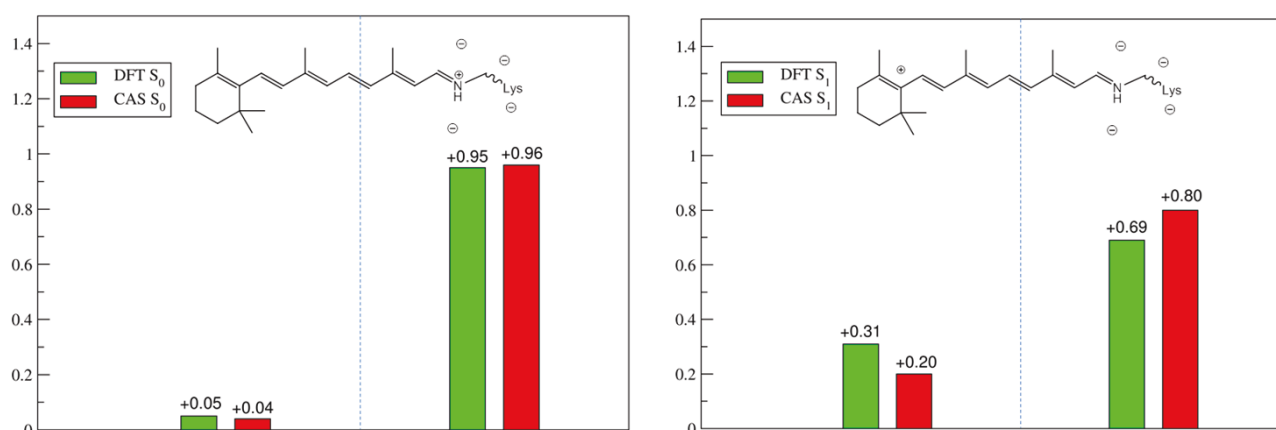


Figure S9: Charges of chromophore halves in ground and HOMO-LUMO excited state for DFT and CASSCF optimized structures in protein environment.

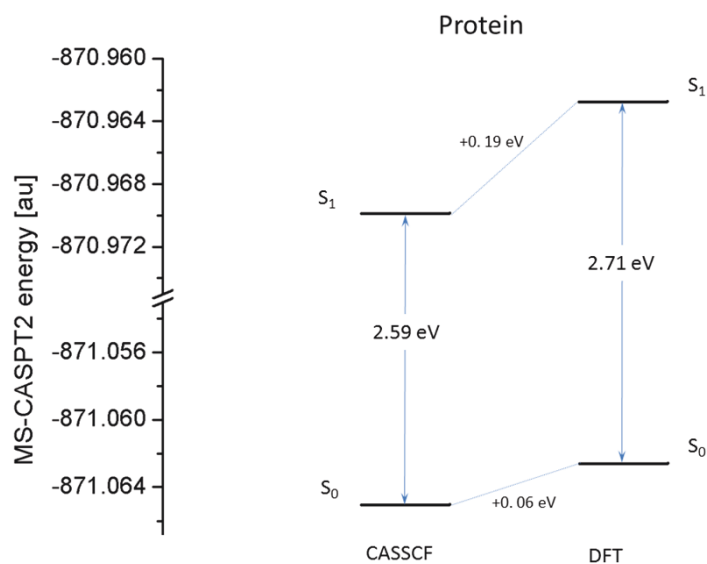


Figure S10: MS-CASPT2 energies of retinal ground state and first excited state in CASSCF and DFT-optimised C1C2 structures in the protein environment.

SS-CASPT2 vs. MS-CASPT2

Figure S11 compares the excitation energies in C1C2 retinal obtained by single and multistate CASPT2 computations. The single state energies strongly deviate from the MS-CASPT2 description. In single state, the HOMO-LUMO excitation to the state is strongly blue-shifted, and it appears in the S_2 state. The lowest state in this description is the 2A_g -like state associated with HOMO-LUMO double excitation, S_3 and S_4 states are stabilised. If the states are allowed to interact in the multistate protocol, S_1 becomes the bright HOMO-LUMO excited state with good agreement to the experimental value. S_2 is upshifted, and also the oscillator strengths of this and higher states differ from the single state values. Table S10 shows the MS-CASPT2 coupling matrix, describing the contribution of the corresponding single state wavefunctions in the multistate treatment. As apparent from the off-diagonal elements, strong mixing occurs in comparison to the same values computed for the retinal chromophore without protein surrounding (Table S11), especially for the S_1 , S_2 and S_3 states. The situation does not significantly change, when the nearby protein residues are allowed to relax. The diagonal elements in the vacuum computation indicate much more similarities in single and multistate treatment and thus only minor contribution due to state mixing.

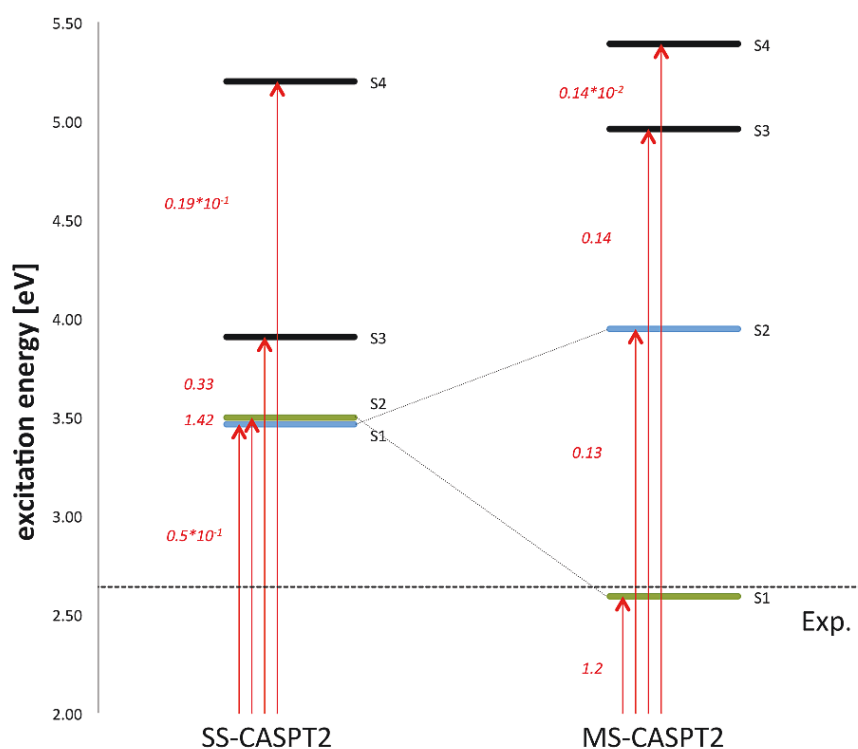


Figure S11: Excitation energies and oscillator strengths of retinal in C1C2 at single state and multistate CASPT2 level.

Table S10: SS-CASPT2 coupling coefficients for CASSCF-optimised retinal structure computed in protein environment

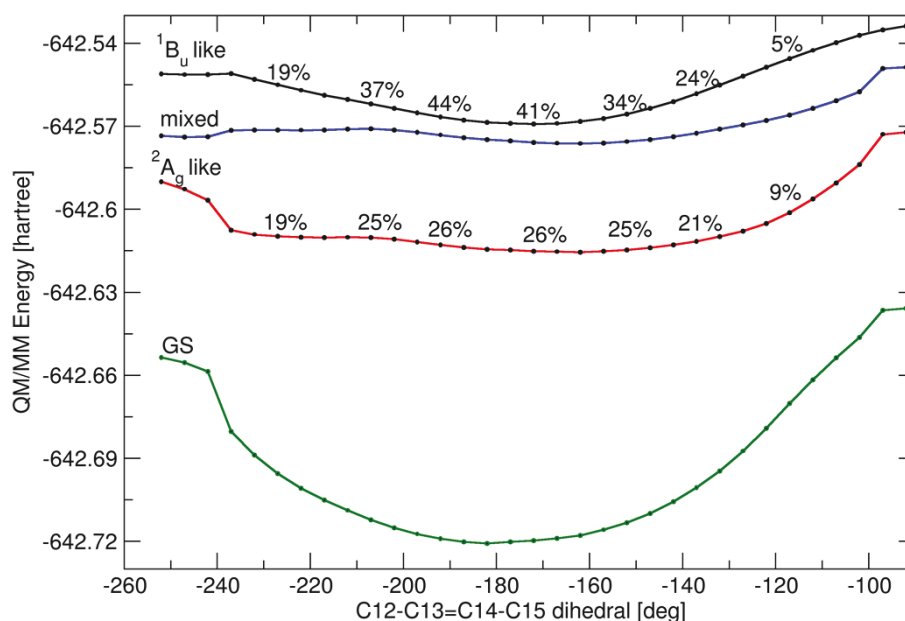
State	1	2	3	4	5
1	0.96875851	0.24296495	0.03527090	0.03434745	0.00715471
2	0.08306823	-0.46878480	0.87637558	0.07282017	0.00188344
3	-0.12107917	0.53785097	0.36161432	-0.75002472	-0.05248017
4	0.19915028	-0.65321348	-0.31464418	-0.64169374	-0.15125973
5	0.01690000	-0.07236937	-0.03091807	-0.13859851	0.98707223

Table S11: SS-CASPT2 coupling coefficients for CASSCF-optimized retinal structure computed in vacuum

State	1	2	3	4	5
1	0.97906369	-0.18127462	-0.08849572	-0.02360742	-0.01360126
2	-0.18485031	-0.96193594	-0.09829880	0.14541878	-0.09849024
3	0.07885328	-0.02755976	0.89388707	0.31081453	-0.31206225
4	-0.02841510	-0.14291553	0.25731834	-0.93620336	-0.18994280
5	0.01547226	-0.14364667	0.34243105	-0.07220566	0.92555581

CASSCF potential energy surfaces at CASSCF level

The CASSCF energy profiles for the S_0 - S_4 states along the computed torsion path in the S_3 state is shown in Figure S12. Also indicated is the percentage of the leading configuration for S_3 (HOMO-LUMO single excitation) and S_1 (HOMO-LUMO double excitation). The S_2 state is highly mixed at the CASSCF level. The S_3 state was used for geometry optimisation.

**Figure S12:** Computed CASSCF energy profiles for torsion in S_3 along the C13=C14 bond.

Charge translocation along 1B_u -like isomerisation paths

Figure S13 provides CASPT2 computed charges along the isomerisation paths of the chromophore fragments right and left of the isomerising C13=C14 double bond. Notice that this is a different fragmentation as the one shown in figure S9, where for illustration the cut was placed among the C11=C12 bond. The S_0 and S_1 values for the FC geometry are also given in table S12. The Figure shows that the positive charge, which resides in the β -ionone fragment after excitation, is effectively transported back towards the Schiff base tail along both isomerisation coordinates. In the negative path, the β -ionone fragment becomes initially more positive, in-line with the occurrence of a barrier in this direction.

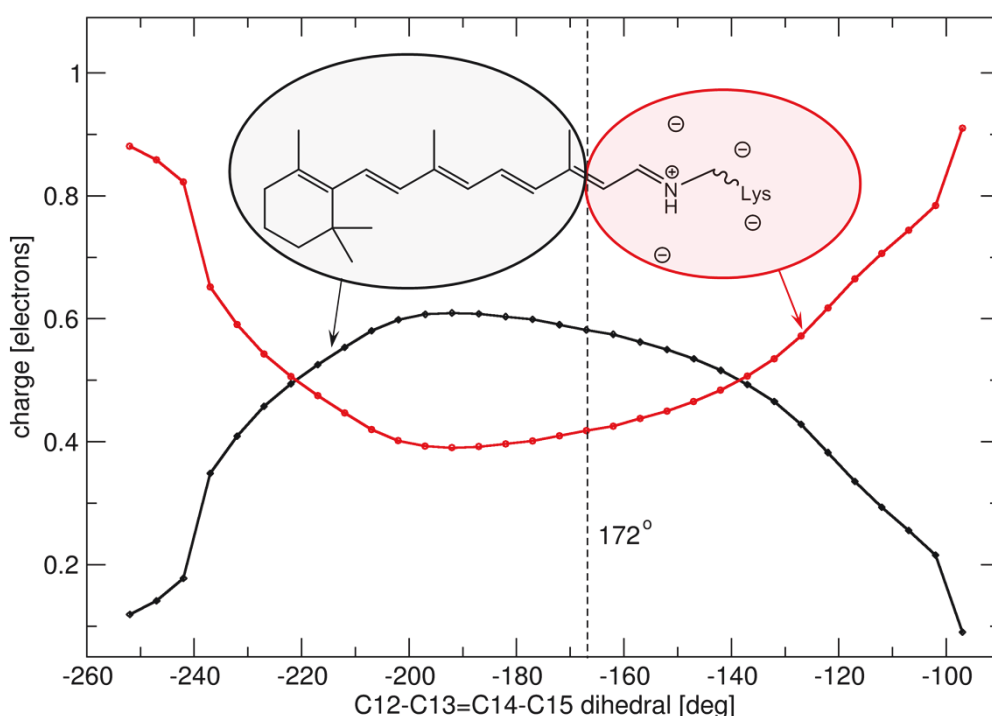


Figure S13: Charge of chromophore fragments along computed isomerisation pathways. The vertical line denotes the start of the computed path at -172° twist.

Table S12: Charges of chromophore fragments in S_0 and S_1 states at the FC point and bond-relaxed 1B_u -like structure (starting point in Figure S13).

Structure	FC		1B_u -like	
State	S_0	S_1	S_0	S_1
β -ionone fragment [e]	+0.227	+0.499	+0.274	+0.590
Schiff-base fragment [e]	+0.773	+0.501	+0.726	+0.410

Resonance Raman spectra of all-*trans* and 13-*cis* retinal in C1C2

Assigned resonance Raman modes and frequencies are listed in the Table S12-S13.

Table S13 Frequencies and mode assignments for optimised all-*trans*-retinal in C1C2 (Intensity > 0.05)

Freq. [cm ⁻¹]	Scaled (* 0.877)	Intensity	Mode
852	747	0.14	H2O wagging
908	797	0.07	H11-12 OOP asym.
920	807	0.06	NH asym. OOP
938	823	0.14	H12-H14 ip. bending, C1-Me wagging
978	858	0.97	H14 OOP, H11 asym. OOP
993	870	0.13	Lys-C18/C19-CH2 rocking
1012	887	0.07	Lys-C18-C19 stretching, C18/C19-CH2 wagging, H17 ip bending
1059	929	0.09	N-H ip bending, H15 OOP, C17-CH2 wagging
1066	935	0.34	H11-H12 OOP sym.
1105	969	0.53	C5-C9-C13-Me rocking, C1-Me rocking, H7 OOP
1106	970	0.32	H15/N-H sym. OOP
1112	975	0.31	H7 OOP, C1-Me rocking
1129	990	2.75	C13-Me rocking;
1151	1009	0.08	C9-Me rocking
1167	1023	0.40	C13-Me rocking, C16- methylene wagging,
1176	1031	0.16	C13-Me rocking, C16- methylene wagging
1200	1052	0.46	C5-Me und C-9 Me wagging, C6-C7 stretching, C1-Me rocking
1219	1069	0.17	Lys-Me rocking
1222	1072	0.41	H8/H7 OOP, C2-CH2 twisting, C1-Me wagging
1242	1089	1.38	C9-Me rocking, C10-C11 stretching, H10-H11 asym. bending, H11- H12 asym. bending, C13-Me rocking, H14- H15 bending, C1-Me

			rocking
1268	1112	3.65	C4-methylene twisting, Lys-C17-methylene twisting, Lys-C19-methylene twisting,
1258	1103	0.48	Lys-C17/C19-CH2 twisting
1263	1108	0.59	C2-CH2 twisting
1266	1110	1.06	C14-C15 stretching, Lys-Lys-C17-CH2 twisting, C4-CH2 twisting
1268	1112	1.59	C4-CH2 twisting, Lys-C17/C19-CH2 twisting
1272	1115	1.04	Lys-C19-CH2 twisting
1284	1126	0.10	H8/H7 symm. OOP, C4-CH2 twisting, C1-C(Me) stretching, C2-CH2 wagging, C1-Me wagging
1325	1162	0.51	C5-Me stretching, C4-methylene wagging,
1342	1177	0.69	Lys-H20 bending
1343	1178	2.51	H10,H11,H12 bending
1361	1193	6.04	H14 bending
1392	1221	0.14	Lys-C18-CH wagging, Lys-C19-CH2 wagging, Lys-C20-CH wagging
1414	1240	1.16	H11 bending
1419	1244	0.27	C4-CH2 scissoring, C3-CH2 wagging, C2-CH2 wagging
1451	1272	0.09	C12 C13 C17- +C18-methylene wagging +bending
1466	1286	0.26	H7,H8, H10 ip. bending; C1-Me scissoring
1481	1299	0.19	H7 ip. bending
1487	1304	0.06	Lys-C17-CH2 scissoring
1490	1307	0.23	C1 asymm. methylene scissoring, C4-methylene scissoring
1496	1312	0.08	C1-Me scissoring
1497	1313	0.11	C4-CH2 scissoring, C1-Me scissoring
1504	1319	0.74	H10-H12 bending, H12-

			H14 asym. Bending, (H14-H15-NH bending)
1517	1330	0.07	C4-CH2 wagging + scissoring, C3/C2-CH2 scissoring,
1543	1353	0.10	H8-H10 asymm. Bending, C9-Me HOOP, C2-Me scissoring, H19- NH bending, H15 bending
1559	1367	0.09	C2/C3-CH2 scissoring, C1-Me (left) scissoring,
1561	1369	0.29	C9-Me scissoring, Lys CH2 bending, H15 ip bending
1574	1380	0.20	Lys-C18-methylen scissoring, H14-H15- bending, Lys-C16- methylene scissoring
1597	1401	0.12	C13-Me scissoring
1622	1422	0.11	C16-methylene twisting, C18- + C19-methylene wagging
1631	1431	0.10	C13-Me scissoring
1640	1438	0.13	C13-Me scissoring, Lys- C16-CH2 scissoring
1641	1440	0.28	C13-Me asym. scissoring, Lys-C16- methylen scissoring
1645	1443	0.12	C9-Me scissoring
1653	1450	0.11	C1-Me (left) scissoring
1655	1452	0.15	C9-Me asym. scissoring
1667	1462	0.09	C5-Me scissoring
1671	1466	0.10	N-H bending, Lys-C16- CH2 twisting
1673	1467	0.75	C1-Me scissoring, Lys- C16-methylene twisting, H7-H8 bending, H8-H10 asymm. bending, C7=C8 stretching
1737	1523	0.36	C4-C5 C6-C1 stretching C4-methylen wagging
1765	1548	4.93	C9=C10/C11=C12 stretching, H10/H11/H12 ip.

			bending, C13=C14 stretching (weak)
1769	1551	13.79	C13=C14 stretching
1797	1576	0.22	H7/H8 asymm. ip. bending, C7=C8/C9=C10 stretching
1848	1621	0.18	C5=C6 stretching, H7/H8 asymm. ip. bending
1891	1658	1.95	C15=N stretching, Lys-C16-CH2 scissoring
1959	1718	0.97	C15=N stretching, Lys-C16-CH2 scissoring

Table S14: Frequencies and mode assignments for optimised 13-cis retinal in C1C2 (intensity > 0.05).

Freq. [cm ⁻¹]	Scaled (*0.877)	Intensity	Mode
839	735	0.07	HN OOP, Lys-C17-C18-C19-CH2 rocking
873	766	0.06	C2-CH2 rocking, C4-CH2 rocking, C9-C(Me) stretching
892	782	0.22	C2/C4-Me rocking, H12 OOP (weak)
940	824	1.34	H12 ip. bending, C1-Me scissoring
954	836	1.10	H2O wagging, H12 ip. bending
959	841	0.12	H2O wagging
973	853	0.21	H10 OOP
985	864	0.11	H14-OOP
998	875	0.09	Lys-C17-CH2 twisting, Lys-C18-C19 stretching, Lys-C18/C19-CH2 wagging
1018	893	0.13	Lys-C18-CH2 twisting
1063	933	0.11	H11/H12/H15 sym. OOP
1076	944	0.09	H15-OOP; C19 (lys.) scissoring
1112	975	0.18	C5/C9-Me rock; H7 OOP
1137	997	1.28	H15/N-H- OOP; Lys-C16-CH2 twisting, Lys-C17-CH2 twisting
1150	1008	0.51	C9-Me scissoring
1155	1013	0.40	C9-Me rocking
1176	1032	1.21	C13-Me rock, C14-C15 stretching, H15 ip bending
1210	1061	0.20	C13-Me rock, C3-CH2 twisting, C1-Me scissoring
1226	1075	0.69	C13-Me rock, C2-CH2 twisting, C1-Me scissoring
1235	1083	0.80	C13-Me rock, C3-CH2 twisting
1253	1099	1.49	H15 ip bending, Lys-C16/C17-CH2 bending
1286	1128	1.18	H7/H8 OOP

1295	1136	2.08	C10-C11 stretching, C4-CH2 twisting, Lys-C18-CH2 twisting
1325	1162	3.35	H10/H12 ip. bending, C4-CH2 scissoring
1332	1168	1.16	C4-CH2 wagging, H10/H12 ip bending, H20 ip. bending
1333	1169	0.15	H20 ip. bending
1383	1213	0.55	C13-C(Me) stretching, H12/H14 bending, C13-Me-H bending
1441	1264	0.12	H15 ip bending, Lys-C17-CH2 scissoring
1459	1280	0.14	H15 ip bending, Lys-C18/C19-CH2 scissoring, H20 ip bending
1481	1299	0.78	C1-Me rocking (right), H7 ip. bending, Lys-C17-CH2 scissoring
1489	1305	1.07	H7/H11 ip bending
1497	1313	0.61	H11 ip. bending, C4/C2-CH2 scissoring, C1-Me scissoring
1510	1325	0.41	H10-H11 ip. bending, C1-Me scissoring
1569	1376	0.74	NH ip. bending, Lys-C18 scissoring
1577	1383	1.63	H14/NH ip. bending, Lys-C16/C18-CH2 scissoring
1631	1430	0.77	H14 ip. bending, C16-CH2 scissoring, C13-Me scissoring
1650	1447	0.37	C13-Me scissoring; C16-CH2 scissoring, NH ip bending
1655	1452	0.13	C9-Me scissoring, C13-Me scissoring (weak)
1670	1464	0.11	C1-Me scissoring (left)
1683	1476	0.79	C1-Me scissoring (right)
1718	1506	1.09	H14/N-H ip. bending; C13-Me scissoring

1738	1524	0.25	β -ion.: C4-C5 + C1-C6 stretching, C4-methylene wagging
1778	1560	7.94	C11-C12 stretching, C9-C10 stretching; H10-H11-H12 ip. bending
1799	1578	5.73	C13=C14 stretching, H14 ip. bending, H7-H8 asym. ip. bending
1805	1583	2.71	H8 ip bending, C13-C14 stretching, H14 ip bending
1871	1641	2.79	C15-N stretching, H15-NH asym. bending, Lys-C16-CH2 scissoring
1901	1667	0.94	C16-CH2 scissoring; C15-N stretching, H-N ip bending

Table S15 shows modes assigned by Nack et al. for the experimental ChR2 spectrum.

Table S15: ChR2 modes assigned by Nack et al. (ref. 12 in the paper).

Freq. [cm ⁻¹]	Mode
823	HOOP
845	HOOP
1001	in-plane rocking of two methyl groups at C9 and C13
1061	all- <i>trans</i> C-C stretching
1088	C-C stretching of gauche modes
1125	all- <i>trans</i> C-C stretching
1183	contributions of 13- <i>cis</i> retinal
1200	all- <i>trans</i> characteristic band
1270	CCH in plane rocking of vinyl hydrogens
1440	methylene scissoring
1551	symmetric C=C stretching mode of retinal
1657	C=N stretching of the Schiff base

IR difference spectra of all-trans and 13-cis C1C2

Figure S14 shows the computed IR difference spectra of the optimised 13-*cis* and all-*trans* retinal structures in C1C2 in comparison with the experimental data by Furutani et al. [ref 40 in the paper] The computed line spectra were convoluted with a Lorentz shaped function (5 cm⁻¹ FWHM) before taking the difference. Prominent bands representing changes in Schiff base C=N and C=C stretching modes match well with the experimental findings. For band assignments see also resonance Raman data.

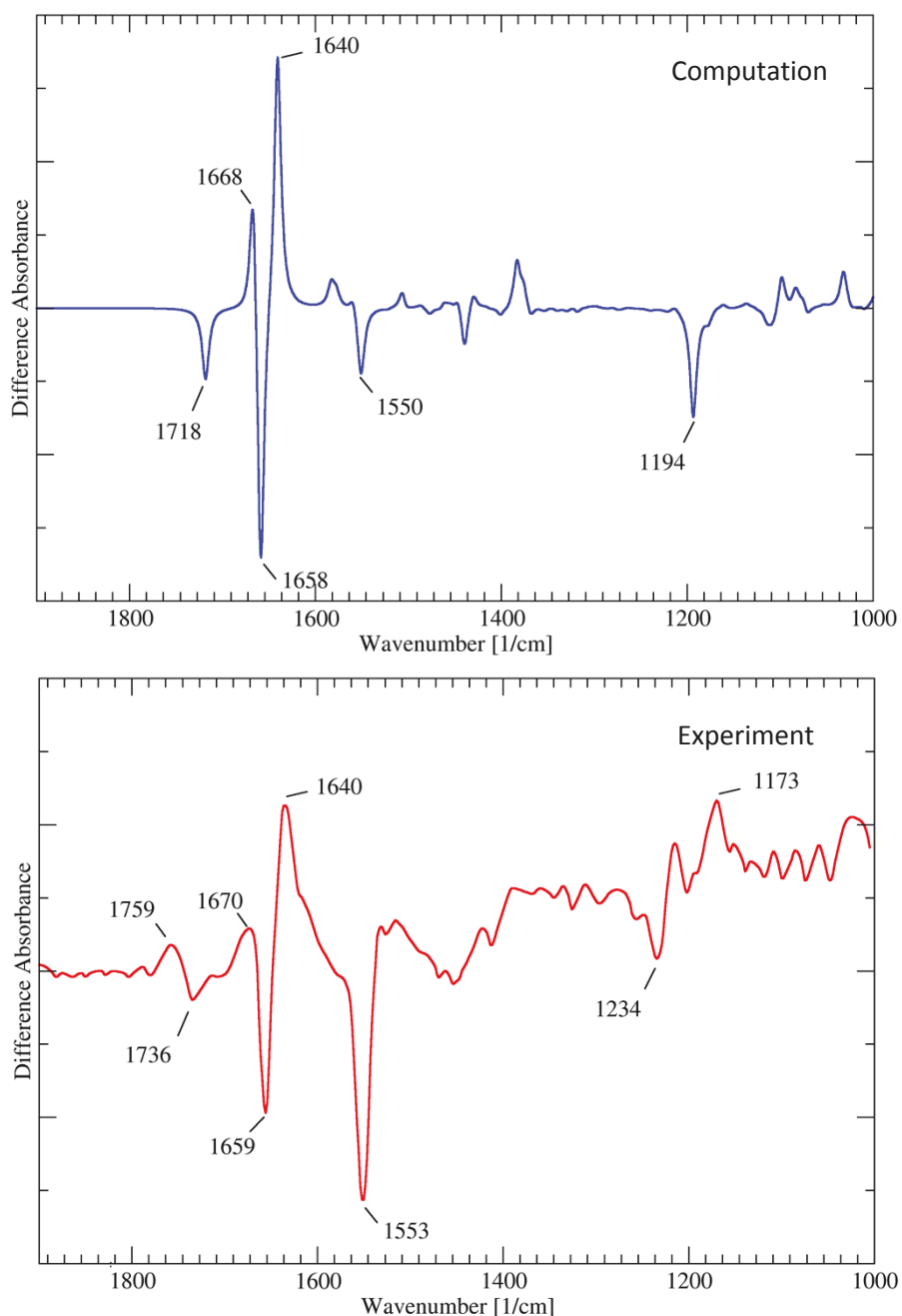


Figure S14: computed IR difference spectrum (above) and experimental result from Furutani and coworkers, reproduced from data in Fig. 8a of reference [40] in the paper.

Forces after HOMO-LUMO excitation

Figure S15 and Figure S16 show force vectors in all-*trans* and 13-*cis* retinal after vertical excitation to the 1B_u -like state.

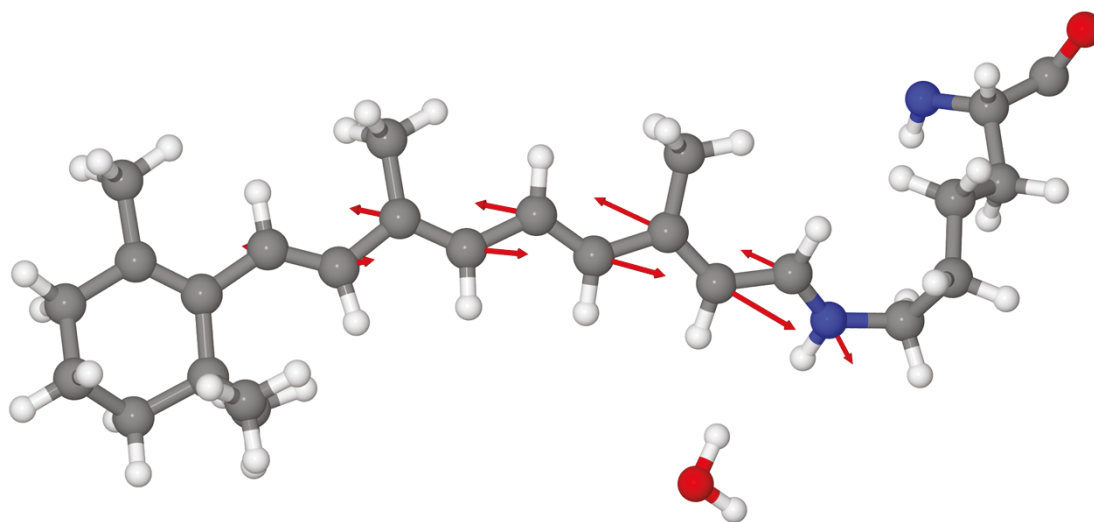


Figure S15: Atomic forces after vertical excitation in the all-*trans* retinal optimised geometry.

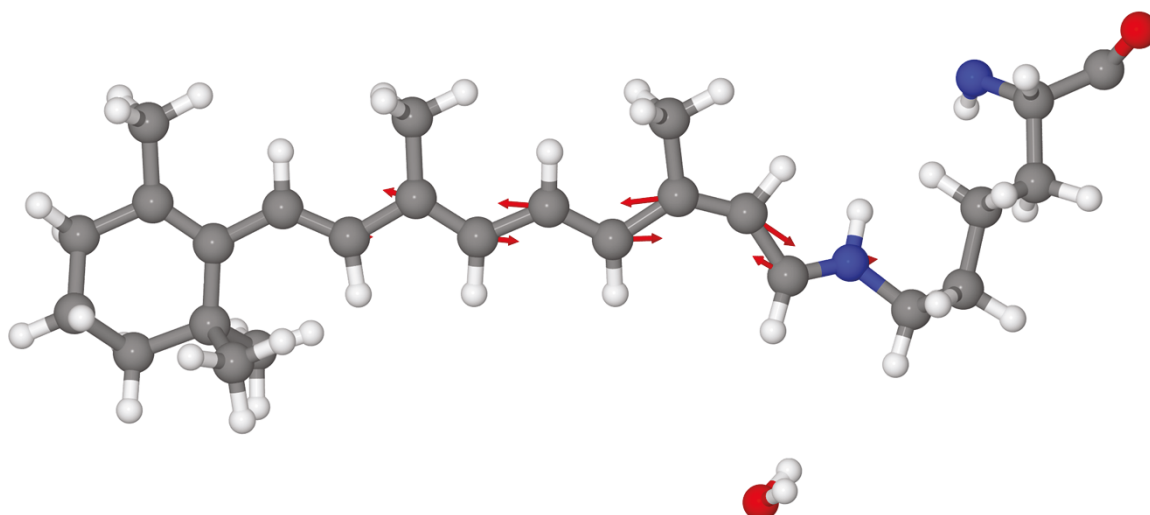


Figure S16: Forces after vertical 1B_u -like excitation in the 13-*cis* retinal optimised geometry.

Influence of the electrostatic environment on retinal absorption

To investigate the influence of the electrostatic environment on retinal absorption, twelve protein residues in the vicinity of the retinal chromophore were selected (Figure S17).

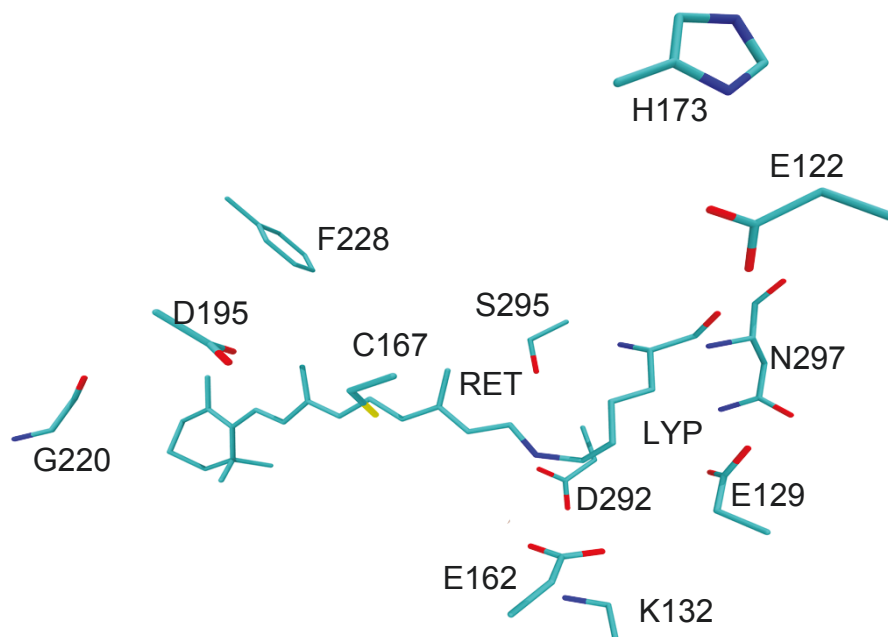


Figure S17: Location of amino acids regulating retinal absorption.

Coordinates of optimised geometries

Table S16: Coordinates of the all-*trans* S₀-CASSCF optimised geometry

Residue Name	Atom Name	x	y	z
E162	N	7.466808	32.349233	19.920353
	H	7.62654	31.408851	19.575532
	C	8.362653	33.416995	19.41741
	H	8.872387	33.87274	20.268553
	C	9.439089	32.891995	18.437634
	H	10.130509	32.242384	18.978521
	H	8.951324	32.27688	17.683998
	C	10.237212	34.02715	17.734696
	H	9.575569	34.759678	17.267194
	H	10.839796	34.562812	18.473617
	C	11.09998	33.488621	16.618971
	O	10.575063	32.700912	15.799877
	O	12.292736	33.852806	16.556715
	C	7.580934	34.510728	18.726468
	O	7.848195	35.683059	18.903266
D292	N	12.042608	34.837237	9.046666
	H	12.152875	33.844232	8.879225
	C	12.503682	35.333422	10.366398
	H	11.700542	35.926977	10.804745
	C	12.833045	34.192913	11.374376
	H	13.539447	33.501744	10.913209
	H	13.344053	34.619582	12.239616
	C	11.61555	33.407137	11.894049
	O	11.756639	32.232354	12.31082
	O	10.515263	34.012584	11.996968
	C	13.695178	36.267867	10.207756
	O	13.769484	37.332394	10.790316
LYP	N	12.695722	39.898305	11.282804
	C	13.285196	40.322242	12.57453
	C	14.600764	41.028471	12.375266
	O	14.757054	42.178544	12.735863
	C	13.370285	39.162551	13.591247
	C	11.981871	38.911138	14.215166
	C	11.883204	37.625107	15.051082
	H	12.656804	41.081609	13.048634
	H	11.769411	39.732698	14.893497
	H	11.221817	38.91852	13.432641
	H	11.022398	37.689059	15.717828
	H	12.762109	37.568008	15.687469
	H	12.987417	39.00472	10.898897
	H	14.062763	39.434506	14.38919

	H	13.765713	38.265405	13.116523
	C	11.75357	36.362092	14.180327
	N	10.399182	36.130845	13.615741
	H	12.440914	36.414446	13.350451
	H	11.981502	35.48992	14.772976
	H	10.337523	35.214501	13.163105
	C	-0.570526	32.616672	10.481896
	C	-1.677281	31.599262	10.117566
	C	-3.069135	32.201215	10.098807
	C	-3.090784	33.264899	9.018764
	C	-1.949069	34.256404	9.134737
	C	-0.797647	33.990833	9.812878
	C	0.227038	35.058086	9.956395
	C	1.395134	35.050684	10.644256
	C	2.381582	36.141492	10.700439
	C	3.535848	35.905584	11.384553
	C	4.686475	36.784715	11.592398
	C	5.836888	36.29476	12.120501
	C	7.061232	37.038285	12.449803
	C	8.115216	36.32779	12.940523
	C	9.348567	36.864895	13.496754
	C	0.740278	31.95696	10.017124
RET	C	-0.576355	32.79861	12.010376
	C	-2.299925	35.566036	8.457608
	C	2.087465	37.422612	9.962167
	C	7.105404	38.518563	12.173817
	H	9.365236	37.861693	13.890488
	H	8.052803	35.257141	12.947428
	H	5.867746	35.245707	12.344206
	H	4.603212	37.823976	11.331823
	H	3.644807	34.928315	11.824305
	H	1.685662	34.191913	11.210284
	H	-0.010102	35.968643	9.446315
	H	2.945195	38.077214	9.959835
	H	1.834226	37.219614	8.927291
	H	1.249046	37.950498	10.411718
	H	8.013277	38.98688	12.530105
	H	7.060439	38.673356	11.101352
	H	6.258229	39.030614	12.607782
	H	-2.970933	36.135282	9.096236
	H	-1.459439	36.194552	8.204348
	H	-2.836583	35.363086	7.533572
	H	1.629467	32.494414	10.296841
	H	0.815162	30.964017	10.446547

	H	0.748919	31.842726	8.940758
	H	0.119732	33.547801	12.353735
	H	-1.553865	33.106462	12.348315
	H	-0.321176	31.860876	12.497401
	H	-4.023087	33.809755	9.0432
	H	-3.070356	32.78602	8.040675
	H	-3.819006	31.445724	9.890718
	H	-3.310974	32.639564	11.063218
	H	-1.478249	31.191511	9.129855
	H	-1.612564	30.758307	10.80278
WAT619	O	9.093849	31.860766	12.361789
	H	9.420837	32.781281	12.119944
	H	9.998057	31.479714	12.525309

Table S17: Coordinates of the 1B_u -CASSCF optimised geometry

Residue Name	Atom Name	x	y	z
LYP	N	12.699263	39.908138	11.279383
	C	13.282307	40.29548	12.582986
	C	14.603443	41.002965	12.405384
	O	14.764759	42.143053	12.796293
	C	13.344225	39.099929	13.56503
	C	11.952862	38.8701	14.195767
	C	11.813246	37.562826	14.996525
	H	12.651584	41.044789	13.068448
	H	11.769064	39.681353	14.896111
	H	11.192163	38.926768	13.418483
	H	10.991875	37.660696	15.707749
	H	12.713833	37.438762	15.59091
	H	12.966587	39.006769	10.899405
	H	14.050369	39.325753	14.364778
	H	13.70192	38.200805	13.060777
RET	C	11.584336	36.330045	14.113384
	N	10.18552	35.937498	14.053585
	H	11.979312	36.512183	13.119946
	H	12.118091	35.49701	14.543517
	H	10.082768	34.942358	13.969365
	C	-0.590269	32.626758	10.48202
	C	-1.709471	31.597217	10.134055
	C	-3.093117	32.241964	10.110961
	C	-3.07624	33.285069	8.997811
	C	-1.958155	34.297205	9.146756
	C	-0.795266	33.999002	9.832199
	C	0.20568	35.049361	9.972815

	C	1.402757	35.060412	10.655224
	C	2.337048	36.140046	10.67336
	C	3.538622	35.904682	11.36204
	C	4.622368	36.766037	11.537511
	C	5.800947	36.280656	12.129919
	C	6.988494	36.992133	12.403862
	C	8.053303	36.241601	12.977844
	C	9.254371	36.71966	13.478818
	C	0.708801	31.92469	10.011799
	C	-0.581147	32.789048	12.020622
	C	-2.298333	35.617243	8.488909
	C	2.072958	37.416488	9.920527
	C	7.121324	38.448006	12.040364
	H	9.444308	37.7706	13.478351
	H	7.927001	35.179745	13.004415
	H	5.790069	35.243756	12.401736
	H	4.571048	37.788064	11.214033
	H	3.64954	34.922367	11.785911
	H	1.723864	34.226777	11.244239
	H	-0.027556	35.960504	9.46805
	H	2.964444	38.016289	9.841322
	H	1.73506	37.204469	8.914423
	H	1.305639	37.998804	10.423127
	H	8.150118	38.767816	12.091188
	H	6.780835	38.620861	11.025028
	H	6.536461	39.086001	12.694331
	H	-3.135109	36.073712	9.011655
	H	-1.50179	36.34135	8.453428
	H	-2.627468	35.451503	7.465076
	H	1.618418	32.469974	10.224602
	H	0.81105	30.953278	10.496464
	H	0.687095	31.756352	8.937041
	H	0.122034	33.53131	12.386917
	H	-1.55895	33.113336	12.360743
	H	-0.338519	31.847059	12.510718
	H	-4.036803	33.788926	9.00334
	H	-3.009662	32.795933	8.022564
	H	-3.860335	31.493332	9.906461
	H	-3.328363	32.705692	11.070838
	H	-1.530482	31.161369	9.147834
	H	-1.701728	30.761629	10.837433
WAT619	O	9.295719	32.050658	12.444386
	H	9.620581	32.981191	12.222325
	H	10.197998	31.681238	12.630337

Table S18: Coordinates of the S_0 -CASSCF optimised geometry in alternative protonation

Residue Name	Atom Name	x	y	z
E162	N	7.454168	32.338145	19.894443
	H	7.594888	31.385416	19.576235
	C	8.372837	33.370888	19.378397
	H	8.895933	33.81765	20.227404
	C	9.423348	32.7361	18.429873
	H	10.056049	32.057481	19.006508
	H	8.898369	32.113268	17.713873
	C	10.330742	33.737398	17.660375
	H	9.862453	34.707891	17.514624
	H	11.236877	33.91737	18.243062
	C	10.695253	33.236953	16.276585
	O	10.063621	32.492956	15.55614
	O	11.895784	33.66456	15.844987
	H	11.998762	33.398403	14.893406
	C	7.61186	34.479975	18.673354
	O	7.942691	35.643367	18.800039
D292	N	12.077763	34.838838	9.053952
	H	12.200762	33.849263	8.87478
	C	12.527734	35.326959	10.380488
	H	11.717672	35.907626	10.822158
	C	12.933212	34.205847	11.393635
	H	13.419633	33.38825	10.865817
	H	13.687379	34.611665	12.073176
	C	11.80867	33.646209	12.284256
	O	12.092512	33.07285	13.365946
	O	10.611971	33.946964	12.071426
	C	13.707058	36.272349	10.205243
	O	13.783344	37.320359	10.816756
LYP	N	12.716996	39.891522	11.290442
	C	13.306607	40.334454	12.579475
	C	14.623892	41.024192	12.376881
	O	14.789398	42.170613	12.743492
	C	13.39578	39.206442	13.62444
	C	12.002092	38.961658	14.239907
	C	11.882973	37.614455	14.953654
	H	12.684822	41.10591	13.042271
	H	11.794494	39.735444	14.972956
	H	11.244323	39.0327	13.459561
	H	11.018133	37.609511	15.620342
	H	12.762193	37.477956	15.57688
	H	13.027982	39.008642	10.897604

	H	14.080206	39.499127	14.422635
	H	13.81446	38.308735	13.170124
RET	C	11.760094	36.481007	13.927851
	N	10.373782	36.18388	13.507806
	H	12.340191	36.70239	13.045638
	H	12.146995	35.578276	14.361164
	H	10.293085	35.250467	13.103633
	C	-0.589657	32.611923	10.476188
	C	-1.696573	31.597415	10.105664
	C	-3.086569	32.203901	10.079702
	C	-3.1034	33.267721	8.999767
	C	-1.959794	34.254829	9.11943
	C	-0.81379	33.984307	9.803844
	C	0.208477	35.049162	9.951137
	C	1.368411	35.039335	10.650085
	C	2.353923	36.126744	10.705666
	C	3.499529	35.890083	11.402001
	C	4.644424	36.770255	11.596972
	C	5.797957	36.291821	12.126318
	C	7.012452	37.052606	12.414271
	C	8.090515	36.361584	12.883136
	C	9.331483	36.927379	13.347701
	C	0.723866	31.949874	10.021939
	C	-0.601778	32.79502	12.004546
	C	-2.298571	35.566739	8.440578
	C	2.075263	37.403012	9.953983
	C	7.032679	38.528561	12.122882
	H	9.383396	37.957681	13.633762
	H	8.043069	35.289379	12.901671
	H	5.838703	35.24656	12.365361
	H	4.562063	37.804214	11.318656
	H	3.60476	34.916582	11.850484
	H	1.651847	34.182584	11.222033
	H	-0.019328	35.958228	9.434766
	H	2.933197	38.057359	9.957487
	H	1.837169	37.191496	8.917848
	H	1.231607	37.934722	10.387731
	H	7.928349	39.02106	12.47583
	H	6.988031	38.668736	11.049047
	H	6.17418	39.025109	12.550881
	H	-2.961805	36.143777	9.080068
	H	-1.452329	36.186262	8.183474
	H	-2.838313	35.367831	7.518106
	H	1.613645	32.482395	10.309953

	H	0.790944	30.957009	10.45258
	H	0.744278	31.835485	8.94625
	H	0.078017	33.558077	12.350219
	H	-1.585293	33.084154	12.34029
	H	-0.327894	31.862877	12.492149
	H	-4.033493	33.81557	9.024049
	H	-3.081853	32.790798	8.02103
	H	-3.83752	31.451193	9.868589
	H	-3.330803	32.643299	11.042866
	H	-1.492496	31.189735	9.119131
	H	-1.638721	30.756135	10.790895
WAT619	O	9.330669	31.866132	12.953708
	H	9.850373	32.551087	12.454745
	H	9.593001	32.106618	13.864621

Table S19: Coordinates of the ${}^1\text{B}_u$ -CASSCF optimised geometry at $\text{P}_{[-97]}^+$

Residue Name	Atom Name	x	y	z
LYP	N	12.697681	39.892355	11.268457
	C	13.277806	40.235143	12.584291
	C	14.603047	40.955093	12.443931
	O	14.76394	42.069567	12.898338
	C	13.313324	39.010698	13.538448
	C	11.922265	38.807145	14.184493
	C	11.793909	37.571018	15.09247
	H	12.643948	40.975293	13.076974
	H	11.744166	39.66183	14.8305
	H	11.143997	38.795719	13.417254
	H	10.940773	37.697201	15.755262
	H	12.670106	37.519709	15.734574
	H	12.976601	39.009465	10.855149
	H	14.033068	39.200835	14.334338
	H	13.644223	38.110844	13.018182
	C	11.635023	36.269193	14.313418
RET	N	10.388247	36.192892	13.560394
	H	12.432405	36.177204	13.592908
	H	11.687396	35.428982	14.964063
	H	10.472368	35.577331	12.753421
	C	-0.642928	32.602616	10.501469
	C	-1.806515	31.615381	10.171025
	C	-3.166235	32.311885	10.131982
	C	-3.12445	33.305383	8.974414
	C	-1.9842	34.290769	9.099658
	C	-0.791652	33.96982	9.82992

	C	0.225601	34.987909	9.973427
	C	1.418287	34.976256	10.672838
	C	2.381072	36.058254	10.749327
	C	3.59388	35.795788	11.503767
	C	4.635695	36.670017	11.71445
	C	5.871732	36.301284	12.338505
	C	6.989535	37.164754	12.376272
	C	8.359803	36.717754	12.661629
	C	9.186865	36.675015	13.793315
	C	0.627705	31.845264	10.04229
	C	-0.623463	32.791345	12.037892
	C	-2.2674	35.611574	8.421951
	C	2.203944	37.310884	9.93236
	C	6.898148	38.575323	11.858549
	H	8.957824	37.112097	14.743589
	H	8.927526	36.446021	11.803908
	H	5.977347	35.29456	12.683184
	H	4.546704	37.67517	11.350221
	H	3.688957	34.800842	11.904636
	H	1.714764	34.127648	11.257567
	H	0.013854	35.908595	9.472844
	H	3.150094	37.800397	9.748309
	H	1.771026	37.088026	8.964328
	H	1.549795	38.025231	10.430717
	H	7.829419	39.107817	12.004838
	H	6.664958	38.593828	10.798562
	H	6.119465	39.114909	12.37956
	H	-3.050878	36.143693	8.958618
	H	-1.419517	36.272119	8.340628
	H	-2.643088	35.450125	7.412113
	H	1.550583	32.393652	10.181614
	H	0.728019	30.912367	10.596096
	H	0.567825	31.600237	8.983267
	H	0.059969	33.558881	12.388417
	H	-1.604804	33.100835	12.375412
	H	-0.355465	31.866282	12.545502
	H	-4.066735	33.837483	8.954889
	H	-3.06493	32.779078	8.018762
	H	-3.965621	31.587598	9.969911
	H	-3.367754	32.828292	11.071904
	H	-1.64733	31.153263	9.19421
	H	-1.831269	30.793448	10.888053
WAT619	O	9.315831	31.957441	12.448486
	H	9.622643	32.899203	12.30005

H 10.217794 31.564522 12.551456

Table S20: Coordinates of the $^1\text{B}_u$ -CASSCF optimised geometry at $\text{P}_{[-247]}^-$

Residue Name	Atom Name	x	y	z
LYP	N	12.700757	39.915145	11.280009
	C	13.280374	40.31623	12.579004
	C	14.598392	41.025491	12.38345
	O	14.756504	42.171745	12.756142
	C	13.405387	39.116882	13.537939
	C	12.035176	38.653536	14.087398
	C	12.163349	37.292462	14.807039
	H	12.63888	41.044661	13.083659
	H	11.67986	39.390566	14.809025
	H	11.294464	38.600541	13.285398
	H	11.752373	37.356885	15.818053
	H	13.21588	37.068747	14.902462
	H	12.995771	39.029763	10.886056
	H	14.03396	39.405102	14.382619
	H	13.910801	38.293145	13.030491
RET	C	11.51894	36.136858	14.043499
	N	10.07221	36.18836	14.217257
	H	11.747679	36.228031	12.9987
	H	11.88017	35.193424	14.422811
	H	9.72467	36.143718	15.156813
	C	-0.551165	32.605182	10.428974
	C	-1.635986	31.556864	10.037506
	C	-3.047042	32.144886	10.053907
	C	-3.104952	33.222791	8.97437
	C	-2.018913	34.265648	9.124667
	C	-0.796754	33.985741	9.813862
	C	0.165401	35.056818	9.963166
	C	1.369081	35.09012	10.632033
	C	2.311343	36.190767	10.65944
	C	3.566653	35.919689	11.317628
	C	4.648941	36.759407	11.474996
	C	5.810304	36.335675	12.182681
	C	6.903608	37.16198	12.519175
	C	7.892439	36.703557	13.51343
	C	9.234119	36.392793	13.240659
	C	0.773098	31.957246	9.949446
	C	-0.554365	32.720183	11.971447
	C	-2.384219	35.595214	8.504775
	C	2.047126	37.472537	9.919351

	C	6.924825	38.634629	12.196385
	H	9.629509	36.389359	12.246478
	H	7.625457	36.767514	14.557136
	H	5.834301	35.315659	12.512378
	H	4.610827	37.762459	11.094794
	H	3.663138	34.932504	11.737211
	H	1.707591	34.260184	11.21803
	H	-0.099462	35.973741	9.482232
	H	2.965452	38.008709	9.728003
	H	1.581422	37.28635	8.958754
	H	1.38658	38.126881	10.487535
	H	7.558797	39.180685	12.885104
	H	7.303781	38.806139	11.19441
	H	5.931845	39.054503	12.255001
	H	-3.178609	36.067804	9.078306
	H	-1.573532	36.301465	8.43217
	H	-2.771332	35.452948	7.496722
	H	1.666753	32.488541	10.246637
	H	0.874238	30.947843	10.347665
	H	0.791844	31.888679	8.86561
	H	0.159149	33.44297	12.357494
	H	-1.530308	33.045652	12.322035
	H	-0.320089	31.759133	12.428022
	H	-4.080665	33.692652	9.026983
	H	-3.048793	32.770192	7.981373
	H	-3.783657	31.369968	9.840029
	H	-3.287764	32.567405	11.031334
	H	-1.442141	31.178763	9.030661
	H	-1.592335	30.690782	10.702175
WAT619	O	9.324117	32.024062	12.433215
	H	9.619035	32.976631	12.282609
	H	10.242439	31.655257	12.51773

Table S21: Coordinates of the 13-*cis* S₀-CASSCF optimised geometry

Residue Name	Atom Name	x	y	z
E162	N	7.476169	32.353561	19.922919
	H	7.647083	31.413108	19.582817
	C	8.375742	33.424104	19.419453
	H	8.875477	33.885882	20.27363
	C	9.475116	32.911436	18.453296
	H	10.133789	32.222074	18.986466
	H	8.994917	32.34255	17.658467
	C	10.333035	34.053851	17.83038
	H	9.709864	34.841779	17.403657
	H	10.944933	34.517125	18.608265

	C	11.20061	33.562537	16.693723
	O	10.699379	32.747723	15.886572
	O	12.364419	34.005756	16.597631
	C	7.591936	34.511513	18.716547
	O	7.859022	35.684672	18.887027
D292	N	12.060627	34.844974	9.040786
	H	12.175477	33.851702	8.879329
	C	12.514936	35.352081	10.356524
	H	11.709889	35.945361	10.784809
	C	12.847322	34.222541	11.367216
	H	13.55264	33.529497	10.906037
	H	13.36272	34.648752	12.229107
	C	11.619986	33.447365	11.873598
	O	11.774205	32.303274	12.369999
	O	10.496299	34.01303	11.83427
	C	13.699281	36.28358	10.202187
	O	13.76908	37.352849	10.779872
LYP	N	12.698514	39.906623	11.287779
	C	13.293131	40.351013	12.572467
	C	14.60401	41.052854	12.352634
	O	14.761425	42.212013	12.682455
	C	13.406377	39.2074	13.603518
	C	12.043767	38.968133	14.286091
	C	11.946442	37.638598	15.065349
	H	12.669624	41.116476	13.044208
	H	11.861139	39.76691	14.996252
	H	11.245952	39.049848	13.547609
	H	11.068491	37.687738	15.710922
	H	12.808859	37.558199	15.717871
	H	13.030677	39.036596	10.883897
	H	14.131058	39.487143	14.370392
	H	13.780939	38.303156	13.124618
RET	C	11.804895	36.405234	14.147099
	N	10.701371	36.667098	13.220045
	H	12.709975	36.273272	13.577256
	H	11.62014	35.518674	14.731519
	H	10.894454	37.410703	12.570629
	C	-0.399278	32.630029	10.481953
	C	-1.559916	31.725422	10.014693
	C	-2.912891	32.412403	10.059444
	C	-2.877407	33.599909	9.111805
	C	-1.63905	34.460808	9.25572
	C	-0.51224	34.049302	9.897421
	C	0.583166	35.02825	10.095723

	C	1.729935	34.922216	10.809704
	C	2.751339	35.967624	10.934012
	C	3.903156	35.66289	11.598538
	C	5.000364	36.586037	11.806985
	C	6.240287	36.311892	12.28806
	C	7.245847	37.369998	12.332739
	C	8.553831	37.321538	12.6879
	C	9.450328	36.36452	13.301632
	C	0.88052	31.918908	10.010904
	C	-0.445183	32.724649	12.018198
	C	-1.823413	35.855942	8.683935
	C	2.47819	37.29833	10.272126
	C	6.812054	38.727483	11.849663
	H	9.167358	35.525407	13.906089
	H	9.072135	38.246666	12.531348
	H	6.521133	35.313667	12.550938
	H	4.794271	37.596371	11.533629
	H	4.02468	34.66218	11.968468
	H	1.968298	34.027203	11.340846
	H	0.417231	35.970599	9.616382
	H	3.296746	37.990919	10.378482
	H	2.308604	37.162822	9.208787
	H	1.589584	37.76342	10.691566
	H	5.941613	39.067446	12.390134
	H	7.590512	39.46385	11.989274
	H	6.561453	38.690256	10.796861
	H	-2.02325	36.569284	9.479283
	H	-0.967733	36.212157	8.123385
	H	-2.673954	35.871878	8.009854
	H	1.790445	32.409968	10.302556
	H	0.914225	30.914877	10.417269
	H	0.894476	31.828249	8.932909
	H	0.301018	33.39542	12.41792
	H	-1.40242	33.094921	12.355857
	H	-0.281247	31.742017	12.454463
	H	-3.739068	34.234111	9.271811
	H	-2.965998	33.246147	8.084951
	H	-3.704557	31.729426	9.767111
	H	-3.13376	32.749299	11.068381
	H	-1.373575	31.409778	8.990946
	H	-1.553913	30.817903	10.613601
WAT619	O	9.138441	31.897296	12.397368
	H	9.420287	32.814175	12.075477
	H	10.069598	31.592059	12.592993

Paper II

New perspectives on an old issue—a comparative MS-CASPT2 and OM2-MRCI study of polyenes and protonated Schiff bases

I. Dokukina, C. M. Marian, O. Weingart

Photochem. Photobiol. **2017**, *93*, 1345-1355

New Perspectives on an Old Issue: A Comparative MS-CASPT2 and OM2-MRCI Study of Polyenes and Protonated Schiff Bases[†]

Irina Dokukina, Christel M. Marian and Oliver Weingart*

Institut für Theoretische Chemie und Computerchemie, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany

Received 30 June 2017, accepted 13 August 2017, DOI: 10.1111/php.12833

ABSTRACT

Polyenic systems are involved in light perception of numerous living organisms. Although a π -conjugated backbone is a common feature of all polyenes, their photophysics may vary. We provide a comparative quantum mechanical study of low-lying S_1 and S_2 excited states in short (3–5 double bonds) symmetric all-*trans* linear polyenes and corresponding protonated Schiff bases. In our investigation, we use the well-established *ab initio* multireference CASPT2 approach and benchmark the efficient semiempirical OM2-MRCI approach against it. For all protonated Schiff bases, MS-CASPT2 results in two distinct S_1 minima with inverted and noninverted bond length pattern, respectively. We find that OM2-MRCI is a computationally affordable and reliable alternative to MS-CASPT2 for investigations of polyenic systems, particularly when highly demanding calculations (e.g. excited-state dynamics) need to be performed.

INTRODUCTION

Linear π -conjugated systems, generally named polyenes, belong to a widespread class of natural compounds. One prominent example is the class of carotenoids. They play an essential role in light harvesting (LH) complexes of the photosystem found in photosynthetic species (e.g. plants, algae, cyanobacteria and photosynthetic bacteria) where they enhance light absorption (1). In photoreceptors ranging from mammalian visual rhodopsins (2) to microbial rhodopsins like bacteriorhodopsin or channel-rhodopsins (3,4), a carotenoid derivative—protonated retinal Schiff base (RPSB)—triggers protein activity. Despite structural similarities, the photochemical properties of different polyenic systems may significantly vary. In the current study, we investigate the properties of all-*trans* protonated Schiff base models (PSBs) and corresponding symmetric linear polyenes (termed linear polyenes throughout the article). The light-absorbing bright state in PSBs is described in the literature as “ 1^1B_u -like” (5–8). This term originates from the nomenclature of symmetric linear polyenes, where the optically bright 1^1B_u excited state with ionic nature is dominated by the singly excited HOMO→LUMO configuration (9). There is, however, a significant difference in the two systems. In linear polyenes, excitation to the 1^1B_u state leads to bond length equalization (9–11), whereas, according to an earlier

accepted mechanism, the ionic 1^1B_u -like excitation in PSBs transports the positive NH_2^+ charge toward the CH_2 terminus and *inverts* the bond ordering, that is former single bonds gain double-bond character and vice versa (12–14). The bond-inverted state in linear polyenes rather corresponds to the optically dark covalent 2^1A_g state (9,15). The influence of the bond length pattern on the photophysical properties of the carotenoid peridinin was investigated by Knecht *et al.* (16). In their study, they were able to show a correlation between state composition and bond length alternation (BLA) due to the charge transfer character of the low-lying S_1 and S_2 excited states. For peridinin, small BLA values (delocalized π -density) lead to a dark S_1 state with dominant double excitation character and a bright S_2 state with single excitation character. An increase in BLA results in S_1 and S_2 states with similar brightness and dipolar character.

Another difference between linear polyenes and PSBs concerns the state ordering of the low-lying excited states. In all-*trans* linear polyenes with more than three double bonds, the 2^1A_g state is reported to lie below the 1^1B_u state (17,18), while in protonated Schiff bases the bright state in vacuum is dominantly the S_1 state.

As isomerization in retinal Schiff bases involves torsion about formal double bonds, the bond length pattern can play a significant role in this process. Both bond inversion and equalization would facilitate the isomerization process, but bond inversion would be favorable. Although for many years there was an agreement on a bond-inverted 1^1B_u -like S_1 state in PSBs, there is a lack of high-level calculations which consequently document the change of the bond length pattern and the configurations of the low-lying excited states with respect to the iso-electronic polyenes. Most studies of these aspects rely on complete active space self-consistent field (CASSCF) computed potential energy surfaces (PES), a method which—due to the lack of dynamic correlation—is prone to fail in the prediction of correct excitation energies and state orderings. Higher correlated methods were applied, for example in the studies of Olivucci *et al.* (19) or in the recent work of Martinez *et al.* (20). Tuna *et al.* (21) studied reaction paths of a retinal model with ADC2 and CC2 methods. In these works, it becomes apparent that CASSCF indeed tends to overestimate the bond length inversion and the steepness of the computed PES with respect to the torsion coordinate.

The view of a bond-inverted state in PSBs has been challenged by coupled-cluster/time-dependent density functional theory (CC/TD-DFT) and multistate complete active space second-order perturbation (MS-CASPT2) computations on Schiff base models. In the combined CC and TD-DFT study of Send

*Corresponding author email: Oliver.Weingart@hhu.de (Oliver Weingart)

[†]This article is a part of the Special Issue dedicated to Prof. Volker Buß on the occasion of his 75th birthday.

© 2017 The American Society of Photobiology

et al. (22), a significant difference in the electronic structure between polyenes and protonated Schiff bases was reported, and a bond-inverted minimum for the investigated PSBs was, however, not identified. Upon excitation to the bright state, the bond lengths changed only little.

The study of Valsson and Filippi (23) used quantum Monte Carlo, second-order approximate coupled-cluster model (CC2), and CASSCF and CASPT2 techniques to investigate *cis*-Schiff base models with three, four and five double bonds and the full retinal chromophore. For the small three-double-bond models, primarily bond equalization was observed upon S_1 excitation. The four-double-bond model showed no significant bond inversion, and only in the five-double-bond model (full retinal without β -ionone ring), two CASPT2 minima were identified with inverted and unchanged bond order. Moreover, taking the basis set dependency of the second minimum into account, the authors concluded that the occurrence of the bond-inverted structure must be an artifact and could not be reached upon excitation. Nevertheless, one of the crucial statements in their study is the inability of CASSCF to properly describe the photophysics and photochemistry of polyenes.

Two different S_1 state minima were also reported in the CASSCF study of Conti *et al.* (24). Fluorine-substituted PSBs showed the existence of an “anomalous” higher energy minimum on the S_1 excited-state surface with no bond inversion. Related aspects are also discussed in the recent Amber/CASSCF//CASPT2 work of Gozem *et al.* (25) studying 11-*cis* locked chromophores in opsin. In the computed models, two S_1 minima with inverted and noninverted configuration coexist which are separated by a small barrier of ca. 0.09 eV. The CASSCF study of El-Tahawy also indicates the existence of the two minima in the 11-*cis* retinal chromophore, when an electromagnetic field is applied (26).

Clearly, a thorough investigation on the discussed aspects must be based on geometries including both static and dynamic correlation contributions. In this work, we study the primary effects upon excitation in linear polyenes and analogous protonated Schiff bases with three, four and five double bonds (Fig. 1). We identify ground- and excited-state minima with MS-CASPT2/cc-pVDZ optimization methods and investigate and compare their electronic structure. With this study, we aim to close the gap of missing high-level comparative polyene and Schiff base studies to consequently document their different photochemical behavior and clarify the excited-state surface topology

in the immediate vicinity of the Franck–Condon point. Furthermore, we benchmark the promising semiempirical orthogonalization model 2 with multireference configuration interaction (OM2-MRCI) strategy (27–29) against the obtained MS-CASPT2 data. Valid data obtained with this method may pave the way toward reliable molecular dynamics simulations of bio-chromophores within their protein environment, which are currently unaffordable at the MS-CASPT2 level of theory.

MATERIALS AND METHODS

Computational details. All *ab initio* calculations were performed with Dunning’s correlation-consistent (cc)-pVDZ basis set. Ground-state geometries were first optimized with the Møller–Plesset expansion truncated at second order (MP2) in Gaussian09 (30). Optimized MP2 geometries were used for further ground- and excited-state CASSCF and MS-CASPT2 geometry optimizations with Molcas 8.1 (31) applying the resolution of the identity (RI) approximation and using numerical gradients for MS-CASPT2. All MS-CASPT2 calculations were performed with the IPEA shift set to 0 and an imaginary shift of 0.2 for the zero-order Hamiltonian. The CASSCF calculations were performed with state averaging (SA) over the three lowest states in the given symmetry with equal weighting. All π -orbitals and π -electrons were included in the active space. In MS-CASPT2 calculations, the three lowest states of A' symmetry were allowed to interact. S_1 geometries of PSBs were optimized using two starting structures, the ground state and a manually pre-inverted geometry obtained by elongating double and shortening single bonds. Due to convergence problems, the decapentaene 1^1B_u geometry was optimized with SS-CASPT2 and a small atomic natural orbital set (ANO-S).

Semiempirical OM2-MRCI calculations were performed with the MNDO program using the standard OM2 Hamiltonian without reparameterization (32,33). Three configurations with single and double excitations from a restricted open-shell Hartree–Fock (ROHF) wave function were used as reference for the MRCI expansion. The active space consisted of all π -orbitals and π -electrons. For automatic maintenance of the π -orbital space during optimization, the π -orbital tracking option was enabled. No symmetry constraint was used and the ground-state optimization was performed with the RHF OM2 Hamiltonian. Additionally, OM3 and OM3-MRCI calculations were performed (34). As OM3 results qualitatively reproduce OM2, they will not be discussed in the main article and can be found in the Data S1 and S2.

RESULTS AND DISCUSSION

Notations for ground and excited states

Two different notations can be used to classify the ground and excited states of the investigated structures. Throughout the

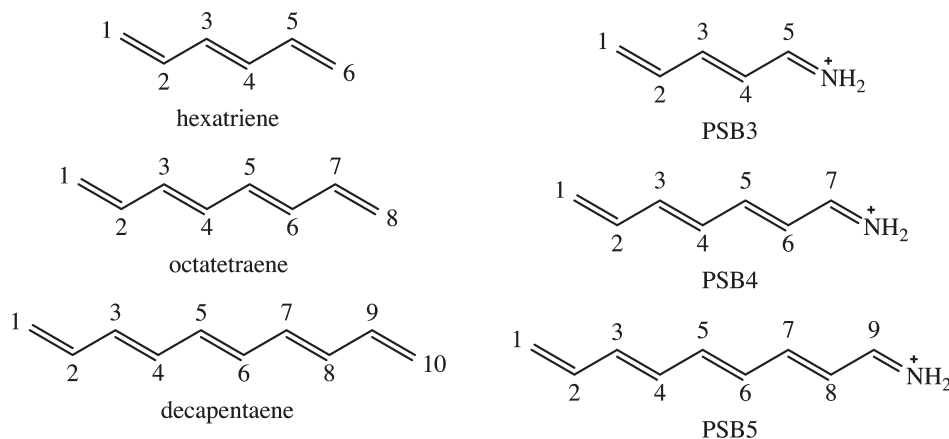


Figure 1. Linear polyenes and protonated Schiff base models used for current calculations.

paper, we make use of both standards, the notation describing the pure energetic ordering of the states (i.e. S_0 , S_1 , S_2), and the one describing their symmetry (i.e. 1^1A_g , 1^1B_u , 2^1A_g). Clearly, these notations are not synonymous; nevertheless, we decided to use the symmetry-related classification for the polyene ground and excited states pointing out that in these systems, the energetic ordering of the states changes as a function of chain length. The polyene ground state has always 1^1A_g symmetry, the state corresponding to HOMO \rightarrow LUMO transition (with largest oscillator strength) is the 1^1B_u state, and the spectroscopically dark state with HOMO $^2\rightarrow$ LUMO 2 double excitation as the leading configuration is denoted as 2^1A_g . In all Schiff bases, HOMO \rightarrow LUMO excitation is always the first excited singlet state (S_1 , largest oscillator strength) with A' symmetry, while the doubly excited HOMO $^2\rightarrow$ LUMO 2 state is always the second (S_2 , weakly allowed) excited state ($3A'$ in the given C_s symmetry). We particularly avoid of speaking about “ 1^1B_u –” or “ 2^1A_g –like” excited PSB states used earlier in the literature, as the type of excitation in PSBs and the structural consequences are *not* related to the corresponding polyene states; this will also be shown in the current work.

Figure 2 reports the results of MS-CASPT2 and OM2-MRCI ground- and excited-state geometry optimizations of hexatriene, octatetraene, decapentaene and the corresponding Schiff bases PSB3 to PSB5. The calculated bond lengths for the CASSCF optimizations can be found in the Data S1 and S2. We will start the discussion with the ground-state geometries of polyenes and PSBs and their vertical excitation energies, followed by outlining the excited-state bond length parameters including adiabatic and emission properties. In the next section, we discuss the electronic relocation in the excited states with difference density plots and provide linearly interpolated paths between the S_1 excited-state minima of the PSBs. The chapter closes with static rotations of the central single and double bonds of PSB5 in the bright S_1 excited state performed at MS-CASPT2 and OM2-MRCI levels of theory.

Bond lengths in the ground state

For the linear polyenes, the calculated bond length pattern agrees well with the results reported in the literature (22). The 1^1A_g geometries show a pronounced bond length alternation with short double bonds and long single bonds. The inner double bonds are slightly shorter than at the termini. The largest deviation (up to 0.02 Å) between the applied methods is found in the double-bond lengths. OM2-MRCI delivers the shortest values, followed by CASSCF (see Data S1 and S2) and MS-CASPT2. The situation changes in the optimized Schiff bases. With 1.28–1.31 Å, the C–N double bond is the shortest bond in the polyene chain. The alternation decreases toward the N-terminus. MS-CASPT2 gives significantly less alternating structures, especially near the N-tail. In the case of PSB5, we even note equal bond lengths for the C5–C6 and C6–C7 bonds and an elongated C7–C8 bond with respect to the other methods. A similar result has also been reported for the CASPT2-optimized PSB5 model for 11-*cis* retinal (19). OM2-MRCI stays closer to the CASSCF computed bond length values.

Vertical excitation energies

Vertical excitation energies to the 2^1A_g and 1^1B_u states of the linear polyenes computed with MS-CASPT2 and OM2-MRCI

can be found in Table 1. The values estimated with the MS-CASPT2 approach are in good agreement with vertical excitation energies previously reported (see Table S1 in the Supporting information). For species with more than three double bonds, the 2^1A_g state is always below the 1^1B_u state. In hexatriene, the 1^1B_u state drops below the 2^1A_g state, which corresponds to earlier reported CASPT2 values (35) and is in line with experimental results (36–38). The 2^1A_g – 1^1B_u energy difference is, however, very small (0.08 eV in the experiment and 0.06 eV in MS-CASPT2, see Table 1). CASSCF values for the 2^1A_g state are also in the range of the values found in the literature (see Data S1). However, the energy of the 1^1B_u state is highly overestimated, which makes this state the 2^1B_u state on the CASSCF level. This results in the 1^1B_u state always lying above the 2^1A_g state. OM2-MRCI places in all polyenes the 1^1B_u state above the 2^1A_g state with a much larger energy difference. Silva-Junior and Thiel (39) found the same state ordering for hexatriene and octatetraene with OM2 and OM3. The energy values show deviations up to 0.7 eV from the computed MS-CASPT2 values, with largest differences in the long decapentaene. Deviations of this magnitude are not uncommon and were also observed by Silva-Junior and Thiel (39) in their benchmark sets.

The state configurations, however, are very similar in the MS-CASPT2 and OM2-MRCI approaches, with the HOMO \rightarrow LUMO configuration dominating the 1^1B_u state and the typical HOMO $^2\rightarrow$ LUMO 2 double excitation combined with HOMO \rightarrow LUMO+1 and HOMO–1 \rightarrow LUMO single excitations in 2^1A_g .

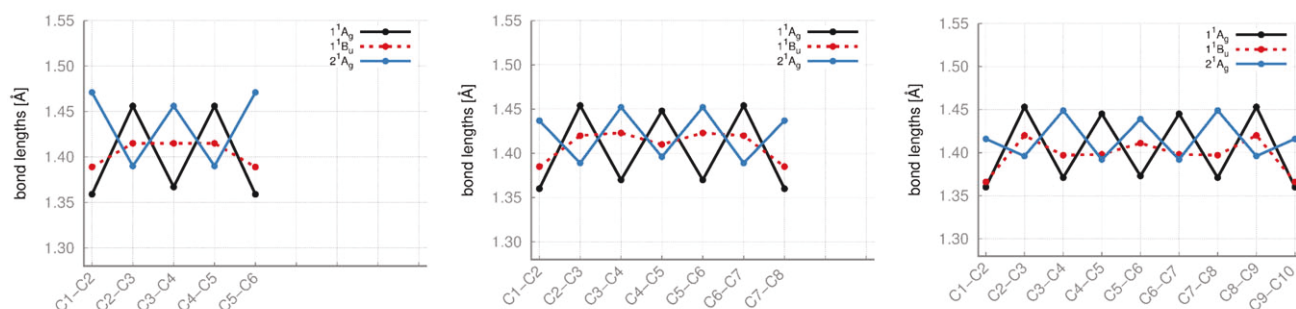
Table 2 lists MS-CASPT2 and OM2-MRCI computed vertical excitation energies to the S_1 and S_2 states of PSB3 to PSB5. The calculated energies deviate from those reported in the literature. Except for PSB5, the computed S_1 and S_2 values in this study are lower in energy, but the literature CASPT2 energies were based on CASSCF-optimized structures. Geometry optimizations with this method (see Data S2) yield the same trends as reported in earlier works. We ascribe the much higher MRCI-Q excitation energies reported by Aquino *et al.* to the small active space (5,6) with only four π -orbitals and one σ -orbital in the reference space. As expected, the bright S_1 state with dominating HOMO \rightarrow LUMO configuration lies below the S_2 state throughout all the calculations. The S_2 state is dark considering MS-CASPT2 calculations and weakly allowed according to the CASSCF results due to the presence of the HOMO \rightarrow LUMO determinant in its configuration (see Data S2). In contrast to the results obtained for the linear polyenes, there were no state ordering problems at the CASSCF level, that is the bright state always corresponds to the S_1 state. The OM2-MRCI computed excitation energies are in remarkable agreement with the MS-CASPT2 obtained data, with less deviations than for the previously discussed linear polyenes.

Excited-state geometries

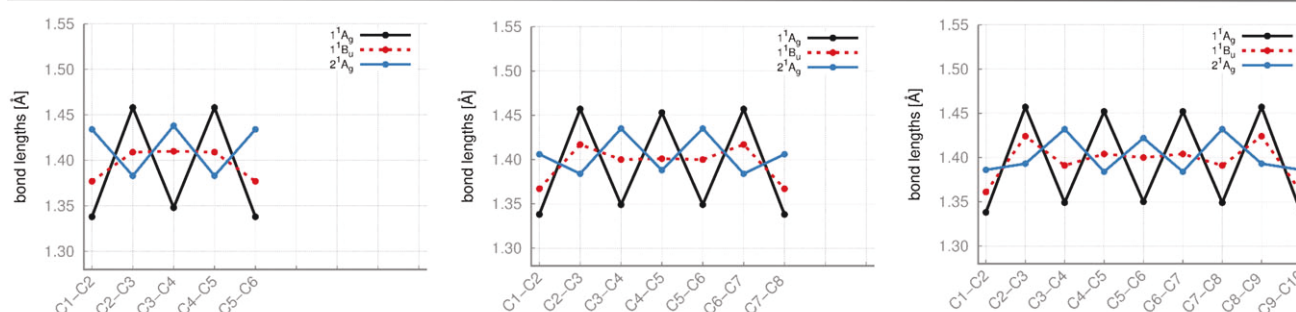
As expected for linear polyenes, relaxation to the bright 1^1B_u state minimum leads to only small changes at the termini and an equalization of the inner single and double bonds with all methods (10). The optically dark 2^1A_g state is bond-inverted with respect to the ground state throughout all performed geometry optimizations (9,48). For the Schiff bases, we obtain different results for the bright state, depending on the used optimization strategy and methods. In the smallest Schiff base, CASSCF optimization in the HOMO \rightarrow LUMO singly excited S_1 state always

Polyenes

MS-CASPT2

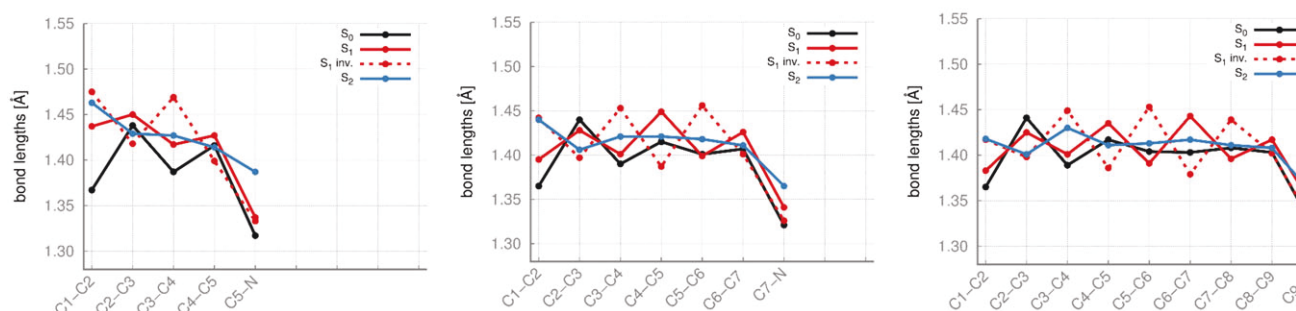


OM2-MRCI



Schiff Bases

MS-CASPT2



OM2-MRCI

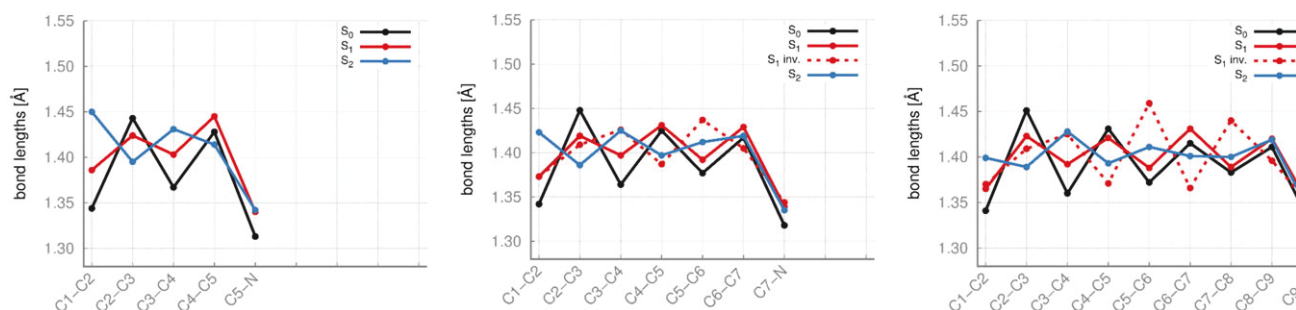


Figure 2. Bond lengths of MS-CASPT2- and OM2-MRCI-optimized linear polyenes (upper panel) and PSBs (lower panel) in ground and excited states.

yields a strongly bond-inverted structure, regardless of the starting geometry. MS-CASPT2 delivers two distinct S_1 minima. Using the S_0 geometry as a starting point, we find a geometry

with a bond length pattern similar to the ground state but with slightly enhanced bond lengths, similar to previous observations (19,22,23). The pre-inverted starting structure leads to a bond-

Table 1. Vertical excitation energies, configuration weights and oscillator strengths of MS-CASPT2- and OM2-MRCI (in parentheses)-optimized 1^1A_g ground-state geometries in computed linear polyenes.

State	Weight [%]	Energy [eV]	f	Lit. [eV]
Hexatriene				
2^1A_g (S_2)				
$H^2 \rightarrow L^2$	34 (33)	5.04	–	5.21*
$H-1 \rightarrow L$	22 (22)	(5.09 S_1)		
$H \rightarrow L+1$	16 (26)			
1^1B_u (S_1)				
$H \rightarrow L$	86 (88)	4.98 (5.40 S_2)	0.93 (1.11)	5.13 † 4.93 ‡
Octatetraene				
2^1A_g (S_1)				
$H^2 \rightarrow L^2$	29 (35)	4.20	–	4.10 §
$H-1 \rightarrow L$	19 (21)	(4.49)		
$H \rightarrow L+1$	16 (24)			
1^1B_u (S_2)				
$H \rightarrow L$	85 (85)	4.27 (4.86)	1.17 (1.44)	4.41 ¶
Decapentaene				
2^1A_g (S_1)				
$H^2 \rightarrow L^2$	29 (36)	3.62	–	3.48 **
$H-1 \rightarrow L$	18 (19)	(4.14)		
$H \rightarrow L+1$	15 (22)			
1^1B_u (S_2)				
$H \rightarrow L$	77 (83)	3.83 (4.50)	1.36 (1.79)	4.02 **

*Values from ref (36), † ref (37), ‡ ref (38), § ref (40), ¶ ref (41), ** ref (42).

Table 2. Vertical MS-CASPT2 and OM2-MRCI (in parentheses) excitation energies and oscillator strengths of MS-CASPT2- and OM2-optimized S_0 geometries in PSB3 to PSB5.

State	Weight [%]	Energy [eV]	f	Lit. [eV]
PSB3				
S_1 ($2A'$)				
$H \rightarrow L$	78 (74)	3.97 (3.84)	1.10 (0.80)	4.73* 4.12 ‡
S_2 ($3A'$)				
$H^2 \rightarrow L^2$	32 (35)	5.15 (4.96)	0.16×10^{-1}	5.59*
$H-1 \rightarrow L$	29 (30)		(0.13)	5.42 †
$H \rightarrow L+1$	14 (17)			
PSB4				
S_1 ($2A'$)				
$H \rightarrow L$	72 (74)	3.28 (3.11)	1.59 (1.09)	4.14* 3.23 §
S_2 ($3A'$)				
$H^2 \rightarrow L^2$	33 (37)	4.05 (4.04)	0.15×10^{-1}	4.81*
$H-1 \rightarrow L$	23 (25)		(0.16)	4.40 §
$H \rightarrow L+1$	14 (18)			
PSB5				
S_1 ($2A'$)				
$H \rightarrow L$	68 (72)	2.84 (2.64)	2.11 (1.38)	2.72 ¶ 2.69 **
S_2 ($3A'$)				
$H^2 \rightarrow L^2$	32 (38)	3.35 (3.48)	0.12×10^{-1}	3.73 ¶
$H-1 \rightarrow L$	20 (23)		(0.19)	3.68 **
$H \rightarrow L+1$	15 (17)			

MRCI+Q energies from ref (43), † CASPT2 values from CASSCF-optimized geometries in ref (44), ‡ CASPT2 values from CASSCF-optimized geometries in ref (45), § CASPT2/6-31G SA3 values from single-state CASSCF-optimized geometry in ref (46), ¶ CASPT2/6-31G* CASSCF-optimized geometry in ref (44)-Data S2, ** CASPT2/6-31G* SA3 values from SA2 CASSCF-optimized geometry in ref (47).

inverted S_1 geometry, where the bond inversion is more pronounced for the C1-C2 and C3-C4 bonds. C2-C3 and the terminal C4-C5 and C-N bonds stay similar to their ground-state

values. In contrast, OM2-MRCI reveals only one minimum with ground-state like bond lengths, which are slightly enhanced close to the NH_2^+ terminus. Also for the longer PSBs, CASSCF only converges to the strongly inverted geometry, while now both MS-CASPT2 and OM2-MRCI result in two minima on the S_1 potential energy surface. The inverted and the noninverted structures have, in comparison with the ground-state geometry, a more regular BLA. The inverted S_1 Schiff base geometries show remarkable resemblance to those of the polyene 2^1A_g state structures. To clarify the relationship between the two distinct PSB- S_1 minima, linearly interpolated paths connecting inverted and non-inverted geometries were calculated. These will be discussed in the next chapter.

All applied methodologies converge to an S_2 geometry with equalized inner bonds (Fig. 2). Moreover, the C-terminal double bond and its neighboring single bond are inverted with respect to the ground state, while the bonds at the N-terminus do not change their length substantially. OM2-MRCI leads to geometries with a smaller number of equalized bonds compared to CASSCF and MS-CASPT2. Overall, the bond equalized S_2 structure exhibits a bond length pattern very similar to the one observed in the optimized 1^1B_u linear polyenes.

A rationale of the change in bond length pattern can be provided through electron difference density plots between S_0 and the excited states. This documents the immediate change in electron distribution upon excitation. Figure 3 provides such plots for optimized ground- and excited-state geometries of PSB3 and hexatriene. The red color denotes reduced electron density, and blue describes enhanced density with respect to the ground-state electronic structure. We will first discuss the situation at the respective ground-state geometry. The hexatriene shows reduced π -density at the center of the double bonds and enhanced π -density at the single bonds. Furthermore, the π -density surrounding the central double bond is enhanced—this observation is in line with an overall bond equalization, as already noted from 1^1B_u geometry optimization. For the Schiff base, at first no evidence for bond length inversion upon S_0 - S_1 excitation is found. The density is rather shifted from the π -orbitals at the C-terminus and the p-type orbital at C4 toward p-type orbitals at N, C3 and C5. This observation agrees well with the results found by Sundholm *et al.* (22) and corresponds to an overall positive charge migration from the N-terminus toward the C-terminus. Starting from this structure, MS-CASPT2 optimization in the S_1 state in fact mostly preserves the ground-state bond alternation pattern, and the system stays trapped in a noninverted configuration.

In the 1^1B_u -optimized geometry, hexatriene shows basically the same electron difference density features as in the ground-state geometry. This is different for PSB3, where a pronounced relocation of electron density from the double bonds C1-C2 and C3-C4 toward the single bonds C2-C3 and C4-C5 is observed, still resulting in positive net charge migration toward the C-terminus. This relocation closely resembles the difference density of the 2^1A_g state in hexatriene, which features bond inversion.

The S_2 difference density of PSB3 shows a balanced situation, where density is transferred equally from p-type orbitals located at C2, C4 and N toward C1, C3 and C5, thus reflecting an equalization of bond lengths in this state, although with a slightly different density pattern as in hexatriene.

In Table 3, we compare MS-CASPT2 and OM2-MRCI energies for the corresponding 2^1A_g - and 1^1B_u -optimized geometries of linear polyenes with various theoretical and experimental data.

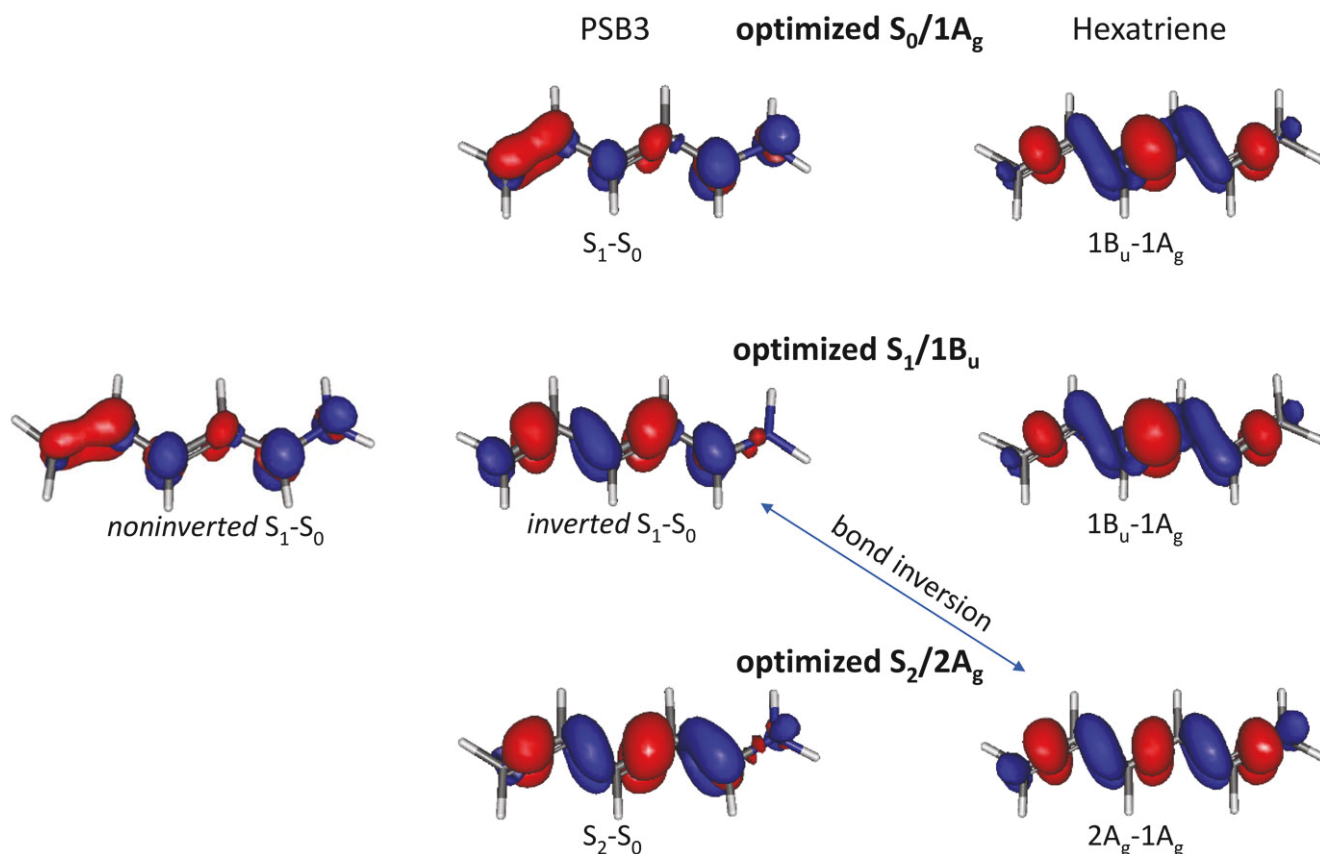


Figure 3. Difference densities of PSB3 (left) and hexatriene (right) at ground- and excited-state optimized geometries based on Molcas Pt2Orb files generated after MS-CASPT2 computation.

Values for 0–0 transitions are available for all studied polyenes from fluorescence excitation (42,49,50) and jet-cooled absorption measurements (38,51,52). Vertical emissions are, however, only reported for octatetraene (53,54). In our calculations, MS-CASPT2 yields adiabatic energies that are somewhat underestimated compared to the experimental 0–0 transitions.

OM2-MRCI performs reasonably well in the computation of vertical and adiabatic energies, with deviations of 0.1–0.4 eV to the experimental data. Vertical excitation energies lie in the range of OM2-MRCI values computed by Silva-Junior and Thiel (39) for MP2-optimized ground-state geometries for hexatriene and octatetraene. Minor deviations can be explained by the differences in optimized ground-state geometries. Consideration of the zero-point vibrational energy (ZPVE) yields a better agreement of adiabatic energies with the experimental values. This has been demonstrated by Tuna *et al.*, (28) who computed OM2-MRCI 0–0 energy transitions for the 1^1B_u of hexatriene and octatetraene with ZPVE correction, perfectly matching the experiment. For hexatriene and octatetraene, the OM2-MRCI energies yield a better match to the experiment than MS-CASPT2. We note, however, slight differences in the state compositions. The typical 2^1A_g excitations have a larger weight with this method.

Several trends can be concluded from the performed MS-CASPT2 and OM2-MRCI calculations:

- 1 the difference between MS-CASPT2 and OM2-MRCI adiabatic and emission energies is smaller for the 2^1A_g state than for the 1^1B_u for all investigated linear polyenes.

- 2 OM2-MRCI energies for the 1^1B_u state are clearly higher than MS-CASPT2 values.

- 3 All estimated energies are in the range of values reported in the literature.

The corresponding data for the Schiff bases PSB3 to PSB5 are summarized in Table 4. Experimental gas-phase data as well as computed values for adiabatic or vertical transitions from the S_1 and S_2 states are not yet available for these compounds.

For the S_1 state of the smallest Schiff base PSB3, MS-CASPT2 yields the same value (3.79 eV) as adiabatic energy for both inverted and noninverted equilibrium geometries. Also for longer PSBs, the adiabatic energies of the two minima remain almost degenerate. In contrast, the vertical emissions differ by at least 0.1 eV. This trend is also reproduced with OM2-MRCI. This method delivers the largest difference in emission for the two PSB5 S_1 structures with a value of 0.34 eV (Table 4). Despite their different geometrical structures, there is no significant difference in the MS-CASPT2 state configurations of the S_1 inverted and noninverted minima, that is they cannot be distinguished solely by their state compositions. The weight of the S_1 HOMO→LUMO excitation in MS-CASPT2 decreases with growing polyene chain. The character of the dark S_2 state is less sensitive to the conjugated chain length.

OM2-MRCI yielded only one S_1 minimum for PSB3. Its adiabatic and emission energies deviate only by 0.11 eV and 0.02 eV from MS-CASPT2, respectively. However, the S_2

Table 3. Adiabatic state energies, configuration weights and vertical emissions of MS-CASPT2- and OM2-MRCI-optimized S₁ and S₂ geometries of linear polyenes. OM2-MRCI computed values are given in parentheses.

State	Weight [%]	ΔE_{adia} [eV]	ΔE_{0-0} [eV] (Lit.)	ΔE_{vert} [eV]	ΔE_{vert} [eV] (Lit.)	f
Hexatriene						
2^1A_g (S ₁)						
H ² →L ²	31 (35)	4.13	4.26 [*]	3.41	3.62 ^{††}	–
H-1→L	22 (35)	(3.98)	4.17 ^{††}	(3.29)		
H→L+1	17 (23)					
1^1B_u (S ₂)						
H→L	81 (86)	4.70	4.93 ^{†,‡}	4.44	4.59 ^{††,†††}	0.8
		(5.07)	4.84 ^{††}	(4.91)		(1.1)
Octatetraene						
2^1A_g (S ₁)						
H ² →L ²	28 (35)	3.39	3.59 [§]	2.79	2.80 ^{††}	–
H-1→L	21 (23)	(3.36)	3.50 ^{††}	(2.77)	3.10 ^{§§}	
H→L+1	17 (24)		3.34 ^{***}		2.94 ^{***}	
1^1B_u (S ₂)						
H→L	81 (83)	4.05	4.41 [¶]	3.81	3.80 ^{††}	1.06
		(4.52)	4.34 ^{††}	(4.39)	4.20 ^{¶¶}	(1.44)
			3.85 ^{***}		3.96 ^{†††}	
Decapentaene						
2^1A_g (S ₁)						
H ² →L ²	26 (34)	2.86	3.10 ^{**}	2.35	2.37 ^{††}	–
H-1→L	21 (22)	(2.98)	2.99 ^{††}	(2.43)		
H→L+1	17 (24)					
1^1B_u (S ₂)						
H→L	76 (79)	3.72	3.98 ^{**}	3.54	3.38 ^{††}	1.34
		(4.15)		(4.02)	3.50 ^{†††}	(1.81)

^{*}Fluorescence excitation of jet-cooled cis-1,3,5-hexatriene in ref (49), [†]absorption of isolated hexatriene in ref (51), [‡]absorption of jet-cooled hexatriene in ref (52), [§]fluorescence excitation of jet-cooled octatetraene in ref (50), [¶]absorption of jet-cooled octatetraene in ref (38), ^{**}0-0 band in various solvents in ref (42), ^{††}origin of gas-phase absorption in ref (42), ^{†††}MRMP corrected 0-0 transitions in ref (55), ^{§§}in ref (53), ^{¶¶}in ref (54), ^{***}MS-CASPT2 (PT2F) energies in ref (15), ^{†††}DFT/MRCI calculated vertical emission in ref (10), ^{††††}due to convergence problems, the 1^1B_u structure for decapentaene was obtained from a single-state CASPT2 optimization. Experimental values are denoted in bold font.

energies differ by 0.3 eV. The deviation of the S₂ energies becomes smaller for longer PSBs, while for S₁ energies an opposite trend can be found.

Except for PSB5, OM2-MRCI also reproduces MS-CASPT2 S₁ state compositions remarkably well. For the longest PSB, we note larger differences in the weights of the HOMO→LUMO excitation in S₁. In the inverted structure, the weight of this configuration is 9% lower than for MS-CASPT2. Also for S₂, some deviations between MS-CASPT2 and OM2-MRCI can be observed. For example, the HOMO²→LUMO² double excitation has a 7–9% larger weight in OM2-MRCI. Furthermore, in OM2-MRCI the order of the H→L+1 and H-1→L configuration is inverted, that is H→L+1 becomes the dominant configuration in S₂ with this method. For PSB3, the corresponding configurations are equally represented. In accordance with MS-CASPT2, the overall composition of the S₂ state does not change with the chain length.

Interpolated paths for PSB S₁

To analyze the relationship between the two S₁ minima found by MS-CASPT2 and OM2-MRCI, we performed linearly interpolated scans which connect the structures. Interpolation was performed in Cartesian coordinates within 20 steps, starting from the bond-inverted structure. Figure 4 shows the results of these interpolations for the PSB4 and PSB5 systems. The computed values for PSB3 can be found in the Data S2.

For all PSBs, MS-CASPT2 separates the almost iso-energetic inverted and noninverted minima by a small barrier of

ca. 0.1–0.2 eV. The CASSCF reference energies in the MS-CASPT2 computation always yield a lower energy for the inverted minimum (see Data S2), that is dynamic correlation tends to favor the noninverted geometry. Considering the composition of the reference CASSCF wave function (see Data S2) provides a better understanding of the aforementioned stabilizing effect. The inverted S₁ minimum exhibits a larger fraction of the polyenic 2^1A_g -like configuration including GS, H-1→L, H→L+1 and H²→L² excitations, and a lower amount of the 1^1B_u -like HOMO→LUMO contribution. In contrast to the HOMO→LUMO excitation, the states characterized by excitations typical for polyenic A_g symmetry are already sufficiently described by CASSCF, resulting in minor energy lowering by subsequent MS-CASPT2 calculation. Consequently, the states with more pronounced HOMO→LUMO character are stabilized more, leading to almost iso-energetic inverted and noninverted minima.

OM2-MRCI delivers a barrierless path for PSB4, that is the inverted structure appears as a transition state and is ca. 0.02 eV higher in energy. In PSB5, both S₁ minima are almost iso-energetic, similar to MS-CASPT2, but the barrier separating these structures is lower.

The impact of the two S₁ minima on the photoreactivity of PSBs was probed by computing unrelaxed PES scans along single- and double-bond torsions of PSB5 (see Data S2) using the S₁-optimized geometry as starting point. Figure 5 depicts MS-CASPT2 and OM2-MRCI computed torsion profiles along the central C5-C6 bond of PSB5 for the inverted and noninverted S₁ structures. In our static MS-CASPT2 computations,

Table 4. Adiabatic state energies, configuration weights and vertical emissions of MS-CASPT2- and OM2-MRCI-optimized S_1 and S_2 geometries of PSB3 to PSB5. OM2-MRCI values are given in parentheses.

State	Weight [%]	ΔE_{adia} [eV]	ΔE_{vert} [eV]	f
PSB3				
S_1 ($2A'$)				
H \rightarrow L	76 (76)	3.79 (3.68)	3.56 (3.58)	1.06 (0.83)
S_1 inv. ($2A'$)				
H \rightarrow L	76 (–)	3.79 (–)	3.40 (–)	1.09 (–)
S_2 ($3A'$)				
H $^2\rightarrow$ L 2	33 (40)	4.59	4.18	0.02
H-1 \rightarrow L	24 (24)	(4.28)	(3.88)	(0.02)
H \rightarrow L+1	15 (24)			
PSB4				
S_1 ($2A'$)				
H \rightarrow L	70 (72)	3.16 (2.98)	3.02 (2.92)	1.50 (1.12)
S_1 inv. ($2A'$)				
H \rightarrow L	71 (72)	3.16 (3.00)	2.88 (2.80)	1.56 (1.23)
S_2 ($3A'$)				
H $^2\rightarrow$ L 2	34 (41)	3.70	3.42	0.02
H-1 \rightarrow L	21 (21)	(3.46)	(3.17)	(0.03)
H \rightarrow L+1	15 (24)			
PSB5				
S_1 ($2A'$)				
H \rightarrow L	65 (69)	2.73 (2.53)	2.61 (2.48)	1.95 (1.4)
S_1 inv. ($2A'$)				
H \rightarrow L	67 (58)	2.75 (2.52)	2.51 (2.14)	2.01 (1.26)
S_2 ($3A'$)				
H $^2\rightarrow$ L 2	32 (41)	3.09	2.88	0.024
H-1 \rightarrow L	18 (20)	(2.95)	(2.71)	(0.03)
H \rightarrow L+1	15 (24)			

only rotation among this bond leads to a crossing of the S_1 and S_0 states, thus indicating radiationless and efficient photoisomerization. In the inverted structure scan, the S_1 PES swiftly rises until ca. 140° by ca. 0.09 eV, and then decreases until at ca. 90° the S_1/S_0 surface crossing is reached. With decreasing dihedral angle, the adiabatic energy of the covalent S_2 state stays well separated from S_1 . An S_1/S_0 crossing is also reached when starting from the noninverted structure, and the increase in S_1 energy is, however, considerably larger. The barrier height along this state is ca. 0.4 eV, and the S_1/S_2 energy

difference decreases until ca. 150° . Not surprisingly, all double-bond torsions have increased barriers for the noninverted case, and in all scans, we note a rise in S_0 energy, which is considerably pronounced for double-bond rotations (see Data S2).

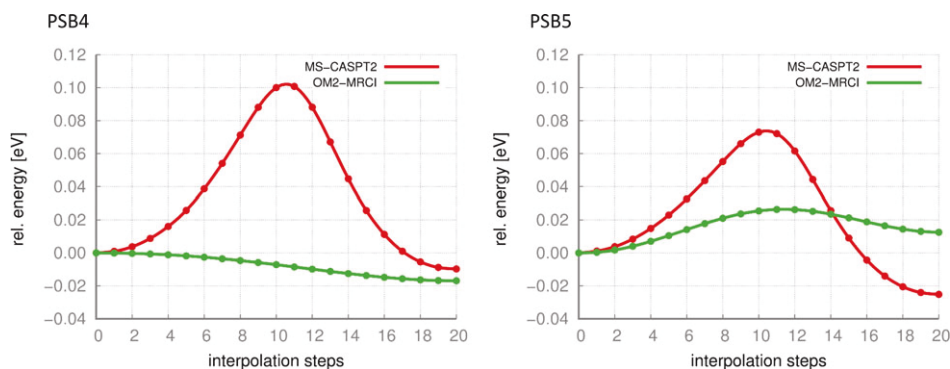
The unrelaxed OM2-MRCI scan required a larger active space (20 electrons in 20 orbitals) compared to previous geometry optimizations. Consideration of exclusively π -electrons and π -orbitals led to discontinuities of the PES (see Data S2) at highly twisted geometries (130 – 85°). OM2-MRCI resembles the MS-CASPT2 results considerably well. Starting with the inverted structure, the S_1 torsion path leads without any barrier to the S_1/S_0 crossing at 90° . For the noninverted minimum, the crossing point is approached after overcoming a barrier of ca. 0.15 eV at 145° .

CONCLUSIONS

In the present work, we have characterized the ground state and two low-lying excited states of the protonated all-*trans* Schiff base models PSB3 to PSB5 and the corresponding linear polyenes with a high-level MS-CASPT2 approach. Furthermore, we have evaluated the performance of OM2-MRCI for the investigated π -conjugated systems. We have shown that excited-state properties of PSBs differ considerably from linear polyenes. The most prominent differences refer to the excited-state geometries of both compounds. For the HOMO \rightarrow LUMO S_1 equilibrium geometry of all PSBs, CASSCF always yields a bond-inverted minimum relative to the ground state. With MS-CASPT2, we found two minima on the S_1 PES, one with an inverted and one with a noninverted S_1 equilibrium geometry. Bond-inverted structures at this level could only be located by manually modifying the geometries into this direction prior to optimization. The difference between CASSCF and MS-CASPT2 results from a lack of dynamic correlation in the CASSCF approach. As a result, the noninverted geometry becomes more stabilized by MS-CASPT2 and is almost equal to the energy of the inverted minimum.

The two minima are connected by a small barrier of ca. 0.1 eV, an energy value which may easily be overcome by kinetic energy when the system is excited from its electronic ground state. The existence of two minima rather than a single one may change the topology of the PES substantially, suggesting a different S_1/S_0 relaxation path.

Filippi and Valsson (23) investigated *cis*-PSBs and found two CASPT2 minima only for the methylated *cis*-PSB5, and hence,

**Figure 4.** MS-CASPT2 and OM2-MRCI interpolated paths for S_1 inverted and noninverted equilibrium geometries. The left minimum is assigned to the inverted and the right minimum to the noninverted geometry.

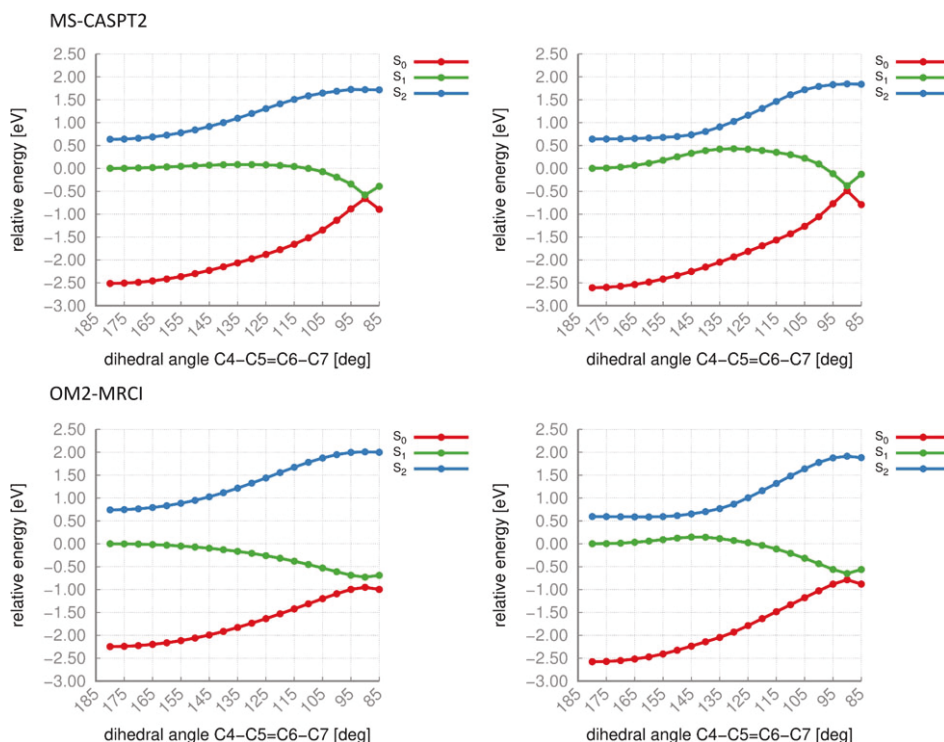


Figure 5. Torsion profiles along the central double bond of PSB5, starting with inverted (left) and noninverted (right) S_1 minima.

the occurrence of the bond-inverted structure was interpreted as an artifact. Similar observations for *cis*-PSB3 have also been documented by Keal *et al.* (56) and also by Page and Olivucci (19). However, as we consequently observe two minima for all investigated *trans*-PSBs with MS-CASPT2 and for PSB4-PSB5 with OM2-MRCI, we have no reason to suppose one of the minima to be a computational or methodological artifact. The results of Filippi *et al.* rather indicate that at least at the (MS-)CASPT2 level of theory, the role of *cis*-PSB3 and *cis*-PSB4 as minimal models for the 11-*cis* retinal chromophore should be critically reconsidered.

In the polyenes, bond inversion is only observed in the dark 2^1A_g state featuring $HOMO^2 \rightarrow LUMO^2$ double excitation. 1^1B_u excitation consequently and for all methods leads to an equalization of the inner bonds. Therefore, the terminology “ 1^1B_u -like” for PSB is somewhat misleading, as it only refers to the type of excitation and the brightness of the optical transition but not to the immediate structural consequences. Bond equalization, although with a different density pattern, is observed in the “dark” S_2 state of PSBs.

The OM2-MRCI method is able to reproduce the main trends for the well-studied linear polyenes and also for the longer PSBs investigated in this study. This method therefore offers a computationally inexpensive and valuable alternative for the investigation of large polyenic systems, for example in a QM/MM environment or in excited-state dynamics.

Acknowledgements—The authors thank Dr. Marko Schreiber and Prof. Marco Garavelli for inspiring discussions on the topic. I. D. is grateful for financial support through “Strategischer Forschungsfonds” of the Heinrich-Heine-University Düsseldorf, project no. F 2013 – 442-9. Further financial support by Deutsche Forschungsgemeinschaft (DFG)

through funds (INST 208/704-1 FUGG) to purchase the hybrid computer cluster used in this study is gratefully acknowledged.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Data S1. Linear Polyenes.

Table S1. Literature values for CASPT2 vertical excitation energies for the ground state of linear polyenes.

Table S2. CASSCF and MS-CASPT2 vertical excitation energies calculated for MP2, CASSCF and MS-CASPT2 optimized 1^1A_g geometries of linear polyenes.

Table S3. CASSCF reference excited state compositions of 1^1A_g optimized MS-CASPT2 geometries in linear polyenes.

Figure S1. Bond lengths of CASSCF optimized ground and excited state geometries in linear polyenes.

Figure S2. MS-CASPT2 unrelaxed torsion scan for decapentaene.

Figure S3. Bond lengths of OM3-MRCI optimized ground and excited state geometries in linear polyenes.

Data S2. Protonated Schiff bases.

Table S4. Literature values for vertical excitation energies for the optimized ground state of protonated Schiff bases PSB3 to PSB4 from Supporting Information ref. (3).

Table S5. CASSCF and MS-CASPT2 vertical excitation energies and oscillator strengths calculated for MP2, CASSCF and MS-CASPT2 optimized 1^1A_g geometries of protonated Schiff bases.

Table S6. CASSCF excited state compositions of S_0 optimized geometries in protonated Schiff bases.

Table S7. CASSCF reference and MS-CASPT2 excited state compositions for MS-CASPT2 S_1 and S_2 equilibrium geometries.

Figure S4. Bond lengths of CASSCF optimized ground and excited state geometries of protonated Schiff bases.

Figure S5. Bond lengths of OM3-MRCI optimized ground and excited state geometries of protonated Schiff bases.

Figure S6. CASSCF reference energies for MS-CASPT2 linear interpolated paths between inverted and noninverted S_1 geometries.

Figure S7. MS-CASPT2 interpolated path for S_1 inverted and noninverted equilibrium geometries of PSB3.

Figure S8. Retinal Schiff base.

Figure S9. MS-CASPT2 unrelaxed dihedral angle scans for optimized inverted and non-inverted PSB5 geometries.

Figure S10. OM2-MRCI unrelaxed torsion scan of optimized and inverted PSB5 geometry.

Table S8. xyz-coordinates of optimized minima in Cartesian coordinates (Å).

REFERENCES

- Polivka, T. and V. Sundstrom (2004) Ultrafast dynamics of carotenoid excited states – From solution to natural and artificial systems. *Chem. Rev.* **104**, 2021–2071.
- Litman, B. J. and D. C. Mitchell (1996) Rhodopsin structure and function. *Biomembranes*, **2**, 1–32.
- Stoeckenius, W., R. H. Lozier and R. A. Bogomolni (1979) Bacteriorhodopsin and the purple membrane of halobacteria. *BBA Rev. Bioenerg.* **505**, 215–278.
- Nagel, G., D. Ollig, M. Fuhrmann, S. Kateriya, A. M. Musti, E. Bamberg and P. Hegemann (2002) Channelrhodopsin-1: a light-gated proton channel in green algae. *Science* **296**, 2395–2398.
- Mohseni, M., Y. Omar, G. S. Engel and M. B. Plenio (2014) *Quantum Effects in Biology*. Cambridge University Press, Cambridge, UK.
- Chattopadhyay, A. (2012) Comparative study of spectroscopic properties of the low-lying electronic states of 2,4-pentadien-1-iminium cation and its N-substituted analogues. *J. Chem. Sci.* **124**, 985–994.
- Cembran, A., F. Bernardi, M. Olivucci and M. Garavelli (2005) The retinal chromophore/chloride ion pair: structure of the photoisomerization path and interplay of charge transfer and covalent states. *Proc. Natl Acad. Sci. USA* **102**, 6255–6260.
- Robb, M. A., M. Garavelli, M. Olivucci and F. Bernardi (2000) A Computational Strategy for Organic Photochemistry, in *Reviews of Computational Chemistry*. Vol. 15, (Edited by K. B. Lipkowitz and D. B. Boyd Lipkowitz), pp. 87–146. John Wiley & Sons, Inc., Hoboken.
- Kurashige, Y., H. Nakano, Y. Nakao and K. Hirao (2004) The $\pi \rightarrow \pi^*$ excited states of long linear polyenes studied by the CASCI-MRMP method. *Chem. Phys. Lett.* **400**, 425–429.
- Marian, C. M. and N. Gilka (2008) Performance of the density functional theory/multireference configuration interaction method on electronic excitation of extended π -systems. *J. Chem. Theory Comput.* **4**, 1501–1515.
- Blomgren, F. and S. Larsson (2003) Bond length alternation in ground and HOMO-LUMO excited states in polyenes. Dynamic Stokes shift? *Theor. Chem. Acc.* **110**, 165–169.
- Garavelli, M., F. Negri, M. Olivucci, C. G. Ciamician, V. F. Selmi and V. Uni (1999) 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. *J. Am. Chem. Soc.* **121**, 1023–1029.
- Schapiro, I., M. N. Ryazantsev, L. M. Frutos, N. Ferré, R. Lindh and M. Olivucci (2011) The ultrafast photoisomerizations of rhodopsin and bathorhodopsin are modulated by bond length alternation and HOOP driven electronic effects. *J. Am. Chem. Soc.* **133**, 3354–3364.
- Schapiro, I., M. N. Ryazantsev, W. J. Ding, M. M. Huntress, F. Melaccio, T. Andruniow and M. Olivucci (2010) Computational photobiology and beyond. *Aust. J. Chem.* **63**, 413–429.
- Angeli, C. and M. Pastore (2011) The lowest singlet states of octatetraene revisited. *J. Chem. Phys.* **134**, 1–10.
- Knecht, S., C. M. Marian, J. Kongsted and B. Mennucci (2013) On the photophysics of carotenoids: a multireference dft study of peridinin. *J. Phys. Chem. B* **117**, 13808–13815.
- Hsu, C.-P., S. Hirata and M. Head-Gordon (2001) Excitation energies from time-dependent density functional theory for linear polyene oligomers: butadiene to decapentaene. *J. Phys. Chem. A* **105**, 451–458.
- Christensen, R. L., M. G. I. Galinato, E. F. Chu, J. N. Howard, R. D. Broene and H. A. Frank (2008) Energies of low-lying excited states of linear polyenes. *J. Phys. Chem. A* **112**, 12629–12636.
- Page, C. S. and M. Olivucci (2003) Ground and excited state CASPT2 geometry optimizations of small organic molecules. *J. Comput. Chem.* **24**, 298–309.
- Liu, L., J. Liu and T. J. Martinez (2016) Dynamical correlation effects on photoisomerization: Ab initio multiple spawning dynamics with MS-CASPT2 for a model trans-protonated schiff base. *J. Phys. Chem. B* **120**, 1940–1949.
- Tuna, D., D. Lefrançois, Ł. Wolański, S. Gozem, I. Schapiro, T. Andruniow, A. Dreuw and M. Olivucci (2015) Assessment of approximate coupled-cluster and algebraic-diagrammatic-construction methods for ground- and excited-state reaction paths and the conical-intersection seam of a retinal-chromophore model. *J. Chem. Theory Comput.* **11**, 5758–5781.
- Send, R., D. Sundholm, M. P. Johansson and F. Pawłowski (2009) Excited State potential energy surfaces of polyenes and protonated Schiff bases. *J. Chem. Theory Comput.* **5**, 2401–2414.
- Valsson, O. and C. Filippi (2010) Photoisomerization of model retinal chromophores: insight from quantum monte carlo and multiconfigurational perturbation theory. *J. Chem. Theory Comput.* **6**, 1275–1292.
- Conti, I. and M. Garavelli (2007) Substituent-controlled photoisomerization in retinal chromophore models: fluorinated and methoxy-substituted protonated Schiff bases. *J. Photochem. Photobiol. A Chem.* **190**, 258–273.
- Laricheva, E. N., S. Gozem, S. Rinaldi, F. Melaccio, A. Valentini and M. Olivucci (2012) Origin of fluorescence in 11-cis locked bovine rhodopsin. *J. Chem. Theory Comput.* **8**, 2559–2563.
- El-Tahawy, M. M. T., A. Nenov and M. Garavelli (2016) Photoelectrochromism in the retinal protonated schiff base chromophore: photoisomerization speed and selectivity under a homogeneous electric field at different operational regimes. *J. Chem. Theory Comput.* **12**, 4460–4475.
- Dral, P. O., X. Wu, L. Spörkel, A. Kosłowski, W. Weber, R. Steiger, M. Scholten and W. Thiel (2016) Semiempirical quantum-chemical orthogonalization-corrected methods: theory, implementation, and parameters. *J. Chem. Theory Comput.* **12**, 1082–1096.
- Tuna, D., Y. Lu, A. Kosłowski and W. Thiel (2016) Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks of electronically excited states. *J. Chem. Theory Comput.* **12**, 4400–4422.
- Dral, P. O., X. Wu, L. Spörkel, A. Kosłowski and W. Thiel (2016) Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks for ground-state properties. *J. Chem. Theory Comput.* **12**, 1097–1120.
- Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, R. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, I. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox *Gaussian 09 Revision A.02*, Gaussian, Inc., Wallingford, CT.
- Aquilante, F., J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez, N. Galván, L. M. Ferré, L. Frutos, M.

- Gagliardi, A., Garavelli, C. E., Giussani, G., Li, Hoyer, H., Manni, D., Lischka, P., Å. Ma, T., Malmqvist, A., Müller, M., Nenov, T. B., Olivucci, D., Pedersen, F., Peng, B., Plasser, M., Pritchard, I., Reiher, I., Rivalta, J., Schapiro, M., Segarra-Martí, D. G., Stenrup, L., Truhlar, A., Ungur, S., Valentini, V., Vancollie, V. P., Veryazov, O., Vysotskiy, F., Zapata, Weingart and R. Lindh (2016) Molcas 8: new capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J. Comput. Chem.* **37**, 506–541.
32. Weber, W. (1996) *Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Implementierung, Parametrisierung und Anwendungen*, Ph.D. Thesis, Universität Zürich, Hartung-Gorre Verlag.
 33. Weber, W. and W. Thiel (2000) Orthogonalization corrections for semiempirical methods. *Theor. Chim. Acta* **103**, 495–506.
 34. Scholten, M. (2003) *Semiempirische Verfahren mit Orthogonalisierungskorrekturen: Die OM3 Methode*, Ph.D. Thesis.
 35. Schreiber, M., M. R. Silva-Junior, S. P. A. Sauer and W. Thiel (2008) Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* **128**, 134110.
 36. Fujii, T., M. Shimizu and S. Maeda (1985) Two-photon absorption study of 1,3,5-hexatriene by CARS and CSRS. *Chem. Phys. Lett.* **115**, 369–372.
 37. Mosher, O. A., W. M. Flicker and A. Kuppermann (1977) Low energy, variable angle electron-impact excitation of the 1,3,5-hexatriene. *Chem. Phys. Lett.* **45**, 492–497.
 38. Leopold, D. G., R. D. Pendley, J. L. Roebber, R. J. Hemley and V. Vaida (1984) Direct absorption spectroscopy of jet-cooled polyenes. II. The $1^1B_u^+ \leftarrow 1^1A_g^-$ transitions of butadienes and hexatrienes. *J. Chem. Phys.* **81**, 4218–4229.
 39. Silva-Junior, M. R. and W. Thiel (2010) Benchmark of electronically excited states for OM3, INDO/S, and INDO/S2. *Excit. States* **6**, 1546–1564.
 40. McDiarmid, R. (1999) On the electronic spectra of small linear polyenes. *Adv. Chem. Phys.* **110**(110), 177–214.
 41. Heimbrook, L. A., B. E. Kohler and I. J. Levy (1984) Fluorescence from the $1^1B_u^+$ state of trans,trans-1,3,5,7-octatetraene in a free jet. *J. Chem. Phys.* **81**, 1592–1597.
 42. Amico, K. L. D., C. Manos and R. L. Christensen (1980) Electronic energy levels in a homologous series. *J. Am. Chem. Soc.*, **102**, 1777–1782.
 43. Aquino, A. J. A., M. Barbatti and H. Lischka (2006) Excited-state properties and environmental effects for protonated schiff bases: a theoretical study. *Chem. Phys. Chem.* **7**, 2089–2096.
 44. Luk, H. L., F. Melaccio, S. Rinaldi, S. Gozem and M. Olivucci (2015) Molecular bases for the selection of the chromophore of animal rhodopsins. *Proc. Natl Acad. Sci.* **112**, 15297–15302.
 45. Garavelli, M. and P. Celani (1997) The $C_5H_6NH_2^+$ protonated Schiff base: an ab initio minimal model for retinal photoisomerization. *J. Am. Chem. Soc.* **7863**, 6891–6901.
 46. Cembran, A., F. Bernardi, M. Olivucci and M. Garavelli (2004) Counterion controlled photoisomerization of retinal chromophore models: a computational investigation. *J. Am. Chem. Soc.* **126**, 16018–16037.
 47. de Vico, L., C. S. Page, M. Garavelli, F. Bernardi, R. Basosi and M. Olivucci (2002) Reaction path analysis of the ‘tunable’ photoisomerization selectivity of free and locked retinal chromophores. *J. Am. Chem. Soc.* **124**, 4124–4134.
 48. Fuß, W., Y. Haas and S. Zilberg (2000) Twin states and conical intersections in linear polyenes. *Chem. Phys.* **259**, 273–295.
 49. Petek, H., A. J. Bell, R. L. Christensen and K. Yoshihara (1992) Fluorescence excitation spectra of the S_1 states of isolated trienes. *J. Chem. Phys.* **96**, 2412–2415.
 50. Pfanstiel, J. F. and D. W. Pratt (1999) Rotationally resolved electronic spectra of trans, trans-octatetraene and its derivatives. *J. Phys. Chem. A* **103**, 2337–2347.
 51. Gavin, R. M., S. Risemberg and S. A. Rice (1973) Spectroscopic properties of polyenes. I. The lowest energy allowed singlet-singlet transition for cis- and trans-1,3,5-hexatriene. *J. Chem. Phys.* **58**, 3160–3165.
 52. Leopold, D. G., V. Vaida and M. F. Granville (1984) Direct absorption spectroscopy of jet-cooled polyenes. I. The $1^1B_u^+ \leftarrow 1^1A_g^-$ transition of trans,trans-1,3,5,7-octatetraene. *J. Chem. Phys.* **81**, 4210–4217.
 53. Bouwman, W. C., A. C. Jones, D. Phillips, P. Thibodeau, C. Friel and R. L. Christensen (1990) Fluorescence of gaseous tetraenes and pentaenes. *J. Phys. Chem.*, **94**, 7429–7434.
 54. Gavin, R. M., C. Weisman, J. K. McVey and S. A. Rice (1978) Spectroscopic properties of polyenes. III. 1,3,5,7-Octatetraene. *J. Chem. Phys.* **68**, 522–529.
 55. Nakayama, K., H. Nakano and K. Hirao (1998) Theoretical study of the $\pi \rightarrow \pi^*$ excited states of linear polyenes: the energy gap between $1B_u$ and $2A_g$ states and their character. *Int. J. Quantum Chem.* **66**, 157–175.
 56. Keal, T. W., M. Wanko and W. Thiel (2009) Assessment of semiempirical methods for the photoisomerisation of a protonated Schiff base. *Theor. Chem. Acc.* **123**, 145–156.

Supporting Information for:

New Perspectives On An Old Issue: A Comparative MS-CASPT2 and OM2-MRCI Study of Polyenes and Protonated Schiff Bases

Irina Dokukina, Christel M. Marian, Oliver Weingart

1. Linear Polyenes

Table S1 lists calculated CASPT2 vertical excitation energies for the optimised ground state structures of linear polyenes. Listed values were calculated based on MP2 geometries for hexatriene and octatetraene [1] and experimental geometry parameters for decapentaene [2]. CASPT2 computes 1^1B_u as the lowest excited state in hexatriene. In octatetraene, 2^1A_g is slightly below 1^1B_u . For decapentaene both excited states are almost isoenergetic. MS-CASPT2 values for the MS-CASPT2 optimised geometries obtained in the current work (see main article) are comparable to CASPT2 results from the literature, although some differences can be noticed. For instance, no degeneracy for 2^1A_g and 1^1B_u in decapentaene was found, furthermore the calculated energies are slightly lower than in the literature.

Table S1. Literature values for CASPT2 vertical excitation energies for the ground state of linear polyenes.

CASPT2/eV	
<u>Hexatriene</u>	
2^1A_g	5.42 ^a
1^1B_u	5.31 ^a
<u>Octatetraene</u>	
2^1A_g	4.64 ^a
1^1B_u	4.70 ^a
<u>Decapentaene</u>	
2^1A_g	3.95 ^b
1^1B_u	3.97 ^b

^aCASPT2 based on (6,6) CASSCF and MP2/6-31G* geometry from ref [1], CASPT2 based on (10,10) CASSCF and experimental geometry parameters from ref. [2]

These discrepancies result from different techniques to obtain the optimised geometries, which were then used for vertical energy calculations. Values for the geometries optimised with the same methodology (except for the basis set) can be found in Table S2 and are in good agreement with the literature data.

In Table S2 the CASSCF and MS-CASPT2 vertical excitation energies calculated in the current work for equilibrium CASSCF, MP2 and MS-CASPT2 ground state geometries of linear polyenes can be found. Here it becomes apparent that 2^1B_u CASSCF excitation energies, which correspond to MS-CASPT2 1^1B_u , are highly overestimated compared to MS-CASPT2 and experimental values, which can be found in Table 1 in the main article. MS-CASPT2 excitation energies for all three ground state optimisation approaches are in a similar range.

Table S2. CASSCF and MS-CASPT2 vertical excitation energies calculated for MP2, CASSCF and MS-CASPT2 optimised 1^1A_g geometries of linear polyenes.

Geometries	State	CASSCF/eV	f	MS-CASPT2	f
<u>Hexatriene</u>					
CASSCF	2^1A_g	5.44		5.17	
	2^1B_u	7.37	1.39	5.05	0.95
MP2.	2^1A_g	5.36		5.11	
	1^1B_u	7.34	1.38	5.00	0.94
MS-CASPT2	2^1A_g	5.29		5.04	
	1^1B_u	7.30	1.39	4.98	0.94
<u>Octatetraene</u>					
CASSCF	2^1A_g	4.77		4.39	
	2^1B_u	6.78	1.84	4.37	1.19
MP2	2^1A_g	4.63		4.27	
	1^1B_u	6.72	1.83	4.30	1.17
MS-CASPT2	2^1A_g	4.55		4.20	
	1^1B_u	6.69	1.84	4.27	1.17
<u>Decapentaene</u>					
CASSCF	2^1A_g	4.33		3.87	
	2^1B_u	6.34	2.21	3.96	1.38
MP2	2^1A_g	4.11		3.69	
	1^1B_u	6.25	2.20	3.86	1.36
MS-CASPT2	2^1A_g	4.01		3.62	
	1^1B_u	6.21	2.20	3.83	1.36

In Table S3 CASSCF state compositions of CASSCF reference wave function corresponding to MS-CASPT2 ground state optimised geometries of linear polyenes are shown. The estimated configurations agree well with those from MS-CASPT2 computations (see Table 1 in the main article).

Table S3. CASSCF reference excited state compositions of 1^1A_g optimised MS-CASPT2 geometries in linear polyenes.

Geometry	Configurations (weights in %)
Hexatriene	
1^1A_g	GS (82)
2^1A_g	$H^2 \rightarrow L^2$ (34), $H-1 \rightarrow L$ (22), $H \rightarrow L+1$ (16)
2^1B_u	$H \rightarrow L$ (85)
Octatetraene	
1^1A_g	GS (74)
2^1A_g	$H^2 \rightarrow L^2$ (29), $H-1 \rightarrow L$ (20), $H \rightarrow L+1$ (16)
2^1B_u	$H \rightarrow L$ (85)
Decapentaene	
1^1A_g	GS (69)
2^1A_g	$H^2 \rightarrow L^2$ (29), $H-1 \rightarrow L$ (18), $H \rightarrow L+1$ (15)
2^1B_u	$H \rightarrow L$ (77)

CASSCF bond lengths

Figure S1 displays bond lengths of CASSCF optimised ground state and excited state linear polyene geometries. For all polyenes, the ground state structures show a pronounced bond length alternation. The 1^1B_u geometry optimisation results, as expected, in bond length equalisation of the inner bonds and for 2^1A_g in bond length inversion. At this point, CASSCF qualitatively reproduces the bond lengths patterns obtained with MS-CASPT2 and OM2-MRCI.

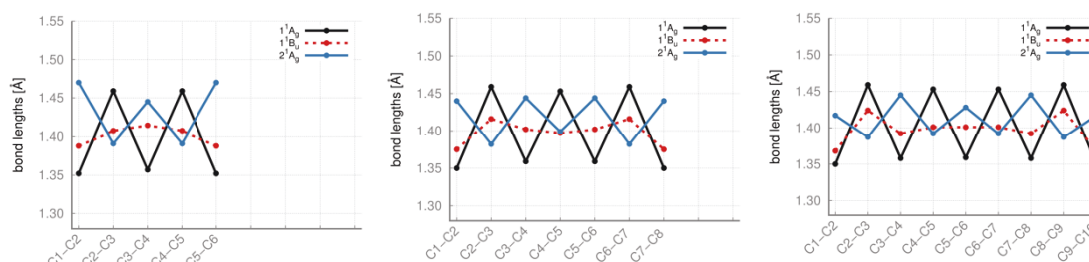


Figure S1. Bond lengths of CASSCF optimised ground and excited state geometries in linear polyenes.

Unrelaxed torsion potentials

Unrelaxed MS-CASPT2 torsion potentials were calculated for decapentaene in the bright 1^1B_u state. In order to follow the 1^1B_u excitation 5 roots were used in the state averaging procedure. As expected for linear polyenes, no S_1/S_0 crossing was detected.

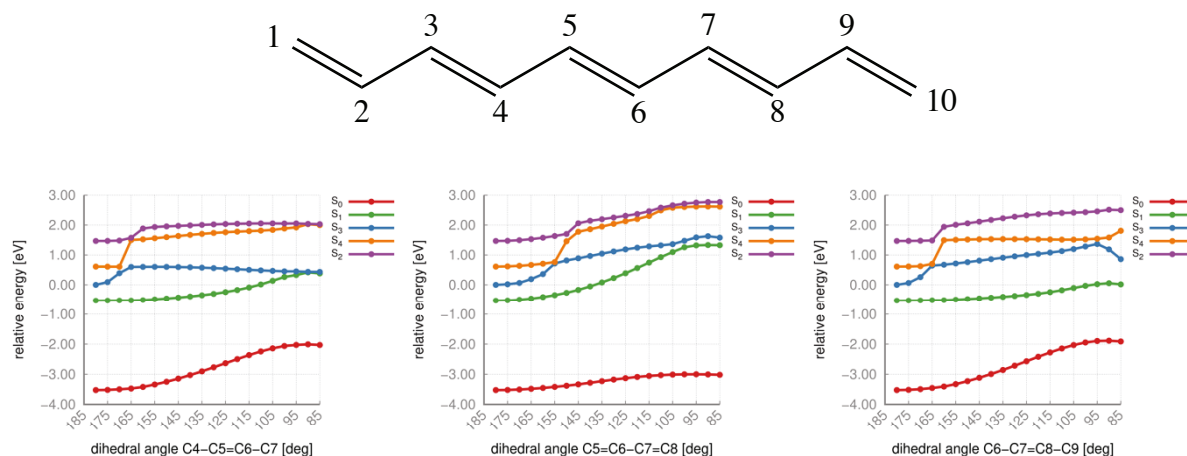


Figure S2. MS-CASPT2 unrelaxed torsion scan for decapentaene.

OM3-MRCI

In addition to OM2-MRCI geometry optimisations, the OM3-MRCI approach was applied. In Figure S3 bond lengths of OM3-MRCI optimised geometries are plotted. Ground state geometries were optimised with OM3 approach. Calculated bond lengths do not differ much from OM2-MRCI. The resulting bond length patterns are identical to those of OM2-MRCI.

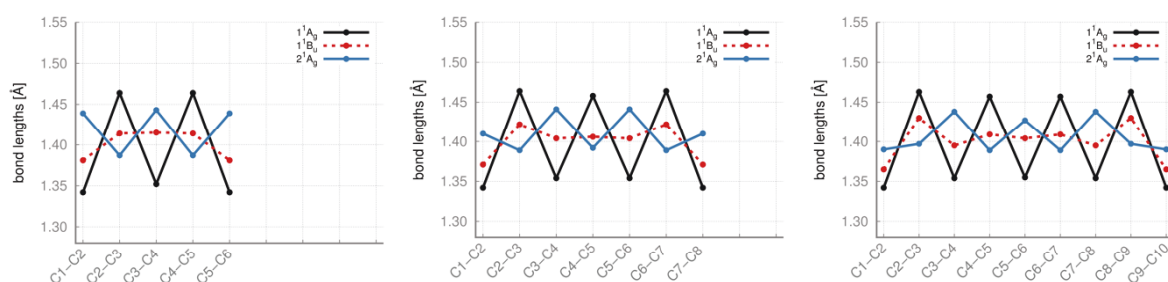


Figure S3. Bond lengths of OM3-MRCI optimised ground and excited state geometries in linear polyenes.

2. Protonated Schiff bases

Table S4 lists the calculated vertical excitation energies for protonated Schiff bases PSB3-PSB4 from ref. [3]. SA-4-CAS (6,5) and the cc-PVDZ basis set were used by the authors for CASSCF, MR-CI and MR-CI+Q. Ground state geometries were optimised with DFT, TDDFT, RI-CC2 and MR-CISD.

Table S4. Literature values for vertical excitation energies for the optimised ground state of protonated Schiff bases PSB3 to PSB4 from ref. [3].

	CASSCF	MRCI	MRCI+Q	RI- CC2	TDDFT
PSB3					
S ₂	5.72	5.65	5.59	6.25	5.92
S ₁	5.64	5.04	4.73	4.20	4.33
PSB4					
S ₂	5.08	4.90	4.81	5.26	4.84
S ₁	5.04	4.47	4.14	3.39	3.59

CASSCF and MS-CASPT2 vertical excitation energies calculated in the current work for equilibrium CASSCF, MP2 and MS-CASPT2 ground state geometries of protonated Schiff bases are listed in Table S5. The computed energies are within the range of the literature values. Deviations result from varying active space sizes and ground state optimisation methodologies.

Table S5. CASSCF and MS-CASPT2 vertical excitation energies and oscillator strengths calculated for MP2, CASSCF and MS-CASPT2 optimised 1^1A_g geometries of protonated Schiff bases.

Geometries		CASSCF/ eV	<i>f</i>	MS- CASPT2/ eV	<i>f</i>
PSB3					
CASSCF	S ₂	5.73	0.13	5.36	0.24·10 ⁻¹
	S ₁	5.06	1.07	4.07	1.07
MP2	S ₂	5.70	0.1	5.31	0.22·10 ⁻¹
	S ₁	5.01	1.11	4.05	1.09
MS-CASPT2	S ₂	5.51	0.78·10 ⁻¹	5.15	0.16·10 ⁻¹
	S ₁	4.92	1.13	3.97	1.10
PSB4					
CASSCF	S ₂	4.79	0.12	4.34	0.35·10 ⁻¹
	S ₁	4.11	1.41	3.34	1.50
MP2	S ₂	4.61	0.71·10 ⁻¹	4.18	0.23·10 ⁻¹
	S ₁	3.99	1.48	3.29	1.54
MS-CASPT2	S ₂	4.44	0.22·10 ⁻¹	4.05	0.15·10 ⁻¹
	S ₁	3.92	1.54	3.28	1.59
PSB5					
CASSCF	S ₂	4.15	0.14	3.68	0.55·10 ⁻¹
	S ₁	3.47	1.68	2.87	1.92
MP2	S ₂	3.90	0.55·10 ⁻¹	3.48	0.25·10 ⁻¹
	S ₁	3.31	1.81	2.82	2.02
MS-CASPT2	S ₂	3.70	0.85·10 ⁻²	3.35	0.12·10 ⁻¹
	S ₁	3.21	1.88	2.84	2.11

Table S6 lists the corresponding state compositions of CASSCF reference wave function for the MS-CASPT2 ground state optimisation of the computed PSBs. The obtained configurations are in good agreement with MS-CASPT2 results from Table 2 in the main article.

Table S6. CASSCF excited state compositions of S_0 optimised geometries in protonated Schiff bases.

Geometry	Configurations (weights in %)
PSB3	
S_0	GS(86)
S_2	$H^2 \rightarrow L^2$ (30), $H-1 \rightarrow L$ (28), $H \rightarrow L+1$ (12), $H \rightarrow L$ (5)
S_1	$H \rightarrow L$ (71)
PSB4	
S_0	GS (79)
S_2	$H^2 \rightarrow L^2$ (33), $H-1 \rightarrow L$ (23), $H \rightarrow L+1$ (14), $H \rightarrow L$ (2)
S_1	$H \rightarrow L$ (67)
PSB5	
S_0	GS (74)
S_2	$H^2 \rightarrow L^2$ (32), $H-1 \rightarrow L$ (20), $H \rightarrow L+1$ (16)
S_1	$H \rightarrow L$ (62)

CASSCF and MS-CASPT2 excited state compositions of MS-CASPT2 optimised S_1 and S_2 geometries in PSBs are summarised in Table S7. We notice considerable differences in the reference CASSCF wavefunctions of the inverted and non-inverted geometries. The inverted structure exhibits a larger amount of double excitation character. The resulting consequences are discussed in the main text.

Table S7. CASSCF reference and MS-CASPT2 excited state compositions for MS-CASPT2 S_1 and S_2 equilibrium geometries.

MS-CASPT2 geometry	Configurations (weights in %)	
	CASSCF	MS-CASPT2
PSB3		
S_2	$H^2 \rightarrow L^2(34)$, $H-1 \rightarrow L(24)$, $H \rightarrow L+1(17)$	$H^2 \rightarrow L^2(33)$, $H-1 \rightarrow L(24)$, $H \rightarrow L+1(15)$
S_1	$H \rightarrow L(55)$, $H \rightarrow L+1(11)$, GS(4), $H^2 \rightarrow L^2(9)$	$H \rightarrow L(76)$
S_1 inv.	$H \rightarrow L(40)$, $H-1 \rightarrow L(14)$, GS(10), $H^2 \rightarrow L^2(13)$	$H \rightarrow L(76)$
PSB4		
S_2	$H^2 \rightarrow L^2(35)$, $H-1 \rightarrow L(20)$, $H \rightarrow L+1(17)$	$^2H \rightarrow ^2L(34)$, $H-1 \rightarrow L(21)$, $H \rightarrow L+1(15)$
S_1	$H \rightarrow L(59)$, $H \rightarrow L+1(8)$, GS(5), $H^2 \rightarrow L^2(6)$	$H \rightarrow L(70)$
S_1 inv.	$H \rightarrow L(43)$, $H-1 \rightarrow L(10)$, GS(14), $H^2 \rightarrow L^2(9)$	$H \rightarrow L(71)$
PSB5		
S_2	$H^2 \rightarrow L^2(33)$, $H-1 \rightarrow L(17)$, $H \rightarrow L+1(17)$	$^2H \rightarrow ^2L(32)$, $H-1 \rightarrow L(18)$, $H \rightarrow L+1(15)$
S_1	$H \rightarrow L(56)$, $H \rightarrow L+1(7)$, GS(5), $H^2 \rightarrow L^2(5)$	$H \rightarrow L(65)$
S_1 inv.	$H \rightarrow L(42)$, $H-1 \rightarrow L(8)$, $H^2 \rightarrow L^2(7)$, GS(16)	$H \rightarrow L(67)$

CASSCF bond lengths of PSBs

In Figure S4, bond lengths of CASSCF optimised ground and excited state geometries in PSBs are given. All PSBs show a similar behaviour: the ground state bond alternation is very pronounced and S_1 optimisation leads only to the bond inverted geometry. S_2 has an equalised bond length pattern.

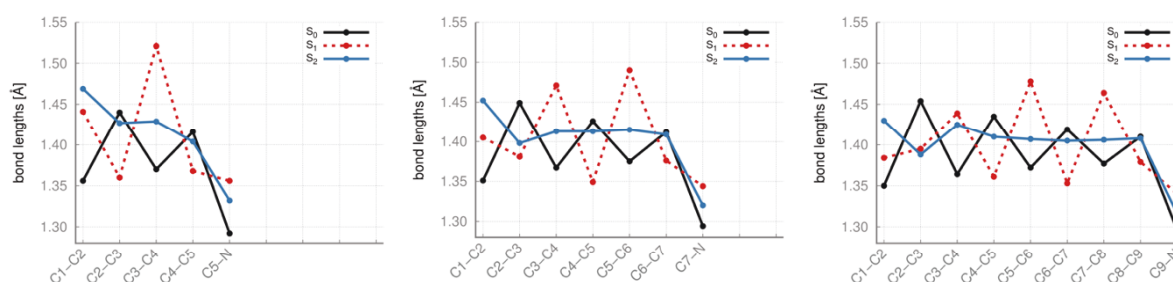


Figure S4. Bond lengths of CASSCF optimised ground and excited state geometries of protonated Schiff bases.

OM3-MRCI optimised PSB geometries

Also for the PSBs, OM3-MRCI geometry optimisations were performed. Similar to results obtained for linear polyenes, no major difference to OM2-MRCI can be noted. Inverted and non-inverted minima could be located for PSB4 and PSB5.

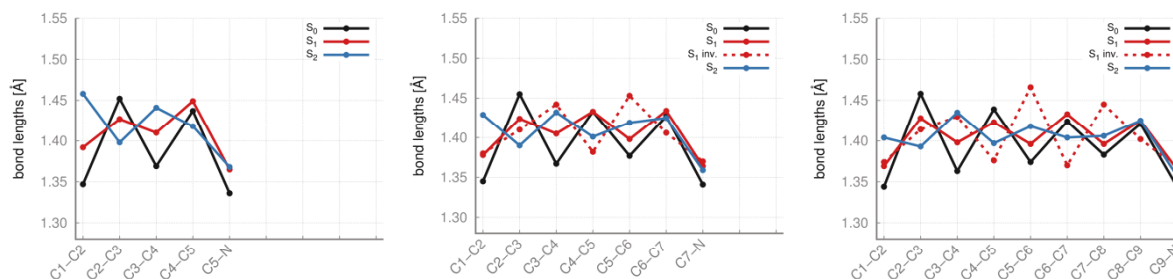


Figure S5. Bond lengths of OM3-MRCI optimised ground and excited state geometries of protonated Schiff bases.

Interpolated paths

The CASSCF reference energies for each point of the linear interpolated PSB MS-CASPT2 paths are plotted in Figure S6. In all paths, the inverted geometry is more favourable compared to the non-inverted one. For PSB3, the barrier from the non-inverted to the inverted structure is rather small (< 0.04 eV). It is even smaller for PSB4 and absent in PSB5. These results may explain the difficulties in stabilising the non-inverted structure at the CASSCF level of theory.

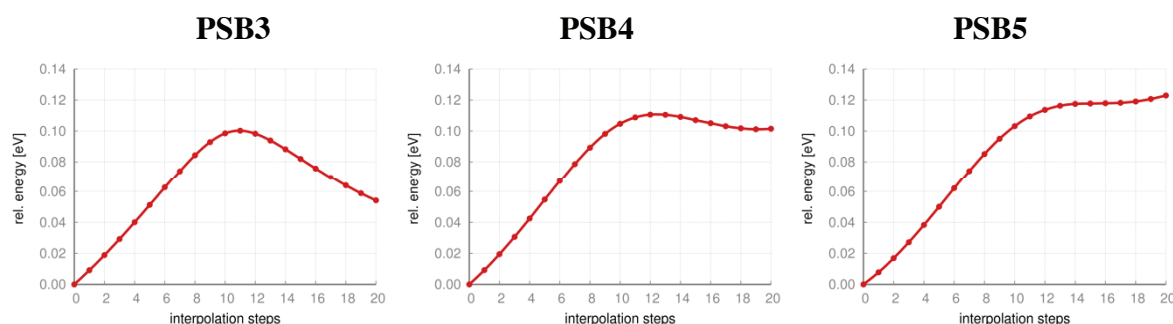


Figure S6. CASSCF reference energies for MS-CASPT2 linear interpolated paths between inverted and non-inverted S_1 geometries.

In Figure S7, the MS-CASPT2 linearly interpolated path for PSB3 is displayed. Inverted and non-inverted minima are almost iso-energetic and are separated by an energy barrier of ca. 0.2 eV.

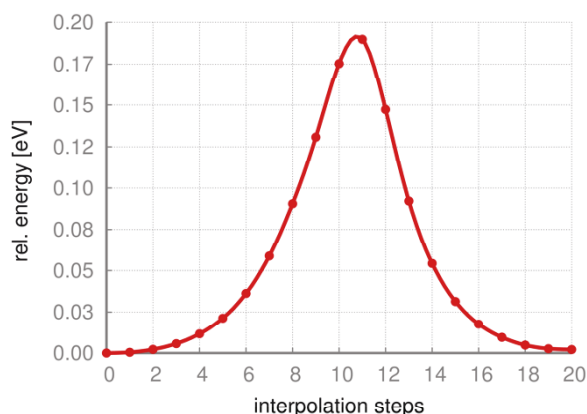


Figure S7. MS-CASPT2 interpolated path for S_1 inverted and non-inverted equilibrium geometries of PSB3.

Unrelaxed MS-CASPT2 torsion potentials

Unrelaxed MS-CASPT2 torsion scans about C3-C4, C4-C5, C6-C7, C7-C8 bonds in PSB5 were performed in addition to the C5-C6 scan in the main article (Figure S9). As expected, barriers on the S_1 PES appear in the non-inverted geometry when following double bond torsions. In the inverted structure torsion about formal double bonds is facilitated. An S_1/S_0 crossing is reached for the torsion along C7-C8, which would correspond to the C13-C14 bond in the retinal protonated Schiff base in Figure S8. Torsion about C13-C14 bond is preferred in microbial rhodopsins during isomerisation.

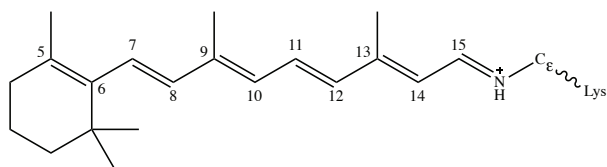


Figure S8. Retinal Schiff base.

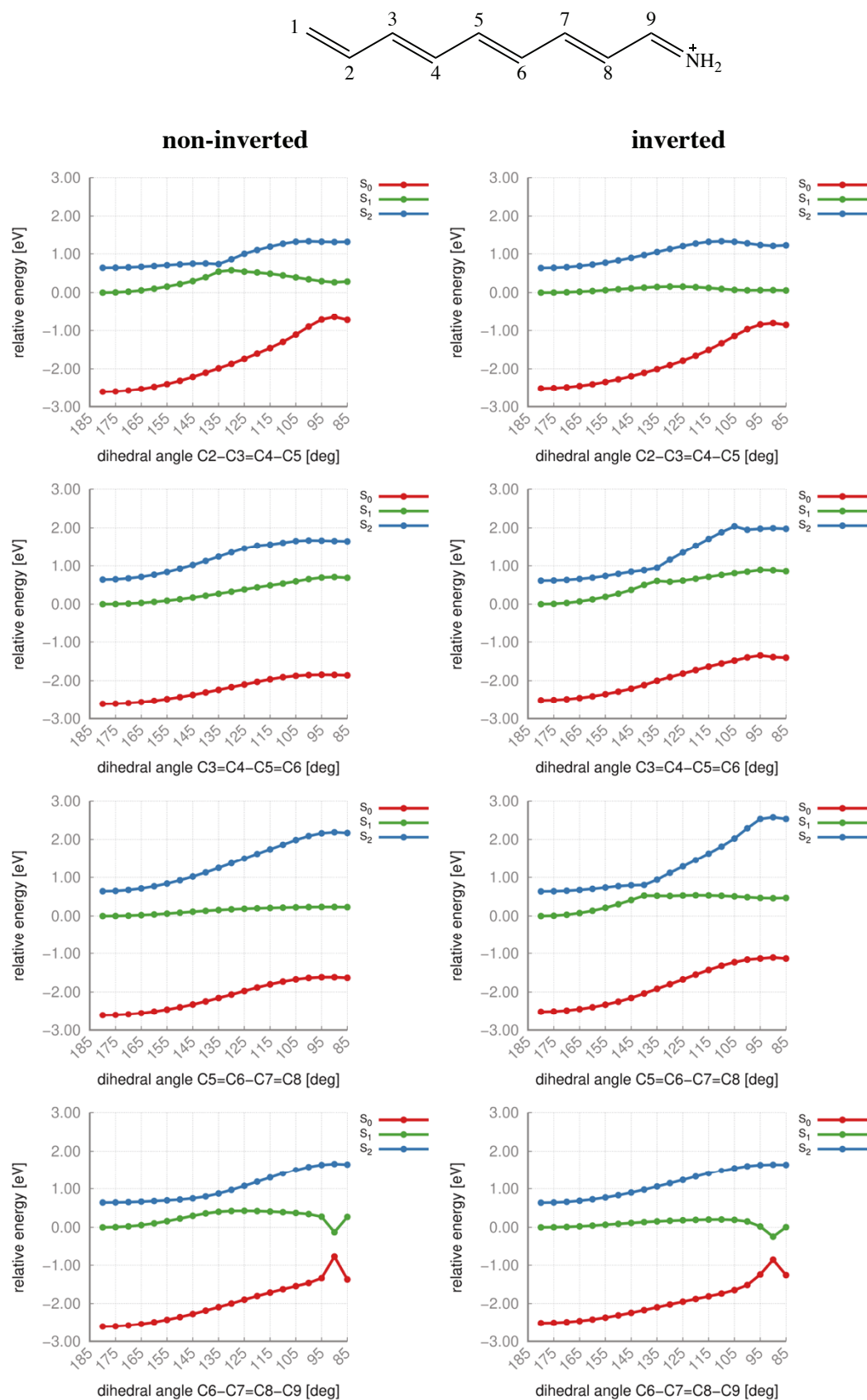


Figure S9. MS-CASPT2 unrelaxed dihedral angle scans for optimised inverted and non-inverted PSB5 geometries.

In Figure S10, an unrelaxed OM2-MRCI scan about the C5-C6 bond in PSB5 is displayed. For this calculation, only π -orbitals were included in the active space, similar to the geometry optimisations discussed in the main text. We conclude, that for a smooth torsion path the usage of σ -orbitals is important.

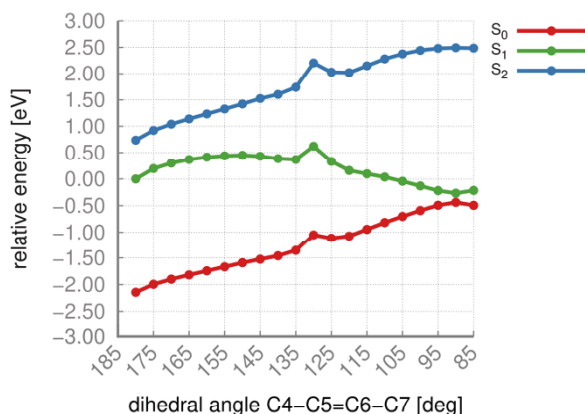


Figure S10. OM2-MRCI unrelaxed torsion scan of optimised and inverted PSB5 geometry.

Note: See next page for Table S8 depicting the xyz-coordinates of optimised minima in Cartesian coordinates.

REFERENCES

- Schreiber, M., M.R. Silva-Junior, S.P.A. Sauer and W. Thiel (2008) Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.*, **128**.
- Nakayama, K., H. Nakano and K. Hirao (1998) Theoretical Study of the $\pi \rightarrow \pi^*$ Excited States of Linear Polyenes: The Energy Gap Between 1Bu and 2Ag States and Their Character. *Int. J. Quantum Chem.*, **66**, 157–175.
- Aquino, A.J.A., M. Barbatti and H. Lischka (2006) Excited-state properties and environmental effects for protonated Schiff bases: A theoretical study. *ChemPhysChem*, **7**, 2089–2096.

Table S8. xyz-coordinates of optimised minima in Cartesian coordinates (\AA).

hexatriene

MP2-GS

C	1.212232	2.847399	0.000000
C	1.212232	1.491835	0.000000
C	0.000137	0.681587	0.000000
C	-0.000137	-0.681587	0.000000
C	-1.212232	-1.491835	0.000000
C	-1.212232	-2.847399	0.000000
H	2.146109	3.416556	0.000000
H	0.272467	3.410901	0.000000
H	2.169489	0.953017	0.000000
H	-0.961946	1.214161	0.000000
H	0.961946	-1.214161	0.000000
H	-2.169489	-0.953017	0.000000
H	-0.272467	-3.410901	0.000000
H	-2.146109	-3.416556	0.000000

CASSCF

1^1A_g

C	1.212188	2.848444	0.000000
C	-1.212188	-2.848444	0.000000
C	-0.000126	0.678356	0.000000
C	0.000126	-0.678356	0.000000
C	-1.207924	-1.496533	0.000000
C	1.207924	1.496533	0.000000
H	2.138192	3.407218	0.000000
H	-2.138192	-3.407218	0.000000
H	0.290688	3.418360	0.000000
H	-0.290688	-3.418360	0.000000
H	2.155498	0.967982	0.000000
H	-2.155498	-0.967982	0.000000
H	0.949442	-1.205060	0.000000
H	-0.949442	1.205060	0.000000

2^1A_g

C	1.224354	2.943895	0.000000
C	-1.224354	-2.943895	0.000000
C	0.019610	0.722038	0.000000

C	-0.019610	-0.722038	0.000000
C	-1.189214	-1.474297	0.000000
C	1.189214	1.474297	0.000000
H	2.165954	3.470461	0.000000
H	-2.165954	-3.470461	0.000000
H	0.309823	3.519191	0.000000
H	-0.309823	-3.519191	0.000000
H	2.140904	0.956314	0.000000
H	-2.140904	-0.956314	0.000000
H	0.930627	-1.245757	0.000000
H	-0.930627	1.245757	0.000000

2^1B_u

C	1.213101	2.882605	0.000000
C	-1.213101	-2.882605	0.000000
C	0.002731	0.706894	0.000000
C	-0.002731	-0.706894	0.000000
C	-1.167793	-1.495435	0.000000
C	1.167793	1.495435	0.000000
H	2.158905	3.402925	0.000000
H	-2.158905	-3.402925	0.000000
H	0.309127	3.476952	0.000000
H	-0.309127	-3.476952	0.000000
H	2.115177	0.965742	0.000000
H	-2.115177	-0.965742	0.000000
H	0.955458	-1.215064	0.000000
H	-0.955458	1.215064	0.000000

MS-CASPT2

1^1A_g

C	1.212080	2.849669	0.000000
C	-1.212080	-2.849669	0.000000
C	0.000540	0.683486	0.000000
C	-0.000540	-0.683486	0.000000
C	-1.211687	-1.490799	0.000000
C	1.211687	1.490799	0.000000
H	2.143853	3.419091	0.000000
H	-2.143853	-3.419091	0.000000
H	0.275058	3.414375	0.000000
H	-0.275058	-3.414375	0.000000
H	2.166050	0.951188	0.000000
H	-2.166050	-0.951188	0.000000

H	0.958439	-1.217051	0.000000
H	-0.958439	1.217051	0.000000

2¹A_g

C	1.225988	2.944974	0.000000
C	-1.225988	-2.944974	0.000000
C	0.018572	0.727700	0.000000
C	-0.018572	-0.727700	0.000000
C	-1.191362	-1.474607	0.000000
C	1.191362	1.474607	0.000000
H	2.175019	3.481050	0.000000
H	-2.175019	-3.481050	0.000000
H	0.296687	3.518247	0.000000
H	-0.296687	-3.518247	0.000000
H	2.149835	0.943363	0.000000
H	-2.149835	-0.943363	0.000000
H	0.943361	-1.255723	0.000000
H	-0.943361	1.255723	0.000000

1¹B_u

C	1.195318	2.839335	0.000000
C	-1.195318	-2.839335	0.000000
C	0.088845	0.702044	0.000000
C	-0.088845	-0.702044	0.000000
C	-1.287960	-1.453544	0.000000
C	1.287960	1.453544	0.000000
H	2.085414	3.478459	0.000000
H	-2.085414	-3.478459	0.000000
H	0.217313	3.337713	0.000000
H	-0.217313	-3.337713	0.000000
H	2.254007	0.935520	0.000000
H	-2.254007	-0.935520	0.000000
H	0.847182	-1.285798	0.000000
H	-0.847182	1.285798	0.000000

OM2-MRCI

1¹A_g

C	1.188734	2.795138	0.000000
C	1.214342	1.457371	0.000000
C	-0.015230	0.673590	0.000000
C	0.015230	-0.673590	0.000000

C	-1.214342	-1.457371	0.000000
C	-1.188734	-2.795138	0.000000
H	2.095915	3.386655	0.000000
H	0.256953	3.349139	0.000000
H	2.162722	0.901673	0.000000
H	-0.964979	1.225679	0.000000
H	0.964979	-1.225679	0.000000
H	-2.162722	-0.901673	0.000000
H	-0.256953	-3.349139	0.000000
H	-2.095915	-3.386655	0.000000

2^1A_g

C	1.191763	2.860736	0.000000
C	1.191864	1.427101	0.000000
C	0.004525	0.718759	0.000000
C	-0.004525	-0.718759	0.000000
C	-1.191864	-1.427101	0.000000
C	-1.191763	-2.860736	0.000000
H	2.121241	3.409712	0.000000
H	0.260053	3.408312	0.000000
H	2.154837	0.904192	0.000000
H	-0.957334	1.248258	0.000000
H	0.957334	-1.248258	0.000000
H	-2.154837	-0.904192	0.000000
H	-0.260053	-3.408312	0.000000
H	-2.121241	-3.409712	0.000000

1^1B_u

C	1.189618	2.823978	0.000000
C	1.185244	1.447277	0.000000
C	-0.011877	0.704808	0.000000
C	0.011877	-0.704808	0.000000
C	-1.185244	-1.447277	0.000000
C	-1.189618	-2.823978	0.000000
H	2.114148	3.384253	0.000000
H	0.266245	3.388658	0.000000
H	2.140384	0.902344	0.000000
H	-0.975073	1.231148	0.000000
H	0.975073	-1.231148	0.000000
H	-2.140384	-0.902344	0.000000
H	-0.266245	-3.388658	0.000000
H	-2.114148	-3.384253	0.000000

octatetraene

MP2-GS

C	0.136440	4.331388	0.000000
C	0.663479	3.081578	0.000000
C	-0.136440	1.864985	0.000000
C	0.396363	0.606908	0.000000
C	-0.396363	-0.606908	0.000000
C	0.136440	-1.864985	0.000000
C	-0.663479	-3.081578	0.000000
C	-0.136440	-4.331388	0.000000
H	-0.948458	4.485312	0.000000
H	0.775878	5.218611	0.000000
H	1.755048	2.957819	0.000000
H	-1.230183	1.979159	0.000000
H	1.490459	0.495385	0.000000
H	-1.490459	-0.495385	0.000000
H	1.230183	-1.979159	0.000000
H	-1.755048	-2.957819	0.000000
H	0.948458	-4.485312	0.000000
H	-0.775878	-5.218611	0.000000

CASSCF

1¹A_g

C	0.138031	4.331688	0.000000
C	-0.138031	-4.331688	0.000000
C	0.655592	3.084503	0.000000
C	-0.655592	-3.084503	0.000000
C	-0.140223	1.862083	0.000000
C	0.140223	-1.862083	0.000000
C	0.392093	0.611750	0.000000
C	-0.392093	-0.611750	0.000000
H	-0.932034	4.502399	0.000000
H	0.932034	-4.502399	0.000000
H	0.776654	5.204685	0.000000
H	-0.776654	-5.204685	0.000000
H	1.733827	2.962786	0.000000
H	-1.733827	-2.962786	0.000000
H	-1.219671	1.976973	0.000000
H	1.219671	-1.976973	0.000000
H	1.471994	0.502557	0.000000
H	-1.471994	-0.502557	0.000000

2¹A_g

C	0.116613	4.406335	0.000000
C	-0.116613	-4.406335	0.000000
C	0.638931	3.064290	0.000000
C	-0.638931	-3.064290	0.000000
C	-0.137746	1.920895	0.000000
C	0.137746	-1.920895	0.000000
C	0.394027	0.578105	0.000000
C	-0.394027	-0.578105	0.000000
H	-0.948992	4.588595	0.000000
H	0.948992	-4.588595	0.000000
H	0.780059	5.257681	0.000000
H	-0.780059	-5.257681	0.000000
H	1.717790	2.952122	0.000000
H	-1.717790	-2.952122	0.000000
H	-1.217665	2.028565	0.000000
H	1.217665	-2.028565	0.000000
H	1.472892	0.467328	0.000000
H	-1.472892	-0.467328	0.000000

2¹B_u

C	0.132428	4.350572	0.000000
C	-0.132428	-4.350572	0.000000
C	0.621283	3.065109	0.000000
C	-0.621283	-3.065109	0.000000
C	-0.160120	1.884365	0.000000
C	0.160120	-1.884365	0.000000
C	0.375787	0.588775	0.000000
C	-0.375787	-0.588775	0.000000
H	-0.931175	4.551954	0.000000
H	0.931175	-4.551954	0.000000
H	0.802121	5.198118	0.000000
H	-0.802121	-5.198118	0.000000
H	1.699048	2.935269	0.000000
H	-1.699048	-2.935269	0.000000
H	-1.239780	1.990735	0.000000
H	1.239780	-1.990735	0.000000
H	1.456935	0.494051	0.000000
H	-1.456935	-0.494051	0.000000

MS-CASPT2

1¹A_g

C	0.134805	4.333205	0.000000
C	-0.134805	-4.333205	0.000000
C	0.665693	3.081192	0.000000
C	-0.665693	-3.081192	0.000000
C	-0.133383	1.866836	0.000000
C	0.133383	-1.866836	0.000000
C	0.398551	0.604686	0.000000
C	-0.398551	-0.604686	0.000000
H	-0.948488	4.486132	0.000000
H	0.948488	-4.486132	0.000000
H	0.769554	5.221761	0.000000
H	-0.769554	-5.221761	0.000000
H	1.755212	2.957789	0.000000
H	-1.755212	-2.957789	0.000000
H	-1.224607	1.983295	0.000000
H	1.224607	-1.983295	0.000000
H	1.489978	0.488360	0.000000
H	-1.489978	-0.488360	0.000000

2¹A_g

C	0.119987	4.408304	0.000000
C	-0.119987	-4.408304	0.000000
C	0.644902	3.070086	0.000000
C	-0.644902	-3.070086	0.000000
C	-0.141098	1.924347	0.000000
C	0.141098	-1.924347	0.000000
C	0.395535	0.574899	0.000000
C	-0.395535	-0.574899	0.000000
H	-0.959082	4.580761	0.000000
H	0.959082	-4.580761	0.000000
H	0.784036	5.273400	0.000000
H	-0.784036	-5.273400	0.000000
H	1.734657	2.947079	0.000000
H	-1.734657	-2.947079	0.000000
H	-1.233067	2.034221	0.000000
H	1.233067	-2.034221	0.000000
H	1.487183	0.463322	0.000000
H	-1.487183	-0.463322	0.000000

1¹B_u

C	0.151568	4.381960	0.000000
---	----------	----------	----------

C	-0.151568	-4.381960	0.000000
C	0.555348	3.057115	0.000000
C	-0.555348	-3.057115	0.000000
C	-0.268928	1.900698	0.000000
C	0.268928	-1.900698	0.000000
C	0.343482	0.615905	0.000000
C	-0.343482	-0.615905	0.000000
H	-0.911726	4.647843	0.000000
H	0.911726	-4.647843	0.000000
H	0.886502	5.192479	0.000000
H	-0.886502	-5.192479	0.000000
H	1.639390	2.866699	0.000000
H	-1.639390	-2.866699	0.000000
H	-1.362128	1.992608	0.000000
H	1.362128	-1.992608	0.000000
H	1.442420	0.580518	0.000000
H	-1.442420	-0.580518	0.000000

OM2-MRCI

1^1A_g

C	-0.023298	-0.045734	0.000000
C	-0.005864	1.292398	0.000000
C	1.247507	2.036187	0.000000
C	1.259918	3.385231	0.000000
C	2.507293	4.129769	0.000000
C	2.519704	5.478813	0.000000
C	3.773075	6.222602	0.000000
C	3.790509	7.560734	0.000000
H	0.890210	-0.629405	0.000000
H	-0.949059	-0.607766	0.000000
H	-0.936333	1.877777	0.000000
H	2.179459	1.454839	0.000000
H	0.325323	3.963363	0.000000
H	3.441888	3.551637	0.000000
H	1.587752	6.060162	0.000000
H	4.703544	5.637223	0.000000
H	2.877001	8.144405	0.000000
H	4.716270	8.122766	0.000000

2^1A_g

C	-0.017421	-0.101176	0.000000
C	0.027803	1.303747	0.000000

C	1.233442	1.982938	0.000000
C	1.278643	3.417641	0.000000
C	2.488605	4.097011	0.000000
C	2.533728	5.531753	0.000000
C	3.739426	6.210948	0.000000
C	3.784728	7.615762	0.000000
H	0.893004	-0.685446	0.000000
H	-0.961243	-0.627151	0.000000
H	-0.915826	1.864395	0.000000
H	2.181665	1.428839	0.000000
H	0.332023	3.972965	0.000000
H	3.435243	3.541758	0.000000
H	1.585521	6.085844	0.000000
H	4.682970	5.650106	0.000000
H	2.874545	8.200429	0.000000
H	4.728350	8.142107	0.000000

1¹B_u

C	-0.018433	-0.072489	0.000000
C	0.028128	1.293512	0.000000
C	1.254794	2.003334	0.000000
C	1.278552	3.402749	0.000000
C	2.486667	4.112994	0.000000
C	2.510415	5.512425	0.000000
C	3.737074	6.222201	0.000000
C	3.783495	7.588239	0.000000
H	0.886015	-0.668194	0.000000
H	-0.959105	-0.606628	0.000000
H	-0.906199	1.872992	0.000000
H	2.197800	1.440598	0.000000
H	0.329631	3.956985	0.000000
H	3.435600	3.558777	0.000000
H	1.567406	6.075206	0.000000
H	4.671396	5.642489	0.000000
H	2.879024	8.183762	0.000000
H	4.724081	8.122616	0.000000

decapentaene

MP2-GS

C	2.412244	5.024976	0.000000
C	2.409502	3.668210	0.000000
C	1.199034	2.860563	0.000000
C	1.199034	1.493020	0.000000

C	-0.001253	0.684917	0.000000
C	0.001253	-0.684917	0.000000
C	-1.199034	-1.493020	0.000000
C	-1.199034	-2.860563	0.000000
C	-2.409502	-3.668210	0.000000
C	-2.412244	-5.024976	0.000000
H	3.347347	5.592100	0.000000
H	1.473517	5.590212	0.000000
H	3.366326	3.128398	0.000000
H	0.236162	3.391825	0.000000
H	2.163619	0.964553	0.000000
H	-0.966758	1.211596	0.000000
H	0.966758	-1.211596	0.000000
H	-2.163619	-0.964553	0.000000
H	-0.236162	-3.391825	0.000000
H	-3.366326	-3.128398	0.000000
H	-1.473517	-5.590212	0.000000
H	-3.347347	-5.592100	0.000000

CASSCF

1¹A_g

C	2.411469	5.021840	0.000000
C	-2.411469	-5.021840	0.000000
C	2.406857	3.671590	0.000000
C	-2.406857	-3.671590	0.000000
C	1.198523	2.854186	0.000000
C	-1.198523	-2.854186	0.000000
C	1.200189	1.496113	0.000000
C	-1.200189	-1.496113	0.000000
C	0.001939	-0.679582	0.000000
C	-0.001939	0.679582	0.000000
H	3.337253	5.581152	0.000000
H	-3.337253	-5.581152	0.000000
H	1.490088	5.592130	0.000000
H	-1.490088	-5.592130	0.000000
H	3.353981	3.142048	0.000000
H	-3.353981	-3.142048	0.000000
H	0.249126	3.380324	0.000000
H	-0.249126	-3.380324	0.000000
H	2.150882	0.972185	0.000000
H	-2.150882	-0.972185	0.000000
H	0.953033	-1.202580	0.000000
H	-0.953033	1.202580	0.000000

2¹A_g

C	2.424882	5.082655	0.000000
C	-2.424882	-5.082655	0.000000
C	2.379713	3.666730	0.000000
C	-2.379713	-3.666730	0.000000
C	1.214444	2.914206	0.000000
C	-1.214444	-2.914206	0.000000
C	1.185007	1.469238	0.000000
C	-1.185007	-1.469238	0.000000
C	-0.016519	-0.713622	0.000000
C	0.016519	0.713622	0.000000
H	3.368023	5.608441	0.000000
H	-3.368023	-5.608441	0.000000
H	1.516978	5.670631	0.000000
H	-1.516978	-5.670631	0.000000
H	3.326673	3.136792	0.000000
H	-3.326673	-3.136792	0.000000
H	0.261918	3.433601	0.000000
H	-0.261918	-3.433601	0.000000
H	2.137341	0.949969	0.000000
H	-2.137341	-0.949969	0.000000
H	0.934137	-1.236901	0.000000
H	-0.934137	1.236901	0.000000

2¹B_u

C	2.387313	5.045429	0.000000
C	-2.387313	-5.045429	0.000000
C	2.344730	3.678231	0.000000
C	-2.344730	-3.678231	0.000000
C	1.158963	2.888972	0.000000
C	-1.158963	-2.888972	0.000000
C	1.153963	1.496494	0.000000
C	-1.153963	-1.496494	0.000000
C	-0.000709	-0.700269	0.000000
C	0.000709	0.700269	0.000000
H	3.330029	5.573521	0.000000
H	-3.330029	-5.573521	0.000000
H	1.482307	5.639981	0.000000
H	-1.482307	-5.639981	0.000000
H	3.287413	3.139758	0.000000
H	-3.287413	-3.139758	0.000000
H	0.206401	3.408229	0.000000

H	-0.206401	-3.408229	0.000000
H	2.113161	0.988168	0.000000
H	-2.113161	-0.988168	0.000000
H	0.960269	-1.204529	0.000000
H	-0.960269	1.204529	0.000000

MS-CASPT2

1^1A_g

C	2.414179	5.028997	0.000000
C	-2.414179	-5.028997	0.000000
C	2.407947	3.668716	0.000000
C	-2.407947	-3.668716	0.000000
C	1.197630	2.864978	0.000000
C	-1.197630	-2.864978	0.000000
C	1.197171	1.493620	0.000000
C	-1.197171	-1.493620	0.000000
C	0.001703	-0.686708	0.000000
C	-0.001703	0.686708	0.000000
H	3.348773	5.593794	0.000000
H	-3.348773	-5.593794	0.000000
H	1.479767	5.597885	0.000000
H	-1.479767	-5.597885	0.000000
H	3.361033	3.126660	0.000000
H	-3.361033	-3.126660	0.000000
H	0.238223	3.397798	0.000000
H	-0.238223	-3.397798	0.000000
H	2.158938	0.964821	0.000000
H	-2.158938	-0.964821	0.000000
H	0.964549	-1.213294	0.000000
H	-0.964549	1.213294	0.000000

2^1A_g

C	2.434148	5.088652	0.000000
C	-2.434148	-5.088652	0.000000
C	2.384774	3.673772	0.000000
C	-2.384774	-3.673772	0.000000
C	1.206578	2.925682	0.000000
C	-1.206578	-2.925682	0.000000
C	1.179455	1.477413	0.000000
C	-1.179455	-1.477413	0.000000
C	-0.012334	-0.719441	0.000000
C	0.012334	0.719441	0.000000

H	3.387000	5.619991	0.000000
H	-3.387000	-5.619991	0.000000
H	1.514548	5.679610	0.000000
H	-1.514548	-5.679610	0.000000
H	3.334783	3.125467	0.000000
H	-3.334783	-3.125467	0.000000
H	0.245197	3.454778	0.000000
H	-0.245197	-3.454778	0.000000
H	2.143822	0.953782	0.000000
H	-2.143822	-0.953782	0.000000
H	0.952086	-1.242818	0.000000
H	-0.952086	1.242818	0.000000

SS-CASPT2

1¹B_u

C	2.386951	5.016847	0.000000
C	-2.386951	-5.016847	0.000000
C	2.363873	3.651072	0.000000
C	-2.363873	-3.651072	0.000000
C	1.174962	2.875471	0.000000
C	-1.174962	-2.875471	0.000000
C	1.165840	1.478308	0.000000
C	-1.165840	-1.478308	0.000000
C	-0.000295	-0.705709	0.000000
C	0.000295	0.705709	0.000000
H	3.319043	5.561312	0.000000
H	-3.319043	-5.561312	0.000000
H	1.469010	5.588493	0.000000
H	-1.469010	-5.588493	0.000000
H	3.304104	3.110641	0.000000
H	-3.304104	-3.110641	0.000000
H	0.222402	3.396432	0.000000
H	-0.222402	-3.396432	0.000000
H	2.120027	0.961043	0.000000
H	-2.120027	-0.961043	0.000000
H	0.957961	-1.214692	0.000000
H	-0.957961	1.214692	0.000000

OM2-MRCI

1¹A_g

C	2.364035	4.936310	0.000000
---	----------	----------	----------

C	2.393340	3.598303	0.000000
C	1.166995	2.811220	0.000000
C	1.201823	1.462358	0.000000
C	-0.017661	0.674967	0.000000
C	0.017661	-0.674967	0.000000
C	-1.201823	-1.462358	0.000000
C	-1.166995	-2.811220	0.000000
C	-2.393340	-3.598303	0.000000
C	-2.364035	-4.936310	0.000000
H	3.269579	5.530337	0.000000
H	1.430698	5.487685	0.000000
H	3.343756	3.045873	0.000000
H	0.215193	3.359377	0.000000
H	2.156488	0.917836	0.000000
H	-0.972397	1.218866	0.000000
H	0.972397	-1.218866	0.000000
H	-2.156488	-0.917836	0.000000
H	-0.215193	-3.359377	0.000000
H	-3.343756	-3.045873	0.000000
H	-1.430698	-5.487685	0.000000
H	-3.269579	-5.530337	0.000000

2^1A_g

C	2.371033	4.972396	0.000000
C	2.367678	3.586752	0.000000
C	1.177689	2.863549	0.000000
C	1.182355	1.431246	0.000000
C	0.000276	0.710978	0.000000
C	-0.000276	-0.710978	0.000000
C	-1.182355	-1.431246	0.000000
C	-1.177689	-2.863549	0.000000
C	-2.367678	-3.586752	0.000000
C	-2.371033	-4.972396	0.000000
H	3.296231	5.531845	0.000000
H	1.445971	5.534590	0.000000
H	3.324066	3.046193	0.000000
H	0.213334	3.388487	0.000000
H	2.148610	0.910009	0.000000
H	-0.963692	1.237917	0.000000
H	0.963692	-1.237917	0.000000
H	-2.148610	-0.910009	0.000000
H	-0.213334	-3.388487	0.000000
H	-3.324066	-3.046193	0.000000
H	-1.445971	-5.534590	0.000000

H	-3.296231	-5.531845	0.000000
1 ¹ B _u			
C	2.373910	4.952413	0.000000
C	2.369583	3.591649	0.000000
C	1.158568	2.841655	0.000000
C	1.174425	1.450995	0.000000
C	-0.012170	0.699707	0.000000
C	0.012170	-0.699707	0.000000
C	-1.174425	-1.450995	0.000000
C	-1.158568	-2.841655	0.000000
C	-2.369583	-3.591649	0.000000
C	-2.373910	-4.952413	0.000000
H	3.296783	5.517122	0.000000
H	1.452181	5.521420	0.000000
H	3.318854	3.036522	0.000000
H	0.202303	3.381773	0.000000
H	2.138439	0.922393	0.000000
H	-0.978633	1.222472	0.000000
H	0.978633	-1.222472	0.000000
H	-2.138439	-0.922393	0.000000
H	-0.202303	-3.381773	0.000000
H	-3.318854	-3.036522	0.000000
H	-1.452181	-5.521420	0.000000
H	-3.296783	-5.517122	0.000000

PSB3

MP2-GS

N	0.154733	4.267870	0.000000
C	0.594842	3.033721	0.000000
C	-0.233904	1.891633	0.000000
C	0.341749	0.649171	0.000000
C	-0.389590	-0.583789	0.000000
C	0.257293	-1.771606	0.000000
H	-0.840425	4.482419	0.000000
H	0.792349	5.059614	0.000000
H	1.677280	2.915724	0.000000
H	-1.315531	2.008723	0.000000
H	1.431426	0.576610	0.000000
H	-1.476392	-0.546594	0.000000
H	-0.283612	-2.711950	0.000000
H	1.342814	-1.827921	0.000000

CASSCF

S₀

N	0.149267	4.264609	0.000000
C	0.591929	3.051146	0.000000
C	-0.226262	1.895727	0.000000
C	0.354392	0.655063	0.000000
C	-0.377190	-0.585314	0.000000
C	0.252261	-1.786628	0.000000
H	-0.836591	4.467804	0.000000
H	0.776158	5.050240	0.000000
H	1.669854	2.942809	0.000000
H	-1.302694	2.010144	0.000000
H	1.437111	0.588684	0.000000
H	-1.458570	-0.537711	0.000000
H	-0.308584	-2.710033	0.000000
H	1.331950	-1.862915	0.000000

S₁

N	0.114332	4.351079	0.000000
C	0.592099	3.082558	0.000000
C	-0.199217	1.967218	0.000000
C	0.379632	0.560168	0.000000
C	-0.364840	-0.578446	0.000000
C	0.257695	-1.878077	0.000000
H	-0.867686	4.546684	0.000000
H	0.738304	5.132647	0.000000
H	1.670240	3.004099	0.000000
H	-1.277493	2.058661	0.000000
H	1.460882	0.483369	0.000000
H	-1.445595	-0.528166	0.000000
H	-0.339254	-2.778092	0.000000
H	1.333934	-1.980076	0.000000

S₂

N	0.025944	4.308598	0.000000
C	0.653427	3.133465	0.000000
C	-0.056388	1.921874	0.000000
C	0.528836	0.618424	0.000000
C	-0.306022	-0.539230	0.000000
C	0.163063	-1.931514	0.000000

H	-0.977233	4.375707	0.000000
H	0.540535	5.170570	0.000000
H	1.733698	3.166124	0.000000
H	-1.140094	1.981239	0.000000
H	1.604812	0.517866	0.000000
H	-1.379399	-0.387798	0.000000
H	-0.555583	-2.732722	0.000000
H	1.217434	-2.158977	0.000000

MS-CASPT2

S₀

N	0.150742	4.285348	0.000000
C	0.597069	3.046440	0.000000
C	-0.233149	1.899943	0.000000
C	0.349663	0.641423	0.000000
C	-0.390414	-0.591289	0.000000
C	0.257335	-1.794966	0.000000
H	-0.849521	4.489820	0.000000
H	0.790986	5.078521	0.000000
H	1.686532	2.932672	0.000000
H	-1.321790	2.016314	0.000000
H	1.445339	0.570996	0.000000
H	-1.483704	-0.544876	0.000000
H	-0.296324	-2.736524	0.000000
H	1.350268	-1.850197	0.000000

S₁

N	0.020295	4.308503	0.000000
C	0.666871	3.138478	0.000000
C	-0.043640	1.900375	0.000000
C	0.554857	0.615859	0.000000
C	-0.309907	-0.547647	0.000000
C	0.141302	-1.911816	0.000000
H	-1.003492	4.357426	0.000000
H	0.529295	5.197223	0.000000
H	1.760970	3.176961	0.000000
H	-1.141627	1.954950	0.000000
H	1.643154	0.502098	0.000000
H	-1.393501	-0.370863	0.000000
H	-0.583313	-2.731004	0.000000
H	1.211769	-2.146918	0.000000

S₁-inv.

N	0.087047	4.346345	0.000000
C	0.673992	3.150043	0.000000
C	-0.059715	1.958399	0.000000
C	0.563362	0.628266	0.000000
C	-0.213512	-0.558398	0.000000
C	0.373055	-1.912162	0.000000
H	-0.937124	4.442177	0.000000
H	0.637797	5.213384	0.000000
H	1.770862	3.152769	0.000000
H	-1.156863	2.012532	0.000000
H	1.658184	0.575187	0.000000
H	-1.306749	-0.469558	0.000000
H	-0.276387	-2.790758	0.000000
H	1.460491	-2.044745	0.000000

S₂

N	0.010790	4.333093	0.000000
C	0.690777	3.124274	0.000000
C	-0.043579	1.915801	0.000000
C	0.525536	0.607461	0.000000
C	-0.316792	-0.547099	0.000000
C	0.155338	-1.931475	0.000000
H	-1.002006	4.368626	0.000000
H	0.527382	5.208947	0.000000
H	1.774722	3.171861	0.000000
H	-1.139580	1.994142	0.000000
H	1.612164	0.480885	0.000000
H	-1.399543	-0.375788	0.000000
H	-0.566889	-2.750122	0.000000
H	1.224710	-2.156980	0.000000

OM2-MRCI

S₀

N	0.108320	4.245386	0.000000
C	0.644052	3.047097	0.000000
C	-0.178686	1.879989	0.000000
C	0.410953	0.646227	0.000000
C	-0.387426	-0.556263	0.000000
C	0.217282	-1.756529	0.000000
H	-0.891298	4.386001	0.000000

H	0.660485	5.089762	0.000000
H	1.748866	2.933925	0.000000
H	-1.260028	1.988704	0.000000
H	1.511970	0.539109	0.000000
H	-1.473672	-0.457953	0.000000
H	-0.357520	-2.676420	0.000000
H	1.299735	-1.865410	0.000000

S₁

N	0.076161	4.272910	0.000000
C	0.674442	3.073625	0.000000
C	-0.146301	1.884613	0.000000
C	0.458360	0.619054	0.000000
C	-0.368198	-0.540125	0.000000
C	0.199628	-1.804039	0.000000
H	-0.923235	4.374036	0.000000
H	0.606240	5.126217	0.000000
H	1.765396	2.995077	0.000000
H	-1.237033	1.988574	0.000000
H	1.547256	0.519299	0.000000
H	-1.458998	-0.426182	0.000000
H	-0.419777	-2.696948	0.000000
H	1.279093	-1.942485	0.000000

S₂

N	0.080749	4.284780	0.000000
C	0.659250	3.073958	0.000000
C	-0.150414	1.914503	0.000000
C	0.443587	0.612197	0.000000
C	-0.366784	-0.523190	0.000000
C	0.224395	-1.847251	0.000000
H	-0.923738	4.406790	0.000000
H	0.630540	5.134439	0.000000
H	1.749831	2.988547	0.000000
H	-1.248351	2.008516	0.000000
H	1.530366	0.517750	0.000000
H	-1.461903	-0.440906	0.000000
H	-0.410657	-2.716872	0.000000
H	1.296163	-1.969634	0.000000

PSB4

MP2-GS

N	0.257102	4.260340	0.000000
C	0.629128	2.995544	0.000000
C	-0.256598	1.896107	0.000000
C	0.257639	0.609178	0.000000
C	-0.528999	-0.578019	0.000000
C	0.066166	-1.821426	0.000000
C	-0.659364	-3.069394	0.000000
C	-0.011919	-4.266618	0.000000
H	-0.726884	4.523870	0.000000
H	0.943255	5.011171	0.000000
H	1.711540	2.816502	0.000000
H	-1.339225	2.066793	0.000000
H	1.351279	0.486701	0.000000
H	-1.622473	-0.493803	0.000000
H	1.165007	-1.877233	0.000000
H	-1.754963	-3.032995	0.000000
H	1.082880	-4.315576	0.000000
H	-0.563567	-5.211147	0.000000

CASSCF

S₀

N	0.237394	4.239561	0.000000
C	0.624554	3.004896	0.000000
C	-0.243014	1.890417	0.000000
C	0.278738	0.618140	0.000000
C	-0.498583	-0.577887	0.000000
C	0.085903	-1.814123	0.000000
C	-0.647726	-3.063371	0.000000
C	-0.036377	-4.268233	0.000000
H	-0.737422	4.487294	0.000000
H	0.899057	4.995042	0.000000
H	1.696514	2.848411	0.000000
H	-1.313156	2.053274	0.000000
H	1.357627	0.506241	0.000000
H	-1.578013	-0.489619	0.000000
H	1.168744	-1.878698	0.000000
H	-1.729536	-3.007537	0.000000
H	1.042719	-4.358128	0.000000
H	-0.607417	-5.185682	0.000000

S₁

N	0.185371	4.321891	0.000000
C	0.618752	3.050049	0.000000
C	-0.203571	1.947357	0.000000
C	0.323930	0.553813	0.000000
C	-0.457216	-0.546462	0.000000
C	0.107850	-1.904560	0.000000
C	-0.649864	-3.058784	0.000000
C	-0.059060	-4.334030	0.000000
H	-0.789695	4.546434	0.000000
H	0.832445	5.083315	0.000000
H	1.693842	2.934211	0.000000
H	-1.278178	2.080652	0.000000
H	1.402394	0.442462	0.000000
H	-1.536437	-0.456924	0.000000
H	1.188001	-1.993425	0.000000
H	-1.729868	-2.985005	0.000000
H	1.015827	-4.452594	0.000000
H	-0.664518	-5.228406	0.000000

S₂

N	0.026810	4.276565	0.000000
C	0.626570	3.101106	0.000000
C	-0.084685	1.884265	0.000000
C	0.517291	0.603694	0.000000
C	-0.275046	-0.566410	0.000000
C	0.219914	-1.890051	0.000000
C	-0.640182	-2.992519	0.000000
C	-0.217874	-4.381597	0.000000
H	-0.973740	4.363945	0.000000
H	0.556660	5.128706	0.000000
H	1.707672	3.112934	0.000000
H	-1.167910	1.940863	0.000000
H	1.595808	0.523450	0.000000
H	-1.352479	-0.440075	0.000000
H	1.291056	-2.047525	0.000000
H	-1.707981	-2.805369	0.000000
H	0.830045	-4.641455	0.000000
H	-0.951923	-5.170533	0.000000

MS-CASPT2

S₀

N	0.246963	4.266230	0.000000
C	0.629638	3.001671	0.000000
C	-0.249233	1.903380	0.000000
C	0.269217	0.601759	0.000000
C	-0.518322	-0.573823	0.000000
C	0.073245	-1.831863	0.000000
C	-0.659978	-3.070459	0.000000
C	-0.020077	-4.276557	0.000000
H	-0.740855	4.520806	0.000000
H	0.925769	5.025213	0.000000
H	1.713017	2.836196	0.000000
H	-1.331519	2.068959	0.000000
H	1.360937	0.482238	0.000000
H	-1.610083	-0.484896	0.000000
H	1.169710	-1.893158	0.000000
H	-1.753604	-3.022755	0.000000
H	1.072471	-4.336956	0.000000
H	-0.577293	-5.215990	0.000000

S₁

N	0.024409	4.291703	0.000000
C	0.644272	3.102853	0.000000
C	-0.073244	1.870201	0.000000
C	0.533890	0.609988	0.000000
C	-0.279672	-0.588945	0.000000
C	0.214790	-1.900283	0.000000
C	-0.670176	-3.021094	0.000000
C	-0.228639	-4.344778	0.000000
H	-0.992964	4.364052	0.000000
H	0.552966	5.162745	0.000000
H	1.739360	3.121352	0.000000
H	-1.170482	1.922773	0.000000
H	1.624431	0.515358	0.000000
H	-1.369391	-0.453328	0.000000
H	1.298161	-2.073148	0.000000
H	-1.749515	-2.824357	0.000000
H	0.841396	-4.579135	0.000000
H	-0.939588	-5.175963	0.000000

S₁-inv.

N	0.130207	4.305891	0.000000
C	0.638027	3.081402	0.000000
C	-0.158897	1.929153	0.000000

C	0.401295	0.585033	0.000000
C	-0.395930	-0.549688	0.000000
C	0.142925	-1.898721	0.000000
C	-0.655397	-3.044628	0.000000
C	-0.116700	-4.382340	0.000000
H	-0.882204	4.466592	0.000000
H	0.731081	5.134651	0.000000
H	1.733453	3.014543	0.000000
H	-1.251041	2.037930	0.000000
H	1.492662	0.486163	0.000000
H	-1.487053	-0.436773	0.000000
H	1.236137	-2.005707	0.000000
H	-1.745999	-2.925123	0.000000
H	0.966166	-4.546550	0.000000
H	-0.778730	-5.251834	0.000000

S₂

N	0.002768	4.301730	0.000000
C	0.652803	3.102018	0.000000
C	-0.059440	1.883638	0.000000
C	0.535734	0.596619	0.000000
C	-0.272574	-0.572250	0.000000
C	0.225657	-1.902997	0.000000
C	-0.649864	-3.003548	0.000000
C	-0.232746	-4.382067	0.000000
H	-1.012312	4.353399	0.000000
H	0.523248	5.173960	0.000000
H	1.743428	3.138352	0.000000
H	-1.157359	1.939310	0.000000
H	1.625733	0.499916	0.000000
H	-1.362014	-0.432131	0.000000
H	1.307646	-2.074011	0.000000
H	-1.726464	-2.795497	0.000000
H	0.828530	-4.642510	0.000000
H	-0.972775	-5.183939	0.000000

OM2-MRCI

S₀

N	0.160449	4.208101	0.000000
C	0.659809	2.987865	0.000000
C	-0.190238	1.853637	0.000000
C	0.364553	0.593770	0.000000

C	-0.453094	-0.572754	0.000000
C	0.125028	-1.808173	0.000000
C	-0.680532	-3.010894	0.000000
C	-0.089664	-4.216067	0.000000
H	-0.832732	4.378428	0.000000
H	0.739361	5.032001	0.000000
H	1.761021	2.846458	0.000000
H	-1.268250	1.988538	0.000000
H	1.464315	0.466994	0.000000
H	-1.536758	-0.457450	0.000000
H	1.223627	-1.924391	0.000000
H	-1.767238	-2.901268	0.000000
H	0.991039	-4.333609	0.000000
H	-0.670693	-5.131189	0.000000

S₁

N	0.105293	4.235689	0.000000
C	0.672479	3.022540	0.000000
C	-0.148810	1.852646	0.000000
C	0.436279	0.589720	0.000000
C	-0.400308	-0.570736	0.000000
C	0.160124	-1.850108	0.000000
C	-0.679086	-2.994716	0.000000
C	-0.132951	-4.254369	0.000000
H	-0.891015	4.357312	0.000000
H	0.652121	5.076915	0.000000
H	1.766266	2.928263	0.000000
H	-1.237697	1.967185	0.000000
H	1.522649	0.472389	0.000000
H	-1.491195	-0.443816	0.000000
H	1.250615	-1.976514	0.000000
H	-1.766365	-2.861902	0.000000
H	0.943148	-4.411296	0.000000
H	-0.761546	-5.139204	0.000000

S₁-inv.

N	0.127729	4.245920	0.000000
C	0.655060	3.009436	0.000000
C	-0.178876	1.879111	0.000000
C	0.403262	0.565470	0.000000
C	-0.416464	-0.553729	0.000000
C	0.156145	-1.859617	0.000000
C	-0.671054	-3.000622	0.000000

C	-0.113768	-4.255618	0.000000
H	-0.864162	4.399716	0.000000
H	0.702296	5.068305	0.000000
H	1.749786	2.893958	0.000000
H	-1.264911	1.988299	0.000000
H	1.495558	0.467755	0.000000
H	-1.504669	-0.449515	0.000000
H	1.252156	-1.970901	0.000000
H	-1.756781	-2.877535	0.000000
H	0.964470	-4.404707	0.000000
H	-0.735777	-5.145731	0.000000

S2

N	0.106096	4.238509	0.000000
C	0.666314	3.026730	0.000000
C	-0.152390	1.867382	0.000000
C	0.427506	0.580438	0.000000
C	-0.392304	-0.551010	0.000000
C	0.165893	-1.862459	0.000000
C	-0.665020	-2.971930	0.000000
C	-0.123113	-4.287626	0.000000
H	-0.894044	4.371493	0.000000
H	0.661932	5.080430	0.000000
H	1.758166	2.926249	0.000000
H	-1.248086	1.977543	0.000000
H	1.512599	0.472120	0.000000
H	-1.488917	-0.439280	0.000000
H	1.252132	-1.981083	0.000000
H	-1.757457	-2.853841	0.000000
H	0.945773	-4.446636	0.000000
H	-0.775078	-5.147032	0.000000

PSB5

MP2-GS

N	0.276213	4.275921	0.000000
C	-0.044700	-4.260962	0.000000
C	-0.687581	-3.042573	0.000000
C	0.038823	-1.814696	0.000000
C	-0.545484	-0.557568	0.000000
C	0.246777	0.616004	0.000000
C	-0.254734	1.913799	0.000000
C	0.635491	3.002514	0.000000

C	-0.724024	-5.536476	0.000000
C	-0.042082	-6.713560	0.000000
H	-0.704136	4.549346	0.000000
H	0.970031	5.018247	0.000000
H	1.716280	2.814435	0.000000
H	-1.336148	2.092750	0.000000
H	1.339541	0.484601	0.000000
H	-1.638498	-0.465493	0.000000
H	1.137581	-1.876555	0.000000
H	-1.784417	-3.007841	0.000000
H	1.055395	-4.274060	0.000000
H	-1.820663	-5.536092	0.000000
H	-0.565409	-7.673955	0.000000
H	1.053616	-6.729683	0.000000

CASSCF

S₀

N	0.240419	4.252146	0.000000
C	-0.025181	-4.254202	0.000000
C	-0.647565	-3.040452	0.000000
C	0.077489	-1.802485	0.000000
C	-0.506313	-0.560750	0.000000
C	0.269814	0.627733	0.000000
C	-0.247487	1.903985	0.000000
C	0.622088	3.014048	0.000000
C	-0.723123	-5.529929	0.000000
C	-0.087587	-6.720681	0.000000
H	-0.732801	4.504088	0.000000
H	0.905373	5.003799	0.000000
H	1.693314	2.853210	0.000000
H	-1.317117	2.069441	0.000000
H	1.348708	0.514086	0.000000
H	-1.585812	-0.473169	0.000000
H	1.160630	-1.863424	0.000000
H	-1.730258	-2.993218	0.000000
H	1.059301	-4.286571	0.000000
H	-1.806714	-5.501270	0.000000
H	-0.638745	-7.650514	0.000000
H	0.993440	-6.787769	0.000000

S₁

N	0.155255	4.332600	0.000000
C	-0.018529	-4.333236	0.000000
C	-0.616373	-3.024634	0.000000
C	0.132746	-1.888767	0.000000
C	-0.432501	-0.523233	0.000000
C	0.337052	0.590053	0.000000
C	-0.203495	1.950704	0.000000
C	0.602258	3.069608	0.000000
C	-0.748875	-5.526418	0.000000
C	-0.134815	-6.767073	0.000000
H	-0.822189	4.543428	0.000000
H	0.791784	5.101844	0.000000
H	1.678839	2.965403	0.000000
H	-1.279759	2.073552	0.000000
H	1.416839	0.489491	0.000000
H	-1.512476	-0.437399	0.000000
H	1.213766	-1.971668	0.000000
H	-1.697639	-2.960501	0.000000
H	1.063458	-4.393554	0.000000
H	-1.830389	-5.473878	0.000000
H	-0.715989	-7.677628	0.000000
H	0.942903	-6.860591	0.000000

S₂

N	-0.064220	4.285050	0.000000
C	0.046846	-4.347665	0.000000
C	-0.450064	-3.012528	0.000000
C	0.350153	-1.852016	0.000000
C	-0.214949	-0.563615	0.000000
C	0.511237	0.639373	0.000000
C	-0.136845	1.886812	0.000000
C	0.545434	3.118804	0.000000
C	-0.794128	-5.451474	0.000000
C	-0.353368	-6.811903	0.000000
H	-1.064969	4.361215	0.000000
H	0.457016	5.141401	0.000000
H	1.626489	3.146212	0.000000
H	-1.220846	1.911304	0.000000
H	1.593102	0.605706	0.000000
H	-1.297617	-0.494006	0.000000
H	1.428042	-1.953133	0.000000
H	-1.526707	-2.882245	0.000000
H	1.119944	-4.496827	0.000000
H	-1.864501	-5.277489	0.000000

H	-1.067645	-7.619964	0.000000
H	0.699469	-7.054908	0.000000

MS-CASPT2

S₀

N	0.258884	4.282981	0.000000
C	-0.038643	-4.273786	0.000000
C	-0.676285	-3.039728	0.000000
C	0.054456	-1.825792	0.000000
C	-0.528598	-0.549097	0.000000
C	0.260440	0.611112	0.000000
C	-0.249299	1.923726	0.000000
C	0.632546	3.015020	0.000000
C	-0.728271	-5.539076	0.000000
C	-0.056957	-6.727397	0.000000
H	-0.727177	4.544504	0.000000
H	0.942345	5.037317	0.000000
H	1.714972	2.842876	0.000000
H	-1.330718	2.094676	0.000000
H	1.351751	0.486645	0.000000
H	-1.620431	-0.457370	0.000000
H	1.151065	-1.890266	0.000000
H	-1.771438	-2.996932	0.000000
H	1.059415	-4.296000	0.000000
H	-1.823307	-5.525391	0.000000
H	-0.589645	-7.680959	0.000000
H	1.036766	-6.758962	0.000000

S₁

N	-0.045989	4.314717	0.000000
C	-0.331943	-6.780814	0.000000
C	-0.825655	-5.488879	0.000000
C	0.022953	-4.343614	0.000000
C	-0.482218	-3.037310	0.000000
C	0.333926	-1.857548	0.000000
C	-0.234835	-0.588224	0.000000
C	0.510871	0.646871	0.000000
C	-0.129274	1.887560	0.000000
C	0.569487	3.120418	0.000000
H	-1.061343	4.389763	0.000000
H	0.486959	5.180657	0.000000
H	1.664698	3.142731	0.000000

H	-1.226804	1.917424	0.000000
H	1.605507	0.596459	0.000000
H	-1.329925	-0.511938	0.000000
H	1.424319	-1.969854	0.000000
H	-1.572045	-2.904131	0.000000
H	1.110465	-4.493945	0.000000
H	-1.910213	-5.328267	0.000000
H	-1.003355	-7.643478	0.000000
H	0.746283	-6.970501	0.000000

S₁-inv.

N	-0.261103	4.345960	0.000000
C	-0.252631	-6.799611	0.000000
C	-0.842984	-5.510489	0.000000
C	-0.103740	-4.324729	0.000000
C	-0.707675	-3.007424	0.000000
C	0.030759	-1.833208	0.000000
C	-0.572747	-0.511910	0.000000
C	0.159460	0.656059	0.000000
C	-0.460752	1.954016	0.000000
C	0.287546	3.140625	0.000000
H	-1.276610	4.471089	0.000000
H	0.310752	5.192310	0.000000
H	1.384019	3.113096	0.000000
H	-1.556107	2.016589	0.000000
H	1.254829	0.607958	0.000000
H	-1.668127	-0.456227	0.000000
H	1.126613	-1.887668	0.000000
H	-1.803403	-2.946993	0.000000
H	0.993885	-4.380648	0.000000
H	-1.938079	-5.440513	0.000000
H	-0.869362	-7.701595	0.000000
H	0.836062	-6.916690	0.000000

S₂

N	-0.086369	4.314285	0.000000
C	0.047177	-4.358019	0.000000
C	-0.456504	-3.019712	0.000000
C	0.350987	-1.863126	0.000000
C	-0.213207	-0.568137	0.000000
C	0.532548	0.636697	0.000000
C	-0.111684	1.891466	0.000000
C	0.567193	3.125399	0.000000

C	-0.808248	-5.467060	0.000000
C	-0.361664	-6.813242	0.000000
H	-1.101483	4.359496	0.000000
H	0.426411	5.190244	0.000000
H	1.658687	3.174785	0.000000
H	-1.210220	1.911727	0.000000
H	1.626579	0.588959	0.000000
H	-1.308677	-0.486127	0.000000
H	1.441143	-1.973301	0.000000
H	-1.545984	-2.884178	0.000000
H	1.131881	-4.514114	0.000000
H	-1.888671	-5.279702	0.000000
H	-1.073928	-7.639959	0.000000
H	0.705905	-7.048280	0.000000

OM2-MRCI

S₀

N	0.128336	4.206783	0.000000
C	-0.009437	-4.228851	0.000000
C	-0.622527	-3.015117	0.000000
C	0.157802	-1.815260	0.000000
C	-0.434244	-0.577511	0.000000
C	0.366557	0.589020	0.000000
C	-0.199624	1.850463	0.000000
C	0.637014	2.986140	0.000000
C	-0.778881	-5.458756	0.000000
C	-0.156548	-6.646401	0.000000
H	-0.864773	4.369010	0.000000
H	0.701857	5.032644	0.000000
H	1.739398	2.856623	0.000000
H	-1.278969	1.972826	0.000000
H	1.468512	0.477544	0.000000
H	-1.519094	-0.478488	0.000000
H	1.259911	-1.909404	0.000000
H	-1.710484	-2.930428	0.000000
H	1.090822	-4.313560	0.000000
H	-1.869365	-5.378679	0.000000
H	-0.710789	-7.577499	0.000000
H	0.926400	-6.732998	0.000000

S₁

N	0.040730	4.234736	0.000000
---	----------	----------	----------

C	0.010693	-4.273673	0.000000
C	-0.569833	-3.008738	0.000000
C	0.238377	-1.840324	0.000000
C	-0.353402	-0.585250	0.000000
C	0.447615	0.600497	0.000000
C	-0.162704	1.848617	0.000000
C	0.625664	3.029681	0.000000
C	-0.801934	-5.441234	0.000000
C	-0.228138	-6.679777	0.000000
H	-0.956856	4.339096	0.000000
H	0.572811	5.084307	0.000000
H	1.723431	2.958679	0.000000
H	-1.253564	1.935874	0.000000
H	1.538253	0.508720	0.000000
H	-1.447102	-0.485010	0.000000
H	1.329018	-1.939953	0.000000
H	-1.662903	-2.904318	0.000000
H	1.104937	-4.380935	0.000000
H	-1.890982	-5.330415	0.000000
H	-0.832869	-7.580078	0.000000
H	0.850632	-6.812401	0.000000

S₁-inv.

N	-0.253470	4.303833	0.000000
C	0.309653	3.084078	0.000000
C	-0.476400	1.930821	0.000000
C	0.151440	0.634731	0.000000
C	-0.607517	-0.500546	0.000000
C	0.031966	-1.812418	0.000000
C	-0.723887	-2.956517	0.000000
C	-0.092180	-4.233674	0.000000
C	-0.858635	-5.415376	0.000000
C	-0.241417	-6.638067	0.000000
H	-1.249576	4.427662	0.000000
H	0.295579	5.143345	0.000000
H	1.409000	3.005597	0.000000
H	-1.567230	1.997182	0.000000
H	1.247242	0.586394	0.000000
H	-1.700805	-0.462179	0.000000
H	1.130873	-1.850406	0.000000
H	-1.815677	-2.909331	0.000000
H	1.008719	-4.292603	0.000000
H	-1.949258	-5.345631	0.000000
H	-0.816803	-7.557279	0.000000

H	0.841563	-6.733997	0.000000
S ₂			
N	0.030607	4.231444	0.000000
C	0.027767	-4.294470	0.000000
C	-0.550080	-2.988441	0.000000
C	0.249082	-1.847602	0.000000
C	-0.338945	-0.564794	0.000000
C	0.447744	0.594439	0.000000
C	-0.166214	1.852463	0.000000
C	0.618438	3.034513	0.000000
C	-0.787710	-5.418445	0.000000
C	-0.237882	-6.704912	0.000000
H	-0.970665	4.338350	0.000000
H	0.563196	5.085699	0.000000
H	1.714625	2.962475	0.000000
H	-1.262544	1.934887	0.000000
H	1.535956	0.513563	0.000000
H	-1.438649	-0.477915	0.000000
H	1.337797	-1.941454	0.000000
H	-1.647932	-2.895530	0.000000
H	1.116589	-4.395941	0.000000
H	-1.880376	-5.305615	0.000000
H	-0.872690	-7.578922	0.000000
H	0.833760	-6.855689	0.000000

Paper III

COBRAMM 2.0 – A software interface for tailoring molecular electronic structure calculations and running nano-scale (QM/MM) simulations

O. Weingart, A. Nenov, P. Altoè, I. Rivalta, J. Segarra-Martí, **I. Dokukina**, M. Garavelli

J. Mol. Model. **2018**, *24*, 271-301



COBRAMM 2.0 — A software interface for tailoring molecular electronic structure calculations and running nanoscale (QM/MM) simulations

Oliver Weingart¹ · Artur Nenov^{2,3} · Piero Altoè² · Ivan Rivalta⁴ · Javier Segarra-Martí^{4,5} · Irina Dokukina¹ · Marco Garavelli^{2,3,4}

Received: 20 April 2018 / Accepted: 27 July 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

We present a new version of the simulation software COBRAMM, a program package interfacing widely known commercial and academic software for molecular modeling. It allows a problem-driven tailoring of computational chemistry simulations with effortless ground and excited-state electronic structure computations. Calculations can be executed within a pure QM or combined quantum mechanical/molecular mechanical (QM/MM) framework, bridging from the atomistic to the nanoscale. The user can perform all necessary steps to simulate ground state and photoreactions in vacuum, complex biopolymer, or solvent environments. Starting from ground-state optimization, reaction path computations, initial conditions sampling, spectroscopy simulation, and photodynamics with deactivation events, COBRAMM is designed to assist in characterization and analysis of complex molecular materials and their properties. Interpretation of recorded spectra range from steady-state to time-resolved measurements. Various tools help the user to set up the system of interest and analyze the results.

Keywords QM · Interface · QM/MM · Excited states · Dynamics · Simulation · Software · Conical intersection · Photochemistry · Spectroscopy

Introduction

In the last 40 years, electronic structure calculations have become valuable tools in the interpretation and analysis of chemical reactivity. This development is not solely owed to the vast

technological advances in computer hardware but also to the evolution of efficient software to treat electronic structure problems. Numerous commercial and academic packages are available by now, some of them offering general-purpose solutions, and some being more specialized to a specific type of electronic structure methodology. The Gaussian series of programs [1], e.g., cover a large range of electronic structure methods, from semi-empirical to Hartree–Fock (HF) and density functional theories (DFT) as well as post HF methods like configuration interaction (CI), coupled cluster (CC), complete active space self-consistent field (CASSCF), and perturbation approaches (Møller–Plesset MP2–MP4 methods, CASMP2). Gaussian combines these methods with efficient tools for obtaining equilibrium and transition state (TS) geometries, enabling also reaction path following in terms of intrinsic reaction coordinate (IRC) or minimal energy path (MEP) computations in ground and excited states. The Turbomole suite [2] offers a variety of options for ground state and time-dependent excited-state density functional computations, second-order Møller–Plesset and explicitly correlated (CC) methods, applying efficiently parallelized and stable algorithms. As an extension to Turbomole, the DFT/MRCI program [3, 4] combines the efforts of density functional theory

Oliver Weingart, Artur Nenov, and Piero Altoè contributed equally to this work.

✉ Marco Garavelli
marco.garavelli@unibo.it

¹ Institut für Theoretische Chemie und Computerchemie, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

² Dipartimento di Chimica “G. Ciamician”, Università degli Studi di Bologna, Via Selmi 2, 40126 Bologna, Italy

³ Dipartimento di Chimica Industriale, Università degli Studi di Bologna, Viale del Risorgimento 4, 40136 Bologna, Italy

⁴ Université de Lyon, École Normale Supérieure de Lyon, CNRS, Université Claude Bernard Lyon 1, Laboratoire de Chimie UMR 5182, 69342 Lyon, France

⁵ Present address: Department of Chemistry, Imperial College London, London SW7 2AZ, UK

and multireference schemes, delivering accurate excitation energies for organic and inorganic compounds. Molcas [5] is specialized in the computation of CASSCF, restricted active space SCF (RASSCF), multi-state (MS)-CASPT2, and RASPT2 excited-state methods including efficient resolution of the identity (RI) routines and frozen natural orbitals (FNO) [6, 7] schemes. MOLPRO [8] offers advanced options for multi-reference CI computations, equation of motion (EOM) CCSD [9], explicitly correlated methods (F12) [10], and local approximations (e.g., LMP2-F12) [11]. The MNDO package [12, 13] collects various programs for computation with semi-empirical approaches, enabling also excited-state calculations with the specially parameterized OMx-MRCI methods [14, 15]. Hence, at this point the user can select between various types of programs, approaches, and functionalities. Often, and especially in the process of finding a suitable method for describing the system of interest, it is necessary to perform comparative studies with two or more programs. The variety of programs with different input structures and options and even different implementations of, e.g., geometry optimization routines or combined QM/MM approaches can make this a challenging task. In some cases, certain functionalities are implemented in one specific program, but the electronic structure method to perform the desired task or the specific option, e.g., for geometry optimization or (CI) gradient calculation is not available. Our implementation of QM and QM/MM interfaces in COBRAMM [16] aims to close the gaps between the existing, commonly applied electronic structure software by providing a common platform with all necessary tools to setup and analyze pure QM and QM/MM calculations in ground and excited states. COBRAMM extends the capabilities and functionalities of the discussed software by providing specific interfaces also *among* the electronic structure programs themselves. A high level of automatization helps the user to define standard calculations, still allowing advanced program options for more specific cases. In the following, we document our implementation of QM and QM/MM functionalities in COBRAMM.

Methods

General functionality

COBRAMM allows the execution of three main tasks on a geometry given in Cartesian coordinates:

- 1) Optimization procedures (including TS search, IRC, and MEP computations)
- 2) Frequency calculations for asserting the stationary nature of optimized geometries
- 3) Molecular dynamics simulations (including surface hopping for RASSCF/RASPT2 and OMx-MRCI methods)

These tasks can be performed within a pure QM or a combined QM/MM environment through the COBRAMM built-in interfaces with various QM software (Table 1), as well as the Amber [17] software for the molecular mechanics part in a QM/MM calculation (details in the section [Methods](#)).

For optimization tasks, connection to the very efficient and robust Gaussian Berny-Schlegel optimization routine [18] is provided using the Gaussian “external” communication facilities. This adaption enables usage of most provided facilities therein, i.e., computation of reaction paths (IRC), transition state optimization, and coordinate scans. Furthermore, the use of generalized redundant internal coordinates for constrained optimizations is facilitated.

Frequency calculations are performed by numerical differentiation of energy gradients. To this aim, gradients computed at each displaced geometry are processed by Gaussian (through the “external” communication facility) which generates and diagonalizes the Hessian in order to obtain normal modes and associated frequencies. Within a QM/MM environment, frequency computations are accelerated by performing explicit QM calculations only when QM atoms are displaced, whereas otherwise MM gradients are employed.

Energy gradients and non-adiabatic couplings (required for conical intersection (CoIn) optimizations and non-adiabatic dynamics simulations) can be computed numerically for methods that lack analytical implementation (such as SS- and MS-CASPT2). This is achieved by a built-in routine in COBRAMM. All numerical computations can be performed in a parallel environment using efficient task-distribution routines handled by external shell scripts, allowing facile porting to various HPC environments. More details are given in the section [Numerical routines and parallel environment](#).

Molecular dynamics simulations can be performed in two flavors: adiabatic and non-adiabatic (employing Tully’s fewest switches surface hopping algorithm) with velocity Verlet [19] and Rattle [20] integrator schemes. Their functionalities and implementation will be discussed in the section [Molecular dynamics implementation](#). Currently, surface hopping is available for CASSCF, SS- and MS-CASPT2 and OMx-MRCI methods, with details in the section [Tully surface hopping](#).

The implementation and functionality of the QM/MM routines are given in the section [QM/MM Implementation](#).

Interfaces to QM programs and Amber

Table 2 lists the current programs that can be used with COBRAMM and the availability of surface hopping and parallel-computing options for the listed modules. The COBRAMM set of routines are written in a mostly object-oriented manner and entirely coded in Python (version 2.7). Interaction between the subroutines and interfaces is realized using the Python shelve functionality, providing a

Table 1 COBRAMM interfaces to QM and MM programs and functionality

Interface	Parallel option	Berny-Optimizer	Velocity Verlet	Tully SH	Development status
Gaussian	SMP/Linda®	Yes	Yes	No	Active
Molcas	Partly MPI	Yes	Yes	Yes	Active
MOLPRO	Partly MPI	Yes	Yes	Yes	Active
ORCA	MPI	Yes	Yes	No	Basic
DFTB(+)	No	Yes	Yes	No	Basic
Turbomole	SMP/MPI	Yes	Yes	No	Active
DFT/MRCI	SMP/MPI	Yes	Yes	No	Basic
Dalton	No	Yes	Yes	No	Basic
MNDO	No	Yes	Yes	Yes	Active
Amber	MPI/GPU	Yes	Yes	–	Active

communication file to store and modify all active variables, vectors, and matrices. This strategy enables simple implementation of new interfaces and subroutines. The *cobram* main routine (Table 2) reads the user input files and generates a list of tasks, which are executed through the central CBF module. The latter decides which interfaces and further subroutines need to be accessed to perform the requested tasks.

In the separate QM and MM interfaces, input files for the corresponding programs are created and the programs are executed. Their outputs are analyzed and the computed quantities (energies and gradients for MM computations and additionally coupling vectors, dipole moments, ESP-fitted charges, and point charge gradients for the QM part) are stored and delivered back to the CBF module, where combined QM/MM quantities for gradients, energies, and charges are produced. These are, depending on the chosen task, subjected either to the optimizer (optxg) to create a new QM/MM geometry for an optimization, to the dynamics module (vverlet), to compute the next MD geometry and velocities or to the frequency module (freqxg) to produce another elongated geometry for numerical frequency

analysis. The flow chart in Scheme 1 provides an overview of the tasks and the data flow between different interfaces and subroutines.

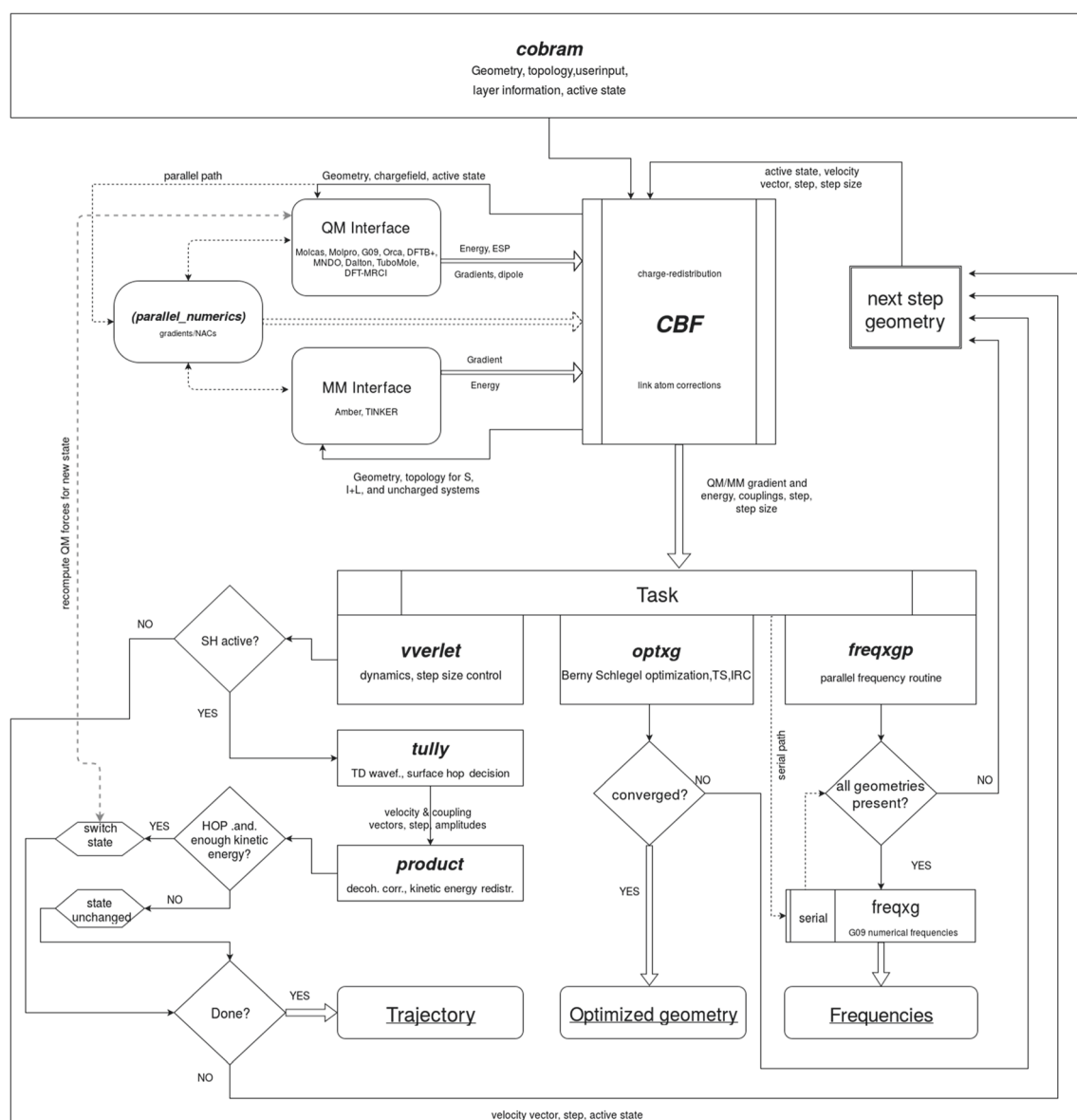
Optimization procedures

The required gradients are supplied to Gaussian in Cartesian coordinates through the *external* module. MEP computations can be initiated in internal reaction coordinates from the Franck–Condon (FC) point or from a TS. In the former case, the gradient is automatically selected as the direction of motion. At a TS, the normal mode vector associated with the imaginary frequency of a preceding vibrational analysis will be chosen. The force constants required to perform a TS optimization can be either estimated using a valence force field or taken from a previous frequency computation. CoIn optimizations are performed employing Bearpark's gradient projection method [21]:

$$\vec{g} = 2(E_1 - E_2) \frac{\vec{X}_1}{\|\vec{X}_1\|} + P_{\perp X_1, X_2}(\nabla_R E_2) \quad (1.1)$$

Table 2 Main subroutines and functions

Subroutine/Call	Function
cobram	Main routine, initialization and task creation
CBF	COBRAMM basic functions, invoke QM and MM interfaces, compute QM/MM energies and gradients, correct link-atoms, distribute charges
freqxg	Numerical frequencies as implemented in Gaussian 03/09 via g09 external command
optxg	Interface to Berny–Schlegel G09 algorithm as implemented in Gaussian 03/09 via the <i>external</i> command
parallel_numerics	Controls the numerical computation of energy gradients, second derivatives, and non-adiabatic coupling vectors (parallelized)
product	Matrix and vector products, decoherence correction, velocity scaling
tully	Tully fewest switches surface hopping algorithm, exponential propagator for time-dependent wavefunction
vverlet	Velocity Verlet and RATTLE propagators



Scheme 1 Schematic task and data flow in the QM/MM program COBRAMM. Subroutines are denoted in *bold italics*. For a clear view, explicit communication lines between *CBF*, *cobram* and the individual tasks are not shown

$E_{1/2}$ are the adiabatic energies, X_1 and X_2 are the gradient difference and the derivative coupling vectors, $\nabla_R E_2$ is the gradient on the higher state, while P is the projection on the 3 $N-8$ complement orthogonal to the intersection space spanned by X_1 and X_2 . The effective gradient is computed inside COBRAMM and forwarded to the Gaussian optimizer. In the case of a QM/MM environment, the algorithm includes the effect of the MM atoms on the structure and energy of the CoIn implicitly via the one-electron Hamiltonian (electrostatic embedding) and explicitly through the difference forces acting on the movable MM atoms in the two states subject to optimization. The only approximation is the neglect of structural deformations in the movable MM

sub-system, which couple the states as the derivative coupling for the MM part is zero due to the missing QM wavefunction description.

Basis set implementation

The interfaces of COBRAMM with Molcas and Molpro offer the capability of a per-atom basis set definition. In the case of generally contracted basis sets, as e.g., the ANO-S and ANO-L available in Molcas, the user can choose the contraction of each atom separately. In the same manner, the interface with Gaussian accepts user-defined and non-standard basis sets specified through a separate section of the COBRAMM input file using the Gaussian native format of the keyword *Gen*.

Turbomole basis sets are selected via the "define" setup routine, which already provides these features.

Numerical routines and parallel environment

COBRAMM's interface with Molcas includes a couple of numerical features. In particular, gradients and NACs can be obtained for methods that do not possess analytical implementation, as e.g., RASSCF, SS- and MS-CASPT2, RICD-MP2, and RICD-CASSCF (prior to Molcas 8). Furthermore, the atomic displacements required for numerical frequency computations can be handled through COBRAMM. Individual displacement steps are stored and can be read in again for a simple restart.

Numerical computations in COBRAMM are performed through finite differences implemented in first (3 N displacements) or second (6 N displacements) order. To ensure total energy conservation, built-in routines monitor the quality of numerical gradients during MD simulations. Gradients are considered acceptable when the ratio of the lengths of consecutive gradients does not overshoot a (user-defined) threshold. In the case of a detected failure, the time step is automatically reduced or enlarged for one step in order to avoid the problematic region. Failures can occur for various reasons: failed wavefunction convergence, divergence of the CP-MCSCF equations, or wavefunction instabilities due to strong mixing in state-average MCSCF computations. In the vicinity of CoIns, the numerics is further complicated by possible state swapping. This is counteracted by monitoring the overlap between the wavefunctions at the reference point and at the displaced geometry.

All numerical computations can also be performed in parallel. COBRAMM uses a simple parallelization scheme that relies on the subprocess module of Python, which allows to spawn processes and outsources the handling of the individual computations to external (Bash) scripts. This allows users to adapt COBRAMM's performance to the architecture of an HPC cluster. A default Bash script (freqext) is hard-linked in COBRAMM, which allows parallel execution on single shared memory multi-processor nodes. Freqext can be replaced by more elaborate scripts which handle distribution among multiple nodes without affecting COBRAMM's functionality.

Spectroscopy analysis

COBRAMM provides several tools that help in creating and analyzing spectral data from QM or combined QM/MM computations. The script-based tools range from static IR, (resonance) Raman (see ref. [22]) and UV/Vis analyses towards time-resolved techniques as e.g., fs-UV/Vis (see the section [Rhodopsin QM/MM photodynamics with](#)

[CASSCF and OM2-MRCI](#) for example). Two-dimensional excited-state spectroscopy (2DES) is a nonlinear optical technique that emerged in the last decade. It is based on sequences of ultrashort laser pulses with high temporal and spectral resolution. This method provides fundamental insight into coherent excited-state dynamics, interstate population transport, solvent reorganization timescales, and noncovalent interactions (π -stacking) [23]. Experimental 2D electronic spectra contain a wealth of (congested) information associated with time-dependent and state-specific signals or fingerprints, such as stimulated emission and excited-state absorption, calling for theoretical interpretations that face several computational challenges. COBRAMM connects to the recently developed computational protocol SPECTRON of Mukamel and coworkers for simulation of linear and non-linear spectra of complex systems like peptides or poly-nucleotides [24–27]. It is based on the sum-over-states (SOS) approach [28], where an ab initio characterization of the excited state manifolds of multichromophoric systems within a quantum mechanics/molecular mechanics (QM/MM) scheme is coupled with nonlinear response theory [23]. An in-depth description of the method can be found in ref. [29]. An example is given in the section [Spectral properties of benzophenone in aqueous solution](#).

Molecular dynamics implementation

General overview

For molecular trajectories, the classical nuclear motions are propagated in time using a velocity Verlet scheme [19]. In photo-excited systems, high-frequency motions may occur, as the system acquires kinetic energy in the excited state and returns, vibrationally hot, to the ground state. These oscillations require small time steps to properly represent their molecular motions. Often these motions do not significantly influence the reaction under study but rather act on the stability and energy conservation of the calculation [30]. To freeze out such motions, an iterative Rattle scheme [20] was implemented. Such a scheme is urgently required, when rigid solvent models as, e.g., the TIP3P model, are used within the mobile MM layer (see the section [Molecular dynamics implementation](#)). The MM parameters of this model are modified to obtain reliable structures in geometry optimizations, and propagation of the corresponding forces would lead to unphysically high frequencies for the O-H bond stretch and H-O-H bending vibrations ($\sim 4000\text{--}5000\text{ cm}^{-1}$). The Rattle algorithm therefore fixes O-H and H-H distances, and the water molecule moves as one rigid entity. Details on the implementation of velocity Verlet and Rattle schemes can be found in the SI.

Tully surface hopping

The transition probability from the current state towards the neighboring state is computed with the method introduced by Tully [31, 32]. His approach elegantly connects the individual probabilities for surface hopping with the electronic properties of all molecules in the ensemble. According to Tully's formulation, the squared coefficients for state i at time t will always reflect the proportion of trajectories that reside in this state:

$$N_i(t) = c_i^*(t)c_i(t)N \quad (2.1)$$

This relation ensures that the average electronic population of the ensemble is always equal to the ratio of trajectories in ground and excited states. Inserting this additional information into the total wavefunction equation over states (see SI) and into the time-dependent Schrödinger equations leads to Tully's surface hopping probability for a hop from state i to j :

$$P_{ij} = 2 \frac{\int_t^{t+\Delta t} \text{Re}(c_i^*(t)c_j(t)\dot{R}\cdot d_{ij})dt}{c_i^*(t)c_i(t)} \quad (2.2)$$

For integration of Eq. (2.2) between two consecutive nuclear MD steps, a unitary propagator is applied. The computed non-adiabatic coupling vectors are linearly interpolated between t and $t + \Delta t$, with a typical step size of $\Delta t/100$. The sign of the coupling vector depends on the sign of the molecular orbitals, which may be arbitrary in consecutive MD steps. It is therefore monitored and adapted accordingly to avoid artificially large numerical changes in between two time steps.

The computed probability P_{ij} for each possible hop from the current state will be compared with a uniform random number. If the random number falls below the corresponding hopping probability, a hop to that state will be invoked, with the molecule propagating from now on the new potential energy surface. Otherwise, the molecule stays in its current state. In COBRAMM, the Tully surface hopping approach is available both in the pure QM and hybrid QM/MM regimes for an arbitrary number of states. These schemes include also the possibility to switch from a state-averaged (SA) calculation to single-state (SS) mode, once the trajectory has left the coupling region and the molecule is in the ground state [33–35]. This strategy will lead to a formal drop in total energy as the wavefunction description changes, but it can avoid the typical SA convergence problems and it allows the user to follow the formation of the ground-state photoproduct with significantly less computational demand.

Coherence effects

An ensemble of independent trajectories cannot fully account for the quantum mechanical properties of a wavepacket. At a conical intersection, a real wavepacket will split, i.e., one part

of the wavepacket will reside on the upper surface, while another part moves to the lower surface [36]. The time correlation between the two separate packets is quickly lost, and they start to progress independently on the upper and lower surfaces. A single trajectory will just hop to the lower state, with no part remaining on the upper surface, and no relation to other trajectories in the ensemble. This will lead to a mismatch of populations in Eq. (2.2), i.e., the average electronic population and the real trajectory state occupation will differ, as the strict hop in each individual trajectory introduces an artificial time/state coherence.

Truhlar and coworkers suggested a scheme for an empirical correction of artificial coherence effects [37]. Such a scheme was later realized by Persico et al. [38]. It has been implemented together with the Tully surface hopping routine into the COBRAMM set of programmes (see SI for further details).

Kinetic energy redistribution

The Tully surface hopping scheme yields stochastic hopping probabilities, but often hops occur at a significant energy difference between the involved states. Excited-state trajectory simulations are typically computed within a microcanonical (NVE) ensemble, i.e., the sum of kinetic and potential energies (the total energy) must stay constant. To avoid loss of energy at a hop, the resulting energy difference is redistributed to guarantee energy conservation. Typically, the energy is scaled along a certain coordinate to adjust the kinetic energy of the system. The coordinate for redistribution is not obvious, and therefore a matter of discussion. It has been argued that the non-adiabatic coupling vector should be used for this procedure, as this coordinate provides the most efficient motion to lift the degeneracy at the conical intersection [32, 39]. If this vector is not available, the second branching space coordinate defined through the state gradient difference vector may be used, or one may simply scale the energy difference along the velocity vector [40]. Often, the coordinates defined by the coupling and gradient difference vectors have a decisive character in organic photoreactions: in the visual chromophore retinal, the coupling vector at the conical intersection contains hydrogen out-of-plane motions, which are deterministic for photoproduct generation [41–43]. As mentioned before, the sign of the coupling vector is arbitrary, thus a random choice of its direction will ultimately influence the computed product distribution when re-scaling becomes necessary. It is therefore sensible to compare the direction of coupling and velocity vectors to avoid scaling into a direction that would artificially disturb the motion of the molecule, possibly pushing it into a different reaction channel. Comparison can easily be achieved by computing the overlap or the angle between the two vectors.

There are two cases where the kinetic energy contribution provided by the non-adiabatic or gradient difference vector is not sufficient to account for the state energy difference. A

trivial situation includes a predicted hop to a state that lies above the total energy level. Such a hop is classically forbidden and is therefore rejected. In another, more rare case, velocity and scaling vectors are orthogonal. In this particular situation, scaling along positive or negative direction cannot contribute to an increase or decrease in kinetic energy, and the hop is then either rejected or the scaling needs to be performed solely along the velocity vector.

All three models for correcting the kinetic energy are available in COBRAMM for non-adiabatic molecular dynamics simulations.

Evaluation of coupling terms

COBRAMM performs mixed quantum-classical dynamics simulations in the adiabatic wavefunction representation, the native representation of quantum-chemical software. Accordingly, propagation of the TDSE (Eq. 2.3 in the SI) requires knowledge of the electronic energies and of the non-adiabatic coupling (NAC) vectors. The latter quantities drive the population transfer between electronic states in regions of near-degeneracy (avoided crossings), as well as at points of degeneracy (conical intersections). Hence, the computation of NACs is a central aspect in non-adiabatic mixed-quantum classical dynamics. It is therefore not surprising that in recent years much effort has been invested into deriving analytical formulations for obtaining the required quantities for various quantum-chemical methods like CASSCF [44], MS-CASPT2 [45, 46], MRCI [47], TDDFT [48, 49], and CC [50, 51]. Among the QM software interfaced with COBRAMM, Molpro and Molcas are capable of computing analytical state-average CASSCF NAC vectors, MNDO provides them at the OMx-MRCI level, while numerical NACs can be computed at the CASPT2 level (both single state and multi state) through the interface with Molcas.¹

In practice, the computation of non-adiabatic coupling terms can be very demanding when a larger number of excited states are involved, e.g., in a CASSCF state-average scheme. For each non-adiabatic coupling vector, a separate set of CP-MCSCF equations needs to be solved. For a system involving five states, the Hamiltonian matrix, which needs to be propagated within the Tully surface hopping scheme, takes the form:

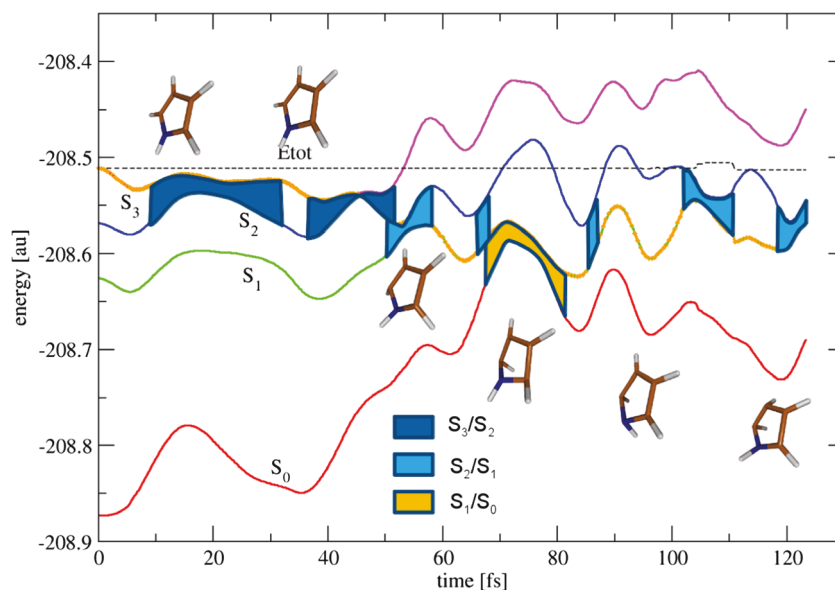
$$\begin{pmatrix} 0 & i\dot{R}d_{12} & i\dot{R}d_{13} & i\dot{R}d_{14} & i\dot{R}d_{15} \\ -i\dot{R}d_{12} & \Delta E_{21} & i\dot{R}d_{23} & i\dot{R}d_{24} & i\dot{R}d_{25} \\ -i\dot{R}d_{13} & -i\dot{R}d_{23} & \Delta E_{31} & i\dot{R}d_{34} & i\dot{R}d_{35} \\ -i\dot{R}d_{14} & -i\dot{R}d_{24} & -i\dot{R}d_{34} & \Delta E_{41} & i\dot{R}d_{45} \\ -i\dot{R}d_{15} & -i\dot{R}d_{25} & -i\dot{R}d_{35} & -i\dot{R}d_{45} & \Delta E_{51} \end{pmatrix} \quad (2.3)$$

The diagonal elements denote the state energy difference to the ground state, and the off-diagonal elements include the scalar products of the time derivative velocity vector dR/dt with the corresponding non-adiabatic coupling vector between the two states involved, multiplied by the complex unit i . The matrix is antisymmetric, i.e., to propagate the full Hamiltonian within two nuclear dynamic timesteps t and $t + \Delta t$, the computation of ten non-adiabatic coupling vectors would be necessary, each one costing about one single nuclear gradient computation. As the computation of excited-state trajectories typically involves many hundreds of single-point computations, this would be unaffordable. The resulting amplitudes for those states, which are energetically well separated from the active state, are, however, expected to be small. They are further damped within the decoherence correction scheme, so that it makes sense to concentrate only on those states that are significantly close to the current potential, and hence compute only the coupling elements for the close interacting states. For the state selection, an energy criterion specified by the user can be employed. Figure 1 shows the application of such a scheme to the CASSCF-computed photoreaction of pyrrole. The system involves four states, starting on the bright S_3 potential energy surface. With a selector threshold of 30 kcal/mol, the computation of non-adiabatic coupling matrix elements starts after ca 10 fs. Before this period, the state amplitudes are set to one for the active state and zero for all other states. With the applied selection threshold for the energy difference, it becomes apparent that no more than three interstate coupling elements have to be computed. The impact on the pyrrole dynamics due to the missing amplitude contributions of the other states is therefore expected to be negligible. The selector threshold must, however, be chosen individually and tested for each system. The presented scheme has been implemented into the COBRAMM set of routines together with a multi-state surface hopping routine.

The description of the population transfer mediated by conical intersections introduces numerical errors when performed in the adiabatic representation, the reason being that real state crossings introduce spikes in the non-adiabatic coupling localized in coordinate space. When the TDSE is numerically integrated with a finite time step and the non-adiabatic coupling exhibits a spike on a time scale shorter than the integration time step, the simulation may step over the spike, thus missing the coupling region completely. The obvious solution to this issue, namely to use sufficiently short time steps (normally around 0.25 fs), would significantly increase the computational cost. Using the same rationale as above, the energy selector

¹ Note, that as Molcas does not create SS-CASPT2 wave functions, the overlaps are obtained using the SA-CASSCF wave functions and subsequently rescaled with the factor $(E_{i,SA-CASSCF} - E_{j,SA-CASSCF})/(E_{i,SS-PT2} - E_{j,SS-PT2})$, which follows from the definition of the NAC. In the case of MS-CASPT2, the overlaps are computed using the wave functions resulting from the linear combination of the original CASSCF wavefunctions utilizing the eigenvectors obtained by diagonalizing the MS-CASPT2 Hamiltonian. See the SI for further details.

Fig. 1 Pyrrole CASSCF photodynamics using the 6-31G* basis set, averaging over four states. The *thick orange line* denotes the active state, i.e., the PES used to compute the forces. The *shaded areas* mark the computation of coupling elements



workaround can be employed in COBRAMM to introduce a shorter time step in regions where two states are energetically close.

An alternative to computing the non-adiabatic couplings (NAC) d_{ij} is to compute the time-derivative couplings (TDC) $\sigma_{ij} = \frac{dR}{dt} d_{ij} = \langle \Psi_i(r, t) | \frac{dR}{dt} \Psi_j(r, t) \rangle$ employing information about the change of the electronic wave function at the each time step. In practice, the change in the electronic wave functions is resolved by computing overlap integrals between the adiabatic wave functions at different times. This formulation elegantly solves the problem of over-stepping a CoIn, thus allowing to work with longer time steps even in regions of near-degeneracy. Originally introduced by Hammes-Schiffer and Tully (HST) as an approximation to the analytical scheme, the overlap-based approach has been shown to produce similar probabilities [47].

The TDC scheme is implemented in COBRAMM generalizing the formulation of Barbatti (Eq. 9 in ref. [47] to arbitrary time steps Δt :

$$\sigma_{ij} = \frac{1}{2\Delta t_{old}} \left(1 - \frac{\frac{1}{2}\Delta t_{old} + \Delta t}{\frac{1}{2}\Delta t_{old} + \frac{1}{2}\Delta t} \right) (\langle \Psi_i(t-\Delta t) | \Psi_i(t-\Delta t - \Delta t_{old}) \rangle - \langle \Psi_i(t-\Delta t) | \Psi_j(t-\Delta t - \Delta t_{old}) \rangle) + \frac{1}{2\Delta t} \left(\frac{\frac{1}{2}\Delta t_{old} + \Delta t}{\frac{1}{2}\Delta t_{old} + \frac{1}{2}\Delta t} \right) (\langle \Psi_j(t) | \Psi_j(t-\Delta t) \rangle - \langle \Psi_j(t) | \Psi_i(t-\Delta t) \rangle) \quad (2.4)$$

where Δt and Δt_{old} are the time steps used to propagate the TDSE between step $i-1$ and i , as well as between step $i-2$ and $i-1$. Note that Eq. 2.4 simplifies to Eq. 9 in ref. [47] when $\Delta t_{old} = \Delta t$. In this way, the TDC scheme can be combined with the adaptive time step approach introduced earlier. In COBRAMM the TDC scheme is available for the single state

and multi state flavors of the CASPT2 approach through the interface with Molcas.

QM/MM implementation

In many systems, the environment plays a crucial role in ground and excited-state chemical reactivity as it modifies the potential energy landscape of the reactive component and may thus enable different reaction routes compared to vacuum [52]. Involvement of a given protein surrounding or solvent is thus inevitable to understand ground-state chemistry, or, for photochemical reactions, the initial excitation process and subsequent energy transfer reactions. A combined quantum mechanical and molecular mechanical (QM/MM) treatment [30, 53–56] applies more accurate quantum mechanical calculations for the reactive component, while the typically much larger environment, which is not directly involved in the initial reactive process, is described by a more approximate strategy based on classical mechanics. A general scheme to treat ground-state reactions within a QM/MM scheme has already been presented in the early 1970s in the pioneering works of Karplus and Warshel [57]. Not much later, this model was extended to simulate also excited-state dynamics with surface hopping [58]. This approach enabled the accurate prediction of reaction timings and quantum yields for the rhodopsin photosystem with findings that, in general, are still valid at present.

The success of the QM/MM methodology led to a quick development of efficient codes. One may categorize these implementations by the underlying software, i.e., into QM codes with additional MM features, MM programs connecting to specific QM codes or modular software not bound to a specific QM or MM code [53]. The first two variants are extensions to existing QM or MM codes. These are often

specific to the underlying software and may not use the full capabilities of the interfaced QM or MM program. Amber, for example, connects to the semi-empirical MOPAC package [59]. The GROMACS MM code [60] has an interface to the Gaussian series of programs; the ONIOM [61] approach in Gaussian allows QM/QM and QM/MM computations within in a subtractive scheme, using all QM methods provided by Gaussian in combination with the Amber force field. MOLCAS offers a connection to the TINKER MM software [62]. A modular implementation on the other hand does not rely on one specific software package. The ChemShell suite of programs [63] e.g. provides interfaces to the QM software Turbomole [2], Gaussian [1], MOLPRO [8], ORCA [64], and MNDO [12], connecting to the MM force fields used in CHARMM [65], Amber [17], GROMOS [66], and GULP [67] through the DL_POLY [68] interface, offering flexible combinations among the programs with a common, TCL-based input language for all modules. Also the Newton-X suite by Barbatti et al. [69] offers QM/MM functionalities using diverse QM programs with the MM program Tinker.

All of the described programs are capable of modeling ground-state reactions within a QM/MM environment. A first hurdle for the user is, however, the generation of a proper molecular mechanics setup, which must be performed with the selected MM software. The preparation and analysis work gets more involved when excited-state QM/MM dynamics are required to compute, e.g., spectral properties or time-resolved spectral traces, although this is generally possible with the described packages. Our QM/MM implementation in the COBRAMM software [5, 70] is a modular approach specifically designed for usage of various excited-state QM programs with the Amber force field. In the following, we will describe the general QM/MM implementation and the specific features added to perform and analyze ground and excited state QM/MM computations.

The QM/MM energy expression

COBRAMM uses a subtractive QM/MM scheme, where the total energy is calculated as

$$E_{QM/MM}(S) = E_{MM}(S) + E_{QM}(I + link) - E_{MM}(I + link) \quad (2.5)$$

S is the whole system under study, I forms the inner region plus link atoms (link, see also Fig. 2). To obtain the total QM/MM energy, three separate computation steps are required:

- 1) Calculation of the entire systems energy at the MM level of theory
- 2) Computation of the inner system and link atom energies using a QM method

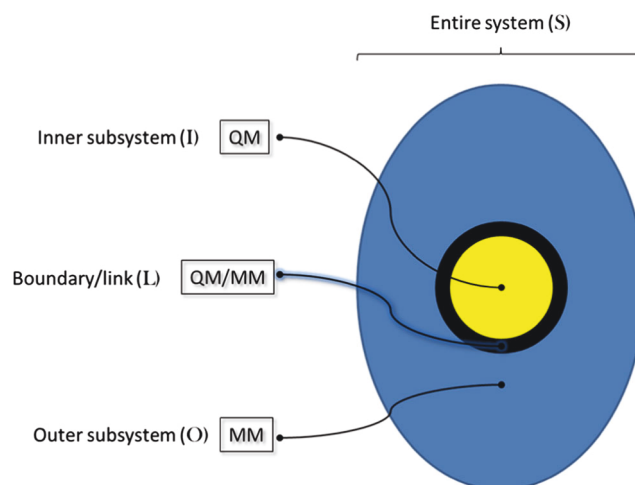


Fig. 2 Division of the total system S into QM (I), boundary (L), and outer (O) regions

- 3) Computation of the inner system and link atom energies with an MM method, and finally subtracting this term from the sum of the latter

Different from additive schemes [53], the suggested formalism requires molecular mechanics parameters also for the QM part of the system. These parameters are easily obtained within our setup procedure (see the section [Methods](#)). Details on the implemented link atom approach are found in the SI.

Layer partitioning

The total system is furthermore partitioned into three interacting regions called layers hereafter:

- The high layer (H), including those atoms that are treated with a quantum mechanics method. This layer is identical to (I + link) in the energy expression.
- The medium layer (M) – this part consists of MM atoms which are movable during optimizations or dynamics runs.
- The low layer (L) – the remaining part of the system, also treated at the MM level.

The first two layers build the mobile portion of interest of the system under study. In addition, the low layer may be frozen or optimized/equilibrated *separately* from the two other parts within a microiterative scheme (Fig. 3). This partitioning enables treatment of the core region (H + M) with an advanced optimization scheme like, e.g., BFGS [71]. Thereafter, the low layer (L) can be optimized with a much faster and more "rough" approach like, e.g., steepest descent or Newton–Raphson, and the procedure continues until the convergence criteria are fulfilled for both, H + M and L regions.

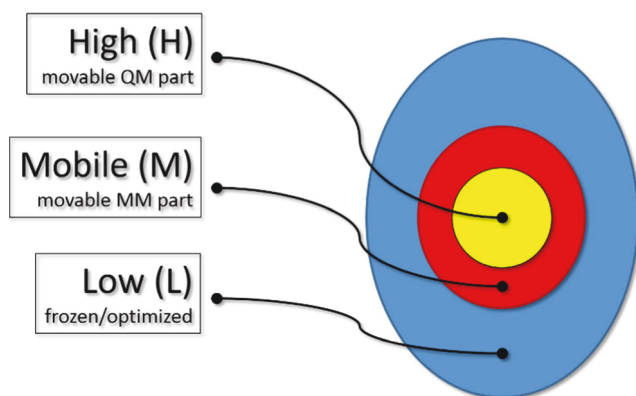


Fig. 3 H, M, and L layers in COBRAMM

Embedding of the high layer

Current QM/MM implementations define two main approaches to treat the electrostatic interactions between QM and MM regions: mechanical embedding and electrostatic embedding [53–56, 72, 73].

In a mechanical embedding scheme, the electrostatic effects between inner and outer layers are solely treated at the molecular mechanics level. This means that the quantum system does not directly interact with the charge distribution of its surrounding, but only indirect through MM electrostatics. While mechanical embedding might be a valuable approach in ground-state conformational reactions, where the charge fluctuations of the QM system are minor, it is a drastic simplification for situations where bonds are formed, broken or the electronic structure suddenly changes through excitation. In these cases, the missing interaction with the inner system wave function cannot be fully recovered by, e.g., simply updating the charges of the QM system in every step of the simulation.

In electrostatic embedding schemes (Fig. 4), the point charges of the MM subsystem enter as additional one-electron terms into the QM Hamiltonian and directly influence

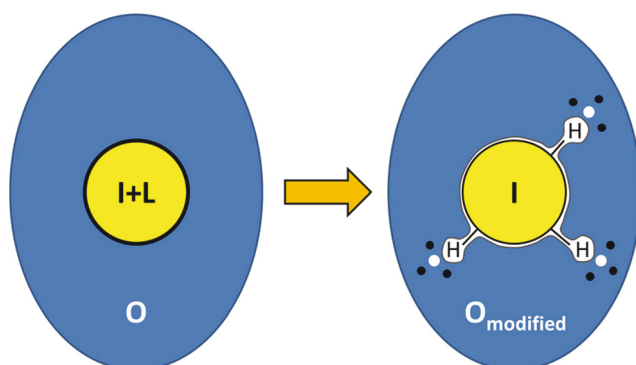


Fig. 4 Embedding of the high layer with hydrogen link-atoms. The charge field of the outer region is adapted near the links to avoid overpolarization of the QM region. Therefore, the charge of the direct link-atom neighbor (M^1) is set to zero (white circles) and the charges are repartitioned among the M^2 atoms (black dots)

the electron distribution in the QM system. Both approaches can be selected in the current version of COBRAMM, although for excited-state computations the latter scheme should be preferred.

The electrostatic interaction between QM and MM parts in the simulation may lead to overpolarization effects, which especially apply for the boundary region. Such effects can be reduced by a charge repartitioning procedure. Such an approach is discussed, together with the treatment of boundary atoms, in the next section.

The QM/MM frontier

In many biochemical systems, the reactive molecule of interest is covalently attached to its environment. This means that the bond between the Q^1 and M^1 atoms, which is not considered in the QM system (Fig. 5), would leave a free valence for this part. In the link-atom approach used in COBRAMM, the QM dangling bond between Q^1 and M^1 atoms is saturated with a hydrogen atom. The link-atoms do not appear in the MM description of the system, thus no van der Waals interactions apply for this part. Electrostatic interactions, however, must be taken care of, since the charge of the close-by M^1 atom may overpolarize the QM system, leading to strong artefacts in the simulation. To avoid this, the charge of the M^1 atom is set to zero (Figs. 4 and 5) [74], and its charge is redistributed among its neighboring M^2 atoms, weighted by the original charge contribution of the M^2 atoms [70, 75]. The motions of the link-atoms are restricted along the M^1 - Q^1 axis. Therefore, the position of a link-atom is expressed as a function of the Q^1 and M^1 positions, and their nuclear gradient components are re-distributed among the Q^1 and M^1 atoms. At this point, COBRAMM adopts a similar strategy as documented for ONIOM [61]. The placement of a link-atom introduces three additional degrees of freedom, which must be eliminated to avoid artefacts in optimization and MD simulations. Details on the implementation of the discussed link-atom scheme can be found in the SI.

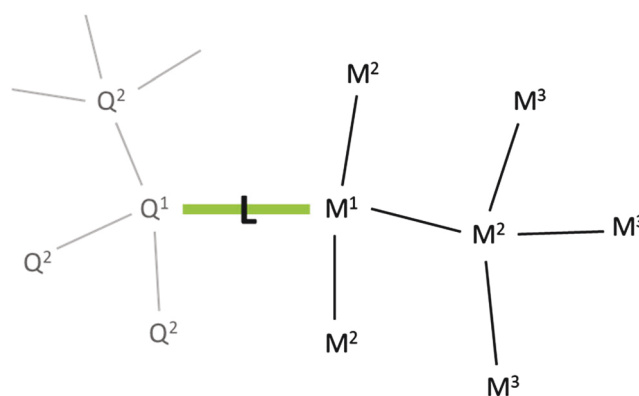


Fig. 5 QM and MM regions with link-atom and numbering

MM interface functionalities

COBRAMM can be used with all Amber distributions starting from version 10, handling also GPU, SMP, and MPI parallel options. Pmemd and CUDA versions of Ambers molecular dynamics and optimization routine “sander” efficiently speed up QM/MM computations in implicit solvation environments. MM gradients are printed using the “dump force” routine of Amber, thus without modification and recompilation of the MM software. In very large MM systems (> 20,000 atoms), this strategy can form a significant bottleneck. The process can be accelerated by computing one MD time step at the MM level. The forces can then be retrieved from the velocities of the mobile part. This option can be selected within the new version of COBRAMM.

QM/MM setup and initial conditions for MD simulations

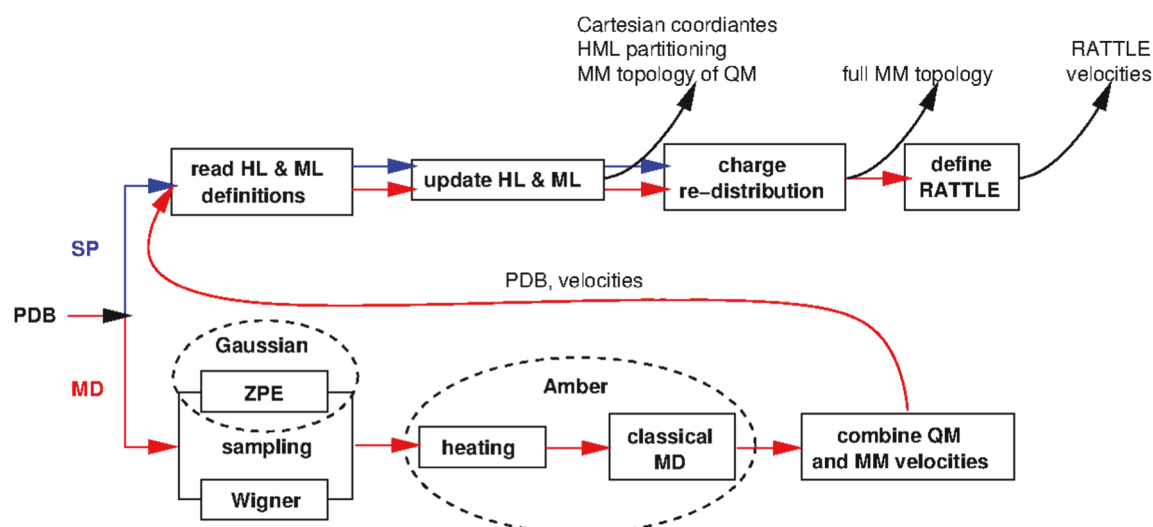
The new version of COBRAMM includes a routine that significantly reduces the user efforts in the generation of the initial QM/MM setup. This script-driven program allows the user to select the QM region of interest and can define in a semi-automatic fashion the mobile region surrounding the QM part through a distance- or H-bonded-based criterion. It also takes care for automatic setup of the Rattle scheme, when rigid solvent molecules are used in the M layer. Missing residues in the MM definition can either be added automatically or they may be loaded from a previously defined library. The distance criteria to define the mobile (M) layer include (a) the center of mass of the H or H + M layers or (b) the distance to the atoms of the H or H + M layers. H-bonds to H and H + M layers are computed and residues forming an H-bond with atoms of the specific layers can be added automatically to the corresponding layer definitions.

Several options were implemented to generate initial conditions for performing MD simulations. Starting geometries and velocities are either generated through a classical Amber MD simulation taking snapshots from a previous MM dynamics run, or they are created using quantum sampling techniques. Both techniques may also be combined. Quantum sampling requires a prior frequency computation. From the resulting normal modes, the user can create (a) a zero-point energy (ZPE) sampling [76, 77], (b) thermal sampling including also a Boltzmann-temperature distribution [76]. For the latter, we provide an interface to the Gaussian ZPE and thermal sampling options. Furthermore, we have interfaced the Wigner sampling technique applied in the quantum molecular dynamics program JADE [78], which also allows to include temperature effects and the excitation of specific normal modes. For all quantum sampling methods, high-frequency modes, which usually do not contribute to the reactions of interest and rather influence the stability of the calculations, may be excluded from sampling. In a combined QM and MM

sampling scheme, the quantum mechanically sampled geometry is subjected to an MM equilibration run, as suggested in a study of Rückenbauer et al. describing solvation effects on the photoisomerization of a retinal model [79]. This strategy allows the mobile MM region to adapt to the QM-sampled geometry and delivers velocities and a new geometry for the MM part, which, combined with the quantum sampling velocities and geometry, forms the starting pointing of a new QM/MM dynamics run. Scheme 2 summarizes the capabilities of the implemented routine.

Applications

The QM interfaces implemented in COBRAMM are mainly controlled through their own program-specific standard keywords, which are provided as a separate section to the COBRAMM input file. We decided against a common input language for all QM programs, as the functionalities and the availability of specific options in the QM programs largely differ. This way, the user can supply the input in a manner he is already used to and does not need to learn an additional language for this part of the setup. For best convenience, COBRAMM offers the capability of automatic routine-specific QM input completion. In most cases, only the definition of the computational level would fall within the user’s responsibilities. Thus, various flavors of geometry optimization (minima, TS, CoIn), minimum energy path computations (IRC), frequencies, and even dynamics simulations can be performed with the same QM input. In this way, COBRAMM lowers significantly the entry barrier for non-experts to a given quantum chemistry software. At the same time, COBRAMM provides experienced users with the freedom to manipulate the QM input according to their needs, hence the user’s input takes precedence over the automatic completion routines. The MM input generation is fully automatized and, as a general rule, does not require the user’s interference. Nevertheless, in the spirit of the COBRAMM’s philosophy, also the Amber input can be manipulated in a designated section to the COBRAMM input file. The output of each (QM and MM) program is, in any case, stored, so that additional data not covered by the COBRAMM analysis facilities can be extracted a posteriori. Table 3 gives a quick overview on the structure of the main input file *cobram.command* for two different tasks: (a) using Gaussian and the Amber interfaces for QM/MM geometry optimization and (b) using Molcas and Amber for excited-state surface hopping dynamics. The implementation of the COBRAMM QM/MM routines itself was validated by performing a comparative DFT/Amberff99 study on the rhodopsin photosystem with the Gaussian 09 ONIOM program. Geometry parameters and excitation energies of retinal in rhodopsin are reported in the SI. In the following, we discuss some features of newly



Scheme 2 Schematic task and data flow in the script-assisted preparation of the input files required for QM/MM computations. The setup for single-point calculations and optimizations is shown in *blue*, the route for initial conditions in molecular dynamics simulations is marked with *red*

implemented and updated QM and MM interfaces by means of specific examples.

Table 3 Input sections in *cobram.command* file. (a) QM/MM geometry optimization with Gaussian and Amber, (b) QM/MM excited-state dynamics with surface hopping using Molcas and Amber. The geometry of the system and the layer partitioning are specified in separate files

(a)	(b)
!keyword type=optxg qm-type= gauss mm-type=amber qmem=1000mb nproc = 8 ?keyword	!keyword type=mdv qm-type=molcas mm-type=amber qmem=1000mb surhop=persico tstep=0.25 ?keyword
!sander real input for whole system S &ctrl imin = 1, maxcyc = 0, ntb = 0, igb = 0, ntr = 0, ibelly = 1, cut = 12, ntxo = 1 / ?sander real	!sander real input for whole system S &ctrl imin = 1, maxcyc = 0, ntb = 0, igb = 0, ntr = 0, ibelly = 1, cut = 12, ntxo = 1 / ?sander real
!gaussian #P CAM-B3LYP/TZVP Nosym 0 1 ?gaussian	!molcas &RASSCF Symmetry = 1 Spin = 1 nActEl = 10 0 0 Clroot = 2 2 1 Inactive = 40 RAS2 = 10 RLXR = 2 LumOrb ?molcas

Rhodopsin QM/MM photodynamics with CASSCF and OM2-MRCI

The photoreceptor protein rhodopsin triggers vision in mammals through light-induced 11-*cis* to all-*trans* isomerization of the retinal chromophore. Due to the size of the system, the simulation of retinal photodynamics within the protein environment forms a challenging task. By now, simulations were performed with full and reduced active-space CASSCF methodology, describing ensembles with up to 60 individual trajectories with retinal in the QM part of the calculation. In a combined experimental and computational study, we were able to trace the time evolution of spectral signatures after rhodopsin photoexcitation and thus could follow the molecular motions of the retinal chromophore within opsin [80].

Figure 6 shows the results of 38 CASSCF/Amber trajectories in comparison with the experiment. Surface hopping MD computations were performed with the COBRAMM Molpro and Amber interfaces (see Ref. [80] for details on the computational setup). The resulting energy differences and oscillator strengths of the individual trajectories were scaled and processed via Gaussian convolution to reproduce the experimental 15-fs time resolution in a pump probe experiment. In such an experiment, the pump pulse excites the molecules, and a probe pulse induced after a certain delay stimulates the molecules to emit (blue signal in Fig. 6). The emission wavelength then gives information about the distance of the potential energy surfaces. When the molecule is in the ground state, the probe pulse is absorbed (orange/red signatures in Fig. 6). Along the time axis in Fig. 6, the blue emission signal shifts to higher wavelengths, as the chromophore undergoes rotation about the reactive bond and the S_1/S_0 state energy gets smaller. After reaching a conical intersection, the molecule returns to

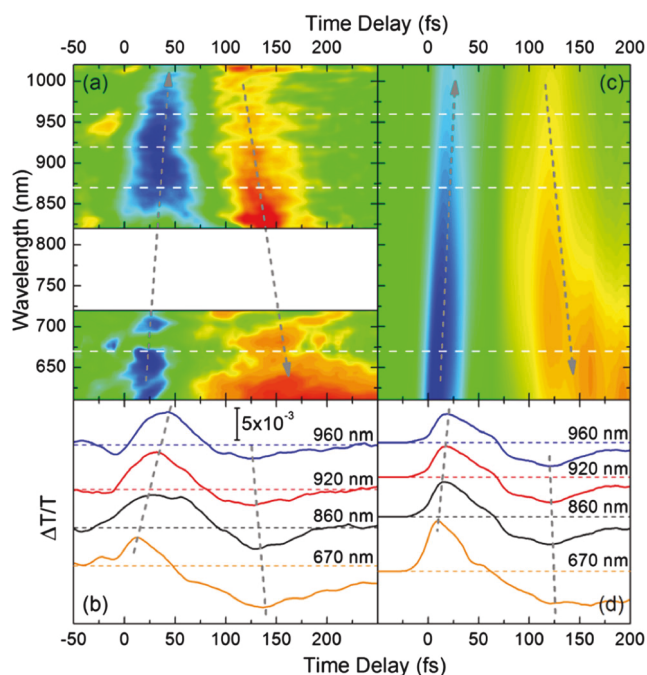


Fig. 6 Experimental (left) and simulated pump-probe spectra of rhodopsin. The blue signals in time-resolved spectra (a) and (c) correspond to stimulated emission, yellow and orange traces denote chromophore photoabsorption. The lower graphs (b) and (d) provide horizontal cuts at selected wavelengths. Figure adapted from Ref. [80] with permission from Nature

the ground state and the state energy difference drops. This can be seen by the occurrence of the orange signal in the right part of the spectra.

Semiempirical methods can significantly reduce the computational effort, provided a suitable method can be found that is able to reproduce the results of the *ab initio* calculations. The OMx methods in the MNDO program by Thiel et al. [12, 13] are, in combination with efficient GUGA-CI [81] routines, specially parameterized for the computation of excited-state properties such as excited-state energies, gradients, and non-adiabatic couplings used for surface-hopping dynamics simulations or geometry optimizations [14]. This has been shown in the latest studies of Thiel and coworkers [15, 82–84]. Orbital overlap procedures make sure that the chosen orbitals remain in the active window during optimization and MD simulation. Within COBRAMM, this option is also available for QM/MM simulations. To reduce a possible drift in the total energy of OMx-MRCI dynamics computations, we have furthermore implemented a variant of the adapted time-step algorithm suggested by Spörkel et al. [85] into our velocity Verlet scheme.

In the following, we report a benchmark study on the photoisomerization of retinal in the rhodopsin protein using OM2-MRCI methodology.

For computations on rhodopsin, the full retinal system including one CH₂ group of Lys296 was used as the QM part, and the remaining system was described by the Amber ff99

force field. The structure of retinal was re-optimized at the OM2 RHF level with a frozen protein environment. As an initial test, a trajectory starting from the S₁ FC region without initial velocities (at 0 K) was computed. This state also corresponds to the HOMO-LUMO single excitation in OM2/MRCI. The protein environment was kept frozen at the crystal structure coordinates during the whole simulation. As active space for excited-state OM2/MRCI computations, the highest ten occupied and the lowest ten unoccupied orbitals were chosen, corresponding to 20 electrons in 20 orbitals. This selection covers occupied and unoccupied π -orbitals, but also includes lower-lying occupied and higher-lying unoccupied σ -orbitals that may become important when the system starts to twist along the reactive coordinates. Single and double excitations were considered in the MRCI expansion. State energies were computed for the ground and the first two excited singlet states. All related derivative couplings were calculated to obtain state amplitudes and hopping probabilities in the decoherence-corrected Tully surface hopping scheme.

Figure 7 shows the evolution of several parameters of this MD in comparison to the CASSCF(10,10) and CASSCF(12,12) MD data from earlier CASSCF work reported in Ref. [80]. We find reasonable agreement between the two approaches. In both, OM2/MRCI and CASSCF computed trajectories, a conical intersection point is reached via the C11 = C12 torsion coordinate and the molecule decays to the all-*trans* isomer. The evolution of torsion angles looks very similar; this is also true for the progression and steepness of the S₁ potential. From the FC point to the CoIn, it drops by about 20 kcal/mol in both descriptions. The distance to the S₀ potential is, however, significantly smaller in OM2/MRCI. At the FC point, its value reduces from about 80 kcal/mol in CASSCF to about 60 kcal/mol in the OM2/MRCI description.

This corresponds to a vertical excitation energy of 2.66 eV (466 nm), which is much closer to the experimental value of 500 nm than in the CASSCF description. The modified distance to S₀ may also be the reason for the faster decay in OM2/MRCI. With this method, the crossing point is reached already after 57 fs, compared to 73 fs in CASSCF. We also note a smaller amplitude in the C=C stretchings. For OM2/MRCI, the first maximum in the C11 = C12 coordinate is reached at 1.48 Å and ca. 15 fs, while CASSCF places it at 1.54 Å and 10 fs. A subsequent OM2 numerical frequency analysis delivered the normal modes to create 40 initial conditions using Wigner sampling, constituting a similar ensemble size as in the earlier study. The high-frequency modes were not considered in this procedure. OM2/MRCI trajectories starting in the S₁ state were followed for up to 250 fs, using a constant time step of 0.25 fs. Figure 8 reports the computed UV spectrum and the population decay in the trajectory ensemble. In the UV spectrum from Wigner sampling, a significant red shift of the maximum absorption occurs with respect to the computed vertical excitation. The observed value (492 nm) is even closer

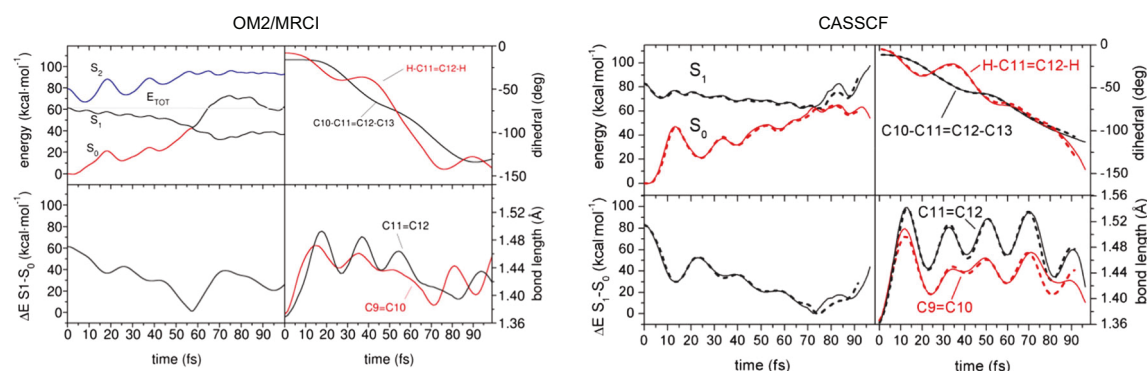


Fig. 7 Selected MD calculated parameters for OM2/MRCI (*left*) and CASSCF (*right*) trajectories. The left graph additionally reports the evolution of the covalent S_2 state and the total energy. The *dotted lines* in the

right graph denote a CASSCF(12,12) active space, the *full lines* correspond to the CASSCF(10,10) trajectory

to the experiment. In this spectrum, also the S_2 absorption corresponding to HOMO-LUMO double transition becomes apparent, peaking at about 370 nm (361 nm for the vertical excitation).

The S_1 population sharply drops after the start of the reaction. This drop is connected with a small raise in S_2 population, i.e., according to the OM2/MRCI description this state participates in a small number of ensemble members. The S_0 population begins to rise slowly after ca. 20 fs. After 50 fs, it quickly increases until ca. 75 fs, then it grows with smaller slope. After 180 fs, all trajectories are in the ground state.

The average excited-state lifetime was ca. 72 fs, and the computed quantum yield corresponds to 0.65. These values are again in good agreement with the data obtained by earlier CASSCF computations, ranging from 0.58–0.75 for the quantum yield and 80–110 fs for the S_1 lifetime. They also compare well with the experiment. In conclusion, the OM2/MRCI method delivers reasonable results for static and dynamic QM/MM computations of rhodopsin. It may provide further insight into the dynamics of more complex photosystems at significantly reduced computational expenditures.

Spectral properties of benzophenone in aqueous solution

Benzophenone is a well-known triplet sensitizer and was already used for DNA sensitization [87, 88]. Its efficiency in triplet sensitizing is explained by fast internal conversion to S_1 , followed by efficient S_1/T_1 intersystem crossing [89]. Figure 9 shows the DFT/MRCI-computed absorption spectrum, obtained from the CAM-B3LYP/TZVP optimized ground-state structure in PCM water solution. Its brightest singlet state absorption in the 225–300 nm range (middle UV, MUV) can be attributed to S_0 - S_4 excitation, providing a mixture of HOMO-3 \rightarrow LUMO, HOMO-1 \rightarrow LUMO and HOMO \rightarrow LUMO π - π^* excitations. The lowest singlet state (S_1) has n - π^* character (HOMO-4 \rightarrow LUMO, see Fig. 9). Experimental evidence for singlet interconversions and S_1/T_1 ISC was provided in the study of Shah et al. [89]. Optimized geometries for benzophenone in aqueous solution have been obtained using the Turbomole [2] interface. QM/MM computations with this interface require the "define" set-up routine to be executed once for the QM part. The

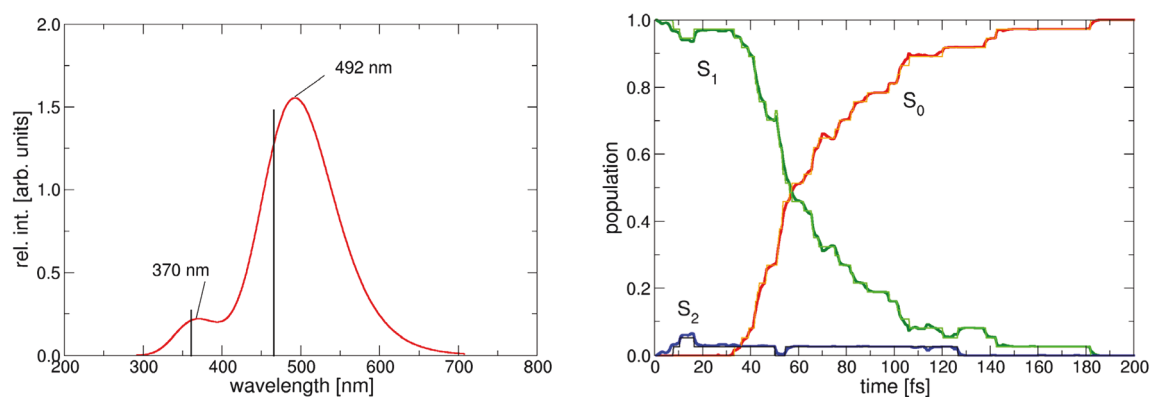


Fig. 8 *Left*: UV spectrum from Wigner sampling including 40 structures. The *lines* denote the computed vertical excitation energies of the S_1 (466 nm) and S_2 states (361 nm). *Right*: Populations of S_0 , S_1 and S_2 states in the trajectory ensemble

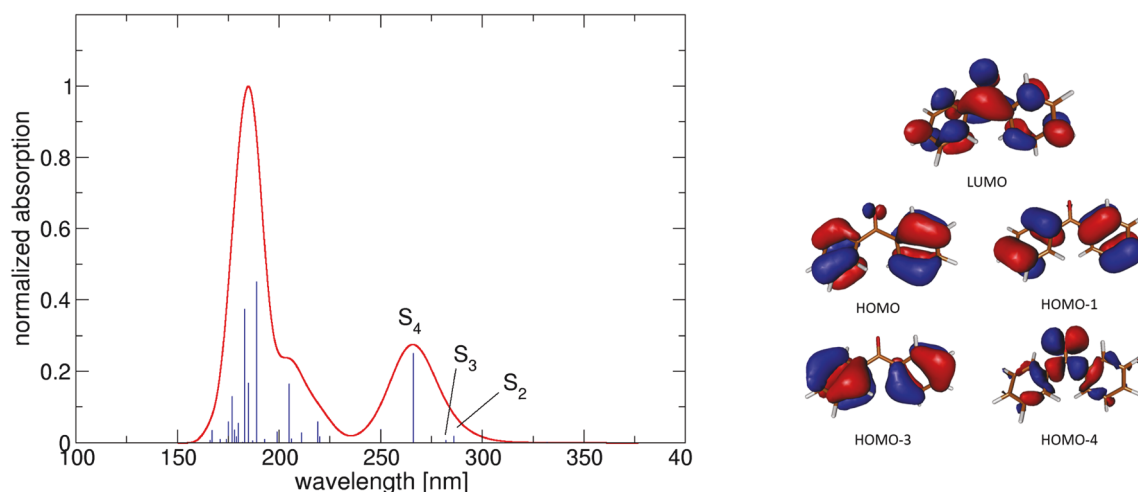


Fig. 9 *Left*: DFT/MRCI computed absorption spectrum of benzophenone in water using the COSMO solvation model [86], convoluted with a Gaussian function of 1500 cm^{-1} (FWHM) and converted to nm. *Right*: BH-LYP orbitals involved in S_1 and S_4 excitations

chargefield of the MM part is then handled by the internal COBRAMM routines.

Figure 10 depicts S_0 , S_1 , and T_1 geometry optimizations of benzophenone (BP) in a water droplet consisting of 500 TIP3P water molecules, computed using the BH-LYP functional and the SVP basis set. Two water molecules H-bonded to the oxygen of BP were included in the QM part of the calculation. Geometry data was obtained from a previous Amber ff99 equilibration run with periodic boundary conditions [90]. All optimizations were started from the same, OM2/Amber-pre-optimized ground-state geometry. S_1 and S_0 state geometries converge quickly within 70–80 optimization cycles, the T_1 geometry taking significantly longer. The DFT/MRCI [3, 4] interface allows subsequent single-point calculations to obtain transient absorption data for benzophenone in the relaxed S_1 ($n\text{-}\pi^*$) and T_1 ($\pi\text{-}\pi^*$) states.

Figure 11 shows the corresponding values in comparison with experimental data. We observe a slight overall blue-shift,

but the computed data clearly reproduce the trend in short and longer time transient absorption. The short time absorption signal stems from S_1 , then the system quickly reaches an S_1/T_1 intersystem crossing, efficiently converting the complete S_1 population towards T_1 .

Further insight into benzophenone photoreactivity in water can be provided by means of 2D electronic spectroscopy. Figure 12 displays a simulated two-dimensional excited-state (2DES) spectrum, based on one excited S_1 state trajectory computed with the TD-B3LYP method (see Ref. [90] for details). The spectrum itself was created by using the RASPT2 method to obtain estimates for vertical excitation energies and associated transition dipole moments on top of the DFT-computed geometries out of the trajectory. The active space included all bonding π valence orbitals and the n_O orbital into the RAS1 subspace, while all anti-bonding valence π^* orbitals comprised the RAS3 subspace. Five holes/electrons were allowed into the RAS1/RAS3 subspaces, respectively, leading

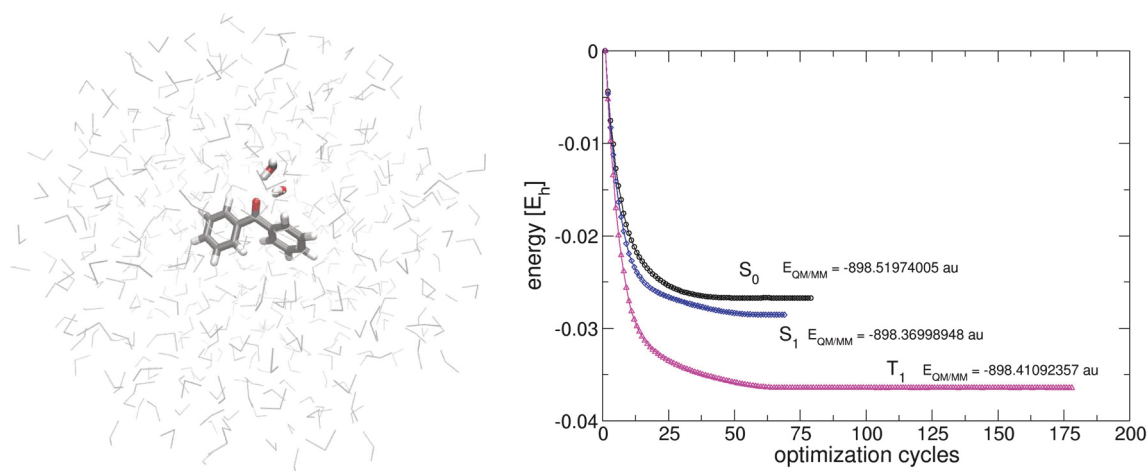


Fig. 10 *Left*: Benzophenone in water droplet with 500 solvent molecules. Two hydrogen bonded water molecules (licorice) were included into the high layer and optimized together with BP. *Right*: Optimization cycles with Bery–Schlegel optimizer and resulting DFT QM/MM energies

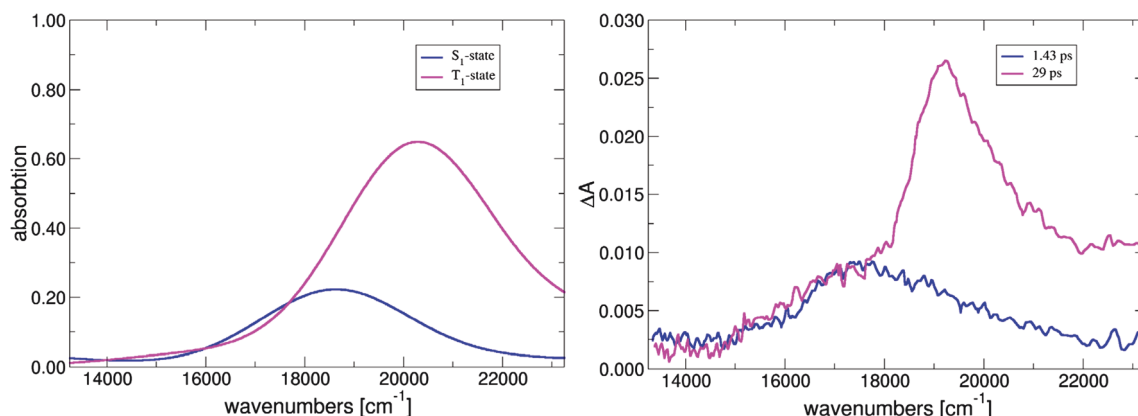


Fig. 11 *Left:* DFT/MRCI computed transient absorption spectrum of BP in a water droplet for S_1 and T_1 states, convoluted with a Gaussian function of 1500 cm^{-1} (FWHM). *Right:* experimental transient absorption

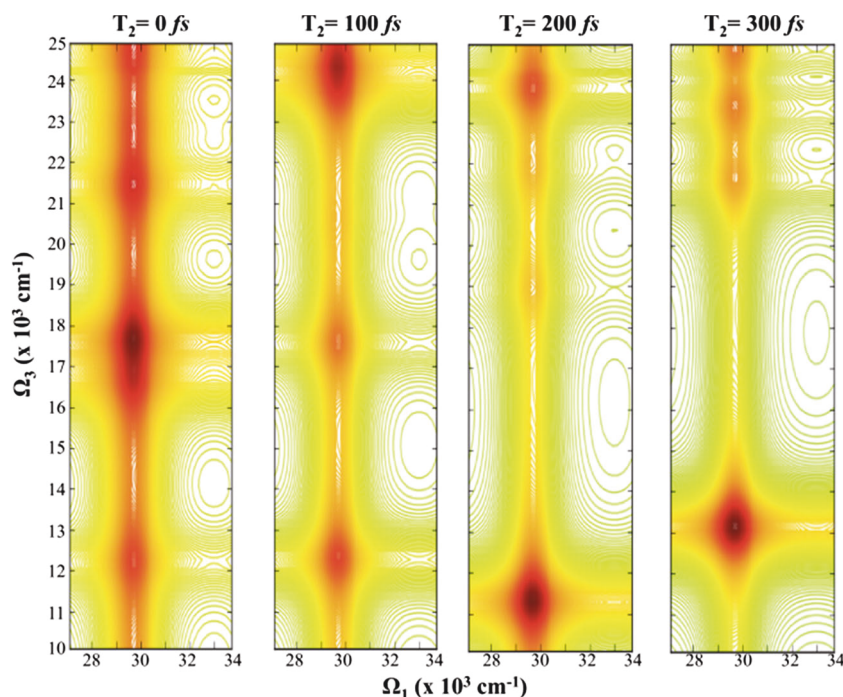
spectra measured at 1.43 and 29 ps after excitation. Experimental data from the right-hand side panel was taken from Ref. [89] and converted to cm^{-1}

to RASPT2(5,8|0,0|5,7). Atomic natural orbital basis of large type (ANO-L) were employed in their triple- ζ (VTZP) contraction. This level of theory was employed as it was found to be the best compromise between accuracy and computational cost in order to represent the electronic excited state manifold of benzophenone [91].

Figure 12 shows 2D spectral traces right after photoexcitation (0 fs) and in three time intervals (100, 200, 300 fs) afterwards. Initially, (at time = 0 fs), the signal consists of slightly shifted analogous fingerprints arisen from the $S_1(n_O\pi^*)$ state and thoroughly characterized in gas phase [91], depicting dipole allowed $n_O\pi^* \rightarrow n_O\pi^*$ transitions, with its main fingerprint being placed at $\sim 17,500\text{ cm}^{-1}$. A focus is taken in the Vis probing window as it provides fingerprints that can also be registered with linear-response-based methods

[90]. After the first 100 fs, this main fingerprint is red-shifted towards the 12–14 k cm^{-1} range, while blue-shifting the other intense transitions registered at the FC region. This strong shift is encompassed by pronounced carbonyl stretchings occurring along the $S_1(n_O\pi^*)$ relaxation path that stabilize (red-shift) the excited state absorption signal while blue-shifting other higher-lying $n_O\pi^*$ contributions. At later times (200 and 300 fs snapshots), it can be seen how the main intense excited-state absorption fingerprint placed at 11–13 k cm^{-1} oscillates with the stretching of the carbonyl as well as the appearance of a series of relatively dark signals in the high-energy window associated to doubly excited $\pi\pi^*$ states, which are partially dipole allowed due to the geometrical deformations along the trajectory that breaks the pure character of the $S_1(n_O\pi^*)$ state making them accessible. Monitoring

Fig. 12 RASPT2 computed 2DES of benzophenone along a representative TD-B3LYP trajectory. Ω_1 refers to the pump frequency whereas Ω_3 is the probe frequency. The red color denotes positive (excited state absorption) signals. Figure reproduced from SI of Ref. [90] with permission from the Royal Society of Chemistry



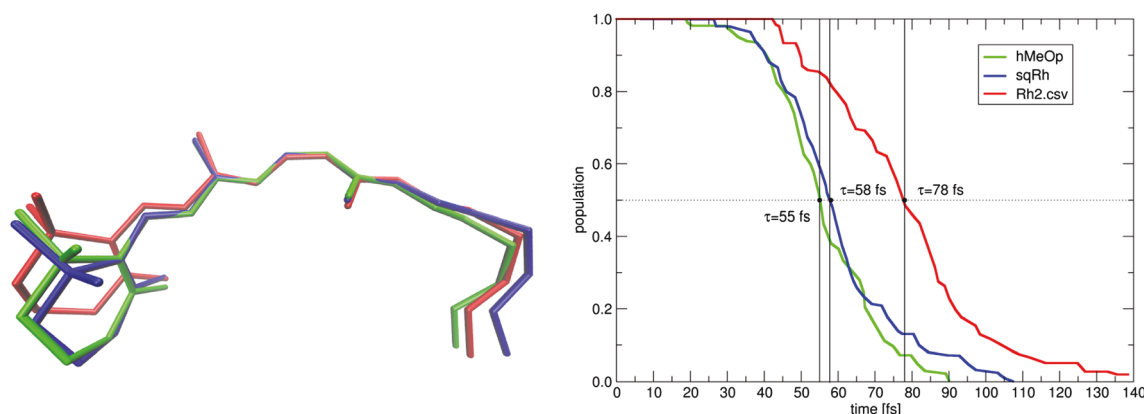


Fig. 13 *Left:* Overlay of CASSCF optimized S_0 retinal structures in bovine Rh (red), sqRh (green) and hMeOp (blue). *Right:* Decay of S_1 state population in each protein from average electronic population of individual trajectories using the same color coding

these excited-state absorption signals thus provides a useful additional way to track excited state dynamics and where their specific broadening, omitted in that study due to computational demands, can provide further information regarding the specific solvation effects and excited state lifetimes.

Excited-state lifetime and quantum yield in animal and human rhodopsins

As a primary photoactive component, the chromophore 11-*cis* retinal is also found in human rod, cone, and ganglion receptor cells as well as in numerous invertebrates. In this study [92], we have investigated the influence of the different environments on the structural, spectral, and dynamic properties of 11-*cis* retinal in the human melanopsin (hMeOp) [93] and squid rhodopsin (sqRh). [94] Although stemming from different species, these two proteins share significant sequence similarities. These analogies allowed the construction of a hMeOp homology model based on the known crystal structure of sqRh [95]. The results of excited-state surface hopping molecular dynamics performed with these systems were compared to the photodynamics of bovine rhodopsin. 11-*cis* retinal was geometry-optimized in the static environments of the sqRh and bovine Rh crystal structures as well as in the hMeOp homology model applying CASSCF methodology. The active CAS space comprised 12 electrons and 12 π -orbitals, using the MOLCAS set of routines and the 6-31G* basis set. The QM model of retinal is the same as already described in the section [Rhodopsin QM/MM photodynamics with CASSCF and OM2-MRCI](#). To describe the fixed protein environment, the Amber FF96 force field was used. The left-hand side of Fig. 13 depicts an overlay of the obtained retinal structures. The geometries were aligned along the central C9-C14 fragment. The retinal chromophores in sqRh (yellow) and hMeOp (blue) are significantly bent with respect to bovine Rh (red). In combination with the electrostatic effects

provided by the protein environment, these deformations lead to shifts in their UV absorption spectra. MS-CASPT2 excitation energies were computed on top of a room-temperature geometry ensemble obtained from thermal sampling using 60 snapshots for each system. The results reproduce the experimental trend: with 481 nm the bright absorption in sqRh is significantly blue-shifted compared to bovine Rh (503 nm). hMeOp (476 nm) is even more blue-shifted.

Earlier single-trajectory computations [95] revealed an additional difference in photoreactivity: for the blue-shifted opsins, a shorter excited state lifetime was observed. According to the one-dimensional Landau–Zener model, this reduction in excited-state lifetime should be connected with an increase in quantum yield for the all-*trans* isomer. To test this hypothesis, 60 excited-state surface-hopping trajectories from the sampled geometries were started using a reduced CASSCF(10,10) active space, averaging equally over the ground and first excited state. These computations were performed by taking advantage of the QM/QM interoperability of COBRAMM. Excited-state gradient computations were done with MOLCAS, while for non-adiabatic coupling vectors the resulting wavefunction was converted and the MOLRPO interface was executed.² The main results of this study are depicted in the right-hand side of Fig. 13. The population graph shows the decay of the excited state in all sets. With 78 fs (using the time, where half of the molecules have decayed), Rh has the longest excited-state lifetime, sqRh and hMeOp decaying 20–23 fs faster. These findings are in line with earlier single trajectory computations started from the Franck–Condon point without initial velocities. Figure 14 gives more detailed information on the S_1 lifetime and includes statistical data collected in periods of 10 fs. The height of the bar graphs denotes the portion of trajectories leading to the all-*trans* photoproduct (green) and back to the 11-*cis* starting configuration (red) within a

² It should be noted at this point, that the new version of MOLCAS includes analytic derivative couplings for CASSCF wavefunctions.

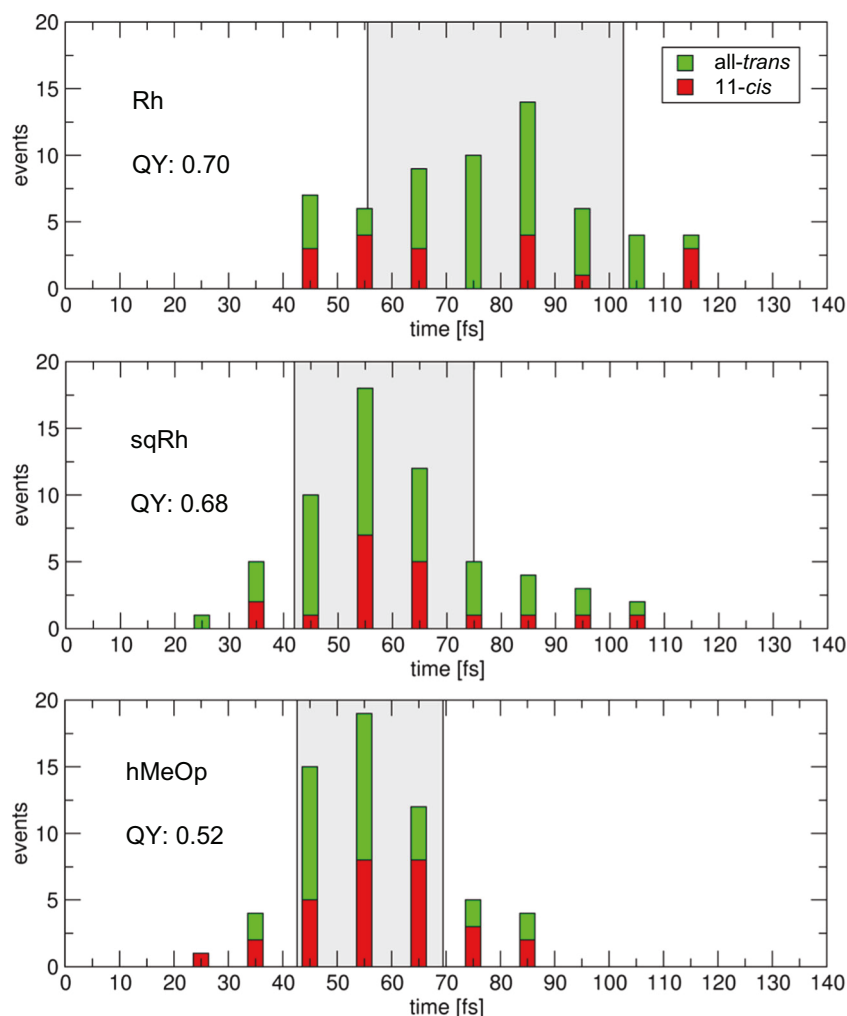
particular 10 fs time frame. As already apparent from the overall all-*trans* retinal quantum yields (QY), which are also given in this figure, the expected increase due to higher reaction speed was not observed. Instead, sqRh and in particular hMeOp feature a reduced QY.

The lifetime statistics graph also reports the decay time window by means of the standard deviation of the hopping times of all ensemble members (grey areas in Fig. 14, right). These time windows start much earlier and have a smaller width for sqRh and hMeOp than in the corresponding bovine Rh, i.e., the hopping area is much more localized and it is shifted towards shorter times with respect to Rh. This behavior is a consequence of the bend introduced in the ground state structures of sqRh and hMeOp, (Fig. 13), which produces an additional strain in the polyene fragment. This strain is more effectively released and converted to torsional motion along the C11 = C12 *cis* bond in the excited state, thus speeding up the approach towards the conical intersection. Product formation involves at least two coordinates: the torsion about the C10-C11-C12-C13 carbon fragment and, as the decisive mode, torsion around the corresponding H-C11 = C12-H

dihedral angle. In other words, the chromophore must be sufficiently twisted around C10-C11-C12-C13 to approach the CoIn, then the direction of the H-C=C-H motion will define, whether the molecule follows this twist in the ground state to lead to all-*trans*, or whether it will return to the 11-*cis* starting configuration. This rule, which was stated also in earlier studies [41–43], is applicable for the vast majority of our computed trajectories and can explain the shift in quantum yield in hMeOp. In this chromophore, the carbon and hydrogen torsions run out of phase as a consequence of the additional strain in the chromophore.

Figure 15, right, shows averaged values of the corresponding carbon and H-C=C-H torsions denoting the increase in reaction velocity along the C-C-C-C motion for hMeOp and sqRh. A speedup can also be seen for the hydrogen torsion, but the oscillations in this mode are mostly unaffected (see min. and max. lines in Fig. 15). This leads to a mismatch of the two angle velocities, i.e., at those carbon torsion angles spanning the decay window, the H-C=C-H mode reverts in many hMeOp trajectories, apparent by the plateau in Fig. 15 left (blue arrow). In the

Fig. 14 Distribution of excited-state lifetimes. The height of the bar graphs denotes the portion of trajectories leading to the all-*trans* photoproduct (*green*) or to the 11-*cis* reactant (*red*). The *grey area* denotes twice the standard deviation of hopping event times



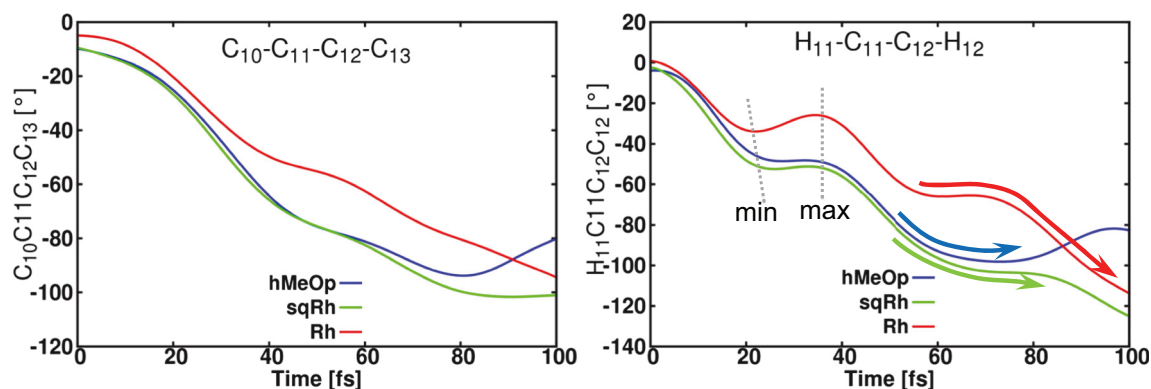


Fig. 15 Averaged values of C11 = C12 carbon (*left*) and hydrogen torsion over all trajectories. The *arrows* denote the direction of the H-C=C-H torsion during the hopping window (Fig. 14)

corresponding time window for Rh (red arrow in Fig. 15), the progression in H-C=C-H is mostly towards larger negative values, similar findings apply to sqRh.

During these calculations, another important parameter controlling product generation was revealed. This parameter regards the change in wavefunction character in the ground state, which is connected with the bond length alternation coordinate (see Ref. [92] for details). In conclusion, the photoreactivity of the studied rhodopsin proteins does not follow a simple one-dimensional Landau-Zener model, which predicts higher quantum yields when the reaction progresses faster along the reactive coordinate. The final outcome is rather related to the phase combination of the decisive modes. A bend in the chromophore structures may speed up the torsion along the carbon fragment, but it has less influence on the involved H-C=C-H motion, leading to phase mismatch and less quantum yield in hMeOp.

Photoinduced H-transfer in protochlorophyllide oxidoreductase

The interface of COBRAMM with Gaussian opens the possibility of utilizing its versatile set of density functionals for ground (DFT) and excited (TD-DFT) state problems. These include a number of hybrid and double-hybrid functionals, as well as those specifically developed to treat dispersion and long-range correlation effects, allowing for QM/MM partitioning with hundreds of atoms in the QM layer. The capabilities of the COBRAMM/Gaussian interface were recently exploited to refine the ternary structure of the light-driven protochlorophyllide oxidoreductase (LPOR) enzyme of *Thermosynechococcus elongatus*. LPOR plays a crucial role in the light-dependent synthesis of bacteriochlorophyll (BChl) in most photosynthetic organisms. It catalyzes the penultimate step of a complex hydride transfer reaction cascade that reduces protochlorophyllide (Pchlide, PCL) to chlorophyllide (Chlide, CHL). The

catalytic reaction involves a light-driven sequential step-wise hydride transfer from the pro-S face of NADPH coenzyme to the C₁₇ position of PCL [96–99] followed by the thermally activated protonation at the C₁₈ position of PCL with a proton, presumably coming from a highly conserved Tyr residue in the vicinity of PCL [98, 100, 101] (see Fig. 16). Due to the lack of an LPOR crystal structure, a comprehensive understanding of the reaction pathway at room temperature is still far from reached. Surprisingly, there is not yet a consensus even on some fundamental questions about the geometry of the intermediate state(s) and about the structural changes that lie at the origin of the photoactivation process. In the past, all structural/mechanistic interpretations of the available spectroscopic data have been based on different homology models, without ever questioning their validity. We developed a protocol that is in many ways superior to homology modeling as it extends beyond the alignment of related amino acid sequences.

Starting from an existing ternary model of NADPH bound LPOR from *Synechocystis* sp. (ssPOR) [102] with docked PCL in the active site, the protocol couples various flavors of molecular modeling techniques—molecular mechanics simulations, hybrid quantum mechanics/molecular mechanics calculations, computational spectroscopy—and incorporates the feedback from experiment in a synergetic pipeline in order to assess the fitness of the model on multiple levels and, fundamentally, to suggest improvements. In particular, we made use of the pronounced sensitivity of the high-frequency stretch mode ($\lambda_{\max} \sim 1700 \text{ cm}^{-1}$) of the keto group in the cyclopentanone ring (at C₁₃) of PCL to changes in the electronic structure of the conjugated system. The keto-mode has been shown to exhibit a 14 cm^{-1} red-shift upon reduction of the C₁₇-C₁₈ double bond in water [103]. The keto group also proves an excellent marker for the hydrogen bonding environment, exhibiting a remarkable 60 cm^{-1} red-shift going from the polar aprotic tetrahydrofuran ($\lambda_{\max} = 1710 \text{ cm}^{-1}$) [104] to the LPOR enzyme environment ($\lambda_{\max} = 1650 \text{ cm}^{-1}$) [103],

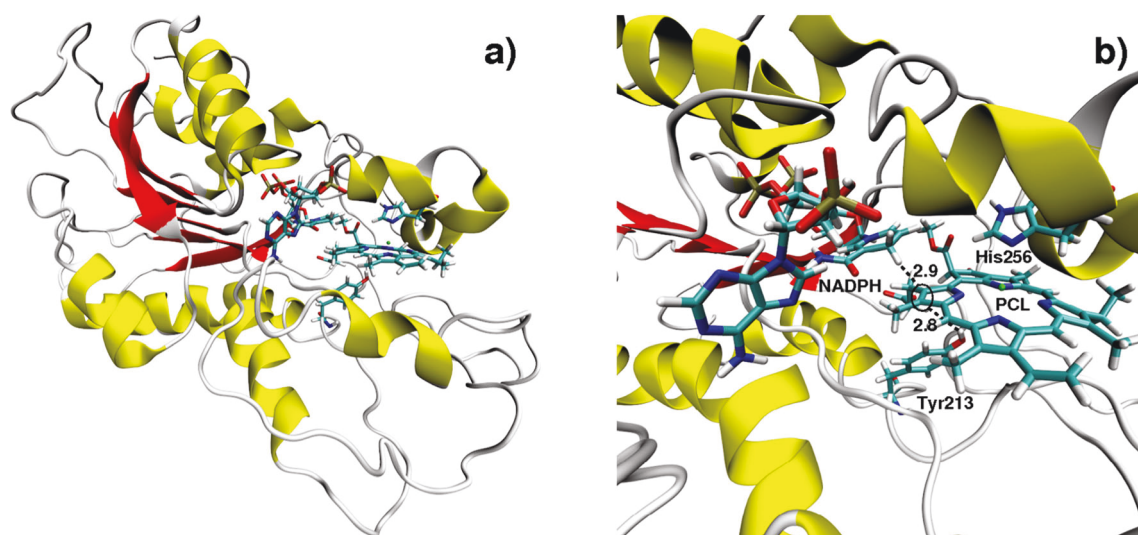


Fig. 16 **a)** Homology model of protochlorophyllide oxidoreductase (LPOR) of *Thermosynechococcus elongatus* in a catalytically competent active site geometry. The structure consists of a central parallel β -sheet (red) surrounded by α -helices (yellow). The structure contains bound NADPH and the docked substrate PCL (shown as licorice

representation). **b)** Catalytically competent active site geometry of PCL in LPOR. Hydride and proton transfer path (PCL(C₁₈)-(H)Tyr213 and PCL(C₁₇)-(H)NADPH, respectively) have been labeled by dashed lines (distances are given in Å)

which indicates a strong hydrogen-bond network between PCL and protein residues. Thus, we used the previously published homology model [102] of LPOR, pre-oriented it in a catalytically competent conformation and subjected it to MD simulations and clustering analysis to obtain representative frames of the binding pocket and the protein environment. These were then refined at the QM/MM level and finally used to compute the IR spectrum of LPOR with cofactor and substrate.

The high layer (QM region) was extended over the pigment (PCL), nicotinamide ring of NADPH, phenol moiety of Tyr213, imidazole ring of His256, axial water to the Mg and waters/residues involved in H-bonds with PCL (Fig. 16a) to a total of nearly 200 atoms. Protein residues in the vicinity of the QM region as well as waters contributing to the hydrogen bond network were allowed to relax (i.e., movable medium layer), while the rest of the system was kept frozen (i.e., low layer), thus building a rigid scaffold. The complex QM/MM partitioning and the generation of the MM topology files were achieved in a semi-automatic fashion through the COBRAMM preparatory routine. It took care of the H/M/L partitioning through automatic recognition of hydrogen-bonded amino acids and waters, handled non-standard residues (i.e., PCL, NADPH), charge redistribution of residues shared between QM and MM layers and link atom definitions. Geometry optimizations and frequency computations were performed through the MM(Amber 12)/QM(Gaussian 09) interface of COBRAMM. The numerical frequency computations were performed on a smaller QM region consisting of PCL, the nicotinamide ring of NADPH, and the phenol moiety of Tyr213 (112 atoms, 1047 basis

functions with 6-31G*) by means of massive parallelization over 100 cores using the COBRAMM native parallel environment in less than 1 day. Hydrogen-bond analysis of LPOR docked PCL and comparison against free (water-solvated) PCL demonstrated that the LPOR model *does not* offer a more structured hydrogen-bond network around the keto group at C₁₃. On the contrary, the network seems to deteriorate as exemplified by the MD simulations where the keto group has no hydrogen bonding partners in considerable intervals of the simulation time.

Accordingly, we chose a snapshot with an average hydrogen-bond network to simulate the IR spectrum, having three residues Ala168, Asn169, and Lys217 binding to the carboxyl, one water binding simultaneously to the ester and to the carboxyl groups, and one water binding to the keto group (Fig. 17a). The IR spectrum (Fig. 17b) does not show the expected 26 cm⁻¹ red-shift of the keto band (as observed experimentally) with respect to water, instead a blue-shift of 15 cm⁻¹ is encountered (1722 cm⁻¹).

The result demonstrates that the modified LPOR enzyme model, while offering a catalytically competent active site, is not able to reproduce the available spectroscopic data. The deficiency of the model is expressed in the lack of a strong hydrogen-bond network around the keto group believed to be responsible for the extraordinary red-shift of its IR frequency. On the basis of these findings we refined the snapshot to introduce a hydrogen bond between the keto group and Lys269, located on the flanking α -helix above the PCL binding site (hereafter referred to as PCL-K269). K269 was selected due its positive charge and proximity to the PCL keto group during the MD simulations. During the refinement of the

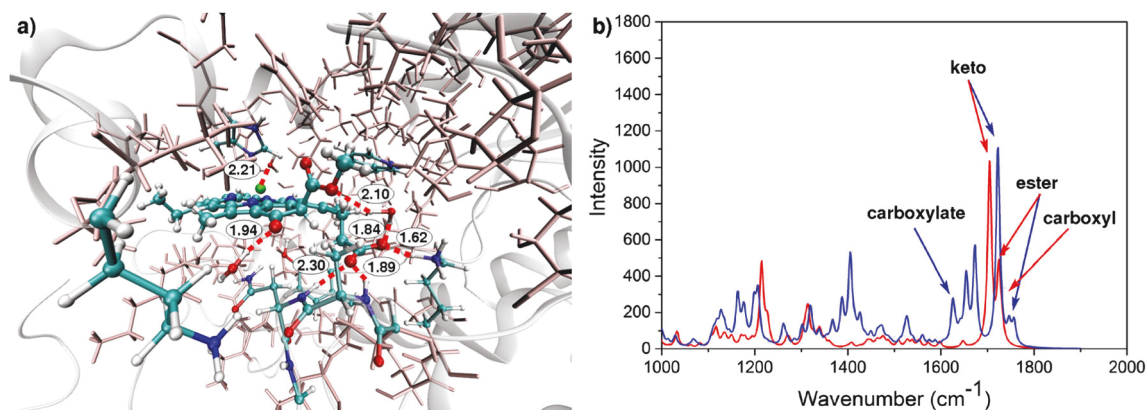


Fig. 17 **a)** Representative snapshot of the catalytic site of the LPOR enzyme with bound NADPH and docked PCL; residues (amino acids and waters) hydrogen-bonded to the carbonyl groups of PCL as well as some relevant distances are highlighted (distances are in Å); residues

models, minimal deformation of the active site was imperative. Figure 18a) documents the changes in the structure (in red) with respect to original snapshot (in blue). Evidently, the largest deformation is exercised on K269 itself and on the water originally involved in the hydrogen bond with the keto group, whereas the nearby residues experience minor deformations. The so-obtained snapshot was further refined using the same QM/MM protocol described above and, subsequently, a frequency computation was performed for the equilibrated structure. The IR spectra are depicted in Fig. 18b).

Remarkably, a moderate red-shift of 14 cm^{-1} with respect to water is encountered for PCL-K269 (1690 cm^{-1}). This observation makes PCL-K269 a possible candidate for a catalytically competent model of the LPOR enzyme. Yet, quantitative agreement was not reached in IR spectra simulations. Thus, it appears likely that inside the enzyme PCL coordinates a water molecule to the keto group in addition to K269. Indeed,

included in the high, medium and low layers given through CPK, licorice (pink) and tube (white) representation. **b)** IR spectra of a snapshot of free (water-solvated) PCL (red) and LPOR-bound PCL (blue) with representative hydrogen bonding networks

geometries with a close contact of the keto group to both Lys269 and a water molecule are sampled during the MD simulations.

Deactivation channels in water-solvated pyrimidine nucleosides

COBRAMM connects to various ground and excited-state features of the Molcas 8 software, including several multiconfigurational approaches such as CASSCF/RASSCF, their second-order perturbation theory extensions (CASPT2/RASPT2) in both single-state and multi-state (MS and XMS) flavors, as well as a wide range of single-reference methods (MP2, CCSD, and CCSD(T) among others), all of those combined with highly efficient resolution of identity schemes framed within the Cholesky decomposition [105, 106] and frozen natural orbitals (FNO) methods [6, 107]. A recent example showcasing the capabilities of the COBRAMM/Molcas

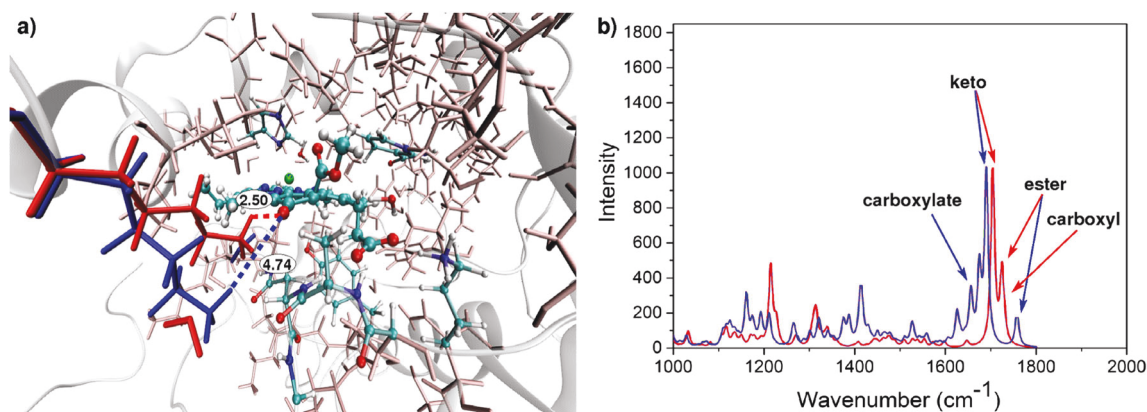


Fig. 18 **a)** Representative snapshot of the catalytic site of the LPOR enzyme with bound NADPH and docked PCL modified to accommodate a hydrogen bond between the keto group at C₁₃ and residue K269. Original position (as in Fig. 16a) of K269 and proximal

water shown in blue, new position shown in red. **b)** IR spectra of a snapshot of free (water-solvated) PCL (red, same as in Fig. 17b) and the refined LPOR-bound PCL (blue)

interface is the studies on deciphering the photoinduced deactivation channels in DNA/RNA water-solvated pyrimidine nucleosides oxy-uridine, deoxy-thymidine, and oxy-cytidine (nucleobases with a ribose sugar group) [108, 109]. Oxy-uridine (referred to as Urd hereafter) will be considered as a representative case, as it displays a range of characterized excited-state minima, transition states (TSs), conical intersections (CoIns), and intersystem crossings (ISCs) depicting the complex photochemical scenario undergone by this species upon excitation. Whilst these excited-state driven properties have been extensively studied theoretically in the gas-phase, less is known about the concrete molecular motions characterizing them in solution. Moreover, most of the simulations carried out in solution are based on polarizable continuum models, where the atomistic details of the specific solvent to solute (short-range) interactions are lost.

Water has been shown to slow down the excited state decay of the DNA/RNA nucleobases and nucleosides with respect to their in vacuo counterparts [110] while also showing a different excited-state population branching along the diverse neighboring states, which strongly modulates its overall photophysics. Given most experimental evidence is recorded in water solution due to being the biologically relevant solvent, QM/MM schemes such as those contained in COBRAMM are appealing to treat these sorts of problems. An example is given in Fig. 19, where the specific QM/MM set-up employed to study Urd embedded in a water droplet is shown. A closer look is given in Fig. 19 right, where the specific high/medium/low layer partitioning employed in the simulations is depicted: the high layer treated at the QM level (CASPT2) consists on the uracil nucleobase, whereas the ribose (sugar) group and all water molecules within 5 Å of the nucleobase center of mass are depicted in grey and belong to the mobile medium layer, the rest of the water molecules being frozen in the low layer.

In this scheme, the first solvation layer of water is included in the optimization procedures, allowing it to relax along the excited state deactivation of the nucleobase and thus accounting for the short-range interactions present due to the hydrogen bonding motifs formed between the surrounding water molecules and uracil.

Figure 20 displays an overall scheme of the many deactivation channels available to water-solvated Urd. White arrows therein denote deactivation routes attributed to occur in an ultrafast (sub- to few-ps) timescale due to initial excitation into $^1\pi\pi^*$ bright states and subsequent decay to the ground state either through a ring-puckering "ethene-like" conical intersection (right-hand side of Fig. 20), associated with a barrier of ~ 0.1 eV (see also ref. [111, 112] for $^1\pi\pi^*$ and $^1n\pi^*$ deactivation channels and ref. [113] for time-resolved IR spectra). Minima, transition states (TSs), and interstate crossing points (CoIns) are characterized at the MS-CASPT2 level employing COBRAMM's numerical routines for computing gradients and non-adiabatic couplings, described in the [Methods](#) section and employing the Gaussian optimizer as described in the [QM/MM implementation](#) section. Frequencies at the MS-CASPT2 level were performed at critical points (minima and TS) to confirm their nature. As mentioned in the [Methods](#) section, absorption cross sections can be computed within the COBRAMM interface by means of Wigner sampling [78]. This option is used here to simulate the absorption spectrum of Urd. Five hundred snapshots were generated at room temperature, thereby excluding high-frequency modes belonging to C-H and N-H stretchings from the sampling. The electronic structure of each snapshot was computed at the MS-CASPT2/SA-CASSCF level using a full valence- π active space comprising ten electrons in eight orbitals (i.e., CAS(10,8)), averaging over four states and using the ANO-L basis set adopting contractions 3s2p1d on non-hydrogen and 2s1p on hydrogen atoms.

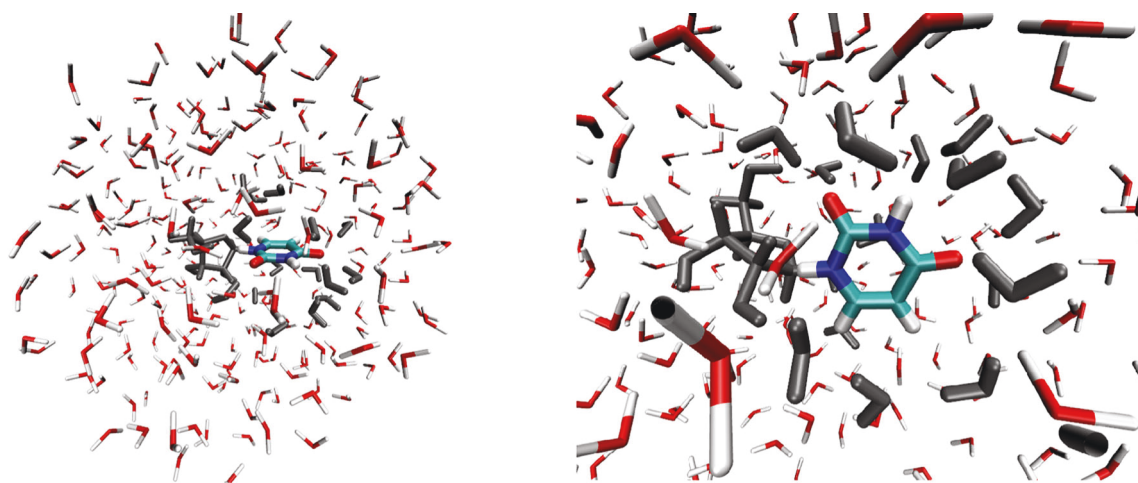


Fig. 19 Right: Urd with a water droplet cut within a 12-Å radius from the Uracil moiety. Left: High/medium/low layer partitioning used for MD simulations; sugar and waters in 5-Å distance to center of mass of uracil (high layer) added to the medium layer are shown in grey

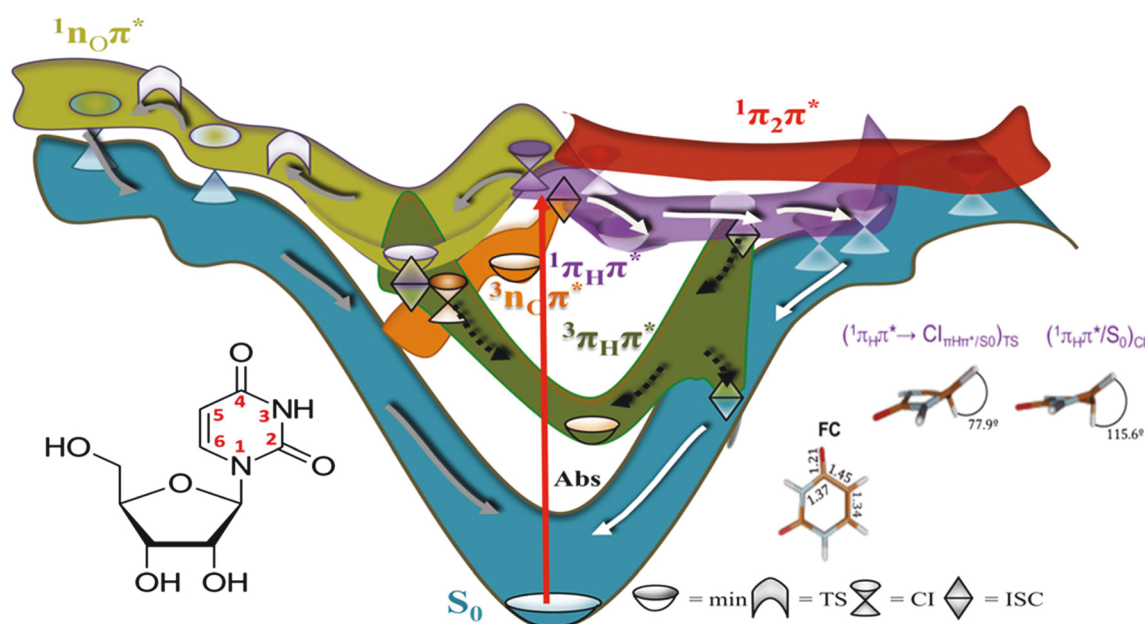


Fig. 20 Schematic representation of the photo-excited phenomena triggered in water-solvated uridine upon UV-light exposure. *White arrows* represent deactivation channels that are ultrafast (sub- to few-ps); *grey* denotes long-lived channels based on optically dark $^1n\pi^*$ states (several to dozens of ps) and *black* shows the longer-lived triplet population

(dozens of ps to ns). On the right-hand side, concrete critical points characterized along the potential energy surfaces of initially accessed and ultrafast $^1\pi\pi^*$ state deactivations are shown. Figure adapted from Ref. [108] with permission from the Royal Society of Chemistry

The result is shown in Fig. 21 left. The experimental spectrum of Urd above 220 nm is characterized by a single intense band peaking at 263 nm (magenta line in Fig. 21). The full valence- π CAS(10,8) reproduces the bandwidth, yet red-shifts the band significantly by ca. 20 nm (~ 0.35 eV at this wavelength). In a series of studies on various chromophoric systems, we have demonstrated recently that this effect is general and due to the overestimation of the correlation energy with perturbative methods like CASPT2 with full valence- π active spaces [26, 27, 114, 115]. Increasing the active space by including extra-valence virtual orbitals rectifies the problem as seen in Fig. 21 right.

In this particular case, by extending the full valence- π active space by an additional eight extra-valence virtual π -orbitals treated at the RASSCF level, where up to two simultaneous excitations are allowed, we achieve a near quantitative agreement with the experimental data. The treatment of such large active spaces is made possible through the restricted active space feature available in Molcas coupled to the Cholesky utility, which reduces the number of RASPT2 terms and speeds up their calculation [7].

The interface of COBRAMM with Molcas allows running trajectory-based non-adiabatic mixed quantum-classical dynamics simulations applying Tully's fewest switches surface

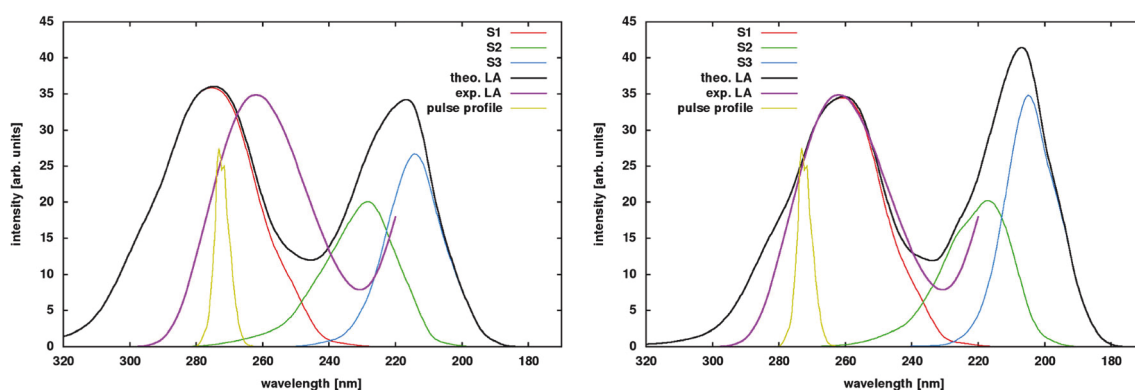


Fig. 21 Comparison of the simulated (*black*) and experimental (*magenta*) linear absorption spectra of uridine in aqueous solution. Simulations performed at the MS-4-CASPT2/SA-4-CASSCF(10,8) (*left*) and MS-4-RASPT2/SA-4-RASSCF(10,8|2,8) (*right*) levels of theory. Also shown

are the individual contributions to the total spectrum from the adiabatic states S_1 - S_3 and a typical pump pulse envelope used in time-resolved studies of the systems

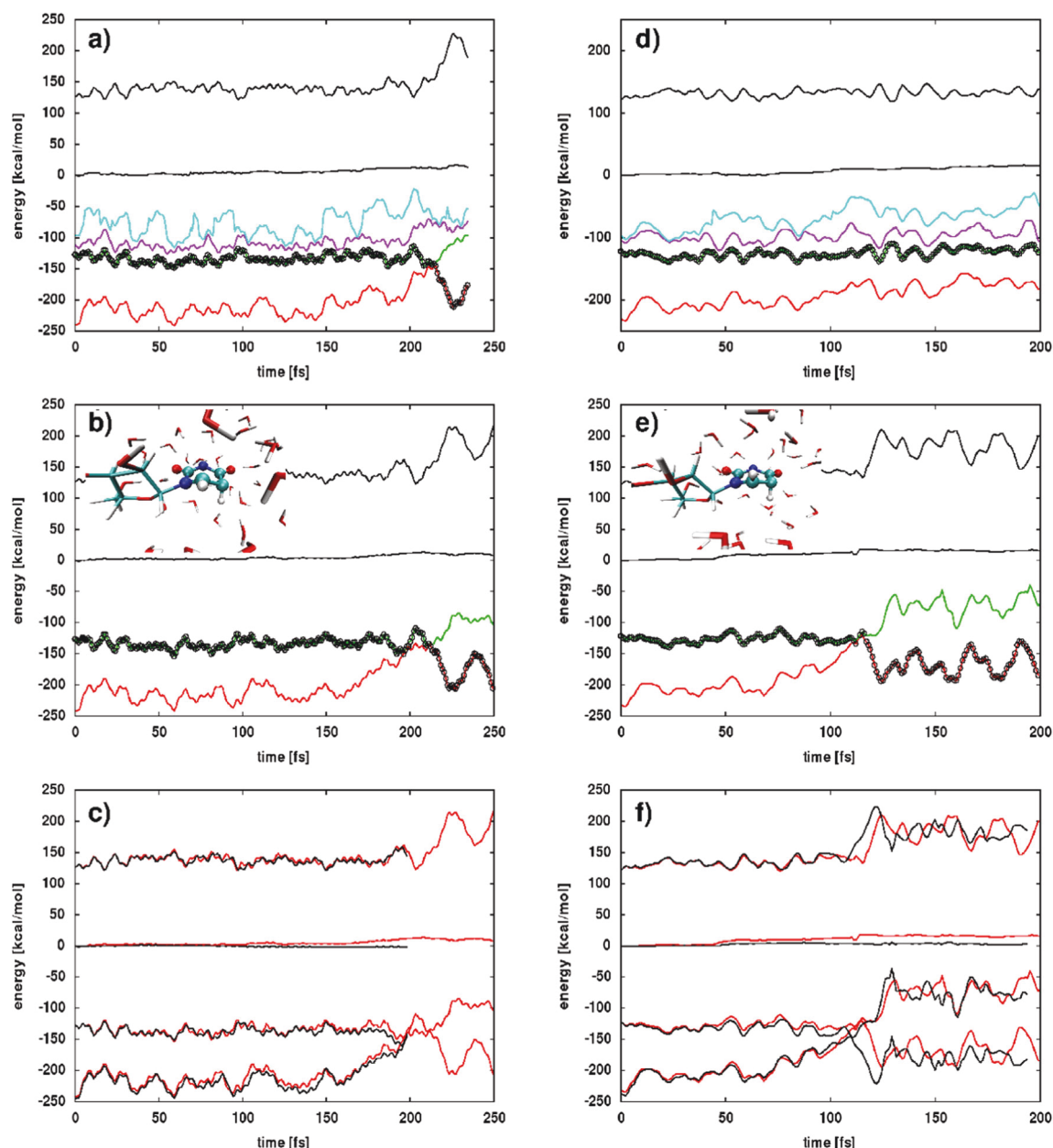


Fig. 22 Comparison of simulations run at the MS-4-CASPT2/SA-4-CASSCF(10,8) (**a**, **d**), the SS-CASPT2/SA-2-CASSCF(10,8) (**b**, **e**) and the SS-CASPT2/SA-2-CASSCF(10,10) (**c**, **f**) levels of theory for two snapshots. Color code for **a**, **b**, **d**, and **e**: S_0 (red), S_1 (green), S_2

(magenta), S_3 (cyan), photoactive state (black dotted), kinetic and total energy (black); Color code for **c** and **f**: SS-CASPT2/SA-2-CASSCF(10,8) level (red), SS-CASPT2/SA-2-CASSCF(10,10) (black). The insets show the uridine conformation at the crossing points

hopping algorithm [116]. In the following, we use Urd to showcase these features. As an example, Fig. 22 compares simulations at the MS-4-CASPT2/SA-4-CASSCF(10,8) (Fig. 22a,d), the SS-CASPT2/SA-2-CASSCF(10,8) (Fig. 22b,e) and the SS-CASPT2/SA-2-CASSCF(10,10) (Fig. 22c,f) levels of theory propagated for a total of 200 fs with a time-step of 1.0 fs. To afford this relatively long time-step, we kept the high-frequency modes frozen during the sampling procedure. Furthermore, we used the Tully–Hammer–Schiffer (THS) variation [117] of the fewest switches surface hopping scheme, which is more suitable to

avoid overstepping confined regions in coordinate space with non-adiabatic coupling peaks. The movable layer of the system was increased to include several water layers (on average 45 waters per snapshot) to allow the water solvation shell to adapt to the puckering deformation of the uracil moiety. Gradients were computed numerically through a two-point finite differentiation formula using the COBRAMM's parallel environment. To speed up the numerical computations of the SS-CASPT2 gradients, we made use of the fact that the perturbative correction is applied independently on each root. Thus, while at the reference point all states are corrected at the

SS-CASPT2 level, during the computation of the gradient only the state that drives the nuclear dynamics (i.e., the photoactive state) was taken into account. Care was taken for possible state swapping at each displaced geometry in case of near-degeneracy at the CASSCF level by following the nature of the wave function. To speed up the numerical computations in the case of the MS-CASPT2 gradients, instead, the MS-CASPT2 eigenvector of the photoactive state at the reference geometry was analyzed and only CASSCF roots that contribute considerably to the MS-CASPT2 wave function were taken into account. Internal routines monitor the stability of the simulations. If

- individual contributions to the numerical gradient exceed a threshold (default 0.5)
- the ratio of the lengths of the gradient vectors at two consecutive steps (i.e., $\|g(t)\|/\|g(t-dt)\|$) and, at the same time, the deviation of the total energy from its value at the previous step (i.e., $E_{\text{tot}}(t)-E_{\text{tot}}(t-dt)$) exceed certain thresholds (defaults 1.3 and 0.5 kcal/mol, respectively)
- a single point computation fails to converge the gradient is discarded and the Velocity Verlet step is repeated at the last stable geometry, first with a half step (in this case 0.5 fs) and, if the thresholds are exceeded again, with a double step (in this case 2.0 fs). Furthermore, the wave function of the photoactive state is followed by computing the overlap between two consecutive time steps. A warning is issued in case the overlap is below 0.9, thus indicating possible intruder states (e.g., in this particular case the intrusion of $n\pi^*$ states due to lone pair rotation in the active space).

A set of trajectories was run with both CASPT2 flavors (SS and MS). The decay to the ground state occurs exclusively through the ring-puckering "ethene-like" CoIn (compare insets in Fig. 22 to structure number 5 in Fig. 20b). Thereby, we observe splitting of the wavepacket towards two ring-puckering CoIns with contrariwise orientation of hydrogens H9 and H10. In the limited set of trajectories run for this study, we find no evidence for decay through bond breaking. The comparison between the two CASPT2 flavors shows a considerably slower dynamics in the case of MS-CASPT2, whereas virtually all trajectories have already decayed to the GS after 200 fs in the case of SS-CASPT2. While in some cases SS and MS simulations agree well (Fig. 22a,b), in many cases we observe that the system quickly leaves the FC region in order to roam then for hundreds of fs in the CoIn region at the MS-CASPT2 level (Fig. 22d) opposite to trajectories run at the SS-CASPT2 level, which decayed in a ballistic fashion (Fig. 22e). The reason for this discrepancy is an artifact of the MS-CASPT2 approach with small active spaces that gives rise to relatively large off-diagonal S_0 - S_1 elements in the MS-CASPT2 Hamiltonian (an

order of magnitude larger than the value of 0.002 suggested in the seminal paper on MS-CASPT2 [118]). While not posing a problem in the FC region where the S_0 - S_1 energy gap is large these become important in the case when the two states come energetically close. In a direction in which the CASSCF wave functions mix, the MS-CASPT2 routine over-estimates the level of un-mixing, thus pushing the two states apart, introducing an artificial barrier in the excited state and preventing an efficient decay. The only way for the excited state to decay to the ground state is in the form of adiabatic crossing where the two states do not mix.

The MS- and SS-CASPT2 dynamics with the full valence- π AS are overall stable, showing a tolerable increase of the total energy over the course of the simulation, which reaches ca. 10 kcal/mol. There are several factors that play a role in this issue. Apart from the 1.0-fs time step, which is insufficient to describe fast stretching dynamics or the large amplitude dynamics of the "hot" ground state, Fig. 22c,f exemplifies the issues arisen due to active space size. Increasing the full valence- π active space by two extra-valence orbitals reduces substantially the deviation of the total energy. The improved accuracy comes at the price of an increased cost in the computation, only affordable with restricted active spaces (i.e., RASPT2). We note, however, that the two sets of trajectories, SS-CASPT2/SA-2-CASSCF(10,8) (red) and SS-CASPT2/SA-2-CASSCF(10,10) (black), behave qualitatively similarly, i.e., the active space increase does not affect the overall photophysics displayed by the system.

Conclusions

We have reported on new developments and additions to our modular QM and QM/MM interface program COBRAMM, enabling efficient electronic structure computations in ground and excited states. The new developments include algorithms to perform surface hopping dynamics with an arbitrary number of states and decoherence correction using CASSCF, CASPT2 and OMx-MRCI methods. A selector scheme allows the computation of only those gradients and derivative couplings strictly necessary for the current step of the MD. Efficient numerical procedures allow the parallelization of numerical frequency calculations for all implemented methods and speed up gradient and coupling calculations for single state and multistate CASPT2 methods. The modified atom-link implementation restricts the movement of the link atom to the bond axis of the QM/MM boundary. A Rattle version of the velocity Verlet algorithm can freeze out high-frequency motions like C-H or O-H, thus reducing total energy fluctuations and enabling the use of larger time steps. Several tools were added to simplify the QM/MM setup. A script-driven routine helps to select QM and MM portions as well as to define mobile and static regions. Within these tools, computations defining the initial conditions for MD simulations

can be performed. Further routines allow simplified generation of static and time-resolved absorption spectra.

The capabilities of the implementations have been demonstrated with a number of examples. CASSCF/Amber studies of the rhodopsin photoisomerization could reproduce time-resolved spectral traces of a pump-probe experiment. A benchmark calculation applying OMx-MRCI methods to the same reaction revealed good overall agreement with the CASSCF QM/MM trajectory computations. In combined DFT and DFT/MRCI studies, we find indications for a fast singlet and triplet crossing in benzophenone. Photodynamics of the human melanopsin protein and squid rhodopsin were simulated by combining features from different QM programs: wavefunction computations with Molcas and non-adiabatic couplings from Molpro, demonstrating the interoperability of these routines. Both sqRh and hMeOp show faster decay in comparison to bovine rhodopsin, but lead to a lower quantum yield. The drastic reduction in hMeOp quantum yield stems from a phase mismatch of the product generating H-C=C-H motion at the isomerizing double bond. DFT-based QM/MM computations assisted in refining the structure of the LPOR enzyme. The results indicate an advanced hydrogen bonded network that involves a keto-group of the PCL chromophore, a water molecule and a neighboring lysine residue. The numerical routines available within the MOLCAS interface have been extensively used to revisit the photophysics and photochemistry of water-solvated uridine, enabling the unprecedented mapping of potential energy hypersurfaces at the MS-CASPT2 level within a QM/MM scheme that considers environmental effects. These have also allowed the simulation of non-adiabatic molecular dynamics within different schemes. The latter feature a series of on-the-fly checks that ensure the quality of the simulation and simplify the tasks required by the user, facilitating their use for non-specialists and highlighting COBRAMM's philosophy of extending the applicability of the different features available in a user-friendly manner.

Acknowledgements The COBRAMM project was started by Piero Altoé and Marco Stenta under guidance of Marco Garavelli and Andrea Bottoni at the University of Bologna. The current development of COBRAMM is led by Marco Garavelli and his group in Bologna, in collaboration with Oliver Weingart, Heinrich-Heine University of Düsseldorf. Major contributions to the code were provided by Piero Altoé, Marco Stenta, Oliver Weingart, Artur Nenov, Irina Dokukina, Ivan Rivalta, Javier Segarra-Martí, Emiliano Poli, and Salvatore F. Altavilla. We thank Baptiste Demoulin, Irene Conti, Ana Julieta Pepino and Mohsen El-Tahawy for extensive testing and help with debugging. We furthermore acknowledge the kind help of Francesco Aquilante when implementing certain options of the MOLCAS interface.

I.R. gratefully acknowledges the use of HPC resources of the "Pôle Scientifique de Modélisation Numérique" (PSMN) at the ENS-Lyon (France).

Author contributions The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding sources I.D. is grateful for financial support through "Strategischer Forschungsfonds" of the Heinrich-Heine-University Düsseldorf, project no. F 2013–442-9. J.S.M. acknowledges support from the European Commission through the Marie Curie actions (FP8-MSCA-IF, grant n° 747,662). I.R., M.G. and J.S.M. thank the Agence National de la Recherche project FEMTO-2DNA (ANR-15-CE-29-0010).

Abbreviations 2DES, Two-dimensional electronic spectroscopy; 2DUV, Two-dimensional UV spectroscopy; B3LYP, Becke's 3 parameter exchange in combination with Lee, Young and Parr correlation functional; BChl, Bacteriochlorophyll; BFGS, Broydon–Fletcher–Goldfarb–Shannon; BH-LYP, Becke's half-and-half exchange in combination with LYP correlation functional; BP, Benzophenone; CAM, Coulomb attenuated method; CAS, Complete active space; CC, Coupled cluster; CCSD, Coupled cluster singles and doubles; CHL, Chlorophyllide; CI, Configuration interaction; CoIn, Conical intersection; CP-MCSCF, Coupled perturbed multiconfigurational SCF; DFT, Density functional theory; EOM, Equation-of-motion; FC, Franck–Condon; FNO, Frozen natural orbitals; Fs, Femtosecond; HF, Hartree–Fock; hMeOp, Human melanopsin; HOMO, Highest occupied molecular orbital; HPC, High-performance computing; IR, Infrared; IRC, Intrinsic reaction coordinate; ISC, Intersystem crossing; LMP2, Local MP2; LPOR, Protochlorophyllide oxidoreductase; LUMO, Lowest unoccupied molecular orbital; MD, Molecular dynamics; MEP, Minimal energy path; MM, Molecular mechanics; MNDO, Modified neglect of differential overlap (a semi-empirical method and a program package distributed by Thiel et al.); MP, Moller–Plesset (e.g., MP2, MP4); MRCI, Multi-reference CI; MS, Multi state; NAC, Non-adiabatic coupling; NADPH, Reduced nicotinamide adenine dinucleotide phosphate; NVE, Microcanonical ensemble (constant number of particles, volume and energy); OMx, Orthogonalization method X; PCL, Protochlorophyllide; PCM, Polarizable continuum model; Ps, Picosecond; QM, Quantum mechanics; QY, Quantum yield; RAS, Restricted active space; Rh, Rhodopsin; SA, State-average; SCF, Self-consistent field; SOS, Sum-over-states; SqRh, Squid rhodopsin; SS, Single state; ssPOR, LPOR from *Synechocystis* sp.; SVP, Single zeta valence basis set including polarization functions; TCL, Tool command language; TD, Time-dependent; TDC, Time-derivative coupling; TDSE, Time-dependent Schrödinger equation; THS, Tully–Hammes–Schiffer (surface hopping scheme); TIP3P, Three-site transferable intermolecular potential (MM water model); TS, Transition state; Urd, Oxy-uridine; ZPE, Zero-point energy

References

1. Frisch MJ, Trucks GW, Schlegel HB, et al Gaussian 09 Revision E.01
2. TURBOMOLE V7.0 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>
3. Grimme S, Waletzke M (1999) A combination of Kohn–Sham density functional theory and multi-reference configuration interaction methods. *J Chem Phys* 111:5645–5655. <https://doi.org/10.1063/1.479866>
4. Lyskov I, Kleinschmidt M, Marian CM (2016) Redesign of the DFT/MRCI Hamiltonian. *J Chem Phys* 144:34104. <https://doi.org/10.1063/1.4940036>
5. Aquilante F, Autschbach J, Carlson RK et al (2016) Molcas 8: new capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J Comput Chem* 37:506–541. <https://doi.org/10.1002/jcc.24221>
6. Segarra-Martí J, Garavelli M, Aquilante F (2015) Multiconfigurational second-order perturbation theory with

- frozen natural orbitals extended to the treatment of photochemical problems. *J Chem Theory Comput* 11:3772–3784. <https://doi.org/10.1021/acs.jctc.5b00479>
7. Aquilante F, Malmqvist P-Å, Pedersen TB et al (2008) Cholesky decomposition-based multiconfiguration second-order perturbation theory (CD-CASPT2): application to the spin-state energetics of CoIII(diiminato)(NPh). *J Chem Theory Comput* 4:694–702. <https://doi.org/10.1021/ct700263h>
 8. Werner H-J, Knowles PJ, Knizia G, et al (2010) MOLPRO, version 2010.1, a package of ab initio programs
 9. Watts JD (2008) An introduction to equation-of-motion and linear-response coupled-cluster methods for electronically excited states of molecules. In: Shukla MK, Leszczynski J (eds) *Radiation induced molecular phenomena in nucleic acids*. Springer Netherlands, Dordrecht, pp 65–92
 10. Kong L, Bischoff FA, Valeev EF (2012) Explicitly correlated R12/F12 methods for electronic structure. *Chem Rev* 112:75–107. <https://doi.org/10.1021/cr200204r>
 11. Werner H, Köppl C, Ma Q, Schwilk M (2017) Explicitly correlated local electron correlation methods. *Fragmentation*. John Wiley & Sons, Ltd, Chichester, UK, pp 1–79
 12. Dewar MJS, Thiel W (1977) Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J Am Chem Soc* 99:4899–4907. <https://doi.org/10.1021/ja00457a004>
 13. Kolb M, Thiel W (1993) Beyond the MNDO model: methodical considerations and numerical results. *J Comput Chem* 14:775–789. <https://doi.org/10.1002/jcc.540140704>
 14. Tuttle T, Thiel W (2008) OMx-D: semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application. *Phys Chem Chem Phys* 10:2159–2166. <https://doi.org/10.1039/B718795E>
 15. Tuna D, Lu Y, Kosłowski A, Thiel W (2016) Semiempirical quantum-chemical Orthogonalization-corrected methods: benchmarks of electronically excited states. *J Chem Theory Comput* 12: 4400–4422. <https://doi.org/10.1021/acs.jctc.6b00403>
 16. Altoè P, Stenta M, Bottoni A, Garavelli M (2007) A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction. *Theor Chem Accounts* 118:219–240. <https://doi.org/10.1007/s00214-007-0275-9>
 17. Case DA, Babin V, Berryman JT et al (2014) *Amber 14*. University of California, San Francisco
 18. Hratchian HP, Schlegel HB (2005) Finding minima, transition states, and following reaction pathways on ab initio potential energy surfaces. In: Frenking G, Kim KS, Scuseria G (eds) *Theory and applications of computational chemistry: the first 40 years*. Elsevier, Amsterdam, pp 195–249
 19. Verlet L (1967) Computer experiments on classical fluids. *Phys Rev* 159:98–103
 20. Andersen HC (1983) Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations. *J Comput Phys* 52:24–34. [https://doi.org/10.1016/0021-9991\(83\)90014-1](https://doi.org/10.1016/0021-9991(83)90014-1)
 21. Bearpark MJ, Robb MA, Bernhard Schlegel H (1994) A direct method for the location of the lowest energy point on a potential surface crossing. *Chem Phys Lett* 223:269–274. [https://doi.org/10.1016/0009-2614\(94\)00433-1](https://doi.org/10.1016/0009-2614(94)00433-1)
 22. Dokukina I, Weingart O (2015) Spectral properties and isomerisation path of retinal in C1C2 channelrhodopsin. *Phys Chem Chem Phys* 17(38):25142–25150. <https://doi.org/10.1039/c5cp02650d>
 23. Mukamel S (1995) *Principles of nonlinear optics and spectroscopy*. Oxford University Press, Oxford, UK
 24. Nenov A, Rivalta I, Cerullo G et al (2014) Disentangling peptide configurations via two-dimensional electronic spectroscopy: ab initio simulations beyond the Frenkel exciton Hamiltonian. *J Phys Chem Lett* 5:767–771. <https://doi.org/10.1021/jz5002314>
 25. Nenov A, Beccara S, Rivalta I et al (2014) Tracking conformational dynamics of polypeptides by nonlinear electronic spectroscopy of aromatic residues: a first-principles simulation study. *ChemPhysChem* 15:3282–3290. <https://doi.org/10.1002/cphc.201402374>
 26. Nenov A, Rivalta I, Mukamel S, Garavelli M (2014) Bidimensional electronic spectroscopy on indole in gas phase and in water from first principles. *Comput Theor Chem* 1040–1041:295–303. <https://doi.org/10.1016/j.comptc.2014.03.031>
 27. Nenov A, Giussani A, Segarra-Martí J et al (2015) Modeling the high-energy electronic state manifold of adenine: calibration for nonlinear electronic spectroscopy. *J Chem Phys* 142:212443. <https://doi.org/10.1063/1.4921016>
 28. Ivan R, Artur N, Giulio C et al (2013) Ab initio simulations of two-dimensional electronic spectra: the SOS//QM/MM approach. *Int J Quantum Chem* 114:85–93. <https://doi.org/10.1002/qua.24511>
 29. Segarra-Martí J, Nenov A, Mukamel S, et al (2018) Towards accurate simulations of two-dimensional electronic spectroscopy. *Top Curr Chem* submitted
 30. Weingart O (2017) Combined quantum and molecular mechanics (QM/MM) approaches to simulate ultrafast photodynamics in biological systems. *Curr Org Chem* 21:586–601. <https://doi.org/10.2174/1385272821666161108150421>
 31. Tully JC, Preston RK (1971) Trajectory surface hopping approach to nonadiabatic molecular collisions: the reaction of H+ with D2. *J Chem Phys* 55(2):562–572
 32. Tully JC (1990) Molecular dynamics with electronic transitions. *J Chem Phys* 93:1061–1071
 33. Hayashi S, Tajkhorshid E, Schulten K (2009) Photochemical reaction dynamics of the primary event of vision studied by means of a hybrid molecular simulation. *Biophys J* 96:403–416. <https://doi.org/10.1016/j.bpj.2008.09.049>
 34. Weingart O, Garavelli M (2012) Modelling vibrational coherence in the primary rhodopsin photoproduct. *J Chem Phys* 137:523. <https://doi.org/10.1063/1.4742814>
 35. Polli D, Weingart O, Brida D et al (2014) Wavepacket splitting and two-pathway deactivation in the photoexcited visual pigment isorhodopsin. *Angew Chemie - Int Ed* 53(9):2504–2507. <https://doi.org/10.1002/anie.201309867>
 36. Quenneville J, Martí TJ (2000) Ab initio multiple spawning: photochemistry from first principles quantum molecular. *J Phys Chem A* 104:5161–5175. <https://doi.org/10.1021/jp994174i>
 37. Zhu C, Jasper AW, Truhlar DG (2005) Non-Born–Oppenheimer Liouville-von Neumann dynamics. Evolution of a subsystem controlled by linear and population-driven decay of mixing with decoherent and coherent switching. *J Chem Theory Comput* 1: 527–540. <https://doi.org/10.1021/ct050021p>
 38. Granucci G, Persico M (2007) Critical appraisal of the fewest switches algorithm for surface hopping. *J Chem Phys* 126: 134114. <https://doi.org/10.1063/1.2715585>
 39. Doltsinis NL (2002) Nonadiabatic dynamics: mean-field and surface hopping. In: Grotendorst J, Marx D, Muramatsu A (eds) *Quantum simulations of complex many-body systems: from theory to algorithms, lecture notes*. NIC series. John von Neumann Institute for Computing, Jülich, Germany, pp 377–397
 40. Fabiano E, Keal TW, Thiel W (2008) Implementation of surface hopping molecular dynamics using semiempirical methods. *Chem Phys* 349:334–347. <https://doi.org/10.1016/j.chemphys.2008.01.044>
 41. Weingart O (2008) The role of HOOP-modes in the ultrafast photo-isomerization of retinal models. *Chem Phys* 349(1):348–355
 42. Schapiro I, Ryazantsev MN, Frutos LM et al (2011) The ultrafast photoisomerizations of rhodopsin and bathorhodopsin are modulated by bond length alternation and HOOP driven electronic

- effects. *J Am Chem Soc* 133:3354–3364. <https://doi.org/10.1021/ja1056196>
43. Weingart O, Altoè P, Stenta M et al (2011) Product formation in rhodopsin by fast hydrogen motions. *Phys Chem Chem Phys* 13: 3645–3648. <https://doi.org/10.1039/c0cp02496a>
 44. Fdez Galván I, Delcey MG, Pedersen TB et al (2016) Analytical state-average complete-active-space self-consistent field nonadiabatic coupling vectors: implementation with density-fitted two-electron integrals and application to conical intersections. *J Chem Theory Comput* 12:3636–3653. <https://doi.org/10.1021/acs.jctc.6b00384>
 45. Park JW, Shiozaki T (2017) Analytical derivative coupling for multistate CASPT2 theory. *J Chem Theory Comput* 13:2561–2570. <https://doi.org/10.1021/acs.jctc.7b00018>
 46. Liu L, Liu J, Martinez TJ (2016) Dynamical correlation effects on photoisomerization: ab initio multiple spawning dynamics with MS-CASPT2 for a model trans-protonated Schiff base. *J Phys Chem B* 120:1940–1949. <https://doi.org/10.1021/acs.jpcc.5b09838>
 47. Pittner J, Lischka H, Barbatti M (2009) Optimization of mixed quantum-classical dynamics: time-derivative coupling terms and selected couplings. *Chem Phys* 356:147–152. <https://doi.org/10.1016/j.chemphys.2008.10.013>
 48. Ou Q, Bellchambers GD, Furche F, Subotnik JE (2015) First-order derivative couplings between excited states from adiabatic TDDFT response theory. *J Chem Phys* 142:064114. <https://doi.org/10.1063/1.4906941>
 49. Zhang X, Herbert JM (2014) Analytic derivative couplings for spin-flip configuration interaction singles and spin-flip time-dependent density functional theory. *J Chem Phys* 141:64104. <https://doi.org/10.1063/1.4891984>
 50. Faraji S, Matsika S, Krylov AI (2018) Calculations of non-adiabatic couplings within equation-of-motion coupled-cluster framework: theory, implementation, and validation against multi-reference methods. *J Chem Phys* 148:44103. <https://doi.org/10.1063/1.5009433>
 51. Tajti A, Szalay PG (2009) Analytic evaluation of the nonadiabatic coupling vector between excited states using equation-of-motion coupled-cluster theory. *J Chem Phys* 131:124104. <https://doi.org/10.1063/1.3232011>
 52. Toniolo A, Olsen S, Manohar L, Martinez TJ (2004) Conical intersection dynamics in solution: the chromophore of green fluorescent protein. *Faraday Discuss* 127:149–163. <https://doi.org/10.1039/b401167h>
 53. Senn HM, Thiel W (2007) QM/MM methods for biological systems. *Top Curr Chem* 268:173–290. <https://doi.org/10.1007/128>
 54. Bo C, Maseras F (2008) QM/MM methods in inorganic chemistry. *Dalton Trans* 22:2911–2919
 55. van der Kamp MW, Mulholland AJ (2013) Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* 52:2708–2728. <https://doi.org/10.1021/bi400215w>
 56. Lin H, Truhlar DG (2006) QM/MM: what have we learned, where are we, and where do we go from here? *Theor Chem Accounts* 117:185–199. <https://doi.org/10.1007/s00214-006-0143-z>
 57. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* 103:227–249. [https://doi.org/10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9)
 58. Warshel A (1976) Bicycle-pedal model for the first step in the vision process. *Nature* 260:679–683
 59. Stewart JJP (2016) MOPAC2016
 60. Abraham MJ, Murtola T, Schulz R et al (2015) GROMACS: high-performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
 61. Dapprich S, Komáromi I, Byun KS et al (1999) A new ONIOM implementation in Gaussian 98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J Mol Struct* 461–462:1–21. [https://doi.org/10.1016/S0166-1280\(98\)00475-8](https://doi.org/10.1016/S0166-1280(98)00475-8)
 62. Ponder Jay W, Richards Frederic M (2004) An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J Comput Chem* 8:1016–1024. <https://doi.org/10.1002/jcc.540080710>
 63. Metz S, Kästner J, Sokol AA et al (2014) ChemShell-a modular software package for QM/MM simulations. *Wiley Interdiscip Rev Comput Mol Sci* 4:101–110. <https://doi.org/10.1002/wcms.1163>
 64. Neese F (2012) The ORCA program system. *Wiley Interdiscip Rev Comput Mol Sci* 2:73–78. <https://doi.org/10.1002/wcms.81>
 65. Brooks BR, Brooks CL, Mackerell AD et al (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545–1614. <https://doi.org/10.1002/jcc.21287>
 66. Reif MM, Winger M, Oostenbrink C (2013) Testing of the GROMOS force-field parameter set 54A8: structural properties of electrolyte solutions, lipid bilayers, and proteins. *J Chem Theory Comput* 9:1247–1264. <https://doi.org/10.1021/ct300874c>
 67. Gale JD, Rohl AL (2003) The general utility lattice program (GULP). *Mol Simul* 29:291–341. <https://doi.org/10.1080/0892702031000104887>
 68. Smith W, Yong CW, Rodger PM (2002) DL_POLY: application to molecular simulation. *Mol Simul* 28:385–471. <https://doi.org/10.1080/08927020290018769>
 69. Barbatti M, Ruckebauer M, Plasser F et al (2014) Newton-X: a surface-hopping program for nonadiabatic molecular dynamics. *Wiley Interdiscip Rev Comput Mol Sci* 4:26–33. <https://doi.org/10.1002/wcms.1158>
 70. Altoè P, Stenta M, Bottoni A, Garavelli M (2007) A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction. *Theor Chem Accounts* 118(1): 219–240
 71. Fletcher R (1987) Practical methods of optimization, 2nd edn. Wiley, New York
 72. Friesner RA, Guallar V (2005) Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Ann Rev Phys Chem* 56:389–427. <https://doi.org/10.1146/annurev.physchem.55.091602.094410>
 73. Brunk E, Rothlisberger U (2015) Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem Rev* 115:6217–6263. <https://doi.org/10.1021/cr500628b>
 74. Eurenium KP, Chatfield DC, Brooks BR, Hodoscek M (1996) Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. I. Theoretical considerations. *Int J Quantum Chem* 60:1189–1200. [https://doi.org/10.1002/\(SICI\)1097-461X\(1996\)60:6<1189::AID-QUA7>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-461X(1996)60:6<1189::AID-QUA7>3.0.CO;2-W)
 75. König PH, Hoffmann M, and Th. Frauenheim, Cui Q (2005) A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *J Phys Chem B* 109:9082–9095. doi: <https://doi.org/10.1021/jp0442347>
 76. Sloane CS, Hase WL (1977) On the dynamics of state selected uni-molecular reactions: chloroacetylene dissociation and predissociation. *J Chem Phys* 66(4):1523–1533
 77. Lan Z, Fabiano E, Thiel W (2009) Photoinduced nonadiabatic dynamics of pyrimidine nucleobases: on-the-fly surface-hopping study with semiempirical methods. *J Phys Chem B* 113:3548–3555. <https://doi.org/10.1021/jp809085h>
 78. Du L, Lan Z (2015) An on-the-fly surface-hopping program JADE for nonadiabatic molecular dynamics of polyatomic systems: implementation and applications. *J Chem Theory Comput* 11:1360–1374. <https://doi.org/10.1021/ct501106d>

79. Ruckebauer M, Barbatti M, Müller T, Lischka H (2013) Nonadiabatic photodynamics of a retinal model in polar and non-polar environment. *J Phys Chem A* 117:2790–2799. <https://doi.org/10.1021/jp400401f>
80. Polli D, Altoè P, Weingart O et al (2010) Conical intersection dynamics of the primary photoisomerization event in vision. *Nature* 467:440–443. <https://doi.org/10.1038/nature09346>
81. Koslowski A, Beck ME, Thiel W (2003) Implementation of a general multireference configuration interaction procedure with analytic gradients in a semiempirical context using the graphical unitary group approach. *J Comput Chem* 24:714–726
82. Fabiano E, Thiel W (2008) Nonradiative deexcitation dynamics of 9H-adenine: an OM2 surface hopping study. *J Phys Chem A* 112: 6859–6863
83. Silva-Junior MR, Thiel W (2010) Benchmark of electronically excited states for semiempirical methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J Chem Theory Comput* 6:1546–1564. <https://doi.org/10.1021/ct100030j>
84. Sanchez-Garcia E, Doerr M, Thiel W (2010) QM/MM study of the absorption spectra of DsRed.M1 chromophores. *J Comput Chem* 31:1603–1612. <https://doi.org/10.1002/jcc.21443>
85. Spörkel L, Thiel W (2016) Adaptive time steps in trajectory surface hopping simulations. *J Chem Phys* 144:194108. <https://doi.org/10.1063/1.4948956>
86. Klamt A (1995) Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J Phys Chem* 99:2224–2235. <https://doi.org/10.1021/j100007a062>
87. Marazzi M, Mai S, Roca-Sanjuán D et al (2016) Benzophenone ultrafast triplet population: revisiting the kinetic model by surface-hopping dynamics. *J Phys Chem Lett* 7:622–626. <https://doi.org/10.1021/acs.jpclett.5b02792>
88. Cuquerella MC, Lhiaubet-Vallet V, Cadet J, Miranda MA (2012) Benzophenone photosensitized DNA damage. *Acc Chem Res* 45: 1558–1570. <https://doi.org/10.1021/ar300054e>
89. Shah BK, Rodgers MAJ, Neckers DC (2004) The S2 – S1 internal conversion of benzophenone and *p*-iodobenzophenone. *J Phys Chem A* 108:6087–6089. <https://doi.org/10.1021/jp049615z>
90. Zvereva E, Segarra-Martí J, Marazzi M et al (2018) The effect of solvent relaxation in the ultrafast time-resolved spectroscopy of solvated benzophenone. *Photochem Photobiol Sci* 17:323–331. <https://doi.org/10.1039/c7pp00439g>
91. Segarra-Martí J, Zvereva E, Marazzi M et al (2018) Resolving the singlet excited state manifold of benzophenone by first-principles simulations and ultrafast spectroscopy. *J Chem Theory Comput* <https://doi.org/10.1021/acs.jctc.7b01208>
92. El-Tahawy M, Nenov A, Weingart O et al (2018) On the relationship between excited state lifetime and isomerization quantum yield in animal rhodopsins: beyond the one-dimensional Landau–Zener model. *J Phys Chem Lett* 9(12):3315–3322
93. Arendt D (2003) Evolution of eyes and photoreceptor cell types. *Int J Dev Biol* 47:563–571
94. Murakami M, Kouyama T (2008) Crystal structure of squid rhodopsin. *Nature* 453:363–367. <https://doi.org/10.1038/nature06925>
95. Rinaldi S, Melaccio F, Gozem S et al (2014) Comparison of the isomerization mechanisms of human melanopsin and invertebrate and vertebrate rhodopsins. *Proc Natl Acad Sci* 111:1714 LP–1711719
96. Archipowa N, Kutta RJ, Heyes DJ, Scrutton NS (2018) Stepwise hydride transfer in a biological system: insights into the reaction mechanism of the light-dependent protochlorophyllide oxidoreductase. *Angew Chemie Int Ed* 57:2682–2686. <https://doi.org/10.1002/anie.201712729>
97. Begley TP, Young H (1989) Protochlorophyllide reductase. 1. Determination of the regiochemistry and the stereochemistry of the reduction of protochlorophyllide to chlorophyllide. *J Am Chem Soc* 111:3095–3096. <https://doi.org/10.1021/ja00190a071>
98. Heyes DJ, Ruban AV, Wilks HM, Hunter CN (2002) Enzymology below 200 K: the kinetics and thermodynamics of the photochemistry catalyzed by protochlorophyllide oxidoreductase. *Proc Natl Acad Sci* 99:11145 LP–11111150
99. Heyes DJ, Heathcote P, Rigby SEJ et al (2006) The first catalytic step of the light-driven enzyme protochlorophyllide oxidoreductase proceeds via a charge transfer complex. *J Biol Chem* 281: 26847–26853
100. Heyes DJ, Ruban AV, Hunter CN (2003) Protochlorophyllide oxidoreductase: "dark" reactions of a light-driven enzyme. *Biochemistry* 42:523–528. <https://doi.org/10.1021/bi0268448>
101. Heyes DJ, Hunter CN (2004) Identification and characterization of the product release steps within the catalytic cycle of protochlorophyllide oxidoreductase. *Biochemistry* 43:8265–8271. <https://doi.org/10.1021/bi049576h>
102. Townley HE, Sessions RB, Clarke AR et al (2001) Protochlorophyllide oxidoreductase: a homology model examined by site-directed mutagenesis. *Proteins Struct Funct Genet* 44:329–335. <https://doi.org/10.1002/prot.1098>
103. Sytina OA, Alexandre MT, Heyes DJ et al (2011) Enzyme activation and catalysis: characterisation of the vibrational modes of substrate and product in protochlorophyllide oxidoreductase. *Phys Chem Chem Phys* 13:2307–2313. <https://doi.org/10.1039/C0CP01686A>
104. Sytina OA, Van Stokkum I, Heyes DJ et al (2010) Protochlorophyllide excited-state dynamics in organic solvents studied by time-resolved visible and mid-infrared spectroscopy. *J Phys Chem* 114(12):4335–4344
105. Aquilante F, Pedersen TB, Lindh R (2007) Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals. *J Chem Phys* 126: 194106. <https://doi.org/10.1063/1.2736701>
106. Ke S-H, Baranger HU, Yang W (2007) Electron transport through single conjugated organic molecules: basis set effects in ab initio calculations. *J Chem Phys* 127:144107. <https://doi.org/10.1063/1.2770718>
107. Rohrdanz MA, Herbert JM (2008) Simultaneous benchmarking of ground- and excited-state properties with long-range-corrected density functional theory. *J Chem Phys* 129:34107. <https://doi.org/10.1063/1.2954017>
108. Pepino AJ, Segarra-Martí J, Nenov A et al (2018) UV-induced long-lived decays in solvated pyrimidine nucleosides resolved at the MS-CASPT2/MM level. *Phys Chem Chem Phys* 20:6877–6890. <https://doi.org/10.1039/C7CP08235E>
109. Pepino AJ, Segarra-Martí J, Nenov A et al (2017) Resolving ultrafast photoinduced deactivations in water-solvated pyrimidine nucleosides. *J Phys Chem Lett* 8:1777–1783. <https://doi.org/10.1021/acs.jpclett.7b00316>
110. Improta R, Santoro F, Blancafort L (2016) Quantum mechanical studies on the photophysics and the photochemistry of nucleic acids and nucleobases. *Chem Rev* 116:3540–3593. <https://doi.org/10.1021/acs.chemrev.5b00444>
111. Kong W, He Y, Wu C (2008) Decay pathways of pyrimidine bases: from gas phase to solution BT - radiation induced molecular phenomena in nucleic acids: a comprehensive theoretical and experimental analysis. In: Leszczynski J (ed) Shukla MK. Springer Netherlands, Dordrecht, pp 301–321
112. Hare PM, Crespo-Hernández CE, Kohler B (2007) Internal conversion to the electronic ground state occurs via two distinct pathways for pyrimidine bases in aqueous solution. *Proc Natl Acad Sci* 104:435 LP–435440
113. Hare PM, Middleton CT, Mertel KI et al (2008) Time-resolved infrared spectroscopy of the lowest triplet state of thymine and

- thymidine. Chem Phys 347:383–392. <https://doi.org/10.1016/j.chemphys.2007.10.035>
114. Nenov A, Mukamel S, Garavelli M, Rivalta I (2015) Two-dimensional electronic spectroscopy of benzene, phenol, and their dimer: an efficient first-principles simulation protocol. J Chem Theory Comput 11:3755–3771. <https://doi.org/10.1021/acs.jctc.5b00443>
115. Nenov A, Borrego-Varillas R, Oriana A et al (2018) UV-light-induced vibrational coherences: the key to understand kasha rule violation in *trans*-azobenzene. J Phys Chem Lett 9:1534–1541. <https://doi.org/10.1021/acs.jpcllett.8b00152>
116. Tully JC (1998) Mixed quantum-classical dynamics. Faraday Discuss 110:407–419. <https://doi.org/10.1039/A801824C>
117. Hammes-Schiffer S, Tully JC (1994) Proton transfer in solution: molecular dynamics with quantum transitions. J Chem Phys 101:4657–4667. <https://doi.org/10.1063/1.467455>
118. Serrano-Andrés L, Merchán M, Lindh R (2005) Computation of conical intersections by using perturbation techniques. J Chem Phys 122:104107. <https://doi.org/10.1063/1.1866096>

Notes

The COBRAMM set of routines can be obtained free of charge upon request to the corresponding author.

SUPPORTING INFORMATION

COBRAMM 2.0 – A Program Package for Ground and Excited State QM/MM simulations

Oliver Weingart^{1,‡}, Artur Nenov^{2,3,‡}, Piero Altoè^{2,‡}, Ivan Rivalta⁴, Javier Segarra-Martí^{4,†}, Irina Dokukina¹ and Marco Garavelli^{2,3,4}*

¹Institut für Theoretische Chemie und Computerchemie, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany.

²Dipartimento di Chimica "G. Ciamician", Università degli Studi di Bologna, Via Selmi, 2, I - 40126 Bologna, Italy.

³Dipartimento di Chimica Industriale, Università degli Studi di Bologna, Viale del Risorgimento 4, I-40136 Bologna, Italy.

⁴Université de Lyon, École Normale Supérieure de Lyon, CNRS, Université Claude Bernard Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France.

Table of Contents

S1. QM/MM link atom scheme	page S2
S2. Velocity Verlet and RATTLE MD-Schemes	page S3
S3. Evaluation of SS-CASPT2 non-adiabatic couplings	page S6
S4. Tully surface hopping	page S6
S5. Decoherence correction	page S7
S6. Comparison of ONIOM and COBRAMM QM/MM schemes	page S7
References	page S11

S1. QM/MM link atom scheme

To eliminate the resulting three degrees of freedom per defined link-atom, the link-atom position R_L can be defined as a function of the border atoms Q_1 and M_1 with positions R_{Q_1} and R_{M_1} : [1]

$$R_L(R_{Q_1}, R_{M_1}) = R_{Q_1} + g(R_{M_1} - R_{Q_1}) \quad (S1)$$

The link-atom is now positioned along the bond between the M^1 and Q^1 atoms. The factor g is typically defined as the ratio between the optimal Q^1 -L and Q^1 - M^1 bond lengths. In this definition, the link-atom is allowed to move along the M^1 - Q^1 bond axis as the distance between M^1 and Q^1 may vary, but it is invisible for geometry optimization or MD routines. In the QM calculation, however, a force contribution results for the atom L. This may play only a minor role for geometry optimizations. In MD simulations, this can lead to artificial velocity components for the bound Q^1 atom, which will certainly affect energy conservation. To eliminate these components we must know how the QM/MM energy changes with respect to a change in link-atom position. This relation is given by the chain rule: [2]

$$\frac{\partial \tilde{E}}{\partial R_K} = \frac{\partial E}{\partial R_K} + \frac{\partial E}{\partial R_L} \frac{\partial R_L}{\partial R_K} \quad (S2)$$

Apparently, there is no effect when L and K are unrelated. A force contribution only applies, when K appears in the coordinate definition of L. This is the case for the Q^1 and M^1 atoms. In Cartesian coordinates, $\frac{\partial R_L}{\partial R_K}$ is a 3x3 Jacobian matrix of the partial derivatives of R_L with respect to R_K , i.e.:

$$\frac{\partial R_L}{\partial R_K} = \begin{pmatrix} xx & xy & xz \\ yx & yy & yz \\ zx & zy & zz \end{pmatrix} \quad \text{with } xx = \frac{\partial R_L(x)}{\partial R_K(x)}, \quad xy = \frac{\partial R_L(x)}{\partial R_K(y)} \dots \quad (S3)$$

For a constant g , the coordinate derivatives of Q^1 and M^1 coordinates take the diagonal form:

$$J_{Q_1}^{(\alpha\beta)} = \frac{\partial R_{L(\alpha)}}{\partial R_{Q_1(\beta)}} = (1 - g)\delta_{\alpha\beta}$$

$$J_{M_1}^{(\alpha\beta)} = \frac{\partial R_{L(\alpha)}}{\partial R_{M_1(\beta)}} = g\delta_{\alpha\beta} \quad (\text{S4})$$

with $\delta_{\alpha\beta}$ being the Kronecker delta function. The force contribution to the Q^1 and M^1 atoms in x-direction can then be calculated as:

$$\frac{\partial \tilde{E}}{\partial R_{Q_1(x)}} = \frac{\partial E}{\partial R_{Q_1(x)}} + (1 - g) \quad (\text{S5})$$

and

$$\frac{\partial \tilde{E}}{\partial R_{M_1(x)}} = \frac{\partial E}{\partial R_{M_1(x)}} + g \quad (\text{S6})$$

Similar equations apply for the corresponding y and z directions. This scheme has been implemented into the QM/MM routines of COBRAMM.

S2. Velocity Verlet and RATTLE MD-Schemes

The velocity Verlet code [3] is a self-starting algorithm which needs three quantities: the velocity (V_A), the geometry (R_A) and the forces (F_A) of the molecule at the current time step. The geometry for the next time step is then estimated through:

$$R_A(t + \Delta t) = R_A(t) + V_A(t + \Delta t)\Delta t \quad (\text{S7})$$

with

$$V_A(t + \Delta t) = V_A(t) + \frac{1}{2M_A} [F_A(t) + F_A(t + \Delta t)]\Delta t \quad (\text{S8})$$

The quantity $F_A(t+\Delta t)$ is not known and, therefore, the half step velocity is computed by:

$$V_A(t + \Delta t_{1/2}) = V_A(t) + \frac{1}{2M_A} F_A(t) \Delta t \quad (\text{S9})$$

This quantity is used to compute the new full step geometry, resulting as:

$$R_A(t + \Delta t) = R_A(t) + V_A(t) \Delta t + \frac{1}{2M_A} F_A(t) \Delta t^2 \quad (\text{S10})$$

Due to the half step velocity approximation, the sum of computed kinetic and potential energies (the total energy) undergoes algorithm specific fluctuations. The timestep Δt must be chosen small enough to warrant energy conservation and minimize artefacts. Typically, the time step is chosen to model high-frequency vibrations of the molecule resulting, e.g., from C-H stretching modes ($\sim 2500\text{-}3000 \text{ cm}^{-1}$) with enough points. Proper modeling of one half period requires at least five points, which can be realized with the choice of 0.25-0.5 fs time steps.

Consideration of high frequency motions in ultrafast semiclassical photodynamics simulations is a problematic issue. As mentioned in the molecular dynamics implementation section of the paper, the ZPE attributes a large amount of kinetic energy to these modes. In photodynamics reactions, the molecule often gains more kinetic energy while running down the excited-state PES. Due to the classical nuclei description, the energy will not stay in the corresponding modes, but will rather re-distribute and lead to an artificial thermalisation of the molecule in long time simulations. This can be avoided by eliminating the mode couplings in the Hamiltonian or forcing at least the ZPE to stay in the corresponding modes. This, however, requires computation of energy second derivatives and is often too costly for standard applications. In many photochemical reactions, high frequency C-H and N-H modes do not play a role in the reactive mechanism, it is therefore sensible to exclude these modes from the initial sampling and freeze them out during dynamics by using RATTLE [4] or SETTLE [5] schemes. In these approaches, the high frequency oscillations are eliminated by redistributing the velocity components such, that the C-H, N-H or O-H fragments move as one rigid entity. In the RATTLE approach, the constraint condition

$$\sigma_{AB}(t)(\{R(t)\}) \equiv [R_A(t) - R_B(t)]^2 - d_{AB}^2 = 0 \quad (\text{S11})$$

must be fulfilled. d_{AB} is the fixed bond length between the two involved atoms A and B . Velocity restriction applies through

$$[V_A(t) - V_B(t)] \cdot [R_A(t) - R_B(t)] = 0 \quad (\text{S12})$$

The equation of motion for constrained dynamics can be written as

$$M_A \frac{d^2 R_{A(t)}}{dt^2} = F_A(t) + G_A(t) \quad (\text{S13})$$

where $G_A(t)$ corresponds to the force on atom A due to the constraint. This restrictive force is given by

$$G_A(t) = - \sum_{B=1}^{N_r} \lambda_{AB}(t) \frac{\partial}{\partial R_A} \sigma_{AB}(\{R(t)\}) \quad (\text{S14})$$

The term $\lambda_{AB}(t)$ is a Lagrange Multiplier. The corresponding equation for the next geometry within the velocity Verlet scheme is then, including the bond constraint between atoms A and B , given as

$$R_A(t + \Delta t) = R_A(t) + V_A(t)\Delta t + \frac{1}{2M_A} [F_A(t) - 2 \sum_{B=1}^{N_r} \lambda_{RR,AB}(t) R_{AB}(t)] \Delta t^2 \quad (\text{S15})$$

with

$$R_{AB}(t) = R_A(t) - R_B(t) \quad (\text{S16})$$

for the next geometry and

$$V_A(t + \Delta t) = V_A(t) + \frac{1}{2M_A} [F_A(t) - 2 \sum_{B=1}^{N_r} \lambda_{RR,AB}(t) R_{AB}(t) + F_A(t + \Delta t) - 2 \sum_{B=1}^{N_r} \lambda_{RV,AB}(t + \Delta t) R_{AB}(t + \Delta t)] \Delta t \quad (\text{S17})$$

for the corresponding velocity. The sum runs over those atoms B which have a restrictive bond with atom A . The Lagrange-multipliers $\lambda_{RR,AB}(t)$ and $\lambda_{RV,AB}(t + \Delta t)$ are chosen such that they fulfill the conditions in equations (S12) and (S14). This requires iterative solution of geometry and velocity equations (S15) and (S17), where each restriction on atom pairs A and B is treated separately. Application of this scheme is ultimately necessary, when e.g. rigid-type solvent models such as the TIP3P model are used in molecular dynamics simulations. The parameters for this water model are defined to deliver reliable ground state structures through optimization. When used in dynamics, unphysically high frequency oscillations (ca. 6000 cm^{-1}) will govern the water dynamics. The RATTLE approach for TIP3P water therefore fixes both, the distances between O and H, and the H-O-H angle by restraining the distance between the two H atoms. Application of RATTLE or SETTLE schemes will minimize the energy flow from the high to the low frequency modes and allows for larger time steps without major loss of accuracy,

increasing energy conservation. Both, the velocity Verlet and the RATTLE scheme have been implemented into the QM/MM package COBRAMM.

S3. Evaluation of SS-CASPT2 non-adiabatic couplings

The individual terms are obtained from WF overlaps. We use the CASSCF wavefunctions and correct the final expression by the ratio of the SS-CASPT2 and CASSCF energy gaps ($E_j^{CASSCF} - E_i^{CASSCF}$)/($E_j^{SSPT2} - E_i^{SSPT2}$) following:

$$\begin{aligned}
\sigma_{ij} &= \langle \Psi_i^{SSPT2}(t) | \frac{d}{dt} | \Psi_j^{SSPT2}(t) \rangle = \frac{dR}{dt} \cdot \langle \Psi_i^{SSPT2}(t) | \nabla_R | \Psi_j^{SSPT2}(t) \rangle = \\
&= v \cdot d_{ij}^{SSPT2} = v \cdot \frac{\langle \Psi_i^{SSPT2}(t) | \nabla_R \hat{H} | \Psi_j^{SSPT2}(t) \rangle}{E_j^{SSPT2} - E_i^{SSPT2}} \approx \\
&\approx v \cdot \frac{\langle \Psi_i^{CASSCF}(t) | \nabla_R \hat{H} | \Psi_j^{CASSCF}(t) \rangle}{E_j^{CASSCF} - E_i^{CASSCF}} \cdot \frac{E_j^{CASSCF} - E_i^{CASSCF}}{E_j^{SSPT2} - E_i^{SSPT2}} = \\
&= v \cdot d_{ij}^{CASSCF} \cdot \frac{\Delta(E_{ji}^{CASSCF})}{\Delta(E_{ji}^{SSPT2})}
\end{aligned} \tag{S18}$$

Above chain following equality is assumed

$$\langle \Psi_i^{SSPT2}(t) | \nabla_R \hat{H} | \Psi_j^{SSPT2}(t) \rangle \approx \langle \Psi_i^{CASSCF}(t) | \nabla_R \hat{H} | \Psi_j^{CASSCF}(t) \rangle \tag{S19}$$

Effectively, when SS-CASPT2 correction decreases the energy gap with respect to CASSCF the value of the time-derivative coupling increases.

S4. Tully surface hopping

The total wavefunction of the system can be described by a linear combination of all involved states accompanied by a linear coefficient c for each state:

$$\Psi(r, R, t) = \sum_i c_i(t) \phi_i(r; R(t)) \tag{S20}$$

The squared state coefficients give the weight of the corresponding state in the full wavefunction. Inserting this ansatz into the time-dependent Schrödinger equation

$$\left(i\hbar \frac{\partial}{\partial t} - H_e \right) \Psi(r, R, t) = 0 \tag{S21}$$

and recasting the equations yields the time evolution of the (complex) state coefficients:

$$i\hbar \frac{\partial c_j(t)}{\partial t} = \epsilon_j c_j(t) - i\hbar \sum_i d_{ji} \cdot \dot{R} c_i(t) \quad (\text{S22})$$

$$d_{ji} = \langle \Psi_j | \frac{\partial}{\partial R} \Psi_i \rangle = d_{ij} \quad (\text{S23})$$

In these coupled equations, d_{ji} is the interstate coupling vector (S23), ϵ_j is the energy of state j and \dot{R} is the velocity vector along the trajectory. Integration of this equation yields the state amplitudes, which are connected with the state probabilities.

S5. Decoherence correction

According to equations S22 and 2.2 in the paper, each nuclear MD step yields new state amplitudes c_j . In Persico's decoherence correction algorithm, the obtained state amplitudes will be empirically modified to correct the mismatch between electronic populations and trajectory ensemble occupations. The amplitudes for all states ($k \neq m$), except for the currently populated one (m), are damped by an exponential factor:

$$c'_k = c_k e^{-\Delta t / \tau_{km}} \quad (\text{S24})$$

With

$$\tau_{km} = \frac{\hbar}{|\epsilon_k - \epsilon_m|} \left(1 + \frac{C}{E_{kin}} \right) \quad (\text{S25})$$

where e_k and e_m denote the energies of the corresponding states, and E_{kin} is the kinetic nuclear energy along the trajectory. The empirical correction factor C is typically chosen to be 0.1 E_h. The corrected amplitudes for the current state are computed as:

$$c'_m = c_m \left[\frac{1 - \sum_{k \neq m} |c'_k|^2}{|c_m|^2} \right]^{\frac{1}{2}} \quad (\text{S26})$$

This correction is applied after estimation of the probabilities P_{ij} and showed good performance in numerous applications.

S6. Comparison of ONIOM and COBRAMM QM/MM schemes

To verify the implementation of QM/MM routines in COBRAMM we performed a comparative DFT/Amber study starting from the crystal structure 1U19 of the rhodopsin photosystem. Prior to QM/MM computations the crystal structure was geometry-optimized with the Amber ff99 force field for 1000 steps to remove effects from crystal packing and close contacts after adding hydrogens. The whole retinal chromophore including one methylene group of Lys296 was selected as the QM part. With this setup, geometry optimizations of the retinal chromophore within a fixed protein environment were performed at the B3LYP/6-31G* QM level and with the Amber ff99 MM method. Vertical excitation energies for both programs were computed with TD-CAM-B3LYP/6-31G*. Figure S1 gives an overview on the numbering of the retinal chromophore. MM optimized and QM/MM optimized bond lengths are collected in Table S1. Angle and torsion parameters are listed in Tables S2 and S3. Excitation energies and state compositions are found in Table S4. Geometry parameters as well as excitation energies and state configurations are in valuable agreement between ONIOM and COBRAMM.

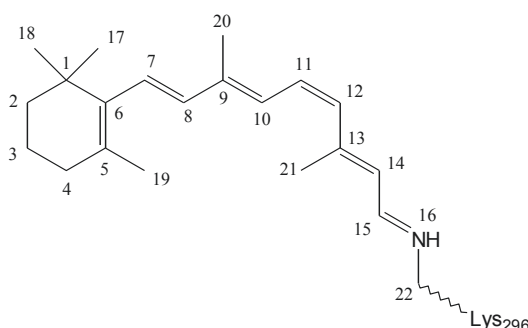


Figure S1: Atom numbering in retinal chromophore.

Table S1: Retinal bond lengths at QM/MM optimized geometries.

Bond	ONIOM [Å]	COBRAMM [Å]	Difference [Å]
C1-C2	1.544	1.545	0.001
C1-C6	1.544	1.546	0.002
C2-C3	1.526	1.527	0.001
C3-C4	1.531	1.531	0.000
C4-C5	1.514	1.514	0.000
C5-C6	1.358	1.358	0.000
C6-C7	1.462	1.465	0.003
C7-C8	1.359	1.360	0.001
C8-C9	1.440	1.444	0.004
C1-C17	1.545	1.545	0.000
C1-C18	1.547	1.547	0.000
C5-C19	1.506	1.508	0.002
C9-C10	1.382	1.382	0.000
C9-C20	1.502	1.503	0.001
C10-C11	1.415	1.418	0.003
C11-C12	1.382	1.383	0.001
C12-C13	1.419	1.424	0.005
C13-C14	1.402	1.402	0.000
C13-C21	1.509	1.508	-0.001
C14-C15	1.387	1.392	0.005
C15-N16	1.331	1.332	0.001
N16-C22	1.465	1.476	0.011

Table S2: Retinal bond angles at QM/MM optimized geometries.

Bond angle	ONIOM [°]	COBRAMM [°]	Difference [°]
C1-C2-C3	112.109	112.429	0.320
C1-C6-C5	122.173	122.053	-0.120
C1-C6-C7	115.181	115.606	0.425
C2-C1-C6	109.921	109.785	-0.136
C2-C3-C4	110.510	110.573	0.063
C3-C4-C5	115.216	115.515	0.299
C4-C5-C6	122.598	122.771	0.173
C5-C6-C7	122.297	122.093	-0.204
C6-C7-C8	125.632	125.362	-0.270
C7-C8-C9	125.877	126.303	0.426
C2-C1-C17	107.354	107.258	-0.096
C2-C1-C18	110.970	110.837	-0.133
C4-C5-C19	112.021	112.004	-0.017
C6-C1-C17	111.224	111.336	0.112
C6-C1-C18	108.038	108.666	0.628
C6-C5-C19	125.280	125.177	-0.103
C8-C9-C10	117.713	117.735	0.022
C8-C9-C20	119.482	119.321	-0.161
C10-C9-C20	122.764	122.890	0.126
C17-C1-C18	109.349	108.961	-0.388
C9-C10-C11	125.443	125.786	0.343
C10-C11-C12	129.743	130.160	0.417
C11-C12-C13	130.886	131.007	0.121
C12-C13-C14	116.102	116.702	0.600
C12-C13-C21	122.399	121.698	-0.701
C13-C14-C15	124.296	124.646	0.350
C14-C13-C21	121.496	121.601	0.105
C14-C15-N16	122.832	123.035	0.203
C15-N16-C22	123.673	123.093	-0.580

Table S3: Retinal dihedral angles at QM/MM optimized geometries.

Dihedral angle	ONIOM [°]	COBRAMM [°]	Difference [°]
C1-C2-C3-C4	59.308	58.642	-0.666
C1-C6-C7-C8	138.774	137.677	-1.097
C2-C1-C6-C5	25.184	25.273	0.089
C2-C1-C6-C7	-161.447	-160.356	1.091
C2-C3-C4-C5	-38.639	-37.111	1.528
C3-C4-C5-C6	13.082	11.572	-1.510
C4-C5-C6-C1	-6.443	-5.742	0.701
C4-C5-C6-C7	-179.368	-180.000	-0.632
C5-C6-C7-C8	-47.864	-47.955	-0.091
C6-C1-C2-C3	-51.433	-51.638	-0.205
C6-C7-C8-C9	176.694	177.315	0.621
C17-C1-C2-C3	-172.538	-172.742	-0.204
C17-C1-C6-C5	143.941	143.893	-0.048
C17-C1-C6-C7	-42.690	-41.737	0.953
C18-C1-C2-C3	68.018	68.422	0.404
C18-C1-C6-C5	-96.040	-96.097	-0.057
C18-C1-C6-C7	77.329	78.271	0.942
C19-C5-C6-C1	169.637	171.553	1.916
C19-C5-C6-C7	-3.259	-2.451	0.808
C3-C4-C5-C19	-163.468	-166.046	-2.578
C7-C8-C9-C10	178.049	174.479	-3.570
C7-C8-C9-C20	0.316	-2.953	-3.269
C8-C9-C10-C11	172.954	174.416	1.462
C20-C9-C10-C11	-9.421	-8.255	1.166
C9-C10-C11-C12	174.026	172.151	-1.875
C10-C11-C12-C13	-15.884	-15.284	0.600
C11-C12-C13-C14	171.405	171.672	0.267
C11-C12-C13-C21	-9.096	-8.400	0.696
C12-C13-C14-C15	-178.481	-178.999	-0.518
C13-C14-C15-N16	171.812	171.146	-0.666
C14-C15-N16-C22	177.806	178.387	0.581
C21-C13-C14-C15	2.027	1.049	-0.978

Table S4: CAM-B3LYP/6-31G**//Amberff99 computed excitation energies, oscillator strengths, configurations and their coefficients with ONIOM and Amber.

ONIOM					
<i>state</i>	$\Delta E(S_0)$ [eV]	λ [nm]	<i>f</i>	<i>configuration</i>	<i>coeff.</i>
S ₁	2.68	463	1.44	H→L	+0.70
S ₂	3.85	322	0.24	H-1→L	+0.69
S ₃	4.66	266	0.11	H-2→L, H→L+1	+0.45 +0.52
COBRAMM					
<i>state</i>	$\Delta E(S_0)$ [eV]	λ [nm]	<i>f</i>	<i>configuration</i>	<i>coeff.</i>
S ₁	2.71	457	1.44	H→L	+0.70
S ₂	3.90	318	0.23	H-1→L	+0.69
S ₃	4.67	266	0.13	H-2→L, H→L+1	+0.44 +0.52

References

1. Dapprich S, Komáromi I, Byun KS, et al (1999) A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. J Mol Struct 461–462:1–21 . doi: [http://dx.doi.org/10.1016/S0166-1280\(98\)00475-8](http://dx.doi.org/10.1016/S0166-1280(98)00475-8)
2. Senn HM, Thiel W (2007) QM/MM Methods for Biological Systems. Top Curr Chem 268:173–290 . doi: 10.1007/128
3. Verlet L (1967) Computer experiments on classical fluids. Phys Rev, 159:98–103
4. Andersen HC (1983) Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. J Comput Phys 52:24–34 . doi: [http://dx.doi.org/10.1016/0021-9991\(83\)90014-1](http://dx.doi.org/10.1016/0021-9991(83)90014-1)
5. Miyamoto S, Kollman PA (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13:952–962 . doi: 10.1002/jcc.540130805

Paper IV

QM/MM photodynamics of Retinal in the Channelrhodopsin Chimera C1C2 with OM3/MRCI

I. Dokukina, A. Nenov, M. Garavelli, C. M. Marian, O. Weingart

The manuscript was submitted to ChemPhotoChem (2018)

QM/MM photodynamics of Retinal in the Channelrhodopsin Chimera C1C2 with OM3/MRCI

Irina Dokukina¹, Artur Nenov², Marco Garavelli², Christel M. Marian¹, and Oliver Weingart^{*1}

¹Institut für Theoretische Chemie und Computerchemie, Heinrich-Heine-Universität
Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

²Dipartimento di Chimica Industriale "Toso Montanari", Università degli Studi di Bologna,
Viale del Risorgimento, 4, 40136 Bologna, Italia

September 6, 2018

Abstract

The photoisomerization of all-*trans* retinal in channelrhodopsins triggers the opening of a cation channel in the membrane of certain green algae. The stimulus resulting from cell depolarization enables the algae to perceive and react to the present lighting conditions. This property makes channelrhodopsins especially interesting for neuroscientific applications in optogenetics. We investigate the initial photoreaction in the channelrhodopsin chimera C1C2 with a combined quantum mechanical/molecular mechanical (QM/MM) strategy. For geometry optimizations, the OM3/RHF and OM3/MRCI/Amber protocols are used. Trajectory surface hopping calculations are performed with OM3/MRCI/Amber and a statistically relevant number of molecules to obtain reaction rates and quantum yields. The utilized models include a direct hydrogen bond of the retinal Schiff base to the proposed counterions in C1C2, i.e. either toward glutamate or aspartate. Static and dynamic aspects are compared to experimental findings including time constants and spectral traces. The calculations provide a rationale for the observed short and long time components in the photokinetics of C1C2 and yield detailed insight into the photoreaction mechanisms that include bicycle pedal and hula twist motions.



Figure 1: Initial all-*trans* to 13-*cis* photoreaction of retinal in microbial rhodopsins

1 Introduction

Channelrhodopsins (ChRs) are light-gated ion-channels in green algae such as e.g. *Chlamydomonas reinhardtii* or *Haematococcus pluvialis*. [1–3] In these organisms, they play a central role in mediating phototaxis by inducing motion in response to the exterior light conditions. [4–6]

A common feature of ChRs and other microbial (e.g. bacteriorhodopsin, halorhodopsin) and animal rhodopsins (e.g. visual rhodopsin) is their activation by a cofactor, the retinal chromophore. [7] Retinal is covalently bound to a lysine (LYS) of the protein backbone via a protonated Schiff base (RPSB) linkage. Upon light excitation, RPSB changes its conformation from 11-*cis* to all-*trans* in animal rhodopsins and from all-*trans* to 13-*cis* in microbial rhodopsins (Figure 1). [8–11] Retinal isomerization induces conformational changes within the protein, which in ChRs lead to the formation of a channel in the cell membrane, enabling immediate cation influx followed by cell depolarization. [12] In green algae, depolarization is responsible for movement toward or away from a light source. [4–6]

The discovery of ChRs formed a major breakthrough in optogenetics, a scientific field combining methods of optics and genetics. [13, 14] Optogenetical tools enable scientists to easily control single neurons by light and caused a radical change at a technical level of neurobiological research. [15–18] Excitable cells (e.g. muscle fibers or neuron cells) equipped with ChRs can be rapidly and selectively depolarized and thus switched on by light. [7] The demand for faster and more precise tools to investigate and influence neuronal activity (e.g. brain circuits) further boosted the developments in the field of optogenetics. Meanwhile, further types of ChRs have been discovered and engineered. [17, 19] Herein, the modifications of the protein environment plays a pivotal role whereas the cofactor remains unchanged. Blue light absorbing ChR2 found a widespread application in neuroscientific research due to its rapid on-rate and higher conductance. Red light absorbing variants are of special interest, since red light provides deeper tissue penetration with less cell damage. [20]

The pivotal impact of optogenetics becomes apparent in medicine, where promising strategies for effective treatment of several diseases (e.g., macular degeneration, Parkinson’s disease) evolve. [15, 16, 21]

Numerous experimental and theoretical studies on ChRs were performed to reveal their action mechanism or to modify properties such as absorption wavelength or ion conductance. While experimental studies can involve any available ChR, structure-based theoretical research requires reliable geometry information. Until recently, the only known ChR crystal structure was C1C2, a fully functional chimera protein, containing

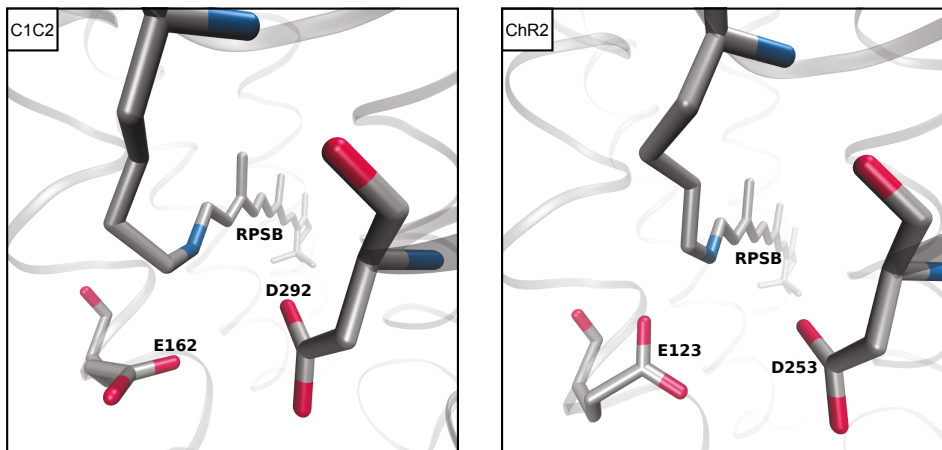


Figure 2: Protonated retinal Schiff base and negatively charged counter ions glutamate (E) and aspartate (D) in C1C2 (left) and ChR2 (right).

5 helices from ChR1 and 2 helices from ChR2. [22] This structure was used in numerous theoretical studies and also provided a template for homology modelling of other ChRs (e.g. ChR2). The crystal structure of ChR2 has been obtained very recently by Volkov et al. [23]

The binding pocket in C1C2 and also in ChR2 (Figure 2) contains two counterions in the vicinity of the RPSB linkage, namely glutamate (E162 in C1C2, E123 in ChR2) and aspartate (D292 in C1C2, D253 in ChR2). These negatively charged residues significantly influence the excited states of the chromophore and therefore the absorption properties of retinal, as has been demonstrated in recent work. [24] Both counterions are in principle able to accept the Schiff base proton after completion of the photoisomerization. A comprehensive picture of the primary proton acceptor and the hydrogen bonded network in the vicinity of retinal in C1C2 does not yet exist. Ito et al. suggest a model of hydrogen bonding between retinal and the glutamate counterion based on low-temperature Fourier transform infra red (FTIR) spectroscopy measurements. [25] In the publication related to the crystal structure of C1C2, D292 is assumed to play the role of the primary proton acceptor. This assumption is supported by geometrical data and mutation studies. In the crystal structure, the retinal Schiff base is located closer to aspartate (by ca. 0.4 \AA) than to glutamate. Furthermore, mutation of D292 to alanine resulted in a substantial decrease in photocurrents, while the E162A mutant showed moderately decreased photocurrents. [22]

VanGordon et al. [26] investigated intramolecular interactions in the closed state of C1C2 by performing atomistic molecular dynamics (MD) simulations. Their study reveals hydrogen bonds between RPSB and both counterions. Based on these findings, the authors propose aspartate to be the primary proton acceptor, at the same time emphasizing that glutamate cannot be completely neglected in this sense. A recent paper by Adam and Bondar [27] discusses the stability of the protonated all-*trans* RPSB in C1C2. In their dynamics simulations, water molecules enter the RPSB binding pocket.

The authors computed retinal de-protonation paths via direct hydrogen bonding to the counterions and via water-mediated hydrogen bonded networks, stating that both forms (water mediated or directly bound to any of the counterions) are stable in the ground state. De-protonation of retinal is facilitated, when all water molecules in the active site, even those from the crystal structure, are removed. So far, both starting configurations, i.e., with RPSB hydrogen bonded toward E162 or toward D292 are plausible.

Time-resolved UV/Vis absorption measurements for wild type (wt) ChR2 and several mutants were recorded by Scholz et al. at a resolution of 50-90 fs. [28] These experiments indicated a pH dependence of the initial photoreaction and a significant influence of E123 mutations. Replacement of this residue by a negatively charged aspartate (E123D) leads to a speedup of the photoisomerization, whereas the neutral threonine mutant (E123T) slowed down the initial reaction kinetics. The measured spectra lack some features that are, e.g., known from bovine rhodopsin. [29] In the latter, a quick red-shift of the chromophore stimulated emission (SE) is observed, when retinal moves toward the conical intersection (CoIn) along the torsional C11=C12 coordinate and the S_1/S_0 energy difference decreases. This indicates that motion toward the CoIn in ChRs rather occurs in a state with lower oscillator strength. Traces of the red shifting SE are, however, barely visible in the wt ChR2 spectra at short times. [28] This SE signal gained further intensity in the mutation experiments.

First insights into the photodynamics of C1C2 were given in a recent, combined experimental and computational study by Hontani et al. [30] The authors report three major components in the decay of excited retinal with time constants of 450 fs, 2 ps and 11 ps, estimated by transient absorption and pump-dump-probe experiments. The study suggests that the fast 450 fs component leads to the 13-*cis* photoproduct with 0.3 quantum yield, while the slower components are not productive. These findings are supported by quantum mechanical/molecular mechanical (QM/MM) calculations of the complete active space with second-order perturbational corrections//complete active space self-consistent field with 12 active electrons in 12 active orbitals (CASSCF(12,12)//CASPT2) profiles for torsion about the C13=C14 bond and a CASSCF trajectory without initial velocities (denoted as 0K-trajectory). [30] This single trajectory can, however, only provide qualitative insight into the photodynamics of retinal as it does not account for the statistical effects occurring in a molecular ensemble. The computational study considered two configurations of hydrogen-bonded networks as starting structures. In one, RPSB was hydrogen bonded directly to E162. The second structure involves RPSB in a network with two water molecules, with one water bound to D292. Direct hydrogen bonding to this residue was not considered. Computed excited-state torsion paths about the C13=C14 bond resulted in two conical intersections (CI1 and CI2). CI1 leads, by clockwise torsion, nearly barrierlessly, to the 13-*cis* product and was thus assigned to the fast 450 fs component. After isomerization, retinal is in the 13-*cis*,15-*anti* conformation, i.e. the NH^+ -terminus points upwards and hydrogen bonding of RPSB is not preserved. CI2 was suggested to provide non-productive counterclockwise torsion assigned to the 2 ps and 11 ps time constants, depending on the starting ground state structure (RPSB bonded with water or E162). The authors propose that through CI2 retinal relaxes back to the initial all-*trans* conformation.

Our current work aims at clarifying the initial retinal isomerization mechanism in C1C2 with a sufficiently large number of trajectories to obtain meaningful statistical results. The environment of the retinal chromophore and especially the two counterions near RPSB (E162 and D292), play a crucial role in this process and in following thermal reactions. Therefore, we use combined QM/MM strategies to describe retinal and the surrounding opsin protein. We apply the semiempirical orthogonalization model 3 (OM3) RHF Hamiltonian [31, 32] and the AMBER force field [33] for retinal geometry optimization within C1C2. The photodynamics of the resulting models are studied with the OM3 multireference configuration interaction (MRCI) method and a trajectory surface hopping approach to obtain quantum yields, reaction times and possible photoproducts. Three sets of trajectories are computed, where two sets represent the orientation of the NH^+ -proton toward E162 and one set where the NH^+ -terminus is involved in a direct hydrogen bond with D292.

2 Computational methods

All calculations were carried out with a recent version of the program package COBRAMM, developed in Bologna, Lyon and Düsseldorf. [34, 35] It provides interfaces to several QM- and MM-programs, enabling hybrid QM/MM calculations in a subtractive scheme with electrostatic embedding and the hydrogen link atom approach.

QM/MM setup. For the QM/MM calculations, a previously established protein setup was employed (see ref. [24] for details). The QM-region, which includes retinal and the ε - CH_2 group of the bound lysine, was described by the semiempirical OM3 Hamiltonian [31, 32] complemented by multireference configuration interaction (MRCI) based on the graphical unitary group approach (GUGA) [36] when applied to excited state calculations. The remaining protein was treated at the MM-level by the ff99SBildn-Amber force field. [37] We used OM3 instead of OM2, because OM3 has shown better convergence during geometry optimizations for our system. In a previous study of linear polyenes [38] we have shown that OM2 and OM3 deliver qualitatively similar results.

Geometry optimizations. Ground-state geometry optimizations were performed with the OM3 restricted Hartree–Fock (RHF) Hamiltonian. For excited-state calculations, the OM3/MRCI method was used with molecular orbitals generated by the restricted open-shell Hartree–Fock single (ROHF) approach. The reference space included single and double excitation within an active space of 20 electrons in 20 orbitals. The same strategy was also applied by Guo et al. [39]. All excited-state geometry optimizations were performed with only retinal (RET, QM-part) and the remaining part of lysine 296 (LYP, MM-part) movable, whilst keeping the remaining portion of the protein fixed.

Excited-state torsion path. For the calculation of the S_1 -torsion path, the OM3-optimized ground-state geometry of the corresponding model was used as a starting point. The gradients were calculated for the first excited state (S_1), characterized by the HOMO-LUMO excitation at the Franck-Condon point. A relaxed scan was performed along the C12-C13=C14-C15 torsion coordinate. The torsion angle was altered in 2

degree steps considering both clockwise and counterclockwise directions.

Geometry sampling. To generate initial geometries and velocities for MD trajectories we performed a zero-point energy (ZPE) [40] sampling using Gaussian09 [41]. The OM3-optimized ground state geometry including retinal and the lysine residue was used to obtain the required displacements via vibrational frequency calculations. High-frequency C-H and N-H modes were excluded from the sampling procedure as they may lead to total energy fluctuations and do not influence the reaction. [29, 42–44]

Excited-state dynamics. Excited-state dynamics calculations were performed using a velocity-Verlet integrator scheme with an adapted time step [45] starting at 0.1 fs. All trajectories were started in the first singlet excited state corresponding to the HOMO–LUMO excitation. An active space consisting of 20 electrons in 20 orbitals was defined. State populations were computed by propagating the time-dependent wavefunction in each individual trajectory with a step size of 1/100 of the MD step. Tully’s fewest-switches surface-hopping algorithm [46, 47] with decoherence correction [48, 49] was applied to provide hopping probabilities. State populations and state occupations resulting from trajectory surface hopping were collected each full fs and averaged over the whole ensemble. In the same way, excitation energies and transition dipole moments between S_0 and the S_1 and S_2 states were collected to produce time-resolved UV/Vis spectra. For spectra generation, energy differences and oscillator strengths for all trajectories of the corresponding model were Gaussian weighted and convoluted with a Gaussian function of a) 1 fs or b) 16 fs full width at half maximum (FWHM) to obtain data comparable to the highest experimental resolution.

Excited-state lifetime. To obtain the lifetimes of the S_1 and S_2 states, we applied exponential fitting using the following equation:

$$y(t) = A \cdot e^{(-t/\tau)} + y_0 \quad (1)$$

Herein, A is the amplitude denoting the initial population, τ corresponds to the fitted time constant and y_0 is an offset parameter of the fit function.

3 Results & Discussion

3.1 Investigated models

We performed our computations with three models: ASP, GLU1 and GLU2, denoting hydrogen bonds of RPSB to the corresponding counterions. The models were obtained by including different numbers of amino acid residues into the optimizable MM region during QM/MM ground state geometry optimizations. A complete list of optimized residues is given in Table 1. Figure 3 gives an overview over the orientation of the NH^+ moiety in the separate models. Figure 4 depicts the retinal structures in the ASP and GLU2 models.

The ASP model was obtained by relaxing only retinal and the connected lysine. The remaining surrounding was kept frozen at crystal structure positions. In this case, the NH^+ -terminus points toward the counterion D292 after geometry optimization. This

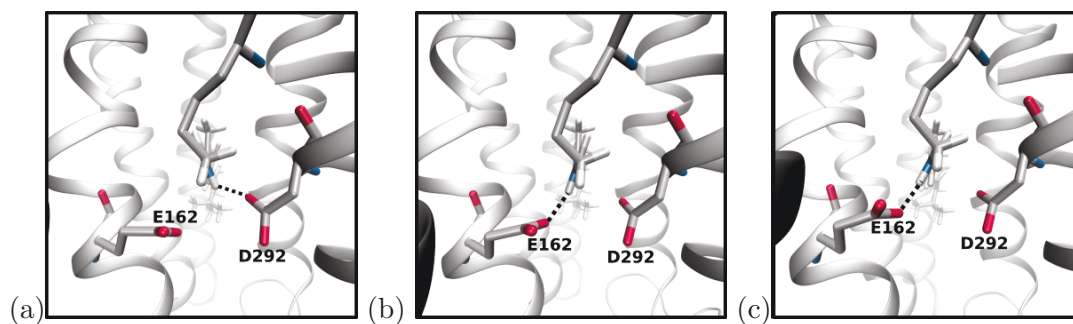


Figure 3: Orientation of the NH^+ -terminus in ground-state optimized minima of (a) ASP, (b) GLU1 and (c) GLU2 models. The hydrogen bond with the corresponding counter ion is depicted by a dashed line.

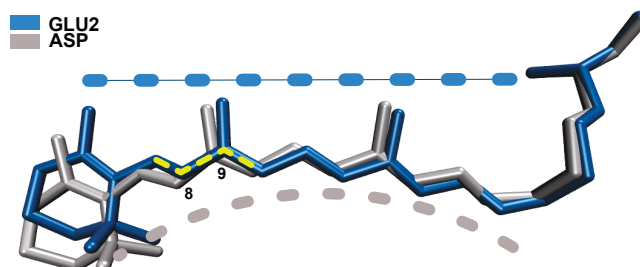


Figure 4: Overlay of optimized retinal in the ground state in ASP (grey) and GLU2 (blue) models. The yellow dashed line denotes an additional twist in the corresponding dihedral angle (see text).

finding resembles the results of CASSCF, RHF and DFT optimizations performed in a previous C1C2-study by some of us. [24] The overall retinal structure is almost planar but somewhat bent along the polyene chain (see Figures 3 (a) and the grey structure in Figure 4).

The GLU1 model included the two counterions E162 and D292 as well as a close-by lysine (K132) and tyrosin (T166) in the mobile part of the optimization. The retinal chromophore in this model shares the same structural features with the ASP model and has a nearly planar and bent structure.

The GLU2 model had 24 residues within a 5 Å distance from the retinal chromophore which were additionally minimized. We note a significant twist about the C8-C9 bond after optimization (ca. 27°, see Figure 4) and a different, more linear shape in the polyene chain than in the latter two models (blue structure in Figure 4).

Table 1: Residues optimized in the QM and MM regions during ground state calculations. The lysine residue K296 (bold font) was optimized in all calculations.

Model	residues QM	residues MM
ASP	RET	K296
GLU1	RET	K132; E162; T166; D292; K296
GLU2	RET	E129; K132; Y160; E162; W163; T166; C167; I170; T198; I199; G202; T203; F216; F217; G220; L221; G224; F228; W262; F265; P266; F269; D292; S295; K296

3.2 Static calculations

3.2.1 Ground-state geometries

To characterize the created models prior to dynamics calculations, we analyzed bond lengths and electronic properties of the obtained ground-state (GS) models. Additionally, we performed S_1 and S_2 excited-state geometry optimizations of the models, with retinal and lysine in the movable region. The obtained results are presented in the following. In Figure 5 we show bond lengths of the computed ground- and excited-state minima in the ASP, GLU1 and GLU2 models. All three models show similar bond-length patterns. The GS geometries have alternating double and single bonds, with the terminal C=N showing the shortest bond length.

A detailed description of excitation energies and excited-state character is provided in Table 2. Some general trends can be noted for all models. All $S_0 \rightarrow S_1$ excitation energies are 0.64-0.84 eV higher than the experimental value (ca. 2.63-2.64 eV [22, 25]). Similar observations have been made by Guo et al. [39] for calculated OM2/MRCI absorption spectra of ChR2, where the authors obtained ca. 0.56 eV deviation from the experimental value. (Figure 7 in the ref. [39], red value). The leading configuration of the S_1 state refers to the HOMO–LUMO excitation, thus to the spectroscopically

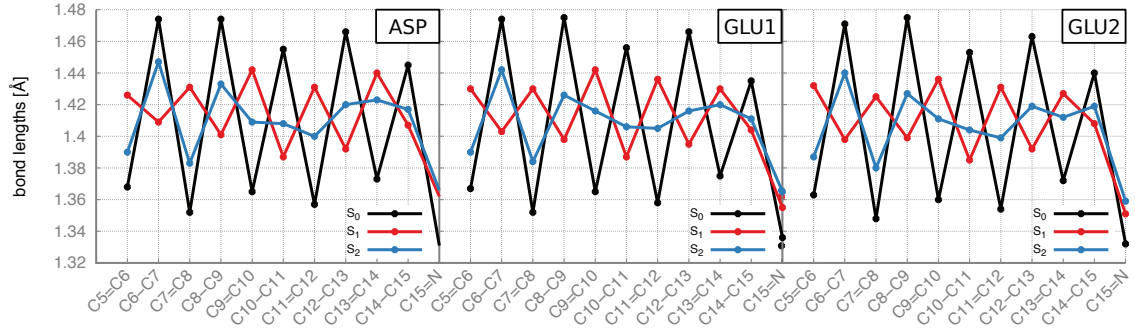


Figure 5: Bond lengths of the S_0 , S_1 and S_2 optimized retinal geometries in ASP, GLU1 and GLU2 models.

bright state. Excitation to the S_2 state is weakly allowed and is characterized by the $\text{HOMO}^2\text{-LUMO}^2$ double excitation accompanied by the typical $\text{HOMO-LUMO}+1$ and $\text{HOMO}-1\text{-LUMO}$ configurations.

Table 2: Vertical excitation energies, oscillator strengths, electronic configurations and weights of OM3-optimized geometries.

State	ΔE <i>eV</i>	<i>f</i>	conf.	weight %
ASP				
S ₀	0	-	GS	89
S ₁	3.32	1.68	<i>H</i> → <i>L</i>	71
S ₂	3.88	0.46	<i>H</i> ² → <i>L</i> ²	31
			<i>H</i> → <i>L</i> + 1	21
			<i>H</i> − 1 → <i>L</i>	20
GLU1				
S ₀	0	-	GS	89
S ₁	3.28	1.63	<i>H</i> → <i>L</i>	72
S ₂	3.90	0.47	<i>H</i> ² → <i>L</i> ²	30
			<i>H</i> → <i>L</i> + 1	20
			<i>H</i> − 1 → <i>L</i>	21
GLU2				
S ₀	0	-	GS	90
S ₁	3.48	1.72	<i>H</i> → <i>L</i>	76
S ₂	4.15	0.30	<i>H</i> ² → <i>L</i> ²	30
			<i>H</i> → <i>L</i> + 1	28
			<i>H</i> − 1 → <i>L</i>	16

3.2.2 Excited-state properties

An important observation refers to the state ordering at the excited state minima. At the ground-state minimum geometry, the lowest excited singlet state has HOMO–LUMO single-excitation character, resembling the electronic structure of the optically bright B_u state in the closely related polyenes. S_2 corresponds to the doubly-excited A_g -like $\text{HOMO}^2\text{--LUMO}^2$ configuration in all considered models. Upon geometry minimization in S_1 , the doubly-excited configuration drops below the singly-excited HOMO–LUMO state. Consequently, the S_1 state is now weakly allowed and S_2 is spectroscopically bright. A corresponding effect occurs when optimizing the geometry in the S_2 state, i.e. it changes from initially $\text{HOMO}^2\text{--LUMO}^2$ to singly-excited HOMO–LUMO. Hence, there must be a crossing between the S_1 and S_2 potential energy surfaces in all models which may facilitate population of the dark $\text{HOMO}^2\text{--LUMO}^2$ state right after photoexcitation. To locate the crossing, we performed a linear interpolation between the S_0 and S_1 equilibrium geometries. Figure 6 shows the interpolated path for the ASP model. All energies are plotted relative to the S_1 energy at the initial ground state geometry. The relative weights of the leading configurations in the S_1 and S_2 states are displayed in percent for each second step. Red color denotes the HOMO–LUMO singly excited state, while blue color corresponds to the $\text{HOMO}^2\text{--LUMO}^2$ double excitation. At step seven (denoted by an arrow), the weight of the HOMO–LUMO configuration becomes slightly larger than the weight of the $\text{HOMO}^2\text{--LUMO}^2$ excitation (25% vs. 27%). However, considering also the contribution of the HOMO-1–LUMO and HOMO–LUMO+1 determinants, the overall character of this state resembles the 2^1A_g state of the related polyenes. The crossing appears almost in the center (step nine) of our interpolated path. At this point, $\text{HOMO}^2\text{--LUMO}^2$ together with the corresponding single excitations contribute to the A_g -like character of the S_1 state, while the S_2 state is now dominated by HOMO–LUMO excitation. The crossover between the S_1 and S_2 states is also confirmed by MS-CASPT2 (see SI). The GLU1 and GLU2 models exhibit the same behaviour and are presented in the SI.

In the optimized S_1 geometries (Figure 5), the bond order inverts relative to the ground state, except for the C=N bond, i.e. former double bonds gain single-bond character and vice versa. The C=N bond elongates slightly but remains shorter than the average single bond in the S_1 state. The bond-length alternation is more pronounced in the C9–C14 fragment. The S_2 geometries show two distinct bond-length patterns. The first fragment, which is closer to the β -ionone ring (between the atoms C5–C9), exhibits a bond-length pattern similar to the ground state, although with slightly longer double bonds and shorter single bonds. In the fragment closer to the positively charged NH^+ -terminus bond-length equalization can be noted. The bond-length pattern of the S_2 geometry resembles the observations already made for the CASSCF-optimized, bond-equalized structure. [24] In both, S_1 and S_2 , geometries, the C13=C14 bond involved in all-*trans* to 13-*cis* isomerization is longer than in the ground state and has more pronounced single bond character.

Considering both, bond length pattern and excitation character, we conclude, that the doubly-excited $\text{HOMO}^2\text{--LUMO}^2$ configuration involves bond-length inversion while the

singly-excited HOMO–LUMO configuration refers to partial bond-length equalization. This observation resembles the results of our previous CASSCF study on C1C2 [24] and the behavior of other polyenes [50–52] and carotenoids [53].

Table 3: Adiabatic and vertical emission energies, oscillator strenghts, electronic configurations and weights for optimized S_1 and S_2 minima

State	$\Delta E_{adia.}$ eV	$\Delta E_{em.}$ eV	f	conf.	weight %
ASP					
S_1	2.53	2.02	0.12	$H^2 \rightarrow L^2$	30
				$H - 1 \rightarrow L$	25
				$H \rightarrow L + 1$	20
S_2	2.92	2.81	2.16	$H \rightarrow L$	72
GLU1					
S_1	2.61	2.09	0.09	$H^2 \rightarrow L^2$	33
				$H - 1 \rightarrow L$	24
				$H \rightarrow L + 1$	21
S_2	2.99	2.82	2.21	$H \rightarrow L$	73
GLU2					
S_1	2.67	2.08	0.07	$H^2 \rightarrow L^2$	33
				$H - 1 \rightarrow L$	25
				$H \rightarrow L + 1$	22
S_2	3.13	2.93	2.22	$H \rightarrow L$	74

Adiabatic and vertical emission energies for the S_1 - and S_2 -optimized structures of all three models can be found in Table 3. No significant differences between the models can be noted for configurations of the optimized states, their oscillator strengths and vertical emission energies. However, their adiabatic energies vary slightly. The most pronounced difference is observed between the GLU2 and ASP models. For GLU2, the S_1 energy is 0.14 eV and S_2 0.21 eV higher than for ASP.

3.2.3 Relaxed S_1 torsion paths

Figure 7 presents relaxed S_1 state torsion scans about the C12-C13=C14-C15 dihedral angle in both, clockwise and counterclockwise directions for the considered models.

In all models, the electronic character of the excited state changes upon torsion. The torsion starts in the S_1 state with HOMO²–LUMO² as leading configuration, while the corresponding S_2 state is dominated by HOMO–LUMO excitation. This picture inverts at the near degeneracy points, where the S_1 state adopts formal HOMO–LUMO and the S_2 state HOMO²–LUMO² character. The orbitals at this point are, however, significantly different from those at the planar FC geometry. After twisting, the formerly delocalized π -orbitals begin to localize, with the HOMO density concentrated near the NH⁺-fragment and the LUMO density on the C7-C13 part of the chromophore (see SI).

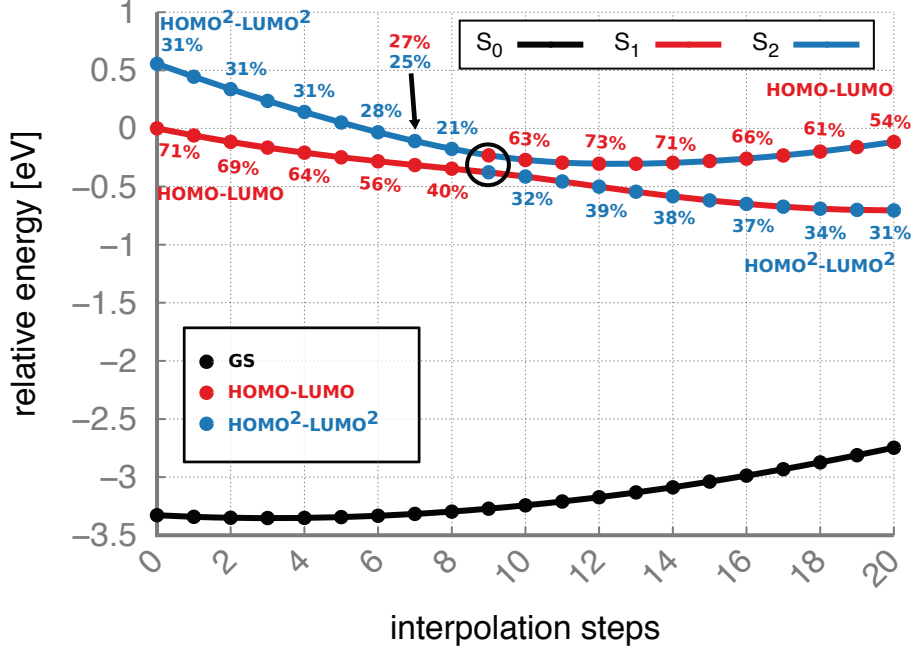


Figure 6: Linear interpolation between optimized S_0 (step 0) and S_1 (step 20) geometries. The leading configurations of the excited states are represented in percent for each second interpolation step. The black arrow denotes a point, where the dominant configuration of S_2 changes from doubly to singly excited HOMO–LUMO configuration. Existence of a conical intersection between these two states is indicated by a black circle. Colored circles on the potential energy surfaces additionally indicate the state character at each interpolation step. The red color refers to HOMO–LUMO single and blue color to HOMO²–LUMO² double excitation. Colors of potential energy surfaces are kept according to their initial dominant configuration: the HOMO–LUMO excitation is represented in red and the HOMO²–LUMO² double excitation in blue colors.

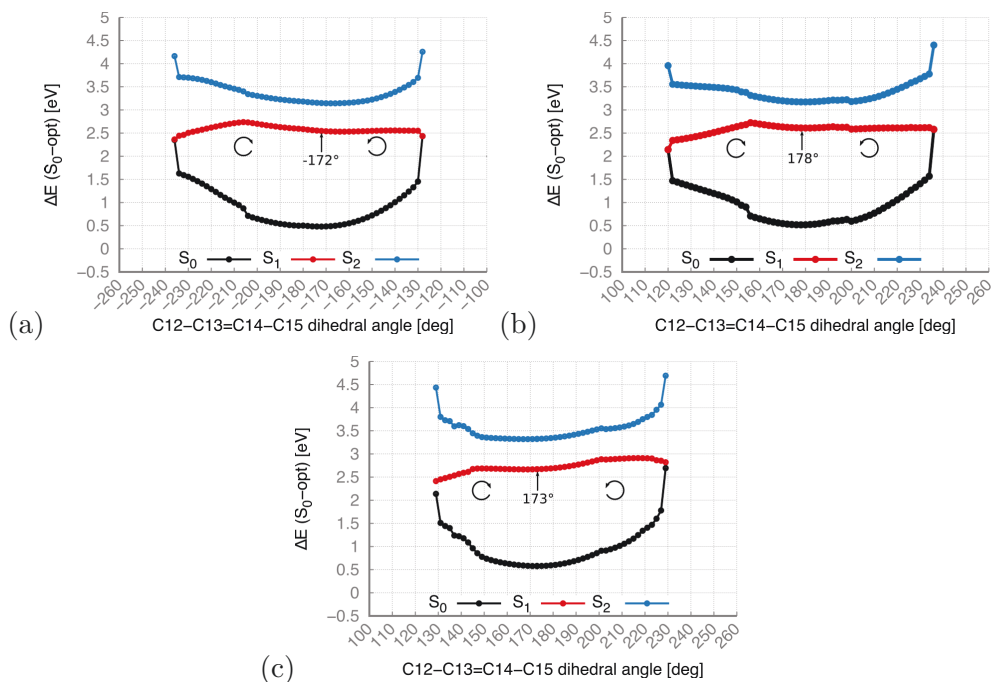


Figure 7: Calculated relaxed torsion scans along C13=C14 reactive coordinate of retinal in (a) ASP, (b) GLU1 and (c) GLU2 models. Straight arrows denote the starting dihedral angle at the C13=C14 bond. Circular arrows correspond to the torsional direction (clockwise or counterclockwise) for the particular torsion path. All energies are plotted relative to the optimized ground-state geometry of the corresponding model.

In the computed models, there is always one torsion direction that exhibits a small barrier (2.65-4.42 kcal/mol (see Table 4), while in the other direction the crossing can be reached almost barrierlessly. Torsion in the ASP and GLU1 models is favoured in counterclockwise directions. In the latter model, the barrier for clockwise direction is ca. 1.8 kcal/mol lower than in the ASP model, which would increase the probability of bi-directional torsion. In the GLU2 model clockwise torsion appears preferred. In contrast to ASP and GLU1, in the GLU2 model the energies of the S_1 and S_0 states are not degenerate at the cross-over (129 and -129 degrees). Root flipping at these points, however, indicates, that the electronic character has changed and an avoided crossing has been passed.

Table 4 summarizes the results obtained from the calculation of the retinal relaxed torsion paths in the S_1 state in the ASP, GLU1 and GLU2 models, compared to values known from the literature.

Table 4: Reaction-coordinate parameters estimated from the retinal S_1 torsion scan in ASP, GLU1 and GLU2 models.

Parameters	ASP	GLU1	GLU2	Lit.
$CI_{clock.}$ [$^\circ$]	124	120	129	110 ^a); 130 ^b)
$CI_{counterclock.}$ [$^\circ$]	-128	-124	-129	-100 ^a); -130 ^b)
barrier _{clock.} [kcal/mol]	4.42	2.65	0.36	3.8 ^a); 1.96 ^b)
barrier _{counterclock.} [kcal/mol]	0.23	0.59	5.55	ca. 0 ^a); 7.42 ^b)
distance from $FC_{clock.}$ [$^\circ$]	42	58	44	-78 ^a); -50 ^b)
distance from $FC_{counterclock.}$ [$^\circ$]	62	58	58	72 ^a); 50 ^b)
excess of energy [kcal/mol]	18	15	19	3.6 ^a); 12 ^b)

^a from ref. [24]

^b from ref. [30]

The bi-directional torsion path for retinal in C1C2 was already investigated by some of us in a previous CASSCF//CASPT2 study (based on an ASP-like model) and also in the CASSCF//CASPT2 study conducted by Hontani et al. Our CASSCF/CASPT2 calculation resulted in a counterclockwise torsion as the preferred direction. Hontani et al. performed their simulations in GLU-like models and obtained higher barriers for the counterclockwise torsion. The torsion path calculations within the current study at first sight confirm the bi-directionality of the C13=C14 torsion. With respect to a preferred torsion direction, the ASP and GLU1 models reproduce the results obtained in our previous CASSCF//CASPT2 study, while the GLU2 model agrees with the conclusions of Hontani et al. Torsion-path parameters listed in Table 4 are in reasonable agreement with the literature. A more detailed description of the computed torsion paths can be found in the SI.

The computed barriers in all models are in an energy range which may easily be overcome by the excess energy of the molecules after photoexcitation. This excess energy is defined as the difference of the excited state energies at the Franck–Condon point and

the excited state equilibrium. The photochemical reaction happens on a timescale that is significantly faster than the thermal equilibration of the chromophore, thus it is plausible to assume that the full excess energy is at disposal for the photoisomerization. For our models, this energy is in a range between 15 and 19 kcal/mol.

3.3 Molecular dynamics calculations

In the current section, we discuss the results of excited-state trajectory calculations on the previously described models. A summary on the trajectory results is given in Table 5.

3.3.1 Reaction time

In Figure 8, we show the averaged electronic state populations and the trajectory occupations resulting from surface hopping, plotted against the simulation time. The S_1 half-life times are also given in these graphs.

For all three models, the S_1 relaxation process can be subdivided into three components: a) rapid initial decay of the S_1 state toward S_2 (ca. 50 % of the starting population), followed by b) quick repopulation of S_1 and finally c) a much slower relaxation to the ground state. a) and b) originate from the identified S_1/S_2 crossing. Each of these processes can be fitted exponentially, the entire S_1 de-excitation process is, however, not exponential. This means, the three components do not contribute to the S_1 decay simultaneously, but rather in a sequential manner. Table 5 summarizes the exponentially fitted parameters for the processes b) and c). The initial rapid S_1 decay and S_2 population increase (a) were not considered, since these processes are very fast (<10 fs), and the exponential fitting procedure will be inaccurate for these timings. The time constants for S_2 de-population and the associated S_1 re-population are in a similar range in all models (between 18 and 42 fs). Minor deviations are due to the fast oscillations also apparent in the population graphs, resulting in lower r^2 parameters. Further details on the fitting procedure are available in the SI. The longest time component, which corresponds to the final S_1 decay (c), has the highest probability to be resolved in the experiment and will thus be compared with available experimental data.

For ASP, we note quick de-population of S_1 within ca. 7 fs toward S_2 for ca 48 % of the molecules. De-population of the S_2 state occurs after ca. 85 fs. Complete S_1 population decay is reached after ca. 2 ps (see SI). Exponential fit of the final S_1 population decay reveals, with a time constant of ca. 284.6 ± 1.6 fs, a reasonable agreement with the experiment. The population decay in the GLU1 model is considerably different. The S_2 state gains a slightly higher population in the initial relaxation period (up to 0.5, see Figure 8) than in the previous model. Within 100 fs the S_1 state is repopulated and decays only slowly. Ca. 88 % of trajectories still reside in this state at the end of the simulation. An exponential fit of the population indicates an S_1 lifetime beyond ca. 6 ps, which is rather comparable to the longer reaction time components in the experiment (2 ps and 11 ps). [30] A subset of those trajectories that hop to the ground state is further analyzed in the SI.

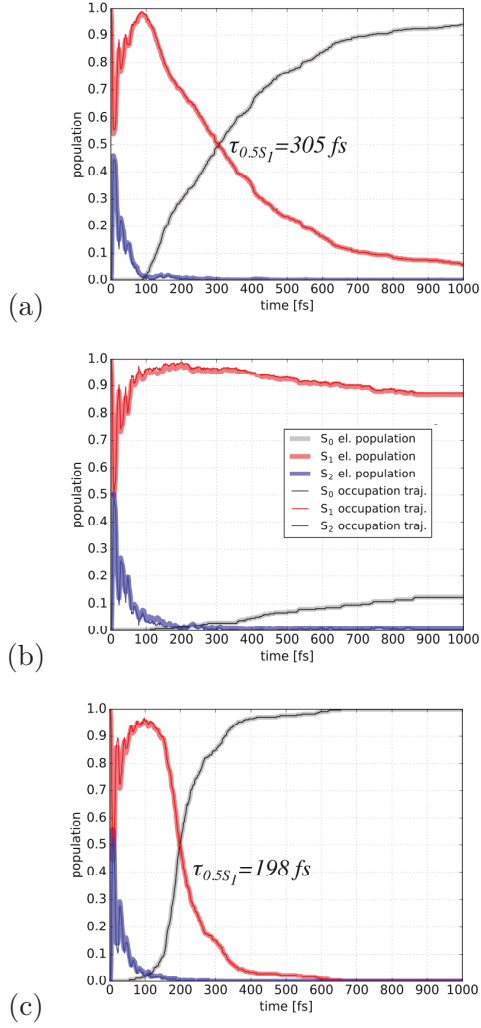


Figure 8: Averaged electronic populations and occupations for (a) ASP. (b) GLU1. (c) GLU2 models. Each plot includes S_1 halflives, except for the GLU1, where only S_2 halflife can be estimated. Fitted S_1 and S_2 times can be found in Table 5

Table 5: Summary of parameters for excited state dynamics of all three models: ASP, GLU1 and GLU2

Parameters	ASP	GLU1	GLU2	Lit.
number of traj.	538	324	346	-
$\tau_{0.5}$, [fs]	305	-	198	-
$\tau_{i(S_1)}$, [fs]	26.3 ± 2.1	28.7 ± 1.6	18.1 ± 2.0	-
$\tau_{d(S_1)}$, [fs]	284.6 ± 1.6	$6.2 \cdot 10^3 \pm 37.1$	95.2 ± 1.0	450, 2000, 11000
$\tau_{d(S_2)}$, [fs]	28.8 ± 0.4	41.7 ± 0.8	29.8 ± 0.6	-
number of react. channels	1	2	1	1
clockwise tors.	-	+	+	+
c. clockwise torsion	+	+	-	+
13- <i>cis</i> QY [%]	30	5	66	30

In contrast, the GLU2 model reacts much faster. At ca. 10 fs, 55% of the trajectories are in the S_2 state. The decay of this state is somewhat slower than in GLU1, and it takes ca. 190 fs to deplete. The S_0 population begins to rise earlier at ca. 60 fs and grows exponentially until ca. 200 fs. With 95.2 ± 0.9 fs the exponential fit does not deliver a value which compares with the experiment.

The population graphs confirm the interaction between the close-lying S_1 and S_2 states. From the interpolated S_1/S_2 paths it is clear that the character of S_1 changes upon relaxation from HOMO–LUMO to HOMO²–LUMO². The nature of the actually occupied states in the trajectories can, however, not easily be deduced from these graphs. This requires an analysis of the transition dipole moments and is provided, along with the simulation of UV/Vis pump–probe spectra, in chapter 3.3.4.

3.3.2 Photoproducts and reaction channels

We could identify up to four different photoproducts in the computed model systems, differing in the configuration of the reactive C13=C14 bond and in the final position of the NH^+ proton. The photoreaction of the ASP model generates all these four photoproducts. In this model, the largest fraction of excited trajectories (68 %) returns to the all-*trans* configuration after passing the S_1/S_0 crossing seam. 32 % undergo all-*trans* to 13-*cis* isomerization. This result perfectly reproduces the experimental quantum yield of 30 %. The 13-*cis* fraction consists of two conformers, which vary with regard to the NH^+ orientation. 29 % represent a conformation with the hydrogen atom in the *anti*-position, i.e. pointing upwards (Figure 9) and only 3 % represent the *syn* conformation with the hydrogen atom pointing downward. The remaining 12 % react to the all-*trans*,15-*syn* photoproduct. This type of isomerization occurs through torsion about the C15=N double bond.

The splitting of the 13-*cis* photoproduct into two conformers results from two different isomerization mechanisms (Figure 9) observed during the simulation. One is known as

the so called bicycle pedal (BP) mechanism described by Warshel. [54, 55] It involves two double bonds (the C13=C14 and the C15=N bonds), one rotating in clockwise, the other in counterclockwise directions (Figure 9). In our computations, this motion leads to the 13-*cis*,15-*syn* conformation with the NH⁺ proton still pointing downwards toward the counterion (see Table 6). The second possibility for isomerization implies hula twist (HT) motion as suggested by Liu and Asato. [56,57] This mechanism involves concerted twisting of the isomerizing double and one single bond (C13=C14 and the N-Lys single bond), leading to a conformation where the hydrogen bond between the NH⁺-terminus and ASP is broken and the NH⁺ proton points upward (13-*cis*,15-*anti*). Different from the original formulation, the mechanism observed here involves two bonds that are not direct neighbors. Since HT in this case is less space-conserving than BP and most of the isomerizing molecules follow this path, we conclude, that there is enough space in the retinal binding pocket to perform the rather extensive HT motion. The hydrogen atom pointing downward (*syn*) would facilitate proton transfer to one of the counterions. The hydrogen in the upward or *anti* position might not be preferable for a direct proton transfer pathway. However, several proton transfer mechanisms were suggested for bacteriorhodopsin with the same NH⁺ orientation. [58, 59] One of them refers to a decrease of the pK_A value for the NH⁺ when moving away from the counterions, which would facilitate proton release. Also in the C1C2 study of Hontani et al., NH⁺ is oriented upwards after geometry optimization.

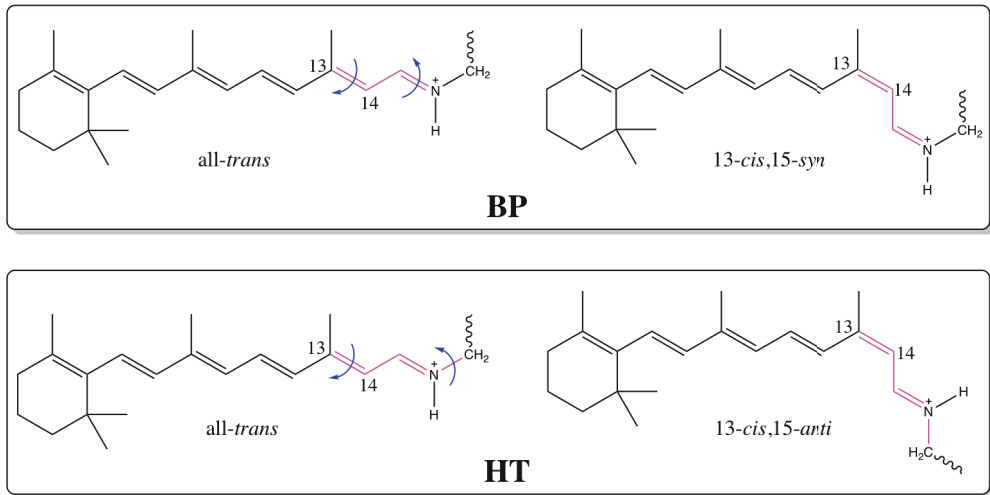


Figure 9: Bicycle pedal (BP) and hula-twist (HT) mechanisms observed in C1C2 surface hopping dynamics.

Figure 10 shows the distribution of photoproducts in the ASP model with respect to simulation time. 90 % of all trajectories hop to the ground state between 100 and 800 fs. Most productive with regard to 13-*cis* formation are hops between 100-200 fs (30 % of all 13-*cis* photoproducts) and between 300-400 fs (23 %). The maximum number of hop events can be observed in the 100-200 fs time frame. The remaining 10 % of the

trajectories hop between 800-2000 fs, while the major part (52 % of all trajectories in this timeframe) ends up in the initial *all-trans* conformation.

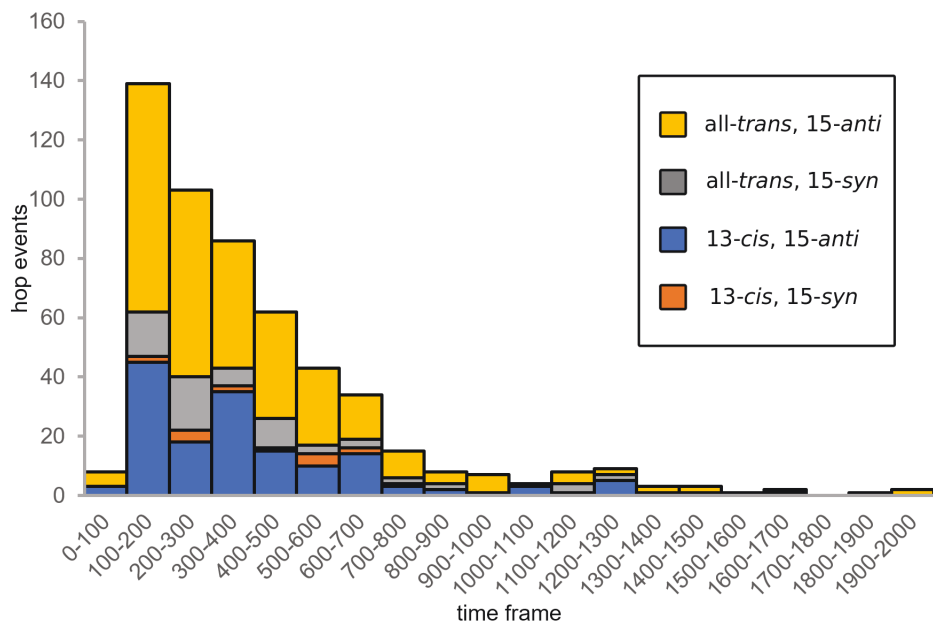


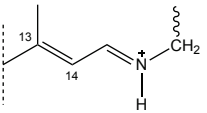
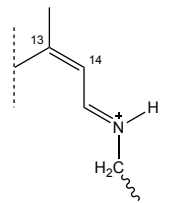
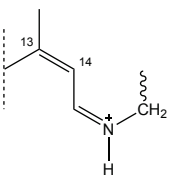
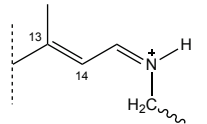
Figure 10: Distribution of hopping times with fractions of the resulting photoproducts in the ASP model.

In Figure 11 the evolution of the C13=C14 dihedral angles is reported for all computed trajectories and models. Green lines denote the trajectory being in the excited state, red lines stand for the ground state. Hopping events are indicated by black stars. Average dihedral angle values denoting productive *trans-cis* and unproductive *trans-trans* isomerization events are given by black arrows.

In the ASP model we observe torsion exclusively in the counterclockwise direction along the C13=C14 bond, with ca. 70% returning to *all-trans* and 30% yielding *13-cis* conformers

The GLU1 trajectory set has two deactivation channels (clockwise and counterclockwise C13=C14 torsion directions), which are only sparsely populated. Trajectories which relax to the ground state result in three photoproducts. Ca. 5 % of the entire set undergo *all-trans* to *13-cis* isomerization by performing HT motion. Ca. 1 % react to the *all-trans*,*15syn* photoproduct. 7 % return to the starting *all-trans* structure. The remaining 87 % stayed in the excited state until the end of the simulation and did not progress toward a conical intersection (CoIn). The preferred direction in the few trajectories that hop is counterclockwise, with mostly *all-trans* as the photoproduct. *13-cis* retinal is, however, generated in both torsion directions. This observation agrees with the results obtained for the S₁ torsion scan, where we noted overall small barriers for both torsions with a slightly smaller activation energy for counterclockwise torsion. This slower model is, however, not entirely unproductive as suggested by Hontani et al. [30]

Table 6: Summary of photoproducts obtained for excited-state dynamics of all three models: ASP, GLU1 and GLU2

Photoproducts	Quantum Yield [%]		
	ASP	GLU1	GLU2
all- <i>trans</i> ,15- <i>anti</i>			
	56	7	34
13- <i>cis</i> , 15- <i>anti</i>			
	29	5	66
13- <i>cis</i> , 15- <i>syn</i>			
	3	-	-
all- <i>trans</i> , 15- <i>syn</i>			
	12	1	-

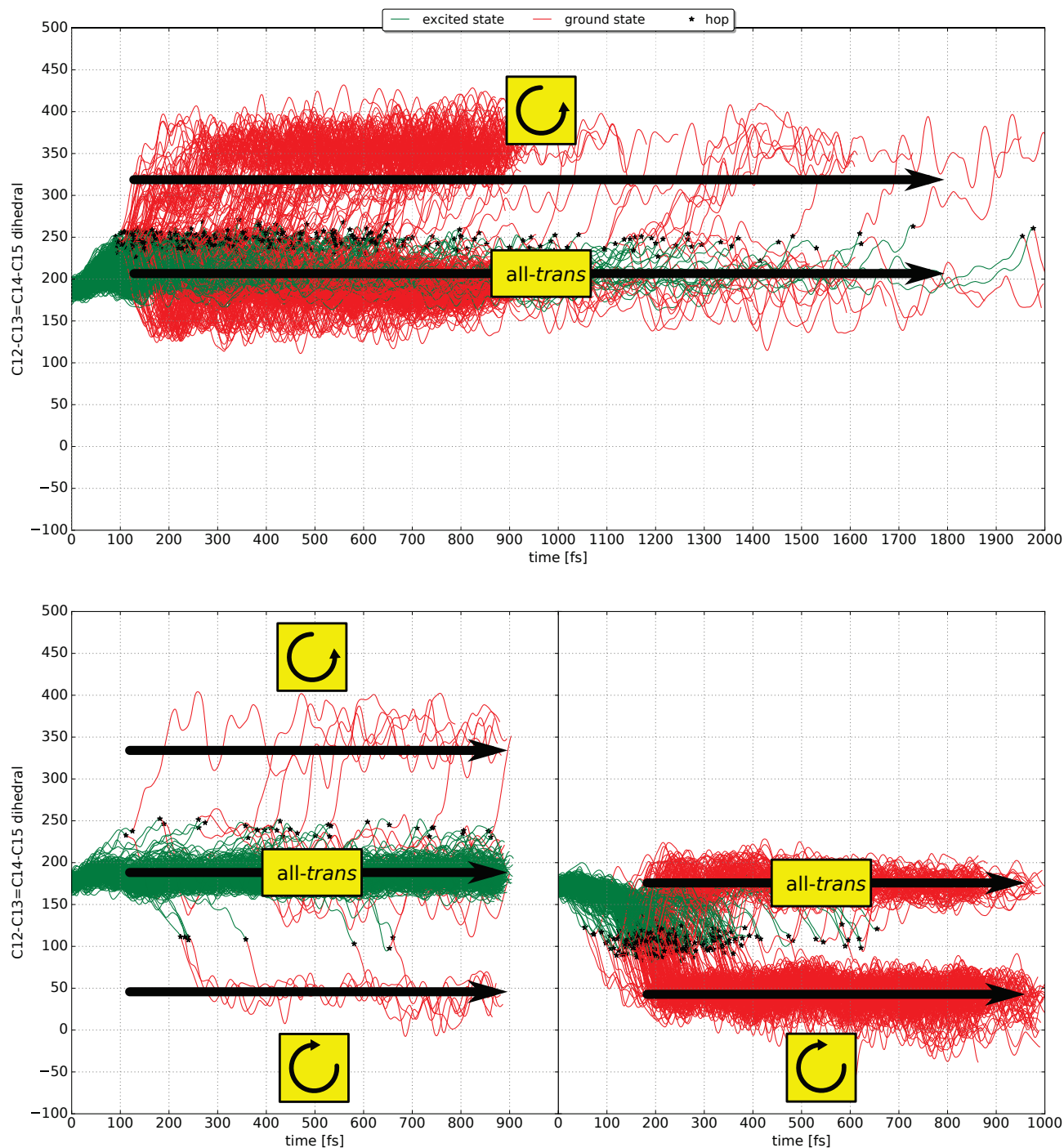


Figure 11: Evolution of C12-C13=C14-C15 dihedral in all computed trajectories for the ASP (upper panel), GLU1 (lower left panel) and GLU2 (lower right panel). Green color denotes excited state, red color denotes relaxation to the ground state. Black stars are assigned to the hopping points. Black arrows highlight the identified reaction routes toward *trans* or *cis* products. The arrows indicate the torsion directions for *cis*-productive routes.

The photoproduct distribution in the GLU2 model significantly differs from the previously described ASP model. With few exceptions in the beginning, the primary torsion direction is clockwise, in agreement with the S_1 torsion scan. In this trajectory set, only two photoproducts were recognized. One fraction of the excited trajectories returns to the all-*trans* structure. The second portion evolves to the 13-*cis* configuration. In contrast to the ASP model, no separation into two 13-*cis* conformers occurs, only 13-*cis* with the hydrogen pointing in the *anti* (upward) position was detected. Consequently, only the HT mechanism is involved in the isomerization reaction of this model.

3.3.3 Influence of the starting geometry

The calculation of excited-state trajectories with surface hopping for the three models ASP, GLU1 and GLU2 yielded three different reaction times. The reaction time computed for the ASP model is in fair agreement with the experimental 450 fs component. The GLU1 model reacts significantly slower. The extrapolated ca. 6 ps reaction time in this set fits better to the long-living species reported by Hontani et al. The excited-state lifetime in the GLU2 model, which had more residues movable in the initial optimization, is unexpectedly short and is not in accordance with any of the reported experimental results. The difference in the reactivity of these models may be related to a) changes in the electrostatic environment of the retinal NH^+ -terminus and b) different structural parameters in the photoactive retinal chromophore. The distances to the counterions in ASP and GLU models are in a similar range in all models (see SI). The position of the positively charged lysine K132, however, largely differs. In ASP, the distance between the NH^+ proton and N of K132 is 5.35 Å, i.e. ca 1.3 Å longer than between the same atoms in the GLU1 model (4.07 Å, see also SI). This closer interaction disturbs the charge partitioning in the chromophore and is apparently connected with the reduced reaction speed — the molecules stay trapped in the $HOMO^2-LUMO^2$ state, and no torsion along the C13=C14 bond occurs. The effect is visible when plotting the charges of the chromophore fragments separated by the isomerizing C13=C14 bond (see SI). In the slow GLU1 model, the positive charge localizes at the iminium fragment in the $HOMO^2-LUMO^2$ doubly excited state, with a larger value than in the S_0 state, thus featuring an inverse CT character compared to the $HOMO-LUMO$ singly excited state. In the GLU2 model, which has an even shorter distance between NH^+ and K132 (3.78 Å), this effect must be largely compensated by other factors. A first indication of the influence of structural parameters can be derived from Figure 12, showing bond and dihedral angles of the optimized model systems.

For bond angles, the parameters in ASP and GLU1 are very similar along the polyene chain, except for the terminus involving the Schiff base nitrogen. The alternation in bonding angles is less pronounced in the GLU2 model. Overall, this leads to a less curved structure than in the two other models (see also chapter 3). The dihedral angles in ASP and GLU1 start to deviate after the C12-C13 bond. The reactive C13=C14 bond is slightly pre-twisted into the photoproduct direction in ASP by ca. 10°, while it is almost planar in GLU1. GLU2 shows major deviations in the C6-C10 region with pronounced twists around the C6-C7 and C8-C9 single bonds. These twists lead to

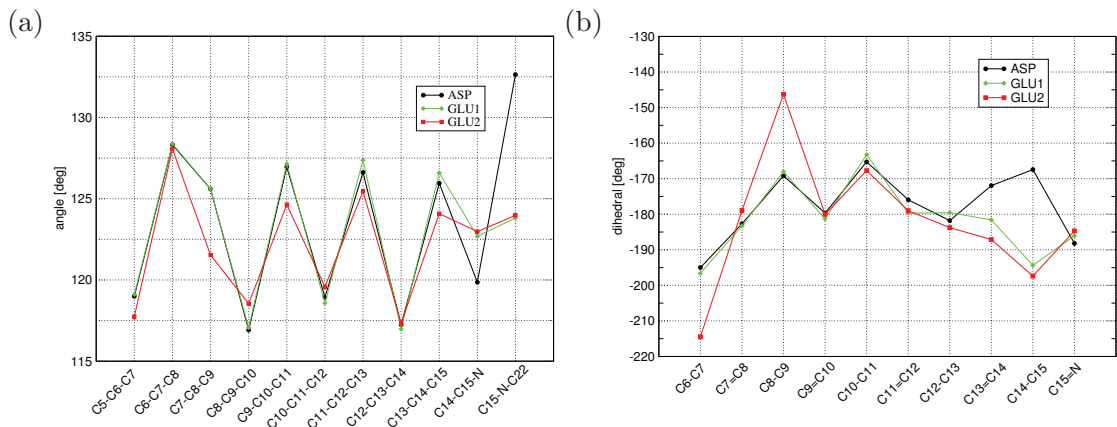


Figure 12: Bond angles and dihedral angles in the optimized ASP, GLU1 and GLU2 models

distortions in the π -system, an effect which also becomes apparent in the slight blue-shift (ca 0.2 eV, see chapter 3.2) in the maximum absorption of this chromophore. OM3/RHF computations with vacuum retinal PSB models reveal, that in fact the torsion about C6-C7 is overestimated with growing chain length. This effect is not solely restricted to application of the OMx methods, but also appears with other semiempirical treatments. It is possibly a result of the small basis used in these methods and the neglect of three-center two-electron terms (see SI).

Furthermore, we also note a slight pre-twist into the photoproduct direction for the C13=C14 bond. These factors consequently speed up the reaction in GLU2. We could verify this hypothesis by a) introducing a similar twist in the GLU1 retinal chromophore and b) strengthening the chromophore of the GLU2 model (see SI). 0K trajectories started from these modified structures in fact reveal faster occurrence of a conical intersection in a), while for b) the conical intersection vanished and the trajectory stayed in the S_1 state until 1 ps, the end of the simulation.

3.3.4 Simulation of pump-probe spectra

The computation of the S_1/S_0 and S_2/S_0 transition dipole moments along the trajectories allows us to follow ground-state bleach (GSB) and photoproduct absorption (PA) signals of the chromophore generated, e.g., in a pump-probe experiment. Transient excited-state absorption data that originate from excitation of the S_1 or S_2 states toward higher excited states, were, however, not included in the computations. Consequently, the generated spectra can not represent all features of a pump-probe experiment. The computed signatures provide valuable assistance in understanding the nature of the spectroscopic states involved right after photoexcitation and along the photoisomerization coordinate.

Figure 13 shows the stimulated emission (SE, red/yellow colors) and PA sections (blue color) in a simulated pump-probe experiment generated from the trajectories of the models and sorted by their reaction speed. The theoretical resolution of our computed

trajectories is ca. 0.1 fs, corresponding to the time step used in the molecular dynamics simulations. These raw data would be incompatible with experimental measurements which include apparatus response times in a range of 50-100 fs. The highest documented experimental resolution for rhodopsin proteins is ca. 20 fs. [29] To enable a comparison with experimental spectra, we use a Gaussian broadening procedure documented in detail elsewhere. [29, 60] At a (still theoretical) resolution of 1 fs, we note common oscillatory features in all models which result from coherent C-C stretching bond vibrations after excitation. These features remain until ca. 350 fs for the fast GLU2 model. In the ASP and GLU1 models they are still present until ca. 500 fs. The oscillatory signatures are barely visible at a simulated resolution of 16 fs. At this resolution, the fast GLU2 set shows two areas for SE: one centered around 450 nm lasting until ca 100 fs. The other, lower energy and wider area around 700 nm appears shortly (ca 30-50 fs) after excitation centered and remains for another 100 fs. At ca 150 fs, the signal quickly shifts toward higher wavelengths and is no longer visible after 200 fs. At this point, a strong PA signal is observed that lasts until the end of the simulation. The separation into two defined areas of SE indicates occupation of a state with low oscillator strength and thus almost no signal intensity between 500 and 550 nm. Comparison with the population graphs in Figure 8 leads to the conclusion, that the first period with intense SE marks the interplay between S_1 and S_2 states, driven by C-C bond stretching vibrations; it takes almost 100 fs for S_2 to decay nonadiabatically to S_1 where the SE signal disappears because of the A_g -like, and hence optically dark HOMO²-LUMO² character. The SE signal rises again while the molecules twist further about the C13=C14 bond which exhibits single-bond character in the S_1 state (compare Figure 5). The twist about the C13=C14 bond causes the ground-state potential energy to rise, thus leading to a red shift of the signal. The rise in the signal intensity may be explained by the localization of the HOMO and LUMO orbitals on either side of the isomerizing double bond, producing a large transition dipole moment (see SI). This effect, named sudden polarization, is a well-known phenomenon occurring in many olefins upon isomerization. [61] The motion toward lower emission energies is very peaked and happens between ca 140 and 190 fs. Then, the molecules are channeled toward the S_1/S_0 conical intersection. Shortly before they reach the intersection seam, the state character changes again to an open shell configuration and finally the molecules decay to the ground state.

The signatures in the ASP ensemble are somewhat different. Here, the areas peaking at 450 and 700 nm are connected between ca. 60 and 80 fs, a clear indication that the molecules reside in a state with sufficient oscillator strength in this time frame. The population graph of this model shows that fewer molecules reach the S_2 state than in the former model and that this population almost fully reverts back to S_1 before the molecules move along the C13=C14 twist coordinate toward lower emission energies. The latter process starts at ca. 60 fs, but the SE signal is much broader (from 60 to ca 150 fs) than in the previous model. Weak traces of PA can be observed starting at ca. 200 fs, together with SE signals around 700-750 nm from molecules with slower C13=C14 torsion. In this ensemble, the molecules apparently stay longer in the vicinity of the crossing seam before they finally decay and yield a strong PA signal starting around 300 fs.

Two almost separated SE regions can also be detected for the slowly reacting GLU1 model. Also in this model, we note quick population transfer to S_2 for nearly half of the molecules, then back transfer to S_1 . At 200 fs, almost all molecules reside back in the S_1 state, now with double excitation character, and oscillate there, visible by the broad and long term SE signal around 650 nm. Very weak traces of PE appear from the small amount of hopping trajectories. These are barely visible around 300 fs.

Some of these observations can also be followed in the intensity plots for the 650 nm wavelength. In the fast GLU2 model we note a moderate build-up of the 650 nm intensity until ca 100 fs. After a period of ca. 50 fs, the SE intensity of this band sharply drops and oscillates around zero. The medium speed ASP set features a sharp growth in SE intensity during the first 50 fs, then a sharper drop until ca. 150 fs followed by slower intensity decay in the next 350 fs. The slow GLU1 ensemble shows an initial raise in intensity but then oscillates around the same value until the end of the simulation.

The simulated spectra indicate that the strong SE signal between 400 and 500 nm is a result of wavepacket oscillations in the HOMO-LUMO state, while the red shifted SE between 600 and 700 nm arises from wavepacket oscillations in the HOMO²-LUMO² state. This is also apparent from Figure 6 and the corresponding Figures in the SI. The minimum of a diabatic parabola connecting the S_1 and S_2 HOMO-LUMO potentials has an energy difference of ca. 3 eV (413 nm, point 10 in Figure 6) to the ground state, and the corresponding minimum of the HOMO²-LUMO² adiabatic parabola results in a gap of ca. 2 eV (620 nm, point 20 in Figure 6).

The features discussed in the simulated spectra, primarily the red shift in SE when the chromophore approaches the CoIn, can also partly be observed in the pump-probe experiments by Scholz et al. [28] These spectra, however, include transient excited-state absorption data that obscure the corresponding SE signal. Traces of these signatures are still visible in the 650-750 nm range in the ChR2 wt spectra, with brighter emission at lower pH values. The low signal intensity in this area indicates that also this chromophore evolves toward the CoIn in a state with lower oscillator strength. This situation changes upon mutation of the E123 counterion to a charged aspartate. In this case, the red shifted SE between 650 and 750 nm becomes visible again, still with lower intensity than the 480 nm SE signal at the start of the experiment. The SE signal in the lower energy region becomes broader along the time axis, when a neutral threonine is used instead. This apparently slows down the isomerization reaction in a similar manner as in our experiment with the GLU1 model.

3.3.5 Influence of hydrogen motion on isomerization

It is well known that out-of-plane motions of hydrogen atoms, attached to the carbon atoms of the isomerizing bond, play a key role in the product-determining step in PSB photoisomerizations. This has been shown for small retinal models and also for retinal in bovine rhodopsin. [42,62–65] In particular, the number of surface approaches prior to the hop event correlates with the product formation during isomerization. This dependency arises from a correlation between the CCCC-torsions, defining the reaction coordinate, and the hydrogen out-of-plane motions at the reactive double bond described through

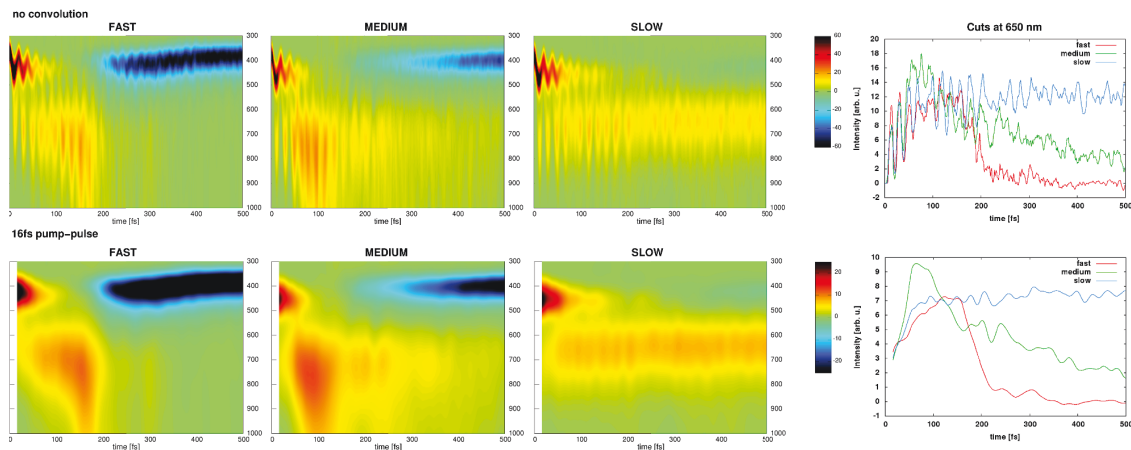


Figure 13: Simulated pump probe spectra for GLU2 (fast), ASP (medium) and GLU1 (slow) models at a resolution of 1 and 16 fs. The right graph displays horizontal cuts along the time axis at 650 nm.

HCCH-torsion. A clear relationship between surface approaches and product formation was shown by Klaffki et al. [66] for the *cis*-PSB3 model. Positive HCCH-torsion in this study was responsible for generation of the *trans* photoproduct. After the first S_1/S_0 approach the HCCH-torsions mostly have a negative phase, which preferably yielded starting material when trajectories hop at this point. After three approaches both, CCCC- and HCCH-torsion are dephased, generating equal probabilities to obtain *cis* or *trans* photoproducts.

We analyzed this aspect in our models by computing the angular velocities of the two dihedrals C12-C13=C14-C15 (CCCC-torsion) and C17-C13=C14-H14 (out-of-plane CCCH-torsion) involved in the photoisomerization (see Figure 14). In Figure 15 we plot the velocities of the CCCC dihedrals against the corresponding CCCH dihedrals at the hopping step. Separate plots distinguish between hops occurring at first, second and n -th ($n > 2$) approaches of the S_0 and S_1 potential energy surfaces.

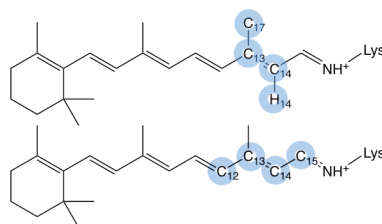


Figure 14: Selected dihedrals for analysis of angular velocities.

For the first approach of the surfaces, the product distribution for the ASP and GLU2 models is dominated by the *cis*-photoproduct, i.e. the phase correlation between CCCH- and CCCC- torsion at this point is still preserved, even when the trajectories spend a long

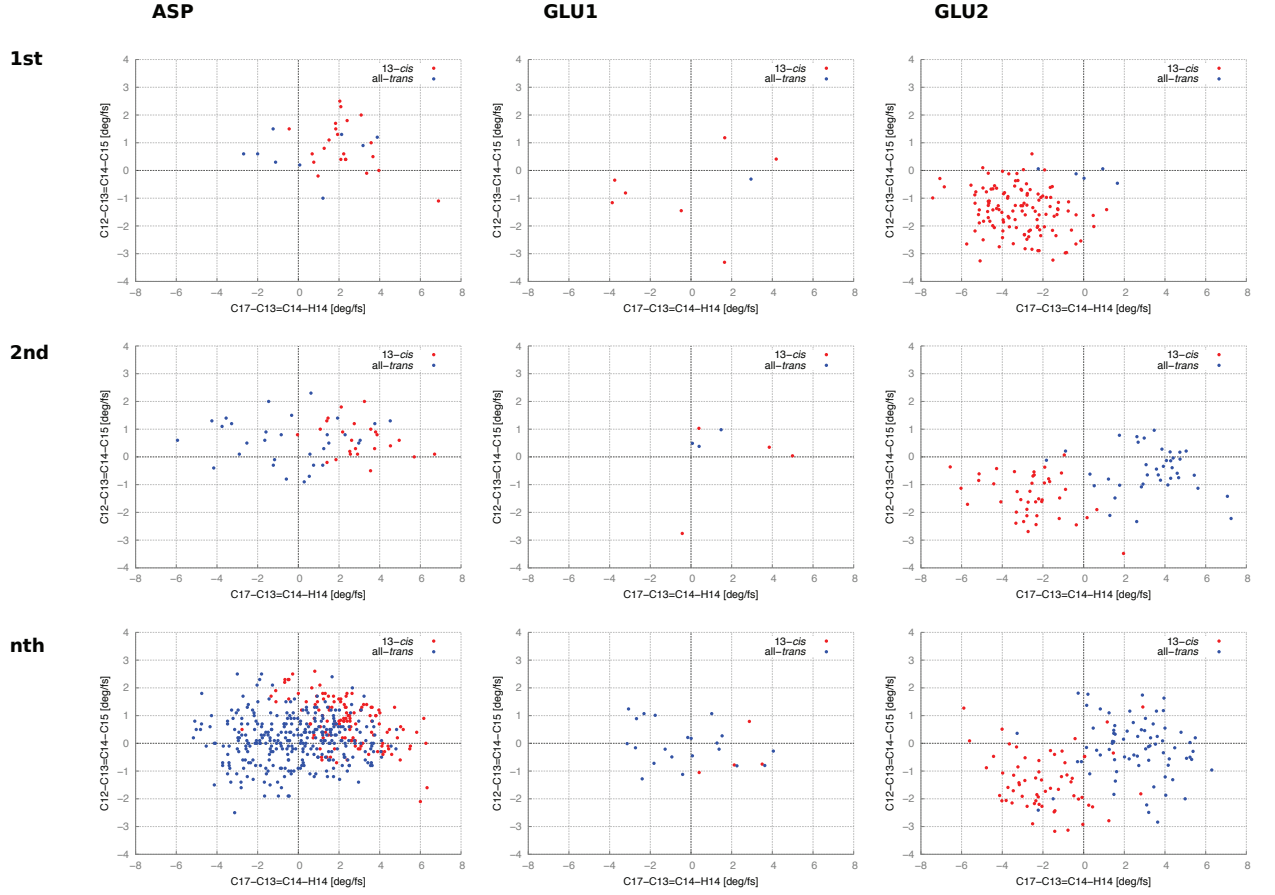


Figure 15: CCCC and CCCH angular velocities at hopping point for ASP, GLU1 and GLU2 models. Red spots correspond to successful all-*trans* to 13-*cis* events, blue spots mark back reaction to the all-*trans* configuration. Product distributions are plotted according to the number of PESs approaches before passing through the conical intersection.

time in the excited state. The ASP model shows mostly positive in-phase CCCC- and CCCH-torsion and *cis* products. In the GLU2 model, the product population almost exclusively contains 13-*cis* with negative in-phase torsion (lower left section in GLU2 graph of Figure 15). This set also has the largest number of decays at the first approach. Due to the small number of hopping events, no clear trend can be derived for the GLU1 model.

In the second S_1/S_0 approach, the ASP and GLU2 models generate an almost equal amount of *cis* and *trans* photoproducts. While the phase rule for product generation still holds for GLU2, we note some deviations in ASP. These become even larger at the n -th approach, where most of the ASP trajectories hop to the ground state. The large number of hops at $n > 2$ shows that these trajectories in fact stay longer in the vicinity of the crossing seam before finally hopping, an effect that was already observed in the simulated pump probe spectra in Figure 13.

Apparently, after a period of motion close to the hopping seam, other factors than hydrogen torsion start to influence the photoproduct generation in these longer running trajectories. El-Tahawy et al., e.g., observe a certain influence of the bond-length alternation coordinate and the charge repartitioning when the system changes from the excited charge-transfer state to the covalent ground state. [67] The much faster GLU2 conserves the trend even after several approaches, with mostly *cis* products for negative CCCH-torsion and all-*trans* in the other direction.

4 Conclusions

We performed a detailed QM/MM dynamical study of the retinal photoreaction in the chimeric channelrhodopsin C1C2. In total, 1208 trajectories were computed with three different models, constituting three different initial geometries of the retinal chromophore. The difference in the structures lie in the orientation of the NH^+ -terminus and the shape of the polyenic chain. All three models share similar electronic configurations and bond length patterns in optimized ground and excited state geometries. At the ground-state geometry, the S_1 state can be characterized by the B_u -like HOMO-LUMO single excitation while the S_2 state has pronounced A_g -like doubly excited HOMO²-LUMO² character. Right after excitation to the bright S_1 state, bond-length equalization sets in which drives the system toward a crossing with the S_2 state. As a result, the state ordering changes, and the system can cross over to the spectroscopically weakly allowed HOMO²-LUMO² state. Geometry relaxation in this state leads to bond inversion. Such behaviour, which is more typical for linear polyenes, was already observed in our earlier CASSCF//CASPT2 study. [24]

To clarify whether there is a preferential torsion direction of retinal in C1C2, torsion scans along the C13=C14 reaction coordinate were carried out. In each model, one torsion path is almost barrierless, and the other one exhibits a small barrier. According to these static calculations, ASP and GLU1 are expected to favor counterclockwise torsion, while the GLU2 model would rather promote clockwise reaction. This behaviour agrees with our previous CASSCF//CASPT2 and with the results obtained by Hontani

et al. The computed barrier heights are, however, small compared to the excess energy of the chromophore after photoexcitation. Hence, both torsion directions and thus bi-directional deactivation appear plausible. In the model of Hontani et al, both torsion directions are populated, but only one specific conical intersection yields successful isomerization events. The other channel leads back to all-*trans*. Our dynamics computations yield a different picture. We find fast and exclusive counterclockwise torsion in ASP and exclusive clockwise torsion in GLU2 with both productive and unproductive isomerization events. The reaction in GLU1 is bi-directional, also showing successful and unsuccessful isomerizations.

The simulation of pump-probe spectra reveals two separate regions of stimulated emission (SE) in all models. The first, shorter-wavelength region is associated with oscillations between S_1 and S_2 states caused by C-C stretching vibrations. The high SE signal intensity indicates that the HOMO-LUMO excitation character is largely preserved at this stage. Geometry relaxation in the excited-state potential eventually leads to a change of the electronic structure toward the HOMO²-LUMO² doubly excited configuration. This state has low oscillator strength, notable by the gap in SE intensity between 500 and 550 nm. As soon as the molecules start to twist about the C13=C14 bond, which rather is a single bond in the excited state, the SE regains intensity due to localization of the π -orbitals. The intensity is, however, lower than the initial SE signal. At the same time, the signal shifts toward longer wavelength as the torsion about this bond is unfavourable in the electronic ground state. In the fast GLU2 model, we note collective motion toward the S_1/S_0 crossing seam in a rather short time interval between 120 and 190 fs. The trajectories in the slower ASP model stay longer in the vicinity of the crossing seam before they finally hop to the ground state. These findings are also verified by the analysis of the hydrogen torsions at the surface hopping point. The majority of molecules hop already at the first approach of the S_1/S_0 surfaces in GLU2, while ASP has most hopping events after at least two close S_1/S_0 approaches. The almost unreactive GLU1 model stays trapped in a local minimum with formal HOMO²-LUMO² double excitation character and oscillates there yielding moderate SE intensity.

The excited-state lifetime of the ASP model is in accordance with the experimental fast decay component. The same is true for its quantum yield. With an estimated lifetime of ca. 6 ps, the GLU1 model would rather match the slower components in the experiment. The reduction in reaction speed is caused by geometric and electrostatic effects: The chromophore is more planarized about the C13=C14 bond, and the positively charged K132 residue moves closer toward the Schiff base. As a consequence, the charge distribution in the chromophore is disturbed. No torsion sets in, and the molecules are trapped in the S_1 doubly excited minimum, from where they can escape only slowly, in accord with the slow components observed in the C1C2 photodynamics. Both paths exhibit productive and unproductive isomerization events. Thus, the reaction kinetics in C1C2 are largely modulated by the flexible binding pocket of the retinal chromophore. They are slowed down, when the negative charge of E162 and D292 is counterbalanced by the positive charge of K132. A similar observation was already made in the work of Scholz et al. [28] The reaction of the GLU2 model appears to be too fast and to occur with a too high 13-*cis* quantum yield. This model, could, however, exemplify the rela-

tion between the retinal geometry and the speed of the reaction. In this conformation of the protein pocket, the retinal chromophore is pre-twisted about the C6-C7, the C8-C9 and the C13=C14 bonds whereas the slower GLU1 model exhibits a planar chromophore backbone. These twists speed up the photoreaction and may indicate a route for efficient modulation of retinal dynamics through substitution with residues that can induce such twists.

Four different photoproducts were detected throughout our calculations. The simulations yielded all-*trans* and 13-*cis* retinal in both, *syn* and *anti* forms. The same isomers were also reported in the literature. [59, 68] ASP yields all mentioned photoproducts, the GLU1 model all except 13-*cis*,15-*syn*, and GLU2 only the *anti* forms of 13-*cis* and all-*trans* retinal. The 13-*cis* photoproducts are generated by two different mechanisms: bicycle pedal (BP) and hula-twist (HT). HT places the NH^+ proton to the *anti* position, i.e. away from the counterion. BP motion retains the hydrogen bond between the counterion and retinal, delivering 13-*cis*,15-*syn*. The HT mechanism appears preferred, indicating that there is enough space in the protein pocket. The BP mechanism may furthermore be hindered by the double bond nature of the C15=N bond also in the excited state.

The fast reaction of the GLU2 model apparently leads to high photoproduct selectivity with only HT reactions, suppressing the corresponding *syn*- channels, although with no correlation to the experiment. This model had the highest 13-*cis* quantum yield, an effect also connected with the large amount of trajectories hopping at the first approach of the S_1/S_0 surfaces. At this point, the CCH-torsion is mostly in-phase with the reaction coordinate.

To capture the described effects, molecular dynamics simulations are essential, as static calculations like the computation of torsion paths can only give a one-dimensional picture. Such methods leave out dynamic effects and disregard the impact of the ground-state vibrational distribution in a molecular ensemble. The deficits of a static picture became apparent in the computed GLU1 torsion path and the actual dynamics of this model. The path itself gives no indication on its rather weak reactivity. The surface hopping photodynamics in this study also revealed that only one model for the description of the flexible C1C2 binding pocket is not enough to achieve a complete understanding of its photoreactivity. After all, we found two configurations matching experimental time constants, the GLU1 model representing slow kinetics, and the ASP model describing the faster component. In our simulations, the K132 residue relocated its position and moved closer toward the N-terminus of retinal when the protein pocket was allowed to relax. This situation apparently favors hydrogen bonding toward the GLU counterion. Further configurations with a similar reactivity as the ASP or GLU1 models but oriented toward the corresponding other counterion cannot be excluded and may be located with a combination of geometry sampling and optimization techniques.

The complexity of the photoreaction requires methods that deliver accurate results within reasonable time. The applied OMx/MRCI strategy performed very well in this respect. Care must be taken with regard to the different description of geometry parameters, as observed for the torsion about the C6-C7 bond in retinal. Our study may encourage the application of the OMx/MRCI method for the investigation of other bio-

logically relevant systems.

5 Acknowledgments

The authors thank Prof. Peter Gilch (Universität Düsseldorf) for valuable discussions on the topic. I. D. is grateful for financial support through “Strategischer Forschungsfonds” of the Heinrich-Heine-University Düsseldorf, project no. F 2013 – 442-9. Further financial support by Deutsche Forschungsgemeinschaft (DFG) through funds (INST 208/704-1 FUGG) to purchase the hybrid computer cluster used in this study is gratefully acknowledged. Computational support and infrastructure was also provided by the “Centre for Information and Media Technology” (ZIM) at the University of Düsseldorf (Germany).

References

- [1] M. Beckmann, P. Hegemann, *Biochemistry*, 1991, **30**, 3692–3697.
- [2] G. Nagel, D. Ollig, M. Fuhrmann, S. Kateriya, A. M. M. Musti, E. Bamberg, P. Hegemann, *Science*, 2002, **296**, 2395–2398.
- [3] S. Ehlenbeck, D. Gradmann, F.-J. Braun, P. Hegemann, *Biophys. J.*, 2002, **82**, 740–751.
- [4] E. M. Holland, H. Harz, R. Uhl, P. Hegemann, *Biophys. J.*, 1997, **73**, 1395–1401.
- [5] O. A. Sineshchekov, F. F. Litvin, L. Keszthelyi, *Biophys. J.*, 1990, **57**, 33–39.
- [6] O. A. Sineshchekov, E. G. Govorunova, *Trends Plant Sci.*, 1999, **4**, 58–63.
- [7] O. P. Ernst, D. T. Lodowski, M. Elstner, P. Hegemann, L. S. Brown, H. Kandori, *Chem. Rev.*, 2014, **114**, 126–163.
- [8] T. Takahashi, K. Yoshihara, M. Watanabe, M. Kubota, R. Johnson, F. Derguini, K. Nakanishi, *Biochem. Biophys. Res. Commun.*, 1991, **178**, 1273–1279.
- [9] B. J. Litman, D. C. Mitchell, *Biomembranes*, 1996, **2**, 1–32.
- [10] R. R. Birge, *Annu. Rev. Phys. Chem.*, 1990, **41**, 683–733.
- [11] H. Kandori, Y. Shichida, T. Yoshizawa, *Biochemistry*, 2001, **66**, 1197–1209.
- [12] G. Nagel, T. Szellas, W. Huhn, S. Kateriya, N. Adeishvili, P. Berthold, D. Ollig, P. Hegemann, E. Bamberg, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 13940–13945.
- [13] P. Hegemann, G. Nagel, *EMBO Mol. Med.*, 2013, **5**, 173–176.
- [14] K. Deisseroth, G. Feng, A. K. Majewska, G. Miesenböck, A. Ting, M. J. Schnitzer, *J. Neurosci.*, 2006, **26**, 10380–10386.

- [15] G. P. Dugué, W. Akemann, T. Knöpfel, *Prog. Brain Res.*, 2012, **196**, 1–28.
- [16] S. B. G. Correia, J. F. Ribeiro, A. F. Silva, R. M. Costa, J. H., *J. Neural Eng.*, 2017, **14**, 41001.
- [17] N. C. Klapoetke, Y. Murata, S. S. Kim, S. R. Pulver, A. Birdsey-Benson, Y. K. Cho, T. K. Morimoto, A. S. Chuong, E. J. Carpenter, Z. Tian, J. Wang, J. Wang, Y. Xie, Z. Yan, Y. Zhang, B. Y. Chow, B. Surek, M. Melkonian, V. Jayaraman, M. Constantine-Paton, G. K.-S. Wong, E. S. Boyden, *Nat. Methods*, 2014, **11**, 338–346.
- [18] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, K. Deisseroth, *Nat. Neurosci.*, 2005, **8**, 1263–1268.
- [19] J. Y. Lin, *Exp. Physiol.*, 2010, **96**, 19–25.
- [20] E. G. Govorunova, O. A. Sineshchekov, R. Janz, X. Liu, J. L. Spudich, *Science*, 2015, **349**, 647–650.
- [21] A. Bi, J. Cui, Y.-P. Ma, E. Olshevskaya, M. Pu, A. M. Dizhoor, Z.-H. Pan, *Neuron.*, 2006, **50**, 23–33.
- [22] H. E. Kato, F. Zhang, O. Yizhar, C. Ramakrishnan, T. Nishizawa, K. Hirata, J. Ito, Y. Aita, T. Tsukazaki, S. Hayashi, P. Hegemann, A. D. Maturana, R. Ishitani, K. Deisseroth, O. Nureki, *Nature*, 2012, **482**, 369–374.
- [23] O. Volkov, K. Kovalev, V. Polovinkin, V. Borshchevskiy, C. Bamann, R. Astashkin, E. Marin, A. Popov, T. Balandin, D. Willbold, G. Büldt, E. Bamberg, V. Gordeliy, *Science*, 2017, **358**, 8862.
- [24] I. Dokukina, O. Weingart, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25142–25150.
- [25] S. Ito, H. E. Kato, R. Taniguchi, T. Iwata, O. Nureki, H. Kandori, *J. Am. Chem. Soc.*, 2014, **136**, 3475–3482.
- [26] M. R. VanGordon, G. Gyawali, S. W. Rick, S. B. Rempe, *Biophys. J.*, 2017, **112**, 943–952.
- [27] S. Adam, A.-N. Bondar, *PLoS One*, 2018, **13**, 1–28.
- [28] F. Scholz, E. Bamberg, C. Bamann, J. Wachtveitl, *Biophys. J.*, 2012, **102**, 2649–2657.
- [29] D. Polli, P. Altoè, O. Weingart, K. M. Spillane, C. Manzoni, D. Brida, G. Tomasello, G. Orlandi, P. Kukura, R. A. Mathies, M. Garavelli, G. Cerullo, *Nature*, 2010, **467**, 440–443.
- [30] Y. Hontani, M. Marazzi, K. Stehfest, T. Mathes, I. H. Stokkum, M. Elstner, P. Hegemann, J. T. Kennis, *Sci. Rep.*, 2017, **7**, 7217.

- [31] W. Weber, W. Thiel, *Theor. Chim. Acta*, 2000, **103**, 495–506.
- [32] P. O. Dral, X. Wu, L. Spörkel, A. Kosłowski, W. Weber, R. Steiger, M. Scholten, W. Thiel, *J. Chem. Theory Comput.*, 2016, **12**, 1082–1096.
- [33] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, Walker, J. Wang, R. M. Wolf, X. Wu, P. A. Kollman, *Amber 14 OR - University of California, San Francisco*, 2014.
- [34] P. Altoè, M. Stenta, A. Bottoni, M. Garavelli, *Theor. Chem. Acc.*, 2007, **118**, 219–240.
- [35] O. Weingart, A. Nenov, P. Altoè, I. Rivalta, J. Segarra-Martí, I. Dokukina, M. Garavelli, *J. Mol. Model.*, 2018, 271.
- [36] E. Fabiano, T. Keal, W. Thiel, *Chem. Phys.*, 2008, **349**, 334–347.
- [37] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.
- [38] I. Dokukina, C. M. Marian, O. Weingart, *Photochem. Photobiol.*, 2017, **93**, 1345–1355.
- [39] Y. Guo, F. E. Beyle, B. M. Bold, H. C. Watanabe, A. Kosłowski, W. Thiel, P. Hegemann, M. Marazzi, M. Elstner, *Chem. Sci.*, 2016, **7**, 3879–3891.
- [40] C. S. Sloane, W. L. Hase, *J. Chem. Phys.*, 1977, **66**, 1523–1533.
- [41] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian09 Revision E.01*, Gaussian Inc. Wallingford CT 2009.
- [42] O. Weingart, P. Altoè, M. Stenta, A. Bottoni, G. Orlandi, M. Garavelli, *Phys. Chem. Chem. Phys.*, 2011, **13**, 3645–3648.

- [43] D. Polli, I. Rivalta, A. Nenov, O. Weingart, M. Garavelli, G. Cerullo, *Photochem. Photobiol. Sci.*, 2015, **14**, 213–228.
- [44] O. Weingart, *Curr. Org. Chem.*, 2017, **21**, 586–601.
- [45] L. Spörkel, W. Thiel, *J. Chem. Phys.*, 2016, **144**, 194108.
- [46] J. C. Tully, R. K. Preston, *J. Chem. Phys.*, 1971, **55**, 562–572.
- [47] J. C. Tully, *J. Chem. Phys.*, 1990, **93**, 1061–1071.
- [48] C. Zhu, A. W. Jasper, D. G. Truhlar, *J. Chem. Theory Comput.*, 2005, **1**, 527–540.
- [49] G. Granucci, M. Persico, *J. Chem. Phys.*, 2007, **126**, 134114.
- [50] F. Blomgren, S. Larsson, *Theor. Chem. Acc.*, 2003, **110**, 165–169.
- [51] C. M. Marian, N. Gilka, *J. Chem. Theory Comput.*, 2008, **4**, 1501–1515.
- [52] Y. Kurashige, H. Nakano, Y. Nakao, K. Hirao, *Chem. Phys. Lett.*, 2004, **400**, 425–429.
- [53] T. Polívka, V. Sundström, *Chem. Rev.*, 2004, **104**, 2021–2072.
- [54] A. Warshel, *Nature*, 1976, **260**, 679.
- [55] A. Warshel, N. Barboy, *J. Am. Chem. Soc.*, 1982, **104**, 1469–1476.
- [56] R. Liu, A. E. Asato, *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 259–263.
- [57] R. S. Liu, *Acc. Chem. Res.*, 2001, **34**, 555–562.
- [58] S. Subramaniam, R. Henderson, *Nature*, 2000, **406**, 653.
- [59] H. Luecke, *Biochim. Biophys. Acta, Bioenerg.*, 2000, **1460**, 133–156.
- [60] D. Polli, O. Weingart, D. Brida, E. Poli, M. Maiuri, K. M. Spillane, A. Bottoni, P. Kukura, R. A. Mathies, G. Cerullo, M. Garavelli, *Angew. Chem. Int. Ed.*, 2014, **53**, 2504–2507.
- [61] V. Bonačić-Koutecký, P. Bruckmann, P. Hiberty, J. Koutecký, C. Leforestier, L. Salem, *Angew. Chem. Int. Ed.*, 1975, **14**, 575–576.
- [62] O. Weingart, I. Schapiro, V. Buss, *J. Mol. Model.*, 2006, **12**, 713–721.
- [63] A. Migani, M. A. Robb, M. Olivucci, *J. Am. Chem. Soc.*, 2003, **125**, 2804–2808.
- [64] O. Weingart, *Chem. Phys.*, 2008, **349**, 348–355.
- [65] I. Schapiro, M. N. Ryazantsev, L. M. Frutos, N. Ferré, R. Lindh, M. Olivucci, *J. Am. Chem. Soc.*, 2011, **133**, 3354–3364.

- [66] N. Klaffki, O. Weingart, M. Garavelli, E. Spohr, *Phys. Chem. Chem. Phys.*, 2012, **14**, 14299–14305.
- [67] M. M. El-Tahawy, A. Nenov, O. Weingart, M. Olivucci, M. Garavelli, *J. Phys. Chem. Lett.*, 2018, **9**, 3315–3322.
- [68] S. Subramaniam, R. Henderson, *Nature*, 2000, **406**, 653–657.

Supporting Information

QM/MM photodynamics of Retinal in the Channelrhodopsin Chimera C1C2 with OM3/MRCI

Irina Dokukina¹, Artur Nenov², Marco Garavelli², Christel M. Marian¹, and Oliver Weingart^{*1}

¹Institut für Theoretische Chemie und Computerchemie, Heinrich-Heine-Universität
Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

²Dipartimento di Chimica Industriale "Toso Montanari", Università degli Studi di Bologna,
Viale del Risorgimento, 4, 40136 Bologna, Italia

1 Static calculations

1.1 Interpolation paths

Figure 1 shows linearly interpolated paths between the S_0 and S_1 geometries in the GLU1 and GLU2 models. The interpolation results show trends similar to the ASP model. The initial configuration of the S_1 state shows HOMO–LUMO singly excited character. After passing the intersection region, the S_1 state becomes dominated by the HOMO²–LUMO² double excitation, while the S_2 state has now HOMO–LUMO character.

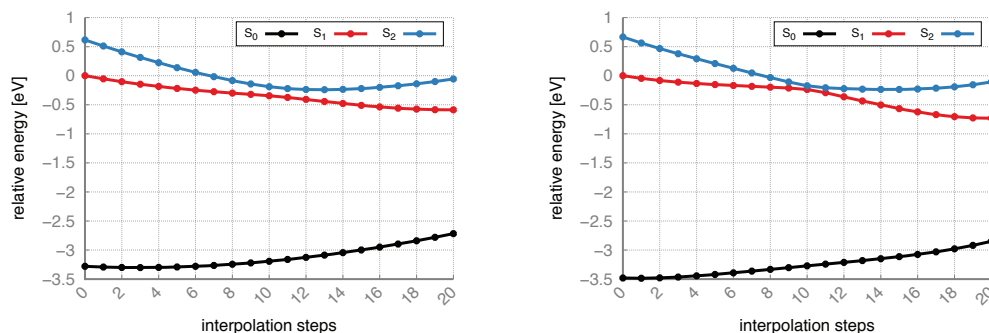


Figure 1: Linear interpolation paths in GLU1 (left) and GLU2 (right) models.

Figure 2 shows MS-CASPT2 energies computed on top of the OM2/MRCI computed interpolation path in ASP. The state crossing appears also here, but somewhat later along the path.

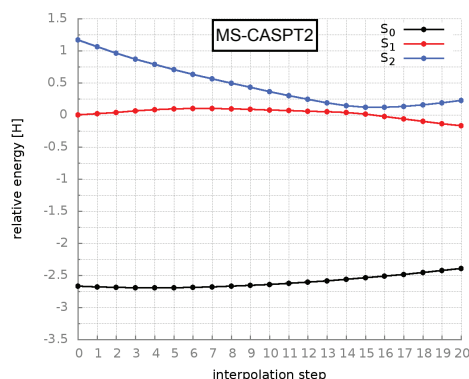


Figure 2: MS-CASPT2 computed energies along OM2/MRCI interpolated S_0 and S_1 geometries.

1.2 Torsion Scans

ASP Figure 7 in the main article shows the calculated torsion path in the S_1 state. At the starting geometry, the value of the C13=C14 dihedral angle is ca. -172° . Torsion in both directions results in a crossing between the first excited singlet state and the ground state. In negative direction, the crossing seam is reached at -128° almost barrierlessly (barrier of 0.23 kcal/mol (0.01 eV)). To approach the second CI at 124° in the clockwise direction, the system has to overcome a barrier of 4.42 kcal/mol (0.2 eV). This is generally possible with the 18 kcal/mol excess energy resulting from relaxation of the FC-geometry. It also delivers similar trends as the CASSCF//CASPT2 torsion path calculated by Hontani et al., however the assignment of the barriers differs. In their work, the authors propose the counterclockwise path having a higher barrier (12 kcal/mol) than the clockwise one (1.96 kcal/mol). In the FC-region, the leading configuration of the S_1 state is HOMO–LUMO (71%). As already mentioned above, the configuration changes to doubly excited upon relaxation. The configuration of the first optimized S_1 geometry at -172° is dominated by the HOMO²–LUMO² determinant. Figure 3 shows the swap of HOMO–LUMO and HOMO²–LUMO² characters in the S_1 state during torsion.

For both directions, the amount of double excitation overall decreases while for the single excitation it increases. At -148° for the counterclockwise path and at -206° for the clockwise path, single and double excitations are equally represented. After passing this point, the HOMO–LUMO configuration dominates the S_1 state. At both CoIns the double excitation vanishes completely.

GLU1 Similar to the ASP and GLU2 model, the torsion about the C13=C14 bond leads to two CoIns in the GLU1 model. The barrier to reach the CI in the clockwise direction (ca. 2.6 kcal/mol) is higher than for the counterclockwise torsion (0.59 kcal/mol),

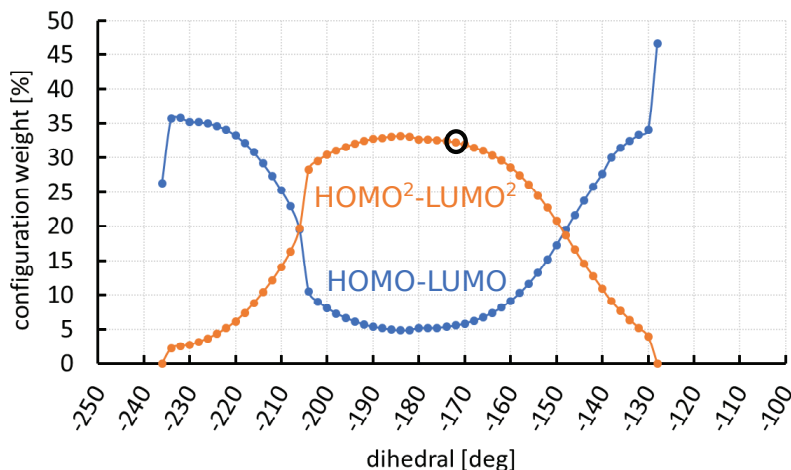


Figure 3: Change of the S_1 state character during clockwise and counterclockwise torsion in the ASP model. The black circle denotes the starting point of the torsion.

similar to the ASP model. However, both barriers are lower than those computed for ASP. The excess energy after excitation is ca. 15 kcal/mol and would be sufficient to overcome the barriers during dynamics simulation. Also Hontani et al. report two CoIns in equal distance (50° from the FC-region). [1]

GLU2 The torsion path for the GLU2 model starts with a C13=C14 dihedral angle of $+173^\circ$ and, similar to the ASP model, results in two possible torsion directions. The clockwise direction is almost barrierless (0.36 kcal/mol barrier). The counterclockwise torsion exhibits a 5.6 kcal/mol barrier, which could be overcome with the excess energy after excitation (about 19 kcal/mol). Both barriers are slightly higher than those for the ASP model. Torsion in both directions results in avoided crossings between the ground and the first excited singlet states. In the clockwise direction, the crossover is reached at 129° . However, at this point, the energies of the S_0 and S_1 states are not degenerate, (the energy gap is ca. 0.27 eV). Root flipping indicates a nearby intersection of S_0 and S_1 potential energy surfaces. In the counterclockwise direction, the crossover is approached at -129° , i.e. ca. 58° from the starting dihedral. Also at this crossing the energies of the interacting states are not equal, although the energy gap (0.12 eV) is smaller than for the first crossing. Similar to the clockwise CoIn, ground and S_1 states interchange their configurations. The computed torsion scan for the GLU2 model shows general trends estimated by Hontani et al. for their CASSCF//CASPT2 torsion scan. [1]

1.3 HOMO and LUMO molecular orbitals at CoIn

Figure 4 shows HOMO and LUMO orbitals corresponding to the geometry at the S_1/S_0 conical intersection. In contrast to the starting geometry, these orbitals are localized on opposite sides of the C13=C14 bond.

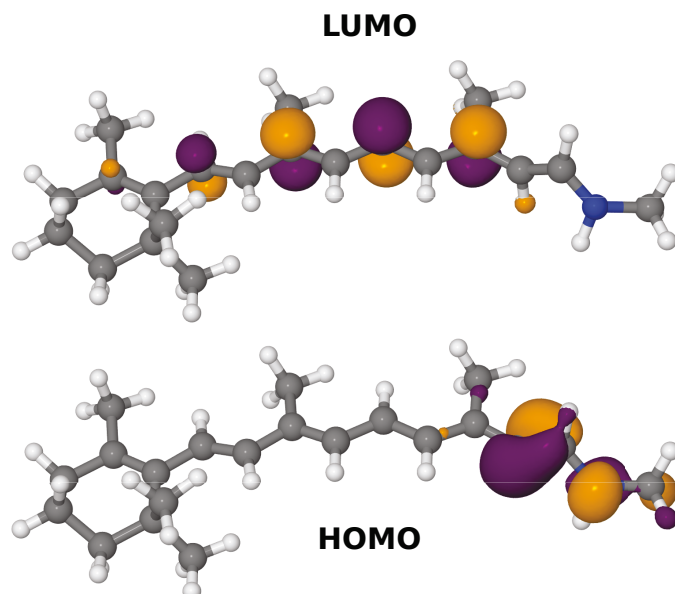


Figure 4: HOMO-LUMO orbitals at the CoIn in the ASP model.

2 Dynamics

2.1 Energy gap at hop

Figure 5 shows the distribution of energy gaps at the hopping point in all calculated models. In the ASP and GLU2 models, we note the typical behavior expected in trajectory surface hopping calculations involving conical intersections. A large number of molecules hop at small S_1/S_0 state energy differences, i.e. close to the crossing seam. This number drops exponentially for hops at large energy differences, i.e. for hops further away from the crossing seam. Due to the small number of hopping events in GLU1, the trend is somewhat difficult to spot.

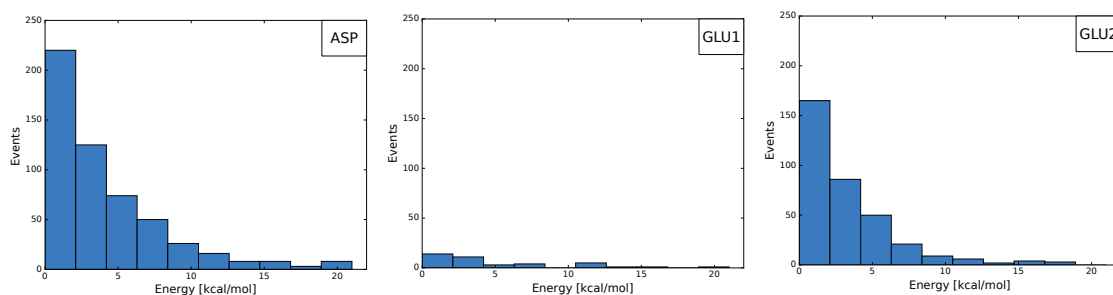


Figure 5: Distribution of energy differences at hop to the ground state for ASP, GLU1 and GLU2 models.

2.2 Reaction time

In Table 1, exponential fit parameters are summarized. Figure 6 provides an example of exponential fit for the S_1 state in the ASP model.

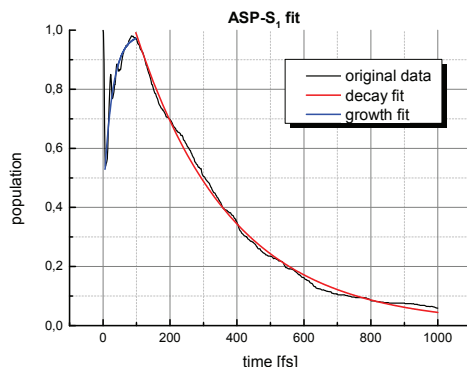


Figure 6: Exponential fit for the S_1 state in the ASP model.

A negative amplitude denotes exponential growth, while a positive amplitude describes a decay. The sum of individual amplitudes corresponding to the S_1 processes is expected to be one, since the initial S_1 population is 100%. However, the first rapid S_1 decay was not fitted, which explains that the total amplitude deviates from one. The offset parameter y_0 is almost zero for the decay components, which indicates entire decomposition of the initial population. For the increase, the y_0 parameter corresponds to the level of repopulation. For the fitting procedure, the Levenberg-Marquardt-algorithm was applied [2] implemented in Origin (OriginLab, Northampton, MA).

2.3 GLU1 population subset

Most of the trajectories in the GLU1 set stayed in the excited state. In order to investigate 40 trajectories that relaxed to the ground state, their populations and occupations were plotted separately (see Figure 7). Analysis of the averaged population and occupation decay for the trajectories relaxed to the GS, reveals an image similar to the ASP and GLU2 models. Within a short time frame (ca. 5 fs), the sub-population of the S_1 state decays, populating S_2 . In the next ca. 100 fs, the population of the S_2 state starts to decrease repopulating the S_1 state. Within next ca. 800 fs the S_1 state starts decaying to the ground state. At ca. 900 fs, the population of the S_0 state has reached 100% of the subset. At ca. 450 fs, half of the starting S_1 population decayed and half of the GS population is achieved. Ca. 37.5% undergo all-*trans* to the 13-*cis*,15-*anti* isomerization by performing HT motion. Ca. 7.5 % react to the all-*trans*,15-*syn* photoproduct. The remaining 55 % return to the starting all-*trans* structure.

Table 1: Exponentially fitted time constants, amplitudes and offset of S_1 and S_2 states in computed models. The quality of the fit is provided by the r^2 parameter.

fit parameters	ASP	GLU1	GLU2
S_1 increase			
A	-0.57	-0.52	-0.68
τ_i , [fs]	26.30 ± 2.07	28.68 ± 1.63	18.1 ± 1.99
y_0	0.98	0.96	0.94
r^2	0.93	0.88	0.79
S_1 decay			
A	1.40	1	3.89
τ_d , [fs]	284.57 ± 1.62	$6.22 \cdot 10^3 \pm 37.1$	95.21 ± 0.94
y_0	$3.58 \cdot 10^{-3}$	0	$-4.88 \cdot 10^{-3}$
r^2	1	0.97	0.98
S_2 decay			
A	0.57	0.44	0.58
τ_d , [fs]	28.78 ± 0.35	41.66 ± 0.83	29.78 ± 0.64
y_0	$2.05 \cdot 10^{-3}$	$9.66 \cdot 10^{-3}$	$1.61 \cdot 10^{-3}$
r^2	0.97	0.91	0.90

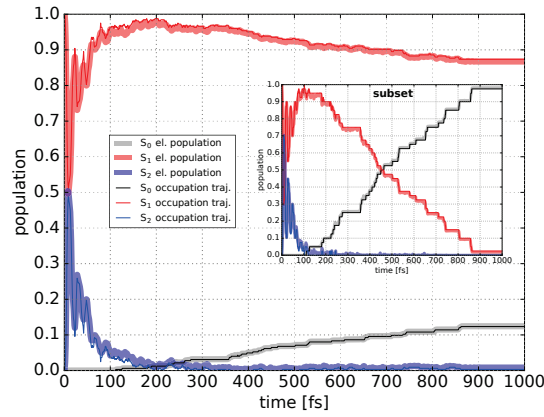


Figure 7: Excited-state populations and occupations in the GLU1 model depicting separately the subset of trajectories hopping to the ground state.

2.4 Product distribution of the GLU1 and GLU2 models

In Figure 8, photoproduct distributions with corresponding hopping times in the GLU1 and GLU2 models are shown. For the GLU1 model, the 13-*cis* product resulting from clockwise and counterclockwise torsion are plotted separately and denoted by (cc) for counterclockwise and (c) for clockwise torsion. The results for this model are difficult to analyze since only a minor population of trajectories relaxes to the ground state. Most of the relaxed trajectories result in an all-*trans* geometry. In the GLU2 model we note that the product population is dominated by the 13-*cis* photoproduct. Most of the trajectories hop between 100-400 fs.

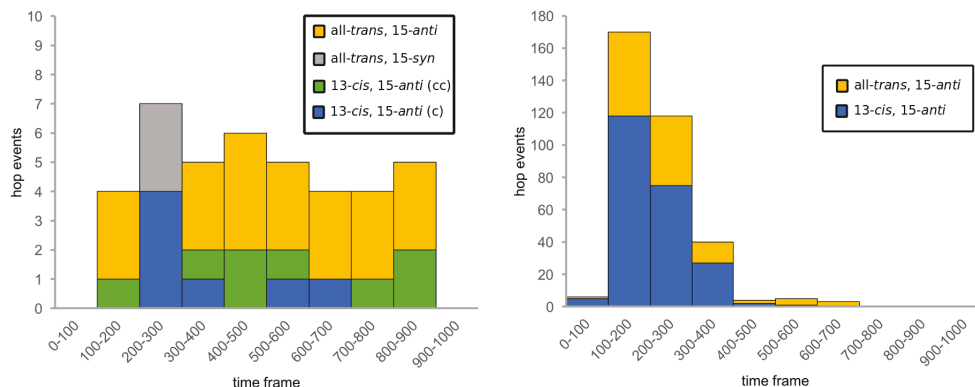


Figure 8: Product distribution along corresponding hopping times in GLU1 (left) and GLU2 (right) models. In the GLU1 model, 13-*cis* (cc) denotes products resulting from counterclockwise torsion and 13-*cis* (c) denotes products resulting from clockwise torsion.

2.5 Influence of starting geometries

To verify the influence of the C8-C9 torsion on the speed of photoisomerization we calculated 0K trajectories based on the original (o) and a modified (m) retinal geometry for the GLU1 and GLU2 models, respectively. The idea was to introduce a C8-C9 twist into the almost planar (original C7=C8-C9=C10: -168°) retinal geometry in the GLU1 model, and to planarize the C7=C8-C9=C10 dihedral (original C7=C8-C9=C10: -146°) in the GLU2 model. The twist (modified C7=C8-C9=C10: -108°) in GLU1 was introduced through partial optimization in the GS. The modified retinal with almost planar C7=C8-C9=C10 (172°) dihedral was obtained through Hartree-Fock S_0 geometry optimization. 0K trajectories with original and modified geometries are shown in Figure 9. The initial simulation time was set to 1 ps.

As expected, introduction of a twist leads to a significant speed-up of the reaction, while the opposite effect is achieved with planarization. We continued the simulations of the non-intersecting trajectories (original 0K trajectory of the GLU1 (oGLU1) and modified trajectory of the GLU2 (mGLU2)) to monitor their further behaviour. While

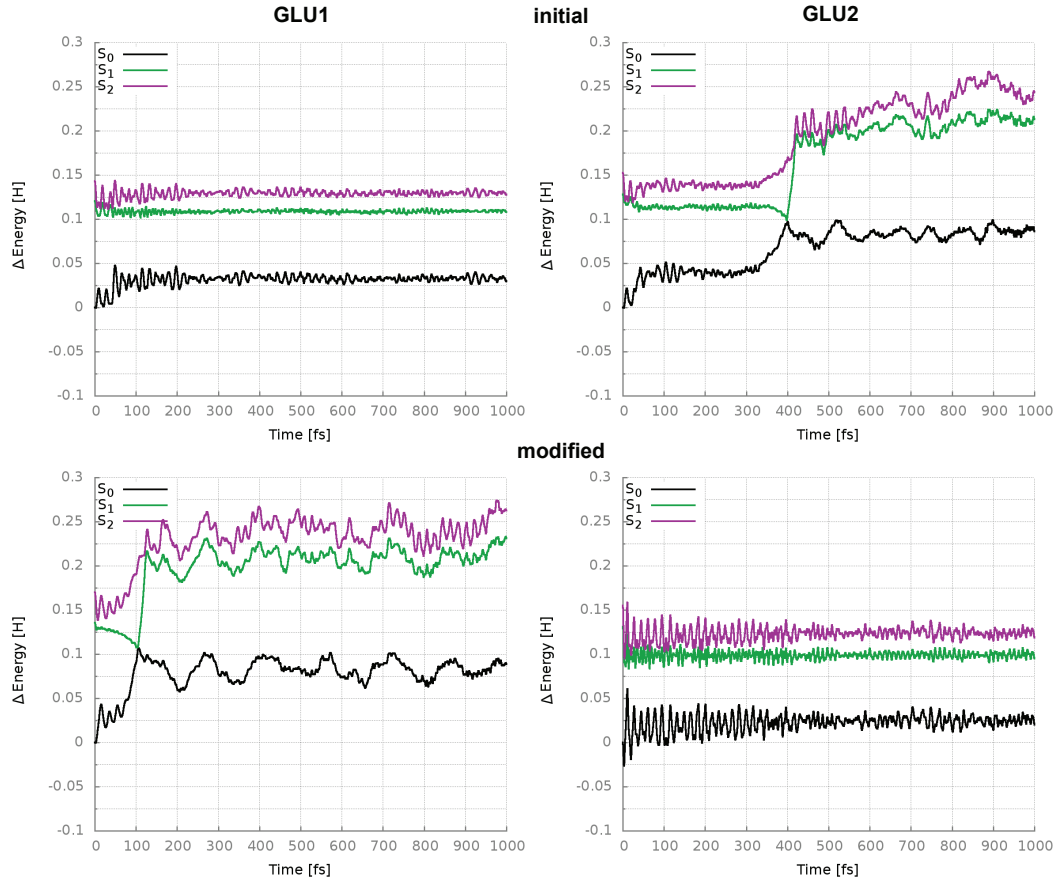


Figure 9: 0K trajectories in GLU1 and GLU2 models based on original and modified retinal geometries.

mGLU2 reaches the intersection after ca. 1.8 ps, the oGLU1 trajectory still resides in the excited state after 3 ps.

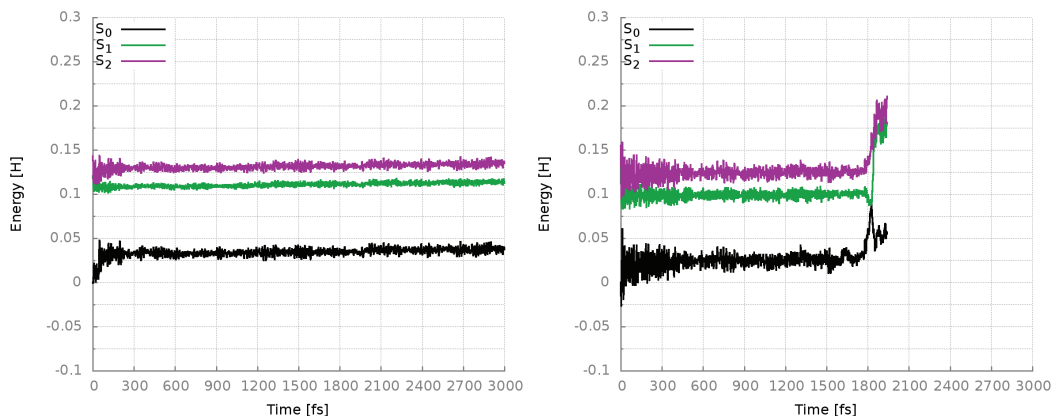


Figure 10: Extended 0K trajectory calculations for the oGLU1 (original GLU1) and mGLU2 (modified) models.

Despite a planar C7=C8-C9=C10 dihedral in both structures, the overall geometries are not completely identical. The GLU2 retinal exhibits a larger NH⁺ twist compared to the chromophore in the GLU1 model (see Figure 11).

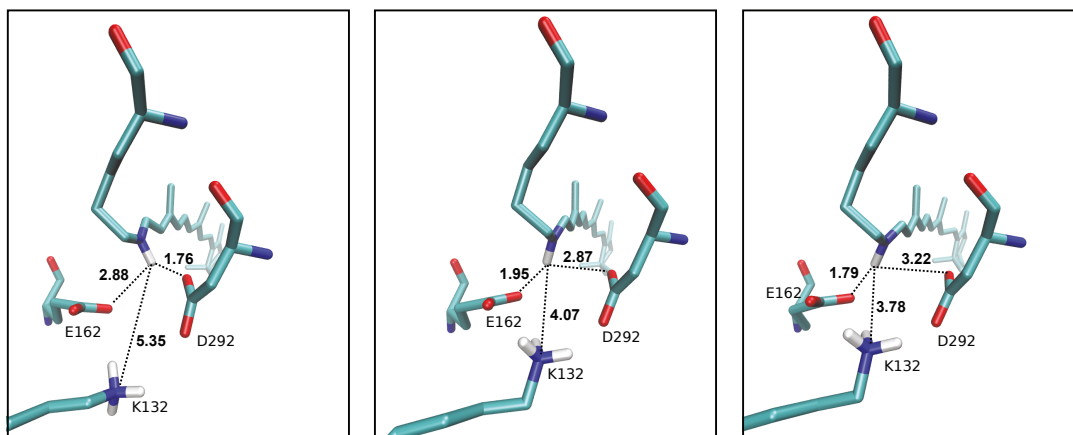


Figure 11: Residues in close proximity of the NH⁺-terminus with corresponding distances.

Thus, this additional twist might be responsible for the higher reactivity of the mGLU2 with planar dihedral angle about the C8-C9 bond, compared to the original GLU1 model with planar C7=C8-C9=C10 dihedral and NH⁺ almost in the molecular plane.

2.6 Retinal C6-C7 torsion potentials

In the GLU2 model, we note a twist about the C6-C7 and C8-C9 single bonds that is not present in the ASP and GLU1 models. Vacuum calculations of retinal models including the β -ionone ring with various semiempirical methods reveal that the minima of the C6-C7 torsion potentials are significantly shifted from those computed with wavefunction or DFT methods. Figure 12 shows that the effect is enhanced with the increasing length of the polyene chain. The freely optimized OM3 geometry with six double bonds has the minimum at ca 130 °, while the RHF and B3LYP methods stay closer to planarity. Apart from the neglect of three-center two-electron integrals, also the basis set plays a role. Semiempirical methods use a minimal basis - the STO-3G basis at the RHF level already demonstrates this effect. An MRCI treatment on top of the OM3/RHF calculation with only the π -orbitals in the active space partially corrects the dihedral toward the RHF and DFT values.

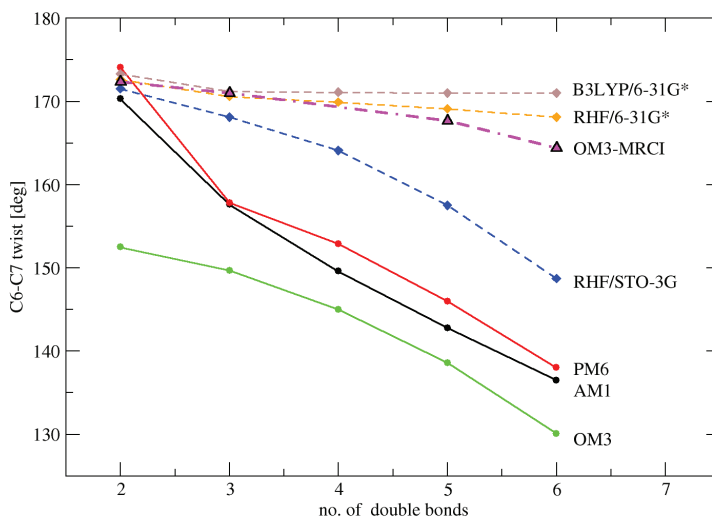


Figure 12: Minimum C5=C6-C7=C8 ground state dihedral angle in PSBs with 2-6 double bonds computed with various methods

The partially optimized OM3 potential along the C6-C7 bond twist is shown in Figure 13, together with the corresponding B3LYP/6-31G* computation. It has distinct minima at -50° and $+45^\circ$ and a shallow minimum at ca. 130° . Unlike for B3LYP, the absolute minimum is located at -50° , i.e. at the 6s-*cis* conformation rather than 6s-*trans*. The potential itself is rather flat, with a barrier of only ca. 2 kcal/mol when going from the 130° local minimum toward 6s-*cis*, compared to more than 7.5 kcal/mol at the B3LYP level. Consequently, the relaxation of the binding pocket in GLU2 may shift residues while keeping the angle in a planar position, generating enough room for a C6-C7 torsion

to occur. The retinal chromophore will then account for this twist by also twisting along the next single bond (C8-C9), as the C9-C13 fragment is still kept in place by the protein residues, thus generating the observed twists in GLU2.

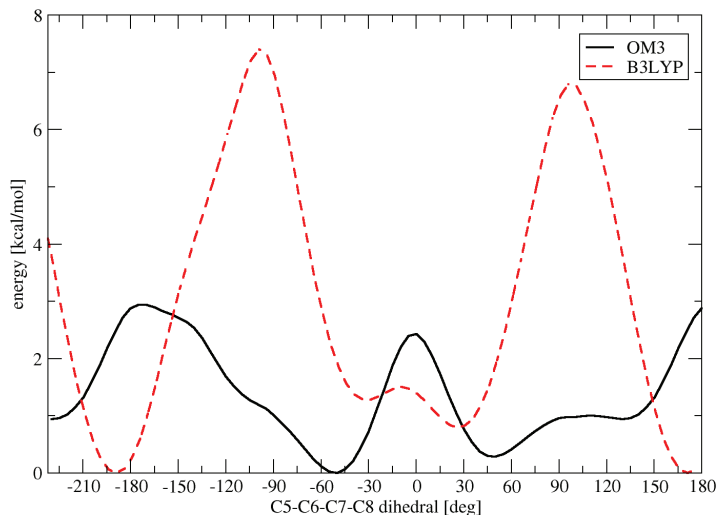


Figure 13: Torsion potentials along the C6-C7 single bond computed with B3LYP and the 6-31G* basis set and with OM3/RHF

2.7 Charge analysis of ASP and GLU1 0K trajectories

Figure 14 shows the charges of the C14- ϵ -CH₂ fragment of retinal during the 0K trajectories of ASP and GLU1 models. It is worth noting that, different from the ensemble computations in the main paper, these simulations were performed without initial velocities and result in slower de-excitation. At the beginning of the simulations, ASP and GLU1 show similar features, i.e. larger fluctuations of charge in S₁ and S₂ excited states. In the GLU1 trajectory, the oscillations become weaker after ca. 250 fs, while they are retained in ASP. In this trajectory we also note fluctuations between S₁ and S₀, i.e. these states have a more similar charge partitioning. After ca. 580 fs, the S₁ state charge significantly drops, thus transporting electron density from the β -ionone ring toward the NH⁺ fragment. In this phase, the single excitation character of the chromophore is restored. When the hop is induced at ca. 600 fs, the charges in S₁ and S₀ states are inverted.

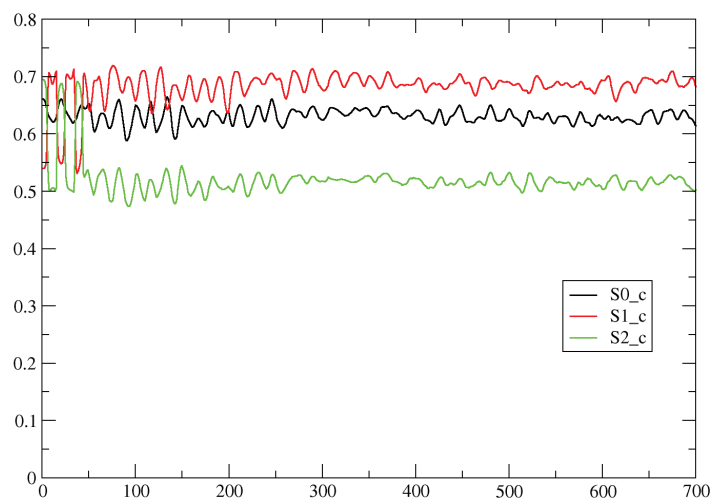
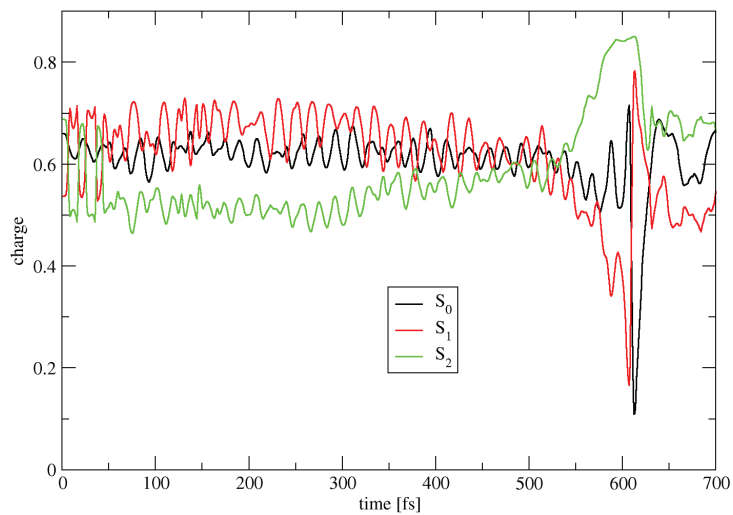


Figure 14: Evolution of charge in the C14- ϵ -CH₂ fragment of retinal in ASP (upper graph) and GLU1 (lower graph) 0K trajectories.

References

- [1] Y. Hontani, M. Marazzi, K. Stehfest, T. Mathes, I. H. Stokkum, M. Elstner, P. Hege-
mann, J. T. Kennis, *Sci. Rep.*, 2017, **7**, 7217.
- [2] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, et al., *Numerical
recipes, Vol. 2*, Cambridge university press Cambridge, 1989.