Article

Frontiers in Multiscale Modeling of Photoreceptor Proteins

MROGINSKI, Maria-Andrea, *et al*.

Abstract

This perspective article highlights the challenges in the theoretical description of photoreceptor proteins using multiscale modeling, as discussed at the CECAM workshop in Tel Aviv, Israel. The participants have identified grand challenges and discussed the development of new tools to address them. Recent progress in understanding representative proteins such as green fluorescent protein, photoactive yellow protein, phytochrome, and rhodopsin is presented, along with methodological developments.

Reference

MROGINSKI, Maria-Andrea, *et al*. Frontiers in Multiscale Modeling of Photoreceptor Proteins. *Photochemistry and Photobiology*, 2021, vol. 97, no. 2, p. 243-269

DOI : 10.1111/php.13372

Frontiers in Multiscale Modelling of Photoreceptor Proteins

Maria-Andrea Mroginski[4], Suliman Adam [17], Gil S. Amoyal [17], Avishai Barnoy [17], Ana-Nicoleta Bondar [9], Veniamin Borin [17], Jonathan R. Church [17], Tatiana Domratcheva[16], Bernd Ensing [6], Francesca Fanelli[14], Nicolas Ferré [5], Ofer Filiba [17], Laura Pedraza González [13], Ronald González [4], Cristina E. González-Espinoza [1], Rajiv K. Kar [17], Lukas Kemmler [9], Seung Soo Kim[12], Jacob Kongsted [11] , Anna I. Krylov [8], Yigal Lahav [17,18], Michalis Lazaratos [9], Qays NasserEddin [17], Isabelle Navizet [7], Alexander Nemukhin[2,3], Massimo Olivucci [13,15], Jógvan Magnus Haugaard Olsen [10], Alberto Pérez de Alba Ortíz [6], Elisa Pieri [5], Aditya G. Rao [17], Young Min Rhee[12] , Niccolò Ricardi [1], Saumik Sen [17], Ilia Solov'yov [19], Luca De Vico [13] Tomasz A. Wesolowski[1], Christian Wiebeler [17], Xuchun Yang [15], Igor Schapiro [17]

[1]. Université de Genève, Département de Chimie Physique 30, quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland.[2]. Department of Chemistry, Lomonosov Moscow State University, Leninskie Gory 1/3, Moscow, 119991, Russia. [3]. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Kosygina 4, Moscow, 119334, Russia. [4]. Technische Universität Berlin, Institut für Chemie, Sekr. PC14, Straße des 17. Juni 135, D-10623 Berlin, Germany. [5]. Aix-Marseille Univ, CNRS, Institut de Chimie Radicalaire, Marseille, France. [6]. Van 't Hoff Institute for Molecular Science and Amsterdam Center for Multiscale Modeling, University of Amsterdam, 1098 XH Amsterdam . [7].MSME, Univ Gustave Eiffel, UPEC, CNRS, 77454 Marne-la-Vallée, France. [8]. University of Southern California, Los Angeles, California 90089, USA. [9]. Freie Universität Berlin, Department of Physics, Theoretical Molecular Biophysics Group, Animallee 14, D-14195 Berlin, Germany. [10]. Hylleraas Centre

for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, Tromsø, N-9037, Norway. [11]. JK: Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Denmark. [12]. Department of Chemistry, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea. [13].Biotechnology, Chemistry and Pharmacy Department, University of Siena, Siena, Italy. [14].Department of Life Sciences, Center for Neuroscience and Neurotechnology, Università degli Studi di Modena e Reggio Emilia, I-41125 Modena, Italy. [15].Chemistry Department, Bowling Green State University, Ohio, USA [16].Department Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahnstrasse 29, 69118 Heidelberg, Germany. [17].Chemistry Institute, The Hebrew University of Jerusalem, Israel. [18]. MIGAL – Galilee Research Institute, S. Industrial Zone, Kiryat Shmona, Israel. [19]. Department of Physics, Carl von Ossietzky University Oldenburg, Carl-von-Ossietzky-Str. 9-11, 26129 Oldenburg, Germany

*Corresponding author e-mail: andrea.mroginski@tu-berlin.de (Maria-Andrea Mroginski), igor.schapiro@mail.huji.ac.il (Igor Schapiro)

ABSTRACT

This perspective article highlights the challenges in the theoretical description of photoreceptor proteins using multiscale modelling, as discussed at the CECAM workshop in Tel Aviv, Israel. The participants have identified grand challenges and discussed the development of new tools to address them. Recent progress in understanding representative proteins such as green-fluorescent protein, photoactive yellow protein, phytochrome, and rhodopsin are presented, along with methodological developments.

INTRODUCTION

Photoreceptor proteins are light-sensitive proteins involved in sensing and response to light in a variety of organisms.[1] In nature, these proteins fulfill important biological functions, such as regulation of circadian rhythms, phototaxis, and light-oriented growth in plants. Photoreceptor proteins absorb light through small organic chromophores embedded within the protein matrix. The chromophore typically absorbs light at a specific wavelength and uses this radiant energy to trigger the protein response, which ultimately leads to completing a biological function. From a biotechnology viewpoint, these proteins represent potential candidates for use as efficient biological light converters. They have already been successfully utilized in a number of technological applications.[2,3] For example, the green fluorescent protein and its derivatives are used to visualize spatial and temporal information in cells with molecular-level resolution. More recently, photoreceptor proteins have been used in the field of optogenetics, which allows light-activation of specific cells in living organisms.[4] In this context they have been successfully aiding researchers investigating biological conditions such as depression, Parkinson's disease, sleep disorders and schizophrenia. Despite the ground-breaking nature of this utilization in life science and other disciplines, our understanding of photoreceptors' function at molecular level is still incomplete. These gaps in knowledge, which hinder the development of new technologies, can be filled with the help of computer simulations of photoreceptors using multiscale modelling.[5]

Due to the large size of these proteins, the hybrid quantum-mechanics/molecular mechanics (QM/MM) embedding has been the main tool used to model photoreceptor proteins. QM/MM allows one to accurately model the chromophore and surrounding environment while remaining

computational feasible.[6–8] In this scheme, the protein environment is treated with classical force fields and the chemically active part of the chromophore and its local environment are treated with more expensive quantum methods. Hence, the first step in using this multiscale approach is to determine how to appropriately partition the QM and MM regions. An additional challenge is to identify an appropriate QM method and force field parameters, if any exist at all. The QM/MM embedding has been widely applied to many families of photoreceptors involving retinal proteins [9], green fluorescent proteins[10], photoactive yellow protein[11], phytochromes[12], and flavin binding proteins[13,14] (Fig. 1).

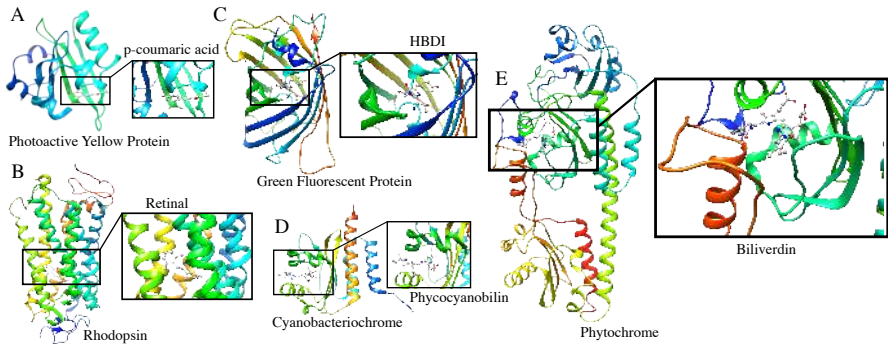


Fig. 1 Photoreceptor proteins addressed in this perspective article. A) Photoactive Yellow Protein with p-coumaric acid as a chromophore; B) Rhodopsin with a retinal chromophore; C) Green Fluorescent Protein with a HBDI chromophore; D) Cyanobacteriochrome with phycocyanobilin as a chromophore; E) Phytochrome with a Biliverdin chromophore.

Furthermore, the photocycle of typical photoreceptor proteins involves multiple competing processes[15,16] illustrated in Fig. 2. Photoexcitation promotes the chromophore into an electronically excited state characterized by a different electron distribution than in the ground state. Different electron distributions result in different bonding patterns and, consequently,

different shapes of the PESs. Ensuing excited-state dynamics often entail isomerization, conformational changes, proton and electron transfer, as well as breaking and forming bonds. The relaxation pathways include fluorescence, internal conversion, and intersystem crossing. Function of natural and engineered photoactive proteins is determined by the interplay between these processes, which entail coupled electronic and nuclear dynamics. Understanding how these quantum processes unfold in systems with many degrees of freedom and coupled to the environment, is of great fundamental and practical importance and quantitative theoretical modeling[10,14] is the key tool for deriving mechanistic insights. Thus, theory is instrumental for understanding these fascinating species and for the design of novel motifs for practical applications. But in order to be useful, the theory should be able to describe multiple interacting electronic states, include the effect of the environment, and be able to provide not only accurate energies, but also nuclear gradients, interstate properties (i.e., non-adiabatic and spin-orbit couplings), as well as other properties relevant for spectroscopy. Furthermore, for a complete description of the photocycle, one needs to be able to execute dynamical simulations including electronic transitions between multiple states. The main challenge here lies in the electronic structure theory and software: despite a tremendous progress[16], much more needs to be done in terms of devising more robust and more versatile electronic structure models for excited states and implementing them in efficient and practical software.[17]

In September 2019 leading experts in the field of computational modelling of photoreceptor proteins met for a Centre of Europeen de Calcul Atomique et Moleculaire (CECAM) meeting which took place in Tel Aviv (Israel). During this CECAM workshop titled “Frontiers in Multiscale Modelling of Photoreceptors” the participants identified the challenges currently facing the field and discussed the development of new tools to address them. Many specific

points were examined in detail including how to determine the correct protonation states of the chromophore within the protein, the effects of electronic polarization on the chromophore and resulting absorption spectra, QM/MM protocols for partitioning systems and sampling, and the computational software used in these types of simulations. In the following we present several contributions that were presented at the CECAM workshop Tel Aviv. The contributions are organized in three categories: 1) challenges in modelling of photoactive proteins, 2) application of multiscale methods to photoactive proteins and 3) methodological development and software updates.

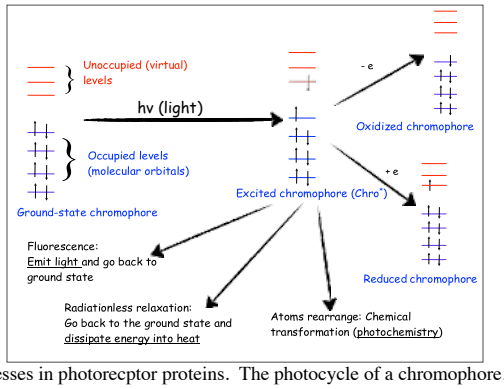


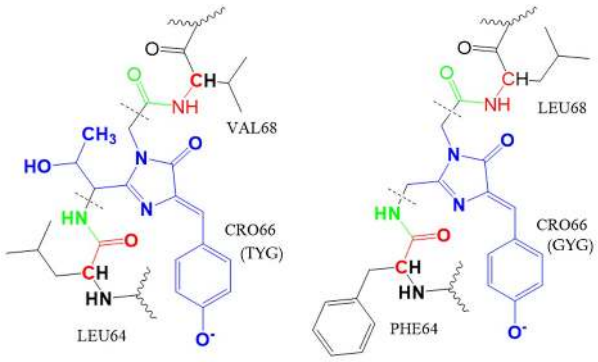
Fig. 2 Excited-state processes in photoreceptor proteins. The photocycle of a chromophore, an acting core of a photoreceptor, involves various competing processes: fluorescence, radiationless relaxation, inter-system crossing (not shown), excited-state chemical transformations and electron transfer. Reproduced with permission from Ref. 15

1. Challenges in modelling of photoactive proteins

1.1 Lessons from recent computational studies of GFP

To illustrate some of the persisting challenges in modelling of photoactive proteins, we consider a recent study[18] on the influence of the first chromophore-forming residue (in position 65) on brightness, photobleaching, and oxidative photoconversion of fluorescent proteins from the GFP family. The

goal of modeling was to explain stark differences in brightness and photostability of EGFP, EYFP, and their mutants with reciprocally substituted chromophore residues, EGFP-T65G and EYFP-G65T. The key quantities responsible for fluorescent quantum yield, extinction coefficients, and bleaching yield are the rates of radiationless and radiative relaxation of the photoexcited proteins. Their first- principle modeling would have required quantum-dynamical simulations of the photoexcited proteins using on-the-fly generated *ab initio* PES and couplings computed using high level of theory (e.g., equation-of-motion coupled-cluster methods[19,20], which is currently impractical. Instead, the authors[18] carried out classical molecular dynamics simulations on the ground and electronically excited states. To study excited-state state dynamics, the force-field parameters of the chromophores were modified to fit the results of electronic structure calculations, most importantly, the change in the torsional potential along the phenolate twist (see Fig. 3) and the changes in partial charges.



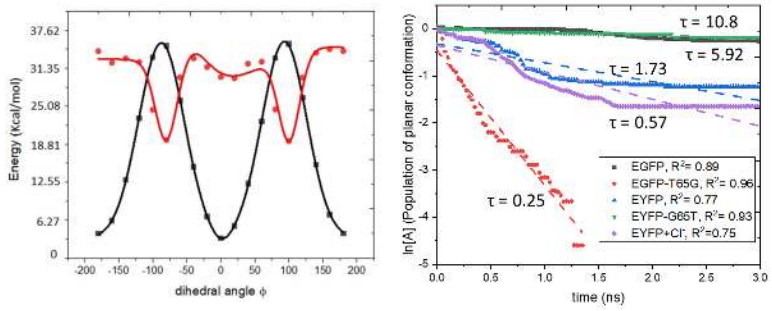


Fig. 3 Top: Structures of the model proteins with the TYG (EGFP, YFP-G65T) (left) and GYG (YFP, EGFP-T65G) (right) chromophores and the definition of the QM/MM partitioning (the QM part is shown in blue and the MM part in black). The key difference between the TYG and GYG chromophores is the -C(OH)CH3 tail in the latter. Bottom left: Potential energy along torsional angle ϕ (phenolate flip) in the ground and excited states. Bottom right: First-order kinetics of the chromophore's twisting in the excited state in four model proteins. Reproduced with permission from Ref. 18.

The rate of twisting along torsional coordinate ϕ, was used as a proxy for the rate of radiationless relaxation. To evaluate brightness and the rate of radiative relaxation, the authors computed excitation energies and oscillator strengths by the QM/MM protocol with high-level electronic structure[21] using the snapshots from the molecular dynamics trajectories. Despite the simplicity of this approach, the calculations were able to pinpoint the role of residue 65 on the photochemical properties of the proteins. The absence of the -C(OH)CH_3 tail in the GYG chromophore affects hydrogen-bond pattern and results in an increased flexibility, which facilitates radiationless relaxation leading to the reduced fluorescence quantum yield in the T65G mutant. The computed lifetimes (see Fig. 3) were in a reasonable agreement with experiment. Although not conjugated with the π-system of the chromophore, the -C(OH)CH_3 tail also affects its electronic properties. The GYG chromophore also has larger oscillator strength as compared to TYG, which leads to a shorter radiative lifetime (i.e., a faster rate of fluorescence). The faster fluorescence rate partially compensates for the loss of quantum efficiency due to radiationless relaxation. The shorter excited-state lifetime of the GYG chromophore is responsible for its increased photostability. One

important aspect left out in this study[18] was the effect of mutation 65 on the rate of photoexcited electron transfer, one of the main bleaching channels.[15,22] To evaluate the rates of electron transfer, one needs to compute Gibbs free energies and couplings, which requires extensive sampling using QM/MM PES of reduced and oxidized forms (as was done, for example, in Ref. 22). Presently, such calculations are rather labor-intensive and also computationally expensive, which precludes their large-scale applications. This study[18] illustrate the need for devising faster electronic structure codes for excited-state treatments and more robust and automated protocols for QM/MM simulations.

1.2 Challenges in modeling bioluminescent systems

Other systems that are related to photoreceptor proteins are the bioluminescent systems. The most studied bioluminescent system is the one responsible for the light emission in fireflies. The emitting light in fireflies arises from the electronic relaxation of oxyluciferin, an organic compound resulting from the oxidation of the D-luciferin substrate inside an enzyme called luciferase. As the fireflies' bioluminescent system is already used as a marker in biology, [23] man needs to understand what are the chemical and physical important factors responsible for the emitted light's color. In order to have insight of the mechanism of the light emission, both experimental and theoretical joint studies have been performed.

In order to theoretically study such systems, the use of quantum mechanical/molecular mechanical (QM/MM) methods is required. Taking into account the surrounding protein at the MM level is essential in order to understand the influence of the enzyme in the color modulation.

The increase of the capacity of the computer allows to deals nowadays with systems of twenty atoms like the firefly oxyluciferin surrounding by thousands of atoms from protein and/or

solvent. The first QM/MM study on this system published in 2010 [24] showed that the protein surrounding is able to modulate the color of the emitter. This first study based on second-order multiconfigurational perturbation calculations using a coupling between two programs MOLCAS, for the QM calculation describing the emitter molecule and TINKER for the MM part describing the protein and solvent environment, as long as using an additive QM/MM scheme with Electrostatic Potential Fitted (ESPF)[25] method, showed that H-bonding network in the cavity was very important and explained how a single mutation of one residue, even far from the light emitter could dramatically changes the H-bond network and therefore the color of the light emission. This study and a latter study [26] demonstrated that increasing the H-bonding network involving the phenolate oxygen of the benzothiazole moiety was inducing a blue-shift of the light emitted whereas the increase of the network in the other side of the molecule was inducing a red-shift. This is easily understandable looking at electronic transition that has a character of LUMO-HOMO and comes with a negative charge transfer from the thiazolone moiety to the benzothiazole moiety during the emission. H-bonding stabilizing the HOMO will not stabilize the LUMO and increase the energy HOMO-LUMO gap, leading to a blue shift (Fig. 4).

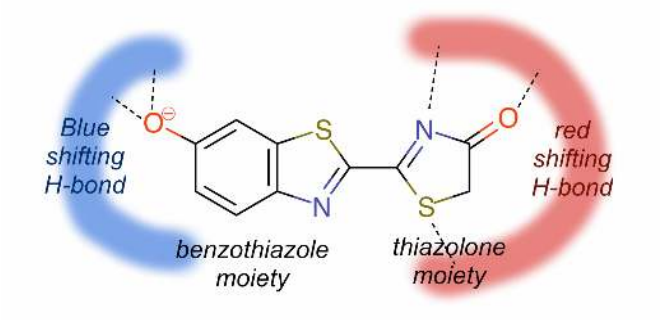


Fig. 4 Oxylyciferin and the H-bonding networks that induced blue or red-shifting of the light emission.

The studies on modulation of the light emission from protein surrounding studies give hope to the computational community and fruitful collaboration with experimentalists. Computation of the light emission of modified emitters and analysis of factors responsible for the color are also very promising.[27,28] Theoretical studies can give complementary insights to the experimental results for the understanding of such complex phenomena. However, all computational studies done in the last decade also point out the challenges to meet. For instance, the calculations are contributing to the availability of crystallographic structures of the protein. Only few structures have the ligands inside the cavity. Some have missing loops that can be very important to the correct description of the protein environment. [26] Models have to be constructed based on hypotheses. This is the case for all systems involving excited states in proteins. From the discussion in the CECAM meeting in Tel Aviv, there is a need to set up robust protocols to face theoretical results up to experimental results.

Protonation state and nature of the emitter are other challenges that are shared with photoreceptor proteins. The nature of the emitter can depend on the pH and the "local" pH inside the cavity is not an experimental measurable data. Joint experimental and numerical

spectroscopic studies of emitters in solvent and of analogues that block the reaction like the keto-enol tautomerization of the thiazolone moiety or the deprotonation of the phenolate group shows that the experimental data can be reproduced by protocols taking into account the dynamic of the protein and that the use of analogues is helpful to a better comprehension of the nature of the light emitter.[29–31] The fluctuating protonation states of the protein residues are still to be better taking into account.

2. Application of multiscale methods to photoactive proteins

Hybrid QM/MM simulations have been instrumental for gaining molecular level insights into the mechanism of light energy conversion and subsequent reactions. These simulations can be used to elucidate reaction pathways directly or in tandem with complementary spectroscopic studies. Often these studies go hand in hand with new method developments or derivation of new unifying concepts that advance our understanding of this light-triggered proteins. In the following sections we present state-of-the-art studies on four different photoreceptor proteins: photoactive yellow protein (PYP), the Green Fluorescent Protein (GFP), Phytochrome and Rhodopsin.

2.1 Resonance interactions of ionic chromophores play a key role in biological photoreception

Ionic chromophores bound to their receptor proteins in protonated or deprotonated forms, are efficiently modulated by interactions with the protein. Such tuning is typically linked to the charge transfer occurring in the chromophore upon photoexcitation and electrostatic interactions with the protein[32–36]. In the S_0 state, the molecular charge is typically localized on the protonated (or deprotonated) moiety of the chromophore whereas in the excited state, the charge is translocated to the opposite end of the conjugated π-system leading to a considerable charge-

transfer character of the S_0-S_1 transition. The electrostatic environment constituted by the protein interacts with these two different charge distributions and thus by modulates the S_0-S_1 energy. Moreover, charge separation at highly twisted geometries enables electrostatic control of the energies of conical intersections (CoIns) and saddle points mediating photochemical or thermal isomerization[36,37]. Although the protein-chromophore interactions are not limited to electrostatic effects[38–40], considering the protein and solvent as a collection of point charges remains a popular approach which, in practice, is often sufficient for reproducing experimentally observed protein effects even quantitatively[41–43].

In the photoactive yellow protein (PYP) photoreceptor, the anionic *p*-coumaric-acid thioester (pCT^-) chromophore is profoundly affected by hydrogen bonding[44] or electrostatic interactions[37]. In the native protein environment, pCT^- photochemically isomerizes around its central double bond (DB)[45,46]. In addition, computational studies suggested that pCT^- may undergo rotation around its central single bond (SB)[37]. Properties of the pCT^- chromophore are rationalized by invoking chemical resonance[10,44] in addition to considering the S_0-S_1 charge transfer. Four resonance forms with the negative charge either on the phenolic or carbonyl groups (Fig. 5) are stabilized depending on the charge localization by hydrogen bonds that pCT^- forms with the protein[47]. Naturally, the contributions of the resonance structures depend on the difference in their energies[48]. The extend of the resonance structure mixing in the S_0 wave function is reflected by the difference in the length of the SB and DB at the S_0 optimized geometry (the bond length alternation, BLA). The larger the energy difference, the larger the BLA and vice versa. Accordingly, the pCT^- chromophore tuned by interactions with water molecules shows an increased BLA when hydrogen bonds are formed with the phenolic group and a decreased BLA when a hydrogen bond is formed with the carbonyl group[49]. The extend of the resonance mixing

determines the S_0-S_1 excitation energy and amount of charge transfer, as demonstrated by the linear correlation plots of these properties (Fig. 6).

Twisting around a central bond increases contributions of the resonance forms with the twisted bond being a single bond (in the S_0 state) and biradicaloid (in the S_1 state), which determines the localization of the molecular charge. At the geometry 90° twisted around the SB, the negative charge is localized on the carbonyl fragment in the S_1 state and on the phenolate fragment in the S_0 state. In contrast, the DB twist increases the negative charge on the phenolate in the S_1 state and on the carbonyl in the S_0 state[37,44,49]. This opposite charge localization in the S_0 and S_1 states enables efficient stabilization of the SB-twisted and DB-twisted CoIns by the carbonyl and phenolic hydrogen bonds, respectively[37]. The energies of the SB and DB CoIns show linear correlations with the BLA (Fig. 6a); the signs of the energy correlations are determined by the charge localization and charge transfer (Fig. 6b).

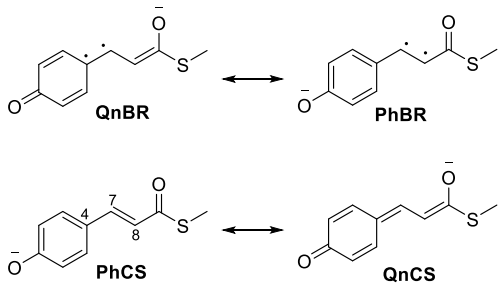


Fig. 5 Resonance structures explaining the interdependent properties of the anionic pCT- chromophore of PYP derived in ref. 49. C4-C7 and C7=C8 are the central single bond (SB) and double bond (DB), respectively.

As suggested by the linear correlations, any property presented in Fig. 6 can be regarded as a descriptor[40] characterizing tuning of the pCT$^-$ chromophore by interactions with its environment in the protein or solvent. The increased mixing of the resonance structures reduces the BLA and charge transfer character, shifts the absorption maximum to the red and activates the SB twist in

the S_1 state. In contrast, the decreased mixing of the resonance structures increases the BLA and charge transfer character, blue-shifts the absorption maximum and favors DB isomerization. In accord with these dependences derived from a computational study, recently published theoretical analysis of systematically tuned green fluorescent protein (GFP) variants[48] has suggested employing the difference in the energies of the resonance forms as a linear scale for analyzing and predicting optical properties. Among the chromophore properties obtained computationally, the BLA is a convenient descriptor as it is derived from the S_0 state geometry optimization. In fact, models examining electrostatic tuning of biological chromophores by their protein environment highlight correlations among properties[36], and in particular, the correlation between the BLA and excitation energy[32–34,50,51]. The correlations of photochemical properties strongly suggest that the theory of resonance could be generally applicable to rationalizing tuning mechanism of photoreceptor proteins.

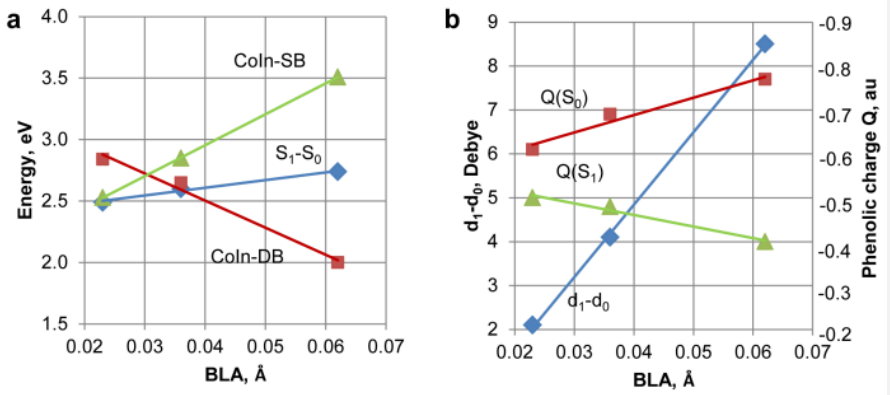


Fig. 6 The linear correlation plots summarize the XMCQDPT2/cc-pvdz results for the pCT- chromophore interacting with water molecules[49]. Panels a and b show correlations for the energies and charge transfer, respectively. The bond length alternation (BLA) value corresponds to the difference in the length of the C4-C7 and C7=C8 bonds at the geometries fully optimized in the S0 state.

2.2 Modeling photochemical reactions

Studies of chemical reactions occurring with chromophores or molecular groups in chromophore-containing pockets in the ground and excited electronic states, constitute an important field of the photoreceptor protein research. Application of methods of multiscale modeling is an essential step in computational simulations of these reactions.

To illustrate the approaches, we consider a reaction of the recovery of the fluorescence state of the reversibly photoswitchable protein Dreiklang[52]. The unique properties of this protein from the Green Fluorescent Protein (GFP) family are due to a reversible hydration/dehydration reaction at the imidazolinone ring of the chromophore. Recovery of the fluorescent state, which is associated with a chemical reaction of chromophore dehydration, is an important part of the photocycle of this protein.

In Fig. 7 a model system composed of the protein surrounded by solvent water molecules is shown in the left panel. The dark balls specify the atoms of the hydrated chromophore, and the side chains of the critical amino acid residues Arg96 and Glu222. The corresponding molecular groups are assigned to the quantum subsystem (QM-part) shown in the right panel in the figure. The remaining molecular groups are included to the molecular mechanics (MM) subsystem.

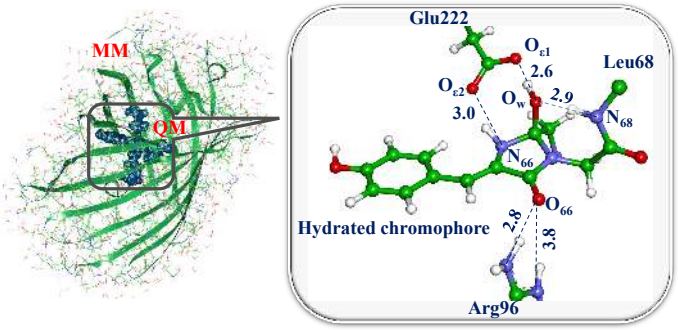


Fig. 7 Left: a model system for simulations of the recovery reaction of the fluorescent state in Dreiklang. Right: a part of the system selected for QM/MM calculations of the reaction energy profile.

In QM/MM calculations, energies and forces in the QM-part are computed using conventional quantum chemistry methods, while the MM-subsystem is described by force field parameters. Usually, the electrostatic embedding scheme is applied to relate the QM and MM parts, assuming contributions of the partial charges from all MM atoms to the one-electron QM Hamiltonian.

The first step of the dehydration reaction in Dreiklang is a proton transfer from the nitrogen atom N_{68} to the oxygen atom O_W, leading to the cleavage of the bond between the hydroxyl and the imidazolinone ring and formation of the water molecule. To compute the corresponding energy profile for this step, a series of QM/MM constrained minimizations along the assumed reaction coordinate (here, the N_{68} – H distance) should be carried out. When the saddle point on the energy surface is located using the conventional transition state search, calculations of harmonic vibrational frequencies should confirm that single imaginary frequency (here, 810i cm^{-1}) characterizes the obtained structure. The computed energy difference between the levels of the reactant and the transition state allows one to estimate the activation energy (here, about 25

kcal/mol), which is consistent with the measured rate constant of the reaction of thermal recovery of the fluorescent state in Dreiklang.

In a similar fashion, a full cycle of chemical transformations in the chromophore maturation in the wild-type GFP,[53] as well as reactions of the photo-induced decomposition of the GFP chromophore upon photobleaching of the protein are considered.[54] Also, we can describe the competing reactions of covalent binding of the biliverdin chromophore to cysteine residues in the bacterial phytochrome domains upon assembly a prospective variant of the near-infrared fluorescent protein miRFP670.[55]

The primary goal of all these simulation is to establish mechanisms of chemical transformations in the chromophore-containing pockets.

2.3 pKa calculations of Phytochromes with Poisson Boltzmann electrostatics

The accurate determination of pKa values of amino acid residues buried in a protein environment remains a challenging task. [56] Since, burial of titratable moieties in proteins usually leads to an increase of the pKa value of acidic groups and the decrease of the pKa value of basic groups, compared to their values in aqueous solution. However, depending on the concrete microscopic description of the protein environment, unusual titration may occur as a result of specific charge-charge interactions which can shift the pKa values of titratable groups in any direction.[57] This effect is often significant at active sites where perturbed pKa values of specific groups are of biological relevance.[58]

Specifically, photoactivation of photoreceptor proteins is often coupled with protonation changes of the chromophore and/or key residues of the protein matrix.[59–62] These protonation

changes, in turn, facilitate proton transfer reactions involved in signal transduction and functional activation.[63] Thus, the precise determination of the protonation states of chromophores and titratable amino acids in their vicinity constitutes a major step in computational modeling of photoreceptors proteins.

Various approaches for computing pKa values in proteins have been developed over the last six decades starting from the pioneering work of Tanford and Kirkwood. [64] Since pKa shifts are of electrostatic origin, much of the effort has been laid on the accurate description of electrostatics using either microscopic or macroscopic models or even a combination of both.[56] For example, approaches based on the combination of electrostatic energy computations based on the solution of the Poisson Boltzmann equation (PBE) with classical molecular dynamics (MD) simulations have been used in different proteins for predicting pKa values.[65] This hybrid approach allows taking into account protein flexibility, hydrogen-bond network rearrangements, side-chain reorientations and water molecules inside protein cavities.

To illustrate this methodology, we will focus on the photoactivation process of the Agp2 phytochrome structure. This process is initiated by a double bond isomerization of the biliverdin (BV) chromophore which is covalently bound to the protein matrix thereby triggering conformational changes and proton transfer reactions between protein and cofactor. It is important to mention that the BV molecule contains six titratable sites, two propionic side chains (pscB and pscC) and four pyrrole rings (A, B, C, and D) (see Fig. 8) Furthermore, two conserved histidine residues are in direct contact with the tetrapyrrole chromophore. Interestingly, spectroscopic data indicate that the propionic side chain at ring C (pscC) of the BV molecule of the Agp2 phytochrome is protonated in the Pfr and Meta-F states and remains protonated even up to pH of 11.[66] Proton release is observed to take place at the step from Meta-F to Pr state

when the photoreceptor becomes activated. [67] Recently, it has been demonstrated that the pscC deprotonation of BV chromophore is essential for triggering a secondary structure change.[68] In the case of prototypical phytochromes, it has been spectroscopically observed that one of the inner pyrrole rings transiently loses a proton during the transition from Meta-R to Pfr states.[62,69] These experimental results highlight the importance of considering the bilin chromophore as a titratable site for computational modeling.

Since pKa calculations rely upon an accurate description of protein electrostatics, derivation of atomic partial charges represents an essential step.[70] This is true in particular for many prosthetic groups such as bilin molecules which were not included in the standard parametrization protocols of classical protein force fields like CHARMM[71], AMBER[72] or GROMOS[73]. Such atomic partial charges can be generated from the electrostatic potential of the molecule computed using quantum chemical approaches and employing the two-stage restraint-electrostatic-potential (RESP) method.[74]

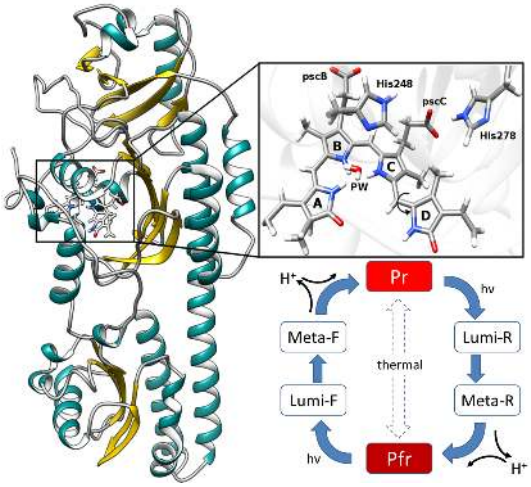


Fig. 8 Left: crystal structure of Agp2 phytochrome in Pfr state, Right: BV chromophore, pyrrole water (PW) and histidines located in the chromophore binding pocket. Bottom: generic phytochrome photocycle with the red light absorbing parent state (Pr) and far-red light absorbing parent state (Pfr).

Predicted pKa values of proteins are highly sensitive to the atomic arrangement of the input structure.[75] This fact may have serious consequences when trying to gain mechanistic insights out of these calculations since the crystal structure represents the protein arrangement at the pH in which it was crystallized and not necessarily the active conformation. In the case of Agp2 phytochrome the pH at which the protein was crystallized was estimated to lie between 5.5 and 9.[76] Since variations in pH can alter the protonation states of titratable groups including the bilin chromophore, slight conformational changes, such as propionic side chains reorientations (pscB and pscC) or rearrangements in the hydrogen bond network can be expected in the crystalline state. These minor structural distortions may significantly affect the computed pKa values.

In order to sample a wider conformational space and thereby account for protein flexibility, Monte Carlo (MC) or Molecular Dynamics (MD) simulations are carried out considering different discrete protonation patterns.[77] These approaches have been applied successfully on different proteins even using very short MD simulations (20 ns). [65] For example, the pKa calculation of the pscC of Agp2 in the Pfr state requires at least three MD trajectories: one trajectory with deprotonated pscC and two trajectories with singly protonated pscC. On each 20 ns long trajectory, time frames are extracted every 100 ps and used as input for electrostatic energy computations performed by solving the linearized Poisson-Boltzmann equation using the BV atomic partial charges derived in the previous step. An equilibrium pKa value of the pscC can be then obtained by averaging the results of electrostatic energy computations for each of the three MD simulations weighted by their respective Boltzmann factor.[38] The pKa values of

the pscC in the Pfr state obtained with this methodology are predicted within the range between 4.5 and 9.6 units, which is in perfect agreement with experimental spectroscopic data suggesting a protonation of the of the carboxylic group. Furthermore, Meyer et al.,[38] showed that the root mean square deviations (RMSD) between measured and computed pKa values of 194 titratable residues in 13 proteins can be improved from 0.96 to 0.79 (in pH units) if energy minimizations with weaker electrostatic interactions ($\varepsilon = 4$) of the structures extracted from MD simulations, is preformed prior to electrostatic calculations.

In a similar fashion, this methodology can be applied to determine the protonation state of histidine residues located in the chromophore binding pocket. His248 and His278 (Fig. 8) are highly conserved residues that are involved in the proton transfer events in the chromophore binding pocket. There have been some efforts to determine precisely protonation states of both histidines. Velazquez et al.,[78] identified one of these histidines as the key residue controlling pH-dependent equilibria in the Cph1 phytochrome, suggesting a pKa below 6.0. Additionally, Takiden et al.,[79] performed electrostatic energy calculations in combination with MD simulations in the Agp1 phytochrome for determining the most likely protonation states of both histidines. The pKa values obtained for both histidines were below 7.0, indicating that both histidines are deprotonated. These results are in agreement with previous PROPKA calculations.[62,80]

Crystal structures are often dehydrated, that means that functional water molecules may be missing. Therefore, some proteins, require the inclusion of additional internal water molecules for the electrostatic energy calculations. This can be achieved implicitly by filling the volume occupied by water molecules with a dielectric medium with a dielectric constant $\varepsilon = 80$. [42] In deed, [81] applied a new cavity-algorithm to 20 titratable groups introduced as point mutations in

Staphylococcal nuclease (SNase) variants for which crystal structures and pKa values are available. This methodology led to a better agreement between computed and measured pKa values in a set of nine mutants, as reflected by an RMSD of 2.04 for pKa obtained with a cavity algorithm compared to 8.8 predicted using standard approach.

In summary, since photoactivation of photoreceptor proteins is often coupled to proton transfer events between chromophore and key residues of the protein matrix, precise determination of their protonation states becomes a crucial step in computational modeling. In this respect, the combination of electrostatic energy calculations with classical MD simulations extended by proper description of protein hydration offers an efficient and reliable approach for investigating pKa shifts of chromophores and amino acid residues of photoreceptor proteins.

2.4 Investigation of the Photoproduct Color Tuning in the Cyanobacteriochrome Slr1393g3

Cyanobacteriochromes (CBCRs) were recently discovered and categorized as a subfamily of phytochrome photoreceptor proteins. They are distinct from the typical phytochrome due to their compact size, because they only require one domain for chromophore incorporation and complete photochemistry, whereas three domains are required in case of canonical phytochromes [82]. Like all the representatives of this superfamily, they are photochromic meaning they have two stable forms which can be interconverted into each other by light of different wavelength. What makes CBCRs special is their variety in absorption maxima in contrast to the canonical phytochromes that absorb in red or far-red. Therefore, it is of high interest to understand the molecular regulation of the absorption in these proteins. One measure for this quantity is the difference in the positions of the lowest energy absorption maxima between the thermodynamically stable dark state and the photoproduct state. To investigate this, application

of multiscale modelling is the natural choice as it allows to determine the excited states of the chromophore, which is described with quantum mechanics, and to explicitly include the effect of the apo-protein, which is treated via molecular mechanics.

One example of a CBCR is the cyanobacteriochrome Slr1393g3 where the absorption is shifted from the red in the dark form Pr, to the green in the photoproduct form Pg. For this protein, crystal structures of both forms are available [83]. In addition, a hybrid form Ph was reported, in which the tetrapyrrolic phycocyanobilin (PCB) chromophore is found in Pr conformation, but its protein environment is in the Pg form. First insights into the absorption of the chromophore inside the protein for these forms could already be gained by calculations on optimized structures [84].

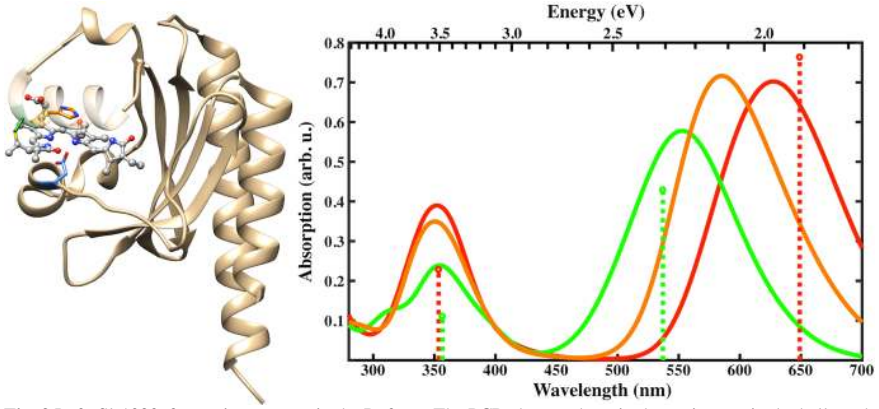


Fig. 9 Left: Slr1393g3 protein structure in the Pr form. The PCB chromophore is shown in gray in the balls and sticks representation and the colors of selected sidechains are: CYS-528 in green, HIS-529 in orange and ASP-498 in blue. Right: Absorption spectra for the Pr (red), Pg (green) and Ph (orange) forms calculated with sTD-DFT based on CAM-B3LYP ground state calculations with a QM region consisting of PCB and the sidechains shown on the left. The spectra are based on 100 snapshots from a DFTB2+D/AMBER trajectory taken every 10 ps. The sticks represent the positions and relative absorption maxima for the Pr and Pg forms extracted from measured spectra [85].

For a systematic study of the absorption in the Pr and Pg forms, a benchmark study was conducted [86]. One focus was on the efficient description of the chromophore in the ground state

geometry. For this purpose, semiempirical methods were employed to optimize the PCB structures of both forms in the ground state and the resulting geometries were compared with higher level *ab initio* calculations. It was found that DFTB2+D is the best performing method for this purpose. In addition, DFTB2+D/AMBER molecular dynamics (MD) simulations were realized to extract snapshots for excited state calculations, for which a variety of semiempirical and *ab initio* methods were benchmarked. Also the results for methods based on *ab initio* time-dependent density functional theory were included in the benchmark, the semiempirical methods ZINDO/S and sTD-DFT as well as the *ab initio* method RI-ADC(2) turned out to be most promising.

The photoproduct tuning was studiedsystematically, using DFTB2+D/AMBER for sampling of 100 snapshots via a 1 ns trajectory and focusing on the three aforementioned methods for the excited state calculations, [87]. It was found that the electrostatic interactions of the protein with the chromophore induce similar shifts in absorption for both forms. In contrast to this, wavefunction analysis showed that the length of the conjugated system decreases when going from the Pr to the Pg form explaining the unusual blue shift observed in this protein. In particular, the tilt of the D-ring is correlated with the energy of the lowest excited state (S1), which is responsible for the absorption in the visible range.

In conclusion, the computational studies lead to a molecular understanding of the photoproduct tuning in the CBCR Slr1393 supporting the trapped-twist mechanism proposed by Lagarias and coworkers for this class of red/green CBCRs [88]. On the technical note, we established a computational protocol for spectra simulations, in particular the QM/MM partitioning, as well as with a benchmark of approaches for efficient conformational sampling. We expect that the derived protocol is applicable to further phytochrome-like photoreceptors.

2.5 Application of constant pH molecular dynamics simulations to sensory rhodopsin

Most of the time, QM/MM models of photoactive proteins are based on a particularly drastic assumption: each and every titratable amino-acid keeps a user-defined protonation state which is determined according to chemical intuition and/or (semi-) empirical titration procedures [89–91]. However, whenever the property of interest is experimentally found to be pH-dependent, this assumption no longer holds. The protein has to be considered as a poly-acid with a very large number of interacting titrating sites. Accordingly, even a slight pH shift may induce numerous protonation changes, leading to some structural reorganization and modified electrostatic interactions, eventually altering the property. The recently designed protocol based on Constant-pH Molecular Dynamics followed by thousands of QM/MM calculations (CpHMD-then-QM/MM) is specifically meant to study how pH can tune the photophysical and photochemical properties of a chromophore embedded in an extended (macro-)molecular environment [92]. In a nutshell, it consists in (i) extracting thousands of statistically independent structures from MD trajectories which are sampling both the conformation and the protonation state spaces of the protein [93] and (ii) averaging the desired property obtained from the corresponding number of QM/MM calculations. The CpHMD-then-QM/MM protocol has been successfully applied to the elucidation of the molecular origin of the pH-dependent absorption spectrum of Anabaena Sensory Rhodopsin (ASR) [94], a trans-membrane protein featuring the retinal chromophore in either the all-*trans* or 13-*cis* conformations [95,96]. Both the tiny pH=3 to pH=5 red-shift and the small pH=5 to pH=7 blue-shift have been reproduced and their molecular origin analyzed.

Here we want to stress out that the number of populated protonation microstates is always large. In the case of ASR, assuming only aspartic acid (9 residues), glutamic acid (5 residues) and histidine (4 residues) can titrate between pH=3 and pH=8, the maximum number of microstates

is $2^9 \times 2^5 \times 3^4 = 1327104$! Even if most of them are not significantly populated at a given pH value, it must be realized that hundreds or thousands of microstates can still co-exist (Table 1)

Table 1. N: number of populated microstates at 3 different pH values (40 ns long CpHMD trajectories, ASR with 13-cis retinal) [94]. #1 to #8: probabilities of the 8 most probable protonation microstates. Note that proton positions are considered indistinguishable in protonated ASP and GLU residues, as well as in deprotonated HIS.

pH	N	1	2	3	4	5	6	7	8
3.5	492	0.31	0.13	0.08	0.06	0.05	0.04	0.03	0.02
5.5	3600	0.03	0.03	0.03	0.02	0.01	0.01	0.01	0.01
7.5	1161	0.08	0.07	0.06	0.06	0.05	0.04	0.03	0.03

While pH=3.5 can be characterized by a few important protonation microstates, it is no longer the case for pH=5.5 and pH=7.5. Accordingly, the ASR absorption spectrum must be calculated as the weighted average of all the most important microstates (Equation 1). The weight (w_i) of a given microstate is pH-dependent: it can be evaluated by the number of structures featuring its proton distribution and extracted from a particular CpHMD trajectory, relatively to the total number of structures.

$\lambda_{max} = \sum_i w_i \lambda_{max}^i$ (1)

In Equation (1), λ_{max}^i represents the maximum absorption wavelength of a given protonation microstate, as calculated as an average of all the corresponding structures. Because the number of populated microstates is usually large, it is possible that the most abundant one is

characterized by a maximum absorption λ^i_{max} that differs significantly from the overall λ_{max}. In equation (1) both λ_{max} and the 8 first λ^i_{max} are reported together for the same 3 pH values.

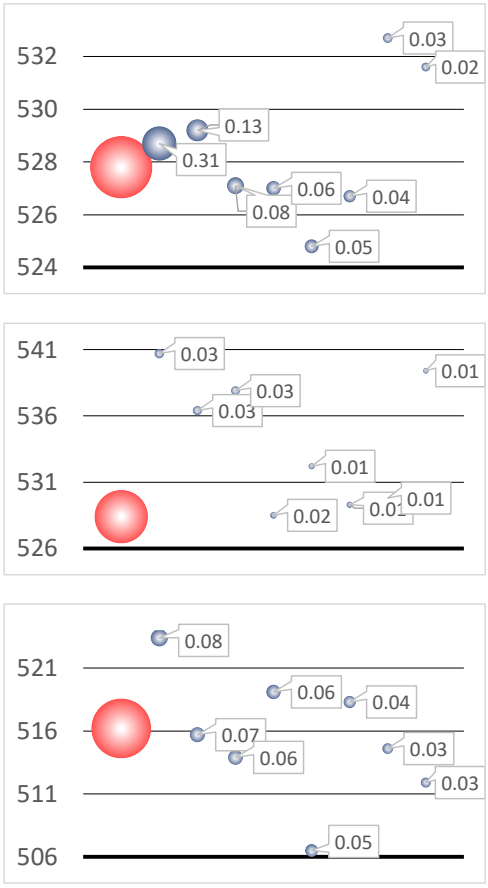


Fig. 10 pH 3.5 (top), 5.5 (middle), 7.5 (bottom) maximum absorption wavelength (in red) and the 8 most populated protonation microstate individual contributions (in grey). Wavelengths are given in nm. Bubble surfaces are proportional to microstate weights (squared labels), relatively to the complete ensemble.

At pH=3.5, only a few ASP and GLU residues are titrating. This translates to a dominant contribution (31%) whose λ^i_{max} is in good agreement with the corresponding λ_{max}. At neutral

Commented [IS1]: What does it mean? Accessible? Titratable?

pH, the most populated microstate features a λ_{max}^{i} value which is 7 nm red-shifted. However, its probability is low (8%) and is balanced by other contributions which are closer to λ_{max}. Accordingly, this is the perfect example where the most abundant protonation microstate is not a good representative of the average. Finally, at the intermediate pH=5.5 value, most of the ASP, GLU and HIS residues are titrating, resulting in a huge number of populated microstates, i.e. to very low individual probabilities. In this case, it is virtually impossible to manually pick a protonation microstate whose λ_{max}^{i} would be close to λ_{max}.

3. Methodological Development and Software Updates

Progress in algorithm and software developments is tightly coupled to the success of multiscale modeling in application to photoreceptor proteins. As a result higher accuracy can be achieved and larger systems can be studied. In this section seven contributions are presented which aim the implementation of data-driven approaches in QM/MM methodologies, characterization of intricate hydrogen bonding networks in large macromolecular systems, the development of new sampling algorithms for the characterization ground- and excited state dynamics of biological chromophores and the implementation of user friendly software interfaces for efficient and automated modeling of complex biological systems, such as rhodopsin, using multiscale methods.

3.1 Extending the capabilities of QM/MM by a data-driven approach

Naturally, the essential purpose of applying any multiscale modeling technique is to reduce the associated computational cost. As explained earlier, QM/MM has been the most straightforward approach for studying photoreceptor and related proteins in this respect, as it does not suffer

from any limitations from the *a priori* selected potential models as in conventional force fields. As a matter of fact, describing photoactive systems requires one to adopt excited state calculations, and applying QM/MM is often very costly. One possible remedy against this cost issue will of course be employing computationally more economic approaches for the QM part such as semi-empirical techniques. However, matching the reliabilities of high level theories such as CASPT2 with semi-empiricism is a daunting task even after intensive re-parametrizing efforts. Thus, somehow constructing the potential energy surface in an explicit manner based on data from high level quantum chemical theories will be a desirable path. It can also be quite useful for studying protein mutation effects as the constructed surface can often be simply re-used with the protein mutation without any further quantum chemical calculations, whereas directly using QM/MM for mutants necessitates re-performing costly QM calculations. The greatest hurdle against explicitly constructing PES is its reliability. When an analytic form of a surface is employed with parametric fitting with the high level computational data, the range of high fidelity region is rather limited. This is especially troublesome for treating photo-activated processes, as the molecular system after photon absorption tends to possess a large amount of vibrational energy such that the chromophore and potentially its neighboring units can wander around an ample configurational space.

As a practical alternative of direct QM/MM calculations without assuming an intrinsically limited analytic form of the surface, using the interpolation mechanics/molecular mechanics (IM/MM) technique can be considered. The interpolation technique was originally designed for describing gas phase reaction dynamics[97], and it was recently extended for describing excited state surface hopping dynamics of chromophore-embedded protein systems such as GFP[98]. Of course, the interpolation itself depends on a relatively large data set with energies, gradients, and

Hessians at multiple configurations. For example, for describing the GFP chromophore, a data set with more than 1000 configurations were needed for a reliability. This burden also increases with more flexible chromophore units. Even for the relatively flexible chromophore in the photoactive yellow protein (PYP) with multiple torsional degrees of freedom, an interpolated PES with high fidelity can still be constructed (Fig. 11)

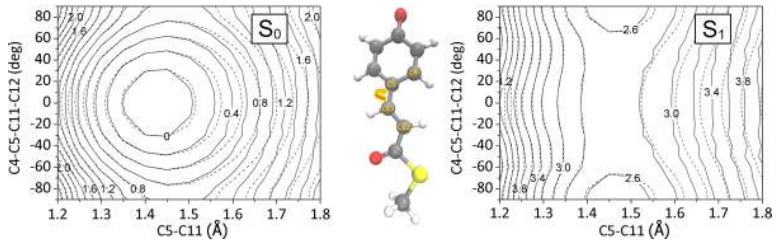


Fig. 11 Reliability of interpolated PES: contour maps of the interpolated S_0 and S_1 state surfaces (solid lines) of the PYP chromophore in comparison with the reference quantum chemical data (dashed lines). Energy values are denoted in eV units. The contours were drawn by varying a torsional angle and its coupled bond length around the S_0-optimized geometry as denoted with the molecular structure. The interpolation data points were sampled in an iterative manner by adopting excited state molecular dynamics simulations. The size of the interpolation data set was 2100.

Interpolation-based PES will be a typical example of data-driven surfaces. As such, how to expedite the data collection processes and how to filter out more important data from a large set will be an important task for generalizing the technique itself, and there are on-going improvements in this regard[99]. In addition, presently adopted Shepard interpolation scheme based on Euclidean distance in the configurational space does not have to be the most reliable approach. Machine learning algorithms which have come into much fashion these days may also contribute significantly in this regard in the future.

3.2 Long-distance proton transfers via dynamic hydrogen(H)-bond networks

In protein environments, the transfer of protons across long distances is thought to occur via H-bond paths composed of protein groups and water molecules. Such H-bond paths could be sampled transiently as the protein changes conformation along its reaction cycle; moreover changes in protonation states during the proton transfer reaction are likely to couple to changes in local protein and water dynamics. Prominent examples here are retinal proteins, for which changes in the retinal isomeric state associate with rearrangements of internal water molecules [100–102], and photosystem II, in which dynamic water molecules help establish proton-conduction path [103].

The dynamic nature of the water-mediated proton-transfer paths and the complexity of the bio-systems, bring about the challenge of how to identify and characterize H-bond paths between putative proton-transfer groups. We have thus designed and implemented Bridge [104], a set of Phyton graph-based algorithms that relies on efficient analyses of water-mediated H-bond networks that derives graphs of the H bonds of the bio-system; this graph of H bonds can be interrogated to identify, for example, all H-bonded paths starting from a protein group, or all H-bonded paths between two protein groups [104].

Particularly important for long-distance proton transfers is to identify, in an ensemble of protein conformations, those events characterized by a continuous connection between the proton donor and acceptor group – such conformations could, for example, be used as starting point for quantum mechanical computations to find out whether the energetics of proton transfer along that path is compatible with experiments [104].

To illustrate the usefulness of Bridge with identifying of H-bond paths in complex bio-systems we present the network of protein-water H bonds in a dimer of chimera

channelrhodopsin, C1C2 (Fig. 12A). The extracellular halves of the two protein monomers participate in a remarkable network of H bonds that included some 48 charges and polar protein groups, and numerous water molecules (Fig. 12A). An unexpected observation from the Bridge analyses was that, in this network, the two retinal Schiff bases can bridge transiently via continuous H-bond paths of 12-13 H-bonds [104]. This long-distance network between the two retinals is rapidly perturbed by mutations that alter H bonding [104].

Once we computed a protein's graph of H bonds, we can use centrality measures to identify protein groups that are common to H-bond paths of particular interest for the functioning of that protein [105]. With betweenness centrality, for example, we evaluate how often a protein group participates in the short-distance paths that connect two other protein groups.[105] Such an analysis reveals that, on the surface of photosystem II, there is a carboxylate group (PsbU-E93 in (Fig. 12B) central to a dense network of protein-water H bonds that included the putative proton-binding site PsbO-D102 [106,107]. Although most of the waters that participate in the H-bond network are very dynamic [107] and visit the surface of the protein just shortly, for picoseconds or less, close to PsbU-E93 waters can stay for as long as 320.6 ± 0.6 ps [107].

Pursuant to the considerations above, we suggest that graph-based analyses of protein H-bond networks provide valuable tools to analyze efficiently large data sets arising from numerical simulations of complex bio-systems, to identify long-distance H-bond paths that could conduct protons, explore the response to mutations, and predict protein groups with central role in long-distance connection networks of a protein.

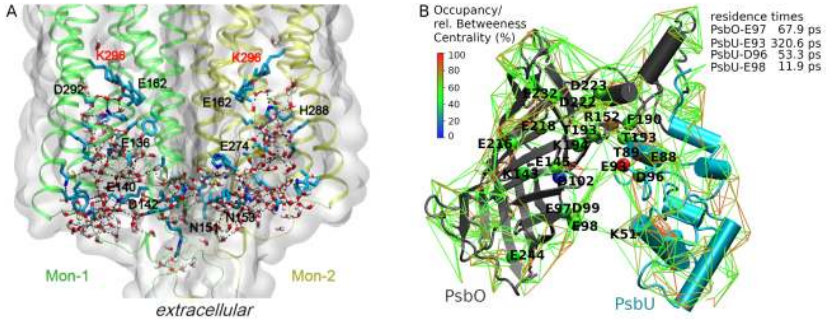


Fig. 12 Dynamic H-bond networks in complex bio-systems. Left: Extensive protein-water H-bond network in the extracellular halves of monomers Mon-1 and Mon-2 of the C1C2 dimer. The protein is shown as ribbons and molecular surfaces, and selected protein groups are shown as bonds with carbon atoms colored cyan, nitrogen – blue, and oxygen – red. For clarity, we label only protein groups part of a shortest-distance path that connects the two retinal Schiff bases. The molecular graphics and the path analyses are based on ref.[104] Right: Protein-water H-bond network at the surface of the soluble PsbO and PsbU subunits of photosystem II. Lines inter-connect charged and polar sidechains via H-bonded water bridges, for clarity, we only show water bridges that are present during at least 50 % of a simulation of the PsbO-PsbU complex in aqueous solution. Cα of amino acid residues are colored according to relative betweeness centrality values. Lines are color-coded, according to occupancy values. The image and the centrality computations are based on ref. [107].

3.3 Toward efficient sampling of photoactivation mechanisms with path-based methods

Even with the computational cost efficiency of multiscale modeling techniques, such as hybrid quantum mechanics/molecular mechanics (QM/MM) simulations, sampling the key steps of a photocycle can still be a daunting task. The pathway from the dark state to the signaling state, and back, is typically comprised of several somewhat activated processes–e.g., electron transfers, proton transfers, and conformational changes–which occur rarely in affordable simulation timescales. Enhanced sampling techniques are commonly used to tackle such challenges in protein systems. However, these schemes traditionally require the definition of a sensible reaction coordinate and/or stable states, both of which are not trivial to formulate and often unknown in photoreceptors. The application of more robust sampling techniques, not subject to numerous trials and errors, is vital to resolve intricate photoactivation mechanisms.

A first step toward increasing sampling efficiency and making free energy calculations and dynamics simulations more affordable, is offered by path-based methods. In these schemes, the handling of a high-dimensional reaction coordinate, composed of several collective variables (CVs), is simplified by the introduction of an optimizable curve that connects two known states in the space of the CVs. Then, the progress along this path CV can effectively be used as a reaction coordinate. Examples of this kind of methods can be found in Refs ([108–110]). The benefits are three-fold: 1) free energy calculations along adaptive paths are not subject to the exponential increase in cost with the dimensionality of the reaction coordinate, and can reach a linear performance scaling [111]; 2) with the diminished penalty for dimensionality, one can introduce more candidate CVs in a single attempt to increase the chance of success; and 3) with the directionality provided by the path, the sampling is focused into the transition region of interest. Furthermore, most standard biasing methods (e.g. umbrella sampling [112] or metadynamics [113]) and algorithmic extensions can be employed either along the path [114], or in the direction perpendicular to it in order to find alternative mechanisms [115]. In Fig. 13 we show an illustrative example of an adaptive path CV capturing a transition channel on the Müller-Brown [116] potential energy surface.

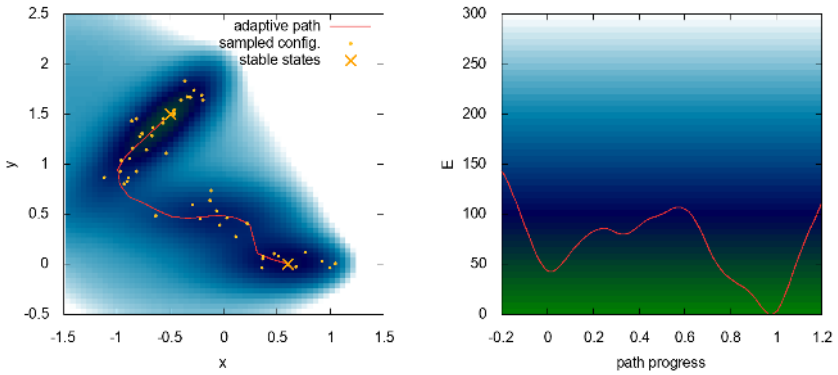


Fig. 13 An adaptive path CV captures the transition channel on the Müller-Brown [116] potential energy surface.

Path CVs have been successfully used in BLUF photoreceptors, to efficiently extract mechanistic details and free energies [117,118]. Other path-based methods that do not require biasing, such as transition path sampling (TPS) [119], have also been used in PYP photoreceptors [120], and similar principles have been applied to rhodopsin by tracking the time evolution of an excited-state population [121].

Path-based methods still require two stable state definitions–e.g., the dark and light states–as well as a set of, even if many, somewhat correct CVs. In cases where these aspects are unknown or debated, a new generation of data-driven and machine learning-based sampling methods holds the key to speed up the exploration of photoactivation mechanisms. CVs can be discovered with novel combinations of clustering, time-lagged independent component analysis, slow-mode separation, autoencoder-based dimensionality reduction, and many more techniques [122–126]. Cutting-edge advances in deep learning also yield free energy differences without the need for reaction coordinates, by mapping atomic configurations to a reference latent representation [127].

3.4 *FDET-Based Simulation of Vertical Excitation Energies of Chromophores Embedded in Proteins*

Frozen-Density Embedding Theory (FDET) [128,129] based multi-level simulations provide an alternative to conventional QM/MM simulations (both polarizable or not). In FDET, the total energy functional is expressed as a functional depending on two independent variables: N_A-electron wavefunction (embedded wavefunction Ψ_A) and a user-chosen density $\varrho_B(\mathbf{r})$ associated with the environment. Ψ_A is thus obtained from the constrained minimization of the Hohenberg-Kohn density functional for the energy of the total system (Fig. 14). We refer the reader to Ref. [129] for the FDET definitions and formulas for the energy and the embedding potential applicable to variational methods for ground state and Refs. [130,131] for FDET extensions. These are omitted here for the sake of brevity. Compared to QM/MM, setting up a FDET simulations involves similar steps: a) selecting the subsystem to be treated at the quantum mechanics level, b) choosing the suitable method for the quantum part, c) generating the embedding potential, d) solving the "embedded QM problem", e) evaluating the properties. The most important differences between FDET-based based and QM/MM simulations concern steps a), c) and e). Concerning a), the commonly used universal system-independent approximations for the non-Coulombic components of the FDET embedding potential given in Eq. (44) in Ref. [129] are adequate only for weakly overlapping Ψ_A and ϱ_B. Thus, the applicability of FDET with such approximations is limited to models where the chromophores are not covalently bound to the protein. Concerning b), generating the FDET embedding potential involves choosing electron density $\varrho_B(\mathbf{r})$ for the environment of the quantum part. This step corresponds to choosing parameters for atom-centered potentials in QM/MM describing the interactions with the

Commented [IS2]: B or c? the test below describes b).

environment. Concerning e), in FDET the non-Coulombic interactions with the environment are taken into account in a self-consistent manner in both energy and in the embedding potential. This results in the dependence on the FDET embedding potential on Ψ_A. For electronic excitations, this numerically inconvenient feature of FDET can be efficiently treated by linearized FDET [130] or by performing additional iterations (see interface E in Fig. 16).

The representation of not only the Coulomb part but also all quantum effects in the FDET embedding potential (not as *a posteriori* energy contributor to the energy, which usually made in QM/MM methods), makes FDET-based methods especially suitable for evaluation of changes in the properties evaluated as expectation values of the embedded wavefunction. The total FDET energy given in Eq. (30) in Ref. [129], includes a term depending solely on $\varrho_B(\mathbf{r})$. For practical applications targeting the energy of the total system, this term must be approximated as well. The essence of multi-level modeling is that this energy contribution is approximated using some simpler method. This offers a large number of possibilities for practical realizations. In this perspective, we focus on such practical applications of FDET where the explicit evaluation of this contribution to the total energy is not needed. This concerns studies targeting the environment induced shifts of observables evaluated as expectation values for a given $\varrho_B(\mathbf{r})$.

Turning back to the choice of $\varrho_B(\mathbf{r})$, the simplest protocol consists in using the electron density evaluated as a ground state density of the environment without the embedded species (*level 0*). If the environment does not comprise molecules which are H-bonded one to another, this protocol can be simplified even further by means of generating $\varrho_B(\mathbf{r})$ as a superposition of densities of individual molecules [128,132]. We recommend *level 0* as the starting point for any large scale FDET based simulation. Benchmark studies on model systems indicate its great usefulness (MAE on 351 excitations is about 0.04 eV). Hydrogen bonding-induced shifts of excitation

energies in organic chromophores lie usually in the range of -1.5 to 1.5 eV. For excitations with magnitude exceeding 0.1eV, FDET-based protocols represent a valuable modeling tool (see Ref. [133] and Fig. 16). More sophisticated protocols to generate $\varrho_B(\mathbf{r})$ take into account such effects as *i*) mutual polarization of different parts of the environment [132], *ii*) implicit or explicit treatment of electronic polarization of $\varrho_B(\mathbf{r})$ by the chromophore [134], *iii*) fluctuations of the structure of the environment [135,136]. We refer the reader also to the work by Neugebauer and collaborators [137,138] on different protocols to generate $\varrho_B(\mathbf{r})$ and their effect on the observables obtained from the embedded wavefunctions for chromophores embedded in proteins. We rather see going beyond *level 0* as a possible option decided on the case-by-case basis. The user should be given the possibility to estimate these effects using smaller model systems to determine if going beyond *level 0* is needed in large scale simulations. Fig. 16 shows a flow diagram indicating essential steps and tools available for setting up and performing large scale FDET-based based computations of electronic excitations for a chromophore embedded in a protein environment. Our previous report on chromophores embedded in proteins [35], used the LR-TDDFT strategy for excited states. The tools presented in Fig. 16 allow the user to: a) use methods going beyond LR-TDDFT if the nature of the excited states requires it and b) more flexible and controllable choices for $\varrho_B(\mathbf{r})$ if going beyond *level 0* is desired.

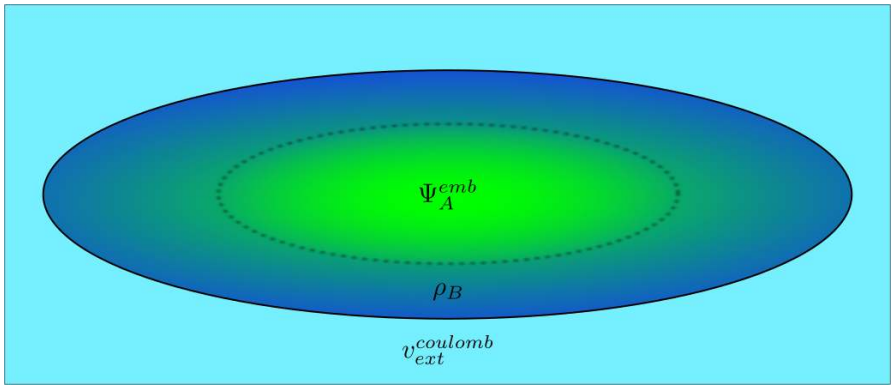


Fig. 14 Scheme of the FDET model of a chromophore embedded in a protein. Different colors represent regions in 3D space which are described using different descriptors: embedded Ψ_A for the chromophore (green) and density $\varrho_B(r)$ for its nearest neighbors (dark blue). Note that these regions can overlap. If needed, the long range effects on the embedded wavefunction can be accounted for by means of a Coulombic potential $v_{ext}^{Coulomb}$.

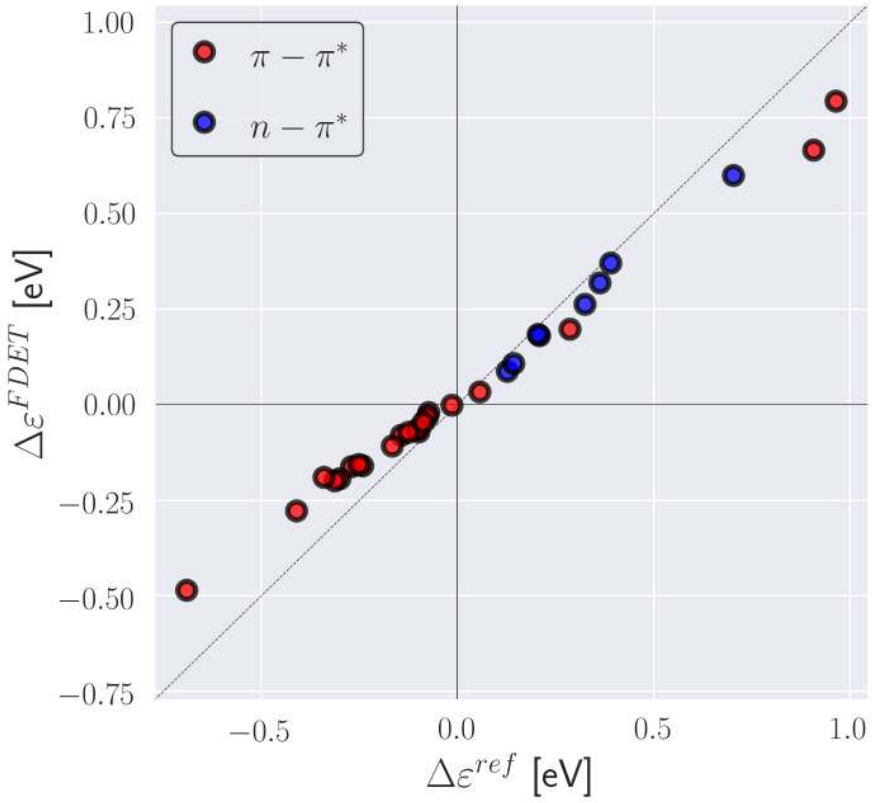


Fig. 15 Environment-induced shifts of the lowest vertical excitation energy for organic chromophores in hydrogen-bonded environments (XH-27 dataset from Ref. [139]). Reference shifts ($\Delta\varepsilon^{ref}$) are taken from Ref. [139] (excitation energy shifts obtained from ADC(2) calculations for the whole clusters). FDET shifts ($\Delta\varepsilon^{FDET}$) are obtained from embedded ADC(2) calculations as described in Ref. [139] except the reduction of the number of centers in the basis sets used for Ψ_A and ϱ_B (monomer expansion is used here).

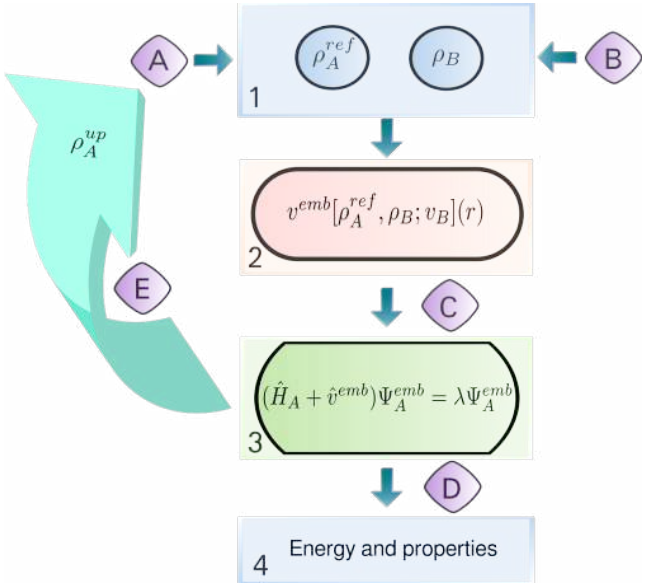


Fig. 16 General work-flow of a FDET-based simulation. The main steps, given in square boxes, can be performed using various standard quantum chemistry codes: 1 – generation of $\varrho_A(r)$ and $\varrho_B(r)$ in real space, 2 – generation of the embedding potential in real space, 3 – obtaining embedded N_A-electron wavefunction (variational or not) from a user chosen quantum chemistry method and code; 4 – a posteriori evaluation of the FDET energy components which depend on the method used in step 2 2–4 and other properties. Interfacing is performed by subroutines indicated with capital letters: A – generation of initial N_A-electron density $\varrho_A^{ref(r)}$; B – generation of $\varrho_B(r)$ (superposition of atomic or molecular densities, statistical ensemble averaging, pre-polarization , freeze-and-thaw optimization, etc.); C – generation of the embedding potential in atomic basis set representation (can include additional electric field component as shown in Fig. 14); D – extracting quantities obtained in step 2 for step 3; E – iterative update of the embedding potential for verification of the linearization approximation (optional).

Commented [IS3]: What is this? Is there some sort of punctuation missing?

Commented [IS4]: What is meant here? V_ext_coulomb? It is an electrostatic field, not electric.

Formatted: Font: Not Bold, Complex Script Font: Not Bold

Deleted: Fig.

3.5 Polarizable Embedding as a Tool to Address Photoreceptor Proteins

Photoreceptor proteins are activated by their interaction with light. In order to understand the working mechanisms of photoreceptors at an atomistic level at least a partial quantum mechanical description is needed. This, unfortunately, is significantly hampered by the fact that the size of photoreceptor proteins in their natural environments quickly becomes out of reach for conventional quantum chemistry methods. Thus, in order to gain atomistic insight into the

functioning, and eventually to be able to define rational design strategies of novel photoreceptor proteins, development of suitable quantum chemistry methods is of significant importance.

The polarizable embedding (PE) model [140] is a fragment-based quantum-classical approach belonging to the class of discrete embedding models, i.e. a central part of the system in question is described at the level of quantum chemistry whereas the remaining part of the system – the environment - is described effectively. For photoreceptor proteins, the part treated using quantum chemistry would typically be chosen as the chromophoric part of the protein. In the PE model, the environment is divided into a number of fragments, and the permanent charge distribution of each fragment is modeled by a multicenter multipole expansion. In addition, distributed dipole–dipole polarizabilities are assigned to each of the fragments thus introducing an explicit account of polarization in the environment. One of the strengths of the PE model is exactly this account of environment polarization in addition to the possibility to calculate the fragment multipole moments and polarizabilities based on separate quantum chemistry calculations, i.e. the model does not rely on the use of a predefined force field. Fig. 17 contains an illustration of the PE model indicating the part of the system treated using either quantum chemistry or by multipoles and polarizabilities.

The PE model has been designed for calculation of spectroscopic properties and excited states in particular. Thus, the model is centered around a formulation building on quantum chemical response theory. Both linear and non-linear properties, such as one-, two-, and three-photon absorption processes, may be described based on the PE model. In addition, PE has been formulated within both time-dependent density functional theory (TD-DFT) as well as correlated wave function approaches, such as coupled cluster (CC) and multi-configurational self-consistent field (MCSCF), and can thus be used to describe situations where TD-DFT is

known to possess problems for example in relation to excited states dominated by double excitations. For a recent discussion of the capabilities of PE model, we refer to Refs. [141,142].

PE is generally applicable to any kind of environment ranging from simple solvents to highly heterogeneous systems like a protein matrix. For the latter, and other biological environments as well, the fragmentation of the environment becomes more involved since covalent bonds need to be broken in order to define the fragments making up the environment. For this we have found the method of molecular fractionation with conjugate caps to be very efficient [143,144].

Through applications, we have generally found the PE model to represent a rather robust computational procedure providing results in close agreement with full quantum chemistry-based calculations. However, care should be taken when considering especially negatively charged molecules (chromophores) or excited states of even partial Rydberg character [145]. In such situations, PE-based calculations may suffer from so-called electron spill-out errors meaning that electron density from the part of the system treated using quantum chemistry is leaking into the environment thereby leading to an over-stabilization of the ground and especially the excited states [146]. In order to address such issues, we recently formulated the polarizable density embedding (PDE) approach [147]. In this model, the fragments in the environment are described by their full charge densities, replacing the multipoles, while still keeping the atom-centered polarizabilities to efficiently account for polarization effects. Importantly, the PDE model contains, in addition, a term in the embedding operator that accounts for Pauli repulsion thereby preventing electron spill-out [146].

Both the PE and PDE models are available through the Polarizable Embedding library (PElib) and are currently interfaced to a number of electronic structure programs – for details we refer to a recent tutorial paper on the use of PE [141].

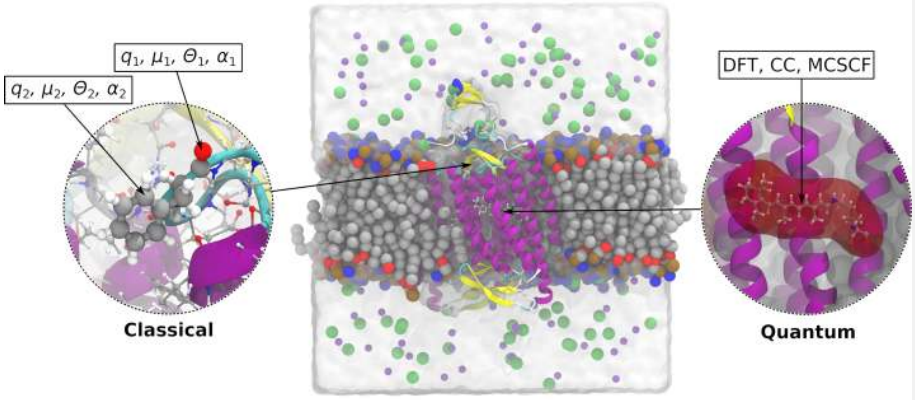


Fig. 17 Illustration of the PE model applied on the membrane-embedded C1C2 channelrhodopsin. The active part, a protonated retinylidene Schiff base, is modeled using DFT/WFT, while the effect from the chromophore environment is modeled classically using atom-centered multipoles and polarizabilities.

3.6 Towards automated population dynamics simulations of light-responsive proteins

At the molecular level, light sensitivity is controlled by two photoreceptor properties: (i) activation quantum efficiency and (ii) dark noise.[121] A complete theory of light sensitivity in biological photoreceptors must therefore describe the relationship between each property and the photoreceptor electronic and molecular structure. Years ago, we reported[36] on the theory of dark noise in a specific family of biological photoreceptors (i.e. light-responsive proteins): type II (animal) rhodopsins.[9] However, even when limiting our interest to such rhodopsin family, a theory of quantum efficiency has not yet been established. As a consequence, we still ignore,

for instance, the mechanism enabling rod cell rhodopsin, the vertebrate retina most sensitive dim-light photoreceptor, to utilize almost 70% of the absorbed photons for visual transduction. Reveling such a mechanism will impact not only our understanding of light sensitivity, but also the design of rhodopsin mutants leading to controllable receptor responses with obvious implications in biology/medicine, [148,149] optogenetics[150,151] as well as in the emerging field of synthetic biology.[152]

As for the case of rhodopsin dark noise, to be of biological interest the validity of a quantum efficiency theory/mechanism has to assess not only for a single photoreceptor but for an entire array of related photoreceptors. When this investigation has to be carried out computationally, it is necessary to construct a full array of photoreceptor models of the same class or, in the case of proteins, homologues or/and mutants. There is another reason for focusing on arrays of models. Rarely chemists and biologists are interested in properties of a specific molecular system but, rather, in trends. In fact, trends are not only more significant for predictions and applications but are less affected by systematic errors in the property calculation. This also apply to the prediction of light sensitivity.

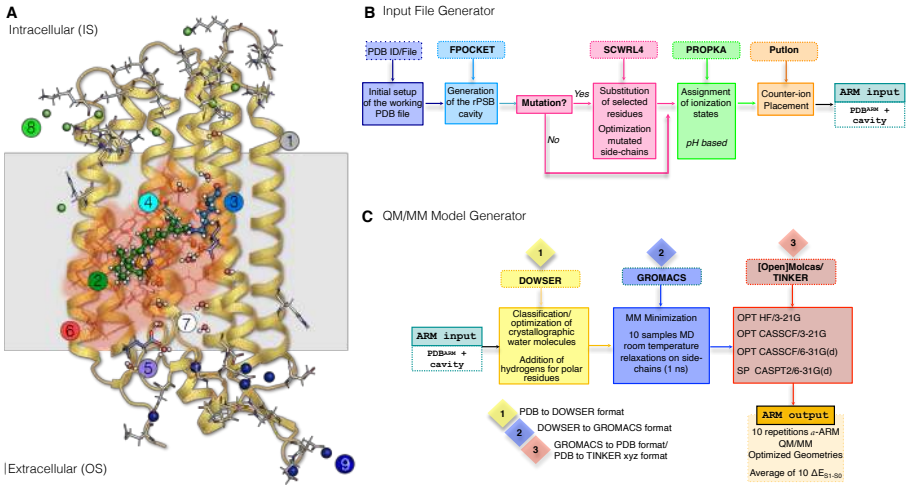


Fig. 18 Structure of *a*-ARM. (A) General scheme of a QM/MM model generated by *a*-ARM for a Type II rhodopsin. This is composed of: (1) environment subsystem (gold cartoon), (2) retinal chromophore (green tubes), (3) Lys side-chain covalently linked to the retinal chromophore (blue tubes), (4) main counter-ion MC (cyan tubes), (5) protonated residues, (6) residues of the chromophore cavity subsystem (red tubes), (7) water molecules, and external (8) Cl^- (green balls) and (9) Na^+ (blue balls) counterions. The external OS and IS charged residues are shown in frame representation (B). (right) General workflow of the *a*-ARM protocol for the generation of QM/MM models of wild-type and mutant rhodopsins. (B) The *a*-ARM protocol comprises two phases: input file preparation phase and (C) QM/MM model generator phase.

The discussion above indicates that the investigation of light sensitivity in rhodopsins (or, actually, any other biological photoreceptor) poses a formidable computational chemistry challenge. On one hand, it is apparent that the complexity of the unavoidable atomistic and multiscale (QM/MM) photoreceptor models (a seven α-helices transmembrane protein incorporating a light responsive retinal chromophore. See Fig18A) and of the protocol for building them, limits the number of models, possibly a few tens, that can be built manually for each given investigation. Therefore, one has to employ an effective automated QM/MM model building protocol if one plans to study hundreds of mutants as it seems necessary for either establishing/simulating a trend, a general mechanism or a mechanistic spectrum. A second issue originates from the fact that the property to be calculated, i.e. the quantum efficiency of the

rhodopsin (Rh) activation, requires the simulation of the light-triggered dynamics of a sizable molecular population. For Rh this corresponds to the dynamics of the light-triggered photoisomerization of the retinal chromophore from its dark (i.e. equilibrium) form containing the 11-*cis* stereoisomer (rPSB11) to its transient bathorhodopsin (bathoRh) primary photoproduct (see Fig. 19A) containing a distorted form of the all-*trans* stereoisomer ($rPSB_{AT}$) as illustrated in Fig. 19B. Thus, the quantum efficiency can be defined as the fraction of photoexcited Rh molecules that after absorption of a photon successfully form bathoRh. Such fraction is indicated with the symbol $\Phi_{cis\text{-}trans}$.

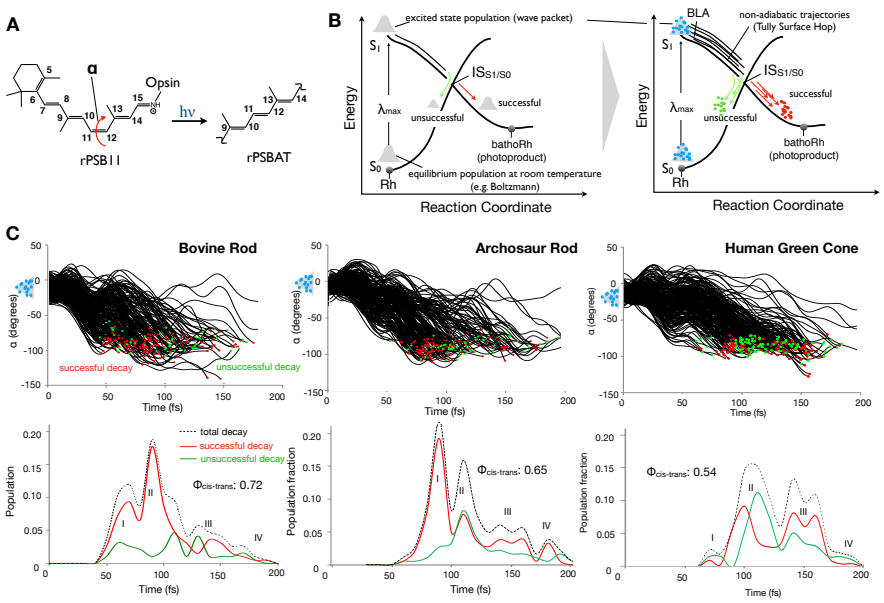


Fig. 19 Rhodopsin population dynamics. (A) 11-*cis* retinal chromophore (rPSB11 for Type II rhodopsins such as Rh) photoisomerization and isomerizing torsional angle α (C10-C11-C12-C13 dihedral angle). (B) Schematic representation of the light-triggered ultrafast population dynamics of Rh. $IS_{S1/S0}$ [J,K] stands for intersection space between the ground state (S_0) and the first singlet excited state (S_1) representing collectively the points of decay (hop) to the ground state (S_0). The reaction coordinate is complex but it is mainly driven by the α angle. The diagram on the right represents a non-adiabatic trajectory calculation where the initial vibrational wave-packet (or population) is represented by a collection of initial conditions (structures and velocities) indicated by light-blue

circles and one trajectory is propagated from each initial condition point. (C) The time progression of α along a set of 200 non-adiabatic trajectories simulating the S_1 population dynamics of bovine rhodopsin at room temperature is given in the top-left panel. The bottom-left panel give the statistic of successful and unsuccessful hops as a function of time. The computed quantum efficiency value is also given. The circles represent decay from S_1 to S_0 with a red circle representing successful decays leading to the photoproduct while green circles lead to the reactant. Center. Same results for a model reconstructed from an amino acid sequence obtained via phylogenetic analysis and ancestral sequence reconstruction techniques [L,M]. Right. Same data for the opsin from a human green cone receptor cell.

In this section we report on the prospective systematic investigation of the $\Phi_{cis\text{-}trans}$, and therefore of light sensitivities of an entire arrays (say hundreds) of rhodopsins. While this is still an unpractical research endeavor, we show that the basic technology necessary to do so is rapidly becoming at hand. Such technology is based on two paradigms: (i) the automatic building of rhodopsin QM/MM models (see Fig. 18 B and C) and (ii) the use of such models for the automated generation of room temperature (actually, any temperature) Boltzmann distributions providing the initial conditions (geometries and velocities) for successive quantum-classical (non-adiabatic) trajectory calculations (see Fig 19B). The resulting trajectory bundle corresponds to a simulation of the light-triggered population dynamics describing the rhodopsin photoisomerization and necessary for $\Phi_{cis\text{-}trans}$ calculations (Fig. 19 C for three different cases).

Automatic building of QM/MM models of rhodopsins. A specialized protocol for the automated construction of QM/MM models of rhodopsins, which uses OpenMolcas[153] as the electronic structure calculation engine, has been introduced. This is the Automatic Rhodopsin Modeling (*a*-ARM) protocol designed to produce congruous and reproducible monomeric, gas-phase and globally uncharged models of rhodopsins based on electrostatic embedding and the hydrogen-link-atom frontier between the QM and MM subsystems.[154,155] Although *a*-ARM currently only constructs rhodopsin-like models (see the model structure in Fig 18 A.), it provides a template for future development and generation of an automatic QM/MM building

strategy for other, more general systems. The building protocol is illustrated and detailed in Fig 18 B and C.

a-ARM has already been benchmarked for several rhodopsins from different organisms and for different functions. In fact, members of the rhodopsin family are found in diverse organisms and, thus, constitute an exceptionally widespread class of light-responsive proteins, driving fundamental functions in vertebrates, invertebrates and microorganisms.[9,156,157] *a*-ARM has been shown to be able to generate models suitable for the prediction of trends in spectroscopic properties (i.e., maximum absorption ($\lambda_{max}{}^{a}$) and emission ($\lambda_{max}{}^{f}$) wavelengths) of wild type rhodopsin-like photoreceptors and their variants, with an error bar of 3.0 (0.13).[154,155] See Fig. 20 A and B. These two critical wavelengths are approximately calculated and expressed in terms of vertical excitation energies $\Delta E_{S1\text{-}S0}$ from S_0 and S_1 energy minima respectively.

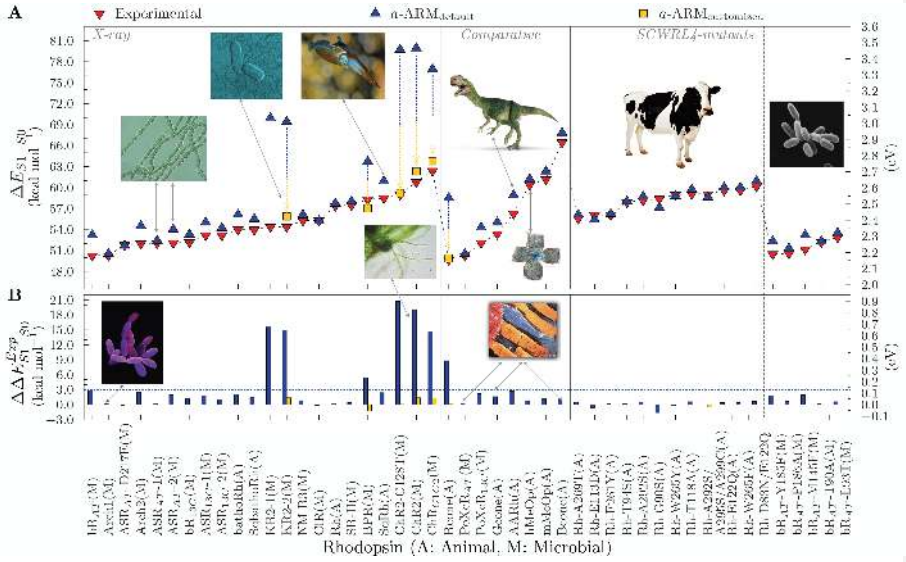


Fig. 20 Benchmarking of *a*-ARM. (A) Computed excitation energies $\Delta E_{S1\text{-}S0}$ in both kcal mol^{-1} (left axis) and eV (right axis) for various rhodopsins. The employed protein structures where obtained from X-ray crystallography (left panel) or through comparative modeling (center panel). Two sets of variants for bovine rhodopsin (Rh) and bacteriorhodopsin (bR) are also reported (right panel). The computed data was obtained using the *a*-ARM$_{default}$ (blue up-turned triangles) and *a*-ARM$_{customized}$ (gold squares). Experimental data, as energy difference corresponding to the wavelength of the absorption maxima, is also reported (red down-turned triangles). (B) Differences between computed and experimental excitation energies $\Delta\Delta E^{Exp}{}_{S1\text{-}S0}$ in both kcal mol^{-1} (left axis) and eV (right axis).

The employed protein structures used as template for the model constructions where obtained from X-ray crystallography (green panel, left) or through comparative modeling (white panel, center). Two sets of variants for bovine rhodopsins (Rh) and bacteriorhodopsin (bR) are also reported (red panel, right). The computed data was obtained using the *a*-ARM$_{default}$[155,158](blue up-turned triangles) and *a*-ARM$_{customized}$[155] (yellow squares). Experimental data, as energy difference corresponding to the wavelength of the absorption maxima, is also reported (red down-turned triangles). (B) Differences between computed and experimental excitation energies $\Delta\Delta E^{Exp}{}_{S1-S0}$ in both (left axis) and (right axis). All computed data is within a 3.0 kcal mol^{-1} (0.13) error, apart from a number of outliers that were corrected using a-ARM$_{customized}$ which required manual changes in, for instance, the conformation of a residue side chain or a change in the ionization state of an ionizable residue. Further details can be found in ref. [158].

Population dynamics during the light-triggered isomerization of homologue rhodopsins. Rod rhodopsin, is the light-sensitive G-protein coupled receptor responsible for dim-light vision in vertebrates. As anticipated above and illustrated in Fig 19. A and B, its activation is driven by a vibrationally coherent 11-*cis* to all-*trans* double-bond photoisomerization of the retinal chromophore which occurs with a 67% quantum efficiency, triggers the receptor photocycle and, ultimately, visual transduction.[9] From the above discussion the first step in investigating such a mechanistic problem must be the construction of the QM/MM model of the photoreceptor capable to describe its spectroscopic and photochemical properties and, most importantly, which

could allow to calculate $\Phi_{cis\text{-}trans}$. This is possible by running a sufficiently large number of non-adiabatic trajectories (e.g. with the surface-hop methodology of Tully and with decoherence correction, see ref. [159–162] and then counting the trajectories that reach the product all-*trans* configuration after the hop to S_0: a process schematically illustrated in Fig 19. B as the fraction of red-circles with respect to the blue circles). The fraction of successful trajectories with respect to the total provides the $\Phi_{cis\text{-}trans}$ value.[13] Of course, the trajectories require initial conditions (nuclear positions and velocities) consistent with a Boltzmann distribution. In the near future, we hope to be able to implement an automated initial condition generator as well as an automated way to start the required number of trajectories (few hundreds at least) directly in the a-ARM QM/MM model generator so that to automate the full procedure of computing $\Phi_{cis\text{-}trans}$. The use of the constantly increasing number of CPU cores available either locally or at regional, national or international computer centers, would make such a research possible even for hundreds of rhodopsins hopefully representing different organisms and mutants. We are convinced that, soon in the future studying systematically $\Phi_{cis\text{-}trans}$ in many diverse organisms will lead to new and fundamental knowledge on how proteins control photochemical reactions in general and what are the mechanistic spectrum achievable. Most importantly we hope to "extract" from these calculations the general mechanistic rules, which must be based on factors such as the steric and electrostatic interactions between protein and chromophore, controlling the mechanistic spectrum and, most importantly, the $\Phi_{cis\text{-}trans}$.

Commented [IS5]: There is something wrong, this sentence is too long

In Fig. 19 C, we report, as demonstrative examples, the comparison between the population dynamics of three different Type II rhodopsin proteins all corresponding to visual pigments (one ancestral to indicate that the proposed calculations can also be applied to studies attempting to reconstruct the evolution of biological photoreceptors[163]). The result of each calculation is

described in terms of progression along the α coordinate spanning the potential energy surface connecting the S_1 vertical excitation region to a decay region in the vicinity of $IS_{S1/S0}$. where the decay occurs at geometries where α is comprised between 60 and 120 degrees on timescales and the timescale goes from 30 to 180 fs. All results shown are based on automatically constructed a-ARM models and 200 trajectory simulations.

3.7 Unifying computational protocols for multiscale modeling of photoreceptor proteins

Computational protocols for multiscale modeling of photoreceptor proteins involves a large number of computer programs and protocols that are generally highly specialized for a particular modeling technique and scale of modeling. The new platform VIKING (Scandina**vi**an Online **Ki**t for **N**anoscale Modellin**g**, viking-suite.com) integrates a number of these tools in a single easy-to-use multiscale platform, that provides tools for setting up simulations, data analysis and visualization.[164] VIKING not only alleviates the need for specialized know-how, which is traditionally required for each individual modeling technique, but also provides a standardized workflow, making the elaborate work of integrating multiple methods in a single study significantly more tractable and reproducible. The primary goal of VIKING is to deliver a set of standard protocols that researchers could use to study complex functioning of biomolecular systems, where photoreceptors naturally provide a good example. Furthermore, VIKING has been developed as a platform where new methods and protocols could be implemented with ease once they become available to the broad research community.

VIKING lowers the barrier of entry and time investment for computational studies of biomolecular processes occurring on sub-atomic to macromolecular scales and beyond. By

making it easy to set up multiscale molecular models and employ a range of industry standard tools, VIKING provides a rapid workflow and illustrates simulation results in a 3D web viewer.

VIKING serves the purpose of a computational microscope, i.e. a unique instrument for researchers. In particular, it provides a computational workflow for intuitive linking of existing modeling software, which so-far existed as stand-alone programs. VIKING utilizes the established programs as “engines” to obtain scientific data and provides unique algorithms that are able to set up all the needed files for computations; earlier the process was often tedious even for experienced users. VIKING algorithms take the user through a carefully thought workflow (VIKING wizard). This VIKING wizard is relying on a unique approach as it integrates more than 15 years of research experience in computational biophysics which allow the system to provide protocol templates to address practically any possible simulation that involves spatial scales ranging from electronic to the macromolecular assemblies. The simulation protocols include justified combination of simulation parameters that throughout decades were shown to be optimal for different types of biophysical simulations.

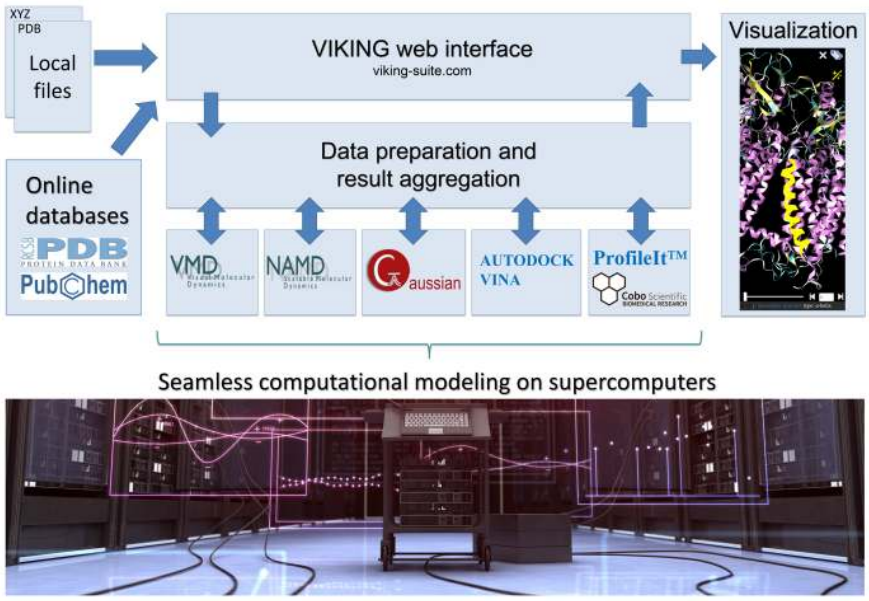


Fig. 21 Concept and workflow of VIKING. Computational tasks are configured in the web interface by supplying the input data (structures, potentials, input field values etc), from the local computer or an online database. The simulation is then performed on a supercomputer (Stampede2, Marconi and Abacus 2.0 are currently supported), the results are aggregated and represented visually in the web browser. Supercomputer photograph courtesy of iStockphoto LP. Copyright 2012.

VIKING wizard addresses all the key questions that one has to answer to prepare needed input for simulations (provide a structure for the simulation, define temperature, pressure, level of theory, etc.); many technical parameters for simulations are then automatically determined by a complex algorithm based on the input provided by the user. This permits unexperienced users, unaware of the algorithmic backgrounds, not to worry about the technicalities. The automatic determination of simulation parameters relies on those simulation protocols that have been shown appropriate for a given simulation through decades of research experience. For example, classical molecular dynamics (MD) simulations of biomolecules are handled through the so-

called molecular mechanics (MM) potential. In MD simulations, a complex molecular assembly is modeled by a set of interacting particles, whose evolution in time and space is calculated by numerical integration of Newton's second law. In most cases the parameters for numerical solutions of the Newton's equation are standard, and the protocol for its solution in the case of a multimillion molecule can be automatized. MD simulations of biological systems in VIKING interfaces to the popular programs NAMD[165], AMBER[166], MBN Explorer[167], or GROMACS[168]. Quantum chemical (QC) codes, such as, e.g., Gaussian[169], GAMESS[170], DALTON,[171] ORCA[172] and Firefly, or the spin dynamics code MolSpin[173] allow studying electronic processes in molecular systems and chemical reactions, which becomes an important issue in most problems dealing with photoreceptive proteins. Naturally, VIKING permits linking MD and QC simulations such that input for one calculation type can be used for another calculation. The multiscale nature of VIKING goes well beyond linking of MD and QC, as it provides the key framework to link any possible scale ranging from electronic to the scale of protein complexes. VIKING effectively prepares parameters for simulation on one scale from other complete simulations and thereby permits programs to exchange data in the most efficient way. VIKING handles the relevant chunks of data from one program into special file formats that permit effective communication between the codes on the dedicated VIKING server.

VIKING is likely the first tool to enable streaming support of biomedical data from supercomputers. High performance computing (HPC) increasingly opens new frontiers for diverse research areas as well as for industrial applications; in this respect, computational investigations of photoreceptors is probably one of the highlight examples. A traditional area of research making extensive use of HPC is computational biophysics, where MD and QC are typically employed to study the workings of complex biomolecular machineries in the smallest

detail. While the increasing HPC power enables researchers to study ever more complex molecular systems, the amount of data produced is becoming a serious challenge of its own – both in terms of storage and processing but also in terms of visual exploration; researchers need large data storage arrays, a fast network connection to the supercomputer and powerful computer workstations with large amounts of memory to explore and analyze the resulting datasets. VIKING is designed as powerful web-based visualization toolkit for atomistic simulations that will allow the user to visually explore results using any PC without the need to download the data in full. Through invention of a unique specialized file format for storing simulation results, the VIKING toolkit will stream the compressed atomic coordinates from a server during "playback" of a simulation to the web browser on a client PC, which will use the data to generate an animated visual representation of the studied molecular structure on the fly.

In summary, we hope that in the future, VIKING will be a handy platform to (i) alleviate the growing logistic challenges when working with large-scale simulation data, (ii) support broader adoption of biophysical simulations through easy-to-use and modern web-based tools and (iii) enable direct sharing of simulation data through the web to any target audience.

CONCLUSIONS

Multiscale modelling has matured as a vital research method since its foundation more than 40 year ago. It has been widely applied in the field of photoreceptor proteins. Specifically, the hybrid QM/MM method have become essential for addressing electronic properties of biological chromophores such as those present in biological photoreceptors. Despite the immense knowledge and experience collected over the last decades, there are still numerous challenges

remaining in the field. The high computational demand associated with these hybrid techniques, the large degree of freedom of the protein matrix and environment, the lack of structural and chemical information and the complex nature of the electronic structure, significantly limits their applicability leaves opportunity for future development. Practical applications of QM/MM approaches to various photoactive proteins, such as GFP, PYP, phytochrome, rhodopsin and luciferase, clearly illustrate the urgent need for devising faster electronic structure codes for excited state description, comprehensible protocols for transparent handling of structural data input and user friendly software for the analysis and evaluation of computational output as well as more robust and automated protocols for QM/MM simulations in general. These challenges in the field as well as progress in the method development were presented and thoroughly discussed at the CECAM meeting in Tel Aviv.

Several conceptual as well as technical strategies for reducing the computational cost of QM/MM approaches were introduced by the participants. For instance, the potential energy surfaces of ground and excited states of chromophores in proteins can be more efficiently performed by combining QM/MM methodology with data-driven approaches such as those based on interpolation mechanics. The implementation of these data driven approaches provides opportunities for further improvements with machine learning techniques. In addition, high computational demand can be also overcome by rational simplification of the system under study. This simple approach has been shown to be effecive for the anionic p-coumaric-acid thioester chromophore of PYP, for which linear correlations were found between excitation energies as well as charge transfer nature and BLA resulting from the combination of a set of resonance structures.

The relevance of conformational sampling when computing any molecular property has been suggested in many contributions. The computational cost of any QM/MM calculation, however, increases enormously upon exploration of the conformational space, which is, for biological systems, still being mostly performed at a classical level. Thus, development and implementation of efficient and accurate hybrid QM/MM sampling techniques are urgently requested. The computational cost of stuyding reaction mechanisms can be drastically reduced, for example, by using path-based methods which replaces the high dimensional space of reaction coordinates by an optimizable curve in the collective variable space connecting two known states. Although the definition of adequate collective variables is not trivial, several cutting-edge methods based on clustering techniques are currently under development.

The complexity of the protein environment usually makes the computation of any ground or excited state property of biological chromophores demanding. Here, the fine tuning of protein electrostatics by protonation states of titratable amino acids and the dynamics of hydrogen bond networks should be adequately described. While the protonation states of titratable side chain can be accurately predicted by Poisson Boltzman based approaches combined with molecular dynamics simulations or in the form of constant pH MD simulations, the analysis of the intricate H-bond network is possible with the help of Bridge, a newly developed tool.

Finally, two strategies for user friendly automation of multiscale modeling protocols of photoreceptor proteins were presented: the Automatic Rhodopsin Modeling (α-ARM) which is able to successfully reproduce and predict excited state properties of rhodopsin-like photoreceptor. Further, the VIKING project which offers a general online platform interfacing various quantum chemical codes and molecular dynamics packages for multiscale modeling of complex biological systems.

The rapid growth of high-performance computing technologies, including the emergence of new and potent mathematical algorithms and their implementations in more user friendly and automated codes, will certainly support the development and implementation of multiscale modeling methods towards higher accuracy and expand their applicability to larger and more complex dynamical biological systems.

ACKNOWLEDGMENTS*:*

A. N. is thankful for discussion with Bella Grigorenko (Department of Chemistry, Lomonosov Moscow State University, Russia) and acknowledges funding from Russian Science Foundation (Grant Number 17-13-01051). A.P.A.O. received funding from the Mexican National Council for Science and Technology (CONACYT). Financial support was provided in part by the DFG Collaborative Research Center SFB 1078 Project C4 (to A.-N.B.) and by the Freie Universität Berlin within the Excellence Initiative of the German Research Foundation. Computing time was provided to A.-N.B. by the HLRN, the North-German Supercomputing Alliance. A.-N.B. thanks Jens Dreger (Physics Department, FU Berlin, Germany) for excellent technical support. Y.M.R. was financially supported by the Mid-career Researcher Program of the National Research Foundation of Korea (Grant 2017R1A2B3004946). J. M. H. O. acknowledges financial support from the Research Council of Norway through its Centres of Excellence scheme (Project ID: 262695). T.A.W. thanks the Swiss National Science Foundation (Grant No. 200020-172532). N.F. thanks the French Agence Nationale de la Recherche for funding (grant ANR-14-CE35-0015-02, project FEMTO-ASR). N.F. acknowledges Centre de Calcul Intensif d'Aix-Marseille for granting access to its high performance computing resources. M.O.

acknowledges funding form NSF Grant No. CHE-CLP-1710191, NIH Grant No. 1R15GM126627 and MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca) Grant "Dipartimento di Eccellenza 2018–2022. A.I.K. was supported by the U.S. National Science Foundation (No. CHE-1856342). A.I.K. is a grateful recipient of the Simons Fellowship in Theoretical Physics and Mildred Dresselhaus Award from CFEL/DESY, which supported her sabbatical stay in Germany. Ilia S. acknowledges financial support by the Lundbeck Foundation, the Danish Councils for Independent Research, Volkswagen Stiftung (Lichtenberg professorship to IAS), and the DFG (GRK1885). J. K. acknowledges the Danish Council for Independent Research for financial support (Grant ID: DFF--7014-00050B) and the H2020-MSCA-ITN-2017 COSINE Training network for Computational Spectroscopy In Natural sciences and Engineering (Project ID: 765739) for financial support. M.A.M. and R.G. acknowledge financial support by the DFG via the SFB 1078. I.S. thanks the SFB 1078 for support within the Mercator program. I.S. gratefully acknowledges funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant No. 678169, "PhotoMutant"). R.K.K. acknowledges support from the Lady Davis Trust for the Arskin postdoctoral fellowship. S.A. acknowledges support by the Minerva Stiftung in the form of a postdoctoral fellowship. J.C. acknowledges the Zuckerman STEM leadership program. CW acknowledges funding by the Deutsche Forschungsgemeinschaft (WI 4853/2-1 and 4853/1-1).

Bibliography

(1) Van Der Horst, M. A.; Hellingwerf, K. J. Photoreceptor Proteins, "Star Actors of Modern Times": A Review of the Functional Dynamics in the Structure of Representative Members of Six Different Photoreceptor Families. *Acc. Chem. Res.* **2004**, *37* (1), 13–20. https://doi.org/10.1021/ar020219d.

(2) Zimmer, M. Green Fluorescent Protein (GFP): Applications, Structure, and Related Photophysical Behavior. *Chem. Rev.* **2002**, *102* (3), 759–781. https://doi.org/10.1021/cr010142r.

(3) Reis, J. M.; Andrew Woolley, G. Photo Control of Protein Function Using Photoactive Yellow Protein. In *Methods in Molecular Biology*; Humana Press Inc., 2016; Vol. 1408, pp 79–92. https://doi.org/10.1007/978-1-4939-3512-3_6.

(4) Hegemann, P.; Nagel, G. From Channelrhodopsins to Optogenetics. *EMBO Mol. Med.* **2013**, *5* (2), 173–176. https://doi.org/10.1002/emmm.201202387.

(5) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chemie Int. Ed.* **2009**, *48* (7), 1198–1229. https://doi.org/10.1002/anie.200802019.

(6) Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angew. Chemie Int. Ed.* **2014**, *53* (38), 10020–10031. https://doi.org/10.1002/anie.201403689.

(7) Karplus, M. Development of Multiscale Models for Complex Chemical Systems: From H+H_2 to Biomolecules (Nobel Lecture). *Angew. Chemie Int. Ed.* **2014**, *53* (38), 9992–

10005. https://doi.org/10.1002/anie.201403924.

(8) Levitt, M. Birth and Future of Multiscale Modeling for Macromolecular Systems (Nobel Lecture). *Angew. Chemie Int. Ed.* **2014**, *53* (38), 10006–10018. https://doi.org/10.1002/anie.201403691.

(9) Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.; Brown, L. S.; Kandori, H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical Reviews*. American Chemical Society January 8, 2014, pp 126–163. https://doi.org/10.1021/cr4003769.

(10) Bravaya, K. B.; Grigorenko, B. L.; Nemukhin, A. V.; Krylov, A. I. Quantum Chemistry behind Bioimaging: Insights from Ab Initio Studies of Fluorescent Proteins and Their Chromophores. *Acc. Chem. Res.* **2012**, *45* (2), 265–275. https://doi.org/10.1021/ar2001556.

(11) Groenhof, G.; Schäfer, L. V.; Boggio-Pasqua, M.; Grubmüller, H.; Robb, M. A. Arginine52 Controls the Photoisomerization Process in Photoactive Yellow Protein. *J. Am. Chem. Soc.* **2008**, *130* (11), 3250–3251. https://doi.org/10.1021/ja078024u.

(12) Rockwell, N. C.; Lagarias, J. C. Phytochrome Evolution in 3D: Deletion, Duplication, and Diversification. *New Phytol.* **2019**, *225* (6), nph.16240. https://doi.org/10.1111/nph.16240.

(13) Goings, J. J.; Hammes-Schiffer, S. Early Photocycle of Slr1694 Blue-Light Using Flavin Photoreceptor Unraveled through Adiabatic Excited-State Quantum Mechanical/Molecular Mechanical Dynamics. *J. Am. Chem. Soc.* **2019**, *141* (51),

20470–20479. https://doi.org/10.1021/jacs.9b11196.

(14) Nemukhin, A. V.; Grigorenko, B. L.; Khrenova, M. G.; Krylov, A. I. Computational Challenges in Modeling of Representative Bioimaging Proteins: GFP-Like Proteins, Flavoproteins, and Phytochromes. *J. Phys. Chem. B* **2019**, *123* (29), 6133–6149. https://doi.org/10.1021/acs.jpcb.9b00591.

(15) Acharya, A.; Bogdanov, A. M.; Grigorenko, B. L.; Bravaya, K. B.; Nemukhin, A. V.; Lukyanov, K. A.; Krylov, A. I. Photoinduced Chemistry in Fluorescent Proteins: Curse or Blessing? *Chem. Rev.* **2017**, *117* (2), 758. https://doi.org/10.1021/acs.chemrev.6b00238.

(16) Matsika, S.; Krylov, A. I. Introduction: Theoretical Modeling of Excited State Processes. *Chemical Reviews*. American Chemical Society August 2018, p 6925. https://doi.org/10.1021/acs.chemrev.8b00436.

(17) Krylov, A.; Windus, T. L.; Barnes, T.; Marin-Rimoldi, E.; Nash, J. A.; Pritchard, B.; Smith, D. G. A.; Altarawy, D.; Saxe, P.; Clementi, C.; Crawford, T. D.; Harrison, R. J.; Jha, S.; Pande, V. S.; Head-Gordon, T. Perspective: Computational Chemistry Software and Its Advancement as Illustrated through Three Grand Challenge Cases for Molecular Science. *J. Chem. Phys.* **2018**, *149* (18). https://doi.org/10.1063/1.5052551.

(18) Sen, T.; Mamontova, A.; Titelmayer, A.; Shakhov, A.; Astafiev, A.; Acharya, A.; Lukyanov, K.; Krylov, A.; Bogdanov, A. Influence of the First Chromophore-Forming Residue on Photobleaching and Oxidative Photoconversion of EGFP and EYFP. *Int. J. Mol. Sci.* **2019**, *20* (20), 5229. https://doi.org/10.3390/ijms20205229.

(19) Krylov, A. I. Equation-of-Motion Coupled-Cluster Methods for Open-Shell and Electronically Excited Species: The Hitchhiker's Guide to Fock Space. *Annu. Rev. Phys. Chem.* **2008**, *59* (1), 433. https://doi.org/10.1146/annurev.physchem.59.032607.093602.

(20) Bartlett, R. J. Coupled-cluster Theory and Its Equation-of-motion Extensions. *WIREs Comput. Mol. Sci.* **2012**, *2* (1), 126. https://doi.org/10.1002/wcms.76.

(21) Krylov, A. I.; Gill, P. M. W. Q-Chem: An Engine for Innovation. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (3), 317. https://doi.org/10.1002/wcms.1122.

(22) Bogdanov, A. M.; Acharya, A.; Titelmayer, A. V.; Mamontova, A. V.; Bravaya, K. B.; Kolomeisky, A. B.; Lukyanov, K. A.; Krylov, A. I. Turning On and Off Photoinduced Electron Transfer in Fluorescent Proteins by π-Stacking, Halide Binding, and Tyr145 Mutations. *J. Am. Chem. Soc.* **2016**, *138* (14), 4807. https://doi.org/10.1021/jacs.6b00092.

(23) Yan, Y.; Shi, P.; Song, W.; Bi, S. Chemiluminescence and Bioluminescence Imaging for Biosensing and Therapy: In Vitro and in Vivo Perspectives. *Theranostics* **2019**, *9* (14), 4047–4065. https://doi.org/10.7150/thno.33228.

(24) Navizet, I.; Liu, Y.-J.; Ferré, N.; Xiao, H.-Y.; Fang, W.-H.; Lindh, R. Color-Tuning Mechanism of Firefly Investigated by Multi-Configurational Perturbation Method. *J. Am. Chem. Soc.* **2010**, *132* (2), 706–712. https://doi.org/10.1021/ja908051h.

(25) Ferré, N.; Ángyán, J. G. Approximate Electrostatic Interaction Operator for QM/MM Calculations. *Chem. Phys. Lett.* **2002**, *356* (3–4), 331–339.

https://doi.org/10.1016/S0009-2614(02)00343-3.

(26) Carrasco-López, C.; Ferreira, J. C.; Lui, N. M.; Schramm, S.; Berraud-Pache, R.; Navizet, I.; Panjikar, S.; Naumov, P.; Rabeh, W. M. Beetle Luciferases with Naturally Red- and Blue-Shifted Emission. *Life Sci. Alliance* **2018**, *1* (4). https://doi.org/10.26508/lsa.201800072.

(27) Berraud-Pache, R.; Navizet, I. QM/MM Calculations on a Newly Synthesised Oxyluciferin Substrate: New Insights into the Conformational Effect. *Phys. Chem. Chem. Phys.* **2016**, *18* (39), 27460–27467. https://doi.org/10.1039/c6cp02585d.

(28) Zemmouche, M.; García-Iriepa, C.; Navizet, I. Light Emission Colour Modulation Study of Oxyluciferin Synthetic Analogues: Via QM and QM/MM Approaches. *Phys. Chem. Chem. Phys.* **2019**, *22* (1), 82–91. https://doi.org/10.1039/c9cp04687a.

(29) García-Iriepa, C.; Gosset, P.; Berraud-Pache, R.; Zemmouche, M.; Taupier, G.; Dorkenoo, K. D.; Didier, P.; Léonard, J.; Ferré, N.; Navizet, I. Simulation and Analysis of the Spectroscopic Properties of Oxyluciferin and Its Analogues in Water. *J. Chem. Theory Comput.* **2018**, *14* (4), 2117–2126. https://doi.org/10.1021/acs.jctc.7b01240.

(30) Garcia-Iriepa, C.; Navizet, I. Effect of Protein Conformation and AMP Protonation State on Fireflies' Bioluminescent Emission. *Molecules* **2019**, *24* (8). https://doi.org/10.3390/molecules24081565.

(31) García-Iriepa, C.; Zemmouche, M.; Ponce-Vargas, M.; Navizet, I. The Role of Solvation Models on the Computed Absorption and Emission Spectra: The Case of Fireflies Oxyluciferin. *Phys. Chem. Chem. Phys.* **2019**, *21* (8), 4613–4623.

https://doi.org/10.1039/C8CP07352J.

(32) Pal, R.; Sekharan, S.; Batista, V. S. Spectral Tuning in Halorhodopsin: The Chloride Pump Photoreceptor. *J. Am. Chem. Soc.* **2013**, *135* (26), 9624–9627. https://doi.org/10.1021/ja404600z.

(33) Fujimoto, K.; Hayashi, S.; Hasegawa, J. Y.; Nakatsuji, H. Theoretical Studies on the Color-Tuning Mechanism in Retinal Proteins. *J. Chem. Theory Comput.* **2007**, *3* (2), 605–618. https://doi.org/10.1021/ct6002687.

(34) Cheng, C.; Kamiya, M.; Uchida, Y.; Hayashi, S. Molecular Mechanism of Wide Photoabsorption Spectral Shifts of Color Variants of Human Cellular Retinol Binding Protein II. *J. Am. Chem. Soc.* **2015**, *137* (41), 13362–13370. https://doi.org/10.1021/jacs.5b08316.

(35) Zhou, X.; Sundholm, D.; Wesołowski, T. A.; Kaila, V. R. I. Spectral Tuning of Rhodopsin and Visual Cone Pigments. *J. Am. Chem. Soc.* **2014**, *136* (7), 2723–2726. https://doi.org/10.1021/ja411864m.

(36) Gozem, S.; Schapiro, I.; Ferré, N.; Olivucci, M. The Molecular Mechanism of Thermal Noise in Rod Photoreceptors. *Science (80-.).* **2012**, *337* (6099), 1225–1228. https://doi.org/10.1126/science.1220461.

(37) Boggio-Pasqua, M.; Burmeister, C. F.; Robb, M. A.; Groenhof, G. *Photochemical Reactions in Biological Systems: Probing the Effect of the Environment by Means of Hybrid Quantum Chemistry/Molecular Mechanics Simulations*; The Royal Society of Chemistry, 2012; Vol. 14, pp 7912–7928. https://doi.org/10.1039/c2cp23628a.

(38) Yanai, K.; Ishimura, K.; Nakayama, A.; Hasegawa, J. Y. First-Order Interacting Space Approach to Excited-State Molecular Interaction: Solvatochromic Shift of p-Coumaric Acid and Retinal Schiff Base. *J. Chem. Theory Comput.* **2018**, *14* (7), 3643–3655. https://doi.org/10.1021/acs.jctc.7b01089.

(39) Guareschi, R.; Valsson, O.; Curutchet, C.; Mennucci, B.; Filippi, C. Electrostatic versus Resonance Interactions in Photoreceptor Proteins: The Case of Rhodopsin. *J. Phys. Chem. Lett.* **2016**, *7* (22), 4547–4553. https://doi.org/10.1021/acs.jpclett.6b02043.

(40) Khrenova, M. G.; Nemukhin, A. V.; Tsirelson, V. G. Origin of the π-Stacking Induced Shifts in Absorption Spectral Bands of the Green Fluorescent Protein Chromophore. *Chem. Phys.* **2019**, *522*, 32–38. https://doi.org/10.1016/j.chemphys.2019.02.010.

(41) Yu, J. K.; Liang, R.; Liu, F.; Martínez, T. J. First-Principles Characterization of the Elusive i Fluorescent State and the Structural Evolution of Retinal Protonated Schiff Base in Bacteriorhodopsin. *J. Am. Chem. Soc.* **2019**. https://doi.org/10.1021/jacs.9b08941.

(42) Schnedermann, C.; Yang, X.; Liebel, M.; Spillane, K. M.; Lugtenburg, J.; Fernández, I.; Valentini, A.; Schapiro, I.; Olivucci, M.; Kukura, P.; Mathies, R. A. Evidence for a Vibrational Phase-Dependent Isotope Effect on the Photochemistry of Vision. *Nat. Chem.* **2018**, *10* (4), 1–7. https://doi.org/10.1038/s41557-018-0014-y.

(43) Suomivuori, C. M.; Gamiz-Hernandez, A. P.; Sundholm, D.; Kaila, V. R. I. Energetics and Dynamics of a Light-Driven Sodium-Pumping Rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (27), 7043–7048. https://doi.org/10.1073/pnas.1703625114.

(44) Gromov, E. V.; Burghardt, I.; Köppel, H.; Cederbaum, L. S. Native Hydrogen Bonding Network of the Photoactive Yellow Protein (PYP) Chromophore: Impact on the Electronic Structure and Photoinduced Isomerization. *J. Photochem. Photobiol. A Chem.* **2012**, *234*, 123–134. https://doi.org/10.1016/j.jphotochem.2012.01.007.

(45) Genick, U. K.; Soltis, S. M.; Kuhn, P.; Canestrelli, I. L.; Getzoff, E. D. Structure at 0.85 Å Resolution of an Early Protein Photocycle Intermediate. *Nature* **1998**, *392* (6672), 206–209. https://doi.org/10.1038/32462.

(46) Schotte, F.; Cho, H. S.; Kaila, V. R. I.; Kamikubo, H.; Dashdorj, N.; Henry, E. R.; Graber, T. J.; Henning, R.; Wulff, M.; Hummer, G.; Kataoka, M.; Anfinrud, P. A. Watching a Signaling Protein Function in Real Time via 100-Ps Time-Resolved Laue Crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (47), 19256–19261. https://doi.org/10.1073/pnas.1210938109.

(47) Gromov, E.; Domratcheva, T. Four Resonance Structures Elucidate Double-Bond Isomerisation of a Biological Chromophore. *Phys. Chem. Chem. Phys.* **2020**. https://doi.org/10.1039/D0CP00814A.

(48) Lin, C. Y.; Romei, M. G.; Oltrogge, L. M.; Mathews, I. I.; Boxer, S. G. Unified Model for Photophysical and Electro-Optical Properties of Green Fluorescent Proteins. *J. Am. Chem. Soc.* **2019**, *141* (38), 15250–15265. https://doi.org/10.1021/jacs.9b07152.

(49) Gromov, E. V., Domratcheva, T. Four Resonance Structures Elusidate Double-Bond Isomerization of a Biological Chromophore. *To be Publ.*

(50) Sekharan, S.; Morokuma, K. Drawing the Retinal out of Its Comfort Zone: An

ONIOM(QM/MM) Study of Mutant Squid Rhodopsin. *J. Phys. Chem. Lett.* **2010**, *1* (3), 668–672. https://doi.org/10.1021/jz100026k.

(51) Sekharan, S.; Altun, A.; Morokuma, K. Photochemistry of Visual Pigment in a Gq Protein-Coupled Receptor (GPCR)-Insights from Structural and Spectral Tuning Studies on Squidrhodopsin. *Chem. - A Eur. J.* **2010**, *16* (6), 1744–1749. https://doi.org/10.1002/chem.200903194.

(52) Grigorenko, B. L.; Polyakov, I. V.; Krylov, A. I.; Nemukhin, A. V. Computational Modeling Reveals the Mechanism of Fluorescent State Recovery in the Reversibly Photoswitchable Protein Dreiklang. *J. Phys. Chem. B* **2019**, *123* (42), 8901–8909. https://doi.org/10.1021/acs.jpcb.9b06988.

(53) Grigorenko, B. L.; Krylov, A. I.; Nemukhin, A. V. Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *J. Am. Chem. Soc.* **2017**, *139* (30), 10239–10249. https://doi.org/10.1021/jacs.7b00676.

(54) Grigorenko, B. L.; Nemukhin, A. V; Polyakov, I. V; Khrenova, M. G.; Krylov, A. I. A Light-Induced Reaction with Oxygen Leads to Chromophore Decomposition and Irreversible Photobleaching in GFP-Type Proteins. *J. Phys. Chem. B* **2015**, *119* (17), 5444–5452. https://doi.org/10.1021/acs.jpcb.5b02271.

(55) Khrenova, M. G.; Kulakova, A. M.; Nemukhin, A. V. Competition between Two Cysteines in Covalent Binding of Biliverdin to Phytochrome Domains. *Org. Biomol. Chem.* **2018**, *16* (40), 7518–7529. https://doi.org/10.1039/C8OB02262C.

(56) Alexov, E.; Mehler, E. L.; Baker, N.; M. Baptista, A.; Huang, Y.; Milletti, F.; Erik

Nielsen, J.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. Progress in the Prediction of PKa Values in Proteins. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (12), 3260–3275. https://doi.org/10.1002/prot.23189.

(57) Koumanov, A.; Rüterjans, H.; Karshikoff, A. Continuum Electrostatic Analysis of Irregular Ionization and Proton Allocation in Proteins. *Proteins Struct. Funct. Bioinforma.* **2002**, *46* (1), 85–96. https://doi.org/10.1002/prot.10034.

(58) Harris, T. K.; Turner, G. J. Structural Basis of Perturbed PKa Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life (International Union Biochem. Mol. Biol. Life)* **2002**, *53* (2), 85–98. https://doi.org/10.1080/15216540211468.

(59) Yan, E. C. Y.; Kazmi, M. A.; Ganim, Z.; Hou, J. M.; Pan, D.; Chang, B. S. W.; Sakmar, T. P.; Mathies, R. A. Retinal Counterion Switch in the Photoactivation of the G Protein-Coupled Receptor Rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (16), 9262–9267. https://doi.org/10.1073/pnas.1531970100.

(60) Gil, A. A.; Laptenok, S. P.; Iuliano, J. N.; Lukacs, A.; Verma, A.; Hall, C. R.; Yoon, G. E.; Brust, R.; Greetham, G. M.; Towrie, M.; French, J. B.; Meech, S. R.; Tonge, P. J. Photoactivation of the BLUF Protein PixD Probed by the Site-Specific Incorporation of Fluorotyrosine Residues. *J. Am. Chem. Soc.* **2017**, *139* (41), 14638–14648. https://doi.org/10.1021/jacs.7b07849.

(61) Borg, O. A.; Durbeej, B. Relative Ground and Excited-State PKa Values of Phytochromobilin in the Photoactivation of Photochrome: A Computational Study. *J.*

Phys. Chem. B **2007**, *111* (39), 11554–11565. https://doi.org/10.1021/jp0727953.

(62) Modi, V.; Donnini, S.; Groenhof, G.; Morozov, D. Protonation of the Biliverdin IXα Chromophore in the Red and Far-Red Photoactive States of a Bacteriophytochrome. **2019**, *123* (10), 2325–2334. https://doi.org/10.1021/acs.jpcb.9b01117.

(63) Mroginski, M. A.; Von Stetten, D.; Escobar, F. V.; Strauss, H. M.; Kaminski, S.; Scheerer, P.; Günther, M.; Murgida, D. H.; Schmieder, P.; Bongards, C.; Gärtner, W.; Mailliet, J.; Hughes, J.; Essen, L. O.; Hildebrandt, P. Chromophore Structure of Cyanobacterial Phytochrome Cph1 in the Pr State: Reconciling Structural and Spectroscopic Data by QM/MM Calculations. *Biophys. J.* **2009**, *96* (10), 4153–4163. https://doi.org/10.1016/j.bpj.2009.02.029.

(64) Tanford, C.; Kirkwood, J. G.; Tanford, C. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. *J. Am. Chem. Soc.* **1957**, *79* (20), 5333–5339. https://doi.org/10.1021/ja01577a001.

(65) Meyer, T.; Knapp, E. W. PKa Values in Proteins Determined by Electrostatics Applied to Molecular Dynamics Trajectories. *J. Chem. Theory Comput.* **2015**, *11* (6), 2827–2840. https://doi.org/10.1021/acs.jctc.5b00123.

(66) Zienicke, B.; Molina, I.; Glenz, R.; Singer, P.; Ehmer, D.; Escobar, F. V.; Hildebrandt, P.; Diller, R.; Lamparter, T. Unusual Spectral Properties of Bacteriophytochrome Agp2 Result from a Deprotonation of the Chromophore in the Red-Absorbing Form Pr. *J. Biol. Chem.* **2013**, *288* (44), 31738–31751. https://doi.org/10.1074/jbc.M113.479535.

(67) Velazquez Escobar, F.; Piwowarski, P.; Salewski, J.; Michael, N.; Fernandez Lopez,

M.; Rupp, A.; Muhammad Qureshi, B.; Scheerer, P.; Bartl, F.; Frankenberg-Dinkel, N.; Siebert, F.; Andrea Mroginski, M.; Hildebrandt, P. A Protonation-Coupled Feedback Mechanism Controls the Signalling Process in Bathy Phytochromes. *Nat. Chem.* / **2015**, *7*, 423. https://doi.org/10.1038/NCHEM.2225.

(68) Kraskov, A.; Nguyen, A. D.; Goerling, J.; Buhrke, D.; Velazquez Escobar, F.; Fernandez Lopez, M.; Michael, N.; Sauthof, L.; Schmidt, A.; Piwowarski, P.; Yang, Y.; Stensitzki, T.; Adam, S.; Bartl, F.; Schapiro, I.; Heyne, K.; Siebert, F.; Scheerer, P.; Mroginski, M. A.; Hildebrandt, P. Intramolecular Proton Transfer Controls Protein Structural Changes in Phytochrome. *Biochemistry* **2020**, *59* (9), 1023–1037. https://doi.org/10.1021/acs.biochem.0c00053.

(69) Borucki, B.; Von Stetten, D.; Seibeck, S.; Lamparter, T.; Michael, N.; Mroginski, M. A.; Otto, H.; Murgida, D. H.; Heyn, M. P.; Hildebrandt, P. Light-Induced Proton Release of Phytochrome Is Coupled to the Transient Deprotonation of the Tetrapyrrole Chromophore. *J. Biol. Chem.* **2005**, *280* (40), 34358–34364. https://doi.org/10.1074/jbc.M505493200.

(70) Woelke, A. L.; Galstyan, G.; Galstyan, A.; Meyer, T.; Heberle, J.; Knapp, E. W. Exploring the Possible Role of Glu286 in CcO by Electrostatic Energy Computations Combined with Molecular Dynamics. *J. Phys. Chem. B* **2013**, *117* (41), 12432–12441. https://doi.org/10.1021/jp407250d.

(71) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain X1 and X2 Dihedral Angles.

J. Chem. Theory Comput. **2012**, *8* (9), 3257–3273. https://doi.org/10.1021/ct300400x.

(72) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85. https://doi.org/10.1016/S0065-3233(03)66002-X.

(73) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25* (13), 1656–1676. https://doi.org/10.1002/jcc.20090.

(74) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280. https://doi.org/10.1021/j100142a004.

(75) Nielsen, J. E.; Vriend, G. Optimizing the Hydrogen-Bond Network in Poisson-Boltzmann Equation-Based PKa Calculations. *Proteins Struct. Funct. Genet.* **2001**, *43* (4), 403–412. https://doi.org/10.1002/prot.1053.

(76) Schmidt, A.; Sauthof, L.; Szczepek, M.; Lopez, M. F.; Escobar, F. V.; Qureshi, B. M.; Michael, N.; Buhrke, D.; Stevens, T.; Kwiatkowski, D.; von Stetten, D.; Mroginski, M. A.; Krauß, N.; Lamparter, T.; Hildebrandt, P.; Scheerer, P. Structural Snapshot of a Bacterial Phytochrome in Its Functional Intermediate State. *Nat. Commun.* **2018**, *9* (1), 1–13. https://doi.org/10.1038/s41467-018-07392-7.

(77) Mongan, J.; Case, D. A.; McCammon, J. A. Constant PH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25* (16), 2038–2048.

https://doi.org/10.1002/jcc.20139.

(78) Escobar, F. V.; Lang, C.; Takiden, A.; Schneider, C.; Balke, J.; Hughes, J.; Alexiev, U.; Hildebrandt, P.; Mroginski, M. A. Protonation-Dependent Structural Heterogeneity in the Chromophore Binding Site of Cyanobacterial Phytochrome Cph1. *J. Phys. Chem. B* **2017**, *121* (1), 47–57. https://doi.org/10.1021/acs.jpcb.6b09600.

(79) Takiden, A.; Velazquez-Escobar, F.; Dragelj, J.; Woelke, A. L.; Knapp, E. W.; Piwowarski, P.; Bart, F.; Hildebrandt, P.; Mroginski, M. A. Structural and Vibrational Characterization of the Chromophore Binding Site of Bacterial Phytochrome Agp1. *Photochem. Photobiol.* **2017**, *93* (3), 713–723. https://doi.org/10.1111/php.12737.

(80) Falklöf, O.; Durbeej, B. Modeling of Phytochrome Absorption Spectra. *J. Comput. Chem.* **2013**, *34* (16), 1363–1374. https://doi.org/10.1002/jcc.23265.

(81) Meyer, T.; Kieseritzky, G.; Knapp, E. W. Electrostatic PK a Computations in Proteins: Role of Internal Cavities. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (12), 3320–3332. https://doi.org/10.1002/prot.23092.

(82) Rockwell, N. C.; Lagarias, J. C. A Brief History of Phytochromes. *ChemPhysChem* **2010**, *11* (6), 1172–1180. https://doi.org/10.1002/cphc.200900894.

(83) Xu, X.; Höppner, A.; Wiebeler, C.; Zhao, K.-H.; Schapiro, I.; Gärtner, W. Structural Elements Regulating the Photochromicity in a Cyanobacteriochrome. *Proc. Natl. Acad. Sci.* **2020**, *117* (5), 2432–2440. https://doi.org/10.1073/pnas.1910208117/-/DCSupplemental.

(84) Xu, X.; Port, A.; Wiebeler, C.; Zhao, K.-H.; Schapiro, I.; Gärtner, W. Phycocyanobilin Chromophore Distortion as Origin of Color Tuning in Red/Green Cyanobacteriochrome. unpublished.

(85) Slavov, C.; Xu, X.; Zhao, K. H.; Gärtner, W.; Wachtveitl, J. Detailed Insight into the Ultrafast Photoconversion of the Cyanobacteriochrome Slr1393 from Synechocystis Sp. *Biochim. Biophys. Acta - Bioenerg.* **2015**, *1847* (10), 1335–1344. https://doi.org/10.1016/j.bbabio.2015.07.013.

(86) Wiebeler, C.; Schapiro, I. QM/MM Benchmarking of Cyanobacteriochrome Slr1393g3 Absorption Spectra. *Molecules* **2019**, *24* (9), 1720. https://doi.org/10.3390/molecules24091720.

(87) Wiebeler, C.; Gopalakrishna Rao, A.; Gärtner, W.; Schapiro, I. The Effective Conjugation Length Is Responsible for the Red/Green Spectral Tuning in the Cyanobacteriochrome Slr1393g3. *Angew. Chemie Int. Ed.* **2019**, *58* (7), 1934–1938. https://doi.org/10.1002/anie.201810266.

(88) Lim, S.; Yu, Q.; Gottlieb, S. M.; Chang, C.-W.; Rockwell, N. C.; Martin, S. S.; Madsen, D.; Lagarias, J. C.; Larsen, D. S.; Ames, J. B. Correlating Structural and Photochemical Heterogeneity in Cyanobacteriochrome NpR6012g4. *Proc. Natl. Acad. Sci.* **2018**, *115* (17), 4387–4392. https://doi.org/10.1073/pnas.1720682115.

(89) Bombarda, E.; Ullmann, G. M. PH-Dependent PKa Values in Proteins-a Theoretical Analysis of Protonation Energies with Practical Consequences for Enzymatic Reactions. *J. Phys. Chem. B* **2010**, *114* (5), 1994–2003. https://doi.org/10.1021/jp908926w.

(90) Olsson, M. H. M.; SØndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537. https://doi.org/10.1021/ct100578z.

(91) Wang, L.; Li, L.; Alexov, E. PKa Predictions for Proteins, RNAs, and DNAs with the Gaussian Dielectric Function Using DelPhi PKa. *Proteins Struct. Funct. Bioinforma.* **2015**, *83* (12), 2186–2197. https://doi.org/10.1002/prot.24935.

(92) Pieri, E.; Ledentu, V.; Huix-Rotllant, M.; Ferré, N. Sampling the Protonation States: The PH-Dependent UV Absorption Spectrum of a Polypeptide Dyad. *Phys. Chem. Chem. Phys.* **2018**, *20* (36), 23252–23261. https://doi.org/10.1039/c8cp03557a.

(93) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant PH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10* (3), 1341–1352. https://doi.org/10.1021/ct401042b.

(94) Pieri, E.; Ledentu, V.; Sahlin, M.; Dehez, F.; Olivucci, M.; Ferré, N. CpHMD-Then-QM/MM Identification of the Amino Acids Responsible for the Anabaena Sensory Rhodopsin PH-Dependent Electronic Absorption Spectrum. *J. Chem. Theory Comput.* **2019**, *15* (8), 4535–4546. https://doi.org/10.1021/acs.jctc.9b00221.

(95) Tahara, S.; Kato, Y.; Kandori, H.; Ohtani, H. PH-Dependent Photoreaction Pathway of the All-Trans Form of Anabaena Sensory Rhodopsin. *J. Phys. Chem. B* **2013**, *117* (7), 2053–2060. https://doi.org/10.1021/jp311217a.

(96) Rozin, R.; Wand, A.; Jung, K. H.; Ruhman, S.; Sheves, M. PH Dependence of

Anabaena Sensory Rhodopsin: Retinal Isomer Composition, Rate of Dark Adaptation, and Photochemistry. *J. Phys. Chem. B* **2014**, *118* (30), 8995–9006. https://doi.org/10.1021/jp504688y.

(97) Ischtwan, J.; Collins, M. A. Molecular Potential Energy Surfaces by Interpolation. *J. Chem. Phys.* **1994**, *100* (11), 8080–8088. https://doi.org/10.1063/1.466801.

(98) Park, J. W.; Rhee, Y. M. Electric Field Keeps Chromophore Planar and Produces High Yield Fluorescence in Green Fluorescent Protein. *J. Am. Chem. Soc.* **2016**, *138* (41), 13619–13629. https://doi.org/10.1021/jacs.6b06833.

(99) Cho, K. H.; Chung, S.; Rhee, Y. M. Efficiently Transplanting Potential Energy Interpolation Database between Two Systems: Bacteriochlorophyll Case with FMO and LH2 Complexes. *J. Chem. Inf. Model.* **2019**. https://doi.org/10.1021/acs.jcim.9b00472.

(100) Kouyama, T.; Nishikawa, T.; Tokuhisa, T.; Okumura, H. Crystal Structure of the L Intermediate of Bacteriorhodopsin: Evidence for Vertical Translocation of a Water Molecule during the Proton Pumping Cycle. *J. Mol. Biol.* **2004**, *335*, 531–546. https://doi.org/10.1016/j.jmb.2003.10.068.

(101) Jardón-Valadez, E.; Bondar, A. N.; Tobias, D. J. Coupling of Retinal, Protein, and Water Dynamics in Squid Rhodopsin. *Biophys. J.* **2010**, *99* (7), 2200–2207. https://doi.org/10.1016/j.bpj.2010.06.067.

(102) Wolter, T.; Elstner, M.; Fischer, S.; C. Smith, J.; Bondar, A.-N. Mechanism by Which Untwisting of Retinal Leads to Productive Bacteriorhodopsin Photocycle States. *J. Phys. Chem. B* **2015**, *119* (6), 2229–2240. https://doi.org/10.1021/jp505818r.

(103) Guerra, F.; Siemers, M.; Mielack, C.; Bondar, A.-N. Dynamics of Long-Distance Hydrogen-Bond Networks in Photosystem II. *J. Phys. Chem. B* **2018**, *122* (17), 4625–4641. https://doi.org/10.1021/acs.jpcb.8b00649.

(104) Siemers, M.; Lazaratos, M.; Karathanou, K.; Guerra, F.; S. Brown, L.; Bondar, A.-N. Bridge: A Graph-Based Algorithm to Analyze Dynamic H-Bond Networks in Membrane Proteins. *J. Chem. Theory Comput.* **2019**, *15* (12), 6781–6798. https://doi.org/10.1021/acs.jctc.9b00697.

(105) Karathanou, K.; Bondar, A.-N. Using Graphs of Dynamic Hydrogen-Bond Networks To Dissect Conformational Coupling in a Protein Motor. *J. Chem. Inf. Model.* **2019**, *59* (5), 1882–1896. https://doi.org/10.1021/acs.jcim.8b00979.

(106) Bondar, A. N.; Dau, H. Extended Protein/Water H-Bond Networks in Photosynthetic Water Oxidation. *Biochim. Biophys. Acta - Bioenerg.* **2012**, *1817*, 1177–1190. https://doi.org/10.1016/j.bbabio.2012.03.031.

(107) Kemmler, L.; Ibrahim, M.; Dobbek, H.; Zouni, A.; Bondar, A.-N. Dynamic Water Bridging and Proton Transfer at a Surface Carboxylate Cluster of Photosystem II. *Phys. Chem. Chem. Phys.* **2019**, *21* (45), 25449–25466. https://doi.org/10.1039/C9CP03926K.

(108) Onsson, H. J.; Mills, G.; Jacobsen, K. W.; Jónsson, H.; Mills, G.; Jacobsen, K. W. *Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions*; Berne, B. J., Ciccotti, G., Coker, D. F., Eds.; World Scientific, 1998; pp 385–404.

(109) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. **2006**,

125 (2). https://doi.org/10.1063/1.2212942.

(110) Díaz Leines, G.; Ensing, B. Path Finding on High-Dimensional Free Energy Landscapes. *Phys. Rev. Lett.* **2012**, *109* (2), 020601. https://doi.org/10.1103/PhysRevLett.109.020601.

(111) Pérez De Alba Ortíz, A.; Tiwari, A.; Puthenkalathil, R. C. C.; Ensing, B. Advances in Enhanced Sampling along Adaptive Paths of Collective Variables. **2018**, *149* (7). https://doi.org/10.1063/1.5027392.

(112) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199. https://doi.org/10.1016/0021-9991(77)90121-8.

(113) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* **2002**, *99* (20), 12562–12566. https://doi.org/10.1073/pnas.202427399.

(114) Pérez de Alba Ortíz, A.; Vreede, J.; Ensing, B. The Adaptive Path Collective Variable: A Versatile Biasing Approach to Compute the Average Transition Path and Free Energy of Molecular Transitions; 2019; pp 255–290. https://doi.org/10.1007/978-1-4939-9608-7_11.

(115) Branduardi, D.; Gervasio, F. L. L.; Parrinello, M. From A to B in Free Energy Space. **2007**, *126* (5). https://doi.org/10.1063/1.2432340.

(116) Müller, K.; Brown, L. D. Location of Saddle Points and Minimum Energy Paths by a Constrained Simplex Optimization Procedure. *Theor. Chim. Acta* **1979**, *53* (1), 75–93.

https://doi.org/10.1007/BF00547608.

(117) Goyal, P.; Hammes-Schiffer, S. Role of Active Site Conformational Changes in Photocycle Activation of the AppA BLUF Photoreceptor. **2017**, *114* (7), 1480–1485. https://doi.org/10.1073/pnas.1621393114.

(118) Goings, J. J. J.; Reinhardt, C. R. R.; Hammes-Schiffer, S. Propensity for Proton Relay and Electrostatic Impact of Protein Reorganization in Slr1694 BLUF Photoreceptor. **2018**, *140* (45), 15241–15251. https://doi.org/10.1021/jacs.8b07456.

(119) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem* **2002**, *53*, 291–318. https://doi.org/10.1146/annurev.physchem.53.082301.113146.

(120) Vreede, J.; Juraszek, J.; Bolhuis, P. G. Predicting the Reaction Coordinates of Millisecond Light-Induced Conformational Changes in Photoactive Yellow Protein. *Proc. Natl. Acad. Sci.* **2010**, *107* (6), 2397–2402. https://doi.org/10.1073/pnas.0908754107.

(121) Gozem, S.; Luk, H. L.; Schapiro, I.; Olivucci, M. Theory and Simulation of the Ultrafast Double-Bond Isomerization of Biological Chromophores. *Chem. Rev.* **2017**, *117* (22), 13502–13565. https://doi.org/10.1021/acs.chemrev.7b00177.

(122) Tiwary, P.; Berne, B. J. J. Spectral Gap Optimization of Order Parameters for Sampling Complex Molecular Systems. **2016**, *113* (11), 2839–2844. https://doi.org/10.1073/pnas.1600917113.

(123) M. Sultan, M.; Pande, V. S. TICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput.* **2017**, *13* (6), 2440–2447. https://doi.org/10.1021/acs.jctc.7b00182.

(124) Chiavazzo, E.; Covino, R.; Coifman, R. R. R.; Gear, C. W. W.; Georgiou, A. S. S.; Hummer, G.; Kevrekidis, I. G. G. Intrinsic Map Dynamics Exploration for Uncharted Effective Free-Energy Landscapes. **2017**, *114* (28), E5494–E5503. https://doi.org/10.1073/pnas.1621481114.

(125) Chen, W.; Tan, A. R. R.; Ferguson, A. L. L. Collective Variable Discovery and Enhanced Sampling Using Autoencoders: Innovations in Network Architecture and Error Function Design. **2018**, *149* (7). https://doi.org/10.1063/1.5023804.

(126) Wehmeyer, C.; Noé, F. Time-Lagged Autoencoders: Deep Learning of Slow Collective Variables for Molecular Kinetics. **2018**, *148* (24). https://doi.org/10.1063/1.5011399.

(127) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann Generators: Sampling Equilibrium States of Many-Body Systems with Deep Learning. *Science (80-.).* **2019**, *365* (6457). https://doi.org/10.1126/science.aaw1147.

(128) Wesolowski, T. A.; Warshel, A. Frozen Density Functional Approach for Ab Initio Calculations of Solvated Molecules. *J. Phys. Chem.* **1993**, *97* (30), 8050–8053. https://doi.org/10.1021/j100132a040.

(129) Wesołowski, T. A. Embedding a Multideterminantal Wave Function in an Orbital-Free Environment. *Phys. Rev. A - At. Mol. Opt. Phys.* **2008**, *77* (1), 1–9. https://doi.org/10.1103/PhysRevA.77.012504.

(130) Zech, A.; Aquilante, F.; Wesolowski, T. A. Orthogonality of Embedded Wave Functions for Different States in Frozen-Density Embedding Theory. *J. Chem. Phys.* **2015**, *143* (16). https://doi.org/10.1063/1.4933372.

(131) Zech, A.; Dreuw, A.; Wesolowski, T. A. Extension of Frozen-Density Embedding Theory for Non-Variational Embedded Wavefunctions. *J. Chem. Phys.* **2019**, *150* (12). https://doi.org/10.1063/1.5089233.

(132) Fradelos, G.; Kaminski, J. W.; Wesolowski, T. A.; Leutwyler, S. Cooperative Effect of Hydrogen-Bonded Chains in the Environment of a Π→π Chromophore. *J. Phys. Chem. A* **2009**, *113* (36), 9766–9771. https://doi.org/10.1021/jp906483z.

(133) Fradelos, G.; Lutz, J. J.; Wesołowski, T. A.; Piecuch, P.; Włoch, M. Embedding vs Supermolecular Strategies in Evaluating the Hydrogen-Bonding-Induced Shifts of Excitation Energies. *J. Chem. Theory Comput.* **2011**, *7* (6), 1647–1666. https://doi.org/10.1021/ct200101x.

(134) Ricardi, N.; Zech, A.; Gimbal-Zofka, Y.; Wesolowski, T. A. Explicit vs. Implicit Electronic Polarisation of Environment of an Embedded Chromophore in Frozen-Density Embedding Theory. *Phys. Chem. Chem. Phys.* **2018**, *20* (41), 26053–26062. https://doi.org/10.1039/c8cp05634j.

(135) Kaminski, J. W.; Gusarov, S.; Wesolowski, T. a; Kovalenko, A. Modeling Solvatochromic Shifts Using the Orbital-Free Embedding Potential at Statistically Mechanically Averaged Solvent Density. *J. Phys. Chem. A* **2010**, *114* (i), 6082–6096. https://doi.org/10.1021/jp100158h.

(136) Laktionov, A.; Chemineau-Chalaye, E.; Wesolowski, T. A. Frozen-Density Embedding Theory with Average Solvent Charge Densities from Explicit Atomistic Simulations. *Phys. Chem. Chem. Phys.* **2016**, *18* (31), 21069–21078. https://doi.org/10.1039/c6cp00497k.

(137) Goez, A.; Jacob, C. R.; Neugebauer, J. Modeling Environment Effects on Pigment Site Energies: Frozen Density Embedding with Fully Quantum-Chemical Protein Densities. *Comput. Theor. Chem.* **2014**, *1040–1041*, 347–359. https://doi.org/10.1016/j.comptc.2014.02.009.

(138) Daday, C.; König, C.; Neugebauer, J.; Filippi, C. Wavefunction in Density Functional Theory Embedding for Excited States: Which Wavefunctions, Which Densities? *ChemPhysChem* **2014**, *15* (15), 3205–3217. https://doi.org/10.1002/cphc.201402459.

(139) Zech, A.; Ricardi, N.; Prager, S.; Dreuw, A.; Wesolowski, T. A. Benchmark of Excitation Energy Shifts from Frozen-Density Embedding Theory: Introduction of a Density-Overlap-Based Applicability Threshold. *J. Chem. Theory Comput.* **2018**, *14* (8), 4028–4040. https://doi.org/10.1021/acs.jctc.8b00201.

(140) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6* (12), 3721–3734. https://doi.org/10.1021/ct1003803.

(141) Steinmann, C.; Reinholdt, P.; Nørby, M. S.; Kongsted, J.; Olsen, J. M. H. Response Properties of Embedded Molecules through the Polarizable Embedding Model. *Int. J. Quantum Chem.* **2019**, *119* (1), 1–20. https://doi.org/10.1002/qua.25717.

(142) List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited States in Large Molecular Systems through Polarizable Embedding. *Phys. Chem. Chem. Phys.* **2016**, *18* (30), 20234–20250. https://doi.org/10.1039/c6cp03834d.

(143) Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Protein-Molecule Interaction Energy. *J. Chem. Phys.* **2003**, *119* (7), 3599–3605. https://doi.org/10.1063/1.1591727.

(144) Olsen, J. M. H.; List, N. H.; Kristensen, K.; Kongsted, J. Accuracy of Protein Embedding Potentials: An Analysis in Terms of Electrostatic Potentials. *J. Chem. Theory Comput.* **2015**, *11* (4), 1832–1842. https://doi.org/10.1021/acs.jctc.5b00078.

(145) Hršak, D.; Olsen, J. M. H.; Kongsted, J. Polarizable Density Embedding Coupled Cluster Method. *J. Chem. Theory Comput.* **2018**, *14* (3), 1351–1360. https://doi.org/10.1021/acs.jctc.7b01153.

(146) Reinholdt, P.; Kongsted, J.; Olsen, J. M. H. Polarizable Density Embedding: A Solution to the Electron Spill-Out Problem in Multiscale Modeling. *J. Phys. Chem. Lett.* **2017**, *8* (23), 5949–5958. https://doi.org/10.1021/acs.jpclett.7b02788.

(147) Olsen, J. M. H.; Steinmann, C.; Ruud, K.; Kongsted, J. Polarizable Density Embedding: A New QM/QM/MM-Based Computational Strategy. *J. Phys. Chem. A* **2015**, *119* (21), 5344–5355. https://doi.org/10.1021/jp510138k.

(148) Ajith Karunarathne, W. K.; O'Neill, P. R.; Gautam, N. Subcellular Optogenetics - Controlling Signaling and Single-Cell Behavior. *J. Cell Sci.* **2015**, *128* (1), 15–25. https://doi.org/10.1242/jcs.154435.

(149) El-Shamayleh, Y.; Horwitz, G. D. Primate Optogenetics: Progress and Prognosis. In *Proceedings of the National Academy of Sciences of the United States of America*; National Academy of Sciences, 2019; Vol. 116, pp 26195–26203. https://doi.org/10.1073/pnas.1902284116.

(150) Adamantidis, A.; Arber, S.; Bains, J. S.; Bamberg, E.; Bonci, A.; Buzsáki, G.; Cardin, J. A.; Costa, R. M.; Dan, Y.; Goda, Y.; Graybiel, A. M.; Häusser, M.; Hegemann, P.; Huguenard, J. R.; Insel, T. R.; Janak, P. H.; Johnston, D.; Josselyn, S. A.; Koch, C.; Kreitzer, A. C.; Lüscher, C.; Malenka, R. C.; Miesenböck, G.; Nagel, G.; Roska, B.; Schnitzer, M. J.; Shenoy, K. V.; Soltesz, I.; Sternson, S. M.; Tsien, R. W.; Tsien, R. Y.; Turrigiano, G. G.; Tye, K. M.; Wilson, R. I. Optogenetics: 10 Years after ChR2 in Neurons - Views from the Community. *Nature Neuroscience*. Nature Publishing Group August 26, 2015, pp 1202–1212. https://doi.org/10.1038/nn.4106.

(151) Kandori, H. Retinal Proteins: Photochemistry and Optogenetics. *Bull. Chem. Soc. Jpn.* **2020**, *93* (1), 76–85. https://doi.org/10.1246/bcsj.20190292.

(152) Schmidt, D.; Cho, Y. K. Natural Photoreceptors and Their Application to Synthetic Biology. *Trends Biotechnol.* **2015**, *33* (2), 80–91. https://doi.org/10.1016/j.tibtech.2014.10.007.

(153) Fdez. Galván, I.; Vacher, M.; Alavi, A.; Angeli, C.; Aquilante, F.; Autschbach, J.; Bao, J. J.; Bokarev, S. I.; Bogdanov, N. A.; Carlson, R. K.; Chibotaru, L. F.; Creutzberg, J.; Dattani, N.; Delcey, M. G.; Dong, S. S.; Dreuw, A.; Freitag, L.; Frutos, L. M.; Gagliardi, L.; Gendron, F.; Giussani, A.; González, L.; Grell, G.; Guo, M.; Hoyer, C. E.; Johansson, M.; Keller, S.; Knecht, S.; Kovačević, G.; Källman, E.; Li Manni, G.;

Lundberg, M.; Ma, Y.; Mai, S.; Malhado, J. P.; Malmqvist, P. Å.; Marquetand, P.; Mewes, S. A.; Norell, J.; Olivucci, M.; Oppel, M.; Phung, Q. M.; Pierloot, K.; Plasser, F.; Reiher, M.; Sand, A. M.; Schapiro, I.; Sharma, P.; Stein, C. J.; Sørensen, L. K.; Truhlar, D. G.; Ugandi, M.; Ungur, L.; Valentini, A.; Vancoillie, S.; Veryazov, V.; Weser, O.; Wesołowski, T. A.; Widmark, P.-O.; Wouters, S.; Zech, A.; Zobel, J. P.; Lindh, R. OpenMolcas: From Source Code to Insight. *J. Chem. Theory Comput.* **2019**, *15* (11), 5925–5964. https://doi.org/10.1021/acs.jctc.9b00532.

(154) Melaccio, F.; Del Carmen Marín, M.; Valentini, A.; Montisci, F.; Rinaldi, S.; Cherubini, M.; Yang, X.; Kato, Y.; Stenrup, M.; Orozco-Gonzalez, Y.; Ferré, N.; Luk, H. L.; Kandori, H.; Olivucci, M. Toward Automatic Rhodopsin Modeling as a Tool for High-Throughput Computational Photobiology. *J. Chem. Theory Comput.* **2016**, *12* (12), 6020–6034. https://doi.org/10.1021/acs.jctc.6b00367.

(155) Pedraza-González, L.; De Vico, L.; MarÍn, M. D. C.; Fanelli, F.; Olivucci, M. A-ARM: Automatic Rhodopsin Modeling with Chromophore Cavity Generation, Ionization State Selection, and External Counterion Placement. *J. Chem. Theory Comput.* **2019**, *15* (5), 3134–3152. https://doi.org/10.1021/acs.jctc.9b00061.

(156) Nilsson, D. E. Photoreceptor Evolution: Ancient Siblings Serve Different Tasks. *Current Biology*. Cell Press February 8, 2005, pp R94–R96. https://doi.org/10.1016/j.cub.2005.01.027.

(157) Gushchin, I.; Shevchenko, V.; Polovinkin, V.; Borshchevskiy, V.; Buslaev, P.; Bamberg, E.; Gordeliy, V. Structure of the Light-Driven Sodium Pump KR2 and Its Implications for Optogenetics. *FEBS J.* **2016**, *283* (7), 1232–1238.

https://doi.org/10.1111/febs.13585.

(158) Pedraza-González, L.; Marín, M. D. C.; Jorge, A. N.; Ruck, T. D.; Yang, X.; Valentini, A.; Olivucci, M.; De Vico, L. Web-ARM: A Web-Based Interface for the Automatic Construction of QM/MM Models of Rhodopsins. *J. Chem. Inf. Model.* **2020**. https://doi.org/10.1021/acs.jcim.9b00615.

(159) Schapiro, I.; Ryazantsev, M. N.; Frutos, L. M.; Ferré, N.; Lindh, R.; Olivucci, M. The Ultrafast Photoisomerizations of Rhodopsin and Bathorhodopsin Are Modulated by Bond Length Alternation and HOOP Driven Electronic Effects. *J. Am. Chem. Soc.* **2011**, *133* (10), 3354–3364. https://doi.org/10.1021/ja1056196.

(160) Tully, J. C. Molecular Dynamics with Electronic Transitions. *J. Chem. Phys.* **1990**, *93* (2), 1061–1071. https://doi.org/10.1063/1.459170.

(161) Granucci, G.; Persico, M. Critical Appraisal of the Fewest Switches Algorithm for Surface Hopping. *J. Chem. Phys.* **2007**, *126* (13). https://doi.org/10.1063/1.2715585.

(162) Gozem, S.; Schapiro, I.; Ferre, N.; Olivucci, M. The Molecular Mechanism of Thermal Noise in Rod Photoreceptors. *Science (80-.).* **2012**, *337* (6099), 1225–1228. https://doi.org/10.1126/science.1220461.

(163) Luk, H. L.; Bhattacharyya, N.; Montisci, F.; Morrow, J. M.; Melaccio, F.; Wada, A.; Sheves, M.; Fanelli, F.; Chang, B. S. W.; Olivucci, M. Modulation of Thermal Noise and Spectral Sensitivity in Lake Baikal Cottoid Fish Rhodopsins. *Sci. Rep.* **2016**, *6* (1), 1–9. https://doi.org/10.1038/srep38425.

(164) Korol, V.; Husen, P.; Sjulstok, E.; Nielsen, C.; Friis, I.; Frederiksen, A.; Salo, A. B.; Solov'Yov, I. A. Introducing VIKING: A Novel Online Platform for Multiscale Modeling. *ACS Omega* **2020**, *5* (2), 1254–1260. https://doi.org/10.1021/acsomega.9b03802.

(165) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. John Wiley and Sons Inc. 2005, pp 1781–1802. https://doi.org/10.1002/jcc.20289.

(166) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *Journal of Computational Chemistry*. December 2005, pp 1668–1688. https://doi.org/10.1002/jcc.20290.

(167) Solov'yov, I. A.; Yakubovich, A. V.; Nikolaev, P. V.; Volkovets, I.; Solov'Yov, A. V. MesoBioNano Explorer-A Universal Program for Multiscale Computer Simulations of Complex Molecular Structure and Dynamics. *J. Comput. Chem.* **2012**, *33* (30), 2412–2439. https://doi.org/10.1002/jcc.23086.

(168) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry*. December 2005, pp 1701–1718. https://doi.org/10.1002/jcc.20291.

(169) Frisch, M.; Trucks, G.; Schlegel, H.; … G. S.-T. is no corresponding; 2013, undefined. Gaussian 03, Revision C. 02; Gaussian, Inc.: Wallingford, CT, 2004.

(170) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14* (11), 1347–1363. https://doi.org/10.1002/jcc.540141112.

(171) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I. M.; Iozzi, M. F.; Jansík, B.; Jensen, H. J. A.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.; Packer, M. J.; Pawlowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V. V.; Sałek, P.; Samson, C. C. M.; de Merás, A. S.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. The Dalton Quantum Chemistry Program System. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (3), 269–284. https://doi.org/10.1002/wcms.1172.

(172) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (1), 73–78. https://doi.org/10.1002/wcms.81.

(173) Nielsen, C.; physics, I. S.-T. J. of chemical; 2019, undefined. MolSpin—Flexible and Extensible General Spin Dynamics Software. *aip.scitation.org*.