

An Average Solvent Electrostatic Configuration Protocol for QM/MM Free Energy Optimization: Implementation and Application to Rhodopsin Systems

Yoelvis Orozco-Gonzalez,^{*,†,‡,§,||} Madushanka Manathunga,[§] María del Carmen Marín,^{||} Damianos Agathangelou,[†] Kwang-Hwan Jung,[∇] Federico Melaccio,^{||} Nicolas Ferré,^{⊥,||} Stefan Haacke,[†] Kaline Coutinho,[#] Sylvio Canuto,^{*,#} and Massimo Olivucci^{*,†,‡,§,||}

[†]Université de Strasbourg-CNRS, UMR 7504, Institut de Physique et Chimie des Matériaux de Strasbourg, F-67034 Strasbourg, France

[‡]USIAS Institut d'Études Avancées, Université de Strasbourg, 5 allée du Général Rouvillois, F-67083 Strasbourg, France

[§]Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, United States

^{||}Department of Biotechnology, Chemistry e Pharmacy, Università di Siena, via A. Moro 2, I-53100 Siena, Italy

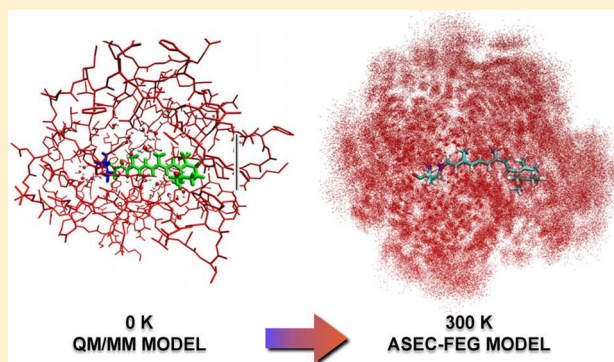
[⊥]Aix-Marseille Univ, CNRS, ICR, 13013 Marseille, France

[#]Instituto de Física, Universidade de São Paulo, 05508-090 Cidade Universitária, São Paulo/SP, Brazil

[∇]Department of Life Science and Institute of Biological Interfaces, Sogang University 35 Baekbeom-Ro, Mapo-Gu, Seoul 04107, Korea

Supporting Information

ABSTRACT: A novel atomistic methodology to perform free energy geometry optimization of a retinal chromophore covalently bound to any rhodopsin-like protein cavity is presented and benchmarked by computing the absorption maxima wavelengths (λ_{max}) of distant rhodopsin systems. The optimization is achieved by computing the Nagaoka's Free Energy Gradient (FEG) within an Average Solvent Electrostatic Configuration (ASEC) atomistic representation of the thermodynamic equilibrium and minimizing such quantity via an iterative procedure based on sequential classical MD and constrained QM/MM geometry optimization steps. The performance of such an ASEC-FEG protocol is assessed at the CASPT2//CASSCF/Amber level by reproducing the λ_{max} values observed for 12 mutants of redesigned human cellular retinol binding protein II (hCRBP2) systems; a set of 10 distant wild-type rhodopsins from vertebrates, invertebrates, eubacteria, and archaea organisms; and finally a set of 10 rhodopsin mutants from an eubacterial rhodopsin. The results clearly show that the proposed protocol, which can be easily extended to any protein incorporating a covalently bound ligand, yields correct λ_{max} trends with limited absolute errors.



computed by using *ab initio* calculations while the rest of the protein is described by a molecular mechanics force field, corresponding to the MM subsystem. These computational models have been employed not only for the investigation of the molecular mechanisms responsible for the λ_{max} variations but also for describing the chromophore relaxation along the electronically excited states (e.g., the associated transient fluorescence and photoreactivity), for predicting the structure and spectral features of the primary ground state photoproduct, for estimating the ground state barriers controlling the

1. INTRODUCTION

Due to their diversity, natural and artificial rhodopsins constitute ideal benchmark systems for the development of computer models of light-responsive supramolecular systems. These models must primarily be able to reproduce the variations of the λ_{max} values along different series of homologous systems before being employed in systematic computational studies of properties, which may be difficult to measure. Due to the wide set of observed λ_{max} values, the available structural information, and their biological/technological importance, several groups have focused on rhodopsins for assessing the quality of hybrid quantum mechanical/molecular mechanical (QM/MM) models.^{1–7} In such models, the retinal chromophore corresponds to the QM subsystem,

computed by using *ab initio* calculations while the rest of the protein is described by a molecular mechanics force field, corresponding to the MM subsystem. These computational models have been employed not only for the investigation of the molecular mechanisms responsible for the λ_{max} variations but also for describing the chromophore relaxation along the electronically excited states (e.g., the associated transient fluorescence and photoreactivity), for predicting the structure and spectral features of the primary ground state photoproduct, for estimating the ground state barriers controlling the

Received: August 12, 2017

Published: November 7, 2017



chromophore thermal isomerization^{1,3,8} and, recently, for computing the photoisomerization quantum yields.⁹

In order to properly understand the molecular-level origin of the λ_{max} tuning in rhodopsins, as well as in other light-responsive proteins, it is desirable to build QM/MM models providing a suitable description of the thermodynamic equilibrium conditions of the apoprotein and the external environment when needed. The description of such an equilibrium condition is commonly not properly accounted for, in which case, a single configuration of the chromophore environment is considered to obtain the final model, constructed via geometry optimization on the potential energy surface of the ground singlet electronic state (S_0). As a result, the final optimized model can be highly dependent on the selected initial configuration and may lead, for instance, to a higher potential energy minimum with side chain configurations not representative of the thermodynamic ensemble. Therefore, it is highly desirable to perform the geometry optimization on the free energy surface of the system. Aiming at such a target, Hayashi and Kosugi¹⁰ have recently developed a methodology called QM/MM reweighting free energy self-consistent field (QM/MM-RWFE-SCF), which allows the carrying out of QM/MM free energy optimizations. In QM/MM-RWFE-SCF, the QM subsystem is optimized at DFT level in the presence of an *average potential* created by considering a molecular dynamic (MD) sampling of the MM environment. A similar methodology, known as Average Solvent Electrostatic Potential/Molecular Dynamic (ASEP/MD), was previously proposed by Aguilar and co-workers^{11–13} and applied to the free energy optimization of a QM solute in an MM solvent environment.

In the present manuscript, we introduce a novel protocol for performing QM/MM free energy optimizations of protein chromophores. In order to do so, we capitalize on the idea of Herbert et al.^{13,14} of combining the Average Solvent Electrostatic Configuration (ASEC) model^{15,16} and the free energy gradient (FEG) method proposed by Nagaoka et al.^{12,17,18} to optimize the structure of a molecular system in solution. This protocol, named hereafter ASEC-FEG, is not oriented to the calculation of the *average potential* created by the MM environment but instead, the generation of an ensemble of configurations that, according to statistical thermodynamics, represents the time-averaged environment interaction with the QM subsystem. As discussed below, the possibility to directly link the interaction on the QM subsystem with the precise configurational (i.e., atomistic) description of the MM environment has several advantages.

As anticipated above, in order to benchmark the present implementation of the ASEC-FEG protocol, we specifically focus on the absorption spectroscopy of rhodopsin systems. In fact, the rhodopsin protein family represents a case of regulation of the maximum absorption wavelength (λ_{max}) playing a role in fundamental biological processes such as vision, chromatic adaptation, and ion pumping.¹⁹ In these proteins, the λ_{max} values of different stereoisomers of the same chromophore are modulated by the surrounding protein environment from 420 nm (human short wave-sensitive pigment, hSWS) to 587 nm (sensory rhodopsin I). Indeed, in all cases, the chromophore is formed by a retinylidene stereoisomer bound to a lysine residue via an iminium (also known as a protonated Schiff base, PSB) linkage.^{1,19} The detailed knowledge of the molecular mechanisms that control the spectral properties of rhodopsins is important not only for

understanding the functions of these photoreceptors but also for the rational design of novel genetically encodable tools. In fact, there is a growing interest in employing rhodopsins to switch “on” and “off” metabolic pathways, gene expression, and ion channels.^{20–23} Rhodopsins are also central to the field of optogenetics,²⁴ where DNA expressing rhodopsins functioning as light-gated ion-channels is delivered via a viral vector to a mammal brain cell. Most recently, the engineering of rhodopsin mimics based on a redesigned (mutated) human cellular retinol binding protein II (hCRBP II) capable of spontaneously linking the rhodopsin chromophore has led to the preparation in the laboratory of genetically encodable colorimetric probes with a λ_{max} change spanning a 425 to 644 nm range.²⁵ In section 2, we present the details of the ASEC-FEG protocol, while in section 3 we describe the actual computational implementation and the characteristics of the QM/MM models. Section 3 is concluded by describing the automatic code created that allows performance of the free energy geometry optimization and computation of spectroscopic properties of any rhodopsin-like protein. Finally, in section 4, we assess the quality of the ASEC-FEG models by investigating three different sets of proteins: (i) the whole set of reported hCRBP II rhodopsin mimics (we also provide a detailed rationalization of the effect of the mutations on the λ_{max} variation), (ii) a set of wild-type rhodopsins from distant organisms (vertebrates, invertebrates, eubacteria, and archaea), and (iii) a set of mutants of a sensory rhodopsin from the cyanobacterium *Anabaena* PCC 7120. For each group, the predicted *trend* of computed λ_{max} values is compared with the experimental data and used to evaluate the method performance.

2. THEORETICAL DETAILS

2.1. General Structure of the Model. In the ASEC-FEG protocol, a mathematical construct called the “ASEC configuration” corresponds to a set of superimposed configurations of the MM subsystem (the environment) surrounding the QM subsystem (the chromophore). More specifically, the ASEC configuration is built from a selected sampling of environment configurations, obtained via extensive MD, to mimic the effect of thermodynamic equilibrium conditions at the selected temperature. The ASEC configuration is then used to account for the average electrostatic and van der Waals interaction between the MM subsystem and the QM subsystem and also to carry out QM/MM geometry optimization of the QM subsystem. This methodology has four main advantages:

(i) The coupling of this methodology with the MOLCAS code²⁶ allows us to use several wave-function-based methods already implemented in such a quantum chemical package, like for instance, CASSCF,²⁷ Møller–Plesset perturbation theory,²⁸ CASPT2,²⁹ Coupled Cluster,³⁰ and also DFT,³¹ both to perform geometry optimizations and to evaluate spectroscopic properties. In addition, other tools can be used to search for transition states, minimum energy paths, and even surface crossings (e.g., conical intersections and singlet–triplet crossings) either in the ground or in the excited electronic states.

(ii) The configurations of the MM subsystem contributing to the ASEC configuration form a Boltzmann distribution since they are representative configurations of the system in thermodynamic equilibrium conditions. Therefore, these configurations can be used to generate initial conditions to simulate laser-induced excited state population dynamics or to reproduce the absorption or emission band of the molecule.

(iii) A detailed study of the specific interactions between side chains of the environment and the QM subsystem considering many configurations of the environment at the same time can be easily achieved (see section 3).

(iv) As mentioned in point ii, once the QM subsystem optimization is converged, the spectroscopic properties of the system can be computed directly from the ASEC configuration, but in addition, it is possible to extract from the ASEC configuration one specific configuration, or one specific group of configurations, closest to the average. Therefore, one can select a “structure” representative (in an approximated way) of the average environment, which can be used for successive analysis to derive chemical information.

Finally, we have to stress that the ASEC-FEG code has been coupled to the Automatic Rhodopsin Model (ARM) generator,⁴ to offer semiautomatic machinery to systematically generate the optimized ASEC-FEG model starting from an “experimental” structure (like X-ray crystallographic structure or a structure obtained via comparative modeling) with predefined ionization states as a template for building the model. As we will discuss in section 3, this greatly facilitates and speeds up the modeling of the systems. The code workflow of the implemented ASEC-FEG protocol is detailed in Figure S4 of the Supporting Information.

2.2. Theory. The fundamental equation of free energy perturbation theory is expressed in terms of the variations of the enthalpy of the system, H , as shown in the following equation:^{32,33}

$$G_B - G_A = \Delta G = -k_B T \ln \langle e^{-\Delta H/k_B T} \rangle_A \quad (1)$$

In this expression, k_B and T are the Boltzmann constant and temperature, respectively, while A and B represent two different, but *close*, configurations of the chromophore embedded in the binding pocket, being $\Delta H = H_B - H_A$. Note that large values of ΔH , due to very different configurations of the chromophore, would lead to an unreliable free energy difference calculation. As can be noted from eq 1, $\langle \dots \rangle_A$ indicates that the average of the ΔH in the exponential is computed in the ensemble of configurations generated for A . For the purpose of localizing a stationary structure on the free energy surface using free energy perturbation theory, small geometrical changes of the chromophore in the direction of the average energy gradient (or average forces) would lead to a minimum on the free energy surface. In that case, it is assumed that the distribution of microstates for two consecutive geometry optimization steps yields negligible environmental energy differences. Accordingly, only the interactions between the chromophore and environment vary. Therefore, we will use the free energy gradient method, proposed by Nagaoka et al.,^{12,17,18} for localizing stationary structures on the free energy surface of the systems.

Before starting the presentation of the theoretical details, it is necessary to define several terms that will be used throughout this work. In our QM/MM model, the entire protein can be divided into two subsystems: (i) the *QM-Lys subsystem*, which is defined by the quantum mechanical part of the protein (the retinal chromophore plus the C_ϵ atom of the chromophore-bound Lys side-chain; notice that this part constitutes the “usual” QM subsystem of QM/MM methods) and the rest of the same Lys side-chain described at the MM level, details shown in Figure 1, and (ii) the *QM-Lys environment*, defined by all the atoms of the rest of the protein, treated at the MM level. In the case of hCRBP II rhodopsin mimics, where the protein is

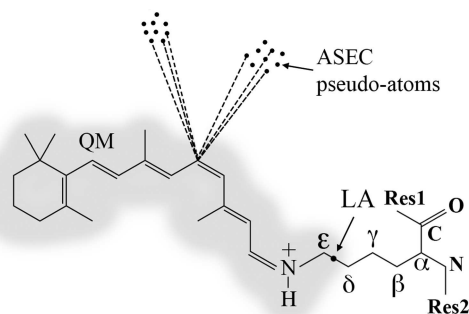


Figure 1. Schematic representation of the interactions between the atoms of the QM-Lys subsystem (represented by a chemical formula) and the pseudoatoms of the QM-Lys environment (represented by black dots). The electrostatic and van der Waals interactions are schematically represented by the dashed lines. The ordinary QM subsystem terminates with the link-hydrogen-atom (LA), and it is marked with a gray background. The backbone is represented by Res1 and Res2, indicating the other residues.

embedded into a solvent box, the solvent atoms are also considered part of the QM-Lys environment.

Let us now consider the QM-Lys subsystem at the thermodynamic equilibrium. Under such conditions, one wants to locate stationary structures (e.g., minima) of the QM-Lys subsystem on the free energy surface. As previously noted, we will use the free energy gradient method proposed by Nagaoka et al.^{12,17,18} In such a method, the average force acting on each atom of the QM-Lys subsystem is calculated using the free energy gradient, which is obtained through the following relationship:

$$F(q) = -\frac{\partial G(q)}{\partial q} = -\left\langle \frac{\partial V}{\partial q} \right\rangle \approx -\frac{\partial \langle V \rangle}{\partial q} \quad (2)$$

In this equation, q represents the nuclear coordinates of the QM-Lys subsystem, G is the free energy of the system, and V is the potential energy of the QM-Lys subsystem plus the interaction energy with the QM-Lys environment. These forces are equal to the time-averaged forces acting on each atom of the QM-Lys subsystem over the equilibrium distribution of the total system, as obtained from a molecular dynamics (MD) simulation of suitable length at the desired temperature.

In order to compute the potential energy of the system in eq 2, this is decomposed in the following way:

$$V = V_{QM} + V_{QM/MM} \quad (3)$$

The term V_{QM} represents the potential energy of the QM part computed using quantum mechanical methods. The term $V_{QM/MM}$ comprises several components:

$$V_{QM/MM} = V_{ele}(QM/MM) + V_{vdw}(QM/MM) + V_{bond}(QM/MM) \quad (4)$$

The first term, $V_{ele}(QM/MM)$, is the electrostatic interaction energy between the QM-Lys subsystem and the rest of the MM atoms of the system (i.e., protein plus solvent box). This term is considered in this work through the ElectroStatic Potential Fitted (ESPF) method, which includes some one-electron operators in the Hamiltonian for computing the QM/MM electrostatic interaction in a uniquely defined way, see details in refs 34 and 35. $V_{vdw}(QM/MM)$ represents the van der Waals interaction energy between the QM-Lys subsystem and the

MM atoms of the environment where van der Waals parameters are assigned both to the QM and to the MM atoms. The last term of eq 4 deals with some empirical bonded terms that should be used when the frontier between the QM and MM subsystems involves bonding interactions. The system to be optimized here is the QM-Lys subsystem, and therefore, the QM/MM frontier incorporating the link atom³⁶ needs to be properly described.

The statistical contribution of the QM-Lys subsystem to the total free energy of the system could be, in principle, accounted for by performing a straightforward QM/MM molecular dynamic where the geometry and wave function of the chromophore are computed at each step of the simulation, but this option is computationally unpractical, especially when large sampling of the environment is needed. This contribution could also be estimated by approximately computing the partition function of the vibrational motion using the harmonic approximation. In any case, large computational resources are needed, and we are more interested in finding the equilibrium structure of the QM-Lys subsystem rather than its vibrational contribution. Therefore, this is not considered in our calculations, but it is important to highlight that the entropy contribution, due to the interaction between the chromophore and the QM-Lys environment, is already considered in the average of $V_{\text{QM/MM}}$ along the MD.

A critical point for calculating the average total energy of eq 2 is to find an efficient way of calculating the average interactions energy between the atoms of the QM-Lys subsystem and the atoms of the QM-Lys environment. For this purpose, we use the ASEC model.¹⁵ Accordingly, we perform a sampling of representative configurations of the system via an extensive MD run and generate the mathematical construct mentioned above and called the ASEC configuration. This corresponds to a superposition of all uncorrelated configurations selected during the sampling and where a “cloud” of identical pseudoatoms replaces each MM atom of the environment. A visualization of the ASEC configuration for the case of a soluble hCRBPII rhodopsin mimic is given in Figure 2 (the pseudoatoms are represented by tiny spheres forming distinct but still localized clouds). A more schematic but general representation is given in Figure 1.

In our ASEC model, each pseudoatom has associated a $1/N$ fraction of the original point charge where N is the number of selected (sampled) MD configurations. In contrast with the original ASEC^{15,16} formulation, each pseudoatom also carries a fractional (scaled) van der Waals interaction. The procedure followed for scaling the van der Waals interactions is shown in the Supporting Information. It can be observed that the interaction energy (electrostatic and van der Waals) between the atoms of the QM-Lys subsystem and the atoms of the QM-Lys environment, calculated using the ASEC configuration, is exactly the same as the average interaction energy taken over all the individual sampled configurations.

The optimization protocol (which is a shell scripting based code) for finding a stationary structure of the Lys-QM molecular subsystem on the free energy surface is achieved by constructing the ASEC configuration and then by performing a full QM/MM geometry optimization of the QM-Lys subsystem using the quasi-Newton–Raphson method implemented in the MOLCAS-TINKER interface.^{26,35} These two steps, construction of the ASEC configuration and QM/MM geometry optimization in the field generated by the fixed ASEC configuration, are repeated recursively until the energy

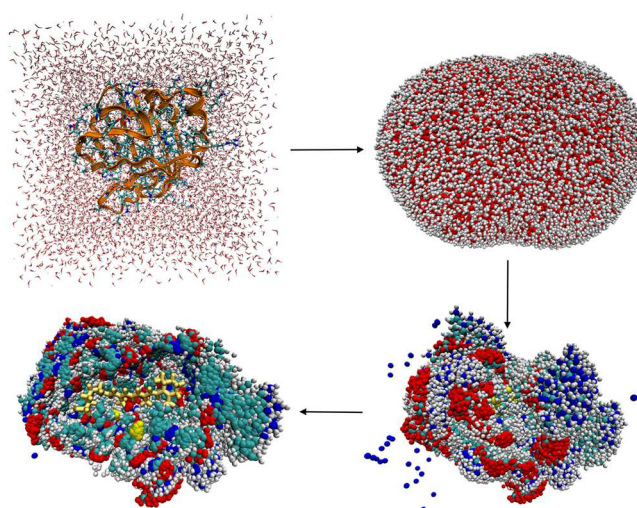


Figure 2. Representation of the ASEC model for a soluble rhodopsin mimic. Top-left, a snapshot of the 70 Å solvent box as obtained in a MD simulation using periodic boundary conditions. Top-right, the ASEC configuration formed by a superposition of 100 uncorrelated configurations of the QM-Lys environment selected along the MD run. A selected 30 Å shell, based on the minimum distance criterion from the chromophore, has been chosen to define the QM-Lys environment necessary to compute the spectroscopic properties and/or perform the geometry optimization of the QM-Lys subsystem. The tiny spheres represent the ASEC pseudoatoms of the entire QM-Lys environment (only the oxygen and hydrogen of the water molecules of the solvent are visible). Bottom-right, the same ASEC configuration of the top-right with the solvent molecules removed in order to expose the pseudoatoms of the protein. The blue spheres scattered around represent the ASEC representation of the ions added to neutralize the system. Bottom-left, the same ASEC configuration with the solvent and part of the protein residues removed to expose the chromophore (in yellow) or QM-Lys subsystem.

difference, relative to the previous step, is less than a defined threshold. In this way, the ASEC configuration representing the QM-Lys environment in a statistical fashion, is gradually adapting to the geometry and charge redistribution of the chromophore. On the other hand, the chromophore is relaxed in the recursively improved environment. A schematic representation of the iterative procedure is shown in Figure 3. The full QM/MM geometry optimizations performed in each step of the iterative cycles follow the standard thresholds of MOLCAS-TINKER. However, the convergence of the entire iterative procedure is determined by comparing the energy difference between two successive iterations, setting the convergence threshold at 0.5 kcal/mol. It is worth highlighting that (i) the MD step is carried out entirely at the MM level with a parametrized retinal chromophore (see computational details) and (ii) at every step of the iterative procedure a new reparametrization of the retinal charges is performed before starting the new MD. This reparametrization is performed by following the standard RESP ESP charge derived method,³⁷ in which case, a surrounding grid of points around the chromophore, reflecting the electrostatic potential, is used to fit a set of point charges in the atom positions of the chromophore.

3. COMPUTATIONAL DETAILS

As a benchmark set for testing the method performance in reproducing observed λ_{max} variations, we selected three sets of

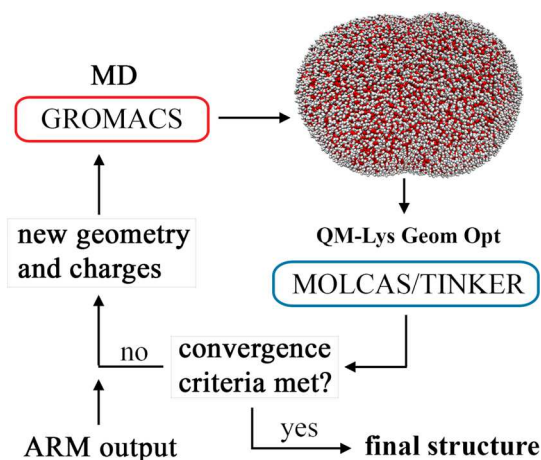


Figure 3. Schematic representation of the iterative procedure for performing the geometry optimization. The ASEC configuration is generated using a sample of configurations from the MD run (performed using the GROMACS molecular dynamics package and treating the QM-Lys subsystem at the MM level as described in the text), and such configuration is then passed to the MOLCAS/TINKER interface, which carries out a QM/MM optimization of the QM-Lys subsystem at the CASSCF/AMBER level within the constrained ASEC configuration of the QM-Lys environment. The ARM output refers to the initial guess structure, see also section 3.

proteins. The first set comprises 12 hCRBP II rhodopsin mimics in a water solution,²⁵ one of which has been demonstrated to be able to undergo photoisomerization.³⁸ As previously mentioned, these synthetic soluble proteins show $\lambda_{\max}$ values from 425 to 644 nm depending on specific mutations performed near the binding pocket of the chromophore. A second set features, with respect to the rhodopsin mimics, wild type vertebrate, invertebrate, and microbial rhodopsins, for which the observed $\lambda_{\max}$ values are reported in the literature and computational results using different methods have been reported.⁴ More specifically, the second set comprises the all-*trans* and 13-*cis* isomers of Anabaena Sensory Rhodopsin (ASR) already mentioned above, bovine rhodopsin (Rh),³⁹ squid rhodopsin (SqR),⁴⁰ human melanopsin (hMeOp),^{41,42} the light- and dark-adapted forms of bacteriorhodopsin (bR_{LA} and bR_{DA}),^{43,44} blue proteorhodopsin (PR),⁴⁵ and batho-

rhodopsin (bathoRh).⁴⁶ Finally a third set comprises five ASR mutants (S86D, S214D, L83Q, V112N, and W76F), each of which has an all-*trans* and a 13-*cis* form yielding a total of 10 $\lambda_{\max}$ values. See refs 47 and 48 and the Supporting Information for the original measurements.

The description of hCRBP II rhodopsin mimics and “natural” rhodopsin (wild types and mutants) proteins was different due to the following reasons:

(i) hCRBP II are small soluble proteins, which are quite flexible and internally hydrated. Therefore, the QM-Lys environment subsystem needs to be extended beyond the protein itself and must comprise the effect of the solvent (represented by the square solvent box in Figure 2).

(ii) Being transmembrane proteins, rhodopsins are instead relatively rigid and large with respect to the chromophore size. Furthermore, they are mostly surrounded by a nonpolar environment (the membrane fatty acid chains) with polar groups localized at the cytoplasmic and extracellular membrane sides. Therefore, in this case, and because we are focusing on $\lambda_{\max}$ values, the QM-Lys environment subsystem only comprises the protein cavity (as shown in Figure 4), which is defined by the CASTp protocol.⁴⁹

These differences are detailed in sections 3.1 and section 3.2.

3.1. Human Cellular Retinol Binding Protein II Rhodopsin Mimics. In order to compute the $\lambda_{\max}$ values of the 12 hCRBP II proteins²⁵ (hereafter referred to as M1 to M11 and “isomer” for the photoisomerizing rhodopsin mimic), the X-ray crystallographic structures of M4, M8, and M10, obtained from the protein data bank (PDB IDs: 4EXZ, 4EFG, and 4EEJ, respectively), were used as initial templates for generating the corresponding mutants. For “isomer,”³⁸ the 4YFP X-ray crystallographic structure was used. The list of mutations performed to generate each mutant is shown in Table 1. The acetate ion seen in the X-ray crystallographic structure of M4 has been removed after assuming that it enters the protein during the crystallization process; also chain B of the PDB files has been removed. The automatic procedure for introducing the mutations in the X-ray crystallographic structure and adding the hydrogen atoms is detailed in ref 4 and is part of the ARM protocol, to which the ASEC-FEG code has been coupled. The charged residues were kept in their standard protonation state

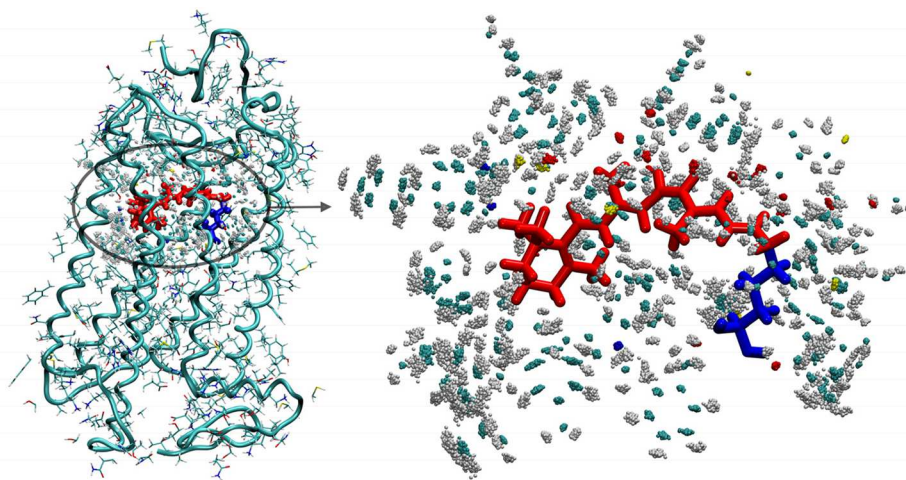


Figure 4. Schematic representation of the ASEC configuration generated for rhodopsin proteins. On the right, it is evident that each atom of the cavity (i.e., the QM-Lys environment subsystem) is replaced by a cloud of pseudoatoms.

Table 1. ΔE_{S1-S0} Values for the hRCBP II Rhodopsin Mimics Computed Using the ASEC-FEG Method, Other QM/MM Methods reported in Literature, and the Experimentally Observed Results^a

	ASEC-FEG	Hayashi et al.	Kaila et al.	exptl. ²⁵
M1(4EXZ/L40K,T51D,L117E)	72.0(397)	70.8(404)	62.5(457)	67.3(425)
M2(4EXZ/L40K,T51D)	67.0(427)	66.9(427)		60.3(474)
M3(4EXZ/L40S)	64.2(445)			59.3(482)
M4(4EXZ)	62.6(457)	61.4(466)	54.7(523)	56.3(508)
M5(4EXZ/T51 V)	58.4(490)	57.9 (494)		53.6(533)
M6(4EXZ/T51 V,R58W)	57.5(497)	55.1(519)		50.2(570)
Isomer (4YFP ^b)	54.4(526)		49.1(582)	50.1(571)
M7(4EXZ/T51 V,Y19W,R58W)	54.8(522)			49.6(576)
M8(4EFG)	53.8(531)	49.8(574)	46.8(611)	48.4(591)
M9(4EEJ/W4R)	52.0(550)	48.2(593)		46.6(614)
M10(4EEJ)	51.9(551)			46.0(622)
M11(4EEJ/A33W)	52.2(548)			45.9(623)

^aThe energies are expressed in kilocalories per mole, and the corresponding $\lambda_{\max}$ values in nanometers are given in parentheses. ^b4YFP corresponds to the crystallographic structure of the photoisomerizing rhodopsin mimic³⁸ indicated as “Isomer” in Figure 5.

at neutral pH, with the exception of Glu72, which is more likely to be neutral according to the PROPKA code.⁵⁰

To properly describe these molecular systems, each mutant has been embedded in a 70 Å cube water box. Under these conditions, the molecular dynamics, previously described in the iterative procedure, are carried out using the GROMACS code,⁵¹ using the AMBER94⁵² and TIP3P⁵³ force fields respectively for the proteins and water molecules. The constrained QM-Lys subsystem interacts through the parametrized van der Waals forces⁵⁴ and QM-derived RESP³⁷ point charges, which are recomputed for each system along the iterative procedure. The whole system is relaxed along the MD, except the QM-Lys subsystem, which is treated at the MM level (see above) and kept geometrically fixed. Before starting the iterative procedure, an initial pre-equilibration of the entire volume has been performed in the *NPT* ensemble, heating the system from 0 to 300 K in 300 ps followed by 1000 ps of thermalization. Then, using the equilibrated volume, the MD of the iterative procedure is performed in the *NVT* ensemble using 5000 ps for thermalization and 5000 ps for production under standard room conditions. The production stage is executed in 10 different computer nodes with different starting seeds in order to get more uncorrelation in the sampled environment. Periodic boundary conditions (PBC) have been used to avoid boundary effects. In order to neutralize the total negative charge of some of these proteins, Na⁺ ions were added to the solvent box, in which case, using the “genion” procedure implemented in the GROMACS code,⁵¹ random molecules of the solvent are replaced by the ions. These ions are free to move along the MD; therefore, when the configurations are selected from the MD to generate the ASEC configuration, many different ion positions will be obtained (as represented by the external blue points in Figure 2) with their respective charge and van der Waals parameters scaled in the same way as for any atom of the system. In this way, we expect to get a well-balanced description of the effect of the ions since in the QM/MM calculations we will have an ion charge distribution weighted by the frequency of the position of the ions in different regions of the configuration space.

Regarding the QM-Lys geometry optimization stage, the QM part is defined by all the atoms of the retinal chromophore plus the C_ε of the bonded lysine (see Figure 1). In this scheme, the total charge of the MM atoms of the Lys is equal to zero, while the net charge of +1 belongs to the QM part. In addition, since

the hydrogen atom is added between the carbons C_ε and C_δ, the charge of the C_δ is set to be zero in order to not overpolarize the QM/MM frontier (see details in ref 54). The QM part is computed at the complete-active-space self-consistent field (CASSCF) level²⁷ and using the 6-31G* basis set. In these calculations, 12 electrons are correlated in 12 orbitals, which comprise the full π system. The remaining MM atoms of the bonded lysine are also represented by the molecular mechanics AMBER94 force field. As discussed in the previous section, these QM calculations are performed with the QM-Lys subsystem embedded in the ASEC configuration. In these soluble rhodopsin mimic systems, where the full protein and the solvent is relaxed along the MD, we have decided to generate the ASEC configurations considering all the atoms within 30 Å from the retinal. Since it is not viable to use PBC for the QM calculations, this 30 Å selection is assumed to be a better compromise than using a cube in order to avoid the nonsymmetric effect of the corners. It is also expected that the electrostatic interaction after 30 Å is weak enough to be disregarded. Therefore, the QM-Lys environment subsystem necessary for computing the QM/MM geometry optimization and vertical excitation energies is defined as the 30 Å cloud of pseudoatoms schematically represented at the top-right of Figure 2.

The absorption $\lambda_{\max}$ values are calculated after the full iterative procedure has converged using the multiconfigurational complete active space second-order perturbation theory (CASPT2)²⁹ in order to account for a substantial part of the dynamical correlation energy. The 6-31G* and ANO-L-VDZP basis sets⁵⁵ were used for these calculations with the QM part embedded into the ASEC configuration. A detailed discussion of the obtained results, as well as the effect of the basis set on the geometry optimization, is presented in section 4.1.

3.2. Rhodopsin Proteins. The definition of the environment surrounding the retinal in the rhodopsin proteins studied in this work is similar to the definition used for hRCBP II-based models, but for rhodopsins an external environment is also defined. In fact, we followed the idea⁴ of dividing the protein into three parts constituted by an external environment, QM-Lys environment subsystem (which corresponds to the chromophore cavity), and a QM-Lys subsystem incorporating, as mentioned above, the QM part. In fact, the QM-Lys subsystem is defined in the same way as for the hRCBP II mimics. The QM-Lys environment is then defined by the side

chains located in close vicinity of the QM-Lys subsystem, as selected using the CASTp online server (i.e., the side chains creating the chromophore cavity).⁴⁹ Notice that only such a QM-Lys environment is relaxed via extensive MD. Finally, the external environment is defined by the rest of the atoms of the protein located outside the QM-Lys environment, and it is kept fixed. The employed models should be embedded into a solvated lipid bilayer. However, in the present benchmarking study, we consider the basic gas-phase models generated by the ARM protocol.⁴ In the ARM models, all atoms of the protein backbone and the side chains that do not belong to the cavity define the external environment are kept fixed along the MD. This has the aim to retain the structural information on the initial X-ray crystallographic structure or homology model. The side chains (not the Ca) of the defined cavity are then relaxed along the MD keeping the QM-Lys subsystem fixed. The same MM force field as in hCRBPPII systems is used here. An initial preheating of 50 ps is performed in the NPT ensemble, followed by 150 ps of thermalization and 5000 ps for production at standard room conditions. Again, the production stage is divided into 10 different nodes to better uncorrelate the sampling.

The starting X-ray crystallographic structures were provided by the Protein Data Bank, identified by the following PDB IDs: ASR (1XIO), Rh (1U19), SqR (2Z73), bR_{LA} (1C3W), bR_{DA} (1X0S), PR (4JQ6), and bathoRh (2G87). For hMeOp, the comparative model obtained from ref 41 was used. The protonation states of charged residues are determined based on the more likely characterization provided by the PROPKA code.⁵⁰ Again, the automatic procedure of adding the hydrogen atoms and introducing mutations are detailed in ref 4 as part of the ARM protocol.

Similar to the hCRBPPII systems, to compute the QM/MM geometry optimizations, the CASSCF/6-31G*²⁷ level of calculation was used to describe the QM-Lys part, while the MM atoms of the bonded lysine and the rest of the protein were considered through the AMBER94 force field.⁵² The active space also comprises the full π system. The vertical excitation energies were then computed at the CASPT2²⁹ level and 6-31G* basis set. This level of calculation, known as CASPT2//CASSCF/6-31G*/MM, has been extensively used^{4,56,57} to study the rhodopsin type proteins in the gas phase, following the proposed description of the model. As should be noted, the ASEC configuration generated in these cases is different from the 30 Å ASEC configuration generated for the hCRBPPII systems. Here, it will be defined by the side chains defining the cavity, as shown schematically in Figure 4. In particular, the intermolecular interactions acting on the QM-Lys subsystem are generated by the scaled electrostatic and van der Waals interaction of the ASEC configuration pseudoatoms of the cavity (i.e., the QM-Lys environment subsystem) plus the interactions with the rest of the atoms of the protein (i.e., the external environment).

3.3. Differences between the Rhodopsin and the hCRBPPII Protein Models. The models described in sections 3.1 and 3.2 are very different and, in fact, represent limiting cases. The main motivation for constructing the rhodopsin models of section 3.2 is to be able to save simulation time. In fact, the construction of the ASEC-FEG model starting from an ARM model allows refinement of the model at a limited cost. Due to the fact that only the atoms of the QM-Lys environment are sampled during the MD calculations, the ASEC-FEG procedure can be carried out relatively quickly.

More specifically, we estimate a ca. 48 h wall clock time for converging the ARM model plus an additional 48 h for the three iterations necessary to converge the ASEC-FEG calculation on a 2.4 GHz Intel Xeon E5-2680 processor (with the MD step carried out in parallel on eight cores and the QM/MM optimization step carried out on a single core). An ARM-type model is not applicable to the hCRBPPII proteins. In fact, in this case, the protein is small, highly flexible, and a hydrated QM-Lys environment in direct contact with the solvent. This means that the MD sampling should include a solvent shell plus the entire unconstrained protein (i.e., with both mobile backbone and side chains). However, the fact that such an extended QM-Lys environment contains unbound solvent molecule prevents the partition in a fixed solvent external environment and mobile solvent–protein QM-Lys environment of the MM subsystem. This is because the mobile water of the QM-Lys environment would diffuse away during the MD step of the ASEC-FEG procedure. Therefore, for the hCRBPPII proteins, we adopt, during the MD step, a fully unconstrained model with the QM-Lys environment comprising the entire protein and solvent box and treated via PBC. Such an environment is then trimmed in the way explained in section 3.1 during the QM/MM geometry optimization step. In spite of the limited size of the protein, the application of the ASEC-FEG procedure on such a model turns out to be more expensive mainly due to the cost of the MD run. Indeed, on the same computer system, the convergence, reached in about three or four iterations, requires about 120 h of wall clock time.

4. RESULTS AND DISCUSSION

In order to achieve both convergence and statistical uncorrelation in the average interaction energy between the QM-Lys subsystem and the QM-Lys environment represented by the ASEC configuration, two factors need to be carefully selected. The first one is the time step for selecting energetically (statistically) uncorrelated configurations from the MD run, and the second one is the number of configurations to be selected. Accordingly, the autocorrelation function⁵⁸ of the electrostatic and van der Waals interactions was computed for one of the studied systems. Details can be found in the Supporting Information, indicating that configurations temporarily separated by about 1 ps show less than 10% of energy autocorrelation. In addition, it is also shown that 80 uncorrelated configurations yield a statistical error less than 0.5 kcal/mol. As previously stated, in our calculations the molecular dynamics has been parallelized in 10 different nodes, starting with different seeds, in order to ensure even more uncorrelation between the selected configurations. Finally, 100 statistically uncorrelated configurations, out of the 5000 ps of production MD, are selected in each step of the iterative cycle (Figure 3) for generating the ASEC configuration.

In the following subsections, we report and discuss the results of absorption $\lambda_{\max}$ calculations expressed in terms of vertical excitation energy $\Delta E_{S_1-S_0}$ (i.e., the electronic energy difference between the first singlet excited state (S_1) and the S_0 state at the converged S_0 equilibrium geometry) using the optimized QM-Lys subsystem embedded into the ASEC configuration. We first present the results of the hCRBPPII rhodopsin mimics in water solution²⁵ and the photoisomerizing rhodopsin mimic³⁸ and then the rhodopsin type proteins, wild types, and mutants.

4.1. hCRBPPII Rhodopsin Mimics. The $\Delta E_{S_1-S_0}$ values of the hCRBPPII rhodopsin mimics computed at the CASPT2

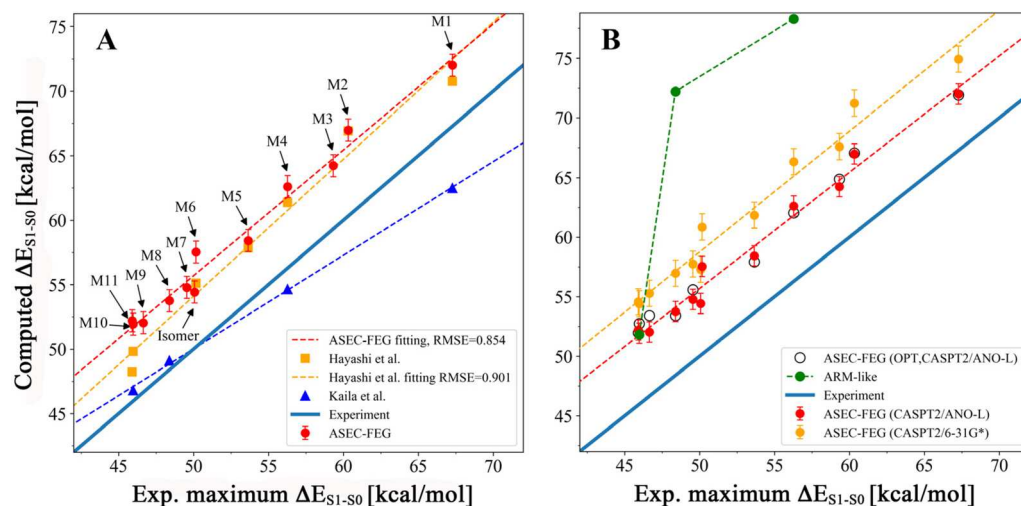


Figure 5. hCRBP II rhodopsin mimics. Panel A shows the computed vertical excitation energies (ΔE_{S1-S0}) using the ASEC-FEG model (red points) and comparison with the observed experimental results and with other theoretically reported results. Panel B shows the effect of the basis set in the calculation of the vertical excitation energy and geometry optimization. In addition, the results obtained following an ARM-like protocol are shown. The dashed lines represent the trend of the corresponding models, determined by a linear regression fitting (except the green one in panel B, which is just connecting the points), while the full bold line is the perfect trend.

(ipea = 0)/ANO-L-VDZP level of calculation are shown in Figure 5A. In the same figure, we also present the results obtained by Hayashi et al.⁷ for some of these mutants by using the free energy gradient method previously described¹⁰ and the results reported by Kaila et al.,⁵⁹ who performed a QM/MM cluster study. It can be observed that the ASEC-FEG points (obtained in this work) closely reproduce the experimental trend, displaying a linear regression almost parallel to the perfect correlation line. The error bars indicate a statistical error of about 0.9 kcal/mol related to the trend line, while the average total error, obtained by comparing the computed and experimental results, is about 5 kcal/mol corresponding to the systematic energy difference between the ASEC-FEG fitting line and the perfect correlation line. It is important to highlight that the CASPT2 (ipea = 0)//CASSCF/6-31G* protocol (i.e., geometry optimization at CASSCF/6-31G* followed by a single point CASPT2 (ipea = 0)/6-31G* calculation for computing the vertical excitation energy), which has been widely used for evaluating vertical excitation energies of different proteins yielding a systematic error of a few kilocalories per mole,^{4,56,57} has been also tested. These results are shown in Figure 5B by the yellow points. The obtained results are still parallel to the perfect correlation line (i.e., they reproduce the trend correctly), but they are now blue-shifted by about 7.5 kcal/mol relative to the experimental results. This behavior is related to the documented cancellation of errors operating during a CASPT2 (ipea = n)//CASSCF calculation. For instance, it has been reported that the large red shift observed in a CASPT2 (ipea = 0.00) calculation with respect to the recommended CASPT2 (ipea = 0.25) calculation is counterbalanced by the use of the CASSCF optimized geometry and 6-31G* basis set (both shifting to the blue) with respect to the recommended, but computationally more demanding, CASPT2 optimized geometry and ANO-L-VDZP basis set. For this reason, the CASPT2 (ipea = 0)//CASSCF/6-31G* protocol has been reported to yield vertical excitation energies close to the CASPT2 (ipea = 0.25)/ANO-L-VDZP//CASPT2 (ipea = 0)/6-31G*. More details about this can be found in the ref 60. On the other hand, the magnitude of such a

cancellation is expected to be system specific. Having commented on this, we can stress that the CASPT2 (ipea = 0)/ANO-L-VDZP//CASSCF/6-31G* level of calculation used in the results shown in Figure 5A is consistent with the documented error cancellation. In fact, the increase of the basis set from 6-31G* to ANO-L-VDZP will just provoke the previously documented red shift in ΔE_{S1-S0} values.⁶⁰ The effect of ANO-L-VDZP on the CASSCF geometry optimization was also studied (see the empty circles in Figure 5B). As can be observed, there is not a large effect in ΔE_{S1-S0} when using this basis set, but in addition, a considerable increase in the computation time was observed as expected. The ASEC-FEG excitation energy values, as well as the experimentally derived vertical excitation energies, are shown in Table 1.

The parallel trend seen in the ASEC-FEG results is consistent with the trend generated by the QM/MM-RWFE-SCF method but extends to a larger number of mutants. When using the ANO-L-VDZ basis set, we also obtain numerical values of ΔE_{S1-S0} very close to those obtained with the QM/MM-RWFE-SCF method which uses the 6-31G** basis (i.e., relatively close to the 6-31G* used in our calculations). The close excitation energies generated by different basis sets are explained by compensating differences in the equilibrium geometries used in the two protocols (CASSCF and B3LYP respectively). In fact, as mentioned above, more correlated QM methods, such as CASPT2 and B3LYP, yield equilibrium geometries with a lower excitation energy due to a more extended delocalization of the π electrons (i.e., the chromophore geometry has a longer double bond and shorter single bonds).

It is also worth highlighting that, as stated in section 3, the model used for describing the hCRBP II rhodopsin mimics is different from the model used for describing the rhodopsin proteins due to several reasons. Indeed, the model used for describing the rhodopsin type systems (the one described by the ARM protocol⁴) was also tested for the description of some of the hCRBP II rhodopsin mimics (the three with available crystallographic structures). These results are shown in Figure 5B by the green points. As can be observed, in addition to the

lack of a trend, a large blue shift is observed for two out of three tested systems, indicating the need for a different description of the modeling approach.

4.1.1. Analysis of the Color Variation. In this section, we discuss the effect of specific amino acid replacements on the $\Delta E_{S_1-S_0}$ values based on the ASEC-FEG models. As a reference (i.e., the “wild-type” system), we will use M4, since all the mutants from M1 to M7 are generated from the M4 crystallographic structure and also because its excitation energy is close to the middle of the observed excitation energy range.

The ASEC configuration concept makes it easy to rationalize the effect of specific mutations. However, before discussing why, it is useful to remind the reader that the S_1 electronic state of the chromophore, which has been shown to correspond to the absorbing spectroscopic state, is characterized by a charge transfer from the positively charged Schiff base toward the β -ionone ring. Therefore, any mutation changing the protein electrostatic interaction to one favoring a charge transfer state is expected to stabilize S_1 and/or destabilize S_0 and therefore to decrease the excitation energy value. Even though both electronic states can be stabilized (e.g., a negatively charged residue would stabilize both ground and excited states), the different distance and/or more diffuse character of the excited state would lead to a different stabilization and, therefore, to a change in the energy gap. The opposite effect is expected for mutations disfavoring the chromophore charge transfer. In Figure 6, we report an analysis of the ASEC configuration

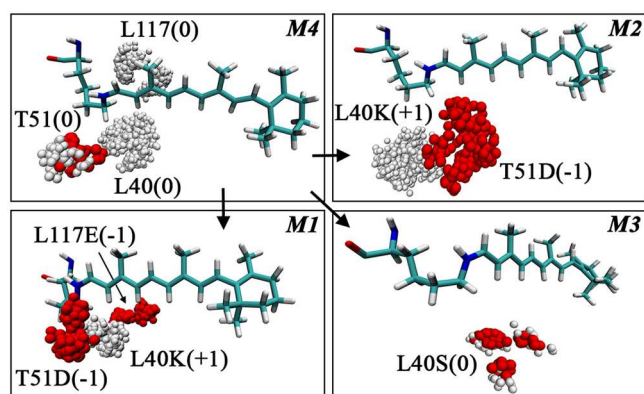


Figure 6. ASEC representation of the polar and charged part of the amino acid replacements introduced to generate the M1, M2, and M3 mutants generated from M4. The clusters of small spheres correspond to all the sampled positions of each corresponding atom. The red and white colors correspond to oxygen and hydrogen atoms, respectively. The total charge of the corresponding side chain is indicated in parentheses.

structure of the M1, M2, and M3 mutants of M4 in the region of the Schiff base. Notice that the clusters of small spheres correspond to the ASEC configuration of specific atoms in the mobile side chains. For instance, in the M4 panel, the red and white clusters correspond to the sampled positions of the oxygen and hydrogen atoms respectively of the T51 side chain (for clarity, we only visualize the parts of the side chain replaced with parts of different polarity in the mutant).

In order to generate M1 (the most blue-shifted mutant of the set), three mutations were performed (L40K, T51D, L117E), with respect to M4. As can be observed in the M1 panel of Figure 6, the replacements create a conformation near the Schiff base of the chromophore, which stabilizes S_0 . In fact,

negative and well localized charged regions (seen as small clouds or clusters) are created on both sides of the Schiff base, which features a positive charge localized on the Schiff base $C=N$ bond in S_0 . This stabilization is additionally enhanced by the L40K replacement, which places a positive charged cloud between the two negative clouds generated by the T51D and L117E replacements. Therefore, in M1 the chromophore S_0 charge distribution/state is stabilized relative to the S_1 leading to an increase in $\Delta E_{S_1-S_0}$ value as observed (see Figure 5). When analyzing M2, the S_0 stabilizing interactions seen in M1 are reduced due to the absence of the negative region generated by the L117E replacement. As a consequence, a more diffuse negative cloud is generated, and the resulting $\Delta E_{S_1-S_0}$ turns out to be more red-shifted with respect to M1 but still blue-shifted with respect to M4. In contrast, in M3, the neutral side chain L40 of the M4 reference is replaced by a serine, and the corresponding ASEC-FEG model yields a relative minor blue-shift. This is supported by the oxygen atom of the S40 side chain of M3, which points toward the Schiff base, inducing a stabilization of S_0 and therefore a blue shift.

In Figure 7, we investigate the effect of the side chain replacements of the M5–M11 mutants. These mutants are all red-shifted with respect to M4 (see Figure 5). As can be observed in panel M5, the T51V replacement of a polar threonine with the nonpolar valine is predicted to lead to a red

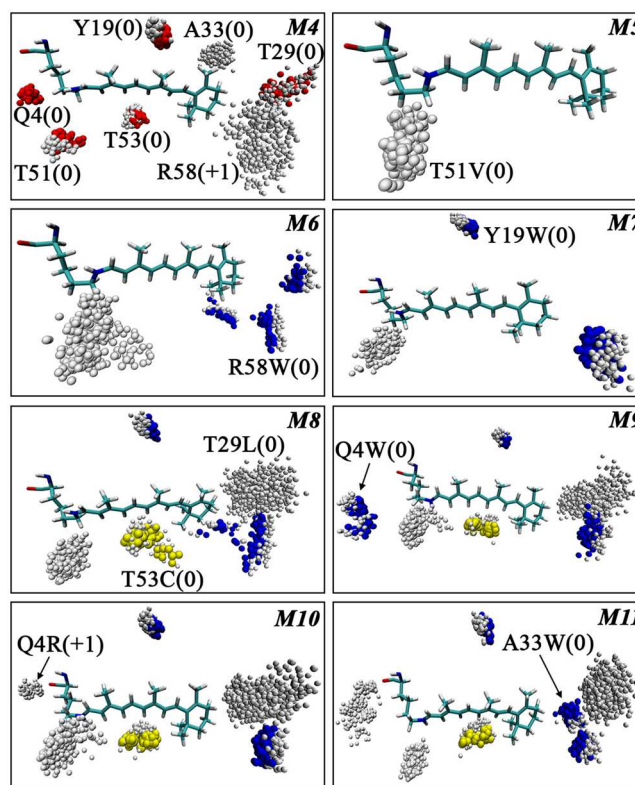


Figure 7. ASEC representation of the polar and charged part of the mutations introduced to generate the mutants from M5 to M11 from the crystallographic structure of M4. The clusters of small spheres correspond to all the sampled positions of each corresponding atom. For clarity, only the mutated side chain is displayed in each panel; the remaining side chains are the same as in M4. The red, white, blue, and yellow colors correspond to oxygen, hydrogen, nitrogen, and sulfur atoms, respectively. The total charge of the corresponding side chain is indicated in parentheses.

shift. In fact, the stabilization of the Schiff base positive charge through the interaction with the OH of the threonine vanishes, resulting in a destabilization of S_0 with respect to S_1 and, consequently, in a vertical excitation energy decrease. In M6, in addition to the effect of the T51V replacement, one has the R58W replacement of a positively charged arginine with the polar tryptophan. This mutation favors the charge transfer to the β -ionone ring, since the repulsive interaction or the positive arginine is replaced by the NH dipole moment of tryptophan with the negatively polarized nitrogen atom pointing to the β -ionone ring. Therefore, in this case, the S_1 state is stabilized with respect to S_0 , yielding a red shift with respect to M5. It is interesting to highlight that three different conformations of W58 are observed in the sampling. These are indicated by the three delocalized clusters on panel M6.

For the M7 mutant, the explanation provided for M6 still applies. However, the effect of the Y19W replacement is not evident since this replaces the OH dipole moment of a tyrosine by the NH dipole moment of tryptophan, both pointing almost in the same direction. The most significant difference is in the distance from the chromophore, which is larger for NH. Therefore, in order to elucidate the role of the mutation, we need to look at its effect on the hydrogen bond network (HBN) in the region of the chromophore. If we look at Figure 8, it is apparent that in the ASEC configuration the water

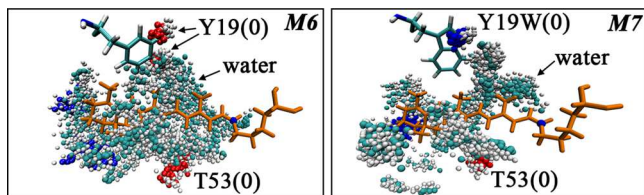


Figure 8. ASEC representation of the change in the water molecules distribution upon Y19W mutation. The cyan and white colors represent the oxygen and hydrogen atoms of the water molecules.

molecules are located within 6 Å from the chromophore both in M6 and M7. A well structured HBN is clearly observed in M6, which involves the interaction of the water molecules with Y19 and T53. Such HBN reaches the Schiff base. Thus, one expects an S_0 stabilization relative to M7 where, due to the Y19W mutation, the HBN is weaker, with almost no interaction with the Schiff base region.

The Q4W replacement generates M9 from M8. In this case, the negative charge distribution of the carbonyl oxygen of glutamine, which stabilizes a positively charged Schiff base, is replaced by the NH dipole moment of the tryptophan. Essentially, this substitution tends to destabilize the Schiff base charge in M9 compared to M8, thus decreasing the excitation energy. Following the same analysis, we can see that for generating M10, the same glutamine is replaced with a positively charged arginine residue, destabilizing the Schiff base even more and consequently continuing the red shifting trend. The same reasoning is applied to the β -ionone side. When looking at the panel M11 of Figure 7, A33W is generating a dipole moment near the β -ionone ring with the tryptophan nitrogen pointing to the ring. This mutation, which is reinforcing the effect of the R58W replacement, favors the charge transfer state, thus stabilizing S_1 , leading to a red shift.

4.2. Rhodopsins and Their Mutants. The results obtained for the wild-type rhodopsin set and the corresponding mutant set are displayed in Figure 9. In this case, the $\Delta E_{S_1-S_0}$

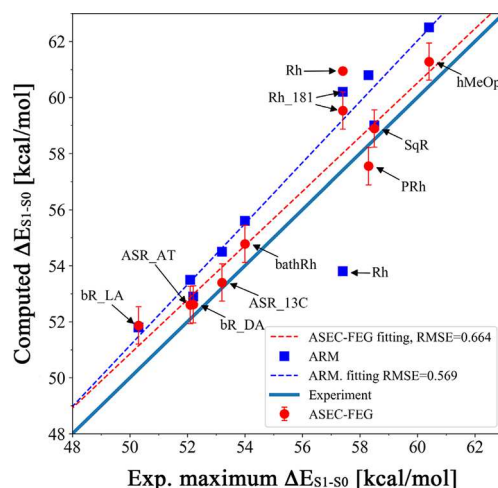


Figure 9. Vertical excitation energies ($\Delta E_{S_1-S_0}$) computed for the wild-type rhodopsin systems using the ASEC-FEG free energy geometry optimization (red points) and the ARM protocol⁴ (blue points).

values are computed using the optimized structures of the QM-Lys subsystem embedded in the electrostatic fields generated by the ASEC configurations of the QM-Lys environment subsystem and by external protein environment. For comparison, we also include the vertical excitation energy obtained with the ARM protocol.⁴ As we will discuss below, both the ASEC-FEG and the ARM procedures yield comparable results in the wild-type systems, except for one model of bovine rhodopsin (Rh) and two mutants of the Anabaena Sensory Rhodopsin. The ARM protocol is based on a potential energy minimization of the entire Lys-QM and Lys-QM environment subsystems (the chromophore cavity) in the field of the external environment producing, substantially, a 0 K model. This is not a bad approximation when the system is relatively rigid such as for transmembrane proteins and when, consequently, we can afford a relatively small Lys-QM environment subsystem. In fact, in this case, the average representation of the environment is assumed to be correctly provided by the X-ray crystallographic structure. However, there are cases where more than one stable cavity configuration of the system is “explored” by the short preparatory molecular dynamics run in the ARM protocol. In these cases, an iterative electrostatic and geometrical equilibration of the system such as the one offered by the ASEC-FEG method is needed. This is the case of the Rh model featuring a protonated (i.e., neutral) E181 residue. For this particular protonation state of Rh, as shown in Figure 10A and B, two largely different HBNs exist that are not quickly interconverted into each other. The ASEC-FEG model computed for such a specific protonation state of Rh leads to a very large increase (7 kcal/mol) in the excitation energy which is now comparable with the experimental data and also with the excitation energy computed with the alternative model (Rh_181) featuring a deprotonated (i.e., charged) E181. Interestingly, the second cycle of the ASEC iteration is needed for the observation of this distinctively HBN, in comparison with the first cycle and with the ARM result. The better agreement of the energy difference of both the Rh models with the observed value is expected since E181 is located almost exactly at the midpoint of the chromophore backbone, and therefore the E181 ionization status should not greatly affect the excitation energy. The correlation lines

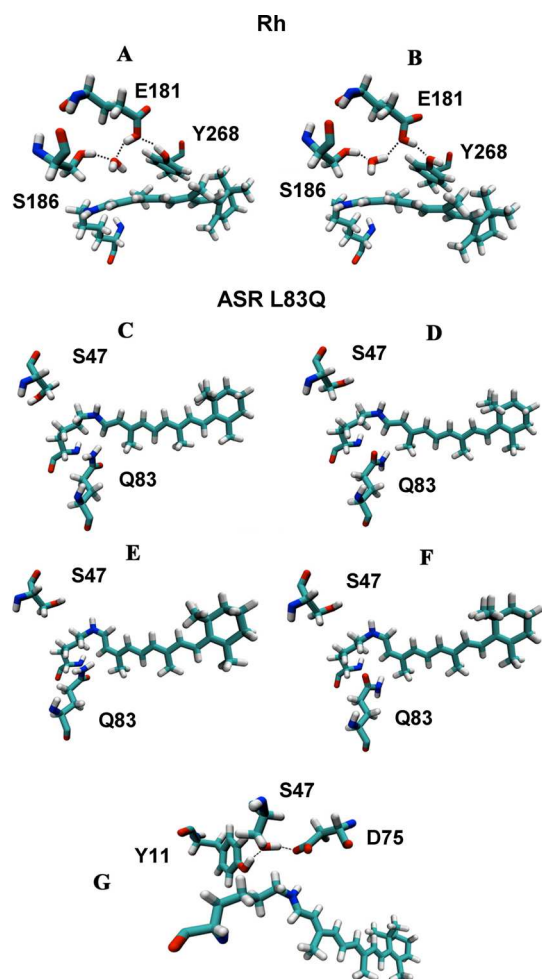


Figure 10. Comparison of the hydrogen bond network and side chain orientations of the Rh neutral bovine rhodopsin model (panels A and B) and L83Q ASR mutants (panels C and D for the all-*trans* conformation and panel E and F for the 13-*cis* conformation) obtained by the ARM (panels A, C, and E) and ASEC-FEG (panels B, D, and F) protocols. In panel G, the S47 side chain is represented in an orientation consistent with the displayed HBN.

corresponding to the dashed lines in Figure 9 do not consider the protonated forms of Rh, since the unprotonated Rh₁₈₁ form seems to be more favorable for describing the spectroscopic properties.

The results obtained for the ASR mutants are shown in Figure 11. As can be seen, some improvement has been achieved in the description of the spectroscopic properties of the mutants when using the ASEC-FEG model and indicating that a more elaborate description of the chromophore cavity is needed for mutants. This is mainly for the L83Q mutant (see below), which represents the most problematic case featuring ARM models unusually red-shifted and displaying an all-*trans*-13-*cis* ΔE_{S1-S0} trend opposite of the observed one. Such improvement is achieved in two ways: first, 100 statistically uncorrelated configurations of the cavity are used to describe the statistical properties of the system, and second, the sequential iterations of QM/MM optimizations and resampling of the cavity (updating geometry and charges of the QM part) allow reaching a better equilibrated description of the cavity of the system (i.e., the electrostatic and geometrical effect of the mutation is iteratively equilibrated). As previously stated,

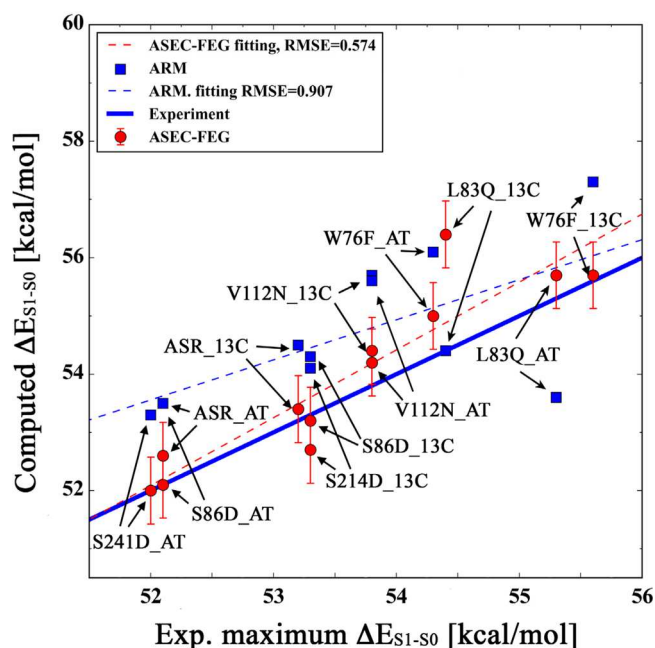


Figure 11. Vertical excitation energies (ΔE_{S1-S0}) computed for some mutants of the Anabaena Sensory Rhodopsin using the ASEC free energy geometry optimization (red points) and the ARM protocol⁴ (blue points).

special attention deserves the mutant L83Q of ASR where (see Figure 10) an opposite orientation of the mutated Q83 side chain is observed when comparing the ARM (left side) and the ASEC-FEG (right side) models. The consequence of the orientation change is a large blue-shift (see Figure 11) for both the all-*trans* and 13-*cis* isomer, producing ΔE_{S1-S0} values more consistent with the general trend (although the mutant all-*trans*-13-*cis* ΔE_{S1-S0} trend is not reversed and remains incorrect). It is also interesting to highlight the role of residue S47. As can be observed, a different orientation of this residue is observed in panel C of Figure 10, corresponding to the ARM model of the all-*trans* isomer. Such a conformation, which appears incorrect as it breaks the HBN shown in panel G, is changed during the corresponding ASEC-FEG model generation. In other words, we found that the better sampling allows generation of the proper side chain orientation and, therefore, increases the chances to reproduce the observed ΔE_{S1-S0} trend. Our results along with the previously reported ARM results and the corresponding experimental values are collected in Table 2 for both wild type systems and ASR mutants.

It is interesting to highlight that no qualitative geometrical changes are observed in the structures of the QM-Lys subsystem in comparison with the ones obtained by using ARM. For such a reason, we conclude that the improvements are mainly due to the interaction of the QM-Lys subsystem with the more realistic ASEC environment accounting for the statistical sampling.

5. CONCLUSIONS AND PERSPECTIVES

In this work, we focused on the development of a QM/MM protocol to optimize the retinal chromophore on rhodopsin-type systems by combining the free energy gradient of the system and the ASEC model. The resulting ASEC-FEG protocol has been benchmarked by determining the ground

Table 2. ASEC-FEG Computed Vertical Excitation Energies (ΔE_{S1-S0}) of the Wild-Type Rhodopsin Systems and of the Mutants of the Anabaena Sensory Rhodopsin^a

	ASEC-FEG	ARM [†]	Exp.
wild-type			
ASR_13C	53.4 (535)	54.5 (526)	53.2 (537) ⁴⁸
ASR_AT	52.6 (543)	53.5 (534)	52.1 (549) ⁴⁸
bathRh	54.8 (522)	55.6 (514)	54.0 (529) ⁶¹
bR_DA	52.6 (544)	52.9 (540)	52.2 (548) ⁶²
bR_LA	51.9 (551)	51.8 (552)	50.3 (568) ⁶²
hMeOp	61.3 (466)	62.5 (457)	60.4 (473) ^b
PRh	57.6 (496)	60.8 (470)	58.3 (490) ⁶³
Rh	61.0 (469)	53.8 (531)	57.4 (498) ⁶⁴
SqR	58.9 (485)	59.0 (485)	58.5 (489) ⁶⁵
Rh_181	59.5 (481)	60.2 (475)	57.4 (498) ⁶⁴
ASR mutants			
V112N_AT	54.2 (527)	55.6 (513)	53.8 (531) ⁴
L83Q_AT	55.7 (513)	53.6 (502)	55.3 (517) ^c
W76F_AT	55.0 (519)	56.1 (509)	54.3 (527) ⁴
S214D_AT	52.0 (549)	53.3 (536)	52.0 (550) ⁴
S86D_AT	52.1 (548)	53.5 (533)	52.1 (549) ⁴
V112N_13C	54.4 (525)	55.8 (512)	53.8 (531) ⁴
L83Q_13C	56.4 (502)	54.4 (496)	54.4 (526) ^c
W76F_13C	55.7 (513)	57.3 (498)	55.6 (514) ⁴
S214D_13C	52.7 (542)	54.3 (526)	53.3 (536) ⁴
S86D_13C	53.2 (537)	54.1 (528)	53.3 (536) ⁴

^aThe ARM results (second column) and the observed values (third column) are also shown. The energies are expressed in kcal/mol and the corresponding $\lambda_{\max}$ values in nm are given in parentheses. ^bAverage of available values from refs 62 and 63. ^cThe experimental results are shown in the Supporting Information.

state equilibrium geometries and the vertical excitation energies of a set of hCRBPPII rhodopsin mimics, a set of wild-type rhodopsins, and a set of mutants of a sensory rhodopsin from the cyanobacterium Anabaena PCC 7120.⁴⁸ Above, we have shown that ASEC-FEG is successfully dealing with systems with high mobility (e.g., the hCRBPPII rhodopsin mimics) where extended HBN configuration changes may be sampled during relatively short MD runs. The result is an ASEC-FEG structure representing a minimum on the system free energy surface at the selected temperature and thus offering a more robust, but not dramatically more expensive, alternative to potential energy optimizations. Furthermore, the integration of the ASEC-FEG model¹⁵ into the ARM protocol^{1,2} has allowed computation of the free energy gradient^{12,17,18} of the hybrid QM/MM Hamiltonian in a practical and efficient way, generating novel and more robust models where incorrect HBN configurations (e.g., like in the cases of the Rh and ASR L83Q models discussed above) generated by ARM are, at least partially, repaired. Ultimately, the benchmarks show excitation energy trends substantially parallel to the experimental trend.

Although the present benchmarking has been focused on the study of the ground electronic state geometry optimization and vertical excitation energies, it is also possible to compute excited electronic state equilibrium geometries by using the ASEC-FEG protocol for studying the fluorescent and phosphorescent spectra at different temperatures. Furthermore, one of the most promising features of the present ASEC-FEG approach is that, since we are able to compute the gradient and also an improved model Hessian,³⁵ it can be applied, without further changes, to the geometry optimization of transition

states, i.e., first order saddle-points on the free energy surface of protein systems.^{8,66} In this way, both the free energy and, therefore, entropy variations in isomerization processes can be studied as a function of the temperature. In those cases, the relative stability between the stationary structures (i.e., between two isomers or between a minimum and a transition state) can be computed using the thermodynamics perturbation theory (eq 1). Although, in order to properly account for free energy differences, several small intermediate steps need to be performed,³² as well as other components, like the vibrational motions, should be taken into account (see details about several methodologies developed to this end in refs 67 and 68 and refs therein). These types of developments of the ASEC-FEG protocol are presently being pursued in our lab.

Finally, due to the need for a fast QM/MM model construction, necessary for working on sizable benchmarking sets as well as for effectively interfacing of ARM and ASEC-FEG, above we have focused exclusively on models of rhodopsin systems and on a “minimal” QM subsystem (i.e., the retinal chromophore itself). These limitations of the presented implementation are not intrinsic features of the ASEC-FEG protocol and will be removed by future work.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00860.

Computational and experimental details for scaling of the van der Waals parameters and charges to be used in the ASEC configuration, autocorrelation function, preparation of the models, experimental determination of $\lambda_{\max}$ of the ASR L83Q mutant, full ref 26 of the main text (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: yoelvis.orozco@gmail.com.

*E-mail: canuto@if.usp.br.

*E-mail: molivuc@bgsu.edu.

ORCID

Yoelvis Orozco-Gonzalez: 0000-0001-7225-2424

Nicolas Ferré: 0000-0002-5583-8834

Massimo Olivucci: 0000-0002-8247-209X

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported mainly by the Italian MIUR for funding (PRIN 2015) and, in part, by the National Science Foundation under grant no. CHE-1710191. M.O. is grateful to the Ohio Supercomputer Center for granted computer time. M.O. and Y.O.-G. are grateful to the University of Strasbourg for a USIAS 2015 fellowship. Y.O.-G. and S.C. acknowledge the support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under the grants 2012/15161-7 and 2013/11634-0. N.F., S.H., and D.A. acknowledge support from the French Agence Nationale de la Recherche (grant for project FEMTO-ASR, no. ANR-14-CE35-0015-02).

■ REFERENCES

- (1) Gozem, S.; Melaccio, F.; Luk, H. L.; Rinaldi, S.; Olivucci, M. Learning from Photobiology How to Design Molecular Devices Using a Computer. *Chem. Soc. Rev.* **2014**, *43*, 4019–4036.
- (2) Montisci, F. Benchmarking and Applications of a New Computational Photobiology Tool for Automatic Rhodopsin Modeling. M.S. Dissertation, Università di Siena, 2014.
- (3) Laricheva, E. N.; Gozem, S.; Rinaldi, S.; Melaccio, F.; Valentini, A.; Olivucci, M. Origin of Fluorescence in 11-*cis* Locked Bovine Rhodopsin. *J. Chem. Theory Comput.* **2012**, *8*, 2559–2563.
- (4) Melaccio, F.; Del Carmen Marin, 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. Towards Automatic Rhodopsin Modeling as a Tool for High-Throughput Computational Photobiology. *J. Chem. Theory Comput.* **2016**, *12*, 6020.
- (5) Tomasello, G.; Olaso-Gonzalez, G.; Altoe, P.; Stenta, M.; Serrano-Andres, L.; Merchan, M.; Orlandi, G.; Bottoni, A.; Garavelli, M. Electrostatic Control of the Photoisomerization Efficiency and Optical Properties in Visual Pigments: on the Role of Counterion Quenching. *J. Am. Chem. Soc.* **2009**, *131*, 5172–86.
- (6) Altun, A.; Yokoyama, S.; Morokuma, K. Mechanism of Spectral Tuning Going from Retinal in Vacuo to Bovine Rhodopsin and its Mutants: Multireference *Ab Initio* Quantum Mechanics/Molecular Mechanics Studies. *J. Phys. Chem. B* **2008**, *112*, 16883–90.
- (7) 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*, 13362–70.
- (8) Gozem, S.; Schapiro, I.; Ferré, N.; Olivucci, M. The Molecular Mechanism of Thermal Noise in Rod Photoreceptors. *Science* **2012**, *337*, 1225–8.
- (9) Manathunga, M.; Yang, X.; Luk, H. L.; Gozem, S.; Frutos, L. M.; Valentini, A.; Ferré, N.; Olivucci, M. Probing the Photodynamics of Rhodopsins with Reduced Retinal Chromophores. *J. Chem. Theory Comput.* **2016**, *12*, 839–50.
- (10) Kosugi, T.; Hayashi, S. QM/MM Reweighting Free Energy SCF for Geometry Optimization on Extensive Free Energy Surface of Enzymatic Reaction. *J. Chem. Theory Comput.* **2012**, *8*, 322–334.
- (11) Galván, I. F.; Sánchez, M. L.; Martín, M. E.; Olivares del Valle, F. J.; Aguilar, M. A. Geometry Optimization of Molecules in Solution: Joint Use of the Mean Field Approximation and the Free-Energy Gradient Method. *J. Chem. Phys.* **2003**, *118*, 255–263.
- (12) Okuyama-Yoshida, N.; Kataoka, K.; Nagaoka, M.; Yamabe, T. Structure Optimization via Free Energy Gradient Method: Application To Glycine Zwitterion In Aqueous Solution. *J. Chem. Phys.* **2000**, *113*, 3519–3524.
- (13) Georg, H. C.; Canuto, S. Electronic Properties of Water in Liquid Environment. A Sequential QM/MM Study Using the Free Energy Gradient Method. *J. Phys. Chem. B* **2012**, *116*, 11247–54.
- (14) Bistafa, C.; Georg, H. C.; Canuto, S. Combining *Ab Initio* Multiconfigurational and Free Energy Gradient Methods to Study the Π – Π^* Excited State Structure and Properties of Uracil in Water. *Comput. Theor. Chem.* **2014**, *1040–1041*, 312–320.
- (15) Coutinho, K.; Georg, H. C.; Fonseca, T. L.; Ludwig, V.; Canuto, S. An Efficient Statistically Converged Average Configuration for Solvent Effects. *Chem. Phys. Lett.* **2007**, *437*, 148.
- (16) Coutinho, K.; Rivelino, R.; Georg, H. C.; Canuto, S. In *Solvation Effects on Molecules and Biomolecules: Computational Methods and Applications*; Canuto, S., Ed.; Springer: London, 2008; Vol. 6, pp 159–189.
- (17) Okuyama-Yoshida, N.; Nagaoka, M.; Yamabe, T. Transition-State Optimization on Free Energy Surface: Toward Solution Chemical Reaction Ergodography. *Int. J. Quantum Chem.* **1998**, *70*, 95–103.
- (18) Hirao, H.; Nagaoka, Y.; Nagaoka, M. Transition-State Optimization by the Free Energy Gradient Method: Application to Aqueous-Phase Menshutkin Reaction Between Ammonia And Methyl Chloride. *Chem. Phys. Lett.* **2001**, *348*, 350–356.
- (19) Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.; Brown, L. S.; Kandori, H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chem. Rev.* **2014**, *114*, 126–163.
- (20) Gorostiza, P.; Isacoff, E. Y. Optical Switches for Remote and Noninvasive Control of Cell Signaling. *Science* **2008**, *322*, 395–399.
- (21) Boyden, E. S.; Zhang, F.; Bamberg, E.; Nagel, G.; Deisseroth, K. Millisecond-timescale, Genetically Targeted Optical Control of Neural Activity. *Nat. Neurosci.* **2005**, *8*, 1263–1268.
- (22) Han, X.; Boyden, E. S. Multiple-Color Optical Activation, Silencing, and Desynchronization of Neural Activity, with Single-Spike Temporal Resolution. *PLoS One* **2007**, *2*, e299.
- (23) Schoenenberger, P.; Gerosa, D.; Oertner, T. G. Temporal Control of Immediate Early Gene Induction by Light. *PLoS One* **2009**, *4*, e8185.
- (24) Fenno, L.; Yizhar, O.; Deisseroth, K. The Development and Application of Optogenetics. *Annu. Rev. Neurosci.* **2011**, *34*, 389–412.
- (25) Wang, W.; Nossoni, Z.; Berbasova, T.; Watson, C. T.; Yapici, I.; Lee, K. S. S.; Vasileiou, C.; Geiger, J. H.; Borhan, B. Tuning the Electronic Absorption of Protein-Embedded *All-trans*-Retinal. *Science* **2012**, *338*, 1340–1343.
- (26) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P. A.; Neogrady, P.; Pedersen, T. B.; Pitonak, M.; Reiher, M.; Roos, B. O.; et al. MOLCAS 7: The Next Generation. *J. Comput. Chem.* **2010**, *31*, 224–47.
- (27) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. A Complete Active Space SCF Method (CASSCF) Using a Density-Matrix Formulated Super-CI Approach. *Chem. Phys.* **1980**, *48*, 157.
- (28) Möller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
- (29) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. 2nd-Order Perturbation-Theory with a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.
- (30) Bartlett, R. J.; Musial, M. Coupled-cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291–352.
- (31) Parr, R. G. Density Functional Theory. *Annu. Rev. Phys. Chem.* **1983**, *34*, 631–656.
- (32) Kollman, P. Free energy calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (33) Schleyer, P. v. R. In *Encyclopedia of Computational Chemistry*; J. Wiley: New York, 1998.
- (34) Ferré, N.; Ángyán, J. G. Approximate electrostatic interaction operator for QM/MM calculations. *Chem. Phys. Lett.* **2002**, *356*, 331–339.
- (35) Melaccio, F.; Olivucci, M.; Lindh, R.; Ferré, N. Unique QM/MM Potential Energy Surface Exploration Using Microiterations. *Int. J. Quantum Chem.* **2011**, *111*, 3339–3346.
- (36) Singh, U. C.; Kollman, P. A. A Combined *Ab Initio* Quantum Mechanical and Molecular Mechanical Method for Carrying Out Simulations on Complex Molecular Systems: Applications to the CH₃Cl + Cl[−] Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718–730.
- (37) Bayly, C. I.; Cieplak, P.; Cornell, W.; 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*, 10269–10280.
- (38) Nosrati, M.; Berbasova, T.; Vasileiou, C.; Borhan, B.; Geiger, J. H. A Photoisomerizing Rhodopsin Mimic Observed at Atomic Resolution. *J. Am. Chem. Soc.* **2016**, *138*, 8802–8.
- (39) Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V. The Retinal Conformation and its Environment in Rhodopsin in Light of a New 2.2 Å Crystal Structure. *J. Mol. Biol.* **2004**, *342*, 571–83.
- (40) Murakami, M.; Kouyama, T. Crystal Structure of Squid Rhodopsin. *Nature* **2008**, *453*, 363–367.
- (41) Rinaldi, S.; Melaccio, F.; Gozem, S.; Fanelli, F.; Olivucci, M. Comparison of the Isomerization Mechanisms of Human Melanopsin and Invertebrate and Vertebrate Rhodopsins. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1714–1719.

- (42) Provencio, I.; Jiang, G.; De Grip, W. J.; Hayes, W. P.; Rollag, M. D. Melanopsin: An Opsin in Melanophores, Brain, And Eye. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 340–5.
- (43) Luecke, H.; Schobert, B.; Richter, H. T.; Cartailier, J. P.; Lanyi, J. K. Structure of Bacteriorhodopsin at 1.55 Å Resolution. *J. Mol. Biol.* **1999**, *291*, 899–911.
- (44) Nishikawa, T.; Murakami, M.; Kouyama, T. Crystal Structure of the 13-*cis* Isomer of Bacteriorhodopsin in the Dark-Adapted State. *J. Mol. Biol.* **2005**, *352*, 319–28.
- (45) Ran, T.; Ozorowski, G.; Gao, Y.; Sineshchekov, O. A.; Wang, W.; Spudich, J. L.; Luecke, H. Cross-Protomer Interaction with the Photoactive Site in Oligomeric Proteorhodopsin Complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1965–80.
- (46) Yoshizawa, Y.; Miyazaki, Y.; Inase, N.; Otani, Y.; Isogai, S.; Furuie, M.; Kuramochi, J.; Kishi, M.; Jinta, T. Chronic Hypersensitivity Pneumonitis. *Nippon Naika Gakkai Zasshi* **2006**, *95*, 1904–11.
- (47) Jung, K. H.; Trivedi, V. D.; Spudich, J. L. Demonstration of a Sensory Rhodopsin in Eubacteria. *Mol. Microbiol.* **2003**, *47*, 1513–22.
- (48) Vogeley, L.; Sineshchekov, O. A.; Trivedi, V. D.; Sasaki, J.; Spudich, J. L.; Luecke, H. Anabaena Sensory Rhodopsin: A Photochromic Color Sensor at 2.0 Å. *Science* **2004**, *306*, 1390–1393.
- (49) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated Residues. *Nucleic Acids Res.* **2006**, *34*, W116–W118.
- (50) Olsson, M. H.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical Pka Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–37.
- (51) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.
- (52) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, And Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
- (54) Ferré, N.; Cembran, A.; Garavelli, M.; Olivucci, M. Complete-Active-Space Self-Consistent-Field/Amber Parameterization of the Lys296–Retinal–Glu113 Rhodopsin Chromophore-Counterion System. *Theor. Chem. Acc.* **2004**, *112*, 335–341.
- (55) Widmark, P. O.; Malmqvist, P. A.; Roos, B. O. Density-Matrix Averaged Atomic Natural Orbital (ANO) Basis-Sets for Correlated Molecular Wave-Functions. 1st Row Atoms. *Theor. Chim. Acta* **1990**, *77*, 291.
- (56) Altoe, P.; Cembran, A.; Olivucci, M.; Garavelli, M. Aborted Double Bicycle-Pedal Isomerization with Hydrogen Bond Breaking is the Primary Event of Bacteriorhodopsin Proton Pumping. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20172–7.
- (57) Coto, P. B.; Marti, S.; Oliva, M.; Olivucci, M.; Merchan, M.; Andres, J. Origin of the Absorption Maxima of the Photoactive Yellow Protein Resolved via *Ab Initio* Multiconfigurational Methods. *J. Phys. Chem. B* **2008**, *112*, 7153–6.
- (58) Allen, M. P.; Tildesley, D. J. In *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.
- (59) Suomivuori, C. M.; Lang, L.; Sundholm, D.; Gamiz-Hernandez, A. P.; Kaila, V. R. Tuning the Protein-Induced Absorption Shifts of Retinal in Engineered Rhodopsin Mimics. *Chem. - Eur. J.* **2016**, *22*, 8254–61.
- (60) Gozem, S.; Huntress, M.; Schapiro, I.; Lindh, R.; Granovsky, A. A.; Angeli, C.; Olivucci, M. Dynamic Electron Correlation Effects on the Ground State Potential Energy Surface of a Retinal Chromophore Model. *J. Chem. Theory Comput.* **2012**, *8*, 4069–80.
- (61) Kawanabe, A.; Furutani, Y.; Jung, K. H.; Kandori, H. FTIR Study of the Photoisomerization Processes in the 13-*cis* and All-*trans* Forms of Anabaena Sensory Rhodopsin at 77 K. *Biochemistry* **2006**, *45*, 4362–70.
- (62) Matsuyama, T.; Yamashita, T.; Imamoto, Y.; Shichida, Y. Photochemical Properties of Mammalian Melanopsin. *Biochemistry* **2012**, *51*, 5454–5462.
- (63) Bailes, H. J.; Lucas, R. J. Human Melanopsin Forms a Pigment Maximally Sensitive to Blue Light (λ_{max} Approximately 479 nm) Supporting Activation of G(q/11) and G(i/o) Signalling Cascades. *Proc. R. Soc. London, Ser. B* **2013**, *280*, 20122987.
- (64) Furutani, Y.; Kawanabe, A.; Jung, K. H.; Kandori, H. FTIR Spectroscopy of the All-*trans* Form of Anabaena Sensory Rhodopsin At 77 K: Hydrogen Bond of A Water Between the Schiff Base and Asp75. *Biochemistry* **2005**, *44*, 12287–96.
- (65) Stein, C. J.; Reiher, M. Automated Selection of Active Orbital Spaces. *J. Chem. Theory Comput.* **2016**, *12*, 1760–71.
- (66) Guo, Y.; Sekharan, S.; Liu, J.; Batista, V. S.; Tully, J. C.; Yan, E. C. Y. Unusual Kinetics of Thermal Decay of Dim-Light Photoreceptors in Vertebrate Vision. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 10438–10443.
- (67) Lima, M. C. P.; Coutinho, K.; Canuto, S.; Rocha, W. R. Reaction Mechanism and Tautomeric Equilibrium of 2-Mercaptopurine in the Gas Phase and in Aqueous Solution: A Combined Monte Carlo and Quantum Mechanics Study. *J. Phys. Chem. A* **2006**, *110*, 7253–7261.
- (68) García-Prieto, F. F.; Galván, I. F.; Aguilar, M. A.; Martín, M. E. Study on the Conformational Equilibrium of the Alanine Dipeptide in Water Solution by Using the Averaged Solvent Electrostatic Potential from Molecular Dynamics Methodology. *J. Chem. Phys.* **2011**, *135*, 194502.