JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Origin of the Spectral Shifts among the Early Intermediates of the Rhodopsin Photocycle

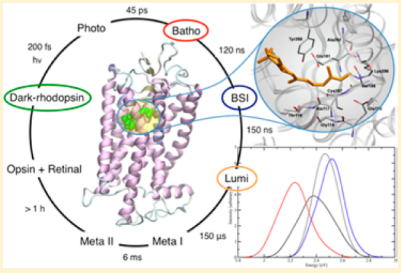
Pablo Campomanes,[†] Marilisa Neri,[†] Bruno A. C. Horta,[†] Ute F. Röhrig,[‡] Stefano Vanni,[†] Ivano Tavernelli,[†] and Ursula Rothlisberger*[,†]

[†]Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale Lausanne, CH-1015 Lausanne, Switzerland

[‡]Molecular Modeling Group, Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland

Supporting Information

ABSTRACT: A combined strategy based on the computation of absorption energies, using the ZINDO/S semiempirical method, for a statistically relevant number of thermally sampled configurations extracted from QM/MM trajectories is used to establish a one-to-one correspondence between the structures of the different early intermediates (dark, batho, BSI, lumi) involved in the initial steps of the rhodopsin photoactivation mechanism and their optical spectra. A systematic analysis of the results based on a correlation-based feature selection algorithm shows that the origin of the color shifts among these intermediates can be mainly ascribed to alterations in intrinsic properties of the chromophore structure, which are tuned by several residues located in the protein binding pocket. In addition to the expected electrostatic and dipolar effects caused by the charged residues (Glu113, Glu181) and to strong hydrogen bonding with Glu113, other interactions such as π-stacking with Ala117 and Thr118 backbone atoms, van der Waals contacts with Gly114 and Ala292, and CH/π weak interactions with Tyr268, Ala117, Thr118, and Ser186 side chains are found to make non-negligible contributions to the modulation of the color tuning among the different rhodopsin photointermediates.



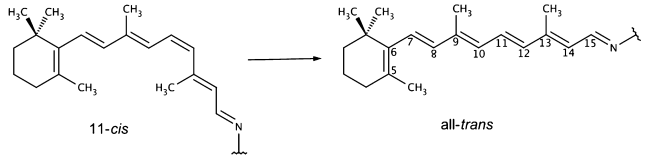
1. INTRODUCTION

G protein-coupled receptors (GPCRs) constitute a large family of transmembrane proteins, whose primary function consists in mediating cellular responses to a wide range of extracellular stimuli, thus being key components of a broad variety of biological signal transduction pathways. These receptors share a highly homologous fold characterized by a common seven transmembrane helix architecture, and present a conformational equilibrium between inactive and active conformations that is modulated by the selective binding of different ligands.[1]

The visual pigment rhodopsin is a highly specialized member of the GPCR family found in vertebrate rod cells. Rhodopsin is able to capture and convert light into a chemical signal, in what constitutes the first step of vision.[2] In the dark-state, its natural ligand (11-*cis*-retinal) acts as a strong inverse agonist and is covalently linked to residue Lys296 of transmembrane helix 7 (TM7) of the opsin protein via a protonated Schiff base (PSB).[3,4] The ultrafast and efficient photoinduced isomerization of the retinal chromophore, from 11-*cis*- to all-*trans* (Scheme 1), inside the binding pocket initiates a cascade of conformational changes that ultimately leads to receptor activation and subsequent downstream signaling.[5]

Several spectroscopically distinguishable intermediates involved in the rhodopsin photoactivation mechanism have first been detected using UV/visible spectroscopy. Although the sequences of spectroscopically distinct species that have been identified either employing time-resolved techniques at physiological temperature[6,7] or steady-state experiments at low temperature[8,9] are not completely identical, a mechanism involving a sequential decay between the experimentally detected intermediates is the most commonly accepted model (Figure 1).[10] Light absorption by the chromophore triggers the transition from dark state rhodopsin to a first photointermediate, photorhodopsin, whose formation is characterized by a very fast rate (within 200 fs)[11] and a very high quantum yield (0.65).[12] Subsequently, photorhodopsin thermally relaxes within a few picoseconds to a new short-lived intermediate, bathorhodopsin, which in turn, on a nanosecond time scale, gives rise to the so-called blue-shifted intermediate, BSI, before it decays to form lumirhodopsin. Lumirhodopsin's structural relaxation takes place on a longer time scale (microseconds) to give rise to another intermediate, metarhodopsin I, which is the precursor of the active conformation,

Scheme 1. 11-*cis*- to all-*trans*-Isomerization of the Retinal Chromophore with the Atom Numbering for the Conjugated Chain



Received: November 5, 2013
Published: February 10, 2014

dx.doi.org/10.1021/ja411303v | J. Am. Chem. Soc. 2014, 136, 3842−3851

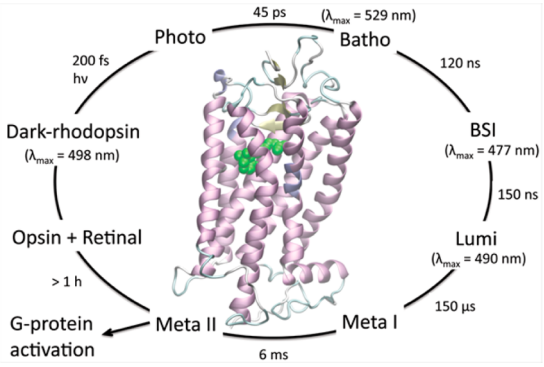


Figure 1. Rhodopsin photoactivation mechanism. The experimental absorption maxima corresponding to the early photointermediates are reported within parentheses.

metarhodopsin II, required for the coupling of the G protein transducin to the receptor.[13,14] On basis of the experimental measurements at physiological temperature,[6] the early intermediates in the photocycle can be characterized and easily distinguished by their absorption maxima; while bathorhodopsin (529 nm, 2.34 eV) displays a red-shifted maximum with respect to the one of dark-state rhodopsin (498 nm, 2.49 eV), BSI has a maximum at 477 nm (2.60 eV), and lumirhodopsin absorbs at a similar wavelength (490 nm, 2.53 eV) as the dark-state.

The publication of the first crystal structure of rhodopsin in its dark-adapted state has been a major contribution toward a better understanding of the first steps of the visual cascade.[3,4] The structures of some of the intermediates in the rhodopsin photocycle have also been determined more recently.[15−17] These crystallographic studies have provided a structure-based framework, contributing to the interpretation of spectroscopic measurements as well as allowing the rationalization and validation of previous theoretical predictions. In addition, insightful structural information in realistic environments and mimicking physiological conditions has been extracted from molecular dynamics (MD) simulations.[18−22] Furthermore, the detailed atomistic understanding gained from these studies has been combined with comprehensive information on the specific thermal stability ranges and temporal intervals corresponding to each of the early photointermediates of the visual transduction pathway to infer many relevant structural details regarding the orientation and adjustment of the retinal PSB inside the protein binding pocket upon photon absorption. Several biophysical techniques such as Fourier transformed infrared (FTIR),[23−25] circular dichroism (CD),[26] nuclear magnetic resonance (NMR),[27−29] and resonance Raman spectroscopy[30,31] have been utilized to this end. On the other hand, rhodopsin crystal structures, while rich in 3D structural information, represent single conformational states and therefore must be viewed as static pictures forming part of a yet to be built "dynamic movie" able to increase our understanding of the factors that regulate activation and spectral tuning along the rhodopsin photocycle.[32]

Color tuning in retinal proteins, especially the so-called opsin shift,[33] has been the subject of numerous experimental[34−41] and theoretical studies.[42−50] These studies, in general, consider three main contributions to explain the origin of the spectral shifts: (i) structural distortions of the chromophore enforced by steric constraints inside the protein binding pocket, (ii) electrostatic influence of the counterion on the chromophore to balance its positive charge, and (iii) interactions between the chromophore and the remaining polar amino acids constituting the active site. Several variables have been used to quantify the relevance of these different contributions. In particular, it has been suggested that the degree of conjugation of the chromophore unsaturated polyene chain (measured by the single/double bond length alternation, BLA), the deviation from planarity (reflected as the variation of different torsion angles forming part of the chromophore molecular structure), the deviation from linearity (calculated from the end-to-tail distance), and the alteration of the distance between the counterion (Glu113) and the retinal PSB are the dominant contributors affecting the color tuning mechanism in rhodopsin.

Nevertheless, not all of these studies agree on the relative importance displayed by individual contributions and, even more, the weight of different residues to those contributions. Experimental studies are limited, as they require protein structural modifications through individual mutations, while theoretical studies are usually restricted to the investigation of only one configuration, without proper statistical sampling of the different contributions. Moreover, although a great part of them have focused on the differences in color tuning mechanisms between rhodopsin and other retinal proteins, a systematic study of the factors that selectively modulate retinal absorption in rhodopsin during the initial events of the visual cascade has not been undertaken, and therefore, a direct relation between the structural changes caused by the relaxation of the photoexcited chromophore upon isomerization and the corresponding optical shifts observed for the intermediates involved in the very early steps of the rhodopsin photoactivation mechanism has not been established yet.

To shed light on the factors that determine spectral tuning along the different intermediates in the early steps of the rhodopsin photocycle, we used a strategy based on combining classical and quantum mechanical/molecular mechanical (QM/MM) MD simulations to take advantage of the strengths of both methodologies. This combined approach allowed us to obtain a dynamical picture of the structural changes involved in the various transitions between the early intermediates (by means of classical MD) as well as to properly describe (using QM/MM MD) the influence of environmental effects created by the surrounding protein residues on the geometrical features of the chromophore in the different photointermediates, the latter being a necessary condition to obtain accurate absorption energies. Furthermore, this strategy permitted us to take into account thermal fluctuations in the structure of both chromophore and protein, and therefore to adequately describe the effects caused by these fluctuations on the spectral bands of the different intermediates while allowing a proper sampling of all statistically relevant configurations of the system.

2. COMPUTATIONAL DETAILS

For each of the four early photointermediates identified in our previous classical MD study,[19] a 5 ps QM/MM Car−Parrinello MD simulation was carried out using the CPMD package.[51] In all the simulations, the electronic structure of the QM subsystem, formed by the retinal PSB and the Lys296 side chain up to the Cδ atom, was described within the density functional theory (DFT) formalism using the PBE exchange-correlation functional,[52] while the remaining atoms, belonging to the MM part, were considered at the classical level. The AMBER/parm99 force field[53] was employed to model standard protein residues, whereas water molecules were described using the TIP3P model.[54] Force field parameters for nonstandard residues and

dx.doi.org/10.1021/ja411303v | J. Am. Chem. Soc. 2014, 136, 3842−3851

lipids from previous studies[22,55] were also used. Valence electrons were described using a plane-wave expansion up to a kinetic energy cutoff of 75 Ry, and soft norm-conserving Martins−Troullier pseudopotentials[56] were employed to represent the interactions between the valence electrons and the ionic cores. A capping hydrogen atom was used to saturate the electronic density of the QM region at the Cδ atom. Long-range electrostatic effects between MM atoms were described using the P3M method[57] with a real space cutoff of 10 Å. Electrostatic interactions between QM and MM regions were taken into account by means of a fully Hamiltonian hierarchical coupling scheme,[58−60] whereas bonded and van der Waals interactions between both subsystems were described at the force field level. The inherent periodicity in the plane-wave calculations was circumvented, solving Poisson's equation for nonperiodic boundary conditions,[61] while periodic boundary conditions were retained for the classical simulation box. The propagation of the equations of motion was performed within the Car−Parrinello scheme[62] using a fictitious electron mass of 500 au and a time step of 0.12 fs. Simulations were carried out in the canonical (NVT) ensemble. A Nosé−Hoover chain of thermostats[63] with a coupling frequency of 2000 cm^{-1} was employed to maintain the temperature at 300 K. Data for analyses, 200 equally spaced snapshots, were collected over the last part of each trajectory.

The accuracy of density functional theory (DFT) functionals for retinal systems has been assessed in some recent theoretical studies.[64,65] In particular, it has been shown[65] for a series of structurally modified 11-*cis*-retinal chromophores that multiconfigurational second-order perturbation theory (CASPT2)-based geometries agree reasonably well with the ones obtained using other correlated *ab initio* methods such as Møller−Plesset second-order perturbation theory (MP2), local second-order approximate coupled cluster singles and doubles (LCC2), as well as DFT approaches. Theoretical studies have also been performed on the dark state of 11-*cis*-retinal in rhodopsin using Variational Monte Carlo (VMC) methodologies.[64] In these studies, it has been reported that VMC and DFT-based geometries present similar features if a reliable QM/MM coupling scheme is employed.

Vertical excitation energies for the different photointermediates were calculated using the ZINDO/S semiempirical method[66] implemented in *Gaussian09*.[67] This method has been calibrated on a large set of compounds, and its parameters have been optimized to give accurate excitation energies for the calculation of absorption spectra in the visible range. In particular, it has been reported that the application of the ZINDO/S methodology to the computation of vertical excitation energies and oscillator strengths on retinal and related polyenals provides results in very good agreement with the experiments.[68] The excitation energies of the four lowest roots were computed and weighted with their corresponding oscillator strengths to build the ZINDO absorption spectra.

A two-step protocol was employed to identify the electronic and geometrical features that determine the spectral shifts between the different photointermediates. In a first step, the dimensionality of the set of variables originally considered was reduced using a feature selection algorithm, and subsequently, the causal structure of this reduced subset of variables was investigated.

The Correlation Based Feature Selection (CBFS) algorithm implemented in the Weka machine learning package[69] was employed to perform attribute selection in order to filter irrelevant, redundant, and noisy geometrical features, thus enabling the subsequent application of causality inference techniques. CBFS is a filter algorithm able to identify the best subset of features from a given data set such that variables highly correlated with the target, yet being uncorrelated to each other, are selected.[70] It ranks feature subsets according to a correlation-based evaluation function:

$$\mathrm{CBFS}_S = \frac{k\langle r_{\mathrm{ft}}\rangle}{\sqrt{k + k(k-1)\langle r_{\mathrm{ff}}\rangle}} \quad (1)$$

where CBFS_S is the heuristic *merit* of the subset S containing k features, $\langle r_{\mathrm{ft}}\rangle$ is the mean feature-target correlation ($f \in S$), and $\langle r_{\mathrm{ff}}\rangle$ is the average feature−feature correlation. Therefore, the numerator of this equation provides an indication of the predictive ability of a given subset of features while the denominator gives a measure of the redundancy among that group of features. In our analysis, we followed a *BestFirst* forward selection strategy to search the attribute space and find optimal subsets. We used an initial set of descriptors that included the single/double bond length alternation, BLA, of the unsaturated polyene chain, as well as all intramolecular (within chromophore), and intermolecular distances, angles, and dihedrals that involve heavy atoms in the chromophore and the surrounding residues. BLA was calculated as the sum of all formal single-bond lengths minus the sum of all formal double-bond lengths along the conjugated chain between C5 and N atoms.

We used the PC-LiNGAM algorithm[71] to infer causal relationships between variables and to estimate the underlying causal structure of our models. This algorithm combines two different approaches, the weakness of the one being the strength of the other: PC (Peter−Clark),[72] which assumes that variables follow multivariate normal distributions and fails in distinguishing between conditional independence equivalent models, and LiNGAM (Linear Non-Gaussian Acyclic Model),[73] which fails if applied to data that are partially Gaussian. We estimated conditional independences between variables by means of the Fisher's z-transform, all statistical tests being performed using a significance level of 0.001.

To determine the physical nature of some of the more relevant interactions between the chromophore and the surrounding protein residues, we employed the NCI (Non-Covalent Interactions) index,[74,75] which provides an intuitive and visual way to detect weak interactions in chemical systems. This approach is based on the analysis of the electron density, ρ, and its reduced gradient, s, where

$$s = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}} \quad (2)$$

and permits highlighting interactions characterized by a low-density regime. The analysis of the second derivatives of the density along the main axis of variation allows distinguishing between different interaction types. In particular, nonbonded interactions, such as steric repulsion, are identified by the positive sign ($\lambda_2 > 0$) of the second eigenvalue, λ_2, of the electron density Hessian matrix; bonding interactions, such as hydrogen bonds, are characterized by a negative sign ($\lambda_2 < 0$); van der Waals interactions present a negligible density overlap that gives $\lambda_2 \cong 0$.

3. RESULTS AND DISCUSSION

In our previous study about the role of dimerization in rhodopsin signal transduction,[19] we have provided a dynamical picture of the geometrical changes that take place in the protein binding pocket as a consequence of chromophore relaxation upon light absorption. We were able to identify several intermediates within a few tens of nanoseconds following photoexcitation by following the time-evolution of appropriately defined structural order parameters. In particular, the distance between the ionone ring of the chromophore and Ala169 was used to this end (Figure 2). Remarkably, the time scales corresponding to the lifetimes of these inactive intermediates involved in the early steps of the rhodopsin signal transduction pathway are in good agreement with those experimentally determined.[76,77] However, in spite of this good agreement, a conclusive and unambiguous assignment of the different intermediates identified in the simulations with the ones experimentally detected requires the calculation of the optical absorption as a primary experimental quantity. Furthermore, the demonstration of this correspondence between the experimentally measured and the theoretically identified intermediates would also support the validity of the structural model obtained for BSI, for which no X-ray structure is available.

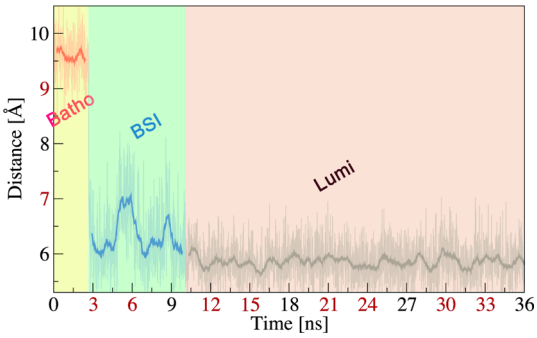


Figure 2. Time evolution of the distance between the ionone ring of the chromophore and Ala169 upon photoinduced cycloisomerization (see ref 19 for details).

In the current study, the protocol described below was followed as an attempt to establish a one-to-one correspondence between these intermediates and the most reliable primary experimental quantity available: the absorption energy. Initially, DFT-based Quantum Mechanical/Molecular Mechanical (QM/MM) molecular dynamics (MD) simulations, one for each of the early photointermediates (dark, batho, BSI, lumi), were carried out using selected snapshots extracted from our previous classical MD study[19] as starting configurations. This approach permits the generation of a significant number of thermally sampled conformations while maintaining the accuracy required to obtain reliable results in the geometrical features of the retinal PSB.[78,79] Subsequently, vertical electronic excitation energies between the ground state (S_0) and the first excited singlet state (S_1) were computed utilizing the ZINDO/S method, whose ability to get reliable results in this system has recently been validated,[80] for a converged number of configurations extracted from the QM/MM trajectories. Finally, the computed excitation energies were used to build the absorption spectra for the different intermediates, from which the values corresponding to the different absorption maxima for each photointermediate were determined and directly compared to the ones measured in experiments at physiological temperature.

3.1. Optical Spectra of Rhopodsin Early Photointermediates. ZINDO/S excitation energies for the different intermediates (200 snapshots were extracted from each QM/MM trajectory) were calculated on model systems composed by the chomophore together with all protein residues at a distance shorter than 5 Å from every atom of the retinal PSB (about 40 residues comprising 720 atoms). The optical spectra that correspond to these extended models (EM) are displayed in Figure 3, while the excitation energy values associated with the absorption maxima for each intermediate, and their relative shifts are reported in Table 1. Remarkably, the broadening of the spectral bands displayed in Figure 3 is in good agreement with the one obtained when the same number of equally spaced snapshots extracted from much longer classical MD trajectories is employed to compute the absorption spectra (see Figure S1 in Supporting Information), which corroborates the adequate convergence of the spectral properties obtained from the QM/MM MD simulations in spite of their relatively shorter sampling time. As shown in Table S1 in Supporting Information, using a single configuration for each photointermediate does not give accurate results and the calculation of the absorption maxima by averaging along the dynamics is required. Inspection of the data presented in Table 1 reveals

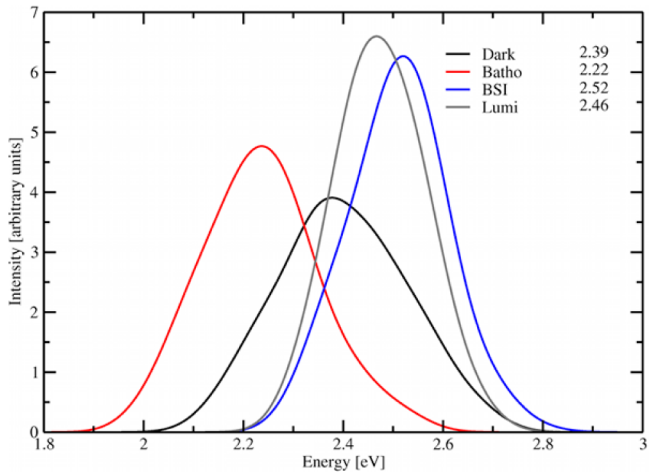


Figure 3. ZINDO/S optical spectra for the early rhodopsin photointermediates. Values of the absorption maxima (in the upper right corner) are given in eV.

Table 1. Experimental and ZINDO/S Absorption Maxima and Relative Shifts Given in eV; Gaussian Broadened Transition Width As a Function of the Size of the QM Model; Chromophore-Lys296 Model (M0), Reduced Model (M9), and Extended Model (EM)

	M0	M9	EM	Exp
λ_{max}(Batho)	2.04	2.23	2.22	2.34
$\Delta\lambda_{max}$(Batho/Dark)	0.10	0.17	0.17	0.15
λ_{max}(Dark)	2.14	2.40	2.39	2.49
$\Delta\lambda_{max}$(Dark/Lumi)	0.04	0.07	0.07	0.04
λ_{max}(Lumi)	2.18	2.47	2.46	2.53
$\Delta\lambda_{max}$(Lumi/BSI)	0.04	0.04	0.06	0.07
λ_{max}(BSI)	2.22	2.51	2.52	2.60

that, in spite of the approximate nature of the Hamiltonian employed by ZINDO/S, the agreement between the values computed using the extended models and the ones experimentally determined is remarkably good. The excitation energies are consistently underestimated by about 0.1 eV. Thus, not only qualitative trends, going from the red-shifted absorption maximum computed for bathorhodopsin to the blue-shifted one found for BSI, but also relative optical shifts between intermediates are well reproduced, which strongly supports the assignment of the structurally different species identified in the simulations with the experimentally detected ones.

To estimate the influence of individual residues on the ZINDO/S electronic excitation energies and therefore probe their effect on the vertical absorption, we performed a depletion analysis according to which the residues included in the extended models (EM) were removed one at a time. The results of this analysis for every intermediate, depicted in Figure 4 as energy differences relative to the original model without deletions, reveal that the two negatively charged amino acids (Glu113 and Glu181) and also Ala292 make significant contributions to the total excitation energies, their presence causing a blue-shift in the optical spectra as evident from their negative ΔE values. An opposite and smaller effect is observed upon deletion of residues Gly114, Ala117, Thr118, Ser186, Cys187, and Tyr268, which present positive ΔE values, while the contributions of all remaining amino acids is very small (lower than 0.01 eV). The negligible effect caused by the latter

dx.doi.org/10.1021/ja411303v | J. Am. Chem. Soc. 2014, 136, 3842−3851

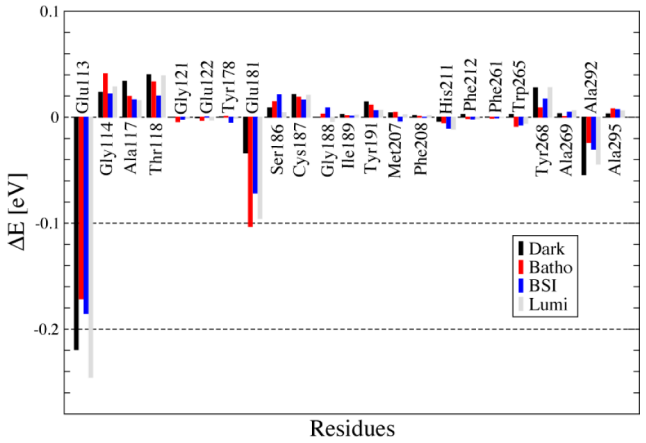


Figure 4. Contribution of individual residues to the ZINDO/S electronic excitation energies shown as energy differences relative to the original model without deletions (EM). Residues are removed one at a time from each of the photointermediate structures.

set of residues was further verified by means of additional ZINDO/S spectral calculations performed on the reduced model (M9) displayed in Figure 5 in which only the amino

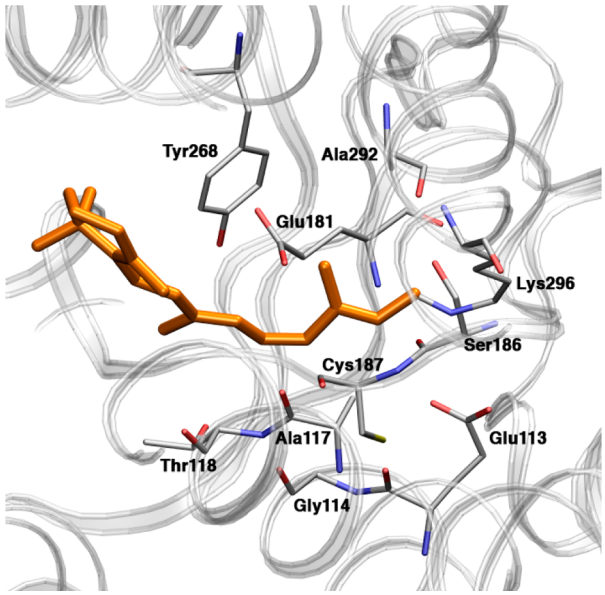


Figure 5. Amino acids included in the reduced model (M9). The chromophore is depicted in orange. Hydrogen atoms are omitted for clarity.

acids that contribute with values greater than 0.01 eV were included. The results from these calculations are also collected in Table 1, where the values associated with the absorption maxima for the various intermediates and their relative shifts are reported. Interestingly, in a previous theoretical study,[81] many of the residues mentioned above were found to modify the absorption spectra in dark state rhodopsin in a similar way as reported here, thus serving as further validation of the methodology here employed, the only exceptions being Gly114, Cys187, and Ala295 that were not considered in that study.

The convergence of the absorption energy values with respect to the system size is displayed in Figure 6, which clearly shows that the spectral shifts between different intermediates

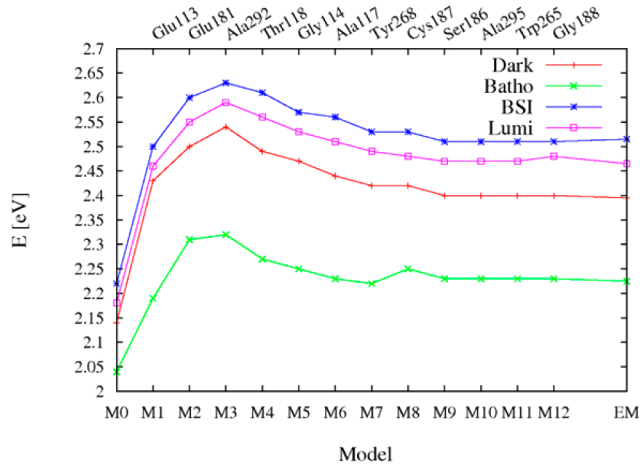


Figure 6. Convergence of the ZINDO/S vertical excitation energies with respect to the system size. Residues are added to the smallest (M0) model, which only includes Lys296 and the chromophore, according to their average contribution in the depletion analysis shown in Figure 4. The residues that are consecutively added to the M0 model in the sequence of decreasing contributions are shown on the upper axis.

are already present to a large extent when the minimal M0 model that only contains the Lys296 and the chromophore itself is used in the calculations. This fact suggests that the origin of the color tuning between the rhodopsin photointermediates can be mainly ascribed to changes in intrinsic properties of the retinal PSB, its structural features and charge distribution being modulated by the nearest protein environment. On the other hand, it is also evident from Figure 6 that the enlargement of the original model through the inclusion of Glu113 causes a blue-shifting effect in the optical spectra that is gradually increased by the consecutive addition of Glu181 and Ala292, and finally compensated in part by the presence of other residues (Gly114, Ala117, Thr118, etc.) that promotes a smooth red-shift in the spectra until system size convergence is reached at model size M9.

3.2. Determinants of the Spectral Shifts between Rhodopsin Early Photointermediates. The above-mentioned results imply that all structural determinants that modulate the color tuning between the rhodopsin early intermediates are likely to be present in our reduced model (M9). Therefore, we decided to investigate the electronic and geometrical features that cause the spectral shifts between the different photointermediates based on this model. This is an ambitious objective mainly due to two reasons. First, because the problem of inferring causal relations between variables from purely observational data is a very challenging issue that requires the usage of sophisticated algorithms. Second, owing to the very large number of geometrical features (distances, angles, and dihedrals) that emerges even from our reduced model, which makes it nonviable to reliably solve the causality problem unless using attribute selection techniques to reduce the dimensionality of the original set. In particular, the set of attributes that fully characterizes the spatial relationships between the constituents of this model is formed by more than 4×10^6 descriptors. This set includes the single/double bond length alternation (BLA) of the unsaturated polyene chain, as well as all intramolecular (within chromophore), and intermolecular distances, angles, and dihedrals that involve heavy atoms in the chromophore and the surrounding residues.

dx.doi.org/10.1021/ja411303v | J. Am. Chem. Soc. 2014, 136, 3842−3851

We used Correlation-Based Feature Selection (CBFS),[70] a filter algorithm that ranks feature subsets according to a correlation based heuristic evaluation function (see Computational Details section for details), to solve the dimensionality reduction problem. The bias of this evaluation function is toward subsets that contain attributes that are highly correlated with the target, the absorption spectra, while uncorrelated with each other. In this way, irrelevant, redundant, and noisy attributes are screened out, therefore allowing the identification of a representative subset of features that globally maintains a predictive ability similar to the one presented by the original set. Using this strategy, we were able to identify the relevant features contained in the highly correlated original set. Only 13 descriptors survived after application of the attribute selection protocol. The degree of correlation between the ZINDO/S electronic excitation energies and each of these descriptors is displayed in Figure 7 (features 1−13). The largest correlation

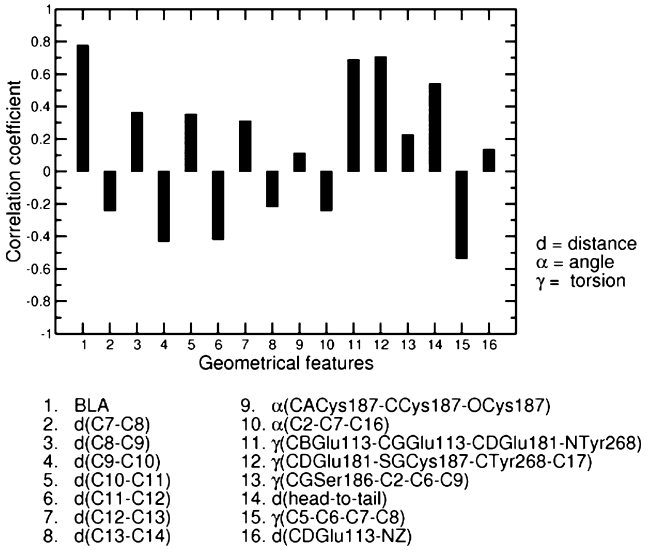


Figure 7. Degree of correlation between the ZINDO/S electronic excitation energies and various geometrical parameters.

coefficients were found for the BLA (0.78), for the γ(CDGlu181−SGCys187−CTyr268-C17) torsion angle, which gives a measure of the arrangement of the β-ionone ring relative to the position of the polyene chain and the location of the Glu181 (0.71), and for the γ(CBGlu113−CGGlu181−CDGlu181-NTyr268) torsion angle, which gives an idea of the orientation of Glu181 with respect to the chromophore and the other residues in the binding site (0.69).

Other geometrical features that, in previous studies, have been suggested to affect the spectral tuning mechanism in the rhodopsin family of photoreceptors: (i) the head-to-tail distance, (ii) the γ(C5−C6−C7−C8) torsion angle, and (iii) the distance between the Cδ atom of the counterion (Glu113) and the N atom of the retinal PSB were also added to this reduced set of descriptors in order to investigate causal relationships between them, even if no relevant correlation, 0.14, (see Figure 7, features 14−16) was found for the latter. Causality was inferred using the PC-LiNGAM algorithm,[71] which has been reported to work well to estimate dependency structures in linear acyclic causal networks with arbitrary (both Gaussian and non-Gaussian) distributions. This algorithm determines whether a particular variable influences another and, therefore, can be used to discover the Markov Blanket (MB) of a given variable, which is formed within this context by its parents, its children, and its children's other parents. The MB of our target variable (the absorption spectra) together with the edge strengths, which characterize how important the relationship between different features is, is displayed in Figure 8. The causal graph displayed in this figure, which was built

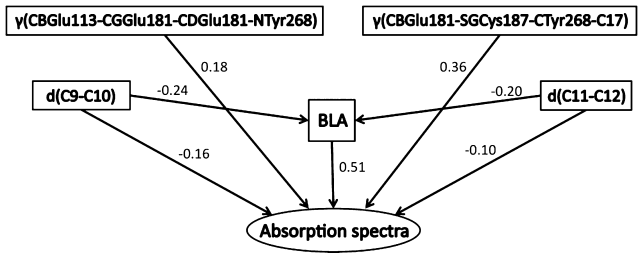


Figure 8. Markov Blanket of the absorption spectra. The edge strength indicates the relative magnitude of the dependency between two variables, given the other inter-relationships.

after standardization (zero mean and unit variance) of all the variables, clearly indicates the important role played by the BLA and some of the C−C bond distances that appear in its definition in tuning the spectral properties of the early intermediates along the rhodopsin photocycle. A statistically significant influence on the absorption spectra was also found for the γ(CDGlu181−SGCys187−CTyr268−C17) and γ(CBGlu113−CGGlu181−CDGlu181−NTyr268) torsion angles that, as mentioned above, give an indication of the arrangement of the β-ionone ring and the Glu181 with respect to the location of the polyene chain of the chromophore and the other residues in the binding pocket. The average value of the characteristic features shown in Figure 8 in the different photointermediates, as well as their effects on the corresponding optical spectra relative to the dark-state, are displayed in Table 2.

Table 2. Mean Values and Standard Deviations (within parentheses) of the Characteristic Features Shown in Figure 8 for the Different Photointermediates[a]

	Dark	Batho	BSI	Lumi
BLA	-1.05 (0.10)	-1.11 (0.10)	-1.03 (0.09)	-1.06 (0.07)
d_1	1.39(0.03)	1.39(0.02)	1.38(0.02)	1.39(0.02)
d_2	1.38(0.02)	1.39(0.02)	1.38(0.01)	1.38(0.01)
γ_1	115.0 (3.1)	93.4 (6.7)	123.4 (4.0)	121.7 (2.0)
γ_2	144.3 (7.4)	126.8 (6.7)	163.7 (3.9)	158.5 (4.1)

[a]Colors indicate the direction of the spectral shifts caused by each feature with respect to their values in dark-state rhodopsin. Red shifting effect (in red), blue shifting effect (in blue). d_1, d_2 are the d(C9−C10) and d(C11−C12) distances (in Å) while γ_1, γ_2 are the γ(CBGlu113−CGGlu181−CDGlu181−NTyr268) and γ(CDGlu181−SGCys187−CTyr268−C17) torsion angles (in deg).

To get a deeper understanding on how the protein environment influences the absorption spectra through effects on the BLA, we also built the MB of that variable (BLA). Moreover, we also inferred the MB of all geometrical features that affect BLA or any of the other variables that were found (Figure 8) to tune the spectral properties of the rhodopsin early photointermediates. To this end, we used a procedure similar

dx.doi.org/10.1021/ja411303v | J. Am. Chem. Soc. 2014, 136, 3842−3851

to the one outlined above when the absorption spectra were employed as target variable, i.e., feature selection followed by causality inference. The addition of these new variables and their connection strengths (Figure 9) to the original graph give

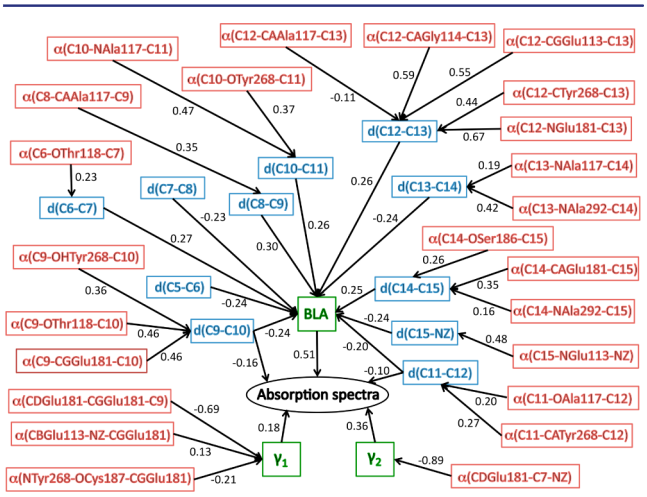


Figure 9. Causal graph inferred from the data using the PC-LiNGAM algorithm. Intramolecular features that cause a direct effect on BLA are depicted in blue, while intermolecular ones, whose effect is indirect, are colored in red. The edge strength indicates the relative magnitude of the dependency between two variables, given the other interrelationships. γ_1 and γ_2 are used to abbreviate γ(CBGlu113−CGGlu181−CDGlu181−NTyr268) and γ(CDGlu181−SGCys187−CTyr268−C17) torsion angles, respectively.

a more complete picture of the mechanisms that cause the color tuning between rhodopsin photointermediates and evidence the role played by the residues included in our final model (M9). As shown in Figure 9, key distances between the atoms that constitute the polyene chain of the chromophore are altered by the location and orientation of various of these active site residues, which in turn directly or indirectly (through BLA modifications) modulate the optical spectra.

The quantitative knowledge of the dependencies between the variables that compose this causal graph allows the generation of a structural model able to link the structural parameters obtained from the MD simulations with their corresponding excitation energies. Thus, the ZINDO/S excitation energy for a given snapshot extracted from the MD trajectories can be inferred from a simple linear combination of the geometrical descriptors that, according to our analysis, cause a direct effect on the spectra:

$$E = a^{*}\mathrm{BLA} + b^{*}d_1 + c^{*}d_2 + d^{*}\gamma_1 + e^{*}\gamma_2 + f \qquad (3)$$

where d_1, d_2 are the d(C9−C10) and d(C11−C12) distances (measured in Å) and γ_1, γ_2 are the γ(CBGlu113−CGGlu181−CDGlu181−NTyr268) and γ(CDGlu181−SGCys187−CTyr268−C17) torsion angles (given in degrees), respectively. E is the ZINDO/S predicted excitation energy (in eV), and the coefficients required for its estimation are reported in Table 3.

The predictive ability of this model was investigated using frames extracted from both the QM/MM and classical MD simulations reported above. The tests performed (see Supporting Information for details) indicate a high predictive power and reliability of this structural model and, therefore, support the usage of the aforementioned equation to estimate ZINDO/S excitation energies from this small set of geometrical descriptors within the range in which this equation has been validated.

Table 3. Values for the Coefficients in the Structural Equation Model

coefficient	value	units
a	0.6498	eV/Å
b	−0.7574	eV/Å
c	−0.5545	eV/Å
d	0.0022	eV/deg
e	0.0029	eV/deg
f	4.2460	eV

3.3. Nature of the Interactions Responsible for the Color Tuning between Rhodopsin Early Photointermediates.

As already suggested by the results from the depletion analysis previously described, the connections between the different geometrical features that form the causal graph in Figure 9 demonstrate that the spectral tuning between the early intermediates in the rhodopsin photocycle is mainly governed by modifications in intrinsic properties of the chromophore structure itself, which is altered by changes in the intermolecular interactions between the retinal PSB and the surrounding residues as a consequence of the structural relaxation that takes place in the chromophore upon photoinduced isomerization. The effect of the charged residues (Glu113, Glu181) on the C9−C10, C12−C13, C14−C15, and C15−NZ distances (see Figure 9) can be expected to be mostly electrostatic, although dipolar interactions could also play an important role. To reveal the nature of the interactions between the chromophore and the other protein residues identified as relevant in our study, we carried out an NCI (noncovalent interactions) analysis[74] on the structures in the absorption maxima of the various photointermediates. The NCI index is based on the electron density, ρ, and its derivatives and enables visualizing and distinguishing both stabilizing and destabilizing noncovalent interactions. As shown in Figure 10 for the case of dark-state rhodopsin, the results of the NCI analysis evidence that, in addition to the strong hydrogen bond formed with Glu113, other interactions such as π-stacking with Ala117 and Thr118 backbone atoms, van der Waals contacts with Gly114 and Ala292, and CH/π weak hydrogen-bond interactions with Tyr268, Ala117, Thr118, and Ser186 side chains are also involved in chromophore stabilization. Similar kinds of interactions were found when the same analysis was performed on the other intermediates (batho, BSI, and lumi).

Remarkably, although the similarity in amino acid sequence between rhodopsin and other visual pigments such as red, green, and blue cones pigments is low (about 40%),[82] many of the residues that according to our analysis modulate the spectral tuning between the early rhodopsin photointermediates are highly conserved in all pigments. As shown in Figure 11, only a few but crucial differences are present in key residues. These differences can provide a rationale for the spectral shifts between the visual pigments. On the one hand, the experimentally found blue-shifting effect of A292S and A117G mutations[83] contributes to the rationalization of the absorption maximum reported for blue cone (2.91−2.99 eV),[84−86] which is blue-shifted with respect to the one of dark-state rhodopsin (2.49 eV). On the other hand, the mutation of Glu181 into either His or Tyr should diminish the important blue-shifting effect in the optical spectra we found associated to the Glu residue. This mutation could partially

dx.doi.org/10.1021/ja411303v | J. Am. Chem. Soc. 2014, 136, 3842−3851

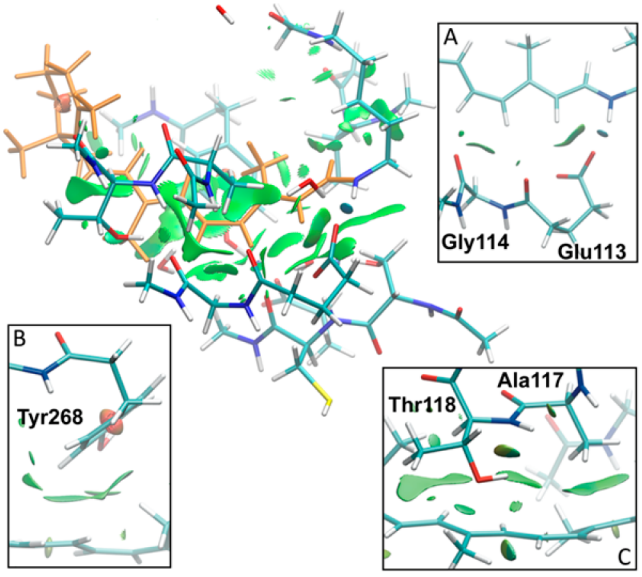


Figure 10. Noncovalent interactions between the retinal chromophore and the residues around it. The chromophore is depicted in orange while the noncovalent interactions are represented by gradient isosurfaces (s = 0.3 au) colored according to the interaction strength (values of sign (λ_2) ranging from −0.05 to +0.05 au) using a blue−green−red scale, such that *blue*, *green*, and *red* indicate strong attractive interactions, weak van der Waals contacts, and strong nonbonded overlaps, respectively. The interactions between some selected residues and the chromophore are enlarged in insets A−C.

	113	118	181	187	268	292
Rhodopsin	E	T	E	C	Y	A
Blue cone	E	T	E	C	Y	S
Green cone	E	S	Y	C	Y	S
Red cone	E	S	H	C	Y	A
	Helix III		E-II loop		Helix VI	Helix VII

Figure 11. Partial sequence alignment between rhodopsin, and blue, green, and red cone pigments. Only residues belonging to the final M9 model are included. Differences are highlighted in yellow. The residue numbers correspond to the rhodopsin sequence.

explain the red-shifted absorption maxima, compared to the one in rhodopsin, that have been reported for green (2.33 eV),[84] and red (2.21 eV)[84] cones. The blue-shifting effect created by the A292S mutation would compensate in some degree the opposite effect caused by Glu181 mutation and explain why the absorption maximum experimentally found for green cones is blue-shifted with respect to the one measured for red cones. Therefore, the picture that emerges from the sequence alignment between rhodopsin, and blue, green, and red cone pigments insinuates that the residues which we found to be responsible for the color tuning between rhodopsin early photointermediates could also play a similar role in the spectral shifts observed between different visual pigments. This suggests that modifications in the interactions between these key amino acids and the chromophore, either due to mutations of some key residues or simply owing to alterations in the surrounding residues able to change the hydrogen-bond pattern between the active site members, may be the characteristic mechanism used by retinal proteins to tune their spectral properties.

4. CONCLUSIONS

In the present work, a one-to-one correspondence between the structures of the intermediates involved in the early steps of the rhodopsin photocycle (dark, batho, BSI, lumi) and their optical spectra was established. To this end, a combined strategy based on the computation of absorption energies, using the ZINDO/S semiempirical method, for a statistically relevant number of configurations extracted from QM/MM trajectories was employed. The use of this strategy yields a remarkably good agreement between computed and experimentally determined spectral bands, the excitation energies corresponding to the absorption maxima being consistently underestimated by only 0.05−0.11 eV. Moreover, a systematic statistical analysis of the data extracted from the aforementioned trajectories allowed us to identify the main features affecting the absorption spectra.

Three main results emerge as a consequence of the analysis of our data: (i) the origin of the color tuning between the early intermediates along the rhodopsin photocycle is mainly governed by modifications in intrinsic properties of the structure of the retinal protonated Schiff base, which is modulated by changes that take place in the nearest protein environment as a consequence of chromophore relaxation upon photoinduced isomerization; (ii) several residues located in the protein-binding pocket shift the absolute value of the absorption maxima in all intermediates to a similar extent. In particular, the blue-shifting effect caused by Glu113, Glu181, and Ala292 residues is compensated in part by the presence of Gly114, Ala117, Thr118, Ser186, Cys187, Tyr191, and Tyr268, which promote a red-shift in the optical spectra; (iii) in addition to the expected electrostatic and dipolar interactions caused by Glu113 and Glu181, and the strong hydrogen bond formed with Glu113, other interactions such as π-stacking with Ala117 and Thr118 backbone atoms, van der Waals contacts with Gly114 and Ala292, and CH/π weak interactions with Tyr268, Ala117, Thr118, and Ser186 side chains are also involved in chromophore stabilization and, as a consequence, play a non-negligible role in modulating the spectral properties of the early photointermediates involved in the rhodopsin signal transduction pathway.

■ ASSOCIATED CONTENT

Supporting Information

ZINDO/S optical spectra computed from classical MD trajectories, excitation energies calculated from annealed structures, and validation of the structural equation model. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

ursula.roethlisberger@epfl.ch

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The CSCS (Manno, Switzerland) is acknowledged for providing access to its supercomputing facilities. Support from the Swiss National Science Foundation (Grant No. 200020-130082) and the NCCR-MUST interdisciplinary research program is also gratefully acknowledged.

REFERENCES

(1) Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K. *Nature* **2009**, *459*, 356−363.
(2) Palczewski, K. *Annu. Rev. Biochem.* **2006**, *75*, 743−767.
(3) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739−745.
(4) Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V. *J. Mol. Biol.* **2004**, *342*, 571−583.
(5) Gomperts, B. D.; Kramer, I. M.; Tatham, P. E. R. *Signal Transduction*; Elsevier/Academic Press: Amsterdam, 2009; pp 159−183.
(6) Randall, C. E.; Lewis, J. W.; Hug, S. J.; Björling, S. C.; Eisner-Shanas, I.; Friedman, N.; Ottolenghi, M.; Sheves, M.; Kliger, D. S. *J. Am. Chem. Soc.* **1991**, *113*, 3473−3485.
(7) Hug, S. J.; Lewis, J. W.; Einterz, C. M.; Thorgeirsson, T. E.; Kliger, D. S. *Biochemistry* **1990**, *29*, 1475−1485.
(8) Sasaki, N.; Tokunaga, F.; Yoshizawa, T. *FEBS Lett.* **1980**, *114*, 1−3.
(9) Sasaki, N.; Tokunaga, F.; Yoshizawa, T. *Biochim. Biophys. Acta* **1983**, *722*, 80−87.
(10) Jager, S.; Lewis, J. W.; Zvyaga, T. A.; Szundi, I.; Sakmar, T. P.; Kliger, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8557−8562.
(11) Schoenlein, R. W.; Peteanu, L. A.; Mathies, R. A.; Shank, C. V. *Science* **1991**, *254*, 412−415.
(12) Kim, J. E.; Tauber, M. J.; Mathies, R. A. *Biochemistry* **2001**, *40*, 13774−13778.
(13) Knierim, B.; Hofmann, K. P.; Ernst, O. P.; Hubbell, W. L. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20290−20295.
(14) Mahalingam, M.; Martínez-Mayorga, K.; Brown, M. F.; Vogel, R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17795−17800.
(15) Nakamichi, H.; Okada, T. *Angew. Chem., Int. Ed.* **2006**, *45*, 4270−4273.
(16) Schreiber, M.; Sugihara, M.; Okada, T.; Buss, V. *Angew. Chem., Int. Ed.* **2006**, *45*, 4274−4277.
(17) Nakamichi, H.; Okada, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12729−12734.
(18) Pitman, M. C.; Grossfield, A.; Suits, F.; Feller, S. E. *J. Am. Chem. Soc.* **2005**, *127*, 4576−4577.
(19) Neri, M.; Vanni, S.; Tavernelli, I.; Rothlisberger, U. *Biochemistry* **2010**, *49*, 4827−4832.
(20) Röhrig, U. F.; Guidoni, L.; Laio, A.; Frank, I.; Rothlisberger, U. *J. Am. Chem. Soc.* **2004**, *126*, 15328−15329.
(21) Saam, J.; Tajkhorshid, E.; Hayashi, S.; Schulten, K. *Biophys. J.* **2002**, *83*, 3097−3112.
(22) Röhrig, U. F.; Guidoni, L.; Rothlisberger, U. *Biochemistry* **2002**, *41*, 10799−10809.
(23) Ganter, U. M.; Kashima, T.; Sheves, M.; Siebert, F. *J. Am. Chem. Soc.* **1991**, *113*, 4087−4092.
(24) Ganter, U. M.; Gartner, W.; Siebert, F. *Biochemistry* **1988**, *27*, 7480−7488.
(25) Kandori, H.; Maeda, A. *Biochemistry* **1995**, *34*, 14220−14229.
(26) Thomas, Y. G.; Szundi, I.; Lewis, J. W.; Kliger, D. S. *Biochemistry* **2009**, *48*, 12283−12289.
(27) Concistre, M.; Gansmuller, A.; McLean, N.; Johannessen, O. G.; Montesinos, I. M.; Bovee-Geurts, P. H. M.; Verdegem, P.; Lugtenburg, J.; Brown, R. C. D.; DeGrip, W. J.; Levitt, M. H. *J. Am. Chem. Soc.* **2008**, *130*, 10490−10491.
(28) Carravetta, M.; Zhao, X.; Johannessen, O. G.; Lai, W. C.; Verhoeven, M. A.; Bovee-Geurts, P. H. M.; Verdegem, P. J. E.; Kiihne, S.; Luthman, H.; de Groot, H. J. M.; deGrip, W. J.; Lugtenburg, J.; Levitt, M. H. *J. Am. Chem. Soc.* **2004**, *126*, 3948−3953.
(29) Smith, S. O.; Courtin, J.; deGroot, H.; Gebhard, R.; Lugtenburg, J. *Biochemistry* **1991**, *30*, 7409−7415.
(30) Pan, D. H.; Ganim, Z.; Kim, J. E.; Verhoeven, M. A.; Lugtenburg, J.; Mathies, R. A. *J. Am. Chem. Soc.* **2002**, *124*, 4857−4864.
(31) Palings, I.; Pardoen, J. A.; Vandenberg, E.; Winkel, C.; Lugtenburg, J.; Mathies, R. A. *Biochemistry* **1987**, *26*, 2544−2556.
(32) Vanni, S.; Rothlisberger, U. *Curr. Med. Chem.* **2012**, *19*, 1135−1145.
(33) Nakanishi, K.; Balogh-Nair, V.; Arnaboldi, M.; Tsujimoto, K.; Honig, B. *J. Am. Chem. Soc.* **1980**, *102*, 7945−7947.
(34) Zhukovsky, E. A.; Oprian, D. D. *Science* **1989**, *246*, 928−930.
(35) Smith, S. O.; Palings, I.; Miley, M. E.; Courtin, J.; DeGroot, H.; Lugtenburg, J.; Mathies, R. A.; Griffin, R. G. *Biochemistry* **1990**, *29*, 8158−8164.
(36) Asenjo, A. B.; Rim, J.; Oprian, D. D. *Neuron.* **1994**, *12*, 1131−1138.
(37) Chan, T.; Lee, M.; Sakmar, T. P. *J. Biol. Chem.* **1992**, *267*, 9478−9480.
(38) Janz, J. M.; Farrens, D. L. *Biochemistry* **2001**, *40*, 7219−7227.
(39) Kochendoerfer, G. G.; Wang, Z.; Oprian, D. D.; Mathies, R. A. *Biochemistry* **1997**, *36*, 6577−6587.
(40) Lin, S. W.; Kochendoerfer, G. G.; Carroll, K. S.; Wang, D.; Mathies, R. A.; Sakmar, T. P. *J. Biol. Chem.* **1998**, *273*, 24583−24591.
(41) Neitz, M.; Neitz, J.; Jacobs, G. H. *Science* **1991**, *252*, 971−974.
(42) Bravaya, K.; Bochenkova, A.; Granovsky, A.; Nemulkhin, A. *J. Am. Chen. Soc.* **2007**, *129*, 13035−13042.
(43) Coto, P. B.; Strambi, A.; Ferre, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17154−17159.
(44) Sekharan, S.; Sugihara, M.; Buss, V. *Angew. Chem., Int. Ed.* **2007**, *46*, 269−271.
(45) Sekharan, S.; Katayama, K.; Kandori, H.; Morokuma, K. *J. Am. Chem. Soc.* **2012**, *134*, 10706−10712.
(46) Frähmcke, J. S.; Wanko, M.; Elstner, M. *J. Phys. Chem. B* **2012**, *116*, 3313−3321.
(47) Trabanino, R. J.; Vaidehi, N.; Goddard, W. A., III. *J. Phys. Chem. B* **2006**, *110*, 17230−17239.
(48) Fujimoto, K.; Hasegawa, J. Y.; Hayashi, S.; Kato, S.; Nakatsuji, H. *Chem. Phys. Lett.* **2006**, *414*, 239−242.
(49) Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. *J. Am. Chem. Soc.* **2006**, *128*, 10808−10818.
(50) Altun, A.; Yokoyama, S.; Morokuma, K. *J. Phys. Chem. B* **2008**, *112*, 6814−6827.
(51) *CPMD*, V3.13.2; http://www.cpmd.org, Copyright IBM Corp 1990−2008, Copyright MPI für Festkörperforschung Stuttgart 1997−2001.
(52) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(53) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5191.
(54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Comput. Chem.* **1983**, *79*, 926−935.
(55) Vanni, S.; Neri, M.; Tavernelli, I.; Rothlisberger, U. *Biochemistry* **2009**, *48*, 4789−4797.
(56) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993−2006.
(57) Hünenberger, P. H. *J. Chem. Phys.* **2000**, *113*, 10464−10476.
(58) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941−6947.
(59) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 7300−7307.
(60) Laio, A.; Gervasio, F. L.; VandeVondele, J.; Sulpizi, M.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 7963−7968.
(61) Martyna, G. J.; Tuckerman, M. E. *J. Chem. Phys.* **1999**, *110*, 2810−2821.
(62) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.
(63) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511−519.
(64) Coccia, E.; Varsano, D.; Guidoni, L. *J. Chem. Theory Comput.* **2013**, *9*, 8−12.
(65) Walczak, E.; Szefczyk, B.; Andruniow, T. *J. Chem. Theory Comput.* **2013**, *9*, 4915−4927.
(66) Ridley, J.; Zerner, M. *Theor. Chim. Acta* **1973**, *32*, 111−134.

(67) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian09*, Revision A.2; Gaussian, Inc.: Wallingford, CT, 2009.
(68) López, C. S.; Faza, O. N.; Estévez, S. L.; de Lera, A. R. *J. Comput. Chem.* **2006**, *27*, 116−123.
(69) Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. A. *SIGKDD Explorations* **2009**, *11*, 10−18.
(70) Hall, M. A. Correlation-based feature selection for machine learning. Ph.D. Dissertation, University of Waikato: Hamilton, NZ, 1999.
(71) Hoyer, P. O.; Hyvarinen, A.; Scheines, R., Spirtes, P.; Ramsey, J.; Lacerda, G.; Shimizu, S. In *Proceedings of the 24th Annual Conference on Uncertainty in Artificial Intelligence, Helsinki, July 9−12, 2008*; McAllester, D., Myllymaki, P., Eds.; AUAI Press: OR, 2008; pp 282−289.
(72) Spirtes, P.; Glymour, C.; Scheines, R. *Causation, Prediction, and Search*, 2nd ed.; The MIT Press: Cambridge, MA, 2000.
(73) Shimizu, S.; Hyvarinen, A.; Hoyer, P.; Kerminen, A. *J. Mach. Learn. Res.* **2006**, *7*, 2003−2030.
(74) Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. *J. Am. Chem. Soc.* **2010**, *132*, 6498−6506.
(75) Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. *J. Chem. Theory Comput.* **2011**, *7*, 625−632.
(76) Okada, T.; Ernst, O. P.; Palczewski, K.; Hofmann, K. P. *Trends Biochem. Sci.* **2001**, *26*, 318−324.
(77) Shichida, Y.; Imai, H. *Cell. Mol. Life Sci.* **1998**, *54*, 1299−1315.
(78) Blomgren, F.; Larsson, S. *J. Comput. Chem.* **2005**, *26*, 738−742.
(79) Altun, A.; Yokoyama, S.; Morokuma, K. *Photochem. Photobiol.* **2008**, *84*, 845−854.
(80) Valsson, O.; Campomanes, P.; Tavernelli, I.; Rothlisberger, U.; Filippi, C. *J. Chem. Theory Comput.* **2013**, *9*, 2441−2454.
(81) Tomasello, G.; Olaso-González, G.; Altoè, P.; Stenta, M.; Serrano-Andrés, L.; Merchán, M.; Orlandi, G.; Bottoni, A.; Garavelli, M. *J. Am. Chem. Soc.* **2009**, *131*, 5172−5186.
(82) Stenkamp, R. E.; Filipek, S.; Driessen, C. A. G. G.; Teller, D. C.; Palczewski, K. *Biochim. Biophys. Acta* **2002**, *1565*, 168−182.
(83) Lin, S. W.; Kochendoerfer, G. G.; Carroll, K. S.; Wang, D.; Mathies, R. A.; Sakmar, T. P. *J. Biol. Chem.* **1998**, *273*, 24583−24591.
(84) Oprian, D. D. *Biochemistry* **1991**, *30*, 11367−11372.
(85) Merbs, S. L.; Nathans, J. *Nature* **1992**, *356*, 433−435.
(86) Fasick, J. I.; Lee, N.; Oprian, D. D. *Biochemistry* **1999**, *38*, 11593−11596.