**RSC**Publishing

# Excitation energies of a water-bridged twisted retinal structure in the bacteriorhodopsin proton pump: a theoretical investigation†

Tino Wolter,[a] Kai Welke,[a] Prasad Phatak,[b] Ana-Nicoleta Bondar*[c] and Marcus Elstner*[a]

The first proton transfer in the bacteriorhodopsin photocycle takes place during the L → M transition. Structural details of the pre proton transfer L intermediate have been investigated using experiments and computations. Here, we assess L-state structural models by performing hybrid quantum mechanical/molecular mechanical molecular dynamics and excitation energy calculations. The computations suggest that a water-bridged twisted retinal structure gives the closest agreement with the experimental L/bR shift in the excitation energy.

## 1 Introduction

The light-driven proton pump bacteriorhodopsin (bR) is a model system for understanding how transmembrane proteins translocate ions across biological membranes. The proton pumping cycle starts with the all-*trans* to 13-*cis* photo-isomerization of the retinal, which is covalently bound to Lys216 *via* a protonated Schiff base. The first proton transfer step of the cycle involves the transfer of the Schiff base proton to the nearby Asp85 during the L → M transition. In spite of its apparent simplicity of a short-distance proton transfer step, the mechanism of the Schiff base deprotonation has been a subject of significant controversy from both experiments and theory.[1–5] Discrepancies between the published crystal structures of L have largely contributed to the controversies over the mechanism. Indeed, as we summarize below, the four crystal structures proposed for L[6–9] have given conflicting information on the conformation of the retinal Schiff base and its interactions with the surrounding protein and the water environment (see also discussion in ref. 2).

In the bR resting state, the Schiff base of the all-*trans* retinal is hydrogen-bonded to a water molecule (w402, according to the

X-ray structure of ref. 10, PDB code: 1c3w) that further connects to the negatively-charged carboxylates Asp85 and Asp212. Both carboxylates are hydrogen-bonded to additional water molecules (Fig. 1). The first crystal structure proposed for L (PDB code: 1e0p)[6] was resolved with a resolution of 2.1 Å from crystals cryotrapped at 170 K. Assignment of this structure to L was debated owing to the possible contamination with K and M,[11–13] and a new structure was published (PDB code: 1vjm)[7] based on crystals cryotrapped at 150 K. In the structures from both ref. 6 and 7, the retinal was modeled planar, with a Schiff base orientation towards the cytoplasmic side and w402 being absent. The absence of w402 was interpreted as an indication of coupled movements of w402 and nearby protein groups, most notably displacement of Arg82 towards the extracellular side.[6] Combined quantum mechanical/molecular mechanical (QM/MM) computations on the L structure from ref. 6 indicated that for this geometry of the active site, the M-like protonation state (Schiff base deprotonated, Asp85 protonated) is energetically favored by approximately 7 kcal mol$^{-1}$.[14]

A cytoplasmic-oriented Schiff base is also indicated by the crystal structure from ref. 8 (PDB code: 1ucq), solved at 160 K with a resolution of 2.4 Å. Based on the analysis of electron densities in a set of K, L, and M state structures,[8,15,16] the intriguing proposal was made[8] that during the K to L transition w402 is translocated to a new location on the cytoplasmic side of the Schiff base; this new location was labeled w602 in ref. 8, and water B in ref. 17. Whether or not w402 can indeed translocate to the cytoplasmic side is unclear, because the energetics of water translocation indicated a rate-limiting barrier of 13–15 kcal mol$^{-1}$, which is too high for being compatible with the L-to-M transition.[17]

[a] *Karlsruhe Institute of Technology, Institute of Physical Chemistry, Kaiserstraße 12, Karlsruhe, Germany. E-mail: marcus.elstner@kit.edu; Fax: +49 721 608 45710; Tel: +49 721 608 45700*

[b] *Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA*

[c] *Theoretical Molecular Biophysics Group, Department of Physics, Freie Universitaet Berlin, Arnimallee 14, D-14195 Berlin, Germany. E-mail: nbondar@zedat.fu-berlin.de; Fax: +49 30 38-56510; Tel: +49 30/838-53583*

† Electronic upplementary information (ESI) available: Retinal dihedral angles for X-ray and QM/MM optimized structures. See DOI: 10.1039/c3cp44280b
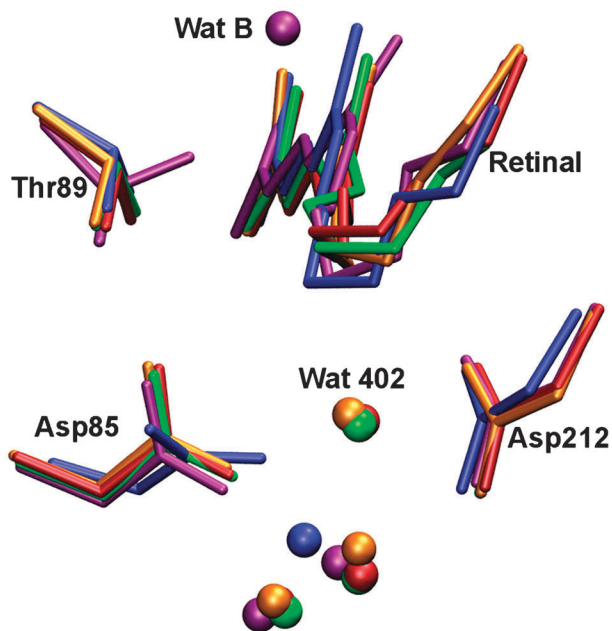
**Fig. 1** Comparison of crystal structures proposed for the L state. The crystal structures from ref. 7 (1vjm), ref. 8 (1ucq), ref. 18 (1o0a), ref. 9 (2ntw) and ref. 10 (1c3w) are shown in blue, purple, orange, green and red, respectively. For simplicity, the retinal chain is shown without its β-ionone ring. Hydrogen atoms are not depicted.

In contrast to the cytoplasmic-oriented Schiff base from ref. 6–8, the L state models from ref. 9 and 18 (PDB code: 1o0a and 2ntw, respectively) indicate a Schiff base orientation towards the extracellular side. In both models with the extracellular-oriented retinal (ref. 9 and 18), w402 is present close to its original location in the bR resting state (Fig. 1), which led to the proposal that w402 acts as an intermediate carrier for the proton.[18] But the retinal geometry from these structures appears difficult to reconcile with other experimental and computational observations.[17,19] As noted by Bondar et al.,[17] in the structure from ref. 18 the C14–C15=N–CE dihedral angle is 43.7° (i.e., the retinal is 15-syn), which is not consistent with the 15-anti conformation deduced in other spectroscopic experiments;[20–22] in the structure from ref. 9 the C12–C13=C14–C15 dihedral angle is −107.9°, i.e., the retinal has a twisted all-trans geometry.[17] (A planar all-trans retinal would be characterized by the C12–C13–C14–C15 dihedral angle being 180 degrees, and planar 13-cis by an angle of 0 degrees.)

Understanding the geometry of the proton transfer complex in the L intermediate is further complicated by observations from time-resolved FTIR[23] studies indicating different vibrational bands at low temperature compared to room temperature. The differences in the vibrational frequencies interpreted to indicate differences in the conformation of the retinal and in the position of water molecules close to the Schiff base. Whereas at room temperature the 15-H HOOP mode at 983 cm$^{-1}$ suggests that retinal is twisted around the C14–C15 bond,[23] the absence of the 15-H HOOP fingerprint at low temperatures supports a planar retinal geometry.[23] Likewise, FTIR measurements were used to assign different vibrational bands to the

active site water molecules at low vs. room temperature.[5] The observations from experiments that the retinal geometry and water interactions depend on the temperature are supported by results from theory: for example, the free energy for inserting a water molecule in the water B position is −17 kcal mol$^{-1}$ at 150 K, and −12 kcal mol$^{-1}$ at 300 K.[17]

Starting from a L-like model with a water molecule in the water B position (Fig. 1), it was found that at room temperature the geometry rapidly converges to a structure in which the water molecule bridges the twisted retinal Schiff base and the negatively-charged Asp85.[17] This model is consistent with recent NMR experiments[19] as well as calculations, which estimated the proton transfer barrier to be 7 kcal mol$^{-1}$, which matches well the 10 μs timescale of the L to M transition.[17]

To understand which of the various geometries proposed for the active site of L reproduces the absorption maxima measured in experiments, we computed here the excitation energies starting from four different L-state models constructed from three different X-ray structures. We performed combined QM/MM molecular dynamics simulations at room temperature, and then computed excitation energies of the minimum energy structures. As first demonstrated in ref. 2, we confirm here via independent molecular dynamics simulations that at room temperature water B relocates to bridge the Schiff base and Asp85. The important new observation from the current computations is that relative to the bR ground state, the water-bridged structure gives an excitation energy shift that is compatible with experiments. The other structural models considered here give absorption maxima which, within the error of our computations, are inconsistent with the experiments.

## 2 Methods

### Protein models

We performed three independent sets of simulations using the L state crystal structures from ref. 7–9 (PDB codes: 1vjm, 1ucq and 2ntw). Due to the reasons discussed in the Introduction, here we do not perform new computations on the crystal structures from ref. 6 and 18. The missing internal coordinates and hydrogen atoms were added using the CHARMM software.[24,25] The amino acid residues Asp96, Asp115[26,27] and Glu204[28] were considered protonated, all other titrable residues were modeled in their standard protonation state.

Each of the protein structures was geometry-optimized as described below. Throughout the manuscript, we denote L models A, B and C the optimized structures derived from the crystal structures of ref. 7–9, respectively. In the case of the crystal structure from Kouyama et al., the Thr89 N–C$_\alpha$–C$_\beta$–C$_\gamma$ dihedral angle was changed from the trans-conformation as shown in the crystal structure to a gauche-conformation, because previous computations demonstrated that the gauche-conformation is energetically favored.[17]

### Force field and the combined QM/MM approach

Geometry optimization and molecular dynamics simulations were performed using a combined QM/MM approach as described in

ref. 29, with SCC-DFTB[30] for the QM region and the CHARMM22[31] force field for the description of the remaining protein.

SCC-DFTB[30] is an approximate Density Functional Theory (DFT) method which is about 2–3 orders of magnitude faster than DFT with medium-sized basis sets. The standard SCC-DFTB is derived from a second order taylor expansion of the DFT total energy around a reference density. Extensive tests demonstrated that the accuracy of the standard SCC-DFTB for describing heats of formation, molecular geometries, etc.[32,33] is similar to modern Hartree–Fock based semi-empirical methods such as PM6, OM2 or PDDG-MNDO.[34–36] Importantly, it was shown that SCC-DFTB describes the ground state structural properties of protonated Schiff bases with a performance comparable to full DFT, e.g. structure, bond length alternation and torsional barriers,[37–39] even the geometry of the bacteriorhodopsin active site.[39]

Recently, SCC-DFTB has been extended to third order. This improvement allows a much more accurate description of proton affinities and hydrogen bonding energies.[40,41] For example, whereas the standard SCC-DFTB underestimates hydrogen bonding energies by 2 kcal mol$^{-1}$, the improved third order SCC-DFTB reduces the error to less than 1 kcal mol$^{-1}$ for a large test set of hydrogen bonding complexes.[41]

### Quantum mechanical region

The MD simulations were performed with a QM region that included the retinal molecule, Lys216, Asp85, Asp212, Thr89 and three water molecules close to the retinal Schiff base region (Fig. 1). This QM region is denoted here QM6. The QM/MM frontiers for the amino acid residues were set at the $C_\alpha$–$C_\beta$ bond, and a link atom scheme was used to saturate the QM host atoms.[42] The electrostatic interactions at the boundary are treated using the "divided frontier charge" scheme proposed by König et al.[43]

To maintain the shape of the protein, all atoms further than 14 Å from the Schiff base nitrogen atom were restrained harmonically to their initial crystal structure coordinates based on their B-factors. Atoms within 6 Å from the Schiff base nitrogen were not subjected to any constraints. In the region between 6 Å and 14 Å, atoms were restrained harmonically with force constants scaled by a sigmoid function to smoothen the transition between the regions without and with harmonic constraints. To account for the effect of bulk solvent, a charge scaling scheme was used as proposed by Dinner et al.[44] The partial charges of charged residues which were exposed to the bulk solvent are scaled using scaling factors determined from solutions of the Poisson–Boltzmann[45] equation.

To investigate the impact of the protein environment and the size of the QM region on the excitation energies, we performed additional sets of computations in which coordinates for the retinal were extracted from the QM/MM trajectory obtained with QM6. In that manner, one can separate the impact of electrostatic from steric influences on the excitation energy of the chromophore. We also assessed a smaller QM region, termed QM1, where the retinal and Lys216 are described quantum mechanically, while the rest of the protein is included at the MM level.

### Molecular dynamics simulations

For each L-state model we performed an energy optimization followed by heating, equilibration, and production run. Heating of the systems to 300 K was performed in 20 steps, allowing for 1 ps of equilibration at each temperature value. Heating, equilibration (300 ps) and production runs (2.5 ns) were performed using the Nose–Hoover[46,47] thermostat with an integration time step of 1.0 fs.

### Computation of excitation energies

To predict changes in excitation energy with respect to external electrostatic field perturbations, we used the *ab initio* method SORCI[48] (spectroscopy oriented configuration interaction) as implemented in the ORCA program package.[49] The various thresholds of the SORCI method were set in accordance with previous studies.[38,50] The SV(P) basis set[51] was used with diffuse *s*- and *p*-functions on carboxylate oxygen atoms of anionic residues. As one-electron basis for the SORCI calculation, averaged approximate natural orbitals were used, resulting from a MR-DDCI3 calculation with a CAS(4,4) reference space after the initial RHF calculation and generation of improved virtual orbitals.

### Accounting for protein polarization in excitation energy computations

Previous studies have shown that protein polarization may influence the excitation energies.[52,53] We have implemented a polarizable force field model which allows us to include the effect of protein electronic polarization in the calculation of excitation energies for static structures.[54,55] This model combines polarization-free point charges with an interactive induced atomic dipole model in response to the charge distribution of the ground and first excited states of the QM region. To improve the electrostatic representation of the protein, the CHARMM22 point charges are substituted by the explicitly polarized "polar.h" model, which was benchmarked for peptides and applied to various retinal proteins.[54,55] The electronic polarization affects not only absolute excitation energies, but also relative excitation energies.[56–58]

Using our methodology, we are able to compute the excitation energy of bR between ground and first excited states (2.18 eV)[58] in agreement with the experimentally determined absorption maximum. Note that this agreement is the result of a fortunate error cancellation. It has been shown previously that the bond length alternation, *i.e.* the difference between average single- and double-bond lengths, is underestimated using SCC-DFTB and DFT, so that excitation energies obtained from these structures are consistently too low by about 0.1 eV compared to hybrid DFT.[56]

## 3 Results and discussion

### Features of QM/MM optimized L state models

The structural features of QM/MM-optimized geometries of several L state models have been discussed before (see, for example, ref. 2, 17 and 59). Here we discuss briefly the aspects

of the geometry of the active site that are potentially important determinants of the excitation energies.

**The QM/MM-optimized L model A.** As noted before,[17] the retinal in ref. 9 is in a twisted all-*trans* conformation. Upon QM/MM geometry optimization, the retinal relaxes to a geometry closer to a planar all-*trans*: the dihedral angle around the C13═C14 double bond increases from $-107.9°$ in the crystal structure,[9] to $-150.7°$ in the QM/MM-optimized model (Fig. 2). The dihedral angles of the polyene chain for the crystal and optimized structure are shown in the ESI.† Although DFT-based methods tend to overestimate the planarity of the retinal molecule,[37] our observations from two independent computations (here and ref. 17) confirm that the QM/MM-refined geometry of ref. 9 is bR-like all-*trans*. After QM/MM optimization (Fig. 3a), the distance between the Schiff base nitrogen and the w402 oxygen atom is 2.8 Å, which is close to the crystal structure (2.9 Å). The optimized structure of the binding pockets of L model A and the resting state of bR[50] are very similar. The distance between Arg82 and Glu204 is slighly shortened from 5.2 Å in the crystal structure to 4.3 Å in the QM/MM-optimized structure and the distance between Glu204 and Glu194 is shortened from 3.8 Å to 3.6 Å. Glu204 is connected to Arg82 *via* a bridging water molecule.

**The QM/MM-optimized L model B.** In the crystal structure,[7] the retinal was modeled as a planar polyene chain. Upon QM/MM geometry optimization the retinal becomes twisted around the N16═C15, C15–C14 and C14═C13 bonds by 20.03°, 24.36° and 20.99°, respectively. As a result of this twist, the retinal Schiff base is oriented toward Thr89 (Fig. 3b). The distance between Thr89 and Asp85 shortened from 4.5 Å in the crystal structure to 2.8 Å in the QM/MM-optimized model.
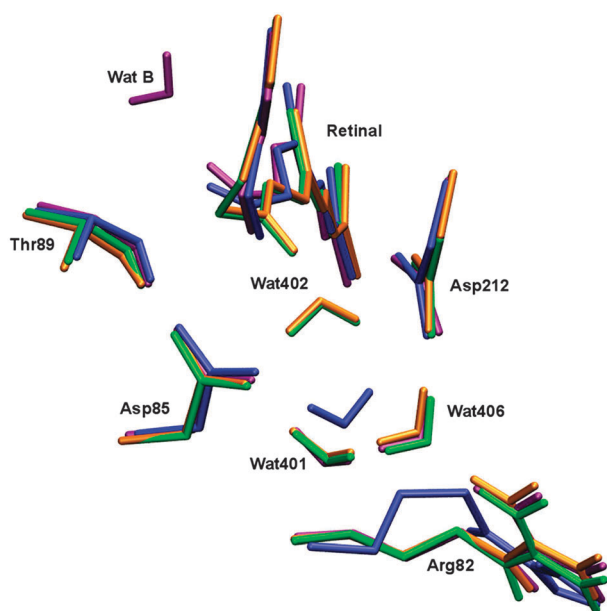
Similar observations on retinal Schiff base twisting and the formation of a hydrogen bond between Asp85 and Thr89 upon QM/MM optimization were made in previous work[14] on the crystal structure from ref. 6. Taken together, these observations suggest that a planar retinal geometry in the polar environment of the binding pocket may be less favorable than a geometry where both the retinal and Asp85 form a hydrogen bond to Thr89. The hydrogen bond between Thr89 and the negatively-charged Asp85 has been revealed by experiments,[60] and observed in computations on L state models.[2,59]

**The QM/MM-optimized L model C.** Geometry optimization of the model prepared from the crystal structure of ref. 8 preserved the overall starting geometry (Fig. 3c). Several hydrogen-bonding distances in the active site became shorter upon optimization: between Asp85 and Thr89 by 0.03 Å, Asp85 and water401 by 0.02 Å, and between water B and the Schiff base by 0.3 Å.

It must be noted, that all X-ray structures for the L-intermediate are taken at low temperatures, and therefore may not resemble the actual structure, that is functional under physiological conditions. We therefore perform MD simulations at room temperature to allow for a better description with conditions closer to the native environment of bR. However, due to limited computation time, the simulations are still somewhat biased to the X-ray structure, which was taken as a starting point, and large conformational changes that are connected to large energy barriers may not be captured.
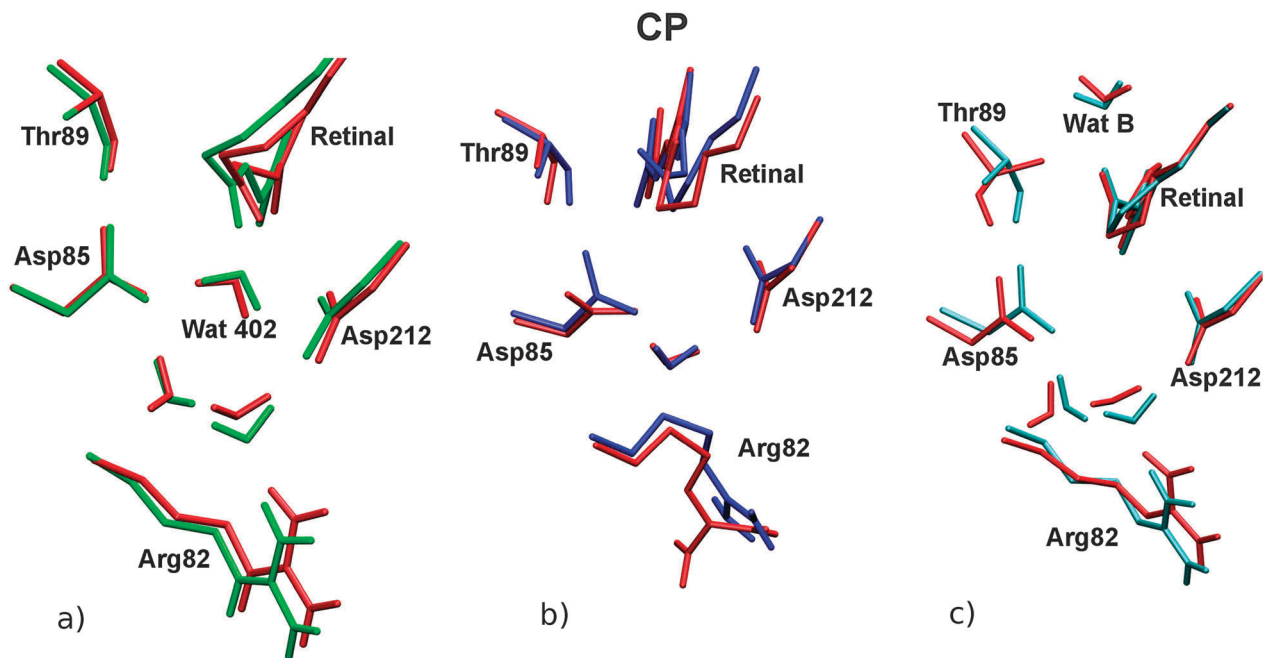
### QM/MM MD simulations at 300 K

In the following, we briefly describe the main observations from the QM/MM MD simulations at room temperature. During the 2.5 ns simulation starting from L model A, the interactions in the retinal binding pocket remain stable and the pentagonal water cluster as well as the hydrogen bonded network close to the chromophore remain intact. During heating, a hydrogen bond forms between Glu194 and Glu204, and the water-bridged hydrogen bond between Arg82 and the proton release group changes from Glu204 to Glu194. The MD trajectory of L model B indicates that the water molecule from the active site (Fig. 3b) is stable between Asp85 and Asp212. The retinal is twisted towards Thr89, and the hydrogen bond between Arg82 and the proton release group is stable during dynamics.

**MD-simulation of L model C.** A recent QM/MM study[17] using second-order SCC-DFTB with a specific parametrization of the retinal Schiff base demonstrated that a water molecule placed in the location of water B can rapidly relocate during dynamics at 300 K to the region between the twisted retinal Schiff base and the negatively-charged carboxylate of Asp85. This geometry with a twisted retinal Schiff base bridged to Asp85 *via* water B was found compatible with the proton transfer kinetics of the reaction cycle. As described below, the observation that water B rapidly relocates to bridge the Schiff base to Asp85[17] is reproduced with the setup used here.

Early during our equilibration of L model C, water B reorients and forms a hydrogen bond with the Schiff base and Thr89 (Fig. 4b). This stabilizes transiently the retinal Schiff base in a geometry twisted towards Thr89. Within approximately 1.8 ns, the



**Fig. 2** Comparison of the QM/MM-geometry optimized L models A (green), B (blue), C (purple) and the bR resting state (orange). Note that upon geometry optimization of L model B, the retinal Schiff base segment becomes slightly twisted towards Thr89. L model A becomes almost identical with the bR resting state.

**Fig. 3** Comparison of the starting crystal structures and the QM/MM-optimized L-state geometries. (a) L model A (green), (b) L model B (blue) and (c) L model C (cyan). The crystal structures are colored red. CP: cytoplasmic side.
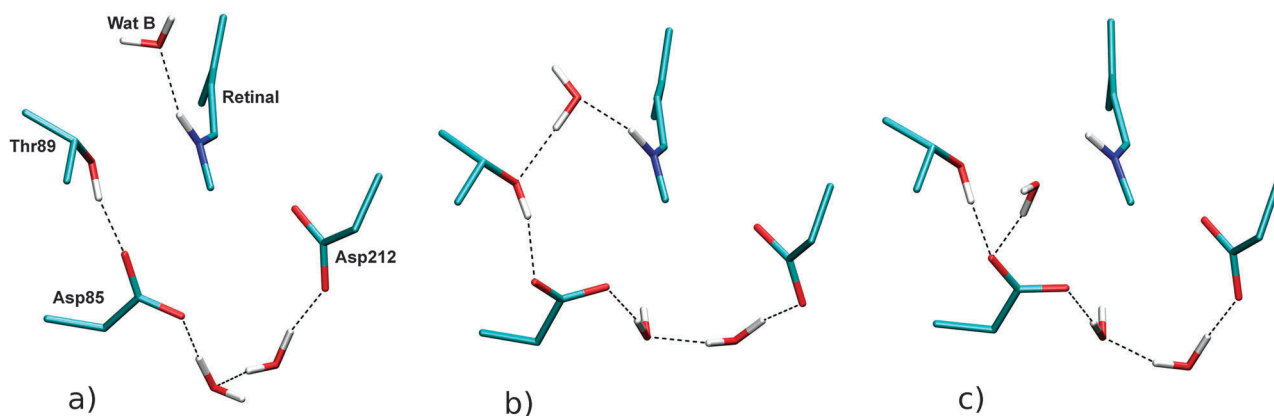


**Fig. 4** Coordinate snapshots illustrating Schiff base:water interactions sampled for L model C. (a) L model C after QM/MM optimization showing Thr89 in the *gauche* conformation, and water B as a hydrogen-bonding partner for the Schiff base. (b) Snapshot from the QM/MM MD at 300 K. The retinal is twisted, and water B bridges the Schiff base and Thr89. (c) After ~1.8 ns, the dynamics converges to a protein conformation where water B bridges the twisted retinal Schiff base and Asp85. We denote this bridged structure as L model D. The coordinate snapshots depicted in panels a–c are similar to those sampled in previous QM/MM MD and reaction path computations.[17]

retinal and water B undergo reorientations leading to a geometry very similar to that observed in ref. 17 (L model D, Fig. 4). The longer simulation time required for the formation of the water-bridged structure in our current simulation as compared to ref. 17 may be attributed to the stochastic nature of the transition to the water-bridged structure or to differences in the setup (different starting coordinates for the protein, different order of the SCC-DFTB expansion, different protocol for the harmonic constraints).

### Excitation energies

**Impact of retinal geometry on the vertical excitation.** The bond length alternation (BLA) of the retinal polyene chain and the deviation from planarity of the π-conjugated system are major determinants of the excitation energy. For example, increasing the BLA leads to a blue-shift of the excitation energy. A strong twist around the C13=C14 bond leads to a red-shift, because the rotation destabilizes the electronic ground state ($S_0$) and stabilizes the first excited state ($S_1$).

In a first step, we evaluate the excitation energies for the isolated chromophore under vacuum by removing the protein, but keeping retinal in its geometry from the QM/MM optimized protein. The data summarized in Table 1 show that the gas-phase excitation energies vary by 0.07 eV, indicating that the difference in excitation energies results from the interaction with the protein environment. The major structural determinant of the variation is the BLA of the retinal polyene chain.

**Table 1** Connection between structural parameters and the excitation energy of the chromophore in the gas phase (in eV and nm)

|                          | bR          | L model A   | L model B   | L model C   | L model D   |
|--------------------------|-------------|-------------|-------------|-------------|-------------|
| $\Delta E$               | 1.86 (667)  | 1.86 (667)  | 1.88 (659)  | 1.92 (646)  | 1.93 (642)  |
| Planarity[a]             | 10.01       | 9.43        | 11.07       | 9.65        | 7.05        |
| Max. rotation[b]         | 18.6        | 19.3        | 20.9        | 22.6        | 21.4        |
| BLA[c] (in pm)           | 5.39        | 5.41        | 6.39        | 6.71        | 6.08        |

[a] Average deviation of dihedral angles from planarity. [b] Max. rotation is around C13=C14 in all cases. [c] Bond length alternation = difference between averaged C–C single bonds and C=C double bonds.

The deviation from planarity is similar for all models assessed here (Table 1). The similar values of the excitation energies computed for L model A and the bR ground state further supports our proposal that the structural model from ref. 9 is a twisted all-*trans* structure incompatible with L. The excitation energy of the retinal from L model B is slightly higher than that of the bR ground state. The increased BLA has a larger impact on the excitation energy than the slightly stronger twist of the chromophore, which leads to a 0.02 eV blue-shift compared to the bR resting state.

The blue-shift of 0.06 eV of the L model C with respect to the bR resting state can be explained by the larger BLA in L model C as compared to L model A and L model B. The marginal difference between L models C and D can be explained by the compensation of two opposing effects: a more twisted retinal L model C (around C13=C14), and a smaller BLA of the retinal in L model D.

**Impact of the protein environment and size of the QM-region.** The shift of the excitation energies induced by the protein varies from 0.47 eV in the case of model A to 0.63 eV with model C (Table 2). This shift can be explained by the interactions with (a) the charges of the counterions which stabilize the $S_0$ state relative to $S_1$, since the positive charge of the retinal is shifted from the Schiff base to the β-ionone ring during the $S_0$–$S_1$ excitation; (b) the remainder of the retinal binding pocket groups. For all our models we obtain vertical excitation energies that are blue-shifted compared to experiment. This artefact arises from neglect of dispersion interaction and polarization of the protein environment, if describing the electrostatic protein environment by point charges. The differences between QM1 and QM6 arise from two effects: (a) the polarization of the counterions and (b) a small charge transfer from the Schiff base towards the counterions and the water.[58]

For bR and L model A, where the Schiff base is oriented towards the extracellular side, the protein blue-shifts (0.08 eV and 0.05 eV, QM6–QM1 in Table 4) are comparable. L models B, C and D, in which the Schiff base is oriented towards the cytoplasm, the protein blue-shifts range from 0.11 eV for

L models B and D, to 0.23 eV for model C. These relatively large blue-shifts observed for these models are rather suprising, because the polarization of the counterions is small.

Similar to previous studies,[50] we observe a correlation between the BLA and the response of the excitation energy to the electric field induced by the point charges of the protein (Table 2).

The Arg82 adopts a different conformation in the L model B compared to the other models. To estimate the impact of the different orientations on the excitation energy, we mutated Arg82 into Gly, which effectively removes the charges of the arginine side chain. In general, the replacement has only a small effect on the excitation energies (Table 3), which was also found for the bR resting state previously.[50,58]

Unlike Arg82, w402 and water B have larger effects on the excitation energies computed for all models except L model C. The change in the excitation energy upon removing water402 ranges from −0.08 eV in the case of bR and L model A, to −0.03 eV in the case of L model C. Note that neither w402 nor water B is present in the L model B. Water B becomes an important determinant of the excitation energy when it moves to the bridging position between the retinal Schiff base and Asp85 (L model D). Removing the point charges of water B in L model D leads to a red-shift of 0.06 eV.

When the computations are performed with the larger QM6 region, removing w402 from L model A leads to a significant red-shift of 0.19 eV. Water402 in bR and L model A, where it is located between the Schiff base and the Asp85/Asp212 counterions, has a slighly larger influence than water B in L models C and D, where it is positioned on the cytoplasmic side of the Schiff base. The shifts reflect the stronger polarization of the water molecule by the protein environment in the former case.

**Table 2** The retinal bond length alternation, the protein opsin shift, and the difference dipole moments measured from the QM/MM optimized structures

| L model | BLA  | Protein shift (eV) |
|---------|------|--------------------|
| A       | 5.41 | 0.47               |
| B       | 6.39 | 0.5                |
| C       | 6.71 | 0.63               |
| D       | 6.08 | 0.49               |

**Table 3** Perturbation analysis: influence of Arg82 and water molecules water402 and water B on the excitation energy of the retinal chromophore

|           | Schiff base orient. | Orient. of Arg82[a] | R82G-shift[b] | wat402-shift[c] | wat402-shift[d] |
|-----------|---------------------|---------------------|---------------|-----------------|-----------------|
| bR        | Extracellular       | 9.98                | −0.03         | −0.08           | −0.16           |
| L model A | Extracellular       | 10.27               | −0.01         | −0.08           | −0.19           |
| L model B | Cytoplasmic         | 11.22               | −0.03         | —               | —               |
| L model C | Cytoplasmic         | 10.82               | ±0            | −0.03           | −0.12           |
| L model D | Cytoplasmic         | 10.47               | +0.02         | −0.06           | −0.13           |

[a] Distance between chromophore ($N_{16}$) and Arg82 ($C_Z$). [b] Effect of mutation from arginine to glycine on excitation energy (QM6). [c] Effect of deleting water402/water B on excitation energy (QM1). [d] Effect of deleting water402/water B on excitation energy (QM6).

**Table 4** Effect of different QM fragments and polar.h on excitation energy (in eV and nm)

|              | bR          | L-model A   | L-model B   | L-model C   | L-model D   |
|--------------|-------------|-------------|-------------|-------------|-------------|
| QM1          | 2.32 (534)  | 2.33 (532)  | 2.38 (521)  | 2.55 (486)  | 2.42 (512)  |
| QM6          | 2.40 (517)  | 2.38 (521)  | 2.49 (498)  | 2.75 (451)  | 2.52 (492)  |
| QM6+ polar.h | 2.18 (568)  | 2.15 (577)  | 2.15 (577)  | 2.41 (514)  | 2.25 (551)  |
| Exp[a]       | 2.18 (568)  | 2.28 (544)  |             |             |             |

[a] Absorption maxima at room temperature.[63]

**Role of the electronic polarization.** When the point charges of the protein are replaced using a polarizable model, we observe a red shift of the excitation energy. In the case of the bR resting state, the computations using a polarizable model result in excellent agreement with the experimental value. However, this seemingly perfect agreement can be attributed to an error cancellation:[58] SCC-DFTB underestimates the bond length alternation, leading to excitation energies in the gas-phase, that are about 0.1 eV too low compared to experiment. Using PBE0, a hybrid density functional, the geometry optimization leads to a structure of the chromophore with a BLA in good agreement with CASPT2 benchmark calculations.[56,61,62] The neglect of dispersion on the other hand leads to excitation energies, which are too high. Thus, both effects, the neglect of dispersion and the BLA being underestimated with SCC-DFTB compensate each other. We expect these approximations to be systematic and similar for each structure.

Including the effect of mutual instantaneous polarization of the protein environment and the QM region results in red-shifts of the excitation energies (Table 4). In the case of an extracellular-oriented Schiff base, the polarization-induced red shift is 0.22 eV for bR and 0.23 eV for L model A, respectively. The red-shifts obtained using L models B, C and D, where the retinal is cytoplasmic-oriented, are larger (−0.23 to −0.34 eV). Therefore, the large protein-induced blue-shifts are counterbalanced by polarization-induced red-shifts.

# 4 Conclusions

We evaluated the compatibility of different models for the L intermediate of bacteriorhodopsin's photocycle with the experimentally obtained absorption maxima. In the case of the bR resting state, we demonstrated previously that our approach to the calculation of excitation energies with SCC-DFTB/MM gives values in agreement with experiments (2.18 eV). This excellent agreement is definitely accidental and relies on an error cancellation of the method, since none of the existing excited state methods applicable to large molecules like retinal show an intrinsic accuracy better than 0.1 eV.

Error cancellation is not an unusual feature of methods in computational chemistry and is very often exploited in applications. The main requirement is, that this error cancellation is systematic, *i.e.* that it applies for different molecules to the same degree. For application to color tuning in retinal proteins that means that changes in the chromophore and the protein

environment are accurately reproduced. We have checked our methodology in this respect in a series of prior publications,[38,54–58] as discussed above. Therefore, we are sufficiently confident that shifts in excitation energies are reproduced reliably. This approach has been successfully used in identifying the shifts due to mutations in Rhodopsin (Rh)[56] and in identifying a reasonabe O state model of bR.[57] We have shown that error cancellation appears in a series of proteins and their intermediates to a similar degree. Therefore, we can apply this methodology to the L-intermediate of the bR photocycle and can expect a reliable estimate of its excitation energy.

We started from three crystal structures and generated L-state models A, B and C using QM/MM geometry optimization. Model A (obtained from ref. 9) is very similar to the structure of the bR resting state; model B (derived from ref. 7) differs from the bR resting state only slightly in structural details of the binding pocket. The excitation energies computed for models A and B are thus very close to the bR resting state, suggesting that models A and B are unlikely to represent L.

Model C (derived from X-ray structure 1ucq[8]) gives an excitation energy too high compared to the experiment. An important observation here is that, as observed previously,[17] during QM/MM MD simulations at 300 K. Model C relaxes into model D. During this relaxation, water B relocates from hydrogen bonding to Thr89 and the Schiff base to the cavity between Asp85 and the Schiff base, such that the Schiff base bridges to Asp85 *via* water B. Relative to the bR resting state, model D gives an absorption maximum shift in close agreement with experiments which show a range of 541 to 550 nm (2.29 to 2.25 eV) depending on temperature and spectral reconstruction.[64,65] This may be due to the increased structural flexibility of the L state. The observation that the bridged-water structure gives an absorption maximum consistent with L is compatible with NMR data[1] and previous molecular dynamics simulations and proton transfer calculations.[17]

In a pervious work, we have in detail investigated the influence of retinal geometry on excitation energies.[38] We can investigate the effect of retinal conformation on the excitation energy by omitting the atomic charges from the opsin calculation. Models A and B, which resemble the bR resting state, are slightly red-shifted by 0.05 eV with respect to models C and D. Therefore, the retinal twist in models C and D is responsible for roughly half of the shift between bR and L. The other 0.05 eV then comes from the interaction of the retinal with the opsin electrostatic environment.

Excitation energies are quite sensitive to the active site structure, conformational changes in the presence/absence of single water molecules in the active site can have some measurable influence, as discussed in detail of our previous work.[38,54–58] Therefore, the calculation of excitation energies can give some further information to evaluate structural models. It is definitely not that specific to local interactions, as *e.g.* IR spectra are, since they condense all structural details of the active site into one single number, the excitation energy. For the evaluation of structural models, one has to keep in mind that the excitation energy of a wrong structural model can by accident lead to the

right excitation energy. Therefore, a positive proof is not possible, since "false positive" results cannot be excluded based on the excitation energy alone. However, a wrong excitation energy severely questions a structural model, provided a well calibrated excited state method is used. In this respect, excited state calculations can help in identifying structural models, which we have used in this work to elucidate the bR L-state model.

## Acknowledgements

## References

1 M. L. Mak-Jurkauskas, V. S. Bajaj, M. K. Hornstein, M. Belenky, R. G. Griffin and J. Herzfeld, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 883–888.

2 A. Bondar, M. Elstner, S. Suhai, J. Smith and S. Fischer, *Structure*, 2004, **12**, 1281–1288.

3 J. E. Morgan, A. S. Vakkasoglu, R. B. Gennis and A. Maeda, *Biochemistry*, 2007, **46**, 2787–2796.

4 A. Morgan and J. E. Maeda, *Photochem. Photobiol.*, 2008, **84**, 1038–1045.

5 V. A. Lorenz-Fonfria, Y. Furutani and H. Kandori, *Biochemistry*, 2008, **47**, 4071–4081.

6 A. Royant, K. Edman, T. Ursby, E. Pebay-Peyroula, E. M. Landau and R. Neutze, *Nature*, 2000, **406**, 645–648.

7 K. Edman, A. Royant, G. Larsson, F. Jacobson, T. Taylor, D. van der Spoel, E. M. Landau, E. Pebay-Peyroula and R. Neutze, *J. Biol. Chem.*, 2004, **279**, 2147–2158.

8 T. Kouyama, T. Nishikawa, T. Tokuhisa and H. Okumura, *J. Mol. Biol.*, 2004, **335**, 531–546.

9 J. K. Lanyi, *J. Mol. Biol.*, 2007, **365**, 1379–1392.

10 H. Luecke, B. Schobert, H. Richter, J. Cartailler and J. K. Lanyi, *J. Mol. Biol.*, 1999, **291**, 899–911.

11 S. P. Balashov and T. G. Ebrey, *Photochem. Photobiol.*, 2001, **73**, 453.

12 J. K. Lanyi and H. Luecke, *Curr. Opin. Struct. Biol.*, 2001, **11**, 415–419.

13 R. Neutze, E. Pebay-Peyroula, K. Edman, A. Royant, J. Navarro and E. M. Landau, *Biochim. Biophys. Acta –Biomembr.*, 2002, **1565**, 144–167.

14 A. N. Bondar, M. Elstner, S. Suhai, S. Fischer and J. C. Smith, *Phase Transitions*, 2005, **78**, 5–9.

15 Y. Matsui, K. Sakai, M. Murakami, Y. Shiro, S. Adachi, H. Okumura and T. Kouyama, *J. Mol. Biol.*, 2002, **324**, 469–481.

16 K. Takeda, Y. Matsui, N. Kamiya and S. Adachi, *J. Mol. Biol.*, 2004, 1023–1037.

17 A.-N. Bondar, J. Baudry, S. Suhai, S. Fischer and J. C. Smith, *J. Phys. Chem. B*, 2008, **112**, 14729–14741.

18 J. K. Lanyi, *J. Mol. Biol.*, 2003, **328**, 439–445.

19 V. S. Bajaj, M. L. Mak-Jurkauskas, M. Belenky, J. Herzfeld and R. G. Griffin, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 9244–9249.

20 S. O. Smith, A. B. Myers, J. A. Pardoen, C. Winkel, P. P. J. Mulder, J. Lugtenburg and R. Mathies, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**, 2055–2059.

21 S. P. A. Fodor, J. B. Ames, R. Gebhard, E. M. M. Van Den Berg, W. Stoeckenius, J. Lugtenburg and R. A. Mathies, *Biochemistry*, 1988, **27**, 7097–7101.

22 J. B. Ames, S. P. A. Fodor, R. Gebhard, J. Raap, E. M. M. Van Den Berg, J. Lugtenburg and R. A. Mathies, *Biochemistry*, 1989, **28**, 3681–3687.

23 C. Rodig, I. Chizhov, O. Weidlich and F. Siebert, *Biophys. J.*, 1999, **76**, 2687–2701.

24 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.

25 A. T. Brünger and M. Karplus, *Proteins: Struct., Funct., Bioinf.*, 1988, **4**, 148–156.

26 L. S. Brown, G. Varo, M. Hatanaka, J. Sasaki, H. Kandori, A. Maeda, N. Friedman, M. Sheves, R. Needleman and J. K. Lanyi, *Biochemistry*, 1995, **34**, 12903–12911.

27 G. Metz, F. Siebert and M. Engelhard, *FEBS Lett.*, 1992, **303**, 237–241.

28 P. Phatak, N. Ghosh, H. Yu, Q. Cui and M. Elstner, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 19672–19677.

29 Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim and M. Karplus, *J. Phys. Chem. B*, 2001, **105**, 569–585.

30 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**, 7260–7268.

31 A. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

32 K. W. Sattelmeyer, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. A*, 2006, **110**, 13551–13559.

33 N. Otte, M. Scholten and W. Thiel, *J. Phys. Chem. A*, 2007, **111**, 5751–5755.

34 M. Repasky, J. Chandrasekhar and W. Jorgensen, *J. Comput. Chem.*, 2002, **23**, 1601–1622.

35 J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173–1213.

36 W. Weber and W. Thiel, *Theor. Chem. Acc.*, 2000, **103**, 495–506.

37 H. Zhou, E. Tajkhorshid, T. Frauenheim, S. Suhai and M. Elstner, *Chem. Phys.*, 2002, **277**, 91–103.

38 M. Wanko, M. Hoffmann, P. Strodel, A. Koslowski, W. Thiel, F. Neese, T. Frauenheim and M. Elstner, *J. Phys. Chem. B*, 2005, **109**, 3606–3615.

39 A.-N. Bondar, S. Suhai, S. Fischer, J. C. Smith and M. Elstner, *J. Struct. Biol.*, 2007, **157**, 454–469.

40 M. Elstner, *Theor. Chem. Acc.*, 2006, **116**, 316–325.

41 Y. Yang, H. Yu, D. York, Q. Cui and M. Elstner, *J. Phys. Chem. A*, 2007, **111**, 10861–10873.

42 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, 718–730.

43 P. H. König, M. Hoffmann, T. Frauenheim and Q. Cui, *J. Phys. Chem. B*, 2005, **109**, 9082–9095.

44 A. R. Dinner, X. Lopez and M. Karplus, *Theor. Chem. Acc.*, 2003, **109**, 118–124.

45 T. Simonson, G. Archontis and M. Karplus, *J. Phys. Chem. B*, 1997, **101**, 8349–8362.

46 S. Nose, *Mol. Phys.*, 1984, **52**, 255–268.

47 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

48 F. Neese, *J. Chem. Phys.*, 2003, **119**, 9428–9443.

49 F. Neese, *ORCA an ab initio*, density functional and semi-empirical program package, Version 2.3 – Revision09, February 2004, Max Planck Institut fur Strahlenchemie, Mulheim, technical report, 2004.

50 M. Hoffmann, M. Wanko, P. Strodel, P. König, T. Frauenheim, K. Schulten, W. Thiel, E. Tajkhorshid and M. Elstner, *J. Am. Chem. Soc.*, 2006, **128**, 10808–10818.

51 A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571–2577.

52 R. R. Birge, T. M. Cooper, A. F. Lawrence, M. B. Masthay, C. Vasilakis, C. Zhang and R. Zidovetzki, *J. Am. Chem. Soc.*, 1989, **111**, 4063–4074.

53 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.

54 M. Wanko, M. Hoffmann, T. Frauenheim and M. Elstner, *J. Phys. Chem. B*, 2008, **112**, 11462–11467.

55 M. Wanko, M. Hoffmann, J. S. Frähmcke, T. Frauenheim and M. Elstner, *J. Phys. Chem. B*, 2008, **112**, 11468–11478.

56 J. S. Främcke, M. Wanko, P. Phatak, M. A. Mroginski and M. Elstner, *J. Phys. Chem. B*, 2010, **114**, 11338–11352.

57 P. Phatak, J. S. Frähmcke, M. Wanko, M. Hoffmann, P. Strodel, J. C. Smith, S. Suhái, A. N. Bondar and M. Elstner, *J. Am. Chem. Soc.*, 2009, **131**, 7064–7078.

58 K. Welke, J. S. Fraehmcke, H. C. Watanabe, P. Hegemann and M. Elstner, *J. Phys. Chem. B*, 2011, **115**, 15119–15128.

59 A.-N. Bondar, S. Fischer, J. C. Smith, M. Elstner and S. Suhai, *J. Am. Chem. Soc.*, 2004, **126**, 14668–14677.

60 H. Kandori, Y. Yamazaki, Y. Shichida, J. Raap, J. Lugtenburg, M. Belenky and J. Herzfeld, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 1571–1576.

61 T. W. Keal, M. Wanko and W. Thiel, *Theor. Chim. Acta*, 2009, **123**, 145–156.

62 C. S. C. Page and M. M. Olivucci, *J. Comput. Chem.*, 2003, **24**, 298–309.

63 R. Lozier, R. Bogomolni and W. Stoeckenius, *Biophys. J.*, 1975, **15**, 955–962.

64 S. P. Balashov and T. G. Ebrey, *Photochem. Photobiol.*, 2001, **73**, 453–462.

65 R. W. Hendler, *J. Phys. Chem. B*, 2005, **109**, 16515–16528.