Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics

Electronic Supplementary Information

Quantum Chemical Modeling of Rhodopsin Mutants Displaying Switchable Colors

Federico Melaccio, Nicolas Ferré and Massimo Olivucci

Supplementary Methods

QM/MM setup

The QM/MM setup used in this work was already described elsewhere[1, 2] and it was used for both Model I and Model II. Briefly, the protein system was split into two regions: the retinal chromophore bound to lysine and the rest of the protein. The first region is evaluated at the QM level, while a MM force field is employed for the remaining part. The QM wavefunction is polarized by MM point charges and the QM/MM interaction energy is added to the QM energy to obtain the total energy[1]. The QM region is described using a protocol based on CASSCF [3, 4]. This is an *ab initio* multiconfigurational method offering maximum flexibility for an unbiased description (i.e. with no empirically derived parameters and avoiding single-determinant wavefunctions) of the electronic and equilibrium structures of the ground and excited states of a molecule. Furthermore, the CASSCF wavefunction can be used for subsequent multiconfigurational second-order perturbation theory[5] computations (CASPT2) ultimately allowing for an evaluation of energy gap between different electronic states [4]. The QM/MM boundary was placed across the K210 Cδ – Cε bond to ensure it was far enough from the photochemically important part. The hydrogen link atom scheme was employed. [6] Its position was restrained along the frontier C-C bond, at 1.0 Å from the first QM atom (e.g. the Cε). The full QM region was relaxed during the QM/MM geometry optimization. In the MM region, the lysine sidechain, the 2 H atoms belonging to its backbone and the water placed between the PSB and the counterion D75 (HWat) were allowed to move, whereas D75 was kept frozen in all models. Starting from mutated Model I structures (see *Model I mutants' models preparation* below), a QM/MM CASSCF(12,12)/6-31G*/MM single point was performed to check if the correct active space was preserved. The active orbitals were visualized with Molden [7] to check them. In case of wrong active orbitals, the active space was adjusted. Once that the active space was verified to be correct, a QM/MM geometry optimization at the same level of theory as before was run. After full convergence, a CASPT2/6-31G* single point was executed on the lowest 3 roots (no averaging), starting from a 3-states average CASSCF(12,12)/6-31G* single point on the final geometry and evaluating the oscillator strengths for the 1-2 and 1-3 transitions (using the CASSCF 3-states average wavefunction).

All the initial single points and the geometry optimizations were performed by using Gaussian 03[8] for the QM region and Tinker 4.2 [9] for the MM subsystem, interfaced with customized routines.[1] For the final CASPT2 single point, Molcas 6 [10] was used. Gaussian 03 orbitals were used as an initial guess for the 3-states average CASSCF single point, upon which the 2nd order perturbation is applied to get the CASPT2 vertical excitation energy gaps. The total QM/MM energy was evaluated according to the following formula:

$$E_{total} = E_{QM} + E_{QM/MM} + E_{MM}$$

$E_{QM/MM}$ includes the mixed terms. The wavefunction was polarized by the MM force field point charges with a corresponding Coulomb potential operator added into the Hamiltonian (electrostatic embedding). [11] To avoid artifacts at the boundary, a zero charge was placed on the Cδ MM frontier atom, and its former force field partial charge was redistributed on the lysine.[12] QM/MM Van der Waals interactions were treated at the MM level. All the torsional potentials involving both QM and MM atoms were modeled according following the previously reported approach. [12] The chosen force field was AMBER, [13] as it is implemented in Tinker, [9] and it was modified as stated above. Additionally, a corrected TIP3P water model [14] was employed: Van der Waals parameters were added to water hydrogen atoms to guarantee a correct treatment of hydrogen bonding directionality, which was not necessarily required in the condensed liquid water models the TIP3P model was developed for.

Model I mutants' models preparation

The starting ASR_{AT} and ASR_{13C} wild type, QM/MM optimized models were taken from ref.[2]. For Model I, residues were mutated according to the following procedure:

1. The xyz file in Tinker[9] was opened with Molden; [7]
2. The protein Z-Matrix (Molden control window, ZMAT Editor button) was created in order to build up the primary sequence;
3. The selected residue was changed using the Replace function, keeping the default rotamer. The conformation was checked by visual inspection, and the chosen rotamer was eventually changed in case of evident problems;
4. The new residue coordinates for the sidechain were saved and manually pasted into the wild type structure. The backbone was always left unchanged, except for proline mutants;
5. The numbering was reordered by using Tinker xyzedit routine;
6. A Tinker key file was created to specify the active atoms, which were the ones from the new added sidechain, and the MM parameter file (the same as for QM/MM calculations);
7. The Tinker routine minimize was run with the default gradient convergence threshold (0.01 kcal mol^{-1} $Å^{-1}$), in order to optimize the geometry for the new sidechain;

A different rotamer was chosen for V112N (the Molden best rotamer), because the default rotamer caused the N sidechain to clash with the nearby residue Q109.

This protocol underwent small variations when dealing with proline mutants and neutral aspartic and glutamic acids. Since proline features a 5-membered ring, its backbone was included in the active atoms to be relaxed with minimize after the mutation. By doing this, larger scale conformational changes were not taken into account. Whenever a neutral acid residue was required (P206D, P206E, S86D and S214D mutants), a usual charged aspartate or glutamate was introduced with Molden. Then an extra hydrogen atom was manually added, by typing appropriate coordinates to place it halfway between the two carboxyl oxygens, and it was connected to the second oxygen atom in the Tinker file atom ordering. After that, the usual MM minimization of the sidechain was employed for local relaxation.

Model II models preparation

To prepare Model II models, which feature 11 extra water molecules in addition to HWat (the water molecule bridging PSB and the D75 counterion), the corresponding oxygen atoms coordinates were extracted from ASR crystallographic structure (1XIO.pdb) [15] into an empty PDB file. Molden was used to open it, and place one H at a time by using ZMAT Editor and the Add line feature. The resulting 11 water molecules coordinates were saved as Tinker xyz file, their numbering was shifted to match the protein file and then they were placed at the bottom of the Model I wild type models. At this stage, the MM minimization (Tinker minimize, default threshold) was performed on the added H atoms, while O atoms were kept frozen at their crystallographic position. After that, the wild type models were optimized via the usual QM/MM protocol (see below). Regarding the mutation protocol, once that Model II wild type structures were obtained, the same protocol was applied to obtain the mutated structure, except for step 6. In fact, to account for the perturbation induced by the mutant on the water hydrogen bond network in an approximate way, the water H atoms were included among the active atoms and relaxed with minimize, together with the new sidechain.

Electrostatic potential analysis

APBS (Adaptive Poisson-Boltzmann Solver) [16] software was used to compute the electrostatic potential 3D grid for all the models, by execution through the VMD (Visual Molecular Dynamics) [17] plugin, on the required PQR files which were set up by PDB2PQR.[18] Default values were used, except for the solute and solvent dielectric constant (chosen to be equal to 1.0 and 2.0 respectively): since the QM/MM computations were performed *in vacuo*, the protein interior is usually assigned a low dielectric constant. Moreover rhodopsins are membrane proteins, so their environment has low polarity. Force field point charges were used for the protein, whereas the QM region got

zero partial charges, in order to see the effect of the protein environment onto the chromophore. The mapping of electrostatic potential onto the chromophore surface was plotted with VMD (Surf representation, 1.4 Å probe radius, colored by Volume, after loading the APBS-generated electrostatic potential 3D grid). Whenever the external influence on some protein part was evaluated, point charges were put at zero on such a part; otherwise the potential due to the local point charges would be the large contribution to what appears in the plot.
VMD was the software of choice to generate all the other molecular images.

Previous mutant studies

Hoffmann *et al.*[19] looked at the λ^{a}_{max} changes occurring upon replacement of the cavity residues of pharaonis phoborhodopsin with glycine or with the corresponding bR residues. The spectral tuning in bR was addressed by Houjou *et al.*[20], where tryptophanes and tyrosine in the cavity were mutated with alanines to assess their effect on λ^{a}_{max}. In a 2008 study,[21] the absorption maxima of human cone pigments and their mutants were calculated and compared to the available experimental data. Similarly, Altun *et al.*[22, 23] and Rajamani *et al.*[24] compared computed λ^{a}_{max} values for different Rh mutants with the experimental data. Rh mutants have also been modeled to investigate the ionization status of E181[25, 26]. Two mutants of bR and sRII were investigated by Fujimoto *et al.*[27] to further analyze the spectral tuning in rhodopsins. While all the above studies have been carried out with different levels of QM descriptions (e.g. semi-empirical and TD-DFT methods), currently CASPT2//CASSCF is the only practical method allowing a description of spectra, reaction paths, conical intersections and trajectories on a consistent basis[28-32].

References

1. N. Ferré and M. Olivucci, *J Am Chem Soc*, 2003, **125**, 6868-6869.
2. A. Strambi, B. Durbeej, N. Ferré and M. Olivucci, *Proc Natl Acad Sci U. S. A.*, 2010.
3. B. O. Roos, in *Ab initio Methods in Quantum Chemistry*, ed. K. P. Lawley, Wiley, New York, 1987, pp. 399-446.
4. B. O. Roos, M. P. Fülscher, P.-A. Malmqvist, M. Merchan and L. Serrano-Andrés, in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*, ed. S. R. Langhoff, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995, pp. 357-438.
5. K. Andersson, P.-A. Malmqvist, B. O. Roos, A. J. Sadlej and K. J. Wolinski, *J. Phys. Chem.*, 1990, **94**, 5483-5488.
6. U. C. Singh and P. A. Kollman, *J Comp Chem*, 1986, **7**, 718-730.
7. G. Schaftenaar and J. H. Noordik, *J Comput-Aided Mol Des*, 2000, **14**, 123-134.
8. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, A. Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, 2003.
9. J. W. Ponder and F. M. Richards, *J Comp Chem*, 1987, **8**, 1016-1024.
10. G. Karlström, R. Lindh, P.-Å. Malmqvist, B. O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrady and L. Seijo, *Comp Mat Sci*, 2003, **28**, 222-239.
11. H. Senn and W. Thiel, in *Atomistic Approaches in Modern Biology*, ed. M. Reiher, Springer Berlin / Heidelberg, 2007, vol. 268, pp. 173-290.
12. N. Ferré, A. Cembran, M. Garavelli and M. Olivucci, *Theor Chem Acc*, 2004, **112**, 335-341.
13. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J Am Chem Soc*, 1995, **117**, 5179-5197.
14. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J Chem Phys*, 1983, **79**, 926-935.
15. L. Vogeley, O. A. Sineshchekov, V. D. Trivedi, J. Sasaki, J. L. Spudich and H. Luecke, *Science*, 2004, **306**, 1390-1393.
16. N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc Natl Acad Sci U. S. A.*, 2001, **98**, 10037-10041.
17. W. Humphrey, A. Dalke and K. Schulten, *J Mol Graph*, 1996, **14**, 33-38.
18. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Res*, 2004, **32**, W665-W667.
19. M. Hoffmann, M. Wanko, P. Strodel, P. H. König, T. Frauenheim, K. Schulten, W. Thiel, E. Tajkhorshid and M. Elstner, *J. Am. Chem. Soc.*, 2006, **128**, 10808–10818.
20. H. Houjou, Y. Inoue and M. Sakurai, *J. Phys. Chem. B*, 2001, **105**, 867-879.
21. K. Fujimoto, J.-y. Hasegawa and H. Nakatsuji, *Chem. Phys. Lett.*, 2008, **462**, 318-320.
22. A. Altun, S. Yokoyama and K. Morokuma, *J. Phys. Chem. B*, 2008, **112**, 16883-16890.

23. A. Altun, S. Yokoyama and K. Morokuma, *J. Phys. Chem. B*, 2008, **112**, 6814-6827.
24. R. Rajamani, Y.-L. Lin and J. Gao, *J. Comput. Chem.*, 2010, **32**, 854-865.
25. K. F. Hall, T. Vreven, M. J. Frisch and M. J. Bearpark, *J. Mol. Biol.*, 2008, **383**, 106-121.
26. J. S. Främcke, M. Wanko, P. Phatak, M. A. Mroginski and M. Elstner, *J. Phys. Chem. B*, 2010, **114**, 11338-11352.
27. K. Fujimoto, S. Hayashi, J. Hasegawa and H. Nakatsuji, *J. Chem. Theory Comput.*, 2007, **3**, 605–618.
28. L. M. Frutos, T. Andruniow, F. Santoro, N. Ferré and M. Olivucci, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 7764-7769.
29. A. Migani and M. Olivucci, in *Conical Intersections: Electronic Structure, Dynamics and Spectroscopy*, eds. W. Domcke, D. R. Yarkony and H. Köppel, World Scientific, Singapore, 2004, pp. 271-320.
30. M. Olivucci, *Computational Photochemistry*, Elsevier, Amsterdam, 2005.
31. A. Sinicropi, T. Andruniow, L. De Vico, N. Ferré and M. Olivucci, *Pure Appl. Chem.*, 2005, **77**, 977-994.
32. A. Sinicropi and M. Olivucci, in *Theory and Applications of Computational Chemistry: The First 40 Years A Volume of Technical and Historical Perspectives*, eds. C. E. Dykstra, G. Frenking, K. S. Kim and G. E. Scuseria, Elsevier, Amsterdam,, 2005.

Table S1. λ^{a}_{max} expressed in nm and corresponding vertical excitation energy in kcal mol^{-1} of retinal *in vacuo*, at the geometry extracted from the QM/MM optimized Model I structures.

Structure	ASR_{AT} λ^{a}_{max}	ASR_{AT} S_0-S_1 ΔE	ASR_{13C} λ^{a}_{max}	ASR_{13C} S_0-S_1 ΔE
WT	596.0	47.97	573.3	49.87
S214D	596.0	47.97	571.7	50.01
V112N	594.1	48.12	575.0	49.72
L83Q	595.7	47.99	575.6	49.67
W76F	596.2	47.96	571.6	50.02
Y73Q	596.0	47.97	573.3	49.87
P206Q	596.8	47.91	573.4	49.86
P206D	594.0	48.13	572.3	49.96
P206E	595.8	47.99	571.5	50.03
S86D	596.0	47.97	573.2	49.88
E36Q	596.6	47.92	573.3	49.87
D217N	596.9	47.90	573.1	49.89
D217E	596.0	47.97	573.2	49.88

Table S2. Distance of the protonated Schiff base nitrogen from the two carboxyl oxygens in D75 counterion for ASR_{AT}.

Structure	Distance (a)	Distance (b)
WT	4.14	4.07
S214D	4.15	4.08
V112N	4.15	4.08
L83Q	4.15	4.06
W76F	4.15	4.07
Y73Q	4.15	4.08
P206Q	3.99	4.04
P206D	4.17	4.10
P206E	3.99	4.04
S86D	4.15	4.08
E36Q	4.15	4.07
D217N	4.15	4.07
D217E	4.15	4.08

Values are reported in Å. Distance (a) is from the PSB N to the D75 atom labeled OD1, distance (b) is from the PSB N to the D75 atom labeled OD2.

Table S3. Distance of the protonated Schiff base nitrogen from the two carboxyl oxygens in D75 counterion for ASR_{13C}.

Structure	Distance (a)	Distance (b)
WT	3.58	3.89
S214D	3.59	3.90
V112N	3.59	3.90
L83Q	3.59	3.90
W76F	3.59	3.89
Y73Q	3.59	3.89
P206Q	3.59	3.89
P206D	3.60	3.90
P206E	3.38	3.81
S86D	3.59	3.89
E36Q	3.59	3.89
D217N	3.59	3.89
D217E	3.59	3.89

Values are reported in Å. Distance (a) is from the PSB N to the D75 atom labeled OD1, distance (b) is from the PSB N to the D75 atom labeled OD2.

Table S4. CASPT2 energies in atomic units of all roots for Model II structures.

	ASR_{AT}			ASR_{13C}		
Structure	S_0	S_1	S_2	S_0	S_1	S_2
WT	-870.96716	-870.88334	-870.84626	-870.96824	-870.88193	-870.84199
S214D	-870.96796	-870.88402	-870.84691	-870.97933	-870.89006	-870.85178
V112N	-870.97111	-870.88860	-870.85094	-870.97183	-870.88709	-870.84620
L83Q	-870.96271	-870.88124	-870.84362	-870.97104	-870.88530	-870.84383
W76F	-870.97333	-870.88640	-870.85103	-870.97478	-870.88510	-870.84698
Y73Q	-870.96979	-870.88547	-870.84867	-870.96942	-870.88213	-870.84257
P206Q	-870.96405	-870.88072	-870.84274	-870.96019	-870.87615	-870.83593
P206D	-870.98903	-870.89711	-870.86693	-870.98865	-870.89518	-870.86117
P206E	-870.96232	-870.88091	-870.84274	-870.96144	-870.887735	-870.83644
S86D	-870.96815	-870.88422	-870.84710	-870.96952	-870.88283	-870.84299
E36Q	-870.93918	-870.85849	-870.81998	-870.94043	-870.85647	-870.81537
D217N	-870.93431	-870.85457	-870.81584	-870.94132	-870.85778	-870.81653
D217E	-870.96830	-870.88447	-870.84734	-870.96973	-870.88315	-870.84329