

## Supporting Information

For

### Color modeling of protein optical probes

**N. Arul Murugan<sup>†</sup>, Jacob Kongsted<sup>‡</sup>, Zilvinas Rinkevicius<sup>†</sup> and Hans Ågren<sup>†</sup>**

<sup>†</sup>Department of Theoretical Chemistry and Biology

School of Biotechnology

Royal Institute of Technology

SE-10691 Stockholm, Sweden

<sup>‡</sup>Department of Physics and Chemistry

University of Southern Denmark

Campusvej 55

DK-5230 Odense M, Denmark

<sup>§</sup>Swedish e-Science Research Center (SeRC)

Royal Institute of Technology

SE-100 44 Stockholm, Sweden

# 1 Computational details

## 1.1 Structure modeling using hybrid QM/MM molecular dynamics simulations

The protein chosen for this study is bovine BLG present in cow milk and is known to possess a single hydrophobic cavity. The structure for this protein has been taken from the NMR structure available in the protein data bank<sup>1</sup> entry 1B0O. This protein consists of 162 residues.<sup>2</sup> We aim to study this in normal aqueous solvent with pH corresponding to 7. So, the charges for the residues were chosen as per this pH condition. Thereby, the residues such as LYS and ARG carry a positive charge while the residues such as ASP and GLU carry a negative charge. However, both the histidine residue has been chosen to be protonated since they were both exposed to the water solvent. The total charge of the protein is -8.

In all simulations we assume that nile red exists in the deprotonated form. A  $pK_a$  calculation performed using the  $pK_a$  module as implemented in Jaguar<sup>3</sup> revealed a  $pK_a$  value of 5.1 for N,N-diethylaniline, which can be seen as a model system for nile red. Thus the assumption of only simulating the fully deprotonated form of nile red is justified. We have aimed at studying the structure and properties of nile red in three different situations (i) in bulk water (ii) on the BLG protein surface (iii) and within the hydrophobic cavity of the protein. For this we have carried out three different molecular dynamics simulations. First, a single molecule of nile red was optimized at the HF level using the 6-31+G\* basis set using the gaussian03 software.<sup>4</sup> This optimized structure along with the GAFF<sup>5</sup> force-field has been used to describe nile red in all three sets of molecular dynamics calculations. The partial charges used for nile red in these simulations were obtained using the CHELPG<sup>6</sup> procedure as implemented in the gaussian03 software. The first set of molecular dynamics simulations consisted of nile red in water solution. Here one molecule of nile red was solvated in 3876 water molecules and the size of initial simulation box was approximately, 54.4, 48.3 and 45.2 Å. For the water solvent the TIP3P<sup>7</sup> force-field has been used. In the second set of simulations, nile red was placed on the surface of the BLG protein closer to the hydrophobic cavity. An initial configuration is shown in Figure 2. In this simulation, the force-field used to describe the protein was PARM99<sup>8</sup> while the

water molecules were still described by the TIP3P water model. Since the protein carries a total charge of -8, the entire system was neutralized by adding 8 Na<sup>+</sup> ions. In this case, the simulation box contained 12308 water molecules and the initial simulation box had dimensions of 71.2, 75.2 and 74.1 Å. As in the first molecular dynamics simulation Nile Red was described using the GAFF force-field. In the third and last set of molecular dynamics simulations we have studied the Nile Red within the protein cavity. The simulation box size was 70.5, 74.5 and 73.3 Å. All other descriptions for the simulation box remain the same as the two previously described cases. All simulations were initiated with a minimization to remove any unwanted hot-spots in the initial configuration due to any distorted geometry. Following this minimization, a scaling run has been carried out to bring the system to the temperature of 300 K. Finally, the simulations were carried out in the isothermal-isobaric ensemble where the temperature was controlled by Nose thermostat while the pressure was controlled by Barendsen barostat. The molecular dynamics simulations were carried out using the Sander module of Amber08 software.<sup>9</sup> The time step used to solve the equation of motion for these systems was 1 fs. The long range electrostatic interaction within the system was computed using PME Ewald sum method. During all three runs, the Nile Red was kept rigid in its optimized geometry while the remaining system was allowed to relax and evolve in the given temperature and pressure. We have done this since we wanted to treat the Nile Red as a QM region in the hybrid QM/MM runs. The simulations were carried out until the system is equilibrated.

The final configuration from these equilibrated MD runs were used as input configurations for the subsequent hybrid QM/MM simulations. In these simulations, the Nile Red was described using density functional level of theory while the remaining subsystem was described using similar force-fields as detailed for the initial classical simulations described above. The QM/MM implementation<sup>10</sup> as in CPMD/GROMOS<sup>11</sup> software used here includes the coupling between the electron density of the QM subsystem and the instantaneous electrostatic field arising due to the dynamic MM environment. The interaction between the QM and MM subsystems involves electrostatic, short range repulsion and long range dispersion interaction terms (using the empirical van der Waals parameters). In our present calculations, we have used the Becke, Lee, Yang and Parr (BLYP) gradient corrected functional<sup>12</sup> and the Troullier-Martins norm conserving pseudopotentials.<sup>13</sup> Here, the electronic wavefunction is expanded in a plane wave basis set. The energy

cutoff used was 80 Ry. We have used 5 au as the time step for the integration of the equation of motion and 800 amu as the fictitious electronic mass. The QM/MM molecular dynamics calculation starts with a quenching run that relaxes the initial structure within the QM/MM setup. Subsequently, a temperature scaling run was carried out for 0.5 ps to bring the system temperature to 300 K. Finally, the system was connected to the Nose-Hoover thermostat to carry out the Nose run. The length of the production runs were approximately between 20-30 ps for the three systems. During this production run a number of configurations were saved for the QM/MM calculations (see the discussion below).

## **1.2 Property modeling using QM/MM response approach**

### **1.2.1 Calculations on the isolated NR molecule**

Optical and spin probes display substantial changes in molecular geometry depending upon the nature of the microenvironment. These structural changes alone contribute significantly to properties such as absorption maximum, fluorescence maximum and hyperfine coupling constant. It is therefore important to quantify the contribution to properties solely due to these structural changes. For this purpose, we have calculated the absorption spectra for the isolated Nile red molecule but taken from the trajectory corresponding to hybrid QM/MM simulation for Nile red in water and within the protein environment. The results from this set of calculations will be referred to as QM/MM-0. To obtain the absorption spectra, we have used time-dependent density functional theory at the CAM-B3LYP<sup>14</sup> level and the Turbomole-TZVP<sup>15</sup> basis set has been used for NR in these calculations.

### **1.2.2 Calculations on the NR molecule in protein and bulk solvent environment**

In the second approach the Nile red and the discrete nature of environment is included in the property calculation. For this we have used the QM/MM response technique which retains the discrete description of the solvent and includes the solute-solvent interaction through different coupling terms in the QM effective Hamiltonian. For all the QM/MM

calculations, we have used the development version of DALTON2.0<sup>16</sup>. In the present work the solvent molecules are described either using atom centered partial charges or with atom centered partial charges and polarizabilities. The use of charges alone in the solvent model only accounts for the polarization of the solute due to the solvent and the solvent back-polarization is neglected. However, the solvent polarization is usually included implicitly through the use of enhanced values for the charges. On the other hand, when both atomic charges and polarizabilities are used in the description of the solvent molecules both the solute polarization and the explicit back polarization effects are included. The calculations using the solvent description based on only charges located on the atomic sites of the solvent molecules will be referred as MM-1 while the model that places the charges and polarizability on atomic sites of solvents will be referred as MM-2. In MM-1 the water model used is TIP3P and in MM-2 it is the modified Ahlström potential.<sup>17</sup> In the third model which will be referred to as MM-3, we also include a few solvent molecules into the QM region explicitly. Here, we have included 4 water molecules in the QM region explicitly, while the rest of the solvent molecules have been described using the modified Ahlström potential. This description allows for the inclusion of charge transfer between the solute and a few "influential" solvent molecules. The selection of water molecules is based on the notion that the solvent molecules close to the acceptor and donor groups are more "influential" and involved in charge transfer. Thereby, we have included 1 water molecule closest to each of the nitrogen and oxygen atoms of the NR molecule. So, overall 4 water molecules were included explicitly. In the case of modeling properties of NR in the protein environment, we have adopted only the MM-1 level. Based on the fact that the cavity is hydrophobic in nature, it was not attempted to include explicitly polarization of the environment since from our earlier studies we find that backpolarization effect and explicit inclusion of the environment does not contribute significantly to properties in a non-polar environment. The validity of this approximation will be detailed more in the following section. During the molecular dynamics simulation production runs a number of configurations were saved for the QM/MM calculations. The specific number of configurations used for the property modeling varies depending on the specific type of QM/MM calculations (as detailed below). In the most inexpensive cases, e.g. Nile red simulated in TIP3P water, we have used up to 140 configurations. In practice a convergence in the calculated properties is observed after inclusion of 40 configurations and therefore we have in the more expensive QM/MM calculations used

this number of configurations. The estimated statistical error related to the predicted absorption maxima (the standard error of the mean) is approximately 5 nm.

For the preparation of the configurations to the QM/MM calculations Nile red was first centered in the simulation box. Next a center-of-mass based spherical cut-off was used to include all water molecules within 20 Å from the Nile red. If the protein was included in the simulation all atoms within the protein were included in the QM/MM calculations.

## References

- [1] S.Y. Wu, M.D. Perez, P. Puyol, L. Sawyer, *J.Biol.Chem.*, **274**, 170(1999).
- [2] S. Uhrinova, M.H. Smith, G.B. Jameson, D. Uhrin, L. Sawyer, P.N. Barlow, *Biochemistry*, **39(13)**, 3565(2000).
- [3] Jaguar, version 7.7, Schrödinger, LLC, New York, NY, 2010
- [4] M.J. Frisch, *et al.*, *Gaussian 03, Rev. B.05*. **2003**, Gaussian, Inc. Pittsburgh, PA.
- [5] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, *J. Comp. Chem.*, **25(9)**, 1157(2004).
- [6] C.M. Breneman, K.B. Wiberg, *J. Comp. Chem.*, **11**, 361(1990).
- [7] W.L. Jørgensen, *J. Am. Chem. Soc.*, **103**, 335(1981).
- [8] S. Chalmet, H. Harb, M.F. Ruiz-Lopez, *J. Phys. Chem. A*. **2001**, *105*, 11574.
- [9] D.A. Case, *et al.*, AMBER8. 2004. University of California, San Francisco, CA.
- [10] (a) A. Warshel, M. Levitt, *J. Mol. Biol.* **103**, 22724 (1976). (b) H.M. Senn, W. Thiel, *Angew. Chem. Int. Ed.*, **48**, 1229(2009). (c) J.L. Gao, D.G. Truhlar, *Annu. Rev. Phys. Chem.*, **53**, 467(2002). (d) P.K. Biswas, V. Gogonea, *J. Chem. Phys.*, **129**, 154108(2008). (e) M. Eichinger, P. Tavan, J. Hutter, M. Parrinello, *J. Chem. Phys.*, **110**, 10452(1999). (f) U.C. Singh, P.A. Kollman, *J. Comp. Chem.*, **7(6)**, 718(1986). (g) M.J. Field, P.A. Bash, M. Karplus, *J. Comp. Chem.*, **11(6)**, 700(1990).

- [11] (a) J. Hutter, M. Parrinello, D. Marx, P. Focher, P.; M. Tuckerman, W. Andreoni, A. Curioni, E. Fois, U. Röthlisberger, P. Giannozzi, T. Deutsch, A. Alavi, D. Sebastiani, A. Laio, J. VandeVondele, A. Seitsonen, S. Billeter, *Computer code CPMD, version 3.11*, Copyright IBM Corp. and MPI-FKF Stuttgart 1990–2002. (b) A. Laio, J. VandeVondele, U. Röthlisberger, *J. Phys. Chem. B.*, **106**, 7300(2002). (c) A. Laio, J. VandeVondele, U. Röthlisberger, *J. Chem. Phys.*, **116**, 6941(2002).
- [12] (a) A.D. Becke, *Phys. Rev. A.*, **38**, 3098(1988). (b) C. Lee, W. Yang, R.C. Parr, *Phys. Rev. B.*, **37**, 785(1988).
- [13] N. Trouiller, J.L. Martins, *Phys. Rev. B.*, **43**, 1993(1991).
- [14] (a) T. Yanai, D. P. Tew, N. C. Handy, *Chem. Phys. Lett.*, **393**, 51 (2004). (b) M. J. G. Peach, T. Helgaker, P. Salek, T. W. Keal, O. B. Lutns, D. J. Tozer, N. C. Handy, *Phys. Chem. Chem. Phys.*, **5**, 558 (2006).
- [15] (a) A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.*, **97**, 2571(1992). (b) A. Schäfer, C. Huber, R. Ahlrichs, *J. Chem. Phys.*, **100**, 5829 (1994).
- [16] "DALTON, a molecular electronic structure program, Release 2.0 (2005), see <http://www.kjemi.uio.no/software/dalton/dalton.html>"
- [17] P. Ahlström, A. Wallqvist, S. Engström, B. Jönsson, *Mol. Phys.*, **68**, 563(1989).