Many-body meets QM/MM: Application to indole in water solution

Adriano Mosca Conte*, Emiliano Ippoliti2, Rodolfo Del Sole1, Paolo Carloni2, and Olivia Pulci1

1 NAST, ETSF, CNR-INFM-SMC, Physics Dep., University of Rome Tor Vergata, Via della Ricerca Scientifica 1, Roma, Italy
2 Democritos, SISSA Scuola Internazionale Superiore di Studi Avanzati, via Beirut 2–4, 34014 Trieste, Italy

Received 9 November 2009, revised 13 January 2010, accepted 22 February 2010
Published online 15 June 2010

Keywords electronic structure calculations, molecular physics, optical absorption spectra

*Corresponding author: e-mail adriano.mosca.conte@roma2.infn.it, Phone: +39 06 72594741, Fax: +39 06 2023507

Aromatic chromophores are used as optical probes to investigate the structure of complex biological systems. For this purpose, it is crucial to understand the role of the bio-chemical environment in the modification of their optical properties. Here we present a theoretical method to analyze this aspect on large scale systems, like biologically relevant molecules in aqueous solution. We combine many body perturbation theory with a quantum-mechanics/molecular-mechanics (QM/MM) approach. We include quasi-particle and excitonic effects for the calculation of optical absorption spectra in a QM/MM scheme. In agreement with the experiments, we find that the solvent induces a redshift in the main spectral peak of indole, and we demonstrate this shift originates both from electrostatic and geometrical effects.

1 Introduction

16.6% of the human body weight is due to proteins. Proteins are the second most abundant body constituent after water (61% in weight). They are mostly present inside the cell membranes and regulate a huge variety of functions. They act as enzymes in all the chemical reactions such as the ones involved in the respiration or in metabolism, they play a crucial role in cellular signaling and molecular transport. These are just a small number of the possible functions. Proteins are essential macromolecules without which our bodies would be unable to repair, regulate, assimilate food, or protect themselves.

A protein is essentially a long chain of amino acids. The final shape of the protein depends on the aminoacidic sequence and on the formation conditions. The amino acids forming the proteins in living beings are only 20. Three of them include aromatic rings in the side chain: tryptophan, tyrosin, and phenylalanine. This chemical characteristic enables these three so-called aromatic amino acids to extend their absorption optical spectrum range up to the close ultraviolet region. The optical active part of the molecule is called chromophore. The chromophore of tryptophan is indole and is formed by two aromatic carbon rings (one exagonal and the other pentagonal containing a nitrogen atom, see Figure in abstract). Due to its optical properties, that strongly depend on the particular bio-chemical environment, indole is used as a probe in proteins.

Actually, the analysis of the optical properties of biological systems is often the only possible method to characterize them in their specific bio-chemical environments: the absorption and emission lines of a biomolecule, called spectral signatures, are useful for its identification, and the shift of the signatures induced by the interaction with the environment (called solvatochromic shift or solvatochromism) gives information on the full system.
Theoretical simulations to predict optical properties can help to better understand these systems. A completely ab initio, parameter-free, determination of the excited state properties of materials is very important to interpret the experimental spectra and the chemical features of the system. A problem of biological systems is their large size making the quantum calculations computationally very expensive, and in many cases impossible. A solution to this problem is to combine the quantum mechanics (QM) theoretical methods with molecular mechanics (MM) one by the so-called QM/MM techniques [1, 2]. This theoretical scheme has been applied in the past for the calculation of ground state and optical properties by interfacing methods based on TDDFT [3, 4] with molecular mechanics [5–7]. In this work we show that it is possible to implement a similar interface also for many-body methods (GW and Bethe–Salpeter equation [8–10]) and we apply this scheme to indole in water solution. We investigate the physical reasons of the red shift induced on the lower part of the indole spectrum by the presence of the water solution and compare our results with the one obtained in this work by TDDFT/MM methods.

2 Methods To obtain the optical spectrum we follow a classical scheme involving DFT [11] calculations providing Kohn–Sham solutions, then we obtain the quasi-particle energies by the GW method [8], and finally we solve the Bethe–Salpeter equation (BSE) [9]. The DFT calculations are performed in a QM/MM scheme, by expanding wavefunctions and charge density in plane waves, and by using Tuckerman–Martyna method [12] to avoid Coulomb interactions between periodic quantum replicas. Further details concerning DFT calculations are discussed later.

The GW approximation consists into set the vertex in the Hedin equations [13] equal to a delta function. Under this condition, the time Fourier transform of the proper self-energy $\Sigma(x, x', \omega)$ is a convolution of the Green’s function $G(x, x', \omega)$, with the screened Coulomb potential $W(x, x', \omega)$. The electronic bands are obtained solving the following eigenproblem:

$$\left[ -\frac{\hbar^2 \nabla^2}{2m} + V_H(x) + U_{QM}(x) + U_{QM/MM}(x) \right] \phi_i(x) + \int d^3x_2 \Sigma(x_1, x_2, \epsilon_{OP}^i) \phi_j(x_2) = \epsilon_{QP}^i \phi_j(x_1).$$

(1)

This expression is derived from the Dyson equation for $G$ in Lehman representation [8]. $U_{QM}$ is the electron–ion potential of the QM part, while $U_{QM/MM}$ is the potential felt by the electrons due to the point charges of the MM part. Finally, $\epsilon_{QP}^i$ are the quasi-particle eigenvalues. The main point to be stressed here is that $U_{QM/MM}$ includes the potential induced by the atoms of the classical region. This is a first important assumption to insert the GW method in the QM/MM scheme. Equation (1) has the same form of a KS equation, [11] where the DFT exchange-correlation potential $V_{xc}(x)$ is replaced by the self-energy $\Sigma(x_1, x_2, \epsilon_{OP}^i)$ that acts as a nonlocal potential. Therefore the eigenvalue problem described above can be solved perturbatively considering the KS equation as an unperturbed Hamiltonian and $\Sigma - V_{xc}$ as a perturbative term. The quasi-particle eigenvalues are obtained through the following expression:

$$\epsilon_{QP}^i = \epsilon_{KS}^i + \frac{\langle KS | \Sigma(\epsilon_{OP}^i) - V_{xc} \rangle | KS \rangle}{1 - \langle KS | \frac{\partial^2 \Sigma}{\partial \omega^2} | KS \rangle}.$$

(2)

All the Coulomb interactions, and hence also the one induced by the classical part, are included in the KS eigenvalues $\epsilon_{KS}^i$, and eigenvectors $|KS\rangle$, that are used to calculate the $\Sigma$ in the GW/MM scheme. In Eq. (2) we use a cutoff in the real space for the self-energy to prevent the periodic images of the quantum part to interact with each other, following Ref. [14].

The GW method not only provides a way to calculate the energy bands, but it also allows the calculation of some of the main ingredients necessary to solve the BSE giving the four-point polarization $P^4$ and consequently the optical spectrum ($^4P = ^4P_0 + ^4P_0(v - W)^4P$). These main ingredients are the four-point screened Coulomb interaction $^4W(1234) = ^4W(12)\delta(13)\delta(24)$, and the four-point independent particle polarizability, that in transition space takes the form:

$$p_0^{0123}(n_1n_2n_3n_4)(\omega) = \frac{f_{n_2} - f_{n_1}}{\epsilon_{QP}^i - \epsilon_{QP}^j - \omega} \delta_{n_1n_3}\delta_{n_2n_4}.$$  

(3)

Given these two quantities, the BSE is the same [15] of a all-QM case. The main approximation of this procedure is that we do not consider the excitonic effect due to the presence of electrons in water unoccupied states or holes in water occupied states. We demonstrate in the next section that this approximation can be done for the low absorption energy peaks of indole, and does not induce errors in the spectrum calculation for the specific case of indole in water.

3 Results and discussion We performed a 20 ps hybrid QM/MM Car–Parrinello [16] simulation of an indole molecule (QM part) [17] surrounded by 1376 water molecules treated classically (MM part), with the Amber force field. Since the Full Hamiltonian QM/MM scheme we used [2] is a dual simulation box method, it is not obvious to understand how many water molecules are necessary to correctly reproduce the physical properties of a disordered system such as liquid water at room temperature (300 K). On the other hand, it was clear already at the DFT-independent particle approach (DFT-IPA) that an underestimation of such a number could take to incorrect results for the optical properties: in fact we calculated the absorption spectrum for indole with no water (with indole in the same “distorted” geometrical configuration as if in water), for indole with 2 water molecules, and for indole with 1376 water molecules (Fig. 1). The three spectra are all different. This is due to the long range electrostatic potential of water which acts on indole. Indeed 30–60 water molecules may be enough for...
many hydration properties. However, we decided to use 1376 H2O in order to be on the safe side without the need of convergence tests.

Next, we tested our assumption that the solvent can be treated classically in the calculation of absorption spectra by performing TDDFT calculations [25] for a system where two water molecules were treated at quantum level, and the remaining 1374 classically. The position of the absorption peak is the same (within 0.02 eV) as in the case where all waters were treated within MM. Our result supports the use of this approach for solutes in water [6]. A complete discussion on this subject including an analysis of the hydrogen bonds between solute and solvent is reported in Ref. [26].

For ten snapshots of the QM/MM dynamics (one every 2 ps) we computed the optical spectra at the independent particle level (DFT-IPA) and within TDDFT. The final spectrum was obtained by an average over the snapshots. The convergence of the spectrum was reached already after 6 snapshots, hence the subsequent GW and BSE calculations have been performed on only 6 snapshots.

The calculated DFT and GW HOMO–LUMO gaps [30], averaged over the QM/MM configurations, are 3.8 eV (with standard deviation $\sigma = \pm 0.1$ eV) and 7.2 eV ($\sigma = \pm 0.2$ eV), respectively. The HOMO–LUMO gap calculated by GW method corresponds to the electronic gap, and not to the optical gap. The GW corrections to the gap turned out to be practically constant in all the snapshots considered ($3.4 \pm 0.1$ eV). This fact, already found for liquid water [31], confirms that one can in principle strongly reduce the computational effort, by performing a GW calculation for just one snapshot.

We finally calculate the low energy range of the optical spectrum of indole, by GW–BSE and TDDFT, always as result of an average over the QM/MM snapshots. In Figs. 2 and 3 we report our results together with the calculated absorption spectrum in gas phase. We notice that, in both approaches, the most intense peak ($^1L_a$) is red-shifted on passing from gas phase to the water solution. The same conclusion was obtained by previous theoretical calculations of indole in water based on CASSCF and CASPT2 methods [33]. But, the value we calculate for such a redshift is $\sim 0.2$ eV, in good agreement with experiment (0.18 eV) [32]. On the contrary the CASSCF and CASPT2 prediction for the solvent shift is about 0.06 eV only. Such an underestimation may depend on the geometrical distortion of indole molecule caused by temperature effects due to the solvent and by an explicit H-bonding between water molecules, which were not considered explicitly therein. To quantify the effect of the geometry distortion on such shift, we performed calculations of indole switching on and off the water field in order to separate the geometry effect from the electrostatic ones. The results, presented for a single snapshot in Fig. 4, are obtained by performing additional calculations for the same QM/MM configuration without the water field. The corresponding solvent-shift goes from $-0.1$ eV with water field to $+0.2$ eV (hence, a blue-shift) without water field. This emphasizes the importance of taking into account explicitly the electrostatic interaction with the solvent, since the geometry distortion alone would...
The optical shift enables to understand the nature of their proteins and cell membranes, for which the evaluation of MBPT/MM to other bio-relevant molecules, such as optical calculations.

one snapshot (several, indeed) for carrying out accurate environment.

dipoles. Both effects, and their sum, depend on the particular the electrostatic interaction with the water molecules electric geometrical distortion of indole molecule in the solvent and shift is a consequence of the combination of two effects: the Our GW–BSE calculations further show that the solvent

0.25 eV. As expected [34], CASSCF is much worse, it overestimates by ~1 eV or more, whilst CASPT2 is more accurate (~0.13 eV for the gas phase).

In conclusion, we have included many-body perturbative techniques in a QM/MM scheme. We have applied it, together with a TDDFT/MM approach, to study the optical properties of indole in water solution. Both methods reproduce quantitatively the redshift induced by the solvent.

give, at least for this snapshot, a wrong sign in the solvatochromoshift.

TDDFT underestimates the energy position of the $^1L_a$ peak both in gas phase and in solution by ~0.4 eV, and BSE–GW overestimates them by ~0.3 eV. As expected [34], CASSCF is much worse, it overestimates by ~1 eV or more, whilst CASPT2 is more accurate (~0.13 eV for the gas phase).

In conclusion, we have included many-body perturbative techniques in a QM/MM scheme. We have applied it, together with a TDDFT/MM approach, to study the optical properties of indole in water solution. Both methods reproduce quantitatively the redshift induced by the solvent. Hence the GW–BSE method can be applied to biomolecules in aqueous solution (i.e., in laboratory-realizable conditions) in cases where the TDLDA/GGA approach does not work [35, 36]. Moreover, a better exchange and correlation kernel can be derived from MBPT to improve TDLDA/GGA [37]. Our GW–BSE calculations further show that the solvent shift is a consequence of the combination of two effects: the geometrical distortion of indole molecule in the solvent and the electrostatic interaction with the water molecules electric dipoles. Both effects, and their sum, depend on the particular configuration of the system; this emphasizes the need of including both altogether, and of averaging over more than one snapshot (several, indeed) for carrying out accurate optical calculations.

This work opens the way to further applications of MBPT/MM to other bio-relevant molecules, such as proteins and cell membranes, for which the evaluation of the optical shift enables to understand the nature of their environment.

Acknowledgements The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no 211956. Computer resources from INFM “Progetto Calcolo Parallelo” at CINECA are gratefully acknowledged. We also thank L. Guidoni for interesting discussions.

References

[1] E. Carter and J. Hynes, J. Chem. Phys. 94, 5961 (1991).
[2] A. Laio, J. VandeVondele, and U. Röthlisberger, J. Chem. Phys. 116, 6943 (2002).
[3] E. Runge and E. K. U. Gross, Phys. Rev. Lett. 52, 997–1000 (1984).
[4] M. A. L. Marques and E. K. U. Gross, Time-Dependent Density Functional Theory (Springer-Verlag, Berlin, Heidelberg, 2003).
[5] M. Sulpiz, P. Carloni, J. Hutter, and U. Röthlisberger, Phys. Chem. Chem. Phys. 5, 4798–4805 (2003).
[6] M. Sulpiz, U. F. Röhrig, J. Hutter, and U. Röthlisberger, Int. J. Quantum Chem. 101, 671–682 (2004).
[7] M.-E. Moret, E. Tapavicza, L. Guidoni, U. F. Röhrig, M. Sulpiz, I. Tavernelli, and U. Röthlisberger, CHIMIA Int. J. Chem. 59, 493–498 (2005).
[8] A. L. Fetter and J. D. Walecka, Quantum Theory of Many-Particle Systems (McGraw-Hill, San Francisco, 1971).
[9] For a review see G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. 74, 601–659 (2002).
[10] Y. Ma, M. Rohlfing, and C. Molteni, J. Chem. Theory Comput. 6(1), 257–265 (2010).
[11] W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
[12] G. J. Martyna and M. E. Tuckerman, J. Chem. Phys. 110, 2810–2821 (1999).
[13] L. Hedin, Phys. Rev. 139, A796 (1965).
[14] G. Onida, L. Reining, R. W. Godby, R. Del Sole, and W. Andreoni, Phys. Rev. Lett. 75, 818–821 (1995).
[15] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. 74, 601–659 (2002).
[16] R. Car and M. Parrinello, Phys. Rev. Lett. 55, 2471–2474 (1985).
[17] The QM/MM code combines the CPMD3.11.1 [18] and Gromos [19] codes. The time step chosen for the simulations was about 0.1 fs. Since the indole’s N–H group forms H-bonds with water, we used an energy cutoff of 70 Ry with the BLYP recipe [20] for the exchange-correlation functional, which have widely been used for biophysical applications. Norm-conserving pseudopotentials of Troullier–Martins type [21] corrected for a better description of Van der Walls interactions [22] have been used. A Nose–Hoover thermostat [23, 24] is applied throughout all simulations to keep the temperature constant.
[18] J. Hutter, A. Alavi, T. Deutsch, P. Ballone, M. Bernasconi, P. Focher, S. Goedecker, M. Tuckerman, and M. Parrinello, CPMD, Max-Planck-Institute für Festkörperforschung, Stuttgart and IBM Research Laboratory Zürich, 1995–1999.
[19] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, Biomolecular Simulation: The GROMOS96 Manual and User Guide; Vdf Hochschulverlag AG an der ETH Zürich, Zürich, 1996.
[20] A. D. Becke, Phys. Rev. A 38, 3098 (1988); C. Lee, W. Yang, R. G. Parr, Phys. Rev. B 37, 785 (1988).
[21] N. Troullier and J. L. Martins, Phys. Rev. B 43, 1993 (1991).
[22] O. A. von Lilienfeld, I. Tavernelli, U. Röthlisberger, and D. Sebastiani, Phys. Rev. Lett. 93, 153004 (2004).
[23] S. J. Nosé, J. Chem. Phys. 81, 511 (1984).
[24] D. J. Evans and B. L. Holian, J. Chem. Phys. 83, 4069 (1985).
[25] TDDFT calculations are obtained within the Tamm-Dancoff approximation [27, 28] as implemented in CPMD 3.11.1 package, using the PBE [29] exchange-correlation functional. The same corrected Trouiller–Martins pseudopotentials as for the dynamics have been used.
[26] A. Mosca Conte, E. Ippoliti, R. Del Sole, P. Carloni, and O. Pulci, J. Chem. Theory Comput. 2009(5), 1822 (2009).
[27] S. Hirata and M. Head-Gordon, Chem. Phys. Lett. 314, 291–2299 (1999).
[28] J. Hutter, J. Chem. Phys. 118, 3928 (2003).
[29] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).
[30] GW calculations have been done by using 12077 plane-waves and 500 electronic bands. Following [14], we used a cutoff in the real space for the self-energy to prevent that periodic images of the quantum part interact with each other. The screen function is calculated within a plasmon pole approximation.
[31] V. Garbuio, M. Cascella, L. Reining, R. Del Sole, and O. Pulci, Phys. Rev. Lett. 97, 137402 (2006).
[32] H. Lami, J. Chem. Phys. 67, 3274–3281 (1977).
[33] L. Serrano-Andres and B. O. Roos, J. Am. Chem. Soc. 118, 185–195 (1996).
[34] D. M. Rogers and J. D. Hirst, J. Phys. Chem. 107, 11191–11200 (2003).
[35] Z.-L. Cai, K. Sendt, and J. R. Reimers, J. Chem. Phys. 117, 5543 (2002).
[36] A. Dreuw and M. Head-Gordon, J. Am. Chem. Soc. 126, 4007 (2004).
[37] S. Botti, A. Shindlmayr, R. Del Sole, and L. Reining, Rep. Prog. Phys. 70, 357 (2007), and references therein.