Electrostatic considerations affecting the calculated HOMO–LUMO gap in protein molecules

FAST TRACK COMMUNICATION

Electrostatic considerations affecting the calculated HOMO–LUMO gap in protein molecules

Greg Lever[1], Daniel J Cole[1,2], Nicholas D M Hine[3], Peter D Haynes[3] and Mike C Payne[1]

[1] Theory of Condensed Matter group, Cavendish Laboratory, 19 JJ Thomson Ave, Cambridge CB3 0HE, UK
[2] Department of Chemistry, Yale University, New Haven, CT 06520-8107, USA
[3] Departments of Materials and Physics, Imperial College London, Exhibition Road, London SW7 2AZ, UK

E-mail: gl319@cam.ac.uk

Abstract
A detailed study of energy differences between the highest occupied and lowest unoccupied molecular orbitals (HOMO–LUMO gaps) in protein systems and water clusters is presented. Recent work questioning the applicability of Kohn–Sham density-functional theory to proteins and large water clusters (Rudberg 2012 *J. Phys.: Condens. Matter* **24** 072202) has demonstrated vanishing HOMO–LUMO gaps for these systems, which is generally attributed to the treatment of exchange in the functional used. The present work shows that the vanishing gap is, in fact, an electrostatic artefact of the method used to prepare the system. Practical solutions for ensuring the gap is maintained when the system size is increased are demonstrated. This work has important implications for the use of large-scale density-functional theory in biomolecular systems, particularly in the simulation of photoemission, optical absorption and electronic transport, all of which depend critically on differences between energies of molecular orbitals.

S Online supplementary data available from stacks.iop.org/JPhysCM/25/152101/mmedia

1. Introduction

Density-functional theory (DFT) [1, 2] simulations are becoming increasingly widespread for simulating biological systems at the level of individual atoms and electrons. The advantages of DFT over, for example, molecular mechanics (MM) force fields include correct treatment of long-range polarization, charge transfer, bond breaking and forming, electronic states and, by association, spectroscopy. In addition, transition metals, unusual ligands or functional groups that are difficult to correctly parameterize within force fields are also accurately treated. A limiting factor in applying conventional first-principles approaches to large systems is the unfavourable computational requirements which typically increase as the third (or greater) power of the number of atoms in the system. However, methods to overcome this bottleneck using linear-scaling approaches to DFT have been under development for over a decade [3] and are still advancing today [4], allowing system sizes on the order of tens of thousands of atoms to be routinely accessed [5].

As larger systems become accessible, the application of density-functional methods to systems of biological interest is becoming increasingly common. The FHI-aims package [6] has been applied to folding processes in helices and polypeptides [7]. The TeraChem package [8] has recently been used to optimize the structures of more than 50 polypeptides of sizes ranging up to 590 atoms, predicting protein structure with the same accuracy as empirical force fields parametrized extensively for this purpose [9]. Linear-scaling DFT calculations have been reported on a dry DNA model comprising 715 atoms using the SIESTA code [10] and a B-DNA decamer using the CONQUEST code [11] with explicit water molecules and counter-ions [12]. The ONETEP code [13], applied in the present investigation, has previously been used to perform geometry optimizations to characterize binding energetics of small molecules to the metalloprotein myoglobin [14] and to measure binding of small molecules to T4 lysozyme in solution [15]. Recent developments in the simulation of optical spectroscopy [16], dynamical mean field theory with applications to human respiration [17] and methods to aid interpretation of the electronic structure [18] broaden the scope of biomolecular simulations still further.

Despite considerable interest in the use of *ab initio* simulations for the study of complex biomolecular systems, there is a growing concern that DFT, in conjunction with pure exchange–correlation functionals (those defined as containing no Hartree–Fock exchange), may be inappropriate for the simulation of large molecular clusters. In particular, it has recently been shown that there are unphysical vanishing HOMO–LUMO gaps in systems such as proteins [19] and even water clusters [20]. This can lead to poor convergence during the self-consistent electronic structure optimization procedure. Poor self-consistent field (SCF) convergence has been found for BLYP and B3LYP DFT functionals [21] and for LDA and PBE functionals when simulating large glutamic acid–alanine helices [22]. Attempts to optimize polypeptide structures *in vacuo* using the TeraChem package with BLYP and B3LYP functionals reported a lack of SCF convergence for many of the peptides simulated [9]. Similar problems have been observed when using molecular fractionation with conjugated caps to compute *ab initio* binding energies in vacuum for protein–ligand complexes of between 1000 and 3000 atoms [23].

SCF convergence issues are widely blamed upon the well-known phenomenon of pure functionals underestimating the HOMO–LUMO gap [24–26]. However, while the lack of the derivative discontinuity of the exchange–correlation potential at integer particle numbers and errors in the single-particle eigenvalues resulting from the approximate nature of the functional itself will indeed reduce the gap, there is no obvious reason why the effect should worsen at larger system sizes [25]. Indeed, it has been shown that recovery of a sizeable gap and consequently robust self-consistent convergence of the electronic energy levels is possible by including electrostatically embedded point charges to represent water molecules around an inner cluster treated with quantum mechanics (QM) [19, 27] or

by simulating the system in a dielectric medium [23]. These results point to the possibility that the vanishing gap is a surface effect and not an inherent difficulty with pure Kohn–Sham DFT.

In this communication, we use the linear-scaling DFT code ONETEP [13] to investigate the HOMO–LUMO gap of water clusters and protein systems. As in previous studies [19] we find that the gap often vanishes *in vacuo*. However, we provide conclusive evidence that the issue is not related to the use of a pure exchange–correlation functional but rather is a result of the approach used to prepare the system. We suggest a number of practical measures for preparing large systems that do not lead to a vanishing HOMO–LUMO gap and thus open the way for continued investigations of biomolecular systems with Kohn–Sham DFT.

2. Computational method

The present work uses ONETEP [13], a linear-scaling DFT package designed for use on parallel computers [28] that combines near-complete basis set accuracy with a computational cost that scales linearly with the number of atoms. This allows an accurate QM description of systems of thousands of atoms [5], including entire proteins [14, 29, 30]. The ONETEP formalism is based on a reformulation of conventional Kohn–Sham DFT [1, 2] in terms of the single-particle density matrix

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}') \qquad (1)$$

where $\phi_\alpha(\mathbf{r})$ are non-orthogonal generalized Wannier functions (NGWFs) [31] that are localized in real space. The density kernel $K^{\alpha\beta}$ is a representation of the density matrix in the biorthogonal duals of the NGWFs. ONETEP achieves linear scaling by exploiting the 'nearsightedness' of the single-particle density matrix in non-metallic systems [32, 33]. In practice, linear scaling arises from enforcing strict localization of the NGWFs onto atomic regions and through the optimization of the density kernel and NGWFs, subject to localization constraints. Optimizing the NGWFs *in situ* allows for a minimal number of atom-centred orbitals to be used whilst maintaining plane-wave accuracy. The basis set underlying the NGWFs consists of periodic cardinal sine (psinc) functions [34]. The use of a plane-wave basis allows an unbiased approach to DFT calculations with systematically improvable accuracy by varying a single parameter similar to the energy cutoff in conventional plane-wave DFT packages. QM calculations have been performed in ONETEP using the PBE gradient corrected exchange–correlation functional [35]. ONETEP will converge to a minimized electronic energy irrespective of the filling of the states that is applied at the beginning of a simulation. No density kernel truncation has been performed in the current calculations. The kernel is optimized using a combination of methods [36], which has the effect of ensuring integer occupancies of zero or one for all single-electron eigenstates of the Hamiltonian, which is appropriate in bulk insulators with a clearly defined band gap or, as is the case in the systems in the present work, in isolated

systems with a clear HOMO–LUMO gap. In cases where the systems have a zero gap, traditional linear-scaling algorithms for the optimization of the density kernel are not suitable and often converge to unphysical solutions not corresponding to integer occupation of the valence eigenstates. In such cases we have indicated that the system did not converge.

Where indicated in the text the use of implicit solvent has been made. The implicit solvent model is fully self-consistent and based on direct solution of the inhomogeneous Poisson equation. The approach involves defining the solute cavity in terms of an isosurface of the electron density [37] and has been implemented in ONETEP [15, 38]. In addition, the use of electrostatic point charges (denoted by QM/EE in the text) has been made. Electrostatic embedding can significantly reduce the computational cost associated with density-functional calculations by representing a selected portion of the total system in terms of highly localized classical charge distributions that are electrostatically coupled with the quantum system, representing the effect of the environment in which the quantum system is embedded [39]. ONETEP parameters are described in detail in the supplementary methods (available at stacks.iop.org/JPhysCM/25/152101/mmedia).

3. Water clusters

The correct treatment of water is vital for realistic simulations of biomolecular environments. However, recent large-scale DFT simulations with pure functionals have encountered SCF convergence problems when simulating isolated water clusters, due to the HOMO–LUMO gap decreasing to zero when the cluster radius becomes larger than approximately 10 Å [19]. We begin by using the linear-scaling DFT code ONETEP to calculate the band gap of a 2010-atom periodic supercell of bulk water, equilibrated using the classical molecular dynamics package AMBER [40] (supporting methods available at stacks.iop.org/JPhysCM/25/152101/mmedia). Consistent with previous DFT calculations of water employing the PBE functional and as expected for a bulk insulator, the system has a clearly defined band gap of 4.2 eV. We have also calculated the HOMO–LUMO gap of spherical water clusters of increasing radius extracted from a larger 50 Å cube of water equilibrated at 300 K with classical molecular dynamics (MD). For small clusters such as these, quantum confinement would generally be expected to result in a HOMO–LUMO gap greater than that of the bulk, although one might expect the value of the cluster gap to tend to the bulk band gap with increasing system size. However, in agreement with previous work [19], figure 1(a) (black line) shows that the HOMO–LUMO gap quickly approaches zero for systems of more than approximately 200 atoms. This observation has therefore led many to question the applicability of pure DFT functionals to large systems, such as water clusters and proteins [9, 19, 23, 27, 41].

However, our results showing the insulating nature of very large supercells of bulk water indicate that the system size in itself is not a problem but, rather, the small
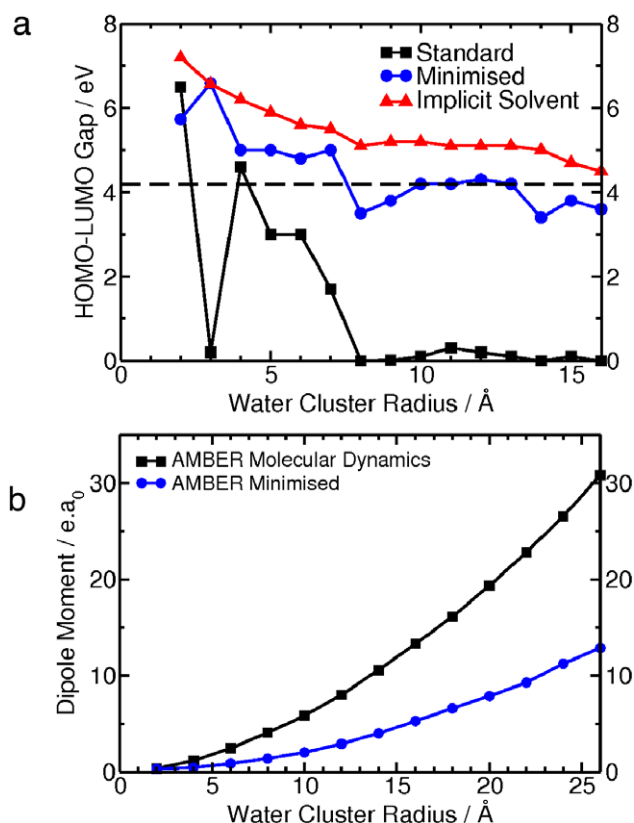
Figure 1. (a) DFT HOMO–LUMO gaps of water clusters of increasing radius extracted from a larger 50 Å cube of water equilibrated at 300 K using classical MD. Black line: extracted straight from bulk water. Blue line: after classical minimization is performed on each extracted cluster. Red line: simulating the extracted clusters in implicit solvent. Dashed line: HOMO–LUMO gap of bulk water. (b) Average dipole moments of water clusters of increasing radius calculated using the TIP3P point charge model. Black line: averaged over 5000 snapshots extracted from a larger 50 Å cube of water. Blue line: averaged over 1800 snapshots extracted from the bulk and minimized using an MM force field.

HOMO–LUMO gap is caused by the vacuum–water interface. Bulk water consists of a continuous hydrogen-bonded network of water molecules, each of which has an intrinsic dipole moment. The process of extracting a cluster, freezing the atomic positions and surrounding with vacuum potentially results in a large surface dipole being created. A rough estimate illustrates this: the dipole moment of a single isolated water molecule ($0.73\ e.a_0$) produces a potential difference of 0.4 V between opposing points on a sphere of radius 5 Å with the molecule at its centre. While the molecular dipoles inside the cluster are oriented so as to mostly cancel any long-ranged effect, those on the surface are not compensated by their neighbours, so a large cluster will retain a large net dipole moment. Indeed, figure 1(b) (black line), which shows the average dipole moment of clusters of water molecules calculated using a TIP3P point charge model for water [42], reveals that the dipole moment increases with system size as the created surfaces become larger in area. A similar trend in dipole moment is observed for QM calculations of single snapshots of water clusters of increasing radius (figure
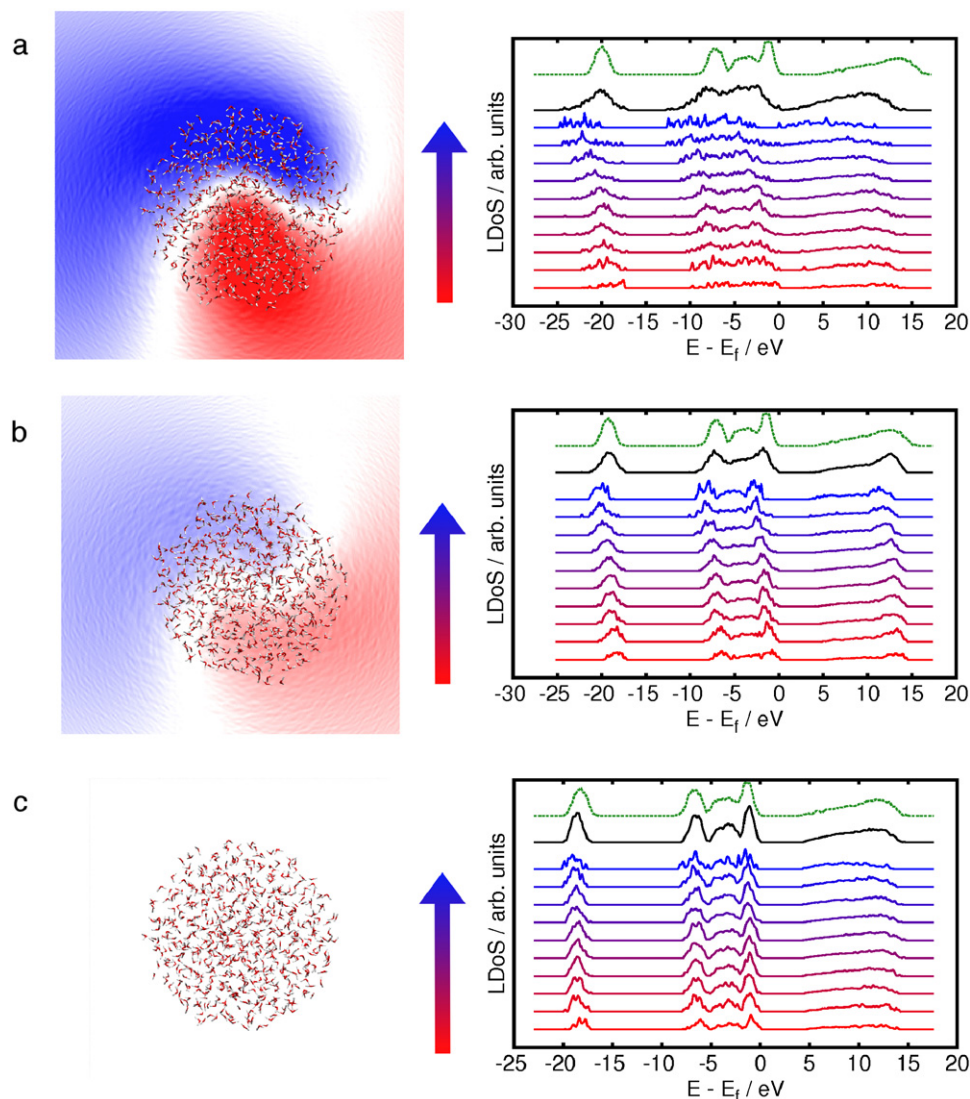
Figure 2. Electrostatic potential for a 16 Å radius water cluster and local density of electronic states for groups of atoms as a function of position along the dipole moment vector of the cluster. The dipole moment vector (coloured arrow) runs from the red line to blue. The black line is the total density of states and the green dashed line is the total density of states for bulk water. Each line in the LDoS plot is normalized by the number of molecules contained in the slab. The electrostatic potential ranges from −0.3 V (red) to +0.3 V (blue). The slice is 24.6 Å behind the water cluster. (a) Snapshot extracted from bulk water. The dipole moment is high, the LDoS is strongly dependent on position relative to the dipole moment vector, and the total range of states is much wider than for bulk water. (b) After classical minimization of the same snapshot. (c) Simulated using the ONETEP implicit solvent model. In both cases, the dipole moment is reduced and the DoS closely resembles that of the bulk.

S1 available at stacks.iop.org/JPhysCM/25/152101/mmedia) and figure 2(a) plots the DFT electrostatic potential on a plane behind the 16 Å water cluster, revealing a clear dipolar potential.

We have demonstrated that water clusters extracted from the equilibrated bulk display large multipole moments, as measured both by MM and QM. What effect does this have on the computed HOMO–LUMO gap? Figure 2(a) also shows the local density of electronic energy states (LDoS) for a 16 Å water cluster. To determine this quantity, clusters are nominally divided into 10 slabs perpendicular to the dipole moment, defined as the z-direction, and the slab LDoS is defined as the sum of the contributions to the total DoS from the local orbitals centred on the atoms in each slab.

We observe a clear shift in the LDoS as a function of the position along the dipole moment vector, with the electric field pushing some states higher in energy and some lower. Analogous to the concept of Fermi-level pinning in polar semiconductor nanorods, where the Fermi energy coincides with a finite density of states at either end of the rods [43, 44], the Fermi energy coincides with a non-zero density of states on opposite surfaces of the water cluster. We expect the HOMO–LUMO gap to disappear completely when the radius of the water cluster increases to the point where the variation of the surface potential is of sufficient magnitude to bridge the gap.

Given the apparent electrostatic origin of the vanishing gap, we hypothesize that there is no fundamental problem in

Table 1. HOMO–LUMO gaps for a range of proteins from the PDB. Atom number in parentheses includes a 5 Å solvation shell of water used in classical minimization and QM/EE simulations. Systems that did not converge are indicated by n/a. Vacuum calculations and implicit solvent simulations did not include any explicit water molecules.

PDB ID	Atoms	Charge	HOMO–LUMO Gap (eV)			
			In vacuo	QM water	Implicit solvent	QM/EE
1PLW	75(456)	0	0.0	3.7	3.7	3.5
1FUL	135(453)	−1	n/a	2.7	2.6	2.6
1RVS	172(670)	0	n/a	3.4	3.7	2.9
1EDW	399(978)	−1	n/a	3.1	3.7	2.6
1FDF	419(1526)	3	n/a	1.9	3.3	1.6
1UBQ	1231(2386)	0	n/a	2.6	3.4	2.4

the use of pure functionals in simulations of large systems, but simply that the issue manifests itself at smaller system sizes than it would for hybrid functionals which have an inherently larger gap. We would, therefore, expect any method that corrects these surface effects to also restore the HOMO–LUMO gap, which is consistent with observations made in the literature. It has been shown previously that the HOMO–LUMO gap may be restored by embedding classical point charges outside the electron distribution to represent, for example, the aqueous environment of the water or protein cluster [19]. In these cases, no significant changes occurred to the electronic density of the inner water molecules and the only significant changes in electronic density were observed on water molecules close to the surface [27]. Further, findings that a dielectric medium with a relative permittivity of 4 leads to robust SCF convergence of proteins in vacuum [23] imply that screening of the surface dipole is sufficient to restore the HOMO–LUMO gap. In the following, we test a number of methods for setting up a QM cluster in such a way that DFT calculations employing either pure or hybrid functionals may be readily applied.

The electric field across the cluster will be reduced if the atom positions are allowed to relax, either by geometry optimization or an annealing procedure. Figure 1(b) (blue line) reveals that, following classical minimization, via fast conjugate gradient and Newton–Raphson optimization (supplementary methods available at stacks.iop.org/JPhysCM/25/152101/mmedia), the average dipole moment of the extracted water clusters is substantially reduced (as measured by classical point charges). Furthermore, the HOMO–LUMO gaps of clusters that have undergone MM minimization are all restored to values close to the bulk water value of 4.2 eV (figure 1(a) (blue line)). In addition, an implicit solvent model is expected to reduce the shift in electronic states on opposite surfaces by screening the electrostatic potential across the cluster. Figure 1(a) (red line) shows that when the extracted water cluster is simulated with implicit solvent in ONETEP the HOMO–LUMO gap is again restored to the bulk value. Figure 2(b), (c) confirm that the dipole moment of the 16 Å water cluster is reduced following MM minimization and is negligible in the dielectric medium. In both cases, the density of electronic states is less dependent on the *z*-coordinate and closely resembles the bulk DoS (green line), as expected for a large cluster.

4. Protein systems

Turning our attention to polypeptide systems, six protein conformations (1PLW [45], 1FUL [46], 1RVS [47], 1EDW [48], 1FDF [49] and 1UBQ [50]) from the Brookhaven National Laboratory Protein Data Bank (PDB) [51] were used as starting configurations for our calculations. Table 1 shows the computed HOMO–LUMO gaps for these structures *in vacuo*, using the PBE functional. In agreement with [19] and as expected for this method of system preparation, the calculations did not produce HOMO–LUMO gaps (and hence did not converge) for any of the proteins apart from the smallest system considered. Figure 3 reveals that the problem is similar in nature to that of the water cluster. Plotting the electrostatic potential far from the ubiquitin (1UBQ) protein reveals a strong dipole moment. The LDoS reveals the dipole moment has a significant effect on the valence states, shifting them to where the HOMO–LUMO gap should lie.

In order to recover the expected HOMO–LUMO gaps, we use similar techniques to those described in the previous section (supplementary methods available at stacks.iop.org/JPhysCM/25/152101/mmedia). Classical minimization was performed on the 1FDF structure *in vacuo*, but again the electronic structure calculation failed to converge. Classical optimization, though able to restore the HOMO–LUMO gap in water clusters, is unsuitable for protein systems. This may be understood from considering the reduced opportunity for mobility in proteins, compared to water, as residues are fixed in secondary structure conformations. In addition, proteins may contain residues that are charged at physiological pH, yet when these are simulated *in vacuo* the charges are unscreened, and thus have a stronger and longer-ranged effect on the electrostatics than they do when solvated. Such residues will contribute significantly to the molecular dipole moment and may cause an undesired shift in the energies of the surface electronic states. We hypothesize that a more suitable strategy is to include the effects of the environment through the use of explicit water molecules or implicit solvent, in order to screen the effect of charged residues.

To this end, each protein structure was solvated by a classically minimized 5 Å layer of water and simulated in ONETEP using full DFT for the entire system (up to 2386 atoms) to re-calculate the electronic structure. In addition, implicit solvent calculations have been used for the protein structures in their vacuum configurations. Using
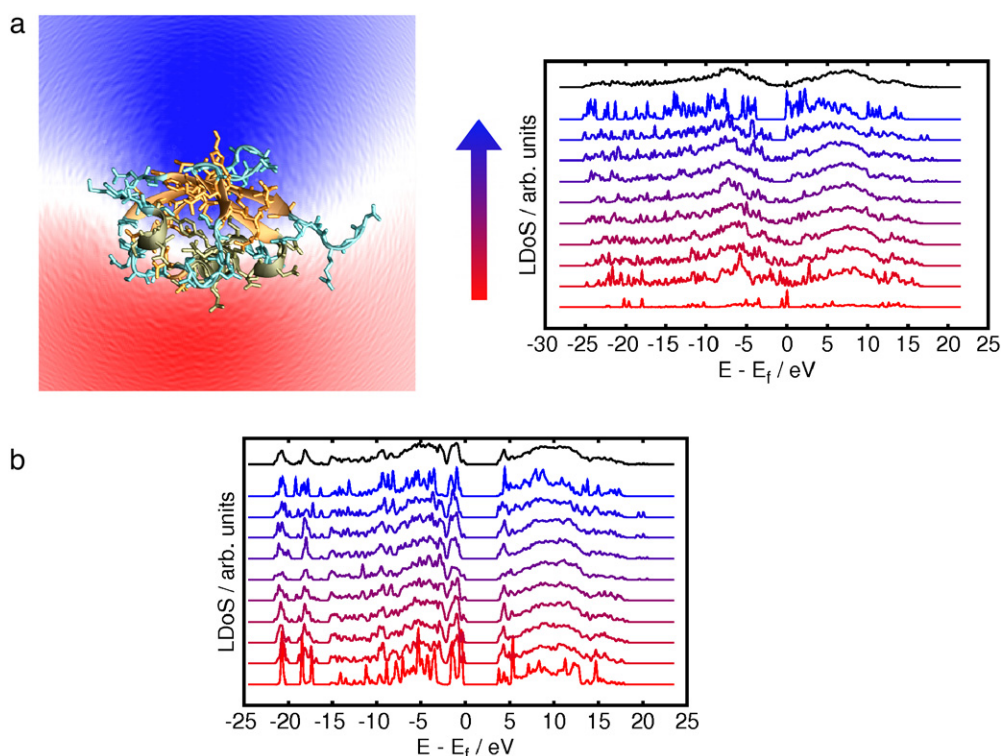
Figure 3. Electrostatic potential and LDoS for groups of atoms as a function of position along the dipole moment vector (coloured arrow) of ubiquitin. The black line is the total density of states. Each line in the LDoS plot is normalized by the number of molecules contained in the slab. (a) The experimental structure with no solvation. The electrostatic potential ranges from −0.2 V (red) to +0.2 V (blue). The slice is 42.9 Å behind the water cluster. (b) LDoS along the dipole moment of the same structure simulated with implicit solvent.

either an explicit or implicit solvation strategy restores the HOMO–LUMO gaps to similar values. Figure 3(b) shows that the DoS for the implicitly solvated system more closely resembles that of an insulator. The associated variation in the electrostatic potential is negligible, as in the case of the water cluster from figure 2(c) above, and is not shown. To explore the possibility of reducing the computational cost of the calculation, we have also represented the explicit water layer by embedded point charges with a TIP3P charge distribution. In this case, the HOMO–LUMO gap is very similar to that of the full QM water layer, implying that classical charges produce an electrostatic environment appropriate to reduce the net dipole of the system.

5. Conclusions

We have confirmed recent findings [19] that DFT electronic structure optimization is hindered by vanishing HOMO–LUMO gaps in large water and protein clusters— systems that should display insulating behaviour. This issue manifests itself in clusters prepared with improper treatment of the vacuum/system interface. In the examples presented here, unequilibrated vacuum/water interfaces and x-ray protein crystal structures exhibit strong dipole moments. Our LDoS calculations, decomposed by slabs along the direction of the dipole moment, reveal that electronic states are shifted by the electric field across the cluster. The Fermi energy is, hence, pinned by states on opposite surfaces, which

leads to closure of the HOMO–LUMO gap. The presence of a 4.2 eV band gap in a 2010-atom model of bulk water calculated with the PBE exchange–correlation functional, emphasizes that this effect has an electrostatic origin, rather than being peculiar to any particular class of functionals. Hybrid functionals that tend to have an intrinsically wider gap appear to be able to converge clusters comprising thousands of atoms [9, 23], but we would expect HOMO–LUMO gap closure, even for functionals containing Hartree–Fock exchange, once DFT methodological advances allow access to still larger systems.

In this communication, we have demonstrated practical solutions for restoring the HOMO–LUMO gap, with methods ranging from classical structural optimization of water/vacuum interfaces, to screening of molecular dipole moments via implicit solvation of protein structures. In general, implicit solvation gives the best correspondence between HOMO–LUMO gaps of large water clusters and the bulk band gap, and restores larger HOMO–LUMO gaps for proteins than does explicit solvation. In the present work we have looked at systems comprising up to 2386 atoms, and the practical solutions demonstrated here will allow the continued investigation of biomolecular systems through the use of Kohn–Sham DFT.

References

[1] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864
[2] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
[3] Goedecker S 1999 *Rev. Mod. Phys.* **71** 1085
[4] Bowler D R and Miyazaki T 2012 *Rep. Prog. Phys.* **75** 036503
[5] Hine N D M, Haynes P D, Mostofi A A, Skylaris C-K and Payne M C 2009 *Comput. Phys. Commun.* **180** 1041
[6] Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K and Scheffler M 2009 *Comput. Phys. Commun.* **180** 2175
[7] Tkatchenko A, Rossi M, Blum V, Ireta J and Scheffler M 2011 *Phys. Rev. Lett.* **106** 118102
[8] Ufimtsev I S and Martínez T J 2008 *Comput. Sci. Eng.* **10** 26
[9] Kulik H J, Luehr N, Ufimtsev I S and Martínez T J 2012 *J. Phys. Chem.* B **116** 12501
[10] de Pablo P J, Moreno-Herrero F, Colchero J, Gómez Herrero J, Herrero P, Baró A M, Ordejón P, Soler J M and Artacho E 2000 *Phys. Rev. Lett.* **85** 4992
[11] Bowler D R, Miyazaki T and Gillan M J 2002 *J. Phys.: Condens. Matter* **14** 2781
[12] Otsuka T, Miyazaki T, Ohno T, Bowler D R and Gillan M J 2008 *J. Phys.: Condens. Matter* **20** 294201
[13] Skylaris C-K, Haynes P D, Mostofi A A and Payne M C 2005 *J. Chem. Phys.* **122** 084119
[14] Cole D J, O'Regan D D and Payne M C 2012 *J. Phys. Chem. Lett.* **3** 1448
[15] Dziedzic J, Fox S J, Fox T, Tautermann C S and Skylaris C-K 2013 *Int. J. Quantum Chem.* **113** 771
[16] Ratcliff L E, Hine N D M and Haynes P D 2011 *Phys. Rev.* B **84** 165131
[17] Weber C, O'Regan D D, Hine N D M, Littlewood P B, Kotliar G and Payne M C 2013 *Phys. Rev. Lett.* **110** at press
[18] Lee L P, Cole D J, Payne M C and Skylaris C-K 2013 *J. Comput. Chem.* **34** 429
[19] Rudberg E 2012 *J. Phys.: Condens. Matter* **24** 072202
[20] Rubensson E H and Rudberg E 2011 *J. Comput. Chem.* **32** 1411
[21] Isborn C M, Luehr N, Ufimtsev I S and Martínez T J 2011 *J. Chem. Theory Comput.* **7** 1814
[22] Rudberg E, Rubensson E H and Sałek P 2011 *J. Chem. Theory Comput.* **7** 340
[23] Antony J and Grimme S 2012 *J. Comput. Chem.* **33** 1730
[24] Perdew J P and Zunger A 1981 *Phys. Rev.* B **23** 5048
[25] Godby R W, Schlüter M and Sham L J 1988 *Phys. Rev.* B **37** 10159
[26] Seidl A, Görling A, Vogl P, Majewski J A and Levy M 1996 *Phys. Rev.* B **53** 3764
[27] do Couto P C, Estácio S G and Cabral B J C 2005 *J. Chem. Phys.* **123** 054510
[28] Skylaris C-K, Haynes P D, Mostofi A A and Payne M C 2006 *Phys. Status Solidi* b **243** 973
[29] Cole D J, Skylaris C-K, Rajendra E, Venkitaraman A R and Payne M C 2010 *Europhys. Lett.* **91** 37004
[30] Cole D J, Rajendra E, Roberts-Thomson M, Hardwick B, McKenzie G J, Payne M C, Venkitaraman A R and Skylaris C-K 2011 *PLoS Comput. Biol.* **7** e1002096
[31] Skylaris C-K, Mostofi A A, Haynes P D, Diéguez O and Payne M C 2002 *Phys. Rev.* B **66** 035119
[32] Kohn W 1996 *Phys. Rev. Lett.* **76** 3168
[33] Prodan E and Kohn W 2005 *Proc. Natl Acad. Sci. USA* **102** 11635
[34] Mostofi A A, Haynes P D, Skylaris C-K and Payne M C 2003 *J. Chem. Phys.* **119** 8842
[35] Perdew J P, Burke K and Enzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
[36] Haynes P D, Skylaris C-K, Mostofi A A and Payne M C 2008 *J. Phys.: Condens. Matter* **20** 294207
[37] Fattebert J-L and Gygi F 2002 *J. Comput. Chem.* **23** 662
[38] Dziedzic J, Helal H H, Skylaris C-K, Mostofi A A and Payne M C 2011 *Europhys. Lett.* **95** 43001
[39] Fox S J, Pittock C, Fox T, Tautermann C S, Malcolm N and Skylaris C-K 2011 *J. Chem. Phys.* **135** 224107
[40] Case D A *et al* 2010 *AMBER 11* (San Francisco: University of California)
[41] Grimme S, Hujo W and Kirchner B 2012 *Phys. Chem. Chem. Phys.* **14** 4875
[42] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 *J. Chem. Phys.* **79** 926
[43] Avraam P W, Hine N D M, Tangney P and Haynes P D 2012 *Phys. Rev.* B **85** 115404
[44] Hine N D M, Avraam P W, Tangney P and Haynes P D 2012 *J. Phys.: Conf. Ser.* **367** 012002
[45] Marcotte I, Separovic F, Auger M and Gagne S M 2004 *Biophys. J.* **86** 1587
[46] Assa-Munt N, Jia X, Laakkonen P and Ruoslahti E 2001 *Biochemistry* **40** 2373
[47] Jaroniec C P, MacPhee C E, Bajaj V S, McMahon M T, Dobson C M and Griffin R G 2004 *Proc. Natl Acad. Sci. USA* **101** 711
[48] Yeagle P L, Salloum A, Chopra A, Bhawsar N, Ali L, Kuzmanovski G, Alderfer J L and Albert A D 2000 *J. Peptide Res.* **55** 455
[49] Yeagle P L, Danis C, Choi G, Alderfer J L and Albert A D 2000 *Mol. Vis.* **6** 125
[50] Vijay-Kumar S, Bugg C E and Cook W J 1987 *J. Mol. Biol.* **194** 531
[51] Bernstein F C, Koetzle T F, Williams G J, Meyer E E Jr, Brice M D, Rodgers J R, Kennard O, Shimanouchi T and Tasumi M 1977 *J. Mol. Biol.* **112** 535