Paper II

Analysis of computational models for an accurate study of electronic excitations in GFP

T. Schwabe, M. T. P. Beerepoot, J. M. H. Olsen and J. Kongsted

CrossMark
click for updates

Analysis of computational models for an accurate study of electronic excitations in GFP

Tobias Schwabe,*[a] Maarten T. P. Beerepoot,[b] Jógvan Magnus Haugaard Olsen[cd] and Jacob Kongsted[d]

Using the chromophore of the green fluorescent protein (GFP), the performance of a hybrid RI-CC2/ polarizable embedding (PE) model is tested against a quantum chemical cluster approach. Moreover, the effect of the rest of the protein environment is studied by systematically increasing the size of the cluster and analyzing the convergence of the excitation energies. It is found that the influence of the environment of the chromophore can accurately be described using a polarizable embedding model with only a minor error compared to a full quantum chemical description. It is also shown that the treatment of only a small region around the chromophore is only by coincidence a good approximation. Therefore, such cluster approaches should be used with care. Based on our results, we suggest that polarizable embedding models, including a large part of the environment to describe its effect on biochromophores on top of an accurate way of describing the central subsystem, are both accurate and computationally favourable in many cases.

1 Introduction

The description of photochemical and photophysical processes in biomolecules has recently seen an increased interest because of promising new applications in (multiphoton) spectroscopy, optogenetics, and therapeutic tools.[1–4] For a thorough understanding of these processes, quantum chemistry is a valuable tool. Indeed, the challenging nature of these processes requires the development and assessment of computational methods to complement experimental research. A good overview of the recent progress that has been achieved in this field can be obtained from some recent reviews.[5–9]

It would surely be desirable to describe the whole system of interest—typically a relatively small chromophore and its protein surroundings in a physiological solution—using quantum chemistry at a sufficient level of theory. However—given the huge number of particles in such a system—this is hardly feasible even for the most advanced implementations of quantum chemistry methods and the best high-performance computing systems.

[a] *Center for Bioinformatics and Physical Chemistry Institute, Bundesstraße 43, D-22148 Hamburg, Germany. E-mail: schwabe@zbh.uni-hamburg.de; Fax: +49 40 42838 7352; Tel: +49 40 42838 7333*
[b] *Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø – The Arctic University of Norway, N-9037 Tromsø, Norway*
[c] *Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*
[d] *Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense, Denmark*

Obviously, approximations have to be introduced. One standard approach[10,11] in this case is the division of the molecular system into a small part that is treated fully quantum mechanically (QM) and a larger remaining part that is described on the basis of classical molecular mechanics (MM). The treatment of MM is very efficient and therefore large systems can be studied with this hybrid QM/MM approach.

Although this scheme is quite a strong approximation, good results can be obtained with QM/MM. For the computation of electronic excitation properties, this can even be enhanced by applying a polarizable embedding (PE) method instead of the standard point charge model.[12] The PE method allows for a direct coupling between subsystems through the computation of induced dipoles in the MM region depending on the charge distribution in both the QM and MM region. The PE method is related to the similar effective fragment potential method.[13–15] This coupling improves the description of the environmental effect on electronic excitations.[16] The PE method has been combined with various standard quantum chemistry methods. Among those, the PE-DFT[17] and the PE-CC[18] variants have already been applied successfully to study electronic excitations in biomolecular systems.[19–22] Even though the computational effort is slightly larger than for standard QM/MM approaches, PE methods can be applied to complete protein systems taking even the aqueous solution into account.

Another strategy to study protein systems on a quantum chemical level is the truncation of the protein surroundings to a small cluster around the region of interest. The basic assumption is that only the closest environment has a significant effect

on the property of interest. Smaller clusters can be treated fully with a quantum chemistry method. Normally, these systems are still too large to be treated with any other method than the most efficient ones, *i.e.*, density functional theory (DFT) is chosen in almost all cases if one does not want to resort to semi-empirical approaches.

DFT is not free of flaws and this complicates the study of electronic excitations in biomolecules.[23] Therefore, an alternative approach to compare against is highly beneficial. Recently, Sundholm and co-workers introduced the reduced-virtual-space approach for approximate second-order coupled cluster (RVS-CC2).[24] With RVS-CC2, relatively large (<150 atoms) QM systems can be addressed because of a reduced computational cost, which is gained by a slight loss of error control. RVS-CC2 has an efficiency that makes protein cluster model computations feasible with a satisfactory wavefunction treatment.

The RVS-CC2 approach has been applied to compute excitation energies in cluster models of the green fluorescent protein (GFP) in its neutral and anionic state (also called A and B state).[25] This system is known as a challenging test case to accurately predict the effect of the protein environment.[26–28] The problem could in part be related to the complex electronic structure that has been reported for the free neutral chromophore.[29–31] For their study, Kaila *et al.*[25] resorted to a cluster model approach, for which the 13 nearest residues and the chromophore *p*-hydroxybenzylideneimidazolinone (pHBDI) and the amino acid remainders from which it is formed were chosen. Good agreement was obtained for the excitation energy shift between the two states as well as for the absolute excitation energies.

Despite the impressive cluster size for a CC2 computation, most of the protein surrounding is ignored in the above described study. This raises the question whether it is really safe to neglect the contributions from the major part of the system. To answer this question, one cannot apply RVS-CC2 because it is impossible to further increase the system size significantly. Additional computations in the same work with the remaining protein modelled with point charges already showed a non-negligible effect of the remaining environment.[25] It was suggested that an improved description of the protein like the PE method would be better suited to check to which extent the full protein model is needed.[25] Also Filippi *et al.* identified the description of polarization effects in the MM region as one of the critical points (besides the uncertainties of protonation states in the protein) in the accurate computation of excitation energies in GFP.[26] This is supported by recent additional work on GFP which has shown that amino acid residues up to a distance of 20 Å from the chromophore can influence its excitation energy through polarization interactions of the many (partial) charges in the protein.[32]

Recently, an efficient implementation of the PE method for RI-CC2 has been introduced, allowing CC2 level treatment with polarizable embedding even for relatively large QM system such as the GFP chromophore.[33] This allows to evaluate the effect of the complete protein in comparison to a truncated cluster model. The aim of this work is thus to evaluate the performance

of PERI-CC2 compared to RVS-CC2 results for smaller GFP cluster models and demonstrate its use in accurately describing the chromophore surroundings. In addition, PERI-CC2 will be applied to truncation models of systematically increased size up to the full protein size to study the effect of more and more distant residues on the experimentally relevant electronic excitation in the GFP chromophore.

2 Computational details

2.1 Protein model preparation

Cluster complexes from the Supplementary Information of ref. 25 are used to compare PERI-CC2 to RVS-CC2 results. They are shown in Fig. 1.

Beside the full model with the chromophore and 13 residues/crystal water molecules, smaller cluster models have also been prepared from these clusters. These smaller clusters consist of 2, 4, 7, and 10 residues/crystal water molecules in the surroundings. The actual fragments included for the two series, the neutral and anionic states, are slightly different, see Table 1 for details.
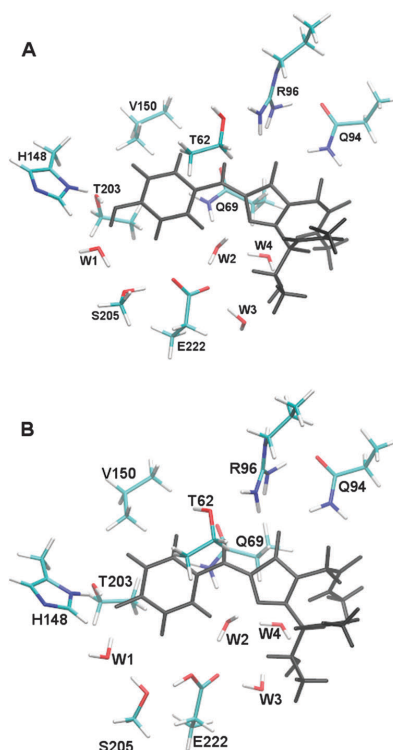

Fig. 1 The neutral (A) and anionic (B) GFP cluster complexes and residue labeling. Structures were obtained from ref. 25. The chromophore is indicated in black, otherwise the colour code is: carbon (green), nitrogen (blue), oxygen (red), hydrogen (white). The figure has been prepared with VMD.[34]

Table 1 Definition of the cluster model series for the neutral and anionic state based on the cluster models from ref. 25 (see Fig. 1 for labeling)

System	Additional fragments	
	Neutral	Anionic
Chromophore	—	—
+2 residues	T203, Q94	H148, Q94
+4 residues	H148, R96	T203, R96
+7 residues	W1, W2, W4	W1, W2, W4
+10 residues	S205, Q69, T62	S205, Q69, T62
Full cluster	V150, E222, W3	V150, E222, W3

In addition, a full protein model was used to investigate the influence of the rest of the protein. For this, chain A of the 1GFL crystal structure[35] including crystal water molecules was prepared with a neutral and an anionic chromophore as described in detail in ref. 32. Residue orientations and protonation states were predicted with the Schrödinger protein preparation wizard[36] except for Glu222, which is protonated for the anionic form of the chromophore and deprotonated for the neutral form as is well-accepted in the literature.[37] The orientation of Thr203 was adjusted for the anionic structure in accordance with experimental and theoretical evidence[37] such that its hydroxyl group points towards the anionic chromophore (thereby donating a hydrogen bond) and away from the neutral chromophore. We note that this structural effect was neither correctly taken into account in ref. 25, where the hydroxyl group points towards the chromophore in both the neutral and anionic forms, nor in ref. 19 and 20, where the hydroxyl group points away from the chromophore in both the neutral and anionic forms (as in the 1GFL and 1EMB crystal structures).

Another difference compared to ref. 25 is that the 1GFL crystal structure (used in this work) is from wild-type GFP whereas the 1EMB crystal structure (used in ref. 25) is from a GFP mutant and is composed of L-allo-threonine instead of serine at position 65 in the chromophore, which affects the hydrogen-bonding network around the chromophore.

The chromophore and parts of its surroundings (Gln94, Arg96, Gln148, Thr203, Ser205, Glu222 and the three closest water molecules) were quantum chemically geometry optimised separately for the two states within the electrostatic embedding of the rest of the frozen protein. This geometry optimization was done using the QSite program[38,39] with the B3LYP exchange–correlation functional[40–43] and the 6-31+G* basis set[44–46] in presence of point charge embedding based on the OPLS force field.[47] The chromophore that was cut out as the QM region for the PERI-CC2 calculations on the whole protein consisted of residues 65–67 including a carbonyl group at the N-terminus and a nitrogen at the C-terminus that come from residues 64 and 68, respectively, in agreement with earlier work.[19,20,32] Both sides were capped with an extra hydrogen atom. For comparison, the chromophore in the clusters from ref. 25 is capped with methyl instead of hydrogen, with the methyl carbon at the place of the C_z that was removed.

2.2 Potential generation for the polarizable embedding

The embedding potential parameters for the two structures, *i.e.* the cluster and full protein models, were calculated by dividing the full system into smaller fragments. For each fragment we then calculated atom-centred permanent multipole moments (up to quadrupoles) and anisotropic dipole–dipole polarizabilities. This was done using the LoProp method by Gagliardi *et al.*[48] as implemented in MOLCAS.[49] The level of theory employed was B3LYP using an ANO-type recontraction of the 6-31+G* basis set.

The cluster models do not include the backbone structure and each fragment thus consists of single amino acid side chains or water molecules. For the full protein model we used the molecular fractionation with conjugate caps (MFCC) scheme developed by Zhang and Zhang[50] to fragment the protein and derived the embedding potential parameters using the procedure reported by Söderhjelm and Ryde.[51]

The embedding potentials were created using a Python script that automates and parallelizes the generation of potentials used in polarizable embedding calculations and will be released in the near future.[52]

For comparison, electrostatic embedding potentials based on point charges from the Amber 94 force field[53] have also been generated. The charges on the capping hydrogens at the QM/MM interface have been chosen such that the total charge of the corresponding residues is zero.

2.3 Details for RVS- and PERI-CC2 computations

All RI-CC2[54,55] computations have been carried out with the RICC2 module of a local version of the TURBOMOLE program package[56] following a Hartree–Fock calculation with the direct SCF module.[57] The def2-SVP or the def2-TZVP basis sets[58] have been applied along with their auxiliary basis sets for the resolution-of-the-identity (RI) approximation.[59] For the RVS calculations, all virtual orbitals with energies higher than 60 eV have been discarded, following ref. 25. In addition, the RI approximation has also been applied. A critical discussion of the error introduced by reducing the virtual space is found in ref. 24.

If not otherwise stated for PERI-CC2 computations, the embedding potential was described by electric multipole moments up to quadrupoles and anisotropic polarizabilities. Polarizabilities and multipole moments at MM sites closer than 1.4 Å to any QM nucleus have been moved to the nearest MM neighbour site to avoid overpolarization.

3 Results

First, RVS-CC2 and PERI-CC2 computations are compared for two series of small cluster models of the neutral and anionic state of GFP. Here, the same systems are treated fully based on the wavefunction method (RVS-CC2) or based on a QM/MM scheme (PERI-CC2) with only the chromophore treated by RI-CC2. The results can be found in Table 2.

The methods show the same trends in both cluster series with actual excitation energies not differing by more than 0.08 eV, well within the estimated error of both methods. Compared to RVS-CC2, PERI-CC2 excitation energies are slightly higher in the neutral and slightly lower in the anionic case. For the full cluster, the deviation of the relative shift between the two states is 0.08 eV.

Table 2 Comparison of RVS-CC2/def2-TZVP and PERI-CC2/def2-TZVP for the GFP cluster models with increasing number of residues considered (for cluster definitions see Table 1). Excitation energies in eV

| System	Neutral		Anionic	
	RVS	PE	RVS	PE
Chromophore	3.44	3.44	2.77	2.74
+2 residues	3.46	3.48	2.76	2.71
+4 residues	3.35	3.37	2.76	2.70
+7 residues	3.21	3.25	2.72	2.64
+10 residues	3.17	3.21	2.70	2.63
Full cluster	3.13	3.17	2.72	2.68

Note that the RVS-CC2 result for the anionic state of the full cluster exceeds those reported in ref. 25 by 0.04 eV, so that the RVS-CC2 literature results are even in better agreement with PERI-CC2. In any case, the trends of the influence of the surrounding residues are in good agreement.

For further validation, the basis set size is reduced to allow a direct comparison to full RI-CC2 computations. RI-CC2 excitation energies have been computed for those systems that were feasible and results are listed in Table 3. The effect of the smaller basis set is a systematic blue shift for the neutral and the anionic state. The change is about 0.05 eV larger for the neutral case. Overall, this results in a larger relative shift between the neutral and anionic state excitation energies. Nevertheless, effects are very systematic and the choice of basis set should allow to estimate the quality of RVS-CC2 and PERI-CC2. Both approximations are within 0.1 eV of the full RI-CC2 results, yet the RVS-CC2 error is smaller and more systematic. PERI-CC2 results again fully recover all trends and the deviation to RVS-CC2 for the relative shift of the full cluster only increases by 0.01 eV. To highlight the agreement of all three methods, excitation energies for the neutral and anionic cluster models are plotted in Fig. 2.

In the next step, PERI-CC2 is applied to protein models where the included parts are systematically increased. As no geometrical data of the whole protein model from the study from Kaila *et al.* are available, this test has been carried out with another structure derived from a different protein crystal structure. As discussed in Section 2.1, this model also differs in the orientation of the Thr203 residue and in the residue at position 65 in the chromophore. While this restricts the direct

Table 3 Comparison of RI-CC2, RVS-CC2 and PERI-CC2 with the def2-SVP basis set for the GFP cluster models with increasing number of residues considered (for cluster definitions see Table 1). Excitation energies in eV

System	Neutral			Anionic		
	CC2	RVS	PE	CC2	RVS	PE
Chromophore	3.60	3.60	—	2.87	2.91	—
+2 residues	3.64	3.64	3.66	2.84	2.88	2.82
+4 residues	3.48	3.49	3.54	2.81	2.84	2.81
+7 residues	3.32	3.34	3.41	2.77	2.80	2.75
+10 residues	—	3.30	3.38	—	2.78	2.73
Full cluster	—	3.25	3.32	—	2.80	2.77
↳PE(M2P0)			3.42			2.92
↳PE(M0P0)			3.45			2.92
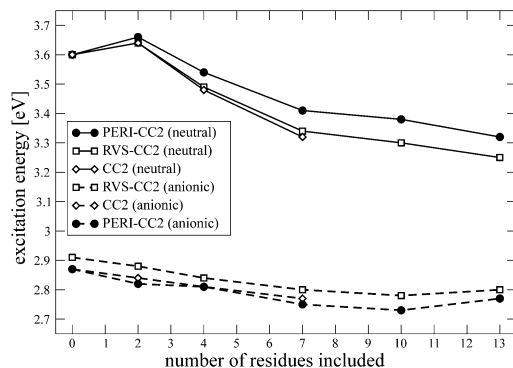
Fig. 2 Excitation energies for the neutral and anionic GFP cluster models with RI-CC2, RVS-CC2, and PERI-CC2 and the def2-SVP basis set.

comparison of the absolute results, general trends in the effect of going to a larger protein environment should be similar.

Excitation energies have been calculated for different sizes of the embedding region based on a cut-off radius R_{cut} to study the convergence of the excitation energy. An amino acid residue is completely included in the embedding model if one of its atoms falls within R_{cut}. The results are shown in Table 4.

The largest cluster model from the first part of this study consists of 113 MM sites and therefore falls between a cut-off radius of 1.5 and 2.0 Å in terms of its number of MM sites. Considering the included residues next to the chromophore, the 2.0 Å truncation model is in closer agreement to the cluster model. At that size, neither the excitation energy for the neutral and anionic states nor their relative shift is fully converged. Instead, one can find a minimum for the excitation energies for the truncated models with a cut-off radius of 6.0 Å (neutral state)

Table 4 PERI-CC2/def2-TZVP excitation energies (E_{exc}) and relative shift of the neutral and anionic state for increasing GFP truncation model size, characterised by their cut-off radius (R_{cut}) and the number of MM sites (# sites), up to the whole protein model. Distances in Å, energies in eV

R_{cut}	Neutral		Anionic		Shift
	# Sites	E_{exc}	# Sites	E_{exc}	
0.0	0	3.55	0	2.73	−0.82
1.5	79	3.40	79	2.65	−0.75
2.0	228	3.30	282	2.64	−0.66
2.5	427	3.30	420	2.61	−0.69
3.0	604	3.27	605	2.69	−0.58
4.0	826	3.26	827	2.65	−0.61
6.0	1207	3.25	1208	2.62	−0.63
8.0	1723	3.30	1724	2.68	−0.62
10.0	2140	3.34	2138	2.73	−0.61
12.0	2697	3.35	2695	2.75	−0.60
14.0	2983	3.35	3001	2.79	−0.56
16.0	3513	3.35	3514	2.77	−0.58
18.0	3683	3.35	3687	2.75	−0.60
20.0	3889	3.34	3887	2.72	−0.62
All	3991	3.33	3992	2.72	−0.61
↳PE(M2P0)		3.51		2.88	−0.63
↳PE(M0P0)		3.40		2.88	−0.52
↳AmberFF94		3.46		2.85	−0.61

and 2.5 Å (anionic state), while the relative shift has a minimum for a cut-off radius of 14.0 Å but does not level off even for larger cut-off radii. Thus, from the 2.0 Å cut-off radius truncated protein model to the full protein model, the relative shift drops only from −0.66 to −0.61 eV in the end. However, it varies considerably along the series of truncated protein models.

4 Discussion

The data for the GFP model in Table 4 clearly show that more distant residues also have an effect on the excitation energies of the chromophore and cannot generally be discarded. The effect from such distant MM sites is most likely related to the appearance of charged residues. The effect of surrounding residues converges faster if no charged residues are present because the (weaker) effect of an induced charge is damped away faster than that of a net charge.[32] Because some of these long-range interactions have an opposite effect and cancel each other to some extent in the present case, the full protein model yields results similar to the ones with the size of the largest cluster model from Kaila et al.[25] Thus, the analysis shows that the properties are affected by the more distant protein residues, and it is therefore a coincidence that such a good agreement with experiment can be reached with a minimal GFP cluster model. The cluster approach can therefore not be used safely in calculations on (other) biomolecular systems without knowing beforehand the influence of the part of the molecular system that is truncated.

Again, the models in the first and second part of the present study are not fully comparable. This is also clearly reflected in the different excitation energies of the chromophores without surrounding (compare Tables 2 and 4). The discrepancy is larger for the neutral state: 3.44 eV for the first vs. 3.55 eV for the second model, while for the anionic state the values are 2.74 eV and 2.73 eV, respectively. Nevertheless, the trend in going to larger cluster models should be very similar and fully reproducible with the present approach.

The final relative shift of the full protein model is −0.61 eV. From the comparison of RVS-CC2 and PERI-CC2 (compare Tables 2 and 3), it can be assumed that the relative shift is overestimated by around 0.1 eV which brings the shift down to −0.51 eV. This is in good agreement to the experimental relative shift of −0.42 to −0.53 eV.[60]

Comparing absolute excitation energies, the neutral state shows less agreement with experiment and therefore might be more problematic to describe quantitatively, which can also be deduced from its larger sensitivity to basis set quality and is in line with previous findings.[26–28] The challenging character can been seen from two of the most recent and advanced studies on the system, in which even two very similar setups and QM/MM treatments of the neutral state (XMCQDPT2/cc-pVDZ on a CASSCF(11,10)/cc-pVDZ reference wavefunction) yielded different excitation energies of 3.40 eV (ref. 27) or 3.19 eV,[28] respectively. A further extensive study on the problem is presented in ref. 26. The CASPT2/ANO-S-PVDZ results presented in that

paper (neutral state: 3.53 eV, anionic state: 2.82 eV) have been obtained with a non-polarizable point-charge MM potential only. To estimate the effect of an explicitly polarizable surrounding, the PERI-CC2 results were re-computed with only static multipole moments up to quadrupoles (PE(M2P0)) and with point charges only (PE(M0P0)), which means environmental polarization is completely neglected (see Table 4). Removing the explicit polarization increases the excitation energies in both cases. In the anionic case, the excitation energy is increased by about 0.16 eV for either level of the electrostatic embedding. This is different for the neutral state, where the PE(M2P0) result shows an increase by 0.18 eV but the PE(M0P0) result only by 0.07 eV.

The point charges used in the PE(M0P0) GFP model are only the first term in the series expansion of multipole moments and should not give rise to the same quality as the electrostatic potential that includes up to quadrupole moments. This is also different from the CASPT2/MM results, which where obtained with point charges from a standard protein force field. Such point charges have been fitted and therefore compensate, to some extent, the effects of missing higher-order multipole moments. So for comparing CASPT2 and CC2 results the PE(M2P0) should be more realistic. To demonstrate this, the Amber force field point charges (AmberFF94) have also been used as embedding potential. Indeed, a very good agreement between PE(M2P0) and AmberFF94 can be found. In any case, the results indicate that polarizable embedding would also change the results for the CASPT2 approach. Furthermore, the results in the first part of the present study, where PE(M2P0)RI-CC2 and PE(M2P2)RI-CC2 are compared against RVS-CC2 (see Table 3), show that a polarizable embedding approach is needed for a quantitative agreement to a full QM treatment—at least for the absolute excitation energies.

Of course, another possible way to compute accurate environmental effects is to keep a point charge embedding and to increase the QM region instead. For the GFP related photoactive yellow protein, it has been shown that this approach converges very slowly.[61] This accords with findings about long-ranged polarization effects in GFP.[32] Even in a more homogeneous, less charged environment such as an aqueous solution, it was found for acrolein that at least the complete first solvation shell has to be included into the QM region to compensate for missing polarization effects.[62] In any case, the PE method is the more efficient alternative.

Kaila et al. concluded that the inclusion of the crystallographic water into the QM region is more important for an accurate treatment than including residues into the QM region.[25] This is in line with a PE-DFT study on GFP where the crystallographic water was found to be important for a correct description of the anionic state.[19] From the results presented here, a similar conclusion is less apparent. Inspecting the smaller cluster models from the Kaila structures, one can find that the model with four residues includes no water molecules and in the next step the nearest three water molecules are included (see Table 1). The deviation between RVS-CC2 and PERI-CC2 (or between full RI-CC2 and PERI-CC2) is not

significantly increased for the latter. This shows that water is not too problematic for the polarizable embedding approach used here which has also been found for studies on solvatochromic effects in aqueous solutions.[62]

5 Conclusions

In this study, PERI-CC2 has been applied to minimal cluster models of the neutral and anionic states in GFP and assessed against RVS-CC2 and full RI-CC2 results. The three methods agree well among each other with deviations not exceeding 0.1 eV. In addition, PERI-CC2 has been used to investigate the effect of systematically increasing the number of protein residues included in the calculations showing that most of the environment cannot *per se* be neglected when computing electronic excitation energies in the GFP chromophore. The good agreement between results obtained on small cluster calculations and experimental excitation energies thus relies on coincidence. The results indicate that calculations on a truncated part of a biomolecular system should be preceded by a careful investigation of the effect of the rest of the environment.

This work has demonstrated once more how difficult it is to compute accurate molecular properties of biomolecular systems. Not only are the absolute excitation energies hard to reproduce but also their relative shift. Most likely, different factors contribute to the error in quantum chemical approaches: uncertainties in the molecular structure (especially in the chromophore region), necessary approximations to describe the protein environment, protonation states of amino acid residues that are not fully resolved, or the size of the QM region. While a systematic error control under such circumstances is truly challenging, it should be emphasized that the observed error is not significantly larger than what can be found for a full QM treatment. RI-CC2 is still only an approximate method of second order and it might be that even higher-order correlation effects play a role for electronic excitations in GFP.

Nevertheless, RI-CC2 is a valuable alternative to (TD-)DFT or even more approximate semi-empirical approaches and PERI-CC2 is a very efficient method to carry out studies like the present one to analyse trends from varying surroundings (even for very small clusters). Due to its efficiency, it should be preferred over RVS-CC2, which is still computationally demanding. Yet, the latter is more accurate and an acceptable approximation to full RI-CC2. In case that a larger QM region should be treated it is a viable option. It can also easily be combined with the PE approach such that there does not have to be a trade-off between a larger QM region and a better description of the classical MM region.

Our understanding of a best-practice protocol to compute molecular properties of biomolecular systems is increasing with every study on this topic and a lot has already been learned. New computational tools and a thorough understanding of the environmental effects that are essential for a reliable description of a biological system contribute towards routine application of such computations by non-expert users.

References

1 A. Specht, F. Bolze, Z. Omran, J.-F. Nicoud and M. Goeldner, *HFSP J.*, 2009, **3**, 255–264.
2 D. Warther, S. Gug, A. Specht, F. Bolze, J.-F. Nicoud, A. Mourot and M. Goeldner, *Bioorg. Med. Chem.*, 2010, **18**, 7753–7758.
3 G. C. R. Ellis-Davies, *ACS Chem. Neurosci.*, 2011, **2**, 185–197.
4 C. Brieke, F. Rohrbach, A. Gottschalk, G. Mayer and A. Heckel, *Angew. Chem.*, 2012, **124**, 8572–8604.
5 J. Neugebauer, *ChemPhysChem*, 2009, **10**, 3148–3173.
6 V. Tozzini, *Acc. Chem. Res.*, 2010, **43**, 220–230.
7 I. Schapiro, M. N. Ryazantsev, W. J. Ding, M. M. Huntress, F. Melaccio, T. Andruniow and M. Olivucci, *Aust. J. Chem.*, 2010, **63**, 413–429.
8 J. Hasegawa, K. J. Fujimoto and H. Nakatsuji, *ChemPhysChem*, 2011, **12**, 3106–3115.
9 K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin and A. I. Krylov, *Acc. Chem. Res.*, 2012, **45**, 265–275.
10 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.
11 H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.
12 J. M. H. Olsen and J. Kongsted, in *Advances in Quantum Chemistry*, ed. J. R. Sabin and E. Brändas, Academic Press, 2011, vol. 61, pp. 107–143.
13 L. V. Slipchenko, *J. Phys. Chem. A*, 2010, **114**, 8824–8830.
14 D. Kosenkov and L. V. Slipchenko, *J. Phys. Chem. A*, 2011, **115**, 392–401.
15 A. DeFusco, N. Minezawa, L. V. Slipchenko, F. Zahariev and M. S. Gordon, *J. Phys. Chem. Lett.*, 2011, **2**, 2184–2192.
16 K. Sneskov, T. Schwabe, O. Christiansen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2011, **13**, 18551–18560.
17 J. M. Olsen, K. Aidas and J. Kongsted, *J. Chem. Theory Comput.*, 2010, **6**, 3721–3734.
18 K. Sneskov, T. Schwabe, J. Kongsted and O. Christiansen, *J. Chem. Phys.*, 2011, **134**, 104108.
19 A. H. Steindal, J. M. H. Olsen, K. Ruud, L. Frediani and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2012, **14**, 5440–5451.
20 M. T. P. Beerepoot, A. H. Steindal, J. Kongsted, B. O. Brandsdal, L. Frediani, K. Ruud and J. M. H. Olsen, *Phys. Chem. Chem. Phys.*, 2013, **15**, 4735–4743.

21 K. Sneskov, J. M. H. Olsen, T. Schwabe, C. Hättig, O. Christiansen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2013, **15**, 7567–7576.

22 N. H. List, F. M. Pimenta, L. Holmegaard, R. L. Jensen, M. Etzerodt, T. Schwabe, J. Kongsted, P. R. Ogilby and O. Christiansen, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9950–9959.

23 A. J. Cohen, P. Mori-Sánchez and W. Yang, *Chem. Rev.*, 2012, **112**, 289–320.

24 R. Send, V. R. I. Kaila and D. Sundholm, *J. Chem. Phys.*, 2011, **134**, 214114.

25 V. R. I. Kaila, R. Send and D. Sundholm, *Phys. Chem. Chem. Phys.*, 2013, **15**, 4491–4495.

26 C. Filippi, F. Buda, L. Guidoni and A. Sinicropi, *J. Chem. Theory Comput.*, 2012, **8**, 112–124.

27 B. L. Grigorenko, A. V. Nemukhin, D. I. Morozov, I. V. Polyakov, K. B. Bravaya and A. I. Krylov, *J. Chem. Theory Comput.*, 2012, **8**, 1912–1920.

28 B. L. Grigorenko, A. V. Nemukhin, I. V. Polyakov, D. I. Morozov and A. I. Krylov, *J. Am. Chem. Soc.*, 2013, **135**, 11541–11549.

29 K. B. Bravaya, A. V. Bochenkova, A. A. Granovskii and A. V. Nemukhin, *Russ. J. Phys. Chem. B*, 2008, **2**, 671–675.

30 C. Filippi, M. Zaccheddu and F. Buda, *J. Chem. Theory Comput.*, 2009, **5**, 2074–2087.

31 I. V. Polyakov, B. L. Grigorenko, E. M. Epifanovsky, A. I. Krylov and A. V. Nemukhin, *J. Chem. Theory Comput.*, 2010, **6**, 2377–2387.

32 M. T. P. Beerepoot, A. H. Steindal, K. Ruud, J. M. H. Olsen and J. Kongsted, *Comput. Theor. Chem.*, 2014, **1040**, 304–311.

33 T. Schwabe, K. Sneskov, J. M. H. Olsen, J. Kongsted, O. Christiansen and C. Hättig, *J. Chem. Theory Comput.*, 2012, **8**, 3274–3283.

34 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

35 F. Yang, L. G. Moss and G. N. Philips Jr., *Nat. Biotechnol.*, 1996, **14**, 1246–1251.

36 *Schrödinger Suite 2013 Protein Preparation Wizard; Epik version 2.5*, Schrödinger, LLC, New York, NY, 2013, *Impact version 6.0*, Schrödinger, LLC, New York, NY, 2013, *Prime version 3.3*, Schrödinger, LLC, New York, NY, 2013.

37 M. Zimmer, *Chem. Rev.*, 2002, **102**, 759–781.

38 *QSite, version 5.8*, Schrödinger, LLC, New York, NY, 2012.

39 R. B. Murphy, D. M. Philipp and R. A. Friesner, *J. Comput. Chem.*, 2000, **21**, 1442–1457.

40 S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.

41 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.

42 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

43 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.

44 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.

45 P. C. Hariharan and J. A. Pople, *Theor. Chem. Acc.*, 1973, **28**, 213–222.

46 T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. Von Ragué Schleyer, *J. Comput. Chem.*, 1983, **4**, 294–301.

47 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

48 L. Gagliardi, R. Lindh and G. Karlström, *J. Chem. Phys.*, 2004, **121**, 4494–4500.

49 F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-A. Malmqvist, P. Neogády, T. B. Pedersen, M. Pitoňák, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov and R. Lindh, *J. Comput. Chem.*, 2010, **31**, 224–247.

50 D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2003, **119**, 3599–3605.

51 P. Söderhjelm and U. Ryde, *J. Phys. Chem. A*, 2009, **113**, 617–627.

52 J. M. H. Olsen, PhD thesis, University of Southern Denmark, Odense, Denmark, 2012.

53 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, J. K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 2013, **117**, 5179–5197.

54 O. Christiansen, H. Koch and P. Jørgensen, *Chem. Phys. Lett.*, 1995, **243**, 409–418.

55 C. Hättig and F. Weigend, *J. Chem. Phys.*, 2000, **113**, 5154–5161.

56 TURBOMOLE V6.4 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007, available from http://www.turbomole.com.

57 M. Häser and R. Ahlrichs, *J. Comput. Chem.*, 1989, **10**, 104–111.

58 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

59 F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, *Chem. Phys. Lett.*, 1998, **294**, 143–152.

60 T. M. H. Creemers, A. J. Lock, V. Subramaniam, T. M. Jovin and S. Völker, *Nat. Struct. Biol.*, 1999, **6**, 557–560.

61 C. M. Isborn, A. W. Götz, M. A. Clark, R. C. Walker and T. J. Martínez, *J. Chem. Theory Comput.*, 2012, **8**, 5092–5106.

62 T. Schwabe, J. M. H. Olsen, K. Sneskov, J. Kongsted and O. Christiansen, *J. Chem. Theory Comput.*, 2011, **7**, 2209–2217.

Correction: Analysis of computational models for an accurate study of electronic excitations in GFP

Tobias Schwabe,*[a] Maarten T. P. Beerepoot,[b] Jógvan Magnus Haugaard Olsen[cd] and Jacob Kongsted[d]

Correction for 'Analysis of computational models for an accurate study of electronic excitations in GFP' by Tobias Schwabe *et al.*, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2582–2588.

On page 2585, Table 4, the results in the last row are incorrect. The correct values are shown below:

Table 4 PERI-CC2/def2-TZVP excitation energies (E_{exc}) and relative shift of the neutral and anionic state for increasing GFP truncation model size, characterised by their cut-off radius (R_{cut}) and the number of MM sites (# sites), up to the whole protein model. Distances in Å, energies in eV

R_{cut}	Neutral		Anionic		Shift
	# Sites	E_{exc}	# Sites	E_{exc}	
(all/AmberFF94)	3991	3.48	3992	2.90	−0.58

These revised values do not affect any conclusions drawn in our paper. In fact, the absolute excitation energies of the corrected AMBER potential are even closer to the results obtained with the PE(M2P0) potential and underline the observation that neglect of polarization leads to blue-shifted results.

The Royal Society of Chemistry apologises for these errors and any consequent inconvenience to authors and readers.

[a] Center for Bioinformatics and Physical Chemistry Institute, Bundesstraße 43, D-22148 Hamburg, Germany. E-mail: schwabe@zbh.uni-hamburg.de; Fax: +49 40 42838 7352; Tel: +49 40 42838 7333

[b] Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø – The Arctic University of Norway, N-9037 Tromsø, Norway

[c] Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

[d] Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense, Denmark