All-Atom Quantum Mechanical Calculation of the Second-Harmonic Generation of Fluorescent Proteins

Beaujean, Pierre; Champagne, Benoît; Grimme, Stefan; de Wergifosse, Marc

Published in:
The Journal of Physical Chemistry Letters

DOI:
10.1021/acs.jpclett.1c02911

Publication date:
2021

Document Version
Peer reviewed version

All-Atom Quantum Mechanical Calculation of the Second-Harmonic Generation of Fluorescent Proteins

Pierre Beaujean,[†] Benoît Champagne,[†] Stefan Grimme,[‡] and Marc de Wergifosse[*,‡]

†*Laboratory of Theoretical Chemistry, Unit of Theoretical and Structural Physical Chemistry, Namur Institute of Structured Matter, University of Namur, Rue de Bruxelles 61, B-5000 Namur, Belgium*

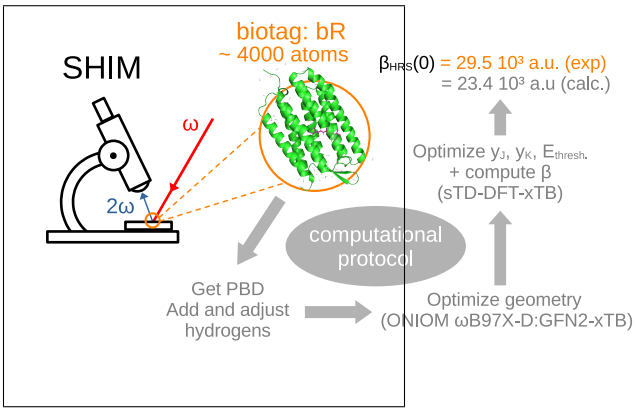
‡*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie der Universität Bonn, Berings. 4, D-53115 Bonn, Germany*

E-mail: mdewergifosse@gmail.com

Abstract

Fluorescent proteins (FPs) are biotags of choice for second-harmonic imaging microscopy (SHIM). Because of their large size, computing their second-harmonic generation (SHG) response represents a great challenge for quantum chemistry. In this contribution, we propose a new all-atom quantum mechanics methodology to compute SHG of large systems. This is now possible because of two recent implementations: the tight-binding GFN2-xTB method to optimize geometries and a related version of the simplified time-dependent density functional theory (sTD-DFT-xTB) to evaluate quadratic response functions. In addition, a new dual-threshold configuration selection scheme is introduced to reduce the comptational costs while retaining overall similar accuracy. This methodology was tested to evaluate the SHG of the proteins iLOV and bacteriorhodopsin (bR). In the case of bR, quantitative agreement with respect to experiment was reached for the out-of-resonance low-energy part of the β_{HRS} frequency dispersion. This work paves the way towards an accurate prediction of the SHG of large structures, a requirement for the design of new and improved SHIM biotags.

Graphical TOC Entry



Fluorescent proteins (FPs)[1–4] are used as genetically engineered biotags for the second-harmonic imaging microscopy (SHIM).[5–7] SHIM is a high-resolution bioimaging technique that provides contrast for non-centrosymmetric molecular arrangements. The phenomenon is the second-harmonic generation (SHG)[8,9] for which the response is governed by the first hyperpolarizability (β). While β is very sensitive to the local non-centrosymmetry and the polarization,[8,9] SHIM presents a low phototoxicity, less out-of-focus photobleaching, and higher penetration in biological tissues with respect to traditional fluorescence.[10–12] Reeve et al.[6] described requirements for good SHIM biotags, in particular a strong SHG response at the laser frequency (usually in the cell transparent region) and a high affinity for the hydrophobic cell membrane. The green fluorescent protein (GFP)-like family perfectly fits these requirements.[1–4]

Reports[13,14] showed that GFP was already employed as biotag for SHIM a couple of decades ago. Quantitative β values from hyper-Rayleigh scattering (HRS) experiments were obtained for a full rainbow of FPs.[1–4] These experimental findings were supported by quantum mechanics (QM) calculations. SHardonnay[3] was specifically engineered to remove eYFP local centrosymmetry and to enhance its SHG signal. HRS measurements were also reported for the bacteriorhodopsin[15] as well as other GFP-like proteins and several channel rhodopsins.[16]

All-atom QM calculations on such large biological systems are challenging. Most of the theoretical studies on photoreceptor proteins applies multi-scale modeling in which only a small fraction of the protein is treated at the QM level.[17–21] Specific difficulties on this subject were recently reviewed by Mroginski et al.[21] To evaluate β for FPs or other complex systems, previous attempts pinpointed the importance of considering the environment:[3] either implicitly (polarizable continuum[22,23] or charge embedding[24]) or explicitly (ONIOM schemes[3,4] or fragmentation methods[25]). The partitioning between different parts of the structure and their levels of approximation is also important choices to make.[21]

In this communication, we propose a new all-atom QM methodology to generally compute

β for FPs. This is now possible because of two recent developments and implementations: the GFN2-xTB method[26] to optimize geometries with the `xtb` program[27] and the evaluation of quadratic response functions[28] with the simplified time-dependent density functional theory (sTD-DFT)[28–30] and in particular its tight-binding version (sTD-DFT-xTB)[31] available in the freely available `stda` program.[32] The geometry of large proteins can now be optimized fully quantum mechanically[33] before computing their β values[28,34] with modest CPU requirements. This QM protocol retains most of the quality expected from higher levels of theory as it was demonstrated for the dynamical structural effects on β of tryptophan-rich amino peptides.[34] The reader interested in simplified quantum chemistry methods for evaluating response properties and excited states can consult our recent perspective articles[35,36] on the subject. Shortly, three approximations are introduced in the simplified scheme: a) two-electron integrals are approximated by damped short-range Coulomb interactions with two globally fitted y_J and y_K parameters, b) the configuration state function (CSF) space is truncated to cover a spectral range up to $E_{thresh.}$, and c) the response of the exchange-correlation potential is neglected. The game-changer strategy[28,34] is to fine-tune the y_J and y_K parameters to reproduce affordable high-level calculations for the chromophore only. This gives to the simplified calculations on FPs a similar accuracy at many orders of magnitude lower computational cost. $E_{thresh.}$ is adjusted to provide a sufficient but still tractable expansion space consisting of typically thousands of CSF.

Considering systems such as FPs, for which only few protein parts contribute significantly to the β response (mostly the chromophore), on top of the previous developments, we propose here a new dual-threshold method. Its motivation is to drastically reduce the configuration space and thereof the memory. In the dual-threshold method, the occupied (occ.) molecular orbitals (MOs) of the protein are partitioned into two layers. The high layer includes occ. MOs that are mostly located on the chromophore and selected important residues (with an electron density $\zeta_i > 0.1$). The remaining occ. MOs constitute the low layer. To determine the truncated space of CSFs, a tighter energy threshold E_{high} is employed for the high layer

Single threshold method

- Truncation of the MO space

$$\epsilon > \epsilon_{HOMO} - 2(1 + 0.8a_x)E_{thr.}$$
$$\epsilon < \epsilon_{LUMO} + 2(1 + 0.8a_x)E_{thr.}$$

- Selection of P-CSFs $i \rightarrow a$

$$A_{ia,ia} \leq E_{thr.}$$

- Selection of S-CSFs $j \rightarrow b$ for $A_{jb,jb} > E_{thr.}$

$$E_{jb}^{(2)} = \sum_{ia}^{P-CSFs} \frac{|A_{ia,jb}|^2}{A_{ia,ia} - A_{jb,jb}} > 10^{-4} E_h$$

#CSFs = #P-CSFs + #S-CSFs

Dual threshold method

- Truncation of the MO space

$$\epsilon > \epsilon_{HOMO} - 2(1 + 0.8a_x)E_{high}$$
$$\epsilon < \epsilon_{LUMO} + 2(1 + 0.8a_x)E_{high}$$

- Select occ. MOs for both threshold

$$\zeta_i = \sum_{\alpha \in high\ layer} C_{\alpha i}^2$$
$$\zeta_i > 0.1 \rightarrow E_{high}$$
$$\zeta_i \leq 0.1 \rightarrow E_{low}$$

1. For the high layer:

- Selection of P-CSFs-H $i \rightarrow a$

$$A_{ia,ia} \leq E_{high}$$

- Range of S-CSFs-H $j \rightarrow b$ for $A_{jb,jb} > E_{high}$

2. For the low layer:

- Selection of P-CSFs-low $i \rightarrow a$

$$A_{ia,ia} \leq E_{low}$$

- Range of S-CSFs-H $j \rightarrow b$ for $E_{low} < A_{jb,jb} < 2(1 + 0.8a_x)E_{low}$

3. For both layers:

- Selection of S-CSFs

$$E_{jb}^{(2)} = \sum_{ia}^{P-CSFs} \frac{|A_{ia,jb}|^2}{A_{ia,ia} - A_{jb,jb}} > 10^{-4} E_h$$

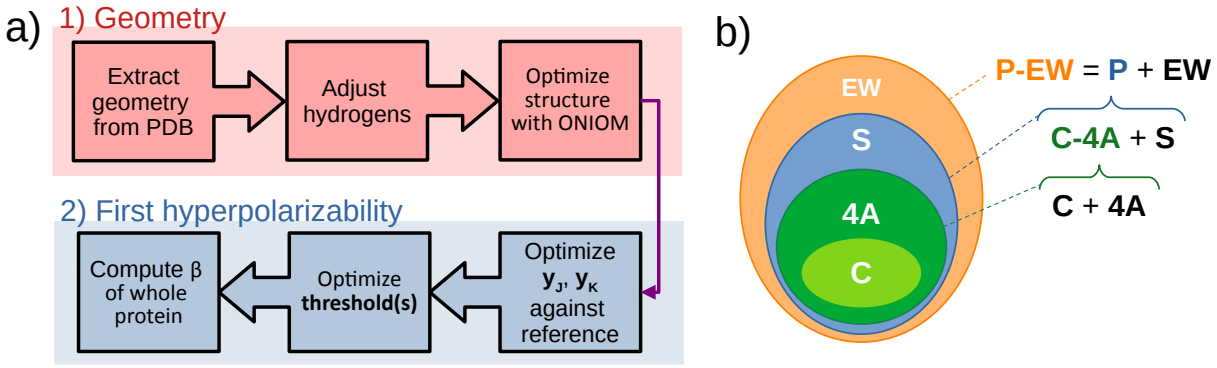
#CSFs = #P-CSFs-H + #P-CSFs-L + #S-CSFs

Scheme 1: Details of the sTD-DFT single and dual-threshold response methods. a_x is the Hartree-Fock exchange percentage, i and j refer to occupied molecular orbitals (MOs) while a and b to unoccupied ones, $C_{\alpha i}$ is the LCAO coefficient considering atomic orbital (AO) α and MO i, $A_{ia,jb}$ is an element of the linear response matrix A. P- and S-CSFs stand form primary and secondary configuration state functions, respectively.[35]

than for the low layer: $E_{low} < E_{high}$. Thus, with respect to a usual sTD-DFT calculation at a given $E_{thresh.}$, considering that $E_{low} = E_{thresh.}$, the dual-threshold method is increasing the active space but only for parts included in the high layer while the low layer stays unchanged. This allows to keep computational costs reasonable with respect to simply increasing $E_{thresh.}$. The molecular response property is then computed considering this extended set of CSFs. Scheme 1 summarizes details of the whole implementation.

Scheme 2a presents the all-atom QM methodology used to compute β of FPs. This procedure is divided in two parts: the protein geometry optimization and the evaluation of β. Starting geometries are usually obtained from the protein data bank (PDB).[37] Hydrogen atoms are added to the PDB geometry with the PlayMolecule web interface[38] at the experimental pH and manually for the chromophore to comply with its pK_a. The global charge of the system (Table S1) is determined according to the amino acid protonation states and inherent charges from other parts (chromophore, ions,...). Because β is highly sensitive to

structural details, we used an ONIOM[39] QM/QM scheme to optimize the protein geometry. This approach is similar to a QM/MM mechanical embedding but the use of the GFN2-xTB method for the low layer improves the treatment with respect to a MM method as it was demonstrated by Schmitz et al.[33] The chromophore (**C**, Scheme 2b) and the surrounding amino acids in a 4 Å radius (**4A**) are treated within the high layer at the ωB97X-D/6-31G* level (in gas phase). This was chosen to correctly account for the non-covalent interaction with the chromophore, while keeping a reasonable number of atoms within the high layer. We expect that this protocol should be applicable to other fluorescent proteins where one chromophore dominates the response but also more generally to large systems with a central NLOphore. The rest of the structure, including a few external water molecules, is optimized with the GFN2-xTB method[26] (in water, treated with the GBSA model[40]). These calculations were done with the Gaussian 16 A03 package[41] and the `xtb` 6.2.2 program.[27,42]



Scheme 2: Left: the all-atom QM methodology to compute the first hyperpolarizability of a fluorescent protein. Right: different parts of the protein and their acronyms: chromophore (**C**), 4-Å amino acids surrounding the chromophore (**4A**), rest of the surrounding amino acids and internal water molecules (**S**) and external water molecules (**EW**). Acronyms are also provided for combinations of parts.

The second part concerns the evaluation of the β_{HRS} and the depolarization ratio (DR) for the FP as defined by the mean and ratio of the β-tensor orientations,[9] respectively,

according to:

$$\beta_{HRS} = \sqrt{\langle\beta^2_{ZZZ}\rangle + \langle\beta^2_{ZXX}\rangle} \qquad \text{DR} = \frac{\langle\beta^2_{ZZZ}\rangle}{\langle\beta^2_{ZXX}\rangle}, \tag{1}$$

where in a HRS experiment, both incident and scattered photons are polarized, either parallel to the X (= horizontal) or to the Z (= vertical) axes for the incident photons and parallel to Z for the scattered photons. The y_J and y_K parameters in the sTD-DFT method are fine-tuned with respect to MP2/6-31+G* results to provide sTD-DFT-xTB values for the chromophore with a similar accuracy. The frequency dispersion is obtained by a multiplicative scheme[43] with either ωB97X-D or M06-2X exchange-correlation functionals using TD-DFT for the frequency dependence. Convergence of the β_{HRS} for the chromophore as a function of $E_{thresh.}$ is then assessed to select a sufficiently large number of CSFs. With these y_J, y_K and $E_{thresh.}$ parameters, β_{HRS} values are then computed for the optimized FP structure at the sTD-DFT-xTB level. Note that solvation effects are accounted for by the implict GBSA solvation model[40] but only for the generation of MOs. The reference values are obtained with the Gaussian 16 A03, while a development version of the `stda` program[32] is used for the sTD-DFT-xTB calculations.

To illustrate this new methodology, we selected two example FPs of increasing size: iLOV ($\sim$ 2000 atoms) and the bacteriorhodopsin (bR, $\sim$ 3850 atoms). Figure 1 displays their chromophore structures. iLOV is an engineered extrinsically fluorescent protein that binds the flavine mononucleotide (FMN).[44,45] bR is a light-driven transmembrane protein pump. Its retinal chromophore is covalently linked via a Schiff base to the protein backbone.[46] From the PDB, we used as input geometries 4EES for iLOV[47] and 6G7H for bR.[46] Both structures were protonated considering an experimental pH of 5. A full discussion about their optimizations is provided in the SI. Shortly, structural deviations (Figures S2-S3) for the optimized geometries with respect to X-ray data (0.52 and 0.32 Å for iLOV and bR, respectively) are within the experimental uncertainty of 0.5 Å. For iLOV, the FMN undergoes

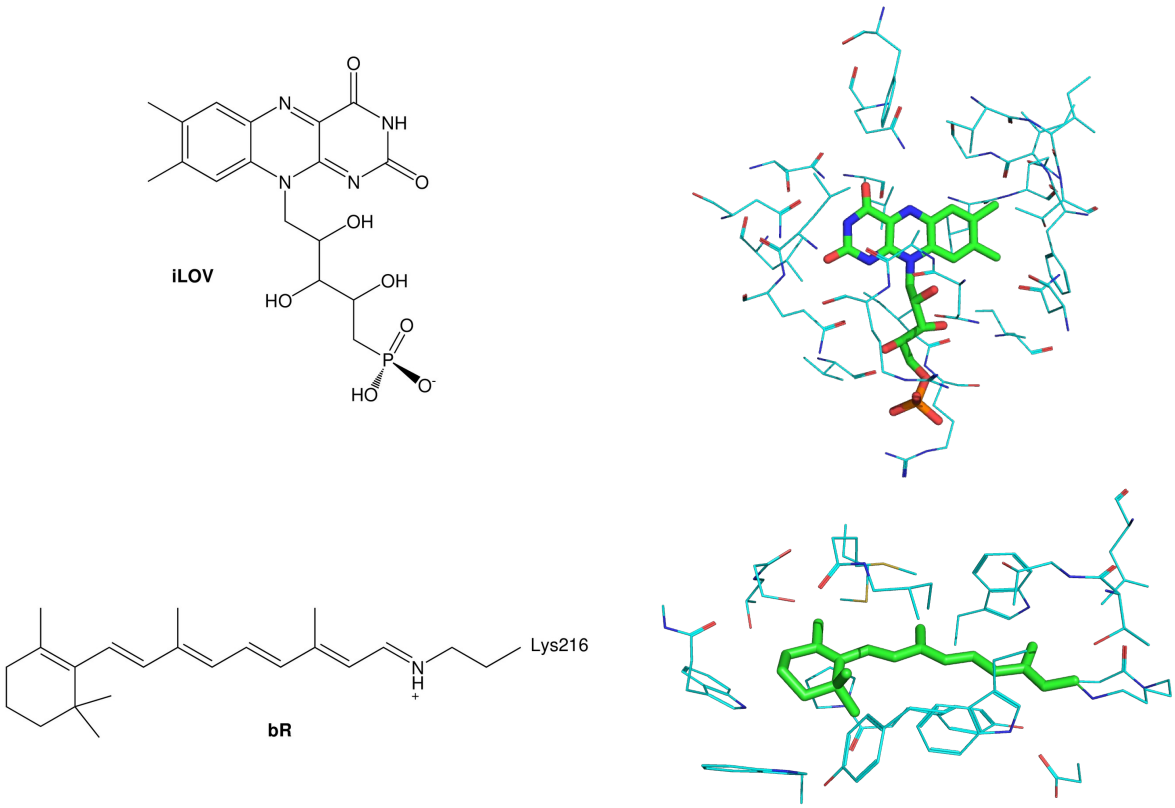


Figure 1: NLO-active chromophores (**C**) of the different proteins, in their experimental protonation state (left) and with their first shell of surrounding amino acids (right, chromophore in green, hydrogens hidden for clarity): iLOV (top, flavine mononucleotide) and bR (bottom, retinal schiff-base, in its native all-trans state conformation).

a displacement of its ribytil tail but without much modification of its π-conjugated pathway. The π-conjugation is also well preserved for the retinal schiff-base of bR (Figure S4).

At these geometries, we obtained sTD-DFT-xTB optimized parameters of $y_J = 2.0$ and $y_K = 0.15$ for both systems (Figure S5). Concerning the truncation of the CSF space (Figure S6), the β_{HRS} value of bR rapidly converges so that a $E_{thresh.}$ of 9 eV was selected. For iLOV, a larger value is required. To balance computational cost and accuracy, we selected $E_{thresh.} =$ 10 eV. Note that for bR with $E_{thresh.} = 9$ eV, 35 701 CSFs are included in the computation which took less than 74 hours on a AMD Epyc CPU with 64 cores (2.0 GHz). Moving now towards the dual-threshold scheme, the high layer contains only the chromophore. We use the following notation to refer to those calculation: E_{High}-E_{Low}. For example, the "9-7" calculation on bR used 3399 CSFs when $E_{High} = 9$ eV and $E_{Low} = 7$ eV. This calculation run only 5 hours (instead of 74).

Figure 2 presents the impact of the two thresholds on the static β_{HRS} value for both proteins as well as the number of included CSFs. For iLOV, the β_{HRS} value is gradually improved with the number of CSFs in comparison to the value obtained at $E_{thresh.} = 10$ eV (62 882 CSFs). The β_{HRS} value with $E_{thresh.} = 9$ eV is already converged within 10 % for a smaller configuration space (14 978 CSFs). With the dual-threshold method, including important CSFs for the chromophore with $E_{High} = 10$ eV but smaller E_{Low} drastically improves the efficiency of the treatment while maintaining its accuracy. For example, the β_{HRS} value at 10-7 is only 5 % lower than the value with a unique threshold of 10 eV while only accounting for 17 203 CSFs. Going from 10-8 to 10-9, a small increase of β_{HRS} is observed similar to the one from $E_{thresh.} = 8$ to 9 eV. The convergence with E_{Low} could even be smoother by including all tyrosine and tryptophan amino acids into the high layer (Figure S7), though at a slightly higher computational cost. For bR, the calculation with a threshold of 7 eV (1522 CSFs) already retains most of the physics with only 3 % difference with respect to the β_{HRS} value at a threshold of 9 eV (35 701 CSFs). Using the dual-threshold method, the convergence is even smoother.

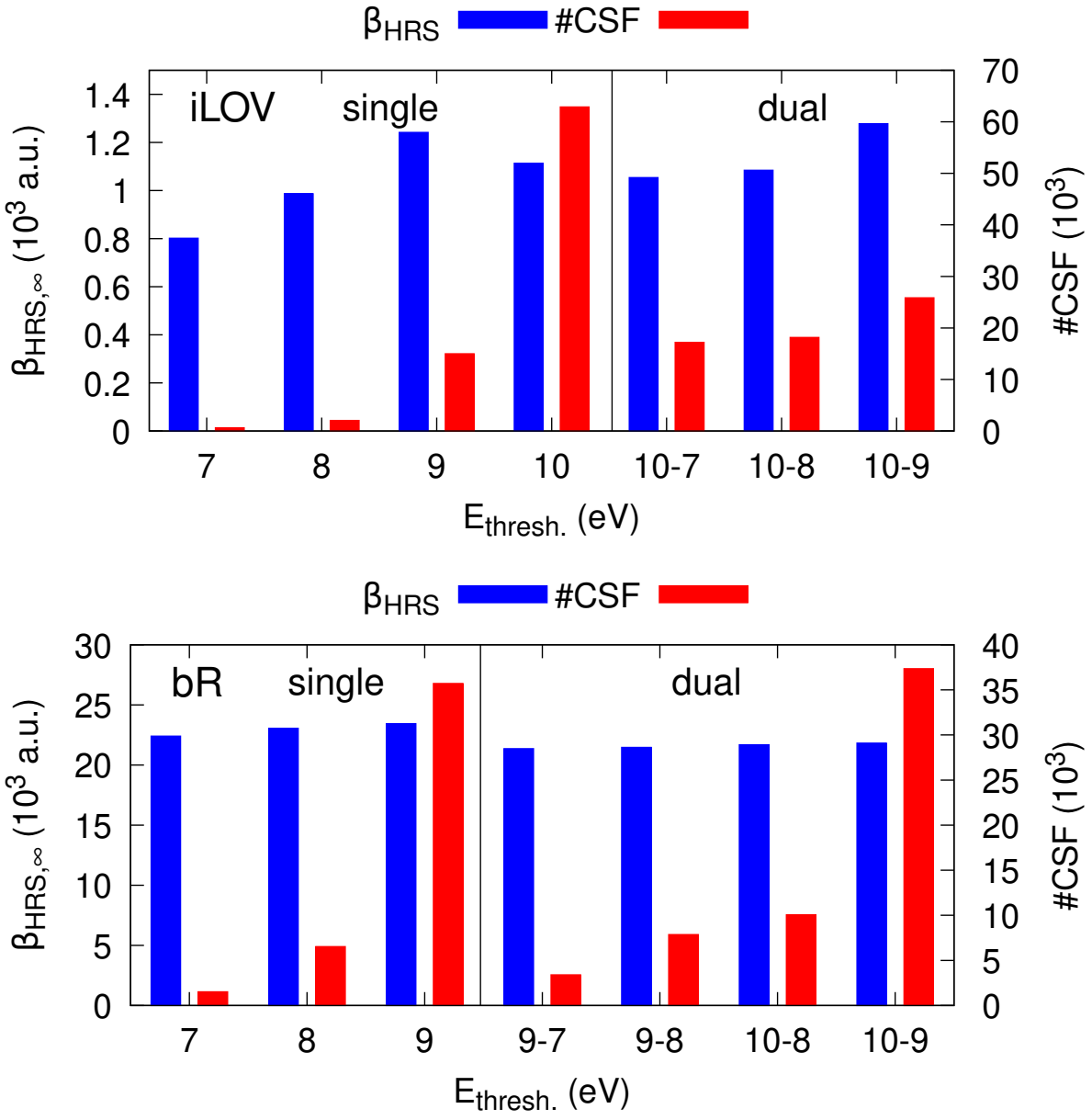


Figure 2: Influence of $E_{thresh.}$ on the static β_{HRS} of iLOV (top) and bR (bottom), as computed at the sTD-DFT-xTB level of theory (with $y_J = 2.0$ and $y_K = 0.15$) in water (GBSA), and corresponding numbers of CSFs. For the dual-threshold scheme, the first number indicates E_{High} and the second E_{Low}.

Table 1: Static β_{HRS} (in 10^3 a.u., DR in parentheses) of the chromophore (C), of its surroundings (4A, with extra hydrogens to saturate bonds) the C-4A region (ONIOM high layer, with extra hydrogens to saturate bonds), and of the whole protein (P-EW), as computed at the sTD-DFT-xTB level of theory (with $y_J = 2.0$, $y_K = 0.15$) in water (GBSA) with a threshold value of 10 eV for iLOV and 9 eV for bR. ONIOM MP2:sTD-DFT-xTB results are also provided.

	sTD-DFT-xTB				MP2:sTD-DFT-xTB	
	C	**4A**	**C-4A**	**P-EW**	**C-4A**	**P-EW**
iLOV	1.09 (4.8)	0.16 (2.3)	1.37 (6.2)	1.11 (3.6)	1.57 (5.7)	1.26 (3.6)
bR	17.32 (4.7)	0.66 (5.9)	21.08 (4.9)	23.43 (5.0)	22.04 (5.0)	24.40 (5.2)

To assess the impact of the chromophore surroundings on the response, Table 1 presents the β_{HRS} and DR for different parts of both structures. For iLOV, the β_{HRS} for the FMN in water is close to the one for the whole protein but the DR goes from 4.8 to a more octupolar value of 3.6 for the full protein. For bR, the β_{HRS} increases monotonically with the increasing size of the surroundings but its DR is almost unchanged. The sTD-DFT-xTB calculations are compared to ONIOM MP2:sTD-DFT-xTB results to assess their accuracy and demonstrate an excellent agreement (Table 1) for both structures. This confirms the suitability of the empirically fine-tuned y_J and y_K parameters to emulate higher-level QM methods.

Fig. 3 (Table S2) presents the β_{HRS} frequency dispersions for both FPs, which are mostly impacted by the first two-photon resonance. The β_{HRS} spectrum for bR was recorded by Clays and coworkers.[15] Usually, β_{HRS} is extrapolated to the static limit by different levels of refinement based on the two-state approximation.[48,49] We used a simple vibronic model (SVM), of which the key parameters were determined such that the experimental UV-visible spectrum is reproduced. Details about the SVM are given in the SI (Figure S8). Figure 4 compares the computed β_{HRS} spectrum to the experimental one as well as to SVM results. The sTD-DFT-xTB β_{HRS} frequency dispersion reproduces quantitatively the first three low-energy experimental points (those below the two-photon resonance) and follows well the extrapolation to the static limit by the SVM. Because of the divergent nature of our response theory in the resonance regime, it was expected that this frequency region could not be

reproduced. Nevertheless, for low energy values, quantitative agreement with experiment is striking showing the suitability of this methodology. We obtained a static β_{HRS} value of 23.4×10^3 a.u. close to the extrapolated experimental value of 29.5×10^3 a.u..

In conclusion, the proposed methodology enables computing the SHG of proteins (here with about 4000 atoms) fully quantum mechanically in a reasonable amount of computation time. The key concept is to refit only two empirical parameters in the sTD-DFT-xTB method to reproduce higher level β results for parts of the system (mostly the chromophore), providing a similar accuracy. In addition, a dual-threshold method is introduced to truncate specifically the single-excitation space for two different layers of the system, reducing the computational costs. We tested this approach for iLOV and the bacteriorhodopsin. For bR where experimental data are available, the agreement between sTD-DFT-xTB and experimental low-energy β_{HRS} frequency dispersion is excellent. This kind of comparison could not be achieved by only considering parts of the protein. This substantiates the importance to account for the whole protein (or at least large parts of it) into the calculation and the suitability of this workflow. In a near future, we should extend this methodology to the characterization of dynamical structural effects, e.g protein conformations as well as the impact of the truncation of the explicit solvation shell around the system.

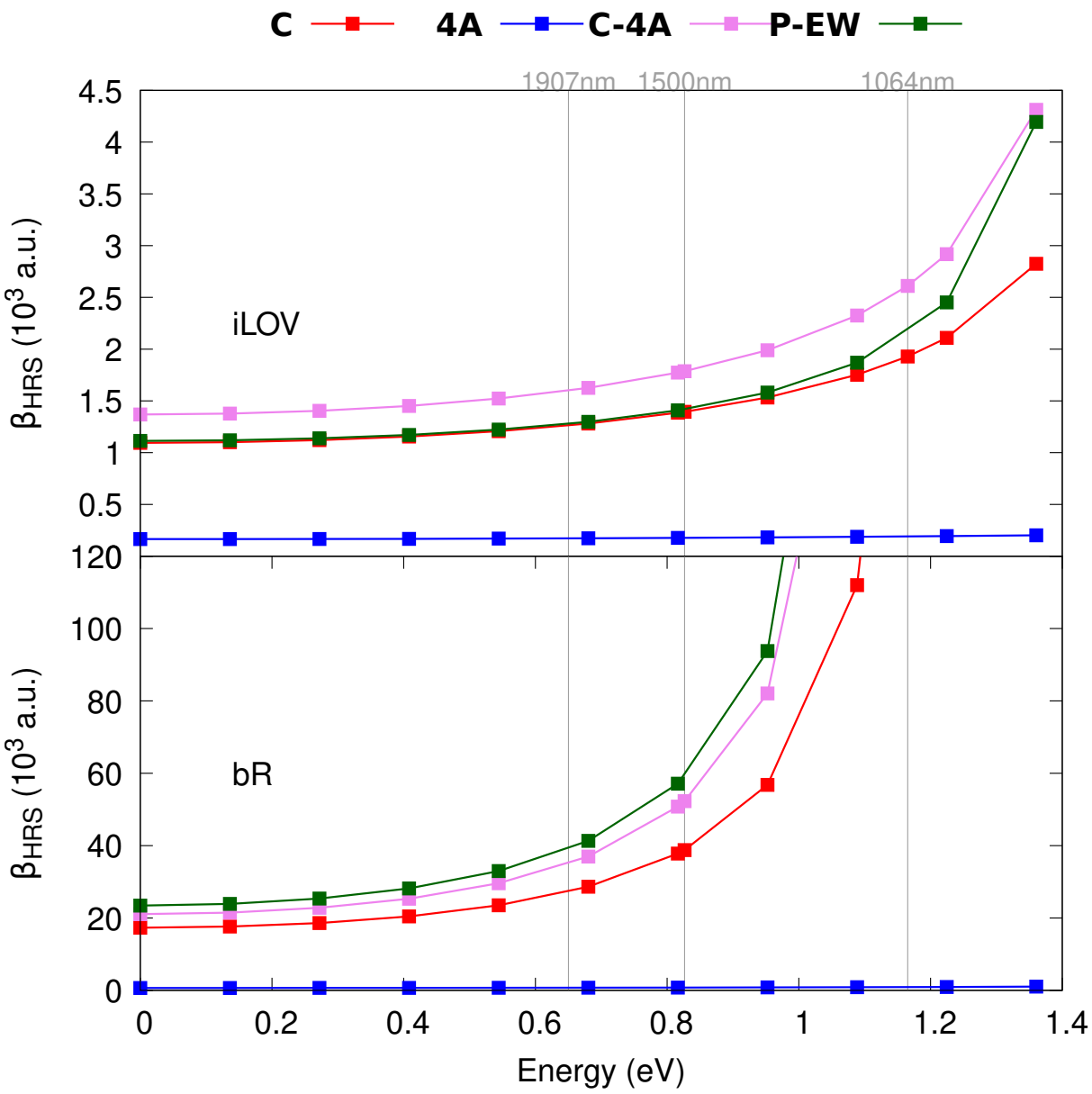


Figure 3: Comparison between the β_{HRS} of the chromophore (**C**), of its surrounding amino acids (**4A**), of the ONIOM "high" layer (**C-4A**), and of the total protein (**P-EW**), as computed at the sTD-DFT-xTB level (with $y_J = 2.0$, $y_K = 0.15$) in water (GBSA) with a threshold value of 10 eV for iLOV (top) and 9 eV for bR (bottom).

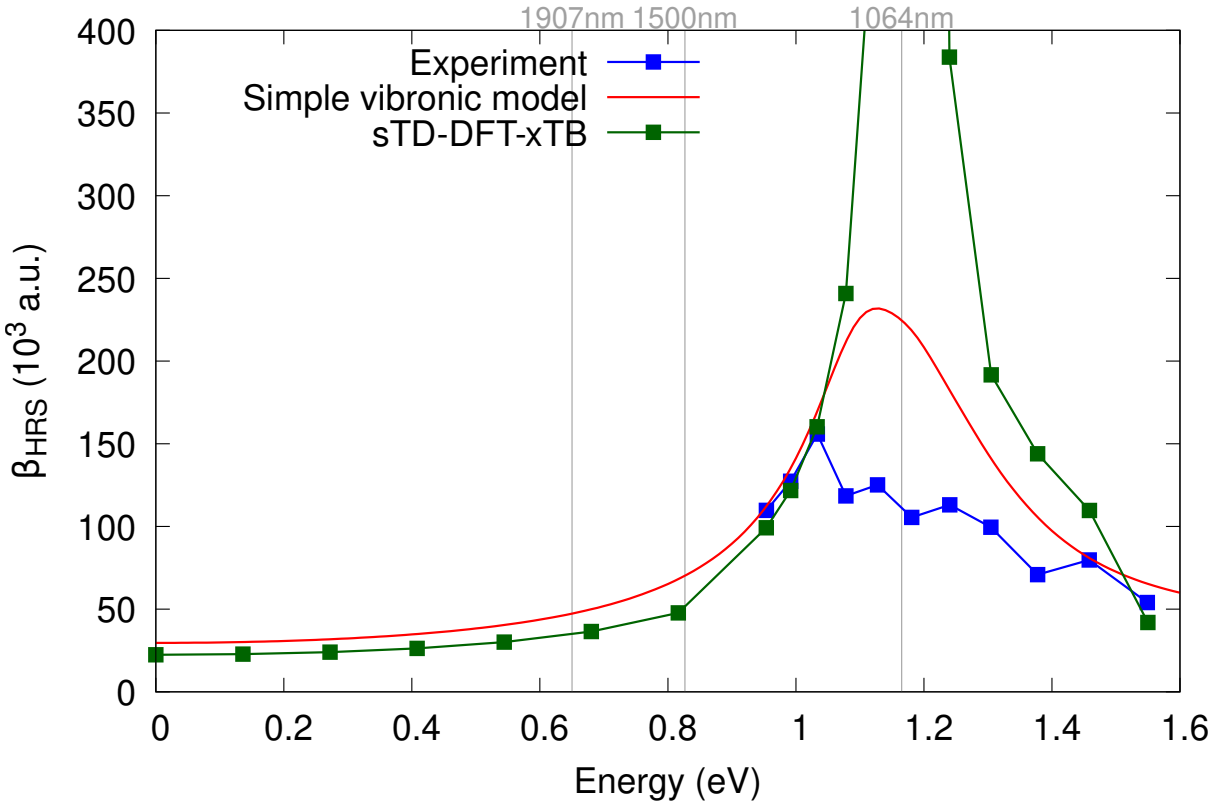


Figure 4: Experimental versus calculated β_{HRS} frequency dispersion of bR. The experimental one[15] has been extrapolated (red curve) to the static limit by using a vibronic model. The calculations were carried out at the sTD-DFT-xTB level ($y_J = 2.0$, $y_K = 0.15$, and $E_{thresh.} = 9\,\text{eV}$) in water (GBSA).

Acknowledgement

The authors thank Prof. K. Clays for fruitful discussions. This work was supported by the DFG in the framework of the project "Theoretical studies of nonlinear optical properties of fluorescent proteins by novel low-cost quantum chemistry methods" (Nr. 450959503). The calculations were performed on the computers of the Consortium des Équipements de Calcul Intensif, including those of the Technological Platform of High-Performance Computing, for which we gratefully acknowledge the financial support of the FNRS-FRFC, of the Walloon Region, and of the University of Namur (Conventions No. 2.5020.11, GEQ U.G006.15, 1610468, and RW/GEQ2016).

Supporting Information Available

See Supporting Information for: a) details on the geometry optimizations (RMSD and MAE of the structures, probability density of important bonds and BLA of the chromophores), including justification of the ONIOM scheme, b) optimization of the parameters for the sTD-DFT-xTB calculations (y_J, y_K and $E_{thresh.}$), c) numerical results of Figure 3, d) dual-threshold results for iLOV when including Tyr and Trp amino acids, and e) description of the SVM model and its application to bR.

References

(1) Asselberghs, I.; Flors, C.; Ferrighi, L.; Botek, E.; Champagne, B.; Mizuno, H.; Ando, R.; Miyawaki, A.; Hofkens, J.; der Auweraer, M. V. et al. Second-Harmonic Generation in GFP-like Proteins. *J. Am. Chem. Soc.* **2008**, *130*, 15713–15719.

(2) De Meulenaere, E.; Asselberghs, I.; de Wergifosse, M.; Botek, E.; Spaepen, S.; Champagne, B.; Vanderleyden, J.; Clays, K. Second-Order Nonlinear Optical Properties of Fluorescent Proteins for Second-Harmonic Imaging. *J. Mater. Chem.* **2009**, *19*, 7514.

(3) De Meulenaere, E.; Nguyen Bich, N.; de Wergifosse, M.; Van Hecke, K.; Van Meervelt, L.; Vanderleyden, J.; Champagne, B.; Clays, K. Improving the Second-Order Nonlinear Optical Response of Fluorescent Proteins: The Symmetry Argument. *J. Am. Chem. Soc.* **2013**, *135*, 4061–4069.

(4) de Wergifosse, M.; Botek, E.; De Meulenaere, E.; Clays, K.; Champagne, B. ONIOM Investigation of the Second-Order Nonlinear Optical Responses of Fluorescent Proteins. *J. Phys. Chem. B* **2018**, *122*, 4993–5005.

(5) Campagnola, P. J.; Loew, L. M. Second-Harmonic Imaging Microscopy for Visualizing Biomolecular Arrays in Cells, Tissues and Organisms. *Nat. Biotechnol.* **2003**, *21*, 1356–1360.

(6) Reeve, J. E.; Anderson, H. L.; Clays, K. Dyes for Biological Second Harmonic Generation Imaging. *Phys. Chem. Chem. Phys.* **2010**, *12*, 13484–13498.

(7) Campagnola, P. Second Harmonic Generation Imaging Microscopy: Applications to Diseases Diagnostics. *Anal. Chem.* **2011**, *83*, 3224–3231.

(8) Franken, P. A.; Hill, A. E.; Peters, C. W.; Weinreich, G. Generation of Optical Harmonics. *Phys. Rev. Lett.* **1961**, *7*, 118–119.

(9) Verbiest, T.; Clays, K.; Rodriguez, V. *Second-Order Nonlinear Optical Characterization Techniques: An Introduction*; Taylor & Francis, 2009.

(10) Sheppard, C. J. R.; Gu, M. Image Formation in Two-Photon Fluorescence Microscopy. *Optik* **1990**, *86*, 104–106.

(11) Denk, W.; Strickler, J. H.; Webb, W. W. Two-Photon Laser Scanning Fluorescence Microscopy. *Science* **1990**, *248*, 73–76.

(12) So, P. T. C.; Dong, C. Y.; Masters, B. R.; Berland, K. M. Two-Photon Excitation Fluorescence Microscopy. *Annu. Rev. Biomed. Eng.* **2000**, *2*, 399–429.

(13) Lewis, A.; Khatchatouriants, A.; Treinin, M.; Chen, Z.; Peleg, G.; Friedman, N.; Bouevitch, O.; Rothman, Z.; Loew, L.; Sheres, M. Second-Harmonic Generation of Biological Interfaces: Probing the Membrane Protein Bacteriorhodopsin and Imaging Membrane Potential around GFP Molecules at Specific Sites in Neuronal Cells of C. Elegans. *Chem. Phys.* **1999**, *245*, 133–144.

(14) Khatchatouriants, A.; Lewis, A.; Rothman, Z.; Loew, L.; Treinin, M. GFP Is a Selective Non-Linear Optical Sensor of Electrophysiological Processes in Caenorhabditis Elegans. *Biophys. J.* **2000**, *79*, 2345–2352.

(15) de Coene, Y.; Van Cleuvenbergen, S.; Van Steerteghem, N.; Baekelandt, V.; Verbiest, T.; Bartic, C.; Clays, K. Fluorescence-Free Spectral Dispersion of the Molecular First Hyperpolarizability of Bacteriorhodopsin. *J. Phys. Chem. C* **2017**, *121*, 6909–6915.

(16) De Meulenaere, E.; de Coene, Y.; Russier-Antoine, I.; Vanpraet, L.; Van den Haute, C.; Thevissen, K.; Baekelandt, V.; Bartic, C.; Hofkens, J.; Brevet, P.-F. et al. Fluorescence-Free First Hyperpolarizability Values of Fluorescent Proteins and Channel Rhodopsins. *J. Photochem. Photobiol. A* **2020**, *400*, 112658.

(17) List, N. H.; Olsen, J. M. H.; Jensen, H. J. A.; Steindal, A. H.; Kongsted, J. Molecular-Level Insight into the Spectral Tuning Mechanism of the DsRed Chromophore. *J. Phys. Chem. Lett.* **2012**, *3*, 3513–3521.

(18) Steindal, A. H.; Olsen, J. M. H.; Ruud, K.; Frediani, L.; Kongsted, J. A Combined Quantum Mechanics/Molecular Mechanics Study of the One- and Two-Photon Absorption in the Green Fluorescent Protein. *Phys. Chem. Chem. Phys.* **2012**, *14*, 5440–5451.

(19) Ding, L.; Chung, L. W.; Morokuma, K. Reaction Mechanism of Photoinduced Decarboxylation of the Photoactivatable Green Fluorescent Protein: An ONIOM(QM:MM) Study. *J. Phys. Chem. B* **2013**, *117*, 1075–1084.

(20) Nemukhin, A. V.; Grigorenko, B. L.; Khrenova, M. G.; Krylov, A. I. Computational Challenges in Modeling of Representative Bioimaging Proteins: GFP-Like Proteins, Flavoproteins, and Phytochromes. *J. Phys. Chem. B* **2019**, *123*, 6133–6149.

(21) Mroginski, M.-A.; Adam, S.; Amoyal, G. S.; Barnoy, A.; Bondar, A.-N.; Borin, V. A.; Church, J. R.; Domratcheva, T.; Ensing, B.; Fanelli, F. et al. Frontiers in Multiscale Modeling of Photoreceptor Proteins. *Photochem. Photobiol.* **2021**, *97*, 243–269.

(22) Ferrighi, L.; Frediani, L.; Cappelli, C.; Sałek, P.; Ågren, H.; Helgaker, T.; Ruud, K. Density-Functional-Theory Study of the Electric-Field-Induced Second Harmonic Generation (EFISHG) of Push–Pull Phenylpolyenes in Solution. *Chem. Phys. Lett.* **2006**, *425*, 267–272.

(23) Marrazzini, G.; Giovannini, T.; Egidi, F.; Cappelli, C. Calculation of Linear and Non-Linear Electric Response Properties of Systems in Aqueous Solution: A Polarizable Quantum/Classical Approach with Quantum Repulsion Effects. *J. Chem. Theory Comput.* **2020**, *16*, 6993–7004.

(24) Bouquiaux, C.; Tonnelé, C.; Castet, F.; Champagne, B. Second-Order Nonlinear Optical Properties of an Amphiphilic Dye Embedded in a Lipid Bilayer. A Combined Molecular Dynamics–Quantum Chemistry Study. *J. Phys. Chem. B* **2020**, *124*, 2101–2109.

(25) Zahariev, F.; Gordon, M. S. Nonlinear Response Time-Dependent Density Functional Theory Combined with the Effective Fragment Potential Method. *J. Chem. Phys.* **2014**, *140*, 18A523.

(26) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

(27) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *11*, e01493.

(28) de Wergifosse, M.; Grimme, S. Nonlinear-Response Properties in a Simplified Time-Dependent Density Functional Theory (sTD-DFT) Framework: Evaluation of the First Hyperpolarizability. *J. Chem. Phys.* **2018**, *149*, 024108.

(29) de Wergifosse, M.; Grimme, S. Nonlinear-Response Properties in a Simplified Time-Dependent Density Functional Theory (sTD-DFT) Framework: Evaluation of Excited-State Absorption Spectra. *J. Chem. Phys.* **2019**, *150*, 094112.

(30) de Wergifosse, M.; Grimme, S. A Unified Strategy for the Chemically Intuitive Interpretation of Molecular Optical Response Properties. *J. Chem. Theory Comput.* **2020**, *16*, 7709–7720.

(31) Grimme, S.; Bannwarth, C. Ultra-Fast Computation of Electronic Spectra for Large Systems by Tight-Binding Based Simplified Tamm-Dancoff Approximation (sTDA-xTB). *J. Chem. Phys.* **2016**, *145*, 054103.

(32) de Wergifosse, M.; Bannwarth, C.; Shushkov, P.; Grimme, S. stda program for computing excited states and response properties via the simplified TD-DFT methods, version 1.6.2. Code available at https://github.com/grimme-lab/stda.

(33) Schmitz, S.; Seibert, J.; Ostermeir, K.; Hansen, A.; Göller, A. H.; Grimme, S. Quantum Chemical Calculation of Molecular and Periodic Peptide and Protein Structures. *J. Phys. Chem. B* **2020**, *124*, 3636–3646.

(34) Seibert, J.; Champagne, B.; Grimme, S.; de Wergifosse, M. Dynamic Structural Effects on the Second-Harmonic Generation of Tryptophane-Rich Peptides and Gramicidin A. *J. Phys. Chem. B* **2020**, *124*, 2568–2578.

(35) de Wergifosse, M.; Grimme, S. Perspective on Simplified Quantum Chemistry Methods for Excited States and Response Properties. *J. Phys. Chem. A* **2021**, *125*, 3841–3851.

(36) Grimme, S.; Bohle, F.; Hansen, A.; Pracht, P.; Spicher, S.; Stahn, M. Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules. *J. Phys. Chem. A* **2021**, *125*, 4039–4054.

(37) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(38) Martínez-Rosell, G.; Giorgino, T.; De Fabritiis, G. PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2017**, *57*, 1511–1516.

(39) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field derivatives. *Comp. Theor. Chem.* **1999**, *461-462*, 1–21.

(40) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian 16 Revision A.03. 2016; Gaussian Inc. Wallingford CT.

(42) Ehlert, S.; Bannwarth, C.; Grimme, S. Semiempirical Extended Tight-Binding Program Package xtb, version 6.2.2. `https://github.com/grimme-lab/xtb/`, Accessed: 2019-12-19. See also xTB documentation at `https://xtb-docs.readthedocs.io/en/latest/`.

(43) Sekino, H.; Bartlett, R. J. Molecular Hyperpolarizabilities. *J. Chem. Phys.* **1993**, *98*, 3022–3037.

(44) Christie, J. M.; Hitomi, K.; Arvai, A. S.; Hartfield, K. A.; Mettlen, M.; Pratt, A. J.; Tainer, J. A.; Getzoff, E. D. Structural Tuning of the Fluorescent Protein iLOV for Improved Photostability. *J. Biol. Chem.* **2012**, *287*, 22295–22304.

(45) Cardoso Ramos, F.; Cupellini, L.; Mennucci, B. Computational Investigation of Structural and Spectroscopic Properties of LOV-Based Proteins with Improved Fluorescence. *J. Phys. Chem. B* **2021**, *125*, 1768–1777.

(46) Nogly, P.; Weinert, T.; James, D.; Carbajo, S.; Ozerov, D.; Furrer, A.; Gashi, D.; Borin, V.; Skopintsev, P.; Jaeger, K. et al. Retinal Isomerization in Bacteriorhodopsin Captured by a Femtosecond X-Ray Laser. *Science* **2018**, *361*, eaat0094.

(47) Christie, J. M.; Hitomi, K.; Arvai, A. S.; Hartfield, K. A.; Mettlen, M.; Pratt, A. J.; Tainer, J. A.; Getzoff, E. D. Structural Tuning of the Fluorescent Protein iLOV for Improved Photostability. *J. Biol. Chem.* **2012**, *287*, 22295–22304.

(48) Oudar, J. L.; Chemla, D. S. Hyperpolarizabilities of the Nitroanilines and Their Relations to the Excited State Dipole Moment. *J. Chem. Phys.* **1977**, *66*, 2664–2668.

(49) Campo, J.; Wenseleers, W.; Goovaerts, E.; Szablewski, M.; Cross, G. H. Accurate Determination and Modeling of the Dispersion of the First Hyperpolarizability of an Efficient Zwitterionic Nonlinear Optical Chromophore by Tunable Wavelength Hyper-Rayleigh Scattering. *J. Phys. Chem. C* **2008**, *112*, 287–296.