

Faculté de Sciences

Département de Chimie

Laboratoire de Chimie Théorique

**Quantum chemical investigations of nonlinear
optical compounds: from model to complex
systems for second harmonic imaging
microscopy**

Dissertation présentée par
Marc de Wergifosse
en vue de l'obtention du grade
de Docteur en Sciences

Composition du Jury:

Professeur Benoît Champagne (Promoteur)

Professeur Micheal Bearpark

Professeur Koen Clays

Professeur Tom Leyssens

Docteur Vincent Liégeois

Professeur Johan Wouters

2014

Quantum chemical investigations of nonlinear optical compounds: from model to complex systems for second harmonic imaging microscopy

by Marc de Wergifosse

Abstract: This thesis is focused on the characterization and the interpretation of the nonlinear optical (NLO) response of complex systems by quantum chemistry methods. The main goal is the reduction of the gap between real systems characterized experimentally and model systems usually employed in simulations and therefore the introduction of the structural complexity of the studied compounds. Computationally, the main targets are biological structures useful for second harmonic generation microscopy (SHIM), fluorescent proteins and collagen, of which the responses have different origins. The NLO response of fluorescent proteins originates mostly from a push-pull π -conjugated chromophore whereas, for collagen, small π -conjugated peptide bonds are at the origin of the NLO response.

In the case of fluorescent proteins, theoretical simulations carried out at different levels of approximation are accompanied by experimental investigations in order to design proteins that fluoresce over the whole visible spectra and exhibit large second-order NLO responses. In particular, different applications of the ONIOM scheme have been assessed in comparison with experiment, demonstrating the need for an explicit description of the close surroundings of the chromophore and the inclusion of electron correlation. This study has substantiated the importance of the symmetry argument for the second-order nonlinear optical properties of fluorescent proteins and the sensitivity of this principle.

In the investigation on collagen, a short peptide triple-helix model [(Pro-Pro-Gly)₁₀]₃ (PPG10) has been considered and its first hyperpolarizability has been calculated using the ONIOM method in combination with time-dependent Hartree-Fock and DFT calculations. Detailed analysis of the first hyperpolarizability tensors has unraveled that β originates from the individual peptide bonds and that the

triple-helix structure gives the relative orientation of these bonds. So, the total response is mostly determined by summing the β tensors of each bond as well as by accounting for (hyper)polarization effects between the chains.

The selection of quantum chemistry methods to describe the NLO responses of fluorescent proteins and collagen has been substantiated by investigations of the impact of electron correlation on the hyperpolarizabilities of small model systems, i) push-pull π -conjugated systems, which model the chromophore of fluorescent proteins and ii) open-shell systems such as the *p*-quinodimethane and two of its derivatives. For closed-shell systems, these studies validate the choice of the MP2 method for estimating the first hyperpolarizability and show the lack of reliability of DFT approaches with most exchange-correlation functionals. In parallel, several technical aspects associated with the evaluation of the hyperpolarizabilities have been addressed, *i.e* i) the automatization of the Romberg scheme to improve the numerical accuracy in the finite field method, ii) the evaluation of the frequency dispersion at correlated levels by approximated schemes, and iii) the deviations from Kleinman's conditions.

Voici venu le temps des remerciements. C'est avec une très vive émotion que ces quelques lignes sont rédigées. Ce travail n'aurait jamais pu aboutir sans l'aide d'innombrables personnes qui m'ont accompagnées durant ces dures années de labeur. Dans ces quelques phrases, j'espère n'oublier personne.

Tout d'abord, mes proches à qui je dédie cette thèse : mon aimée Coralie dont le soutien inébranlable m'a accompagné au jour le jour ; mon frère Michaël qui a toujours été un modèle d'enrichissement intellectuel et de soutien ; mes parents qui malgré une vie dure et tumultueuse m'ont tellement donné.

Je remercie le Pr. Benoît Champagne qui m'a donné l'opportunité de réaliser ce travail. Son aide, son enseignement ont été la source d'un apprentissage scientifique de très haute qualité.

Un tout grand merci au Dr. Edith Botek pour son soutien, sa grande disponibilité et pour les discussions scientifiques qui m'ont permis d'avancer à grand pas.

Mes remerciements vont aussi aux membres présents et passés du laboratoire ainsi qu'à ces chers Laurent Demelenne, Frédéric Wautelet et Jérôme de Ruycq.

Je remercie les membres de mon comité d'accompagnement ainsi que les membres de mon jury : Pr. Koen Clays, Pr. Tom Leyssens, Dr. Vincent Liégeois, Pr. Johan Wouters, Pr. André Peremans et plus particulièrement, Pr. Michael Bearpark pour les corrections de mon anglais.

Ces années de thèse ont été l'occasion d'initier des collaborations et des amitiés de part le monde, source d'enrichissement. Il m'est donc impossible de ne pas les remercier chaleureusement : tout d'abord, le Pr. Masayoshi Nakano qui m'a donné l'opportunité de séjourner à Osaka et dont l'accueil restera pour moi inoubliable ; mon ami, le Dr. Hitoshi Fukui, pour les bons moments passés dont la visite inoubliable de Kyoto ; le Pr. Ryohei Kishi pour le beau repas de départ et bien sûr, sa collaboration scientifique.

Parmi mes autres collaborateurs, je désire aussi témoigner un grand merci, aux Dr. Evelien De Meulenaere, Dr. Stijn Van Cleuvenbergen et au Pr. Thierry Verbiest.

Cette thèse n'aurait jamais pu voir le jour sans l'enseignement de M^{me} Viviane Demarteau durant mes jeunes années. Merci de m'avoir donné le goût de la chimie et d'avoir cru en moi, malgré mon parcours tumultueux. Sans vous, je n'en serai pas arrivé là.

Je remercie mes amis pour tous ces bons moments passés, en particulier, Jean-Philippe Moiny, pour nos temps de midi.

Finalement, il me reste à témoigner mes remerciements à mes collègues assistants, ainsi qu'aux Dr. Diane Baillieul et Dr. Isabelle Ravet, et bien entendu Daniel Lemaire, notre rocker dans l'âme malheureusement disparu trop vite.

Liège, le 13 avril 2014

Table of contents

I. <u>Introduction</u>	19
I.A. <i>Nonlinear optical effects</i>	20
I.A.1. General aspects	20
I.A.2. Hyper-Rayleigh scattering (HRS)	22
I.B. <i>Second-harmonic imaging microscopy</i>	26
I.B.1. Principles	26
I.B.2. Useful compounds for SHIM biological applications	27
I.C. <i>Objectives</i>	33
II. <u>Theoretical methods</u>	39
II.A. <i>Wave function-based methods</i>	39
II.A.1. The Hartree-Fock method	39
II.A.2. The Møller-Plesset perturbation theory	42
II.A.3. Coupled Cluster method	45
II.B. <i>Density functional theory</i>	49
II.C. <i>Second order nonlinear optical properties</i>	53
II.C.1. Time-dependent response theory for an exact state	53
II.C.2. Time-dependent Hartree-Fock method	55
II.C.3. Time-dependent DFT method	59
II.C.4. Frequency-dependent first hyperpolarizabilities using coupled cluster quadratic response theory	60
II.C.5. The Finite Field method and frequency dispersion.	64
II.C.5.a. The finite field method	64
II.C.5.b. Frequency dispersion	66
II.C.5.c. Kleinman's symmetry	67
II.C.5.d. Molecular structures and computational procedure	68
II.C.5.e. Towards an automatic Romberg differentiation procedure	70

II.C.5.e.1. Theory	70
II.C.5.e.2. Applications	79
II.C.5.f. Approximate frequency dispersions of the first hyperpolarizability from correlated calculations	85
II.C.5.g. Deviation from Kleinman's symmetry conditions	90
II.C.5.h. Conclusions	92
II.C.6. The ONIOM method	93
II.D. Solvent effects	94
II.E. Unit sphere representation of the first hyperpolarizability tensor	96
II.F. Geometry optimizations	97
II.G. Conclusions	98
III. <u>Second-Order Nonlinear Optical Properties of Fluorescent Proteins for Second-Harmonic Imaging</u>	107
III.A. The case of eGFP, eYFP, and DsRed	108
III.A.1. Introduction	108
III.A.2. Materials and methods	109
III.A.2.a. Materials	109
III.A.2.b. Instrumental Methods	109
III.A.2.c. Computational methods and geometrical structures	111
III.A.3. Results	112
III.A.4. Discussion	118
III.A.5. Conclusions	119
III.B. The case of mStrawberry and mCherry	121
III.B.1. Introduction	121
III.B.2. Materials and methods	122
III.B.2.a. Materials	122
III.B.2.b. Instrumental methods	122
III.B.2.c. Computational methods	122

III.B.3. Results	123
III.B.4. Discussion	127
III.B.4.a. mStrawberry and mCherry	127
III.B.4.b. mFruits	129
III.B.5. Conclusions	130
III.C <i>Improving the second-order nonlinear optical response of fluorescent proteins: the symmetry argument</i>	131
III.C.1. Introduction	131
III.C.2. Methods	133
III.C.2.a. Optical characterization	133
III.C.2.b. Computational methods and geometrical structures	134
III.C.3. Results	135
III.C.3.a. Experimental linear optical characterization	135
III.C.3.b. X-ray diffraction	136
III.C.3.c. Experimental first hyperpolarizability	139
III.C.3.d. Calculated first hyperpolarizability	139
III.C.4. Discussion	140
III.C.5. Conclusions	143
III.D. <i>Predicting and interpreting the second-order nonlinear optical responses of fluorescent proteins</i>	144
III.D.1. Introduction	144
III.D.2. Theoretical and computational aspects	146
III.D.3. Results and discussion	152
III.D.3.a. Comparison between full CPHF/TDHF results and the ONIOM values	152
III.D.3.b. Solvent effects	157
III.D.3.c. Comparison of HF to MP2 results on small systems	160
III.D.3.d. Assessment of DFT exchange-correlation functionals	161

III.D.4. Overall comparison with experiment	162
III.D.5. Conclusions	171
IV. <u>How the second-order nonlinear optical response of the collagen triple helix appears: a theoretical investigation</u>	181
IV.A. Introduction	181
IV.B. Materials and methods	183
IV.B.1. Theoretical and computational aspects	183
IV.B.2. Materials and experimental spectroscopic characterizations	186
IV.C. Results and discussion	188
IV.C.1. Assessment of the ONIOM method and of electron correlation effects	188
IV.C.2. Structure-property relationships	190
IV.C.3. [(Pro-Pro-Gly) ₁₀] ₃ : results and comparison to experiment	198
IV.D. Conclusions	201
V. <u>Electron Correlation Effects on the Hyperpolarizabilities of Model Systems</u>	209
V.A. <i>Introduction</i>	210
V.B. <i>Electron Correlation Effects on the First Hyperpolarizability of Push-Pull π-Conjugated Systems</i>	217
V.B.1. Theoretical and computational aspects	217
V.B.2. Results and discussions	220
V.B.2.a. Optimized geometries	220
V.B.2.b. Basis set effects	220
V.B.2.c. Electron correlation effects using wavefunction based methods	222
V.B.2.d. Structure-property relationships	227
V.B.2.e. Assessment of DFT exchange-correlation functionals.	232
V.B.3 Further discussion, conclusion, and outlook	237

V.C. <i>Electron correlation effects on the hyperpolarizabilities of p-quinodimethane derivatives</i>	239
V.C.1. Theoretical and computational approaches	239
V.C.2. Results and discussion	240
V.C.2.a. Wavefunction methods	240
V.C.2.b. Density functional theory calculations	245
V.C.2.c. Polarizability and first hyperpolarizability	247
V.C.3. Conclusions and outlook	248
VI. <u>Conclusions and outlooks</u>	259

Publications List :

de Wergifosse M, Liégeois V and Champagne B, On the evaluation of the molecular static and dynamic first hyperpolarizabilities. *International Journal of Quantum Chemistry*. **2014**; accepted, DOI: 10.1002/qua.24685.

de Wergifosse M, de Ruyck J, and Champagne B, How the second-order nonlinear optical response of the collagen triple helix appears: a theoretical investigation *Journal of Physical Chemistry C*. **2014**; accepted, DOI: 10.1021/jp501602a.

De Meulenaere E, Nguyen Bich N, de Wergifosse M, Van Hecke K, Van Meervelt L, Vanderleyden J, Champagne B, and Clays K, Improving the second-order nonlinear optical response of fluorescent proteins: The symmetry argument. *Journal of the American chemical society*. **2013**; 135 (10) : 4061-4069.

de Wergifosse M, Wautelet F, Champagne B, Kishi R, Fukuda K, Matsui H, and Nakano M, Challenging compounds for calculating hyperpolarizabilities: *p*-quinodimethane derivatives. *Journal of Physical Chemistry A*. **2013**; 117 (22) : 4709-4715.

de Wergifosse M, Champagne B. Electron correlation effects on the first hyperpolarizability of push-pull- π -conjugated systems. *The journal of chemical physics*. **2011**; 134 (7) : 074113.

De Meulenaere E, de Wergifosse M, Botek E, Spaepen S, Champagne B, Vanderleyden J, and Clays K, Nonlinear optical properties of mStrawberry and mCherry for second harmonic imaging. *Journal of Nonlinear Optical Physics and Materials*. **2010**; 19 (1) : 1-13.

De Meulenaere E, Asselberghs I, de Wergifosse M, Botek E, Spaepen S, Champagne B, Vanderleyden J, and Clays K, Second-order nonlinear optical properties of fluorescent proteins for second-harmonic imaging. *Journal of Materials Chemistry*. **2009**; 19 (40) : 7514-7519.

Abbreviations List:

BLA	Bond length alternation
CC	Coupled cluster
CCSD	Coupled cluster singles-and-doubles
CCSD-QRF	CCSD quadratic response functions
CCSD(T)	CCSD with perturbative estimate of the triples
CI	Configuration interaction
CPHF	Coupled-perturbed Hartree-Fock
D/A	Donor/acceptor
DFT	Density functional theory
DR	Depolarization ratio
eGFP	Enhance green fluorescent protein
eYFP	Enhance yellow fluorescent protein
FF	Finite field
FP	Fluorescent protein
GFP	Green fluorescent protein
GGA	Generalized gradient approximation
GGG3	$[(\text{Gly-Gly-Gly})_3]_3$
HF	Hartree-Fock
HRS	Hyper-Rayleigh scattering
HWHM	Half-width at half-maximum
IEF	Integral equation formalism
LC	Long-range corrected
LDA	Local density approximation
MM	Molecular mechanics
MP _n	Moller-Plesset n^{th} -order
NLO	Nonlinear optical
ONIOM	Our own N-layered integrated molecular orbital and molecular mechanics
PDB	Protein data base
PCM	Polarizable continuum model
PPG3	$[(\text{Pro-Pro-Gly})_3]_3$
PPG10	$[(\text{Pro-Pro-Gly})_{10}]_3$
PMISP	Polarized multipolar interactions with supramolecular pairs
QM	Quantum mechanics
RPA	Random phase approximation
SCF	Self-consistent-field
SCS-MP2	Spin-component-scaled MP2
SDQ-MP4	Singles-doubles-quadruples MP4
SHG	Second harmonic generation

SHIM	Second harmonic imaging microscopy
TDDFT	Time-dependent DFT
TDHF	Time-dependent HF
TPEF	Two-photon-excitation fluorescence
TSA	Two states approximation
U	Unrestricted
XC	Exchange-correlation

“Mes compagnons et moi, nous fûmes alors témoins d’un curieux spectacle. Les panneaux du salon étaient ouverts, et comme le fanal du *Nautilus* n’était pas en activité, une vague obscurité régnait au milieu des eaux.

Le ciel orageux et couvert d’épais nuages ne donnait aux premières couches de l’Océan qu’une insuffisante clarté.

J’observais l’état de la mer dans ces conditions, et les plus gros poissons ne m’apparaissaient plus que comme des ombres à peine figurées, quand le *Nautilus* se trouva subitement transporté en pleine lumière. Je crus d’abord que le fanal avait été rallumé, et qu’il projetait son éclat électrique dans la masse liquide. Je me trompais, et après une rapide observation, je reconnus mon erreur.

Le *Nautilus* flottait au milieu d’une couche phosphorescente, qui dans cette obscurité devenait éblouissante. Elle était produite par des myriades d’animalcules lumineux, dont l’étincellement s’accroissait en glissant sur la coque métallique de l’appareil. Je surprénais alors des éclairs au milieu de ces nappes lumineuses, comme eussent été des coulées de plomb fondu dans une fournaise ardente, ou des masses métalliques portées au rouge blanc ; de telle sorte que par opposition, certaines portions lumineuses faisaient ombre dans ce milieu igné, dont toute ombre semblait devoir être bannie. Non ! ce n’était plus l’irradiation calme de notre éclairage habituel ! Il y avait là une vigueur et un mouvement insolites ! Cette lumière, on la sentait vivante !

En effet, c’était une agglomération infinie d’infusoires pélagiens, de noctiluques miliaires, véritables globules de gelée diaphane, pourvus d’un tentacule filiforme, et dont on a compté jusqu’à vingt-cinq mille dans trente centimètres cubes d’eau. Et leur lumière était encore doublée par ces lueurs particulières aux méduses, aux astéries, aux aurélias, aux pholadesdattes, et autres zoophytes phosphorescents, imprégnés du graissin des matières organiques décomposées par la mer, et peut-être du mucus sécrété par les poissons.”

Vingt mille lieues sous les mers

Jules Verne

I. Introduction

Nonlinear optics studies the phenomena where the light modifies the properties of matter and those where the properties of light are changed by matter [for example, in the second harmonic generation (SHG) phenomenon].¹ This area of science is in unceasing development and one of its main goals is focusing on new materials with large nonlinear optical (NLO) susceptibilities or, from a molecular point of view, molecules with large hyperpolarizabilities. Nonlinear optics is also the basis of new characterization techniques such as the second-harmonic imaging microscopy (SHIM)², which needs endogenous or exogenous probes with large hyperpolarizabilities. In this particular field, the theoretical tools for the prediction and the interpretation of hyperpolarizabilities are the foundation of a multidisciplinary approach including also the synthesis of exogenous dyes, the genetic engineering of probes, and experimental characterizations³. Nowadays, the hyperpolarizabilities of organic compounds containing 20 to 50 atoms can be predicted with sufficient accuracy to guide the synthesis in the direction of better performing compounds.⁴ Those simulations take into account the effects of the electron correlation, of the solvent, and in some cases, of the vibrational motions. However, in the case of supramolecular assemblies, the characterization and the prediction of hyperpolarizabilities are still a tremendous task for quantum chemistry (coordination complexes⁵, helical structures⁶, ...). The main goal of this thesis is to overcome this by predicting and characterizing hyperpolarizabilities of systems containing up to 1000 atoms, particularly those with a biological interest for SHIM. Those compounds are well known to be quite challenging as already recognized by Loison *et al.*⁷ in 2010: “For large polymers such as proteins, precise quantum chemical calculations of the hyperpolarizability are beyond the reach of today’s computer power”.

I.A. *Nonlinear optical effects*

I.A.1. General aspects

Applying light to matter, owing to its electric field component, results in a perturbation of the motion of electrons and nuclei, which has the effect to create different phenomena, such as absorption, emission or scattering of the light. These optical phenomena can be classified into linear or nonlinear processes regarding the dependence of the intensity of the optical response with respect to the intensity of the electric field. For example, in Rayleigh scattering, when a photon is scattered, there is no formation of an excited state, which is commonly viewed as the interaction with a so-called virtual state. This phenomenon is a linear process. When a scattered photon has the same frequency as the sum of two incident ones ($\omega_\sigma = \omega_1 + \omega_2$), this is a second-order NLO process, as the intensity of the optical response with respect to the intensity of the electric field is quadratic. If these two incident photons have the same frequency it is known as second harmonic generation (see figure 1). The frequency conversion is then noted $(-2\omega; \omega, \omega)$.

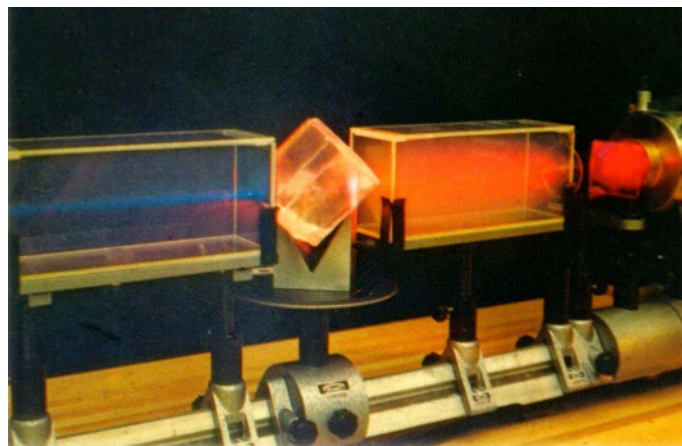


FIG. 1. Conversion of red (700 nm) into blue (350 nm) light by a crystal of ammonium dihydrogen phosphate $\text{NH}_4\text{H}_2\text{PO}_4$.

In all these processes, the motion of the electrons and nuclei is perturbed by the external electric field and this interaction can be characterized by the variations of the molecular dipole moment under the form of a Taylor expansion:

$$\vec{\mu}(\vec{F}) = \vec{\mu}_0 + \vec{\alpha} \cdot \vec{F} + \frac{1}{2!} \vec{\vec{\beta}} : \vec{F}\vec{F} + \frac{1}{6!} \vec{\vec{\gamma}} : \vec{F}\vec{F}\vec{F} + \dots \quad (\text{I.A.1})$$

The first term in the expansion, $\vec{\mu}_0$, is the intrinsic dipole moment vector, the second term $\vec{\alpha}$ is the polarizability tensor, and the higher-order terms are the nonlinear ones where $\vec{\vec{\beta}}$ and $\vec{\vec{\gamma}}$ are respectively the first and second hyperpolarizability tensors. $\vec{\vec{\beta}}$ is a rank-3 tensor with 27 components:

$$\vec{\vec{\beta}} = \begin{pmatrix} \beta_{xxx} & \beta_{xyy} & \beta_{xzz} & \beta_{xyz} & \beta_{xzy} & \beta_{xxz} & \beta_{xzx} & \beta_{xxy} & \beta_{xyx} \\ \beta_{yxx} & \beta_{yyy} & \beta_{yzz} & \beta_{yyz} & \beta_{yzy} & \beta_{yxz} & \beta_{yzx} & \beta_{yxy} & \beta_{yyx} \\ \beta_{zxx} & \beta_{zyy} & \beta_{zzz} & \beta_{zyz} & \beta_{zzy} & \beta_{zxz} & \beta_{zzx} & \beta_{zxy} & \beta_{zyx} \end{pmatrix}. \quad (\text{I.A.2})$$

The Taylor expansion is sometimes replaced by a power series, of which the consequence is to implicitly include the numerical $1/n!$ factor into the hyperpolarizabilities. These definitions correspond to different conventions, T when using the Taylor expansion and B for the power series expansion.⁸ Common nonlinear optical processes are presented in table 1. In this thesis, we are mainly interested in SHG, which is the process involved in SHIM.

<i>Second-order effects</i>		
Static		$\vec{\vec{\beta}}(0;0,0)$
EOPE	Electro-optical Pockels effect	$\vec{\vec{\beta}}(-\omega;\omega,0)$
SHG	Second-harmonic generation	$\vec{\vec{\beta}}(-2\omega;\omega,\omega)$

Table 1. Common second-order nonlinear optical effects

By symmetry, some of the $\vec{\beta}$ components are identical or zero. In the case of second-harmonic generation, the following permutation can be applied on the components, as the two incident photons are indiscernible:

$$\beta_{ijk} = \beta_{ikj}. \quad (\text{I.A.3})$$

18 independent elements remain. If the frequencies implied are far from resonance condition as for a static field, the Kleinman's conditions are satisfied:

$$\beta_{ijk} = \beta_{ikj} = \beta_{jik} = \beta_{kij} = \beta_{jki} = \beta_{kji}. \quad (\text{I.A.4})$$

Therefore, only 10 independent components need to be characterized. All quantities presented in this thesis are in atomic units (or in 10^{-30} esu for β). Conversion factors are presented in table 2.

μ	$1 \text{ a.u.} = 8.478 \cdot 10^{-30} \text{ Cm}$
α	$1 \text{ a.u.} = 0.16487 \cdot 10^{-40} \text{ C}^2 \text{ m}^2 \text{ J}^{-1}$
β	$1 \text{ a.u.} = 3.206361 \cdot 10^{-53} \text{ C}^3 \text{ m}^3 \text{ J}^{-2}$
	$1 \text{ a.u.} = 8.6392 \cdot 10^{-33} \text{ esu}$
γ	$1 \text{ a.u.} = 6.360 \cdot 10^{-64} \text{ C}^4 \text{ m}^4 \text{ J}^{-3}$

Table 2. Unit conversions table

I.A.2. Hyper-Rayleigh scattering (HRS)

The incoherent hyper-Rayleigh scattering^{1b, 9} generates light at the second harmonic frequency 2ω . As it is an incoherent process, this effect is weak but it avoids the need to order the chromophores, which provides a useful route to determine β . Indeed, it is assumed that for the characterization of the SHG response of chromophores in solution, there are no specific phase relations between individual molecules. They can therefore be treated as individual, uncorrelated scattering units. The HRS scattering intensity, $I_{\text{HRS}}^{2\omega}$, is a linear function of the number N of chromophores in solution. The slope of this linear function is a

quadratic form of the first hyperpolarizability tensor components, $\langle \beta_{\text{HRS}}^2 \rangle$, and it is also proportional to G , and to I_0^2 , the square of the incident light:

$$I_{\text{HRS}}^{2\omega} = GN \langle \beta_{\text{HRS}}^2 \rangle I_0^2. \quad (\text{I.A.5})$$

Thus, it is experimentally possible to determine the first hyperpolarizability of a given compound by comparing its $\langle \beta_{\text{HRS}}^2 \rangle$ slope to reference compounds ones. In a genuine HRS experiment where the incident light is vertically polarized (along the Z axis), the scattered light is analyzed at 90° from the incident one (Y axis) (fig. 2). The Intensity of the horizontally polarized scattered light $I_{\text{HV}}^{2\omega}$ is proportional to the orientational average $\langle \beta_{\text{ZXX}}^2 \rangle$, while for the vertically polarized $I_{\text{VV}}^{2\omega}$ to $\langle \beta_{\text{ZZZ}}^2 \rangle$. The $\langle \beta_{\text{ZZZ}}^2 \rangle$ and $\langle \beta_{\text{ZXX}}^2 \rangle$ molecular averages are defined as¹⁰

$$\begin{aligned} \langle \beta_{\text{ZZZ}}^2 \rangle = & \frac{1}{7} \sum_{\zeta}^{x,y,z} \beta_{\zeta\zeta\zeta}^2 + \frac{4}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\eta}^2 + \frac{2}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\zeta\eta\eta} + \frac{4}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\eta\zeta\zeta} \beta_{\zeta\zeta\eta} \\ & + \frac{4}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\eta\eta\zeta} + \frac{1}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\eta\zeta\zeta}^2 + \frac{4}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\zeta\eta} \beta_{\eta\zeta\xi} + \frac{1}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\eta\zeta\zeta} \beta_{\eta\zeta\xi} \\ & + \frac{4}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\zeta\eta} \beta_{\xi\xi\eta} + \frac{2}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi}^2 + \frac{4}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi} \beta_{\eta\zeta\xi} \end{aligned} \quad (\text{I.A.6})$$

$$\begin{aligned} \langle \beta_{\text{ZXX}}^2 \rangle = & \frac{1}{35} \sum_{\zeta}^{x,y,z} \beta_{\zeta\zeta\zeta}^2 + \frac{4}{105} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\zeta\eta\eta} - \frac{2}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\eta\eta\zeta} + \frac{8}{105} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\eta}^2 \\ & + \frac{3}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\eta\eta}^2 - \frac{2}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\eta} \beta_{\eta\zeta\zeta} + \frac{1}{35} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\eta} \beta_{\zeta\xi\xi} - \frac{2}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\zeta\xi} \beta_{\eta\eta\xi} \\ & - \frac{2}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\zeta\eta} \beta_{\eta\xi\xi} + \frac{2}{35} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi}^2 - \frac{2}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi} \beta_{\eta\zeta\xi} \end{aligned} \quad (\text{I.A.7})$$

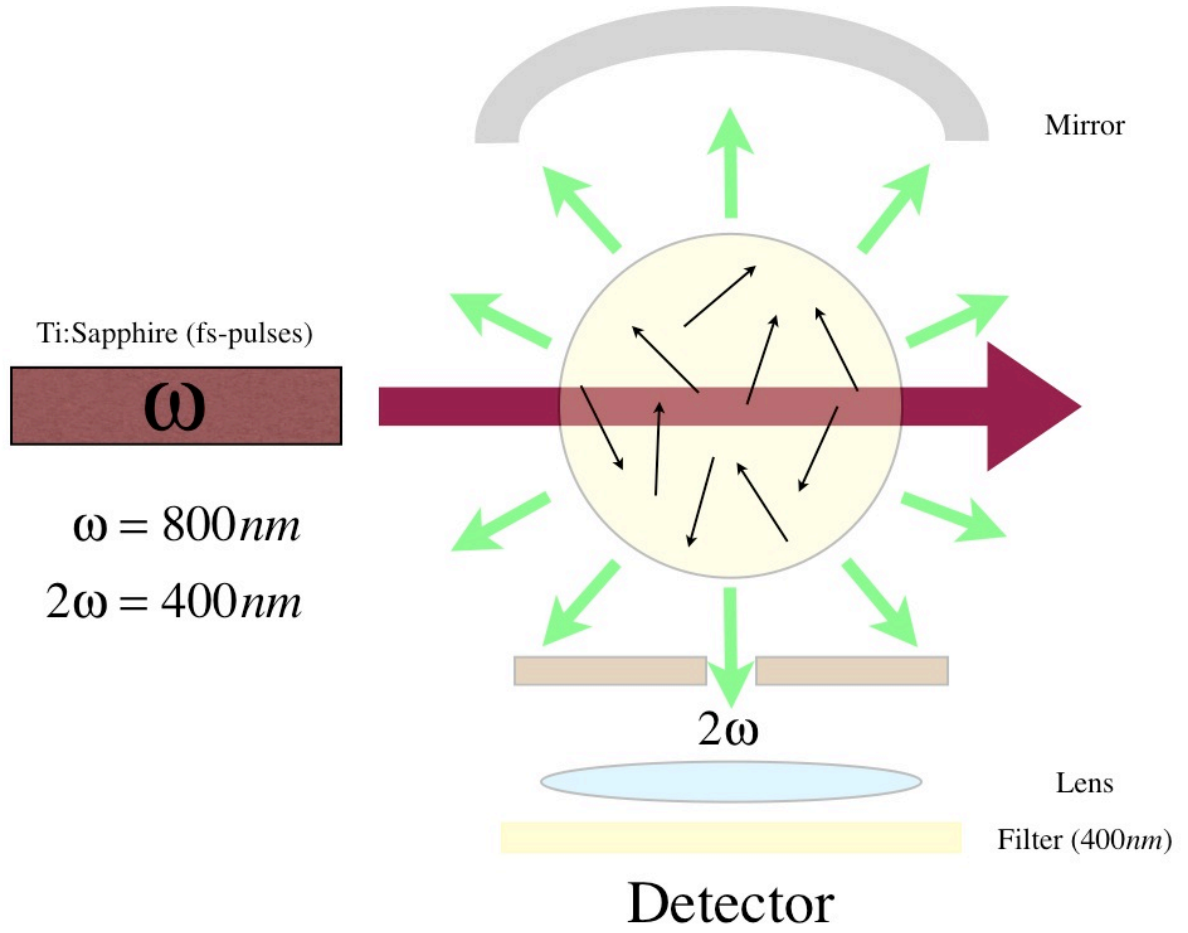


FIG. 2. Hyper-Rayleigh scattering experiment setup.

The “full” HRS first hyperpolarizability reads

$$\beta_{\text{HRS}}(-2\omega; \omega, \omega) = \sqrt{\langle \beta_{\text{ZZZ}}^2 \rangle + \langle \beta_{\text{ZXX}}^2 \rangle}. \quad (\text{I.A.8})$$

An interesting quantity in HRS is the depolarization ratio (DR), defined as

$$\text{DR} = \frac{I_{\text{VV}}^{2\omega}}{I_{\text{HV}}^{2\omega}} = \frac{\langle \beta_{\text{ZZZ}}^2 \rangle}{\langle \beta_{\text{ZXX}}^2 \rangle}. \quad (\text{I.A.9})$$

DR reveals the shape of the NLO-phore. Indeed, the molecular symmetry reduces considerably the number of β tensor components, and by consequence the number of terms intervening in equations (I.A.6) and (I.A.7). For example, for the CCl_4 molecule belonging to the T_d point group, the only non-vanishing tensor component considering the Kleinman conditions is β_{xyz} . The corresponding DR is therefore equal to 1.5, which is typical of an octupolar response. On the other hand, molecules

belonging to the C_{2v} point group have three independent components: β_{zxx} , β_{zyy} , and β_{zzz} . In a limit C_{2v} case, by assuming that $\beta_{zyy}=0$ and considering the ratio β_{zxx}/β_{zzz} tends to zero, the DR becomes equal to 5, which corresponds to a one-dimensional system.

Alternatively, the HRS measurements can be discussed in terms of their multipolar components¹¹, where the β tensor is decomposed as the sum of a dipolar ($J=1$),

$$|\beta_{J=1}|^2 = \frac{3}{5} \sum_{\zeta}^{x,y,z} \beta_{\zeta\zeta\zeta}^2 + \frac{6}{5} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\zeta\eta\eta} + \frac{3}{5} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\eta\zeta\zeta}^2 + \frac{3}{5} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\eta\zeta\zeta} \beta_{\eta\xi\xi}, \quad (\text{I.A.10})$$

and an octupolar ($J=3$) tensorial β_J – components,

$$|\beta_{J=3}|^2 = \frac{2}{5} \sum_{\zeta}^{x,y,z} \beta_{\zeta\zeta\zeta}^2 - \frac{6}{5} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\zeta\eta\eta} + \frac{12}{5} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\eta\zeta\zeta}^2 - \frac{3}{5} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\eta\zeta\zeta} \beta_{\eta\xi\xi} + \frac{2}{35} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi}^2. \quad (\text{I.A.11})$$

(The expressions (I.A.10) and (I.A.11) are given considering the Kleinman's symmetry.) Then, the $\langle \beta_{zzz}^2 \rangle$ and $\langle \beta_{zxx}^2 \rangle$ molecular averages can be rewritten:

$$\langle \beta_{zzz}^2 \rangle = \frac{9}{45} |\beta_{J=1}|^2 + \frac{6}{105} |\beta_{J=3}|^2, \quad (\text{I.A.12})$$

$$\langle \beta_{zxx}^2 \rangle = \frac{1}{45} |\beta_{J=1}|^2 + \frac{4}{105} |\beta_{J=3}|^2, \quad (\text{I.A.13})$$

whereas the HRS first hyperpolarizability reads

$$\begin{aligned} \beta_{\text{HRS}} &= \sqrt{\frac{10}{45} |\beta_{J=1}|^2 + \frac{10}{105} |\beta_{J=3}|^2} \\ &= |\beta_{J=1}| \sqrt{\frac{2}{3} \left(\frac{1}{3} + \frac{1}{7} \rho^2 \right)}, \end{aligned} \quad (\text{I.A.14})$$

and the DR is given by:

$$DR = 9 \frac{(1 + \frac{2}{7} \rho^2)}{(1 + \frac{12}{7} \rho^2)}, \quad (\text{I.A.15})$$

where $\rho = |\beta_{J=3}|/|\beta_{J=1}|$ is called the nonlinear anisotropy parameter, which compares the relative contribution of the octupolar and dipolar β_J – components. If there is no octupolar contribution ($|\beta_{J=3}|=0$), i.e. for pure dipolar molecules, the nonlinear anisotropy ρ is equal to 0 and the DR to 9, which is higher than for a one-dimensional system with a DR of 5 and a ρ equal to $\sqrt{2/3}$.

I.B. *Second-harmonic imaging microscopy*

I.B.1. Principles

Second-harmonic imaging microscopy is a high-resolution biomedical imaging technique, which has been developed to get contrast enhancements of non-centrosymmetric molecular arrangements.^{2, 12, 13} This technique is highly complementary to two-photon-excitation fluorescence (TPEF) microscopy, which involves the relaxation of an excited state. These two types of microscopy imply the use of the same near-infrared incident laser beam. Since SHIM only involves a scattered light process, it does not damage the living environment. Information about the molecular organization of chromophores, including dyes and structural proteins, can be extracted from SHG imaging data because the SHG signal is polarization-dependent, and thus the SHG polarization anisotropy can be used to determine the absolute orientation and the degree of organization of proteins in tissues. In addition TPEF images can be collected in a separate data channel simultaneously to SHG, where polarization anisotropy can also be used, depending of the time decay of the involved excited state. Correlation between SHG and TPEF provides complementary structural information: i) with TPEF, the location of chromophores, and ii) with SHIM, noncentrosymmetric molecular arrangements.

SHIM is a very powerful non-invasive diagnostic tool, which can probe biological functions and monitor molecular structures in vivo. High contrast imaging depends on the molecular orientation of ordered and disordered structures, which

can be quantitatively analyzed. In a natural biological environment, SHG signals are present from endogenous proteins that have ordered structures: myosin, mainly constituted by α -helices, collagen triple-helix, and microtubules.¹³⁻¹⁴ Many other biological structures can be enlightened by the use of dyes, as for example in the imaging of voltage sensitive membranes via exogenous labeling.

I.B.2. Useful compounds for SHIM biological applications

In the endogenous SHIM emitters class, collagen is one of the most studied for its application in the diagnosis of diseases as reviewed in the book edited by Pavone and Campagnola¹³. Indeed, diseases alter the tissues and modify the collagen structure. In the case of a fibrosis, the three-dimensional distribution of collagen is modified and fibrillar collagen accumulates in the tissue (fig. 3). During cancer, tumor-associated collagen signatures are observed, associated with changes of collagen alignment (fig. 4). An alteration of the collagen structural arrangement of an organ leads to its functional failure. In their review, Campagnola and Loew's² report that the first biological second harmonic generation imaging experiment was done in 1986 by I. Freund *et al.*¹⁵ on collagen of rat-tail tendon (fig. 5). They revealed for the first time the existence of a discrete network of fine, polar, filamentous or columnar, structures, and, also, the presence of strongly polar surface, or near-surface patches. It is now known that biological materials can be highly polarized and often assemble into large, ordered noncentrosymmetric structures. The determination of the molecular origin of this particular response is however difficult to elucidate but type-1 collagen has been clearly identified as one of the principal endogenous SHG emitters.¹⁵⁻¹⁶

Microtubules are other endogenous SHG emitters, which are present in the cell cytoskeleton and serve numerous vital functions. A single microtubule is constituted of 13 parallel protofilaments, each of which is a linear chain of tubulin proteins connected end to end. During SHIM of microtubules, the SHG intensity is

sensitive to the orientation of the microtubules and reflects its polarity. A direct application of SHIM to microtubules can tackle the embryogenesis. Indeed, from the initial single cell state, the fertilized egg goes through a series of cell divisions, which can be monitored at subcellular level with temporal resolution finer than the duration of each phase of cell division.

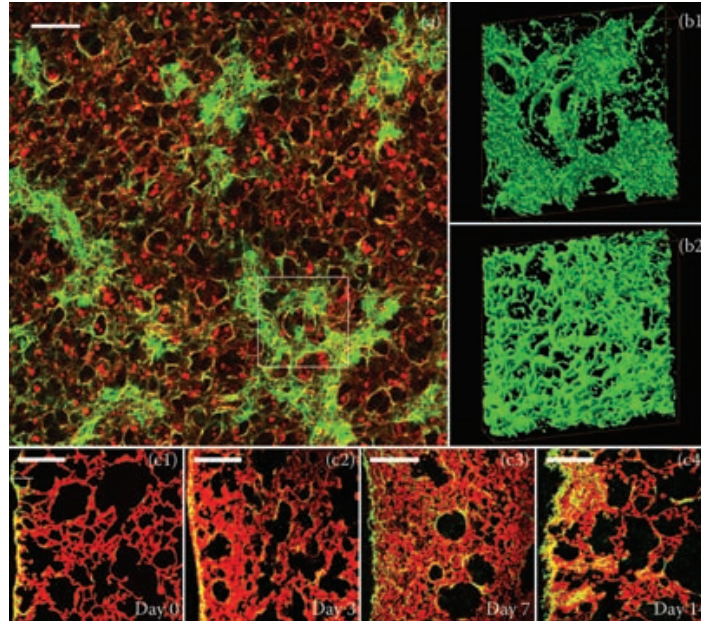


FIG. 3. SHG/TPEF imaging of lung fibrosis in bleomycin-treated mice. SHG (green: collagen) and TPEF (red: elastin, macrophages) signals excited at 860 nm. Yellow color: SHG/2PEF colocalization.

(From Pena, A.-M. et al. Three-dimensional investigation and scoring of extracellular matrix remodeling during lung fibrosis using multiphoton microscopy. *Microsc. Res. Tech.* 2007. 70:162–170, with the permission of John Wiley and Sons.)

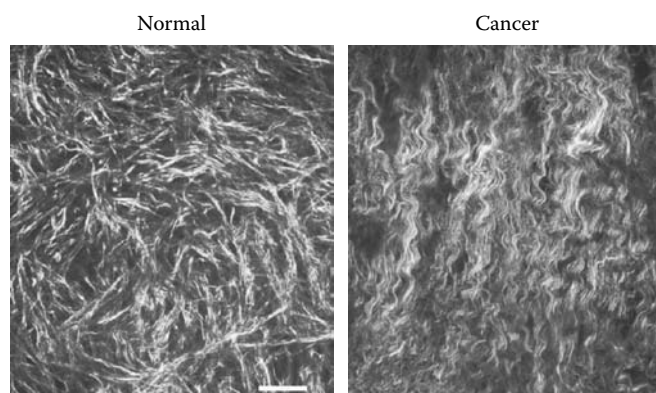


FIG. 4. SHG single optical sections from normal and malignant human ovarian tissues. Scale bar = 25 microns. (from Nadiarnykh, O. et al. 2010. *BMC Cancer* 10:94., open access.)

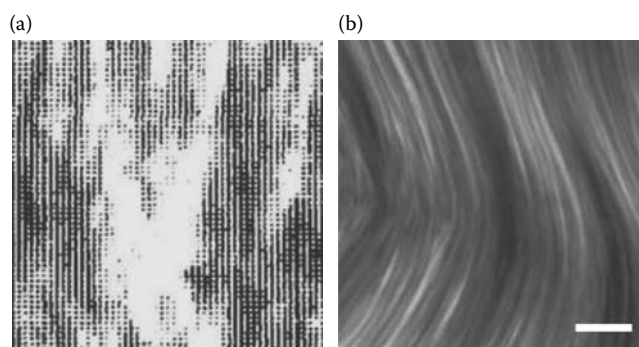


FIG. 5. SHG from rat-tail tendon collagen. (a) First coherent second-harmonic image of rat-tail tendon. (from Freund, I. and Deutsch, M. 1986. *Opt Lett* 11, 94., with the permission of the Optical Society of America.) (b) Contemporary image of rat-tail tendon collagen. (from *Biophys J*, 88, Williams, R.M. Zipfel, W.R., and Webb, W.W. 1377, 2005, with the permission of Elsevier.) Scale bar 10 μm . For comparison, the image size is $2.5 \times 2.5 \text{ mm}^2$ and $56 \times 56 \mu\text{m}^2$ for panels a and b, respectively.

Exogenous SHG emitters can also be employed. In 1993, Bouevitch *et al.*¹⁷ showed that the electric potential of a lipid membrane can be monitored by surface SHG. They used a hemispherical bilayer of oxidized cholesterol stained by styryl dyes with chiral centers for this demonstration; and they showed the path to image living cells. Dyes are also used for imaging membrane potential, which is a crucial element to understand the functions of the central nervous system. Indeed, SHIM allows monitoring the SHG intensity that is generated by the ordered dyes present in the membrane. With different potentials (electric fields), different responses are observed, as a result of an electrooptic mechanism.

The fine-tuning of a number of molecular properties is required for the design of dyes for SHG imaging.^{12a} Firstly, a high molecular first hyperpolarizability at the wavelength of illumination is needed for the chromophore in order to maximize the signal. Secondly, SHG dyes with resonance enhancement of β used in a biological environment need to have charge transfer bands at approximately the wavelength (one-photon enhancement) and half the wavelength (two-photon enhancement) of the biological tissue transparency window (700-900 nm). Thirdly, a high affinity for the biological membrane is also needed for an efficient chromophore. These

considerations need to be tuned in order to reduce the phototoxicity. Different types of dyes have been designed during the past decades, which satisfy simple practical conditions:^{12a}

- 1°, an optimized highly-polarizable π -conjugated pathway to provide a good coupling between the donor and the acceptor groups. Among these are the popular polymethine bridges, due to the strong π -conjugation and the ease of synthesis. This is why Bouevitch *et al.*¹⁷ have succeeded with the extended FM4-64 (fig. 6) polymethine chromophore, which is now a commercially available probe. Alain *et al.*¹⁸ showed also that “the polyenic chain appears as an efficient conjugated linker allowing for long-distance intra-molecular charge transfer. Both lengthening the polyenic chain and increasing the donor/acceptor strength result in bathochromic shift and enhancement of photo-induced intra-molecular charge transfer. This allows for the design of push-pull diphenylpolyenes exhibiting a highly polarized excited state (with dipole moment close to 40 D).” However as the length of the π -conjugated pathway increases, the solubility of the dyes in aqueous media becomes worse.

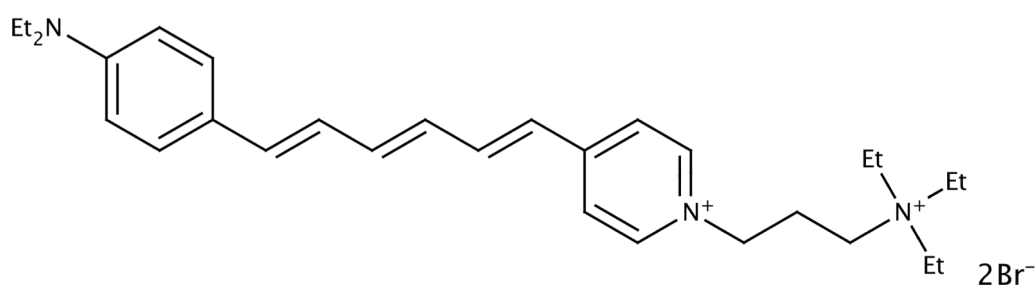


FIG. 6. FM4-64 (*N*-(3-Triethylammoniumpropyl)-4-(6-(4-(Diethylamino) Phenyl) Hexatrienyl) Pyridinium Dibromide)

- 2°, suitable donor and acceptor to exalt the β response. The presence of an intramolecular charge transfer is one of the conditions of an effective SHG.¹⁹ This can be achieved with strong donor-acceptor coupling. Common donor and acceptor groups are listed, donor groups: $-\text{N}(\text{CH}_3)_2$, $-\text{NH}_2$, $-\text{OH}$, $-\text{OCH}_3$, halogens, $-\text{CH}_3$; acceptor groups: $-\text{COOH}$, $-\text{SO}_3\text{H}$, $-\text{NO}_2$.

- 3°, for a set of chromophores in the focal volume, their major β tensor components needs to be aligned along the electric field vector of the incident light in order to have a larger SHG signal coming from constructive interferences. Biological systems contain many membranes and interfacial structures, which provide asymmetric environments for ordering these chromophores.
- 4°, the dyes must be amphiphilic, which enables their insertion into lipid bilayers in a preferential direction, in order to satisfy 3° as well as to avoid diffusion. The choice of polar and nonpolar groups subtly affects the biological localization of the dyes. For example, Barsu *et al.*²⁰ investigated a number of neutral headgroups that gave good membrane localization. This original approach based on carbohydrate head groups is an important advancement in the design of membrane probes, since these highly versatile functional groups confer adequate hydrophilicity and conserve overall molecular neutrality. The dyes can also be attached to proteins or other cellular structures. For example, Salafsky²¹ used the PyMPO dye (fig.7) which was covalently attached to proteins to elucidate the tilt angle of proteins absorbed onto surfaces.

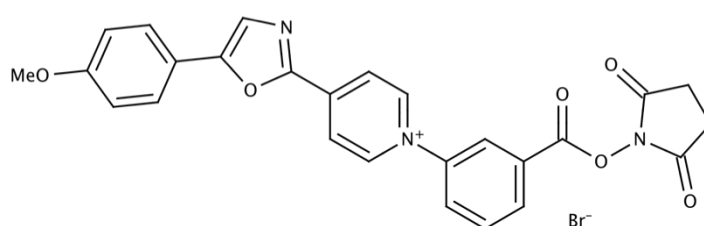


FIG. 7. PyMPO (1-(3-(succinimidylloxycarbonyl)benzyl)-4-(5-(4-methoxyphenyl)oxazol-2-yl)pyridinium bromide)

Furthermore, chiral chromophores ordered in monolayers are also of some interests in the characterization of bilayer cell membranes. In bilayers where the achiral dyes are in opposite direction in each layers, the centrosymmetric packing of chromophores has the effect to cancel their SHG signals, while the use of chiral chromophores leads to a non-centrosymmetric packing and allows for the SHG

signals of the membrane bilayer. This has been shown for example in the case of a chiral dimer of an achiral dye called di-4-ANEPPDHQ (fig.8).²²

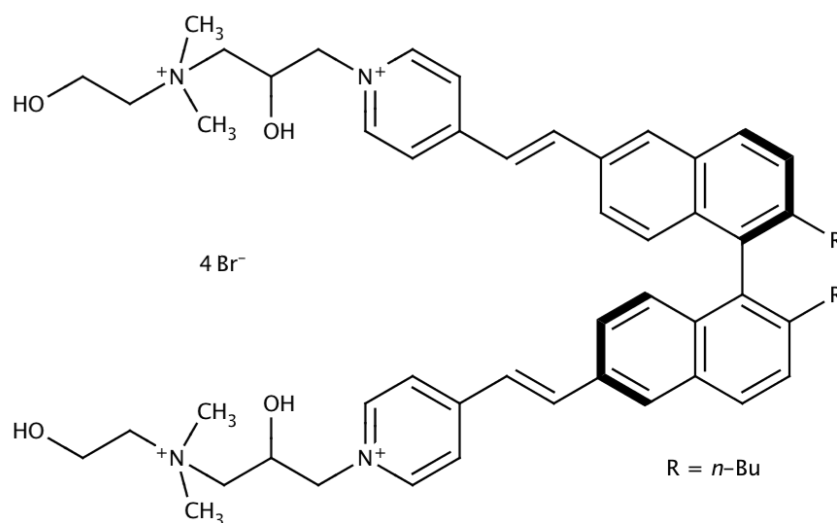


FIG. 8. Dimer of di-4-ANEPPDHQ (1-[2-Hydroxy-3-(N,N-di-methyl-N-hydroxyethyl)ammoniopropyl]-4-[β -[2-(di-*n*-butylamino)-6-naphthyl]vinyl]pyridinium dibromide)

Probes can also be introduced by genetic modification. The development of fluorescent proteins that can be incorporated into organisms by genetic engineering led to the chemistry 2008 Nobel Prize awarded to Shimomura, Chalfie, and Tsien.²³ 238 amino acids forming a β -barrel around a central chromophore is the structural arrangement of the green fluorescent protein (GFP). “In the live specimen of jellyfish *Aequorea*, the light organs contain GFP in addition to aequorin, and the energy of the blue light produced by the aequorin molecule is transferred to the GFP molecule, and GFP emits green light.”^{23b} Several studies have reported on the qualitative use of second-harmonic imaging from Green Fluorescent Protein.²⁴ A quantitative value for the second-order nonlinear optical properties for enhanced Green Fluorescent Protein (eGFP) and for the photoswitchable variant Dronpa has also been determined.²⁵ These results have shown that the chromophore in these GFPs is capable of generating an appreciable second-harmonic signal, which could be used for SHIM. In the

case of GFP, the direct environment reverses the amplitude of the response between protonated and deprotonated forms.

I.C. Objectives

The objective of this thesis is to characterize and interpret by quantum chemistry methods the NLO response of complex systems. The main goal is to bridge the gap between real systems characterized experimentally and model systems employed in simulations and therefore to introduce the structural complexity of the studied compounds in the simulations. The main targets are biological structures useful for SHIM, in particular, fluorescent proteins and collagen. These studies imply the development of new expertises. Indeed, systems of that size are far from the usually investigated systems. Fluorescent proteins and collagen are two types of distinct systems where the origins of the response lead to different strategies. The NLO response of fluorescent proteins originates mostly from a push-pull π -conjugated chromophore whereas, for collagen, small π -conjugated peptide bonds are at the origin of the NLO response. The choice of quantum chemistry methods to describe the NLO responses of fluorescent proteins and collagen must be substantiated by investigations of the impact of electron correlation on the hyperpolarizabilities of small model systems. A first study concerns push-pull π -conjugated systems, motivated by the fact that such compounds are model structures of the chromophores of fluorescent proteins. Open-shell systems such as the *p*-quinodimethane and two of its derivatives are also used to further describe the electron correlation effects. Those compounds are quite small in terms of the number of atoms but with a strong impact of electron correlation, it needs selected methods to achieve accuracy in predicting the hyperpolarizabilities. Several technical aspects associated with the evaluation of hyperpolarizabilities have also be assessed, *i.e* i) the automatization of the Romberg scheme to improve the numerical accuracy, in the finite field method, on the

electric field derivatives of the energy, ii) the evaluation of the frequency dispersion at the correlated level by approximated schemes, and iii) the deviations from Kleinman's conditions. These studies allow mastering the methods used to evaluate the hyperpolarizabilities. Together with the characterization of the size complexity, this knowledge is crucial to reach one goal of this thesis, i.e. the development of new expertizes able to predict and characterize accurately the hyperpolarizabilities of large biological systems.

The manuscript is organized as follows. After a short review on the theoretical methods used to evaluate the hyperpolarizabilities in Chapter II, Chapters III and IV are dedicated to the biological systems, the fluorescent proteins and the collagen, respectively. Then, chapter V, deals with the model systems, i.e. the evaluation of the electron-correlation effects on the hyperpolarizabilities of push-pull π -conjugated systems and open-shell systems. Chapter VI draws conclusions and gives perspectives.

REFERENCES

1. (a) Kuzyk, M. G., *Characterization techniques and tabulations for organic nonlinear optical materials*. CRS Press: 1998; (b) Verbiest, T.; Clays, K.; Rodriguez, V., *Second-order nonlinear optical characterization techniques: An introduction*. CRC Press: New-York, 2009.
2. Campagnola, P. J.; Loew, L. M., *Nat. Biotechnol.* **2003**, 21 (11), 1356-60.
3. Papadopoulos, M. G.; Sadlej, A. J.; Leszczynski, J., *Non-linear optical properties of matter - from molecules to condensed phases*. Springer: 2006; Vol. 1.
4. (a) Mançois, F.; Sanguinet, L.; Pozzo, J. L.; Guillaume, M.; Champagne, B.; Rodriguez, V.; Adamietz, F.; Ducasse, L.; Castet, F., *J. Phys. Chem. B* **2007**, 111 (33), 9795-802; (b) Plaquet, A.; Guillaume, M.; Champagne, B.; Rougier, L.; Mancois, F.; Rodriguez, V.; Pozzo, J. L.; Ducasse, L.; Castet, F., *J. Phys. Chem. C* **2008**, 112 (14), 5638-5645; (c) Sergeyev, S.; Didier, D.; Boitsov, V.; Teshome, A.; Asselberghs, I.; Clays, K.; Vande Velde, C. M.; Plaquet, A.; Champagne, B., *Chemistry* **2010**, 16 (27), 8181-90.
5. Aubert, V.; Guerchais, V.; Ishow, E.; Hoang-Thi, K.; Ledoux, I.; Nakatani, K.; Le Bozec, H., *Angew. Chem. Int. Ed. Engl.* **2008**, 47 (3), 577-580.
6. Botek, E.; Castet, F.; Champagne, B., *Chem. Eur. J.* **2006**, 12 (34), 8687-8695.
7. Loison, C.; Simon, D., *J Phys Chem A* **2010**, 114 (29), 7769-79.
8. Willetts, A.; Rice, J. E.; Burland, D. M.; Shelton, D. P., *J. Chem. Phys.* **1992**, 97 (10), 7590-7599.
9. (a) Clays, K.; Persoons, A., *Phys. Rev. Lett.* **1991**, 66 (23), 2980-2983; (b) Hendrickx, E.; Clays, K.; Persoons, A., *Acc. Chem. Res.* **1998**, 31 (10), 675-683.
10. Bersohn, R.; Pao, Y.-H.; Frisch, H. L., *J. Chem. Phys.* **1966**, 45 (9), 3184-3198.
11. Brasselet, S.; Zyss, J., *J. Opt. Soc. Am. B* **1998**, 15 (1), 257-288.
12. (a) Reeve, J. E.; Anderson, H. L.; Clays, K., *Phys. Chem. Chem. Phys.* **2010**, 12 (41), 13484-98; (b) Abraham, T.; Hirota, J. A.; Wadsworth, S.; Knight, D. A., *Pulm. Pharmacol. Ther.* **2011**, 24 (5), 487-96; (c) Strachan, C. J.; Windbergs,

- M.; Offerhaus, H. L., *Int. J. Pharm.* **2011**, *417* (1-2), 163-72; (d) Huang, S. H.; Hsiao, C. D.; Lin, D. S.; Chow, C. Y.; Chang, C. J.; Liao, I., *PLoS One* **2011**, *6* (9), e24764; (e) Doras, C.; Taupier, G.; Barsella, A.; Mager, L.; Boeglin, A.; Bulou, H.; Bousquet, P.; Dorkenoo, K. D., *Opt. Express* **2011**, *19* (16), 15062-8; (f) Chen, W. L.; Hu, P. S.; Ghazaryan, A.; Chen, S. J.; Tsai, T. H.; Dong, C. Y., *Comput. Med. Imaging Graph.* **2012**, *36* (7), 519-26; (g) Chen, X.; Nadiarynkh, O.; Plotnikov, S.; Campagnola, P. J., *Nat. Protoc.* **2012**, *7* (4), 654-69; (h) Wu, S. L.; Li, H.; Zhang, X. M.; Chen, W. R.; Wang, Y. X., *Las. Med. Science* **2013**, 1-6.
13. Pavone, F. S.; Campagnola, P. J., *Second harmonic generation imaging*. Taylor & Francis: 2013.
 14. Campagnola, P., *Anal. Chem. (Washington, DC, U. S.)* **2011**, *83*, 3224-3231.
 15. Freund, I.; Deutsch, M.; Sprecher, A., *Biophys. J.* **1986**, *50* (4), 693-712.
 16. Roth, S.; Freund, I., *J. Chem. Phys.* **1979**, *70* (4), 1637-1643.
 17. Bouevitch, O.; Lewis, A.; Pinevsky, I.; Wuskell, J. P.; Loew, L. M., *Biophys. J.* **1993**, *65* (2), 672-679.
 18. Alain, V.; Rédoglia, S.; Blanchard-Desce, M.; Lebus, S.; Lukaszuk, K.; Wortmann, R.; Gubler, U.; Bosshard, C.; Günter, P., *Chem. Phys.* **1999**, *245* (1-3), 51-71.
 19. (a) Davydov, B. L.; Derkacheva, L. D.; Dunina, V. V.; Zhabotinskii, M. E.; Zolin, V. F.; Koreneva, L. G.; Samokhina, M. A., *JETP Lett.* **1970**, *12*, 16; (b) Kanis, D. R.; Ratner, M. A.; Marks, T. J., *Chem. Rev.* **1994**, *94* (1), 195-242.
 20. Barsu, C.; Cheaib, R.; Chambert, S.; Queneau, Y.; Maury, O.; Cottet, D.; Wege, H.; Douady, J.; Bretonniere, Y.; Andraud, C., *Org. Biomol. Chem.* **2010**, *8* (1), 142-150.
 21. (a) Salafsky, J. S., *Chem. Phys. Lett.* **2001**, *342* (5-6), 485-491; (b) Salafsky, J. S., *Phys. Chem. Chem. Phys.* **2007**, *9* (42), 5704-5711.
 22. Yan, P.; Millard, A. C.; Wei, M.; Loew, L. M., *J. Am. Chem. Soc.* **2006**, *128* (34), 11030-11031.

23. (a) Chalfie, M., *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (31), 5603-11; (b) Shimomura, O., *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (31), 5590-602; (c) Tsien, R. Y., *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (31), 5612-26.
24. (a) Lewis, A.; Khatchatouriants, A.; Treinin, M.; Chen, Z.; Peleg, G.; Friedman, N.; Bouevitch, O.; Rothman, Z.; Loew, L.; Sheres, M., *Chem. Phys.* **1999**, *245* (1-3), 133-144; (b) Khatchatouriants, A.; Lewis, A.; Rothman, Z.; Loew, L.; Treinin, M., *Biophys. J.* **2000**, *79* (5), 2345-2352.
25. Asselberghs, I.; Flors, C.; Ferrighi, L.; Botek, E.; Champagne, B.; Mizuno, H.; Ando, R.; Miyawaki, A.; Hofkens, J.; Van der Auweraer, M.; Clays, K., *J. Am. Chem. Soc.* **2008**, *130* (46), 15713-9.

II. Theoretical methods

In this chapter, we introduce the principal quantum chemistry methods used in this thesis as well as the different approaches used to determine the hyperpolarizabilities (analytical or numerical differentiations). In order to stay right to the point, only the principal ones are presented in a condensed manner. We also provide specific information about the application of these methods to our systems/compounds of interest. The way to take into account the solvent effects are also described as well as the unit sphere representation of the first hyperpolarizability tensor and the approach used in the geometry optimizations.

II.A. *Wave function-based methods*

Historically, the application of quantum mechanics to chemistry takes its origin in the 20th century with the introduction of the molecular orbital theory. Since then this theory has led to Hartree-Fock and post-Hartree-Fock methods, which are widely used in contemporary quantum chemistry. Here, the Hartree-Fock method¹ is introduced, as well as post-Hartree-Fock methods¹⁻², i.e. the Møller-Plesset perturbation theory and the coupled-cluster theory³.

II.A.1. The Hartree-Fock method

A single Slater determinant, built on spinorbitals, is the simplest antisymmetric wave function, which can be used to describe the ground state of an N -electron system:

$$|\Psi_0\rangle = |\chi_1\chi_2\cdots\chi_N\rangle. \quad (\text{II.A.1})$$

According to the variation principle, the best wave function is the one, which gives the lowest energy:

$$E_0 = \langle\Psi_0|\hat{H}|\Psi_0\rangle \quad (\text{II.A.2})$$

where $\hat{H}$ is the full electronic Hamiltonian:

$$\hat{H} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}}, \quad (\text{II.A.3})$$

where i and j are electron indexes, up to N electrons, and A is the atom index, up to M atoms. Expression (II.A.2) can be rewritten as

$$E_0 = \sum_a \langle a|h|a \rangle + \frac{1}{2} \sum_{ab} \langle ab||ab \rangle \quad (\text{II.A.4})$$

where $\langle a|h|a \rangle = \langle \chi_a|h|\chi_a \rangle = \int d\vec{x}_1 \chi_a^*(\vec{x}_1) \left[-\frac{1}{2} \nabla_1^2 - \sum_{A=1}^M \frac{Z_A}{r_{1A}} \right] \chi_a(\vec{x}_1)$ is a one-electron integral over spinorbitals, and

$$\begin{aligned} \langle ab||ab \rangle &= \langle ab|ab \rangle - \langle ab|ba \rangle \\ &= \int d\vec{x}_1 d\vec{x}_2 \chi_a^*(\vec{x}_1) \chi_b^*(\vec{x}_2) \frac{1}{r_{12}} \chi_a(\vec{x}_1) \chi_b(\vec{x}_2) - \int d\vec{x}_1 d\vec{x}_2 \chi_a^*(\vec{x}_1) \chi_b^*(\vec{x}_2) \frac{1}{r_{12}} \chi_b(\vec{x}_1) \chi_a(\vec{x}_2) \end{aligned}$$

is an antisymmetrized two-electron integral. Minimizing E_0 with respect to the choice of spinorbitals $\chi_a(\vec{x}_i)$, leads to the Hartree-Fock equation, which determines the optimal spinorbitals with the constraint that $\langle \chi_a|\chi_b \rangle = \delta_{ab}$. The Hartree-Fock equation is an eigenvalue equation of the form

$$f(i)|\chi_a(\vec{x}_i)\rangle = \epsilon_a |\chi_a(\vec{x}_i)\rangle \quad (\text{II.A.5})$$

where $f(i)$ is an effective one-electron operator, called the Fock operator, of the form

$$f(i) = \underbrace{-\frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}}}_{h(i)} + v^{\text{HF}}(i). \quad (\text{II.A.6})$$

The Hartree-Fock potential $v^{\text{HF}}(i)$ is an effective one electron potential operator:

$$v^{\text{HF}}(i) = \sum_b^{\text{occupied spinorbitals}} J_b(i) - K_b(i) \quad (\text{II.A.7})$$

where $J_b(\vec{x}_1) \chi_a(\vec{x}_1) = \int d\vec{x}_2 |\chi_b(\vec{x}_2)|^2 \frac{1}{r_{12}} \chi_a(\vec{x}_1)$ is the coulomb operator,

and $K_b(\vec{x}_1)\chi_a(\vec{x}_1) = \int d\vec{x}_2 \chi_b^*(\vec{x}_2) \frac{1}{r_{12}} \chi_a(\vec{x}_2) \chi_b(\vec{x}_1)$ is the exchange operator.

For a closed-shell system where the spinorbitals have the form:

$$\chi_a(\vec{x}) = \begin{cases} \psi_a(\vec{r})\alpha(\omega) \\ \psi_a(\vec{r})\beta(\omega) \end{cases}, \quad (\text{II.A.8})$$

the Fock operator read

$$f(i) = h(i) + \sum_{a=1}^{N/2} 2J_a(i) - K_a(i), \quad (\text{II.A.9})$$

where $N/2$ molecular orbitals are occupied. The calculation of molecular orbitals becomes equivalent to solving the spatial integro-differential equation

$$f(i)|\psi_a(\vec{r}_i)\rangle = \varepsilon_a|\psi_a(\vec{r}_i)\rangle. \quad (\text{II.A.10})$$

The Roothaan equations are solved by introducing a set of K atomic orbitals linearly expanded to form the molecular orbitals basis set:

$$|\psi_a\rangle = \sum_{\mu=1}^K C_{\mu a} |\phi_\mu\rangle. \quad (\text{II.A.11})$$

If the overlap matrix $\mathbf{S}$ has for elements $S_{\mu\nu} = \int d\vec{r}_1 \phi_\mu^*(\vec{r}_1) \phi_\nu(\vec{r}_1)$ and the Fock matrix $\mathbf{F}$ has for elements $F_{\mu\nu} = \int d\vec{r}_1 \phi_\mu^*(\vec{r}_1) f(1) \phi_\nu(\vec{r}_1)$, the Roothaan equations take the form:

$$\sum_{\nu=1}^K F_{\mu\nu} C_{\nu a} = \varepsilon_a \sum_{\nu=1}^K S_{\mu\nu} C_{\nu a} \quad a = 1, 2, \dots, K, \quad (\text{II.A.12})$$

which can be written compactly under the form of a single matrix equation:

$$\mathbf{FC} = \mathbf{SC}\varepsilon. \quad (\text{II.A.13})$$

Since the Fock operator is expressed for a closed-shell system as (II.A.9), an element of the Fock matrix is given by this expression:

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda\sigma} D_{\lambda\sigma} [2(\mu\nu|\sigma\lambda) - (\mu\lambda|\sigma\nu)] \quad (\text{II.A.14})$$

where $H_{\mu\nu}^{\text{core}} = \langle \phi_\mu(\vec{r}_1) | h(1) | \phi_\nu(\vec{r}_1) \rangle$, the element $D_{\lambda\sigma}$ of the density matrix $\mathbf{D}$ reads

$$D_{\lambda\sigma} = \sum_{a=1}^{N/2} C_{\lambda a} C_{\sigma a}^* , \quad (\text{II.A.15})$$

and $(\mu\nu|\sigma\lambda) = \int d\vec{r}_1 d\vec{r}_2 \phi_{\mu}^*(\vec{r}_1) \phi_{\nu}(\vec{r}_1) \frac{1}{r_{12}} \phi_{\sigma}^*(\vec{r}_2) \phi_{\lambda}(\vec{r}_2)$. Since the Fock operator depends on the density matrix, the Hartree-Fock procedure needs to be solved iteratively. For a finite basis set, it is known as the self-consistent-field (SCF) procedure.

For an open-shell system, where the unrestricted set of spinorbitals reads

$$\chi_a(\vec{x}) = \begin{cases} \psi_a^{\alpha}(\vec{r})\alpha(\omega) \\ \psi_a^{\beta}(\vec{r})\beta(\omega) \end{cases} , \quad (\text{II.A.16})$$

the Fock operator has this form for the two sets of spatial orbitals:

$$\begin{cases} f^{\alpha}(i) = h(i) + \sum_{a=1}^{N^{\alpha}} [J_a^{\alpha}(i) - K_a^{\alpha}(i)] + \sum_{a=1}^{N^{\beta}} J_a^{\beta}(i) \\ f^{\beta}(i) = h(i) + \sum_{a=1}^{N^{\beta}} [J_a^{\beta}(i) - K_a^{\beta}(i)] + \sum_{a=1}^{N^{\alpha}} J_a^{\alpha}(i) \end{cases} \quad (\text{II.A.17})$$

The calculation of the two sets of molecular orbitals becomes equivalent to solving two coupled unrestricted Hartree-Fock equations:

$$\begin{cases} f^{\alpha}(i) |\psi_a^{\alpha}(\vec{r}_i)\rangle = \epsilon_a^{\alpha} |\psi_a^{\alpha}(\vec{r}_i)\rangle \\ f^{\beta}(i) |\psi_a^{\beta}(\vec{r}_i)\rangle = \epsilon_a^{\beta} |\psi_a^{\beta}(\vec{r}_i)\rangle \end{cases} . \quad (\text{II.A.18})$$

Then, by the introduction of the molecular orbitals basis set, the unrestricted Pople-Nesbet's equations are obtained in a same way as the Roothaan's equations and are solved iteratively.

II.A.2. The Møller-Plesset perturbation theory

In our quest for a suitable method to determine accurately the hyperpolarizabilities, the Møller-Plesset perturbation theory has been used and tested with respect to more accurate correlated methods, presently coupled-cluster based methods. In this section, the outline of the Møller-Plesset approach is given.

In order to improve the Hartree-Fock energy of an N -electron system by the Møller-Plesset perturbation theory¹⁻², a perturbation expansion of the correlation energy is accounted. First, the Hamiltonian is partitioned according to:

$$\hat{H} = \sum_{i=1}^N h(i) + \sum_{i,j, j>i}^N \frac{1}{r_{ij}} = \hat{H}_0 + \hat{V}, \quad (\text{II.A.19})$$

where $\hat{H}_0$ is the Hartree-Fock Hamiltonian,

$$\hat{H}_0 = \sum_{i=1}^N f(i) = \sum_{i=1}^N \left(h(i) + v^{HF}(i) \right). \quad (\text{II.A.20})$$

The electron correlation operator, $\hat{V}$ is therefore the difference between the exact Hamiltonian and the Hartree-Fock Hamiltonian operator,

$$\begin{aligned} \hat{V} &= \hat{H} - \hat{H}_0 \\ &= \sum_{i=1}^N h(i) + \sum_{i,j, j>i}^N \frac{1}{r_{ij}} - \sum_{i=1}^N \left(h(i) + v^{HF}(i) \right) . \\ &= \sum_{i,j, j>i}^N \frac{1}{r_{ij}} - \sum_{i=1}^N v^{HF}(i) \end{aligned} \quad (\text{II.A.21})$$

Following Rayleigh-Schrödinger perturbation theory, the first-order energy is given by:

$$\langle \Psi_0 | \hat{V} | \Psi_0 \rangle = \langle \Psi_0 | \sum_{i,j, j>i}^N \frac{1}{r_{ij}} | \Psi_0 \rangle - \langle \Psi_0 | \sum_{i=1}^N v^{HF}(i) | \Psi_0 \rangle, \quad (\text{II.A.22})$$

or

$$\langle \Psi_0 | \hat{V} | \Psi_0 \rangle = -\frac{1}{2} \sum_{ab}^{\text{occupied spinorbitals}} \langle ab || ab \rangle. \quad (\text{II.A.23})$$

The Hartree-Fock energy is the sum of the zeroth and first-order energies:

$$E_{HF} = E_0^{(0)} + E_0^{(1)} = \sum_a \epsilon_a - \frac{1}{2} \sum_{ab}^{\text{occupied spinorbitals}} \langle ab || ab \rangle. \quad (\text{II.A.24})$$

Thus, the first correction to the Hartree-Fock energy occurs at the second order of the perturbation theory, which is called the MP2 correction. From the Rayleigh-

Schrödinger perturbation theory, the general expression for the second-order energy is

$$E_0^{(2)} = \sum_{n \neq 0} \frac{|\langle 0 | \hat{V} | n \rangle|^2}{E_0^{(0)} - E_n^{(0)}}, \quad (\text{II.A.25})$$

where the summation runs over all the states of the system except the ground state. At the Hartree-Fock level of theory, any electronic system is described by a set of occupied spin orbitals where there are no excitations since the electrons do not interact explicitly. The correlation interaction between the electrons can be associated with the virtual excitations of electrons from occupied to unoccupied spin orbitals with some probability. Thus, the excited determinants include, in principle, the singles, doubles, triples,... For any single excitation where an electron is excited from the occupied spinorbital a to the unoccupied spinorbital r , $\langle \Psi_0 | \hat{V} | \Psi_a^r \rangle$ is equal to zero because of the Brillouin's theorem. Triple excited states do not mix either with the ground state Slater determinant because of the two-particle nature of the perturbation. So, in (II.A.25) only double excitations remain:

$$\langle 0 | \hat{V} | n \rangle = \langle \Psi_0 | \hat{V} | \Psi_{ab}^{rs} \rangle = \langle \Psi_0 | \sum_{i,j} \frac{1}{r_{ij}} | \Psi_{ab}^{rs} \rangle = \langle ab || rs \rangle. \quad (\text{II.A.26})$$

where two electrons are excited from two occupied spinorbitals a and b to two unoccupied spinorbitals r and s . The second-order energy takes the form

$$E_0^{(2)} = \sum_{\substack{a < b \\ r < s}} \frac{|\langle ab || rs \rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s}, \quad (\text{II.A.27})$$

whereas the MP2 energy reads

$$E_{MP2} = E_{HF} + E_0^{(2)}. \quad (\text{II.A.28})$$

The same scheme is used to obtain higher-order MPn corrections. At the MP4 level, single, double, triple and quadruple excited Slater determinants are involved in the energy expression. In the same way, the Møller-Plesset perturbation theory can also be applied to the unrestricted case.

The MP2 method can now be routinely applied to systems with a hundred of atoms or more, which opens the way to investigate systems of interest for NLO applications. This approach was favored in several recent investigations combining measurements and simulations⁴ because for model systems ($\text{NH}_2\text{-CH=CH-CH=CH-NO}_2$ and $\text{NH}_2\text{-CH=CH-CH=CH-CH=CH-NO}_2$), the MP2 scheme provides first hyperpolarizability values in close agreement with results obtained at the CCSD(T) level of theory⁵. The CCSD(T) method is introduced in the next section. Note that such MP2/CC comparisons also raise the question of the basis set effects. In Ref. 5, the relationships between the MP2 and CCSD(T) results was found rather similar when using the 6-31G(d), 6-31+G(d), cc-pVDZ, and aug-cc-pVDZ basis sets. This is however not always the case, in particular, for smaller compounds as shown in Ref. 6.

II.A.3. Coupled Cluster method

In this section, an introduction to the basic ideas of the coupled-cluster theory³ is presented. This method has been widely used as reference method to determine the hyperpolarizabilities of model systems. This introduction is made in the second quantization formalism.

At the coupled cluster level of theory, the excitation probability between the Hartree-Fock determinant and an excited determinant is associated to an amplitude. The notation $\hat{\tau}_\rho$ is used for the excitation operator of unspecified excitation level [single ($\hat{\tau}_a^r = a_r^\dagger a_a$), double ($\hat{\tau}_{ab}^{rs} = a_r^\dagger a_a^\dagger a_s a_b$), ...] and t_ρ for the corresponding amplitude, where $a_r^\dagger$ is the creation operator of an electron in a spinorbital r and a_a is the annihilation operator of an electron in a spinorbital a . The effect of a given (single or multiple) excitation ρ on any state $|\rangle$ changes the state as follows

$$|\rangle \rightarrow (1 + t_\rho \hat{\tau}_\rho) |\rangle. \quad (\text{II.A.29})$$

If the correlation operator is applied onto the Hartree-Fock ground state wavefunction, the coupled-cluster wave function is obtained:

$$|CC\rangle = \left[\prod_{\rho} (1 + t_{\rho} \hat{\tau}_{\rho}) \right] |\Psi_0\rangle. \quad (\text{II.A.30})$$

This product form leads to a size extensive treatment of the electronic system. Now, the coupled-cluster wave function is analyzed in terms of determinants where

$$|\Psi_{\rho}\rangle = \hat{\tau}_{\rho} |\Psi_0\rangle. \quad (\text{II.A.31})$$

The product operator is expanded in (II.A.30) as

$$|CC\rangle = |\Psi_0\rangle + \sum_{\rho} t_{\rho} |\Psi_{\rho}\rangle + \sum_{\rho > \tau} t_{\rho} t_{\tau} |\Psi_{\rho\tau}\rangle + \dots \quad (\text{II.A.32})$$

In the expansion, the operator $\hat{\tau}_{\rho}$ not only generates the determinant $|\Psi_{\rho}\rangle$ in the second term but also a large number of determinants in the higher-order terms. The presence of product excitations in the coupled-cluster state leads to excited determinants, which can be reached in several distinct ways. For example a double excitation operator is equivalent to two separate single excitations

$$\hat{\tau}_{ab}^{rs} = \hat{\tau}_a^r \hat{\tau}_b^s. \quad (\text{II.A.33})$$

The amplitude of this process is the sum of the individual amplitudes

$$t_{ab}^{rs} \text{ total} = t_{ab}^{rs} + t_a^r t_b^s + t_b^s t_a^r \quad (\text{II.A.34})$$

The amplitude t_{ab}^{rs} is referred to a connected cluster and $t_a^r t_b^s$ as a disconnected cluster. In general higher-order excitations can be reached by a larger number of processes or mechanisms, each contributing to the total amplitude. Note that the spinorbital excitation operators satisfy the relation

$$\hat{\tau}_{\rho}^2 = 0. \quad (\text{II.A.35})$$

Thus, the correlating operators can be written as exponential of the excitation operators since all higher-order terms are cancelled via (II.A.35):

$$1 + t_{\rho} \hat{\tau}_{\rho} = e^{t_{\rho} \hat{\tau}_{\rho}}. \quad (\text{II.A.36})$$

The coupled-cluster equation is rewritten under the form of an exponential ansatz

$$|\text{CC}\rangle = e^{\hat{T}} |\Psi_0\rangle, \quad (\text{II.A.37})$$

where the cluster operator is

$$\hat{T} = \sum_{\rho} t_{\rho} \hat{\tau}_{\rho}. \quad (\text{II.A.38})$$

As an approximation the cluster operator is partitioned into classes, comprising all single excitations, all double excitations, all triple excitations, and so on. The cluster operator takes the form

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_N. \quad (\text{II.A.39})$$

A way to determine the coupled-cluster state is to minimize the expectation value of the Hamiltonian with respect to the amplitudes

$$E_{\min} = \min_{t_{\rho}} \frac{\langle \text{CC} | \hat{H} | \text{CC} \rangle}{\langle \text{CC} | \text{CC} \rangle}. \quad (\text{II.A.40})$$

However, the variational minimization of the coupled-cluster energy is difficult and can be carried out only for small molecular systems.³ As an alternative to this minimization, the projected Schrödinger equations (including the coupled-cluster energy) are used. It consists to project the Schrödinger equation against the Hartree-Fock state as well as the excited determinants. The resulting projected coupled-cluster equations read

$$\langle \Psi_0 | \hat{H} e^{\hat{T}} | \Psi_0 \rangle = E, \quad (\text{II.A.41})$$

$$\langle \Psi_{\rho} | \hat{H} e^{\hat{T}} | \Psi_0 \rangle = E \langle \Psi_{\rho} | e^{\hat{T}} | \Psi_0 \rangle \quad (\text{II.A.42})$$

since $\langle \Psi_0 | \text{CC} \rangle = 1$. Equations (II.A.41) and (II.A.42) are known as the “unlinked coupled-cluster equations”. The equations (II.A.42) must be solved iteratively, substituting in each iteration the coupled-cluster energy as calculated from (II.A.41). In a full coupled-cluster calculation, the CC wave function recovers the full CI wave function. It is convenient to express the projected coupled-cluster equations in a slightly different form by multiplying the Schrödinger equation by the $e^{-\hat{T}}$ operator to obtain

$$e^{-\hat{T}} \hat{H} e^{\hat{T}} |\Psi_0\rangle = E |\Psi_0\rangle, \quad (\text{II.A.43})$$

where the projection of Schrödinger equations gives this set of equations:

$$\langle \Psi_0 | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Psi_0 \rangle = E, \quad (\text{II.A.44})$$

$$\langle \Psi_p | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Psi_0 \rangle = 0, \quad (\text{II.A.45})$$

which is equivalent to the untransformed set [(II.A.41) and (II.A.42)], yielding the same amplitude and the same energy upon solution. These two equations are known as the “linked coupled-cluster equations”.

The advantages of the cluster parameterization become apparent when partitioning and truncating the cluster operator. For a given accuracy (or amount of correlation), it leads to a less resource consuming method. The approximate coupled-cluster state still contains contributions from all the possible determinants, with weight obtained from the different excitation processes leading to these determinants. The most common approximation is to truncate the cluster operator at the doubles level, which leads to the coupled cluster singles-and-doubles (CCSD) model. This scheme includes highly excited determinants generated by disconnected clusters but they are less accurate than those that also contain connected clusters. For example, the $\hat{T}_4$ and $\hat{T}_3\hat{T}_1$ contributions to the quadruples are neglected at the CCSD level, but these determinants are still quite well described by the $\hat{T}_2^2$ contributions. For higher accuracy, the connected triple excitations must be taken into account by truncating the cluster expansion at the $\hat{T}_3$ level but this becomes computationally very demanding and can be applied only to small systems. A different way to include connected triple excitations at lower cost is to obtain the triples energy correction by a perturbative estimation. This method is called CCSD(T).

Nowadays, for small polyatomic molecules, the CCSD(T) level of theory gives the most efficient and accessible description of β but, in general, it cannot be routinely applied to large systems. All along this thesis the CCSD(T) level of theory is

used as reference method to estimate the accuracy of other methods with lower levels of approximation. The unrestricted CCSD and CCSD(T) methods have also been used during this thesis, see reference 7 for more information.

II.B. *Density functional theory*

In this thesis, all the geometry optimizations were performed using the density functional theory approach (DFT) while it was also cautiously employed to determine the hyperpolarizabilities. In this section, the DFT^{8,9} foundations are presented. Its formulation relies on the electronic density $\rho(\vec{r})$, which reads

$$\rho(\vec{r}) = N \int \dots \int |\Psi(\vec{x}_1, \dots, \vec{x}_N)|^2 d\vec{x}_1 d\vec{x}_2 \dots d\vec{x}_N. \quad (\text{II.A.46})$$

It determines, for state Ψ , the probability of finding any of the N electrons within the volume element $d\vec{r}$ but with arbitrary spin while the other $N-1$ electrons have arbitrary positions and spins. Unlike the wave function, the electron density is an observable and can be measured experimentally.

The first theorem of Hohenberg-Kohn¹⁰ states the proof of existence: the ground state density $\rho(\vec{r})$ of a bound system of interacting electrons in some external potential $V(\vec{r})$ determines this potential uniquely. Therefore, the ground state energy is a functional of the ground state electron density ρ_0 , and it reads:

$$E_0[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0], \quad (\text{II.A.47})$$

where T is the kinetic energy functional, E_{ee} is the electron-electron interaction energy functional and E_{Ne} is the potential energy due to the nuclei-electron attraction, known as

$$E_{Ne}[\rho_0] = \int \rho_0(\vec{r}) V_{Ne}(\vec{r}) d\vec{r}. \quad (\text{II.A.48})$$

Regrouping the system independent parts into a new quantity, the *Hohenberg-Kohn functional* $F_{HK}[\rho_0]$, it reads

$$E_0[\rho_0] = F_{HK}[\rho_0] + \int \rho_0(\vec{r}) V_{Ne}(\vec{r}) d\vec{r}; \quad (\text{II.A.49})$$

where the Hohenberg-Kohn functional is expressed as

$$F_{HK}[\rho] = T[\rho] + E_{ee}[\rho] = \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle. \quad (\text{II.A.50})$$

This functional is the Holy Grail of DFT. If it was known exactly the Schrödinger equation would be solved exactly. But unfortunately, the explicit form of this functional remains unknown. However, the classical Coulomb part $J[\rho]$ of $E_{ee}[\rho]$ is well known as

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2. \quad (\text{II.A.51})$$

The remaining non-classical contribution $E_{ncl}[\rho]$ contains the effects of self-interaction correction, exchange and Coulomb correlation.

The second Hohenberg-Kohn theorem states the variational principle: $F_{HK}[\rho_0]$ delivers the ground state energy of the system with the lowest energy, if the input density is the true ground state density ρ_0 :

$$E_0 \leq E[\tilde{\rho}] = T[\tilde{\rho}] + E_{ee}[\tilde{\rho}] + E_{ncl}[\tilde{\rho}]. \quad (\text{II.A.52})$$

Then comes the idea of Kohn and Sham¹¹ who realized, that if it is not possible to accurately determine the kinetic energy through an explicit functional, one should be a little bit less ambitious and compute as much as it is possible of the true kinetic energy exactly and then have to deal with the remainder in an approximate manner. Thus, the kinetic energy of the non-interacting reference system with the same density as the real interacting one reads

$$T_s = -\frac{1}{2} \sum_a^N \langle \psi_a | \nabla^2 | \psi_a \rangle, \quad (\text{II.A.53})$$

where the $|\psi_a\rangle$ are the Kohn-Sham orbitals. The ground state wave function is represented by a Slater determinant. In order to account for that, Kohn and Sham introduced the following separation of the functional $F[\rho]$ as

$$F[\rho(\vec{r})] = T_s[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{xc}[\rho(\vec{r})]. \quad (\text{II.A.54})$$

The residual, correlation, part of the true kinetic energy T_c is simply added to the non-classical electrostatic contributions to form the exchange-correlation functional

$$E_{XC}[\rho(\vec{r})] = T_C[\rho(\vec{r})] + E_{ncl}[\rho(\vec{r})]. \quad (\text{II.A.55})$$

Similarly to the Hartree-Fock approximation, by applying the variational principle under the usual constraint that the orbitals are orthonormalized, the one-particle Kohn-Sham equation is deduced. It reads

$$\left(-\frac{1}{2} \nabla_1^2 + \int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 + V_{XC}(\vec{r}_1) - \sum_A^M \frac{Z_A}{r_{1A}} \right) |\psi_a\rangle = \varepsilon_a |\psi_a\rangle. \quad (\text{II.A.56})$$

V_{XC} is simply defined as

$$V_{XC} \equiv \frac{\delta E_{XC}}{\delta \rho}. \quad (\text{II.A.57})$$

However, the exchange-correlation functional E_{XC} is unknown. In order to obtain exact results, the exact functional is needed. The exchange-correlation functional has been approximated in different ways over the last decades.

By assuming the local density approximation (LDA), the exchange-correlation functional is defined as

$$E_{XC}^{LDA}[\rho] = \int \rho(\vec{r}) \varepsilon_{XC}[\rho(\vec{r})] d\vec{r}, \quad (\text{II.A.58})$$

where $\varepsilon_{XC}[\rho(\vec{r})]$ is the exchange-correlation energy per particle of a uniform electron gas of density $\rho(\vec{r})$. This quantity can be further split into exchange and correlation contributions. In the LDA approximation a typical overestimation of the bond energies and underestimation of bond lengths is observed.^{9b} Improvement can be made by assuming the generalized gradient approximation (GGA), where the exchange-correlation functional is defined with respect to the electronic density as well as to its gradient:

$$E_{XC}^{GGA}[\rho] = \int \rho(\vec{r}) \varepsilon_{XC}[\rho(\vec{r}), \vec{\nabla} \rho(\vec{r})] d\vec{r}. \quad (\text{II.A.59})$$

Using the expressions of Becke¹² for the exchange part and of Lee-Yang-Parr¹³ for the correlation one, BLYP, one of the best known GGA functional, is obtained. The GGA gives excellent atomic ground-state energies and improves significantly dissociation energies and bond lengths.^{9b}

The natural improvement of GGA is then to include the second-order derivative of the electron density, leading to the original form of the meta-GGA exchange-correlation functionals¹⁴. Nowadays, meta-GGA functional rather refers to the inclusion of the kinetic energy density:

$$\tau(\vec{r}) = \sum_a^{\text{occ.}} \frac{1}{2} |\nabla \psi_a(\vec{r})|^2 \quad (\text{II.A.60})$$

In order to include the correct asymptotic $-1/r$ behavior for the exchange part, which is not the case for LDA and GGA functionals, a given percentage of exact Hartree-Fock exchange is added to GGA functionals, which leads the so-called hybrid functionals. In this thesis, the BHandHLYP functional¹² has been used where 50% of exact HF exchange is included and the popular and efficient B3LYP functional¹⁵ with 20% of HF exchange, which depends on three parameters as

$$E_{XC}^{B3LYP} = (1-a)E_X^{HF} + aE_X^{Slater} + bE_X^{Becke} + cE_C^{LYP} + (1-c)E_C^{VWN}, \quad (\text{II.A.61})$$

where $a=0.8$, $b=0.72$ and $c=0.81$ are constants determined by Becke¹⁶ via fitting over 56 atomization energies, 42 ionization potentials, 8 proton affinities, and 10 total atomic energies of first- and second-row systems. Recently, long-range corrected functionals¹⁷ have been proposed, where the electron repulsion operator $1/r_{12}$ is split into short- and long-range parts by using a standard error function:

$$\frac{1}{r_{12}} = \frac{1 - \text{erf}(\mu r_{12})}{r_{12}} + \frac{\text{erf}(\mu r_{12})}{r_{12}}, \quad (\text{II.A.62})$$

where μ is the only (additional) parameter. The first term accounts for the short-range interaction and is combined with GGA exchange while the second term accounts for the long-range interaction and is combined with Hartree-Fock exchange. CAM-B3LYP¹⁸ and LC-BLYP¹⁹ are standard range-separated functionals, which contain 65% and 100% of HF exchange at large inter-electronic distances. In the LC-BLYP functional^{19a}, the initial μ value was set to 0.33 but it has been redetermined and set to $\mu = 0.47$ later^{19b}. After improving the exchange part of the functional, the correlation part can also be improved. One of these strategies

consists in adding MP2 correlation to the pure DFT correlation, leading therefore to double hybrid XC functionals as in the B2-PLYP functional²⁰ (53% E_X^{GGA} , 47% E_X^{HF} , 73% and 27% E_C^{MP2}) and mPW2-PLYP²¹ (55% E_X^{GGA} , 45% E_X^{HF} , 75% E_C^{GGA} and 25% E_C^{MP2}). The performance of several of these XC functionals to calculate the hyperpolarizabilities of our compounds of interest constitutes the core of Chapter 5.

For open-shell compounds, similarly to the unrestricted Hartree-Fock approach, the unrestricted Kohn-Sham picture can also be derived by considering that the total density is decomposed as

$$\rho(\vec{r}) = \rho_\alpha(\vec{r}) + \rho_\beta(\vec{r}). \quad (\text{II.A.63})$$

In that case, spin density functionals that explicitly depend on the α – and β – spin densities are usually employed.

II.C. *Second order nonlinear optical properties*

On the basis of these different levels of approximation, several methods are used to characterize the responses to external electric fields and to evaluate the (hyper)polarizabilities. They are presented in this section.

II.C.1. Time-dependent response theory for an exact state

In this section, all the methods are derived from the time-dependent perturbation theory. This subsection describes the very basics of the perturbation theory for an exact state²².

For the exact wave function $|\Psi_{\text{exact}}(t)\rangle$, the time-dependent Schrödinger equation reads

$$\hat{H}|\Psi_{\text{exact}}(t)\rangle = i\frac{\partial}{\partial t}|\Psi_{\text{exact}}(t)\rangle, \quad (\text{II.B.1})$$

where the Hamiltonian $\hat{H}$ is partitioned into a time-independent term $\hat{H}_0$ describing the unperturbed system and $\hat{V}(t)$, a time-dependent perturbation, which is switched on adiabatically at $t = -\infty$. The perturbation operator reads

$$\hat{V}(t) = \int_{-\infty}^{\infty} d\omega \hat{V}(\omega) e^{(-i\omega + \varepsilon)t} \quad (\text{II.B.2})$$

for all the frequencies, where ε is a small positive infinitesimal quantity that ensures the perturbation is switched on adiabatically and vanishes for $t = -\infty$. This operator is Hermitian. At a finite time t , the perturbed wave function is written as a perturbation expansion,

$$\begin{aligned} |\Psi_{\text{exact}}(t)\rangle &= |\Psi(0)\rangle + \int_{-\infty}^{\infty} d\omega |\Psi(\omega)\rangle e^{(-i\omega + \varepsilon)t} \\ &+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\omega_1 d\omega_2 |\Psi(\omega_1, \omega_2)\rangle e^{(-i(\omega_1 + \omega_2) + 2\varepsilon)t} + \dots \end{aligned} \quad (\text{II.B.3})$$

where $|\Psi(\omega)\rangle$ contains all the linear terms, and $|\Psi(\omega_1, \omega_2)\rangle$ contains all the quadratic terms. However, in response theory, it is not the corrections to the wave function which are of most interest but the time development of the observables.

The expectation values of an operator $\hat{A}$ are expanded as

$$\begin{aligned} \langle \Psi_{\text{exact}}(t) | \hat{A} | \Psi_{\text{exact}}(t) \rangle &= \langle \Psi(0) | \hat{A} | \Psi(0) \rangle + \int_{-\infty}^{\infty} d\omega e^{(-i\omega + \varepsilon)t} \left\langle \left\langle \hat{A}; \hat{V}(\omega) \right\rangle \right\rangle_{\omega} \\ &+ \frac{1}{2!} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\omega_1 d\omega_2 e^{(-i(\omega_1 + \omega_2) + 2\varepsilon)t} \left\langle \left\langle \hat{A}; \hat{V}(\omega_1), \hat{V}(\omega_2) \right\rangle \right\rangle_{\omega_2} \\ &+ \dots \end{aligned} \quad (\text{II.B.4})$$

where all the linear terms are contained into the linear response function $\left\langle \left\langle \hat{A}; \hat{V}(\omega) \right\rangle \right\rangle_{\omega}$ and the quadratic ones into the quadratic response function $\left\langle \left\langle \hat{A}; \hat{V}(\omega_1), \hat{V}(\omega_2) \right\rangle \right\rangle_{\omega_2}$. These response functions determine the changes of observables when the system is subject to one or several perturbations. If the perturbation is an external oscillating electric field expressed as a Fourier expansion

$$\vec{F}(\omega) = \sum_{\omega} \vec{f}(\omega) \underbrace{\left[e^{-i\omega t} + e^{i\omega t} \right]}_{e^{\pm i\omega t}}, \quad (\text{II.B.5})$$

the time-dependent perturbation reads

$$\hat{V}(t) = - \sum_{\omega} \sum_i^{xyz} \hat{\mu}_i(t) f_i(\omega) e^{\pm i\omega t}. \quad (\text{II.B.6})$$

The expectation value of the electric dipole moment at time t is expressed as

$$\begin{aligned} \langle \Psi_{\text{exact}}(t) | \hat{\mu}_i(t) | \Psi_{\text{exact}}(t) \rangle &= \mu_i(0) + \sum_{\omega} \sum_j^{xyz} \alpha_{ij}(-\omega; \omega) f_j(\omega) e^{\pm i\omega t} \\ &+ \frac{1}{2!} \sum_{\omega_1, \omega_2} \sum_{jk}^{xyz} \beta_{ijk}(-\omega_{\sigma}; \omega_1, \omega_2) f_j(\omega_1) f_k(\omega_2) e^{\pm i\omega_{\sigma} t} + \dots \end{aligned} \quad (\text{II.B.7})$$

where $\omega_{\sigma} = \sum_i \omega_i$, the polarizability component $\alpha_{ij}(-\omega; \omega)$ corresponds to the

linear response function $-\langle\langle \mu_i(-\omega); \mu_j(\omega) \rangle\rangle_{\omega}$ to an external electric field and the

first hyperpolarizability component $\beta_{ijk}(-\omega_{\sigma}; \omega_1, \omega_2)$ is a quadratic response

function $-\langle\langle \mu_i(-\omega_{\sigma}); \mu_j(\omega_1), \mu_k(\omega_2) \rangle\rangle_{\omega_1, \omega_2}$. The components of the polarizability

and first hyperpolarizability tensors are defined as the first and second derivatives of the electric dipole moment with respect to the Fourier components:

$$\alpha_{ij}(-\omega; \omega) = \left(\frac{\partial \mu_i}{\partial F_j(\omega)} \right)_{\vec{F}=0}, \quad (\text{II.B.8})$$

$$\beta_{ijk}(-\omega_{\sigma}; \omega_1, \omega_2) = \left(\frac{\partial^2 \mu_i}{\partial F_j(\omega_1) \partial F_k(\omega_2)} \right)_{\vec{F}=0}. \quad (\text{II.B.9})$$

II.C.2. Time-dependent Hartree-Fock method

At the Hartree-Fock level, the responses to a dynamic external electric field are evaluated via the time-dependent Hartree-Fock (TDHF) method.²³ Since at the Hartree-Fock level, the dipole moment components read

$$\begin{aligned}
\mu_i &= \langle \Psi_0 | \hat{\mu}_i | \Psi_0 \rangle = -\langle \Psi_0 | \sum_{k=1}^N i_k | \Psi_0 \rangle + \sum_A^M Z_A i_A \\
&= -2 \sum_{\mu, \nu}^K M_{\mu\nu}(i) D_{\nu\mu} + \sum_A^M Z_A i_A
\end{aligned} \tag{II.B.10}$$

where $M_{\mu\nu}(i) = \langle \phi_\mu | i | \phi_\nu \rangle$ is an element of the electric dipole moment matrix along the i axis in the atomic orbitals basis. The time-dependent perturbation expansion of the electric dipole moment reads

$$\langle \Psi(t) | \hat{\mu}_i(t) | \Psi(t) \rangle = -2 \left[\sum_{\mu, \nu}^K M_{\mu\nu}(i) D_{\nu\mu} + \sum_{\omega} \sum_j^{xyz} \sum_{\mu, \nu}^K M_{\mu\nu}(i) D_{\nu\mu}^{(j)}(\omega) f_j(\omega) e^{\pm i\omega t} \right. \\
\left. + \frac{1}{2!} \sum_{\omega_1, \omega_2} \sum_{jk}^{xyz} \sum_{\mu, \nu}^K M_{\mu\nu}(i) D_{\nu\mu}^{(j)(k)}(\omega_1, \omega_2) f_j(\omega_1) f_k(\omega_2) e^{\pm i\omega_\sigma t} + \dots \right] \tag{II.B.11}$$

Thus, the linear and the quadratic response functions read

$$\alpha_{ij}(-\omega; \omega) = -\langle \langle \mu_i(-\omega); \mu_j(\omega) \rangle \rangle_\omega = -2 \sum_{\mu, \nu}^K M_{\mu\nu}(i) D_{\nu\mu}^{(j)}(\omega), \tag{II.B.12}$$

$$\begin{aligned}
\beta_{ijk}(-\omega_\sigma; \omega_1, \omega_2) &= -\langle \langle \mu_i(-\omega_\sigma); \mu_j(\omega_1), \mu_k(\omega_2) \rangle \rangle_{\omega_1, \omega_2} \\
&= -2 \sum_{\mu, \nu}^K M_{\mu\nu}(i) D_{\nu\mu}^{(j)(k)}(\omega_1, \omega_2)
\end{aligned} \tag{II.B.13}$$

where $D_{\mu\nu}^{(j)}(\omega)$ and $D_{\mu\nu}^{(j)(k)}(\omega_1, \omega_2)$ are respectively components of the first- and the second-order derivatives of the density matrix with respect to the external oscillating fields oriented along the i and j Cartesian axes.

The following expressions describe how to evaluate the derivatives of the density matrix (LCAO coefficients) and, therefore the response functions. From the time-dependent variational principle, using the Lagrange's method, the time-dependent Hartree-Fock equation reads:

$$\mathbf{FC} - i \frac{\partial}{\partial t} \mathbf{SC} = \mathbf{SC} \varepsilon, \tag{II.B.14}$$

where ε is the Lagrangian multipliers matrix. The orthonormalization conditions of the spin orbitals are

$$\mathbf{C}^\dagger \mathbf{S} \mathbf{C} = 1, \quad (\text{II.B.15})$$

$$\frac{\partial}{\partial t} \mathbf{C}^\dagger \mathbf{S} \mathbf{C} = 0, \quad (\text{II.B.16})$$

and the density matrix is defined as

$$\mathbf{D} = \mathbf{C} n \mathbf{C}^\dagger, \quad (\text{II.B.17})$$

where n is a diagonal matrix with elements equal to 1 for an occupied orbital and 0 for an unoccupied orbital. Here, $\mathbf{C}$ is a time-dependent LCAO coefficient matrix. Expanding it with respect to the external oscillating perturbation, it reads

$$\begin{aligned} \mathbf{C}(t) = & \mathbf{C}^{(0)} + \sum_{\omega} \sum_i^{xyz} f_i(\omega) \left[\mathbf{C}^{(i)}(\omega) e^{i\omega t} + \mathbf{C}^{(i)}(-\omega) e^{-i\omega t} \right] \\ & + \sum_{\omega_1, \omega_2} \sum_{i,j}^{xyz} f_i(\omega_1) f_j(\omega_2) \left[\begin{array}{c} \mathbf{C}^{(i)(j)}(\omega_1, \pm\omega_2) e^{i(\omega_1 \pm \omega_2)t} \\ + \mathbf{C}^{(i)(j)}(-\omega_1, \mp\omega_2) e^{-i(\omega_1 \pm \omega_2)t} \end{array} \right] + \dots \end{aligned} \quad (\text{II.B.18})$$

where superscripts (i) , (j) , ... represent the direction of the perturbation while their number indicates the order of the perturbation. For a single monochromatic optical oscillating field as in SHG or dc-Pockels effects, it simplifies to

$$\begin{aligned} \mathbf{C}(\omega) = & \mathbf{C}^{(0)} + \sum_i^{xyz} f_i(\omega) \left[\mathbf{C}^{(i)}(\omega) e^{i\omega t} + \mathbf{C}^{(i)}(-\omega) e^{-i\omega t} \right] \\ & + \sum_{i,j}^{xyz} f_i(\omega) f_j(\omega) \left[\begin{array}{c} \mathbf{C}^{(i)(j)}(\omega, \omega) e^{2i\omega t} \\ + \mathbf{C}^{(i)(j)}(-\omega, -\omega) e^{-2i\omega t} \\ + 2\mathbf{C}^{(i)(j)}(\omega, -\omega) \end{array} \right] + \dots \end{aligned} \quad (\text{II.B.19})$$

Similar perturbation expansions are obtained for the Fock matrix, the density matrix, and the Lagrangian multiplier matrix. The first-order TDHF, normalization, and density matrix equations read respectively

$$\begin{aligned} \mathbf{F}^{(i)}(\omega) \mathbf{C}^{(0)} + \mathbf{F}^{(0)} \mathbf{C}^{(i)}(\omega) + \omega \mathbf{S}^{(0)} \mathbf{C}^{(i)}(\omega) \\ = \mathbf{S}^{(0)} \mathbf{C}^{(i)}(\omega) \epsilon^{(0)} + \mathbf{S}^{(0)} \mathbf{C}^{(0)} \epsilon^{(i)}(\omega) \end{aligned} \quad (\text{II.B.20})$$

$$\mathbf{C}^{\dagger(i)}(-\omega) \mathbf{S}^{(0)} \mathbf{C}^{(0)} + \mathbf{C}^{\dagger(0)} \mathbf{S}^{(0)} \mathbf{C}^{(i)}(\omega) = 0, \quad (\text{II.B.21})$$

$$\mathbf{D}^{(i)}(\omega) = \mathbf{C}^{(i)}(\omega) n \mathbf{C}^{\dagger(0)} + \mathbf{C}^{(0)} n \mathbf{C}^{\dagger(i)}(-\omega). \quad (\text{II.B.22})$$

Introducing the $\mathbf{U}$ unitary matrix to determine the 1th order LCAO coefficient matrix from the zero-order coefficient matrix, it reads

$$\mathbf{C}^{(i)}(\omega) = \mathbf{C}^{(0)}\mathbf{U}^{(i)}(\omega). \quad (\text{II.B.23})$$

Plugging (II.B.23) into (II.B.21), it reads

$$\mathbf{U}^{\dagger(i)}(-\omega) + \mathbf{U}^{(i)}(\omega) = 0, \quad (\text{II.B.24})$$

$$\mathbf{C}^{(i)}(-\omega) = \mathbf{C}^{(0)}\mathbf{U}^{(i)}(-\omega) = -\mathbf{C}^{(0)}\mathbf{U}^{\dagger(i)}(\omega). \quad (\text{II.B.25})$$

The components of the first-order derivative of the density matrix read

$$D_{\mu\nu}^{(i)}(\omega) = \sum_a^{\text{occ. unocc.}} \sum_r \left[C_{\mu r}^{(0)} U_{ra}^{(i)}(\omega) C_{av}^{\dagger(0)} + C_{\mu a}^{(0)} U_{ar}^{\dagger(i)}(-\omega) C_{rv}^{\dagger(0)} \right] \quad (\text{II.B.26})$$

where only the (unocc. occ.) and (occ. unocc.) blocks of the unitary matrix derivative $\mathbf{U}^{(i)}$ are needed to evaluate the density matrix first derivative. If (II.B.20) is multiplied by $\mathbf{C}^{\dagger(0)}$, it reads:

$$\mathbf{C}^{(0)\dagger} \mathbf{F}^{(i)}(\omega) \mathbf{C}^{(0)} + \varepsilon^{(0)} \mathbf{U}^{(i)}(\omega) + \omega \mathbf{U}^{(i)}(\omega) = \mathbf{U}^{(i)}(\omega) \varepsilon^{(0)} + \varepsilon^{(i)}(\omega). \quad (\text{II.B.27})$$

Considering the $\mathbf{G}^{(i)}(\omega) = \mathbf{C}^{\dagger(0)} \mathbf{F}^{(i)}(\omega) \mathbf{C}^{(0)}$ definition, it becomes

$$G_{ar}^{(i)}(\omega) + \sum_c \varepsilon_{ac}^{(0)} U_{cr}^{(i)}(\omega) + \omega U_{ar}^{(i)}(\omega) = \sum_c U_{ac}^{(i)}(\omega) \varepsilon_{cr}^{(0)} + \varepsilon_{ar}^{(i)}(\omega), \quad (\text{II.B.28})$$

$$G_{ar}^{(i)}(\omega) + \varepsilon_a U_{ar}^{(i)}(\omega) + \omega U_{ar}^{(i)}(\omega) = U_{ar}^{(i)}(\omega) \varepsilon_r$$

$$\mathbf{U}_{ar}^{(i)}(\omega) = \frac{G_{ar}^{(i)}(\omega)}{\varepsilon_r - \varepsilon_a - \omega}. \quad (\text{II.B.29})$$

Injecting (II.B.29) into (II.B.23) leads to the possibility to determine the perturbed LCAO matrix $\mathbf{C}^{(i)}(\omega)$ and then the corresponding perturbed density matrix $\mathbf{D}^{(i)}(\omega)$. However, since $\mathbf{F}^{(i)}$ depends of $\mathbf{D}^{(i)}(\omega)$, this needs to be solved iteratively. From the knowledge of $\mathbf{U}^{(i)}(\omega)$ ($\mathbf{D}^{(i)}(\omega)$), the $\alpha_{ij}(-\omega; \omega)$ components can be calculated.

It is possible to take advantage of the “2n+1” rule²⁴, which states that, for a variational procedure, the energy to order 2n+1 can be calculated from the wave function to order n and lower. The first hyperpolarizability being a third-order response of the energy to external electric fields, first-order derivatives of the wave function are sufficient for its evaluation. As a matter of fact, equation (II.B.13) can be rewritten in a form that only contains zeroth- and first-order matrices:

$$\begin{aligned}
\beta_{ijk}(-\omega_\sigma; \omega_1, \omega_2) = & \\
& -tr \left[n \left\{ \begin{aligned} & \mathbf{U}^{(i)}(-\omega_\sigma) \mathbf{G}^{(j)}(\omega_1) \mathbf{U}^{(k)}(\omega_2) + \mathbf{U}^{(k)}(\omega_2) \mathbf{G}^{(j)}(\omega_1) \mathbf{U}^{(i)}(-\omega_\sigma) \\ & + \mathbf{U}^{(j)}(\omega_1) \mathbf{G}^{(k)}(\omega_2) \mathbf{U}^{(i)}(-\omega_\sigma) + \mathbf{U}^{(i)}(-\omega_\sigma) \mathbf{G}^{(k)}(\omega_2) \mathbf{U}^{(j)}(\omega_1) \\ & + \mathbf{U}^{(k)}(\omega_2) \mathbf{G}^{(i)}(-\omega_\sigma) \mathbf{U}^{(j)}(\omega_1) + \mathbf{U}^{(j)}(\omega_1) \mathbf{G}^{(i)}(-\omega_\sigma) \mathbf{U}^{(k)}(\omega_2) \end{aligned} \right\} \right] \quad (\text{II.B.30}) \\
& -tr \left[n \left\{ \begin{aligned} & \mathbf{U}^{(i)}(-\omega_\sigma) \mathbf{U}^{(k)}(\omega_2) \boldsymbol{\varepsilon}^{(j)}(\omega_1) + \mathbf{U}^{(k)}(\omega_2) \mathbf{U}^{(i)}(-\omega_\sigma) \boldsymbol{\varepsilon}^{(j)}(\omega_1) \\ & + \mathbf{U}^{(j)}(\omega_1) \mathbf{U}^{(i)}(-\omega_\sigma) \boldsymbol{\varepsilon}^{(k)}(\omega_2) + \mathbf{U}^{(i)}(-\omega_\sigma) \mathbf{U}^{(j)}(\omega_1) \boldsymbol{\varepsilon}^{(k)}(\omega_2) \\ & + \mathbf{U}^{(k)}(\omega_2) \mathbf{U}^{(j)}(\omega_1) \boldsymbol{\varepsilon}^{(i)}(-\omega_\sigma) + \mathbf{U}^{(j)}(\omega_1) \mathbf{U}^{(k)}(\omega_2) \boldsymbol{\varepsilon}^{(i)}(-\omega_\sigma) \end{aligned} \right\} \right]
\end{aligned}$$

Note that in the static limit of the electric field, the TDHF method reduces to the coupled-perturbed Hartree-Fock (CPHF) method.²⁵

II.C.3. Time-dependent DFT method

With DFT, the evaluation of the responses to dynamic external electric fields is done via the time-dependent DFT (TDDFT) method.²⁶ This proceeds similarly to TDHF method. The time-dependent Kohn-Sham equation reads

$$\mathbf{F}_{KS} \mathbf{C} - i \frac{\partial}{\partial t} \mathbf{S} \mathbf{C} = \mathbf{S} \mathbf{C} \boldsymbol{\varepsilon} \quad (\text{II.B.31})$$

where $\mathbf{F}_{KS} = \mathbf{H}^{core} + \mathbf{D} \mathbf{2J} + \mathbf{V}_{XC}$ is the analogue to the Fock matrix but where the $\mathbf{DK}$ part, where $\mathbf{K}$ is the exchange matrix, has been substituted by $\mathbf{V}_{XC}$. The elements of the exchange-correlation potential matrix in the Kohn-Sham orbitals basis set are defined as

$$(V_{XC})_{ab} = \langle \psi_a | V_{XC}(t) | \psi_b \rangle. \quad (\text{II.B.32})$$

The time-dependent electric dipole moment is expressed as in the TDHF expression (II.B.11). The linear and quadratic perturbation functions have the same form as (II.B.12) and (II.B.13). A unitary matrix is also used to determine the 1st order LCAO coefficient matrix (II.B.23), which is obtained iteratively by the expression (II.B.29) where $G_{ar}^{(i)}(\omega)$ is replaced by its TDDFT counterpart:

$$G_{KS\ ar}^{(i)}(\omega) = \langle \psi_a | \mu_i F_i(\omega) + V_{Coul}^{(i)} + V_{XC}^{(i)} | \psi_r \rangle, \quad (\text{II.B.33})$$

where all the matrix elements are determined by numerical integration. As the “2n+1” rule²⁴ still holds, the first hyperpolarizability determined by TDDFT reads:

$$\begin{aligned} \beta_{ijk}^{\text{TDDFT}}(-\omega_\sigma; \omega_1, \omega_2) = & \\ & -tr \left[n \left\{ \begin{aligned} & \mathbf{U}^{(i)}(-\omega_\sigma) \mathbf{G}^{(j)}(\omega_1) \mathbf{U}^{(k)}(\omega_2) + \mathbf{U}^{(k)}(\omega_2) \mathbf{G}^{(j)}(\omega_1) \mathbf{U}^{(i)}(-\omega_\sigma) \\ & + \mathbf{U}^{(j)}(\omega_1) \mathbf{G}^{(k)}(\omega_2) \mathbf{U}^{(i)}(-\omega_\sigma) + \mathbf{U}^{(i)}(-\omega_\sigma) \mathbf{G}^{(k)}(\omega_2) \mathbf{U}^{(j)}(\omega_1) \\ & + \mathbf{U}^{(k)}(\omega_2) \mathbf{G}^{(i)}(-\omega_\sigma) \mathbf{U}^{(j)}(\omega_1) + \mathbf{U}^{(j)}(\omega_1) \mathbf{G}^{(i)}(-\omega_\sigma) \mathbf{U}^{(k)}(\omega_2) \end{aligned} \right\} \right] \\ & -tr \left[n \left\{ \begin{aligned} & \mathbf{U}^{(i)}(-\omega_\sigma) \mathbf{U}^{(k)}(\omega_2) \boldsymbol{\varepsilon}^{(j)}(\omega_1) + \mathbf{U}^{(k)}(\omega_2) \mathbf{U}^{(i)}(-\omega_\sigma) \boldsymbol{\varepsilon}^{(j)}(\omega_1) \\ & + \mathbf{U}^{(j)}(\omega_1) \mathbf{U}^{(i)}(-\omega_\sigma) \boldsymbol{\varepsilon}^{(k)}(\omega_2) + \mathbf{U}^{(i)}(-\omega_\sigma) \mathbf{U}^{(j)}(\omega_1) \boldsymbol{\varepsilon}^{(k)}(\omega_2) \\ & + \mathbf{U}^{(k)}(\omega_2) \mathbf{U}^{(j)}(\omega_1) \boldsymbol{\varepsilon}^{(i)}(-\omega_\sigma) + \mathbf{U}^{(j)}(\omega_1) \mathbf{U}^{(k)}(\omega_2) \boldsymbol{\varepsilon}^{(i)}(-\omega_\sigma) \end{aligned} \right\} \right] \\ & +tr \left(\mathbf{g}_{xc}(\omega_1, \omega_2) \mathbf{D}^{(a)}(-\omega_\sigma) \mathbf{D}^{(b)}(-\omega_1) \mathbf{D}^{(c)}(-\omega_2) \right) \end{aligned} \quad (\text{II.B.34})$$

where $\mathbf{g}_{xc}(\omega_1, \omega_2)$, the exchange-correlation kernel, is the second-order response functions of the time-dependent exchange-correlation potential with respect to the time-dependent density. The determination of the second-order exchange-correlation kernel is beyond the purpose of this section but can be found in van Gisbergen *et al.*^{26b}. The extra DFT term (with respect to TDHF) is most efficiently calculated through a numerical integration.

II.C.4. Frequency-dependent first hyperpolarizabilities using coupled cluster quadratic response theory

The variational Lagrangian approach based on the time-dependent coupled cluster quasienergy is used to derive the quadratic response function.²⁷ Indeed, in order to obtain the quasienergy Lagrangian expression, several manipulations of the time-dependent Schrödinger equation are needed to determine the quasienergy and the time-dependent amplitudes. After that, with a perturbation expansion of the time-dependent cluster operator, it is possible from the time-average of the quasienergy Lagrangian to derive the response functions.

The time dependence of the coupled cluster state can be parametrized as

$$|\text{CC}(t)\rangle = e^{-iF(t)}|\widetilde{\text{CC}}(t)\rangle = e^{-iF(t)}e^{\hat{T}(t)}|\Psi_0\rangle \quad (\text{II.B.35})$$

where the phase $F(t)$ is a function of time and the time-dependent cluster operator $\hat{T}(t)$ is defined as $\sum_{\rho} t_{\rho}(t)\hat{t}_{\rho}$. Inserting (II.B.35) into the time-dependent Schrödinger equation

$$\left(\hat{H}(t) - i\frac{d}{dt}\right)|\text{CC}(t)\rangle = 0, \quad (\text{II.B.36})$$

gives:

$$\hat{H}(t)e^{-iF(t)}|\widetilde{\text{CC}}(t)\rangle = i\frac{d}{dt}\left(e^{-iF(t)}|\widetilde{\text{CC}}(t)\rangle\right) \quad (\text{II.B.37})$$

$$\Leftrightarrow \hat{H}(t)|\widetilde{\text{CC}}(t)\rangle = \left(\frac{dF(t)}{dt} + i\frac{d}{dt}\right)|\widetilde{\text{CC}}(t)\rangle. \quad (\text{II.B.38})$$

First, projecting (II.B.38) onto the Hartree-Fock reference determines the quasienergy $\langle\Psi_0|\hat{H}(t)|\widetilde{\text{CC}}(t)\rangle$ by isolating the phase factor first derivative:

$$\frac{dF(t)}{dt} = \langle\Psi_0|\left(\hat{H}(t) - i\frac{d}{dt}\right)|\widetilde{\text{CC}}(t)\rangle = \langle\Psi_0|\hat{H}(t)|\widetilde{\text{CC}}(t)\rangle. \quad (\text{II.B.39})$$

Secondly, the time-dependent amplitudes $t_{\rho}(t)$, are obtained by projecting (II.B.37) onto $\langle\Psi_{\rho}|e^{-\hat{T}(t)}$:

$$\underbrace{\langle\Psi_{\rho}|e^{-\hat{T}(t)}\hat{H}(t)e^{\hat{T}(t)}|\Psi_0\rangle}_{e_{\rho}(t)} - i\frac{d}{dt}t_{\rho}(t) = 0 \quad (\text{II.B.40})$$

$$\Leftrightarrow e_{\rho}(t) - i\frac{d}{dt}t_{\rho}(t) = 0$$

For example, in the case of the CCSD model, explicit expressions for CCSD amplitude equations take the form:²⁸

$$\begin{cases} \langle\Psi_a^r|H + [H, \hat{T}_2]|\Psi_0\rangle - i\frac{d}{dt}t_a^r = 0 \\ \langle\Psi_{ab}^{rs}|H + [H, \hat{T}_2] + \frac{1}{2}[[H, \hat{T}_2], \hat{T}_2]|\Psi_0\rangle - i\frac{d}{dt}t_{ab}^{rs} = 0 \end{cases}, \quad (\text{II.B.41})$$

where $\{\langle \Psi_a^r |, \langle \Psi_{ab}^{rs} | \}$ denotes the single and double excitation manifolds and the $\hat{T}_1$ -transformed Hamiltonian is expressed as $H = e^{-T_1} \hat{H} e^{T_1}$. A quasienergy Lagrangian is then defined from the expressions (II.B.39) and (II.B.40):

$$L(t) = \langle \Psi_0 | \hat{H}(t) | \widetilde{CC}(t) \rangle + \sum_{\rho} \bar{t}_{\rho}(t) \langle \Psi_{\rho} | e^{-\hat{T}(t)} \left(\hat{H}(t) - i \frac{d}{dt} \right) | CC(t) \rangle \quad (\text{II.B.42})$$

where $\bar{t}_{\rho}$ are the Lagrangian multipliers. This expression is rewritten as

$$L(t) = \frac{dF(t)}{dt} + \sum_{\rho} \bar{t}_{\rho} \left(e_{\rho}(t) - i \frac{d}{dt} t_{\rho}(t) \right). \quad (\text{II.B.43})$$

Based on a perturbation expansion of the time-dependent cluster operator $\hat{T}(t)$:

$$\begin{aligned} \hat{T}(t) &= \sum_{\rho} t_{\rho}(t) \hat{\tau}_{\rho} \\ &= \sum_{\rho} \left(t_{\rho}^{(0)} + t_{\rho}^{(1)}(t) + t_{\rho}^{(2)}(t) + \dots \right) \hat{\tau}_{\rho}, \end{aligned} \quad (\text{II.B.44})$$

the cluster amplitude can, at each order, be written as a Fourier expansion. For the linear response, it reads:

$$t_{\rho}^{(1)}(t) = \sum_{\omega} \sum_i^{xyz} t^{(i)}(\omega) f_i(\omega) e^{\pm i\omega t}. \quad (\text{II.B.45})$$

The time-average of the quasienergy Lagrangian for a period T ,

$$\{L(t)\}_T = \lim_{t_0 \rightarrow \infty} \frac{1}{2t_0} \int_{-t_0}^{t_0} L(t) dt \quad (\text{II.B.46})$$

may now be required to be stationary with respect to the Fourier component of the cluster amplitudes and the Lagrangian multipliers:

$$\delta \{L(t)\}_T = 0 \quad (\text{II.B.47})$$

Inserting the perturbation and Fourier expansion of the cluster amplitudes (II.B.44) and an analogous expansion of the Lagrangian multipliers into equation (II.B.47), this becomes a time-dependent variational condition for the individual expansion coefficients. The zeroth-order coefficients $t_{\rho}^{(0)}$ are obtained by the usual coupled cluster equations (II.A.45) for the unperturbed system, and the zeroth-order multipliers $\bar{t}_{\rho}^{(0)}$ by

$$\sum_{\rho} \bar{t}_{\rho}^{(0)} \left(\frac{de_{\rho}}{dt_{\nu}} \right)_{\bar{F}=0} = \left(\frac{d}{dt_{\nu}} \frac{\partial F(t)}{\partial t} \right)_{\bar{F}=0}. \quad (\text{II.B.48})$$

In order to determine quadratic response properties, according to the 2n+1 and 2n+2 rules,²⁹ (for 2n+1th energy derivatives only requires derivative integral to order n in the molecular orbitals basis and the Lagrange multipliers satisfy a 2n+2 rule) the first-order response for the amplitudes and for the multipliers is also needed, which are determined from

$$\xi^{(i)} + (\mathbf{A} - \omega_i \mathbf{1}) \mathbf{t}^{(i)}(\omega_i) = 0, \quad (\text{II.B.49})$$

$$\boldsymbol{\eta}^{(i)} + \mathbf{F} \mathbf{t}^{(i)}(\omega_i) + \bar{\mathbf{t}}^{(i)}(\omega_i)(\mathbf{A} - \omega_i \mathbf{1}) = 0, \quad (\text{II.B.50})$$

where the vectors $\boldsymbol{\eta}^{(i)}$ and $\xi^{(i)}$ and the matrices $\mathbf{A}$ and $\mathbf{F}$ are defined by

$$\begin{aligned} \eta_{\rho}^{(i)} &= \left(\frac{\partial^2 \{L\}_T}{\partial f_i(\omega) \partial t_{\rho}^{(j)}(-\omega)} \right)_{\bar{F}=0} \\ \xi_{\tau}^{(i)} &= \left(\frac{\partial^2 \{L\}_T}{\partial f_i(\omega) \partial \bar{t}_{\tau}^{(j)}(-\omega)} \right)_{\bar{F}=0} \\ A_{\tau\rho} &= \left(\frac{de_{\tau}}{dt_{\rho}} \right)_{\bar{F}=0} \\ F_{\tau\rho} &= \left(\frac{\partial^2 \{L\}_T}{\partial t_{\tau}^{(i)}(\omega) \partial t_{\rho}^{(j)}(-\omega)} \right)_{\bar{F}=0} \end{aligned} \quad (\text{II.B.51})$$

The linear and the quadratic response functions are obtained by taking the second and the third derivative of the time-averaged quasi-energy Lagrangian with respect to the Fourier components, respectively. The frequency-dependent polarizability and first hyperpolarizability become

$$\alpha_{ij}(-\omega; \omega) = \frac{1}{2} \hat{P}^{\pm\omega} \left\{ \boldsymbol{\eta}^{(i)} \mathbf{t}^{(j)}(\omega) + \mathbf{F} \mathbf{t}^{(i)}(-\omega) \mathbf{t}^{(j)}(\omega) + \boldsymbol{\eta}^{(j)} \mathbf{t}^{(i)}(-\omega) \right\}, \quad (\text{II.B.52})$$

$$\beta_{ijk}(-\omega_{\sigma}; \omega_1, \omega_2) = \frac{1}{2} \hat{P}^{\pm\omega} \left\{ \begin{aligned} &\mathbf{G} \mathbf{t}^{(i)}(-\omega_{\sigma}) \mathbf{t}^{(j)}(\omega_1) \mathbf{t}^{(k)}(\omega_2) \\ &+ \frac{1}{2} \hat{\mathbf{P}}^{ijk} \mathbf{B} \bar{\mathbf{t}}^{(i)}(-\omega_{\sigma}) \mathbf{t}^{(j)}(\omega_1) \mathbf{t}^{(k)}(\omega_2) \\ &+ \frac{1}{2} \hat{\mathbf{P}}^{ijk} \mathbf{F} \mathbf{t}^{(i)}(\omega_1) \mathbf{t}^{(j)}(\omega_2) \\ &+ \hat{\mathbf{P}}^{ijk} \mathbf{A}^{(i)} \bar{\mathbf{t}}^{(j)}(\omega_1) \mathbf{t}^{(k)}(\omega_2) \end{aligned} \right\}, \quad (\text{II.B.53})$$

where the $\mathbf{G}, \mathbf{B}, \mathbf{F}^{(A)}$ and $\mathbf{A}^{(A)}$ matrix elements are defined as:

$$\begin{aligned}
G_{\tau\rho\gamma} &= \left(\frac{\partial^3 \{L\}_T}{\partial t_\tau^{(i)}(-\omega_\sigma) \partial t_\rho^{(j)}(\omega_1) \partial t_\gamma^{(k)}(\omega_2)} \right)_{\vec{F}=0} \\
B_{\tau\rho\gamma} &= \left(\frac{\partial^3 \{L\}_T}{\partial \bar{t}_\tau^{(i)}(-\omega_\sigma) \partial t_\rho^{(j)}(\omega_1) \partial t_\gamma^{(k)}(\omega_2)} \right)_{\vec{F}=0} \\
F_{\tau\rho}^{(i)} &= \left(\frac{\partial^3 \{L\}_T}{\partial f_i(-\omega_\sigma) \partial t_\tau^{(j)}(\omega_1) \partial t_\rho^{(k)}(\omega_2)} \right)_{\vec{F}=0} \\
A_{\tau\rho}^{(i)} &= \left(\frac{\partial^3 \{L\}_T}{\partial f_i(-\omega_\sigma) \partial \bar{t}_\tau^{(j)}(\omega_1) \partial t_\rho^{(k)}(\omega_2)} \right)_{\vec{F}=0}
\end{aligned} \tag{II.B.54}$$

The operator $\hat{P}^{ijk}$ symmetrizes the matrices with respect to the perturbation indices and their accompanied frequencies.

II.C.5. The Finite Field method and frequency dispersion.

In this section, two aspects related to evaluating the molecular NLO properties are tackled, the numerical differentiations of the finite field procedure and the frequency dispersion. This section is based on an article named: “On the evaluation of the molecular static and dynamic first hyperpolarizabilities”, M. de Wergifosse, V. Liégeois and B. Champagne, *Int. J. Quant. Chem.*; accepted, DOI: 10.1002/qua.24685.

II.C.5.a. The finite field method

Among the different computational schemes, the finite field (FF) method³⁰ is an easy applicable technique, which consists of calculating the energy (or lower-order properties) for different amplitudes of the external field. This method is mostly used when analytical derivatives are not available or if their evaluations are too time consuming. It is usually based on this equation:

$$E(\vec{F}) = E(0) - \vec{\mu}(0) \cdot \vec{F} - \frac{1}{2!} \vec{\alpha} : \vec{F}\vec{F} - \frac{1}{3!} \vec{\beta} : \vec{F}\vec{F}\vec{F} - \dots \tag{II.B.55}$$

where the different tensor components can therefore be determined as derivatives of the dipole moment or of the energy with respect to the external field. For instance,

$$\beta_{ijk} = - \left(\frac{\partial^3 E}{\partial F_i \partial F_j \partial F_k} \right)_0 \quad (\text{II.B.56})$$

where i, j , and k refer to the applied external field directions, i.e. x, y or z . This makes it applicable to a large range of levels of approximation,³¹ provided field-dependent energies can be calculated. Indeed, it only needs an extra electric field term in the Hamiltonian. By design, this method is however limited to static properties or rather to derivatives with respect to static electric field. Besides, an automatic application of the FF method is not obvious. In particular, the selection of an appropriate window of electric field amplitudes is a crucial parameter to achieve high accuracy and this gets more and more stringent when going towards higher and higher order derivatives.³² Ref. **31** has established a range of feasible field strengths that guarantee reliable numerical results. Indeed, owing to the finite difference expressions to calculate the successive derivatives, the results contain contaminations from higher-order hyperpolarizabilities. These can be eliminated by employing the Richardson extrapolation procedure,³³ also known as Romberg differentiation in the quantum chemistry community.^{31d} On the basis of calculations performed for different sets of electric field amplitudes, a Romberg table or triangle is written and can help monitoring the convergence of the numerical differentiations. Still, the analysis of the Romberg triangle is not straightforward and requires understanding the physical phenomena underlying to the application of external fields to chemical compounds. Hereafter, a systematic and automatic procedure is proposed. It is based on three simple criteria, which enable to locate a sub-triangle where the numerical derivative is stable. It provides the targeted values as well as information about the convergence of the Romberg procedure, i.e. a precision. The assessment of the method is carried out at the Hartree-Fock level of

calculation in comparison to benchmark analytical derivatives, which are obtained using the coupled-perturbed Hartree-Fock scheme.²⁵ Further illustrations are performed at a correlated level, using the coupled-cluster singles and doubles (CCSD) method. In that case, the numerical derivatives are estimated from orbital-relaxed field-dependent energies as well as orbital-unrelaxed field-dependent energies. Then, comparisons are made with respect to analytical derivatives obtained with the CCSD quadratic response theory (CCSD-QRF).²⁷

II.C.5.b. Frequency dispersion

The optical characteristic of the molecular hyperpolarizabilities is another aspect to control in order to develop efficient devices or to interpret experimental data. Besides the resonance effects that can strongly impact the NLO responses and make difficult the analysis of the intrinsic responses,^{4b} off-resonance frequency dispersions are usually large in the UV/visible domain for organic π -conjugated molecules and deserve to be accounted for. When adopting strongly-correlated methods, this can turn out to be a real challenge, often because the method is not yet available. To alleviate this issue, approximate schemes have been elaborated. They combine static hyperpolarizabilities calculated at a highly-correlated level with frequency dispersion estimated at a lower one, typically by using the time-dependent Hartree-Fock method.²³ Sekino and Bartlett³⁴ proposed the multiplicative scheme, also previously known as *percentage* correction, which for β reads

$$\begin{aligned}\beta_{\text{correlated}}(-\omega_{\sigma}; \omega_A, \omega_B) &\approx \beta_{\text{TDHF}}(-\omega_{\sigma}; \omega_A, \omega_B) \frac{\beta_{\text{correlated}}(0; 0, 0)}{\beta_{\text{CPHF}}(0; 0, 0)} \\ &= \beta_{\text{correlated}}(0; 0, 0) \frac{\beta_{\text{TDHF}}(-\omega_{\sigma}; \omega_A, \omega_B)}{\beta_{\text{CPHF}}(0; 0, 0)}\end{aligned}\quad (\text{II.B.57})$$

where frequency-dependent TDHF estimates are corrected by the ratio between the correlated and Hartree-Fock static response or, equivalently, where the static correlated response is corrected for the ratio between the dynamic and static

Hartree-Fock values. Besides, Rice and Handy³⁵ have proposed an additive scheme where any tensor component of β is approximated by:

$$\begin{aligned}\beta_{correlated}(-\omega_\sigma; \omega_A, \omega_B) &\approx \\ \beta_{correlated}(0;0,0) &+ \{\beta_{TDHF}(-\omega_\sigma; \omega_A, \omega_B) - \beta_{CPHF}(0;0,0)\} \\ &= \beta_{TDHF}(-\omega_\sigma; \omega_A, \omega_B) + \{\beta_{correlated}(0;0,0) - \beta_{CPHF}(0;0,0)\}\end{aligned}\quad (II.B.58)$$

The validity of these expressions has been assessed in a few papers, in particular with respect to the range of frequencies where it is valid. Dalskov *et al.*³⁶ concluded that whenever the approximate TDHF frequency dependence is reliable both the multiplicative and the additive schemes work quite well. Pluta and Sadlej³⁷ stated that the range of frequencies for which the scaling can be used should not exceed about half of the lowest computed first excitation energy. For push-pull π -conjugated systems, it has been observed that the multiplicative scheme is superior to the additive one.³⁸ By using the quasi-energy derivative approach at the Møller-Plesset second-order level of approximation, Aiga and Itoh³⁹ calculated the frequency-dependent polarizabilities and hyperpolarizabilities of H₂O and NH₃ and found a good agreement with those of Sekino and Bartlett^{34b} obtained using the multiplicative scheme. In this section, further assessment of both approximate schemes is carried out for π -conjugated systems as well as for small molecules where the correlated dynamic values are evaluated at the CCSD-QRF level of approximation.²⁷

II.C.5.c. Kleinman's symmetry

Kleinman symmetry is defined as the inter-exchangeability of all the indices in the (hyper)polarizability tensor of a system.⁴⁰ Considering the SHG β response, the last two frequencies are identical and therefore the β tensor components are invariant upon permuting the corresponding last two indices, $\beta_{ijk}(-2\omega; \omega, \omega) = \beta_{ikj}(-2\omega; \omega, \omega)$. On the other hand, the following equality

$$\beta_{ijk}(-2\omega; \omega, \omega) = \beta_{kij}(-2\omega; \omega, \omega) = \beta_{jki}(-2\omega; \omega, \omega) \quad (II.B.59)$$

is only enforced by symmetry in the static limit or when the Kleinman symmetry is satisfied. In fact, those conditions are satisfied for a limited range of frequencies and these should be known. Indeed, they are routinely used to simplify the analysis of NLO measurements because they enable to reduce the number of independent components of the hyperpolarizability tensor. In certain circumstances, assuming Kleinman symmetry leads to errors, as pointed out in Ref. 41. They used the Kramers-Kronig dispersion relations to demonstrate that the frequency window where Kleinman's symmetry is satisfied is narrower than usually admitted for a fluorescein chromophore. They showed that the Kramers-Kronig dispersion and Kleinman's relations are mutually exclusive. In this paper, using CCSD-QRF β responses, the range of validity of Kleinman's symmetry conditions is further assessed by calculating the frequency dispersion of the first hyperpolarizability of π -conjugated molecules.

II.C.5.d. Molecular structures and computational procedure

Two molecules were selected to tackle these frequency dispersion and numerical derivative issues, *i.e.* i) *p*-nitroaniline (**1**) and ii) a derivative of the *p*-quinodimethane where the $\text{CH}_2^{(\bullet)}$ terminal units are formally replaced by one $\text{NH}_2^{(+\bullet)}$ and one $\text{BH}_2^{(-\bullet)}$ (**2**) groups (fig.1). As the later compound has a diradical character mostly equal to zero, closed-shell calculations have been performed. See reference ⁴² for the formal expression of the diradical character. Their geometries were optimized at the density functional theory level with the B3LYP exchange-correlation functional and the 6-311G(d) basis set. Though the hyperpolarizabilities depend on the geometries, it is not expected that changing the geometries will modify the conclusions of these investigations. This first set of molecules was extended to two other sets of compounds for the study of the frequency dispersion, i) two additional push-pull π -conjugated systems, $\text{H}_2\text{N}-\text{CH}=\text{CH}-\text{CH}=\text{CH}-\text{NO}_2$ and $\text{H}_2\text{N}-\text{C}\equiv\text{C}-\text{C}\equiv\text{C}-\text{NO}_2$ and ii) small molecules including H_2O , CCl_4 , CHCl_3 , CH_2Cl_2 , CH_3CN , and

CCl_3CN . The geometries of these molecules were all optimized at the same level of theory as the other ones.

Compounds **1** and **2** belong to the C_s and C_{2v} point group, respectively. Note that the two hydrogen atoms of the NH_2 group of *p*-nitroaniline are slightly out-of-plane with respect to the rest of the molecule, implying C_s symmetry. For the C_s point group, there are ten nonzero independent components in the SHG β tensor: $\beta_{xxx}, \beta_{yyy}, \beta_{xxy}, \beta_{xyy}, \beta_{xxz}, \beta_{yyz}, \beta_{yxx}, \beta_{yyx}, \beta_{zzx},$ and β_{zzy} . In the static limit or more generally when the Kleinman conditions are satisfied: $\beta_{xxy} = \beta_{yxx}, \beta_{xyy} = \beta_{yyx}, \beta_{xxz} = \beta_{zzx},$ and $\beta_{yyz} = \beta_{zzy}$, which reduces to 6 the number of independent β tensor components. For the C_{2v} point group, there are five nonzero independent components: $\beta_{zzz}, \beta_{zyy}, \beta_{zxx}, \beta_{yyz},$ and β_{xxz} . Again, in the static limit, some pairs of components are identical: $\beta_{zyy} = \beta_{yyz}$ and $\beta_{zxx} = \beta_{xxz}$.

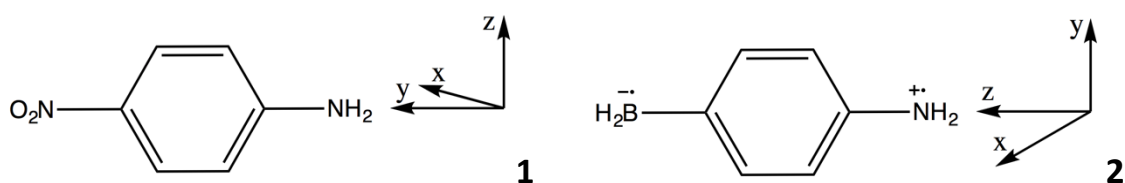


FIG. 1. Structures of compounds **1** and **2** (see page 216 for the drawing) in the frame of Cartesian axes.

The finite field calculations were performed i) at the Hartree-Fock level of approximation with the aug-cc-pVTZ basis set and ii) at the coupled-cluster level of theory with single and double excitations for relaxed orbitals (the external electric field is added directly to the HF Hamiltonian) and non-relaxed ones (the external electric field is added to the CCSD Hamiltonian) with the 6-31+G(d) basis set. The energies were converged to a precision of 10^{-11} a.u. The field amplitudes range from 0.0002 a.u. to 0.0064 a.u. in a geometric progression with a common ratio of $a = 2$, $a = \sqrt{2}$, and $a = \sqrt[4]{2}$. In order to assess the accuracy of the FF calculations, the

numerical values are compared to the analytical ones obtained i) with the Coupled-Perturbed Hartree-Fock method²⁵ with a convergence threshold of 10^{-8} and ii) with the QRF-CCSD method²⁷, when the numerical values are obtained without orbital relaxation.

TDHF and QRF-CCSD calculations of β at wavelengths of 632, 713, 794, 929, 1064, 1500, and 1900 nm were carried out to assess the additive and multiplicative approximations. In addition to the tensor components, the analysis also addresses quantities that can be extracted from hyper-Rayleigh scattering experiments, *i.e.*, β_{HRS} and the depolarization ratios.⁴³ Full expressions for these quantities without assuming Kleinman's conditions can be found in reference 44.

HF and TDHF/CPHF calculations were carried out with the Gaussian09 packages⁴⁵. CCSD FF and QRF calculations were performed with the Dalton2011 program⁴⁶.

II.C.5.e. Towards an automatic Romberg differentiation procedure

II.C.5.e.1. Theory

The Romberg differentiation procedure consists in combining energy values obtained for a succession of k external electric fields, of which the amplitudes form a geometric progression given by $F(k) = a^k F_0$, where F_0 is the smallest field value ($F_0 = 0.0002$ a.u.) and a the common ratio. For the different types of β tensor components the finite difference expressions^{30,31} corresponding to the zero-order iteration read:

$$\beta_{iii}(k,0) = 3 \frac{[E(F_{-i}(k+1)) - E(F_i(k+1))] - a[E(F_{-i}(k)) - E(F_i(k))]}{a(a^2 - 1)(a^k F_0)^3} \quad (\text{II.B.60})$$

$$\beta_{ijj}(k,0) = \frac{\left([E(F_{-i-j}(k)) - E(F_{ij}(k))] + [E(F_{-ij}(k)) - E(F_{i-j}(k))] \right) - 2[E(F_{-i}(k)) - E(F_i(k))]}{2(a^k F_0)^3} \quad (\text{II.B.61})$$

$$\beta_{ijm}(k,0) = \frac{\left(\begin{aligned} & \left[E(F_{-i-j-m}(k)) - E(F_{ijm}(k)) \right] - \left[E(F_{-i-j}(k)) - E(F_{ij}(k)) \right] \\ & - \left[E(F_{-i-m}(k)) - E(F_{im}(k)) \right] - \left[E(F_{-j-m}(k)) - E(F_{jm}(k)) \right] \\ & + \left[E(F_{-i}(k)) - E(F_i(k)) \right] + \left[E(F_{-j}(k)) - E(F_j(k)) \right] \\ & + \left[E(F_{-m}(k)) - E(F_m(k)) \right] \end{aligned} \right)}{2(a^k F_0)^3} \quad (\text{II.B.62})$$

where the $\pm i$, $\pm j$, and $\pm m$ field indexes refer to the field directions, *i.e.* $\pm x$, $\pm y$ or $\pm z$. These expressions are easily obtained by combining Taylor expansions of the energy (II.B.55) with different amplitudes and/or directions of the external fields. These expressions are then truncated to third-order. For example, to obtain Eq. (II.B.61), the corresponding Eq. (II.B.55) reads:

$$\begin{aligned} E(F_{ij}(k)) = E_0 - & \left[\mu_i F_i(k) + \mu_j F_j(k) \right] - \frac{1}{2} \left[\alpha_{ii} F_i(k)^2 + 2\alpha_{ij} F_i(k) F_j(k) + \alpha_{jj} F_j(k)^2 \right] \\ & - \frac{1}{3!} \left[\beta_{iii} F_i(k)^3 + 3\beta_{iij} F_i(k)^2 F_j(k) + 3\beta_{ijj} F_i(k) F_j(k)^2 + \beta_{jjj} F_j(k)^3 \right] - \dots \end{aligned} \quad (\text{II.B.63})$$

and the following field amplitudes are used: $(F_i(k), F_j(k)) = (a^k F_0, a^k F_0)$, $(-a^k F_0, -a^k F_0)$, $(-a^k F_0, a^k F_0)$, $(a^k F_0, -a^k F_0)$, $(a^k F_0, 0)$ and $(-a^k F_0, 0)$. Because the field amplitudes are finite, these expressions provide approximations to the true static β tensor components, and they contain contamination from higher-order hyperpolarizabilities. Eq. (II.B.62) has been designed to use only two additional field amplitudes/directions $F_{ijm}(k)$ and $F_{-i-j-m}(k)$ in addition to those already used for the previous types of components. An alternative and more accurate expression with improved higher-order hyperpolarizability cancelations to get the β_{ijm} tensor components reads:

$$\beta_{ijm}(k,0) = \frac{\left(\begin{aligned} & \left[E(F_{-i-j-m}(k)) - E(F_{ijm}(k)) \right] + \left[E(F_{ij-m}(k)) - E(F_{-i-jm}(k)) \right] \\ & + \left[E(F_{i-jm}(k)) - E(F_{-ij-m}(k)) \right] + \left[E(F_{-ijm}(k)) - E(F_{i-j-m}(k)) \right] \end{aligned} \right)}{8(a^k F_0)^3} \quad (\text{II.B.64})$$

However, this expression requires evaluating the energy at 6 additional fields. To remove the contaminations from higher-order hyperpolarizabilities, successive “Romberg iterations” are carried out using a recursive expression³²⁻³³:

$$\beta(k,n) = \frac{a^{2n}\beta(k,n-1) - \beta(k+1,n-1)}{a^{2n} - 1} \quad (\text{II.B.65})$$

which leads to the so-called “Romberg triangle” where the convergence of the numerical derivative can be monitored. Since these expressions are based on molecular energies, their physico-chemical nature can impact the convergence of the “Romberg procedure”. As mentioned in reference³², for any molecule there exists a field window where the finite differentiation is stable. Upper and lower bounds define this window. Field strengths below that window suffer from a too large round-off error between energy values, which is proportional to the energy convergence threshold. On the other side, the upper bound is imposed by the critical field amplitude corresponding to the intersection between the ground and excited state energies. The latter field depends on the molecule. In the Romberg differentiation procedure, this stability window defines a sub-triangle, which is the “holy grail” for the analysis of the “Romberg triangle”. Up to now, this analysis has mostly been carried out by “hand”. However, by considering some criteria derived from the observations made in reference 32, a window of ideal field strengths can be automatically selected, leading to an automatic Romberg differentiation procedure.

The procedure is based on the analysis of field amplitude errors. The field amplitude error is defined as the differences between β values obtained for consecutive field amplitudes at the same “Romberg iteration”:

$$\varepsilon_k(n) = \beta(k+1,n) - \beta(k,n) \quad (\text{II.B.66})$$

The field error obeys to the same recursive expression as the “Romberg procedure”. Thus, by the use of the recursive expression, the amount of field amplitude errors

along the iterations is expected to decrease. Another useful quantity is the iteration (order) error, which is defined as:

$$\varepsilon_n(k) = \beta(k, n+1) - \beta(k, n) \quad (\text{II.B.67})$$

The iteration error can be seen as a quantity probing the convergence of the “Romberg procedure” between successive iterations.

The analysis starts by considering the $\beta(k, 0)$ values at the iteration $n=0$. If their absolute values are smaller than a particular threshold (10^{-1} or 10^{-2} a.u. for the first hyperpolarizability), the β value is directly obtained and the procedure is stopped. If not, the procedure selects the set of consecutive β values for which the finite differentiation is stable. Stability is defined by field errors of the same sign (of which the magnitude increases with the field amplitude). If there is more than one stability set, the largest set is selected, and then the set with the smallest field error. Note that, for small field amplitudes, a change of sign of the field error is often an indication that below that field amplitude round-off errors are present. If the number of sets is the same as the number of field errors, the smaller field error is picked up and the β value obtained at the next iteration is the final value and the accuracy is given by the corresponding iteration error. If the last condition is not met, a selected set of values is used to go to the next iteration.

If there is no reduction of the field errors between two consecutive iterations via the recursive expression, convergence is achieved and the lowest field error from the previous iteration is selected to determine the final value. In the other case, the “Romberg procedure” reduces the amplitude of the field errors (this corresponds to removing the high-order contaminations), and the iterative procedure is pursued until convergence.

The HF/aug-cc-pVTZ β tensor components determined using the FF approach and field amplitudes ranging from 0.0002 to 0.0064 with $a = 2$, $a = \sqrt{2}$, and $a = \sqrt[4]{2}$, as well as, for each of these components, the final iteration error and the CPHF reference results are listed in Table 1. For example, a value of -993.83 a.u.

is calculated for the β_{yyy} component of compound **1** when $a = 2$. Table **2** presents the “Romberg triangle” for this quantity. At the $n=0$ iteration, the selected set of field errors goes from $k=1$ to 3. Going to the next iteration, the field error signs alternate and the smallest field error corresponds to $k = 2$. The final β_{yyy} value is therefore obtained at iteration $n=2$ for $k=2$. For a step size of $a = \sqrt[4]{2}$ (Table **4**) i) at $n=0$, the region of stability of the field errors goes from $k=8$ to 18, ii) going to $n=1$, the set is restricted to the $k=15-17$ region, and finally iii) $n=2$, there is no reduction of the field errors with respect to the previous iteration and the final value corresponds to $n=2$ and $k=16$.

HF aug-cc-pVTZ	1				2			
	$a = \sqrt[4]{2}$	$a = \sqrt{2}$	$a = 2$	CPHF	$a = \sqrt[4]{2}$	$a = \sqrt{2}$	$a = 2$	CPHF
β_{xxx}	-3.95 0.01	-3.94 0.04	-3.94 0.02	-3.94	0.00	0.00	0.00	0.00
β_{yyy}	-993.82 0.01	-993.83 0.01	-993.83 0.00	-993.81	0.00	0.00	0.00	0.00
β_{zzz}	0.00	0.00	0.00	0.00	-1282.82 0.01	-1282.83 0.00	-1282.81 0.00	-1282.82
β_{xyy}	-8.79 0.00	-8.80 0.00	-8.80 0.00	-8.80	0.00	0.00	0.00	0.00
β_{xzz}	-3.26 0.00	-3.26 0.00	-3.26 0.00	-3.26	0.00	0.00	0.00	0.00
β_{yzz}	176.41 0.00	176.40 0.00	176.40 0.00	176.41	0.00	0.00	0.00	0.00
β_{xxy}	36.03 0.01	36.01 0.00	35.99 0.00	36.01	0.00	0.00	0.00	0.00
β_{xxz}	0.00	0.00	0.00	0.00	-53.38 0.00	-53.36 0.00	-53.36 0.00	-53.37
β_{yyz}	0.00	0.00	0.00	0.00	65.41 0.00	65.41 0.00	65.41 0.00	65.41
β_{xyz}	0.00 0.00	0.00 0.00	0.01 0.00	0.00	0.00 0.00	0.00 0.00	0.01 0.00	0.00
β_{HRS}	387.16	387.17	387.17	387.16	530.85	530.85	530.85	530.85
DR	3.22	3.22	3.22	3.22	4.86	4.86	4.86	4.86

TABLE 1. HF/aug-cc-pVTZ β components (in a.u.) as obtained within the FF and CPHF approaches for compounds **1** and **2**. The reported FF values were determined from analyzing the Romberg table obtained for calculations performed with field amplitudes ranging from 0.0002 to 0.0064 a.u. and common ratios $a = 2$, $a = \sqrt{2}$, and $a = \sqrt[4]{2}$. The corresponding absolute values of the final iteration error are also given.

	n=0		n=1		n=2		n=3		n=4
k=0	-993.69	-0.02	-993.71	-0.02	-993.73	-0.01	-993.73	0.00	-993.73
	0.05		0.32		0.37		0.38		
k=1	-993.64	0.26	-993.39	0.03	-993.36	0.01	-993.35		
	-0.77		-0.45		-0.48				
k=2	-994.41	0.58	-993.83	-0.00	-993.83				
	-1.73		0.02						
k=3	-996.14	2.33	-993.81						
	-7.0								
k=4	-1003.13								

TABLE 2. “Romberg triangle” for the determination of β_{yyy} of compound **1** at the HF/aug-cc-pVTZ level with a step size $a = 2$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

	n=0		n=1		n=2		n=3		n=4
k=4	-994.05	0.41	-993.64	0.01	-993.63	0.00	-993.63	0.00	-993.63
	-0.41		-0.04		0.01		0.05		0.07
k=5	-994.46	0.78	-993.68	0.06	-993.62	-0.04	-993.58	-0.02	-993.56
	-0.78		-0.20		-0.28		-0.34		
k=6	-995.24	1.36	-993.88	-0.02	-993.90	-0.01	-993.91		
	-1.37		0.06		0.07				
k=7	-996.61	2.79	-993.82	-0.01	-993.83				
	-2.78		0.03						
k=8	-999.39	5.60	-993.79						
	-5.61								
k=9	-1005.00								

TABLE 3. Part of the “Romberg triangle” for the determination of β_{yyy} of compound **1** at the HF/aug-cc-pVTZ level with a step size $a = \sqrt{2}$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

	n=0		n=1		n=2		n=3		n=4
k=14	-996.05	2.26	-993.79	0.03	-993.76	0.04	-993.72	0.04	-993.68
	-0.94		-0.03		-0.08		-0.13		-0.18
k=15	-996.99	3.17	-993.82	-0.02	-993.84	-0.01	-993.85	0.01	-993.86
	-1.31		0.01		0.02		0.04		
k=16	-998.30	4.49	-993.81	<u>-0.01</u>	-993.82	0.01	-993.81		
	-1.86		<u>0.01</u>		-0.02				
k=17	-1000.17	6.37	-993.80	-0.04	-993.84				
	-2.64		0.04						
k=18	-1002.81	9.05	-993.76						
	-3.75								
k=19	-1006.56								

TABLE 4. Part of the “Romberg triangle” for the determination of β_{yyy} of compound **1** at the HF/aug-cc-pVTZ level with a step size $a = \sqrt[4]{2}$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

	n=0		n=1		n=2		n=3		n=4
k=0	-1283.58	-0.17	-1283.75	-0.05	-1283.80	-0.01	-1283.81	-0.01	-1283.82
	0.51		0.80		0.84		0.85		
k=1	-1283.07	0.12	-1282.95	-0.01	-1282.96	0.00	-1282.96		
	-0.37		0.14		0.15				
k=2	-1283.44	0.63	-1282.81	<u>0.00</u>	-1282.81				
	-1.88		<u>0.02</u>						
k=3	-1285.32	2.53	-1282.79						
	-7.59								
k=4	-1292.91								

TABLE 5. “Romberg triangle” for the determination of the β_{zzz} component of compound **2** at the HF/aug-cc-pVTZ level with a step size $a = 2$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

	n=0		n=1		n=2		n=3		n=4		n=5
k=3	-1283.11	0.07	-1283.04	-0.09	-1283.13	0.05	-1283.18	-0.02	-1283.20	-0.02	-1283.22
	-0.07		0.26		0.36		0.42		0.44		0.47
k=4	-1283.18	0.40	-1282.78	0.01	-1282.77	0.01	-1282.76	0.00	-1282.76	0.01	-1282.75
	-0.39		-0.04		-0.05		-0.07		-0.07		
k=5	-1283.57	0.75	-1282.82	-0.01	-1282.83	0.00	-1282.83	<u>0.00</u>	-1282.83		
	-0.75		0.01		0.01		<u>0.01</u>				
k=6	-1284.32	1.51	-1282.81	-0.01	-1282.82	0.00	-1282.82				
	-1.50		0.01		0.01						
k=7	-1285.82	3.02	-1282.80	-0.01	-1282.81						
	-3.02		0.05								
k=8	-1288.84	6.09	-1282.75								
	-6.09										
K=9	-1294.94										

TABLE 6. Part of the “Romberg triangle” for the determination of the β_{zzz} component of compound **2** at the HF/aug-cc-pVTZ level with a step size $a = \sqrt{2}$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

	n=0		n=1		n=2		n=3		n=4		n=5
k=14	-1285.23	2.41	-1282.82	-0.04	-1282.86	-0.05	-1282.91	-0.06	-1282.97	-0.06	-1283.03
	-1.00		0.04		0.09		0.18		0.27		
k=15	-1286.23	3.45	-1282.78	0.01	-1282.77	0.04	-1282.73	0.03	-1282.70		
	-1.43		-0.02		-0.06		-0.09				
k=16	-1287.66	4.86	-1282.80	-0.03	-1282.83	<u>0.01</u>	-1282.82				
	-2.02		0.03		<u>-0.01</u>						
k=17	-1289.68	6.92	-1282.77	-0.07	-1282.84						
	-2.86		0.08								
k=18	-1292.55	9.86	-1282.69								
	-4.08										
k=19	-1296.63										

TABLE 7. Part of the “Romberg triangle” for the determination of the β_{zzz} component of compound **2** at the HF/aug-cc-pVTZ level with a step size $a = \sqrt[4]{2}$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

II.C.5.e.2. Applications

Table 1 reports the β tensor components for compounds 1 and 2 as determined using the FF approach and the reference CPHF data. Using $a = 2$, the FF calculations are in a good agreement with the CPHF results with less than 0.06% of difference for any non-zero component. For the dominant β_{yyy} component, the “Romberg triangles” obtained with $a = 2$, $a = \sqrt{2}$, and $a = \sqrt[4]{2}$ are presented in Tables 2, 3 and 4, respectively. For $a = 2$, the algorithm selects the value for $n=2$ and $k=2$ and a final estimate of -993.83 a.u., which differs by $2 \cdot 10^{-3}\%$ from the CPHF value. For $a = \sqrt{2}$ (Table 3) and $a = \sqrt[4]{2}$ (Table 4) many field amplitudes are used, 11 and 21, respectively. Therefore, at $n = 0$, the selected stability zone is larger but, upon iteration, it rapidly shrinks to define the final estimates, which are similar to that with $a = 2$. In view of comparison with experiment, the value of interest is β_{HRS} and Table 1 shows that the same value is obtained using the 3 common ratios.

For the determination of the β_{xxx} component of molecule 2, the analysis of the “Romberg table” is presented in Tables 5, 6, and 7 for step sizes of $a = 2$, $a = \sqrt{2}$, and $a = \sqrt[4]{2}$, respectively. As for molecule 1, the external field window is sufficiently large to include the stability area where the field differentiations are stable. In the case of $a = 2$, the algorithm selects a final value of -1282.81 a.u. for $k=2$ and $n=2$. The difference with respect to the CPHF value is again negligible. Using $a = \sqrt{2}$ gives a converged value for $k=5$ at iteration $n=4$. A similar result is obtained for $a = \sqrt[4]{2}$. Similar analyses can be made for the other tensor components and therefore for the β_{HRS} quantity. Note that i) the numerical approach is also able to predict the zero values for those components, which are zero by symmetry and that ii) such a numerical accuracy on β_{HRS} is largely superior to the precision achieved nowadays experimentally.

In the following, $a = 2$ is used for CCSD calculations because it requires less field-dependent energy calculations. Indeed, the number of calculations needed to

obtain the whole first hyperpolarizability tensor for field amplitudes going from 0.0002 to 0.0064 with $a = 2$, $a = \sqrt{2}$, and $a = \sqrt[4]{2}$, are respectively 120, 220, and 420. With the availability of today's computer architectures, these large numbers of calculations can be run simultaneously, leading to effectiveness comparable to a massively parallel calculation. In these calculations, an *a priori* large field amplitudes window is employed to be certain to enclose the stable sub-window. Then, employing the automatic Romberg differentiation procedure, the range of field amplitudes is narrowed. For instance, for β_{yyy} with $a = 2$, the stable sub-window is defined by field amplitudes ranging from 0.0008 to 0.0064 a.u., at the HF level of calculation. It is therefore chosen for the calculations at the correlated level of theory, in order to decrease the number of calculations.

We now consider the CCSD first hyperpolarizability calculations. Table 8 compares the FF and CCSD-QRF results. For the *p*-nitroaniline, in both relaxed/non-relaxed CCSD calculations, all the components of the first hyperpolarizability tensor converge rapidly with a final iteration order error of ± 0.02 a.u. or less (Tables 9 and 10). As expected, the non-relaxed CCSD FF results and the QRF ones are basically identical. The relaxed CCSD results are slightly smaller than the unrelaxed ones, leading to β_{HRS} (relaxed) of 679 a.u. and β_{HRS} (non-relaxed) of 739 a.u. Comparisons between the non-relaxed and relaxed orbitals approaches in coupled cluster calculations were debated in 1987 by Salter *et al.*,⁴⁷ who demonstrated numerically and proved formally that the CCSD model using non-relaxed orbitals is able to include most of the relaxation effects for the evaluation of first- and second-order properties but some small differences are observed for higher-order properties as the first hyperpolarizability. In the case of first hyperpolarizability, they showed for the HF molecule that the response is 3.2% higher for the non-relaxed orbitals with respect to the relaxed ones. They speculated that this difference could originate from different numerical precisions. Later, O'Neil *et al.*⁴⁸ pointed out that including orbital relaxation of the reference determinant introduces an extra set of unphysical

poles in the CC response, which leads in the case of static electrical properties to results not necessarily more precise. For compound **2**, similar conclusions can be drawn. In Table **11**, it is shown that the non-relaxed CCSD β_{zzz} component converges rapidly. However, in the relaxed case, the “Romberg triangle” is less stable and it might benefit from the use of a smaller step size and/or a smaller energy threshold (Table **12**).

For determining the CCSD β_{xyz} of compounds **1** and **2** Eq. (II.B.64) was preferentially employed. It shows for both compounds a better stability, with a zero value dictated by symmetry. Using Eq. (II.B.62), the relaxed CCSD values of compounds **1** and **2** amount of 0.03 a.u. and 0.02 a.u. while in the non-relaxed case, 0.06 and -0.03 a.u., respectively. The question is of course to know whether, for large systems, this small improvement is worth because it requires additional, resource-expensive, calculations.

CCSD 6-31+G(d)	1			2		
	Relaxed orbitals	Non-relaxed orbitals	Quadratic response functions	Relaxed orbitals	Non-relaxed orbitals	Quadratic response functions
β_{xxx}	-16.23 0.02	-17.97 0.00	-17.92	0.01 0.00	0.00	0.00
β_{yyy}	-1699.26 0.00	-1848.00 0.02	-1848.32	-0.04 0.01	0.00	0.00
β_{zzz}	0.01 0.00	0.00 0.00	0.00	-2387.99 0.31	-2628.69 0.02	-2628.35
β_{xyy}	-20.36 0.00	-25.14 0.00	-25.13	0.00	0.00	0.00
β_{xzz}	-3.26 0.00	-3.31 0.02	-3.31	0.00	0.00	0.00
β_{yzz}	105.05 0.00	110.58 0.00	110.61	0.00	0.00	0.00
β_{xxy}	49.61 0.00	52.86 0.00	52.32	0.00	0.00	0.00
β_{xxz}	0.00	0.00	0.00	-47.23 0.00	-48.33 0.01	-48.40
β_{yyz}	0.00	0.00	0.00	54.19 0.00	68.53 0.00	68.51
β_{xyz}	0.00	0.00	0.00	0.00	0.00	0.00
$\beta_{\text{HRS}, \infty}$	678.89	739.00	739.21	988.19	1085.60	1085.47
DR	4.19	4.21	4.21	4.96	4.92	4.92

TABLE 8. CCSD/6-31+G(d) β components (in a.u.) as obtained within the FF and QRF approaches for compounds **1** and **2**. The reported FF values were determined from analyzing the Romberg table obtained for calculations performed with field amplitudes ranging from 0.0002 to 0.0064 a.u. and common ratios $a = 2$. The corresponding absolute values of the final iteration error are also given.

	n=0		n=1		n=2		n=3		n=4
k=0	-1829.94	6.48	-1823.46	1.71	-1821.75	0.44	-1821.32	0.11	-1821.21
	-19.43		-25.63		-27.41		-27.86		
k=1	-1849.37	0.28	-1849.09	-0.07	-1849.16	-0.02	-1849.18		
	-0.84		1.07		1.16				
k=2	-1850.21	2.19	-1848.02	0.02	-1848.00				
	-6.58		-0.22						
k=3	-1856.79	8.55	-1848.24						
	-25.63								
k=4	-1882.42								

TABLE 9. “Romberg triangle” for the determination of the β_{yyy} component of compound **1** at the non-relaxed CCSD/6-31+G(d) level with a step size $\alpha = 2$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

	n=0		n=1		n=2		n=3		n=4
k=0	-1722.31	-7.08	-1729.39	-1.88	-1731.27	-0.48	-1731.75	-0.12	-1731.87
	21.24		28.27		30.03		30.47		
k=1	-1701.07	-0.05	-1701.12	-0.12	-1701.25	-0.03	-1701.28		
	0.16		1.86		1.99				
k=2	-1700.91	1.65	-1699.26	0.00	-1699.26				
	-4.96		-0.01						
k=3	-1705.87	6.60	-1699.27						
	-19.81								
k=4	-1725.68								

TABLE 10. “Romberg triangle” for the determination of the β_{yyy} component of compound **1** at the relaxed CCSD/6-31G+(d) level with a step size $\alpha = 2$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

	n=0		n=1		n=2		n=3		n=4
k=0	-2647.94	-7.01	-2654.95	-1.94	-2656.89	-0.50	-2657.39	-0.12	-2657.51
	21.03		29.14		31.27		31.82		
k=1	-2626.91	1.10	-2625.81	0.19	-2625.62	0.05	-2625.57		
	-3.31		-2.86		-3.07				
k=2	-2630.22	1.54	-2628.67	-0.02	-2628.69				
	-4.63		0.37						
k=3	-2634.85	6.55	-2628.30						
	-19.62								
k=4	-2654.47								

TABLE 11. “Romberg triangle” for the determination of the β_{zz} component of compound **2** at the non relaxed CCSD/6-31+G(d) level with a step size $a = 2$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

	n=0		n=1		n=2		n=3		n=4
k=0	-2375.50	-3.43	-2378.93	-1.41	-2380.34	-0.39	-2380.73	-0.10	-2380.83
	10.30		21.22		24.65		25.56		
k=1	-2365.20	7.49	-2357.71	2.02	-2355.69	0.52	-2355.17		
	-22.48		-30.28		-32.70				
k=2	-2387.68	-0.31	-2387.99	-0.40	-2388.39				
	0.91		6.02						
k=3	-2386.77	4.80	-2381.97						
	-14.40								
k=4	-2401.17								

TABLE 12. “Romberg triangle” for the determination of the β_{zz} component of compound **2** at the relaxed CCSD/6-31+G(d) level with a step size $a = 2$. The field amplitude errors and the iteration order errors are respectively shown in blue and in red.

II.C.5.f. Approximate frequency dispersions of the first hyperpolarizability from correlated calculations

The performance of the multiplicative and additive schemes to describe the frequency dispersion of the CCSD first hyperpolarizability is tackled in Figure 2 and Table 8 for compounds **1** and **2**, where QRF-CCSD results are used as a benchmark. At large wavelength (1900 nm), the errors made by both approximations remain smaller than 3% and 7%: the additive approximation underestimates by 5.8% and 6.1% the β_{HRS} QRF-CCSD reference value of compounds **1** and **2**, respectively whereas within the multiplicative scheme, these drop to 2.5% and 1.9%. Then, the underestimations grow with the frequency. At 1064 nm, which is a typical laser wavelength, the multiplicative scheme underestimates the HRS first hyperpolarizability of **1** by 9.3% and it goes up to 18.3% when using the additive scheme. Going further towards smaller wavelengths, the errors increase and the additive approximation gets very poor.

An alternative way to determine the correlated dynamic first hyperpolarizabilities consists in applying the two-state approximation (TSA),⁴⁹ which assumes that only one excited state contributes to the first hyperpolarizability and that this excited state governs the frequency dispersion. In the case of SHG β ,

$$\beta(-2\omega; \omega, \omega) \approx \beta(0; 0, 0) \frac{\omega_{ge}^4}{(\omega_{ge}^2 - 4\omega^2)(\omega_{ge}^2 - \omega^2)} \quad (\text{II.B.68})$$

where $\hbar\omega_{ge} = \omega_{ge}$ is the excitation energy. It can thus be employed at different levels of approximation. At the CCSD/6-31+G(d) level of theory the excitation energy of compounds **1** and **2** amounts to 4.72 eV and 4.58 eV and the oscillator strengths to 0.43 and 0.56, respectively. The selection of the *dominant* excited state is not always straightforward but for π -conjugated systems it is the lowest-energy excited state with large oscillator strength. In such a case, Eq. (II.B.68) provides an excellent approximation to the frequency dispersion of the first hyperpolarizability (Figure 2). Since for large systems it is not always possible to perform excitation energy

calculations at a highly-correlated level, an alternative consists in evaluating these at the TDHF (usually called RPA for Random Phase Approximation) level, which then provides a good approximation to the TDHF dispersion applied within the multiplicative scheme (Figure 2) and underestimates the QRF-CCSD values. For compound **1** and **2**, the RPA excitation energies are 5.12 eV and 5.06 eV and the oscillator strengths 0.36 and 0.43, respectively.

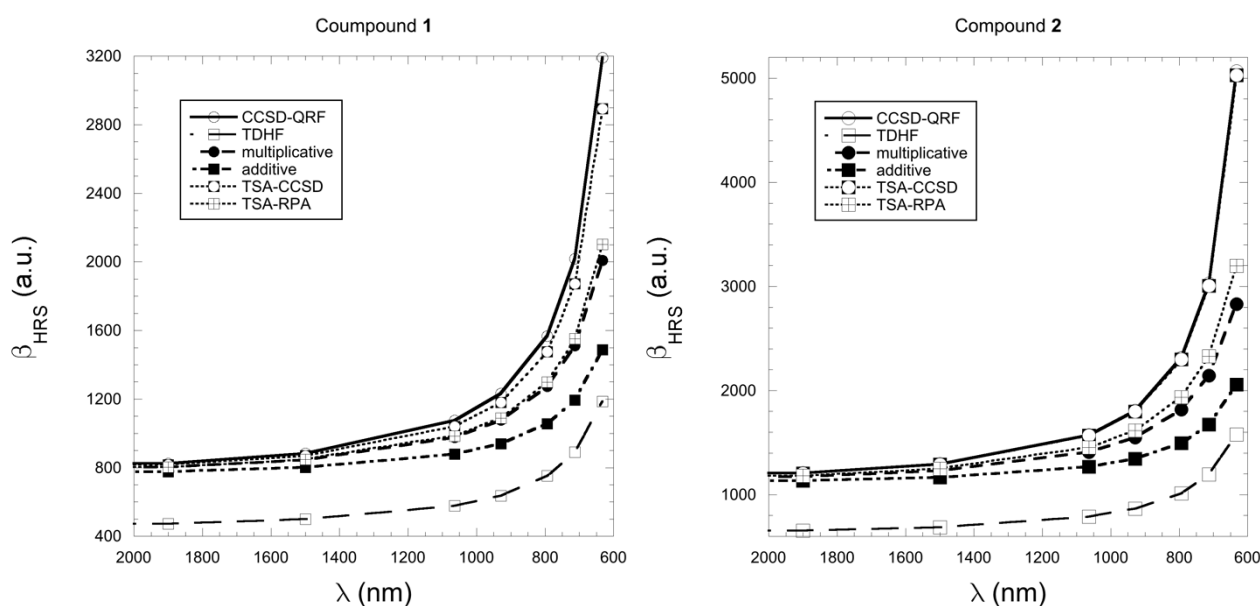


FIG. 2. Frequency dispersion of β_{HRS} for compounds **1** and **2**, as calculated using the multiplicative and additive schemes from unrelaxed-orbital CCSD static values and HF frequency dispersions. Comparisons are made with respect to the reference CCSD-QRF values as well as to TDHF results and two-state approximation dispersions. TSA-CCSD refers to Eq. 15 with excitation energies calculated at the CCSD level of theory whereas TSA-RPA when the excitation energies are calculated with the TDHF/RPA method. Lines are guides for the eyes.

This analysis was then extended to two other sets of molecules, i) two additional push-pull π -conjugated systems, $\text{H}_2\text{N}-\text{CH}=\text{CH}-\text{CH}=\text{CH}-\text{NO}_2$ and $\text{H}_2\text{N}-\text{C}\equiv\text{C}-\text{C}\equiv\text{C}-\text{NO}_2$ (Figure 3) and ii) small molecules including H_2O , CCl_4 , CHCl_3 , CH_2Cl_2 , CH_3CN , and CCl_3CN (Figure 4). For the two π -conjugated systems, the approximate schemes behave like for compounds **1** and **2**. For the polyenic linker, at 1900nm, an error of less than 8% is obtained for the additive scheme and 2% for the multiplicative one

while at 1064nm, the additive/multiplicative approximation becomes worst with 24%/8% of underestimation of the benchmark value. The errors made by the additive and multiplicative approximations for the polyynic linker are smaller than 5% and 1% at 1900nm and they increase to 15% and 5% at 1064nm, respectively. For the polyenic and polyynic linkers, the CCSD excitation energy values are 4.44 eV and 4.55 eV and the oscillator strengths 0.82 and 0.31, respectively. At the TDHF/6-31+G(d) level of theory the excitation energy of the polyenic and polyynic linkers amounts to 4.84 eV and 4.43 eV and the oscillator strengths to 0.98 and 0.16, respectively. Again, the description of the frequency dispersion is improved when employing two-state approximations, with TSA-RPA values close to the full TDHF results whereas the TSA-CCSD nicely approximate the QRF-CCSD first hyperpolarizabilities. Nevertheless, that approximation gets worse when the wavelength decreases.

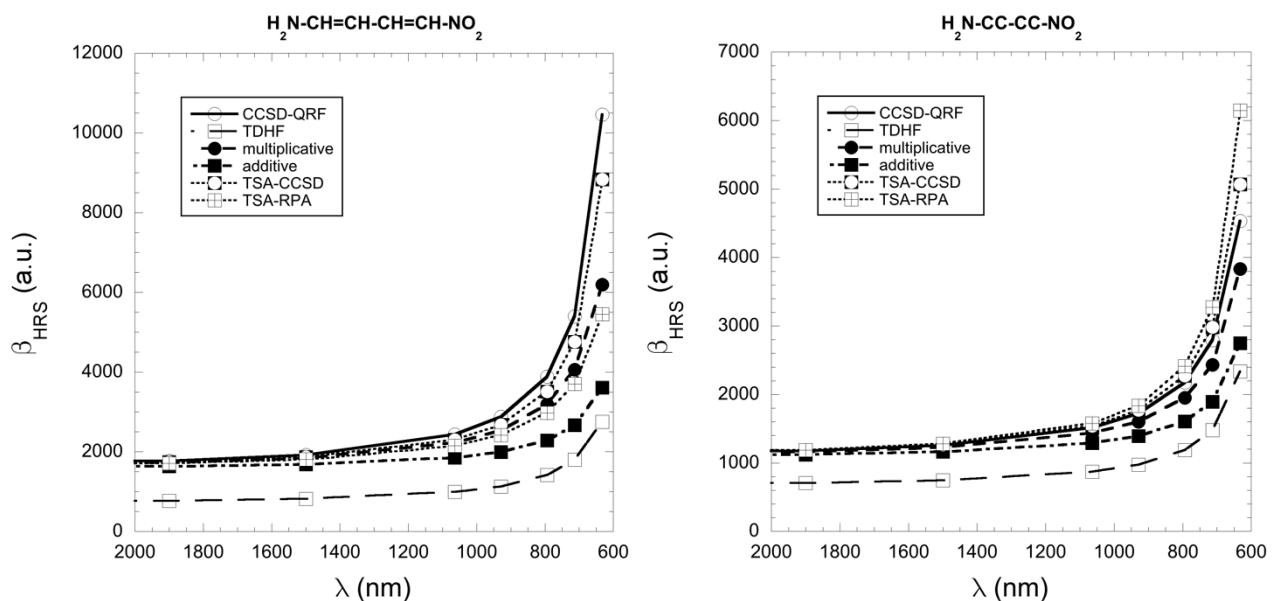


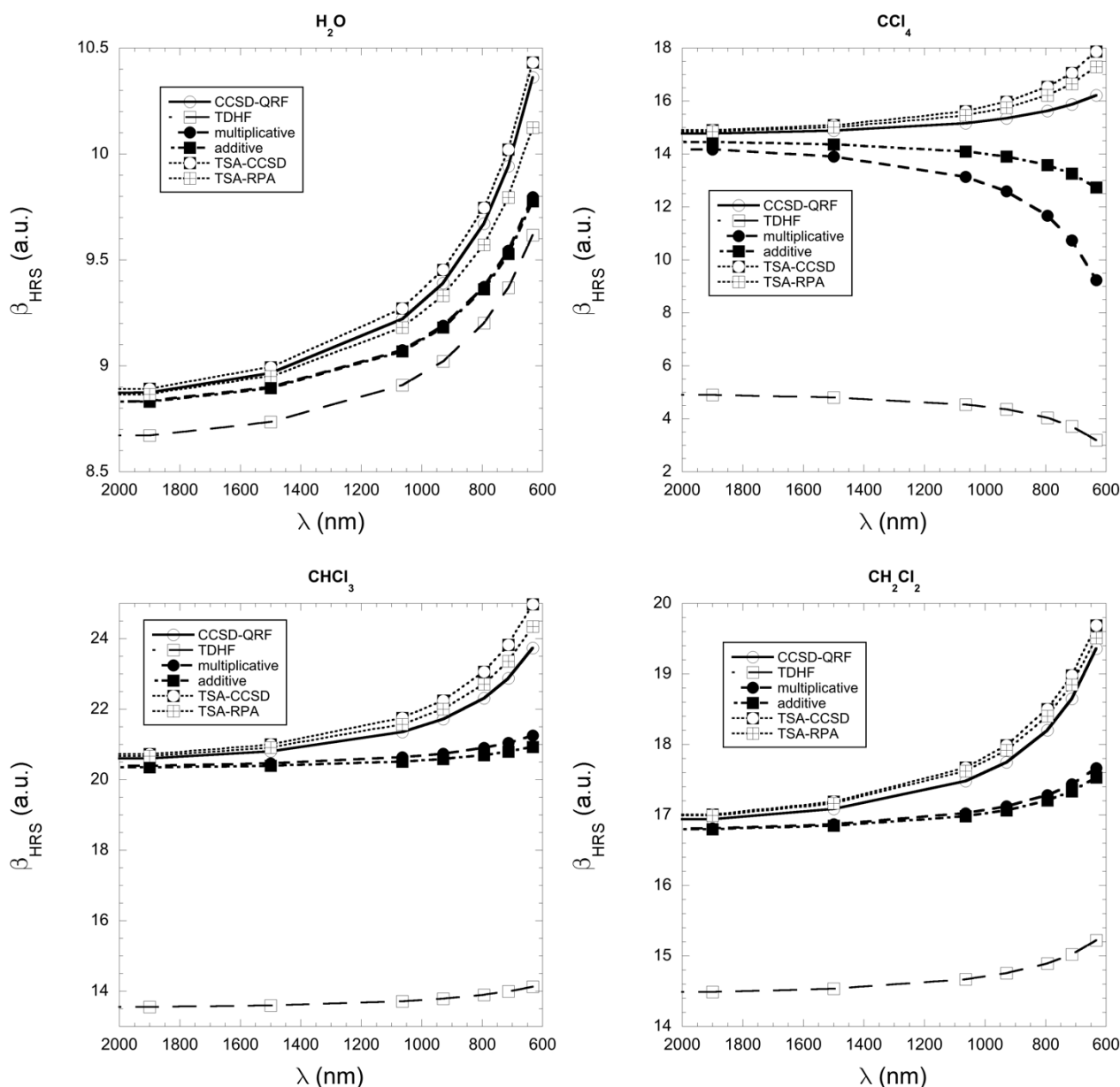
FIG.3. Frequency dispersion of β_{HRS} for $\text{H}_2\text{N-CH=CH-CH=CH-NO}_2$ and $\text{H}_2\text{N-C}\equiv\text{C-C}\equiv\text{C-NO}_2$ as calculated using the multiplicative and additive schemes from unrelaxed-orbital CCSD static values and HF frequency dispersions. Comparisons are made with respect to the reference CCSD-QRF values as well as to TDHF results and two-state approximation dispersions. TSA-CCSD refers to Eq. 15 with excitation energies calculated at the CCSD level of theory whereas TSA-RPA when the excitation energies are calculated with the TDHF/RPA method. Lines are guides for the eyes.

In the case of small molecules, the analysis of the results should account for the fact that the β tensors are no more governed by a single diagonal component, as it is the case for push-pull π -conjugated systems. However, the same approach has been adopted, showing first the limitations of the additive and multiplicative schemes and then illustrating how the two-state approximation can bridge partially the gap and account for the frequency dispersion. In addition to Figure 4, the main data are listed in Table 13. The main findings are: i) the additive and multiplicative schemes underestimate the frequency dispersion for H₂O, CHCl₃, CH₂Cl₂, and CH₃CN but overestimate this in the case of CCl₃CN, ii) for CCl₄ the frequency dispersion as described by the TDHF method leads to smaller and smaller β_{HRS} values when increasing the frequency, and it thus goes in opposite direction to the QRF-CCSD scheme, leading to wrong additive and multiplicative scheme estimates, iii) the TSA schemes improve the description of the frequency dispersion.

	ΔE (f) in eV	% error on β_{HRS} (1900 nm)			% error on β_{HRS} (1064 nm)		
		Additive	Multiplicative	TSA-CCSD	Additive	Multiplicative	TSA-CCSD
1	4.72 (0.43)	-6%	-3%	-1%	-18%	-9%	-3%
2	4.58 (0.56)	-6%	-3%	0%	-19%	-10%	0%
H ₂ N-(CH=CH) ₂ -NO ₂	4.44 (0.82)	-8%	-2%	-2%	-24%	-8%	-5%
H ₂ N-(C≡C) ₂ -NO ₂	4.55 (0.31)	-5%	-1%	0%	-15%	-5%	+2%
H₂O	10.7 (0.13)	0%	0%	0%	-2%	-2%	+1%
CCl₄	10.1 (0.21)	-2%	-4%	+1%	-7%	-13%	+3%
CHCl₃	10.0 (0.45)	-1%	-1%	+1%	-4%	-3%	+2%
CH₂Cl₂	11.1 (0.21)	-1%	-1%	0%	-3%	-3%	+1%
CCl₃CN	10.3 (0.66)	+2%	+10%	+2%	+8%	+37%	+6%
CH₃CN	12.7 (0.46)	-1%	0%	0%	-2%	-1%	-1%

TABLE 13. Performance of approximate schemes to describe the frequency dispersion of β_{HRS} at 1900 and 1064 nm together with the characteristics of the lowest-energy dipole-allowed excitation, evaluated at the CCSD level of approximation

In conclusions, describing at a high-level of theory, *i.e.* by including electron correlation effects, the frequency dispersion of the first hyperpolarizabilities is a difficult challenge when employing approximate schemes. These investigations on small and model compounds have found the following method ordering, from the less reliable to the best one: the additive scheme using TDHF dispersion, the multiplicative scheme using TDHF dispersion, two-state approximation based on RPA excitation energies, and two-state approximation based on CCSD excitation energies.



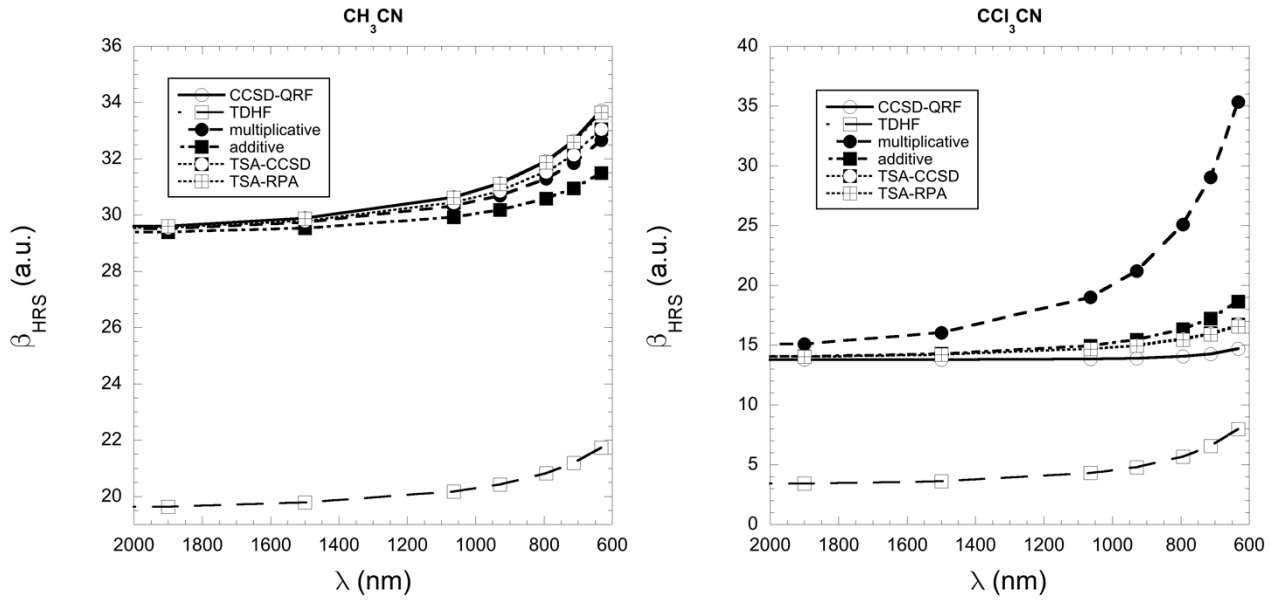


FIG.4. Frequency dispersion of β_{HRS} for H_2O , CCl_4 , CHCl_3 , CH_2Cl_2 , CH_3CN , and CCl_3CN as calculated using the multiplicative and additive schemes from unrelaxed-orbital CCSD static values and HF frequency dispersions. Comparisons are made with respect to the reference CCSD-QRF values as well as to TDHF results and two-state approximation dispersions. TSA-CCSD refers to Eq. 15 with excitation energies calculated at the CCSD level of theory whereas TSA-RPA when the excitation energies are calculated with the TDHF/RPA method. Lines are guides for the eyes.

II.C.5.g. Deviation from Kleinman's symmetry conditions

The importance of deviations from the Kleinman's symmetry conditions is assessed for the SHG β response of compounds **1** and **2** by analyzing the evolution of the following expression as a function of the frequency (Figure 5):

$$(\%)_{ij \rightarrow jii} = \frac{|\beta_{ij} - \beta_{jii}|}{|\beta_{ij} + \beta_{jii}|/2} 100 \quad (\text{II.B.69})$$

If Kleinman's conditions are satisfied, including in the static limit, the two components are identical and $(\%)_{ij \rightarrow jii}$ vanishes. For **1**, a strong deviation is observed for $(\beta_{zzx} \text{ vs. } \beta_{xzz})$ and $(\beta_{zzy} \text{ vs. } \beta_{yzz})$ whereas for $(\beta_{yyx} \text{ vs. } \beta_{xyy})$ and $(\beta_{xxy} \text{ vs. } \beta_{yxx})$ Kleinman's conditions remain valid for wavelengths as large as 1064 nm. However, the deviations amplitudes are not related to the amplitude of the β tensor components since, for example, β_{zzx} is about 17 times smaller than β_{zzy} . For the

overall response, *i.e.* β_{HRS} (see Figure 6), assuming or not the Kleinman's condition, has negligible effects because β_{yyy} is 17 times larger than the next largest component, β_{zzy} , and its frequency dispersion is dominating the response. For compound **2**, deviations as large as 5-15% are already observed at $\lambda = 1907$ nm, precluding the use Kleinman's conditions to simplify the β analysis. As seen in Figure 6, the Kleinman's conditions impact on the β_{HRS} for compound **2**, is also negligible because β_{zzz} is at least 37 times larger than the components that are impacted by Kleinman's conditions.

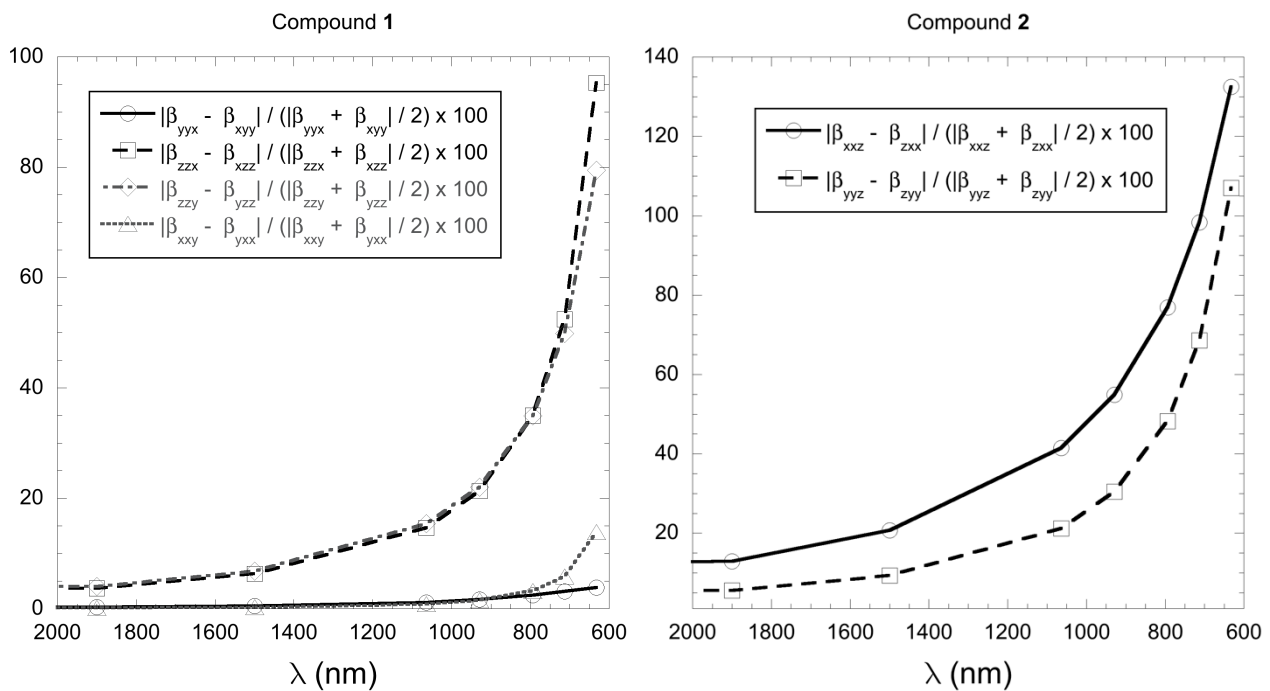


FIG.5. Deviations from Kleinman's symmetry conditions for the non-vanishing independent β tensor components of compounds **1** and **2**, as determined at the QRF-CCSD/6-31+G(d) level of theory. The lines are guides for the eyes.

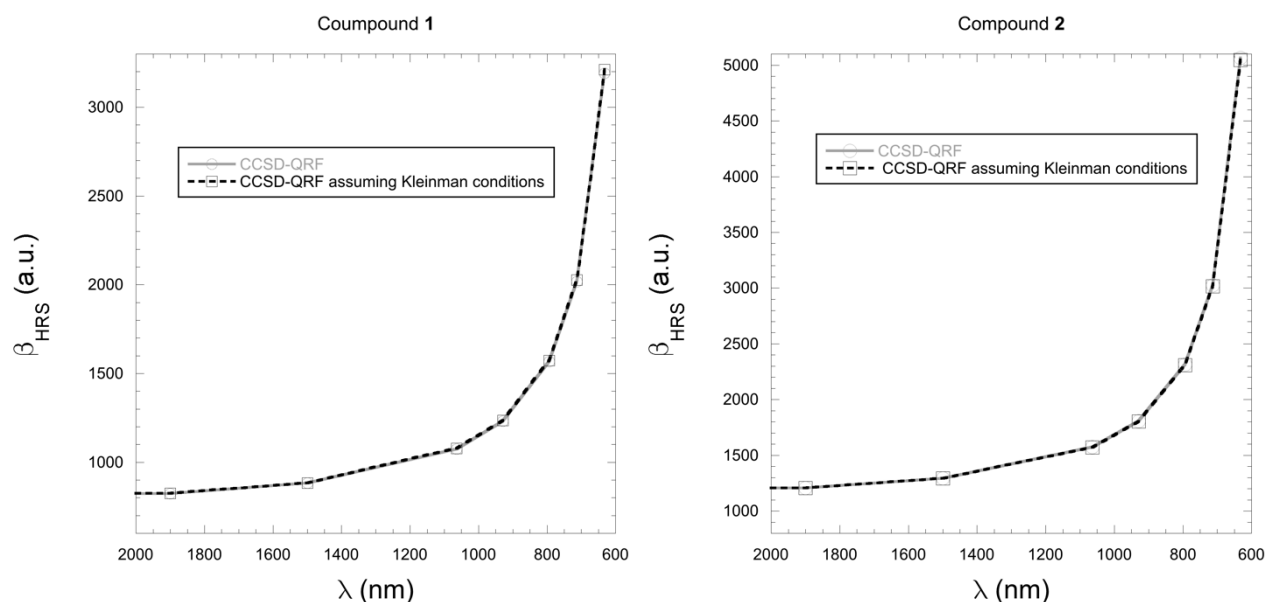


FIG.6. Frequency dispersion of β_{HRS} for compounds **1** and **2**, at the CCSD-QRF level of theory, assuming or not the Kleinman's conditions. The Kleinman's conditions have been assumed by taking the average of the two different components considered: $(\beta_{ij} + \beta_{ji})/2 = \beta_{ij}^{\text{eff}} = \beta_{ji}^{\text{eff}} = \beta_{ji}^{\text{eff}}$.

II.C.5.h. Conclusions

In this section, two aspects of computing the first hyperpolarizabilities of organic compounds have been tackled, i) the selection of a first hyperpolarizability value when employing the finite field procedure and ii) the frequency dispersion of these second-order nonlinear optical responses. It has been shown that accurate numerical derivatives can be obtained by resorting to the Romberg's method and by analyzing the Romberg's table in terms of two quantities, the field error and the iteration error. This leads to selecting in Romberg's table first hyperpolarizability values, which are in close agreement with those obtained within an analytical differentiation procedure. Then, the reliability of the multiplicative and additive approximate schemes to describe the frequency dispersion at correlated levels from using HF frequency dispersion has been confirmed to be limited to large wavelengths or far-from-resonance wavelength regions. Nevertheless, the multiplicative scheme appears to be more accurate than the additive one. Besides these schemes, it was demonstrated that the key quantity is the excitation energy of

the dominant excited state, which can be combined with the two-state approximation to predict rather accurate frequency-dependent values for push-pull π -conjugated systems whereas for small molecules, larger errors are observed, since their absorption spectra and first hyperpolarizabilities are not dominated by a few excited states. Finally, Kleinman's symmetry conditions are assessed, showing that for off-diagonal components, the deviations could be substantial. Therefore, if these off-diagonal components contributions are negligible with respect to the diagonal ones, good accuracy is achieved for experimentally-related quantities like β_{HRS} .

II.C.6. The ONIOM method

For large systems, direct determination of the first hyperpolarizability remains a real challenge. In this thesis, such large systems as fluorescent proteins and collagen-like models have been treated by the ONIOM (our Own N-layered Integrated molecular Orbital and molecular Mechanics) method.⁵⁰ It consists in partitioning the system in several layers treated at different levels of theory. In the two-layer ONIOM approximation, the system is partitioned in a high and a low layer. The high/low ONIOM layer is the part of the system, which is investigated by a high/low level of theory. In particular, to get the extrapolated ONIOM energy of a system, E_{ONIOM} , three calculations are needed. First, the energy of the entire (real) system is calculated with the low level method, $E_{\text{real}, \text{low}}$. The second and third ones provide the energy of the model system at the low and high levels of theory, $E_{\text{model}, \text{low}}$ and $E_{\text{model}, \text{high}}$, respectively. The extrapolated ONIOM energy reads then:

$$E_{\text{ONIOM}} = E_{\text{real}, \text{low}} - E_{\text{model}, \text{low}} + E_{\text{model}, \text{high}} \quad (\text{II.B.70})$$

The extrapolated ONIOM first hyperpolarizability tensor $\vec{\vec{\beta}}_{\text{ONIOM}}$ is obtained in the same way since the ONIOM scheme allows for a consistent treatment of the

molecular properties related to the responses with respect to static and dynamic electric fields⁵⁰:

$$\vec{\beta}_{ONIOM} = \vec{\beta}_{real, low} - \vec{\beta}_{model, low} + \vec{\beta}_{model, high} . \quad (\text{II.B.71})$$

Guidelines and pitfalls about the ONIOM method can be found in reference 50b. In this thesis, the ONIOM(QM:QM) scheme has been used where the electrostatic effects on the electronic structure of the model system are included by the real system but at a lower level of theory. Then, we avoid electronic embedding issues. Indeed, when considering ONIOM(QM:MM), this scheme corresponds to mechanical embedding, the electrostatic effects of the real system atoms are not included for the model system. The ONIOM(QM:MM) formalism has been next extended to include electronic embedding.^{50c}

The ONIOM method was already employed on fluorescent proteins to characterize mechanistic pathways, magnetic exchange coupling constants, as well as absorption, emission, and Raman spectra⁵¹. For example in reference 51a, competitive mechanistic pathways for green-to red photoconversion in the fluorescent protein Kaede has been investigated with explicit consideration of protein environment by ONIOM(B3LYP:AMBER) calculations, the stepwise “E1” and “E1cb” mechanisms have been suggested to occur and to compete with each other in the electronic ground state. In reference 51b, the nature of ON- and OFF-states of reversible photoswitching fluorescent protein Dronpa has been studied, where the vertical absorption and emission energies of the ON- and OFF-states were studied by ONIOM(QM:MM) calculations.

II.D. *Solvent effects*

This thesis investigates biological systems, which are in aqueous solution. Thus, the solvent effects need to be accounted in order to reproduce the proper

first hyperpolarizability of those systems. This was achieved by using the polarizable continuum model⁵² (PCM) in its integral equation formalism⁵³ (IEF). In this section, the very basics of the PCM method are presented without going too much into the details. For more details about this method see the excellent review of Tomasi *et al.*^{53c} In continuum solvation model, the total Hamiltonian is reduced to an effective Hamiltonian:

$$\hat{H}_{eff} = \hat{H}_{solute} + \hat{V}_{int}, \quad (\text{II.C.1})$$

where there are two parts, the Hamiltonian of the solute $\hat{H}_{solute}$ and the solute-solvent interaction potential $\hat{V}_{int}$. The basic concept of all continuum models is to put the solute into a void cavity within a continuous dielectric medium mimicking the solvent. The cavity is defined as an interlocked superposition of atomic spheres with radii close to the van der Waals values. The physics of the electrostatic solute-solvent interaction is simple. The charge distribution $\rho(\vec{r})$ of the solute, inside the cavity, polarizes the dielectric continuum, which in return polarizes the solute charge distribution. Evaluating the solute-solvent interactions corresponds to a self-consistent process, which is numerically solved following an iterative procedure. Reference is often made to the *solvent reaction field* for the interaction potential obtained by continuum models. The apparent surface charge spread on the cavity surface, which depends on the dielectric constant of the solvent, is expressed by

$$\sigma(\vec{s}) = \frac{\epsilon - 1}{4\pi\epsilon} \vec{E} \cdot \vec{n}, \quad (\text{II.C.2})$$

where $\vec{n}$ is a unitary vector perpendicular to the outside surface of the cavity. The total electric field $\vec{E}$ can be partitioned into two terms, one depending on the charge distribution of the solute $\rho(\vec{r})$ and the other relative to the apparent surface charge $\sigma(\vec{s})$ associated to the solvent:

$$\sigma(\vec{s}) = \frac{\epsilon - 1}{4\pi\epsilon} (\vec{E}_\rho + \vec{E}_\sigma) \cdot \vec{n}. \quad (\text{II.C.3})$$

This expression can be rewritten as a function of the directional derivatives of the electrostatic potential:

$$\sigma(\vec{s}) = \frac{\epsilon - 1}{4\pi\epsilon} \frac{\partial}{\partial \vec{n}} (V_\rho + V_\sigma)_{in}, \quad (\text{II.C.4})$$

where $V_\rho(\vec{r}) = \int \frac{\rho(\vec{r}')}{|\vec{r}' - \vec{r}|} d\vec{r}'$ and, since the cavity of the surface is approximated in terms of a set of finite elements (called tesserae), small enough to consider $\sigma(\vec{s})$ almost constant within each tesserae, it is possible to define a set of point charges, q_k , in terms of the local value of $\sigma(\vec{s})$ on each of these tesserae times the corresponding area A_k :

$$V_\sigma(\vec{r}) \simeq \sum_k \frac{\sigma(\vec{s}_k) A_k}{|\vec{r} - \vec{s}_k|} = \sum_k \frac{q_k}{|\vec{r} - \vec{s}_k|}. \quad (\text{II.C.5})$$

In a practical manner, this leads to the self-consistent reaction field where within each SCF cycle, a second iterative procedure is introduced to evaluate $V_\sigma(\vec{r})$ and $\sigma(\vec{s})$. It must be noted that this implementation has been beneficially replaced by a continuous surface charge approach⁵⁴ in Gaussian09⁴⁵.

Note that some applications combine the IEFPCM method within the ONIOM scheme and, in particular, the ONIOM-PCM/X method,⁵⁵ where the reaction field is computed separately in each independent sub-calculation always using the cavity of the real-system.

II.E. Unit sphere representation of the first hyperpolarizability tensor

The need to provide insight and intuition on the relationship between SHG properties and the molecule itself has led Tuer *et al.* to develop a tool for the visualization of the first hyperpolarizability tensor, the so-called unit sphere representation.⁵⁶ An effective first hyperpolarizability dipole

$$\vec{\beta}_{eff} = \vec{\beta} : \hat{E}(\theta, \phi) \hat{E}(\theta, \phi) \quad (\text{II.D.1})$$

is defined to sample all possible polarizations where $\hat{E}(\theta, \phi)$ is the unit vector of the polarization of the incident electric field with a polarization defined in spherical coordinates (θ, ϕ) . A unit sphere is mapped out where at each point, the electric field is considered to be along the radius vector and then the corresponding $\vec{\beta}_{eff}$ vector is plotted. For example figure 1 sketches the unit sphere representation for the β tensor of the CCl_3CN molecule.⁵⁷

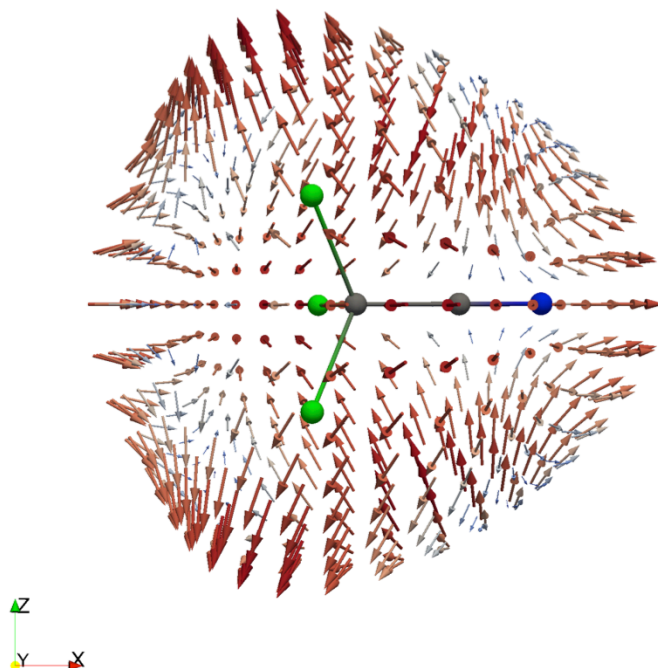


FIG. 7. Unit sphere representation for the molecule of CCl_3CN (CCSD(T)/(d)-aug-cc-pVTZ/PCM level of calculation, $\beta_{\text{HRS}} = 46.35$ a.u. and $\text{DR}=1.51$).

II.F. Geometry optimizations

In this thesis, all the geometry optimizations were performed using the DFT approach. During the optimization procedure, the geometry is adjusted until the global minimum of the potential energy surface is found, the Berny algorithm is used with GEDIIS⁵⁸ in redundant internal coordinates⁵⁹. In the special case of the fluorescent proteins, in order to keep the natural conformation of the chromophore and its immediate surroundings, the structures were taken from X-ray diffraction

data, the missing H atoms has been added and a partial optimization of their positions were performed while the rest of the system is frozen.

II.G. Conclusions

In this second Chapter, different methods have been presented. Among these, the ONIOM scheme is widely used for the large systems considered in the two next Chapters. The accuracy of the quantum chemistry methods selected to describe the different layers of the ONIOM approach is estimated in Chapter V on small model systems. Their time scaling and limitations also motivated the practical choice of methods. In table 14, the time scaling of the different levels of theory considered here is presented where N is a relative measure of the size of the system.⁶⁰ It must be noted that the CCSD and CCSD(T) methods encounter difficulties to converge the iterative procedure when diffuse functions are included in the basis set. To evaluate the first hyperpolarizability, the TDHF and TDDFT methods are used. And in order to have a precise picture of the electron correlation effects, when needed and if it is possible, the MP2 method is used within the finite field approach.

Methods	Scaling
HF	N^3
MP2	N^5
MP3	N^6
MP4	N^7
CCSD	N^6
CCSD(T)	N^6 + one non iterative step N^7
DFT	N^3

TABLE 14. Comparison of the scaling at different levels of theory.

REFERENCES

1. Szabo, A.; Ostlund, N., *Modern quantum chemistry*. Macmillan Publishing Co., Inc, New-York: 1982.
2. Møller, C.; Plesset, M. S., *Phys. Rev.* **1934**, *46* (7), 618-622.
3. Helgaker, T.; Jørgensen, P.; Olsen, J., *Molecular electronic-structure theory*. John Wiley & Sons, LTD: 2000.
4. (a) Bossi, A.; Licandro, E.; Maiorana, S.; Rigamonti, C.; Righetto, S.; Stephenson, G. R.; Spassova, M.; Botek, E.; Champagne, B., *J. Phys. Chem. C* **2008**, *112* (21), 7900-7907; (b) Mancois, F.; Pozzo, J.-L.; Pan, J.; Adamietz, F.; Rodriguez, V.; Ducasse, L.; Castet, F.; Plaquet, A.; Champagne, B., *Chem. Eur. J.* **2009**, *15*, 2560-2571; (c) Sergeyev, S.; Didier, D.; Boitsov, V.; Teshome, A.; Asselberghs, I.; Clays, K.; Vande Velde, C. M.; Plaquet, A.; Champagne, B., *Chemistry* **2010**, *16* (27), 8181-90.
5. Champagne, B.; Kirtman, B., *J. Chem. Phys.* **2006**, *125* (2), 24101.
6. Hammond, J. R.; Kowalski, K., *J. Chem. Phys.* **2009**, *130* (19), 194108.
7. Watts, J. D.; Gauss, J.; Bartlett, R. J., *J. Chem. Phys.* **1993**, *98* (11), 8718-8733.
8. Kohn, W., *Rev. Mod. Phys.* **1999**, *71* (5), 1253-1266.
9. (a) Parr, R. G.; Yang, W., *Density-functional theory of atoms and molecules*. Oxford University Press, New York: 1989; (b) Koch, W.; Holthausen, M. C., *A chemist's guide to density functional theory*. Wiley-VCH: 2001.
10. Hohenberg, P.; Kohn, W., *Phys. Rev.* **1964**, *136* (3B), B864-B871.
11. Kohn, W.; Sham, L. J., *Phys. Rev.* **1965**, *140* (4A), A1133-A1138.
12. Becke, A. D., *J. Chem. Phys.* **1993**, *98* (2), 1372-1377.
13. Lee, C.; Yang, W.; Parr, R. G., *Phys. Rev. B* **1988**, *37* (2), 785-789.
14. (a) Becke, A. D.; Roussel, M. R., *Phys. Rev. A* **1989**, *39* (8), 3761-3767; (b) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E., *Phys. Rev. Lett.* **2003**, *91* (14), 146401.

15. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., *J. Phys. Chem.* **1994**, *98* (45), 11623-11627.
16. Becke, A. D., *J. Chem. Phys.* **1993**, *98* (7), 5648-5652.
17. Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A., *Chem. Phys. Lett.* **1997**, *275* (3–4), 151-160.
18. Yanai, T.; Tew, D. P.; Handy, N. C., *Chem. Phys. Lett.* **2004**, *393* (1–3), 51-57.
19. (a) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K., *J. Chem. Phys.* **2004**, *120* (18), 8425-8433; (b) Song, J.-W.; Hirose, T.; Tsuneda, T.; Hirao, K., *J. Chem. Phys.* **2007**, *126* (15), 154105-7.
20. Grimme, S., *J. Chem. Phys.* **2006**, *124* (3), 034108.
21. Schwabe, T.; Grimme, S., *Phys. Chem. Chem. Phys.* **2006**, *8* (38), 4398-401.
22. Yarkony, D., *Modern electronic structure theory*. World Scientific: 1995.
23. (a) Sekino, H.; Bartlett, R. J., *J. Chem. Phys.* **1986**, *85* (2), 976-989; (b) Karna, S. P.; Dupuis, M., *J. Comput. Chem.* **1991**, *12* (4), 487-504.
24. (a) Silverman, J. N.; van Leuven, J. C., *Phys. Rev.* **1967**, *162* (5), 1175-1194; (b) Nee, T.-S.; Parr, R. G.; Bartlett, R. J., *J. Chem. Phys.* **1976**, *64* (5), 2216-2225.
25. (a) Langhoff, P. W.; Karplus, M.; Hurst, R. P., *J. Chem. Phys.* **1966**, *44* (2), 505-514; (b) Caves, T. C.; Karplus, M., *J. Chem. Phys.* **1969**, *50* (9), 3649-3661.
26. (a) van Gisbergen, S. J. A.; Snijders, J. G.; Baerends, E. J., *Phys. Rev. Lett.* **1997**, *78* (16), 3097-3100; (b) Gisbergen, S. J. A. v.; Snijders, J. G.; Baerends, E. J., *J. Chem. Phys.* **1998**, *109* (24), 10644-10656; (c) Aiga, F.; Tada, T.; Yoshimura, R., *J. Chem. Phys.* **1999**, *111* (7), 2878-2888; (d) Larsen, H.; Jørgensen, P.; Olsen, J.; Helgaker, T., *J. Chem. Phys.* **2000**, *113* (20), 8908-8917; (e) Sałek, P.; Vahtras, O.; Helgaker, T.; Ågren, H., *J. Chem. Phys.* **2002**, *117* (21), 9630-9645.
27. (a) Hättig, C.; Christiansen, O.; Koch, H.; Jørgensen, P., *Chem. Phys. Lett.* **1997**, *269* (5–6), 428-434; (b) Hättig, C.; Jørgensen, P., *Theo. Chem. Acc.* **1998**, *100* (1-4), 230-240.

28. Christiansen, O.; Koch, H.; Jørgensen, P., *Chem. Phys. Lett.* **1995**, *243* (5–6), 409-418.
29. Jorgensen, P.; Helgaker, T., *J. Chem. Phys.* **1988**, *89* (3), 1560-1570.
30. Cohen, H. D.; Roothaan, C. C. J., *J. Chem. Phys.* **1965**, *43*, S34-S38, discussion S39.
31. (a) Bartlett, R. J.; Purvis, G. D., III, *Phys. Rev. A* **1979**, *20*, 1313-22; (b) Kurtz, H. A.; Stewart, J. J. P.; Dieter, K. M., *J. Comput. Chem.* **1990**, *11* (1), 82-87; (c) Sim, F.; Chin, S.; Dupuis, M.; Rice, J. E., *J. Phys. Chem.* **1993**, *97* (6), 1158-1163; (d) Medved', M.; Stachová, M.; Jacquemin, D.; André, J.-M.; Perpète, E. A., *THEOCHEM* **2007**, *847* (1–3), 39-46.
32. Mohammed, A. A. K.; Limacher, P. A.; Champagne, B., *J. Comput. Chem.* **2013**, *34* (17), 1497-1507.
33. Richardson, L. F.; Gaunt, J. A., *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **1927**, *226* (636-646), 299-361.
34. (a) Sekino, H.; Bartlett, R. J., *Chem. Phys. Lett.* **1995**, *234* (1–3), 87-93; (b) Sekino, H.; Bartlett, R. J., *J. Chem. Phys.* **1993**, *98* (4), 3022; (c) Sekino, H.; Bartlett, R. J., *J. Chem. Phys.* **1991**, *94* (5), 3665-3669.
35. (a) Rice, J. E.; Handy, N. C., *Int. J. Quant. Chem.* **1992**, *43* (1), 91-118; (b) Rice, J. E.; Handy, N. C., *J. Chem. Phys.* **1991**, *94* (7), 4959-4971.
36. Dalskov, B. E. K.; Jensen, H. J. A.; Oddershede, J., *Mol. Phys.* **1997**, *90* (1), 3-14.
37. Pluta, T.; Sadlej, A. J., *J Chem Phys* **2001**, *114* (1), 136-146.
38. Jacquemin, D.; Champagne, B.; Hättig, C., *Chem. Phys. Lett.* **2000**, *319* (3–4), 327-334.
39. Aiga, F.; Itoh, R., *Chem. Phys. Lett.* **1996**, *251* (5-6), 372-380.
40. Kleinman, D. A., *Phys. Rev.* **1962**, *126* (6), 1977-1979.
41. Dailey, C. A.; Burke, B. J.; Simpson, G. J., *Chem. Phys. Lett.* **2004**, *390* (1–3), 8-13.

42. Nakano, M.; Kishi, R.; Nitta, T.; Kubo, T.; Nakasuji, K.; Kamada, K.; Ohta, K.; Champagne, B.; Botek, E.; Yamaguchi, K., *J. Phys. Chem. A* **2005**, *109* (5), 885-91.
43. Verbiest, T.; Clays, K.; Rodriguez, V., *Second-order nonlinear optical characterization techniques: An introduction*. CRC Press: New-York, 2009.
44. Bersohn, R.; Pao, Y.-H.; Frisch, H. L., *J. Chem. Phys.* **1966**, *45* (9), 3184-3198.
45. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., *Gaussian 09*, revision a.01. Wallingford CT, 2009.
46. *Dalton, a molecular electronic structure program, release dalton2011*, 2011.
47. Salter, E. A.; Sekino, H.; Bartlett, R. J., *J Chem Phys* **1987**, *87* (1), 502-509.
48. O'Neill, D. P.; Kallay, M.; Gauss, J., *J. Chem. Phys.* **2007**, *127* (13), 134109.
49. Oudar, J. L.; Chemla, D. S., *J. Chem. Phys.* **1977**, *66* (6), 2664-2668.
50. (a) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J., *THEOCHEM* **1999**, *461-462* (0), 1-21; (b) Clemente, F. R.; Vreven, T.; Frisch, M. J., Getting the most out of oniom: Guidelines and pitfalls. In *Quantum biochemistry*, Wiley-VCH Verlag GmbH & Co. KGaA: 2010; pp 61-83; (c)

- Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J., *J. Chem. Theo. Comp.* **2006**, *2* (3), 815-826.
51. (a) Li, X.; Chung, L. W.; Mizuno, H.; Miyawaki, A.; Morokuma, K., *J Phys Chem B* **2010**, *114* (49), 16666-75; (b) Li, X.; Chung, L. W.; Mizuno, H.; Miyawaki, A.; Morokuma, K., *J Phys Chem B* **2010**, *114* (2), 1114-26; (c) Matsuura, A.; Hayashi, T.; Sato, H.; Takahashi, A.; Sakurai, M., *Chem. Phys. Lett.* **2010**, *484* (4-6), 324-329; (d) Ma, Y.; Sun, Q.; Li, Z.; Yu, J. G.; Smith, S. C., *J Phys Chem B* **2012**, *116* (4), 1426-36; (e) Bhattacharya, D.; Panda, A.; Shil, S.; Goswami, T.; Misra, A., *Phys Chem Chem Phys* **2012**, *14* (19), 6905-6913.
 52. Miertuš, S.; Scrocco, E.; Tomasi, J., *Chem. Phys.* **1981**, *55* (1), 117-129.
 53. (a) Cancès, E.; Mennucci, B.; Tomasi, J., *J. Chem. Phys.* **1997**, *107* (8), 3032-3041; (b) Mennucci, B.; Cancès, E.; Tomasi, J., *J. Phys. Chem. B* **1997**, *101* (49), 10506-10517; (c) Tomasi, J.; Mennucci, B.; Cammi, R., *Chem. Rev.* **2005**, *105* (8), 2999-3094.
 54. Scalmani, G.; Frisch, M. J., *J. Chem. Phys.* **2010**, *132* (11), -.
 55. Vreven, T.; Mennucci, B.; da Silva, C. O.; Morokuma, K.; Tomasi, J., *J. Chem. Phys.* **2001**, *115* (1), 62-72.
 56. (a) Tuer, A.; Krouglov, S.; Cisek, R.; Tokarz, D.; Barzda, V., *J. Comput. Chem.* **2011**, *32* (6), 1128-34; (b) Tuer, A. E.; Krouglov, S.; Prent, N.; Cisek, R.; Sandkuijl, D.; Yasufuku, K.; Wilson, B. C.; Barzda, V., *J. Phys. Chem. B* **2011**, *115* (44), 12759-69.
 57. Castet, F.; Bogdan, E.; Plaquet, A.; Ducasse, L.; Champagne, B.; Rodriguez, V., *J. Chem. Phys.* **2012**, *136* (2), 024506.
 58. Li, X.; Frisch, M. J., *J. Chem. Theo. Comp.* **2006**, *2* (3), 835-839.
 59. Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J., *J. Comput. Chem.* **1996**, *17* (1), 49-56.
 60. Jensen, F., *Introduction to computational chemistry*. Wiley: 2006.

Second-Order Nonlinear Optical Properties of Fluorescent Proteins for Second-Harmonic Imaging

E. De Meulenaere; I. Asselberghs; M. de Wergifosse; E. Botek; S. Spaepen; B. Champagne; J. Vanderleyden; K. Clays, *J. Mater. Chem.* **2009**, *19* (40), 7514-7519.

E. De Meulenaere; M. de Wergifosse; E. Botek; S. Spaepen; B. Champagne; J. Vanderleyden; K. Clays, *J. Nonlinear Opt. Phys. Mater.* **2010**, *19* (01), 1-13.

E. De Meulenaere; N. Nguyen Bich; M. de Wergifosse; K. Van Hecke; L. Van Meervelt; J. Vanderleyden; B. Champagne; K. Clays, *J. Am. Chem. Soc.* **2013**, *135* (10), 4061-4069.

M. de Wergifosse; *et al.*; *in preparation*.

III. Second-Order Nonlinear Optical Properties of Fluorescent Proteins for Second-Harmonic Imaging

This Chapter is divided in four sections presenting in a chronological order the different studies on fluorescent proteins done during this thesis. All this work has been done in a close collaboration with the group of Prof. K. Clays from KULeuven, in particular, Dr. E. De Meulenaere carried out all the HRS experiments. In section III.A., a very first set of fluorescent proteins (eGFP, eYFP, and DsRed) is investigated and confronted to the experiments made by our colleagues. With the same approach, in section III.B., a second set of two fluorescent protein (mStrawberry and mCherry) is also presented. These two first studies have just taken into account the structure of the chromophore in the simulations. In order to make better comparisons, we have used another approach with the explicit accounting of the surroundings and the electron correlation, *i.e.* the ONIOM (MP2:HF/6-31+G(d)) method. This approach is applied in the section III.C., on the new SHardonnay fluorescent protein, which has been successfully designed by our colleagues in order to find an alternative to eYFP (Enhanced Yellow Fluorescent Protein) as a dye for SHIM. In section III.D., a theoretical study on a set of six fluorescent proteins is presented, showing the requirement for an accurate modeling and giving quantitative comparisons with experiments.

III.A. *The case of eGFP, eYFP, and DsRed*

The second-order nonlinear optical properties of three fluorescent proteins (green, eGFP; yellow, eYFP; and red, DsRed) have been experimentally determined by frequency-resolved femtosecond hyper-Rayleigh scattering. As expected, DsRed, with its lower-energy bandgap between ground and excited state, exhibits the largest intrinsic hyperpolarizability. The anomalously low first hyperpolarizability for the yellow variant has been rationalized in terms of the centrosymmetrical arrangement between the phenolic Tyr203 residue and the chromophoric Tyr66 moiety, leaving the small imidazolinone moiety as the only effective non-centrosymmetric chromophore for second-order nonlinear effects. The experimental findings are corroborated by high-level computational results and suggest molecular engineering strategies to produce a full rainbow of fluorescent proteins with enhanced nonlinear optical properties.

III.A.1. Introduction

In a recent study, they have reported a quantitative value for the second-order nonlinear optical properties for Enhanced Green Fluorescent Protein (eGFP) and for the photoswitchable variant Dronpa.¹ These results have shown that the chromophore in these GFPs is capable of generating an appreciable second-harmonic signal, which could be used for SHIM. With the availability of a rainbow of fluorescent proteins,² we describe here the extension of this work by screening a first set of three variants, namely eGFP as a benchmark, eYFP and DsRed (for the chromophoric structure, see Fig. 1). While the linear optical properties, *i.e.* the absorption and emission spectra, show the gradual trend from green over yellow to red upon extending the chromophore through larger conjugation or stacking interactions, the second-order NLO properties do not follow this trend. The smallest second-order NLO response is observed for eYFP. This striking observation is then rationalized in terms of the effective symmetry

of the actual chromophore. Our colleagues of the KULeuven have done the experimental part of this work. For consistency of the presentation, we, however, report here the whole experimental and theoretical characterization.

III.A.2. Materials and methods

III.A.2.a. Materials

The expression and purification of the recombinant proteins are not described here but can be found in the published article together with quality control based on mass spectrometry. The nonlinear optical measurements were performed in phosphate buffer (pH 7.3) with chromophore concentrations between 200 and 400 μ M.

III.A.2.b. Instrumental Methods

The linear optical characterization was performed on commercially available instruments. The absorption spectra were obtained on a Perkin Elmer Lambda 900 UV/Vis spectrometer. The fluorescence spectra were obtained with an Ocean Optics setup, consisting of a 430 nm LED (Kingbright L53MBC) with appropriate filter and focusing lens for the excitation. The emitted light was collected at a 90° angle using a fiber (Ocean Optics QP400-2-VIS/BX) fitted with a collimating lens (Ocean Optics 74-UV) and detected by a charge coupled device (CCD) spectrometer (Ocean Optics USB4000).

For the experimental determination of the second-order nonlinear polarizability, or first hyperpolarizability β , of proteins, only HRS can be used.³ For the applicability of this incoherent second-order nonlinear scattering technique, it is essential to discriminate between the signal corresponding to the second-order scattering, and the signal corresponding to two-photon fluorescence, a third-order nonlinear contribution that can occur at the same wavelength. In the

spectral domain, it is not possible to make this distinction, since the narrow HRS signal at the second-harmonic wavelength is superimposed on the broad multiphoton fluorescence spectrum.⁴ Therefore, we have employed an amplitude-modulation (AM) frequency-resolved technique, to show whether or not there is a multiphoton fluorescence contribution at the second-harmonic wavelength in the first place, and if so, to deduce an accurate, fluorescence-free hyperpolarizability value.⁵

This approach is the Fourier-transform in the frequency-domain of the more intuitive separation between immediate (hyper-Rayleigh) scattering and time-delayed (multi-photon) fluorescence in the time domain.⁶ We use the intrinsic high harmonic content of an ultrashort laser pulse and cross-correlation electronics to extract the amplitude and phase of the HRS response at successive higher AM frequencies. Any fluorescence contribution, characterized by a finite response time, exhibits demodulation for high frequencies. This shows up as a reduction of the apparent frequency-dependent hyperpolarizability, with an inflection point at the AM frequency corresponding to the inverse of the fluorescence lifetime. The high-frequency limit of the apparent β is the accurate, fluorescence-free β value. As an additional sign of the contribution of the delayed fluorescence, an increasing phase shift between a scattering reference and the HRS signal is observed. The simultaneous fitting of the amplitude and phase data ensures a precise estimate of the fluorescence-free hyperpolarizability, in addition to estimates for the fluorescence contribution and lifetime.^{5b} This technique has been shown to be useful to deduce the β values of a large number of fluorescent chromophores,⁷ including the different photoswitched or acid-switched states of the fluorescent protein Dronpa.¹

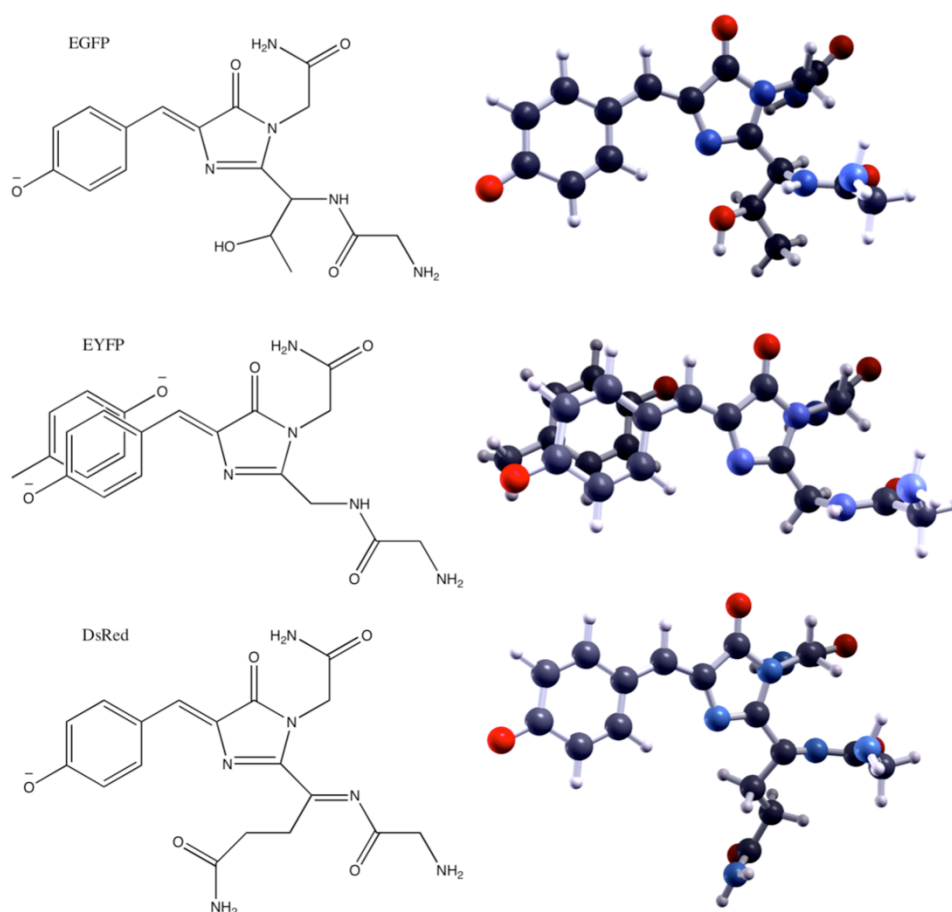


FIG. 1. Models of the isolated eGFP, eYFP, and DsRed chromophores

III.A.2.c. Computational methods and geometrical structures

The geometries of the chromophores surrounded by residues of the surrounding proteins were taken from X-ray diffraction.⁸ The positions of the H atoms were optimized at the DFT level using the B3LYP method and the 6-31+G(d) basis set. The structures are represented in Fig. 1. It should be noted that the model for eYFP and eGFP only differ in taking into account the centrosymmetrical arrangement of the side chain of the amino acid residue tyrosine 203 in close vicinity to the chromophoric backbone imidazolinone moiety for eYFP.

We calculated the dynamic first hyperpolarizability with the TDHF⁹ scheme while the static one was obtained by using the CPHF. The calculations in

TDHF were carried out for a wavelength of 1064/1900 nm as well as in the static limit. To take into account the solvent effects, the IEF-PCM was employed.¹⁰ The reported quantities are related to the actual HRS experiments with plane-polarized incident laser light and observation made perpendicular to the propagation plane.

All reported β values are given within the B convention. Although electron correlation is missing in this approach, the CPHF/TDHF scheme has often been found to provide qualitatively accurate first hyperpolarizabilities, certainly within a class of parent compounds.

The vertical excitation energies [$\Delta E_{ge} = \hbar\omega_{ge} = \hbar(\omega_e - \omega_g)$] and the excited state properties of the compounds were calculated at the Time-dependent DFT (TDDFT) level of approximation using the B3LYP hybrid exchange-correlation functional and the 6-311G(d,p) basis set. This approach has been found to provide UV/visible spectra in, at least, good qualitative agreement with respect to experimental data for both excitation energies and oscillator strength.¹¹

All calculations were performed with the Gaussian package.¹²

III.A.3. Results

The linear optical properties of the fluorescent proteins show the gradual bathochromic shift from the green (for eGFP) over yellow (for eYFP) to red (for DsRed). The UV/Vis spectral characteristics (wavelength of maximal absorption and wavelength of maximal emission) are given in Table 1. The experimental absorption spectra are shown in Figure 2, solid lines.

As for the nonlinear optical properties determined by hyper-Rayleigh scattering, all three fluorescent proteins exhibit the very typical demodulation of an apparent, modulation frequency-dependent β value, together with the phase shift between signal and pure scattering. (see Figure. 3)

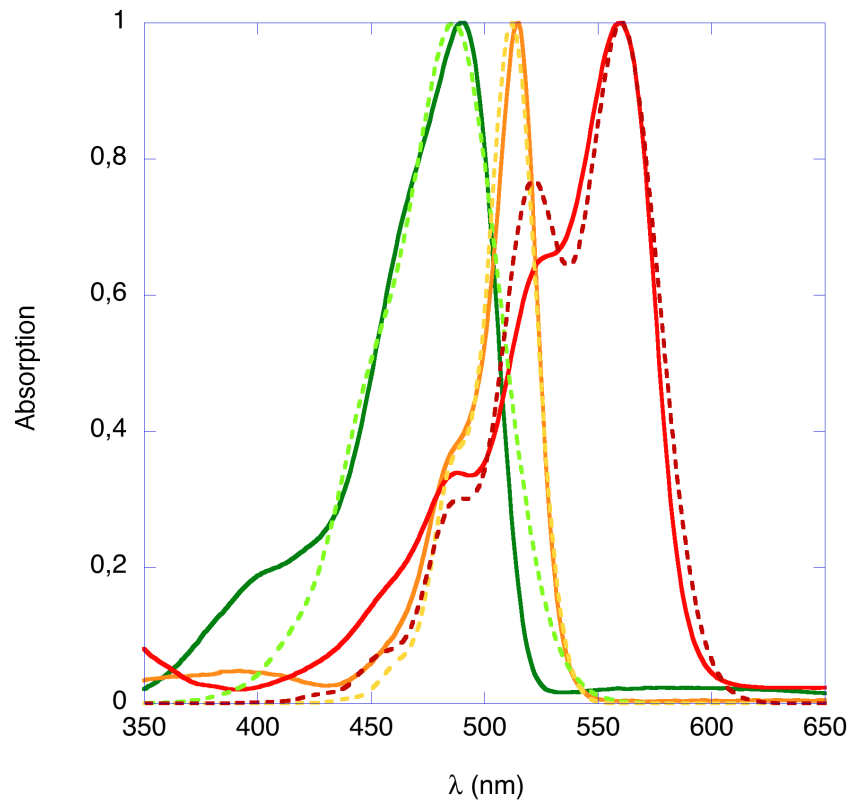


FIG. 2. Experimental (solid line) and simulated (dashed line) absorption spectra for eGFP, eYFP and DsRed.

From the simultaneous analysis of phase and amplitude data,^{5b} it has been possible to deduce the nonlinear optical parameters presented in Table 1, together with the linear optical properties.

An optical property is always dependent on the wavelength at which it has been determined. Because of electronic resonances between ground and excited electronic states, an enhancement of the optical properties occurs near such resonances. This is well-known for the linear optical properties, where there is a one-photon enhancement of both refractive index and absorption near such a resonance. The same occurs in nonlinear optics, both in the one-photon and in the two-photon regime. An enhancement of the first hyperpolarizability is observed due to the proximity of the second-harmonic wavelength with the wavelength of maximal absorption. Since the latter is different for the three proteins, the degree of enhancement is different.

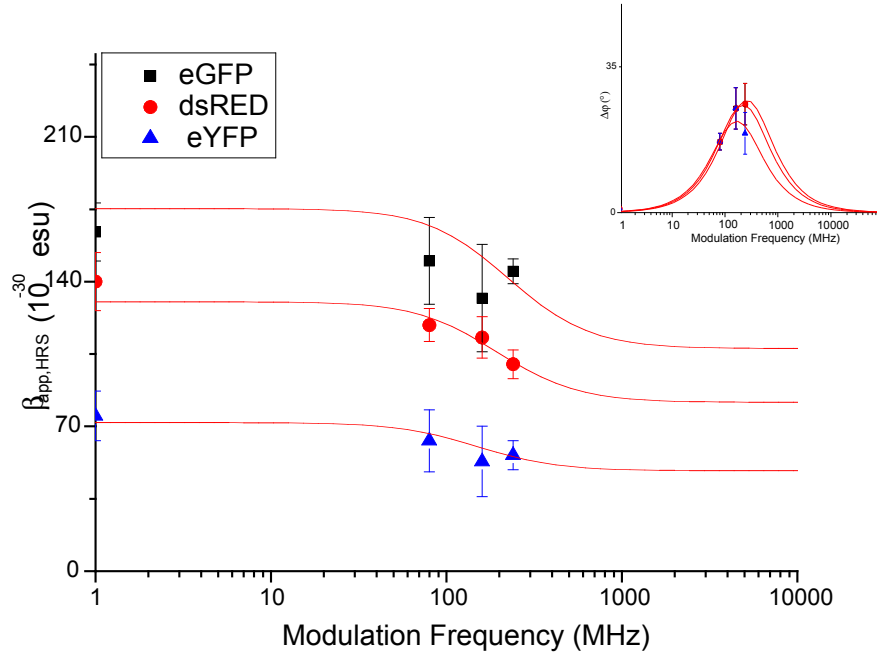


FIG. 3. The apparent, modulation frequency-dependent (or apparent), dynamic first hyperpolarizability $\beta_{\text{HRS},800}$ for eGFP (■), eYFP (▲) and DsRed (●), together with the growing phase difference between the total HRS signal and a scattering reference (shown in the inset).

TABLE 1. Linear and nonlinear optical properties of the 3 fluorescent proteins: wavelength of maximal absorption, $\lambda_{\text{max,abs}}$ (nm); wavelength of maximal emission, $\lambda_{\text{max,em}}$ (nm); dynamic first hyperpolarizability at 800 nm, $\beta_{\text{HRS},800}$ (10^{-30} esu); static first hyperpolarizability, $\beta_{\text{HRS},\infty}$ (10^{-30} esu)

Protein	eGFP	eYFP	DsRed
$\lambda_{\text{max,abs}}$ (nm)	488	513	558
$\lambda_{\text{max,em}}$ (nm)	507	527	583
$\beta_{\text{HRS},800}$ (10^{-30} esu)	107 +/- 17	49 +/- 5	81 +/- 8
$\beta_{\text{HRS},\infty}$ (10^{-30} esu)	33 +/- 5	18 +/- 2	39 +/- 4

In order to compare the experimental and theoretical first hyperpolarizabilities, a pre-treatment of the experimental data is necessary. Indeed, the experimental data are impacted by resonance conditions since the measurements are carried out at a frequency within the absorption band, slightly

above the maximum, whereas this effect is absent from the theoretical values because the poles of the TDHF responses are usually overestimated. Evaluating the magnitude of this effect is, however, difficult.¹³ The simplest way to estimate the impact of frequency dispersion in pseudo-dipolar chromophores is to consider the two-state approximation (TSA),¹⁴ which assumes that only one excited state contributes to the second-order NLO response. Considering a homogeneous damping γ and neglecting the non-resonant terms,¹⁵ the frequency-dispersion factor of the diagonal component of the first hyperpolarizability tensor, $F(\omega, \omega_{ge}, \gamma)$, reads:

$$F(\omega, \omega_{ge}, \gamma) = \frac{\beta_{ZZZ}(-2\omega, \omega, \omega)}{\beta_{ZZZ}(0, 0, 0)} = \frac{\omega_{ge}^2 (\omega_{ge} - i\gamma)^2}{\left([\omega_{ge} - i\gamma]^2 - 4\omega^2\right) \left([\omega_{ge} - i\gamma]^2 - \omega^2\right)}$$

where g and e are the two electronic, ground and excited, states contributing to the first hyperpolarizability, and $\hbar\omega_{ge}$ is the corresponding excitation energy. Thus, for a given γ value, since $\hbar\omega_{ge}$ (eGFP) $>$ $\hbar\omega_{ge}$ (eYFP) $>$ $\hbar\omega_{ge}$ (DsRed) and $\hbar\omega_{ge} = 1.55$ eV, the resonance enhancement will increase from DsRed to eGFP. Table 2 reports the extrapolated static β values obtained by setting γ to zero as well as by choosing γ values matching the HWHM of the corresponding absorption spectra, i.e. $\gamma \approx 1.2$ of the $e^{-[(\omega - \omega_{ge})/\gamma]^2}$ Gaussian function. Using the simple TSA expression without homogeneous damping reverts the β amplitude ordering between eGFP and DsRed, which is consistent with β enhancement according to the length of the π -conjugated path. Including a homogeneous damping in the TSA leads to a small change of these static extrapolated values.

These estimates of the dispersion factors have then been improved by incorporating an inhomogeneous broadening based on the absorption spectrum, which implicitly contains information on the distribution of the transition frequencies as well as by taking into account the vibronic structure of the excited

states. According to Campo et al.¹³ for an incoherent process like HRS, the dispersion factor reads:

$$\left| \frac{\beta^{HRS}(-2\omega; \omega, \omega)}{\beta^{HRS}(0; 0, 0)} \right| = \left\{ \int N(\omega'_{ge}) \left| F(\omega, \omega'_{ge}, \gamma) \right|^2 d\omega'_{ge} \right\}^{1/2}$$

Here, a value of $\gamma = 100 \text{ cm}^{-1}$ was chosen for the homogeneous damping while $N(\omega'_{ge})$, the normalized distribution of the transition frequencies, is approximated by a Gaussian function reproducing the experimental HWHM of the corresponding main absorption band. Moreover, the distribution of the transition frequencies is improved by considering the single-mode vibronic model so that $N(\omega'_{ge})$ is replaced by:

$$N(\omega'_{ge}) = \sum_n^{vib. levels} \left[\frac{S^n e^{-S}}{n!} \right] e^{-[(\omega'_{ge} - \omega_n)/\gamma_{vib}]^2}$$

where $\omega_n = \omega_{ge} + n \omega_{vib}$ corresponds to the transition from the ground state to the n^{th} vibrational level of the excited state. The Huang-Rhys factors (S) were chosen to reproduce the shape of the absorption band. Values of ω_{vib} and γ_{vib} were also chosen such to reproduce the absorption spectra as accurately as possible. This corresponds to the same approach as the one adopted in a recent study on substituted indolinooxazolidine derivatives.¹⁶ The parameters for eGFP, eYFP, and DsRed are listed in Table 3 while the corresponding simulated absorption spectra are compared to the experimental spectra in Figure 2, demonstrating a very close relationship between the two. The DR values are rather close to 5, substantiating the assumption of dipolar structures where one diagonal (longitudinal) β tensor component is dominant, allowing therefore the use of TSA expressions.

TABLE 2. Comparison between calculated, measured, and extrapolated values of the first hyperpolarizabilities

	eGFP	eYFP	DsRed
$\beta_{\text{HRS}, 800} (10^{-30} \text{ esu})$	107 ± 17	49 ± 5	81 ± 8
$\beta_{\text{HRS}, \infty} (10^{-30} \text{ esu})$ TSA	33 ± 5	18 ± 2	39 ± 4
$\beta_{\text{HRS}, \infty} (10^{-30} \text{ esu})$ TSA (homog. damping)	35 ± 6	19 ± 2	41 ± 4
$\beta_{\text{HRS}, \infty} (10^{-30} \text{ esu})$ TSA (homog. + inhomog. damping, vibr.)	19 ± 3	17 ± 2	33 ± 3
$\beta_{\text{HRS}, \infty} (10^{-30} \text{ esu})$ HF/IEFPCM (solvent = water)	31	25	39
DR ($\lambda = \infty$) HF/IEFPCM (solvent = water)	3.99	4.20	5.47
$\beta_{\text{HRS}, 1900} (10^{-30} \text{ esu})$ HF/IEFPCM (solvent = water)	14	12	21
$\beta_{\text{HRS}, 1064} (10^{-30} \text{ esu})$ HF/IEFPCM (solvent = water)	25	21	44

TABLE 3. Best parameters fitting the UV/Vis absorption spectra

	eGFP	eYFP	DsRed
$\hbar\omega_{ge} \text{ (eV)}; \lambda_{eg} \text{ (nm)}$	2.54; 488	2.42; 513	2.21; 561
$\gamma_{\text{homogeneous}} \text{ (eV)}$	0.183	0.078	0.148
$\gamma_{\text{inhomogeneous}} \text{ (cm}^{-1}\text{)}$	100	100	100
S	0.40	0.32	0.7
$\omega_{\text{vib}} \text{ (cm}^{-1}\text{)}$	1700	1050	1380
$\gamma_{\text{vib}} \text{ (cm}^{-1}\text{)}$	1100	500	700

The relative position of the excitation energies was confirmed by TDDFT/B3LYP calculations, which predict vertical excitation energies of 2.90 eV, 2.83 eV, and 2.37 eV for eGFP, eYFP, and DsRed, respectively. Note that this global overestimation has also been noticed for other compounds including merocyanines.¹¹ Using these parameters, improved extrapolated static β responses were calculated (line 4 in Table 2). These show that DsRed presents a second-order linear response about twice as large as eGFP and eYFP. Moreover,

eYFP displays a slightly smaller β value than eGFP. These values are confirmed, at least on a qualitative ground, by the quantum chemical calculations (lines 5, 7 and 8 in Table 2), also showing that DsRed presents a β value about 50% higher than the two other species while β of eYFP is about 20% lower than for eGFP. It is interesting to note that the chromophoric model for DsRed indeed shows a longer conjugation path, but that this is not the case for eYFP.

III.A.4. Discussion

While we do not want to put too much emphasis on the absolute values obtained for the hyperpolarizabilities of these recombinant proteins, the relative values within a series that was prepared and measured under identical circumstances are certainly valuable. A larger first hyperpolarizability is expected for a smaller energy difference between HOMO and LUMO for the chromophore, congruent with a longer wavelength of maximal absorption. Hence, the largest static hyperpolarizability for DsRed with the longest wavelength of maximal absorption and emission was expected. Congruently, an intermediate hyperpolarizability value is expected for the eYFP with the intermediate wavelength of absorption and emission. However, the smallest value is observed, even smaller than for our benchmark. This is rationalized in terms of the strategy that was invoked to induce the redshift of the emission from green to yellow in eYFP. This strategy is not the simple enlargement of the conjugation in the chromophore, since the chromophore itself is identical for eGFP and eYFP. The shift to longer absorption and emission wavelength has been attributed to the π -stacking between the chromophore phenolic ring (from Tyr66) and the phenolic ring of Tyr203.⁹ While this induces a larger effective polarizable system, a sufficient condition for longer wavelength absorption and emission, it also introduces an inversion centre for part of the effective chromophore, namely in

between the two phenolic moieties. Since for second-order nonlinear effects, centrosymmetric structures do not contribute, this results in an effectively shorter chromophore as far as any second-order nonlinear effects are considered. This hypothesis is illustrated in Fig. 4.

It is interesting to observe that, while GFP has been used in second-harmonic imaging,¹⁷ and has been shown to be amenable to second-order nonlinear light scattering,¹ YFP has been reported not to produce a detectable second-harmonic signal.¹⁸ Our experimental results confirm that at 800 nm, the second-order nonlinear optical polarizability of eYFP is less than half that of GFP which would result in a signal that is more than four times weaker. While this is only very qualitative, both the previous report¹⁸ and our recent results do show that the yellow emitting chromophore in YFP is not optimal for second-harmonic imaging.

Finally, the red shift of the absorption spectrum for DsRed suggests larger resonance enhancement when using a longer fundamental wavelength. The electronic transition occurring at 558 nm suggests the use of a (nano- or picosecond) Nd/YAG laser (1064 nm fundamental wavelength). However, this wavelength is already out of the transmission window for aqueous environments such as tissues and cells (700–950 nm). The latter wavelength region is typically covered by femtosecond Ti:sapphire lasers, also capable of providing much higher peak power from lower total energy in femtosecond pulses, highly advantageous for any nonlinear optical effect.

III.A.5. Conclusions

We have experimentally and theoretically evaluated the potential of 3 fluorescent proteins for use in second-harmonic imaging. eGFP, eYFP and DsRed form a small subset of the available rainbow of fluorescent polymers, engineered

to have red-shifted absorption and emission spectra, i.e. linear optical properties. We have shown that the different engineering strategies for red-shifted linear optical properties have drastically different consequences for second-order nonlinear optical properties. For DsRed, the extended conjugation path leads to both a strong red-shift and a larger first hyperpolarizability. However, the centrosymmetric stacking interaction in YFP, resulting in a weak red-shift to the yellow, is counterproductive for second-order NLO properties, where non-centrosymmetry is a crucial prerequisite.

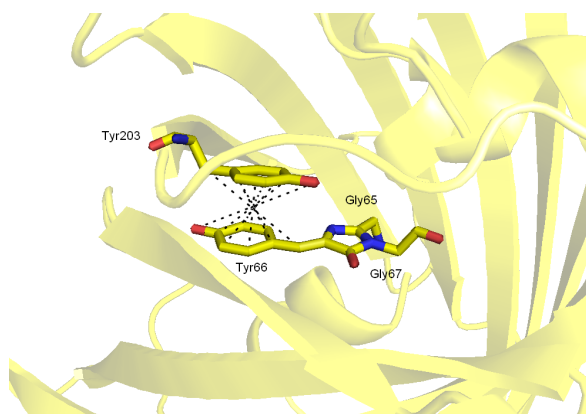


FIG. 4. Presentation of the chromophore of eYFP situated in the protein barrel structure shown as a yellow cartoon (PDB-code 1YFP). The chromophore and tyrosine 203 are shown as yellow sticks and show a clear inversion centre between the two phenolic moieties (indicated by dashed lines).^{8a} This suggests that the remaining effective chromophore for second-order nonlinear optics is limited to the small imidazolinone moiety only. The picture was generated using Pymol.

Our observations provide guidelines for molecular engineering on how to produce a more complete rainbow of fluorescent proteins with optimized second-order nonlinear optical properties for combined TPEF and SHIM microscopy: i) extending the conjugation path even further, e.g. to provide chromophores for working in the NIR transmission window between 700 and 950 nm, will also enhance the first hyperpolarizability; ii) using stacking interaction to red-shift the absorption and emission spectra works, but this stacking should avoid creating an inversion centre. Therefore, the yellow emitting zFP538

protein¹⁹ with close interaction between its tyrosine 67 and histidine 202 should exhibit a larger second-order NLO response. Other monomeric fluorescent proteins will also be evaluated, to investigate the influence of the oligomerization, an effect that was hitherto not taken into account.

III.B. *The case of mStrawberry and mCherry*

The second-order nonlinear optical properties of two monomeric red fluorescent proteins, mStrawberry and mCherry, have been experimentally determined by frequency-resolved femtosecond hyper-Rayleigh scattering by our colleagues of the KULeuven. These proteins were found to exhibit a stronger nonlinear response than the previously described eGFP, eYFP and DsRed, confirming the trend that fluorophores with a more extended conjugated system, or a lower-energy band gap between ground and excited state, exhibit a larger static hyperpolarizability. Furthermore, these experimental data were complemented with quantum chemical calculations. A discrepancy was observed between experimental and theoretical results, but this could be explained by the chromophore model extracted from the available X-ray diffraction data. While eGFP showed a larger dynamic experimental response (β_{HRS}) due to the highest resonance enhancement, we measured an even higher signal for mCherry. Furthermore, mCherry also shows a better separation of the second harmonic signal and two-photon excited fluorescent signal, making this the preferred fluorescent protein for second harmonic imaging at 800 nm so far.

III.B.1. Introduction

In this section, we explore the nonlinear optical properties of “mFruits”,²⁰ an artificially-obtained rainbow of monomeric derivatives of DsRed, like mOrange and mPlum, with their emission maxima between 530 nm and 650 nm, in

particular the earliest described and most commonly used mFruits: mStrawberry and mCherry. As for the preceding section, we report here the whole experimental and theoretical characterization.

III.B.2. Materials and methods

III.B.2.a. Materials

The proteins were expressed and purified as described previously.^{1, 21} After HPLC affinity purification, the samples were dialysed against a 20mM HEPES buffer containing 150mM NaCl at a pH of 7.5. The pRSET plasmids containing the mCherry and mStawberry gene were a generous gift from Roger Y. Tsien's lab. The proteins were checked for purity by SDS-PAGE and mass spectrometry. Concentration of the samples was determined using the extinction coefficient of the chromophore at the wavelength of maximal absorption.²⁰ The HRS experiments were carried out starting from 100 μ M protein solutions.

III.B.2.b. Instrumental methods

The determination of the linear properties as well as the hyper-Rayleigh scattering measurements were carried out using the same instruments and experimental setup as previously described. The HRS measurements were performed using a femtosecond pulsed Ti:sapphire laser emitting light at 800 nm. A 400 +/- 10 nm filter was used to select for signal at 400 nm.

III.B.2.c. Computational methods

The structures of the chromophores represented in Figure 5 were taken from X-ray diffraction data²⁰. For mCherry, the positions of the H atoms were optimized at the DFT level of theory using the B3LYP exchange-correlation functional and the 6-31+G(d) basis set. Based on the values of the XRD bond lengths and bond angles,

we had to consider two potential structures of the mStrawberry chromophore. The first one is the so-called mStrawberry enol form, where the enol group is stabilized by a hydrogen bond as shown in Fig. 5B. The other model is the so-called mStrawberry alcohol form as shown in Fig. 5C. Structure B was optimized like the mCherry chromophore but, since some groups displayed unexpected bond distances, we also optimized the group geometry under several dihedral angle constraints. Structure C was taken from a preliminary optimization step carried out at the PM3 level. It should be noticed that the π -conjugated system is the same for the mCherry and mStrawberry chromophores, which means that only geometrical constraints (differences in dihedral angles) can influence the p-electron conjugation.

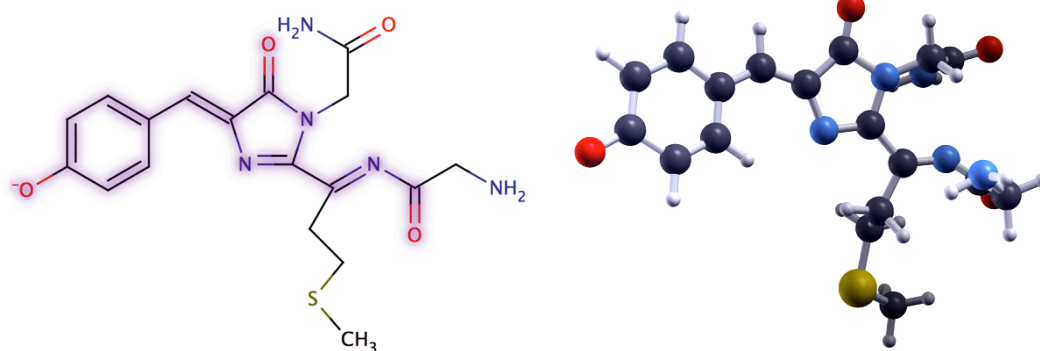
The dynamic first hyperpolarizability values, the vertical excitation energies [$\Delta E_{ge} = \hbar\omega_{ge} = \hbar(\omega_e - \omega_g)$], and the excited state properties of the compounds were calculated with the same methods as in the previous section.

All calculations were performed with the Gaussian 03 package.¹²

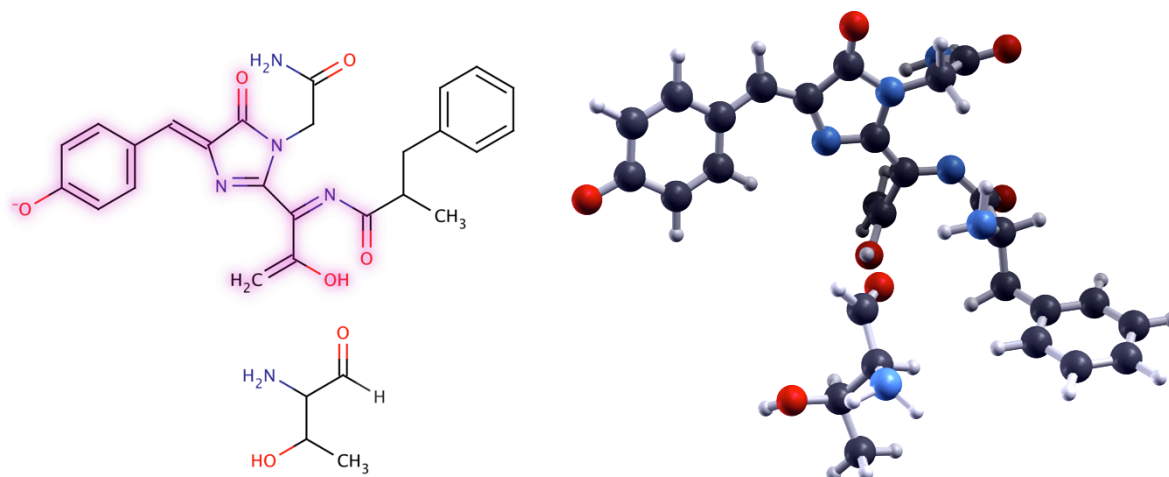
III.B.3. Results

The nonlinear optical properties determined by Hyper Rayleigh scattering are listed in Table 4. While β_{HRS} in this table is the dynamic hyperpolarizability using 800 nm laser light after correction for fluorescence contribution, β_0 , the static hyperpolarizability is β_{HRS} after an additional correction for resonance enhancement, since we are measuring at resonance conditions. When we compare the β values of the red FPs with the previous set of FPs, we notice that the trend in β_0 confirms our expectations and, moreover, that β_{HRS} of mCherry is larger than that of eGFP.

A. mCherry



B. mStrawberry enol form



C. mStrawberry alcohol form

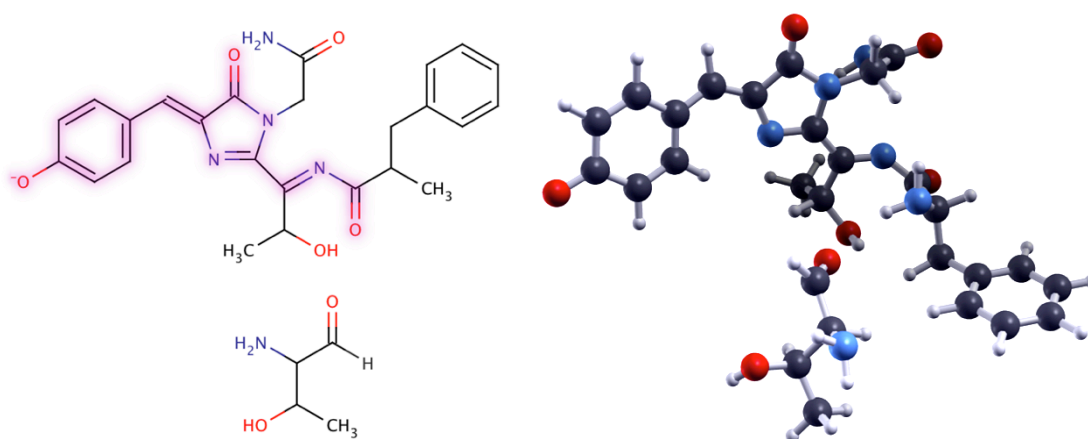


FIG. 5. Models for the mCherry and mStrawberry chromophores. The conjugated systems and hence the supposed chromophores are highlighted in the structures on the left side.

For these specific fluorophores, we did not expect any emission at 400 nm since both the single-photon (Fig. 6) and two-photon absorption spectra²² show very low absorption at 800 nm. The HRS measurements however seem slightly influenced by a small fluorescence contribution. The measurements have been therefore executed so that the demodulation of the apparent β has been performed to get rid of this influence.⁵ The demodulation curves are shown in figure 7.

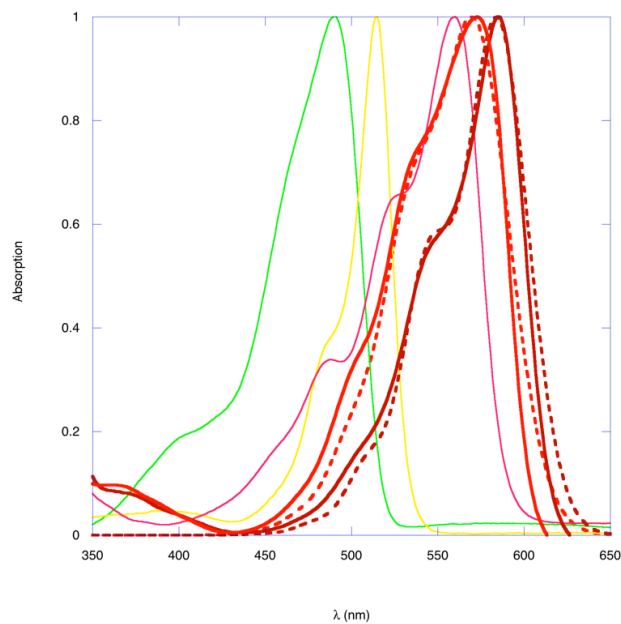


FIG. 6. Experimental (solid line) and simulated (dashed line) absorption spectra for mCherry (dark red) and mStrawberry (light red), and experimental spectra for eGFP (green line), eYFP (yellow line) and DsRed (pink line)

TABLE 4. Wavelengths of maximal absorption and fluorescence, and experimentally determined first hyperpolarizabilities of two fluorescent proteins currently studied. All β values in 10^{-30} esu.

	mStrawberry	mCherry
$\lambda_{\text{max,abs}}$	575 nm	587 nm
$\lambda_{\text{max,flu}}$	596 nm	610 nm
$\beta_{\text{HRS (800 nm)}}$	104 ± 5	134 ± 16
β_0	54 ± 4	71 ± 14

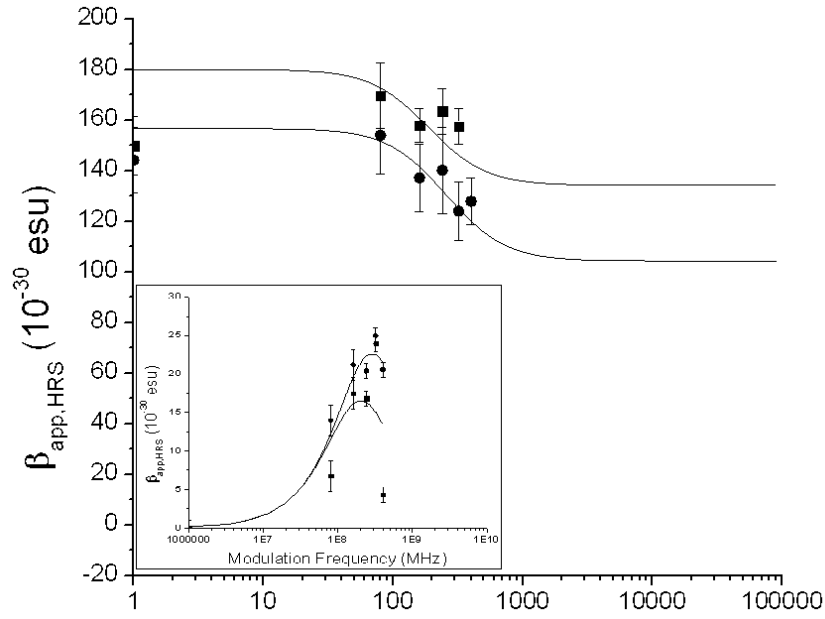


FIG. 7. The apparent, modulation frequency-dependent (or apparent), dynamic first hyperpolarizability $\beta_{\text{HRS},800}$ for mStrawberry (●) and mCherry (■). The inset shows the growing phase difference between the total HRS signal and a scattering reference.

The experimental dynamic β_{HRS} measurements for mCherry and mStrawberry were carried out at a wavelength that is quite far from the absorption band, so that they are only slightly impacted by resonance effects. Subsequently, TSA¹⁴ is a good approximation to extract the static first hyperpolarizabilities from the experimental data, provided β is dominated by a single excited state. Thus, comparison between the experimental and theoretical values does not necessarily require a high-level pre-treatment of the experimental data, contrary to our previous study of eGFP, eYFP, and DsRed. In this previous case we considered 3 levels of refinement within the TSA to rule out influences by resonance effects when evaluating the static values of β from the measured dynamic responses. Nevertheless, for the sake of completeness and consistency with our previous study, we list in Table 5, the static β values as determined with the three levels of refinement of the TSA, for mCherry and for mStrawberry in comparison with the fluorescent proteins of our previous work. The parameters for mCherry and mStrawberry are listed in table 6. As

expected, the static β values of mCherry and mStrawberry are very similar for the three levels of refinement of the TSA.

The relative position of the excitation energies was confirmed by the TDDFT/B3LYP calculations, which predict an excitation wavelength of 553 nm for mCherry and 546 nm for the mStrawberry alcohol form, whereas the mStrawberry enol form has an excitation wavelength of 579 nm.

III.B.4. Discussion

III.B.4.a. mStrawberry and mCherry

We can clearly state that the experimental results match our expectations. The FPs with a larger chromophore have a higher β_0 . The computational results on the other hand reveal an unexpected behaviour for mStrawberry. It is important to understand that mStrawberry was not an evident sample for both the experimental measurements and the computational calculations. Indeed, the poor resolution of the X-ray data is the real bottleneck for the mStrawberry theoretical study. The analysis of all the chromophore bond lengths shows a quite unexpected one, leading to the enol form.

It is known that mStrawberry is very sensitive to pH-changes compared to many other FPs.²⁰ Even the linear properties change dramatically upon changing the pH. Since we based our concentration measurements on the extinction coefficient of mStrawberry from literature, it is very well possible that our calculations of beta from experimental measurements are based on inaccurate concentration data. We do not expect this to be a large error since the differences in concentration will be less than 30%, which would only filter through to the β values as a factor $\sqrt{1 + 30\%}$. In any case the hyperpolarizability of mStrawberry will still be smaller than the result we found for mCherry.

TABLE 5. Comparison between calculated, measured, and extrapolated values of the first hyperpolarizabilities (10^{-30} esu)

	mCherry	mStrawberry	
		enol	alcohol
$\beta_{\text{HRS},800}$	134 ± 16.5	104 ± 5	
$\beta_{\text{HRS},\infty}$ TSA	71 ± 9	54 ± 3	
$\beta_{\text{HRS},\infty}$ TSA (homog. damping)	75 ± 9	56 ± 3	
$\beta_{\text{HRS},\infty}$ TSA (homog. + inhomog. damp., vibr.)	70 ± 9	51 ± 2	
$\beta_{\text{HRS},\infty}$			
HF/IEFPCM (solvent = water)	65	88	73
$\beta_{\text{HRS},1900}$			
HF/IEFPCM (solvent = water)	39	50	41
$\beta_{\text{HRS},1064}$			
HF/IEFPCM (solvent = water)	124	219	143

Comparison between the experimental and theoretical data for mCherry and mStrawberry chromophores enables to exclude the mStrawberry enol form. Indeed, going from mCherry to mStrawberry, a blue shift (+12 nm) of the wavelength of maximum absorption is expected and only the alcohol form is associated with a blue shift of the excitation energy (see table 5). And the alcohol form is the more stable from an energy point of view.

The calculated first hyperpolarizabilities for mCherry and mStrawberry do not exhibit the same order as the experimental data, which is not the case for the three other chromophores. As already mentioned, the π -conjugated system is highly identical for the mCherry and mStrawberry chromophores, but it differs from the other set of chromophores. So, the trends in the first set of 3 systems are more easily unravelled than for the two chromophores investigated in this section. In order to understand the origin of this discrepancy between the β values, several

aspects should be analyzed in more detail. The experiments were performed in aqueous solution while the calculations were carried out using XRD structures, which means on crystallized proteins. Since the π -conjugated pathway is directly influenced by the protein folding, one possible source of discrepancy can be attributed to the folding, that differs in the crystallized proteins versus aqueous solutions and is associated with different geometrical constraints. In this sense, the flawed resolution of the X-ray diffraction data represents a supplementary inconvenience when assigning the optimisation constraints for the mStrawberry structures. Yet, another point that should not be ignored is the level of the calculations, which are done at CPHF-TDHF level, therefore lacking electron correlation effects. The impact of the electron correlation is extensively discussed in the Chapter 5.

TABLE 6. Best parameters fitting the UV/Vis absorption spectra

	mCherry	mStrawberry
$\hbar\omega_{ge}$ (eV); λ_{eg} (nm)	2.12; 585	2.16; 573
$\gamma_{homogeneous}$ (eV)	0.162	0.174
$\gamma_{inhomogeneous}$ (cm ⁻¹)	100	100
S	0.50	0.70
ω_{vib} (cm ⁻¹)	1250	1230
γ_{vib} (cm ⁻¹)	700	800

III.B.4.b. mFruits

The aim of this research is to provide users with a list of fluorescent proteins with good characteristics for both SHG and (two-photon) fluorescence imaging. This will be very dependent on the preferred excitation wavelength and whether a strong SHG signal is required, or whether the separation between SHG and TPEF is more important. It is crucial to see the difference between β_0 and the actually

measured signal. While eGFP will have an effective signal that is stronger than eYFP and DsRed due to resonance enhancement and a significant contribution of fluorescence at 400 nm, the two newly measured FPs have an even stronger effective signal than eGFP without these effects. Furthermore, the TPEF signal will be able to travel further through the biological samples than light at lower wavelengths, which enables researchers to capture signal from cells deeper inside tissues. For the same reason it would be very interesting to measure first hyperpolarizabilities at wavelengths higher than 800 nm.

III.B.5. Conclusions

Even though many more experiments are needed before we can conclude on a specific relationship between the structural, the linear and the nonlinear optical properties of fluorescent proteins, our expectations are quite well met. The simplified rule to expect a higher hyperpolarizability for a larger conjugated π -system is still an acceptable rule of thumb. Furthermore, mCherry exhibits a larger dynamic β_{HRS} than eGFP for which the signal is boosted by quite some resonance enhancement. Combined with a much smaller contribution of fluorescence at 400 nm, this makes mCherry an even better suited chromophore for SHG microscopy.

For future research it is important to investigate the pH-dependency of the hyperpolarizability of the fluorescent proteins and to find methods to determine the concentrations of these proteins as accurately as possible (e.g. based on the Strickler-Berg equation relating the molecular extinction coefficient to the fluorescence radiative lifetime²³). It is also very interesting to measure hyperpolarizabilities at wavelengths higher than 800 nm.

Not only the dubious information from X-ray data, but also the observation of merely identical chromophores in mStrawberry and mCherry complicates the design of a proper model for quantum chemical calculations. Furthermore, finding the ideal balance between calculation time and inclusion of the most important surrounding

residues, molecules, or perhaps even just atoms, remains a challenge for the suitable simulation of these species. This is estimated in the following sections.

III.C Improving the second-order nonlinear optical response of fluorescent proteins: the symmetry argument

Our collaborators of the KULeuven have successfully designed and expressed a new fluorescent protein with improved second-order nonlinear optical properties. It is the first time that a fluorescent protein has been rationally altered for this particular characteristic. On the basis of the specific noncentrosymmetry requirements for second-order nonlinear optical effects, we had hypothesized that the surprisingly low first hyperpolarizability of eYFP could be explained by centrosymmetric stacking of the chromophoric Tyr66 and the neighboring Tyr203 residue. The inversion center was removed by mutating Tyr203 into Phe203, with minor changes in the linear optical properties and even an improved fluorescence quantum yield. Structure determination by X-ray crystallography as well as linear optical characterization corroborate a correct folding and maturation. Measurement of β by means of hyper-Rayleigh scattering as well as their analysis using quantum chemistry calculations validate this hypothesis. This observation can eventually lead to improved red fluorescent proteins for even better performance. On the basis of the specific function (second-harmonic generation), the color of its emission, and in analogy with the “fruit” names, we propose SHardonnay as the name for this Tyr203Phe mutant of eYFP.

III.C.1. Introduction

In the first section, we have observed that eYFP exhibits poor SHG performance compared to other related fluorescent proteins. We speculated that this could be due to the centrosymmetric arrangement of two π -stacked tyrosine residues in the chromophore. In an effort to improve the second-order nonlinear

optical properties of eYFP, we now replace the Tyr203 residue by Phe203, giving rise to a new fluorescent protein with a minor blue shift of 2 nm in the absorption and 3 nm in the emission spectrum compared to the parent eYFP.

Many variations of fluorescent proteins with different optical properties have been developed in the past for varying purposes²⁴ but not for the second-order nonlinear optical properties. Tuning of the linear spectral properties is usually based on the variation of the length of the effective conjugation path in the chromophore. This is usually achieved by extending the covalent conjugated system but also by more subtle alterations such as varying the bond length alternation of single and double bonds in the chromophore or by noncovalent interaction (e.g., π -stacking in eYFP40 and Venus41).

The extremely low β of eYFP could possibly be due to the molecular structure of its chromophore, containing an inversion center, turning half the chromophore into a centrosymmetric structure. In an effort to validate this hypothesis and find a yellow fluorescent protein with an appreciable β , we found zFP538, a natural yellow-to-orange fluorescent protein isolated by Matz and coworkers (chromophore displayed in Figure 8e and 8f).²⁵ Our hypothesis also inspired the generation of an eYFP mutant, eliminating the centrosymmetry in the eYFP chromophore by replacing Tyr203 (Figure 8a and 8b) with Phe203 (Figure 8c and 8d) through site-directed mutagenesis. The mutant was named SHardonnay for its improved second-harmonic properties and for its color appearance similar to white wine of the Chardonnay grape. SHardonnay was thoroughly characterized, both by optical measurements and by X-ray crystallography, and the linear and second-order nonlinear optical properties were corroborated by advanced theoretical methods. As for the two precedent sections, we report here the whole experimental and theoretical characterization.

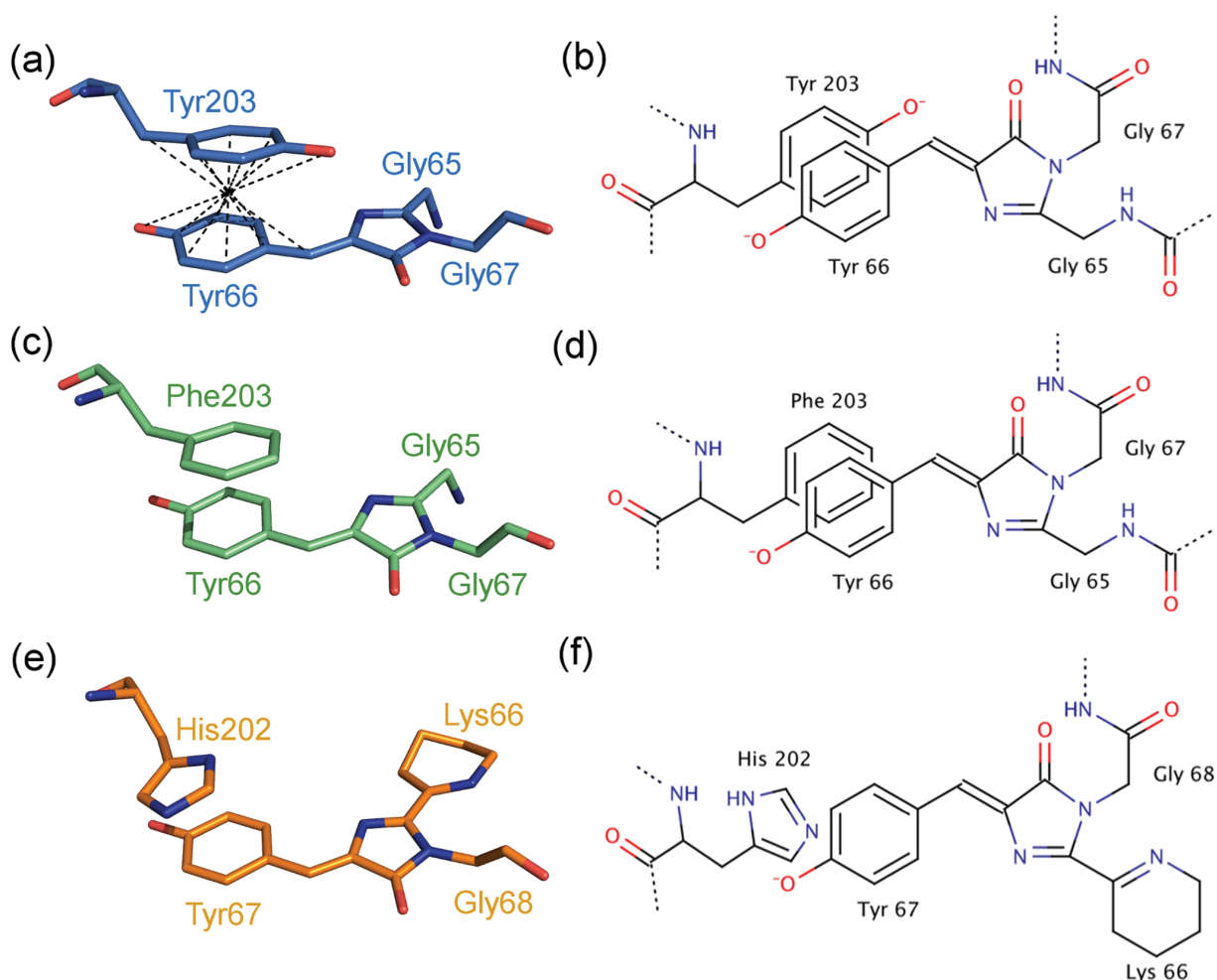


FIG.8. Representation of the different chromophores discussed in this section. (a) The chromophore of eYFP was taken from PDB ID 1YFP,^{8a} (c) the structure of SHardonnay was submitted to the protein database as 3V3D, and (e) the chromophore of zFP538 was taken from PDB ID 2OGR.²⁶ Parts b, d, and f show the respective chemical structures in 2D. The presence of an inversion center in the eYFP chromophore is marked by dashed lines in image a. Images a, c, and e were generated with PyMol.

III.C.2. Methods

III.C.2.a. Optical characterization

The fluorescence spectra and all spectra necessary for the determination of the fluorescence quantum yield were measured on a Synergy MX monochromator-based multimode microplate reader (Biotek). The achieved fluorescence quantum yield of SHardonnay was successfully cross-checked with that of purified eGFP and eYFP. For the determination of the protein concentrations necessary to derive an

extinction coefficient, the Bradford method²⁷ was used. All required absorption values (A_{595} of Bradford solutions, A_{280} of Bovine Serum Albumin, A_{280} and A_{511} of SHardonnay as well as the linear absorption spectra) were measured in 1 cm quartz cuvettes in a Genesys 6 spectrophotometer (ThermoScientific).

The first hyperpolarizability, β , was experimentally determined by frequency-resolved femtosecond hyper-Rayleigh scattering, as described previously. To estimate the intrinsic β response, the experimental data were pretreated to eliminate resonance effects. To apply these extrapolation schemes, the successive $F(\omega, \omega_{ge}, \gamma)$ functions had to be fitted to the UV/vis absorption spectra (fitting parameters are given in table 7). The corresponding simulated absorption spectra are compared with the experimental spectra (Figure 9), demonstrating a very close relationship between the two.

TABLE 7. Best parameters fitting the UV/Vis absorption spectra

	eYFP	SHardonnay	zFP538
$\hbar\omega_{ge}$ (eV); λ_{eg} (nm)	2.42; 512	2.42; 512	2.35; 528
$\gamma_{homogeneous}$ (eV)	0.078	0.083	0.25
$\gamma_{inhomogeneous}$ (cm ⁻¹)	100	100	100
S	0.28	0.35	0.50
ω_{vib} (cm ⁻¹)	1050	1200	1222
γ_{vib} (cm ⁻¹)	500	600	600

III.C.2.b. Computational methods and geometrical structures

The geometries of all chromophores and surrounding residues were extracted from available X-ray diffraction data.^{8a, 19} The position of the H atoms was optimized at the density functional theory level using the B3LYP exchange-correlation functional and the 6-31+G(d) basis set for the chromophores but with the PM6 method for the chromophores surrounded by the first shell of residues. The tautomer form of the histidine of zFP538 used in the simulation and represented in figure 8 is the lower in energy.

The first hyperpolarizabilities determined by the MP2 method recover the largest part of the electron correlation effects. First, the static β values were calculated using the finite field procedure. Then the multiplicative scheme⁹ was used to account for frequency dispersion ($\lambda = 1900$ nm). The static values extrapolated from experiment are best compared with the calculated low-frequency dynamic values to account for the optical dielectric constant of the environment.

The chromophore environment was modeled by employing (i) IEFPCM^{10b} in CPHF/TDHF calculations and (ii) the ONIOM approach,²⁸ which explicitly takes the first shell of residues into account. The outer layer is described at the HF level whereas the MP2 scheme is adopted for the inner region consisting of the chromophore. The 6-31+G(d) basis set was adopted. Taking advantage of using the same basis set for the two layers, there is a nice cancelation of terms in the ONIOM energy expression and similarly to β . The only remaining approximation comes from the use of the MP2 scheme to describe the correlation energy, which is evaluated on the model system.

All calculations were performed with mopac2009²⁹, Gaussian09³⁰ as well as homemade codes.

III.C.3. Results

III.C.3.a. Experimental linear optical characterization

With linear spectroscopy, we determined that the absorption maximum of SHardonnay is at 511 nm, and the emission maximum at 524 nm (Figure 2 and Table 1), a minor hypsochromic shift with respect to eYFP, which is even visible to the eye. This observation already implies that the folding of the protein is not compromised and the chromophore formation is supported by the new microenvironment of the chromophore. The spectral difference between the two yellow FPs and zFP538, respectively, is markedly larger, giving zFP538 a clear orange color. We assessed and

approved the purity of the samples based on SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) and mass spectrometry results. With a more extensive characterization of the fluorescent protein in mind, we also determined a number of other important characteristics of SHardonnay (Table 8). We measured a fluorescence lifetime of 3.4 ns, which is typical for GFP-like proteins, an extinction coefficient at the absorption peak of the chromophore (511 nm) of around 89000 $\text{M}^{-1}\text{cm}^{-1}$, compared to 83400 $\text{M}^{-1}\text{cm}^{-1}$ for eYFP, and a quantum yield of 0.75, remarkably higher than that of eYFP. The small increase in the extinction coefficient could be explained by the absolute absence of a peak around 400 nm (see Figure 9).

III.C.3.b. X-ray diffraction

The structure of SHardonnay has been solved to a resolution of 1.95 Å. The overall structure of the mutant shows the 11-stranded β -barrel fold typical for GFP-like proteins, with the chromophore suspended in the center by a coaxial helix. The protein consists of 238 amino acid residues, of which residues 2–231 could be clearly defined in the electron density map, including four additional residues of the C-terminus, in comparison with its parent eYFP (PDB ID 1YFP).^{8a} These residues are positioned in a hydrophobic cleft of one of the adjacent protomers in the lattice, rendering them more rigid and more easily and unambiguously modeled in the electron density map. The root-mean-square (rms) deviation between the eYFP and the SHardonnay α carbons is 0.28 Å, and no considerable movement of either the backbone or the side chains is detected inside the β -barrel. Superposition of the α carbons of SHardonnay and zFP538 (PDB ID 2OGR),²⁶ results in an rms deviation of 1.51 Å for 210 equivalent positions, with large differences in solvent exposing loop configurations, which is consistent with the low sequence identity of 22% (PDBeFold)³¹ compared to zFP538.^{19, 25}

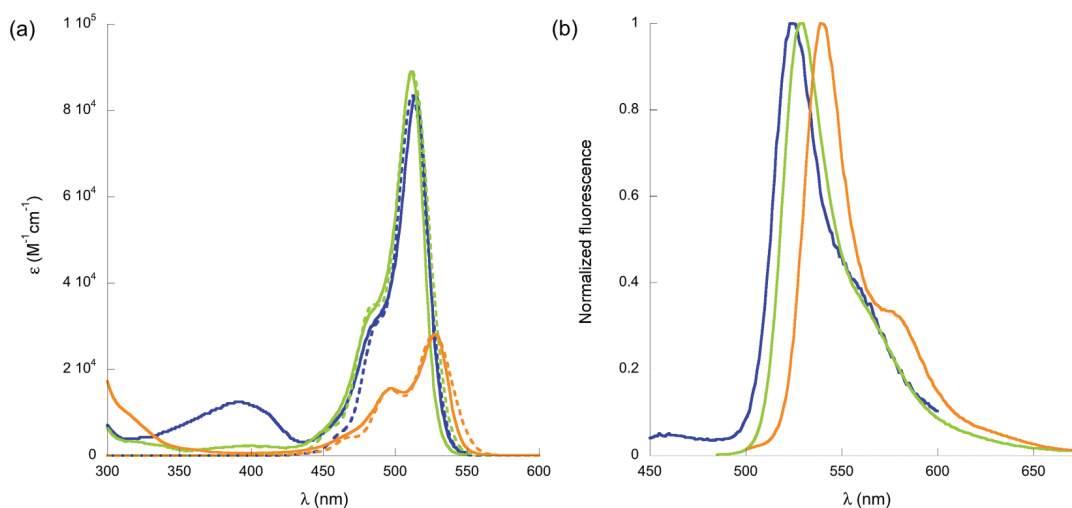


FIG.9. Linear spectral properties of the fluorescent proteins: (a) Extinction coefficients and (b) normalized fluorescence spectra of the fluorescent proteins eYFP (blue), SHardonnay (green), and zFP538 (orange), all measured under the same conditions. Full lines show experimental data, while the dashed lines represent simulated spectra, scaled for the same maximum value.

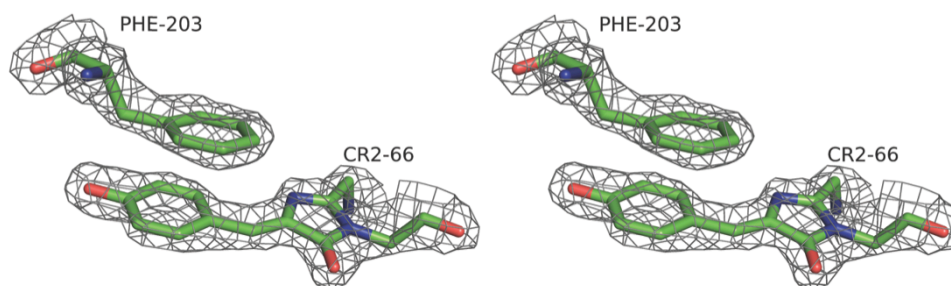


FIG. 10. Stereo view of the SHardonnay chromophore, showing the final $2mF_o-DF_c$ map at 2.0σ level (grey mesh). The image was generated with PyMol.

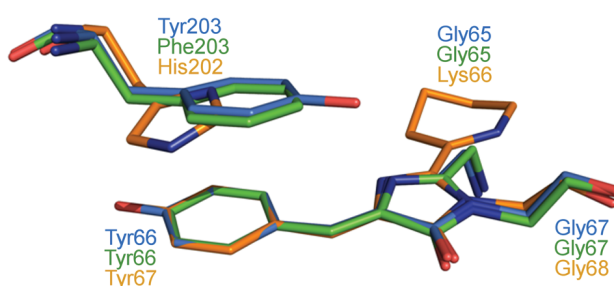


FIG. 11. Superposition of the chromophores of SHardonnay (green) and zFP538 (orange) on eYFP (blue) after pair fitting of the atoms of the corresponding hydroxyphenyl moieties. The chromophores are displayed isolated from the protein barrel but including the (respectively) Tyr203, Phe203, and His202 residue close to the chromophore. The image was generated with PyMol.

TABLE 8. Experimentally determined and calculated linear and nonlinear spectral characteristics.

	eYFP	SHardonnay	zFP538
$\lambda_{\text{abs,max}}$ (nm)	514	511	528
$\lambda_{\text{em,max}}$ (nm)	527	524	538
ϵ at $\lambda_{\text{abs,max}}$ ($\text{M}^{-1}\text{cm}^{-1}$)	83,400	89,000	20,200
τ (ns)	2.9-3.4	3.4	<i>ND</i>
Φ	0.60	0.75	0.42
PDB ID	1YFP	3V3D	2OGR
θ ($^{\circ}$)	14.5	18.1	19.4
γ ($^{\circ}$)	12.1	11.1	15.4
R_{cen} ($\text{\AA}$)	3.69	3.77	3.65
$\beta_{\text{HRS}, 800}$ (10^{-30} esu)	37 ± 4	74 ± 5	90 ± 5
$\beta_{\text{HRS}, \infty}$ (10^{-30} esu) TSA	14 ± 2	28 ± 2	38 ± 2
$\beta_{\text{HRS}, \infty}$ (10^{-30} esu) TSA (homog. damping)	14 ± 2	29 ± 2	41 ± 2
$\beta_{\text{HRS}, \infty}$ (10^{-30} esu) TSA (homog. + inhomog. damping, vibr.)	13 ± 2	25 ± 2	33 ± 2
$\beta_{\text{HRS}, \infty}$ (10^{-30} esu) CPHF, IEFPCM(water)	25	29	35
DR ($\lambda = \infty$) HF/IEFPCM (solvent = water)	4.20	4.42	5.92
$\beta_{\text{HRS}, 1900}$ (10^{-30} esu) TDHF, IEFPCM(water)	12	16	19
$\beta_{\text{HRS}, 1900}$ (10^{-30} esu) ONIOM(MP2:HF)	14	24	40

Unlike most other fluorescent proteins, whose chromophore consists of a single moiety, arising from a post-translational cyclization of three amino acid residues, the chromophore of eYFP and SHardonnay contains a second moiety consisting of the aromatic side chains of Tyr203 and Phe203, respectively. The equivalent residue in zFP538 is His202, although the contribution of His202 to the chromophore and the optical properties was never fully investigated. The red shift of zFP538 can be accounted for by Lys66, creating a third ring with a CN double bond leading to an extension of the conjugated system, because the Lys66Met mutant of zFP538 restores green fluorescence.³² His202 was therefore not

suspected to play a vital role in the optical properties of the chromophore. Both conjugated moieties of the SHardonnay chromophore (the conjugated system originating from the Gly65-Tyr66-Gly67 triad, and Phe203, respectively) were modeled unambiguously in the electron density map (Figure 10). The two coplanar rings of the CR2 moiety adopt the cis conformation, and the two aromatic side chains align in an off-centered parallel displaced orientation.³³ Except for the minor modification, the deletion of the hydroxyl group in the hydroxyphenyl ring of Tyr203, the spatial arrangement of the chromophore atoms is conserved in SHardonnay (Figure 11 and Table 8).

III.C.3.c. Experimental first hyperpolarizability

We obtained the dynamic $\beta_{\text{HRS},800}$ value from measurement with 800 nm femtosecond pulsed laser light and after correction for fluorescence contribution (Table 8).^{5a} The static hyper-polarizabilities, $\beta_{\text{HRS},\infty}$, were then extrapolated from the measured $\beta_{\text{HRS},800}$ by adopting different refinements of the two-state approximation to evaluate the frequency dispersion factor (Table 8).¹⁴

III.C.3.d. Calculated first hyperpolarizability

We obtained the calculated static and dynamic first hyperpolarizabilities by means of coupled-perturbed Hartree–Fock and time-dependent Hartree–Fock calculations,³⁴ Møller–Plesset second-order perturbation theory for deprotonated chromophores, and ONIOM calculations for the chromophores surrounded by the first shell of residues (Table 8). All levels of computation give the same amplitude ordering for β_{HRS} , i.e., $\beta_{\text{HRS}}(\text{eYFP}) < \beta_{\text{HRS}}(\text{SHardonnay}) < \beta_{\text{HRS}}(\text{zFP538})$, which also matches the experimental values, allowing therefore the use of these calculated data for the interpretation. An appropriate comparison in Table 8 should be based on the intrinsic or static β responses. Theoretically, the static responses were directly calculated at the MP2 level ($\beta_{\text{HRS},\infty}$), while frequency-dependent $\beta_{\text{HRS},1900}$

values, far from resonance, were also evaluated to account for a proper treatment of the solvent effects (with dynamic dielectric constant).

III.C.4. Discussion

This report focuses on the effects of a minor modification in eYFP on its linear and nonlinear optical properties. On the basis of previous research, we proposed to mutate the π -stacked Tyr203 into Phe203 to generate a yellow fluorescent protein with an improved β . This mutant, SHardonnay, shows linear spectral properties highly similar to those of eYFP, implying that stacking with Phe203 takes place and causes a comparable spectral shift. The blue shift of 2 nm points to a minimal reduction in the effective conjugation path length of the chromophore, due to elimination of the hydroxyl group of Tyr203. Even though the structural difference is almost negligible in terms of crystal structure and linear optical properties, it brings about a considerable increase in β for SHardonnay with respect to eYFP and restores the originally observed trend. This confirms our earlier conjecture that the local centrosymmetry in the chromophore causes a substantial drop in the second-order optical performance of the fluorescent protein.

The molecular structure of the mutant was determined by X-ray diffraction, providing accurate coordinates of the atoms. Comparison of the chromophores at the molecular level confirmed the conservation of the aromatic ring in residue 203, except for a minimal translation of 0.08 Å. The knowledge of the chromophore structure and its nearby surroundings is also essential for accurate calculation of β by quantum chemical methods such as the ONIOM MP2:HF method applied to large systems (≈ 500 atoms) that was successfully used in this work.

Besides the Tyr203Phe mutant of eYFP, another, natural, fluorescent protein was investigated as an alternative approach for a yellow variant with an appreciable β value. The conjugated system in zFP538 is extended through the covalent bonds of

the Lys66 residue^{26, 32} (Tyr68 being the chromophoric tyrosine equivalent) instead of the noncovalent π -stacking in most other known yellow fluorescent proteins. The conjugated path is slightly longer in zFP538 than in the yellow variants and coincides with a red shift in the linear spectrum and a larger β . This is in line with the general bandgap considerations of stronger nonlinearity (larger β) for a smaller bandgap (red- shifted absorption). The ONIOM MP2:HF method confirms the enhanced second-order nonlinear optical response in our mutant, though slightly overestimates that of zFP538. This may point to the very subtle interplay between conjugation extension and stacking to influence the bandgap for inducing a red shift, but this could also be due to imperfections in the crystal structure.

While SHardonnay was mainly designed to investigate our hypothesis and can be interpreted as a proof of principle, our ultimate goal is to engineer fluorescent proteins that can be used in SHG imaging. SHIM with eGFP as a voltage-sensitive contrast marker was demonstrated in *Caenorhabditis elegans*,^{17b} but no follow-up papers have been published so far. This suggests that fluorescent proteins as we know them may not be ideal markers for SHIM. We estimate that future work to optimize a number of parameters can result in successful SHG imaging with fluorescent proteins. These parameters include the size of the conjugated system of the chromophore, the choice of the fundamental wavelength, and the concentration, or density, of the chromophores as well as their relative orientation.

First of all, the strength of the second-harmonic response is related to the length of the effective conjugation path in the chromophore. The size of the planar chromophores of fluorescent proteins is limited by the eventual folding of the protein. With this report, we show that the conjugated system can also be extended noncovalently, without losing nonlinear optical performance.

Secondly, the use of longer wavelengths should be promoted for cellular SHIM. The contribution of resonance enhancement will then increase significantly,

resulting in a realistic β_λ of 1 order of magnitude higher.³⁵ Good performance is expected for dyes with a β_λ around 300×10^{-30} esu,³⁶ but this could be an overestimation. Di-8-ANEPPS, together with di-4-ANEPPS a popular voltage sensitive marker in TPEF and SHG microscopy near 800 nm,³⁷ performs well with a β_{800} of about $200\text{--}350 \times 10^{-30}$ esu, depending on the medium (unpublished HRS results). While a β_{800} of 134×10^{-30} esu was measured for mCherry, the estimated β_{1000} for mCherry, according to the TSA approximation, would be $290 \pm 60 \times 10^{-30}$ esu, which would make mCherry a good candidate for SHIM. Ideally, the design of SHG probes should aim for chromophores that allow to have all involved wavelengths within the optical window for biological tissue. The second-harmonic effect limits the options, but does not exclude optimization. Regardless of the choice of the wavelengths, the background signal from collagen, micro-tubules, or myosin should not be ignored, even in cultured live cells.

A third point of concern in promoting fluorescent proteins versus small chemical dyes for SHG imaging is the maximal number of chromophores in the focus of the laser beam. A good working concentration for SHIM from membrane-associated dyes is estimated at 1 chromophore per 100 lipids.^{37a} This is harder to achieve with fluorescent proteins than with small chemical dyes because of their size, and for toxicity reasons. However, fluorescent proteins will reduce flip-flopping of the dye, reducing centrosymmetry, and internalization of the chromophore.

Last but not least, a very important issue in SHIM is the control of the relative orientation of the chromophores in the sample. There is a critical need to optimize this because the bulk second-order susceptibility, determining the final signal strength, depends on it. Cellular membranes are obviously convenient scaffolds to force chromophores in a parallel orientation, and as long as the molecules are not subject to flip-flopping or internalization, they could be arranged in a parallel and noncentrosymmetrical way. However, the use of membrane-targeted fluorescent

proteins often involves a flexible linker of about 5 to 20 amino acids to ensure correct folding. This flexibility can be detrimental for the SHG signal, but possible solutions are available.³⁸

III.C.5. Conclusions

On the basis of the findings of our previous research, we have altered eYFP by a single mutation to increase its surprisingly low first hyperpolarizability, β . Analysis by means of X-ray crystallography confirmed a correct folding with conservation of the relative position of the mutated Phe203 residue to the chromophore. The Tyr203Phe mutant of eYFP, dubbed SHardonnay, exhibits a minimal spectral shift, a similar fluorescence lifetime, and even an improved fluorescence quantum efficiency compared to the mother eYFP. Hyper-Rayleigh scattering measurements revealed a substantial increase in its β , hereby confirming our hypothesis as well as the disruption of the centrosymmetry in the chromophore.

We hereby confirm the importance of the noncentrosymmetry requirement for the second-order nonlinear optical properties and show the sensitivity of this principle. The removal of a hydroxyl group resulted in minor or no changes in the linear optical properties but a remarkable increase for the second-order properties. We therefore propose that the noncentrosymmetric stacking strategy could be explored in further research to extend the conjugated system in red fluorescent proteins, because a chromophoric tyrosine residue is present in the majority of fluorescent proteins. For substantial enhancement of the second-order optical properties, however, extension of the conjugated system through covalent interaction seems to be a stronger argument, but the size of the chromophore in fluorescent proteins is restricted by the microenvironment.

The importance of engineering fluorescent proteins with a high first hyperpolarizability becomes clear when considering the other difficulties in

designing the experimental conditions that could result in images with SHG originating from the fluorescent proteins. Not only the strength of the individual molecular response, expressed by β , has to be maximized but also the molecular arrangement of the chromophores, their concentration, or density, will be important for the experiment. Also the choice of the fundamental wavelength is of vital importance, as it should be close enough to resonance conditions yet keep its distance from the fluorescence spectrum for successful SHIM based on fluorescent proteins. We speculate that our results offer a lot of potential to engineer new fluorescent proteins to explore new approaches for SHIM.

III.D. *Predicting and interpreting the second-order nonlinear optical responses of fluorescent proteins*

III.D.1. Introduction

In the previous section, several fluorescent proteins have been considered. Frequency-resolved femtosecond hyper-Rayleigh scattering experiments have been conducted to determine their first hyperpolarizabilities. These experimental characterizations were accompanied by theoretical modeling of their NLO responses, though the calculations were mostly limited to the evaluation of the static and dynamic first hyperpolarizabilities of the chromophores with or without a few surrounding amino acid residues, except for the last section on the new fluorescent protein, SHardonay where the effect of the surroundings and of electron correlation has been accounted for. Thus, in these calculations, the largest part of the surrounding β -barrel is missing while the surrounding effects are mostly described by the polarizable continuum model with water as solvent. In addition, the calculations have been carried out at the Hartree-Fock level, which brings the question about the impact of electron correlation. Nevertheless, most of the preceding theoretical analysis employed the time-

dependent Hartree-Fock scheme and could reproduce and explain the experimental trends.

In order to provide a quantitative description and showing the motivation that led us to use the ONIOM MP2 : HF/6-31+G* method, the previous studies need to be extended. From the chromophore only to the structure where the first shell of surroundings is accounted for, it is quite interesting to assess different levels of treatments and determine the specific requirements for an accurate description.

Nowaday, proteins are still too large to allow the evaluation of their NLO responses using an *ab initio* approach, so approximate schemes have been elaborated. Here the ONIOM method²⁸ is employed. This method appears well suited to characterize a system made of a central chromophore surrounded by a β -barrel.

As a matter of fact, there is an interest to assess the performance of successive versions of the ONIOM scheme to calculate the first hyperpolarizabilities with respect to calculations performed on entire systems. In the case of ONIOM β calculations, all layers must preferentially be described at a quantum mechanics level. This leaves however several combinations of methods or levels of methods to describe the different layers of the system. On the one hand, the low level of treatment used to describe the real system is either the semi-empirical PM6 method³⁹ or the Hartree-Fock method with a relatively small basis set. On the other hand, for the high level of treatment on the model system, the Hartree-Fock method or correlated *ab initio* methods are well suited. The latter calculations should employ more extended basis sets, appropriate to describe the nonlinear response of the molecule to external electric fields, such as the 6-31+G* basis set and a precise description of the electron correlation effects (see Chapter 5). The composition of the two layers are also of most

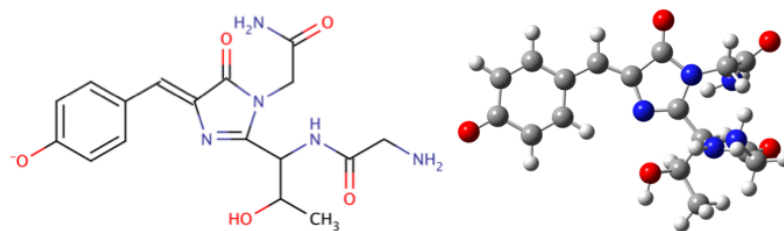
interest and dependent of the level of treatments used.

This section is organized as follows. In Section III.D.2., we present briefly the theoretical and computational approaches that have been employed. In Section III.D.3., the results obtained using successive approximate schemes are presented and discussed. Then, before concluding in Section III.D.5., Section III.D.4. presents comparisons with the available experimental data, highlighting the improvements that could be made by going to higher and higher levels of approximation.

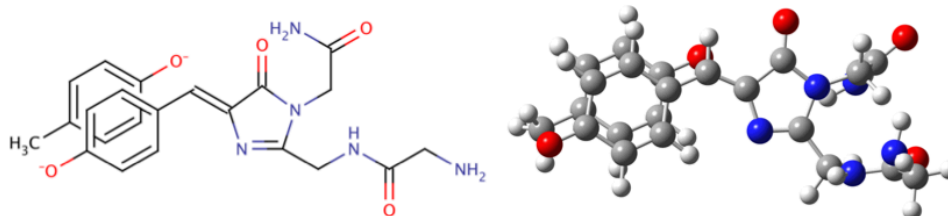
III.D.2. Theoretical and computational aspects

Six fluorescent proteins were studied, namely eGFP, eYFP, SHardonnay, DsRed, zFP538, and mCherry. Initially, the geometries of the chromophores surrounded by the first shell of residues (water molecules and amino acids, including those linking the chromophore to the protein β barrel) were taken from X-ray diffraction^{8, 19-20, 40}. Since the H atoms are missing, they were added by taking into account the structural aspects and the charges of the chromophores surrounded by their residues. The positions of the H atoms were then optimized at the PM6 level of calculation with the MOPAC package²⁹. The rest of the geometry was frozen during the optimization in order to keep the chromophore in its natural conformation. The systems made of the chromophore and the surrounding residues contain 405, 476, 483, 465, 549 and 561 atoms for eGFP, eYFP, SHardonnay, DsRed, zFP538 and mCherry respectively. In a second time, chromophores only were accounted, where the positions of the H atoms were optimized at the density functional theory level using the B3LYP exchange-correlation functional and the 6-31+G(d) basis set. The structures are represented in figure 12.

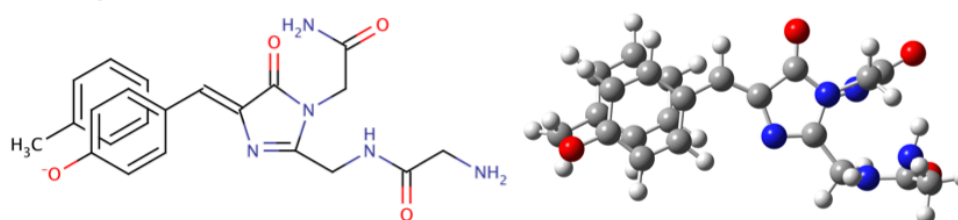
eGFP



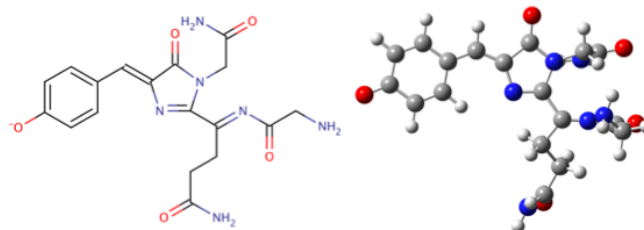
eYFP



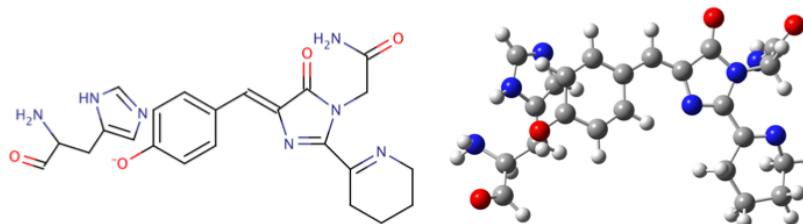
SHardonnay



DsRed



zFP538



mCherry

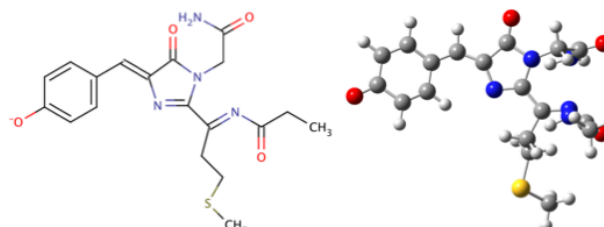


FIG. 12. Models of the isolated eGFP, eYFP, DsRed, mCherry, and zFP538 deprotonated chromophores using in the MP2 calculations.

In order to characterize the first hyperpolarizabilities of the first type of systems, the ONIOM scheme has been used. Several calculated quantities have focused our attention, i) β of the chromophore and of its nearby surrounding (high model of ONIOM), ii) the full ONIOM β , and iii) the environment contribution to β , which is defined as:

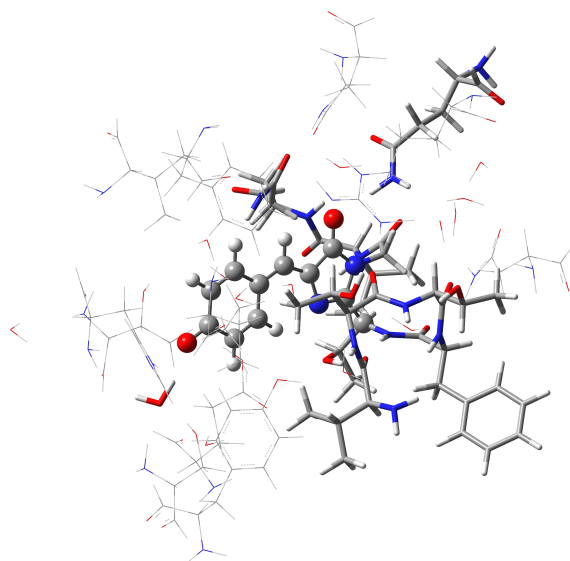
$$\vec{\beta}_{environment} = \vec{\beta}_{real, low} - \vec{\beta}_{model, low} \quad (i.e. \text{ low level layer}).$$

The smallest high ONIOM layer was defined by the chromophore and segments of the amino acids chains directly linked to it. As a consequence, the low ONIOM layer is the rest of the structure. Several levels of high ONIOM layer were however employed, in order to investigate how large it should be. For eGFP, three levels of high layer were considered: i) the high layer consists only of the eGFP chromophore, ii) the GLN 94 amino acid that forms a H-bond with the chromophore is added, and finally iii) a water molecule, which also makes a H-bond with the chromophore, is included (Figure 13). Similarly for eYFP, the high layer i) corresponds to the chromophore, ii) it includes in addition the TYR 203 amino acid in π -stacking interaction with the chromophore as well as the chain attached to TYR 203 as well as iii) the amino acid ARG 96, which forms a H-bond with the chromophore together with water molecule, also forming a H-bond (Figure 13). For SHardonnay, the PHE 203, which is in π -stacking interaction with the chromophore, is already included in the smallest high ONIOM layer. To form the second high layer model, ARG 96 is added. Adding a H-bonded water molecule and part of the HIS 148 formed the largest SHardonnay high layer (Figure 13). ARG 95 is added to the DsRed chromophore to form the second high layer model considered here. Two water molecules are then added as well as a part of the H-bonded SER 146 to obtain the largest DsRed high layer (Figure 13). To form the smallest zFP538 ONIOM high layer, the chromophore is surrounded by HIS 202 together with the chain attached to it. To get the second level of description, ARG 95 is added. By adding a water molecule

and SER 148, the largest zFP538 ONIOM high layer is then formed (Figure 13). Finally, for mCherry, in a second level of treatment, ARG 95 is included in the ONIOM high layer. With two water molecules and SER 146, the mCherry largest ONIOM high layer is formed (Figure 13).

The CPHF/TDHF methods were used at wavelengths of 1064/1900 nm. In order to investigate electron correlation effects on the first hyperpolarizabilities, wavefunction and DFT approaches were also employed. On the one hand, MP2 calculations were performed within the finite field procedure (see Chapter 2). To account for frequency dispersion at the MP2 level, the multiplicative approximation has been considered. These MP2 calculations were carried out by using the 6-31G* and 6-31+G* basis sets. On the other hand, electron correlation effects were incorporated by resorting to TDDFT using a few XC functionals, i) the LC-BLYP XC functional (with a range-separating parameter μ of 0.47), and ii) the widely-used B3LYP XC functional. The semi-empirical PM6 method was also employed. In order to control the efficiency of the PM6 method, the static longitudinal β for push-pull π -conjugated systems from Chapter 5 has been calculated and compared to the reference values. It shows that for doubly bonded systems, PM6 systematically underestimates the reference value from 32 to 38% and in comparison with the Hartree-Fock method, the PM6 response is higher showing that some electron correlation effects are implicitly accounted for. But for triply bonded systems, the PM6 method is unable to reproduce the reference values and so cannot be reliable. Thus, the PM6 method cannot be used without caution especially for structures taken from X-ray diffraction data where there can be some unexpected bond lengths.

eGFP



eYFP



mCherry

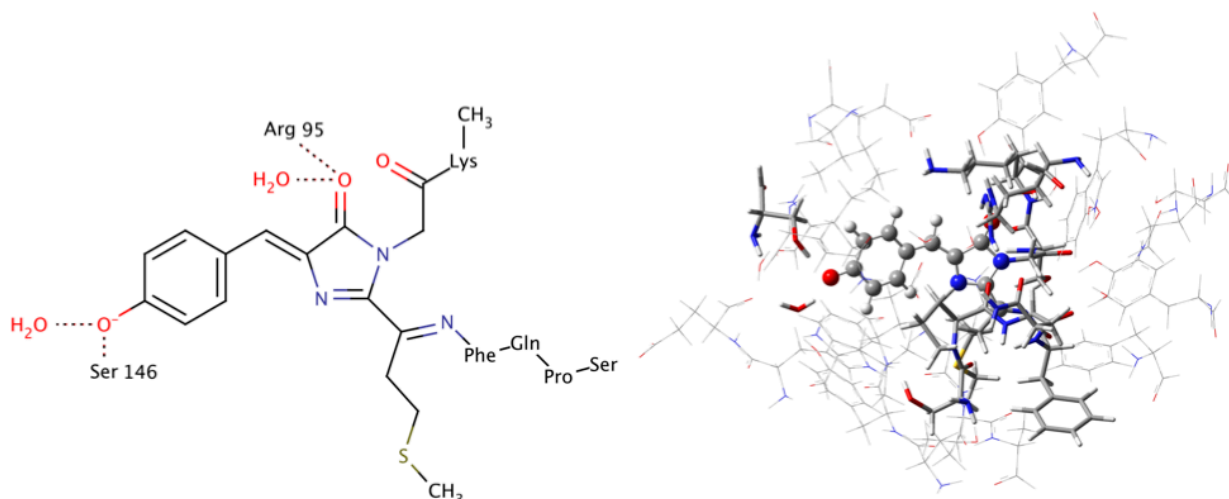


FIG. 13. Model sketches of the maximal high ONIOM layer of eGFP, eYFP, SHardonnay, DsRed, zFP538 and mCherry; the high and low layers are also represented for eGFP and mCherry (chromophores in ball and stick, the rest of the high layers in stick and the low layers in wire).

In the ONIOM calculations, the HF/6-31+G* method was chosen as the high level of theory, to start with. The LC-BLYP/6-31+G* was also employed to provide a first estimate of the impact of electron correlation subsequently. The PM6 method was generally chosen as the low level of theory but comparisons were also performed using HF/3-21G and HF/6-31G* as low-level method. Since β values have been rarely calculated within the ONIOM procedure, TDHF and CPHF calculations were carried out for the entire system and compare to the ONIOM results. This was done with the 6-31G*, and 6-31+G* basis set together with IEFPCM for the solvent description if needed. Furthermore, the MP2 method has also been used as the high level of theory. In that case, if the HF method is used for the low level of theory, taking advantage of using the same basis set for the two layers, there is a nice cancelation of terms in the ONIOM MP2:HF energy expression:

$$E_{\text{ONIOM}} = E_{\text{real, HF}} + E_{\text{model, E2}}$$

$$\vec{\vec{\beta}}_{\text{ONIOM}} = \vec{\vec{\beta}}_{\text{real, HF}} + \vec{\vec{\beta}}_{\text{model, E2}}$$

The only remaining approximation comes from the use of the MP2 scheme to describe the correlation energy, which is evaluated on the model system.

The Polarizable Continuum Model within the integral equation formalism (IEFPCM) was employed to take into account the solvent effects.¹⁰ In the ONIOM method, for each sub-calculation, the cavity of the real system was used⁴¹, this is the so-called ONIOM/PCM-X scheme. This approach was selected, as it is the easier way to include solvent effects for our ONIOM implementation when finite field calculations are needed. Solvent effects were investigated by considering solvents of decreasing static dielectric constant namely, water ($\epsilon = 78.355$), ethanol ($\epsilon = 24.852$), 1-butanol ($\epsilon = 17.332$), 1-heptanol ($\epsilon = 11.321$), 1-decanol ($\epsilon = 7.530$) and without solvent ($\epsilon = 1$).

All reported β values are given in a.u. within the T convention. All calculations were done with the Gaussian09 and Gaussian03 packages^{12, 30}.

III.D.3. Results and discussion

III.D.3.a. Comparison between full CPHF/TDHF results and the ONIOM values

The comparison between HF/6-31+G* : PM6, HF/6-31+G* : HF/3-21G and HF/6-31+G* : HF/6-31G* ONIOM static first hyperpolarizabilities (Table 9 and 10) with respect to full HF/6-31+G* results (Table 11) is presented. In both sets of data, solvent effects were accounted for by using the IEFPCM scheme (solvent = water) and a minimal high layer has been considered. For eGFP, the static β_{ONIOM} overestimates the full result by 21%, 13%, and 9% when the low layer is described with the PM6, HF/3-21G, and HF/6-31G* method, respectively. Thus, the overestimation decreases with increasing the level of treatment. For eYFP where the high layer contains only the chromophore, TYR 203 and its directly linked amino acids chain, the HF/6-31+G* : PM6 value is in good agreement (1%) with the full HF result whereas the HF description of the low layer increases the difference up to 5%.

In the case of SHardonnay, the PM6 method fails to describe well the environment with 36% lower than the reference value. By using HF/6-31+G* : HF/3-21G and HF/6-31+G* : HF/6-31G* approaches, the treatment has been strongly improved, an overestimation of 1% and 4% is obtained, respectively. For DsRed, the HF/6-31+G* : PM6 scheme underestimates the static β by less than 1%, whereas there are 8% and 4% of overestimation when using, respectively the HF/6-31+G* : HF/3-21G and HF/6-31+G* : HF/6-31G* approaches. In the case of zFP538, the HF/6-31+G* : PM6 method overestimates the reference value of 10%. Improvement is obtained by using the HF/6-31+G* : HF/3-21G and HF/6-31+G* : HF/6-31G* methods with respectively 1% and less than 1% of overestimation. For mCherry, less than 2% of overestimation when the HF method is used for the low layer.

β_{HRS} IEFPCM (water)	∞	1900nm	1064nm	∞	1900nm	1064nm	∞	1900nm	1064nm
eGFP				eYFP			SHardonnay		
High model:	<i>chromophore + chain</i>			<i>chromophore + chain + TYR 203</i>			<i>chromophore + chain + PHE 203</i>		
HF/6-31+G*: PM6	6108	3492	5731	4953	3009	5068	2897	3371	5997
HF/6-31+G*: HF/3-21G	5709	3187	5613	5261	3035	5334	4565	3001	4880
HF/6-31+G*: HF/6-31G*	5536	3137	5614	5240	3075	5444	4692	3042	5022
LC-BLYP/6-31+G*: PM6	8012	4357	9526	6701	3870	8788	6498	5266	13027
LC-BLYP/6-31+G*: HF/3-21G	7596	4051	9397	7037	3914	9056	8195	4907	11909
High model:	<i>chromophore + chain + GLN 94</i>			<i>chromophore + chain + TYR 203 + ARG 96</i>			<i>chromophore + chain + PHE 203 + ARG 96</i>		
HF/6-31+G*: PM6	6515	3727	6054	4806	2965	5016	2787	3147	5537
High model:	<i>chromophore + chain + GLN 94 + H₂O</i>			<i>chromophore + chain + TYR 203 + ARG 96 + H₂O</i>			<i>chromophore + chain + PHE 203 + ARG 96 + H₂O + HIS 148</i>		
HF/6-31+G*: PM6	6182	3607	5938	4863	2903	4843	4274	3072	4950
High model:				<i>chromophore + chain</i>					
HF/6-31+G*: PM6				4496	2694	4744			
HF/6-31+G*: HF/3-21G				5361	3085	5360			
HF/6-31+G*: HF/6-31G*				5359	3146	5520			
DsRed				zFP538			mCherry		
High model:	<i>chromophore + chain</i>			<i>chromophore + chain + HIS 202</i>			<i>chromophore + chain</i>		
HF/6-31+G*: PM6	8650	5567	15084	6709	4092	10342	9836	6917	28566
HF/6-31+G*: HF/3-21G	9330	5554	13495	6171	3639	9006	12584	6929	23695
HF/6-31+G*: HF/6-31G*	9041	5442	12951	6153	3730	9160	12610	7121	22892
LC-BLYP/6-31+G*: PM6	11974	7593	42674	9522	5758	23832	12475	8517	113765
LC-BLYP/6-31+G*: HF/3-21G	12620	7534	40935	8917	5288	22455	15264	8509	108785
High model:	<i>chromophore + chain + ARG 95</i>			<i>chromophore + chain + HIS 202 + ARG 95</i>			<i>chromophore + chain + ARG 95</i>		
HF/6-31+G*: PM6	8457	5489	15044	6551	4026	10435	9698	6988	30282
High model:	<i>chromophore + chain + ARG 95 + 2H₂O + SER 146</i>			<i>chromophore + chain + HIS 202 + ARG 95 + H₂O + SER 148</i>			<i>chromophore + chain + ARG 95 + 2H₂O + SER 146</i>		
HF/6-31+G*: PM6	8478	5433	12864	6750	4207	10544	11777	7235	22211

TABLE. 9. Static and dynamic first hyperpolarizabilities (a.u.) of the deprotonated chromophores with a part of their protein environment calculated at different levels of approximation within the ONIOM formalism accounting for the solvent effects (water) via the IEFPCM scheme.

β_{HRS} in vacuo	∞	1900nm	1064nm	∞	1900nm	1064nm	∞	1900nm	1064nm
eGFP				eYFP			SHardonnay		
High model:	chromophore + chain			chromophore + chain + TYR 203			chromophore + chain + PHE 203		
HF/6-31+G* : PM6	2358	2779	4564	1729	2091	3741	1785	2330	4966
HF/6-31+G* : HF/3-21G	1939	2344	4105	1725	2095	3698	2002	2386	4014
HF/6-31+G* : HF/6-31G*	1830	2219	3926	1647	1996	3493	1876	2229	3681
LC-BLYP/6-31+G* : PM6	2501	3120	6675	1939	2553	6501	2397	3431	10460
LC-BLYP/6-31+G* : HF/3-21G	2069	2670	6194	1965	2584	6461	2618	3487	9499
High model:	chromophore + chain + GLN 94			chromophore + chain + TYR 203 + ARG 96			chromophore + chain + PHE 203 + ARG 96		
HF/6-31+G* : PM6	2508	2933	4700	1602	1941	3466	1538	1957	3859
High model:	chromophore + chain + GLN 94 + H2O			chromophore + chain + TYR 203+ ARG 96 + H2O			chromophore + chain + PHE 203 + ARG 96 + H2O + HIS 148		
HF/6-31+G* : PM6	2433	2877	4739	1603	1933	3361	1565	1896	3321
High model:				chromophore + chain					
HF/6-31+G* : PM6				1518	1907	3586			
DsRed				zFP538			mCherry		
High model:	chromophore + chain			chromophore + chain + HIS 202			chromophore + chain		
HF/6-31+G* : PM6	4626	6164	16348	3835	5221	21200	7532	10162	32855
HF/6-31+G* : HF/3-21G	4191	5607	14772	3081	4260	12543	5310	7607	29357
HF/6-31+G* : HF/6-31G*	4080	5485	14564	3155	4399	13133	5520	8033	34142
LC-BLYP/6-31+G* : PM6	4906	7226	40127	4948	7355	42382	7335	10594	98212
LC-BLYP/6-31+G* : HF/3-21G	4437	6620	38411	4207	6411	37873	5058	7980	94687
High model:	chromophore + chain + ARG 95			chromophore + chain + HIS 202 + ARG 95			chromophore + chain + ARG 95		
HF/6-31+G* : PM6	3569	5371	20329	3604	5368	26151	5530	8656	53363
High model:	chromophore + chain + ARG 95 + 2H2O + SER 146			chromophore + chain + HIS 202 + ARG 95 + H2O + SER 148			chromophore + chain + ARG 95 + 2H2O + SER 146		
HF/6-31+G* : PM6	4048	5511	14744	3880	5516	23777	6867	9317	28412

TABLE 10. Static and dynamic first hyperpolarizabilities (a.u.) of the deprotonated chromophores with a part of their protein environment calculated at different levels of approximation within the ONIOM formalism *in vacuo*.

HF	$\beta_{HRS}(\infty)$	$\beta_{HRS}(1900\text{nm})$	$\beta_{HRS}(1064\text{nm})$	$\beta_{HRS}(\infty)$	$\beta_{HRS}(1900\text{nm})$	$\beta_{HRS}(1064\text{nm})$
in vacuo				in water (IEFPCM)		
eGFP						
6-31+G*	1661	2020	3562	5060	2973	5356
6-31G*	1764	2113	3559	4991	2985	5149
eYFP						
6-31+G*	1622	1943	3235	5016	3295	5430
6-31G*	1565	1865	3060	4806	2936	5004
SHardonnay						
6-31+G*	1759	2060	3190	4518	3012	4925
6-31G*	1742	2032	3108	4397	2976	4779
DsRed						
6-31+G*	4143	5511	14026	8675	5385	12453
6-31G*	3935	5224	12797	8202	5138	11366
zFP538						
6-31+G*	3299	4468	12385	6100	4053	8921
6-31G*	3223	4364	11565	5837	3667	8315
mCherry						
6-31+G*	5776	8436	38069	12405	7329	29895
6-31G*	5385	7745	30722	11368	7212	19618

TABLE 11. Basis set effects on the static and dynamic HRS first hyperpolarizabilities (a.u.) of the deprotonated chromophores with a part of their protein environment (full HF results). The calculations were carried out at the HF level in vacuo or using the IEFPCM scheme (solvent = water).

A related analysis can be performed while enlarging the high layer model and using HF/6-31+G* : PM6 calculations. For eGFP, starting from the smallest high layer model, adding GLN 94, and then a water molecule the β_{ONIOM} overestimation goes from 21% to 29%, and then to 22%. This demonstrates the small impact on β of a better description of the H-bonds. Similar conclusions can be drawn for eYFP: as soon as TYR 203 is included in the high layer, the agreement with the full HF

calculation remains within 4% whereas the underestimation attains 10% when TYR 203 is in the low layer. In the case of SHardonnay, it is not straightforward. There is a strong impact of the description of the H-bonds, as the inclusion in the high layer of the H-bonded ARG 96, HIS 148 and a water molecule is needed to get an underestimation of less than 6%. If not, the HF value is underestimated strongly in the order of 38 to 36%. PM6 method does not succeed to describe the stronger H-bond with the HIS 148, which is 1.794 Å between the oxygen and the hydrogen. For DsRed, the underestimation amounts to less than 1% for the smallest high layer, to 2.5% if the H-bonded ARG 95 has been added, and to 2% for the maximal high layer. For zFP538, there is no strong difference if H-bonds are included or not in the high layer; it remains between 7 to 11% of overestimation with respect to the reference value. In the case of mCherry, there is 20% of difference without including the H-bonds in the high layer. When all of it is taken into account, this difference decreases to 5%. An improved description of the H-bonds can be beneficial for the ONIOM calculations when the PM6 method failed to describe well the environment. As intermediate conclusion, the HF : PM6 ONIOM method gives a good compromise between accuracy and calculation resources, in particular when the high layer contains much of the physico-chemical effects. Still, in a number of cases, using the HF method to describe the low layer leads to real improvements.

III.D.3.b. Solvent effects

In Figure 14, Table 9 and 10 the static ONIOM HF/6-31+G* : PM6 first hyperpolarizability of eGFP (green), eYFP (yellow), SHardonnay (purple), DsRed (red), zFP538 (pink) and mCherry (grey) is shown as a function of the dielectric constant of the solvent used within the IEFPCM treatment. $\beta_{model, high}$ and $\beta_{environment}$ are also reported. The focus is set on the relative trends, keeping in mind that there is no inclusion of electron correlation effects, which is important for an accurate description of the β response.

The ONIOM calculations performed with an increasingly large high ONIOM layer show a nice consistency, demonstrating that accurate description is already achieved with the minimal high ONIOM layer except for SHardonnay and mCherry. Indeed, as previously shown, SHardonnay and mCherry need an accurate description of the H-bonds in the high layer, which is only achieved for the maximal high layer considered here. The β_{ONIOM} data reproduce the experimental trends except for SHardonnay, i.e. $\beta(\text{mCherry}) > \beta(\text{DsRed}) \geq \beta(\text{zFP538}) > \beta(\text{SHardonnay}) > \beta(\text{eGFP}) > \beta(\text{eYFP})$. Then, all contributions to β display similar solvent effects where β increases with ϵ , first strongly and then it converges towards a limiting value, with the exception of mCherry. In that case, with a minimal high layer, the environment effects failed to be well described by the PM6 method. Indeed, PM6 poorly described the β response when there are strong H-bonds with the chromophore. Unexpected bond lengths in the structure of mCherry for which the PM6 method is really sensible can cause this behavior. Indeed there is an unexpected bond length in the peptide bond of the chromophore, which is more a triple bond than double one.

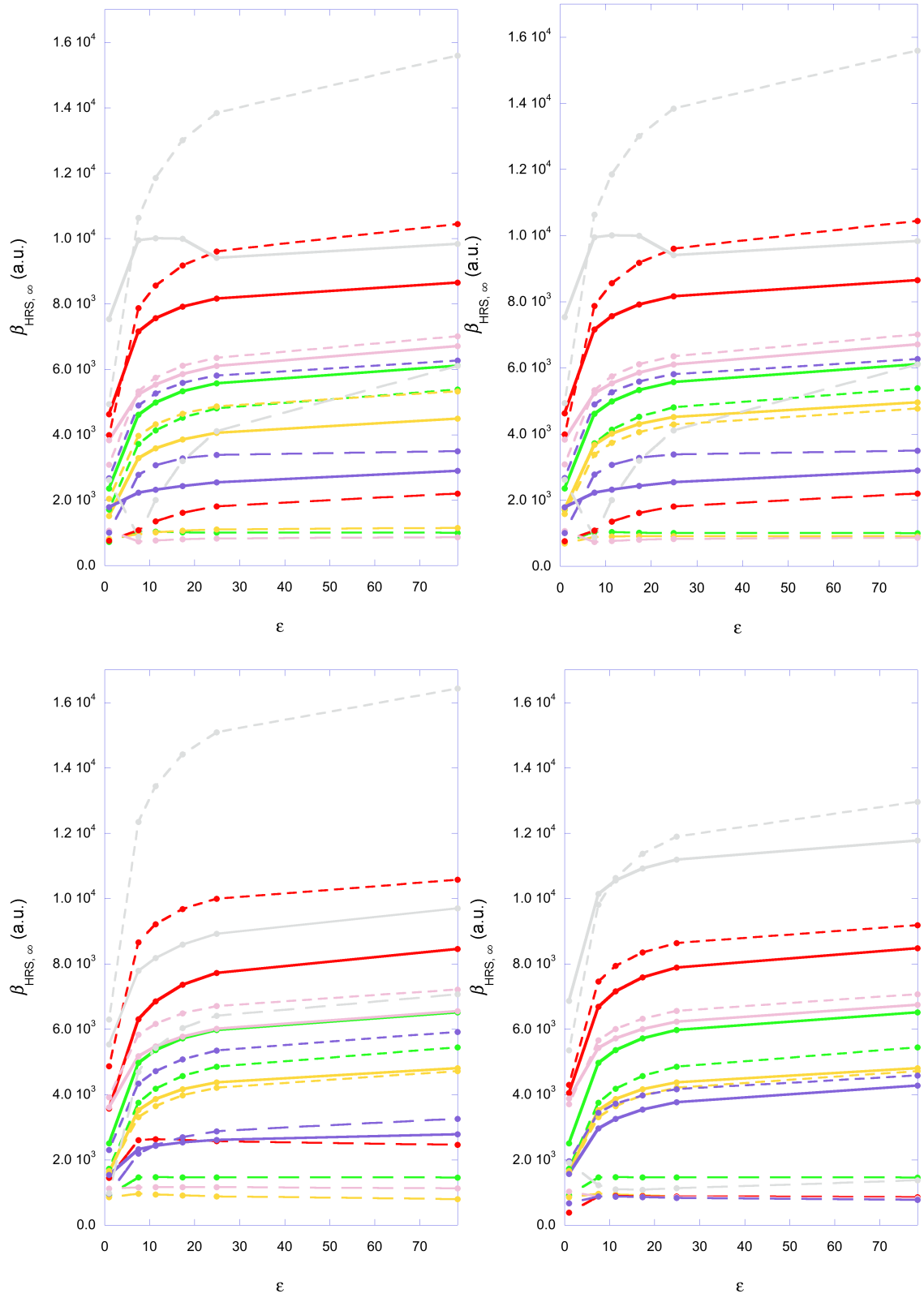


FIG. 14. Static first hyperpolarizability [β_{ONIOM} (plain line), $\beta_{\text{model,high}}$ (small dashed line), $\beta_{\text{environment}}$ (large dashed line)] in eGFP (green), eYFP (yellow), SHardonney (purple), DsRed (red), zFP538 (pink) and mCherry (grey) as a function of the solvent dielectric constant for different types of high level layers. **Top left)** the smallest high ONIOM layers; **Top right)** the smallest high ONIOM layers with the exception of eYFP where TYR 203 and the attached chain are included; **Bottom left)** next

high level models, namely, GLN 94, ARG 96 as well as TYR 203 and its attached chain, ARG 96, ARG 95, ARG 95 and ARG 95 are included for eGFP, eYFP, SHardonnay, DsRed, zFP538 and mCherry respectively; **Bottom right**) the largest high ONIOM layer, i.e. a water molecule is added with regards of the previous models. The straight lines correspond to the extrapolated experimental values.

The successive $\beta_{model, high}$ and $\beta_{environment}$ values are also rich of information: i) though $\beta_{model, high}(mCherry)$ is always the largest, followed by $\beta_{model, high}(DsRed)$ then $\beta_{model, high}(zFP538)$, the $\beta_{model, high}(SHardonnay) > \beta_{model, high}(eGFP)$ except when all the H-bond are in the high layer of SHardonnay, the $\beta_{model, high}(eGFP) > \beta_{model, high}(eYFP)$ condition is only satisfied if the model includes TYR 203 forming locally a centrosymmetric structure due to π -stacking, ii) in absence of TYR 203, $\beta_{model, high}(eGFP) \sim \beta_{model, high}(eYFP)$ since the NLOphore are identical and their geometries are similar, iii) further inclusion of additional amino acid residues and water molecule has a negligible impact except for mCherry and SHardonnay, iv) the $\beta_{environment}$ contribution to the SHardonnay, DsRed and mCherry response are reduced and becomes similar to those of the other FPs provided with the inclusion of all H-bond less than 2Å, demonstrating that the latters are combined with the NLOphore. For the maximal high layer model, the $\beta_{model, high}(eYFP)/\beta_{ONIOM}$ ratio tends towards one and to a lesser extend for the others FPs, pointing out that the remaining of the protein structure has little impact on the first hyperpolarizability.

III.D.3.c. Comparison of HF to MP2 results on small systems

Static HRS first hyperpolarizabilities of the deprotonated chromophore, calculated at the MP2 level are given in Table 12 and 13 together with the $\beta(MP2)/\beta(HF)$ ratios. When employing the IEFPCM/MP2/6-31G* method, the static first hyperpolarizabilities are almost twice larger than at the HF level. This behavior is quite different for the isolated species. Using the 6-31+G* basis set, the increase of

β with the respect to the HF response ranges between 2.159 and 2.683, a little bit more than for the 6-31G* basis set.

III.D.3.d. Assessment of DFT exchange-correlation functionals

In Table 14, the static and dynamic HRS first hyperpolarizabilities of the deprotonated chromophores are displayed, calculated with the 6-31+G* basis set at the DFT level of calculation using the B3LYP and LC-BLYP exchange-correlation functionals, accounting for the solvent effects with the IEFPCM scheme. Using B3LYP underestimates systematically the static MP2 β values with a nonsystematic amount. As it is shown in Chapter 5, the B3LYP method is not reliable for describing second-order NLO properties. In the case of LC-BLYP, systematic underestimations of 33 to 40% are observed, which is consistent with the data reported in Chapter 5.

	in vacuo		in water (IEFPCM)	
	$\beta_{HRS}(\infty)$	$\beta(MP2)/\beta(HF)$	$\beta_{HRS}(\infty)$	$\beta(MP2)/\beta(HF)$
<i>6-31G*</i>				
eGFP	1175	0.896	11890	1.969
eYFP + TYR 203	2161	1.564	10302	2.067
SHardonnay	4976	1.820	16537	2.629
DsRed	5724	1.808	16032	2.019
zFP538 + HIS 202	6650	1.898	14540	1.971
mCherry	6752	1.294	26174	1.934
<i>6-31+G*</i>				
eGFP	1380	0.950	15482	2.159
eYFP + TYR 203	2664	1.801	13134	2.283
SHardonnay	5824	1.980	17897	2.683
DsRed	6041	1.832	19710	2.184
zFP538 + HIS 202	7485	2.021	17481	2.169
mCherry	7683	1.396	33130	2.199

TABLE 12. Static HRS first hyperpolarizabilities of the deprotonated chromophores, calculated at the MP2 level with or without accounting for the effects of the solvent (water) via the IEFPCM scheme. The $\beta_{HRS}(MP2)/\beta_{HRS}(HF)$ ratios are also given

	$\beta_{HRS}(\infty)$	$\beta_{HRS}(1900\text{nm})$	$\beta_{HRS}(1064\text{nm})$
eGFP	15482	7075	12414
eYFP + TYR 203	13134	6242	11146
SHardonnay	17897	9943	17321
DsRed	19710	10745	22482
zFP538 + HIS 202	17481	9656	24845
mCherry	33130	19868	62918

TABLE 13. Static and dynamic HRS first hyperpolarizabilities of the deprotonated chromophores, calculated at the IEFPCM(water)/MP2/6-31+G* level. The dynamic values were calculated using the multiplicative scheme.

β_{HRS}	∞	1900nm	1064nm	∞	1900nm	1064nm
	B3LYP			LC-BLYP		
eGFP	4688	2274	6397	9474	4570	11200
eYFP + TYR 203	4462	2658	5160	8638	4160	10206
SHardonnay	9262	4715	18043	11887	6487	14847
DsRed	13811	9813	194599	11961	7047	23037
zFP538 + HIS 202	10031	7374	816748	10727	6529	27751
mCherry	13260	8718	72699	18709	11683	90351

TABLE 14. IEFPCM(water)/DFT static and dynamic HRS first hyperpolarizabilities of the deprotonated chromophores, calculated with the 6-31+G* basis set using the B3LYP and LC-BLYP XC functionals.

III.D.4. Overall comparison with experiment

The TDHF/CPHF HRS results for the entire systems are shown in Table 11. The measured and extrapolated static values of the HRS first hyperpolarizabilities are presented in Table 15. These extrapolated values have been calculated using the

two-state approximation incorporating an inhomogeneous broadening based on the absorption spectrum. Static extrapolated values are compared to the TDHF/CPHF results for the full systems (Table 16), where solvent effects are modeled at the PCM level. This comparison is however not straightforward because H₂O is a dipolar solvent and therefore its static dielectric constant (78.355) is substantially larger than its optical dielectric constant (1.77 for visible light). Since larger dielectric constants lead to an enhancement of β , it is/might be advantageous to make comparisons with β values calculated at optical frequencies far from electronic resonances, *i.e.* where the frequency dispersion is small if not negligible. Indeed, the IEFPCM/CPHF β values are larger than the experimental ones by 15%, 67%, 72%, and 14% for eGFP, eYFP, SHardonnay, and DsRed, respectively. But it is not the case for zFP538, and mCherry where the IEFPCM/CPHF β values are smaller of 20-24% with the respect to the extrapolated experimental one. In those simulations, the IEFPCM approach is employed to describe the effects of the rest of the protein surrounding on β . Water is therefore the obvious solvent for such modeling but the real environment is also made of organic moieties. It is therefore expected that the water solvent in the IEFPCM approach can only give the upper limit of the solvent effects on the static first hyperpolarizabilities. As a consequence, other solvents need to be employed to assess the range of variations of β . These calculations are performed within the ONIOM scheme to reduce the computational effort and are discussed in the section III.D.3.b.

	eGFP	eYFP	SHardonnay	DsRed	zFP538	mCherry
$\beta_{HRS, 800}$	24800 $\pm$ 3900	8600 $\pm$ 900	18520 $\pm$ 1158	18800 $\pm$ 1900	20835 $\pm$ 1158	31021 $\pm$ 3820
$\beta_{HRS, \infty}$	4400 $\pm$ 700	3010 $\pm$ 460	6250 $\pm$ 460	7640 $\pm$ 695	7640 $\pm$ 460	16200 $\pm$ 2080

TABLE 15. Measured and extrapolated values of the HRS first hyperpolarizabilities for eGFP, eYFP, DsRed, mCherry, SHardonnay, and zFP538.

The calculations have also been done *in vacuo*. As the chromophore is inside the protein, the solvent effects may be minimal and only the very first shell of the surrounding really influences the β response. For completeness, these two approaches have been included in our study.

Qualitatively, the TDHF/CPHF results *in vacuo* keep the experimental trends for the static response as well as for an optical frequency far from resonance i.e., 1900nm. For IEFPCM calculations in water, it is not so obvious. As it is stressed above, a wavelength of 1900nm is needed to make comparisons. The tendency is not good for eYFP, which is larger than eGFP and SHardonnay. Note that there is no difference in the chromophore structures of eGFP, eYFP and SHardonnay, therefore the difference in β values comes from the local centrosymmetric π -stacking interaction between the chromophore and TYR 203 in eYFP, leading also to a different geometry, which lowers the β response and the π -stacking interaction between the chromophore and PHE 203 in SHardonnay. For DsRed, the β response is larger because it has a longer π -conjugated pathway as well as for zFP538 and mCherry.

Furthermore, the CPHF/TDHF HRS static/dynamic first hyperpolarizabilities of the smallest systems, which are the deprotonated chromophores, are listed in Table 16. At the HF level, after accounting for the solvent effects, the same tendency for the static β_{HRS} as the experiments is obtained, with the exception of DsRed and zFP538 in π -stacking interaction with the HIS 202, and the exception of eGFP and SHardonnay. Indeed, the experimental β responses of DsRed and zFP538 are the same or almost the same if the error bars are accounted. In the calculations, for the zFP538 + HIS 202 result, the π -stacking interactions lower β_{HRS} with respect to the situation of an isolated chromophore and reverse the tendency with the respect to the DsRed value. Note that the π -stacking effect is lighter for zFP538 than for eYFP.

	in vacuo			in water (IEFPCM)		
	$\beta_{HRS}(\infty)$	β_{HRS} (1900nm)	β_{HRS} (1064nm)	$\beta_{HRS}(\infty)$	β_{HRS} (1900nm)	β_{HRS} (1064nm)
eGFP	1452	1730	2898	7171	3277	5750
eYFP	1908	2291	3849	7153	3587	6066
eYFP + TYR 203	1479	1843	3478	5752	2734	4882
SHardonnay	2940	3674	6984	6670	3706	6456
DsRed	3297	4356	10016	9026	4920	10294
zFP538	3580	5070	15693	9139	5070	15693
zFP538 + HIS 202	3703	5236	16070	8059	4452	11455
mCherry	5503	7905	27660	15067	9035	28612

TABLE 16. TDHF HRS first hyperpolarizabilities of the deprotonated chromophores using the 6-31+G* basis set.

mCherry exhibits the largest β response within this set of proteins. A careful look at the structures shows that DsRed and mCherry chromophores have the same chemical structure or at least the same π -conjugated pathway. Therefore one might have expected the same or at least a similar β response. However, the chromophore structures, extracted from X-ray diffraction data, are different, which might explain the difference in β response. In fact, this can be rationalized in terms of the average bond length alternation (BLA). Considering the BLA associated with the four bonds at the end of the π -conjugated segment (Figure 15) a BLA values of 0.17Å and 0.00Å are obtained for DsRed and mCherry, respectively. This difference is sufficiently large to explain the largest β response of mCherry.

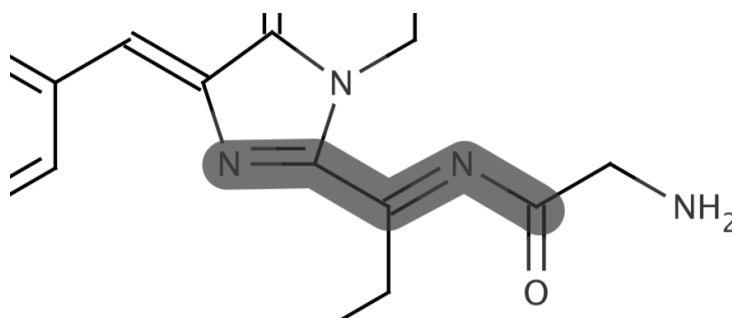


FIG. 15. A sketch of the bonds taken into the calculation of the average bond length alternation (BLA).

Without accounting for the solvent effects with the HF method, the trends differ from experiment. As discussed above, upon considering a larger fraction of the protein environment in vacuo, the calculated values match again the experimental trends. Therefore, one can conclude that it is compulsory to account for the effects of the surroundings on β . A simplified way is to model it using a polarizable continuum, which will provide roughly an upper limit to those effects.

Quantitative comparison with experiment is however a tremendous task for quantum chemistry, as there are a lot of experimental issues that cannot be described in a modeling. All experiments have been carried out in water solution; in our model as it is not possible to take into account all the protein, if the IEFPCM approach is used with water solvent, the solvent effect will be higher than in reality. In solution there is some equilibrium between monomeric, dimeric and tetrameric proteins. mCherry⁴² is a monomer. eGFP²², eYFP²², and SHardonnay are in a weak equilibrium between the monomer and the dimer forms. DsRed^{8b, 40} and zFP538¹⁹⁻²⁰ are tetrameric proteins. The solvent effects will be lighter for dimer and tetramer than for monomeric ones due to a higher screening of the solvent interaction by the protein. In the simulation, it is possible that for monomeric proteins or for proteins (for which the equilibrium with dimer form is more displaced to the monomer form), that accounting the solvent effects is more important to get the best responses. It shall not be the case for dimers as well as for tetramers. The comparison between experiment and theory needs β values calculated at optical frequencies far from electronic resonance i.e. at 1900nm. For calculations that have been done without solvent, the static values can be used as well.

In the case of MP2 calculation on the isolated chromophores (Table 12 and 13), without accounting for the solvent, for eYFP, SHardonnay and DsRed, an overestimation of the experimental values of, respectively, 10%, 16% and 1% is obtained, respectively, and of 38% for zFP538. For eGFP and mCherry, strong

underestimations are obtained. If the solvent effects with the IEFPCM method for water are added, the values are largely higher than the experimental ones. In a second time, the cheaper way to mix correlation and environment effects is to use the long-range corrected exchange correlation functional LC-BLYP for the minimal High layer in an ONIOM calculation (Table 9). Systematic lower values with respect the experimental ones are only obtained for eYFP, DsRed and zFP538 with the LC-BLYP/6-31+G* : HF/3-21G ONIOM scheme, respectively, 14%, 13% and 16%. If solvent effects are accounted, less than 1% of the difference with respect to the experiment for eGFP and DsRed with the LC-BLYP/6-31+G* : PM6 method is obtained.

A strong improvement can be obtained by doing some MP2 calculations for the High layer in an ONIOM scheme. This is applied on reasonable size High layer (see section III.D.2.). In the case of MP2/6-31+G* : PM6 calculations (Table 17), without solvent effects, the static first hyperpolarizability of DsRed is 3% larger than the experiment and 5% larger for zFP538. If the optical frequency is increased to a wavelength of 1900nm, a systematic increase of 36-37% is obtained. This systematic increase is also observed for eYFP and SHardonnay by 32%. In their cases, they are in good agreement with respect to the experimental values, with for eYFP 3% lower and SHardonnay 2% larger. In the case of eGFP and mCherry, strong underestimation is observed without accounting for solvent effects. Moreover, when accounting for the solvent effects, for mCherry at a wavelength of 1900nm, only 1% of underestimation is obtained and for eGFP 31% of overestimation. As mCherry is a monomeric protein, it seems that the solvent effect is important to account, which is not necessarily the case for DsRed and zFP538.

As the PM6 method can strongly fail to describe the first hyperpolarizability for systems with triple bonds or strong H-bonds, it is judicious to stay in a more rigorous approach with Hartree-Fock method to describe the low layer. With the

MP2/6-31+G* : HF/6-31G* ONIOM scheme (Table 17) in vacuo, for eYFP, DsRed and zFP538, less than 8% of underestimation is obtained with respect to the experimental first hyperpolarizabilities. For SHardonnay, there is 22% of underestimation. If it is compared to calculations made at 1900nm, there is a systematic increase of the response for these four proteins, where SHardonnay is less than 6% of underestimation. As seen for the previous method, a strong underestimation of the experimental values is obtained for eGFP and mCherry. By including the solvent effects, 18% of overestimation for eGFP is obtained and less than 1% for mCherry. It seems that eGFP has a more monomeric than dimeric character, which explains the importance of the solvent effects for eGFP and mCherry than for the others. If values with solvent effect are accounted for eGFP and mCherry and without for eYFP, SHardonnay, DsRed and zFP538 the ratios with respect to the experimental values shown in table 18, are mostly in quantitative agreement with experiments.

<i>in vacuo</i>					<i>in water (IEFPCM)</i>			
β_{HRS}	∞	MP2:PM6/ HF:PM6	1900nm	1064nm	∞	MP2:PM6/ HF:PM6	1900nm	1064nm
MP2/6-31+G* : PM6								
eGFP	2516	0.923	2873	4267	10590	1.455	5768	8832
eYFP	2374	1.781	2914	5703	9038	1.833	5356	8824
SHardonnay	4793	2.523	6343	14050	12595	4.007	14016	25404
DsRed	7854	1.683	10374	25210	17731	2.051	11049	25594
zFP538	7986	1.831	10931	42058	15053	2.163	9565	23559
mCherry	8903	1.191	12251	39665	24004	2.353	16069	65962
MP2:HF/ HF:HF					MP2:HF/ HF:HF			
MP2/6-31+G* : HF/6-31G*								
eGFP	1719	0.910	2078	3626	9052	1.583	5180	9199
eYFP	2865	1.729	3491	6299	9606	1.778	5622	9875
SHardonnay	4907	2.525	5886	10082	14314	3.010	9406	15802
DsRed	7356	1.770	9917	25964	18164	2.006	11085	25709
zFP538	7064	2.076	9819	29440	14259	2.267	9227	21085
mCherry	7210	1.261	10538	44960	26858	2.067	16408	50414

TABLE 17. Static and extrapolated dynamic first hyperpolarizabilities (a.u.) of the deprotonated chromophores with a part of their protein environment calculated at different levels of approximation within the ONIOM formalism, where the High layer is calculated at the MP2 level of calculation.

	$\beta_{HRS}(\infty)/\beta_{HRS}(\text{exp})$	$\beta_{HRS}(1900\text{nm})/\beta_{HRS}(\text{exp})$
MP2/6-31+G* : HF/6-31G*		
eGFP (in water)	/	1.178
eYFP	0.952	1.160
SHardonnay	0.785	0.942
DsRed	0.963	1.298
zFP538	0.925	1.285
mCherry (in water)	/	1.013

TABLE 18. $\beta_{HRS}(\text{ONIOM})/\beta_{HRS}(\text{exp})$ ratios for MP2/6-31+G* : HF/6-31G* ONIOM method.

Finally, a better level of treatment can be obtained by using the ONIOM MP2:HF/6-31+G* scheme (Table 19). Similar conclusions can be drawn with this method as for eYFP, DsRed and zFP538, are less than 10% lower than the experimental values. For SHardonnay, there is 24% of underestimation. If the response at 1900nm is calculated, a systematic increase of the β is observed. By accounting the solvent effects, there is for eGFP 13% of overestimation of the experimental value and for mCherry 3% of underestimation. These high level results are really difficult to obtain and cannot be carried out in a reasonable time but it is useful to validate results of other methods as for example the MP2/6-31+G(d) : HF/6-31G(d), which is 15 times faster and gives us similar results. The ONIOM MP2:HF/6-31+G* scheme has success to give an accurate description of the first hyperpolarizability without compromise in the description of the surrounding. The only approximation of this method is the estimation of the correlation energy that has been done on a smaller system. As the NLOphore is responsible for most of the β response and as the correlation effect is important to describe it accurately, it seems the method of choice to characterize such systems.

β_{HRS}	<i>in vacuo</i>				<i>in water (IEFPCM)</i>			
	∞	ONIOM/HF	1900nm	1064nm	∞	ONIOM/HF	1900nm	1064nm
MP2 : HF 6-31+G*								
eGFP	1558	0.938	1895	3341	8448	1.670	4965	8945
eYFP	2784	1.716	3334	5551	9232	1.840	6063	9991
SHardonnay	4721	2.684	5529	8562	14073	3.115	9382	15340
DsRed	7334	1.770	9755	24826	17899	2.063	11109	2569
zFP538	6885	2.087	9325	25847	14165	2.322	9411	20715
mCherry	7251	1.255	10587	47777	26383	2.127	15589	63587

TABLE 19. Static and extrapolated dynamic first hyperpolarizabilities (a.u.) of the deprotonated chromophores with a part of their protein environment calculated at different levels of approximation within the ONIOM formalism, where the High layer is calculated at the MP2 level of calculation.

III.D.5. Conclusions

Predicting and interpreting the second-order nonlinear optical response of fluorescent proteins is a real challenge for quantum chemistry. The performance of successive versions of the ONIOM scheme has been assessed in this study. When the high layer contains much of the physico-chemical effects, the HF : PM6 ONIOM method can be used in order to have similar accuracy than full HF calculations. Some improvements are obtained by using the HF method to describe the low layer with a smaller basis set, taking the advantage that smaller high layer can also be used. In order to have a higher level of treatment, the ONIOM MP2:HF/6-31+G* scheme is very promising to compute first hyperpolarizabilities of such kinds of systems but it is a time consuming method since a full Hartree-Fock calculation is needed with the same large basis set as the correlated method. An alternative is the

ONIOM MP2/6-31+G* : HF/6-31G* method where the accuracy is quite similar but the calculations are 15 times faster. The PM6 method for the description of the low layer in that very case should be avoided since larger high layers need to be taken into account at the MP2 level in order to keep most of the physic-chemical effects, with the consequence to increase the computational time.

Among the methods used to make an overall comparison with experiment, the CPHF/TDHF methods give results on the full systems *in vacuo* that keep the experimental trends for the static response. A simple way to get roughly an upper limit to this qualitative agreement is to take the structure of chromophores only and to model the effects of the surrounding using a polarizable continuum. However, a quantitative comparison to the experiment needs to have an accurate description of the electron correlation effects for the chromophore as well as a precise description of the surroundings. The ONIOM MP2 : HF/6-31+G* method presents all these requirements but the ONIOM MP2/6-31+G* : HF/6-31G* method can be a faster alternative. Indeed, for this set of fluorescent proteins, quantitative agreements can be obtained by using these two methods, considering the hypothesis of the oligomer nature of eYFP, SHardonnay, DsRed and zFP538. A direct follow-up to this study is to demonstrate this hypothesis. The principal limitation in this approach is the use of the unmodified experimental X-ray diffraction structures.

REFERENCES

1. Asselberghs, I.; Flors, C.; Ferrighi, L.; Botek, E.; Champagne, B.; Mizuno, H.; Ando, R.; Miyawaki, A.; Hofkens, J.; Van der Auweraer, M.; Clays, K., *J. Am. Chem. Soc.* **2008**, *130* (46), 15713-9.
2. Pakhomov, A. A.; Martynov, V. I., *Chem. Biol.* **2008**, *15* (8), 755-64.
3. Clays, K.; Hendrickx, E.; Triest, M.; Verbiest, T.; Persoons, A.; Dehu, C.; Brédas, J.-L., *Science* **1993**, *262* (5138), 1419-1422.
4. Hsu, C.-C.; Huang, T.-H.; Zang, Y.-L.; Lin, J.-L.; Cheng, Y.-Y.; Lin, J. T. s.; Wu, H. H.; Wang, C. H.; Kuo, C.-T.; Chen, C.-H., *J. App. Phys.* **1996**, *80* (10), 5996-6001.
5. (a) Olbrechts, G.; Strobbe, R.; Clays, K.; Persoons, A., *Rev. Scient. Instr.* **1998**, *69* (6), 2233-2241; (b) Wostyn, K.; Binnemans, K.; Clays, K.; Persoons, A., *Rev. Scien. Instrum.* **2001**, *72* (8), 3215-3220.
6. Noordman, O. F. J.; van Hulst, N. F., *Chem. Phys. Lett.* **1996**, *253* (1-2), 145-150.
7. (a) Reeve, J. E.; Collins, H. A.; Mey, K. D.; Kohl, M. M.; Thorley, K. J.; Paulsen, O.; Clays, K.; Anderson, H. L., *J. Am. Chem. Soc.* **2009**, *131* (8), 2758-2759; (b) Clays, K.; Olbrechts, G.; Munters, T.; Persoons, A.; Kim, O.-K.; Choi, L.-S., *Chem. Phys. Lett.* **1998**, *293* (5-6), 337-342; (c) Olbrechts, G.; Wostyn, K.; Clays, K.; Persoons, A.; Ho Kang, S.; Kim, K., *Chem. Phys. Lett.* **1999**, *308* (3-4), 173-175; (d) Ho Kang, S.; Kim, K.; Kang, Y.-S.; Zin, W.-C.; Olbrechts, G.; Wostyn, K.; Clays, K.; Persoons, A., *Chem. Comm.* **1999**, (17), 1661-1662; (e) Olbrechts, G.; Wostyn, K.; Clays, K.; Persoons, A., *Opt. Lett.* **1999**, *24* (6), 403-405; (f) Clays, K.; Wostyn, K.; Olbrechts, G.; Persoons, A.; Watanabe, A.; Nogi, K.; Duan, X.-M.; Okada, S.; Oikawa, H.; Nakanishi, H.; Vogel, H.; Beljonne, D.; Brédas, J.-L., *J. Opt. Soc. Am. B* **2000**, *17* (2), 256-265; (g) Zhang, T.-G.; Zhao, Y.; Asselberghs, I.; Persoons, A.; Clays, K.; Therien, M. J., *J. Am. Chem. Soc.* **2005**, *127* (27), 9710-9720; (h) Hennrich, G.; Murillo, M. T.; Prados, P.; Al-

- Saraierh, H.; El-Dali, A.; Thompson, D. W.; Collins, J.; Georghiou, P. E.; Teshome, A.; Asselberghs, I.; Clays, K., *Chem. Eur. J.* **2007**, *13* (27), 7753-7761;
- (i) Duncan, T. V.; Song, K.; Hung, S.-T.; Miloradovic, I.; Nayak, A.; Persoons, A.; Verbiest, T.; Therien, M. J.; Clays, K., *Angew. Chem. Int. Ed. Engl.* **2008**, *47* (16), 2978-2981.
8. (a) Wachter, R. M.; Elsliger, M.-A.; Kallio, K.; Hanson, G. T.; Remington, S. J., *Structure* **1998**, *6* (10), 1267-1277; (b) Elsliger, M.-A.; Wachter, R. M.; Hanson, G. T.; Kallio, K.; Remington, S. J., *Biochem.* **1999**, *38* (17), 5296-5301.
9. Sekino, H.; Bartlett, R. J., *J. Chem. Phys.* **1986**, *85* (2), 976-989.
10. (a) Tomasi, J.; Persico, M., *Chem. Rev.* **1994**, *94* (7), 2027-2094; (b) Tomasi, J.; Mennucci, B.; Cammi, R., *Chem. Rev.* **2005**, *105* (8), 2999-3094.
11. (a) Cavillot, V.; Champagne, B., *Chem. Phys. Lett.* **2002**, *354* (5-6), 449-457; (b) Masunov, A.; Tretiak, S., *J. Phys. Chem. B* **2003**, *108* (3), 899-907; (c) Rappoport, D.; Furche, F., *J. Am. Chem. Soc.* **2004**, *126* (4), 1277-1284; (d) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K., *J. Chem. Phys.* **2004**, *120* (18), 8425-8433; (e) Homem-de-Mello, P.; Mennucci, B.; Tomasi, J.; da Silva, A. B. F., *THEOCHEM* **2005**, *113* (5), 274-280; (f) Guillaume, M.; Champagne, B.; Zutterman, F., *J. Phys. Chem. A* **2006**, *110* (48), 13007-13013; (g) Miura, M.; Aoki, Y.; Champagne, B., *J. Chem. Phys.* **2007**, *127* (8), 084103; (h) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W., *J. Chem. Phys.* **2008**, *128* (13), 134110; (i) Peach, M. J. G.; Sueur, C. R. L.; Ruud, K.; Guillaume, M.; Tozer, D. J., *Phys. Chem. Chem. Phys.* **2009**, *11* (22), 4465-4470.
12. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao,

- O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Laham, A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A., Gaussian 03, revision c.02. 2003.
13. Campo, J.; Wenseleers, W.; Goovaerts, E.; Szablewski, M.; Cross, G. H., *J. Phys. Chem. C* **2008**, *112* (1), 287-296.
 14. Oudar, J. L.; Chemla, D. S., *J. Chem. Phys.* **1977**, *66* (6), 2664-2668.
 15. (a) Kaatz, P.; Shelton, D. P., *J. Chem. Phys.* **1996**, *105* (10), 3918-3929; (b) Pauley, M. A.; Wang, C. H., *Chem. Phys. Lett.* **1997**, *280* (5-6), 544-550; (c) Woodford, J. N.; Wang, C. H.; Asato, A. E.; Liu, R. S. H., *J. Chem. Phys.* **1999**, *111* (10), 4621-4628; (d) Berkovic, G.; Meshulam, G.; Kotler, Z., *J. Chem. Phys.* **2000**, *112* (9), 3997-4003; (e) Wang, C. H., *J. Chem. Phys.* **2000**, *112* (4), 1917-1924.
 16. Mancois, F.; Pozzo, J.-L.; Pan, J.; Adamietz, F.; Rodriguez, V.; Ducasse, L.; Castet, F.; Plaquet, A.; Champagne, B., *Chem. Eur. J.* **2009**, *15*, 2560-2571.
 17. (a) Lewis, A.; Khatchatouriants, A.; Treinin, M.; Chen, Z.; Peleg, G.; Friedman, N.; Bouevitch, O.; Rothman, Z.; Loew, L.; Sheres, M., *Chem. Phys.* **1999**, *245* (1-3), 133-144; (b) Khatchatouriants, A.; Lewis, A.; Rothman, Z.; Loew, L.; Treinin, M., *Biophys. J.* **2000**, *79* (5), 2345-2352.
 18. Nikolenko, V.; Nemet, B.; Yuste, R., *Methods* **2003**, *30* (1), 3-15.

19. Remington, S. J.; Wachter, R. M.; Yarbrough, D. K.; Branchaud, B.; Anderson, D. C.; Kallio, K.; Lukyanov, K. A., *Biochem.* **2004**, *44* (1), 202-212.
20. Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., *Biochem.* **2006**, *45* (32), 9639-9647.
21. Cotlet, M.; Hofkens, J.; Habuchi, S.; Dirix, G.; Van Guyse, M.; Michiels, J.; Vanderleyden, J.; De Schryver, F. C., *Proc. Nat. Acad. Scien.* **2001**, *98* (25), 14398-14403.
22. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y., *Nat. Biotech.* **2004**, *22* (12), 1567-1572.
23. Drobizhev, M.; Tillo, S.; Makarov, N. S.; Hughes, T. E.; Rebane, A., *J. Phys. Chem. B* **2009**, *113* (4), 855-859.
24. (a) Shaner, N. C.; Lin, M. Z.; McKeown, M. R.; Steinbach, P. A.; Hazelwood, K. L.; Davidson, M. W.; Tsien, R. Y., *Nat. Meth.* **2008**, *5* (6), 545-551; (b) Shaner, N. C.; Steinbach, P. A.; Tsien, R. Y., *Nat. Meth.* **2005**, *2* (12), 905-909; (c) Subach, F. V.; Piatkevich, K. D.; Verkhusha, V. V., *Nat. Meth.* **2011**, *8* (12), 1019-1026; (d) Piljić, A.; de Diego, I. a.; Wilmanns, M.; Schultz, C., *ACS Chem. Biol.* **2011**, *6* (7), 685-691; (e) Truong, K.; Sawano, A.; Mizuno, H.; Hama, H.; Tong, K. I.; Mal, T. K.; Miyawaki, A.; Ikura, M., *Nat. Struct. Mol. Biol.* **2001**, *8* (12), 1069-1073; (f) Betzig, E.; Patterson, G. H.; Sougrat, R.; Lindwasser, O. W.; Olenych, S.; Bonifacino, J. S.; Davidson, M. W.; Lippincott-Schwartz, J.; Hess, H. F., *Science* **2006**, *313* (5793), 1642-1645; (g) Xiong, A. S.; Peng, R. H.; Zhuang, J.; Davies, J.; Zhang, J.; Yao, Q. H., *Crit. Rev. Biotechnol.* **2012**, *32* (2), 133-42.
25. Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zarausky, A. G.; Markelov, M. L.; Lukyanov, S. A., *Nat. Biotechnol.* **1999**, *17* (10), 969-73.

26. Pletneva, N. V.; Pletnev, S. V.; Chudakov, D. M.; Tikhonova, T. V.; Popov, V. O.; Martynov, V. I.; Wlodawer, A.; Dauter, Z.; Pletnev, V. Z., *Russ. J. Bioorga. Chem.* **2007**, 33 (4), 390-398.
27. Bradford, M. M., *Anal. Biochem.* **1976**, 72 (1–2), 248-254.
28. Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J., *THEOCHEM* **1999**, 461-462 (0), 1-21.
29. Stewart, J. *Mopac2009*, Stewart Computational Chemistry: Colorado Springs, CO, USA, 2009.
30. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., *Gaussian 09*, revision a.01. Wallingford CT, 2009.
31. Krissinel, E.; Henrick, K., *Acta Cryst. D* **2004**, 60 (12 Part 1), 2256-2268.
32. Remington, S. J.; Wachter, R. M.; Yarbrough, D. K.; Branchaud, B.; Anderson, D. C.; Kallio, K.; Lukyanov, K. A., *Biochem.* **2005**, 44, 202-212.
33. McGaughey, G. B.; Gagne, M.; Rappe, A. K., *J. Biol. Chem.* **1998**, 273, 15458-15463.

34. Karna, S. P.; Dupuis, M., *J. Comput. Chem.* **1991**, *12* (4), 487-504.
35. Mey, K. D.; Perez-Moreno, J.; Reeve, J. E.; Lopez-Duarte, I.; Boczarow, I.; Anderson, H. L.; Clays, K., *J. Phys. Chem. C* **2012**, *116* (25), 13781-13787.
36. (a) De, M. E.; Chen, W.-Q.; Van, C. S.; Zheng, M.-L.; Psilodimitrakopoulos, S.; Paesen, R.; Taymans, J.-M.; Ameloot, M.; Vanderleyden, J.; Loza-Alvarez, P.; Duan, X.-M.; Clays, K., *Chem. Sci.* **2012**, *3*, 984-995; (b) Reeve, J. E.; Anderson, H. L.; Clays, K., *Phys. Chem. Chem. Phys.* **2010**, *12* (41), 13484-98.
37. (a) Reeve, J. E.; Corbett, A. D.; Boczarow, I.; Wilson, T.; Bayley, H.; Anderson, H. L., *Biophys. J.* **2012**, *103*, 907-917; (b) Millard, A. C.; Terasaki, M.; Loew, L. M., *Biophys. J.* **2005**, *88*, L46-L48.
38. (a) Lazar, J.; Bondar, A.; Timr, S.; Firestein, S. J., *Nat. Methods* **2011**, *8*, 684-690; (b) Knopfel, T.; Tomita, K.; Shimazaki, R.; Sakai, R., *Methods* **2003**, *30*, 42-48; (c) Kress, A.; Ferrand, P.; Rigneault, H.; Trombik, T.; He, H.-T.; Marguet, D.; Brasselet, S., *Biophys. J.* **2011**, *101*, 468-476.
39. Stewart, J. J., *J. Mol. Mod.* **2007**, *13* (12), 1173-213.
40. Yarbrough, D.; Wachter, R. M.; Kallio, K.; Matz, M. V.; Remington, S. J., *Proc. Nat. Aca. Sci.* **2001**, *98* (2), 462-467.
41. Vreven, T.; Mennucci, B.; da Silva, C. O.; Morokuma, K.; Tomasi, J., *J. Chem. Phys.* **2001**, *115* (1), 62-72.
42. Shaner, N. C.; Steinbach, P. A.; Tsien, R. Y., *Nat Meth* **2005**, *2* (12), 905-909.

How the second-order nonlinear optical response of the collagen triple helix appears: a theoretical investigation

M. de Wergifosse, J. de Ruyck, and B. Champagne, *J. Phys. Chem. C* **2014**, accepted, DOI: 10.1021/jp501602a.

IV. How the second-order nonlinear optical response of the collagen triple helix appears: a theoretical investigation

In this Chapter, the origin and the nature of the first hyperpolarizability of collagen has been unraveled by performing first principles calculations on the $[(\text{PPG10})_{10}]_3 = \text{PPG10}$ compound, a molecular model for the collagen triple helix structure, and on its building blocks. The first hyperpolarizability of the triple helix originates from the amide groups. Owing to the rigidity of the structure and of the proline units, the β tensor components parallel to the helical axis add to each other whereas the perpendicular components cancel each other, which result in a dipolar first hyperpolarizability and a depolarization ratio close to 9. The calculations have also shown that the resulting β values cannot be viewed as the simple sum of the amide group contributions when considering the triple helix. Indeed, for a given chain, the β per peptide bond increases linearly with the size of the chain whereas the β of the triple helix is smaller than three times the β of a single chain. The calculations, performed at different levels of approximation, demonstrated also the reliability of the ONIOM scheme when combining high and low layers described with different basis sets and the weak impact of electron correlation.

IV.A. Introduction

Collagen is one of the most present proteins in the animal reign and many genetically distinct collagen types exist.¹ Collagen is a three polypeptide α chain where the Gly-X-Y monomer is the repeated unit. Proline and 4-hydroxyproline frequently occupy the X and Y positions, respectively. The glycine units give the

flexibility to the chains to make three left-handed helices, which are held together by interchain hydrogen bonds.

Measurement of the second-order hyperpolarizability of the collagen triple helix by Hyper-Rayleigh scattering (HRS) experiment has been done by Deniset-Besseau *et al.*² who characterized the first hyperpolarizability (β) of collagen from rat-tail tendon as well as a short model of triple-helical peptides [(Pro-Pro-Gly)₁₀]₃ (PPG10). They concluded that the large collagen second-order nonlinear optical (NLO) response originates from the tight alignment of a large number of small and weakly efficient harmonophores, presumably the peptide bonds and not from a particular amino acid, resulting in an amplification of the nonlinear signal. Following this study, in 2010, Loison and Simon³ applied an additive model to theoretically determine β of PPG10. In their study, they showed that the additive model using *N*-methylacetamide as a building block leads to β results that agree qualitatively with HRS data from reference 2. In 2011, Tuer *et al.*⁴ approximated the β tensor of the collagen triple-helix models by taking the sum of the β tensors of the constitutive amino acids with the time-dependent Hartree-Fock method and accounting for the peptide bond contributions. They applied the unit sphere representation for a three-dimensional visualization of the β tensor⁵. Recently, Alparone⁶ investigated the α -helix and β -strand of oligoglycines with *ab initio* methods and showed that for the longest chains, the response of the β -strand is one order of magnitude larger than for the α -helix. Moreover, Wei *et al.*⁷ used the Summation-Over-States (SOS) method combined with Configuration Interaction Singles calculations to highlight the impact of a single amino acid mutation in α -helical polyalanine. They showed that side chains containing benzene, phenol, and indole have large contributions to the polarizability and first hyperpolarizability. However, analyzing the amplitudes and variations of the polarizability raises doubt on the convergence of their SOS expressions, which only includes 50 excited states for a system containing 17 amino acids. After having modeled the β of several fluorescent proteins with the ONIOM

(our own n-layered integrated molecular orbital and molecular mechanics) method where the surrounding of the chromophore as well as the electron correlation effects are accounted for, here we study the first hyperpolarizability of PPG10 with quantum chemical methods. The purpose of this study is i) to further unravel the structure-property relationships associated with the β response, and ii) to validate a first principles method able to accurately describe this large system. This is carried out by using the ONIOM method,⁸ which consists of partitioning the system into several layers described at different levels of theory, where the peptide backbone is treated with the high level method of calculation and the chain ends as well as the remainders of proline rings with the low level of calculation.

IV.B. Materials and methods

IV.B.1. Theoretical and computational aspects

In order to investigate the collagen-like triple helix polypeptide $[(\text{Pro-Pro-Gly})_{10}]_3$ (PPG10), the structure was taken from x-ray diffraction data⁹. After the addition of the hydrogen atoms, the PPG10 geometry (fig.1) was optimized with the semi-empirical PM6 method. Note that the geometry optimization procedure leads to a negligible chain rebuilding as evidenced by the variation of the distance between the chain extremities (taken as terminal N and C atoms) from 82.2 Å to 82.6 Å. To compare different levels of theory, a $[(\text{Pro-Pro-Gly})_3]_3$ (PPG3) fragment was cut from the large PPG10. Its geometry was optimized using the ONIOM B3LYP/6-311G**: PM6 method, where the chain of peptide bonds was treated with the high level of calculation and the remainder with the low level of calculation (fig.2). Then, the PPG3 structure was modified to build the $[(\text{Gly-Gly-Gly})_3]_3$ (GGG3) structure, which was optimized at the B3LYP/6-311G** level of calculation, as well as, the Pro-Pro-Gly monomer.

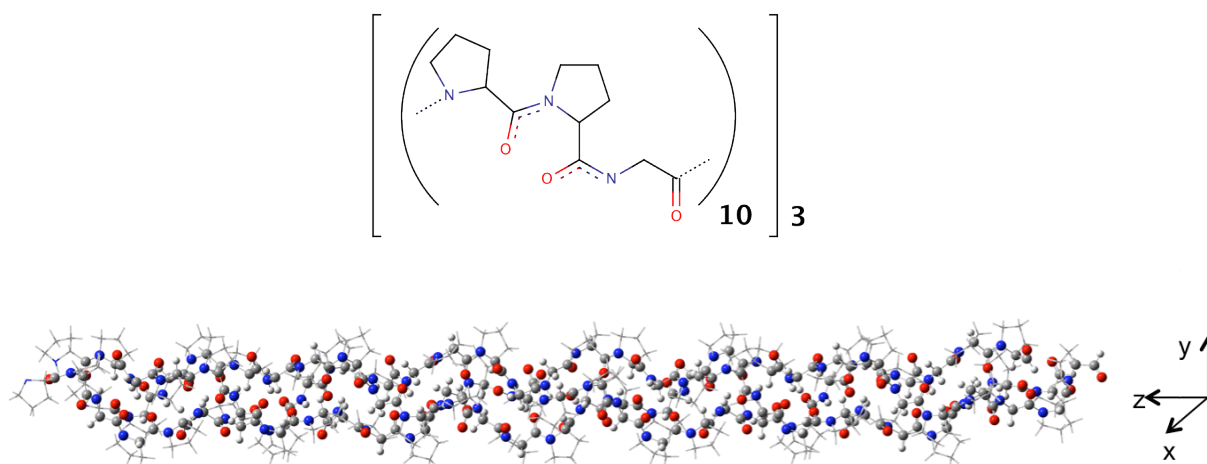


FIG. 1. Sketches of PPG10 as well as its optimized geometry, where the ONIOM high layer is represented with balls and sticks, and the low layer with sticks.

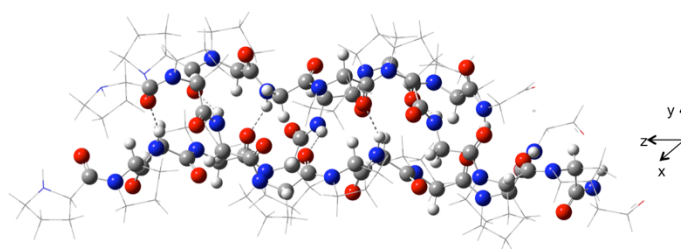


FIG. 2. Optimized geometry of $[(\text{Pro-Pro-Gly})_3]_3$ where the ONIOM high layer is represented with balls and sticks, and the low layer with sticks.

In order to characterize β of PPG10, which for today's computers is still a very large system (1035 atoms), we used the ONIOM scheme in combination with the Time-Dependent Hartree-Fock (TDHF) method to obtain the dynamic β and with the Coupled-Perturbed Hartree-Fock (CPHF) method for the static one. In the two-layer ONIOM approximation, the system is partitioned into high (the peptide backbone) and low (the lateral groups) layers (Figs. 1-2), which are investigated by a high/low level of theory, respectively. For each sub-calculation the static and dynamic β tensors are determined using the TDHF or CPHF schemes, respectively. We employed the Polarizable Continuum Model within the integral equation formalism (IEFPCM) to take into account the solvent (water) effects.¹⁰ In the ONIOM method, for each sub-calculation, the cavity of the real system was used¹¹.

Considering previous investigations⁶, the HF/6-31+G(d) method is initially chosen as the high level of theory for calculating the β tensors while in a second step electron correlation effects were assessed in order to provide more quantitative estimates. The 6-31+G* basis set is a good compromise between accuracy and computational resources whereas the HF method is known to underestimate the first hyperpolarizability of push-pull π -conjugated systems owing to missing electron correlation effects but generally it turns out to be suitable to provide trends among compounds. Moreover, a long-range corrected (LC) exchange-correlation functional, LC-BLYP¹² (with a range-separating parameter μ of 0.47), in combination with the time-dependent density functional theory (TDDFT) scheme was also employed to provide an estimate of the impact of electron correlation, while describing exactly the long-range behavior of exchange. Indeed, this method is reliable for describing the changes of β upon enlarging the π -conjugated linker of push-pull π -conjugated systems; though it also underestimates systematically the CCSD(T) [coupled cluster method including the singles and doubles and a perturbative estimate of the triple excitations] reference values. As it was impossible to perform full TDHF/CPHF calculations with the 6-31+G(d) basis set on PPG10, the validation of the ONIOM method was done on the small PPG3 system by comparing it to TDHF and TDDFT reference calculations. For the low level of theory, the HF/6-31G(d) method was selected. In order to further investigate electron correlation effects on the first hyperpolarizabilities of the PPG monomer, additional Møller-Plesset second-order perturbation theory (MP2) calculations were performed within the finite field (FF) procedure¹³. To remove higher-order contaminations in the finite differentiation procedure and achieve good accuracy (1-10 a.u.) on the numerical derivatives, the Romberg method was applied. The field-dependent energies were calculated for field amplitudes of 0.0004×2^n with $n = 0-4$. In order to account for frequency dispersion at the MP2 level, we employed the multiplicative approximation¹⁴.

Experimentally, the first hyperpolarizabilities and their depolarization ratios (*DR*) were determined from hyper-Rayleigh scattering (HRS) measurements.¹⁵ The analysis of the β tensor also adopts the unit sphere representation.⁴⁻⁵

All reported β values are given in a.u. [1 a.u. of $\beta = 3.6213 \times 10^{-42} \text{ m}^4 \text{ V}^{-1} = 3.2064 \times 10^{-53} \text{ C}^3 \text{ m}^3 \text{ J}^{-2} = 8.639 \times 10^{-33} \text{ esu}$] within the T convention. All calculations were done with the Gaussian09 package¹⁶, except for the geometry optimization of PPG10, which was carried out with the MOPAC package¹⁷.

IV.B.2. Materials and experimental spectroscopic characterizations

In Ref. 2 the first hyperpolarizability of PPG10 was experimentally determined by frequency-resolved femtosecond HRS. To estimate the intrinsic β response, these experimental data were pretreated by ourselves to eliminate the resonance effects.¹⁸ The frequency dispersion factors, $F(\omega, \omega_{ge}, \gamma) = \beta_{zzz}(-2\omega; \omega, \omega) / \beta_{zzz}(0; 0, 0)$, were determined from successive refinements of the two-state approximation (TSA).¹⁹ $\hbar\omega_{ge}$ is the excitation energy of the dominant excited state and γ is the damping factor. Besides the conventional TSA, i) homogeneous damping was taken into account, ii) then, an inhomogeneous broadening based on the absorption spectrum, which implicitly contains information on the distribution of the transition frequencies, and finally iii) the later scheme was extended to account for the vibronic structure of the excited states.¹⁸ To apply these extrapolation schemes, the successive $F(\omega, \omega_{ge}, \gamma)$ functions had to be fitted to the UV absorption spectrum. As a matter of fact, UV/vis absorption spectra were recorded.

The absorption spectrum of PPG10 was obtained using an Uvikon_{XS} spectrophotometer from BIO-TEK instruments. Samples were prepared by mixing (Pro-Pro-Gly)₁₀ (Peptides International, OPG-4006) solutions with acetic acid 0.5 mM in order to obtain a final (Pro-Pro-Gly)₁₀ concentration of 0.6 mg.ml⁻¹, which self-organizes to form triple helices. The solutions were then gently stirred for 24 h at 4°C. A spectrum of the buffer only was recorded as a blank. The difference between

the PPG10 and blank solutions gives the experimental spectrum (figure 3). The $F(\omega, \omega_{ge}, \gamma)$ function was then fitted to get the parameters listed in Table 1.

Dynamic light scattering (DLS) experiments were performed using a DynaPro NanoStar DLS/SLS recorder from Wyatt Technology Corporation. These results demonstrate that there is no oligomerization of the PPG10 triple helices in solution and therefore that the β_{HRS} measurements from Ref. 2 refer to PPG10 and not some of its aggregates.

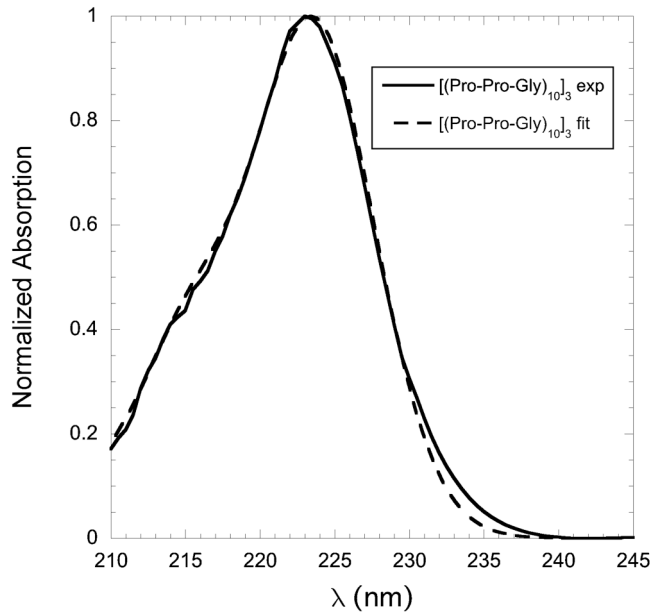


FIG. 3. Normalized absorption spectrum of PPG10. The full line corresponds to the experimental data, while the dashed line the fitted spectrum.

TABLE 1. Parameters fitting the UV/Vis absorption spectrum. See Ref. 31 for more details.

$\hbar\omega_{ge}$ (eV); λ_{eg} (nm)	5.54; 224
$\gamma_{\text{homogeneous}}$ (eV)	0.181
$\gamma_{\text{inhomogeneous}}$ (cm^{-1})	100
S, Huang Rhys factor	0.40
ω_{vib} (cm^{-1})	1700
γ_{vib} (cm^{-1})	1100

IV.C. Results and discussion

IV.C.1. Assessment of the ONIOM method and of electron correlation effects

To reduce the computational costs for calculating the first hyperpolarizability of PPG10, the ONIOM scheme was adopted where the low and high layers are treated with different basis sets, 6-31G(d) for the low level and 6-31+G(d) for the high one. In addition, the low level calculations are carried out using the HF method. This approach was first validated by performing calculations on PPG3, for which full 6-31+G(d) calculations could be carried out (Table 2). ONIOM results obtained at the TDHF and TDDFT/LC-BLYP levels are listed in Table 3, demonstrating that the use of a smaller basis set for the lateral groups of the peptide chains has only a minor effect on the β_{HRS} response. Indeed, the CPHF (CPKS/LC-BLYP) static reference values are at most overestimated by 2.5% (6.5%) *in vacuo* or underestimated by 1.8% (2.8%) when accounting for the solvent whereas for the dynamic responses, the differences are smaller.

TABLE 2. β_{HRS} of $[(\text{Pro-Pro-Gly})_3]_3$ and $[(\text{Gly-Gly-Gly})_3]_3$ calculated at different levels of approximation (in a.u.).

	TDHF/CPHF 6-31+G(d)				TDDFT/LC-BLYP 6-31+G(d)			
	PPG3		GGG3		PPG3		GGG3	
	<i>in vacuo</i>	<i>in water</i> (IEFPCM)	<i>in vacuo</i>	<i>in water</i> (IEFPCM)	<i>in vacuo</i>	<i>in water</i> (IEFPCM)	<i>in vacuo</i>	<i>in water</i> (IEFPCM)
$\beta_{\text{HRS},\infty}$	517	1065	474	870	472	1101	461	948
$\beta_{\text{HRS},1900\text{nm}}$	539	724	492	661	500	721	484	697
$\beta_{\text{HRS},1064\text{nm}}$	591	793	534	718	570	818	541	779

TABLE 3. β_{HRS} of $[(\text{Pro-Pro-Gly})_3]_3$ as calculated with the ONIOM method (in a.u.).

	ONIOM		ONIOM	
	HF/6-31+G(d):HF/6-31G(d)		LC-BLYP/6-31+G(d):HF/6-31G(d)	
	<i>in vacuo</i>	<i>in water</i> (IEFPCM)	<i>in vacuo</i>	<i>in water</i> (IEFPCM)
$\beta_{\text{HRS},\infty}$	530	1046	503	1071
$\beta_{\text{HRS},1900\text{nm}}$	550	731	527	734
$\beta_{\text{HRS},1064\text{nm}}$	599	796	588	816

As shown for push-pull π -conjugated systems, small reference molecules for NLO,²⁰ the electron correlation issue is also briefly addressed here in view of comparison with the experimental data of Deniset-Besseau *et al.*². A first comparison between the HF and LC-BLYP data (Table 2) demonstrates that at most, the HF level underestimates the LC-BLYP β_{HRS} values by 8.2% (when accounting for solvent effects) or it overestimates them by at most 9.5% (*in vacuo*). These effects are very small in comparison to those found for other systems.²⁰ To further substantiate these results, additional calculations have been performed on a unique Pro-Pro-Gly chain, which allows comparisons with the MP2 method (Table 4). The MP2 results are typically 4% larger than the HF one, confirming the small electron correlation effects for β_{HRS} of collagen chains. The HF vs. LC-BLYP results are consistent with those of Table 2. These results demonstrate that qualitative and semi-quantitative estimates of β_{HRS} of collagen chains can already be obtained at the HF level.

Considering a larger basis set for PPG gives similar results with an increase of less than 8% for aug-cc-pVDZ with respect to 6-31+G(d) and of 6% for aug-cc-pVTZ (Table 5).

TABLE 4. β_{HRS} of Pro-Pro-Gly calculated at different levels of theory with the 6-31+G(d) basis set (in a.u.).

<i>in vacuo</i>	TDHF/CPHF	TDDFT LC-BLYP	MP2
$\beta_{\text{HRS},\infty}$	54	48	56
$\beta_{\text{HRS},1900\text{nm}}$	55	50	57
$\beta_{\text{HRS},1064\text{nm}}$	58	54	60

TABLE 5. β_{HRS} of Pro-Pro-Gly calculated at the TDHF/CPHF level of theory with different basis sets (in a.u.).

<i>in vacuo</i>	6-31+G(d)	aug-cc-pVDZ	aug-cc-pVTZ
$\beta_{\text{HRS},\infty}$	54	58	56
$\beta_{\text{HRS},1900\text{nm}}$	55	59	58
$\beta_{\text{HRS},1064\text{nm}}$	58	62	61

IV.C.2. Structure-property relationships

In the previous section, the choice of high layer to be the chain of peptide bonds was based on the hypothesis that it is responsible for most of the HRS response because the C(=O)-N units are the only π -conjugated parts of PPG10. This is now substantiated by first comparing β_{HRS} of PPG3 to β_{HRS} of GGG3, where only the chains of peptide bonds remain (Table 2). For all the dynamic β_{HRS} , the differences between GGG3 and PPG3 are smaller than 10% whereas they could be slightly larger in the static limit when accounting for the solvent effects. This demonstrates that the β_{HRS} response originates from the peptide backbone. Nevertheless, PPG3 and GGG3 differ by their rigidity. The GGG3 chain is more flexible, which allows the formation of 4 additional H-bonds whereas the presence of proline units in PPG3 constraints its conformation.

To further confirm this hypothesis of the peptide backbone origin of β_{HRS} , the response of one chain of PPG3 has been compared to the response of its different building blocks (Table 6). These calculations were performed by keeping the same geometry as in the triple helix. The corresponding unit sphere representations, β (blue) and μ (red) vectors are given in Figure 4. For a single amino acid, the response originates from the asymmetric electron distribution around the heteroatoms and the effective SHG dipole vector is oriented almost parallel to the C=O bond. These responses are largely dipolar, with DR larger than 6.5 or 7.5. By linking two amino acids, a peptide bond is created and the amide moiety brings a charge-transfer contribution to the total first hyperpolarizability. Although the compounds are twice larger than the monomer, the β_{HRS} value remains similar or even decreases. Moreover, the shape of the tensor becomes more complex. This is attributed to the amide function contribution combined with those of C=O group, which are differently oriented. This is also evidenced by the smaller depolarization ratios. Then, going to the tripeptide, the β_{HRS} value increases, which is attributed to the mostly-additive contributions of the two amide bonds. By increasing the size of the

system, the contribution in the other directions becomes more and more negligible leading to an overall response along the backbone. This is evidenced by the increase of the depolarization ratio from the dipeptide to the nonapeptide with a DR of 8.27 for (Pro-Pro-Gly)₃. In parallel, the nonlinear anisotropy parameter $\rho = |\beta_{J=3}|/|\beta_{J=1}|$, which gives the ratio of the octupolar versus dipolar contributions to β ,²⁰ decreases, leading to a mostly dipolar response. In figure 5, the evolution of β_{HRS} as a function of the number of peptide bonds is represented. This shows a mostly linear evolution with the size of the chain.

Thus, in PPG10, the conformation of the triple helix orients the amide π -conjugated segments around the helical axis, so that, when the chain grows, the radial component vanishes and the axial component becomes dominant. This differs from pyridine-pyrimidine and hydrazine-pyrimidine oligomers where the radial component is dominant²¹. To some extent, this constrained conformation can be associated with the rigidity of the proline units. To support this hypothesis, the conformations of PPG3 and GGG3 were compared on the basis of the orientation of key bonds with respect to the helical axis (or z-axis). The selected angles are $\alpha = \widehat{\text{OCz}}$, $\beta = \widehat{\text{NCz}}$, and $\gamma = \widehat{\text{OCN}}$. Note that if the OCN plane formed by the amide function contains the z-axis, γ is equal to $\alpha + \beta$. Data in Tables 7 and 8 show that the successive α and β angles of the three chains of PPG3 display the same pattern, demonstrating more regularity and rigidity than in the case of the 3 chains of GGG3 for which the angle values are more random. This regularity is at the origin of the strong dipolar nature of β in PPG3 and PPG10 and it also explains its larger β_{HRS} value than in GGG3.

TABLE 6. β_{HRS} , DR (depolarization ratio), and ρ (nonlinear anisotropy factor) for (Pro-Pro-Gly)₃ as well as of its different building blocks (in a.u.).

	$\beta_{\text{HRS}, \infty}$ CPHF 6-31+G(d)	DR	ρ
Pro1	41	7.67	0.35
Pro2	46	6.57	0.53
Gly3	40	7.72	0.35
Pro1-Pro2	40	1.68	4.92
Pro2-Gly3	33	3.73	1.17
Pro1-Pro2-Gly3	58	3.21	1.41
Pro1-Pro2-Gly3-Pro4	95	3.91	1.11
Pro1-Pro2-Gly3-Pro4-Pro5	111	4.21	1.01
Pro1-Pro2-Gly3-Pro4-Pro5-Gly6	127	6.32	0.57
(Pro-Pro-Gly) ₃	216	8.27	0.25

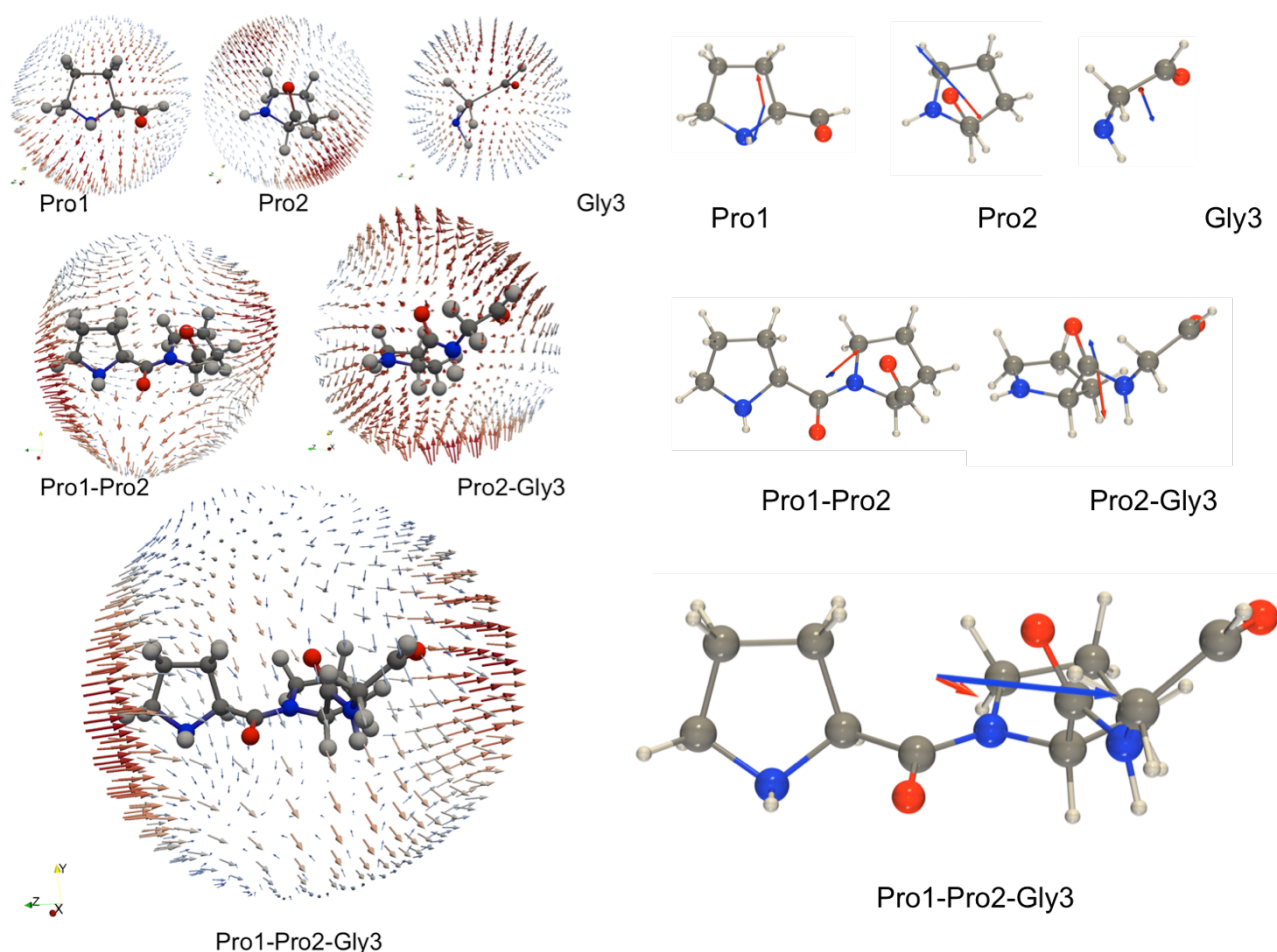


FIG. 4. Unit sphere representations of the first hyperpolarizability tensor, β (blue) and μ (red) vectors obtained at the CPHF/6-31G(d) level of calculation for the (Pro-Pro-Gly) chain and its building blocks. The lengths of the μ and β vectors are proportional to their norms and are consistent for all the systems.

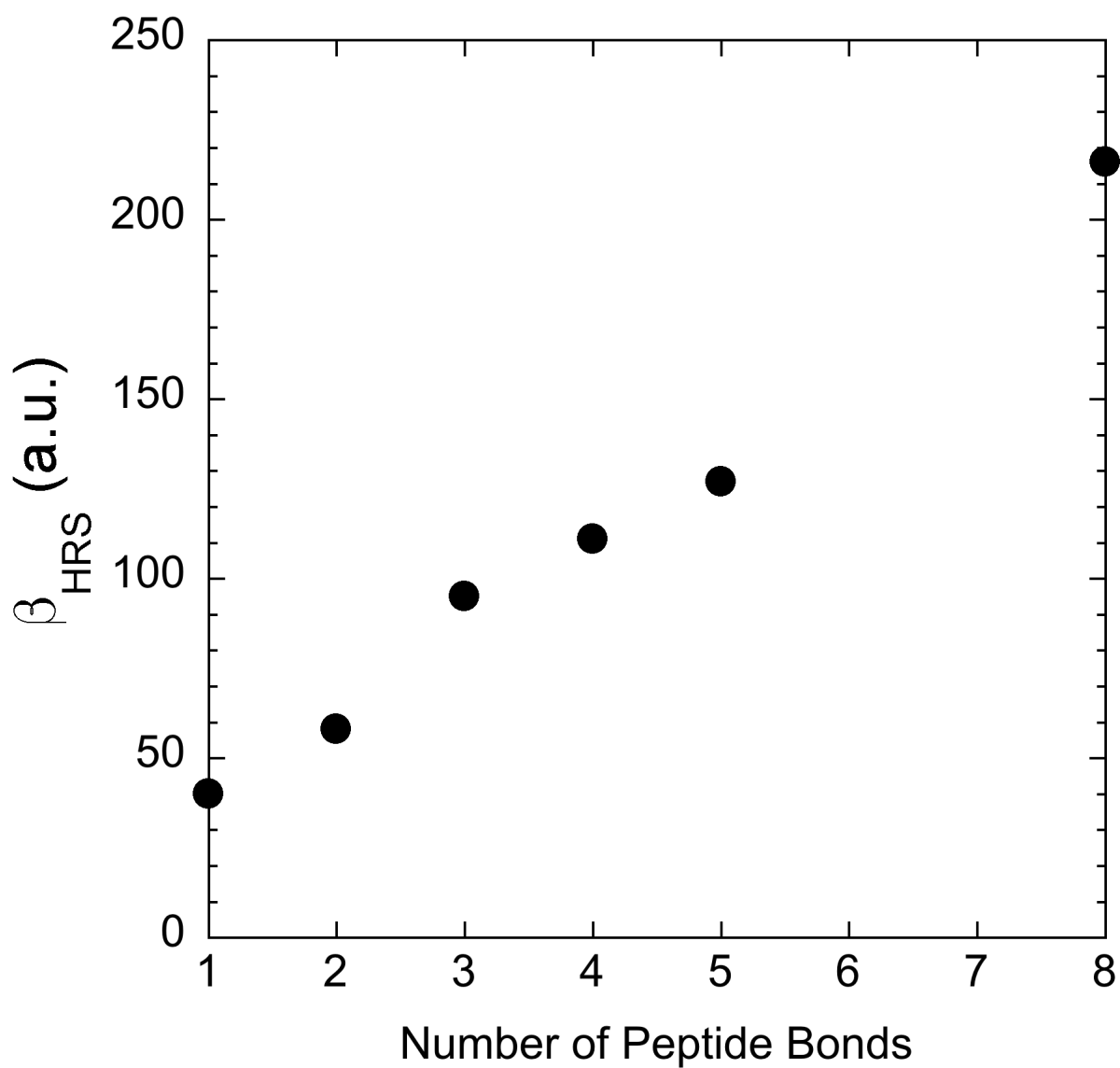


FIG. 5. β_{HRS} values as a function of the number of peptide bonds are represented for $(\text{PPG})_3$.

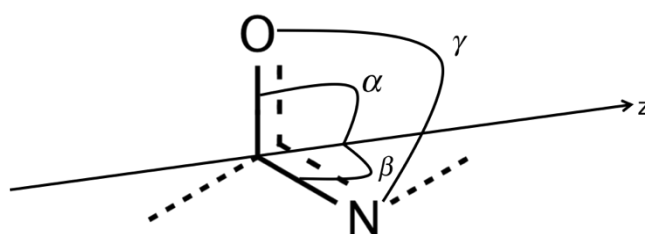


FIG. 6. Representation of typical angles of each amide groups with respect to the z-axis.

TABLE 7. Representative angles (in °) that each amide group of PPG3 forms with the z-axis. The description of the angles can be found in figure 6.

PPG3	α	β	γ	$\alpha + \beta$
1st chain				
Pro1-Pro2	109	19	123	128
Pro2-Gly3	102	56	124	158
Gly3-Pro4	63	62	120	125
Pro4-Pro5	94	34	121	128
Pro5-Gly6	103	60	124	163
Gly6-Pro7	56	69	121	125
Pro7-Pro8	95	29	121	124
Pro8-Gly9	106	59	125	165
2nd chain				
Pro1-Pro2	84	39	122	123
Pro2-Gly3	112	49	124	161
Gly3-Pro4	67	61	121	128
Pro4-Pro5	93	35	121	128
Pro5-Gly6	103	53	124	156
Gly6-Pro7	65	64	121	129
Pro7-Pro8	89	39	121	128
Pro8-Gly9	82	73	125	155
3rd chain				
Pro1-Pro2	85	38	122	123
Pro2-Gly3	112	49	124	161
Gly3-Pro4	65	69	121	134
Pro4-Pro5	87	38	121	125
Pro5-Gly6	114	47	124	161
Gly6-Pro7	64	63	121	127
Pro7-Pro8	95	30	120	125
Pro8-Gly9	118	44	124	162

TABLE 8. Representative angles (in °) that each amide group of GGG3 forms with the z-axis. The description of the angles can be found in figure 6.

GGG3	α	β	γ	$\alpha + \beta$
1st chain				
Gly1-Gly2	65	63	126	128
Gly2-Gly3	113	14	125	127
Gly3-Gly4	92	32	124	124
Gly4-Gly5	97	47	123	144
Gly5-Gly6	106	47	124	153
Gly6-Gly7	81	47	122	128
Gly7-Gly8	91	51	123	142
Gly8-Gly9	67	76	124	143
2nd chain				
Gly1-Gly2	41	90	126	131
Gly2-Gly3	121	12	124	133
Gly3-Gly4	104	29	124	133
Gly4-Gly5	92	33	123	125
Gly5-Gly6	103	51	124	154
Gly6-Gly7	78	57	122	135
Gly7-Gly8	81	42	122	123
Gly8-Gly9	132	39	125	171
3rd chain				
Gly1-Gly2	42	83	126	125
Gly2-Gly3	131	-12	123	143
Gly3-Gly4	70	57	124	127
Gly4-Gly5	97	32	123	129
Gly5-Gly6	104	49	125	153
Gly6-Gly7	70	53	123	123
Gly7-Gly8	64	94	124	158
Gly8-Gly9	102	61	124	163

Consequently for PPG10, β_{zzz} is the major β tensor component. Nevertheless, two other tensor components $\beta_{xxz} = \beta_{xzx} \sim \beta_{zxx}$ and $\beta_{yyz} = \beta_{yzy} \sim \beta_{zyy}$ are not negligible though more than 2 or 3 times smaller than β_{zzz} , respectively. Consistent with this tensor analysis, the depolarization ratio, which gives information on the shape of the NLOphore, amounts to 8.8 [ONIOM HF/6-31+G(d):HF/6-31G(d), *in vacuo*]. Moreover, the nonlinear anisotropy parameter ρ is as low as 0.113 for PPG10. Thus the β response is dominantly dipolar.

In order to unravel the impact of the triple helix, starting from the geometry of PPG3, the dipole moments and first hyperpolarizabilities have been calculated for $[(PPG)_3]_2$ and $(PPG)_3$, which are obtained by removing one or two chains of the $[(PPG)_3]_3 = PPG3$. Going from one to two chains, the β_{HRS} response almost doubles but saturation appears when adding the third chain (Table 9). Similar effects are observed for the $\vec{\beta}$ and $\vec{\mu}$ norms. Figure 7 shows how the $\vec{\beta}$ and $\vec{\mu}$ vectors are closely parallel to the helical axis and how small is the angle between them, in particular for the triple helix. The dipolar behavior of the β response and its orientation are further confirmed in Figure 8 with the unit sphere representation of the β tensor of PPG10, highlighting that the maximum SHG response is obtained for an incident electric field parallel to the triple helix axis. See also Figure 9 for the corresponding unit sphere representations for $(PPG)_3$, $[(PPG)_3]_2$, and $[(PPG)_3]_3 = PPG3$, associated with the results in Table 9, where the orientation of the response is similar but the intensity is increasing. The norm of the $\vec{\beta}$ obtained by the summation of $\vec{\beta}$ of each chain of PPG3 (1565 a.u.) is 11.5% higher than the norm of the $\vec{\beta}$ of PPG3 (1403 a.u.). This is the consequence of the interaction between chains.

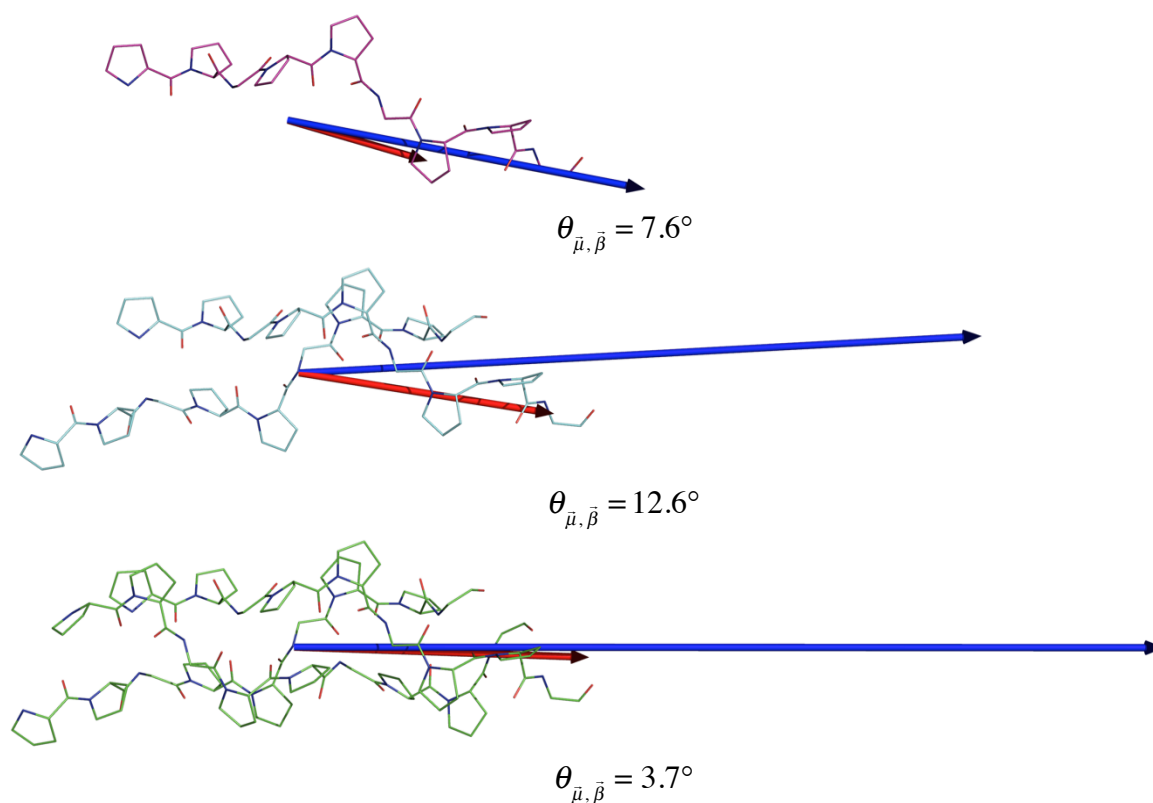


FIG. 7. β (blue) and μ (red) vectors for one, two, and three helices of PPG3. These vectors have been determined at the CPHF level of calculation. The lengths of the μ and β vectors are proportional to their norms and are consistent for the three systems.

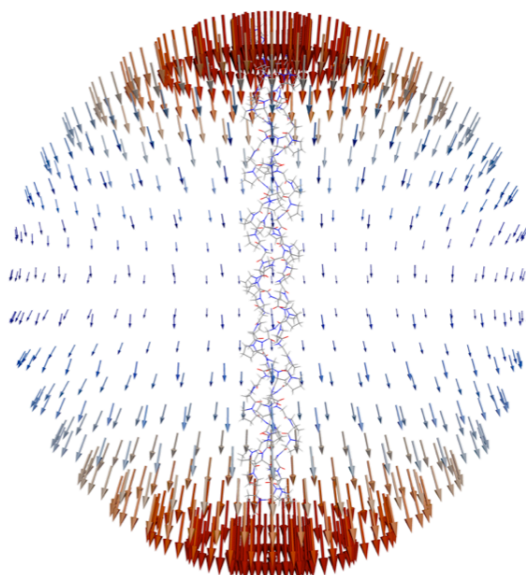


FIG. 8. Unit sphere representation of the first hyperpolarizability tensor of PPG10 obtained with the ONIOM HF/6-31+G(d):HF/6-31G(d) method at 1900nm in water solvent (IEFPCM)

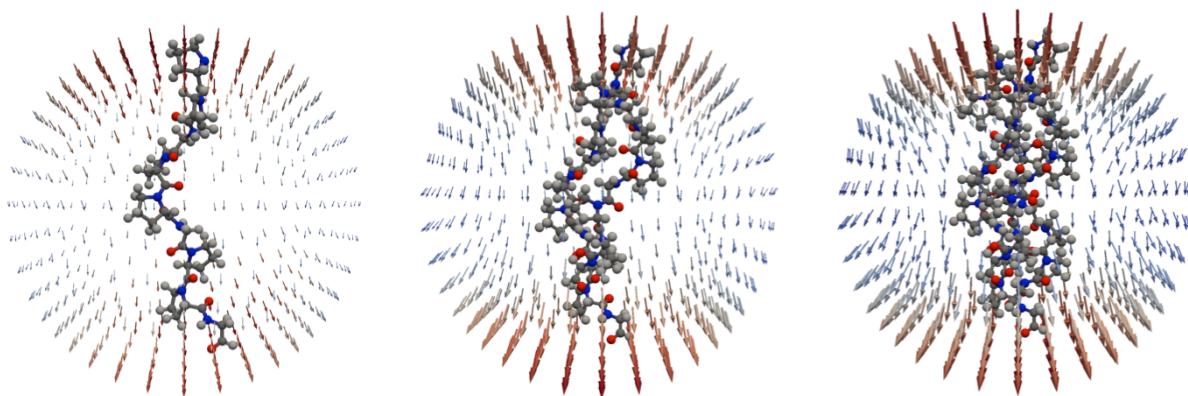


FIG. 9. Unit sphere representations of the first hyperpolarizability tensor of $(\text{PPG})_3$ (left), $[(\text{PPG})_3]_2$ (middle), and $[(\text{PPG})_3]_3$ (right) obtained at the CPHF/6-31+G(d) level in vacuo.

TABLE 9. Static β_{HRS} and related quantities for PPG3 and its derivatives obtained after removing one $[(\text{PPG})_3]_2$ or two $(\text{PPG})_3$ chains as determined at the CPHF/6-31+G(d) level of approximation (in vacuo)

	β_{HRS} (a.u.)	DR	$\ \vec{\beta}\ $ (a.u.)	$\ \vec{\mu}\ $ (a.u.)	$\theta_{\vec{\mu}, \vec{\beta}}$ (degrees)
$[(\text{PPG})_3]_3$	517	8.47	1403	15.43	7.6
$[(\text{PPG})_3]_2$	410	8.29	1109	13.54	12.6
$(\text{PPG})_3$	216	8.27	582	7.30	3.7

IV.C.3. $[(\text{Pro-Pro-Gly})_{10}]_3$: results and comparison with experiment

The PPG10 β_{HRS} values show an increase by more than a 10/3 factor with respect to PPG3 (Table 10). Considering the effects of the solvent, the $\beta_{\text{HRS}}(\text{PPG10})/\beta_{\text{HRS}}(\text{PPG3})$ ratio – evaluated at the ONIOM level – amounts to 4.1 in the static limit (HF and LC-BLYP) and to 4.7-5.1 at $\lambda = 1900$ and 1064 nm. These ratios are slightly larger *in vacuo*. The β_{HRS} is exalted when increasing chain length. This is the result of typical intrachain electrostatic (hyper)polarization effects, which reinforce the electric field parallel to the chain and reduce its components perpendicular to it. Frequency dispersions of β_{HRS} of PPG10 are sketched in Figure 10, exhibiting similar behavior with and without solvent effects, as well as at both the HF and LC-BLYP levels of approximation. Note that at $\lambda = 790$ nm, the difference between HF and LC-BLYP amounts now to about 15%.

TABLE 10. frequency-dispersion of the HRS first hyperpolarizabilities (and depolarization ratios) of [(Pro-Pro-Gly)₁₀]₃ calculated at different levels of theory (in a.u.).

	ONIOM		ONIOM	
	HF/6-31+G(d):HF/6-31G(d)		LC-BLYP/6-31+G(d):HF/6-31G(d)	
	<i>in vacuo</i>	<i>in water</i> (IEFPCM)	<i>in vacuo</i>	<i>in water</i> (IEFPCM)
$\beta_{\text{HRS},\infty}$	2703 (8.84)	4232 (8.12)	2785 (8.85)	4480 (8.08)
$\beta_{\text{HRS},1900\text{nm}}$	2816 (8.86)	3483 (9.85)	2937 (8.88)	3521 (9.90)
$\beta_{\text{HRS},1064\text{nm}}$	3091 (8.91)	3763 (9.11)	3321 (8.92)	4140 (9.10)
$\beta_{\text{HRS},790\text{nm}}$	3472 (8.99)	4230 (9.24)	3881 (8.95)	4558 (9.28)
$\beta_{\text{HRS},694\text{nm}}$	3765 (9.04)	4861 (9.17)	4333 (8.96)	5643 (9.17)

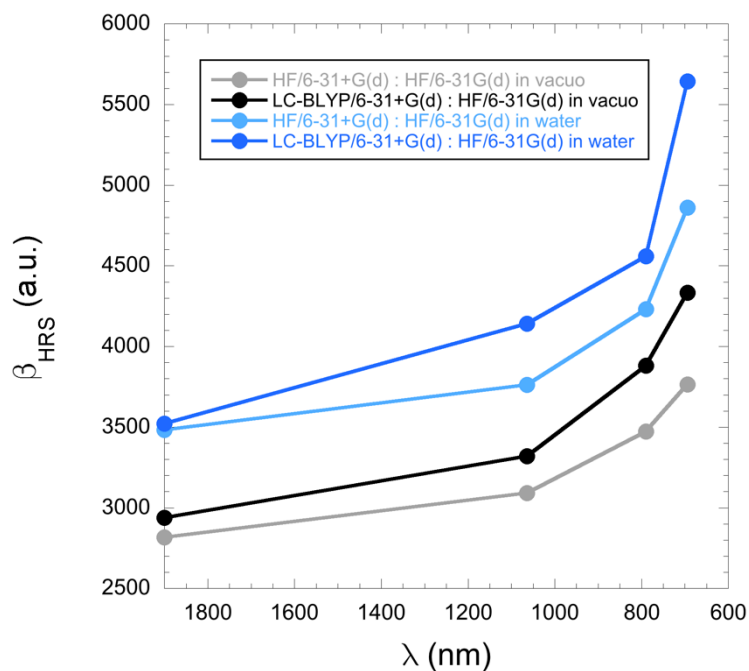


FIG. 10. Frequency dispersion of β_{HRS} calculated with the ONIOM HF/6-31+G(d):HF/6-31G(d) and LC-BLYP/6-31+G(d):HF/6-31G(d) methods.

Experimentally, for $\lambda = 790$ nm, Ref. 2 reports a value of 11460 ± 690 a.u. for the β_{HRS} of PPG10 and a DR of 8.7 (it is assumed that Ref. 2 adopts the T convention). Although, the wavelength of measurement ($790/2 = 395$ nm) is quite far from the first one-photon absorption edge (235 nm), the experimental dynamic response was extrapolated to the static limit by adopting different refinements of the TSA (Table 11). The values obtained with the three TSA levels of refinement are consistent and of the order of 7200 a.u.

TABLE 11. Measured and extrapolated values of β_{HRS} of [(Pro-Pro-Gly)₁₀]₃ (in a.u.).

$\beta_{\text{HRS},790}$	11460 $\pm$ 690
$\beta_{\text{HRS},\infty}$ TSA	7180 $\pm$ 440
$\beta_{\text{HRS},\infty}$ TSA (homog. damping)	7100 $\pm$ 430
$\beta_{\text{HRS},\infty}$ TSA (homog. + inhomog. damping, vibr.)	7420 $\pm$ 450

The experimentally determined DR value of 8.7 is in good agreement with the calculated HF values, from the static DR = 8.8 *in vacuo* to the dynamic DR = 9.0 at 790 nm. For β_{HRS} , the ONIOM/PCM theoretical values underestimate the experimental results by a factor of 2 or more, both for the static and dynamic quantities. Two remarks should however be made because such comparisons are not straightforward. First, water is a dipolar solvent and therefore its static dielectric constant (78.355) is substantially larger than its optical dielectric constant (1.77 for visible light). Since larger dielectric constants lead to an enhancement of β , it is therefore advantageous to compare extrapolated experimental $\beta_{\text{HRS},\infty}$ with β values calculated at optical frequencies far from electronic resonances, where the frequency dispersion is small if not negligible, in our case 1900nm. Thus at 1900nm, β_{HRS} is equal to 3483 a.u. at the HF:HF level and 3521 a.u. at the LC-BLYP:HF level, in comparison to the experimental value of 7420 $\pm$ 450 a.u. . This corresponds to an underestimation of 53% whereas at 790 nm, the underestimation reaches 60-63%. This difference (about 10%) between the underestimates of the static and dynamic experimental values has to be related to the performance of these methods to predict the excitation energy of the states that dominate the UV/vis absorption spectrum (Figure 3). This first excitation energy amounts to 176 nm at the TDDFT/LC-BLYP level (for one chain of PPG3) in comparison to the experimental 224 nm value and this overestimation of the excitation energy explains the smaller frequency dispersion.

Secondly, the experimental value reported in Ref. 2 was determined by considering water as internal reference for the β_{HRS} measurements and a value of 65 a.u. from Ref. 22 was employed for $\beta_{\text{HRS},800}$. Note that other β_{HRS} values for water have been reported in the literature, $\beta_{\text{HRS},800} = 58$ a.u.,²³ and finally $\beta_{\text{HRS},800} = 10$ a.u.²⁴. It is therefore difficult to go further in the comparison because the differences between these reference values are quite large, which is a typical difficulty when comparing theoretical/calculated and experimental/measured NLO responses.²⁵

Thirdly, we ask to our colleagues of the KULeuven (Prof. Th. Verbiest and Dr. S. Van Cleuvenbergen) to redo the measurements in the same conditions but their results were not reproducible.

IV.D. Conclusions

First principles calculations have been performed on the PPG10 compound, a molecular model for the collagen triple helix structure, and its building blocks to unravel the origin and the nature of the first hyperpolarizability of collagen. In agreement with previous investigations, β of the triple helix originates from the amide groups. Their β tensor components parallel to the helical axis add to each other, whereas the perpendicular components cancel each other, which result in a dipolar first hyperpolarizability and a depolarization ratio close to 9. On the other hand, the calculations have also shown that the resulting β values cannot be viewed as the simple sum of the amide group contributions when considering the triple helix. Indeed, for a given chain, the β per peptide bond increases linearly with the size of the chain whereas the β of the triple helix is smaller than three times the β of a single chain. These calculations have adopted the ONIOM scheme, of which the reliability for calculating β has been confirmed when combining high and low layers described with different basis sets. Moreover, comparisons between levels of approximation have demonstrated the weak impact of electron correlation on the first hyperpolarizabilities of collagen building blocks, contrary to many traditional

compounds for nonlinear optics, either push-pull π -conjugated systems or reference molecules like CCl_4 . Comparison with experiment was hampered by the lack of undisputable reference β value for H_2O .

REFERENCES

1. (a) Hulmes, D. J., *J Struct Biol* **2002**, *137* (1-2), 2-10; (b) Kadler, K. E.; Baldock, C.; Bella, J.; Boot-Handford, R. P., *J Cell Sci* **2007**, *120* (Pt 12), 1955-8.
2. Deniset-Besseau, A.; Duboisset, J.; Benichou, E.; Hache, F.; Brevet, P. F.; Schanne-Klein, M. C., *J Phys Chem B* **2009**, *113* (40), 13437-45.
3. Loison, C.; Simon, D., *J Phys Chem A* **2010**, *114* (29), 7769-79.
4. Tuer, A. E.; Krouglov, S.; Prent, N.; Cisek, R.; Sandkuijl, D.; Yasufuku, K.; Wilson, B. C.; Barzda, V., *J. Phys. Chem. B* **2011**, *115* (44), 12759-69.
5. Tuer, A.; Krouglov, S.; Cisek, R.; Tokarz, D.; Barzda, V., *J. Comput. Chem.* **2011**, *32* (6), 1128-34.
6. Alparone, A., *Phys. Chem. Chem. Phys.* **2013**, *15* (31), 12958-12962.
7. Wei, J.; Wang, J.-Y.; Zhang, M.-Y.; Chai, G.-L.; Lin, C.-S.; Cheng, W.-D., *Chemical Physics Letters* **2013**, *556* (0), 260-265.
8. Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J., *THEOCHEM* **1999**, *461-462* (0), 1-21.
9. Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A., *Protein Sci* **2002**, *11* (2), 262-70.
10. (a) Tomasi, J.; Persico, M., *Chem. Rev.* **1994**, *94* (7), 2027-2094; (b) Tomasi, J.; Mennucci, B.; Cammi, R., *Chem. Rev.* **2005**, *105* (8), 2999-3094.
11. Vreven, T.; Mennucci, B.; da Silva, C. O.; Morokuma, K.; Tomasi, J., *J. Chem. Phys.* **2001**, *115* (1), 62-72.
12. (a) Yanai, T.; Tew, D. P.; Handy, N. C., *Chem. Phys. Lett.* **2004**, *393* (1-3), 51-57; (b) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K., *J. Chem. Phys.* **2004**, *120* (18), 8425-8433.
13. Cohen, H. D.; Roothaan, C. C. J., *J. Chem. Phys.* **1965**, *43*, S34-S38, discussion S39.
14. Sekino, H.; Bartlett, R. J., *J. Chem. Phys.* **1991**, *94* (5), 3665-3669.

15. Verbiest, T.; Clays, K.; Rodriguez, V., *Second-order nonlinear optical characterization techniques: An introduction*. CRC Press: New-York, 2009.
16. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., *Gaussian 09*, revision a.01. Wallingford CT, 2009.
17. Stewart, J. *Mopac2009*, Stewart Computational Chemistry: Colorado Springs, CO, USA, 2009.
18. (a) Campo, J.; Wenseleers, W.; Goovaerts, E.; Szablewski, M.; Cross, G. H., *J. Phys. Chem. C* **2008**, *112* (1), 287-296; (b) Mancois, F.; Pozzo, J.-L.; Pan, J.; Adamietz, F.; Rodriguez, V.; Ducasse, L.; Castet, F.; Plaquet, A.; Champagne, B., *Chem. Eur. J.* **2009**, *15*, 2560-2571.
19. Oudar, J. L.; Chemla, D. S., *J. Chem. Phys.* **1977**, *66* (6), 2664-2668.
20. Castet, F.; Bogdan, E.; Plaquet, A.; Ducasse, L.; Champagne, B.; Rodriguez, V., *J. Chem. Phys.* **2012**, *136* (2), 024506.
21. Méreau, R. I.; Castet, F. d. r.; Botek, E.; Champagne, B., *J. Phys. Chem. A* **2009**, *113* (24), 6552-6554.

22. Vance, F. W.; Lemon, B. I.; Hupp, J. T., *J. Phys. Chem. B* **1998**, *102* (50), 10091-10093.
23. Clays, K.; Olbrechts, G.; Munters, T.; Persoons, A.; Kim, O.-K.; Choi, L.-S., *Chem. Phys. Lett.* **1998**, *293* (5–6), 337-342.
24. (a) Campo, J.; Desmet, F.; Wenseleers, W.; Goovaerts, E., *Opt. Express* **2009**, *17* (6), 4587-4604; (b) Duboisset, J.; Matar, G.; Russier-Antoine, I.; Benichou, E.; Bachelier, G.; Jonin, C.; Ficheux, D.; Besson, F.; Brevet, P. F., *J. Phys. Chem. B* **2010**, *114* (43), 13861-13865.
25. (a) Willetts, A.; Rice, J. E.; Burland, D. M.; Shelton, D. P., *J. Chem. Phys.* **1992**, *97* (10), 7590-7599; (b) Reis, H., *J Chem Phys* **2006**, *125* (1), 014506-9.

Electron Correlation Effects on the Hyperpolarizabilities of Model Systems

M. de Wergifosse; B. Champagne, *J. Chem. Phys.* **2011**, *134* (7), 074113-13.

M. de Wergifosse; F. Wautelet; B. Champagne; R. Kishi; K. Fukuda; H. Matsui; M. Nakano, *J. Phys. Chem. A* **2013**, *117*, 4709-4715.

V. Electron Correlation Effects on the Hyperpolarizabilities of Model Systems

In the first part of this Chapter, the first hyperpolarizability of representative push-pull π -conjugated compounds has been calculated at several levels of approximation to assess the effects of electron correlation. First, the 6-31+G(d) basis set has been shown to give the best balance between accuracy and computational resources for a polyene linker whereas for polyyne linker, the 6-31G(d) basis set is already an optimal choice. As a result of cancellations between higher order contributions, the MP2 method turns out to be the method of choice to predict β of push-pull π -conjugated systems since it closely reproduces the values obtained with the reference CCSD(T) scheme. Moreover, the SDQ-MP4 and CCSD approaches provide rarely improved estimates over MP2 while the MP4 method does not represent an improvement over MP4-SDQ or the SCS-MP2 method, over MP2. Among DFT exchange-correlation functionals, LC-BLYP is reliable when characterizing the changes of first hyperpolarizability upon enlarging the π -conjugated linker or upon changing the polyyne linker into a polyene segment. Nevertheless, its reliability is very similar to what can be achieved with the Hartree-Fock method and the MP2 scheme is by far more accurate. On the other hand, the BLYP, B3LYP, and BHandHLYP functionals perform quantitatively better in a number of cases but the trends are poorly described. This is also the case of the B2-PLYP and mPW2-PLYP functionals, which are often the most accurate, though they underestimate the increase of β when going from polyyne to polyene linkers and overestimate the enhancement of β with chain length.

In the second part of this Chapter, the evaluation of hyperpolarizabilities of three *p*-quinodimethane derivatives with low diradical character has been

determined. As electron correlation effects rule the electric field response properties, wavefunction and density functional theory based methods have been compared to the benchmark values calculated with the coupled cluster method including single and double excitations as well as perturbative estimate of the triples [CCSD(T)]. The basis set effects has been also further assessing. This study shows that the determination of the second hyperpolarizability with the CCSD method provides results in closest agreement with the CCSD(T) reference values. The use of MP2 level of theory performs well for closed-shell compound but not for open-shell ones. Spin-projection UMP3 and UMP4 methods reproduce well UCCSD(T) values for the *p*-quinodimethane but not for the charged compound. Without spin projection correction, density functional theory with a large range of exchange-correlation functionals does not perform well for these systems. Similar effects have been observed for polarizability and first hyperpolarizability, though these effects are smaller.

V.A. Introduction

After several decades of intensive research¹, the prediction and the interpretation of the hyperpolarizabilities remain an ambitious and tremendous task for quantum chemistry. Indeed, many aspects need to be considered to provide efficient guidelines for designing new systems with targeted nonlinear optical responses. Among these aspects to master in theoretical investigations, several works have addressed the vibrational contributions² or the effects of the surroundings³ while in this Chapter, we focus on the electron correlation for predicting the first hyperpolarizability of push-pull π -conjugated systems. The effects of electron correlation on the hyperpolarizabilities are known to be large.^{4, 5, 6, 7, 8, 9, 10} In 1979, Bartlett and Purvis^{4a} have already shown that the correlation effects upon the first and the second hyperpolarizabilities is large in the case of the

hydrogen fluoride molecule and also for the water molecule in 1981^{4b}. Maroulis^{4d} has shown that for butadiyne, hexatriyne, and octatetrayne, the second hyperpolarizabilities are more sensitive to geometry changes than to electron correlation. The impact of electron correlation is even more crucial in the case of open-shell systems.^{7a, 7d, 11} Indeed, Nakano *et al*^{7a} have assessed the dependence of the second hyperpolarizability on the diradical character for singlet diradical systems using a model compound, the *p*-quinodimethane. They also showed that the second hyperpolarizability values of singlet diradical systems in the intermediate and somewhat strong correlation regimes are significantly enhanced as compared to those in the weak correlation regime.

Nowadays, in order to estimate the impact of the electron correlation effects, for small polyatomic molecules, high-level wavefunction methods are recommended. Often the most efficient and accessible description is given by CCSD(T) calculations but, in general, this method cannot be routinely applied to large systems. Hierarchies of lower levels of approximation, including the HF and DFT methods, have been proposed for the estimation of β and γ but comparisons with coupled cluster results are quite rare. Among these, a recent CCSD study due to Hammond, Kowalski and coworkers^{10a}, showed that the CAM-B3LYP exchange-correlation functional outperforms the B3LYP, PBE0, and PBE XC functionals with respect to the CCSD result, for evaluating the polarizability and the first hyperpolarizability of acetonitrile while comparisons with the HF, CCS, and CC2 methods demonstrate the key role of electron correlation. In the case of *p*-nitroaniline, the prototype push-pull π -conjugated system, the best agreement is found with the B3LYP exchange-correlation functional whereas CAM-B3LYP underestimates it by about 15%. Moreover, Ref. 10a shows that for *p*-nitroaniline these XC functionals perform better than the CC2 scheme (overestimation by 30%) or the HF scheme (underestimation by 43 %). In their recent work, Suponitsky and coworkers^{10b} showed that two DFT functionals with a larger fraction of Hartree-Fock

exchange (BMK and especially M05-2X) provide the best agreement to experiment, in comparison with the MP2 method. They also highlighted that all hybrid DFT functionals, in their study, overestimate the hyperpolarizability of donor-acceptor π -conjugated molecules while they recommended using the M05-2X functional as a systematic prediction tool for β . In 2012, Castet and Champagne¹² addressed the performance of various exchange–correlation functionals by comparison to CCSD(T) calculations for a series of reference molecules (acetonitrile, dichloromethane, trichloroacetonitrile, chloroform, carbon tetrachloride). In their study, the best functionals turn out to be LC-BLYP, M05-2X, and M06-2X. It has been recognized for a decade that the performance of DFT with LDA, GGA, and hybrid exchange–correlation functionals will generally deteriorate when considering larger systems due to the XC potential shortsightedness¹³. The exchange–correlation requirements differ for small molecules like the reference NLO molecules and for extended systems like polymer chains or large donor-acceptor π -conjugated systems. In the first class of molecules, key issues to describe the hyperpolarization phenomena are related to the description of the most delocalizable and external part of the electron distribution.^{9a, 14} In the second class of compounds, the dominant part of the response is no more attributed to the outer part of the electron distribution but rather to the molecular units/moieties (hyper)polarization and to charge transfer,¹⁵ which requires a proper treatment of intramolecular long-range delocalization. On the other hand, Møller–Plesset second-order perturbation theory (MP2) can now routinely be applied to systems with a hundred atoms or more, which opens the way to investigate systems of interest for NLO applications. This approach was favored in several of recent investigations combining measurements and simulations¹⁶ because for model systems ($\text{NH}_2\text{--CH=CH--CH=CH--NO}_2$ and $\text{NH}_2\text{--CH=CH--CH=CH--CH=CH--CH=CH--CH=CH--NO}_2$), the MP2 scheme provides first hyperpolarizability values in close agreement with the CCSD(T) results.¹⁷ Nevertheless, for polybutatriene chains, the MP2 approach was found to overestimate the second

hyperpolarizability by more than a factor of 2 with respect to CCSD(T) for chains containing at least 16 C atoms^{2h}. In Ref. 10a, at the CCSD level, the importance of using very large basis sets to calculate the first hyperpolarizability has been reaffirmed, with the d-aug-cc-pV5Z basis set considered to provide reference values. So, for acetonitrile and chloroform, Hammond and Kowalski have shown that the d-aug-cc-pVTZ leads to accurate results whereas for *p*-nitroaniline, the Sadlej POL basis set leads to results in close agreement with more extended Dunning basis sets. They have further substantiated the fact that molecules of similar size but different chemical compositions may behave in a very different manner with respect to basis set convergence. Finally, as expected, in the case of *p*-nitroaniline the evaluation of the properties with respect to the basis set is far less demanding.

In a number of cases, long-range corrected XC functionals have appeared appealing to describe the hyperpolarizabilities of π -conjugated systems because the errors observed with LDA and GGA are reduced for a large part^{13a, 18}. These studies have therefore naturally concluded that, after improving the exchange part of the functional, the correlation part should also be improved. One of these strategies consists of adding MP2 correlation to the pure DFT correlation, leading therefore to double hybrid XC functionals¹⁹. Several improvements have been highlighted upon using the double hybrid B2-PLYP functional, for example, in the case of harmonic and anharmonic vibrational frequencies or in the prediction of the barrierless reaction of the benzenium-ethene complex to the ethyl-1H-benzene cation²⁰. This method has also been extended without any further empirical adjustment to electronic excited states in the framework of TDDFT or the closely related Tamm-Dancoff approximation (TDA-DFT)²¹. Other investigations²² have been carried out but not in the case of first hyperpolarizability determinations. Moreover, the new mPW2-PLYP double hybrid XC functional²³ outperforms the B2-PLYP functional in the computation of thermodynamic properties of large molecules. This functional

gives accurate predictions for both the excitation and emission energies of substituted p-quaterphenylenes nanofibers²⁴.

In addition to the MP2 scheme, we focus here also on two other Møller-Plesset perturbation theory schemes to calculate the first hyperpolarizabilities, the MP4 method and the spin-component-scaled Møller-Plesset second-order perturbation theory (SCS-MP2) method, which has been recently proposed by Grimme²⁵. The SCS-MP2 scheme consists in a separate scaling of the parallel and anti-parallel spin components of the MP2 energy expression and, therefore, it can be applied without additional cost to the ordinary MP2 method. The SCS-MP2 method has already demonstrated good reliability for a wide range of molecular properties, including energies of reaction, molecular geometries, harmonic frequencies, non-covalent interactions, heats of formation, as well as excitation energies^{20b, 26}. Preliminary results on the polarizability and second hyperpolarizability of model H₂ chains have also investigated the SCS-MP2 reliability²⁷. All these results call therefore for further investigations of electron correlation effects on the first hyperpolarizability of push-pull π -conjugated systems and for the assessment of MP2 and new DFT schemes with respect to CCSD(T).

In order to carry out this investigation on electron correlation effects, the first part of this chapter is dedicated to six Donor/Acceptor (D/A) π -conjugated systems (Fig. 1). The donor (dimethylamino) and acceptor (nitro) groups are amongst the strongest donating and accepting groups. π -conjugated linkers of different lengths and natures were considered: a polyyne segment of four or six triple bonds forms the backbone of compounds **a** and **c**, whereas in compounds **b** and **d**, it is a polyene segment with four or six double bonds, respectively. In **e** and **f**, the linker is made of two aromatic rings, either two phenyl or two thiophene rings. As a result of electron delocalization along the π -conjugated backbone, the longitudinal component of the first hyperpolarizability, $\beta_{xxx} = \beta_L = \beta$, is by far the largest tensor component. Thus, in the following, we just considered this longitudinal component of the static first

hyperpolarizability. The conclusions of this first part are useful for Chapter 3 on fluorescent proteins.

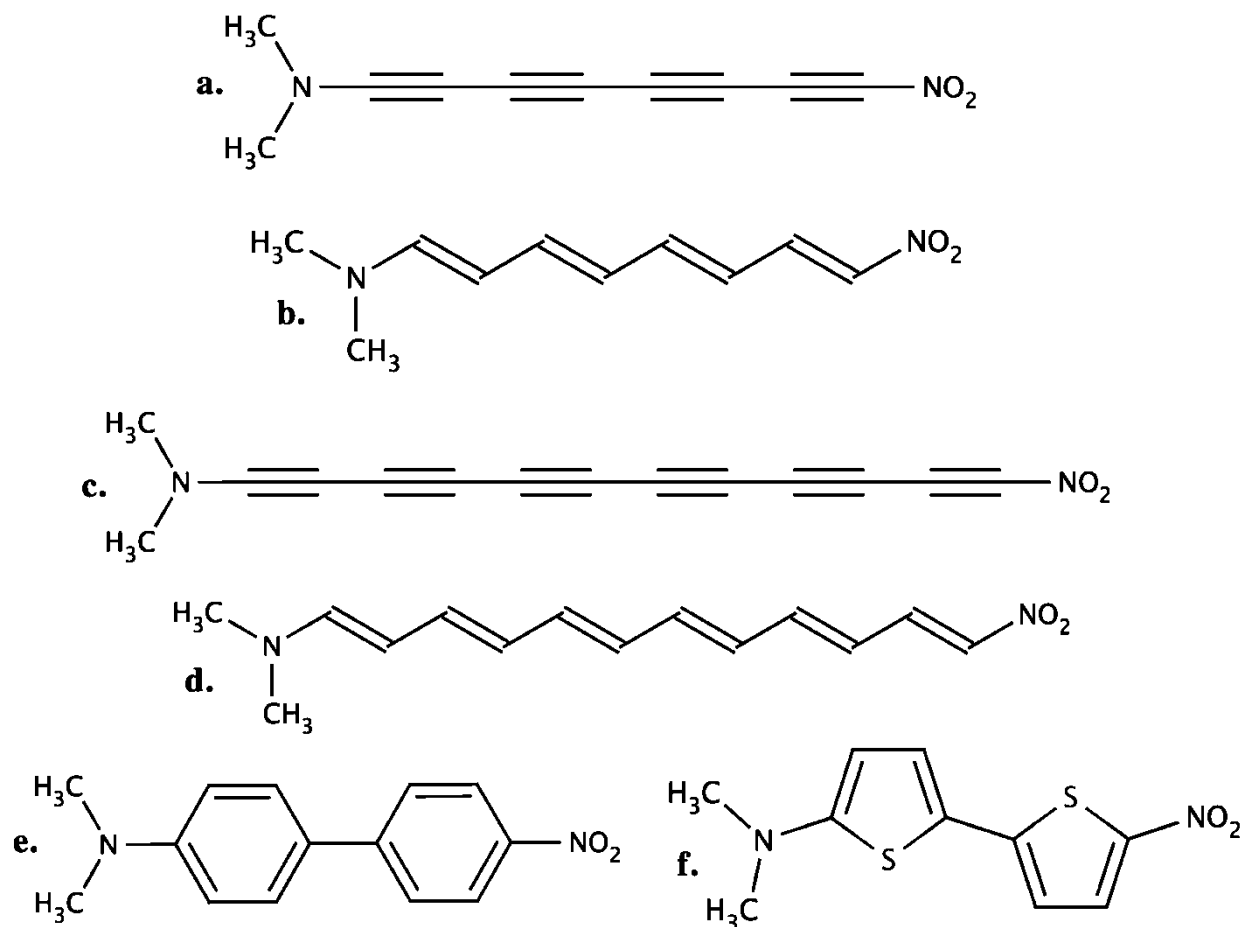


FIG. 1. Chemical structures of the **a-f** D/A π -conjugated systems,

- a**, *dimethyl(8-nitroocta-1,3,5,7-tetrayn-1-yl)amine*,
- b**, *dimethyl(8-nitroocta-1,3,5,7-tetraen-1-yl)amine*,
- c**, *dimethyl(12-nitrododeca-1,3,5,7,9,11-hexayn-1-yl)amine*,
- d**, *dimethyl(12-nitrododeca-1,3,5,7,9,11-hexaen-1-yl)amine*,
- e**, *N,N-dimethyl-4-(4-nitrophenyl)aniline*,
- f**, *N,N-dimethyl-5-(5-nitrothiophen-2-yl)thiophen-2-amine*.

The second part of this Chapter on model systems deals with the evaluation of the second hyperpolarizability of three challenging open-shell molecules, where electron correlation effects govern the electric field response properties. They

encompass *p*-quinodimethane (**1**) and two derivatives where the $\text{CH}_2^\bullet$ terminal units are formally replaced by two $\text{NH}_2^{+\bullet}$ (**2**), or by one $\text{NH}_2^{+\bullet}$ and one $\text{BH}_2^\bullet$ (**3**) (Fig.2). As in the first part, this study addresses the choice of a suitable basis set and the description of electron correlation. So, it goes beyond recent works where γ of **1** has been calculated with rather small basis sets^{7a}, which are not always suitable to describe accurately the evolution of the second hyperpolarizability, especially in the case of poorly polarized sets.²⁸ Electron correlation is treated using the Møller-Plesset perturbation theory approach in its unrestricted form and the effects of spin contamination are assessed using spin projected approaches. Our benchmark values are obtained at the CCSD level as well as at the CCSD(T) level. Then, within DFT, some conventional and less conventional XC functionals are employed and their performance is tested with respect to the UCCSD(T) data. Besides γ , this part also analyzes the polarizability and first hyperpolarizability, of which the estimates depend also on the inclusion of static and dynamic electron correlations. Some of these models belong to the class of “open-shell singlet” species, which have recently attracted much attention due to their unique structural, optical, and magnetic properties.²⁹

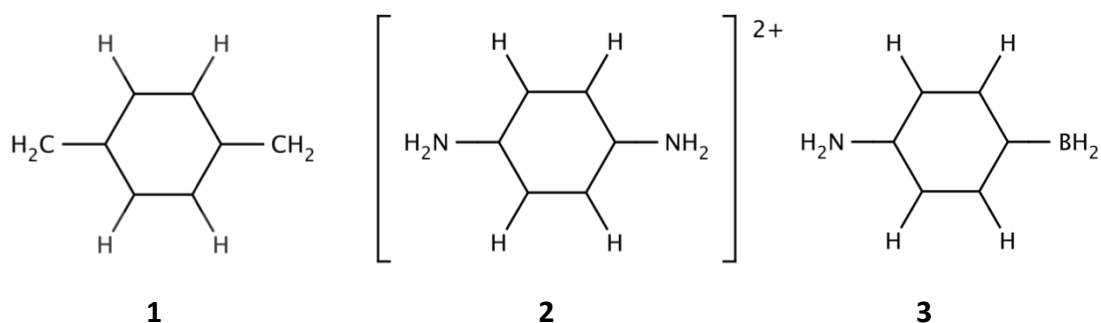


FIG. 2. Structures of three challenging open-shell molecules.

The chapter is divided in two parts as follows. For the first part on push-pull π -conjugated systems, the first Section briefly describes the key computational and theoretical aspects whereas the results are presented and discussed in Section V.B.2. before conclusions are drawn in Section V.B.3. For the second part on open-

shell molecules, the methods and the computational scheme are summarized in Section V.C.1. Section V.C.2. presents and discusses the results, whereas Section V.C.3. concludes and draws some perspectives.

V.B. *Electron Correlation Effects on the First Hyperpolarizability of Push-Pull π -Conjugated Systems*

V.B.1. Theoretical and computational aspects

The geometries were optimized at the DFT level using the B3LYP XC functional and the 6-311G(d) basis set. The longitudinal (x) direction is defined by one of the principal inertial axes. For compounds **a**, **c**, and **e**, this axis is passing by the two nitrogens of the substituents while for molecules **b**, **d**, and **f**, this axis passes nearby the midpoint of the C-N bonds.

The finite field procedure³⁰ was used to evaluate the static longitudinal first hyperpolarizability at the different levels of approximation. These encompass MO based methods: i) the HF method, ii) the MP2 and MP4 approaches, iii) the SCS-MP2 method, iv) the CCSD method, and v) the CCSD(T) method, as well as DFT schemes with a range of XC functionals: i) the GGA BLYP functional, ii) the hybrid B3LYP and BHandHLYP functionals, iii) the LC-BLYP functional ($\mu=0.47$), and iv) the double hybrid B2-PLYP functional and mPW2-PLYP. To achieve an accuracy of 1-10 a.u. on the first hyperpolarizability values, special care was given to the choice of threshold in the iterative procedures. In the SCF calculations, the convergence parameter N, which implies a convergence of 10^{-N} on the energy and $10^{-(N+2)}$ on the wavefunction, was set to 10. On the other hand, in the CC calculations, the convergence parameter M, which implies a convergence of 10^{-M} on the energy and $10^{-(M+2)}$ on the wavefunction, was set to 9. In the latter case, going from M = 9 to M = 10 did not lead to any relevant difference (< 0.05 a.u.) on β . The FF differentiation was used to

determine β . The Romberg procedure was applied in combination with field amplitudes ranging from ± 0.0004 to ± 0.0064 a.u.

Several atomic basis sets were employed, starting from the 6-31G(d) valence double- ζ + polarization basis set and then adding an additional set of valence functions [6-311G(d)] or a set of diffuse functions [6-31+G(d)]. Much larger basis sets, including the 6-311++G(2df,p), aug-cc-pVDZ, and aug-cc-pVTZ basis sets, were also employed at the SCF and MP2 levels of approximation. In addition, the Z3PolX basis set due to Benkova *et al.*³¹ was selected. The Zm PolX [where m denotes the number of primitive GTOs (Gaussian Type Orbitals) in the contracted polarization] basis sets are reduced-size polarized basis sets, which result from the truncation of the original PolX basis sets³² without any major loss of accuracy when calculating the molecular electric properties. This Zm PolX basis sets have primarily been tailored to well reproduce the molecular dipole moments and the dipole polarizabilities but recent calculations have also demonstrated their performance for calculating the NLO properties^{10a}. Simultaneously, the dipole moment and polarizability data remain at almost the same level of accuracy as in the case of the original PolX basis set, which is closely comparable to the d-aug-cc-pVDZ set in the Dunning series, despite being approximately two-thirds of the size. This first part will further assess the performance of the Z3PolX basis set in the case of the first hyperpolarizability of large D/A π -conjugated systems.

All reported β values are given in a.u. within the T convention. All calculations were performed using the GAUSSIAN 03 and 09 program packages³³.

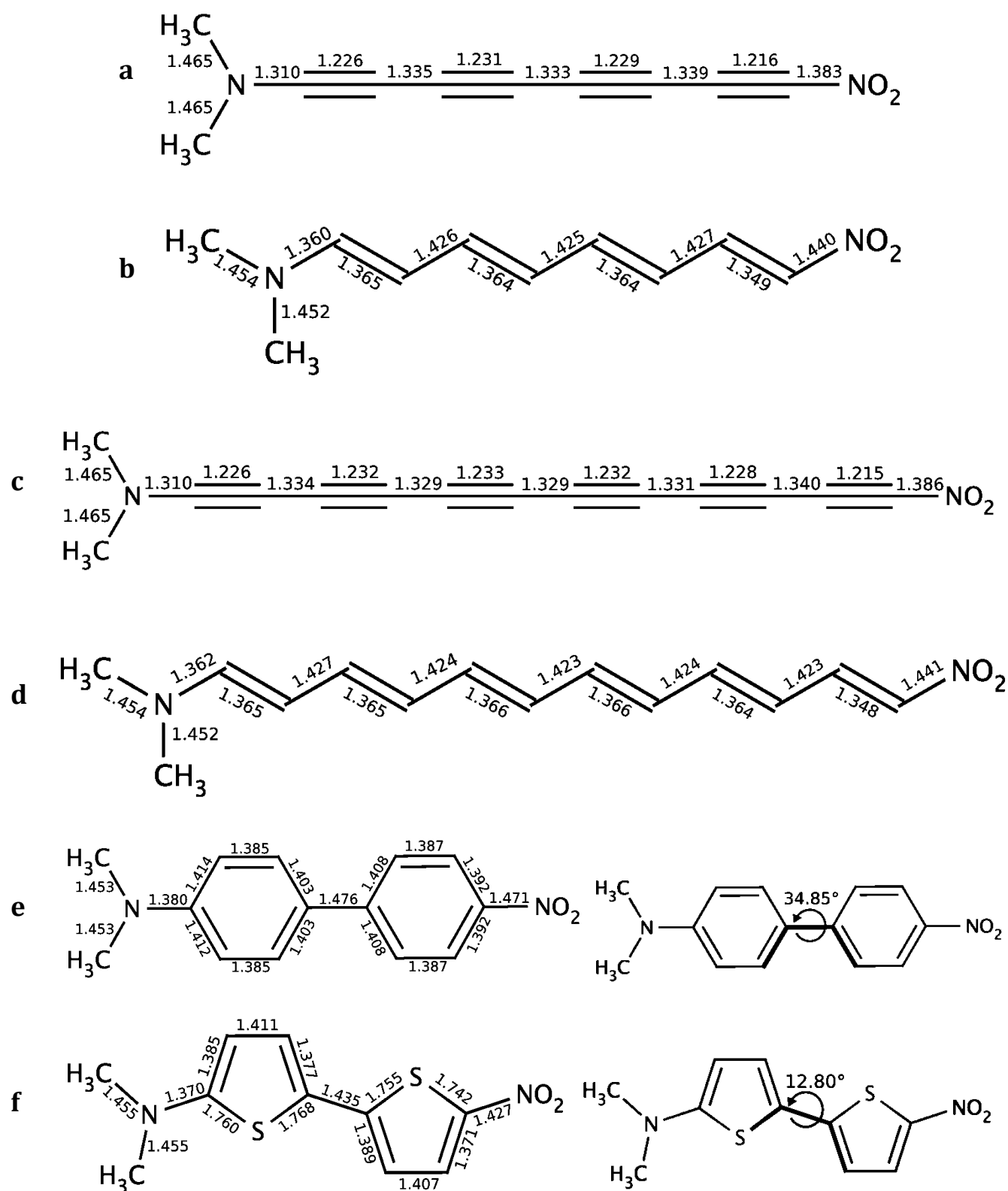


FIG. 3. Principal inter-atomic distances (Å) for the set of six D/A π -conjugated systems. For compounds **e** and **f**, the dihedral angle between the two cycles is also given.

V.B.2. Results and discussions

V.B.2.a. Optimized geometries

The principal geometrical characteristics of the six D/A π -conjugated systems are shown in Fig. 3, i.e. the principal inter-atomic distances and the dihedral angles between the two cycles for molecules **e** and **f**. In the latter cases, the steric constraint between the two cycles prevents the π -conjugated pathway to be planar. Thus, these two molecules are less conjugated than the others. As expected, the average bond length alternation (BLA) is smaller for the polyene linker (0.06 Å for **b** and **d**) than for the polyyne linkers (0.11 Å for **a** and 0.10 Å for **c**). In the case of compounds **e** and **f**, the inter-ring BLA amounts to 0.07 Å and 0.05 Å, whereas the intra-ring BLA is much smaller and attains only 0.02 Å and 0.03 Å, respectively. The smaller inter-ring BLA for compound **f** can be related to the smaller dihedral angle (12.8° for **f** vs. 34.9° for **e**) and therefore a better electron conjugation.

V.B.2.b. Basis set effects

The smaller D/A molecules **a** and **b** are used to investigate the basis set effects on β . Due to basis set compensation effects, basis set effects on the other and larger compounds are expected to be smaller. The results are listed in Table I. The addition of a supplementary set of valence basis functions from the 6-31G(d) to the 6-311G(d) basis set has a rather small effect on β , with a decrease of β not larger than 1.5%, for all levels of calculation. On the other hand, the addition of diffuse functions has a larger effect. At the HF level, going from the 6-31G(d) to the 6-31+G(d) basis set results in an increase of β by about 10% but this increase is about twice larger when using MO correlated methods. Indeed, in the case of compound **a**, the increase of β when going from the 6-31G(d) to 6-31+G(d) basis set amounts to 17% at the MP2 and SCS-MP2 levels, 20% at the MP4 and CCSD levels, and up to 23% for the CCSD(T) method. Similar effects are found for compound **b**, with increases of β by 12.6%, 14.6%, 18.4%, and 22.6% at the HF, MP2, CCSD, and

CCSD(T) levels of approximation, respectively. For compound **a**, the β values obtained with the Z3PolX basis set are another 3-4% larger than the 6-31+G(d) values whereas for compound **b**, the variations are smaller and range from a decrease of β by 2.4 % (MP4) to an increase by 1.8% (SCS-MP2).

Additional assessments of these medium-size basis sets were then carried out by comparison with more extended basis sets. The differences in β between these three reference basis sets [aug-cc-pVDZ, 6-311++G(2df,p), and aug-cc-pVTZ] are generally small (less than 2%). In particular, for compound **a**, going from the aug-cc-pVDZ to the aug-cc-pVTZ basis set leads to a decrease by about 1%, while the number of basis functions more than double. The 6-311++G(2df,p) basis set predicts β values, which are slightly smaller than with the aug-cc-pVTZ basis set. For compound **b**, the differences between the 6-311++G(2df,p) and aug-cc-pVDZ basis sets are slightly smaller than for compound **a** while the differences between the 6-311++G(2df,p) and aug-cc-pVTZ basis sets are very small, and smaller than 0.3% at the MP2 and SCS-MP2 levels of approximation. The ratios between the β values for the medium-size basis sets and these reference values are given in Table I. For molecule **a**, the closest agreement is obtained with the 6-31+G(d) basis set, with overestimations smaller than 2 % or 4% at the HF and MP2/SCS-MP2 levels of approximation, respectively. The Z3PolX basis set gives also good results but the calculations takes longer than with the 6-31+G(d) basis set, due to the slower SCF convergence. For molecule **b**, with respect to the 6-311++G(2df,p) or aug-cc-pVTZ basis sets the 6-31G(d) basis set underestimates less the first hyperpolarizability than the 6-311G(d) basis set while the 6-31+G(d) and Z3PolX basis sets lead to overestimations of similar amplitudes. The results on compound **b** further show that including or not diffuse functions in the basis set has little impact on the agreement with respect to more complete basis sets. Indeed, although the difference in β between the 6-31G(d) and 6-311G(d) basis sets on one side and the 6-31+G(d) and Z3PolX basis sets on the other side attains 10 to 15 %, the amplitude

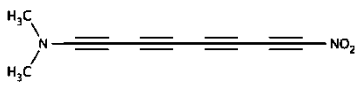
of the deviations with respect to the 6-311++G(2df,p) or aug-cc-pVTZ basis set is similar and fluctuates between 5 and 10 %.

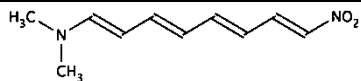
As intermediate conclusion from this section, one can state that, for evaluating the first hyperpolarizability of molecule **a**, the 6-31+G(d) basis set gives the best balance between accuracy and computational resources whereas for molecule **b**, the 6-31G(d) basis set is already an optimal choice. Another interesting comment is related to the usual belief that higher order methods require larger basis sets to achieve a converged description of the electron correlation effects. As far as we consider the data reported in Table I, it turns out that a) the MP2/HF ratios for both compounds **a** and **b** are very stable when going from the smallest 6-31G(d) to the largest aug-cc-pVTZ basis sets with variations smaller than 7 % and b) the variations of the CCSD(T)/MP2 ratios estimated from using the 6-31G(d), 6-311G(d), 6-31+G(d), and Z3PolX basis sets are also within a range of 5-7%, demonstrating that for the longitudinal first hyperpolarizability of push-pull π -conjugated systems, these basis set effects are not exacerbated.

V.B.2.c. Electron correlation effects using wavefunction based methods

Considering the medium-size basis sets discussed above, the impact of electron correlation on the static longitudinal first hyperpolarizabilities of compounds **a**, **b**, **c**, **e** and **f** were addressed. Table II reports the corresponding $\beta(X)/\beta[\text{CCSD(T)}]$ ratios, where X is a lower level of approximation than CCSD(T). The HF method systematically underestimates β . Using basis sets lacking diffuse functions, this underestimation amounts to about 35 to 39% for compounds **a**, **c**, **e** and **f** while 54% for compound **b**. When diffuse functions are included in the basis set, the underestimation increases by 7% for compound **a**, but only by 3 % for compound **b**.

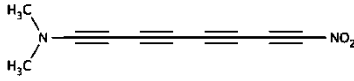
TABLE I. Basis set effects on the static longitudinal first hyperpolarizability of compounds **a** and **b**, at different levels of calculation. The values in parentheses correspond to the ratios, at a given level of approximation, with the respect to the corresponding au-cc-pVDZ, 6-311++G(2df,p), and au-cc-pVTZ basis sets.

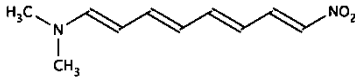
	 a						
	6-31G(d)	6-311G(d)	6-31++G(d)	Z3PolX	aug-cc-pVDZ	6-311++G(2df,p)	aug-cc-pVTZ
Number of basis functions	222	270	278	288	376	518	782
HF	8151	8151	9077	9429			
	(0.90)	(0.90)	(1.00)	(1.04)	9089		
	(0.91)	(0.91)	(1.02)	(1.05)		8946	
	(0.91)	(0.91)	(1.01)	(1.05)			8966
MP2	13255	13068	15576	16011			
	(0.87)	(0.86)	(1.02)	(1.05)	15258		
	(0.88)	(0.87)	(1.04)	(1.07)		14988	
	(0.88)	(0.87)	(1.03)	(1.06)			15096
SCS-MP2	12937	12777	15121	15601			
	(0.87)	(0.86)	(1.02)	(1.05)	14824		
	(0.89)	(0.88)	(1.04)	(1.07)		14566	
	(0.88)	(0.87)	(1.03)	(1.06)			14653
MP4	13783	13607	16498	16925			
CCSD	11490	11260	13700	14069			
CCSD(T)	12685	12380	15628	16017			

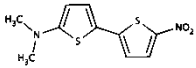
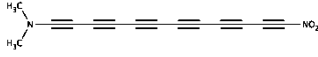
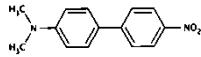
b

	6-31G(d)	6-311G(d)	6-31+G(d)	Z3PolX	aug-cc-pVDZ	6-311++G(2df,p)	aug-cc-pVTZ
Number of basis functions	238	294	294	336	448	574	966
HF	10138	10007	11413	11207			
	(0.96)	(0.95)	(1.08)	(1.06)	10581		
	(0.98)	(0.96)	(1.10)	(1.08)		10389	
	(0.98)	(0.97)	(1.10)	(1.08)			10331
MP2	24206	24016	27745	28126			
	(0.92)	(0.91)	(1.06)	(1.07)	26252		
	(0.94)	(0.93)	(1.07)	(1.08)		25847	
	(0.93)	(0.93)	(1.07)	(1.09)			25905
SCS-MP2	24785	24619	28318	28819			
	(0.93)	(0.92)	(1.06)	(1.08)	26763		
	(0.94)	(0.94)	(1.08)	(1.10)		26291	
	(0.94)	(0.93)	(1.07)	(1.09)			26363
MP4	24943	24574	28574	27897			
CCSD	21190	20874	25079				
CCSD(T)	21646	21549	26539				

TABLE II. Electron correlation effects on the static longitudinal first hyperpolarizability of compounds **a**, **b**, **c**, **e** and **f**, as estimated by the $\beta(X)/\beta[\text{CCSD(T)}]$ ratios, where X is a lower level of approximation than CCSD(T).

a 				
	6-31G(d)	6-311G(d)	6-31+G(d)	Z3PolX
HF	0.643	0.658	0.580	0.589
MP2	1.045	1.056	0.997	1.000
SCS-MP2	1.020	1.032	0.968	0.974
MP4	1.086	1.100	1.056	1.057
SDQ-MP4	0.931	0.933	0.881	0.891
CCSD	0.906	0.910	0.877	0.878

b 			
	6-31G(d)	6-311G(d)	6-31+G(d)
HF	0.464	0.464	0.430
MP2	1.118	1.114	1.045
SCS-MP2	1.145	1.142	1.067
MP4	1.152	1.140	1.076
SDQ-MP4	1.083	1.062	0.999
CCSD	0.980	0.969	0.945

f 			
c 			
e 			
	6-31G(d)	6-31G(d)	6-31G(d)
HF	0.663	0.647	0.618
MP2	1.178	0.992	1.026
SCS-MP2	1.187	0.922	0.998
MP4	1.150	1.174	1.084
SDQ-MP4	0.977	0.887	0.892
CCSD	0.928	0.838	0.887

Including the lowest order electron correlation correction (MP2) leads to β values in very close agreement with the reference CCSD(T) data. Moreover, the difference between the MP2 and CCSD(T) results is smaller i) when the basis set contains diffuse functions and ii) for the polyyne than for the polyene linker. Switching to the SCS-MP2 method does not provide a systematic improvement over the MP2 results: for compound **a**, the SCS-MP2 method is in better agreement with respect to the CCSD(T) reference value provided the basis set does not contain diffuse functions. On the other hand, the MP2 method performs better than the SCS-MP2 scheme for molecules **b**, **c**, and to a lesser extent for molecule **f**. For molecule **e**, SCS-MP2 performs slightly better when using the 6-31G(d) basis set.

Except for compound **f**, moving from MP2 to MP4 leads to a small increase of the first hyperpolarizabilities, which results in larger deviations with respect to the CCSD(T) data. Data in Table II also point out the relatively large contribution from the triples, in particular in the case of the polyyne linker, where the CC triple bonds concentrate a large amount of electronic charge. So, for compounds **a** and **c**, removing the fourth-order triples (T4) contribution switches the difference with respect to CCSD(T) from an overestimation by 5-10 % for **a** and 17% for **c** to an underestimation by 7-12%. This T4 contribution is smaller for compounds **b** and **f**, which also brings the SDQ-MP4 β values in better agreement with the CCSD(T) results than the MP4 results.

Finally, considering the CCSD results, the main conclusions are i) the CCSD method underestimates the CCSD(T) β values by up to 16%, ii) for both compounds **a** and **b**, this difference increases when the basis set includes diffuse functions, and iii) the agreement with CCSD(T) is better for compound **b** than compound **a**, demonstrating again that the triples contribution is larger for systems with CC triple bonds than for these with CC double bonds.

In view of these results, due to some cancellations between higher order contributions, i) the MP2 method turns out to be a method of choice to predict the first hyperpolarizabilities of push-pull π -conjugated systems, ii) the SDQ-MP4 or CCSD approaches provide improved estimates over MP2 only in the case of compounds **b** and **f**, iii) in comparison to MP4-SDQ, the MP4 method does not constitute an improvement, and iv) the SCS-MP2 improves over MP2 only in limited cases.

V.B.2.d. Structure-property relationships

Table III lists the β values of the six D/A π -conjugated systems at different levels of approximation. Since the MP4, CCSD, and CCSD(T) calculations require substantial computational resources, in several cases we have considered the MP2 approximation as the reference, because it leads, in view of the preceding subsection, to results in close agreement with the CCSD(T) values.

In the case of “short” polyene and polyyne chains, β increases nonlinearly with the number (N) of double/triple bonds. In other words, the first hyperpolarizability reported to the number of double/triple bonds increases with N . Indeed, β (MP2, 6-31+G(d))/ N amounts to 3894 a.u. and 5300 a.u. for **a** and **c** and to 6936 a.u. and 12623 a.u. for **b** and **d**, respectively. For a given number of units in the chain, the D/A molecule with doubly-bonded π -conjugated linkers displays a larger β value than its triply-bonded analog. This can be rationalized in terms of the Bond Length Alternation (BLA), following the general relations where β increases with decreasing BLA, attains a maximum for BLA of around 0.05 Å, then, it decreases, becomes negligible, and reverses sign³⁴. Indeed, the BLA for molecule **d** (or **b**) amounts 0.06 Å, which is close to the value maximizing β , while for molecule **c** (and **a**) we have 0.10 Å (0.11 Å).

For the molecules **e** and **f**, the steric constraint between the two cycles prevents the π -conjugated pathway to be planar. Combined with the effect of ring

aromaticity, the electron delocalization from the donor to the acceptor is disturbed, these two molecules are less conjugated than the others, which explains why they have smaller β values. Between molecules **e** and **f**, molecule **f** presents a larger β value than molecule **e**. This can be rationalized in terms of the amplitude of the dihedral angle between the two cycles. For molecule **f**, the dihedral angle is 12.8° while for the molecule **e**, it raises to 34.9°.

Figs. 4 and 5 summarize these structure-property relationships, where β is plotted as a function of the distance between the two nitrogen atoms at the chains ends (Fig. 3) or as a function of the number of the π -electrons in the linker (Fig. 5). These figures also demonstrate that similar conclusions are drawn when using the MP2 and MP4 methods or by employing the 6-31G(d) and 6-31+G(d) basis sets. If adopting the HF and MP2 methods, similar orderings of β values are obtained, though the β underestimation when using the HF method has been shown to be stronger for a polyene linker. This is further evidenced by considering the ratios of β values with respect to compound **a**. At the HF level, the β values are in the ratio 1.00 : 1.26 : 2.16 : 3.49 : 0.43 : 0.74 for compounds **a-f**, respectively whereas 1.00 : 1.78 : 2.04 : 4.86 : 0.43 : 0.80 at the MP2 level.

TABLE III. Static longitudinal first hyperpolarizability for all the D/A π -conjugated systems considered in this study, at different levels of calculation and for different basis sets.

	a	b	c	d	e	f
6-31G*						
Number of basis functions	222	238	282	306	298	268
RHF	8151	10138	18079	28652	3432	5825
MP2	13255	24206	27704	66895	5692	10346
SCS-MP2	12937	24785	25768	66056	5541	10429
MP4	13783	24943	32809	68446	6015	10102
SDQ-MP4	11804	23453	24791	60669	4952	8582
6-311G*						
Number of basis functions	270	294	342	378	366	334
RHF	8151	10007	18175	28421	3498	5904
MP2	13068	24016	27814	67234	5752	10342
SCS-MP2	12777	24619	25930	66592	5602	10429
MP4	13607	24574	32998	68575	6113	10113
SDQ-MP4	11556	22877	24809	60411	4963	8436
6-31+G*						
Number of basis functions	278	294	354	378	370	332
RHF	9077	11413	19607	31701	3895	6693
MP2	15576	27745	31804	75741	6769	12393
SCS-MP2	15121	28318	29454	74657	6555	12413
MP4	16498	28574	38043	/	/	/
SDQ-MP4	13766	26529	28158	/	/	/

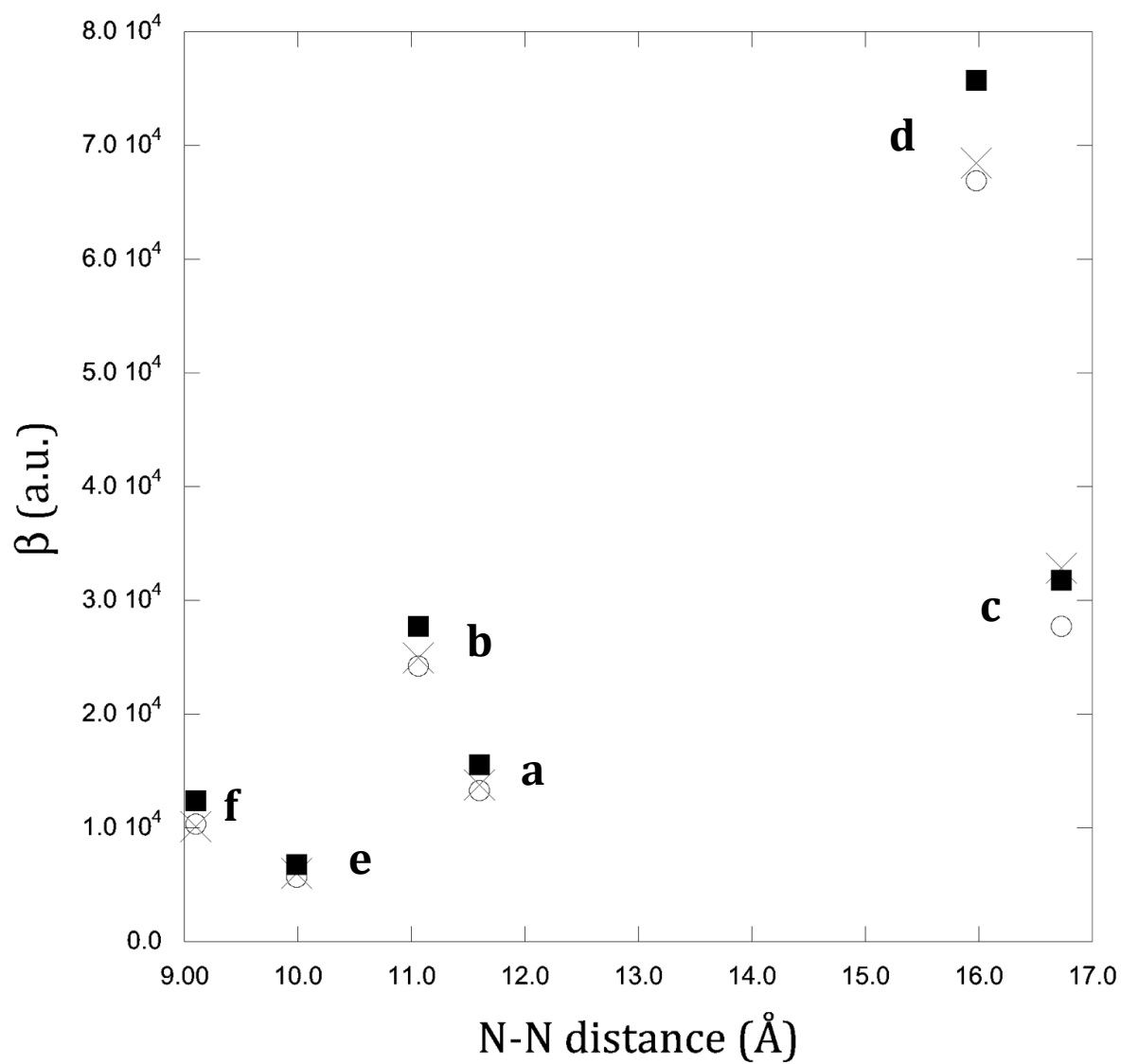


FIG. 4. β as a function of the distance between the two nitrogen atoms at the chain extremities, calculated at different levels of approximation: MP2/6-31+G(d): ■; MP2/6-31G(d): ○; MP4, 6-31G(d): ×.

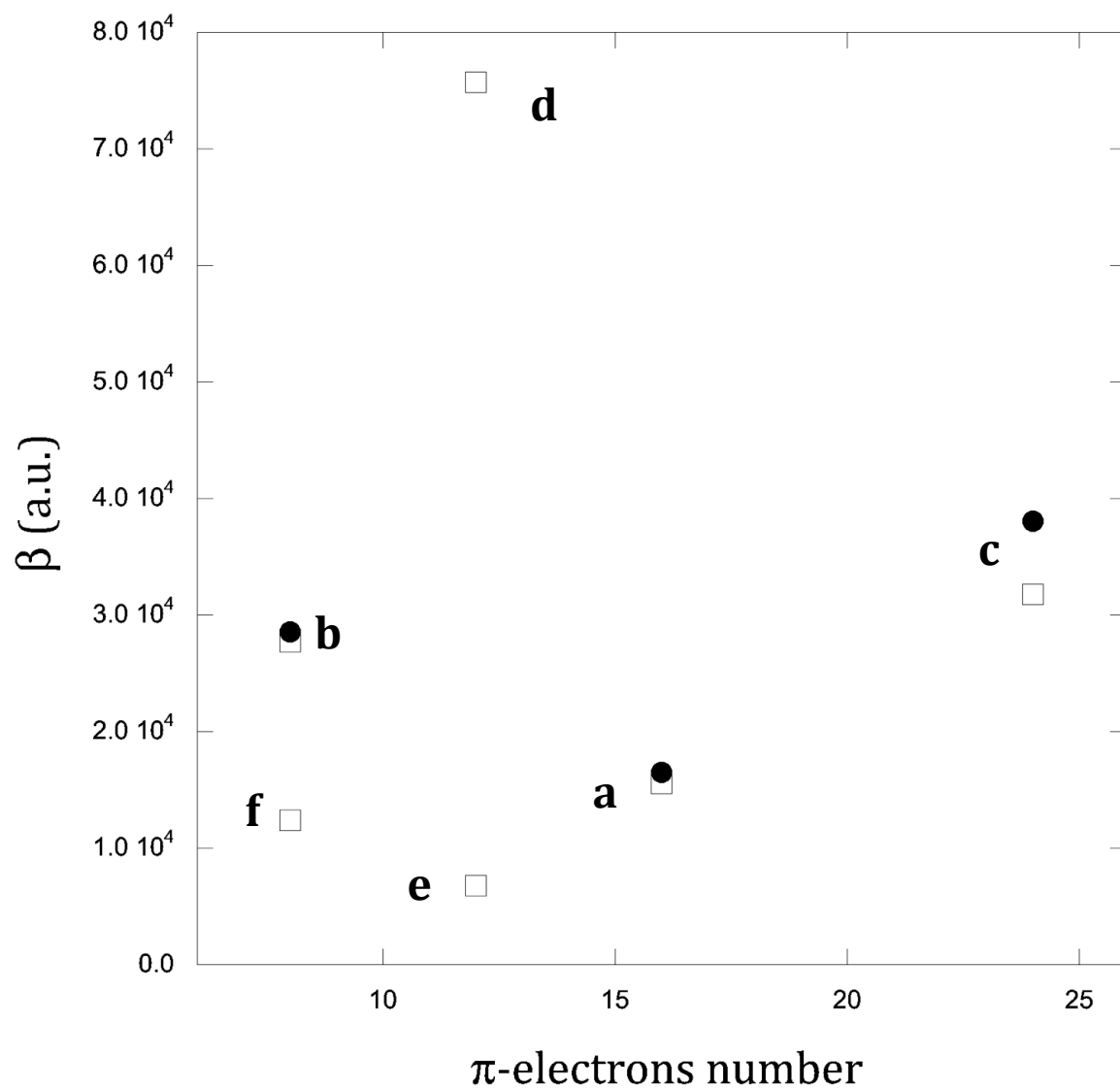


FIG. 5. β as a function of the number of π -electrons in the linker, calculated at different levels of approximation: MP2/6-31+G(d): $\square$; MP4/6-31+G(d): $\bullet$.

V.B.2.e. Assessment of DFT exchange-correlation functionals.

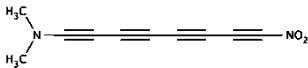
The D/A molecules **a**, **b**, **c** and **d** are used for assessing the performance of DFT XC functionals to account for electron correlation effects in the description of the β responses. As already mentioned in the introduction, the performance of traditional and hybrid XC functionals (here: BLYP, BHandHLYP, and B3LYP) will generally deteriorate when considering larger and larger systems. In order to extend this analysis, we also employed the LC-BLYP functional, which is a long-range corrected XC functional, and the B2-PLYP and mPW2-PLYP functionals, which are double hybrid XC functionals^{19, 24}, where MP2 correlation is added to the pure DFT correlation. Table IV lists the $\beta(X)/\beta[\text{CCSD(T)}]$ ratios or the $\beta(X)/\beta[\text{MP2}]$ ratio when CCSD(T) calculations are not available.

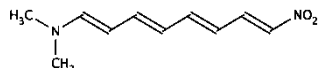
For compound **a**, among the GGA and hybrid XC functionals, the B3LYP functional is very efficient and only underestimates β by 1% or less, for both the 6-31G(d) and 6-31+G(d) basis sets. Using the BHandHLYP and BLYP functionals provides β values smaller than the CCSD(T) reference by 6-10% and 13-15%, respectively. The LC-BLYP functional leads to an underestimation by 19% with the 6-31G(d) basis set or by 20% when adding diffuse functions. The B2-PLYP (mPW2-PLYP) double hybrid XC functional overestimates β by 19% (15%). For molecule **c**, which has a longer polyyne chain than **a**, the β values obtained using the 6-31G(d) basis set are now overestimated when using the BLYP (27%), BHandHLYP (10%), and B3LYP (37%) XC functionals. The smallest overestimation is found for the functional with the largest (50%) amount of HF exchange. The LC-BLYP functional does not perform better than BHandHLYP since it underestimates β by 18% while B2-PLYP overestimates it by about 46% and mPW2-PLYP by 38%. Similar trends are observed when using the 6-31+G(d) basis set, even though the MP2 data are used as reference.

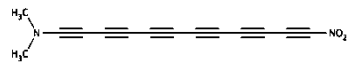
Turning now to the polyene linkers and compound **b**, the BLYP, BHandHLYP, B3LYP, and LC-BLYP functionals strongly underestimate β . The only satisfactory

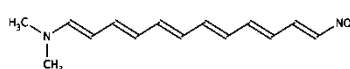
values are obtained with the B2-PLYP and mPW2-PLYP functionals, with underestimations around 10-15%. Again, it is interesting to see the evolution of the method reliability for a push-pull compound with a slightly longer linker, here compound **d**. There is no substantial improvement for the BLYP, BHandHLYP, B3LYP, and LC-BLYP functionals whereas the B2-PLYP and mPW2-PLYP functionals remain accurate. Going from the 6-31G(d) to the 6-31+G(d) basis set has a limited impact on this analysis. Note however that the reliability of the conventional functionals will strongly deteriorate when further enlarging the conjugated linker^{13a, 18}.

TABLE IV. Comparison between MO-based and density-based methods to predict the static longitudinal first hyperpolarizability of compounds **a-d**.

				
Number of basis functions	6-31G(d)	$\beta(X)$	6-31+G(d)	$\beta(X)$
	222	$\beta[\text{CCSD(T)}]$	278	$\beta[\text{CCSD(T)}]$
HF	8151	0.643	9077	0.580
BLYP	10788	0.850	13503	0.864
BHandHLYP	11870	0.936	14021	0.897
B3LYP	12684	0.999	15556	0.995
LC-BLYP	10316	0.813	12471	0.798
B2-PLYP	15121	1.192	18627	1.192
mPW2-PLYP	14607	1.151	17904	1.146
MP2	13255	1.045	15576	0.997
CCSD(T)	12685	1.000	15628	1.000

				
Number of basis functions	6-31G(d)	$\beta(X)$	6-31+G(d)	$\beta(X)$
	238	$\beta[\text{CCSD(T)}]$	294	$\beta[\text{CCSD(T)}]$
HF	10138	0.464	11413	0.430
BLYP	12341	0.570	14845	0.559
BHandHLYP	14067	0.650	16419	0.619
B3LYP	14033	0.648	16706	0.629
LC-BLYP	14847	0.686	17824	0.672
B2-PLYP	19497	0.901	23199	0.874
mPW2-PLYP	18907	0.873	22452	0.846
MP2	24206	1.118	27745	1.045
CCSD(T)	21646	1.000	26539	1.000

				
Number of basis functions	6-31G(d) 282	$\beta(X)$ $\beta[\text{CCSD(T)}]$	6-31+G(d) 354	$\beta(X)$ $\beta[\text{MP2}]$
HF	18079	0.647	19607	0.631
BLYP	35504	1.271	44427	1.397
BHandHLYP	30853	1.104	35596	1.119
B3LYP	38339	1.372	46598	1.465
LC-BLYP	22923	0.821	26612	0.837
B2-PLYP	40902	1.464	49205	1.547
mPW2-PLYP	38235	1.380	46617	1.464
MP2	27704	0.992	31804	1.000
CCSD(T)	27935	1.000	/	/

				
Number of basis functions	6-31G(d) 306	$\beta(X)$ $\beta[\text{MP2}]$	6-31+G(d) 378	$\beta(X)$ $\beta[\text{MP2}]$
HF	28652	0.428	31701	0.418
BLYP	41442	0.619	48946	0.646
BHandHLYP	45090	0.674	51580	0.681
B3LYP	47743	0.714	55781	0.736
LC-BLYP	40287	0.602	47283	0.628
B2-PLYP	64381	0.962	75153	0.992
mPW2-PLYP	61844	0.924	72135	0.952
MP2	66895	1.000	75741	1.000

Rather than just considering the absolute value of the hyperpolarizabilities, the predictive power of these methods can also be assessed by considering trends in the hyperpolarizabilities among chemically-related species. As a matter of fact, we have considered the ratios between β of chains containing the same number of unsaturated bonds but double *versus* triple bonds, *i.e.* $\beta[\mathbf{b}]/\beta[\mathbf{a}]$ and $\beta[\mathbf{d}]/\beta[\mathbf{c}]$,

and ratios between chains containing 6 and 4 unsaturated bonds of the same nature, *i.e.* $\beta [\mathbf{c}]/\beta [\mathbf{a}]$ and $\beta [\mathbf{d}]/\beta [\mathbf{b}]$. The results are presented in Table V. The double *versus* triple bond ratio, estimated by $\beta [\mathbf{b}]/\beta [\mathbf{a}]$ and $\beta [\mathbf{d}]/\beta [\mathbf{c}]$, is larger than 1.5 and increases with the linker size. This increase is reproduced at the HF level but the ratios are underestimated. The performance of B2-PLYP and mPW2-PLYP are similar, while they require the evaluation of cpu-expensive MP2 contributions. Among the DFT schemes, the LC-BLYP functional provides here the best result. On the other hand, the ratios and their evolution with chain length get worse when going from BHandHLYP to B3LYP, and to BLYP. When enlarging the π -linker by 2 unsaturated bonds, the first hyperpolarizability increases by about 100-125% ($\beta [\mathbf{c}]/\beta [\mathbf{a}] \sim 2.2$). This effect is well reproduced at the HF and LC-BLYP levels whereas the increase is strongly overestimated when using the other functionals. A similar situation occurs for the polyene linker, though in that case the $\beta [\mathbf{d}]/\beta [\mathbf{b}]$ ratio is close to 2.7.

Therefore, though they generally underestimates the hyperpolarizability values, the HF and DFT/LC-BLYP methods are suitable to describe the variations of β values induced by chemical modifications. On the other hand, the other DFT XC functionals provide for some compounds an accurate β estimate but the trends are poorly described.

TABLE V. β ratios at different levels of calculation between different molecules

6-31G(d)	$\frac{\beta[\mathbf{b}]}{\beta[\mathbf{a}]}$	$\frac{\beta[\mathbf{d}]}{\beta[\mathbf{c}]}$	$\frac{\beta[\mathbf{c}]}{\beta[\mathbf{a}]}$	$\frac{\beta[\mathbf{d}]}{\beta[\mathbf{b}]}$
HF	1.243	1.585	2.218	2.826
BLYP	1.144	1.167	3.291	3.358
BHandHLYP	1.185	1.461	2.600	3.205
B3LYP	1.106	1.245	3.023	3.402
LC-BLYP	1.439	1.757	2.222	2.713
B2-PLYP	1.289	1.574	2.705	3.302
mPW2-PLYP	1.294	1.617	2.618	3.271
MP2	1.826	2.415	2.090	2.764
CCSD(T)	1.706	/	2.202	/

6-31G+(d)	$\frac{\beta[\mathbf{b}]}{\beta[\mathbf{a}]}$	$\frac{\beta[\mathbf{d}]}{\beta[\mathbf{c}]}$	$\frac{\beta[\mathbf{c}]}{\beta[\mathbf{a}]}$	$\frac{\beta[\mathbf{d}]}{\beta[\mathbf{b}]}$
HF	1.257	1.617	2.160	2.778
BLYP	1.099	1.102	3.290	3.297
BHandHLYP	1.171	1.449	2.539	3.141
B3LYP	1.074	1.197	2.996	3.339
LC-BLYP	1.429	1.777	2.134	2.653
B2-PLYP	1.245	1.527	2.642	3.239
mPW2-PLYP	1.254	1.547	2.604	3.213
MP2	1.781	2.238	2.042	2.730
CCSD(T)	1.698	/	/	/

V.B.3 Further discussion, conclusion, and outlook

In the first part of this chapter, we have investigated the electron correlation effects on the first hyperpolarizabilities of a set of push-pull π -conjugated systems. As a preliminary step, basis set requirements have been tackled, showing that the 6-31+G(d) basis set gives the best balance between accuracy and computational resources for the compounds with alternating single and triple CC bonds whereas for polyene linkers, the 6-31G(d) basis set is already an optimal choice. Among the

molecular orbital-based methods to account for electron correlation, the MP2 method turns out to be the method of choice to predict the first hyperpolarizabilities of push-pull π -conjugated systems. Indeed, as a result of cancellations between higher-order contributions, it closely reproduces the values obtained with the CCSD(T) scheme, considered here as the reference method. In the case of compounds **b** and **f**, however, the SDQ-MP4 and CCSD approaches provide improved estimates over MP2 but the MP4 method does not represent an improvement over MP4-SDQ. Moreover, the use of a separate scaling of the parallel and anti-parallel components of the MP2 energy, as proposed in the SCS-MP2 method, improves over MP2 only in limited cases. Finally, DFT has been employed to provide a further assessment of the reliability of conventional and more recent exchange-correlation functionals. In particular, the LC-BLYP functional is reliable when characterizing the changes of first hyperpolarizability upon enlarging the π -conjugated linker or upon changing the polyyne linker into a polyene segment. Nevertheless, its reliability is very similar to what can be achieved with the Hartree-Fock method while the MP2 scheme is by far more accurate. On the other hand, the BLYP, B3LYP, and BHandHLYP functionals perform quantitatively better in a number of cases but the trends are poorly described. This is also the case of the B2-PLYP and mPW2-PLYP functionals, which are often the most accurate, but they underestimate the increase of β when going from polyyne to polyene linkers and they overestimate the enhancement of β with chain length.

We think that these conclusions – which are also supported by more sporadic results on other push-pull π -conjugated systems including molecular switches – can be extended to many D/A π -conjugated compounds having a potential for NLO applications. On the other hand, twisted tictoid merocyanine dyes, of which the structure results from the delicate balance between diradical and zwitterionic forms, belong to another category of NLO molecules, which remain a great

challenge for theoretical chemistry³⁶. Related systems are treated in the second part of this Chapter.

V.C. Electron correlation effects on the hyperpolarizabilities of p-quinodimethane derivatives

V.C.I. Theoretical and computational approaches

The geometrical structures were fully optimized with density functional theory (DFT) and the UB3LYP XC functional in combination with the 6-311G* basis set. The polarizabilities (α), first (β), and second (γ) hyperpolarizabilities were calculated using the finite field approach.³⁰ In all the FF calculations, the Romberg or Richardson method was employed in combination with selected field amplitudes [$\pm 2^k 0.0004$ or $\pm (\sqrt{2})^k 0.0004$ with k going from 0 to 5] so as to achieve a numerical accuracy within 1 % or less on the γ values while for β and α the accuracy attains at least 0.1 % and 0.01 %, respectively. The use of this method is particularly important for obtaining a stable γ value when large basis sets with many diffuse functions are employed. Tight convergence on the density matrices and on the energies (10^{-10} to 10^{-12} a.u.) was thus mandatory to meet the requested accuracy. The (hyper)polarizabilities were evaluated at different levels of approximation with unrestricted methods, including the HF method, the MP2, MP3, and MP4 approaches as well as SDQ-MP4, the CCSD(T) and CCSD methods. A range of basis sets was employed: the standard 6-31G, 6-31G(d), 6-311G(d), 6-31+G(d), 6-311++G(2df,p), aug-cc-pVDZ, and aug-cc-pVTZ basis set as well as the 6-31G(d) basis set augmented with one set of diffuse p functions [$\zeta = 0.0523, 0.0582$, and 0.0447 , for C, N, and B, respectively], which is denoted 6-31G(d)+ p . Note that in the standard 6-31+G(d) basis set, a set of sp diffuse functions is included, with $\zeta = 0.0438, 0.0639$, and 0.0315 , for C, N, and B, respectively.

The s -fold spin-projection scheme was also employed³⁵. It consists in removing the successive spin contamination contributions from the UHF and UMPn energies. At the UHF level, $s=6$ spin projections were carried out exactly whereas at the UMPn ($n = 2-4$) levels they were done in an approximate way up to $s=4$ following Ref. 36. No projection method was used for the coupled cluster energies since these methods are known to suffer less from spin contamination.³⁷ After the successive removals of the spin contamination contributions the S^2 expectation value drops to 0.001 or less, approaching the exact value for a singlet wavefunction.

Unrestricted DFT calculations with a selection of XC functionals were performed to probe the effect of the functional and particularly the role of HF exchange: BLYP, B3LYP, BHandHLYP, CAM-B3LYP, LC-BLYP ($\mu = 0.33$), LC-BLYP ($\mu = 0.47$), B98, and B2PLYP. BLYP, B3LYP, and BHandHLYP XC functionals. For LC-BLYP, the range-separating parameter μ is a calculation parameter that can be chosen. From comparison with CC results, in the case of diradical species, a $\mu = 0.3-0.5$ value is adequate for calculating γ of organic compounds like *p*-quinodimethane and pentalene whereas for dimetal compounds a larger $\mu = 0.7-0.9$ value is required.³⁸

The calculations were restricted to the diagonal longitudinal component, $\alpha_{zz} = \alpha$, $\beta_{zzz} = \beta$, $\gamma_{zzzz} = \gamma$ (with z the long axis), which are by far dominant in this particular case due to electron delocalization but for other systems, the relative magnitude of other γ components may vary considerably.³⁹ The T convention was used to define β and γ . All the calculations were performed with the Gaussian 09 package^{33b} as well as with homemade codes to perform the Romberg iterations.

V.C.2. Results and discussion

V.C.2.a. Wavefunction methods

The UMPn, UCC, and UDFT γ values calculated using different basis sets are presented in Figures 6-8 for compounds **1-3**, respectively. Additional values,

obtained at the PUHF, PUMP2, PUMP3, and PUMP4 levels with large basis sets are sketched in Figure 9. Table 1 lists the UHF/6-31+G(d) diradical character as well as those γ values determined at different levels of calculation in comparison to the reference γ values, obtained at the UCCSD(T) level with the aug-cc-pVDZ basis set.

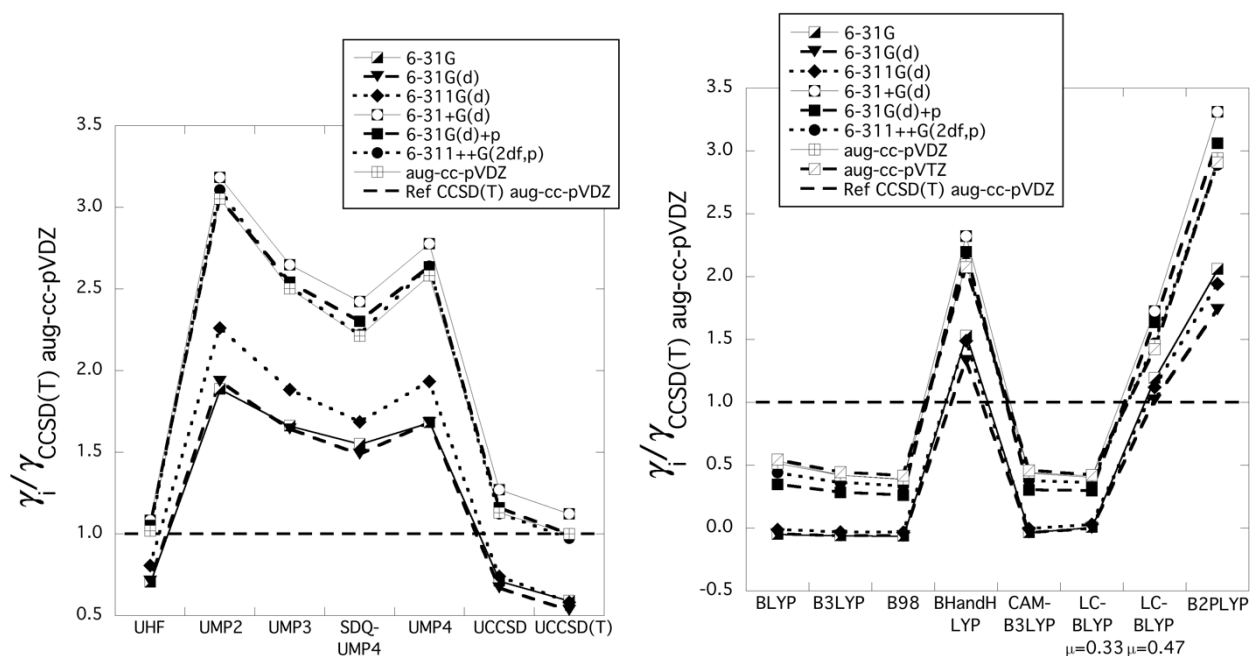


FIG. 6. Basis set and electron correlation effects on the second hyperpolarizability of molecule **1**. Lines are guides for the eyes. The left figure compares wavefunction correlated schemes to UCCSD(T) results as a function of the basis set whereas the right figure concentrates on DFT XC functionals.

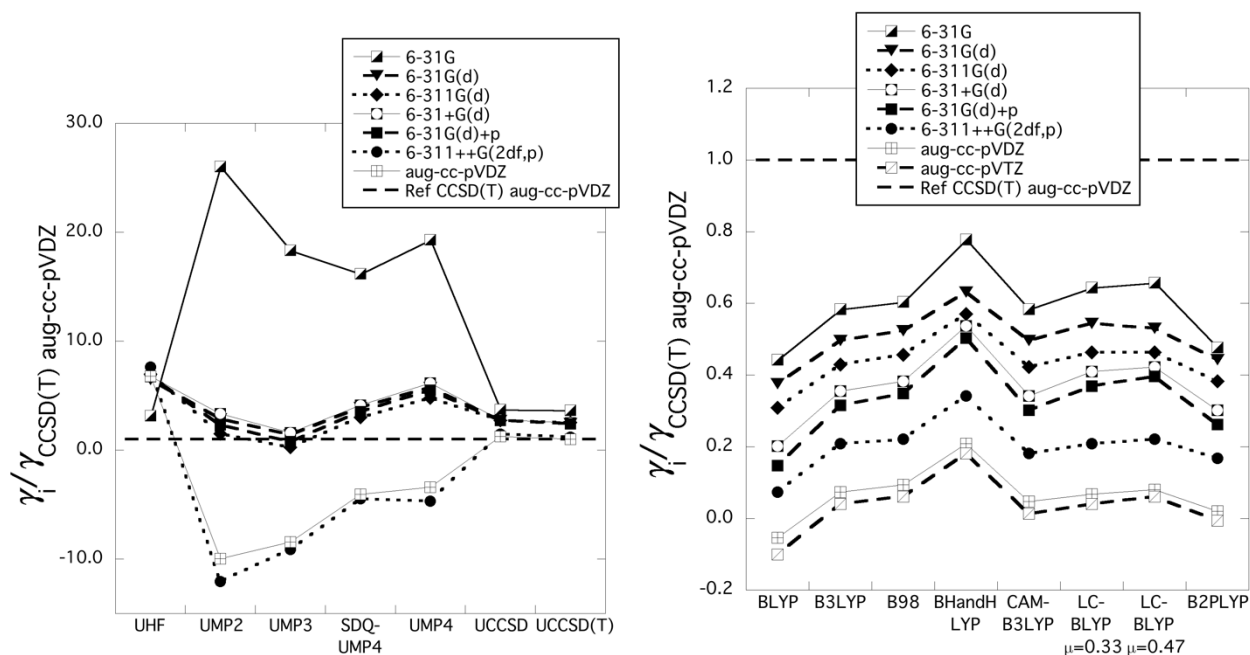


FIG. 7. Basis set and electron correlation effects on the second hyperpolarizability of molecule 2.

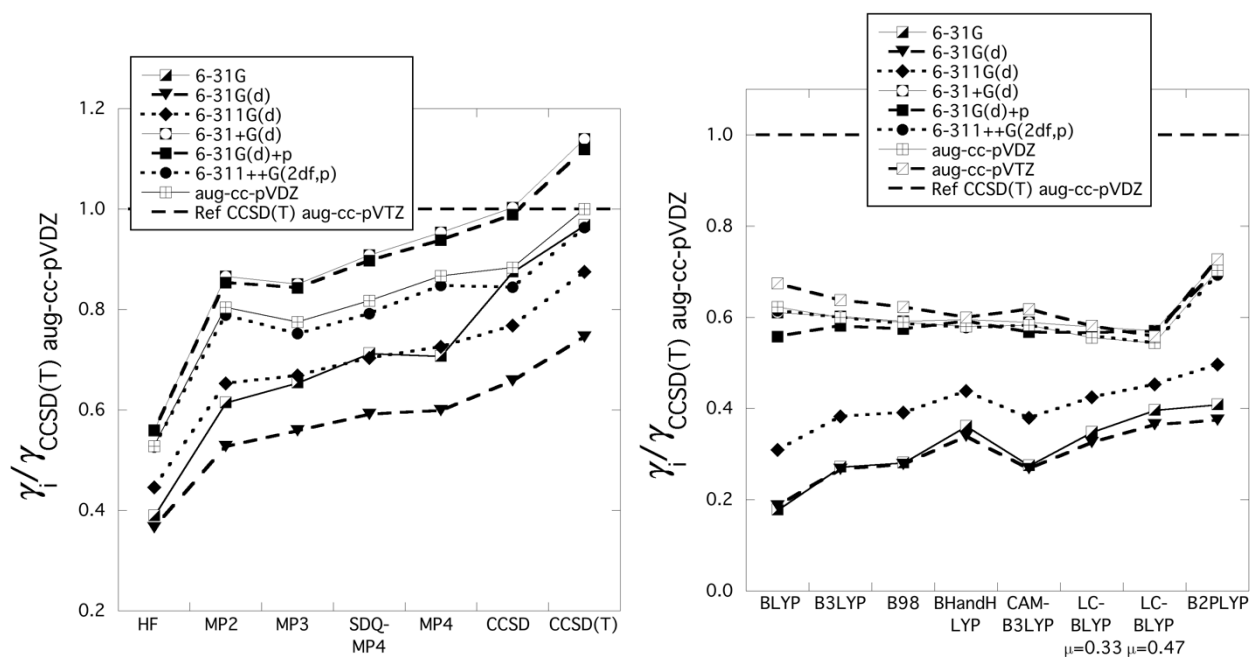


FIG. 8. Basis set and electron correlation effects on the second hyperpolarizability of molecule 3.

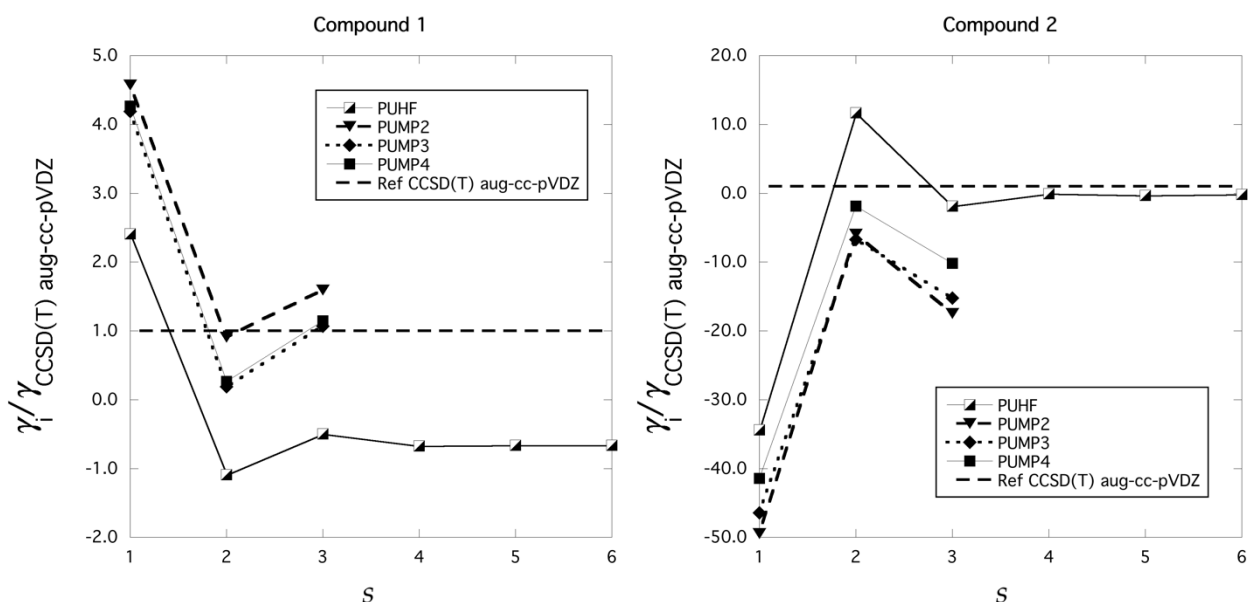


FIG.9. Spin-projected UHF, UMP2, UMP3, and UMP4 results obtained with the aug-cc-pVDZ basis set for compounds **1** and **2** as a function of the order of the spin projection (s).

TABLE VI: UHF/6-31+G(d) diradical characters and aug-cc-pVDZ γ values (in 10^2 a.u.) calculated at different levels of calculation for compounds **1-3**.

γ	1 (U)	2 (U)	3 (R)
γ	0.145	0.061	0.000
HF	1530	-1013	774
MP2	4580	1483	1179
MP3	3750	1255	1136
SDQ-MP4	3320	605	1198
MP4	3870	509	1271
CCSD	1690	-185	1300
CCSD(T)	1500	-149	1470
PUHF/ $s = 6$	-997	20	
PUMP2/ $s = 3$	2400	2600	
PUMP3/ $s = 3$	1600	2300	
PUMP4/ $s = 3$	1700	1500	

The second hyperpolarizability of **1** is underestimated when using basis sets which do not contain diffuse functions [6-31G, 6-31G(d), 6-311G(d)], whereas adding one set of diffuse p functions is sufficient to provide γ values in close agreement to those obtained with the aug-cc-pVDZ basis set. So, the difference between the 6-311++G(2df,p) and aug-cc-pVDZ γ values is negligible (a maximum of

3 %). Note that the 6-31G(d)+*p* basis set provides results in better agreement with aug-cc-pVDZ than 6-31+G(d), which is attributed to the different exponents of the diffuse basis functions. For this compound having a rather small diradical character (Table VI), electron correlation effects are negligible if noticing that the UHF values are close to the UCCSD(T) ones but the situation is more complex. Indeed, on the one hand, the UCCSD values are only 20% larger than the UCCSD(T) results and provide therefore a good estimate. Then, the UMP2, UMP3, SDQ-UMP4, and UMP4 schemes overestimate γ by more than 100%, highlighting the impact of spin contaminations. On the other hand, the PUMP2 method with $s=3$ (2400×10^2 a.u.) leads to a better agreement with the reference UCCSD(T)/aug-cc-pVDZ value (1500×10^2 a.u.) than the UMP2 method (4580×10^2 a.u.). The agreement gets even better at the PUMP3 and PUMP4 levels ($s=3$) with values of 1600×10^2 and 1700×10^2 a.u., respectively. Note also that the PUHF value (-997×10^2 a.u.) is far from the reference value, demonstrating the amplitude of electron correlation effects. Thus, for compound **1**, provided spin contamination is removed, the electron correlation contributions to γ converge quite rapidly with the order of perturbation theory.

Compound **2** is a dication with very small diradical character. It displays a negative UCCSD(T) γ value, about one order of magnitude smaller than compound **1**. When using too small basis sets, the UCCSD and UCCSD(T) γ amplitudes are overestimated by up to a factor of 2-3. Contrary to the case of compound **1** the addition of diffuse functions has a rather limited impact on γ , smaller than the effect of including polarization functions. Electron correlation effects as described by the unrestricted Møller-Plesset scheme are poorly estimated, leading to errors on both the sign and the magnitude of γ . Then, employing the corresponding spin projected methods does not improve the situation. Indeed, spin projection corrections typically lead to increases of the γ values by about 1000×10^2 a.u for $s=3$ and therefore the agreement with UCCSD(T) gets even worse. Only the UCCSD method

can approach the UCCSD(T) results, demonstrating that for this compound with a small γ value, the third-order response depends strongly on electron correlation. The “zwitterionic” compound **3** displays a closed-shell character and a positive γ value similar to that of **1**. The effects of basis sets are however very unusual. At the HF level, the γ value increases in the order 6-31G(d) < 6-31G < 6-311G(d) < 6-311++G(2df,p) < aug-cc-pVDZ < 6-31+G(d) < 6-31G(d)+p. At the MP2 level this order becomes 6-31G(d) < 6-31G < 6-311G(d) < 6-311++G(2df,p) < aug-cc-pVDZ < 6-31G(d)+p < 6-31+G(d) whereas at the CCSD(T) level it is 6-31G(d) < 6-311G(d) < 6-311++G(2df,p) < 6-31G < aug-cc-pVDZ < 6-31G(d)+p < 6-31+G(d). With the exception of the performance of the 6-31G basis set, which is too small, these results show the need for a balance between diffuse and polarization functions and the fact that the basis set effects are the same at most levels of approximation, as evidenced also by the almost parallel lines in Fig. 8. Contrary to the open-shell systems, the MP2 scheme already recovers 80% of the CCSD(T) γ response. The MP4 scheme performs slightly better [86% of CCSD(T) value] whereas the CCSD value differs by only 12%.

V.C.2.b. Density functional theory calculations

We now turn to the comparisons between DFT results obtained with different functionals and the reference UCCSD(T) values. Note that since the diradical character is small for compounds **1** and **2**, not all XC functionals do lead to unrestricted solutions. In fact, for compound **2** ($\gamma = 0.061$) all the solutions are restricted while for compound **1** ($\gamma = 0.145$) only BHandHLYP, LC-BLYP with $\mu=0.47$, and B2PLYP give unrestricted solutions.

From the comparisons sketched in Figs. 6-8, the major conclusion is the non-existence of a XC functional which is able to reproduce the UCCSD(T) γ values of the three compounds. In the case of compound **1**, those functionals with zero or small amount of HF exchange (BLYP, B3LYP, B98) as well as LC-BLYP($\mu = 0.33$) and CAM-

B3LYP predict too small values, off by at least a factor of 2 when using the aug-cc-pVDZ and aug-cc-pVTZ basis sets. On the other hand, the XC functionals, which lead to unrestricted solutions, LC-UBLYP($\mu = 0.47$) and UBHandHLYP, overestimate γ by about 50% and 100%, respectively whereas UB2PLYP leads to even larger values, consistently with the huge γ overestimation noted in Section V.C.2.a. at the UMP2 level. These DFT calculations further substantiate the conclusions drawn above on the basis set requirements, i.e. diffuse functions are important to describe the third-order field response of **1**. In the case of compound **2**, as there are no unrestricted solutions, all DFT results strongly underestimate γ and this underestimation grows when the basis set gets more flexible. Using the aug-cc-pVDZ basis set, BHandHLYP gives the closest γ value to UCCSD(T) but it is off by 79%. For compound **3**, all DFT XC functionals underestimate γ by 25 to 50%. The γ values closest to CCSD(T) are obtained with the B2PLYP XC functional where the value is 10% under the MP2 one. Then, provided the basis set is large enough, larger amounts of HF exchange are associated with smaller γ values, with a typical decrease by 15% from BLYP to LC-BLYP ($\mu = 0.47$).

The DFT calculations have also enabled to assess the quality of the aug-cc-pVDZ basis set with respect to aug-cc-pVTZ. Considering the results obtained with the BHandHLYP XC functional the aug-cc-pVDZ/aug-cc-pVTZ ratio amounts to 1.011, 1.148, and 0.966 for compounds **1**, **2** and **3**, respectively. The above analysis of the performance of DFT XC functionals to calculate γ of different compounds has to be contrasted with previous studies assessing whether DFT could qualitatively describe the change in γ as a function of the diradical character. In that case, it was found that the BHandHLYP XC functional reproduces the UCCSD(T) vs. y curve, though it underestimates γ , especially for large y values and slightly overestimates γ , for small y values whereas LC-BLYP ($\mu = 0.33$) exhibits a similar reliability. However,

improvement can be obtained by using an approximate spin projected scheme, as shown in the case of LC-BLYP ($\mu = 0.47$) calculations on *p*-quinodimethane.⁴⁰

The present set of compounds cover a broad range of *p*-quinodimethane derivatives with quite different responses, such as those that can be modeled by the Valence CI model described in Refs. 41 and 42. Indeed, the second hyperpolarizability does not depend solely on the diradical character but also on the charge. The largest response is attributed to both compounds **1** and **3**. **1** has an intermediate diradical character whereas **3** is closed shell, zwitterionic, and asymmetric. Then, the dicationic compound **2** has the smallest diradical character and also the smallest response. Both the charge and the diradical character appear to be important parameters to explain and more importantly to tune the γ response.

V.C.2.c. Polarizability and first hyperpolarizability

It is further interesting to analyze the performance of the different levels of approximation as well as of the projection corrections for the lower-order properties, α and β , the latter being different from zero only for compound **3**. The results are listed in Table VII. As observed in other studies, the higher the order of the electric field response is, the larger the effects of electron correlation and the spin contamination are. In the case of the polarizability, for compound **1**, the UMPn values are already in good agreement with the reference UCCSD(T) result but the agreement is further improved when removing spin contamination by using the projected schemes, in particular at second order. Contrary to its second hyperpolarizability, the polarizability of compound **2** is little modified by electron correlation. In addition, the unrestricted MPn values are already in good agreement with UCCSD(T) and removing the spin contamination does not lead to any noticeable improvement. Similarly to compound **2**, the CCSD(T) polarizability of compound **3** are accurately estimated at the MP2 level and further inclusion of electron

correlation has a negligible impact. The error introduced by using (U)CCSD instead of (U)CCSD(T) amounts to -2.2, -0.8, and -2.6 %, for compounds **1-3**, respectively.

Looking now at the first hyperpolarizability of compound **3**, its CCSD(T)/aug-cc-pVDZ value attains -2300 a.u. Again, like for α and γ , the MP2 level is already suitable to provide a good estimate of β , with an error of 16 % whereas at the MP4 level this error reduces to 9 %, like at the CCSD level.

In summary, the spin projected technique provides improvements for the polarizability of the neutral compound **1** whereas for compound **2** the spin projection corrections are negligible. Finally, the MP2 method provides good estimates for α and β of compound **3**.

TABLE VII: α and β values calculated at different levels of calculation with the aug-cc-pVDZ basis set. All values are given in a.u.

	α			β
	1 (U)	2 (U)	3 (R)	3 (R)
HF	143	142	142	-1340
MP2	213	138	158	-1930
MP3	198	137	154	-1880
SDQ-MP4	194	137	156	-1980
MP4	205	140	160	-2100
CCSD	185	135	155	-2100
CCSD(T)	189	136	159	-2300
PUHF/s = 6	114	127		
PUMP2/s = 4	184	124		
PUMP3/s = 4	167	124		
PUMP4/s = 4	174	127		

V.C.3. Conclusions and outlook

The second hyperpolarizability of *p*-quinodimethane derivatives has been calculated at different levels of approximation in combination with a large range of basis sets. The main conclusions of this second part are i) the CCSD method provides results in the closest agreement with the CCSD(T) reference values with CCSD/CCSD(T) ratios amounting to 0.85, 1.26, and 0.88 for compounds **1-3**,

respectively, ii) the MP2 level of approximation performs well for the closed-shell compound **3** but not for the open-shell systems **1** and **2**, iii) for compound **1**, the projected UMP3 and UMP4 methods enable to reproduce the unrestricted CCSD(T) value but not for the charged compound **2** where high-order electron correlation effects are needed to predict the sign of γ , and iv) without a spin projection correction, density functional theory with a broad range of XC functionals does not perform well for these systems with low diradical character, in particular for the charged species. Qualitatively, similar effects have been observed for the lower order properties (polarizability and first hyperpolarizability), though the effects are smaller.

REFERENCES

1. (a) Kanis, D. R.; Ratner, M. A.; Marks, T. J., *Chem. Rev.* **1994**, *94* (1), 195-242; (b) Brédas, J. L.; Adant, C.; Tackx, P.; Persoons, A.; Pierce, B. M., *Chem. Rev.* **1994**, *94* (1), 243-278; (c) Kirtman, B., Theoretical and computational modeling of nlo and electronic materials. In *Theoretical and computational modeling of nlo and electronic materials*, Karna, S. P.; Yeates, A. T., Eds. ACS Series: 1996; Vol. 628, p 58; (d) Kirtman, B.; Champagne, B., *Int. Rev. Phys. Chem.* **1997**, *16* (4), 389-420; (e) Bishop, D. M.; Norman, P., Nonlinear optical materials. In *Handbook of advanced electronic and photonic materials and devices* Nalwa, H. S., Ed. Academic Press: San Diego, 2001; Vol. 9; (f) Champagne, B.; Kirtman, B., Nonlinear optical materials. In *Handbook of advanced electronic and photonic materials and devices* Nalwa, H. S., Ed. Academic Press: San Diego, 2001; Vol. 9, p 63; (g) Suponitsky, K. Y.; Timofeeva, T. V.; Antipin, M. Y., *Russ. Chem. Rev.* **2006**, *75* (6), 457-496.
2. (a) Bishop, D. M.; Kirtman, B., *J. Chem. Phys.* **1991**, *95* (4), 2646; (b) Luis, J. M.; Duran, M.; Andrés, J. L.; Champagne, B.; Kirtman, B., *J. Chem. Phys.* **1999**, *111* (3), 875; (c) Kirtman, B.; Champagne, B.; Bishop, D. M., *J. Am. Chem. Soc.* **2000**, *122* (33), 8007-8012; (d) Ingamells, V. E.; Papadopoulos, M. G.; Sadlej, A. J., *J. Chem. Phys.* **2000**, *112* (4), 1645; (e) Luis, J. M.; Duran, M.; Champagne, B.; Kirtman, B., *J. Chem. Phys.* **2000**, *113* (13), 5203; (f) Torrent-Sucarrat, M.; Sola, M.; Duran, M.; Luis, J. M.; Kirtman, B., *J. Chem. Phys.* **2004**, *120* (14), 6346-55; (g) Luis, J. M.; Torrent-Sucarrat, M.; Christiansen, O.; Kirtman, B., *J. Chem. Phys.* **2007**, *127* (8), 084118; (h) Kirtman, B.; Luis, J. M., *J. Chem. Phys.* **2008**, *128* (11), 114101.
3. (a) Bishop, D. M., *Int. Rev. Phys. Chem.* **1994**, *13* (1), 21-39; (b) Hamada, T., *J. Chem. Soc., Faraday Transactions* **1996**, *92* (17), 3165; (c) Malagoli, M.; Munn, R. W., *J. Chem. Phys.* **1997**, *107* (19), 7926; (d) Tomasi, J.; Cammi, R.; Mennucci, B., *Int. J. Quant. Chem.* **1999**, *75* (4-5), 783-803; (e) Kirtman, B.,

Dykstra, C. E.; Champagne, B., *Chem. Phys. Lett.* **1999**, *305* (1-2), 132-138; (f) Luo, Y.; Norman, P.; Macak, P.; Agren, H., *Phys. Rev. B* **2000**, *61*, 3060-3066; (g) Reis, H.; Papadopoulos, M. G.; Hättig, C.; Ángyán, J. G.; Munn, R. W., *J. Chem. Phys.* **2000**, *112* (14), 6161; (h) Panhuis, M. i. h.; Munn, R. W., *J. Chem. Phys.* **2000**, *113* (23), 10685-10690; (i) Bishop, D. M.; Gu, F. L., *Chem. Phys. Lett.* **2000**, *317* (3-5), 322-329; (j) Castet, F.; Champagne, B., *J. Phys. Chem. A* **2001**, *105* (8), 1366-1370; (k) Jansik, B.; Schimmelpfennig, B.; Norman, P.; Mochizuki, Y.; Luo, Y.; Ågren, H., *J. Phys. Chem. A* **2002**, *106* (2), 395-399; (l) Champagne, B.; Bishop, D. M., *Adv. Chem. Phys.* **2003**, *126*, 41-92; (m) Jensen, L.; Sylvester-Hvid, K. O.; Mikkelsen, K. V.; Åstrand, P.-O., *J. Phys. Chem. A* **2003**, *107* (13), 2270-2276; (n) Guillaume, M.; Botek, E.; Champagne, B.; Castet, F.; Ducasse, L., *J. Chem. Phys.* **2004**, *121* (15), 7390-400; (o) Reis, H.; Grzybowski, A.; Papadopoulos, M. G., *J. Phys. Chem. A* **2005**, *109* (44), 10106-20; (p) Ferrighi, L.; Frediani, L.; Cappelli, C.; Salek, P.; Ågren, H.; Helgaker, T.; Ruud, K., *Chem. Phys. Lett.* **2006**, *425* (4-6), 267-272; (q) Nielsen, C. B.; Christiansen, O.; Mikkelsen, K. V.; Kongsted, J., *J. Chem. Phys.* **2007**, *126* (15), 154112; (r) Holtmann, J.; Walczuk, E.; Dede, M.; Wittenburg, C.; Heck, J.; Archetti, G.; Wortmann, R.; Kuball, H. G.; Wang, Y. H.; Liu, K.; Luo, Y., *J. Phys. Chem. B* **2008**, *112* (47), 14751-61; (s) Olejniczak, M.; Pecul, M.; Champagne, B.; Botek, E., *J. Chem. Phys.* **2008**, *128* (24), 244713; (t) Ferrero, M.; Civalleri, B.; Rerat, M.; Orlando, R.; Dovesi, R., *J. Chem. Phys.* **2009**, *131* (21), 214704; (u) Hermet, P.; Ghosez, P., *Phys. Chem. Chem. Phys.* **2010**, *12* (4), 835-43.

4. (a) Bartlett, R. J.; Purvis, G. D., III, *Phys. Rev. A* **1979**, *20*, 1313-22; (b) Purvis, G. D., III; Bartlett, R. J., *Phys. Rev. A* **1981**, *23*, 1594-9; (c) Maroulis, G., *J. Chem. Phys.* **1991**, *94* (2), 1182; (d) Maroulis, G.; Thakkar, A. J., *J. Chem. Phys.* **1991**, *95* (12), 9060; (e) Sekino, H.; Bartlett, R. J., *J. Chem. Phys.* **1993**, *98* (4), 3022; (f) Sim, F.; Chin, S.; Dupuis, M.; Rice, J. E., *J. Phys. Chem.* **1993**, *97* (6), 1158-1163; (g) Archibong, E. F.; Thakkar, A. J., *J. Chem. Phys.* **1994**, *100* (10), 7471;

- (h) Mikkelsen, K. V.; Luo, Y.; Ågren, H.; Jørgensen, P., *J. Chem. Phys.* **1994**, *100* (11), 8240; (i) Aiga, F.; Itoh, R., *Chem. Phys. Lett.* **1996**, *251* (5-6), 372-380; (j) Christiansen, O.; Jørgensen, P.; Hättig, C., *Int. J. Quant. Chem.* **1998**, *68* (1), 1-52.
5. (a) Xenides, D.; Maroulis, G., *Chem. Phys. Lett.* **2000**, *319* (5-6), 618-624; (b) Karamanis, P.; Maroulis, G., *Chem. Phys. Lett.* **2003**, *376* (3-4), 403-410; (c) Pecul, M., *Chem. Phys. Lett.* **2005**, *404* (1-3), 217-226; (d) Sałek, P.; Helgaker, T.; Vahtras, O.; Ågren, H.; Jonsson, D.; Gauss, J., *Mol. Phys.* **2005**, *103* (2-3), 439-450; (e) Benková, Z.; Černušák, I.; Zahradník, P., *Int. J. Quant. Chem.* **2007**, *107* (11), 2133-2152; (f) Pecul, M.; Pawlowski, F.; Jørgensen, P.; Kohn, A.; Hättig, C., *J. Chem. Phys.* **2006**, *124* (11), 114101; (g) O'Neill, D. P.; Kallay, M.; Gauss, J., *J. Chem. Phys.* **2007**, *127* (13), 134109; (h) Maroulis, G., *Chem. Phys. Lett.* **2007**, *442* (4-6), 265-269; (i) Maroulis, G., *J. Chem. Phys.* **2008**, *129* (4), 044314; (j) Karamanis, P.; Pouchan, C.; Leszczynski, J., *J. Phys. Chem. A* **2008**, *112* (51), 13662-71; (k) Holka, F.; Avramopoulos, A.; Loboda, O.; Kellö, V.; Papadopoulos, M. G., *Chem. Phys. Lett.* **2009**, *472* (4-6), 185-189; (l) Sun, S.-L.; Yang, G.-C.; Qin, C.-S.; Qiu, Y.-Q.; Yan, L.-K.; Su, Z.-M.; Wang, R.-S., *Int. J. Quant. Chem.* **2009**, *110*, 865-873.
6. (a) Toto, T. T.; Toto, J. L.; de Melo, C. P.; Hasan, M.; Kirtman, B., *Chem. Phys. Lett.* **1995**, *244* (1-2), 59-64; (b) Jacquemin, D.; Champagne, B.; André, J.-M., *Int. J. Quant. Chem.* **1997**, *65* (5), 679-688; (c) Champagne, B.; Spassova, M., *Chem. Phys. Lett.* **2009**, *471* (1-3), 111-115.
7. (a) Nakano, M.; Kishi, R.; Nitta, T.; Kubo, T.; Nakasuji, K.; Kamada, K.; Ohta, K.; Champagne, B.; Botek, E.; Yamaguchi, K., *J. Phys. Chem. A* **2005**, *109* (5), 885-91; (b) Champagne, B.; Botek, E.; Nakano, M.; Nitta, T.; Yamaguchi, K., *J. Chem. Phys.* **2005**, *122* (11), 114315; (c) Kishi, R.; Nakano, M.; Ohta, S.; Takebe, A.; Nate, M.; Takahashi, H.; Kubo, T.; Kamada, K.; Ohta, K.

- Champagne, B.; Botek, E., *J. Chem. Theo. Comp.* **2007**, *3* (5), 1699-1707; (d) Jha, P. C.; Rinkevicius, Z.; Agren, H., *Chem. Phys. Chem.* **2009**, *10* (5), 817-23.
8. (a) Oliveira, L. N.; Amaral, O. A. V.; Castro, M. A.; Fonseca, T. L., *Chem. Phys.* **2003**, *289* (2-3), 221-230; (b) Chen, W.; Li, Z. R.; Wu, D.; Li, Y.; Sun, C. C.; Gu, F. L., *J. Am. Chem. Soc.* **2005**, *127* (31), 10977-81; (c) Spassova, M.; Champagne, B.; Kirtman, B., *Chem. Phys. Lett.* **2005**, *412* (1-3), 217-222; (d) Ma, F.; Li, Z. R.; Xu, H. L.; Li, Z. J.; Li, Z. S.; Aoki, Y.; Gu, F. L., *J. Phys. Chem. A* **2008**, *112* (45), 11462-7; (e) Li, Z. J.; Wang, F. F.; Li, Z. R.; Xu, H. L.; Huang, X. R.; Wu, D.; Chen, W.; Yu, G. T.; Gu, F. L.; Aoki, Y., *Phys. Chem. Chem. Phys.* **2009**, *11* (2), 402-8; (f) Xu, H. L.; Wang, F. F.; Li, Z. R.; Wang, B. Q.; Wu, D.; Chen, W.; Yu, G. T.; Gu, F. L.; Aoki, Y., *J. Comput. Chem.* **2009**, *30* (7), 1128-34.
 9. (a) Castet, F.; Bogdan, E.; Plaquet, A.; Ducasse, L.; Champagne, B.; Rodriguez, V., *J. Chem. Phys.* **2012**, *136* (2), 024506; (b) Hanauer, M.; Kohn, A., *J. Chem. Phys.* **2009**, *131* (12), 124118; (c) Limacher, P. A.; Li, Q.; Luthi, H. P., *J. Chem. Phys.* **2011**, *135* (1), 014111; (d) Nakano, M.; Minami, T.; Fukui, H.; Kishi, R.; Shigeta, Y.; Champagne, B., *J. Chem. Phys.* **2012**, *136* (2), 024315; (e) Robinson, J. B.; Knowles, P. J., *J. Chem. Phys.* **2012**, *137* (5), 054301.
 10. (a) Hammond, J. R.; Kowalski, K., *J. Chem. Phys.* **2009**, *130* (19), 194108; (b) Sponitsky, K. Y.; Tafur, S.; Masunov, A. E., *J. Chem. Phys.* **2008**, *129* (4), 044109.
 11. Serrano-Andres, L.; Avramopoulos, A.; Li, J.; Labéguerie, P.; Bégue, D.; Kello, V.; Papadopoulos, M. G., *J. Chem. Phys.* **2009**, *131* (13), 134312.
 12. Castet, F.; Champagne, B., *J. Chem. Theor. Comput.* **2012**, *8* (6), 2044-2052.
 13. (a) Champagne, B.; Perpète, E. A.; Jacquemin, D.; van Gisbergen, S. J. A.; Baerends, E.-J.; Soubra-Ghaoui, C.; Robins, K. A.; Kirtman, B., *J. Phys. Chem. A* **2000**, *104* (20), 4755-4763; (b) van Gisbergen, S. J. A.; Schipper, P. R. T.; Gritsenko, O. V.; Baerends, E. J.; Snijders, J. G.; Champagne, B.; Kirtman, B., *Phys. Rev. Lett.* **1999**, *83* (4), 694-697.

14. (a) Vila, F. D.; Strubbe, D. A.; Takimoto, Y.; Andrade, X.; Rubio, A.; Louie, S. G.; Rehr, J. J., *J. Chem. Phys.* **2010**, *133* (3), 034111; (b) Yamada, S.; Nakano, M.; Shigemoto, I.; Yamaguchi, K., *Chem. Phys. Lett.* **1996**, *254* (3-4), 158-161.
15. (a) Nakano, M.; Fujita, H.; Takahata, M.; Yamaguchi, K., *J. Am. Chem. Soc.* **2002**, *124* (32), 9648-9655; (b) Springborg, M.; Tevekeliyska, V.; Kirtman, B., *Phys. Rev. B* **2010**, *82* (16), 165442.
16. (a) Sanguinet, L.; Pozzo, J. L.; Rodriguez, V.; Adamietz, F.; Castet, F.; Ducasse, L.; Champagne, B., *J. Phys. Chem. B* **2005**, *109* (22), 11139-50; (b) Plaquet, A.; Guillaume, M.; Champagne, B.; Rougier, L.; Mancois, F.; Rodriguez, V.; Pozzo, J. L.; Ducasse, L.; Castet, F., *J. Phys. Chem. C* **2008**, *112* (14), 5638-5645; (c) Bossi, A.; Licandro, E.; Maiorana, S.; Rigamonti, C.; Righetto, S.; Stephenson, G. R.; Spassova, M.; Botek, E.; Champagne, B., *J. Phys. Chem. C* **2008**, *112* (21), 7900-7907; (d) Bogdan, E.; Plaquet, A. I.; Antonov, L.; Rodriguez, V.; Ducasse, L.; Champagne, B.; Castet, F. d. r., *J. Phys. Chem. C* **2010**, *114* (29), 12760-12768; (e) Sergeyev, S.; Didier, D.; Boitsov, V.; Teshome, A.; Asselberghs, I.; Clays, K.; Vande Velde, C. M.; Plaquet, A.; Champagne, B., *Chemistry* **2010**, *16* (27), 8181-90.
17. Champagne, B.; Kirtman, B., *J. Chem. Phys.* **2006**, *125* (2), 24101.
18. Bulat, F. A.; Toro-Labbe, A.; Champagne, B.; Kirtman, B.; Yang, W., *J. Chem. Phys.* **2005**, *123* (1), 014319.
19. (a) Grimme, S., *J. Chem. Phys.* **2006**, *124* (3), 034108; (b) Neese, F.; Schwabe, T.; Grimme, S., *J. Chem. Phys.* **2007**, *126* (12), 124115.
20. (a) Biczysko, M.; Panek, P.; Scalmani, G.; Bloino, J.; Barone, V., *J. Chem. Theo. Comp.* **2010**, *6* (7), 2115-2125; (b) Schwabe, T.; Grimme, S., *J. Phys. Chem. A* **2009**, *113* (13), 3005-8.
21. Grimme, S.; Neese, F., *J. Chem. Phys.* **2007**, *127* (15), 154116.
22. (a) Grimme, S.; Steinmetz, M.; Korth, M., *J. Chem. Theo. Comp.* **2007**, *3* (1), 42-45; (b) Wennmohs, F.; Neese, F., *Chem. Phys.* **2008**, *343*, 217-230; (c) Römelt,

- M.; Ye, S.; Neese, F., *Inorg. Chem.* **2008**, *48* (3), 784-785; (d) Seal, P.; Chakrabarti, S., *J. Phys. Chem. A* **2009**, *113* (7), 1377-83; (e) Grimme, S.; Djukic, J. P., *Inorg. Chem.* **2010**, *49* (6), 2911-9.
23. Schwabe, T.; Grimme, S., *Phys. Chem. Chem. Phys.* **2006**, *8* (38), 4398-401.
 24. Menon, A. S.; Wood, G. P.; Moran, D.; Radom, L., *J. Phys. Chem. A* **2007**, *111* (51), 13638-44.
 25. Grimme, S., *J. Chem. Phys.* **2003**, *118* (20), 9095.
 26. (a) Gerenkamp, M.; Grimme, S., *Chem. Phys. Lett.* **2004**, *392* (1-3), 229-235; (b) Hyla-Kryspin, I.; Grimme, S., *Organometallics* **2004**, *23* (23), 5581-5592; (c) Piacenza, M.; Grimme, S., *J. Comput. Chem.* **2004**, *25* (1), 83-99; (d) Parac, M.; Etinski, M.; Peric, M.; Grimme, S., *J. Chem. Theo. Comp.* **2005**, *1* (6), 1110-1118; (e) Mori, T.; Inoue, Y.; Grimme, S., *J. Org. Chem.* **2006**, *71* (26), 9797-806; (f) Distasio Jr, R. A.; Head-Gordon, M., *Mol. Phys.* **2007**, *105* (8), 1073-1083; (g) Sancho-Garcia, J. C.; Perez-Jimenez, A. J., *Phys. Chem. Chem. Phys.* **2007**, *9* (44), 5874-9; (h) Bachorz, R. A.; Bischoff, F. A.; Hofener, S.; Klopper, W.; Ottiger, P.; Leist, R.; Frey, J. A.; Leutwyler, S., *Phys. Chem. Chem. Phys.* **2008**, *10* (19), 2758-66; (i) Goerigk, L.; Grimme, S., *Chem. Phys. Chem.* **2008**, *9* (17), 2467-70; (j) Pluhackova, K.; Grimme, S.; Hobza, P., *J. Phys. Chem. A* **2008**, *112* (48), 12469-74; (k) Bettinger, H. F.; Kar, T.; Sanchez-Garcia, E., *J. Phys. Chem. A* **2009**, *113* (14), 3353-9; (l) Schoenebeck, F.; Ess, D. H.; Jones, G. O.; Houk, K. N., *J. Am. Chem. Soc.* **2009**, *131* (23), 8121-33; (m) Cuypers, R.; Sudholter, E. J.; Zuilhof, H., *Chem. Phys. Chem.* **2010**, *11* (10), 2230-40; (n) Geng, Y.; Takatani, T.; Hohenstein, E. G.; Sherrill, C. D., *J. Phys. Chem. A* **2010**, *114* (10), 3576-82; (o) Huenerbein, R.; Schirmer, B.; Moellmann, J.; Grimme, S., *Phys. Chem. Chem. Phys.* **2010**, *12* (26), 6940-8; (p) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D., *J. Chem. Phys.* **2010**, *132* (14), 144104.
 27. Champagne, B.; Kirtman, B., *Int. J. Quant. Chem.* **2009**, *109* (13), 3103-3111.

28. (a) Maroulis, G.; Karamanis, P.; Pouchan, C., *J. Chem. Phys.* **2007**, *126* (15), 154316-5; (b) Maroulis, G., *Theo. Chem. Acc.* **2010**, *129* (3-5), 437-445.
29. (a) Lambert, C., *Angew. Chem. Int. Ed. Engl.* **2011**, *50* (8), 1756-8; (b) Sun, Z.; Wu, J., *J. Mater. Chem.* **2012**, *22* (10), 4151-4160; (c) Plasser, F.; Pasalic, H.; Gerzabek, M. H.; Libisch, F.; Reiter, R.; Burgdorfer, J.; Muller, T.; Shepard, R.; Lischka, H., *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (9), 2581-4.
30. Cohen, H. D.; Roothaan, C. C. J., *J. Chem. Phys.* **1965**, *43* (10), S34-S39.
31. (a) Benkova, Z.; Sadlej, A. J.; Oakes, R. E.; Bell, S. E. J., *Theo. Chem. Acc.* **2005**, *113* (4), 238-247; (b) Benkova, Z.; Sadlej, A. J.; Oakes, R. E.; Bell, S. E., *J. Comput. Chem.* **2005**, *26* (2), 145-53.
32. (a) Sadlej, A. J., *Collect. Czech. Chem. Commun.* **1988**, *53*, 1995-2016; (b) Sadlej, A. J., *Theor. Chim. Acta* **1991**, *79*, 123-40; (c) Černušák, I.; Kellö, V.; Sadlej, A. J., *Collect. Czech. Chem. Commun.* **2003**, *68* (2), 211-239.
33. (a) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Laham, A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A., *Gaussian 03*, revision c.02. 2003; (b) Frisch, M. J.; Trucks, G. W.; Schlegel, H.

- B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, revision a.01. Wallingford CT, 2009.
34. (a) Bourhill, G.; Bredas, J.-L.; Cheng, L.-T.; Marder, S. R.; Meyers, F.; Perry, J. W.; Tiemann, B. G., *J. Am. Chem. Soc.* **1994**, *116* (6), 2619-2620; (b) Marder, S. R.; Cheng, L. T.; Tiemann, B. G.; Friedli, A. C.; Blanchard-Desce, M.; Perry, J. W.; Skindhoj, J., *Science* **1994**, *263*, 511-14.
 35. Löwdin, P.-O., *Phys. Rev.* **1955**, *97* (6), 1509-1520.
 36. Schlegel, H. B., *J. Chem. Phys.* **1986**, *84* (8), 4530-4534.
 37. Stanton, J. F., *J. Chem. Phys.* **1994**, *101* (1), 371-374.
 38. (a) Bonness, S.; Fukui, H.; Yoneda, K.; Kishi, R.; Champagne, B.; Botek, E.; Nakano, M., *Chem. Phys. Lett.* **2010**, *493* (1-3), 195-199; (b) Fukui, H.; Inoue, Y.; Kishi, R.; Shigeta, Y.; Champagne, B.; Nakano, M., *Chem. Phys. Lett.* **2012**, *523*, 60-64.
 39. Maroulis, G., *Chem. Phys. Lett.* **2011**, *505* (1-3), 5-10.
 40. Nakano, M.; Minami, T.; Fukui, H.; Yoneda, K.; Shigeta, Y.; Kishi, R.; Champagne, B.; Botek, E., *Chem. Phys. Lett.* **2010**, *501* (1-3), 140-145.

41. Nakano, M.; Kishi, R.; Ohta, S.; Takahashi, H.; Kubo, T.; Kamada, K.; Ohta, K.; Botek, E.; Champagne, B., *Phys. Rev. Lett.* **2007**, *99* (3), 033001.
42. Nakano, M.; Champagne, B.; Botek, E.; Ohta, K.; Kamada, K.; Kubo, T., *J. Chem. Phys.* **2010**, *133* (15), 154302.

VI. Conclusions and outlooks

This thesis has focused on the characterization and the interpretation of the nonlinear optical responses of complex systems by quantum chemistry methods. The main goal was to reduce the gap between real systems characterized experimentally and model systems usually employed in simulations and therefore to introduce into the simulations the structural complexity of the studied compounds. The main targets were biological structures useful for second harmonic generation microscopy (SHIM), i.e. fluorescent proteins and collagen, of which the responses have different origins. The NLO response of fluorescent proteins originates mostly from a push-pull π -conjugated chromophore whereas, for collagen, small π -conjugated peptide bonds are at the origin of the NLO response.

Successive investigations on fluorescent proteins have been conducted. From a first set of fluorescent proteins (eGFP, eYFP, and DsRed), evaluated for their potential use in second-harmonic imaging, it was emphasized that the amplitude of the second-order NLO response does not *necessarily* follow the amplitude of the linear one. Indeed, on the one hand, in DsRed a strong red shift of the absorption band maximum and a larger first hyperpolarizability with respect to eGFP are the consequences of a longer π -conjugated pathway. On the other hand, in eYFP, the red shift of the absorption band maximum is not associated with an increase of the first hyperpolarizability. This has been attributed to the centrosymmetric π -stacking interaction in eYFP, which is detrimental to second-order NLO properties. In a second step, the same approach has been adopted to characterize two additional fluorescent proteins, mStrawberry and mCherry, demonstrating larger first hyperpolarizability when the conjugated π -system is larger, making mCherry a suitable chromophore for SHG microscopy. In a third step, a new yellow fluorescent protein, SHardonnay, has been successfully designed by our colleagues from the KULeuven in order to find an alternative to eYFP. In SHardonnay, Tyr203 is replaced

by Phe203. This small structural modification results in minor change of the linear optical response but a substantial enhancement of the second-order NLO response because it destroys the centrosymmetric π -stacking interaction. This study has confirmed the importance of the noncentrosymmetry requirement for the second-order nonlinear optical properties and the sensitivity of this principle. To interpret these results, theoretical simulations carried out at different levels of approximation have been accompanied these design investigations. In particular, different applications of the ONIOM scheme have been assessed in comparison with experiment, demonstrating the need for an explicit description of the close surrounding of the chromophore and a precise representation of the electron correlation.

In the investigation on collagen, a short peptide triple-helix model [(Pro-Pro-Gly)₁₀]₃ (PPG10) has been considered and its first hyperpolarizability has been calculated using the ONIOM method in combination with time-dependent Hartree-Fock and DFT calculations. Detailed analysis of the first hyperpolarizability tensors (tensor components, dipolar and octupolar contributions, depolarization ratios) has been performed to unravel the origin of the β responses and to describe how the three intertwined helices participate to the total β . In particular, it was found that each peptide bond contributes to the total β responses and that the triple-helix structure determines the relative orientation of these bonds. So, the total response is obtained, to a good extent, by summing the β tensors of each bond. This has contributed to propose a molecular interpretation of the SHIM pictures of collagen tissues based in (rat-tail tendon, for example) on its fibers flexibility.

The choice of quantum chemistry methods to describe the NLO responses of fluorescent proteins and collagen has been substantiated by investigations of the impact of electron correlation on the hyperpolarizabilities of small model systems. First, push-pull π -conjugated systems have been considered, motivated by the fact that such compounds are model structures for the chromophores of fluorescent

proteins. This study has shown that the 6-31+G(d) basis set gives the best balance between accuracy and computational resources for triply-bonded π -conjugated systems whereas the 6-31G(d) basis set is already an optimal choice for the doubly-bonded systems. Moreover, the MP2 method turns out to be the method of choice to predict the first hyperpolarizabilities of push-pull π -conjugated systems. Among DFT exchange-correlation functionals, LC-BLYP is the most reliable for that type of systems. Nevertheless, its reliability is very similar to what can be achieved with the Hartree-Fock method while the MP2 scheme is by far more accurate. Open-shell systems such as the *p*-quinodimethane and two of its derivatives have also been used to further describe the impact of electron correlation. The principal conclusions on the determination of the second hyperpolarizability are i) the CCSD method provides results in the closest agreement with the CCSD(T) reference values, ii) the MP2 level of approximation still performs well for the closed-shell compounds but not for the open-shell systems, and iii) without spin projection correction, density functional theory with a broad range of exchange-correlation functionals is not reliable. Qualitatively, similar conclusions can be drawn for the lower order properties.

In parallel, several technical aspects associated with the evaluation of hyperpolarizabilities have been addressed, *i.e.* i) the automatization of the Romberg scheme to improve the numerical accuracy, in the finite field method, on the electric field derivatives of the energy, ii) the evaluation of the frequency dispersion at the correlated level by approximated schemes, and iii) the deviations from Kleinman's conditions. So, a simple two-state approximation has been shown to be preferable over the additive and multiplicative schemes to describe frequency-dispersion of the first hyperpolarizability.

The direct follow-up of this thesis is to apply our gained expertise to modelling of other biological structures useful for SHIM. Indeed, as mentioned in the

introduction, the emitters are classified as exogenous or endogenous. An appealing area to investigate concerns the exogenous dyes for labeling membranes. The molecular requirements for these dyes have been quite extensively discussed in the introduction of this thesis. Several studies proposed the use of styryl dyes as di-4-ANEPPS¹ or di-8-ANEPPS^{1c, 1f, 2}. These dyes have been used for voltage sensitive imaging, where the probes are sensitive to the transmembrane voltage. For example, di-4-ANEPPS, a commercially available dye³, shows in a cultured neuron, a change of 40% of SHG signal per 100mV and only 2% SHG change per 100mV *ex vivo*⁴. Pons *et al.*⁵ proposed that two mechanisms are responsible of the change of response: i) a modification of the molecular hyperpolarizability by an electro-optical mechanism and ii) a modification of the degree of molecular alignment by the induced electric field. The first mechanism is the fastest, ideal for probing the change of neuron action potentials⁶. The dye response is also dependent on the system under study. Indeed, the concentration of lipids and proteins in biological membranes are sensitively different than in the models of membrane used in the laboratory. In Reeve *et al.*⁶, they state: “Dyes that are to be used for neuronal or cardiac voltage sensitive imaging require an SHG response on the millisecond timescale and greater than 10% per 100mV change in SHG signal.” Thus, in view of all those subtle requirements and as a continuation of this thesis, it seems an appealing area of investigation where the theoretical tools can be firstly a source of understanding, but secondly, of a great deal of information about the design of new dyes.

In 2001, Salafsky⁷ introduced the concept of ‘SHG-labels’ for the detection of molecules by second harmonic generation. He proposed to attach SHG dyes to molecules of some interest, which are not SHG-active. This allows the study of structural motion of biological molecules in real time and space.⁸ He proposed, for example, to attach sulfhydryl- or amine-specific dyes to a protein of cytochrome c, which allows to easily detect the ‘undetectable’ molecule. This would be another

interesting area of investigation where theory can help to find the more efficient dye for a particular system.

These new investigations can benefit from what we have learnt during this thesis. First, for the modeling of the dyes alone, standard quantum chemical methods can be used, as was the case for studies on model systems. Secondly, for the interaction of dyes with their surroundings, the ONIOM method can be helpful with some limitations. Indeed, in this thesis, in order to obtain sufficient accuracies (ONIOM QM/QM), we have been limited by the size of the system. Thirdly, in addition to those approaches, other methods can be considered as quantum mechanical/molecular mechanical (QM/MM) methods. In this particular case, Kongsted *et al.*⁹ have proposed QM/MM methods to determine linear response functions (CCSD/MM)^{9a} and quadratic response functions (DFT/MM)^{9b}. This model introduces polarization effects into a QM/MM framework using a self-consistent procedure while electrostatic effects are modeled by assigning partial charges to the MM molecules and a van der Waals potential describes dispersion and short-range repulsion. In order to have a better MM description, Söderhjelm *et al.*¹⁰ have proposed to involve multipole expansion and anisotropic polarizabilities calculations for all atoms and bond centers and to include such a model in the calculation of electronic properties. Another method is the polarized multipolar interactions with supramolecular pairs (PMISP) approach, which used fragment QM calculations for close interactions.¹¹ This method treats electrostatic interactions by multipoles up to quadrupoles, induction by anisotropic polarizabilities, and nonclassical interactions (dispersion, exchange repulsion, etc.) by explicit quantum chemical calculations, using a fragmentation approach, except for long-range interactions that are treated by standard molecular-mechanics Lennard-Jones terms. An extension of the PMISP method to determination of second-order NLO response will be a real asset for the follow-up of this thesis.

The principal shortcoming of the use of dyes for SHIM is the phototoxicity, which is the most problematic aspect to consider. Indeed, during real-time optical recording of neuronal action potential, Sacconi *et al.*¹² have shown that photodamage increased linearly with the dye concentration but grew with the cube of illumination intensity. The key parameter in those considerations is the signal-to-noise since the SHG signal voltage responses are small. It means that a high illumination intensity and/or high dye concentration are needed to attain useful signal-to-noise. Thus, dyes with higher SHG signal limit the photodamage process. Sacconi *et al.* used the FM4-64 dye, where its SHG signal is always accompanied by two-photon fluorescence, which is responsible for the two-photon excited production of singlet oxygen responsible for the photodamage. The understanding of this process is an asset for the design of new dyes. Thus, when studying a particular dye useful for SHIM, it seems important to study its phototoxicity and the underlying mechanism from a theoretical point of view. The question is what is the smallest system in which the observed photodamage could be believably studied? Answering this question could be a training exercise in carrying out excited state calculations, looking at evaluating different computational methods for studying geometry changes in ground and electronic excited states; mapping out decay pathways/reaction paths, most likely via surface crossings. Then this approach could be extended to larger system where the surrounding is accounted by the ONIOM method, with the long-term aims: understanding the mechanism of damage following excitation, developing photoprotecting groups/dyes that minimize damage,...

REFERENCES

1. (a) Millard, A. C.; Jin, L.; Lewis, A.; Loew, L. M., *Opt. Lett.* **2003**, *28* (14), 1221-1223; (b) Millard, A. C.; Jin, L.; Mei-de, W.; Wuskell, J. P.; Lewis, A.; Loew, L. M., *Biophys. J.* **2004**, *86* (2), 1169-1176; (c) Ries, R. S.; Choi, H.; Blunck, R.; Bezanilla, F.; Heath, J. R., *J. Phys. Chem. B* **2004**, *108* (41), 16040-16049; (d) Millard, A. C.; Jin, L.; Loew, L. M., *Proc. SPIE-Int. Soc. Opt. Eng.* **2005**, *5714* (Commercial and Biomedical Applications of Ultrafast Lasers V), 92-98; (e) Millard, A. C.; Jin, L.; Loew, L. M., *Proc. SPIE-Int. Soc. Opt. Eng.* **2006**, *6089* (Multiphoton Microscopy in the Biomedical Sciences VI), 60891G/1-60891G/9; (f) Reeve, J. E.; Corbett, A. D.; Boczarow, I.; Wilson, T.; Bayley, H.; Anderson, H. L., *Biophys. J.* **2012**, *103* (5), 907-917; (g) Reeve, J. E.; Corbett, A. D.; Boczarow, I.; Kaluza, W.; Barford, W.; Bayley, H.; Wilson, T.; Anderson, H. L., *Angew. Chem., Int. Ed.* **2013**, *52* (34), 9044-9048.
2. Millard, A. C.; Terasaki, M.; Loew, L. M., *Biophys. J.* **2005**, *88* (6), L46-L48.
3. Fluhler, E.; Burnham, V. G.; Loew, L. M., *Biochem.* **1985**, *24* (21), 5749-5755.
4. (a) Dombeck, D. A.; Sacconi, L.; Blanchard-Desce, M.; Webb, W. W., *J. Neurophysiol.* **2005**, *94* (5), 3628-3636; (b) Barsu, C.; Cheaib, R.; Chambert, S.; Queneau, Y.; Maury, O.; Cottet, D.; Wege, H.; Douady, J.; Bretonniere, Y.; Andraud, C., *Org. Biomol. Chem.* **2010**, *8* (1), 142-150.
5. Pons, T.; Mongin, O.; Mertz, J.; Blanchard-Desce, M.; Moreaux, L., *J. Biomed. Opt.* **2003**, *8* (3), 428-431.
6. Reeve, J. E.; Anderson, H. L.; Clays, K., *Phys. Chem. Chem. Phys.* **2010**, *12* (41), 13484-98.
7. Salafsky, J. S., *Chem. Phys. Lett.* **2001**, *342* (5-6), 485-491.
8. Salafsky, J. S., *Phys. Chem. Chem. Phys.* **2007**, *9* (42), 5704-5711.

9. (a) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O., *J. Chem. Phys.* **2003**, *118* (4), 1620-1633; (b) Nielsen, C. B.; Christiansen, O.; Mikkelsen, K. V.; Kongsted, J., *J. Chem. Phys.* **2007**, *126* (15), 154112.
10. Söderhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U., *J. Chem. Theo. Comp.* **2009**, *5* (3), 649-658.
11. (a) Söderhjelm, P.; Kongsted, J.; Ryde, U., *Journal of Chemical Theory and Computation* **2010**, *6* (5), 1726-1737; (b) Söderhjelm, P.; Ryde, U., *The Journal of Physical Chemistry A* **2008**, *113* (3), 617-627; (c) Söderhjelm, P.; Aquilante, F.; Ryde, U., *The Journal of Physical Chemistry B* **2009**, *113* (32), 11085-11094.
12. Sacconi, L.; Dombeck, D. A.; Webb, W. W., *Proc Natl Acad Sci U S A* **2006**, *103* (9), 3124-3129.

