Dynamic structural effects on the second-harmonic generation of tryptophane-rich peptides and gramicidin A

Jakob Seibert,[†] Benoît Champagne,[‡] Stefan Grimme,[*,†] and Marc de Wergifosse[*,†]

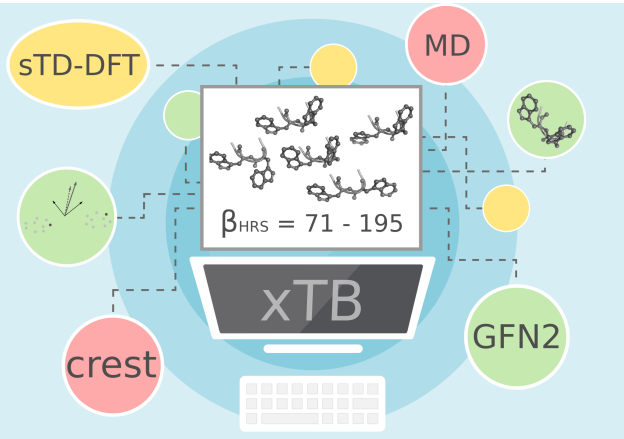
†*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie der Universität Bonn, Beringstr. 4, D-53115 Bonn, Germany*

‡*Laboratoire de Chimie Théorique, Université de Namur, rue de Bruxelles, 61, B-5000 Namur, Belgium*

E-mail: grimme@thch.uni-bonn.de; mdewergifosse@gmail.com

Abstract

Chains of amino acids can model endogenous biotags for applications in second harmonic imaging microscopy. Such structures are inherently flexible which may strongly affect their structure-property relationship. Here, we explore quantum-mechanically the conformational space of a set of relatively large tryptophan-rich model peptides studied experimentally by Duboisset *et al.* [*JPC B* **2014** *118*]. This has become feasible because of the recently proposed meta-dynamics method based on efficient tight-binding (TB) quantum chemical calculations. The TB version of the simplified time-dependent density functional theory (sTD-DFT-xTB) method is used to evaluate the first hyperpolarizability. These new tools enable us to calculate nonlinear optical properties for systems with several thousand atoms and/or to screen large structure ensembles. First, we show that the first hyperpolarizability of these systems is dominated by the indole chromophore in the tryptophan residues. Their relative orientation mostly determines the global β tensor and affects the static first hyperpolarizability response drastically. The results underline the importance of finding low-energy conformers for modeling the first hyperpolarizabilities of flexible molecules. Additionally, we compare calculated and extrapolated experimental static first hyperpolarizabilities. We conclude that the sTD-DFT-xTB method is capable of providing reliable second-harmonic generation values for tryptophan-rich systems at a fraction of the computational cost of the commonly used TD-DFT/TD-HF levels of theory.



Graphical TOC Entry

Introduction

Second harmonic imaging microscopy (SHIM)[1–4] has been developed for contrast enhancement of non-centrosymmetric molecular arrangements where the so-called second-harmonic generation (SHG) occurs. Information about the molecular organization of the chromophores can be extracted from SHG imaging data because the signal is polarization-dependent. SHIM is used as a high-resolution biological imaging technique where the SHG polarization anisotropy yields information about molecular orientation. Furthermore, it enables to analyze the degree of organization of proteins in tissues, related to their healthy or unhealthy state.[3] SHIM biotags could be endogenous like ordered structures of collagen,[5–7] microtubule, or miosin.[1,3] Exogenous biotags on the other hand should be carefully applied to avoid phototoxicity problems.[1]

Generally, the SHG process is not directly photodamaging the living environment because it is a scattering effect. However, at the laser wavelength, two-photon absorption and subsequent emission may occur so that excited state photochemistry could damage the sample. Thus, to avoid this, one must tune the molecular properties of such biotags.[2] A large first hyperpolarizability (β) is required for efficient SHG. Then, one can record the SHG signal of bright dyes with a low laser power, limiting photodamage. The exogenous biotags should also have a large β within the tissue transparency window (700-900 nm) and should involve minor one- and two-photon activities. Biotags such as fluorescent proteins can also be introduced by genetic engineering.

Some of us theoretically characterized the nonlinear optical (NLO) properties of fluorescent proteins at different theoretical levels of theory, including an ONIOM MP2:HF scheme where the first shell of residues around the chromophore was included in.[8–12] These investigations showed the importance to account for the H-bond network close to the chromophore and how the β of FPs strongly depends on the π-conjugation pathway, the degree of bond length alternation, and the presence of π-stacking interactions. Theoretical studies on compounds of interest for SHIM are still scarce.

In this contribution, we calculate and analyze the SHG response of small peptide chains and the peptide gramicidin A as models to understand the NLO properties of tryptophan-rich endogenous dyes. This is a follow-up work of the study of Duboisset *et al.*[13] who investigated experimentally the SHG response of an ensemble of tryptophan-rich peptides and gramicidin A. This set of systems is composed of KWK, KWWK, KWWWK, and KWWKWWK compounds where W and K denote tryptophan and lysine units, respectively. Gramicidin A is a natural peptide with the sequence VGALAVVVWLWLWLW, comprising four W units. Here, V, G, A, and L are the one-letter codes for the amino acids valine, glycine, alanine, and leucine, respectively. SHG signals were measured by hyper-Rayleigh scattering experiments at a wavelength of 784 nm, in Tris buffered aqueous solutions. While providing experimental reference first hyperpolarizabilities for this set of compounds, Duboisset *et al.*[13] showed that the β response follows an additive scheme with increasing number of W units. It also appears that the first hyperpolarizability of a sole tryptophan strongly differs from a KWK one. They argued that the local environment created by both lysine units decreases its SHG response.

From a theoretical point of view, the first hyperpolarizability of these systems is difficult to evaluate not only because of their size but also because of conformational flexibility. This implies a large number of relevant structures at room temperature. Recently, two of us proposed the simplified time-dependent density functional theory (sTD-DFT) method[14] in its tight-binding version to evaluate the frequency-dependent first hyperpolarizability of large compounds with up to about 3000 atoms. With respect to a full TD-DFT treatment, the simplified method applies three approximations: i. the exact Coulomb and exchange integrals are approximated by short-range damped Coulomb interactions of transition density monopoles, ii. the CI excitation space is truncated as controlled by a single energy selection threshold parameter, and iii. the response of the exchange-correlation functional is neglected.[15]

For the evaluation of the first hyperpolarizability, two minor additional approximations are

introduced, i.e., both the Hartree exchange-correlation kernel and the third functional derivative of the exchange-correlation functional are neglected.[14] A semi-empirical tight-binding version called sTD-DFT-xTB also has been developed where instead of using KS-DFT input data, orbitals and eigenvalues from an extended basis set tight-binding calculation are employed.[16] It was shown that the sTD-DFT-xTB method can reproduce reasonably well the β frequency dispersion of a collagen triple helix $[(\text{Pro-Pro-Gly})_{10}]_3$ and of fluorescent proteins with respect to ONIOM reference calculations.[14]

Here, the sTD-DFT-xTB method is used to evaluate and understand dynamical structural effects on the SHG response for the set of tryptophan-rich short peptides and gramicidin A characterized experimentally by Duboisset *et al.*[13]

This article is organized as followed. First, we detail the calculations for this study and then present in the first results part, the analysis of the conformer ensembles for all systems with respect to their first hyperpolarizability properties. Structure-property relationships and the effect of sampling structures along a molecular dynamic (MD) trajectory is discussed. The last part of the results section compares experimental first hyperpolarizabilities to computed values at different levels of theory. The summary section concludes the main findings and possible implications of this work for future applications.

Computational Details

Figure 1 presents the structures of *L*-lysine (K) and *L*-tryptophan (W), which are the building blocks of all tryptophan-rich peptides compounds considered in this study. The secondary (rigid helical) structure of gramicidin A with highlighted tryptophan units is also shown. In order to evaluate the impact of conformational flexibility on the first hyperpolarizability as well as for determining lowest energy conformers, we used the RMSD-based meta-dynamics approach recently proposed by Grimme,[17–19] except for gramicidin A. First, the standard, ground-state tight-binding GFN2-xTB[20] method is used for the generation of the conformer

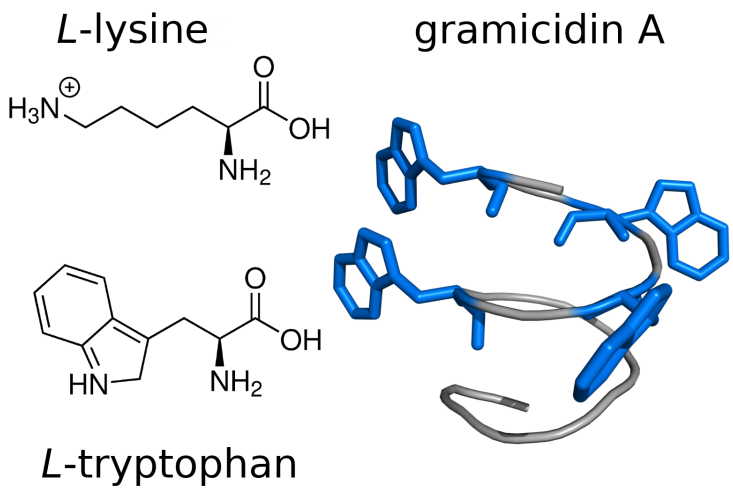


Figure 1: Lewis structures of amino acids lysine and tryptophan, building residues for the model peptides. Secondary structure of gramicidin A with blue highlighted tryptophan units.

ensemble. Solvation effects for water are implicitly accounted for using the GBSA[21–23] continuum model. Second, all conformers within a 6 kcal/mol GFN2-xTB energy window were optimized at the PBEh-3c(COSMO)[24,25] level of theory. Third, within a 4 kcal/mol PBEh-3c energy window all remaining structures were used as input for PW6B95/def2-QZVP[26,27] single point energy calculations. Free energies were computed for the lowest energy conformers by adding solvation free energies with COSMO-RS[28,29] and thermostatistical contributions within the modified[30] rigid-rotor harmonic oscillator approximation based on the GFN2-xTB computed Hessian ($\Delta G = \Delta E_{PW6B95} + \Delta G_{solv}^{COSMO-RS} + \Delta G_{RRHO}^{GFN2-xTB}$). The Boltzmann weights used correspond to a temperature of 298.15 K. The first hyperpolarizability of relevant conformers (population larger than 1.5%) were determined at the sTD-DFT-xTB/GBSA level of theory with modified Coulomb y_J and exchange y_K parameters set to 0.55 and 1.0, respectively. To further explore dynamic structural effects, molecular dynamic (MD) simulations were carried out for 1 ns at the GFN2-xTB level of theory with preceding equilibration. In the simulations, a time step of 4 fs was used and the SHAKE[31,32] algorithm was applied, constraining all covalent bonds. The lowest conformer was used as starting point for the MD simulations. From the resulting trajectory 200 snapshots were taken

equidistantly and used as structural input for the sTD-DFT-xTB/GBSA[14] calculations.

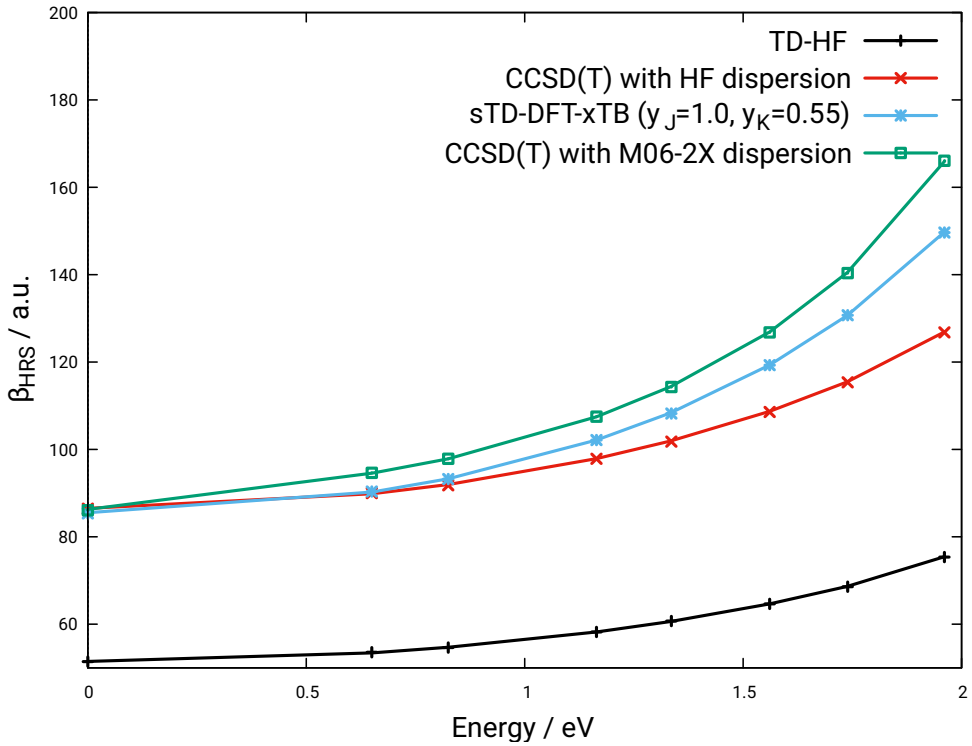


Figure 2: Frequency dispersion for tryptophan computed with CCSD(T)/aug-cc-pVDZ (red), TDHF/aug-cc-pVDZ (black) and sTD-DFT-xTB (blue) methods.

All sTD-DFT-xTB/GBSA calculations were performed with a configuration selection threshold of 10 eV. Note that the GBSA solvation model is only applied in the (ground state) orbitals generation step, meaning that non-equilibrium solvent effects on the hyperpolarizability are not accounted for. These effects usually enhance the SHG response.[33–36] As already mentioned in the original publication of the method,[14] the underlying sTD-DFT-xTB parameterization – originally developed for excitation energies and absorption spectra – is not perfectly suited for nonlinear optical properties which more strongly depend on the high-energy part of the excitation manifold. We employ a small model system for benchmarking and adjustment purposes, i.e., performing high-level calculations for reference β values and then to tune sTD-DFT-xTB Coulomb y_J and exchange y_K parameters accordingly. We computed the static β_{HRS} of tryptophan at the CCSD(T)/aug-cc-pVDZ[37] level of theory using the Romberg's automatic finite-field (FF) differentiation procedure.[38] Its missing frequency dis-

persion was accounted for using the multiplicative approximation($\beta^{\omega}_{CC} = \beta^{0}_{CC}\frac{\beta^{\omega}_{TDHForM06-2X}}{\beta^{0}_{TDHForM06-2X}}$). Figure 2 shows the adjusted sTD-DFT-xTB β_{HRS} curve matching almost perfectly the CCSD(T) one with the M06-2X frequency dispersion when using a y_J parameter of 0.55 instead of the original value of 4.0 and a y_K of 1.00 instead of 2.0. For all following sTD-DFT-xTB calculations, the y_J and y_K parameters are set to 0.55 and 1.0, respectively.

Time-dependent Hartree-Fock (TDHF) calculations were also conducted for lowest energy conformers with the 6-31+G(d) basis set with and without solvent effects accounted for using the IEF-PCM scheme.[39,40] Note that when comparing TDHF results to experiment, the static and dynamic dielectric constants of water differ largely. Thus, we select a value at a large wavelength of 1900 nm instead of the static one. This wavelength is chosen to enable non-biased comparisons and should be large enough to prevent any (near)resonance effects. For the smaller systems W and KWK we also computed the response with the slightly larger aug-cc-pVDZ basis set showing very small differences with respect to 6-31+G(d) (see Table S1 and S2). In the experimental work of Duboisset *et al.*,[13] the hyper-Rayleigh scattering value β_{HRS}[41] was determined. Theoretically, the following definition of β_{HRS} is used as the mean of β-tensor orientations

$$\beta_{\mathrm{HRS}}\left(-2\omega;\omega,\omega\right) = \sqrt{\{\langle\beta^2_{ZZZ}\rangle + \langle\beta^2_{ZXX}\rangle\}}, \tag{1}$$

where molecular averages without assuming Kleinman's conditions[42] are defined in the laboratory frame as

$$\begin{aligned}\left\langle\beta^2_{ZZZ}\right\rangle = \frac{1}{7}\sum_{\zeta}^{x,y,z}\beta^2_{\zeta\zeta\zeta} + \frac{4}{35}\sum_{\zeta\neq\eta}^{x,y,z}\beta^2_{\zeta\zeta\eta} + \frac{2}{35}\sum_{\zeta\neq\eta}^{x,y,z}\beta_{\zeta\zeta\zeta}\beta_{\zeta\eta\eta} + \frac{4}{35}\sum_{\zeta\neq\eta}^{x,y,z}\beta_{\eta\zeta\zeta}\beta_{\zeta\zeta\eta} \\ + \frac{4}{35}\sum_{\zeta\neq\eta}^{x,y,z}\beta_{\zeta\zeta\zeta}\beta_{\eta\eta\zeta} + \frac{1}{35}\sum_{\zeta\neq\eta}^{x,y,z}\beta^2_{\eta\zeta\zeta} + \frac{4}{105}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\zeta\zeta\eta}\beta_{\eta\xi\xi} + \frac{1}{105}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\eta\zeta\zeta}\beta_{\eta\xi\xi} \\ + \frac{4}{105}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\zeta\zeta\eta}\beta_{\xi\xi\eta} + \frac{2}{105}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta^2_{\zeta\eta\xi} + \frac{4}{105}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\zeta\eta\xi}\beta_{\eta\zeta\xi},\end{aligned} \tag{2}$$

and

$$\begin{aligned}\langle\beta^2_{ZXX}\rangle = \frac{1}{35}\sum_{\zeta}^{x,y,z}\beta^2_{\zeta\zeta\zeta} + \frac{4}{105}\sum_{\zeta\neq\eta}^{x,y,z}\beta_{\zeta\zeta\zeta}\beta_{\zeta\eta\eta} - \frac{2}{35}\sum_{\zeta\neq\eta}^{x,y,z}\beta_{\zeta\zeta\zeta}\beta_{\eta\eta\zeta} + \frac{8}{105}\sum_{\zeta\neq\eta}^{x,y,z}\beta^2_{\zeta\zeta\eta} \\ + \frac{3}{35}\sum_{\zeta\neq\eta}^{x,y,z}\beta^2_{\zeta\eta\eta} - \frac{2}{35}\sum_{\zeta\neq\eta}^{x,y,z}\beta_{\zeta\zeta\eta}\beta_{\eta\zeta\zeta} + \frac{1}{35}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\zeta\eta\eta}\beta_{\zeta\xi\xi} - \frac{2}{105}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\zeta\zeta\xi}\beta_{\eta\eta\xi} \\ - \frac{2}{105}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\zeta\zeta\eta}\beta_{\eta\xi\xi} + \frac{2}{35}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta^2_{\zeta\eta\xi} - \frac{2}{105}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\zeta\eta\xi}\beta_{\eta\zeta\xi}.\end{aligned} \tag{3}$$

In addition from these quantities, the depolarization ratio can be obtained

$$\mathrm{DR} = \frac{I^{2\omega}_{VV}}{I^{2\omega}_{HV}} = \frac{\langle\beta^2_{ZZZ}\rangle}{\langle\beta^2_{ZXX}\rangle}. \tag{4}$$

where a value of 1.5 corresponds to a fully octupolar response, of 5 to a one-dimensional push-pull π-conjugated system, and of 9 to a fully dipolar system. Furthermore, it is useful to analyze the β tensor in terms of its dipolar (J=1) and octupolar (J=3) tensorial β_J-components:[43]

$$\beta_{HRS} = \sqrt{\langle\beta^2_{HRS}\rangle} = \sqrt{\frac{10}{45}|\beta_{J=1}|^2 + \frac{10}{105}|\beta_{J=3}|^2}, \tag{5}$$

$$\text{where } |\beta_{J=1}|^2 = \frac{3}{5}\sum_{\zeta}^{x,y,z}\beta^2_{\zeta\zeta\zeta} + \frac{6}{5}\sum_{\zeta\neq\eta}^{x,y,z}\beta_{\zeta\zeta\zeta}\beta_{\zeta\eta\eta} + \frac{3}{5}\sum_{\zeta\neq\eta}^{x,y,z}\beta^2_{\eta\zeta\zeta} + \frac{3}{5}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\eta\zeta\zeta}\beta_{\eta\xi\xi}, \tag{6}$$

$$|\beta_{J=3}|^2 = \frac{2}{5}\sum_{\zeta}^{x,y,z}\beta^2_{\zeta\zeta\zeta} - \frac{6}{5}\sum_{\zeta\neq\eta}^{x,y,z}\beta_{\zeta\zeta\zeta}\beta_{\zeta\eta\eta} + \frac{12}{5}\sum_{\zeta\neq\eta}^{x,y,z}\beta^2_{\eta\zeta\zeta} - \frac{3}{5}\sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta_{\eta\zeta\zeta}\beta_{\eta\xi\xi} + \sum_{\zeta\neq\eta\neq\xi}^{x,y,z}\beta^2_{\zeta\eta\xi}. \tag{7}$$

Using this decomposition, the electronic character of a NLO chromophore can be analyzed. We used a loss-less 3D visualization of the first hyperpolarizability tensor, to have a more intuitive analysis tool for the β tensor.[44] This so-called unit-sphere representation (USR) uses effective SHG dipoles, which are defined as:

$$\vec{\beta}_{\text{eff}} = \overset{\leftrightarrow}{\beta} : \hat{E}(\theta,\phi)\hat{E}(\theta,\phi) \tag{8}$$

Taking all possible incident polarization directions defined by (θ,ϕ), a unit-sphere is mapped out (at a field value of $\hat{E}(\theta, \phi)$ 1 a.u.). At these sampled points on the unit-sphere surface, the corresponding $\vec{\beta}_{\text{eff}}$ is plotted. This scheme allows a three-dimensional visualization of the β tensor without losing any information. Another approach for visualizing the β tensor, but at cost of loosing anisotropic information, consists in defining a vector, having the following components:[45]

$$\beta_x = \beta_{xxx} + \beta_{xyy} + \beta_{xzz}; \beta_y = \beta_{yxx} + \beta_{yyy} + \beta_{yzz}; \beta_z = \beta_{zxx} + \beta_{zyy} + \beta_{zzz} \tag{9}$$

All reported β values are given in atomic units [1 a.u. of β=3.6213 10^{-42} m^4 V^{-1}=3.2064 10^{-53} C^3 m^3 J^{-2} = 8.639 10^{-33} esu] within the Taylor series convention.[46] The CCSD(T) finite-field first hyperpolarizabilities were computed by the T-REX program natively interfaced with QChem Version 5.1.[47] All TDHF calculations were carried out using the Gaussian 09 package.[48] For all DFT calculations TURBOMOLE Version 7.2[49,50] was used with COSMOtherm Version C3.0 release 1601[28,29] for COSMO-RS. The remaining calculations were conducted with the xtb[51] and stda[52] codes.

The first hyperpolarizability values of the tryptophan-rich peptides (β^{800}_{HRS}) were experimentally determined by Duboisset *et al.*[13] To eliminate the resonance effects and extrapolate to the static value,[53,54] we process the experimental data with the two-state approximation (TSA) proposed by Oudar and Chemla.[55]

$$F(\omega, \omega_{ge}, \gamma) = \frac{\beta_{zzz}(-2\omega, \omega, \omega)}{\beta_{zzz}(0, 0, 0)} = \frac{\omega^2_{ge}(\omega_{ge} - i\gamma)^2}{([\omega_{ge} - i\gamma]^2 - 4\omega^2)\,([\omega_{ge} - i\gamma]^2 - \omega^2)} \tag{10}$$

A homogeneous broadening of $\gamma = 0.35$ eV is applied while the experimental excitation energy of tryptophan $\omega_{ge} = 4.44$ eV is used. Aside from the conventional TSA, inhomogeneous broadening or even the vibronic structure of the excited states could be accounted for.[6,53,54] However, using and comparing these refined extrapolation schemes for the studied systems goes beyond the scope of this study and would be inaccessible since some required experi-

mental data are missing. Table 1 shows the dynamic and extrapolated experimental values.

Table 1: Measured and extrapolated (TSA) first hyperpolarizabilites in atomic units.

system	β_{HRS}^{800}	β_{HRS}^{∞}
W	544	240
KWK	100	44
KWWK	396	175
KWWWK	828	365
KWWKWWK	1215	536
gramicidin A	872	384

Results and Discussion

Conformers

Molecular properties depend on details of the geometrical structure under the measurement conditions, and it is important to describe the system as closely as possible to its natural state. Peptides in solution at ambient temperature are represented as a set of various conformers. Depending on the temperature, these are populated differently and the accessed conformational space can become quite large. The objective of this section is to analyze the conformational dependence of the SHG response, shedding light on this structure-property relationship. The above described theoretical multi-level approach provides a very reasonable conformational ensemble including solvation effects. For these sets of conformers, molecular first hyperpolarizabilities (static β_{HRS}) were computed. With the help of the described visualization techniques for the β tensor, the conformers of each system are compared in terms of their electronic and geometrical structures to show how they impact the NLO properties. First, the tryptophan molecule is analyzed. For this system, the indole unit is essentially responsible for the first hyperpolarizability. Therefore, only small variations in β_{HRS} values are expected due to its rigid π-conjugated system. The amine and carboxylic acid moieties are flexible and the obtained conformer ensemble is mainly determined by their different

orientations. Figure 3 shows the most contributing conformers at room temperature for

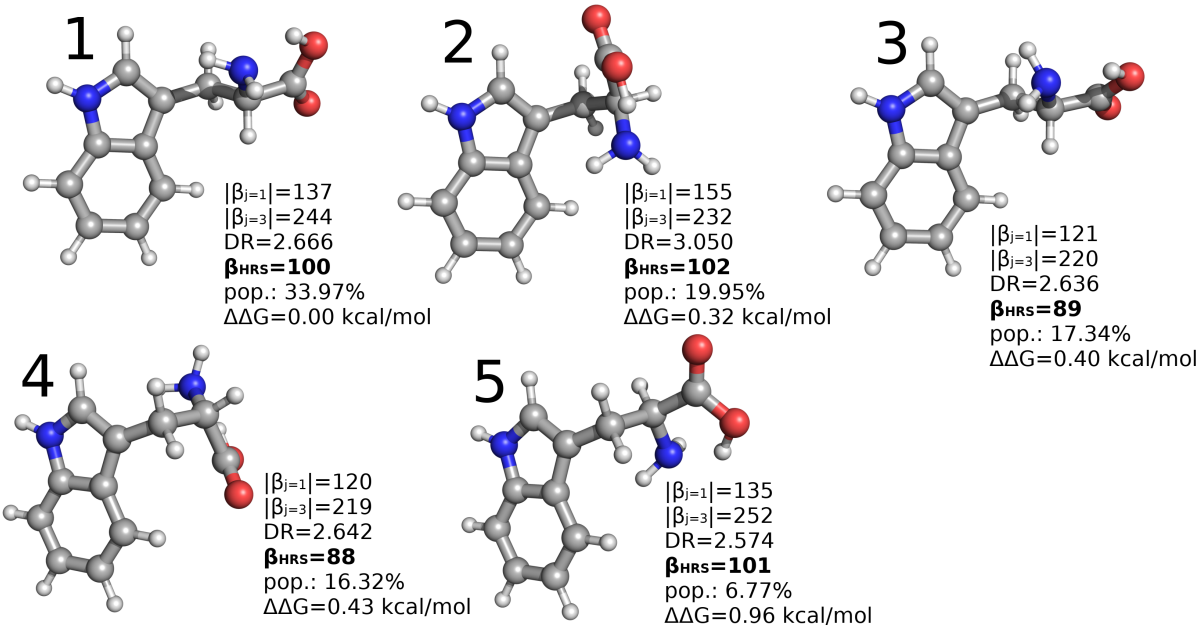


Figure 3: Conformer ensemble for tryptophan. First hyperpolarizability data ($|\beta_{J=1}|$, $|\beta_{J=3}|$, DR, β_{HRS}), population and relative free energies are depicted with the plotted structures.

tryptophan and their respective first hyperpolarizability values which do not differ much. The largest differences are caused by the re-orientation of the amine group. A small SHG enhancement is observed when local dipoles of the indole and amine units are perpendicular (conformer 2). This can be further visualized by the USR (see Figure S1). There, two conformers with the largest and smallest β_{HRS} values are shown (conformers 2 and 4). The pattern of effective SHG dipoles indicates a mix of dipolar and octupolar contributions. According to Equation 5, e.g., the β^2_{HRS} value of conformer 2 contains 51% of dipolar and 49% of octupolar contribution.

The second system is the model peptide KWK, a lysine capped tryptophan. The lysine groups were introduced for solubility reasons.[13] Experimentally, the observed difference in responses between tryptophan and KWK was attributed to the lysine side chains that shield the indole unit from the solvent. A large variety of conformers is found due to the high flexibility of these side chains. However, the rigid indole chromophore is unchanged among these conformers. Figure 4 presents the conformers and their first hyperpolarizability properties.

Note that sTD-DFT-xTB/GBSA calculations cannot account for non-equilibrium solvent

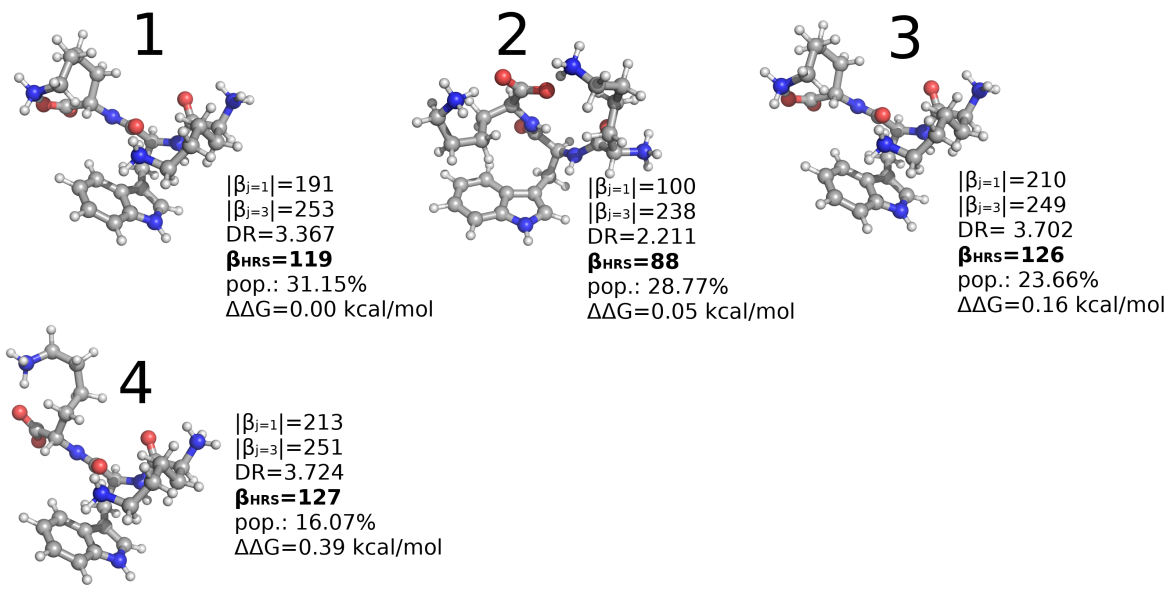


Figure 4: Conformer ensemble for model peptide KWK. First hyperpolarizability data ($|\beta_{J=1}|$, $|\beta_{J=3}|$, DR, β_{HRS}), population and relative free energies are depicted with the plotted structures.

effects on the response and hence part of the solvent-induced difference between tryptophan and KWK is missing. Thus, we conducted TDHF/IEF-PCM calculations for comparison that are discussed in the SI. Mostly, the picture emerging from the calculations is similar for both systems: several conformers of KWK are significantly populated, as for W. The USR for KWK (Figure S2) is more complex to analyze because of small dipolar contributions from both peptide bonds, but overall the response is similar to the one for W (i.e., a mix of dipolar and octupolar contributions). Note, however, that the relative $|\beta_{J=1}|$ and $|\beta_{J=3}|$ contributions to β^2_{HRS} are quite different for, e.g., conformers 2 and 4 (with factors of 2.4 and 0.6 between the two contributions).

The third model peptide KWWK includes two chromophore units. The respective orientation of the two indole groups is mainly responsible for the change of the first hyperpolarizabilities among the ensemble. Figure 5 displays conformers of the KWWK model peptide, where the lysine groups and the peptide backbone are hidden to improve the visibility of chromophores.

Based on the results of tryptophan and KWK systems, we conclude that the effects of the

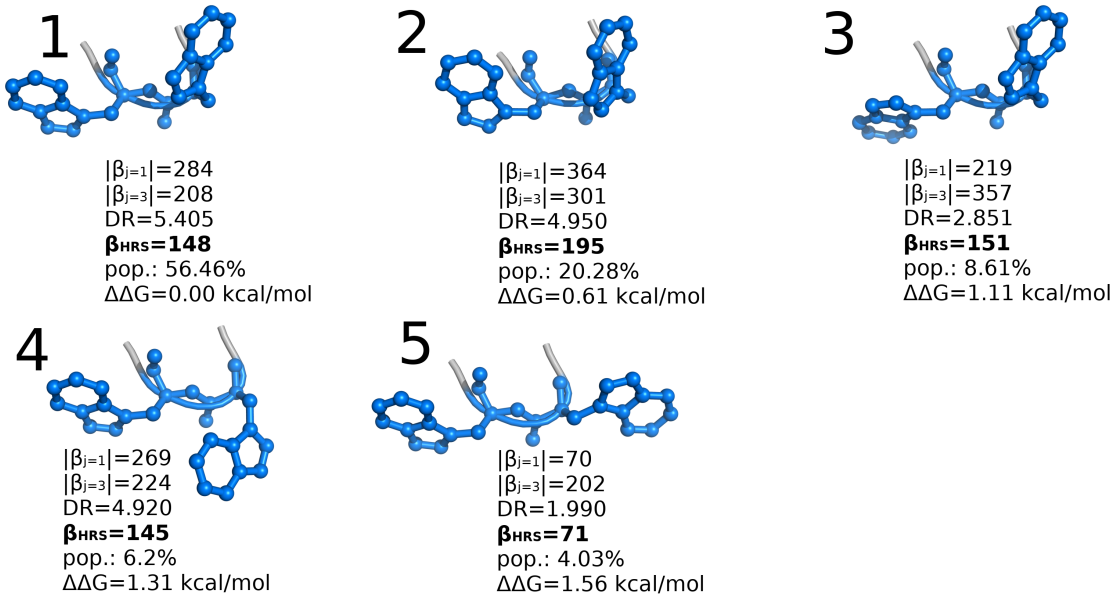


Figure 5: Conformer ensemble for model peptide KWWK. First hyperpolarizability data ($|\beta_{J=1}|$, $|\beta_{J=3}|$, DR, β_{HRS}), population and relative free energies are depicted with the plotted structures.

saturated side chains are negligible for the SHG response. Therefore, the structure-property analysis is focused on the orientation of the indole moieties. For KWWK, the extreme β_{HRS} values differ by more than a factor of two. The values among the ensemble show a larger spread than for the mono-chromophoric examples. The USR of both conformers – with the lowest (5) and highest (2) β_{HRS} values – shows that the orientation of the indole units plays an important role on the SHG response (see Figure 6). The indole units are aligned parallel in the second conformer with the largest first hyperpolarizability (232 a.u.). On the other hand, one can observe a drastic decrease of the β_{HRS} value when the indole units are antiparallel (71 a.u. for conformer 5). The USR shows a very strong dipolar character of the β tensor, when the chromophore units are aligned parallel. In that case, the DR value of 4.95 confirms this. For opposite orientations, the octupolar component dominates the β tensor, as indicated by a DR value of 1.99.

For the larger KWWWK and KWWKWWK systems, similar observations as for KWWK

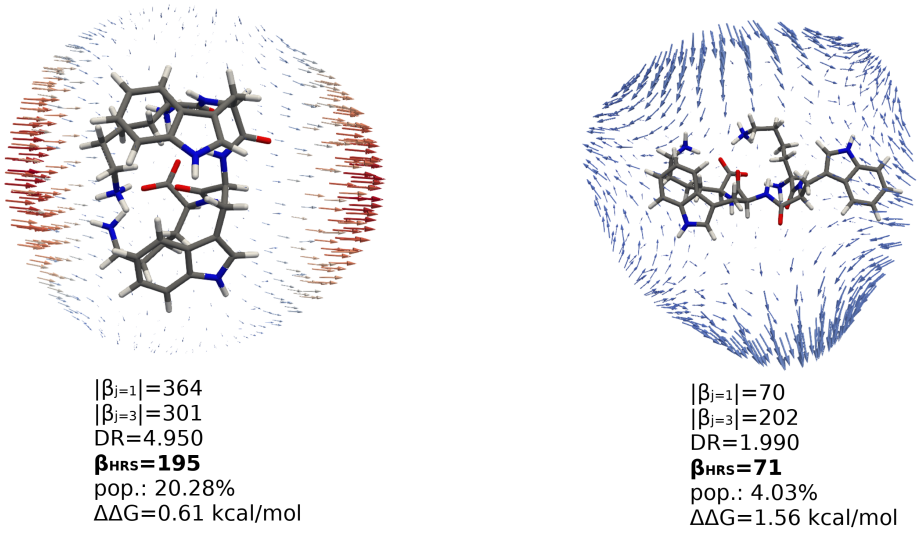


Figure 6: Unit-sphere representation for KWWK conformers 2 (left) and 5 (right). To increase visibility, the vector fields are differently scaled but arrow colors are consistent.

are made. The orientation of the indole groups directly correlates with the magnitude of the first hyperpolarizability. The conformers of KWWWK and KWWKWWK are depicted in Figure 7 and 8, respectively. The two highest populated conformers of KWWWK differ by almost a factor of 1.5 in β_{HRS} values (1 and 2 with β_{HRS}=181 a.u. and 121 a.u., respectively). Conformer 1 has almost equal dipolar $|\beta_{J=1}|$ and octupolar $|\beta_{J=3}|$ values. Considering their weighting factors – $\frac{10}{45}$ and $\frac{10}{105}$, respectively – a strong dipolar contribution to the β_{HRS} is observed, confirmed by a DR of 3.80. For this conformer, two indole units are more or less pointing in the same direction while one points in a different one (but not opposite). For conformer 2, this third indole unit changes its orientation to point in an opposite direction with respect to the first one, canceling each others. The dipolar contribution $|\beta_{J=1}|$ then equals only to 113 a.u. where the octupolar component is more than three times larger ($|\beta_{J=3}|$=349 a.u.). The DR value of 1.94 indicates an octupolar case for this conformer, which is corroborated by the USR depicted in Figure S3. This analysis shows that the inclusion of several conformers changes the β_{HRS} value compared to the one of the lowest energy conformer. In a later section, we will discuss whether this approach improves the

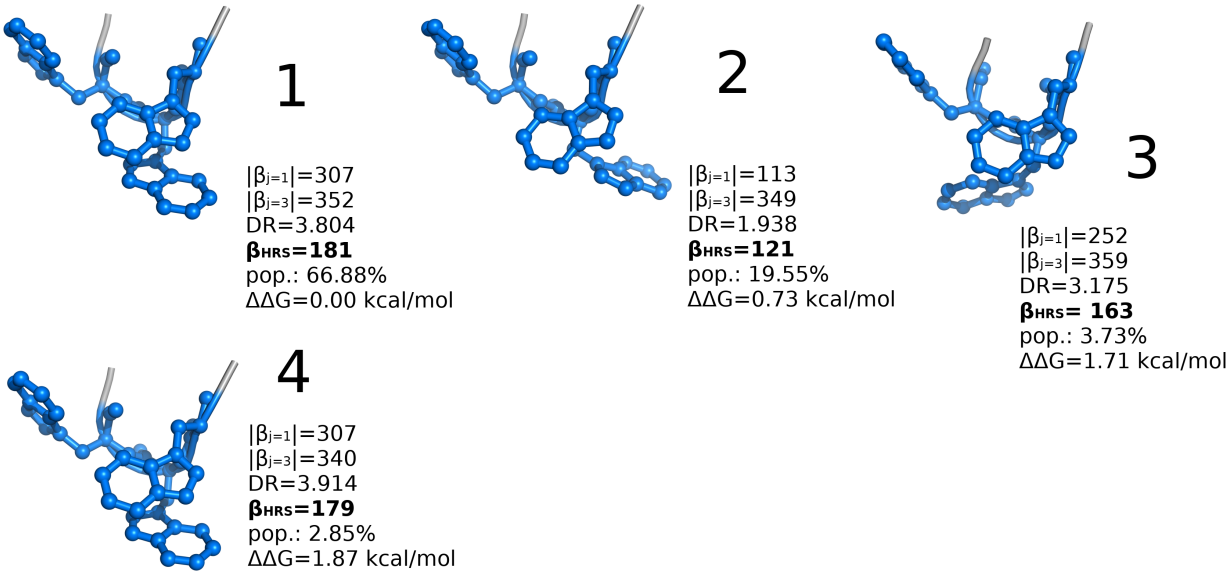


Figure 7: Conformer ensemble for model peptide KWWWK. First hyperpolarizability data ($|\beta_{J=1}|$, $|\beta_{J=3}|$, DR, β_{HRS}), population and relative free energies are depicted with the plotted structures.

first hyperpolarizability values in comparison to experiment.

With four chromophore units, KWWKWWK is the largest example of all model peptides studied here. Figure 8 presents the six significantly populated conformer structures. Among this set, the conformers differ mostly by the orientation of their indole units. The first hyperpolarizability values spread by a factor of two between the conformers 3 and 4 (β_{HRS}=234 and 112 a.u.). For conformers 1, 3, and 6, the indole units are oriented roughly in the same manner which is mainly due to the shared secondary structure. In the second group of trans-like conformers (2, 4 and 5), the indole units are partially oriented anti-parallel. This results in a decrease of the dipolar contribution to the β_{HRS}. The values for $|\beta_{J=1}|$ and $|\beta_{J=3}|$ indicate that for this cis-like group the dipolar contribution is dominant. The secondary structure can bend only because of the flexibility introduced by the third lysine unit. While the KWWWK ensemble is only dominated by a few conformers due to a rigid peptide backbone, the KWWKWWK ensemble is clearly enlarged by this added flexibility. For the interested reader we provide in section S3 an analysis in terms of indole unit β vec-

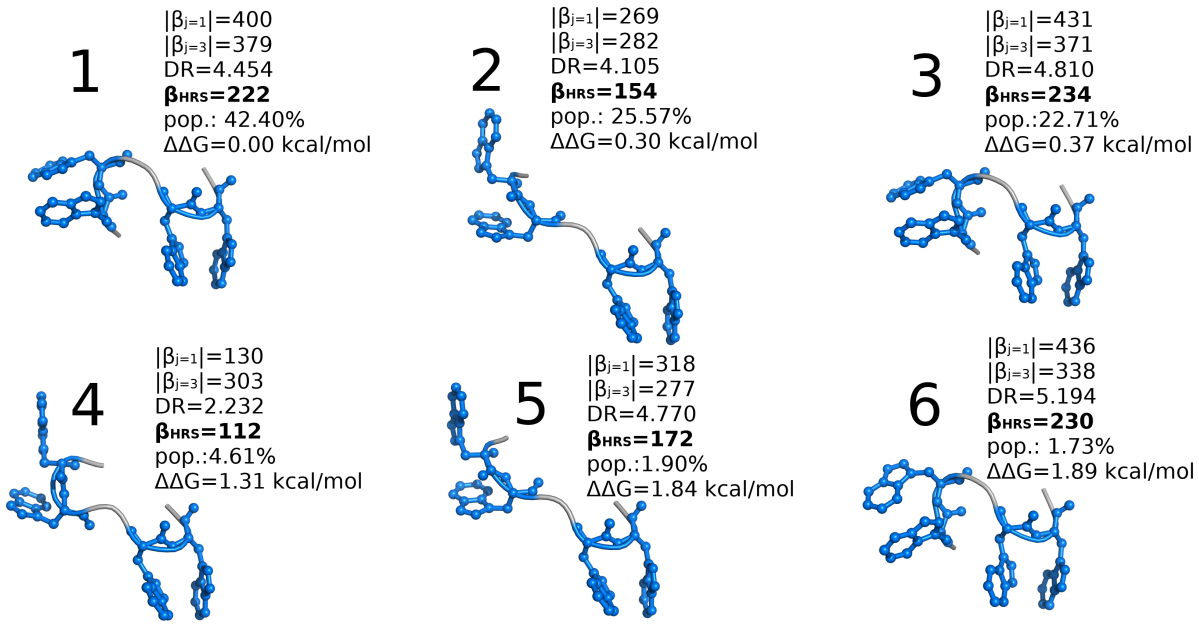


Figure 8: Conformer ensemble for model peptide KWWKWWK. First hyperpolarizability data ($|\beta_{J=1}|$, $|\beta_{J=3}|$, DR, β_{HRS}), population and relative free energies are depicted with the plotted structures.

tors. They explain further the significant differences observed in calculated SHG response among conformers.

Table 2 shows the first hyperpolarizability values for the lowest energy conformer and the Boltzmann-weighted ensemble. For tryptophan and KWK, the two values are very close to each other. As already described above, this is due to the rigidity of the chromophore. For the larger systems, we observed slightly larger differences between a single structure approach and the ensemble average. We explain this with the orientation of multiple chromophore units in ways that enhance or cancel the SHG response. The difference in β_{HRS} values (in a.u.) between KWK and KWWK amounts to 30, to 32 for KWWK and KWWWK, and to 41 a.u. for KWWWK and KWWKWWK. Since not all lowest energy conformers share the same indole orientation, these enhancements are not equal. When considering the Boltzmann weighted β_{HRS}, the differences can be quantified to 41, 13, and 35 a.u. Also here, the enhancement is not perfectly linear. However, such a perfect linear enhancement is not realistic since the different conformers for the individual model peptides have a strong influence

on the SHG response. These results clearly indicate that the addition of a tryptophan unit to these model peptides does not equally enhance the first hyperpolarizability.

Table 2: Static first hyperpolarizability values (in a.u.) for the lowest free energy conformer and the Boltzmann weighted ensemble computed with sTD-DFT-xTB/GBSA.

system	minimum	ensemble
W	100	97
KWK	119	113
KWWK	149	154
KWWWK	181	167
KWWKWWK	222	202

Molecular dynamics sampling

In this section, we analyze the effect of using several structures from a MD simulation as input for the evaluation of β. For the simulation of electronic circular dichroism spectra, some of us have already applied this approach successfully.[56–58] From a MD trajectory, snapshots are taken equidistantly and serve as input structures for the property calculations which are simply averaged over all included snapshots. In this process, no structural relaxations are included and the snapshots are equally weighted. By using a reasonably long simulation time, the considered structures should represent a Boltzmann ensemble.

Figure 9 shows the frequency dispersion of β for each snapshot as well as their average for the largest peptide KWWKWWK. The value of β_{HRS} changes drastically within a factor of six between the most extreme structures. This illustrates the sensitivity of this NLO property with respect to structural subtleties of a flexible system. However, when all snapshots are averaged the resulting frequency dispersion of β_{HRS} does not differ much from the one that considers only the lowest energy conformer with a reduction of around 2-22%. The frequency dispersion of the Boltzmann weighted conformer ensemble does not differ much from the MD averaged ensemble (cf. Figure S8). Figure 10 shows selected snapshots from the MD trajectory giving the lowest, the highest, and intermediate values for the first hyperpolar-

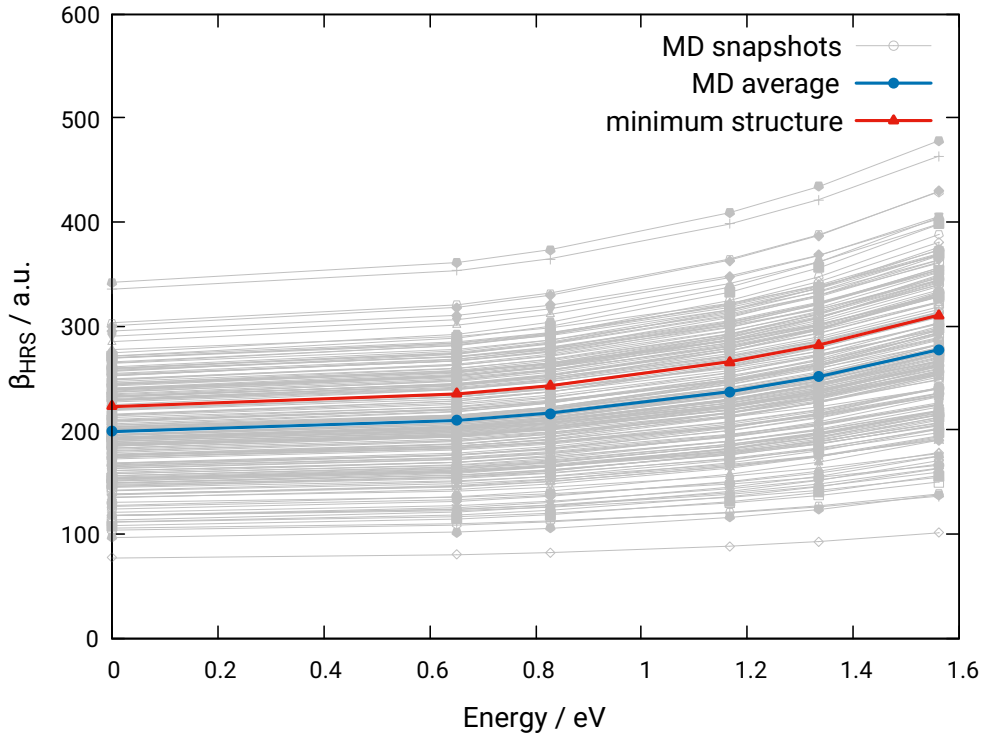


Figure 9: Frequency dispersion computed with sTD-DFT-xTB for 200 snapshots (gray), the MD average (blue) and for the optimized minimum structure (red) for the KWWKWWK peptide.

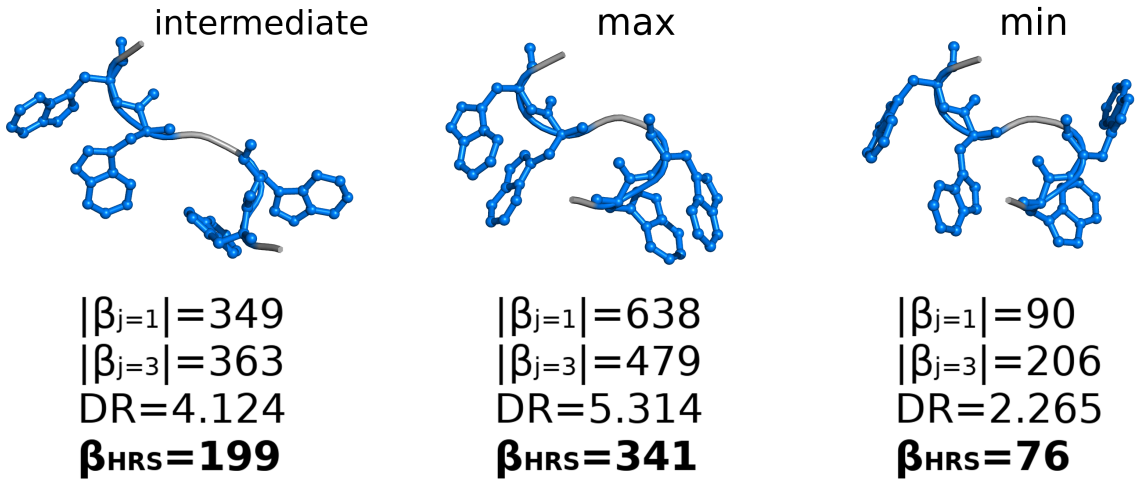


Figure 10: Selected MD structures for model peptide KWWKWWK. First hyperpolarizability data ($|\beta_{J=1}|$, $|\beta_{J=3}|$, DR and β_{HRS}) are depicted with the plotted structures.

izability. Analyzing the structures in terms of indole orientations corroborates the findings from the previous section. When all indole units are oriented along the same direction, the largest β_{HRS} values are obtained. Considering the structure with the minimum response, we observe that two indole units are pointing in opposite directions, canceling out their dipolar contributions to β, leaving a dominant octupolar character.

Table 3: Static first hyperpolarizability values for lowest energy conformer (minimum), Boltzmann weighted ensemble (ensemble) and averaged MD snapshots (MD average) computed with sTD-DFT-xTB/GBSA.

system	minimum	ensemble	MD average
W	100	97	100
KWK	119	113	128
KWWK	149	154	142
KWWWK	181	167	152
KWWKWWK	222	202	198

Table 3 shows MD averaged static first hyperpolarizabilities for all model peptides. The difference in β_{HRS} values between the minimum structure (minimum) and MD averaged structures (MD average) increases with system size. This observation holds also for the difference between the minimum energy conformer and the conformer weighted ensemble (ensemble). The first hyperpolarizability values of the ensemble differ from the ones of the MD average, except for the biggest system. Here, emphasize in the discussion on two systems. First, for tryptophan we obtain similar values for all three approaches. As already stated, this small difference is due to the negligible difference in SHG response among the conformers. Second, for KWWKWWK, the small difference in SHG response of ensemble and MD average can be assigned to a shallow PES. The energy difference between the first and the third most contributing conformers is only 0.35 kcal/mol and those conformers amount up to 91% of the entire population (see Figure 8). Thus, the MD simulation covers an ensemble that is comparable to the one from the equilibrium structure sampling procedure. However, differences between conformer weighted ensemble and MD average are expected and present for the remaining systems, since the simulation time of 1 ns is rather short to

fully explore the conformational PES.

Comparison with extrapolated experimental values

In this section, we compare the experimental values to ones computed at the TDHF level with and without accounting for implicit solvation effects and at both sTD-DFT-xTB and sTD-DFT-xTB/GBSA levels. The frequency dispersion of β for all systems, computed with TDHF and sTD-DFT-xTB, is discussed in section S5.

As geometries, we consider the lowest energy conformers for each system optimized with the PBEh-3c(COSMO) method. In the case of averaging all relevant conformers, the static β_{HRS} values were weighted by their respective Boltzmann weights. The static first hyperpolarizabilites (and at 1900 nm for TDHF/IEF-PCM results) are shown in Figure 11, which is divided in two panels: one for comparing tryptophan with KWK and one for every systems except tryptophan.

First, we compare the β response of tryptophan and KWK. The experimental first hyperpolarizability values reveal a much weaker response (-80%) for KWK compared to tryptophan. This observation is only reproduced by the TDHF/IEF-PCM method, demonstrating the role of non-equilibrium solvent effects, though it yields only a difference of 20%. Due to the flexible side chains of KWK, the chromophore is partially shielded from the solvent. As argued in the experimental paper, this could have a large effect on the electronic structure of the indole unit and thus also on the first hyperpolarizability. However, this could not be corroborated by our calculations. Secondly, the remaining systems are compared. The TDHF method without accounting for solvent effects is not able to reproduce the experimental trends. For a comparison of the frequency dispersion of TDHF and sTD-DFT-xTB see Figure S9. Applying an implicit solvation model improves the quality of the computed data. Except for gramicidin A, the experimental ordering is reproduced. The IEF-PCM scheme improves the β_{HRS} values with respect to the experiment. The sTD-DFT-xTB/GBSA method

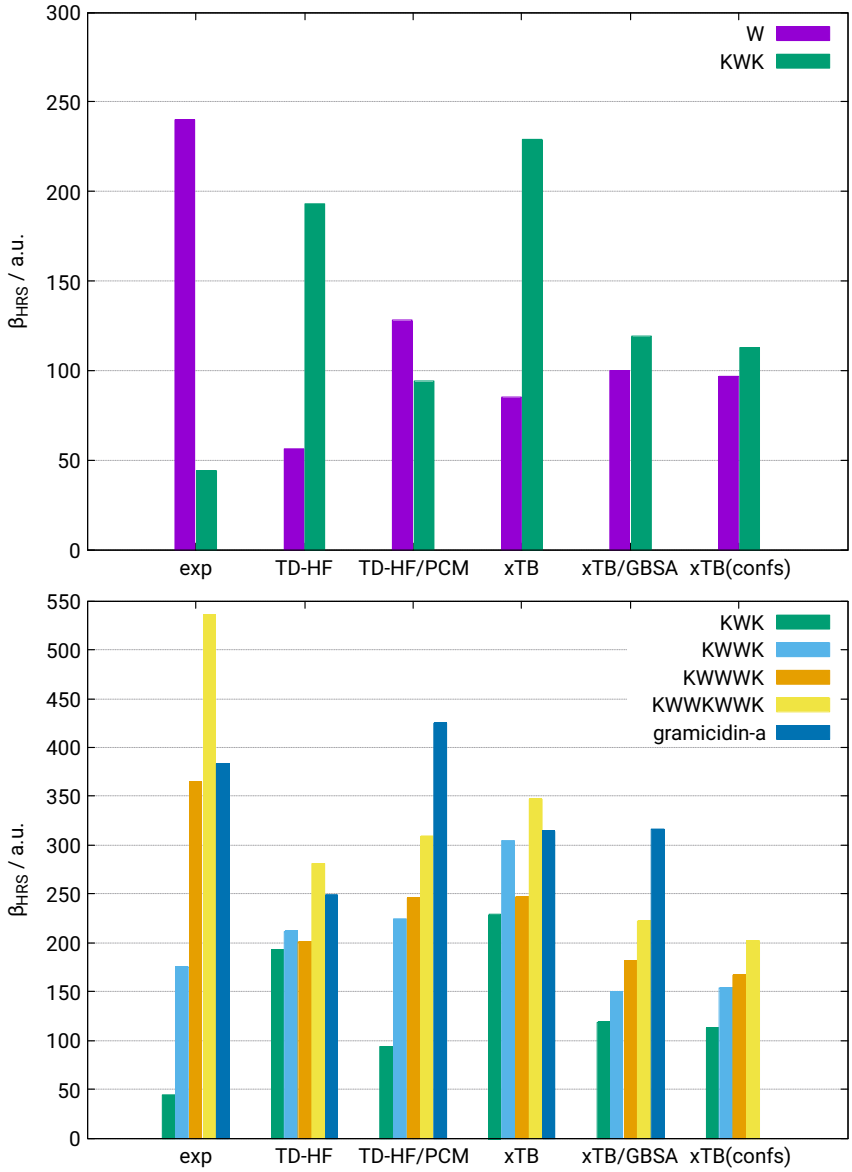


Figure 11: Static first hyperpolarizabilities for tryptophan and KWK (top) and KWK, KWWK, KWWWK, KWWKWWK, gramicidin A (bottom) extrapolated from experiment and computed with TDHF/6-31+G(d), TDHF/6-31+G(d)/PCM, sTD-DFT-xTB, sTD-DFT-xTB/GBSA and Boltzmann weighted ensemble with sTD-DFT-xTB/GBSA.

is also able to provide correct trends except for gramicidin A, but a systematic underestimation with respect to experiment is observed, at least for the largest systems considered here. Note that the value obtained for the rigid gramicidin A is not so far from the experimental one. Including all relevant conformers has a small impact on the β_{HRS} values. The effect of weighting the conformers becomes more pronounced when changes between indole subunit orientations become important. However, accurate computation of free energies in solution for a proper Boltzmann weighting is difficult. Since the individual conformers largely differ in their β_{HRS} values, a slightly miscalculated population could lead directly to a bad result Table S5 shows the relative error for all methods. The relative mean absolute errors (MAEs, in a.u.) of all methods are substantial, ranging from 0.94 (TDHF) to 0.46 (TDHF/IEF-PCM). The sTD-DFT-xTB/GBSA method provides a MAE of 0.62 very close to TDHF/IEF-PCM. Advances in including non-equilibrium solvent effects at the sTD-DFT-xTB/GBSA level could close this gap. This is especially remarkable considering that the sTD-DFT-xTB/GBSA calculations are 3-5 orders of magnitude faster than at the TDHF/IEF-PCM level of theory (see Table S3).

Conclusion

We have presented quantum chemical calculations and a structure-property analysis for NLO properties of tryptophan-rich model peptides. The sTD-DFT-xTB scheme enables computations for systems with up to several thousands of atoms and/or to screen large sets of structures. We used this method to sample the first hyperpolarizabilities with respect to structural changes for the flexible tryptophan-rich peptide chains. For this purpose, molecular dynamics simulations as well as conformational sampling were carried out. This was done with the help of the efficient tight-binding based method GFN2-xTB and a recently proposed approach for exploring the potential energy surface with meta-dynamics. To fine-tune the sTD-DFT-xTB scheme for the desired model peptides, two parameters were adjusted to

reproduce CCSD(T) reference β values for tryptophan.

We first studied the conformer ensemble in terms of the relation between the relative orientations of indole moieties and the first hyperpolarizability. We found that a drastic change in β_{HRS} values relates to different alignments of indole dipoles among conformers. When they are aligned parallel, the value is enhanced while it diminishes when the dipole moments are pointing in opposite directions leading to first hyperpolarizability values of conformers that can differ by a factor of up to two. This highlights the importance of finding the lowest energy conformer when calculating SHG response. The unit-sphere representations clearly showed in most of the case the dipolar character of the β tensor. The assumption that the sum of indole dipole vectors is correlated to the intensity of the first hyperpolarizability was confirmed by model system calculations. The results of the MD simulations indicate additionally a very strong sensitivity of the first hyperpolarizability to details of the molecular structure. Overall it seems essential to properly explore the conformational space of flexible chromophores or when multiple chromophore orientations are possible.

In the second part, we compared sTD-DFT-xTB computed first hyperpolarizabilities to the TDHF values and to experiment. sTD-DFT-xTB and TDHF perform similarly in terms of reproducing the experimental trend of SHG response. Cases where the methods produce incorrect order of values with respect to the size of the system was found at both level of theories. This is mostly when solvent effects are not included. The sTD-DFT-xTB method is able to provide SHG response values for tryptophan-rich systems at a fraction of the cost of the usually-used TDHF level of theory. The comparison of theoretical with experimental values shows that getting a quantitative agreement is a challenging task. TDHF/IEF-PCM performs slightly better than sTD-DFT-xTB/GBSA probably due to the inclusion of additional non-equilibrium solvent effects. The sTD-DFT-xTB scheme can be routinely applied to systems that are inaccessible with TDHF, while still expecting reasonable accuracy. Considering this efficiency, a future study could investigate how the relation between the first hyperpolarizability and the number of tryptophan units evolve for larger peptides. This

relation is not expected to be linear because of the sensitivity of the property on indole unit orientations. Another area of future research is the impact of an explicit solvation on the conformational ensembles and their SHG response.

Disclosure statement

The authors declare no competing financial interest.

Funding

This work was supported by the DFG in the framework of the Gottfried-Wilhelm-Leibniz prize.

References

(1) Campagnola, P. J.; Loew, L. M. *Nature Biotechnology* **2003**, *21*, 1356–1360.

(2) Reeve, J. E.; Anderson, H. L.; Clays, K. *Physical Chemistry Chemical Physics* **2010**, *12*, 13484–13498.

(3) Campagnola, P. *Analytical Chemistry* **2011**, *83*, 3224–3231.

(4) Pavone, F. S.; Campagnola, P. J. *Second harmonic generation imaging*; CRC Press, 2013.

(5) Deniset-Besseau, A.; Duboisset, J.; Benichou, E.; Hache, F.; Brevet, P.-F.; Schanne-Klein, M.-C. *The Journal of Physical Chemistry B* **2009**, *113*, 13437–13445.

(6) de Wergifosse, M.; de Ruyck, J.; Champagne, B. *The Journal of Physical Chemistry C* **2014**, *118*, 8595–8602.

(7) Harczuk, I.; Vahtras, O.; Ågren, H. *Journal of Physical Chemistry Letters* **2016**, *7*, 2132–2138.

(8) De Meulenaere, E.; Asselberghs, I.; de Wergifosse, M.; Botek, E.; Spaepen, S.; Champagne, B.; Vanderleyden, J.; Clays, K. *Journal of Materials Chemistry* **2009**, *19*, 7514–7519.

(9) De Meulenaere, E.; de Wergifosse, M.; Botek, E.; Spaepen, S.; Champagne, B.; Vanderleyden, J.; Clays, K. *Journal of Nonlinear Optical Physics and Materials* **2010**, *19*, 1–13.

(10) De Meulenaere, E.; de Wergifosse, M.; Botek, E.; Vanderleyden, J.; Champagne, B.; Clays, K. **2015**, 522–525.

(11) De Meulenaere, E.; Nguyen Bich, N.; de Wergifosse, M.; Van Hecke, K.; Van Meervelt, L.; Vanderleyden, J.; Champagne, B.; Clays, K. *Journal of the American Chemical Society* **2013**, *135*, 4061–4069.

(12) de Wergifosse, M.; Botek, E.; De Meulenaere, E.; Clays, K.; Champagne, B. *The Journal of Physical Chemistry B* **2018**, *122*, 4993–5005.

(13) Duboisset, J.; Matar, G.; Besson, F.; Ficheux, D.; Benichou, E.; Russier-Antoine, I.; Jonin, C.; Brevet, P. F. *Journal of Physical Chemistry B* **2014**, *118*, 10413–10418.

(14) de Wergifosse, M.; Grimme, S. *Journal of Chemical Physics* **2018**, *149*, 024108.

(15) Grimme, S. *J. Chem. Phys.* **2013**, *138*, 244104.

(16) Grimme, S.; Bannwarth, C. *J. Chem. Phys.* **2016**, *145*, 054103.

(17) Grimme, S. *Journal of Chemical Theory and Computation* **2019**, *15*, 2847–2862.

(18) Grimme, S.; Bannwarth, C.; Dohm, S.; Hansen, A.; Pisarek, J.; Pracht, P.; Seibert, J.; Neese, F. *Angewandte Chemie - International Edition* **2017**, *56*, 14763–14769.

(19) Pracht, P.; Bohle, F.; Grimme, S. *Phys. Chem. Chem. Phys.* **2020**, *submitted*.

(20) Bannwarth, C.; Ehlert, S.; Grimme, S. *Journal of Chemical Theory and Computation* **2018**, *15*, 1652–1671.

(21) Grimme, S.; Bannwarth, C.; Shushkov, P. *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

(22) Shushkov, P.; Grimme, S. manuscript in preparation.

(23) Still, W.; Tempczyk, A.; Hawley, R.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(24) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. *J. Chem. Phys.* **2015**, *143*, 54107.

(25) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *0*, 799–805.

(26) Zhao, Y.; Truhlar, D. G. *Journal of Physical Chemistry A* **2005**, *109*, 5656–5667.

(27) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(28) Eckert, F.; Klamt, A. *AIChE Journal* **2002**, *48*, 369–385.

(29) COSMOtherm, C3.0, release 1601, COSMOlogic GmbH & Co KG, http://www.cosmologic.de.

(30) Grimme, S. *Chemistry - A European Journal* **2012**, *18*, 9955–9964.

(31) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(32) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311–1327.

(33) Yu, J.; Zerner, M. C. *The Journal of Chemical Physics* **1994**, *100*, 7487–7494.

(34) Zhu, W.; Wu, G. S. *Chemical Physics Letters* **2002**, *358*, 1–7.

(35) Yamaguchi, Y.; Yokomichi, Y.; Yokoyama, S.; Mashiko, S. *Journal of Molecular Structure: THEOCHEM* **2002**, *578*, 35–45.

(36) Cammi, R.; Cossi, M.; Mennucci, B.; Tomasi, J. *Journal of Molecular Structure* **1997**, *436-437*, 567–575.

(37) Dunning, T. H. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.

(38) de Wergifosse, M.; Liégeois, V.; Champagne, B. *International Journal of Quantum Chemistry* **2014**, *114*, 900–910.

(39) Cancès, E.; Mennucci, B.; Tomasi, J. *Journal of Chemical Physics* **1997**, *107*, 3032–3041.

(40) Cancès, E.; Mennucci, B. *Journal of Chemical Physics* **2001**, *114*, 4744–4745.

(41) Verbiest, T.; Clays, K.; Rodriguez, V. *Second-order nonlinear optical characterization techniques: an introduction*; CRC press, 2009.

(42) Bersohn, R.; Pao, Y.; Frisch, H. L. *The Journal of Chemical Physics* **1966**, *45*, 3184–3198.

(43) Castet, F. F.; Bogdan, E.; Plaquet, A.; Ducasse, L.; Champagne, B.; Rodriguez, V. *Journal of Chemical Physics* **2012**, *136*, 24506.

(44) Tuer, A.; Krouglov, S.; Cisek, R.; Tokarz, D.; Barzda, V. *Journal of Computational Chemistry* **2011**, *32*, 1128–1134.

(45) Guthmuller, J.; Simon, D. *Journal of Physical Chemistry A* **2006**, *110*, 9967–9973.

(46) Willetts, A.; Rice, J. E.; Burland, D. M.; Shelton, D. P. *The Journal of Chemical Physics* **1992**, *97*, 7590–7599.

(47) Shao, Y. et al. *Molecular Physics* **2015**, *113*, 184–215.

(48) Frisch, M. J. et al. Gaussian 09. 2009.

(49) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100.

(50) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(51) The xtb code can be obtained from:. `https://github.com/grimme-lab/xtb`.

(52) The stda code can be obtained from:. `https://github.com/grimme-lab/stda/`.

(53) Campo, J.; Wenseleers, W.; Goovaerts, E.; Szablewski, M.; Cross, G. H. *Journal of Physical Chemistry C* **2008**, *112*, 287–296.

(54) Mançois, F.; Pozzo, J. L.; Pan, J.; Adamietz, F.; Rodriguez, V.; Ducasse, L.; Castet, F.; Plaquet, A.; Champagne, B. *Chemistry - A European Journal* **2009**, *15*, 2560–2571.

(55) Oudar, J. L.; Chemla, D. S. *The Journal of Chemical Physics* **1977**, *66*, 2664–2668.

(56) Bannwarth, C.; Grimme, S.; Seibert, J.; Grimme, S. *Chirality* **2016**, *28*, 365–369.

(57) Seibert, J.; Bannwarth, C.; Grimme, S. *J. Am. Chem. Soc.* **2017**, jacs.7b05833.

(58) Seibert, J.; Pisarek, J.; Schmitz, S.; Bannwarth, C.; Grimme, S. *Molecular Physics* **2018**, *0*, 1–13.