Ab initio Calculations of the Linear and Nonlinear Optical Properties of Amino Acids

D Tokarz, A Tuer, R Cisek, S Krouglov, and V Barzda
Department of Chemical and Physical Sciences, Department of Chemistry, Department of Physics, and Institute for Optical Sciences, University of Toronto, 3359 Mississauga Road North, Mississauga, ON L5L 1C6, Canada
E-mail: virgis.barzda@utoronto.ca

Abstract. A number of proteins can assemble into chiral structures that display strong nonlinear optical activity. For instance, proteins such as myosin and collagen exhibit intense second harmonic generation (SHG). A large number of experimental studies on the SHG of proteins have been conducted; however, few predictive models have been proposed that reliably relate the macroscopic SHG properties to the amino acids present in the peptidic chain. In this study, the linear polarizability (\(\alpha\)), first (\(\beta\)) and second hyperpolarizability (\(\gamma\)) of all twenty amino acids was investigated by time-dependent Hartree-Fock calculations under physiological conditions. Ab initio calculations were performed using the GAMESSUS computational chemistry package. We have found that the aromatic amino acids give rise to the largest mean \(\alpha\), \(\beta\) and \(\gamma\) values. With this finding, we hope to apply this method to protein structures in order to understand how second harmonic signal is generated from individual amino acids, as well as, recognize how manipulation of the secondary structure of proteins might enhance SHG and third harmonic generation (THG).

1. Introduction
Nonlinear optical microscopy based on second harmonic generation (SHG) and third harmonic generation (THG), is currently an emerging technique used in studying biological structures (see review [1]). Harmonic generation microscopy is used to achieve distinct contrast in living tissue [2] [3]. Although both SHG and THG are coherent light processes, SHG only occurs in media with non-central symmetry. THG can be generated in homogeneous media, but when focusing under tight conditions with a microscope, THG is generated at interfaces [4]. For example, THG has been found to originate from multilayer structures such as biological membranes and cell walls [5] [6] while SHG has been previously found to originate from proteins such as myosin and collagen which consist of ordered helical structures [7] [8] [9]. Due to symmetry restrictions, media exhibits SHG signals at well-defined laser polarizations [10]. This information can be used to determine the orientation of proteins in tissues [10]. Therefore, information related to the organization of the secondary structures of proteins can be extracted from SHG data [10].

1 To whom any correspondence should be addressed.
A large number of experimental studies on the SHG of proteins have been conducted though, few predictive models have been proposed that reliably relate the macroscopic SHG measurements to the structure of proteins. For further understanding of the origin of SHG and THG from various biological structures, modeling of the nonlinear response from molecular structures with various structural organizations is required. In this study, we investigated a relatively simple model for interpreting the nonlinear optical properties of proteins. This was done by examining the influence of the primary structure of proteins to generate both second harmonic and third harmonic signal. 

**2. Theory**

The dipole moment of a molecule when induced by an electric field can be expressed in a Taylor series as [14]

\[
\mu = \mu_o + \alpha_{ij} E_i E_j + \frac{1}{2!} \beta_{ijk} E_i E_j E_k + \frac{1}{3!} \gamma_{ijkl} E_i E_j E_k E_l + \ldots
\]

(1)

where \(\mu_o\) is the permanent dipole moment of the molecule in the absence of an electric field, \(E\) is the electric field, \(\alpha\) is the linear polarizability, and coefficients \(\beta\) and \(\gamma\) are the first and second hyperpolarizabilities, respectively. In equation (1) the repeated indices imply summation.

In general, \(\alpha, \beta,\) and \(\gamma\) are dependent on the frequency, \(\omega\), of the incident electric field. The induced dipole yields linear polarization as well as second harmonic generation at \(2\omega\) and third harmonic generation at \(3\omega\).

The mean linear polarizability can be defined as

\[
\langle \alpha \rangle = \frac{1}{3} (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})
\]

(2)

The mean second hyperpolarizability can be expressed as

\[
\langle \gamma \rangle = \frac{1}{5} (\gamma_{xxxx} + \gamma_{yyyy} + \gamma_{zzzz} + \gamma_{xxyy} + \gamma_{xxzz} + \gamma_{yyzz} + \gamma_{yyyy} + \gamma_{zzzz})
\]

(3)

The first hyperpolarizability is zero for all molecules with inversion symmetry. The first hyperpolarizability is a third rank tensor described by a \(3 \times 3 \times 3\) array. All 27 components of the three dimensional array can be used to find the mean first hyperpolarizability by calculating an average related to the modulus

\[
\langle \beta \rangle = \frac{1}{27} \sqrt{\sum_{ijk} \beta_{ijk}^2}
\]

(4)

The expressions (2), (4), and (5) were used to express the linear and nonlinear optical properties of all amino acids.

**3. Method**

Calculations of the linear polarizability, the second- and third-order hyperpolarizabilities at the fundamental wavelength (1028nm) of our home-built Yb:KGd(WO\(_4\))\(_2\) laser [15] was achieved by applying the \textit{ab initio} time-dependent coupled perturbed Hartree-Fock (TDHF) method at the Restricted Hartree-Fock (RHF) level using GAMESSUS program [12] [13] running parallel on the GPC supercomputer at the SciNet GPC Consortium at the University of Toronto. Calculations were performed...
on a single node which included 8 processors. Geometry optimization of each amino acid did not take longer than a single day. Similarly, the corresponding polarizability and hyperpolarizabilities of each amino acid were calculated within one day. Output geometry optimization files were as large as 430 000 kB whereas hyperpolarizability calculations generated files no larger than 2300 kB.

The ground state geometries of the amino acids were fully optimized without any symmetry restrictions using the 6-311++G** basis set at the DFT/B3LYP level at pH 7 [16]. These molecular structures were then used for nonlinear optical property calculations.

As suggested in literature, reliable calculation of the nonlinear optical properties requires increased convergence parameters which included the following GAMESS options: ITOL=30 (the products of primitives whose exponential factor is less than $10^{-30}$ are omitted), ICUT=20 (integrals less than $10^{-20}$ are overlooked), INTTYP=HONDO (HONDO/Rys integrals evaluated), ATOL=1.0D-07 (the tolerance for convergence of first-order results is 1.0$^{-7}$), BTOL=1.0D-07 (tolerance for convergence of first-order results is 1.0$^{-7}$) and NCONV=10 (when the absolute value of the density change between two consecutive self-consistent field cycles is less than $10^{-10}$ then convergence is achieved) [17]. Nonlinear optical calculations were also done using 6-311++G** basis set. It has been shown previously that a basis set augmented with diffuse functions is necessary when computing the nonlinear optical properties of molecules [18].

4. Results

The mean linear polarizability, first hyperpolarizability, and second hyperpolarizability for all twenty amino acids at pH7 were found by applying equations (2), (4), and (3), respectively. Results are shown in table 1, where a few trends are observed.

Table 1. The average linear polarizability, first hyperpolarizability, and second hyperpolarizability of amino acids organized by chemical derivative.

| Amino Acid         | $<\alpha>$ (10$^{-24}$ esu) | $<\beta>$ (10$^{-32}$ esu) | $<\gamma>$ (10$^{-36}$ esu) |
|--------------------|-----------------------------|-----------------------------|-----------------------------|
| Hydrocarbon        |                             |                             |                             |
| Alanine            | 7.94                        | 5.50                        | 9.80                        |
| Glycine            | 6.24                        | 6.11                        | 8.27                        |
| Isoleucine         | 12.9                        | 6.15                        | 13.5                        |
| Leucine            | 12.9                        | 7.30                        | 14.6                        |
| Proline            | 10.8                        | 8.54                        | 15.8                        |
| Valine             | 11.2                        | 6.83                        | 12.7                        |
| Alcohol            |                             |                             |                             |
| Serine             | 8.33                        | 4.06                        | 8.02                        |
| Threonine          | 10.0                        | 4.73                        | 9.46                        |
| Sulfur-containing  |                             |                             |                             |
| Cysteine           | 10.8                        | 14.1                        | 16.0                        |
| Methionine         | 14.3                        | 8.74                        | 17.0                        |
| Second carboxylic acid |                 |                             |                             |
| Aspartic acid      | 11.1                        | 4.26                        | 18.5                        |
| Glutamic acid      | 12.8                        | 7.52                        | 19.2                        |
| Carboxylic acid derivative |             |                             |                             |
| Asparagine         | 10.7                        | 4.51                        | 10.4                        |
| Glutamine          | 12.3                        | 6.84                        | 10.6                        |
From table 1, larger amino acid structures give rise to increased linear polarizabilities due to increased electron density and delocalization. Of all the different chemical derivatives, the aromatic amino acids, phenylalanine, tryptophan, and tyrosine have the largest linear polarizability values. Aromatic amino acids consist of conjugated residues which contribute to the increased polarizability due to the delocalization of $\pi$-electrons. Aromatic amino acids, may for this reason, play a significant role in the nonlinear response of proteins [19]. In particular, tryptophan has the largest mean linear polarizability as well as the largest mean first and second hyperpolarizability values. In general, a large linear polarizability is necessary to obtain substantial hyperpolarizabilities, though there are additional factors associated with increased hyperpolarizability values [20].

Due to symmetry restrictions, the SHG response of a molecule is highly directionally dependent where each individual $\beta$ tensor varies. Eighteen of twenty-seven unique components of the first hyperpolarizability tensor were calculated where Kleinman symmetry was assumed for the other nine $\beta$ tensor elements. An overall orientational average $\beta$ was calculated using equation (4). In particular, phenylalanine, tryptophan, and histidine demonstrate the largest $<\beta>$ values. These molecules are polarizable in multiple directions and therefore, the overall average of $\beta$ appears to be large. We express an orientational average for the first hyperpolarizability however, when amino acids are aligned, such as in proteins, $\beta$ is best expressed by its individual tensor components.

From calculations of the second hyperpolarizability, aromatic amino acids demonstrate the largest $\gamma$ values due to increased $\pi$-electron delocalization. However, the second carboxylic acid amino acids including aspartic acid and glutamic acid also give rise to large $\gamma$ values. Although these two amino acids are not conjugated, both of them contain two carboxylic acid groups, which are deprotonated at pH 7. Electron density is delocalized amongst the deprotonated carboxylic acid groups at both ends of the amino acid which therefore likely enhances the polarizability of aspartic acid and glutamic acid.

5. Conclusions
We have demonstrated that $ab$ initio calculations can help in interpreting the origin of SHG and THG from the primary structure of proteins. Each amino acid alone does not give rise to high nonlinear optical signals however; increased SHG and THG can be achieved by coherently summing the nonlinear response of many amino acids. In particular, the aromatic amino acids are highly polarizable due to $\pi$-electron delocalization and may be desirable candidates for increasing SHG and THG when forming secondary structures. We hope to use the fundamental understanding of the origin of SHG and THG from amino acids to manipulate the secondary structure of proteins in order to enhance the response of SHG and THG.

Acknowledgements
Computations were performed on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund - Research Excellence; and the University of Toronto.
The authors also gratefully acknowledge support in part by the Natural Sciences and Engineering Research Council of Canada and the Ontario Centres of Excellence.

References

[1] Cisek R, Spencer L, Prent N, Zigmantas D, Espie G and Barzda V 2009 Photosynth Res 102 111-41
[2] Sheppard C J R, Gannaway J N, Kompfner R and Walsh D 1977 Ieee Journal of Quantum Electronics 13 D100-D00
[3] Barad Y, Eisenberg H, Horowitz M and Silberberg Y 1997 Appl Phys Lett 70 922-24
[4] Ward J F and New G H C 1969 Phys Rev 185 57-72
[5] Muller M, Squier J, Wilson K R and Brakenhoff G J 1998 J Microsc-Oxford 191 266-74
[6] Squier J A, Muller M, Brakenhoff G J and Wilson K R 1998 Opt Express 3 315-24
[7] Plotnikov S V, Millard A C, Campagnola P J and Mohler W A 2006 Biophys J 90 693-703
[8] Freund I and Deutsch M 1986 Opt Lett 11 94-96
[9] Stoller P, Reiser K M, Celliers P M and Rubenchik A M 2002 Biophys J 82 3330-42
[10] Campagnola P J, Millard A C, Terasaki M, Hoppe P E, Malone C J and Mohler W A 2002 Biophysical Journal 82 493-508
[11] Schmidt M W, Baldridge K K, Boatz J A, Elbert S T, Gordon M S, Jensen J H, Koseki S, Matsunaga N, Nguyen K A, Su S J, Windus T L, Dupuis M and Montgomery J A 1993 J Comput Chem 14 1347-63
[12] Sekino H and Bartlett R J 1986 J Chem Phys 85 976-89
[13] Karna S P and Dupuis M 1991 J Comput Chem 12 487-504
[14] Buckingham A D and Orr B J 1967 Q Rev Chem Soc 21 195-&
[15] Major A, Cisek R, Sandkuijl D and Barzda V 2009 Laser Phys Lett 6 272-74
[16] Becke A D 1993 J Chem Phys 98 5648-52
[17] Cundari T R, Kurtz H A and Zhou T 2001 J Chem Inf Comp Sci 41 38-42
[18] Shelton D P and Rice J E 1994 Chem Rev 94 3-29
[19] Guthmuller J and Simon D 2006 J Phys Chem A 110 9967-73
[20] Marder S R, Gorman C B, Meyers F, Perry J W, Bourhill G, Bredas J L and Pierce B M 1994 Science 265 632-35