

## New tuned range-separated density functional for the accurate calculation of second hyperpolarizabilities

Pau Besalú-Sala,<sup>a</sup> Sebastian P. Sitkiewicz,<sup>abc</sup> Pedro Salvador,<sup>\*a</sup> Eduard Matito,<sup>\*bd</sup> and Josep M. Luis<sup>\*a</sup>

The calculation of nonlinear optical properties (NLOs) using density functional theory (DFT) remains a challenge in computational chemistry. Although existing range-separated functionals display the best performance for the calculation of this type of properties, their errors strongly depend on the family of molecules studied. Herein, we have explored a new strategy to empirically tune the range-separated LC-BLYP method to improve the accuracy of the calculation of the second hyperpolarizabilities ( $\gamma$ ), which are poorly described by current density functional approximations. First, we benchmarked nine of the most accurate commonly used range-separated hybrid and optimally-tuned functionals (*i.e.* B3LYP, PBE0, BH&HLYP, M06-2X, MN15,  $\omega$ B97XD, CAM-B3LYP, LC-BLYP and OT-LC-BLYP) for the calculation of  $\gamma$  using as reference the CCSD(T) values of a chemically diverse set of 60 molecules. Among these nine functionals, LC-BLYP gives the lowest average errors. We determined the value of the range-separation parameter  $\omega$  required to reproduce the CCSD(T) second hyperpolarizabilities with LC-BLYP functional ( $\omega_{cc}$ ) for the set of 60 molecules. Our new tuned range-separated functional,  $T\alpha$ -LC-BLYP, employs a quadratic correlation between  $\omega_{cc}$  and a molecular descriptor in terms of the linear polarizability and the number of electrons of the molecule. The average error of the  $\gamma$  values obtained with  $T\alpha$ -LC-BLYP is reduced by half or more as compared with the most accurate among the nine density functional approximations benchmarked.

### Introduction

Chemical compounds bearing high nonlinear optical (NLO) features have raised as key building blocks of several devices for a variety of technologies such as electro-optics,<sup>1, 2</sup> optical switching<sup>3-6</sup> or three-dimension fluorescence microscopy.<sup>7-9</sup> In particular, materials bearing high second hyperpolarizabilities ( $\gamma$ ) have gained special interest.<sup>6, 10</sup> Although  $\gamma$  can be measured for instance through third harmonic generation experiments,<sup>11, 12</sup> computational insights are very useful to complement and rationalize experimental data or design new prototypes.

At the molecular level, Hartree-Fock (HF) (hyper)polarizabilities are not accurate enough and, therefore, a correlated wavefunction is required to compute the NLO properties (NLOPs).<sup>13-15</sup> Typically, second-order Møller-Plesset<sup>16</sup> (MP2) perturbation theory, coupled-cluster<sup>17</sup> with singles and doubles (CCSD), or including triple excitations through perturbative treatment, *i.e.* CCSD(T), together with a flexible basis set,<sup>14, 18</sup> are considered appropriate methodologies to obtain accurate values of  $\gamma$ .<sup>19-22</sup> Unfortunately, post-HF approaches become computationally too expensive as the size of the molecule increases.

Kohn-Sham<sup>23</sup> density functional theory (KS-DFT) is a powerful alternative to correlated wavefunction methods due to its good balance between computational cost and accuracy. However, the expression for the exact functional is not known, existing a large variety of alternative approximate definitions. Most common density functional approximations (DFAs) available nowadays are designed and constructed to accurately predict kinetic and thermodynamic properties directly related to the difference of Gibbs energies, such as reaction barriers and reaction energies,<sup>24</sup> regardless of other observables such as electro-optical properties. For this reason, molecular electric properties still remain a challenge for most common DFAs, specially concerning the prediction of NLOPs.<sup>25</sup> This failure is related with the wrong asymptotic behaviour of approximate exchange–correlation (XC) potentials, the “near-sightedness” of the XC response kernels, and the lack of the integer discontinuity.<sup>26, 27</sup> Instead, the HF potential does have the proper asymptotic behaviour and HF exchange does not suffer from self-interaction errors. The range-separated hybrid (RSH) functionals that include 100% of exact HF exchange at large electron-electron distances correctly describe the asymptotic behaviour of XC potentials, and are one of the DFA types providing the most accurate NLOPs up to date.<sup>19, 21, 22</sup>

Further improvement on the results of the RSH functionals can be ideally achieved by tuning the RSH functionals.<sup>26, 28-30</sup> For instance, Kronik and co-workers reproduced experimental band gaps<sup>31</sup> and optical absorption energies<sup>32</sup> for a series of organic solids applying one- and/or two-dimensional optimally tuning (OT) protocols. The most popular strategy for the construction of OT-RSH functionals is based on the optimization of the range-

<sup>a</sup> Departament de Química and Institut de Química Computacional i Catàlisi (IQCC), Universitat de Girona, M. Aurèlia Capmany 69, 17003, Girona, Spain.

<sup>b</sup> Donostia International Physics Center (DIPC), P.K. 1072, 20018 Donostia, Euskadi, Spain.

<sup>c</sup> Kimika Fakultatea, Euskal Herriko Unibertsitatea (UPV/EHU), 20080, Donostia, Euskadi, Spain.

<sup>d</sup> Ikerbasque, Basque Foundation for Science, 48013 Bilbao, Euskadi, Spain.

Electronic Supplementary Information (ESI) available. See DOI: 10.1039/x0xx00000x

## ARTICLE

separation parameter in the RSH to satisfy the DFT version of Koopmans' theorem.<sup>33, 34</sup>

Several research groups have previously explored the behaviour of RSH and OT-RSH functionals to compute the NLOP of a specific type of target molecules such as conjugated polymeric chains.<sup>21, 22, 25, 35</sup> They have shown that a large drawback of DFAs is the presence of the delocalization error. The minimization of the delocalization error together with an optimal tuning of  $\omega$ , is the key to improve the description of some systems.<sup>26</sup> Particularly, for lower-order response properties such as the electric first-hyperpolarizability ( $\beta$ ),  $\omega$  tuning largely improves the results of push-pull systems such as p-nitroaniline (PNA) or dimethylaminonitrostyrene (DANS) oligomers.<sup>36, 37</sup>

Unfortunately, this tuning scheme is, in general, not accurate enough to provide reliable results for  $\gamma$  calculations, giving sometimes even worse results than their corresponding non-optimized RSH functional.<sup>21, 38</sup> Indeed, for cases like conjugated organic oligomers and polymers, which have raised a lot of interest in the field of nonlinear optics due to their very high NLOPs, the usual OT-RSH even decreases the accuracy for  $\gamma$ 's.<sup>19, 21</sup> Therefore, a novel and more specific strategy is required to compute accurate second hyperpolarizabilities at a moderate computational cost.

In this work, we present a new alternative tuning scheme based on the correlations found between the linear polarizability ( $\alpha$ ) obtained at LC-BLYP level of theory and the optimal value of range-separation parameter  $\omega$  that reproduces the values of  $\gamma$  obtained at CCSD(T) level. Furthermore, to the best of our knowledge, we have constructed one of the largest benchmark molecular sets including  $\gamma$  values computed at CCSD(T) level of theory.

## Theory and methodology

### Range-separated functionals

In RSH functionals, the electron-electron repulsion operator  $1/r_{12}$  for the exchange energy is split into short-range (SR) and long-range (LR) terms depending on the interelectronic distance:<sup>39</sup>

$$\frac{1}{r_{12}} \equiv \frac{1 - [\lambda + \kappa \cdot \text{erf}(\omega \cdot r_{12})]}{r_{12}} + \frac{\lambda + \kappa \cdot \text{erf}(\omega \cdot r_{12})}{r_{12}}, \quad (1)$$

where  $\text{erf}$  is the standard error function;  $r_{12}$  is the interelectronic distance between electrons 1 and 2;  $\lambda$  governs the amount of exact exchange (eX) incorporated at both SR and LR;  $\kappa$  controls the extra amount of eX added only at LR; and  $\omega$  is the range-separation parameter, defining the steepness of the error function and the interelectronic distance at which the transition from SR to LR takes place. If the main modification of the exchange potential with respect to the original functional occurs at large interelectronic distances, the functional is called long-range corrected (LC) functional. The implementation of the range-separated electron-electron repulsion operator into the exchange functional leads to the following expression for the exchange-correlation energy:

$$E_{xc}^{RSH} = E_x^{SR} + E_{x,\omega}^{LR} + E_c \quad (2.a)$$

$$E_x^{SR} = \lambda E_{eX}^{SR} + (1 - \lambda) E_{DFAx}^{SR} \quad (2.b)$$

$$E_{x,\omega}^{LR} = (\lambda + \kappa) E_{eX,\omega}^{LR} + (1 - \lambda - \kappa) E_{DFAx,\omega}^{LR} \quad (2.c)$$

where  $E_{eX}$  is the exact HF exchange,  $E_{DFAx}$  denotes the approximate (semi)local exchange energy, and  $E_c$  is the correlation energy.

### Tuned range-separated functionals

In the most popular RSH functionals, the range-separation parameter  $\omega$  is given as a universal constant defined after a fitting procedure against an appropriate dataset. For instance, the parameter values of Eq. 2 for three common RSH functionals are: LC- $\omega$ PBE<sup>40</sup> ( $\lambda=0.0$ ,  $\kappa=1.0$ ,  $\omega=0.40$  Bohr<sup>-1</sup>), CAM-B3LYP<sup>41</sup> ( $\lambda=0.19$ ,  $\kappa=0.46$ ,  $\omega=0.33$  Bohr<sup>-1</sup>) and LC-BLYP<sup>39</sup> ( $\lambda=0.0$ ,  $\kappa=1.0$ ,  $\omega=0.47$  Bohr<sup>-1</sup>). Tuned RSH functionals are usually based on one of these three widely accepted RSH functionals, in which the value of one (usually  $\omega$ ), two ( $\omega$  and  $\kappa/\lambda$ ) or three ( $\omega$ ,  $\kappa$  and  $\lambda$ ) parameters are tuned specifically for each chemical system.<sup>33</sup>

Among the nonempirical tuning schemes, the most popular approach is based on optimizing the value of  $\omega$  to enforce the fulfilment of the DFT version of the Koopmans' theorem (*i.e.* Janak's theorem).<sup>34, 42</sup> This theorem states that for the exact Kohn-Sham functional the energy of the highest occupied molecular orbital (HOMO) is equal to the negative of the ionization potential,  $\epsilon_{\text{HOMO}} = -IP$ .<sup>33</sup> Usually this condition is imposed for both the neutral and the anionic state, where the ionization potential corresponds to the electron affinity of the neutral state.<sup>34</sup> Thus, the optimal tuning of the  $\omega$  parameter is obtained through the minimization of the following expression for a  $N$ -electron system:

$$J^2(\omega) = \sum_{i=N}^{N+1} [IP^i(\omega) + \epsilon_{\text{HOMO}}^i(\omega)]^2 \quad (3)$$

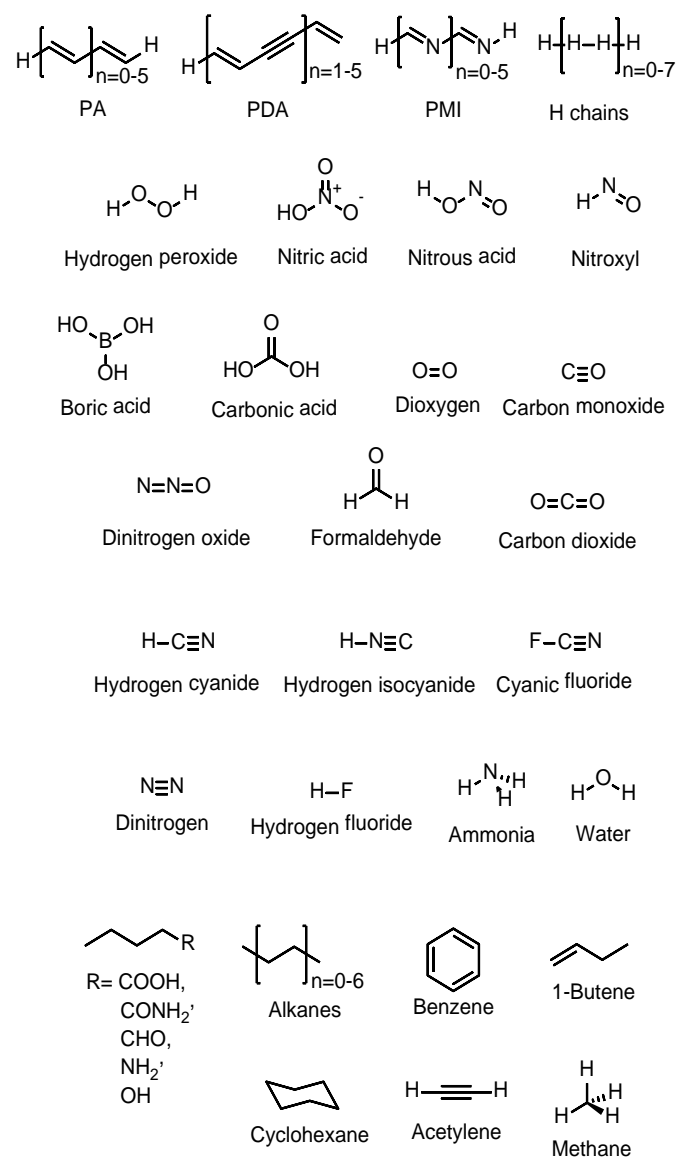
Unfortunately, this approach to design OT-RSH functionals does not systematically improve the accuracy of the second hyperpolarizabilities.<sup>19, 21, 43</sup> Therefore, clearly another approach to tune  $\omega$  should be used to design new RSH functionals for the evaluation of high-order NLOPs.

### Molecular set

The molecular set (Figure 1) used for the development of our new OT-RSH functional for computing static second hyperpolarizabilities ( $\gamma$ -NLO set hereafter) is built up with 60 molecular systems containing from 2 to 36 atoms of the second period and/or hydrogen. The first five/six oligomers of three of the most typical  $\pi$ -conjugated NLO polymers such as all-trans polyacetylene (PA), polydiacetylene (PDA) and all-trans polymethyneimine (PMI) are included in the set. Some small organic and inorganic molecules, and hydrogen chains (which are known to be challenging systems for the calculation of the second hyperpolarizability) are also included. Except dioxygen, all molecules in the set present a closed-shell ground state.

### Computational details

All electronic structure calculations were carried out using Gaussian09 Revision E.01<sup>44</sup> package. Molecular orbitals were constructed using Dunning's<sup>45</sup> correlation-consistent double- $\zeta$  basis set including diffuse functions, aug-cc-pVDZ, which is flexible enough to properly reproduce  $\gamma$  for medium and large organic molecules.<sup>18</sup>



**Fig 1.** Molecular structures of the 60 molecules of the  $\gamma$ -NLO set.

For all DFT calculations, numerical integrations were performed by using the default grid in Gaussian09, which corresponds to a pruned grid of 75 radial shells and 302 angular points per shell. The spin-unrestricted formalism was used for O<sub>2</sub> (triplet ground state) and open-shell singlet *o*-, *m*-, and *p*-benzyne. The geometries of molecular set 2 (*vide infra*) and all systems present in the  $\gamma$ -NLO set but hydrogen chains were optimized at the MP2 level of theory, requesting tight convergence criteria on the root-mean square (RMS) and maxima of the forces and displacements as defined in Gaussian09. Molecular set 1 geometries were optimized using CCSD, which properly reproduces the CASPT2 singlet-triplet split for these diradical molecules.<sup>46</sup> For hydrogen chains, the alternated separation of 2.0 and 3.0 Bohr between hydrogen atoms was set in order to design a system consisting of moderately interacting molecules of dihydrogen, thus avoiding scenarios where the inclusion of strong or nondynamic correlation is required. All minima in the potential energy surface were characterized by analytical vibrational frequency calculations. All molecules were reoriented after the optimization in order to set the Z axis as the

principal inertia axis with the highest moment of inertia, allowing us to focus only on the longitudinal component of the polarizability matrix (*i.e.*  $\alpha_{zz}$ ) and the second hyperpolarizability tensor (*i.e.*  $\gamma_{zzzz}$ ).

The CCSD(T) single-point energy calculations necessary for the finite-field calculations (see SI1, section 6) were done with an energy convergence criterion of 10<sup>-8</sup> atomic units (a.u.). The cutoff of the two-electron integrals accuracy was set to 10<sup>-11</sup> a.u. For all the molecules of  $\gamma$ -NLO set, set 1, and set 2, the T1 diagnostic was computed to confirm that CCSD(T) energies and  $\gamma$  values can be considered a good reference. *O*-, *m*- and *p*-benzyne have an open-shell singlet ground state. The largest CCSD amplitudes for the benzynes are given by t2 amplitudes, but they are lower than 0.6, and therefore the perturbative (T) correction is not problematic.<sup>47</sup> Furthermore, in spite of their multireference character, CCSD(T) properly reproduces the experimental singlet-triplet splitting of *o*-, *m*- and *p*-benzyne.<sup>47</sup> CCSD(T) static linear polarizability and second hyperpolarizabilities were computed as derivatives of the electronic energy with respect to an external electric field using the finite field (FF) approach,

$$\alpha_{zz} = -\left. \frac{\partial^2 E}{\partial F_z^2} \right|_{F_z=0} \quad (4)$$

$$\gamma_{zzzz} = -\left. \frac{\partial^4 E}{\partial F_z^4} \right|_{F_z=0}$$

At the DFT level, both  $\alpha$  and  $\beta$  can be analytically determined. Hence,  $\gamma$  was calculated numerically from the first derivative of  $\beta$  with respect to the external electric field using the FF approach. The number of field strengths chosen in the FF procedure varies for each molecule and method (DFAs or CCSD(T)). In particular, the range of field strength used was  $F=2^j \cdot 10^{-4}$  a.u. with integer values of  $j=0-7$  for all molecules, and the addition of  $j=8-10$  when the former field strengths were not enough to obtain converged numerical derivatives.

The truncation error affecting the numerical estimation of the derivatives, which comes from the higher-order terms neglected in the standard central-differentiation formulae, was minimized thanks to application of the Rutishauser-Romberg scheme:<sup>48,49</sup>

$$R^{i,j} = \frac{4^i R^{(i-1),j} - R^{(i-1),(j+1)}}{4^i - 1} \quad (5)$$

where  $R$  is the calculated property,  $i$  is the iteration number, and  $j$  indicates that the property has been computed using an electric field-strength of  $2^j \cdot 10^{-4}$  a.u. in the FF procedure. By using this methodology, the average numerical errors of CCSD(T)  $\gamma$  is below 2.2% (see SI2, section 1 for further details).

## Results and discussion

### New tuned RSH for NLO properties

RSH functionals usually provide a better description of NLOPs than regular global hybrid DFAs.<sup>25, 43</sup> A possible approach to improve the accuracy of the NLOP obtained with the RSH functionals is tuning the parameters that determine their short-range and long-range regions in the calculation of the exchange energy. Unfortunately, the most popular tuned RSH functionals, which are based on the minimization of  $J^2$ , do not improve the

## ARTICLE

**Table 1.**  $I_\alpha$  descriptor for the  $\gamma$ -NLO set molecules in increasing order and the corresponding  $\omega_{CC}$  and  $\omega_{T\alpha}$  values. All values are given in atomic units.

System	$I_\alpha$	$\omega_{CC}$	$\omega_{T\alpha}$
Hydrogen fluoride	-0.20	0.46	0.49
Nitric acid	-0.04	0.30	0.40
Carbonic acid	-0.03	0.31	0.40
Boric acid	-0.03	0.38	0.39
Dioxygen	-0.03	0.44	0.39
Water	-0.01	0.38	0.38
Hydrogen peroxide	0.02	0.37	0.37
Dinitrogen	0.03	0.44	0.36
Carbon monoxide	0.04	0.37	0.36
Nitroxyl	0.07	0.30	0.35
Cyanic fluoride	0.08	0.30	0.35
Carbon dioxide	0.08	0.49	0.35
Nitrous acid	0.10	0.27	0.34
Ammonia	0.12	0.38	0.33
Formaldehyde	0.15	0.35	0.33
1-Pentanoic acid	0.17	0.32	0.32
Cyclohexane	0.18	0.34	0.32
1-Pentanamide	0.19	0.31	0.32
Dinitrogen oxide	0.19	0.45	0.32
Pentanal	0.20	0.31	0.31
n-Butanol	0.20	0.27	0.31
Hydrogen cyanide	0.22	0.33	0.31
Hydrogen isocyanide	0.22	0.28	0.31
Methane	0.23	0.40	0.31
n-Butylamine	0.23	0.32	0.31
Propane	0.23	0.33	0.31
Ethane	0.24	0.38	0.31
Butane	0.25	0.32	0.30
Pentane	0.25	0.30	0.30
Hexane	0.26	0.30	0.30
PMI1	0.27	0.30	0.30
Heptane	0.27	0.30	0.30
Octane	0.28	0.31	0.30
Benzene	0.29	0.24	0.30
1-Butene	0.31	0.33	0.30
PMI2	0.34	0.28	0.30
PA1	0.35	0.25	0.30
Acetylene	0.37	0.31	0.30
PMI3	0.42	0.30	0.30
PA2	0.45	0.21	0.30
PMI4	0.49	0.29	0.31
PDA1	0.51	0.32	0.31

PMI5	0.54	0.31	0.32
PA3	0.55	0.23	0.32
PMI6	0.58	0.28	0.33
PA4	0.63	0.27	0.34
PDA2	0.64	0.37	0.34
PA5	0.70	0.31	0.37
PDA3	0.72	0.41	0.38
PA6	0.75	0.37	0.39
PDA4	0.77	0.43	0.40
H <sub>2</sub>	0.78	0.52	0.41
PDA5	0.81	0.36	0.42
(H <sub>2</sub> ) <sub>2</sub>	0.92	0.50	0.49
(H <sub>2</sub> ) <sub>3</sub>	0.99	0.58	0.55
(H <sub>2</sub> ) <sub>4</sub>	1.04	0.61	0.58
(H <sub>2</sub> ) <sub>5</sub>	1.07	0.62	0.61
(H <sub>2</sub> ) <sub>7</sub>	1.09	0.63	0.63
(H <sub>2</sub> ) <sub>6</sub>	1.09	0.62	0.63
(H <sub>2</sub> ) <sub>8</sub>	1.12	0.64	0.66

accuracy of  $\gamma$ .<sup>21, 26, 38</sup> The main goal of this work is to propose a new type of tuned RSH replacing the minimization of  $J^2$  by an alternative function that provides more accurate second hyperpolarizabilities.

As starting point to build the new OT-RSH functionals, we have used the LC-BLYP method, which is among the best RSH functionals to compute NLOPs.<sup>43</sup> Our first step was to optimize, for each molecule of the  $\gamma$ -NLO set, the  $\omega$  parameter of LC-BLYP in order to reproduce the  $\gamma$  value computed with CCSD(T). We applied the root-finding bisection method<sup>50</sup> until finding the best value of  $\omega$  with a precision of  $\Delta\omega=0.01$ . This accuracy is enough to reproduce the CCSD(T) values of  $\gamma$  (*i.e.* relative errors with respect to CCSD(T) are smaller than 1%) and obtain smooth profiles describing the dependence of LC-BLYP  $\gamma$  in terms of  $\omega$ . This set of optimized  $\omega$  are labelled as  $\omega_{CC}$  (see Table 1).

Our second goal is to find a correlation between  $\omega_{CC}$  and some molecular descriptor or property that can be easily computed for any molecule. In order to achieve this challenging goal, several possible molecular descriptors were tested, such as HOMO-LUMO gap, the first excitation energies computed within time-dependent DFT formalism or delocalization indices.<sup>51-53</sup> We obtained satisfactory correlations between functions based on these molecular descriptors and  $\omega_{CC}$  within subfamilies of the  $\gamma$ -NLO set, as for instance PA chains (see Figures SI2.5-SI2.6). However, we did not find any function based on these descriptors displaying a sufficiently good correlation with  $\omega_{CC}$  and involving all 60 molecules of the set.

Motivated by the idea of using  $\gamma_{zzzz}$  as self-descriptors to predict the best  $\omega$  to reproduce CCSD(T)  $\gamma$  using LC-BLYP, we found good correlations using as molecular descriptor a function of LC-BLYP  $\gamma_{zzzz}$  (see Figure SI2.1). Considering that  $\alpha$  and  $\gamma$  are both properties of even order (second and fourth derivatives with respect to the external electric field, respectively), and that usually their variation with respect to the size and type of the chemical systems follows similar trends (although variations of

$\gamma$  are usually far larger than variations of  $\alpha$ , we also checked the correlation of  $\omega_{CC}$  with  $\alpha_{zz}$ , the calculation of which bears a lower computational cost (see Figure SI2.2). However, since the  $\gamma$ -NLO set includes molecules of different sizes, one wants to consider size-extensive descriptors. This goal could be achieved by using  $\gamma/\alpha$ ,  $\alpha/N$  or  $\gamma/N^2$  (where  $N$  is the number of electrons in the molecule) instead of  $\alpha$  or  $\gamma$ . For instance, by using as descriptor the ratio  $\log(\gamma/\alpha)$ , we obtained fairly good correlations for all molecules of the set with the exception of the hydrogen chains (see Figure SI2.3). The use of a logarithm in the definition of the descriptors is motivated by the restriction of  $\omega_{CC}$ , which only takes values in the interval  $[0,1]$ . The results were even better using as descriptor  $\log(\gamma/N^2)$ , which provides a sufficiently good correlation for all the molecules of the set (see Figure SI2.4). However, the best correlation was obtained using as descriptor the linear polarizability obtained at LC-BLYP level and the number of electrons in the following way:

$$l_\alpha = \log\left(\frac{\alpha_{LC-BLYP}}{N}\right) \quad (6)$$

This descriptor will be used throughout this work (see Figure 2). LC-BLYP  $\alpha$  can be determined analytically using most computational packages. Therefore,  $l_\alpha$  descriptor bears a low computational cost and it can be fairly easily obtained.

As shown in Figure 2, there is fairly good quadratic correlation between  $\omega_{CC}$  and  $l_\alpha$  descriptor (Eq. 7,  $R^2=0.77$ ), providing a prediction of  $\omega$  that leads to second hyperpolarizability values close to CCSD(T).

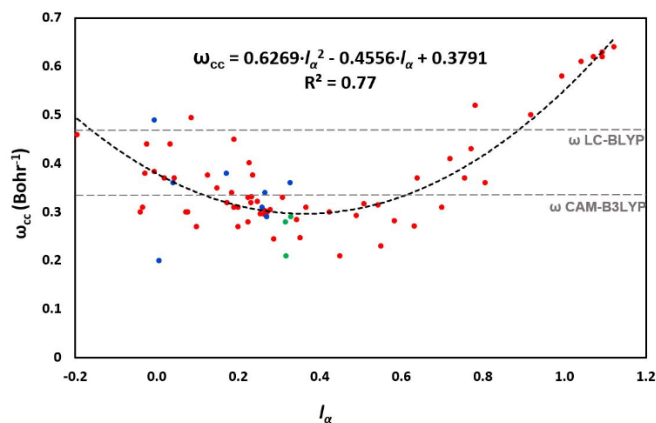
$$\omega_{T\alpha} = 0.6269 \cdot l_\alpha^2 - 0.4556 \cdot l_\alpha + 0.3791 \quad (7)$$

The predicted  $\omega_{T\alpha}$  values obtained with Eq. 7 have a mean absolute error (MAE) with respect to the optimal  $\omega_{CC}$  of only 0.04 Bohr<sup>-1</sup>. The correlation was stable under a leave-N-out ( $N=1, 2, 3$ ) cross-validation tests<sup>54</sup> since the decrease in  $R^2$  was only 0.02 for the three tests, indicating that there is no overfitting of the model.

$\gamma$  values close to CCSD(T) quality could be achieved using the  $\omega_{T\alpha}$  obtained from Eq. 7 within the OT-LC-BLYP framework. We refer to this new tuned range-separated hybrid density functional approach as  $T\alpha$ -LC-BLYP. In order to compute the second hyperpolarizability of a molecule using the  $T\alpha$ -LC-BLYP method, one must first compute the  $\alpha$  value with the LC-BLYP functional, and obtain the corresponding  $\omega_{T\alpha}$  from Eq. 6 and Eq. 7 to be used into the  $T\alpha$ -LC-BLYP functional for the final evaluation of  $\gamma$ .

It is worth mentioning that all tested descriptors based on  $\alpha$  or  $\gamma$  showed two clearly distinct behaviours for molecules with either large or small  $\gamma$ , leading to the classification of the molecules in two subsets. The two subsets present linear correlations between  $\omega_{CC}$  and the (N)LOP-based indices with two very different slopes (see Figure SI2.7). The boundary between the small and large second hyperpolarizability regions is, to some extent, arbitrary.

Nevertheless, in order to keep our scheme as simple as possible, we chose a parabolic function, which can explain the variability of  $\omega_{CC}$  with respect to  $l_\alpha$  for the whole molecular set (see Figure 2). Interestingly, the MAE in the predictions of  $\omega$  using one unique parabolic correlation for the whole set of molecules was lower than using two different linear correlations for each subset. Differences smaller than 0.05 Bohr<sup>-1</sup> between the predicted  $\omega_{T\alpha}$  and the optimal  $\omega_{CC}$  barely affect the final  $\gamma$  obtained for molecules with large second hyperpolarizabilities.



**Fig 2.** Graphical representation of the correlation between  $\omega_{CC}$  and the  $l_\alpha$  descriptor for the  $\gamma$ -NLO set (red). The longitudinal polarizability for the inertia axis with the highest moment of inertia was computed at LC-BLYP level. The points corresponding to set 1 (green) and 2 (blue) molecules are added to the figure but not used in the fitting of the second order polynomial. All values in atomic units.

For this reason, the values of  $\omega$  were rounded at the second decimal. This threshold value could be even increased for molecules presenting low values of  $\gamma$  because then the value of the property becomes much less dependent on the RSH parameter. The small sensitivity of  $\gamma$  to the small variations of  $\omega_{T\alpha}$  is responsible for the good performance of Eq. 7, which accounts for 77% of the variability of the second hyperpolarizability calculated at the CCSD(T) level of theory.

### Benchmarking of DFA for the evaluation of the second hyperpolarizability

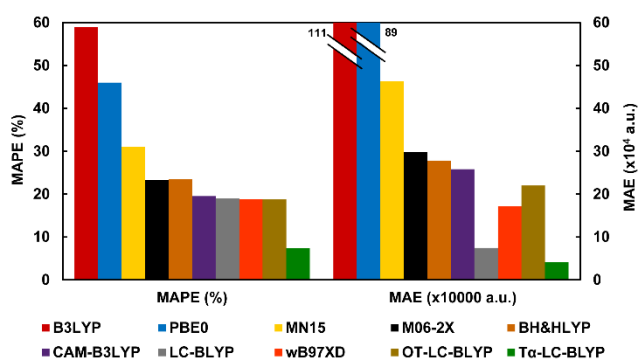
Once the correlation between  $\omega_{CC}$  and  $l_\alpha$  was established, we used the  $T\alpha$ -LC-BLYP functional to compute the static hyperpolarizabilities of the molecules of the  $\gamma$ -NLO set (see Table 2), and compare them to the ones obtained from other nine commonly used DFAs, namely the global hybrid GGAs B3LYP,<sup>55, 56</sup> PBE0,<sup>57</sup> BH&HLYP,<sup>58</sup> the global hybrid meta-GGAs M06-2X,<sup>59</sup> and MN15,<sup>60</sup> the long-range corrected functionals LC-BLYP, CAM-B3LYP, and  $\omega$ B97xD<sup>61</sup> and the optimally-tuned LC-BLYP based on the minimization of  $J^2$  values (Eq. 3), which we labelled as OT-LC-BLYP. We did not include GGA functionals because they usually give far less accurate  $\gamma$ .<sup>43</sup> The results presented in Figure 3 show that the mean absolute percentage error (MAPE) and MAE with respect to CCSD(T) values obtained with  $T\alpha$ -LC-BLYP are clearly lower than the errors obtained with the other nine functionals.

In agreement with the conclusions of previous studies,<sup>43</sup> the average errors in the calculation of  $\gamma$  are lower for the functionals that include more exact exchange. This is true for both hybrid GGAs and hybrid meta-GGAs. In this vein, BH&HLYP and M06-2X provide more accurate values than B3LYP, PBE0, or MN15. As expected, for the  $\gamma$ -NLO set, the most accurate values of  $\gamma$  are obtained by the range-separated functionals LC-BLYP, CAM-B3LYP,  $\omega$ B97xD, and OT-LC-BLYP. Interestingly, the MAPEs obtained for these four DFAs are quite similar and only slightly smaller than for hybrid and meta-hybrids with a high percentage of exact exchange. Although LC-BLYP has a poor performance for the molecules with small  $\gamma$ , it gives more accurate second hyperpolarizabilities than the other RSHs for the molecules of the  $\gamma$ -NLO set with the highest  $\gamma$  values.

**Table 2.**  $\nu_{zzzz}$  for the  $\gamma$ -NLO set molecules in increasing order of the  $I_\alpha$  descriptor computed using CCSD(T) method and various RSHs with the aug-cc-pVDZ basis set. All the results are given in a.u.

System	$\nu_{zzzz}$ (a.u.)				
	CCSD(T)	LC-BLYP	CAM-B3LYP	OT-LC-BLYP	T $\alpha$ -LC-BLYP
Hydrogen fluoride	$2.880 \cdot 10^2$	$2.857 \cdot 10^2$	$3.088 \cdot 10^2$	$2.515 \cdot 10^2$	$2.811 \cdot 10^2$
Nitric acid	$4.032 \cdot 10^3$	$2.932 \cdot 10^3$	$3.629 \cdot 10^3$	$3.074 \cdot 10^3$	$3.296 \cdot 10^3$
Carbonic acid	$4.690 \cdot 10^3$	$3.163 \cdot 10^3$	$4.294 \cdot 10^3$	$3.435 \cdot 10^3$	$3.682 \cdot 10^3$
Boric acid	$3.465 \cdot 10^3$	$2.949 \cdot 10^3$	$3.603 \cdot 10^3$	$3.092 \cdot 10^3$	$3.376 \cdot 10^3$
Dioxygen	$4.097 \cdot 10^2$	$3.877 \cdot 10^2$	$4.430 \cdot 10^2$	$3.617 \cdot 10^2$	$4.632 \cdot 10^2$
Water	$4.740 \cdot 10^2$	$4.369 \cdot 10^2$	$5.015 \cdot 10^2$	$4.179 \cdot 10^2$	$4.757 \cdot 10^2$
Hydrogen peroxide	$1.557 \cdot 10^3$	$1.372 \cdot 10^3$	$1.574 \cdot 10^3$	$1.342 \cdot 10^3$	$1.555 \cdot 10^3$
Dinitrogen	$1.130 \cdot 10^3$	$1.067 \cdot 10^3$	$1.216 \cdot 10^3$	$1.051 \cdot 10^3$	$1.293 \cdot 10^3$
Carbon monoxide	$1.657 \cdot 10^3$	$1.404 \cdot 10^3$	$1.680 \cdot 10^3$	$1.363 \cdot 10^3$	$1.703 \cdot 10^3$
Nitroxyl	$1.504 \cdot 10^3$	$1.124 \cdot 10^3$	$1.331 \cdot 10^3$	$1.326 \cdot 10^3$	$1.375 \cdot 10^3$
Cyanic fluoride	$2.047 \cdot 10^3$	$1.564 \cdot 10^3$	$1.780 \cdot 10^3$	$1.521 \cdot 10^3$	$1.891 \cdot 10^3$
Carbon dioxide	$7.705 \cdot 10^2$	$8.026 \cdot 10^2$	$9.146 \cdot 10^2$	$7.643 \cdot 10^2$	$9.795 \cdot 10^2$
Nitrous acid	$5.117 \cdot 10^3$	$3.282 \cdot 10^3$	$4.248 \cdot 10^3$	$4.068 \cdot 10^3$	$4.266 \cdot 10^3$
Ammonia	$1.049 \cdot 10^3$	$9.405 \cdot 10^2$	$1.129 \cdot 10^3$	$9.405 \cdot 10^2$	$1.125 \cdot 10^3$
Formaldehyde	$2.259 \cdot 10^3$	$1.822 \cdot 10^3$	$2.251 \cdot 10^3$	$1.849 \cdot 10^3$	$2.347 \cdot 10^3$
1-Pentanoic acid	$2.087 \cdot 10^4$	$1.514 \cdot 10^4$	$2.150 \cdot 10^4$	$1.847 \cdot 10^4$	$2.103 \cdot 10^4$
Cyclohexane	$1.670 \cdot 10^4$	$1.125 \cdot 10^4$	$1.457 \cdot 10^4$	$1.286 \cdot 10^4$	$1.446 \cdot 10^4$
1-Pentanamide	$2.569 \cdot 10^4$	$1.776 \cdot 10^4$	$2.570 \cdot 10^4$	$1.839 \cdot 10^4$	$2.488 \cdot 10^4$
Dinitrogen oxide	$1.939 \cdot 10^3$	$1.463 \cdot 10^3$	$2.146 \cdot 10^3$	$1.984 \cdot 10^3$	$2.342 \cdot 10^3$
Pentanal	$1.872 \cdot 10^4$	$1.349 \cdot 10^4$	$1.867 \cdot 10^4$	$1.575 \cdot 10^4$	$1.849 \cdot 10^4$
n-Butanol	$1.504 \cdot 10^4$	$1.002 \cdot 10^4$	$1.386 \cdot 10^4$	$1.100 \cdot 10^4$	$1.355 \cdot 10^4$
Hydrogen cyanide	$2.350 \cdot 10^3$	$1.884 \cdot 10^3$	$2.247 \cdot 10^3$	$1.861 \cdot 10^3$	$2.462 \cdot 10^3$
Hydrogen isocyanide	$5.503 \cdot 10^3$	$3.542 \cdot 10^3$	$4.741 \cdot 10^3$	$1.861 \cdot 10^3$	$5.015 \cdot 10^3$
Methane	$1.720 \cdot 10^3$	$1.595 \cdot 10^3$	$1.867 \cdot 10^3$	$1.612 \cdot 10^3$	$1.968 \cdot 10^3$
n-Butylamine	$1.636 \cdot 10^4$	$1.550 \cdot 10^4$	$1.713 \cdot 10^4$	$1.454 \cdot 10^4$	$1.667 \cdot 10^4$
Propane	$9.545 \cdot 10^3$	$7.547 \cdot 10^3$	$9.977 \cdot 10^3$	$8.096 \cdot 10^3$	$1.002 \cdot 10^4$
Ethane	$5.178 \cdot 10^3$	$4.533 \cdot 10^3$	$5.767 \cdot 10^3$	$4.705 \cdot 10^3$	$5.858 \cdot 10^3$
Butane	$1.424 \cdot 10^4$	$1.096 \cdot 10^4$	$1.487 \cdot 10^4$	$1.217 \cdot 10^4$	$1.465 \cdot 10^4$
Pentane	$1.954 \cdot 10^4$	$1.406 \cdot 10^4$	$1.937 \cdot 10^4$	$1.513 \cdot 10^4$	$1.934 \cdot 10^4$
Hexane	$2.447 \cdot 10^4$	$1.749 \cdot 10^4$	$2.434 \cdot 10^4$	$1.799 \cdot 10^4$	$2.416 \cdot 10^4$
PMI1	$4.167 \cdot 10^3$	$2.991 \cdot 10^3$	$3.746 \cdot 10^3$	$3.324 \cdot 10^3$	$4.198 \cdot 10^3$
Heptane	$2.869 \cdot 10^4$	$2.075 \cdot 10^4$	$2.898 \cdot 10^4$	$2.048 \cdot 10^4$	$2.867 \cdot 10^4$
Octane	$3.319 \cdot 10^4$	$2.428 \cdot 10^4$	$3.402 \cdot 10^4$	$2.311 \cdot 10^4$	$3.360 \cdot 10^4$
Benzene	$1.340 \cdot 10^4$	$9.858 \cdot 10^3$	$1.179 \cdot 10^4$	$1.235 \cdot 10^4$	$1.215 \cdot 10^4$
1-Butene	$1.726 \cdot 10^4$	$1.348 \cdot 10^4$	$1.870 \cdot 10^4$	$2.374 \cdot 10^4$	$1.836 \cdot 10^4$
PMI2	$1.459 \cdot 10^4$	$1.150 \cdot 10^4$	$1.368 \cdot 10^4$	$1.363 \cdot 10^4$	$1.425 \cdot 10^4$
PA1	$4.055 \cdot 10^3$	$2.719 \cdot 10^3$	$3.337 \cdot 10^3$	$3.190 \cdot 10^3$	$3.635 \cdot 10^3$
Acetylene	$3.259 \cdot 10^3$	$2.728 \cdot 10^3$	$3.151 \cdot 10^3$	$2.857 \cdot 10^3$	$3.301 \cdot 10^3$
PMI3	$6.048 \cdot 10^4$	$4.862 \cdot 10^4$	$6.179 \cdot 10^4$	$5.769 \cdot 10^4$	$6.038 \cdot 10^4$
PA2	$3.160 \cdot 10^4$	$2.380 \cdot 10^4$	$2.743 \cdot 10^4$	$2.780 \cdot 10^4$	$2.847 \cdot 10^4$
PMI4	$1.940 \cdot 10^5$	$1.476 \cdot 10^5$	$2.085 \cdot 10^5$	$1.994 \cdot 10^5$	$1.877 \cdot 10^5$
PDA1	$1.133 \cdot 10^5$	$9.602 \cdot 10^4$	$1.201 \cdot 10^5$	$1.158 \cdot 10^5$	$1.142 \cdot 10^5$

PMI5	$4.518 \cdot 10^5$	$3.430 \cdot 10^5$	$5.408 \cdot 10^5$	$5.199 \cdot 10^5$	$4.454 \cdot 10^5$
PA3	$1.501 \cdot 10^5$	$1.192 \cdot 10^5$	$1.379 \cdot 10^5$	$1.420 \cdot 10^5$	$1.364 \cdot 10^5$
PMI6	$9.884 \cdot 10^5$	$6.565 \cdot 10^5$	$1.144 \cdot 10^6$	$1.062 \cdot 10^6$	$8.613 \cdot 10^5$
PA4	$4.890 \cdot 10^5$	$4.007 \cdot 10^5$	$4.821 \cdot 10^5$	$5.010 \cdot 10^5$	$4.532 \cdot 10^5$
PDA2	$7.606 \cdot 10^5$	$6.680 \cdot 10^5$	$9.656 \cdot 10^5$	$9.172 \cdot 10^5$	$7.997 \cdot 10^5$
PA5	$1.240 \cdot 10^6$	$1.035 \cdot 10^6$	$1.319 \cdot 10^6$	$1.364 \cdot 10^6$	$1.147 \cdot 10^6$
PDA3	$2.443 \cdot 10^6$	$2.231 \cdot 10^6$	$3.778 \cdot 10^6$	$2.230 \cdot 10^6$	$2.579 \cdot 10^6$
PA6	$2.482 \cdot 10^6$	$2.223 \cdot 10^6$	$3.027 \cdot 10^6$	$3.136 \cdot 10^6$	$2.429 \cdot 10^6$
PDA4	$5.299 \cdot 10^6$	$4.955 \cdot 10^6$	$9.638 \cdot 10^6$	$9.249 \cdot 10^6$	$5.621 \cdot 10^6$
H <sub>2</sub>	$1.398 \cdot 10^3$	$1.465 \cdot 10^3$	$1.634 \cdot 10^3$	$1.255 \cdot 10^3$	$1.567 \cdot 10^3$
PDA5	$1.110 \cdot 10^7$	$8.676 \cdot 10^6$	$1.883 \cdot 10^7$	$1.861 \cdot 10^7$	$9.538 \cdot 10^6$
(H <sub>2</sub> ) <sub>2</sub>	$1.257 \cdot 10^4$	$1.304 \cdot 10^4$	$1.591 \cdot 10^4$	$1.289 \cdot 10^4$	$1.275 \cdot 10^4$
(H <sub>2</sub> ) <sub>3</sub>	$3.536 \cdot 10^4$	$3.977 \cdot 10^4$	$5.281 \cdot 10^4$	$4.140 \cdot 10^4$	$3.643 \cdot 10^4$
(H <sub>2</sub> ) <sub>4</sub>	$7.404 \cdot 10^4$	$8.629 \cdot 10^4$	$1.235 \cdot 10^5$	$9.340 \cdot 10^4$	$7.583 \cdot 10^4$
(H <sub>2</sub> ) <sub>5</sub>	$1.251 \cdot 10^5$	$1.492 \cdot 10^5$	$2.264 \cdot 10^5$	$1.660 \cdot 10^5$	$1.259 \cdot 10^5$
(H <sub>2</sub> ) <sub>7</sub>	$2.059 \cdot 10^5$	$2.241 \cdot 10^5$	$3.551 \cdot 10^5$	$2.512 \cdot 10^5$	$1.837 \cdot 10^5$
(H <sub>2</sub> ) <sub>6</sub>	$2.498 \cdot 10^5$	$3.067 \cdot 10^5$	$5.016 \cdot 10^5$	$3.457 \cdot 10^5$	$2.492 \cdot 10^5$
(H <sub>2</sub> ) <sub>8</sub>	$3.165 \cdot 10^5$	$3.944 \cdot 10^5$	$6.604 \cdot 10^5$	$4.364 \cdot 10^5$	$3.099 \cdot 10^5$
<b>MAE (a.u.)</b>		$7.44 \cdot 10^4$	$2.58 \cdot 10^5$	$2.20 \cdot 10^5$	$4.08 \cdot 10^4$
<b>RMSE (a.u.)</b>		$3.24 \cdot 10^5$	$1.16 \cdot 10^6$	$1.10 \cdot 10^6$	$2.08 \cdot 10^5$
<b>Maximum error (a.u.)</b>		$2.42 \cdot 10^6$	$7.73 \cdot 10^6$	$7.51 \cdot 10^6$	$1.56 \cdot 10^6$
<b>MAPE (%)</b>		19.5	18.8	18.9	6.6
<b>Maximum error (%)</b>		35.9	108.7	74.5	27.1
		<b>LC-BLYP</b>	<b>CAM-B3LYP</b>	<b>OT-LC-BLYP</b>	<b>T<math>\alpha</math>-LC-BLYP</b>



**Fig 3.** Benchmarking of several DFAs for the evaluation of  $\gamma_{zzzz}$  of  $\gamma$ -NLO set. MAPE: mean absolute percentage error; MAE ( $\times 10^4$ ): mean absolute error  $\times 10^4$ .

That is why the LC-BLYP MAE is the lowest one among the nine studied DFAs. Nevertheless, MAPE and MAE are about half the LC-BLYP ones when T $\alpha$ -LC-BLYP is used. Altogether, the data shows that our tuning strategy is a promising new tuned RSH scheme for the evaluation of second hyperpolarizabilities.

Considering the similarities between LC-BLYP, T $\alpha$ -LC-BLYP, OT-LC-BLYP, and CAM-B3LYP functionals, it is interesting to make a more detailed comparison of these four RSHs. As pointed out above, T $\alpha$ -LC-BLYP clearly reduces the average errors with respect to LC-BLYP, CAM-B3LYP, and OT-LC-BLYP. For instance,

the MAPE of  $\gamma$  for  $\omega$ B97xD, LC-BLYP, CAM-B3LYP, and OT-LC-BLYP are 18.8%, 19.5%, 18.8%, and 18.9%, respectively, while for T $\alpha$ -LC-BLYP is only 6.6%. The MAE of  $\gamma$  obtained with LC-BLYP, CAM-B3LYP, and OT-LC-BLYP are  $7.44 \cdot 10^4$ ,  $2.58 \cdot 10^5$ , and  $2.20 \cdot 10^5$ , respectively, while the MAE value is reduced up to  $4.08 \cdot 10^4$  for T $\alpha$ -LC-BLYP. Hence, T $\alpha$ -LC-BLYP clearly gives the most accurate second hyperpolarizabilities, while CAM-B3LYP is the worst choice among these four DFAs.

For molecules with  $\omega_{cc}$  close to 0.33 (the range-separation value used in CAM-B3LYP),  $\gamma$  values are more accurately predicted by CAM-B3LYP than by LC-BLYP (see, for instance, 1-butene). On the contrary, for molecules with  $\omega_{cc}$  close to 0.47 (the range-separation value used in LC-BLYP), LC-BLYP  $\gamma$  values are closer to CCSD(T) ones than CAM-B3LYP (see, for instance, PDA4). T $\alpha$ -LC-BLYP provides quite accurate second hyperpolarizability for all molecules, regardless their  $\omega_{cc}$  value. Conversely, OT-LC-BLYP does not always improve the performance of LC-BLYP for the calculation of  $\gamma$ . Indeed, for the molecules with the largest  $\gamma$  the accuracy of OT-LC-BLYP is slightly worse than LC-BLYP.

To explore the robustness of T $\alpha$ -LC-BLYP, we have checked its performance for two sets of molecules with electronic structures that differ significantly from the ones included in the  $\gamma$ -NLO molecular set (see Figure 4). Molecular set 1 includes *o*-benzynes, *m*-benzynes, and *p*-benzynes in their diradical singlet ground state. The NLOPs of these systems strongly depend on their diradical character. Molecular set 2 includes 8 hydrogen-

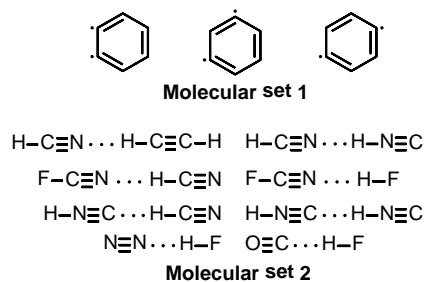


## ARTICLE

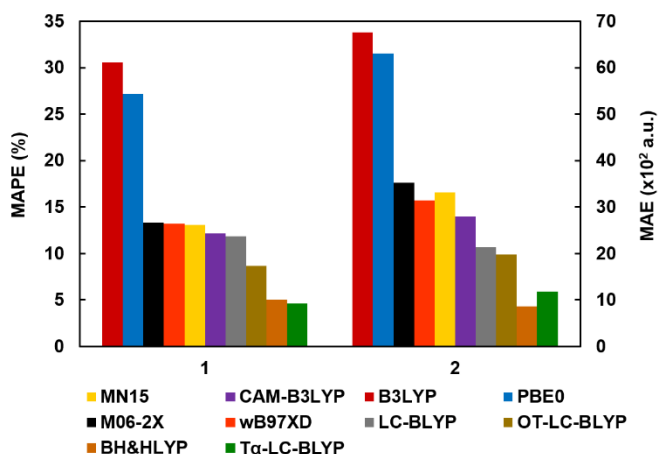
bonded small dimers taken from reference 62. The NLOPs of these molecules arise from the interaction between the two monomers.

For molecular set 1, the most accurate electronic second hyperpolarizabilities are given by  $T\alpha$ -LC-BLYP, BH&HLYP, and OT-LC-BLYP for both MAPEs (4.6, 5.1, and 8.7%, respectively) and MAEs ( $1.2 \cdot 10^3$ ,  $8.7 \cdot 10^2$ , and  $1.9 \cdot 10^3$  a.u., respectively) as can be seen in Figure 5.

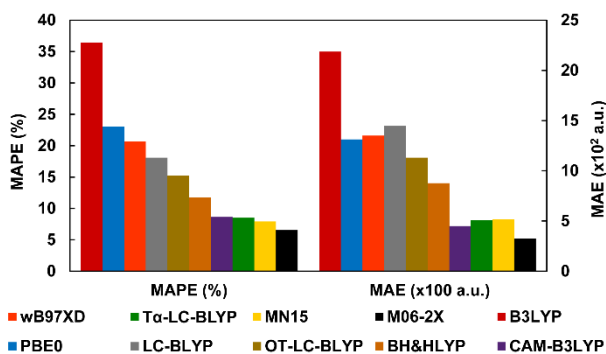
The best results in the calculation of second hyperpolarizabilities of the hydrogen-bonded complexes included in set 2 are given by M06-2X, CAM-B3LYP, and MN15 functionals, with MAEs of  $3.24 \cdot 10^2$ ,  $4.49 \cdot 10^2$ , and  $5.14 \cdot 10^2$  a.u., respectively (see Figure 6).



**Fig 4.** *o*- *m*- and *p*-benzyne (Molecular set 1) and hydrogen bonded species (Molecular set 2).



**Fig 5.** Benchmarking of several DFAs for the evaluation of  $\gamma_{zzzz}$  in the molecular set 1. MAPE: mean absolute percentage error; MAE ( $\times 10^2$ ): mean absolute error  $\times 10^2$ .



**Fig 6.** Benchmarking of several DFAs for the evaluation of  $\gamma_{zzzz}$  in the molecular set 2. MAPE: mean absolute percentage error; MAE ( $\times 10^2$ ): mean absolute error  $\times 10^2$ .

The performance of LC-BLYP is clearly worse because, for all the dimers but one, their  $\omega_{CC}$  are far lower than LC-BLYP  $\omega=0.47$  (MAE =  $14.47 \cdot 10^2$  a.u.). Despite  $T\alpha$ -LC-BLYP is based on LC-BLYP, the average errors of the second hyperpolarizabilities computed with the new tuned DFA are very similar to the most accurate ones (MAE =  $5.06 \cdot 10^2$  a.u.).

The values of  $\omega_{CC}$  and  $l_\alpha$  obtained for the new molecules do not trigger a substantial change in the correlation found for the  $\gamma$ -NLO molecular set (see Figure 2).

Since almost all points of the sets fit properly into the quadratic form that was previously fitted for  $\gamma$ -NLO molecular set (Eq. 7), qualitative good  $\gamma$  results are expected for the molecules in the sets 1 and 2 (see Tables S12.8-S12.10). For both sets,  $T\alpha$ -LC-BLYP clearly improves the results obtained with the LC-BLYP functional, being the  $T\alpha$ -LC-BLYP average errors about half the LC-BLYP ones. Indeed, for both sets,  $T\alpha$ -LC-BLYP is one of the best DFAs among the ones tested.

We also benchmarked the ten studied DFAs for the calculation of  $\mu_z$ ,  $\alpha_{zz}$  and  $\beta_{zzz}$  using as reference the CCSD(T) values of the 60 molecules of  $\gamma$ -NLO set (Tables S11.4-12). Although  $T\alpha$ -LC-BLYP has been designed to give accurate results for second hyperpolarizabilities, this specific tailoring is not detrimental for the calculation of the dipole moment, the linear polarizability and the first hyperpolarizability. As a matter of fact, the MAPE of  $T\alpha$ -LC-BLYP is the smallest for  $\mu_z$ , the second smallest for  $\alpha_{zz}$  and third smallest for  $\beta_{zzz}$ . In order to test whether  $T\alpha$ -LC-BLYP is biased towards the calculation of second hyperpolarizabilities, we have also analyzed other properties. In particular, we have optimized the geometries and computed the vibrational frequencies of the molecules of the  $\gamma$ -NLO set with the B3LYP, LC-BLYP and  $T\alpha$ -LC-BLYP functionals and compared their values with respect to MP2 values (Tables S11.13-15). The root-mean-square-deviation (RMSD) of the bond distances, angles, dihedrals and vibrational frequencies show that, although the best results are obtained with the B3LYP functional, the errors of  $T\alpha$ -LC-BLYP are lower than the errors of its parent functional LC-BLYP.

## Conclusions

In this manuscript, 9 commonly used DFAs (the global hybrid GGAs B3LYP, PBE0, BH&HLYP, the global hybrid meta-GGAs M06-2X and MN15, the long-range corrected functionals LC-BLYP, CAM-B3LYP, and  $\omega$ B97xD, and the optimally-tuned LC-BLYP based on the minimization of the  $J^2$  values, OT-LC-BLYP) have been benchmarked for the evaluation of the second hyperpolarizability using as reference the CCSD(T) values. To this end, we have designed the  $\gamma$ -NLO molecular set, which contains a total of 60 entries. For the hybrid GGAs and hybrid meta-GGAs, the average errors of  $\gamma$  decrease with the increase of the amount of exact exchange included in the DFA. Among these nine functionals, the long-range corrected functionals are the most accurate, LC-BLYP giving the lowest average errors in the calculation of the second hyperpolarizabilities of the  $\gamma$ -NLO set.

For every system of the  $\gamma$ -NLO molecular set, we have determined the value of the range-separation parameter  $\omega$  required to reproduce the CCSD(T)  $\gamma$  with LC-BLYP functional ( $\omega_{CC}$ ). We have shown that  $\omega$  is a key parameter in the RSH construction for  $\gamma$  calculation. Then, the  $\gamma$  for molecules with  $\omega_{CC}$  close to the LC-BLYP  $\omega$  (*i.e.*,  $\omega=0.47$ ) is more accurate if they are



computed with this functional rather than with CAM-B3LYP, whereas the opposite is true for molecules with  $\omega_{\text{CC}}$  close to the CAM-B3LYP  $\omega$ . The most accurate  $\gamma$  computed is obtained by the new tuned T $\alpha$ -LC-BLYP functional, reducing about half the average errors of LC-BLYP, which ranks as the second most accurate DFA tested in this benchmark.

The design of T $\alpha$ -LC-BLYP is based on a quadratic correlation between  $\omega_{\text{CC}}$  and the index given by Eq. 6. This correlation includes both the molecules displaying small and large  $\gamma$ . The tuned range-separated functional T $\alpha$ -LC-BLYP employs this quadratic correlation to predict the tuned  $\omega$  ( $\omega_{\text{T}\alpha}$ ) that reproduces the CCSD(T)  $\gamma$ . T $\alpha$ -LC-BLYP clearly improves the results obtained with the most popular hybrid and range-separated hybrids and bears a computational cost only slightly higher than LC-BLYP, since, in addition to the calculation of  $\gamma$ , it only requires the previous calculation of the linear polarizability with LC-BLYP. Even though T $\alpha$ -LC-BLYP is tailored for the calculation of  $\gamma$ , T $\alpha$ -LC-BLYP performs also well for the calculation of other linear and non-linear optical properties, such as  $\mu_z$ ,  $\alpha_{zz}$ , and  $\beta_{zzz}$ .

The results obtained for set 1 and set 2 are very promising and suggest that the correlation found herein could be generalized to include a wider spectrum of molecular systems and molecules that display high NLOP for reasons other than  $\pi$ -delocalization.

### Author contributions

P.B.-S. performed all the NLOP calculations, wrote the scripts and Fortran codes required for the data generation and analysis, and wrote the first draft of the manuscript. S.P.S. helped with the calculations that required the optimization of J and contributed to the draft. P.S., E.M. and J.M.L. designed and supervised the project and co-wrote the manuscript. All authors have contributed to the results' analysis as well as give approval to the final version of the manuscript.

### Conflicts of interest

There are no conflicts to declare.

### Acknowledgements

This work was financially supported by grants from the Spanish government MICINN (PGC2018-098212-B-C22 to P.S. and J.M.L., and PGC2018-098212-B-C21 and EUR2019-103825 to E.M.), Generalitat de Catalunya (2017SGR39 to J.M.L.), Diputaci3n Foral de Gipuzkoa (2019-CIEN-000092-01 to E.M.), Gobierno Vasco (IT1254-19 and PIBA19-0004 to E.M.). We thank the Spanish government for the predoctoral grant to P.B.-S. (FPU17/02058) and Basque Country government for the predoctoral grant to S.P.S. (PRE 2018 2 02200). We are also grateful for the computational time from the Consorci de Serveis Universitaris de Catalunya (CSUC), DIPC and the SGI/IZO-SGIker UPV/EHU. Finally, we want to acknowledge Dr. Emili Besal3, who provided the subroutines to perform the cross-validation tests.

### Notes and references

1. L. R. Dalton, P. A. Sullivan and D. H. Bale, *Chem. Rev.*, 2010, **110**, 25-55.
2. L. R. Dalton, S. J. Benight, L. E. Johnson, D. B. Knorr, I. Kosilkin, B. E. Eichinger, B. H. Robinson, A. K. Y. Jen and R. M. Overney, *Chem. Mater.*, 2011, **23**, 430-445.
3. T. Ikeda and O. Tsutsumi, *Science*, 1995, **268**, 1873-1875.
4. S. Barlow, J.-L. Br3das, Y. A. Getmanenko, R. L. Gieseking, J. M. Hales, H. Kim, S. R. Marder, J. W. Perry, C. Risko and Y. Zhang, *Mater. Horiz.*, 2014, **1**, 577-581.
5. J. Boixel, V. Guerschais, H. Le Bozec, D. Jacquemin, A. Amar, A. Boucekkine, A. Colombo, C. Dragonetti, D. Marinotto, D. Roberto, S. Righetto and R. De Angelis, *J. Am. Chem. Soc.*, 2014, **136**, 5367-5375.
6. J. M. Hales, J. Matichak, S. Barlow, S. Ohira, K. Yesudas, J. L. Bredas, J. W. Perry and S. R. Marder, *Science (New York, N.Y.)*, 2010, **327**, 1485-1488.
7. M. Drobizhev, N. S. Makarov, S. E. Tillo, T. E. Hughes and A. Rebane, *Nat. Methods*, 2011, **8**, 393.
8. W. R. Zipfel, R. M. Williams and W. W. Webb, *Nat. Biotechnol.*, 2003, **21**, 1369-1377.
9. W. Denk, J. Strickler and W. Webb, *Science*, 1990, **248**, 73-76.
10. J. A. Delaire and K. Nakatani, *Chem. Rev.*, 2000, **100**, 1817-1846.
11. L. Brzozowski and E. H. Sargent, *J. Mater. Sci.: Mater. Electron.*, 2001, **12**, 483-489.
12. V. M. Churikov and C.-C. Hsu, *J. Opt. Soc. Am. B*, 2001, **18**, 1722-1731.
13. B. Champagne, E. Botek, M. Nakano, T. Nitta and K. Yamaguchi, *J. Chem. Phys.*, 2005, **122**, 114315.
14. A. Baranowska and A. J. Sadlej, *J. Comput. Chem.*, 2010, **31**, 552-560.
15. P. A. Limacher, Q. Li and H. P. L3thi, *J. Chem. Phys.*, 2011, **135**, 014111.
16. C. M3ller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618-622.
17. J. Cizek and J. Paldus, *Physica Scripta*, 1980, **21**, 251-254.
18. R. Zaleśny, A. Baranowska-Łączkowska, M. Medveď and J. M. Luis, *J. Chem. Theory Comput.*, 2015, **11**, 4119-4128.
19. M. B. Oviedo, N. V. Ilawe and B. M. Wong, *J. Chem. Theory Comput.*, 2016, **12**, 3593-3602.
20. I. W. Bulik, R. Zaleśny, W. Bartkowiak, J. M. Luis, B. Kirtman, G. E. Scuseria, A. Avramopoulos, H. Reis and M. G. Papadopoulos, *J. Comput. Chem.*, 2013, **34**, 1775-1784.
21. S. N3non, B. Champagne and M. I. Spassova, *Phys. Chem. Chem. Phys.*, 2014, **16**, 7083-7088.
22. R. L. Gieseking, M. K. Ravva, V. Coropceanu and J.-L. Br3das, *J. Phys. Chem. C*, 2016, **120**, 9975-9984.
23. W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133-A1138.
24. N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, 2017, **115**, 2315-2372.
25. M. Chołuj, J. Kozłowska and W. Bartkowiak, *Int. J. Quant. Chem.*, 2018, **118**, e25666.
26. J. Autschbach and M. Srebro, *Acc. Chem. Res.*, 2014, **47**, 2592-2602.
27. O. Gritsenko, R. van Leeuwen, E. van Lenthe and E. J. Baerends, *Phys. Rev. A*, 1995, **51**, 1944-1954.
28. Z. Lin and T. Van Voorhis, *J. Chem. Theory Comput.*, 2019, **15**, 1226-1241.

## ARTICLE

29. T. J. Duignan, J. Autschbach, E. Batista and P. Yang, *J. Chem. Theory Comput.*, 2017, **13**, 3614-3625.
30. M. Srebro and J. Autschbach, *J. Chem. Theory Comput.*, 2012, **8**, 245-256.
31. S. Bhandari, M. S. Cheung, E. Geva, L. Kronik and B. D. Dunietz, *J. Chem. Theory Comput.*, 2018, **14**, 6287-6294.
32. A. K. Manna, S. Refaely-Abramson, A. M. Reilly, A. Tkatchenko, J. B. Neaton and L. Kronik, *J. Chem. Theory Comput.*, 2018, **14**, 2919-2929.
33. N. Kuritz, T. Stein, R. Baer and L. Kronik, *J. Chem. Theory Comput.*, 2011, **7**, 2408-2415.
34. J. F. Janak, *Phys. Rev. B*, 1978, **18**, 7165-7168.
35. K. Garrett, X. Sosa Vazquez, S. B. Egri, J. Wilmer, L. E. Johnson, B. H. Robinson and C. M. Isborn, *J. Chem. Theory Comput.*, 2014, **10**, 3821-3831.
36. H. Sun and J. Autschbach, *ChemPysChem*, 2013, **14**, 2450-2461.
37. A. J. Garza, O. I. Osman, A. M. Asiri and G. E. Scuseria, *J. Phys. Chem. B*, 2015, **119**, 1202-1212.
38. A. K. Pal, T. J. Duignan and J. Autschbach, *Phys. Chem. Chem. Phys.*, 2018, **20**, 7303-7316.
39. H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2001, **115**, 3540-3544.
40. O. A. Vydrov and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 234109.
41. T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51-57.
42. A. Karolewski, L. Kronik and S. Kümmel, *J. Chem. Phys.*, 2013, **138**, 204115.
43. R. Zaleśny, M. Medved', S. P. Sitkiewicz, E. Matito and J. M. Luis, *J. Chem. Theory Comput.*, 2019, **15**, 3570-3579.
44. G. W. T. M. J. Frisch, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Journal*, 2015.
45. T. H. Dunning Jr., *J. Chem. Phys.*, 1989, **90**, 1007-1023.
46. J. Poater, F. M. Bickelhaupt and M. Solà, *J. Phys. Chem. A*, 2007, **111**, 5063-5070.
47. X. Li and J. Paldus, *J. Chem. Phys.*, 2008, **129**, 174101.
48. H. Rutishauser, *Numer. Math. (Heidelb)*, 1963, **5**, 48-54.
49. M. Medved', M. Stachová, D. Jacquemin, J.-M. André and E. A. Perpète, *J. Mol. Struct.*, 2007, **847**, 39-46.
50. W. H. Press, W. T. Vetterling, S. A. Teukolsky and B. P. Flannery, *Numerical Recipes The Art of Scientific Computing*, Cambridge University Press, Second edn., 1992.
51. R. F. W. Bader, *Atoms in Molecules - A Quantum Theory*, Oxford University Press, Oxford, 1990.
52. X. Fradera, M. Duran and J. Mestres, *J. Phys. Chem. A*, 2000, **104**, 8445-8454.
53. E. Matito, M. Solà, P. Salvador and M. Duran, *Faraday Discuss.*, 2007, **135**, 325-345.
54. E. Besalú, *J. Math. Chem.*, 2001, **29**, 191-204.
55. C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785-789.
56. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648-5652.
57. C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158-6170.
58. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372-1377.
59. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215-241.
60. H. S. Yu, X. He, S. L. Li and D. G. Truhlar, *Chem. Sci.*, 2016, **7**, 5032-5051.
61. J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615-6620.
62. R. Zaleśny, M. Medved', R. W. Góra, H. Reis and J. M. Luis, *Phys. Chem. Chem. Phys.*, 2018, **20**, 19841-19849.