Describing the Disulfide Bond: From the Density Functional Theory and Back through the “Lego Brick” Approach

Describing the disulfide bond: from the density functional theory and back through the "Lego brick" approach

Silvia Alessandrini,[†] Hexu Ye,[†] Malgorzata Biczysko,[*,‡] and Cristina Puzzarini[*,†]

†*Dipartimento di Chimica "Giacomo Ciamician", Università di Bologna, Via F. Selmi 2, 40126 Bologna, Italy*

‡*Faculty of Chemistry, University of Wroclaw, F. Joliot-Curie 14, 50-383 Wroclaw, Poland*

E-mail: malgorzata.biczysko@uwr.edu.pl; cristina.puzzarini@unibo.it

Abstract

Selected molecular species containing the disulfide bond, RSSR, have been considered, these ranging from hydrogen disulfide, H_2S_2 (R=H), to diphenyl disulfide with R=C_6H_5. The aim of this work is twofold: (*i*) to investigate different computational approaches in order to derive accurate equilibrium structures at an affordable cost, (*ii*) to employ the results from the first goal in order to benchmark cheaper methodologies rooted in the density functional theory. Among the strategies used for the accurate geometrical determinations, the semi-experimental approach has been exploited in combination with a reduced-dimensionality VPT2 model, without however obtaining satisfactory results. Instead, the so-called "Lego brick" approach turned out to be very effective despite the flexibility of the systems investigated. Concerning the second target of this work, the focus was mainly on the S-S bond and the structural parameters related to it. PBE0(-D3BJ), M06-2X(-D3) and DSD-PBEP86-D3BJ have been found to be the best-performing functionals.

Introduction

Sulfur is an essential element to life. Indeed, it is present in a wide variety of molecules with different physiological functions. Disulfide bonds are present in approximately 10% of the proteins produced by mammalian cells, with cystine, the disulfide derived from the cysteine amino acid, being the first discovered.[1,2] Disulfide (S-S) bonds play a peculiar role in biochemistry because, by breaking and/or forming of disulfide cross-linking, trigger conformational changes, thus influencing protein folding and controlling protein function.[3–6] Notably, in proteins and polypeptides, these disulfide bridges are bonds formed intramolecularly between two sulfur atoms.[7] Rarely, they are established between two vicinal cysteines.[7] Interestingly, these S-S bridges constitute the only natural covalent link between polypeptides strands.[7]

As mentioned above, sulfur is one of life-essential elements. However, moving to the issue of the origin of life, its path from interstellar clouds to planets and their atmospheres is still not well understood. Despite sulfur being the tenth most abundant element in the Galaxy, in the cold regions of the interstellar medium (ISM), S-bearing molecules are not as abundant as expected on the basis of its elemental abundance.[8] This is the so-called "sulfur-depletion" problem.[9–11] So far, among the few S-bearing molecules detected in space, no one contains two sulfur atoms, much less the disulfide moiety. However, although H_2S_2 has not been detected in the ISM and cometary ices, it is postulated to form in H_2S-rich ices and to be one of the molecules able to explain sulfur depletion. For example, very recently, it has been pointed out that temperature-programmed desorption experiments of the co-deposition of C_2H_2, H_2S, and H at 10 K leads to the formation of several sulfur-bearing species.[12] Among the compounds securely identified by means of infrared spectroscopy and mass spectrometry, two molecules containing two sulfur atoms were found: H_2S_2 and $HSCH_2CH_2SH$.[12] This recent work further demonstrated not only the possible presence of H_2S_2 in the IMS, but also the possible formation of more complex species. Therefore, the presence in space of R_2S_2, with R being a small organic fragment, cannot be ruled out.

Sulfur-containing molecules have also been attracting a long-standing interest in the

field of computational chemistry, which however has mainly focused on species containing only one sulfur atoms. Moving to species bearing the HSS moiety, the molecular structures of HSSH (H_2S_2), HSSCN, CH_3SSH, CH_2CHSSH, CH_3SSSH and $CH_2CHSSSH$ have already been studied computationally.[13–15] In this work, we aim at performing an accurate structural characterization of selected molecules containing the disulfide bond and two equal organic substituents, RSSR. More in detail, we have considered a set of molecular species of increasing complexity containing the S-S bond, which have experimentally been studied by means of rotational spectroscopy: dimethyl disulfide ($(CH_3)_2S_2$),[16] diethyl disulfide ($(C_2H_5)_2S_2$),[17] diallyl disulfide ($(C_3H_5)_2S_2$),[18] diisopropyl disulfide ($(C_3H_7)_2S_2$),[19] and diphenyl disulfide ($(C_6H_5)_2S_2$).[20] The RSSR molecules investigated are shown in Figure 1 together with the atom labeling and point-group symmetry. For them, whenever it applies, only the most stable conformer has been considered.

As noted in Figure 1, together with the RSSR species having R = small organic group, H_2S_2 has also been involved in the study as reference species because its equilibrium structure

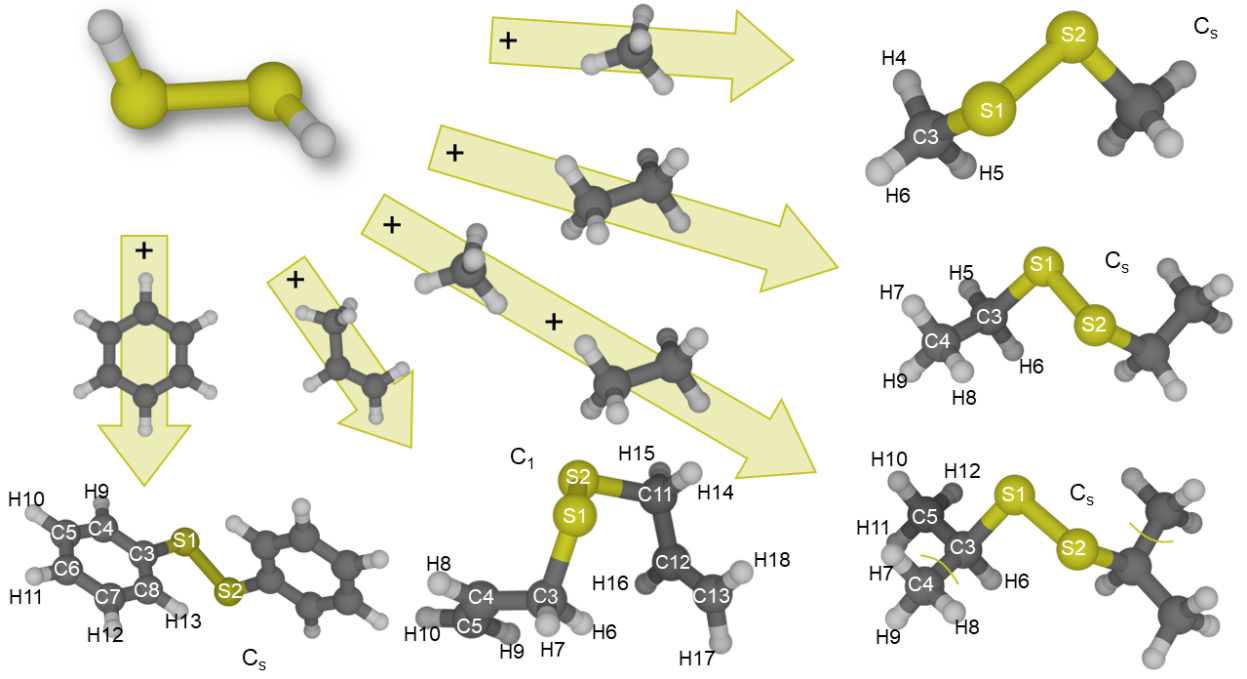


Figure 1: Molecular structures and atom numbering of the disulfide species considered in this study, together with the molecules employed within the "Lego-brick" approach.

(r_e) has recently been obtained by means of accurate composite schemes rooted in the coupled-cluster (CC) theory as well as by exploiting the so-called semi-experimental approach.[13] While for $(CH_3)_2S_2$ composite schemes that combine accuracy and reduced computational cost can be applied, for the larger members of the RSSR series such methodologies become expensive and only density functional theory (DFT) is cost-effective. To improve its performance, the so-called "Lego brick" approach[21,22] has been considered. This is based on the assumption that a molecular system can be seen as formed by smaller fragments whose equilibrium structures are accurately known. The envisaged fragments are shown in Figure 1. Subsequently, the resulting accurate structures have been employed as reference for benchmarking different DFT methodologies. For this second part of the study, several density functionals and different basis sets have been tested.

As mentioned above, the RSSR molecular systems, with R = CH_3, C_2H_5, C_3H_5, C_3H_7, and C_6H_5, have already been investigated. In the corresponding works (refs. 16–20), the rotational spectra of the main isotopologue and various isotopic species have been investigated. This opens the way toward the application of the so-called semi-experimental (SE) approach,[23] which combines experimental ground-state rotational constants for different isotopologues with the corresponding computed vibrational corrections to derive accurate SE equilibrium rotational constants to be used for structural determinations. Indeed, these latter parameters only depend on the equilibrium geometry and the isotopic content.[23–28] In addition to the limited number of experimental data, which prevents any full structural determination, the molecules here considered present an additional problem, which is related to their high flexibility and thus to the presence of large amplitude motions (LAMs). Indeed, this hampered the SE approach to be applied by Demaison and coworkers to diallyl disulfide and diphenyl disulfide.[18,20] In this study, we have tried to overcome these issues.

Methodology

The goal of this work is twofold: (*i*) to determine accurate equilibrium structures for the selected RSSR systems (see Figure 1) and (*ii*) to use the corresponding results for benchmarking different DFT levels of theory. In the following, the computational methodology and details of the two goals are presented separately.

The Gaussian 16,[29] CFOUR[30,31] and PSI4[32] suites of programs have been employed in the present study.

Accurate equilibrium structures

Recent implementations in PSI4 allow the ad hoc definition of gradient-based composite schemes.[33] The one employed in this work, denoted as "gradient-MP2-CCSD(T)", is based on the extrapolation to the complete basis set (CBS) limit of DF-MP2 gradients[34] in conjunction with the aug-cc-pwCVnZ (n=T,Q) basis sets,[35] hereafter shortly denoted as awCnZ. DF-MP2 denotes MP2 combined with the density-fitting approximation, with MP2 standing for Møller-Plesset perturbation theory to second order.[36] The extrapolation to the CBS limit employs the two-point n^{-3} formula[37,38] for the correlation energy gradient (ΔCBS(DF-MP2/awC(T,Q)Z)), while it adopts the energy gradient for the HF-SCF term with the largest basis set (HF/awCVQZ). The contribution due to higher-order electron correlation energy is incorporated as the difference between DF-MP2 and DF-CCSD(T) gradients (DF-CCSD(T)/wCTZ - DF-MP2/wCTZ), both in conjunction with the cc-pwCVTZ basis set (hereafter wCTZ), where CCSD(T) stands for coupled cluster singles doubles augmented by a perturbative treatment of triples.[39] To sum up, the gradient-MP2-CCSD(T) scheme is given

by:

$$\begin{aligned} \text{gradient-MP2-CCSD(T)} = {} & \text{HF/awCVQZ} \\ & + \Delta\text{CBS(DF-MP2/awC(T,Q)Z)} \\ & + [\text{DF-CCSD(T)/wCTZ} - \text{DF-MP2/wCTZ}] \ . \end{aligned} \tag{1}$$

In all the DF-MP2 and DF-CCSD(T) calculations above, all electrons (ae) are correlated. The minimization of the gradient based on eq. 1 leads to the equilibrium structure which, in the text, is named after the composite scheme.

In addition to the gradient-MP2-CCSD(T) scheme, an approach that exploits the additivity approximation at the geometrical parameter level[40] (instead of energy gradient) has also been considered. This is the so-called "Cheap" composite scheme[41] (hereafter shortly denoted as ChS):

$$\begin{aligned} \text{r}_e\text{(ChS)} = {} & \text{r}_e\text{(fc-CCSD(T)/VTZ)} \\ & + [\text{r}_e\text{(MP2/CBS(T,Q))} - \text{r}_e\text{(MP2/VTZ)}] \\ & + [\text{r}_e\text{(ae-MP2/wCTZ)} - \text{r}_e\text{(fc-MP2/wCTZ)}] \ , \end{aligned} \tag{2}$$

where r_e(fc-CCSD(T)/VTZ) denotes a generic optimized structural parameter at the CCSD(T) level, within the frozen-core (fc) approximation, in conjunction with the cc-pV(T+d)Z basis set. The second term on the right-hand side is a corrective term due to the extrapolation to the CBS limit, which is performed at the fc-MP2 level employing the cc-pV(T+d)Z and cc-pV(Q+d)Z basis sets together with the n^{-3} extrapolation formula.[37] In the cc-pV(n+d)Z basis sets,[42–44] "+d" applies to third-period elements. The last term incorporates the core-valence (CV) correction as difference between ae and fc MP2 geometry optimizations, both in the cc-pwCVTZ basis set.

For most of the R-substituent cases, both the gradient-MP2-CCSD(T) and ChS schemes require a high computational cost. For these systems, the so-called "Lego brick" approach,[21,22]

which combines the template-molecule (TM)[45] and linear-regression (LR)[46] approaches, has been employed for obtaining accurate structures. According to the TM model, a medium-/large-sized molecular system can be seen as formed by smaller molecular fragments, whose accurate equilibrium geometries are available. Therefore, a generic equilibrium geometrical parameter (r_e) of the target (T) molecule can be written as:

$$r_e(\mathrm{TM,T}) = r_e^{\mathrm{DFT,T}} + \Delta\mathrm{TM}\ , \tag{3}$$

In the present study, based on previous works,[22,47,48] for $r_e^{\mathrm{DFT,T}}$, the DFT level we employed is the double-hybrid rev-DSDPBEP86-D3BJ[49] functional in conjunction with the jun-cc-pV(T+d)Z[50] basis set, "D3BJ" indicating the incorporation of empirical dispersion corrections.[51,52] Hereafter, this level of theory is shortly denoted as revDSD/junTZ.

The ΔTM correction is defined as:

$$\Delta\mathrm{TM} = r_e^{\mathrm{SE,TM}} - r_e^{\mathrm{DFT,TM}}\ . \tag{4}$$

In the equation above, $r_e^{\mathrm{SE,TM}}$ and $r_e^{\mathrm{DFT,TM}}$ denote the same geometrical parameter of the target system (T) in the selected TM molecule: they are the accurate value and the computed one at the same level of theory as T (i.e. revDSD/junTZ), respectively.[45] In the present case, SE equilibrium structures (see below) have been used for $r_e^{\mathrm{SE,TM}}$.

While the TM approach is able to improve the accuracy of the structural parameters belonging to the envisaged fragments (i.e. the TM molecules), the geometrical parameters connecting them need to be corrected in order to improve the original accuracy of the DFT methodology employed for $r_e^{\mathrm{DFT,T}}$. For this purpose, the LR approach can be used. The LR-corrected r_e(TM,T) is given by:

$$r_e(\mathrm{LR,T}) = (1+a) \times r_e^{\mathrm{DFT,T}} + b\ , \tag{5}$$

where a and b are the linear regression parameters, which depend on the DFT level chosen and have been taken from ref. 46.

While, in principle, the TM and LR approaches can be applied to both distances and angles, as a matter of fact, the DFT methodology here employed, i.e. revDSD/junTZ, already provides very good results for angles.[53–55] Therefore, the TM and LR corrections are only applied to bond lengths. Throughout the manuscript, for the sake of brevity, the "Lego-brick" approach only considering TM corrections is denoted as "TM", while the model also incorporating LR corrections as "TM+LR".

In addition to the pure computational structure determinations described above, as briefly mentioned in the Introduction, the so-called SE approach[23] has also been employed. SE equilibrium structures are obtained from least-squares (LSQ) fits of the molecular structural parameters to the SE equilibrium rotational constants ($B_e^{i,\mathrm{SE}}$) for different isotopologues. The $B_e^{i,\mathrm{SE}}$ values are derived by correcting the experimental ground-state rotational constants ($B_0^{i,\mathrm{exp}}$) for the computed vibrational contributions $\Delta B_{vib}^{i,\mathrm{calc}}$ as follows:

$$B_e^{i,\mathrm{SE}} = B_0^{i,\mathrm{EXP}} - \Delta B_{vib}^{i,\mathrm{calc}} = B_0^{i,\mathrm{exp}} + \frac{1}{2}\sum_n \alpha_n^i \,, \tag{6}$$

where i denotes the inertial axes a, b, c and, thus, B^a gives the rotational constant A, B^b gives B, and B^c gives C. In the equation above, α_n^i is the vibration-rotation interaction constant,[56] with n indicating the vibrational normal mode and, as said above, i the inertial axis; therefore, the sum runs over all vibrational normal modes. To evaluate the vibrational contributions $\Delta B_{vib}^{i,\mathrm{calc}}$, second-order vibrational perturbation theory (VPT2)[56,57] has to be applied to a cubic force field. Since the molecular systems here considered present at least one LAM, the reduced dimensionality VPT2 (RD-VPT2) approach[58–61] has been employed. In such a treatment, the LAM modes are treated at the harmonic level, whereas the remaining vibrational modes are calculated anharmonically. To give an example, for $(CH_3)_2S_2$, the CH_3 torsional mode has been excluded from the VPT2 treatment.

In this study, the required anharmonic force fields have been computed using the B3LYP-

D3BJ functional and Jensen's segmented polarization-consistent aug-pcs-1 basis set;[62] this level of theory is hereafter shortly denoted as B3/apcs-1. Anharmonic and VPT2 calculations have been performed with the Gaussian 16 package, while the `xrefit` module of the CFOUR quantum-chemistry package[30,31] has been employed for the LSQ fits. Since only a partial SE equilibrium structure can be derived for the systems here considered, the non-determinable geometrical parameters have been kept fixed at the TM/TM+LR counterparts (as explained in the Results and Discussion section).

Benchmark study of DFT methodologies

For the benchmark study, different DFT functionals and different basis sets have been considered. These latter have been tested for the reference molecule, i.e. H_2S_2, and include the Dunning series of correlation-consistent polarized basis sets[63] (aug-)cc-pVnZ (n=D,T,Q,5), Jensen's segmented polarization-consistent basis sets[62] (aug-)pcs-n (n=1, 2, 3, 4), and cc-pVnZ-F12 (n=D, T, Q) basis sets[64,65] as well as the "seasonal" jun-cc-pVTZ basis set.[66] For cc-pVnZ-F12, valence functions for cardinal number "n" are generally taken from the cc-pV(n+1)Z basis sets and augmented by additional p or d functions, and diffuse s and p functions from the aug-cc-pV(n+1)Z family.[42,43]

Subsequently, several DFT functionals in conjunction with the best performing basis set, namely aug-pcs-2 (jun-cc-pVTZ for diphenyl disulfide), have been employed to compute equilibrium structures of the selected molecules. In choosing the DFT functionals, the availability of analytical second-derivatives has been considered as a mandatory prerequisite. In particular, all functionals employed in a recent benchmark study on vibrational properties[67] have been considered. The full list of functionals is reported in Table 1, where they are classified according to their hierarchy - "Jacob's Ladder":[68] single-hybrid GGA (X_{HF}+GGA), single-hybrid Meta-GGA (X_{HF}+Meta-GGA), and double-hybrid (X_{HF}+C_{MP2}+GGA) functionals. Dispersion corrections have been incorporated by means of Grimme's dispersion[69] D3 correction[70] with Becke–Johnson (BJ) damping,[52] with the exception of the M06-2X and

APF[71] functionals. For M06-2X, the original D3 correction (with no damping) has been employed, while a purposely developed spherical atom dispersion model has been adopted for APF (APFD[71]).

Table 1: List of the DFT functionals considered in the benchmark study.

X_{HF}+GGA	X_{HF}+Meta-GGA	X_{HF}+C_{MP2}+GGA
B3LYP[72–74]	M06-2X[75]	B2PLYP[76]
B3PW91[72–74]	MN15[77]	DSD-PBEP86[78,79]
CAM-B3LYP[80]		DSD-PBEh95[78]
PBE0[81]		PBE0DH[82]
LC-ωPBE[83–85]		PBEQIDH[86]
B97[87]		
ωB97X[88]		
APF[71]		

Results and discussion

Following the organization of the Methodology section, the determination of accurate structures for the selected molecules containing the disulfide bond are first presented and discussed. In the second part, the benchmark study aiming at investigating the performance of different DFT methodologies is reported.

Accurate equilibrium structures of the RSSR systems

As mentioned in the Introduction, selected molecules containing the S-S bond have been considered for accurate structural determinations to be used in the subsequent benchmark study. Particular emphasis has been given to the S-S bond and to the structural parameters involving the S atoms.

The first and smallest member of this selection is H_2S_2, which –as already mentioned– has been the subject of an accurate structural characterization.[13] In Table 2, its structural parameters determined using the composited schemes of eqs. 1 (gradient-MP2-CCSD(T)) and

Table 2: Structural parameters of H_2S_2 at different levels of theory. Bond lengths in Å, angles in degrees.

	BestEst [a]	r_e^{SE} [a]	gradient- [b] MP2-CCSD(T)	ChS [c]	revDSD/ junTZ
SS (Å)	2.0503	2.0512(1)	2.0504	2.0562	2.0609
SH (Å)	1.3395	1.3408(8)	1.3389	1.3394	1.3422
HSS (deg.)	98.18	98.19(1)	98.17	98.08	98.24
HSSH (deg.)	90.64	90.65(1)	90.77	90.72	90.63

[a] From ref. 13. For r_e^{SE}, vibrational and electronic corrections are at the fc-MP2/cc-pVTZ level. [b] From eq. 1. [c] From eq. 2.

2 (ChS) are reported together with those of the best-estimated equilibrium geometry (hereafter denoted as BestEst) and r_e^{SE} from ref. 13. In detail, the BestEst structure was derived from a composite scheme, entirely based on CC techniques, which incorporates, in addition to extrapolation to the CBS limit and CV correction at the CCSD(T) level, contributions due to high-order excitations in the CC expansion up to quadruples as well as relativistic effects (for all details, the reader is referred to ref. 13). From an inspection of Table 2, the impressive performance of the gradient-MP2-CCSD(T) scheme is noted; the agreement with the BestEst structure is nearly perfect despite the great difference in the computational cost, that of BestEst being extremely high. Concerning the comparison with the SE equilibrium structure, small deviations are observed, the largest ones being about 2 mÅ for distances (S-H bond) and 0.1 degree for angles (HSSH dihedral angle). Moving to the ChS (eq. 2) structure, it is noted that it agrees well with the BestEst, r_e^{SE} and gradient-MP2-CCSD(T) geometries, the only relevant deviation of 5-6 mÅ being noted for the S-S bond. The overall conclusion is that gradient-MP2-CCSD(T) is a very accurate scheme at an affordable computational cost even for larger systems.

In Table 2, the revDSD/junTZ geometry is also reported because this level is the one employed in the "Lego brick" approach. The agreement is rather good for the S-H bond length and the angles, with deviations –with respect to BestEst, r_e^{SE} and gradient-MP2-CCSD(T)– of 2-3 mÅ for the former and well within 0.1 degree for the angles. Instead, the S-S distance is

10 mÅ longer than the BestEst and gradient-MP2-CCSD(T) values, while the overestimation is 9 mÅ with respect to r_e^{SE}.

In view of the performance of the gradient-MP2-CCSD(T) scheme for H_2S_2, it has been applied to the second molecule in terms of increasing dimension, $(CH_3)_2S_2$. The corresponding results are reported in Table 3 together with those from the ChS composite scheme, the "Lego-brick" approach, and the revDSD/junTZ level of theory. First of all, it is noted the good performance of the "Lego-brick" approach: the distances involving the S atom result less than 2 mÅ longer than the gradient-MP2-CCSD(T) ones, while for the C-H distances the discrepancy is well within 1 mÅ. From the inspection of Table 3, it is apparent that the ChS composite scheme performs worse than the "Lego-brick" approach; indeed, for the former the deviations with respect to the gradient scheme are ∼5 mÅ for SS, ∼2 mÅ for SC, and smaller than 0.5 mÅ for the C-H bonds. Moving to the revDSD/junTZ level, overestimations similar to those noted for H_2S_2 are evident. In particular, the revDSD/junTZ S-S bond length is 11 mÅ longer than the gradient-MP2-CCSD(T) one. Concerning the angles, which are kept unchanged in the "Lego-brick" approach (see Methodology section), it is observed that the discrepancy is well within 0.1 degree for valence angles and within 0.5 degrees for dihedral angles. Therefore, the overall conclusion is that the "Lego-brick" approach is well suitable for the evaluation of those structural parameters that cannot be determined in the SE approach. Furthermore, in view of its very limited computational cost, it can also be applied to the largest systems considered in this study and then be used as reference in the benchmark study .

An additional comparison reported in Table 3 is based on the equilibrium rotational constants. Indeed, as stressed in the Introduction, rotational constants only depend on the molecular structure and isotopic composition, this dependency providing the foundation of the SE approach. Taking the gradient-MP2-CCSD(T) structure as reference, it is noted that the average deviation of the revDSD/junTZ rotational constants is 0.97%, which decreases to 0.38% once the TM+LR approach is exploited. The ChS geometry appears a bit improved

with a mean discrepancy of 0.27%.

Table 3: Structural parameters[a] of $(CH_3)_2S_2$ at different levels of theory. Bond lengths in Å, angles in degrees, rotational constants in MHz.

	gradient- [b] MP2-CCSD(T)	ChS [c]	revDSD/ junTZ	TM+LR [d]
S1S2	2.0298	2.0352	2.0409	2.0313
S1C3	1.8077	1.8096	1.8149	1.8094
C3H4	1.0888	1.0890	1.0921	1.0887
C3H5	1.0865	1.0868	1.0893	1.0859
C3H6	1.0856	1.0859	1.0886	1.0852
C3S1S2	102.35	102.23	102.49	–[e]
S1C3H5	110.88	110.85	110.94	–[e]
S1C3H4	106.47	106.46	106.35	–[e]
S1C3H6	110.24	110.21	110.27	–[e]
C3S1S2C3b	84.77	84.72	85.21	–[e]
S2S1C3H4	178.03	177.96	177.54	–[e]
S2S1C3H5	-63.34	-63.42	-63.88	–[e]
S2S1C3H6	59.14	59.06	58.74	–[e]
A_e	8163.61	8131.83	8130.23	8182.68
B_e	2845.22	2839.05	2806.35	2829.29
C_e	2596.58	2591.05	2567.36	2587.61

[a] For atom labeling, see Figure 1. [b] From eq. 1. [c] From eq. 2. [d] Results from the "Lego-brick" approach. The LR approach has been applied to the C-S bond length, while the other distances have been corrected using the TM approach employing H_2S_2 and CH_4 as template molecules (see Figure 1). [e] Angles are kept fixed at the corresponding revDSD/junTZ values.

The accuracy and reliability of the results obtained for $(CH_3)_2S_2$ in the framework of the "Lego-brick" approach can also be verified by comparing the corresponding rotational constants with the experimental counterparts. The results are summarized in first block of Table 4, where the experimental ground-state rotational constants of the main isotopic species are compared with those computed at different levels: revDSD/junTZ, TM, and TM+LR. In order to make such a comparison meaningful, the computed equilibrium rotational constants have been augmented by vibrational corrections at the B3/apcs-1 level (obtained as explained in the Methodology section). It is noted that moving from revDSD/junTZ to TM, and then to TM+LR, the mean relative deviation reduces from 1.00% to 0.64%, and then to 0.49%.

The performance of the "Lego-brick" approach is not as good as noted in the literature,[21,22] but this case with two rotating methyl groups is particularly challenging not only for theory but also for the experimental determination.

As mentioned in the Methodology details, for $(CH_3)_2S_2$ and the larger members of the RSSR series considered, a partial SE equilibrium structure has been determined. For $(CH_3)_2S_2$, this is not available in the literature and has been derived for the first time in this work. To this end, the experimental rotational constants of the parent and five isotopic species, the mono-^{34}S, the mono-^{33}S, mono-^{13}C, and mono- and doubly-CD_3 substituted isotopologues, reported in refs. 16,89 have been employed. Despite the fact that five isotopic species provide 15 rotational constants and the structural parameters to be determined are 13, a full semi-experimental structure could not be derived. This is mainly due to two main reasons. The first one is the high correlation between parameters, which occurs whenever one atom is close to a principal inertial axis. This was already pointed out for dimethyl sulfide, the mono-S analogous of $(CH_3)_2S_2$.[90,91] Indeed, the in-plane H atom is extremely close to the *b*-principal axis, thus leading to a very high correlation between the C-H bond distance and the SCH angle. The second reason is related to the presence of two methyl groups, which are internal rotors giving rise to sizable internal rotation splitting (well resolved experimentally) and also imply the presence of low-frequency LAM vibrations. For them, the VPT2 treatment is not adequate. To overcome this issue, as explained in the Methodology section, the RD-VPT2 approach has been employed. However, excluding one or more vibrational modes from the VPT2 treatment introduces an inaccuracy which affects all spectroscopic quantities.

The SE results are collected in Table 5; the structural parameters not reported in Table 5 are those that could not be determined in the fit and have been kept fixed at the TM counterparts. As mentioned in the Computational details, the LR approach is applied only to bond lengths. This means that the only LR-corrected distance is CS, which is however one of the fitted parameters. From the comparison of the SE results (Table 5) with those reported in Table 3, it is noted a reasonably good agreement, within the statistical uncertainties, of

Table 4: Computed rotational constants of RSSR, with R = CH_3, CH_3CH_2, $(CH_3)_2CH$, CH_2CHCH_2, C_6H_5, compared to the experimental counterparts.[a]

$(CH_3)_2S_2$				
$B_0^{i,\mathrm{exp}}$ [b]	$\Delta B^i_{\mathrm{vib}}$	$B_0^{i,\mathrm{theo}}$ [c]		
ref. 16	B3/apcs-1	revDSD/junTZ	TM	TM+LR
8163.830(1)	-106.49	8023.74 (1.72)	8036.82 (1.56)	8076.19 (1.07)
2816.435(1)	-13.877	2792.48 (0.85)	2809.50 (0.25)	2815.41 (0.04)
2570.352(1)	-7.902	2559.46 (0.42)	2572.93 (0.10)	2579.71 (0.36)
$(CH_3CH_2)_2S_2$				
$B_0^{i,\mathrm{exp}}$	$\Delta B^i_{\mathrm{vib}}$	$B_0^{i,\mathrm{theo}}$ [c]		
ref. 17	B3/apcs-1	revDSD/junTZ	TM	TM+LR
4294.4281(9)	-33.968	4251.84 (0.99)	4270.75 (0.55)	4281.14 (0.31)
1255.5735(4)	-4.388	1251.22 (0.35)	1255.34 (0.02)	1259.41 (0.31)
1182.0645(3)	-4.185	1176.52 (0.47)	1181.49 (0.05)	1184.38 (0.20)
$((CH_3)_2CH)_2S_2$				
$B_0^{i,\mathrm{exp}}$	$\Delta B^i_{\mathrm{vib}}$	$B_0^{i,\mathrm{theo}}$ [c]		
ref. 19	B3/apcs-1	revDSD/junTZ	TM	TM+LR
2559.3242(3)	-28.833	2523.28 (1.41)	2527.88 (1.23)	2536.91 (0.88)
752.3375(2)	-3.106	748.70 (0.48)	751.67 (0.09)	753.85 (0.20)
692.7608(3)	-4.99	688.38 (0.63)	691.01 (0.24)	692.92 (0.02)
$(CH_2CHCH_2)_2S_2$				
$B_0^{i,\mathrm{exp}}$	$\Delta B^i_{\mathrm{vib}}$	$B_0^{i,\mathrm{theo}}$ [c]		
ref. 18	B3/apcs-1	revDSD/junTZ	TM	TM+LR
1784.34618(81)	-8.95	1781.49 (0.16)	1788.00 (0.20)	1791.12 (0.38)
981.02970(50)	-2.22	977.56 (0.35)	981.03 (0.0002)	983.73 (0.28)
879.83311(48)	-1.80	882.87 (0.36)	885.19 (0.61)	888.15 (0.95)
$(C_6H_5)_2S_2$				
$B_0^{i,\mathrm{exp}}$	$\Delta B^i_{\mathrm{vib}}$	$B_0^{i,\mathrm{theo}}$ [c]		
ref. 20	B3/apcs-1	revDSD/junTZ	TM	TM+LR
1106.06772(30)	-2.67	1090.93 (1.37)	1093.71 (1.12)	1095.69 (0.94)
321.36238(15)	-2.77	322.98 (0.51)	324.22 (0.89)	324.83 (1.08)
283.41966(15)	-1.87	283.71 (0.10)	284.70 (0.45)	285.28 (0.66)

[a] In parentheses, the relative deviations with respect to experiment. [b] Principal axis method (PAM) used in II^r representation. [c] Computed equilibrium rotational constants augmented by computed vibrational corrections at the B3/apcs-1 level.

Table 5: Partial semi-experimental equilibrium structures for the disulfide compounds considered in this study.

	$(CH_3)_2S_2$		$(CH_3CH_2)_2S_2$		$((CH_3)_2CH)_2S_2$	
Parameter	r_0	r_e^{SE}	r_0	r_e^{SE}	r_0	r_e^{SE}
SS	2.0472(8)	2.036(5)	2.036(1)	2.028(4)	2.062(6)	1.96(2)
CS	1.8060(5)	1.804(3)	1.824(2)	1.829(8)	1.835(3)	1.837(9)
CC	-	-	1.519(2)	1.504(6)	-	-
CSS	102.47(2)	102.5(1)	102.91(9)	102.7(3)	102.65(2)	105.67(5)
CSSC	84.44(3)	84.7(2)	87.0(1)	87.8(5)	88.6(2)	84.8(6)
σ	8.7×10^{-4}	6.3×10^{-3}	4.1×10^{-3}	4.0×10^{-2}	5.8×10^{-3}	1.8×10^{-2}
	$(CH_2CHCH_2)_2S_2$		$(C_6H_6)_2S_2$			
Parameter	r_0	r_e^{SE}	r_0	r_e^{SE}		
SS	2.00(1)	1.99(1)	2.17(2)	2.12(2)		
CS1	1.84(3)	1.83(3)	1.739(6)	1.74(1)		
CS2	1.87(3)	1.87(3)	-	-		
CSS	-	-	105.2(2)	-		
σ	5.5×10^{-2}	5.4×10^{-2}	1.14×10^{-2}	2.1×10^{-2}		

the former with the gradient-MP2-CCSD(T) and ChS geometries. However, it has to be pointed out that the statistical errors are by far larger than what expected from the SE approach and its intrinsic accuracy (i.e. 1 mÅ for bond lengths and 0.1 degree for angles). This outcome is clearly related to the issues discussed above. In absolute terms, it is observed that, while the SE angles well agree with those at the gradient-MP2-CCSD(T) and ChS levels, the S-S distance is ∼4 mÅ and ∼9 mÅ longer than the gradient-MP2-CCSD(T) and ChS values, respectively. Instead, the SE C-S bond is ∼5 mÅ and ∼7 mÅ shorter than the gradient-MP2-CCSD(T) and ChS counterparts, respectively. Such a behavior of the SE equilibrium values for the S-S and C-S bond lengths suggests a strong correlation between these two parameters. In Table 5, the partial r_0 structure is also given. This is obtained from the LSQ fit of the experimental ground-state rotational constants without accounting for the vibrational corrections. Therefore, it is an effective structure describing the molecule on the vibrational ground state. As expected, the bond lengths are longer than the equilibrium counterparts, but what is interesting is that r_0 is better determined than r_e^{SE}, the statistical uncertainties of the former being smaller at least by one order of magnitude than those of

the latter. This and the deviations from pure theoretical structures, discussed above, suggest that the RD-VPT2 is not so effective in overcoming the LAM issues and, thus, the RD-VPT2 vibrational corrections to rotational constants are not as accurate as the VPT2 ones for semi-rigid molecules.

Going a step further along the RSSR series, the next molecule to be considered is diethyl disulfide. For $(C_2H_5)_2S_2$, because of its size, only the "Lego brick" approach has been considered: the revDSD/junTZ, TM and TM+LR structural parameters are reported in the Supporting Information (SI; Table S1). Their accuracy is here estimated by comparing the corresponding rotational constants with the experimental counterpart. As done for $(CH_3)_2S_2$, for a meaningful comparison with experiment, the revDSD/junTZ, TM and TM+LR equilibrium rotational constants have been augmented for vibrational corrections computed at the B3/apcs-1 level using RD-VPT2. From the inspection of Table 4, it is noted that moving from revDSD/junTZ to TM the average deviation lowers from ∼0.60% to ∼0.21%, with discrepancies as small as 0.02% and 0.05% for B and C, which are in line with what obtainable for semi-rigid molecules.[22] As evident from Figure 1, the fragments envisaged for $(C_2H_5)_2S_2$ are H_2S_2 and ethane (C_2H_6). Therefore, in analogy to $(CH_3)_2S_2$, the unique structural parameter that can be LR corrected is the C-S distance. Incorporating such LR correction improves the agreement for A, but it reduces that for B and C, the overall mean deviation increasing from ∼0.21% to ∼0.27%. The TM structure has then been employed in the determination of the partial SE equilibrium geometry, with the non-determinable parameters kept fixed at the TM values. Since the C-S bond distance is one of the determinable (fitted) parameters, the TM and TM+LR structures provide the same geometrical information. The results are collected in Table 5. It is noted that, similarly to $(CH_3)_2S_2$, the statistical uncertainties are larger than what expected from the SE approach.

For the remaining molecules, namely diisopropyl disulfide ($(C_3H_7)_2S_2$), diallyl disulfide ($(C_3H_5)_2S_2$), and diphenyl disulfide ($(C_6H_5)_2S_2$), the "Lego brick" approach was again the only feasible theoretical model that allows for deriving accurate equilibrium structures at a reduced

computational cost. While the revDSD/junTZ, TM and TM+LR structural parameters are reported in the SI (Tables S2-S4), their accuracy can be addressed from the inspection of Table 4. As done for $(C_2H_5)_2S_2$, the corresponding equilibrium rotational constants have been corrected for B3/apcs-1 vibrational corrections (from RD-VPT2 calculations) in order to meaningfully compare them with experiment. Before proceeding with the discussion of Table 4, it has to be noted that, as evident from Figure 1, the TM fragments envisaged are H_2S_2 for all the three molecules here considered and propylene (CH_3CHCH_2) for diallyl disulfide, benzene (C_6H_6) for diphenyl disulfide, and methane (CH_4) together with ethane for diisopropyl disulfide ($(C_3H_7)_2S_2$). Therefore, for $(C_3H_5)_2S_2$ and $(C_6H_5)_2S_2$, CS is the only parameter that can be LR corrected, while for $(C_3H_7)_2S_2$, in addition to CS, there is also the C3-C4 distance (see Figure 1). Focusing on diisopropyl disulfide, it is noted that the average deviation from the experimental ground state rotational constants of the computed counterparts improves from 0.84% at the revDSD/junTZ level to 0.52% for TM and to 0.37% for TM+LR. Different is instead the behavior for diallyl disulfide and diphenyl disulfide. For the former, at the revDSD/junTZ level, the agreement is better than what usually expected, the mean deviation of the rotational constants being 0.29%. Moving to the TM approach, the improvement is negligible (0.27%), while the introduction of the LR correction leads to a marked worsening (0.54%). For diphenyl disulfide, both the TM and TM+LR approaches lead to an increase of the average deviation: from 0.66% at the revDSD/junTZ level to 0.82% for TM and 0.89% for TM+LR. However, for such a large system, we expected that the spatial distribution of the phenyl rings and the associated LAMs significantly affect rotational constants and the corresponding vibrational corrections. Therefore, we consider the TM-corrected values as accurate estimates. This means that we do not expect a particular degradation of the accuracy of the TM/TM+LR approaches by enlarging the size of the R substituent. The overall conclusion that seems to be drawn from Table 4 is that, for the RSSR systems considered, revDSD/junTZ and, especially, TM and TM+LR perform worse than expected. While the presence of LAMs surely prevents from

obtaining results as accurate as those noted for rather rigid systems like PAHs,[22] we do not expect the degradation of accuracy here noted. However, it is surely difficult to understand to what extent the RD-VPT2 inaccuracy impacts on the outcome obtained.

For diisopropyl disulfide, diallyl disulfide, and diphenyl disulfide, the partial r_e^{SE} has also been determined and the results are collected in Table 5. The non-determinable geometrical parameters have been kept fixed at the TM counterparts, the only exception being $(C_3H_7)_2S_2$. Indeed, for this latter, the LR-corrected value has been considered for the C3-C4 distance. As noted in the discussion of Table 4, diallyl disulfide and diphenyl disulfide are the systems that show the worst results. A similar outcome is obtained when moving to the SE determination. For both r_e^{SE} and r_0, the statistical uncertainties (10 to 30 mÅ for bond distances) are one-two orders of magnitude larger than what is usually expected from such approaches. Focusing on the S-S distance, while for $(C_3H_5)_2S_2$ the SE value is about 40 mÅ shorter that the TM counterpart, for $(C_6H_5)_2S_2$ it results about 100 mÅ longer than the TM value. Finally, for diisopropyl disulfide, we have obtained results that are a little better than those just discussed, but still worse –in terms of statistical errors– than those of $(CH_3)_2S_2$ and $(C_2H_5)_2S_2$. For $(C_3H_7)_2S_2$, the S-S distance is not well determined, 1.96(2) Å, and it seems to be too short based on the revDSD/junTZ (2.0398 Å) and TM (2.0268) values. The conclusion that can be drawn from the inspection of Table 5 is that it becomes more and more difficult to perform a partial SE fit by increasing the dimension of the molecule. Therefore, for the largest members of the RSSR series, further experimental work would be necessary and, at the same time, a variationally-corrected VPT2 model would be required in order to derive accurate (partial) r_e^{SE} structures. Indeed, an improved approach for the derivation of the vibrational corrections to rotational constants is needed and this would also help understand the effective accuracy obtainable from the TM and TM+LR approaches.

The last comment concerns the trend of the S-S distance along the series of the RSSR molecules here considered. At the revDSD/junTZ level, it is noted that, for R = CH_3, C_2H_5, C_3H_5, and C_3H_7, this bond length is about 2.04 Å, while it is longer (∼2.06 Å) for H_2S_2 and

shorter (∼2.03) for R = C_6H_5. Moving to the TM-corrected values, a systematic shortening by about 10 mÅ is observed. From the comparison of these latter results with the r_e^{SE} ones (Table 5), an agreement within the statistical uncertainties is observed only for R = CH_3 and C_2H_5, while for the largest R groups, as already noted, the SE values are either too short (R = C_3H_5 and C_3H_7) or too long (R = C_6H_5).

In the original experimental works,[16–20] the availability of ground-state rotational constants for different isotopic species led to the determination of empirical structures. In details, for dimethyl disulfide, diethyl disulfide and diisopropyl disulfide, the so-called substitution structure (r_s)[92] was reported. Starting from the rotational constants for the vibrational ground state, this approach tries to account for vibrational effects by evaluating the Cartesian coordinates of the atom isotopically substituted using Kraitchman's equations.[93] For $(CH_3)_2S_2$, the partial r_s structure of ref. 16 provides rather good results, with the S-S and C-S distances being about 3 mÅ shorter and ∼5 mÅ longer than the gradient-MP2-CCSD(T) reference values, respectively. Moving to $(C_2H_5)_2S_2$, in ref. 17, Zhang et al. determined 7 geometrical parameters (3 bond lengths, two valence angles and two dihedral angles) using the r_s methodology. While the C-C distance well agrees with our TM value, the S-S and C-S ones result 7 mÅ and 8 mÅ underestimated and overestimated, respectively, thus pointing out a clear correlation of the two parameters. For $(C_3H_7)_2S_2$, the partial r_s structure reported by Zhang et al. in ref. 19 appears rather unreliable, probably because of the too many parameters considered. Indeed, with respect to our TM+LR results, the S-S, S-C3 and C4-C3 bond lengths are 12 mÅ, 11 mÅ and 20 mÅ shorter, while C5-S3 appears 50 mÅ longer. Furthermore, most of the structural parameters were poorly determined with uncertainties that can be as large as 10-20 mÅ for distances and 0.7-1.2 degrees for angles. Noted is also the correlation between the C5C3S1 and C4C3S1 valence angles, whose r_s values are, respectively, 2 degrees larger and 1.5 degrees smaller than the revDSD/junTZ counterparts, whose accuracy is expected to be better than 0.3-0.4 degrees. For diallyl disulfide and diphenyl disulfide, Demaison and coworkers[18,20] instead reported the so-called r_m mass-dependent structure,[94]

which has two variants, namely $r_m(1)$ and $r_m(2)$, with 1 and 2 indicating the number of additional parameters (for each inertial axis) to be evaluated.[94] Among the methodologies that account for vibrational effects in a non-rigorous manner, r_m provides the best approximation to the equilibrium structure. Indeed, the accuracy usually ranges in the 2-7 mÅ interval for bond lengths and 0.4-1° for angles.[25,95–97] For $(C_3H_5)_2S_2$, in ref. 18, a complete $r_m(1)$ structure for heavy atoms (C and S) could be determined owing to the combination of the r_m approach with the method of predicate observations (or mixed regression).[98] This latter employs auxiliary information, denoted as predicates, which come from quantum-chemical computations and are added, with appropriate weights, to the LSQ fit. While the $r_m(1)$ results point out that the SE value of the S-S distance is shorter by about 28 mÅ, they confirm the accuracy of the TM+LR structures, the discrepancies being well within 1-2 mÅ for bond lengths. Furthermore, the $r_m(1)$ values corroborate the validity of fixing the TM/TM+LR angles at the revDSD/junTZ values; in fact, the differences between these latter and the $r_m(1)$ counterparts are well within 0.2-0.3 degrees. Finally, it has to be noted that, in ref. 18, the authors refrain from exploiting the SE approach because of the unreliability of predictions based on VPT2 applied to a cubic force field in presence of LAMs. Similarly, in ref. 20 for $(C_6H_5)_2S_2$, the inspection of the results from anharmonic force-field computations pointed out their inadequacy to calculate accurately vibrational corrections and their small variations upon isotopic substitution. In analogy to $(C_3H_5)_2S_2$,[18] the authors employed the r_m mass-dependent approach combined with the mixed regression method to obtain a complete $r_m(2)$ structure for $(C_6H_5)_2S_2$. Interestingly, the same conclusions as those drawn for diallyl disulfide are obtained, that is the high accuracy of the TM+LR approach together with its use in combination with the revDSD/junTZ angles.

Benchmark study

Basis set effects

As first step of the benchmark study, the basis-set convergence has been addressed using the smallest molecule of the RSSR family, H_2S_2, as a test case. This has been accomplished by considering different hybrid and double-hybrid functionals in conjunction with a wide range of basis sets up to the quintuple-ζ quality. As a starting point, the convergence of the electronic energy at the B2PLYP-D3BJ level has been examined. Three basis-set families, (aug-)cc-pVnZ with n = D, T, Q, 5, (aug-)pcs-n with n = 1, 2, 3, 4, and cc-pVnZ-F12 with n = D, T, Q, have been considered. The corresponding trends are illustrated in Figure 2, where the results obtained with the largest set of each series of bases have been taken as reference for the corresponding family. From the inspection of this figure, it is evident that the cc-pVnZ-F12 series shows the fastest convergence. However, it has to be noted that the $(n-1)$Z-F12 basis is approximately a set of n-ζ quality (despite lacking the highest angular momentum functions of the latest). This is somewhat evident when counting the number of basis functions, which is, for H_2S_2, 178, 194 and 260 for cc-pVQZ, pcs-3 and cc-pVQZ-F12, respectively. Back to Figure 2, it is apparent that the convergence is reached for cc-pVQZ, pcs-3 and cc-pVTZ-F12. However, as noted above, the number of basis functions is already large for a molecule as small as H_2S_2; therefore, they are not affordable for larger systems and/or extensive exploration of potential energy surfaces (PES) as needed, for example, in spectroscopic and chemical reactivity studies.

Since the focus of this work is on structural determinations, the basis-set convergence has been investigated on the geometrical parameters of H_2S_2. Two double-hybrid functionals, namely B2PLYP-D3BJ and DSD-PBEP86-D3BJ, and one global-hybrid functional, namely B3LYP-D3BJ, have been employed. The basis sets considered are of double- and triple-ζ quality, the only exception being the exclusion of the cc-pVTZ-F12 set because of what discussed above; furthermore, it is not affordable for larger molecules. Additionally, the cost-

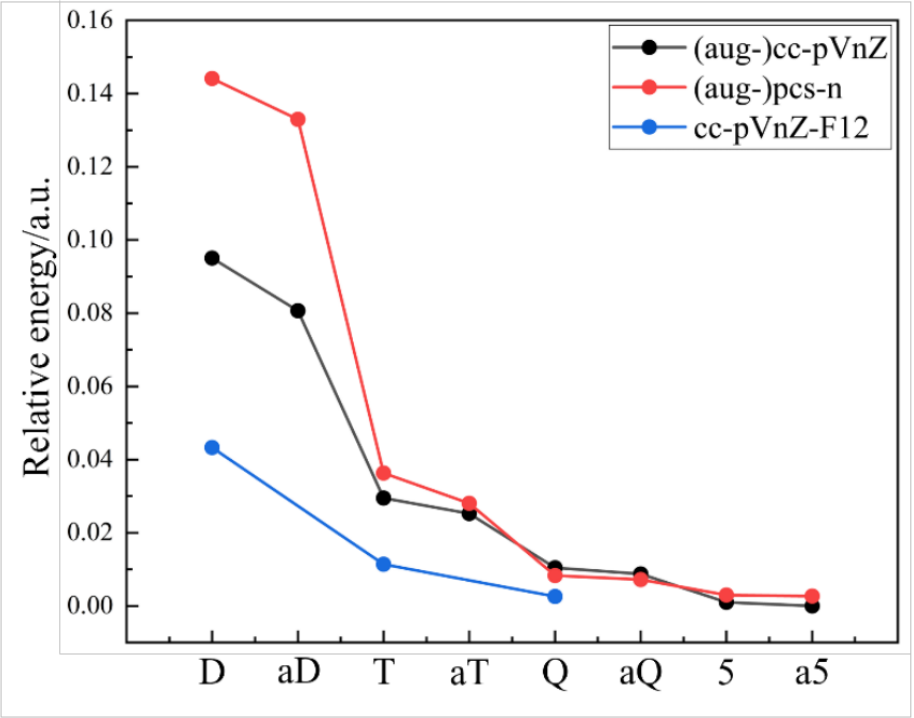


Figure 2: Basis-set convergence of the B2PLYP-D3BJ relative energy for (aug-)cc-pVnZ (with n = D, T, Q, 5), (aug-)pcs-n (with n = 1, 2, 3, 4), and cc-pVnZ-F12 (with n = D, T, Q).

effective "seasonal" jun-cc-pVTZ basis set has also been included. The results are graphically summarized in Figure 3, which reports the mean absolute errors (MAEs) of the H_2S_2 bond lengths and angles with respect to the theoretical BestEst structure. It is noted that all functionals show a similar trend: the largest MAEs are obtained with the double-ζ basis sets. For bond lengths, these are on the order of 30 mÅ for the double-hybrid functionals and 40 mÅ for B3LYP-D3BJ. For angles, the performance of the pcs-1 basis set is instead very good with all functionals, while using aug-cc-pVDZ leads to MAEs as large as 0.4-0.6 degrees. Interestingly, as far as the double-ζ basis sets are concerned, the pcs-1 and aug-pcs-1 sets, despite being smaller, perform better than aug-cc-pVDZ. Therefore, they might be considered a viable choice for expensive computations, as in the case of anharmonic force field calculations. Overall, for B2PLYP-D3BJ and DSD-PBEP86-D3BJ, the best performance, with MAEs smaller than 10 mÅ and 5 mÅ for bond distances, respectively, is obtained with pcs-2 and aug-pcs-2. In contrast, for B3LYP-D3BJ, all basis sets show MAEs larger than 15

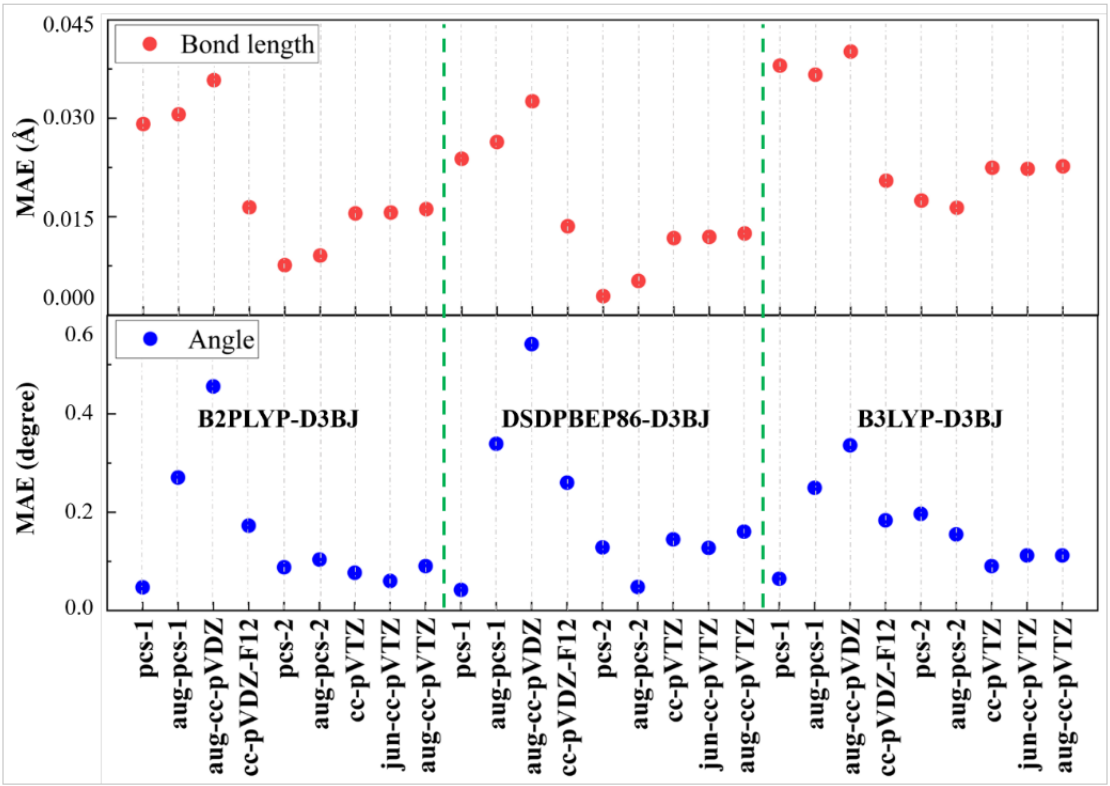


Figure 3: Accuracy of the H_2S_2 structural parameters computed using different DFT functionals in conjunction with selected basis sets. Mean absolute errors (MAEs) with respect to the BestEst reference (see text) for bond lengths (SS and SH, upper panel) and angles (SSH and HSSH, lower panel) are shown.

mÅ. Moreover, among the triple-ζ quality basis sets from Dunning's family, jun-cc-pVTZ provides the best performance at an affordable cost with respect to the standard aug-cc-pVTZ basis. In summary, also based on the MAEs obtained for angles, the aug-pcs-2 basis set is recommended for larger molecules. Owing to the incorporation of diffuse functions, jun-cc-pVTZ offers a good, less expensive, alternative to aug-pcs-2 for the description of the possible weak interactions arising in systems with large substituents.

Benchmark of DFT functionals

Based on the results of the previous section, the performance of different DFT functionals is evaluated by employing the aug-pcs-2 basis set for all molecules except the largest one, $(C_6H_5)_2S_2$. For the latter, in addition to the excessive computational cost, in some cases (for

example, with the B2PLYP functional) aug-pcs-2 yielded problems in the SCF convergence because of the linear dependence of this basis set. These issues have been attributed to the presence of a large number of diffuse basis functions.[66] Therefore, the well-performing (see Figure 3) yet smaller jun-cc-pVTZ basis set has been employed as a feasible and safe alternative.

Figure 4 reports the MAEs with respect to the reference values as well as the largest positive (MAX) and negative (MIN) errors for the key parameters of the disulfide linkage, namely the S-S and C-S bonds, the valence SSC and dihedral CSSC angles. As reference, the most accurate theoretical structures are used. These are: BestEst for H_2S_2, gradient-MP2-CCSD(T) for $(CH_3)_2S_2$, TM+LR for $(CH_3CH_2)_2S_2$, $(CH_2CHCH_2)_2S_2$, $((CH_3)_2CH)_2S_2$, and $(C_6H_6)_2S_2$. The S-S, C-S, SSC and CSSC values for the levels of theory considered in Figure 4 are collected in the SI together with the reference data (Tables S5-S10).

First of all, we note a quite large distribution of both the S-S and C-S bond lengths for the different DFT levels considered (over a range of 35-50 mÅ), with MAEs ranging from a few mÅ to 25 mÅ; the MIN and MAX errors are about -15/-20 and +35/+40 mÅ. The SSC values vary within the range of about 2 degrees, with MAEs ranging from 0.2 degrees to 1.5 degrees, and overestimation (MAX) up to 1.8 degrees and underestimation (MIN) within 1.2 degree. For the CSSC dihedral angle, the variations are in a range of about 4 degrees, which is even larger, about 5 degrees, for the $(C_2H_5)_2S_2$ and, almost 7 degrees, for $(C_6H_5)_2S_2$. The MAEs vary up to 2 degrees, while the MAX error can be as large as $\sim$5 degrees and MIN can reach -3 degrees. These results highlight the difficulty in accurately predict the RSSR structures, and point out the need for a dedicated benchmark study.

Going into details, for the S-S bond length, the best agreement with the reference, with MAE and MAX deviations of about 5 and 10 mÅ respectively, has been obtained for the PBE0(-D3BJ), ω-B97X, M06-2X(-D3) and DSD-PBEP86-D3BJ functionals. Moreover, DSD-PBEh95(-D3BJ) and APF(D) yield MAE below 5 mÅ but in conjunction with MAX values of 12-16 mÅ the outliers being always related to $(C_6H_5)_2S_2$. Among the best performing

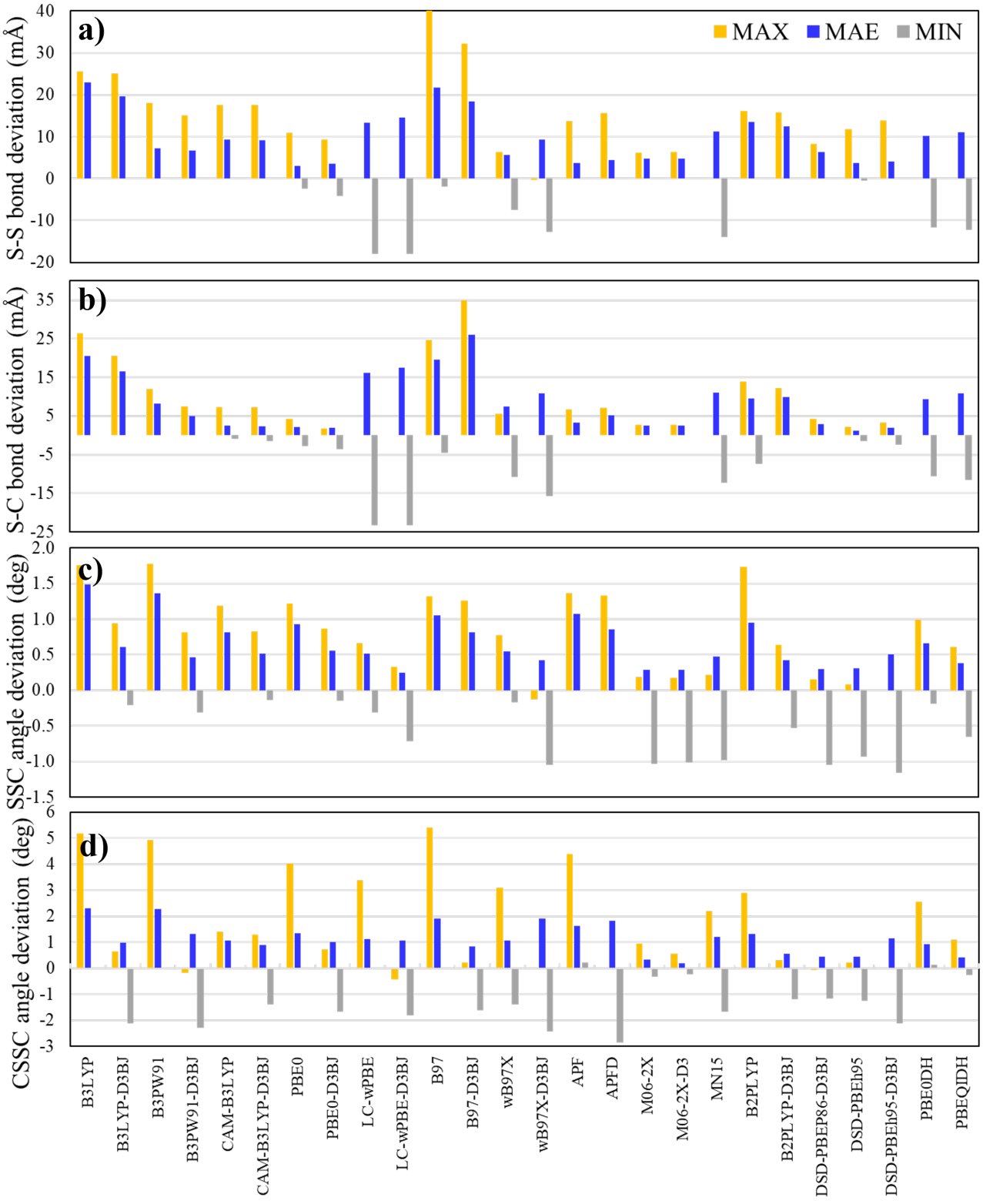


Figure 4: Performance of the DFT functionals for the RSSR disulphide linkage structural parameters. The mean absolute (MAE), largest positive (MAX) and negative (MIN) deviations with respect to the reference (see text) for the (a) S-S bond, (b) S-C bond, (c) SSC angle, and (d) CSSC dihedral angle are reported.

functionals, M06-2X(D3) and DSD-PBEP86-D3BJ also provide the correct (according to the reference) trend for the S-S bond distance, from the longest to shortest, namely $H_2S_2 \gg (CH_3CH_2)_2S_2 > ((CH_3)_2CH)_2S_2 \approx (CH_3)_2S_2 > (CH_2CHCH_2)_2S_2 \gg (C_6H_6)_2S_2$. B2PLYP(-D3BJ) reproduces very well the relative changes along the RSSR series, despite showing rather large MAE and MAX values of 12-14 mÅ and 16 mÅ, respectively. Finally, B3LYP also correctly describes the S-S bond length trend, despite significant larger errors, which are >20 mÅ . For all the molecules considered, most of the DFT functionals predict S-S bond lengths too long (positive deviations). PBE0(-D3BJ) and ω-B97(-D3BJ) show both positive and negative deviations, while the remaining ones from the PBE family (LC-ωPBE(-D3BJ), PBE0DH, PBEQIDH) and MN15 predict S-S bonds too short. It is noted that the long-range correction always leads to S-S bonds shorter than their uncorrected-DFT counterparts. For CAM-B3LYP, this leads to an improvement over B3LYP, while for LC-ωPBE to a worsening because of the already shorter (with respect to the reference) PBE0 distances. Dispersion corrections, as expected, do not have significant effect on the S-S bond length, highlighting once more that D3BJ corrections are suitable for flexible biological molecules, which often require dispersion to be accounted for.

For the C-S bond lengths, a better accuracy of the DFT parameters is observed. Indeed, smaller errors are noted, with MAEs well within 4 mÅ for all the DFT functionals that performed well for the S-S bond, namely PBE0(-D3BJ), M06-2X(-D3), DSD-PBEP86-D3BJ and DSD-PBEh95(-D3BJ). Remarkable is also the good performance of the APF(D) functional with MAE of 3-5 mÅ and MAX of 7 mÅ. Concerning the SSC valence angle, the M06-2X(-D3), DSD-PBEP86-D3BJ and DSD-PBEh95(-D3BJ) functionals are again the best performing, with MAEs within 0.5 degrees and MIN errors up to 1 degree; instead, larger MAEs of 0.6-0.9 degrees are shown by PBE0(-D3BJ). For the CSSC dihedral angle, the improvement obtained by adding dispersion corrections is evident, this indeed lowers the largest discrepancies from 5 degrees to about 2-3 degrees or better. As expected, the most significant effect is observed for $(C_6H_5)_2S_2$, which presents a strong dispersion interaction between the two phenyl rings.

Those functionals that showed the smallest errors for the RSSR systems were already included in previous studies where their performance for structural determinations was analyzed.[77,99–102] Overall, a good accuracy has been noted,[77,99] even for amino acids.[100–102] However, some differences have to be pointed out. In the present study, the double-hybrid functionals from the DSD-PBE family exhibit a significantly better performance compared to B2PLYP, while in the previous works, they were found to provide similar (good) accuracy. Moreover, for the semi-experimental equilibrium structures of 47 organic molecules (SE47 database[103]), the MN15 functional was found to provide slightly better results[77] than M06-2X and ω-B97-D, which –instead– is not the case for the RSSR systems.

Conclusion

For selected RSSR molecules, with R=H, CH_3, C_2H_5, C_3H_5, C_3H_7 and C_6H_5, the structural parameters have been investigated by means of computational and semi-experimental methodologies. The most accurate data have subsequently been used as reference to test the performance of several density functionals. For the smallest member of the RSSR family, H_2S_2, the gradient-MP2-CCSD(T) composite scheme has been validated by comparing its results with those issued by the significantly more expensive one entirely based on CC techniques which also incorporates up to quadruple excitations:[13] a near perfect agreement has been noted. The accurate determination of the S-S bond length carried out in ref. 13 opened the way for the exploitation of the "Lego brick" (TM+LR) approach for disulfide molecules. For dimethyl disulfide, the TM+LR structure shows a very good agreement with the gradient-MP2-CCSD(T) counterpart, with discrepancies well within 1-2 mÅ for bond lengths. For two large members of the RSSR family, namely diallyl disulfide and diphenyl disulfide, the TM+LR geometries could be compared with accurate experimental mass-dependent structures available in the literature.[18,20] Such a comparison together with that for $(CH_3)_2S_2$ allowed us to point out the reliability and accuracy of the "Lego brick"

approach also for flexible systems. For this reason, the TM+LR geometries were used in the subsequent benchmark study of DFT methodologies.

The semi-experimental approach, which is renowned for its great accuracy, instead provided disappointing results both in terms of values and associated errors. This led us to the conclusion that the RD-VPT2 model is unable to overcome the issues related to the presence of LAMs. More precisely, its performance deteriorates significantly by increasing the number of LAMs and thus the flexibility of the system. Therefore, while improving with respect to VPT2, RD-VPT2 is still not suitable for the computation of vibrational corrections to rotational constants with the required accuracy.

Moving to the benchmark DFT study, we note that the challenge of accurately describing the disulphide bond linkage is also evident in the analysis of the results obtained for several DFT methodologies. Indeed, rather large average deviations of about or more than 10 mÅ are noted for the S-S bond. These outcomes can be compared wtih those from other benchmark studies employing highly-accurate geometries as reference.[77,100–102,104] For them, DFT methodologies involving at least a single-hybrid GGA functional in conjunction with a basis set of double-ζ quality typically led to mean errors of about or smaller than 5 mÅ.[77,100–102,104] Among the single-hybrid GGA functionals, we found that PBE0(-D3BJ), ωB97X(-D3BJ) and also APF(D) are the best-performing. Among the single-hybrid Meta-GGA functional, M06-2X(-D3) performs much better than M15. In the group of double-hybrid functionals, we noted that DSD-PBEP86-D3BJ and DSDPBEh95(-D3BJ) provide better results than B2PLYP(-D3BJ). Interestingly, PBE0DH and PBEQIDH are the worst-performing double-hybrid functionals (among those tested here), despite the fact that they are variants of the well-performing PBE0. Finally, it is worthwhile noting that the structural parameters derived in this work represent a useful reference for future benchmark studies of novel DFT-based or other cost-effective methodologies.

Acknowledgement

This work has been supported by MUR (PRIN Grant Numbers 202082CE3T and P2022ZFNBL) and by the University of Bologna (RFO funds). The COST Action CA21101 "COSY - Confined molecular systems: from a new generation of materials to the stars" is acknowledged.

Supporting Information Available

In the SI the following information are reported: (i) revDSD/junTZ and TM/TM+LR structural parameters for diethyl disulfide ($(C_2H_5)_2S_2$; Table S1), diallyl disulfide ($(C_3H_5)_2S_2$; Table S2), diisopropyl disulfide ($(C_3H_7)_2S_2$; Table S3), and diphenyl disulfide ($(C_6H_5)_2S_2$; Table S4); (ii) Structural parameters of the disulfide linkage computed by means of the DFT functionals considered in this work for the RSSR systems, with R = H (Table S5), CH_3 (Table S6), C_2H_5 (Table S7), C_3H_5 (Table S8), C_3H_7 (Table S9), and C_6H_5 (Table S10).

References

(1) Mörner, K. Cystin, ein Spaltungsprodukt der Hornsubstanz. *Biol. Chem.* **1899**, *28*, 595–617.

(2) Cremlyn, R. J. W. *An Introduction to Organosulfur Chemistry*; Wiley, 1996.

(3) Qin, M.; Wang, W.; Thirumalai, D. Protein Folding Guides Disulfide Bond Formation. *P. Natl. Acad. Sci. USA* **2015**, *112*, 11241–11246.

(4) Wedemeyer, W. J.; Welker, E.; Narayan, M.; Scheraga, H. A. Disulfide Bonds and Protein Folding. *Biochem.* **2000**, *39*, 4207–4216.

(5) Sevier, C. S.; Kaiser, C. A. Formation and Transfer of Disulphide Bonds in Living Cells. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 836–847.

(6) Schmidt, B.; Ho, L.; Hogg, P. J. Allosteric Disulfide Bonds. *Biochem.* **2006**, *45*, 7429–7433.

(7) Wiedemann, C.; Kumar, A.; Lang, A.; Ohlenschläger, O. Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. *Front. Chem.* **2020**, *8*, 280.

(8) Fuente, A.; Rivière-Marichalar, P.; Beitia-Antero, L.; Caselli, P.; Wakelam, V.; Esplugues, G.; Rodríguez-Baras, M.; Navarro-Almaida, D.; Gerin, M.; Kramer, C. et al. Gas phase Elemental Abundances in Molecular CloudS (GEMS). VII. Sulfur Elemental Abundance. *Astron. Astrophys.* **2023**, *670*, A114.

(9) Martín-Doménech, R.; Jiménez-Serra, I.; Caro, G. M.; Müller, H.; Occhiogrosso, A.; Testi, L.; Woods, P.; Viti, S. The Sulfur Depletion Problem: Upper Limits on the H_2S_2, HS_2, and S_2 Gas-phase Abundances Toward the Low-mass Warm Core IRAS 16293-2422. *Astron. Astrophys.* **2016**, *585*, A112.

(10) Jiménez-Escobar, A.; Caro, G. M.; Ciaravella, A.; Cecchi-Pestellini, C.; Candia, R.; Micela, G. Soft X-Ray Irradiation of H_2S Ice and the Presence of S_2 in Comets. *Astrophys. J. Lett.* **2012**, *751*, L40.

(11) Zhu, C.; Góbi, S.; Abplanalp, M. J.; Frigge, R.; Gillis-Davis, J. J.; Kaiser, R. I. Space Weathering-induced Formation of Hydrogen Sulfide (H2S) and Hydrogen Disulfide (H2S2) in the Murchison Meteorite. *J. Geophys. Res. Planets* **2019**, *124*, 2772–2779.

(12) Santos, J. C.; Enrique-Romero, J.; Lamberts, T.; Linnartz, H.; Chuang, K.-J. Formation of S-bearing Complex Organic Molecules in Interstellar Clouds via Ice Reactions with C2H2, HS, and Atomic H. *ACS Earth and Space Chemistry* **2024**, *8*, 1646–1660.

(13) Ye, H.; Mendolicchio, M.; Kruse, H.; Puzzarini, C.; Biczysko, M.; Barone, V. The Challenging Equilibrium Structure of HSSH: Another Success of the Rotational Spectroscopy / Quantum Chemistry Synergism. *J. Mol. Struct.* **2020**, *1211*, 127933.

(14) Benabdelkrim, A.; Tourchi, A. E.; Hammoutène, D.; Yaghlane, S. B.; Abdallah, H. H.; Linguerri, R.; Hochlaf, M. Characterization of the Simplest Sulfenyl Thiocyanate: Isomers, Spectroscopy and Implications of Astrophysical and Biological Relevance. *Phys. Chem. Chem. Phys.* **2020**, *22*, 17052–17061.

(15) Nikoo, S.; Meister, P. J.; Hayward, J. J.; Gauld, J. W. An Assessment of Computational Methods for Calculating Accurate Structures and Energies of Bio-relevant Polysulfur/selenium-containing Compounds. *Molecules* **2018**, *23*, 3323.

(16) Meyer, M. Infrared, Raman, Microwave and ab initio Study of Dimethyl Disulfide: Structure and Force Field. *J. Mol. Struct.* **1992**, *273*, 99–121.

(17) Zhang, J.; Li, X.; Gou, Q.; Feng, G. Disulfide Bond in Diethyl Disulfide: A Rotational Spectroscopic Study. *J. Phys. Chem. A* **2018**, *122*, 5597–5601.

(18) Demaison, J.; Vogt, N.; Saragi, R. T.; Juanes, M.; Rudolph, H. D.; Lesarri, A. How Flexible is the Disulfide Linker? A Combined Rotational–Computational Investigation of Diallyl Disulfide. *Phys. Chem. Chem. Phys.* **2019**, *21*, 19732–19736.

(19) Zhang, J.; Ye, H.; Jin, Y.; Gou, Q.; Biczysko, M.; Feng, G. Conformational Equilibria and Molecular Structures of Model Sulfur–Sulfur Bridge Systems: Diisopropyl Disulfide. *J. Phys. Chem. A* **2019**, *123*, 10714–10720.

(20) Demaison, J.; Vogt, N.; Saragi, R. T.; Juanes, M.; Rudolph, H. D.; Lesarri, A. The S- S Bridge: A Mixed Experimental-Computational Estimation of the Equilibrium Structure of Diphenyl Disulfide. *ChemPhysChem* **2019**, *20*, 366–373.

(21) Melli, A.; Tonolo, F.; Barone, V.; Puzzarini, C. Extending the Applicability of the Semi-Experimental Approach by means of "Template Molecule" and "Linear Regression" Models on top of DFT Computations. *J. Phys. Chem. A* **2021**, *125*, 9904–9916.

(22) Ye, H.; Alessandrini, S.; Melosso, M.; Puzzarini, C. Exploiting the "Lego brick" Approach to Predict Accurate Molecular Structures of PAHs and PANHs. *Phys. Chem. Chem. Phys.* **2022**, *24*, 23254–23264.

(23) Pulay, P.; Meyer, W.; Boggs, J. E. Cubic Force Constants and Equilibrium Geometry of Methane from Hartree-Fock and Correlated Wavefunctions. *J. Chem. Phys.* **1978**, *68*, 5077–5085.

(24) Pawłowski, F.; Jørgensen, P.; Olsen, J.; Hegelund, F.; Helgaker, T.; Gauss, J.; Bak, K. L.; Stanton, J. F. Molecular Equilibrium Structures from Experimental Rotational Constants and Calculated Vibration–rotation Interaction Constants. *J. Chem. Phys.* **2002**, *116*, 6482–6496.

(25) Demaison, J. Experimental, Semi-Experimental and Ab Initio Equilibrium Structures. *Mol. Phys.* **2007**, *105*, 3109–3138.

(26) Puzzarini, C.; Stanton, J. F.; Gauss, J. Quantum-Chemical Calculation of Spectroscopic Parameters for Rotational Spectroscopy. *Int. Rev. Phys. Chem.* **2010**, *29*, 273–367.

(27) Demaison, J., Boggs, J. E., Császár, A. G., Eds. *Equilibrium Molecular Structures: From Spectroscopy to Quantum Chemistry*; CRC Press, Taylor & Francis Group: Boca Raton, FL, US, 2011.

(28) Puzzarini, C.; Stanton, J. F. Connections between the Accuracy of Rotational Constants and Equilibrium Molecular Structures. *Phys. Chem. Chem. Phys.* **2023**, *25*, 1421–1429.

(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(30) Stanton, J. F.; Gauss, J.; Cheng, L.; Harding, M. E.; Matthews, D. A.; Szalay, P. G.; and contributors. CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package. For the current version, see http://www.cfour.de.

(31) Matthews, D. A.; Cheng, L.; Harding, M. E.; Lipparini, F.; Stopkowicz, S.; Jagau, T.-C.; Szalay, P. G.; Gauss, J.; Stanton, J. F. Coupled-Cluster Techniques for Computational Chemistry: The CFOUR Program Package. *J. Chem. Phys.* **2020**, *152*, 214108.

(32) Parrish, R. M.; Burns, L. A.; Smith, D. G.; Simmonett, A. C.; DePrince III, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M. et al. Psi4 1.1: An Open-source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185–3197.

(33) Kraus, P.; Frank, I. Validating Additive Correction Schemes Against Gradient-based Extrapolations. *Int. J.Quant. Chem.* **2019**, *119*, e25953.

(34) Bozkaya, U.; Sherrill, C. D. Analytic Energy Gradients for the Coupled-cluster Singles and Doubles with Perturbative Triples Method with the Density-Fitting Approximation. *J. Chem. Phys.* **2017**, *147*, 044104.

(35) Peterson, K. A.; Dunning Jr, T. H. Accurate Correlation Consistent Basis Sets for Molecular Core–valence Correlation Effects: The Second Row Atoms Al–Ar, and the First Row Atoms B–Ne Revisited. *J. Chem. Phys.* **2002**, *117*, 10548–10560.

(36) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618.

(37) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set Convergence of Correlated Calculations on Water. *J. Chem. Phys.* **1997**, *106*, 9639–9646.

(38) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in Correlated Calculations on Ne, N_2, and H_2O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(39) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(40) Puzzarini, C. Extrapolation to the Complete Basis Set Limit of Structural Parameters: Comparison of Different Approaches. *J. Phys. Chem. A* **2009**, *113*, 14530–14535.

(41) Puzzarini, C.; Barone, V. Extending the Molecular Size in Accurate Quantum-Chemical Calculations: the Equilibrium Structure and Spectroscopic Properties of Uracil. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7189–7197.

(42) Dunning Jr, T. H.; Peterson, K. A.; Wilson, A. K. Gaussian Basis Sets for Use in Correlated Molecular Calculations. X. The Atoms Aluminum Through Argon Revisited. *J. Chem. Phys.* **2001**, *114*, 9244–9253.

(43) Dunning Jr, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(44) Woon, D. E.; Dunning Jr, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(45) Piccardo, M.; Penocchio, E.; Puzzarini, C.; Biczysko, M.; Barone, V. Semi-experimental Equilibrium Structure Determinations by Employing B3LYP/SNSD Anharmonic Force Fields: Validation and Application to Semirigid Organic Molecules. *J. Phys. Chem. A* **2015**, *119*, 2058–2082.

(46) Ceselin, G.; Barone, V.; Tasinato, N. Accurate Biomolecular Structures by the Nano-LEGO Approach: Pick the Bricks and Build Your Geometry. *J. Chem. Theory Comput.* **2021**, *17*, 7290–7311.

(47) Spaniol, J.-T.; Lee, K. L. K.; Pirali, O.; Puzzarini, C.; Martin-Drumel, M.-A. A Rotational Investigation of the Three Isomeric Forms of Cyanoethynylbenzene (HCC-C_6H_4-CN): Benchmarking Experiments and Calculations using the "Lego brick" Approach. *Phys. Chem. Chem. Phys.* **2023**, *25*, 6397–6405.

(48) Puzzarini, C.; Alessandrini, S.; Bizzocchi, L.; Melosso, M. Hunting for Interstellar Molecules: Rotational Spectra of Reactive Species. *Faraday Discuss.* **2023**, *245*, 309–326.

(49) Santra, G.; Sylvetsky, N.; Martin, J. M. Minimally Empirical Double-hybrid Functionals Trained against the GMTKN55 Database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4. *J. Phys. Chem. A* **2019**, *123*, 5129–5143.

(50) Papajak, E.; Truhlar, D. G. Convergent Partially Augmented Basis Sets for Post-Hartree-Fock Calculations of Molecular Properties and Reaction Barrier Heights. *J. Chem. Theory Comput.* **2011**, *7*, 10–18.

(51) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 291–309.

(52) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Chem. Theory Comput.* **2011**, *32*, 1456–1465.

(53) Barone, V.; Lazzari, F. Hunting for Complex Organic Molecules in the Interstellar Medium: the Role of Accurate Low-Cost Theoretical Geometries and Rotational Constants. *J. Phys. Chem. A* **2023**, *127*, 10517–10527.

(54) Barone, V. Quantum Chemistry Meets High-Resolution Spectroscopy for Characterizing the Molecular Bricks of Life in the Gas-Phase. *Phys. Chem. Chem. Phys.* **2024**, *26*, 5802–5821.

(55) Uribe, L.; Di Grande, S.; Crisci, L.; Lazzari, F.; Mendolicchio, M.; Barone, V. Accurate Structures and Rotational Constants of Steroid Hormones at DFT Cost: Androsterone, Testosterone, Estrone, β-Estradiol, and Estriol. *J. Phys. Chem. A* **2024**, *128*, 2629–2642.

(56) Mills, I. M. In *Molecular Spectroscopy: Modern Research*; Rao, K. N., Mathews, C. W., Eds.; Academic: New York, 1972; Chapter Vibration-Rotation Structure in Asymmetric- and Symmetric-Top Molecules, p 115.

(57) Barone, V. Anharmonic Vibrational Properties by a Fully Automated Second-order Perturbative Approach. *J. Chem. Phys.* **2005**, *122*, 14108–14108.

(58) Carnimeo, I.; Biczysko, M.; Bloino, J.; Barone, V. Reliable Structural, Thermodynamic, and Spectroscopic Properties of Organic Molecules Adsorbed on Silicon Surfaces from Computational Modeling: the case of Glycine@Si(100). *Phys. Chem. Chem. Phys.* **2011**, *13*, 16713–16727.

(59) Barone, V.; Biczysko, M.; Bloino, J.; Borkowska-Panek, M.; Carnimeo, I.; Panek, P. Toward Anharmonic Computations of Vibrational Spectra for Large Molecular Systems. *Int. J. Quantum Chem.* **2012**, *112*, 2185–2200.

(60) Barone, V.; Biczysko, M.; Bloino, J. Fully Anharmonic IR and Raman Spectra of Medium-size Molecular Systems: Accuracy and Interpretation. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759–1787.

(61) Biczysko, M.; Bloino, J.; Puzzarini, C. Computational Challenges in Astrochemistry. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1349.

(62) Jensen, F. Unifying General and Segmented Contracted Basis sets. Segmented Polarization Consistent Basis Sets. *J. Chem. Theory Comput.* **2014**, *10*, 1074–1085.

(63) Wilson, A. K.; Van Mourik, T.; Dunning Jr, T. H. Gaussian Basis Sets for use in Correlated Molecular Calculations. VI. Sextuple Zeta Correlation Consistent Basis Sets for Boron through Neon. *J. Mol. Struct. THEOCHEM* **1996**, *388*, 339–349.

(64) Peterson, K. A.; Adler, T. B.; Werner, H.-J. Systematically Convergent Basis Sets for Explicitly Correlated Wavefunctions: The atoms H, He, B–Ne, and Al–Ar. *J. Chem. Phys.* **2008**, *128*, 084102.

(65) Kruse, H.; Szabla, R.; Šponer, J. Surprisingly Broad Applicability of the cc-pVnZ-F12 Basis Set for Ground and Excited States. *J. Chem. Phys.* **2020**, *152*, 214104.

(66) Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027–3034.

(67) Xu, R.; Jiang, Z.; Yang, Q.; Bloino, J.; Biczysko, M. Harmonic and Anharmonic Vibrational Computations for Biomolecular Building Blocks: Benchmarking DFT and

Basis Sets by Theoretical and Experimental IR Spectrum of Glycine Conformers. *J. Comp. Chem.* **2024**, *45*, 1846–1869.

(68) Perdew, J. P.; Schmidt, K. Jacob's Ladder of Density Functional Approximations for the Exchange-correlation Energy. AIP Conf. Proc. 2001; pp 1–20.

(69) Grimme, S. Density Functional Theory with London Dispersion Corrections. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211–228.

(70) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(71) Austin, A.; Petersson, G. A.; Frisch, M. J.; Dobek, F. J.; Scalmani, G.; Throssell, K. A Density Functional with Spherical Atom Dispersion Terms. *J. Chem. Theory and Comput.* **2012**, *8*, 4989–5007.

(72) Lee, C.; Yang, W.; Parr, R. Density-functional Exchange-energy Approximation with Correct Asymptotic Behaviour. *Phys. Rev. B* **1988**, *37*, 785.

(73) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio Calculation of Vibrational Absorption and Circular Dichroism Spectra using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(74) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation. *Phys. Rev. B* **1992**, *46*, 6671.

(75) Zhao, Y.; Truhlar, D. G. The M06 suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-class Functionals and 12 Other Functionals. *Theo. Chem. Acc.* **2008**, *120*, 215–241.

(76) Grimme, S. Semiempirical Hybrid Density Functional with Perturbative Second-Order Correlation. *J. Chem. Phys.* **2006**, *124*, 034108.

(77) Haoyu, S. Y.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn–Sham Global-hybrid Exchange–correlation Density Functional with Broad Accuracy for Multi-reference and Single-reference Systems and Noncovalent Interactions. *Chem. Sci.* **2016**, *7*, 5032–5051.

(78) Kozuch, S.; Martin, J. M. Spin-component-scaled Double Hybrids: an Extensive Search for the Best Fifth-rung Functionals Blending DFT and Perturbation Theory. *J. Comput. Chem.* **2013**, *34*, 2327–2344.

(79) Kozuch, S.; Martin, J. M. DSD-PBEP86: in Search of the Best Double-hybrid DFT with Spin-component Scaled MP2 and Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20104–20107.

(80) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange–correlation Functional Using the Coulomb-attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.

(81) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(82) Brémond, E.; Adamo, C. *J. Chem. Phys.* **2011**, *135*, 024106.

(83) Vydrov, O. A.; Scuseria, G. E. Assessment of a Long-range Corrected Hybrid Functional. *J. Chem. Phys.* **2006**, *125*, 234109.

(84) Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. Tests of Functionals for Systems with Fractional Electron Number. *J. Chem. Phys.* **2007**, *126*, 154109.

(85) Vydrov, O. A.; Heyd, J.; Krukau, A. V.; Scuseria, G. E. Importance of Short-range Versus Long-range Hartree-Fock E Exchange for the Performance of Hybrid Density Functionals. *J. Chem. Phys.* **2006**, *125*, 074106.

(86) Brémond, É.; Sancho-García, J. C.; Pérez-Jiménez, Á. J.; Adamo, C. Communication: Double-hybrid Functionals from Adiabatic-connection: The QIDH Model. *J. Chem. Phys.* **2014**, *141*, 031101.

(87) Schmider, H. L.; Becke, A. D. Optimized Density Functionals from the Extended G2 Test Set. *J. Chem. Phys.* **1998**, *108*, 9624–9631.

(88) Chai, J.-D.; Head-Gordon, M. Long-range Corrected Hybrid Density Functionals with Damped Atom–atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(89) Hartwig, H.; Kretschmer, U.; Dreizler, H. The ^{33}S Nuclear Quadrupole Hyperfine Structure in the Rotational Spectrum of ^{32}S, ^{33}S Dimethyl Disulfide. *Z. Naturforsch. A* **1995**, *50*, 131–136.

(90) Demaison, J.; Margules, L.; Rudolph, H. D. Accurate Determination of an Equilibrium Structure in the Presence of a Small Coordinate: The Case of Dimethylsulfide. *J. Mol. Struct.* **2010**, *978*, 229–233.

(91) Vogt, N.; Demaison, J.; Vogt, J.; Rudolph, H. D. Why it is Sometimes Difficult to Determine the Accurate Position of a Hydrogen Atom by the Semiexperimental Method: Structure of Molecules Containing the OH or the CH3 Group. *J. Comput. Chem.* **2014**, *35*, 2333–2342.

(92) Costain, C. C. Determination of Molecular Structures from Ground State Rotational Constants. *J. Chem. Phys.* **1958**, *29*, 864–874.

(93) Kraitchman, J. Determination of Molecular Structure from Microwave Spectroscopic Data. *Am. J. Phys.* **1953**, *21*, 17–24.

(94) Watson, J. K.; Roytburg, A.; Ulrich, W. Least-Squares Mass-Dependence Molecular Structures. *J. Mol. Spectrosc.* **1999**, *196*, 102–119.

(95) Cazzoli, G.; Cludi, L.; Contento, M.; Puzzarini, C. Rotational Spectrum of $CH_3C^{13}CCCH$: Determination of the Equilibrium Structure of Methyldiacetylene from Microwave Spectroscopy and Ab Initio Calculations. *J. Mol. Spectrosc.* **2008**, *251*, 229–234.

(96) Guarnieri, A.; Demaison, J.; Rudolph, H. D. Structure of Ketene – Revisited r_e (Equilibrium) and r_m (Mass-Dependent) Structures. *J. Mol. Struct.* **2010**, *969*, 1–8.

(97) Demaison, J.; Császár, A. G.; Margulès, L. D.; Rudolph, H. D. Equilibrium Structures of Heterocyclic Molecules with Large Principal Axis Rotations upon Isotopic Substitution. *J. Phys. Chem. A* **2011**, *115*, 14078–14091.

(98) Demaison, J., Boggs, J. E., Császár, A. G., Eds. The Method of Least Squares. In *Equilibrium Molecular Structures: From Spectroscopy to Quantum Chemistry*; CRC Press, Taylor & Francis Group: Boca Raton, FL, US, 2011.

(99) Barone, V.; Ceselin, G.; Fuse, M.; Tasinato, N. Accuracy Meets Interpretability for Computational Spectroscopy by Means of Hybrid and Double-Hybrid Functionals. *Front. Chem* **2020**, *8*, 584203.

(100) Shu, C.; Jiang, Z.; Biczysko, M. Toward Accurate Prediction of Amino Acid Derivatives Structure and Energetics from DFT: Glycine Conformers and their Interconversions. *J. Mol. Model.* **2020**, *26*, 1–13.

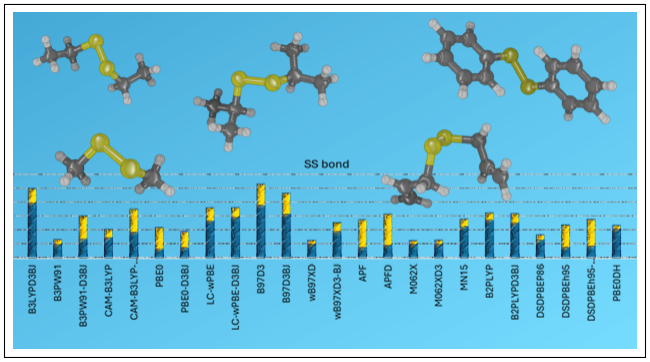
(101) Sheng, M.; Silvestrini, F.; Biczysko, M.; Puzzarini, C. Structural and Vibrational Properties of Amino Acids from Composite Schemes and Double-Hybrid DFT: Hydrogen Bonding in Serine as a Test Case. *J. Phys. Chem. A* **2021**, *125*, 9099–9114.

(102) Wang, P.; Shu, C.; Ye, H.; Biczysko, M. Structural and Energetic Properties of Amino Acids and Peptides Benchmarked by Accurate Theoretical and Experimental Data. *J. Phys. Chem. A* **2021**, *125*, 9826–9837.

(103) Piccardo, M.; Penocchio, E.; Puzzarini, C.; Biczysko, M.; Barone, V. Semi-Experimental Equilibrium Structure Determinations by Employing B3LYP/SNSD Anharmonic Force Fields: Validation and Application to Semirigid Organic Molecules. *J. Phys. Chem. A* **2015**, *119*, 2058–2082.

(104) Brémond, É.; Savarese, M.; Su, N. Q.; Pérez-Jiménez, Á. J.; Xu, X.; Sancho-García, J. C.; Adamo, C. Benchmarking Density Functionals on Structural Parameters of Small-/medium-sized Organic Molecules. *J. Chem. Theory. Comput.* **2016**, *12*, 459–465.

TOC Graphic


SS bond
B3LYPD3BJ
B3PW91
B3PW91-D3BJ
CAM-B3LYP
CAM-B3LYP-...
PBE0
PBE0-D3BJ
LC-wPBE
LC-wPBE-D3BJ
B97D3
B97D3BJ
wB97XD
wB97XD3-BJ
APF
APFD
M062X
M062XD3
MN15
B2PLYP
B2PLYPD3BJ
DSDPBEP86
DSDPBEh95
DSDPBEh95-...
PBE0DH