Chemistry Europe
European Chemical Societies Publishing

WILEY VCH

# Excellence in Chemistry Research


## Announcing our new flagship journal

- Gold Open Access
- Publishing charges waived
- Preprints welcome
- Edited by active scientists

## Meet the Editors of *ChemistryEurope*

**Luisa De Cola**
Università degli Studi di Milano Statale, Italy

**Ive Hermans**
University of Wisconsin-Madison, USA

**Ken Tanaka**
Tokyo Institute of Technology, Japan

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202203990
Chemistry Europe
European Chemical Societies Publishing
www.chemeurj.org

# Bringing Machine-Learning Enhanced Quantum Chemistry and Microwave Spectroscopy to Conformational Landscape Exploration: the Paradigmatic Case of 4-Fluoro-Threonine

V. Barone,*[a] M. Fusè,[a] R. Aguado,[b] S. Potenti,[a, c] I. León,[b] E. R. Alonso,[b] S. Mata,[b] F. Lazzari,[a] G. Mancini,[a] L. Spada,[a] A. Gualandi,[c] P. G. Cozzi,[c] C. Puzzarini,*[c] and J. L. Alonso*[b]

**Abstract:** A combined experimental and theoretical study has been carried out on 4-fluoro-threonine, the only naturally occurring fluorinated amino acid. Fluorination of the methyl group significantly increases the conformational complexity with respect to the parent amino acid threonine. The conformational landscape has been characterized in great detail, with special attention given to the inter-conversion pathways between different conformers. This led to the identification of 13 stable low-energy minima. The equilibrium population of so many conformers produces a very complicated and congested rotational spectrum that could be assigned through a strategy that combines several levels of quantum chemical calculations with the principles of machine learning. Twelve conformers out of 13 could be experimentally characterized. The results obtained from the analysis of the intra-molecular interactions can be exploited to accurately model fluorine-substitution effects in biomolecules.

## Introduction

Fluorinated organic molecules play an important role in medicinal chemistry.[1,2] Before the Covid-19 pandemic, the most prescribed drug in the world was atorvastatin (Lipitor), which features a fluorinated phenyl substituent. Moreover, nearly half of the new small molecule drugs approved by the US Food and Drug Administration in 2018 contained a fluorine atom.[3] Replacement of a hydrogen by the similarly sized fluorine atom tends to have a negligible impact on steric properties and molecular shape, while the vastly different electronic properties of fluorine can have profound effects on binding energy, reactivity, and relative energy of corresponding conformers on the potential energy landscape.[1,4] For the structural and physical chemist, conformers of such fluorinated compounds tend to have geometries qualitatively similar to those of the parent compounds, although the introduction of the fluorine atom can lead to a breaking of equivalence of positions in the parent compound, which leads to a more complicated conformational energy landscape, further exacerbated by additional interactions. For the medicinal or biological chemist, this additional conformational space can provide a means for fine-tuning the mechanism of drug action.[2,5]

Fluorinated $\alpha$-amino-acids have been attracting increasing attention because their introduction in specific domains of proteins can have a profound effect on their stability, folding and biological activity.[6] In this framework, 4-fluoro-threonine (4FT) is the only fluorinated amino acid found in nature[7] and one of the only three fluorometabolites of microbial origin discovered so far.[8] This molecule is obtained from fluorination of the terminal methyl group in the essential amino acid threonine (Thr), a molecule which has already a rich conformational energy landscape.[9] The development of an effective synthesis for producing significant amounts of 4FT[10] opens the way for the first unequivocal characterization of the conformational landscape of a fluorinated aminoacid in the gas phase, i.e., without any perturbation from the environment. In this report, we apply high-resolution gas-phase rotational spectroscopy in conjunction with the tools of modern quantum chemistry and machine learning to explore the potential energy landscape of 4FT. Such an endeavor is extremely challenging. Although the isolated molecule in the gas phase is the appropriate starting point for investigations of this sort, the richness of the potential energy landscape – which is expected to feature up to three times as many minima as the already

[a] Prof. Dr. V. Barone, Dr. M. Fusè, Dr. S. Potenti, F. Lazzari, Dr. G. Mancini, Dr. L. Spada
SMART Laboratory, Scuola Normale Superiore di Pisa, piazza dei Cavalieri 7, 56126 Pisa, Italy
E-mail: vincenzo.barone@sns.it

[b] R. Aguado, Dr. I. León, E. R. Alonso, S. Mata, Prof. Dr. J. L. Alonso
Grupo de Espectroscopia Molecular (GEM), Edificio Quifima, Laboratorios de Espectroscopia y Bioespectroscopia Parque Cientifico UVa, Universidad de Valladolid, 47005 Valladolid, Spain
E-mail: jlalonso@qf.uva.es

[c] Dr. S. Potenti, Dr. A. Gualandi, Prof. Dr. P. G. Cozzi, Prof. Dr. C. Puzzarini
Dipartimento di "Chimica Giacomo Ciamician", University of Bologna, via F. Selmi 2, 40126, Bologna, Italy
E-mail: cristina.puzzarini@unibo.it

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202203990

complicated Thr molecule due to the fluorination of the methyl group – and the size of the molecule significantly complicate the standard methods used in such an investigation. Specifically, the large number of conformers present in an equilibrium mixture leads to significant spectral congestion in the gas-phase rotational spectrum that effectively prevents disentangling this spectrum to reveal the wealth of structural information for each conformer that might otherwise be provided by this technique. Furthermore, brute-force computations with the methods of quantum chemistry are severely compromised here by both the size of the molecule and the significant number of conformers that one expects to find in the equilibrium mixture. On the other hand, predictions (properties and energetics) for rotational spectroscopy require an accuracy that widely-used quantum chemical methodologies (e.g. hybrid density functionals) are not always able to meet.[11] Such an accuracy is also reflected in the correct and exhaustive exploration of the conformational potential energy surface (PES). Specifically, hybrid density functionals and low-order perturbative methods (e.g., MP2[12]) can lead to incorrect and incomplete characterization.[11]

To address the issues mentioned in the preceding paragraph, we have developed a strategy employing modern methods of machine learning to augment quantum chemistry and molecular spectroscopy in solving challenges of this type. The proposed strategy involves a preliminary exploration of the entire conformational PES using a fast semi-empirical method.[13] This step is guided by knowledge-based and evolutionary algorithms with the aim of finding all low-lying energy minima.[14] The results are next refined by last-generation density functionals,[15,16] which are also used for analyzing relaxation paths between pairs of adjacent energy minima[11] and computing thermodynamic functions beyond the rigid rotor – harmonic oscillator model.[17–22] Best estimates of relative electronic energies are finally obtained by means of state-of-the-art wave-function (post-Hartree-Fock) approaches.[23–25] These results are then exploited by accurate high-resolution molecular spectroscopy.

The 4FT molecule is an ideal playground for testing the strategy above, and the success achieved here suggests that the combination of these techniques (i.e. machine learning, quantum chemistry and molecular spectroscopy) can play an important role in studying the conformational energy landscape as well as geometric and electronic structures of the "small molecules" that are playing an increasingly important role in modern pharmacology.

## Results and Discussion

The conformational landscape of isolated amino acids is governed by both backbone ($\phi$ = H–N–$C^{\alpha}$–C′, $\psi$ = N–$C^{\alpha}$–C′–O, $\omega$ = $C^{\alpha}$–C′–O–H) and side-chain ($\chi$) torsional angles, as shown in Figure 1. The non-planarity of the $NH_2$ moiety suggests to replace the customary $\phi$ dihedral angle by $\phi' = \text{LP–N–}C^{\alpha}\text{–C}' = \phi + 120°$, LP being the nitrogen lone-pair.

The most stable conformers result from the formation of hydrogen bonds, which can be classified according to the

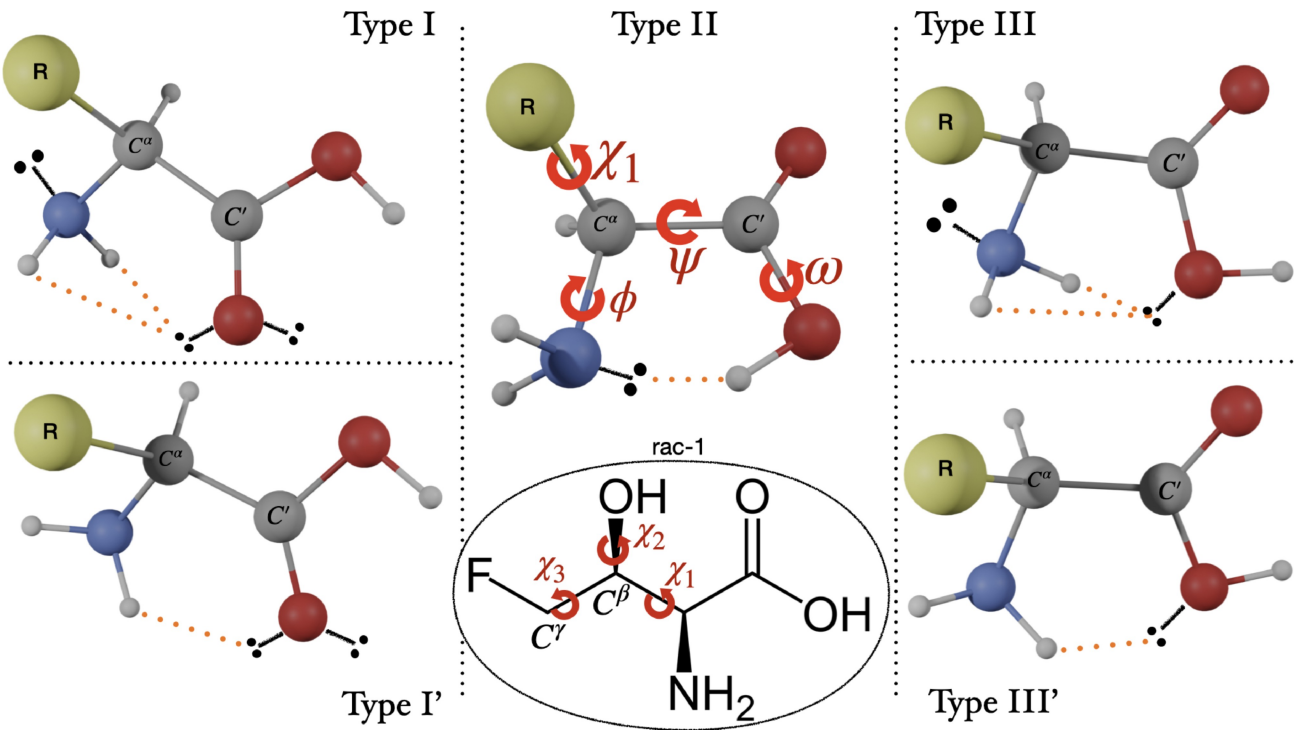


**Figure 1.** Structures and main dihedral angles of a generic amino acid, with R denoting the side-chain. The inset shows the synthesized racemic diastereoisomer (rac-1) of 4FT and its side-chain dihedral angles. See main text for further details.

values of the backbone dihedral angles defined above. Type I is a bifurcated $NH_2$···O=C hydrogen bond having $\phi' \approx 180°$, $\psi \approx 180°$, and $\omega \approx 180°$. Type II is a N···HO hydrogen bond with $\phi' \approx 0°$, $\psi \approx 0°$, and $\omega \approx 0°$. Type III is a bifurcated $NH_2$···OH hydrogen bond with $\phi' \approx 180°$, $\psi \approx 0°$, and $\omega \approx 180°$. The only conformers observed for amino acids with non-polar side chains are those involving hydrogen bonds between the amine and carboxyl moieties of the backbone, which can be either type I or type II. Additional low-energy conformers are observed, when polar side chains are present, because of stabilizing contributions from backbone-side chain hydrogen bonds. In particular, starting from type I structures, rotation of the $NH_2$ moiety by about 90° allows its involvement in two different H-bonds leading to I′ conformers, with $\phi' \approx 90°$, $\psi \approx 180°$, and $\omega \approx 180°$. Noted is that hydrogen bonds of type I and I′ can be formed for either $\omega \approx 180°$ or $\omega \approx 0°$ structures, with the former situation being always preferred by the carboxylic moiety. Conformers of type III (bifurcated) or type III′ (single, $\phi' \approx 180°$, $\psi \approx 90°$, $\omega \approx 180°$) have also been observed in some cases, but they are always the less populated.[26]

For an exhaustive labeling of conformation, the value of each $\chi$ dihedral angle should also be provided. The conventional labels $g^-$, g and t are used to indicate staggered conformations ($\chi \approx -60°$, 60° and 180°, respectively) in place of the **a, b, c** labels employed for $\chi_1$ in previous works.[9] In the case of 4FT, the side-chain dihedral angles are $\chi_1$ = N–$C^\alpha$–$C^\beta$–O, $\chi_2$ = $C^\alpha$–$C^\beta$–O–H and $\chi_3$ = $C^\alpha$–$C^\beta$–$C^\gamma$–F, as shown in the inset of Figure 1. The structure of 4FT is similar to that of Thr, but, as already mentioned, the replacement of a hydrogen atom in the terminal methyl group by fluorine increases the challenge because of a larger number of low-energy conformers due to the loss of equivalence of the $C^\gamma$ substituents (see the inset of Figure 1) and to additional interactions between F and aminic or hydroxylic hydrogens of the backbone.

## Conformational landscape

The identification of the most stable conformers of 4FT, potentially observable in rotational spectroscopy experiments, can start from the exhaustive characterization available for Thr, which provided seven low-energy conformers.[9] Fluorination of its terminal methyl group leads to 21 starting structures for 4FT. However, the empirical procedure based on 'fluorination' of low-energy Thr conformers might lead to a limited sampling of the 4FT conformational PES. Therefore, we resorted to a purposely tailored version of the Island Model Evolutionary Algorithm (IMEA[14,27]), which has proven to be an extremely effective machine-learning algorithm for the structural characterization of several flexible molecules.[10,27]

The exploration step employed the GFN2-xTB semi empirical method[13] and led to, in addition to the 21 structures deriving from fluorination of Thr, several additional conformers. This panel of structures, together with all missing staggered conformers with respect to the $\chi_3$ dihedral for each candidate (see Table S2 and Figure S1 of the Supplementary Information, SI), was selected for further refinement. This step consisted in full geometry optimizations using hybrid (B3LYP-D3BJ/jun-cc-pVDZ,[28–30] hereafter B3) and, subsequently, double-hybrid (revDSD-PBEP86-D3BJ/jun-cc-pvTZ,[31–33] hereafter rDSD) density functionals. Selection of conformers to be retained was based on filtering with increasingly narrower energy thresholds and removal of those conformers relaxing to more stable structures through paths governed by low energy barriers. Based on previous studies,[11,34] the final thresholds were 900 $cm^{-1}$ for relative stabilities (corresponding to populations of about 1% at room temperature) and 400 $cm^{-1}$ for inter-conversion barriers.[35–37] This step reduced the set of structures to the nine conformers deriving from fluorination of Thr shown in Figure 2 and the four additional conformers shown in Figure 3. All of them lie within 750 $cm^{-1}$ with respect to the IIggg$^-$ conformer (structure 3 in Figure 2), which corresponds to the most stable structure of Thr, IIgg (see Figure 2 and Table S2 of the SI). The choice of selecting IIggg$^-$ as reference structure was made because the fluorine atom is not involved in any hydrogen bond. This allows for disentangling the effects of fluorine. Indeed, the interactions between the fluorine atom and the OH group of the side chain explain why the four conformers of Figure 3 were found only in 4FT and not in Thr. Noted is that all low-lying structures detected for Thr have at least one equivalent in 4FT, with each of the IIgg, I′g$^-$g and I′gg$^-$ conformers of Thr generating a pair of observable conformers in 4FT, which correspond to the two possible staggered (g or $g^-$) orientations of $\chi_3$ (see Figure 2).

Since the relative stability of low-lying conformers provides crucial information on their experimental detection and assignment, we have refined relative electronic energies by accurate single-point energy computations, on top of rDSD geometries, using the so-called jun-Cheap-F12 composite scheme (junChS–F12).[25,38] This is detailed in the Computational section and, for relative conformational energies, has an expected accuracy better than 10 $cm^{-1}$.[24,25] The rDSD model provides relative stabilities in good agreement with the junChS–F12 references (the average and maximum absolute error being 26.8 and 58.3 $cm^{-1}$, respectively) without any significant alteration of the conformer stability order (see Figure 4 and Tables S2 and S3 of the SI). Significantly worse results are instead obtained using the B3 and MP2 computational models (MP2 stands for second order Møller-Plesset perturbation theory[12]), which are widely used in the conformational analysis of biomolecule building blocks (see Figure 4). The worst description is provided by the B3 level, which shows absolute average and maximum error of 203 and 390 $cm^{-1}$, respectively, together with a general large over-stabilization of all conformers of type II (see Table S3 of the SI for details).

To determine the relative stability of the 4FT conformers, one has to move from electronic energy differences to the corresponding relative enthalpies at 0 K ($\Delta H_0^\circ$) or free energies at room temperature ($\Delta G^\circ$) depending on the experimental conditions. To this end, harmonic zero-point energies (ZPEs) and vibrational partition functions (see Table S4 of the SI) have been further improved by incorporating in ZPEs anharmonic contributions, obtained with second-order vibrational perturbation theory (VPT2),[39] and using the so-called quasi-harmonic

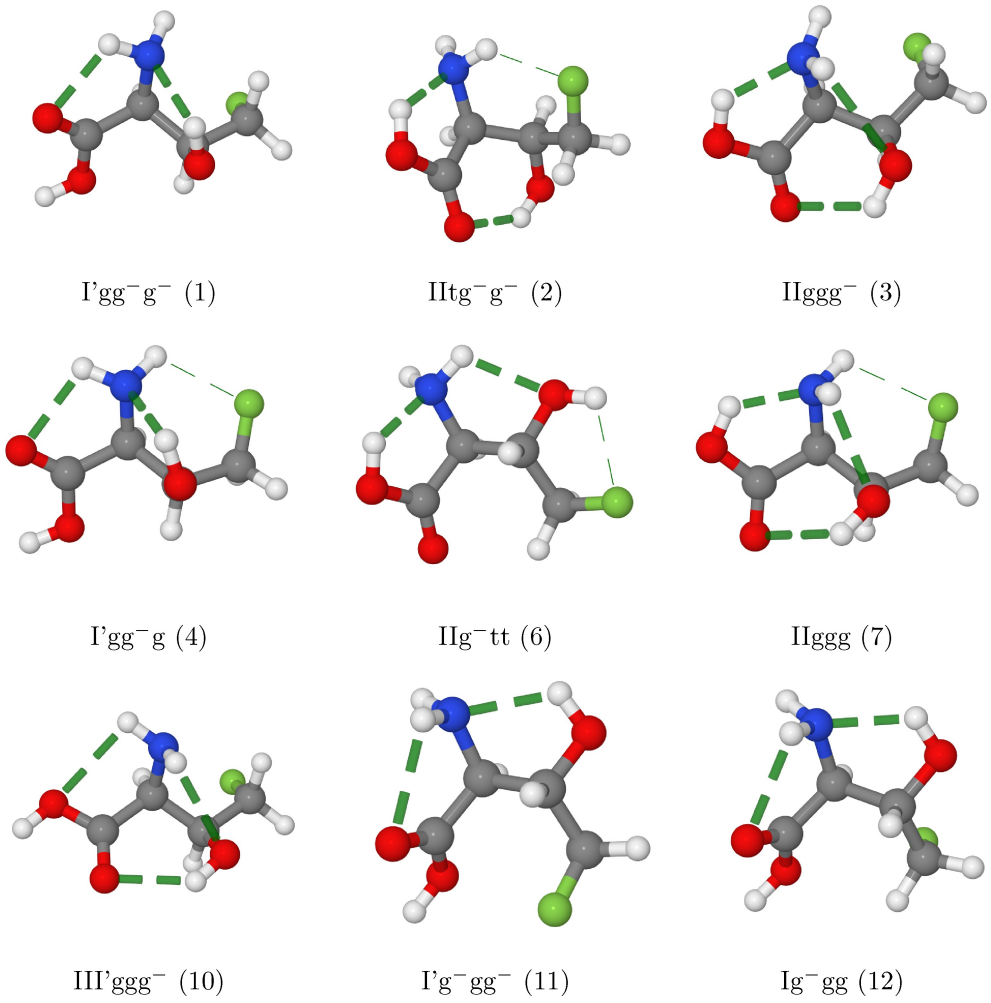


**Figure 2.** Structures and intra-molecular hydrogen-bonds of the nine low-energy conformers of 4FT derived from the observed threonine conformers. Hydrogen bonds involving N or O atoms are made evident by thick broken lines, whereas those involving F atoms by thin broken lines. The numbering of conformers based on relative enthalpies at 0 K is given in parentheses (see main text for details).

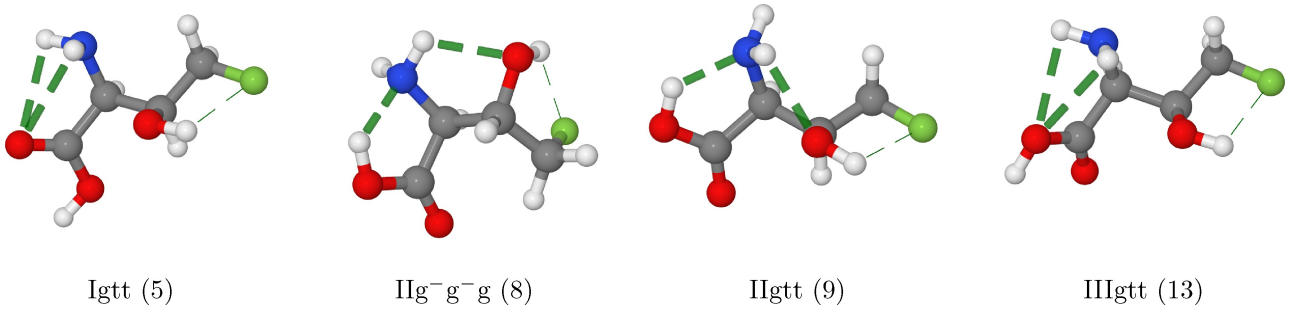


**Figure 3.** Structures and intra-molecular hydrogen-bonds of the four additional low-energy conformers of 4FT not deriving from any observed threonine conformer. For visualization of hydrogen bonds, the convention of Figure 2 applies.

(QH) approximation for entropies[21,40] (see Table S5 of the SI). Further details are provided in the Computational section.

Relative enthalpies and free energies lead to significant differences in the stability order predicted by electronic energies (see Tables 1 and 2), with a general destabilization of those conformers characterized by hydrogen bonds of type II. In any case, the number and type of conformers potentially detectable in the gas-phase rotational spectra of 4FT remain unchanged (i.e. the 13 conformers of Figures 2 and 3). However, the least stable conformer (IIIgtt) for both enthalpy and free energy is hardly observable because of a combination of low population (due to high energy: ~600–700 $cm^{-1}$ above the

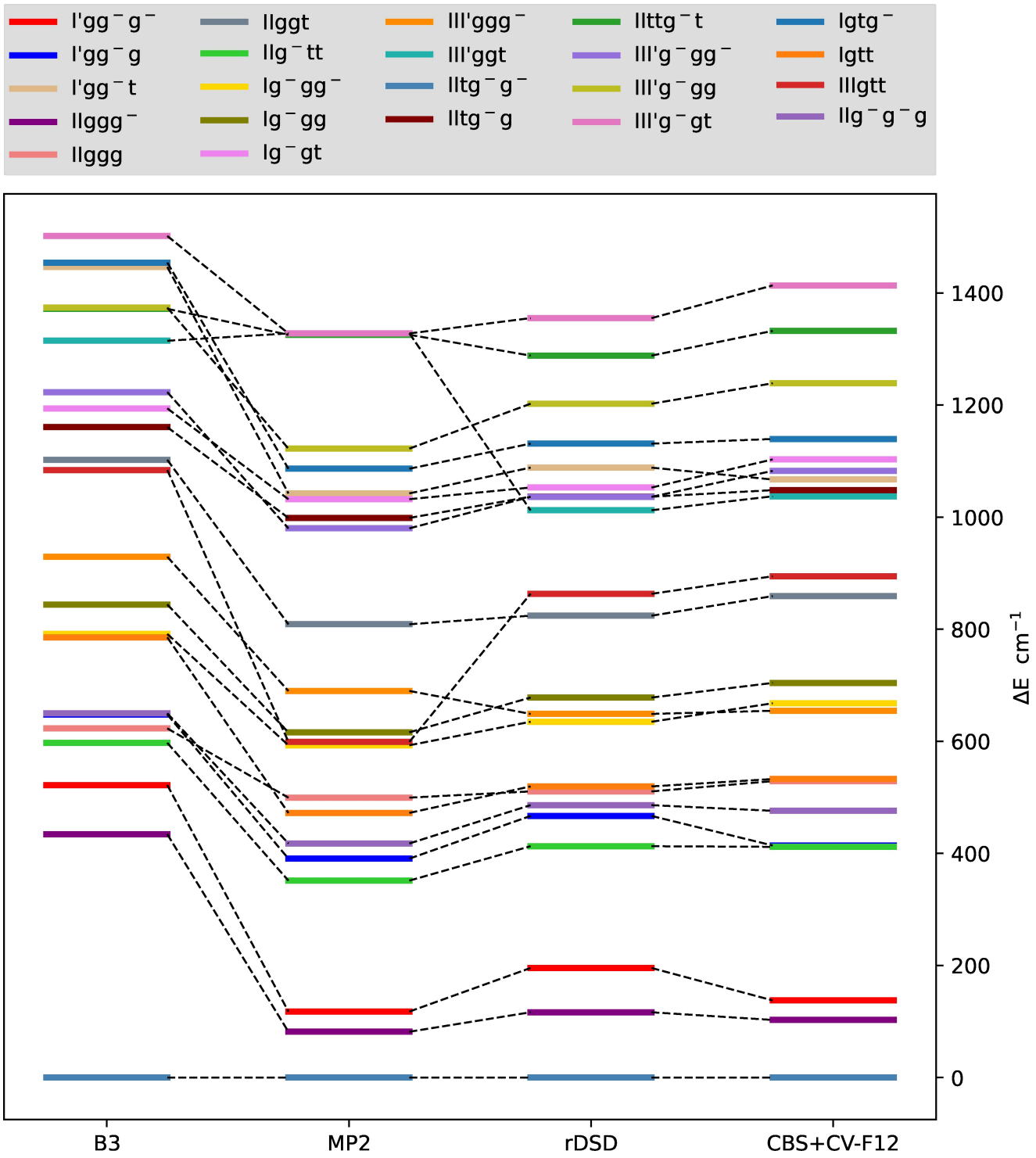


**Figure 4.** Comparison of the relative electronic energies for the low-lying 4FT conformers at the B3, MP2/jun-cc-pVTZ, rDSD and junChS–F12 levels of theory.

most stable structure) and magnitude of the dipole moment components (which are responsible for the intensity of rotational transitions). Several of the most stable conformers are stabilized by type II (N···HO) hydrogen bonds within the backbone, whereas the most stable structure in absolute terms (for both enthalpy and free energy) shows a type I' (NH···O=C) hydrogen bond within the backbone assisted by a second hydrogen bond between the other aminic hydrogen and the side-chain OH moiety. Interestingly, this type of structure is more stable than its I counterpart (bifurcated $NH_2$···O=C hydrogen bond within the backbone), which is instead usually observed in amino acids with 'innocent' side chains. As already

**Table 1.** Ground-state rotational constants ($A_0$, $B_0$, and $C_0$ in MHz), 14 N-nuclear quadrupole coupling constants ($\chi_{ii}$ in MHz) and electric dipole moment components ($\mu_i$ in debye) of the six most stable energy minima of 4FT (in order of increasing enthalpy at 0 K). Relative electronic energies ($\triangle E_{el}$), enthalpies at 0 K ($\triangle H_{0^\circ}$) and room-temperature free energies ($\triangle G^0$), in $cm^{-1}$, are also reported.

| Calc.[a] | I′gg⁻g⁻ | IItg⁻g⁻ | IIggg⁻ | I′gg⁻g | Igtt | IIg⁻tt |
|---|---|---|---|---|---|---|
| $A_0$[c] | 2488.7 | 1918.2 | 2599.5 | 2816.3 | 2951.3 | 2361.6 |
| $B_0$[c] | 1028.7 | 1441.0 | 1069.5 | 1041.1 | 957.1 | 1137.7 |
| $C_0$[c] | 990.1 | 1127.3 | 941.2 | 946.7 | 860.4 | 845.5 |
| $\chi_{aa}$ | 0.466 | −3.277 | −3.032 | 1.278 | −2.515 | −0.265 |
| $\chi_{bb}$ | 1.183 | 2.075 | 0.782 | 2.720 | 1.638 | 2.347 |
| $\chi_{cc}$ | −1.650 | 1.202 | 2.250 | −3.998 | 0.877 | −2.083 |
| $\mu_a$ | −0.160 | 1.324 | −2.634 | −0.629 | 0.361 | −0.727 |
| $\mu_b$ | 0.234 | 2.344 | −1.576 | −1.147 | 0.812 | 3.867 |
| $\mu_c$ | 1.318 | −2.531 | 1.778 | −2.806 | −1.560 | 0.729 |
| $\Delta E$ | 0.0 | −137.7 | −34.9 | 276.4 | 395.0 | 274.0 |
| $\Delta H^\circ_0$ | 0.0 | 61.8 | 68.7 | 285.7 | 307.7 | 347.8 |
| $\Delta G^\circ$ | 0.0 | 285.9 | 127.3 | 366.0 | 215.2 | 437.8 |
| Exp.[b] | 1 | 2 | 3 | 4 | 5 | 6 |
| $A_0$ | 2486.8991(14) | 1906.1362(26) | 2597.9521(34) | 2822.2100(49) | 2966.8253(72) | 2372.1594(36) |
| $B_0$ | 1034.0622(12) | 1433.9976(23) | 1065.7956(13) | 1045.5245(27) | 959.1415(92) | 1139.8048(27) |
| $C_0$ | 993.0311(6) | 1120.3890(27) | 950.1566(9) | 948.9332(68) | 861.4950(46) | 846.0938(14) |
| $\chi_{aa}$ | 0.244(37) | −3.052(36) | −2.70(12) | 1.1541(67) | −2.545(50) | −0.143(13) |
| $\chi_{bb}$ | 1.480(32) | 2.018(57) | 0.54(13) | 2.865(40) | 1.784(64) | 2.237(74) |
| $\chi_{cc}$ | −1.724(32) | 1.034(57) | 2.17(13) | −4.019(40) | 0.760(64) | −2.229(74) |
| $\mu_a$ | Not Obs | Obs | Obs | Not Obs | Not Obs | Not Obs |
| $\mu_b$ | Not Obs | Obs | Obs | Obs | Not Obs | Obs |
| $\mu_c$ | Obs | Obs | Obs | Obs | Obs | Not Obs |
| N[c] | 19 | 43 | 42 | 25 | 19 | 39 |
| RMS/kHz | 29.3 | 30.8 | 31.2 | 40.4 | 31.6 | 22.7 |

[a] Computed data are at the rDSD level (including LRA corrections for equilibrium rotational constants) except for electronic energies (junChS–F12) and vibrational corrections to equilibrium rotational constants (B3). [b] Standard errors are shown in parentheses in units of the last digits. [c] Number of measured transitions or hyperfine components.

**Table 2.** Ground-state rotational constants ($A_0$, $B_0$, and $C_0$ in MHz), 14 N-nuclear quadrupole coupling constants ($\chi_{ii}$ in MHz) and electric dipole moment components ($\mu_i$ in debye) for the seven low-energy minima of 4FT above 450 $cm^{-1}$ (in order of increasing enthalpy at 0 K with respect to the most stable structure). Relative electronic energies ($\triangle E_{el}$), enthalpies at 0 K ($\triangle H_{0^\circ}$) and room temperature free energies ($\triangle G^0$), in $cm^{-1}$, are also reported.

| Calc.[a] | IIggg | IIg⁻g⁻g | IIgtt | III′ggg⁻ | Ig⁻gg⁻ | Ig⁻gg | IIIgtt |
|---|---|---|---|---|---|---|---|
| $A_0$[c] | 2886.7 | 2299.1 | 3029.5 | 2514.3 | 1839.6 | 2067.8 | 3008.5 |
| $B_0$[c] | 1069.0 | 1239.7 | 933.2 | 1049.2 | 1549.0 | 1357.0 | 942.0 |
| $C_0$[c] | 915.5 | 918.5 | 841.9 | 927.5 | 993.6 | 958.3 | 856.4 |
| $\chi_{aa}$ | −2.772 | −2.083 | −4.243 | −3.931 | −3.846 | −2.297 | −2.610 |
| $\chi_{bb}$ | 0.871 | 2.474 | 2.557 | 1.301 | 1.999 | 0.485 | 1.697 |
| $\chi_{cc}$ | 1.902 | −0.391 | 1.686 | 2.630 | 1.846 | 1.812 | 0.913 |
| $\mu_a$ | 2.617 | 1.634 | 2.876 | 2.694 | 1.036 | 2.452 | −1.395 |
| $\mu_b$ | 0.764 | 3.361 | −4.039 | 0.567 | 1.181 | 0.692 | −1.490 |
| $\mu_c$ | −1.075 | 0.655 | 1.860 | 0.353 | 1.726 | 0.263 | 0.345 |
| $\Delta E$ | 391.2 | 338.3 | 508.2 | 516.8 | 530.2 | 566.5 | 756.8 |
| $\Delta H^\circ_0$ | 452.8 | 466.0 | 466.4 | 531.3 | 532.0 | 568.7 | 689.8 |
| $\Delta G^\circ$ | 465.1 | 510.7 | 528.1 | 515.3 | 552.1 | 486.4 | 603.9 |
| Exp.[b] | 7 | 8 | 9 | 10 | 11 | 12 | |
| $A_0$ | 2888.19(30) | 2307.5121(47) | 3036.7178(70) | 2522.28(35) | 1846.7446(68) | 2086.445(16) | |
| $B_0$ | 1071.4955(40) | 1242.2409(38) | 935.2728(73) | 1051.4209(46) | 1551.6071(59) | 1355.8050(10) | |
| $C_0$ | 919.1366(38) | 920.7502(17) | 843.7029(33) | 931.8023(45) | 995.125(74) | 956.60256(93) | |
| $\chi_{aa}$ | Not resolved | −1.83(18) | −4.209(24) | Not resolved | −4.53(20) | −2.3217(86) | |
| $\chi_{bb}$ | Not resolved | 2.61(16) | 2.625(22) | Not resolved | 2.96(21) | 0.883(25) | |
| $\chi_{cc}$ | Not resolved | −0.78(16) | 1.584(22) | Not resolved | 1.57(21) | 1.439(25) | |
| $\mu_a$ | Obs | Obs | Obs | Obs | Not Obs | Obs | |
| $\mu_b$ | Not Obs | Obs | Obs | Not Obs | Not Obs | Not Obs | |
| $\mu_c$ | Not Obs | Not Obs | Obs | Not Obs | Obs | Not Obs | |
| N[c] | 14 | 21 | 29 | 10 | 16 | 10 | |
| RMS/kHz | 44.6 | 59.4 | 20.1 | 40.6 | 32.2 | 3.77 | |

[a] Computed data are at the rDSD level (including LRA corrections for equilibrium rotational constants) except for electronic energies (junChS–F12) and vibrational corrections to equilibrium rotational constants (B3). [b] Standard errors are shown in parentheses in units of the last digits. [c] Number of measured transitions or hyperfine components.

mentioned, although type III and III′ structures are also present, they are among the least populated forms.[9,26,34,41]

### Rotational Spectrum

Comparison with experiment requires, together with equilibrium geometries (straightforwardly providing equilibrium rotational constants), dipole moment components ($\mu_i$, *i* being an inertial axis) and nuclear quadrupole coupling constants ($\chi_{ii}$, for hyperfine structure; *vide infra*), which were both computed at the rDSD level, and anharmonic contributions (vibrational corrections to rotational constants), which were evaluated at the B3 level.[17,42–44] More accurate structures, which means improved equilibrium rotational constants, were obtained by correcting the rDSD geometrical parameters with the so-called linear regression approach (LRA; see the Computational section for details and Table S6 of the SI).[45,46]

To probe the *a priori* description of the conformational landscape of 4FT, a sufficient amount of the racemic diastereoisomer (rac-1; see the inset of Figure 1) was obtained following the recently published route (with further details being given in the SI).[10] Being 4FT a solid compound with a high melting point (182–183 °C),[47] the laser ablation (LA) technique[26] was then employed in combination with a chirped-pulse Fourier-transform Microwave spectrometer (CP-FTMW; details are provided in the SI)[48,49] to record the rotational spectrum of 4FT in the 6.0–14.0 GHz frequency region. In the spectrometer, the gas-phase sample is subjected to supersonic expansion. Depending on the effective vibrational cooling there occurring, the post-expansion abundance of the different conformers is governed by either $\Delta H_0^\circ$ (vibrational relaxation) or room-temperature $\Delta G^\circ$ (no vibrational relaxation). The data collected in Tables 1 and 2 show that this might have some impact on the relative stability of low-lying conformers, but – as already pointed out – does not alter the nature and number of the conformers possibly detectable in microwave (MW) experiments.

As apparent in Figure 5a, the CP-FTMW spectrum is extremely dense and congested, thus suggesting the presence of several species. In the first step of the spectral analysis, the lines due to known photo-fragmentation products and instrumental noise were removed from the spectrum. Most of the remaining features clearly resemble rotational transitions affected by a hyperfine structure (see Figure S3 of the SI). This latter is the result of the interaction between the $^{14}$N quadrupole moment and the electric gradient at the nucleus,[42] and it is thus an indication that these rotational transitions belong to 4FT. In the following analysis of the rotational transitions, the asymmetric-top rigid-rotor Watson Hamiltonian[42,50] combined, if required, with a term accounting for the nuclear quadrupole coupling[42,49] was employed.

Guided by the theoretical predictions, we undertook the search of the low-lying conformers of 4FT, starting from the most stable one, I′gg$^-$g$^-$ (1; see Table 1). From the identification of some rotational transitions, the spectral assignment proceeded iteratively and led to the rotational and nitrogen quadrupole coupling constants collected in the first column of Table 1 (see the Experimental section for details). To further proceed with spectral analysis, the rotational lines assigned to I′gg$^-$g$^-$ were removed from the spectrum and the remaining transitions were analyzed in the search of the second most stable species, the IItg$^-$g$^-$ conformer (2). From an initial assignment of two dozen of lines based on computational predictions, the analysis was again an iterative fit-and-predict process,[51] which led to the experimental spectroscopic parameters given in the second column of Table 1. We then removed the identified transitions of IItg$^-$g$^-$ from the rotational spectrum and moved to systematically search for other species, proceeding in decreasing order of stability. After the removal of all the identified transitions belonging to the six most-stable conformers (see Table 1) from the spectrum, several weak lines were still present (see Figure 5b). Therefore, we searched for the remaining low-lying conformers following the same strategy described above. Five additional conformers were discovered and their obtained experimental spectroscopic parameters are given in Table 2. For all species but conformers 7 and 10, the hyperfine structure was resolved and allowed us to obtain the quadrupole coupling constants, which gave further support to the conformational assignment.

According to our computations, a twelfth conformer – having relative energy and abundance similar to those of conformer 11 – should be accessible. Its quite large *a* component of the dipole moment is expected to lead to rotational transitions sufficiently intense to be observed. Unfortunately, the low signal-to-noise ratio of the CPMW spectrum prevented from the assignment of conformer 12. Instead, exploiting the unprecedented resolution and sensitivity of a molecular-beam Fourier-transform microwave spectrometer (MB-FTMW; see the Experimental section),[52] we proceeded with searching for it. The excellent agreement between theory and experiment allowed us to locate a few transitions and resolve their hyperfine structure (see Figure 5c). The derived spectroscopic parameters are provided in the second-last column of Table 2 and nicely match those predicted for the Ig$^-$gg conformer.

As can be seen in Tables 1 and 2, there is a perfect match between theoretical and experimental results, also for what concerns the type of transitions observed (which are collected in Tables S7–S18 of the SI). The comparison between the experimentally-derived rotational constants and their predicted values for the most stable conformers of 4FT made their identification straightforward. The remarkable accuracy of the computational predictions (average absolute errors of 8.2, 3.2 and 3.2 MHz for $A_0$, $B_0$ and $C_0$, respectively, which correspond to a relative error of about 0.3 %) is on par with that of state-of-the-art quantum chemical methodologies[44,53] and was indeed crucial for such a successful outcome because the rotational constants of several conformers are quite similar one another. Fortunately, all conformers show at least one rotational constant slightly different from the others; this is especially true for the *C* rotational constant. Those conformers that have similar *C* present different *A* (see for example, IIggg$^-$ (3) and I′gg$^-$g (4) conformers) or *B* (see for example IIggg (7) and

Research Article
doi.org/10.1002/chem.202203990
Chemistry Europe
European Chemical Societies Publishing

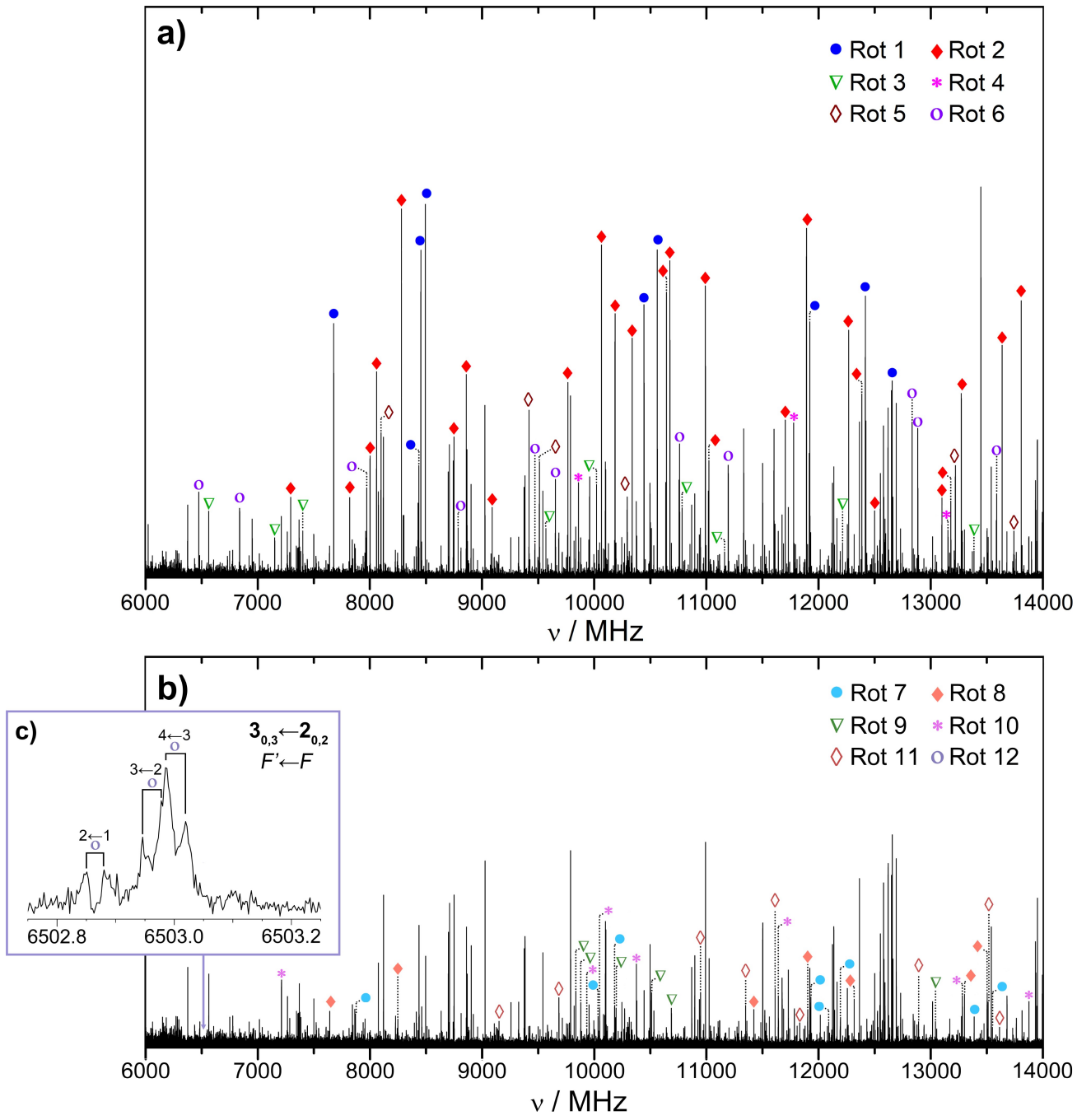


**Figure 5.** a) Broadband spectrum of 4FT in the 6–14 GHz region: the rotational transitions assigned to the six most-stable conformers (1-6) are shown; b) broadband spectrum of 4FT after removing the transitions belonging to 1 to 6: the transitions assigned to the other six conformers (7-12) are shown; c) Inset: narrow-band recording of the $3_{0,3}\leftarrow2_{0,2}$ rotational transition (for notation, see the Experimental section) of the conformer 12 showing its resolved hyperfine structure (all features are split into two components due to the Doppler effect).

IIg⁻g⁻g (8) conformers) constants. However, the differences here addressed lie well within the uncertainties of the low accuracy methodologies that are often employed in the investigation of rotational spectra for flexible molecules of biological interest.

Unfortunately, a quantitative experimental estimate of the conformer populations could not be obtained from the analysis of the CPMW spectra because different experimental conditions were employed in the search and characterization of each 4FT conformer. However, the successful observation of the twelfth conformer is the best demonstration of the reliability of theory in predicting which and how many conformers should be experimentally observed.

### Intra-molecular interactions

To better unravel the role of fluorine in tuning the conformational landscape of 4FT, in Table 3, we compare the relative electronic energies of the Thr conformers with their 4FT counterparts. For those 4FT conformers showing close contacts between F and acidic hydrogens, the corresponding distances are given in the last column and their energies are highlighted in bold. It is apparent that all 4FT conformers involving F···H hydrogen bonds are significantly stabilized with respect to their Thr counterparts. The most significant consequence of this trend is the inversion of stability order for some conformer pairs when moving from Thr to 4FT. For example, in Thr, Igg$^-$ is more stable than I'g$^-$g, but the stability order is reversed in 4FT, with I'gg$^-$g$^-$ lying lower in energy than Ig$^-$gg$^-$ (this also applies to the I'gg$^-$g/Igg$^-$g pair). An analogous example is provided by the IIgg/IItg$^-$ pair of Thr and its IIggg$^-$/IItg$^-$g$^-$ counterpart in 4FT. More generally, in 4FT, the g orientation of both $\chi_2$ and $\chi_3$ dihedrals is strongly disfavoured with respect to its g$^-$ counterpart. In both amino acids, the t and g$^-$ orientations of $\chi_2$ show instead similar relative stabilities, while the t orientation of $\chi_3$ is observed in 4FT only in conjunction with the t orientation of $\chi_2$. The Igt conformer of Thr was not observed because lying higher in energy with respect to the absolute minimum (IIgg) and the same occurs for the IItg$^-$g, III'g$^-$gg$^-$ and III'g$^-$gg conformers of 4FT (the absolute minimum being IItg$^-$g$^-$). The IIgt and IIg$^-$g$^-$ conformers of Thr relax to their IIgg and IIg$^-$t counterparts, respectively, along paths governed by very low energy barriers.

Deeper insight in the stereo-electronic origin of the above trends can be achieved by means of the so-called non covalent interaction (NCI) analysis,[54,55] which is based on the electron density and its derivatives. The results are shown in Figure 6. In this figure, the color code provides us with quantitative information, thus allowing us to infer the role played by each interaction in the overall stabilization of the molecule.

According to Tables 3 and S2 of the SI, the I'gg$^-$g$^-$, IIggg$^-$ and IItg$^-$g$^-$ conformers of 4FT correlate with I'gg$^-$, IIgg and IItg$^-$ of Thr, respectively.[9,10] While the relative stability order remains almost unchanged for I'gg$^-$g$^-$(4FT)/I'gg$^-$(Thr) and IIggg$^-$(4FT)/IIgg(Thr), the introduction of a fluorine atom greatly stabilizes IItg$^-$g$^-$. In all three conformers, fluorine is involved in additional intra-molecular interactions, but Figure 6 points out that, for the latter conformer, such an interaction is more relevant (in terms of both extension of the electron density regions and strength). Together with the three most stable structures, the III'ggg$^-$ conformer of 4FT has also been considered for the NCI analysis. This is the most stable type-III conformer. Its relative stability decreases with respect to its Thr counterpart (III'gg).[9] This is well explained by the reduced role played by fluorine in establishing intra-molecular interactions (see Figure 6).

**Table 3.** Relative electronic energies[a] (in cm$^{-1}$) of selected Thr and 4FT conformers. The reference conformers are the most stable one of Thr (IIgg) and its 4FT counterpart (IIggg$^-$). Close contacts between F and acidic hydrogens in 4FT are given in the last column.

| Conformer | Thr[b] | 4FT[b,c] ($\chi_3$ = g$^-$) | ($\chi_3$ = g) | F interaction |
|---|---|---|---|---|
| I'gg$^-$ | 459.8 | **79.0** | 350.3 | F···HN = 2.36 Å |
| Ig$^-$g | 218.5 | 518.4 | 561.6 | |
| Igt | (962.3) | **402.0**[d] | | F···HO = 2.39 Å |
| IItg$^-$ | 371.5 | **−116.0** | (841.0) | F···HN = 2.12 Å |
| IIgg | 0.0 | 0.0 | 394.1 | |
| IIg$^-$t | 711.2 | **296.0**[d] | | F···HO = 2.27 Å |
| IIg$^-$g$^-$ | (725.9) | (2186.2) | **369.0** | F···HO = 2.21 Å |
| IIgt | (586.1) | **512.9**[d] | | F···HO = 2.41 Å |
| III'g$^-$g | 574.6 | (919.8) | (1085.9) | |
| III'gg | 624.2 | 532.4 | (1982.3) | |

[a] At the rDSD level (1 kJ/mol = 83.59 cm$^{-1}$). [b] Values in parentheses refer to conformers not observed experimentally because unstable and/or because they relax to more stable structures. [c] Energies highlighted in bold refer to 4FT conformers showing F···H hydrogen bonds. [d] t conformer.

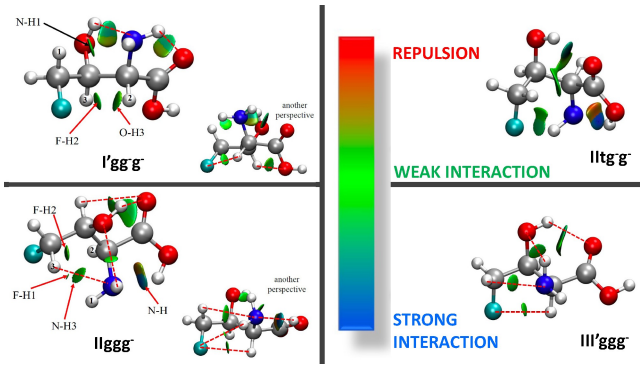


**Figure 6.** Non-covalent interactions in the I'gg$^-$g$^-$ (1), IItg$^-$g$^-$ (2), IIggg$^-$ (3), and III'ggg$^-$ (10) conformers of 4FT.

## Conclusion

An unbiased disentanglement of the conformational bath of 4FT has been successfully obtained by exploiting an integrated experimental/computational approach going well beyond the current state of the art for medium-sized flexible molecules. This allowed us to accurately characterize twelve conformers and understand why a much lower number of conformers (seven) was experimentally detected for the parent Thr amino acid and why fluorination of the latter is not able to account for all 4FT conformers. Overall, the present work provides the first global picture of the conformational behavior of the neutral form of the only natural fluorinated amino acid. These findings together with the NCI analysis performed for 4FT allow a qualitative and quantitative description of the role played by fluorine in tuning the conformational landscape of biomolecules. Therefore, the accurate results now available for 4FT can be used to improve the modeling of fluorine-substitution effects in biomolecules by e.g. providing an accurate description of fluorine atoms in empirical force fields.

Together with the intrinsic relevance of the studied molecule, the results of the present investigation show that, owing to the development of state-of-the-art artificial intelligence and quantum chemical methodologies and their combination in a powerful integrated strategy, a fully *a priori*, highly reliable analysis of the conformational landscape is today

possible for flexible building blocks of biomolecules in the gas phase. Synergistically, rotational spectroscopy provides an unbiased and rigorous validation of the developed computational approaches, and is an integral part of the strategy. This latter, once implemented in a general, user friendly platform, will open the accurate and exhaustive landscape exploration also to non-specialists.

4FT proved to be an optimal playground to show that medium-sized flexible molecules require a suitable strategy for the correct and exhaustive exploration of the conformational landscape. A standard *a posteriori* analysis of the rotational spectrum would hardly have led to the assignment of twelve conformers. Standard quantum chemical methodologies (e.g. hybrid density functionals and low-order perturbative methods) proved to be not suitable for a correct energetic analysis of the conformational bath,[11] offering instead the possibility of an erroneous interpretation of the experimental evidence.

## Experimental

### Sample Synthesis and Preparation

An effective racemic, but diastereomerical synthesis of 4FT was recently proposed by some of the authors in order to reach the maximum possible yield.[10] This involved a fluorination step (see Figure S2 in the SI), which was performed with DAST (diethylaminosulphur trifluoride). To obtain the required amount of 4FT, it was necessary to perform the reaction in a limited scale (200–300 mg of starting material) and to repeat it five times in order to avoid extensive decomposition upon DAST addition (experienced on larger scale). The produced 4FT sample was repeatedly washed with methanol and accurately dried under vacuum, thus leading to a spectroscopically pure compound as white powder. Further details are given in the SI.

Due to the high boiling point, conventional heating methods are not able to vaporize the sample. Thus, as mentioned in the main text, the LA technique has been employed. To this purpose, a small amount of 4FT was finely powdered, mixed with a small amount of copolymeric binder, and finally pressed to form a cylindrical rod. After some time, the sample rod was ready to be placed in the ablation nozzle of the spectrometers.

### Spectral Recording and Assignment

The products of LA were supersonically expanded and analyzed by means of two spectrometers (CP-FTMW and MB-FTMW). The estimated accuracy of the frequency measurements is 50 kHz. A detailed account of the experimental setup and conditions is provided in the SI.

The assigned rotational transitions are classified according to the selection rules for the principal quantum number $J$ ($\Delta J = -1, 0, +1$ leading to *P*-, *Q*- and *R*-branches, respectively) and non-zero dipole moment component (*a*-, *b*- and *c*-type referring to $\mu_a$, $\mu_b$ and $\mu_c$ dipole moment components, respectively). For asymmetric rotors like 4FT, rotational energy levels are identified by two further quantum numbers in addition to $J$, namely $K_{-1}$ and $K_{+1}$: $J_{K_{-1},K_{+1}}$. Transitions are usually denoted reporting the energy levels involved, with the final one given as the first. All measured transitions and hyperfine components are collected in Tables S7 to S18 of the SI.

For the most stable conformer (l′gg$^-$g$^-$), based on computational predictions, a *c*-type *R*-branch progression was easily identified and used to generate an improved set of constants that allowed the identification of new *R* and *Q*-branch *c*-type transitions. No *a*- or *b*-type lines were instead observed. Most of the recorded transitions show clear hyperfine structures, which were used in a mixed fit, also incorporating the frequency centers of some non-resolved transitions. Overall, 14 hyperfine components and 5 unresolved transitions led to the spectroscopic parameters of Table 1. For the lltg$^-$g$^-$ conformer, we relied on the *R*-branch *a*-type transitions because their characteristic pattern is quite straightforward to locate. We easily identified up to 23 *a*-type transitions with $K_{-1}$ values ranging from 0 to 3. This initial assignment produced a first set of rotational constants that allowed us to extend the analysis to *R*-branch *b*- and *c*-type transitions. Since *b*- and *c*-type transitions show resolved hyperfine components, a mixed fit involving 15 hyperfine components and 17 frequency centers was carried out. This led to the experimental spectroscopic parameters reported in Table 1. We then repeated the same procedure for assigning the remaining lines to additional nine conformers in order of decreasing stability, as explained in the main text (spectroscopic parameters given in Tables 1 and 2). As already mentioned, the twelfth conformer was only detected using the MB-FTMW spectrometer,[52] which is based on a Fabry-Perot resonator and is dedicated to maximize performances at low-frequency ranges (2 to 8 GHz). The excellent agreement between theory and experiment allowed us to locate four *R*-branch transitions (Figure 5c) and measure up to 10 hyperfine components, these leading to the spectroscopic parameters in the second-last column of Table 2.

As mentioned in the main text, for all 4FT conformers, the analysis of the rotational transitions was performed using an asymmetric-top rigid-rotor Watson Hamiltonian[42,50] combined, when required, with a term describing the nuclear quadrupole coupling.[42,49] As described above, for those conformers for which hyperfine structures were not completely resolved, we incorporated both hyperfine components and frequency centers of non-resolved transitions in a mixed fit taking into account all assigned transitions in the determination of the spectroscopic parameters.

## Computational

As mentioned in the main text, the conformational PES exploration was carried out using the IMEA algorithm,[14,27] which was driven by the molecular perception PROXIMA library.[56] The initial conformational PES exploration involved four replicas (i.e. runs of the full algorithm starting from different initial conditions), each consisting of about 4000 constrained geometry optimizations, and employed a low-cost (GFN2-XTB[13]) semi-empirical method. Then, a first reduction of the number of structures was obtained by applying a threshold of 1250 cm$^{-1}$ with respect to the absolute energy minimum. The retained candidates were compared with each other in terms of the root-mean-square deviations of heavy atom positions and rotational constants. The surviving structures were further reduced by clustering procedures and their geometries optimized at the B3 level. Finally, the geometries of the most promising candidates were refined at the rDSD level.

The junChS–F12 composite scheme was employed for accurate evaluations of electronic energies. This approach is based on the CCSD(T)-F12b(3 C/FIX) method[57–59] (hereafter CCSD(T)-F12), within the frozen-core (fc) approximation, in

conjunction with the jun-cc-pVTZ basis set.[29,33] To improve the CCSD(T)-F12 energy, extrapolation to the complete basis set (CBS) limit is carried out at the MP2-F12 level[60] employing the jun-cc-pVTZ and jun-cc-pVQZ basis sets and exploiting the standard $n^{-3}$ two-point formula.[61] The core-valence (CV) correlation contribution is then incorporated as difference between all-electron (ae) and fc MP2-F12 calculations, both in conjunction with the cc-pCVTZ–F12 basis set.[62] Further details are given in the SI.

Improved geometrical parameters ($r$), and thus improved equilibrium rotational constants, were obtained by exploiting the LRA methodology. This requires correcting the rDSD bond lengths ($r_{rDSD}$) for systematic errors by means of scaling factors ($a$) and offset values ($b$) depending on the nature of the involved atoms:[46,63]

$$r = (1 + a) \times r_{rDSD} + b \,, \tag{1}$$

where the $a$ and $b$ parameters were derived from a large database of accurate semi-experimental equilibrium geometries, as described in ref. [46]. The intrinsic accuracy of the rDSD model led to, for all types of bonds, $b = 0.0$ and very small values for $a$. The resulting LRA parameters (the type of bond being provided as subscript to $a$) are: $a_{CC} = -0.0018$ Å; $a_{XH}$ (X = C, N, or O) $= a_{CN} = -0.0023$ Å; $a_{CO} = a_{CF} = -0.003$ Å. Several studies have confirmed the remarkable accuracy of the structures issued from this linear regression approach.[11,46,63,64]

The vibrational contributions to thermodynamic functions were first estimated with the harmonic oscillator model employing rDSD analytical Hessians. Next, anharmonic contributions to ZPEs were estimated using a VPT2 model that provides an analytical and resonance-free expression.[39] The contribution of low-frequency vibrations to entropies was instead estimated by the QH approximation[21,40] in which, below a given cut-off value (typically 100 $cm^{-1}$), entropic terms are obtained from the free-rotor approximation and a damping function is used to interpolate between free-rotor and harmonic oscillator expressions close to the cut-off frequency.

The Gaussian package[65] has been employed for all calculations except the junChS–F12, QH and NCI ones, which have been performed using the Molpro,[66] GoodVibes,[40] and NCI/VMD[55,67] software, respectively.

## *Acknowledgements*

The financial fundings from Ministerio de Ciencia e Innovación (PID2019-111396GB-I00), Junta de Castilla y León (Grant VA244P20), Italian Ministry of University and Research (MUR, Grants 2017A4XRCA and 202082CE3T), Italian Space Agency (ASI,'Life in Space' project N. 2019-3-U.0) and European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC-2013-SyG, Grant Agreement n. 610256 NANOCOSMOS, are gratefully acknowledged. We also thank the technical staff at SNS SMART Laboratory for managing the computational facilities and Hexu Ye for her help in the NCI analysis and visualization. Open Access funding provided by Scuola Normale Superiore within the CRUI-CARE Agreement.

## Conflict of Interest

The authors declare no competing financial interest.

## Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

**Keywords:** Machine Learning · Quantum Chemistry · Microwave Spectroscopy · Conformational Landscape · Fluorinated Amino Acids

[1] S. Purser, P. R. Moore, S. Swallow, V. Gouverneur, *Chem. Soc. Rev.* **2008**, *48*, 320.
[2] E. P. Gills, K. J. Eastman, M. D. Hill, D. J. Donnelly, N. A. Meanwell, *J. Med. Chem.* **2015**, *58*, 8315.
[3] Y. Pan, *ACS Med. Chem. Lett.* **2019**, *10*, 1016.
[4] D. O'Hagan, *Chem. Soc. Rev.* **2008**, *37*, 308.
[5] N. A. Meanwell, *J. Med. Chem.* **2018**, *61*, 5822.
[6] J. Moschner, V. Stulberg, R. Fernandes, S. Huhmann, J. Leppkes, B. Koksch, *Chem. Rev.* **2019**, *119*, 10718.
[7] C. D. Murphy, C. Schaffrath, D. O'Hagan, *Chemosphere* **2003**, *52*, 455.
[8] C. Odar, M. Winkler, B. Wiltschi, *Biotechnol. J.* **2015**, *10*, 427.
[9] J. L. Alonso, M. Pérez, M. E. Sanz, S. Blanco, *Phys. Chem. Chem. Phys.* **2009**, *11*, 617.
[10] S. Potenti, L. Spada, M. Fusè, G. Mancini, A. Gualandi, C. Leonardi, P. G. Cozzi, C. Puzzarini, V. Barone, *ACS Omega* **2021**, *6*, 13170.
[11] I. León, M. Fusè, E. R. Alonso, S. Mata, G. Mancini, C. Puzzarini, J. L. Alonso, V. Barone, *J. Chem. Phys.* **2022**, *157*, 074107.
[12] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618.
[13] C. Bannwarth, S. Ehlert, S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 1652.
[14] G. Mancini, M. Fusè, F. Lazzari, B. Chandramouli, V. Barone, *J. Chem. Phys.* **2020**, *153*, 124110.
[15] T. Fornaro, D. Burini, M. Biczysko, V. Barone, *J. Phys. Chem. A* **2015**, *119*, 4224.
[16] E. Penocchio, M. Piccardo, V. Barone, *J. Chem. Theory Comput.* **2015**, *11*, 4689.
[17] V. Barone, *J. Chem. Phys.* **2005**, *122*, 014108.
[18] A. M. Rosnik, W. F. Polik, *Mol. Phys.* **2014**, *112*, 261.
[19] P. R. Franke, J. F. Stanton, G. E. Douberly, *J. Phys. Chem. A* **2021**, *125*, 1301.
[20] M. Mendolicchio, J. Bloino, V. Barone, *J. Chem. Theory Comput.* **2022**, *18*, 7603.
[21] S. Grimme, *Chem. A Eur. J.* **2012**, *18*, 9955.
[22] S.-C. Li, Y.-C. Lin, Y.-P. Li, *J. Chem. Theory Comput.* **2022**, *18*, 6866–6877.
[23] C. Puzzarini, M. Biczysko, V. Barone, L. Largo, I. Pena, C. Cabezas, J. L. Alonso, *J. Phys. Chem. Lett.* **2014**, *5*, 534.
[24] S. Alessandrini, V. Barone, C. Puzzarini, *J. Chem. Theory Comput.* **2020**, *16*, 988.
[25] J. Lupi, S. Alessandrini, V. Barone, C. Puzzarini, *J. Chem. Theory Comput.* **2021**, *17*, 6974.
[26] J. L. Alonso, J. C. López, Microwave spectroscopy of biomolecular building blocks, in *Gas-Phase IR Spec- troscopy and Structure of Biological Molecules*, pages 335–401, Springer **2015**.
[27] G. Mancini, M. Fusè, F. Lazzari, V. Barone, *Digit. Disc.* **2022**, *1*, 10539.
[28] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
[29] T. H. Dunning, *J. Chem. Phys.* **1989**, *90*, 1007.
[30] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[31] G. Santra, N. Sylvetsky, J. M. L. Martin, *J. Phys. Chem. A* **2019**, *123*, 5129.

[32] T. H. Dunning, K. A. Peterson, A. K. Wilson, *J. Chem. Phys.* **2001**, *114*, 9244.
[33] E. Papajak, J. Zheng, X. Xu, H. R. Leverentz, D. G. Truhlar, *J. Chem. Theory Comput.* **2011**, *7*, 3027.
[34] E. R. Alonso, I. León, J. L. Alonso, *Intra- and Intermolecular Interactions Between Non-Covalently Bonded Species*, Elsevier **2020**.
[35] P. D. Godfrey, R. D. Brown, F. M. Rodgers, *J. Mol. Struct.* **1996**, *376*, 65.
[36] G. M. Florio, R. A. Christie, K. D. Jordan, T. S. Zwier, *J. Am. Chem. Soc.* **2002**, *124*, 10236.
[37] R. S. Ruoff, T. D. Klots, T. Emilsson, H. S. Gutowsky, *J. Chem. Phys.* **1990**, *93*, 3142.
[38] J. Lupi, C. Puzzarini, C. Cavallotti, V. Barone, *J. Chem. Theory Comput.* **2020**, *16*, 5090.
[39] J. Bloino, M. Biczysko, V. Barone, *J. Chem. Theory Comput.* **2012**, *8*, 1015.
[40] G. Luchini, J. Alegre-Requena, I. Funes-Ardoiz, R. Paton, *F1000Research* **2020**, *9*.
[41] M. E. Sanz, S. Blanco, J. C. López, J. L. Alonso, *Angew. Chem. Int. Ed.* **2008**, *47*, 6216.
[42] W. Gordy, R. L. Cook, A. Weissberger, *Microwave molecular spectra*, volume 18, Wiley New York **1984**.
[43] I. M. Mills, Vibration-Rotation Structure in Asymmetric- and Symmetric-Top Molecules, in K. N. Rao, C. W. Mathews (Editors), *Molecular Spectroscopy: Modern Research*, pages 115–140, Academic Press, New York **1972**.
[44] C. Puzzarini, J. F. Stanton, J. Gauss, *Int. Rev. Phys. Chem.* **2010**, *29*, 273.
[45] M. Piccardo, E. Penocchio, C. Puzzarini, M. Biczysko, V. Barone, *J. Phys. Chem. A* **2015**, *119*, 2058–2082.
[46] G. Ceselin, V. Barone, N. Tasinato, *J. Chem. Theory Comput.* **2021**, *17*, 7290.
[47] C. Scolastico, E. Conca, L. Prati, G. Guanti, L. Banfi, A. Berti, P. Farine, U. Valcavi, *Synthesis* **1985**, *1985*.
[48] S. Mata, I. Peña, C. Cabezas, J. C. López, J. L. Alonso, *J. Mol. Spectrosc.* **2012**, *280*, 91.
[49] I. León, E. R. Alonso, S. Mata, C. Cabezas, J. L. Alonso, *Angew. Chem. Int. Ed.* **2019**, *58*, 16002.
[50] J. K. G. Watson, Aspects of quartic and sextic centrifugal effects on rotational energy levels, in J. R. Durig (Editor), *Vibrational Spectra and Structure*, volume 6, pages 1–78, Elsevier, New York/Amsterdam **1977**.
[51] H. M. Pickett, *J. Mol. Spectrosc.* **1991**, *148*, 371.
[52] C. Bermúdez, S. Mata, C. Cabezas, J. Alonso, *Angew. Chem. Int. Ed.* **2014**, *53*, 11015.
[53] M. B. Gardner, B. R. Westbrook, R. C. Fortenberry, T. J. Lee, *Spectrochim. Acta Part A* **2021**, *248*, 119184.
[54] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, W. Yang, *J. Am. Chem. Soc.* **2010**, *132*, 6498.
[55] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, W. Yang, *J. Chem. Theory Comput.* **2011**, *7*, 625.
[56] F. Lazzari, A. Salvadori, G. Mancini, V. Barone, *J. Chem. Inf. Model.* **2020**, *60*, 2668.
[57] T. B. Adler, G. Knizia, H.-J. Werner, *J. Chem. Phys.* **2007**, *127*, 221106.
[58] G. Knizia, T. B. Adler, H.-J. Werner, *J. Chem. Phys.* **2009**, *130*, 054104.
[59] H.-J. Werner, G. Knizia, F. R. Manby, *Mol. Phys.* **2011**, *109*, 407.
[60] H.-J. Werner, T. B. Adler, F. R. Manby, *J. Chem. Phys.* **2007**, *126*, 164102.
[61] T. Helgaker, W. Klopper, H. Koch, J. Noga, *J. Chem. Phys.* **1997**, *106*, 9639.
[62] J. G. Hill, S. Mazumder, K. A. Peterson, *J. Chem. Phys.* **2010**, *132*, 054108.
[63] A. Melli, F. Tonolo, V. Barone, C. Puzzarini, *J. Phys. Chem. A* **2021**, *125*, 9904.
[64] H. Ye, S. Alessandrini, M. Melosso, C. Puzzarini, *Phys. Chem. Chem. Phys.* **2022**, *24*, 23254.
[65] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman,, D. J. F. and, Gaussian 16 Revision C.01, gaussian Inc. Wallingford CT 2016.
[66] H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. Heßelmann, D. Kats, A. Köhn, T. Korona, D. A. Kreplin, Q. Ma, T. F. Miller, A. Mitrushchenkov, K. A. Peterson, I. Polyak, G. Rauhut, M. Sibaev, *J. Chem. Phys.* **2020**, *152*, 144107.
[67] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33.

Manuscript received: December 21, 2022
Accepted manuscript online: February 3, 2023
Version of record online: March 20, 2023