



HAL
open science

Intrinsic folding of the cysteine residue: competition between folded and extended forms mediated by the –SH group

Gildas Goldsztejn, Venkateswara Rao Mundlapati, Valérie Brenner, Eric Gloaguen, Michel Mons, Carlos Cabezas, Iker León, José Luis Alonso

► **To cite this version:**

Gildas Goldsztejn, Venkateswara Rao Mundlapati, Valérie Brenner, Eric Gloaguen, Michel Mons, et al.. Intrinsic folding of the cysteine residue: competition between folded and extended forms mediated by the –SH group. *Physical Chemistry Chemical Physics*, 2020, 22 (36), pp.20284-20294. 10.1039/D0CP03136D . hal-03012453

HAL Id: hal-03012453

<https://hal.science/hal-03012453>

Submitted on 14 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Cite this: *Phys. Chem. Chem. Phys.*,
2020, 22, 20284

Intrinsic folding of the cysteine residue: competition between folded and extended forms mediated by the –SH group†

Gildas Goldsztejn, ^{‡,a} Venkateswara Rao Mundlapati, ^a Valérie Brenner, ^a
Eric Gloaguen, ^a Michel Mons, ^{*a} Carlos Cabezas, ^{§b} Iker León ^b and
José Luis Alonso ^{*b}

A dual microwave and optical spectroscopic study of a capped cysteine amino acid isolated in a supersonic expansion, combined with quantum chemistry modelling, enabled us to characterize the conformational preferences of Cys embedded in a protein chain. IR/UV double resonance spectroscopy provided evidence for the coexistence of two conformers, assigned to folded and extended backbones (with classical C7 and C5 backbone H-bonding respectively), each of them additionally stabilized by specific main-chain/side-chain H-bonding, where the sulfur atom essentially plays the role of H-bond acceptor. The folded structure was confirmed by microwave spectroscopy, which demonstrated the validity of the DFT-D methods currently used in the field. These structural and spectroscopic results, complemented by a theoretical Natural Bond Orbital analysis, enabled us to document the capacity of the weakly polar –CH₂–SH side chain of Cys to adapt itself to the intrinsic local preferences of the peptide backbone, *i.e.*, a γ -turn or a β -sheet extended secondary structure. The corresponding local H-bonding bridges the side chain acceptor S atom to the backbone NH donor site of the same or the next residue along the chain, through a 5- or a 6-membered ring respectively.

Received 11th June 2020,
Accepted 27th August 2020

DOI: 10.1039/d0cp03136d

rs.c.li/pccp

1. Introduction

For several decades, high resolution spectroscopy has been used to characterize the intrinsic structure of molecules of increasing size and complexity, including small biomolecules.^{1–4} These investigations carried out on isolated molecules, in the absence of any solvent effect, have been possible thanks to the development of soft vaporization techniques, coupled with supersonic expansion to cool down the neutral species of interest, a necessary condition to overcome spectral congestion effects expected even with small sized molecules.⁵ Under these conditions, microwave spectroscopy provides a direct insight into the structure, through the rotational constants even with a flexible

molecule, whose conformational contributions can be separated thanks to the high resolution.⁶ Alternatively, optical spectroscopy in the IR range enables experimentalists to document the interactions at play in the molecules under examination, in particular their H-bonding content. Moreover, double resonance IR/UV techniques provide conformation resolution, yielding conformation-selective IR spectroscopy. In both cases, however, comparison with structures obtained with high-level quantum chemistry methods is the key for assigning the conformations observed.^{1–6} Conversely, the confrontation also enables experimentalists to assess the theoretical methods used.

These spectroscopic techniques have been widely used in the past decade to investigate isolated biomolecular models, allowing to tackle issues such as competition between folding structures in flexible biomolecules like peptides or sugars.^{1–6} However, the combined approach based on both microwave and IR/UV data remains sparse, and limited so far to a few small molecules of biological interest, *e.g.* ref. 7. Regarding amino acid only a few of them have been investigated by both techniques, although in independent studies: Ala^{8–10} or Pro.^{11,12} One of the interests of these models is to document the role of local main-chain/side-chain (MC/SC) interactions in protein chain folding. Recent gas phase IR/UV experiments carried out on capped amino acids (dipeptide analogues) or

^a LIDYL, CEA, CNRS, Université Paris-Saclay, bât 522, CEA Paris-Saclay,
9119 Gif-sur-Yvette, France. E-mail: michel.mons@cea.fr

^b Grupo de Espectroscopia Molecular (GEM), Edificio Quifima, Laboratorios de
Espectroscopia y Bioespectroscopia, Unidad Asociada CSIC, Parque Científico UVA,
Universidad de Valladolid, 47011, Valladolid, Spain. E-mail: jlalonso@qf.uva.es

† Electronic supplementary information (ESI) available. See DOI: 10.1039/
d0cp03136d

‡ Present address: Institut des Sciences Moléculaires d'Orsay, CNRS, Université
Paris-Saclay, 91405, Orsay, France.

§ Present address: Instituto de Física Fundamental (IFF-CSIC), Group of Molecular
Astrophysics, C/Serrano 121, 28006 Madrid, Spain.

peptides, which mimic a fragment of protein chain, have provided evidence for local interaction linking the MC to the SC of residues like asparagine,^{13,14} glutamine,^{15,16} histidine,¹⁷ serine,^{18,19} methionine,²⁰ cysteine,^{18,19} or selenocysteine;²¹ these side chains playing the role of donor and/or acceptor of a H-bond. Despite uncapped amino acids have been widely studied by microwave techniques, investigations of protein chain models remain seldom,^{22,23} the capped serine, Ac-Ser-NH₂, being the only example involving MC/SC H-bonding.²²

In a context where the presence of a sulfhydryl –SH group at the end of a short side chain has been shown to play a structural role in dipeptide folding,^{18–20} we found interesting to lead a joint microwave and optical study of the same species, the Cys model dipeptide, in order to obtain more precise information on both the structure and interactions, that can be confronted to high level quantum chemistry modeling.

2. Methods

2.1. Experiment

Optical spectroscopy. Conformation-specific IR spectroscopy was carried out using an experimental set-up and procedures, which were described previously.²⁴ It relies on UV spectroscopy to sort out conformers and requires the presence of a convenient chromophore in the model molecule studied.³ For this purpose and in line with previous experimental strategies,^{14,17} a benzyloxycarbonyl (Z-) cap on the N-terminal was used for laser experiments, complemented by an amide cap on the C-terminal, resulting in a Bn-O-CO-Cys-NH₂, or Z-Cys-NH₂, capped peptide. Basically, the phenyl UV spectroscopy, which is sensitive to its close environment, provides specific UV signatures for each conformer. Z-Cys-NH₂ (Genscript) was mixed with graphite (1:4 molecular weight ratio) and pressed into a 6 mm pellet, fixed close to the 1 mm diameter nozzle of a pulsed valve, operating with a 30:70 He:Ne mixture at a backing pressure of 18 bars. The sample was vaporized using the second harmonic of a ns Nd:YAG laser (Continuum, 10 Hz–3 mJ), guided to the sample through an optical fiber. The Z-Cys-NH₂ molecules were entrained and cooled down by the supersonic jet into an expansion chamber, and introduced into the source region of a time-of-flight mass spectrometer, where UV excitation took place. The UV spectroscopy was recorded using the resonant two-photon ionization technique. The output of a frequency-doubled dye laser, pumped by the third harmonic of a ns Nd:YAG laser (Continuum), was scanned in the region of the first π – π^* transition of the phenyl group, which also enabled to photoionise the molecules during the same laser pulse. The ions produced were then mass-analysed in a reflectron time-of-flight mass spectrometer. Double resonance IR/UV experiments were used to record conformer-specific IR spectra. A mid-IR OPO, equipped with a difference frequency generation (DFG) module, was used to excite specific molecular modes: in the NH stretch region (amide A) and in the CO stretch, NH bend (amide I & II) regions. IR spectra are obtained from the

IR-induced depopulation of the conformer-selected ion signal, normalized by this signal in absence of an IR laser.²⁴ Despite a potential interest, one has to note that due to a lack of power of our OPO in the spectral range expected for the SH stretch (2600 cm⁻¹),²⁵ this absorption region could not be investigated.

Microwave spectroscopy. Ac-Cys-NH₂ was produced in a supersonic jet by laser ablation of solid rods. These rods were prepared using commercial samples of Ac-Cys-NH₂ (GeneCust, ~99%, m.p. ~157 °C) without any further purification and pressing the compound's fine powder mixed with a small amount of a commercial binder. These rods were placed in the ablation nozzle and a picosecond Nd:YAG laser (12 mJ per pulse, 20 ps pulse width) was used as a vaporization tool. The resulting products of the laser ablation were supersonically expanded utilizing a flow of neon (10 bar) gas and then probed by Chirp Pulsed Fourier Transform Microwave (CP-FTMW) spectroscopy. Details of the experimental setup have been given elsewhere.²⁶ Chirped-pulses of 4 μ s directly generated by the 24 GS s⁻¹ arbitrary waveform generator were amplified to about 300 W peak power using a traveling wave tube amplifier. The resulting pulses were then transmitted and detected by broadband microwave horn antennas in a high-vacuum chamber, where they interacted with the molecular supersonic expansion. At a repetition rate of 2 Hz, a total of 75 000 free induction decays (4 FID emissions per gas pulse) each with a 10 μ s length were averaged and digitized using a 50 GS s⁻¹ digital oscilloscope. The frequency-domain spectrum in the 6–14 GHz frequency range was obtained by taking a fast Fourier transform (FFT) following the application of a Kaiser–Bessel window to improve the baseline resolution. The resolution achieved was 50 kHz, and the estimated accuracy of the frequency measurements was better than 10 kHz.

2.2. Theory

Quantum chemistry optimized structures have been obtained, after a manual exploration of the conformational landscape of the Ac-Cys-NH₂ and Z-Cys-NH₂ molecules, where the backbone and side chain dihedrals were varied. Geometry optimisations were carried at the B97-D3 level of theory²⁷ using the Becke–Johnson damping and the three-body term options (B97-D3(BJ)-abc), of Turbomole²⁸ with a def2-TZVPPD basis set.^{29,30} The resolution-of-identity (RI) approximation³¹ and the auxiliary associated basis^{32–34} were also used. A *m3* default grid size, a convergence threshold of 10⁻⁵ a.u. on the norm of the cartesian gradient were used. Numerical harmonic frequencies were calculated with a step length of 0.02 a.u., with the central option of the *numforce* module. This RI-B97-D3(BJ)-abc/def2-TZVPPD level already proved to be a good compromise between optimized structures, energetics, vibrational frequency and calculation times, in a context where dispersion interactions are ubiquitous.³⁵ Mode-dependent scaling factors³⁶ obtained from linear fits of the correlation between experimental and harmonic NH stretch frequencies ($f^{\text{scaled}} = a + b \cdot f^{\text{harmonic}}$) on a large library of gas phase peptides (*cf.* Supplementary Information of ref. 37) usually provide an accuracy of typically 20 cm⁻¹ in the NH stretch range.³

The scaling factors (a , b) applied to the harmonic frequencies have been recalculated for the level of theory presently used: for a peptide NH stretch mode: $a = 372.8 \text{ cm}^{-1}$ and $b = 0.86953$; for a carboxamide NH_2 symmetric stretch mode, $a = 1209.8 \text{ cm}^{-1}$ and $b = 0.63115$ and for its antisymmetric counterpart, $a = 1324.1 \text{ cm}^{-1}$ and $b = 0.60872$; in the far infrared (amide I and II regions) we used $a = 0 \text{ cm}^{-1}$ and $b = 1.004$ for the NH bending modes and $a = 0 \text{ cm}^{-1}$ and $b = 1.008$ for the CO stretches, in order to fit experimental frequencies. For the purpose of comparison, calculations were also carried out without the RI option and with a more restricted (def2-TZVP) basis set as well.

For the assignment of the microwave spectra, quantum chemical calculations were carried out, using the Gaussian 16 program package,³⁸ at the B3LYP-D3(BJ)/def2-TZVP^{39–43} level of theory, a method shown to provide reliable values of rotational constants for analog molecules.⁶ Optimization were carried out using an ultrafine grid and tight convergence conditions (4.5×10^{-4} , 3×10^{-4} , 1.8×10^{-4} , 1.2×10^{-4} a.u. for the maximum and RMS forces and maximum and RMS displacements, respectively). The structures obtained were used to estimate the values of the rotational parameters for the Ac-Cys-NH₂ conformers, in particular the rotational constants and dipole moment components, necessary to predict transition frequencies before the spectral survey. Vibrational frequency calculations were carried out to ensure that the optimized geometries are true minima, and to provide the Gibbs free energies for the different conformers. In addition, the barrier heights to internal rotation of the methyl group were estimated at the same level of theory. For purpose of comparison between the B97-D3 and B3LYP-D3 methods used, B3LYP-D3(BJ) calculations were also carried out with the extended def2-TZVPPD basis set.

NBO analysis. The H-bond strengths have been assessed from the theoretical analysis of the electron density in the framework of the Natural Bond Orbital theory,^{44–46} which enables to quantify the stabilizing role of electron delocalization through the H-bond, starting from a Lewis-type description of the electron system. NBO analyses were performed on the B97-D3 structures using the NBO module⁴⁷ of the Gaussian 09 software.⁴⁸ The indicator chosen to assess the strength of the NH-X H-bonds is the sum of the $E(2)$ stabilization energies (noted ΣE_{HB}), resulting from individual interactions between a donor NBO (lone pairs of X = O, S) and the acceptor NBO of the NH group (σ^*_{NH}) involved in the H-bond considered. These energies were calculated at the MP2/TZVPP level of theory (previously shown to be a good compromise between method and computation time⁴⁹) through the second-order perturbation theory analysis of the Fock matrix:

$$E(2) = q_i \frac{F(i,j)^2}{|\epsilon_i - \epsilon_j|}$$

where q_i is the donor orbital occupancy, $F(i,j)^2$ is the off-diagonal matrix element and ϵ_i (resp. ϵ_j) is the donor (acceptor) orbital energy. This ΣE_{HB} indicator was recently proved as a

useful ranking tool for amide-amide interactions in isolated peptides, well correlated to spectroscopic data.^{35,50}

3. Theoretical exploration of the potential energy surface

3.1. Structures

Exploration of the conformational landscape of the capped cysteine residue (Ac-Cys-NH₂ model molecule) led to two competing lowest energy backbone families (Fig. 1 and Table 1): one folded and one extended, stabilized by H-bonds forming respectively a 7- and 5-membered ring³ (classically labelled C7 and C5, 7 and 5 in short), as already mentioned for other residues, like glycine,⁵¹ alanine.^{9,52} In the absence of polar side chain, a recent NBO analysis³⁵ demonstrated that these C7 and C5 remarkable structures are determined by a subtle balance between H-bonding and hyperconjugative effects involving among others the C α H and side chain C α C β covalent bonds. With more polar systems, like histidine,¹⁷ asparagine^{13,14,52} or glutamine,¹⁶ local MC/SC H-bonding also plays a large role by specifically stabilizing one conformation over the other.

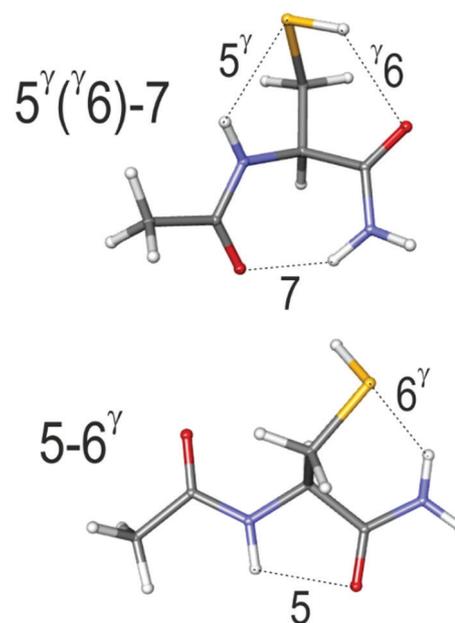


Fig. 1 DFT-D lowest minimum energy structures (RI-B97-D3/def2-TZVPPD level of theory) of Ac-Cys-NH₂, with corresponding H-bonding network. Top: Folded backbone 5^γ(6)-7; bottom: extended 5-6^γ form. H-bond distances are given in Table 1. The terminology adopted for the SC-MC H-bonds, 5^γ, 6^γ, 7^γ, follows that already introduced for histidine and asparagine polar side chains.^{14,17} It is based on that used for the backbone (5, 7), in which the number designates the size of the ring formed by the H-bond, and the greek superscript designates the location of the heavy atom of the SC (relative to the C- α atom) involved in the H-bond. The donor/acceptor character of the SC is indicated by the position of the greek superscript: the SC is acceptor (resp. donor) when the superscript is after (resp. before) the number. The conformations are designated by the H-bonding status of the several NH groups along the main chain, separated by a hyphen, with the SC status indicated between parenthesis.

Table 1 Structures, conformation terminology, energetics, H-bond interatomic distances of relevant conformations, at the RI-B97-D3-BJabc/def2-TZVPPD level of theory, together with NBO H-bonding strength indicators (in italics between parentheses); see text for additional details

Molecule	Ac-Cys-NH ₂		Bn-O-CO-Cys-NH ₂	
Conformation	#1	#2	#1	#2
Backbone Structure	Folded	Extended	Folded	Extended
Z-Cap rotamer	5 ^γ (^γ 6)-7	5-6 ^γ	5 ^γ (^γ 6)-7 <i>gauche</i> ⁺	5-6 ^γ <i>gauche</i> ⁻
Energetics (kJ mol ⁻¹)				
Δ <i>E</i> (electronic)	0	6.5	0	4.9
Δ <i>H</i> (0 K)	0	5.0	0	3.8
Δ <i>G</i> (300 K)	0	1.7	0	1.7
Interatomic distances (pm) (NBO Σ <i>E</i> _{HB}) (kcal mol ⁻¹)				
NH _{Cys} /NH ₂ ···O	7:200 (6.6)	5:214 (1.9)	7:202 (6.2)	5:215 (1.7)
NH _{Cys} /NH ₂ ···S	5 ^γ :257 (2.6)	6 ^γ :249 (5.7)	5 ^γ :261 (2.1)	6 ^γ :249 (5.5)
SH-O	^γ 6:239 (1.3)		^γ 6:238 (1.3)	

The capped Cys behaves in a similar way as these latter residues: it exhibits the two backbone structures, namely C7 and C5, each of them accompanied by a specific MC/SC H-bonding, whose terminology is explicated in the caption of Fig. 1:

- The folded family exhibits an inverse γ -turn backbone structure,³ characterized by a C7 H-bond leaving the Cys side chain in an equatorial position. Its most stable conformation (Fig. 1, top) is additionally stabilized by a specific MC \rightarrow SC 5 ^{γ} interaction complemented by a weaker SC \rightarrow MC ^{γ} 6 interaction, where the sulfhydryl group plays both the role of proton acceptor and donor respectively.

- The extended conformations, similar to the β -strands of proteins,³ are based on a C5 backbone, stabilized by a specific MC \rightarrow SC 6 ^{γ} interaction. Two forms were found: both side chains have an *anti* orientation, but they differ by the orientation of their sulfhydryl group, which can be described by the C α -C β -S-H (χ_2) dihedral. The most stable of them (Fig. 1, bottom) presents a *gauche*⁺ χ_2 dihedral, which enables a short 6 ^{γ} interaction (H-bonding distance 249 pm) to take place between the NH of the C-terminus cap and a lone pair of the sulfur atom. The SH···OC distance in this conformation is large enough (380 pm) to consider the sulfhydryl group free. The next form in the extended backbone family (ΔH energy higher by 4 kJ mol⁻¹) exhibits an *anti* χ_2 dihedral, which does not enable formation of such a short NH···S ^{γ} 6 bond; instead the SH group lies above the C-terminus cap NH₂ group, giving rise to both an elongated ^{γ} 6 bond and a weak SH···NH₂ interaction, penalizing the energetics of the structure.

Energetics. The folded form is found to be the most stable at 0 K, the extended form being less stable by *ca.* 5 kJ mol⁻¹ (Table 1). The situation at 300 K, however, is less contrasted, where the energy difference goes down to \approx 2 kJ mol⁻¹, due to a significant entropy stabilization encountered in the more floppy extended forms.⁵³ Usually such similar stabilities correlates with the coexistence, and observation, of both forms in a supersonic expansion.

Z-Cap model. In order to assess the validity of the Z-capped model for the study, we compared the properties of the most stable Z-cap rotamer for each backbone structure to those of the Ac-Cys-NH₂ molecule. Calculations indicated that the presence of the Z-protection (Table 1 and Fig. S1, ESI[†]) did

not significantly modify the conformational landscape, since structure, energetic order, and H-bonding distances were preserved within the C5-C7 manifold.

4. Optical spectroscopy

4.1. UV spectroscopy of Z-Cys-NH₂

The near UV spectroscopy of the Z-Cys-NH₂, in the origin region of the first $\pi\pi^*$ transition of the phenyl chromophore was obtained from resonant two-photon ionization spectroscopy (Fig. 2). It exhibited three bands labelled A, A' and B, whose conformational assignment was obtained from IR/UV experiments.

IR/UV measurements (Fig. 3) obtained by probing the bands A and A' were identical, showing that these UV bands belonged to the same conformer, labelled A, whereas the B UV feature exhibits a different spectrum and was thus assigned to another conformer.

The IR and UV spectroscopic data can thus be interpreted by the simultaneous presence in the supersonic expansion of:

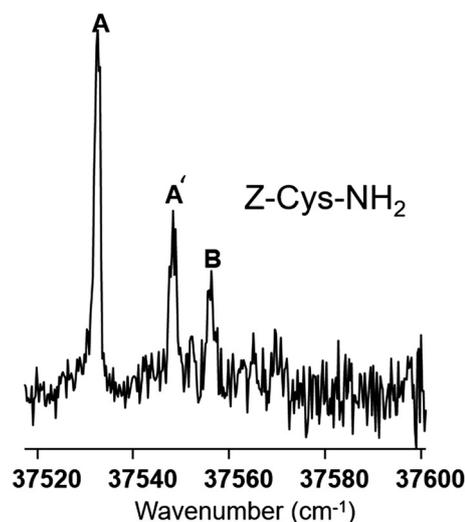


Fig. 2 Mass-resolved ($m/z = 254$ amu) resonant two-photon ionization spectrum of the Z-Cys-NH₂ molecule in the origin region of the first $\pi\pi^*$ transition of the phenyl group.

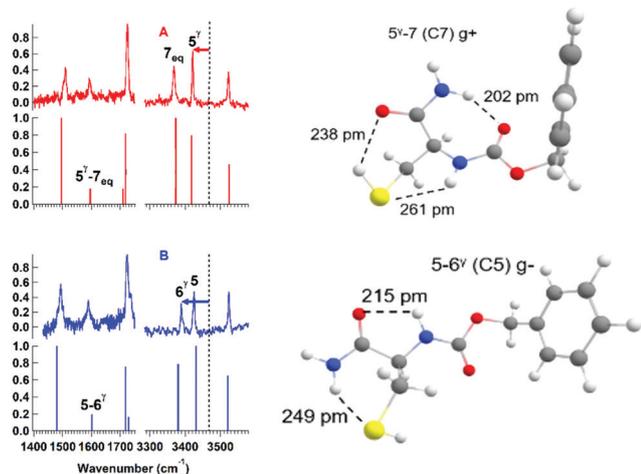


Fig. 3 Left: Conformer-specific IR spectra of conformers A (top) and B (bottom) of Z-Cys-NH₂, in the mid- and far-IR range, plotted as relative depletions measured in the IR/UV experiment, compared with predicted spectra for the most stable Z-tail rotamer of the folded and extended forms, resp. *gauche*⁺ and *gauche*⁻. Depletions in the experimental spectrum are normalized relative to the strongest band in the whole range investigated; theoretical stick spectra are normalized to the maximum intensity, within each region. The dotted line at 3468 cm⁻¹ indicates an average position for the free NHs in related model peptides;²¹ arrows symbolize the red shift associated to the H-bond (see text for a detailed discussion). Right: B97-D3 structures of these conformers.

– One major conformer, labelled A, whose UV spectroscopy exhibits a significant Franck–Condon activity, corresponding to the excitation of a low frequency mode at ~ 18 cm⁻¹ from the UV origin (Fig. 2, band A').

– One minor conformer, labelled B, with a unique UV feature. Assuming comparable photoionization efficiencies for both conformers, the relative population of B was estimated to 15%.

4.2. Conformer-selective IR spectroscopy of Z-Cys-NH₂

The conformer-specific IR spectra of conformers A and B (Fig. 3, and Table 2) looked very similar at a first glance, with three absorption features in the mid-IR (NH stretch) region, due to the presence of three NH oscillators, and three main features in the far IR (Amide I and II) region. They however presented a few differences, in particular, the redmost NH

stretch band seemed to be less intense and less red shifted in B, and the CO stretch band pattern of B exhibited a shoulder to the blue in contrast to its featureless counterpart in A. A close comparison between these experimental features and the theoretical predictions led to a better agreement if A (resp. B) was assigned to the folded (resp. extended) form. The red most band of A in the NH stretch region was assigned to the 7 MC-MC H-bond whereas the next one corresponded to the MC-SC 5^γ NH···S bond. This order is reversed in B, there the MC-SC 6^γ interaction plays the major structuring role. In support of this, one can observe that the 5^γ frequency is similar to what is observed in the capped Cys-Phe system, where the double γ -turn structure also exhibits a 5^γ feature at 3415 cm⁻¹.¹⁹ Regarding the CO stretch region, this assignment also accounts for the red shoulder observed for B. Interestingly, one will note that the nature of the CO stretch modes are very different between A and B. They are strongly coupled in B, because the two carbonyls are oriented in opposite directions (Fig. 1), but not coupled in A, where the carbonyl directions are nearly perpendicular: in this latter case, the Cys carbonyl exhibits the strongest intensity. Finally, the assignment is also supported by the populations as deduced from the UV spectrum: the major A conformer is assigned to the most stable structure of the conformational landscape.

5. Microwave spectroscopy

In the broadband rotational spectrum of Ac-Cys-NH₂ (6 to 14 GHz frequency region), each rotational transition appeared split into two components, A and E, as a result of the interaction between molecular overall and internal rotations (Fig. 4).⁵⁴ Assignments that resulted were very tricky since the number of lines was twice as many and the perturbed E substate appeared up to 120 MHz far from the A substate, for some of the transitions. Firstly, we identified the typical pattern for a-type pairs of rotational transitions $J + 1_{0,J+1} \leftarrow J_{0,J}$ and $J + 1_{1,J+1} \leftarrow J_{1,J}$, with J ranging from 5 to 11. In the employed assignment method, the measured transitions were used to improve the transition frequency predictions and search for new ones. This allowed us to identify b-type rotational transitions as well to reach a total of 95 observed rotational transitions. No c-type rotational transitions could be

Table 2 Conformer-selective IR spectroscopy: experimental frequencies of the two conformers of Z-Cys-NH₂ compared to predicted frequencies from scaled harmonic frequencies at the B97-D3/def2-TZVPPD level of theory. All frequencies are expressed in cm⁻¹

Experiment									
Conformer	Amide A			Amide I			Amide II		
Exp. A	—	—	3368	3421	3523	1725	—	1509	1594
Exp. B	—	—	3390	3425	3524	1724	1737	1493	1589
Theory									
Conformation	ΔH (kJ mol ⁻¹)	ΔG (kJ mol ⁻¹)	NH ₂ sym	NH _{Cys}	NH ₂ anti	CO _{carbamate}	CO _{Cys}	NH ₂ bend	NH bend
5 ^γ (6)–7 folded (<i>g</i> ⁺)	0.0	0.0	3373 (7)	3418 (5 ^γ)	3525	1719	1710	1604	1495
5–6 ^γ extended (<i>g</i> ⁻)	3.8	1.7	3380 (6 ^γ)	3431 (5)	3521	1719 ^a	1730 ^a	1601	1481

^a Modes strongly coupled.

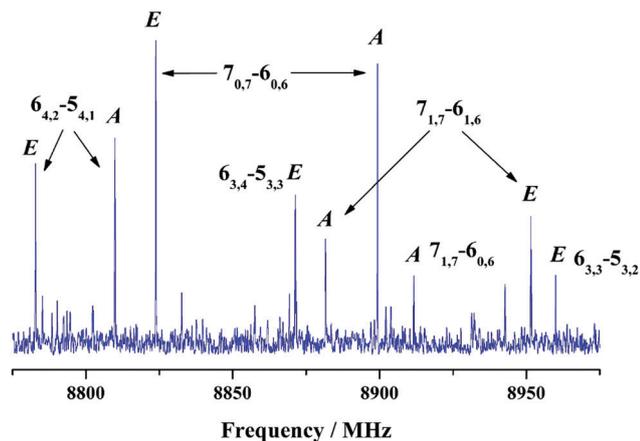


Fig. 4 A 200 MHz section of the broadband spectrum of Ac-Cys-NH₂ recorded in the 6–14 GHz frequency region. Some of the assigned rotational transitions for one rotamer are indicated. A–E internal rotation substates are shown for some of them; for the others, the corresponding components are out from this frequency range.

measured for this conformational species. Although some weak lines remained unassigned in the spectrum, identification of further conformers could not be achieved.

Ac-Cys-NH₂ has two ¹⁴N nuclei so that nuclear quadrupole coupling hyperfine splitting should be expected,⁵⁴ due to the ¹⁴N nuclei non-zero quadrupole moment ($I = 1$) owing to a non-spherical distribution of the nuclear charge. The interaction of these quadrupole moments with the electric field gradient created at the nuclei by the rest of the molecular charges causes the coupling of the nuclear spin moments to the overall rotational momentum. However, the spectral resolution attainable in the CP-FTMW experiments is not sufficient to completely resolve these hyperfine effects, which are smaller than the experimental broadening of the lines and look like the hyperfine pattern collapsed into a single feature (Fig. 4).

All the measured rotational transitions were included in a least-squares fit using the XIAM program,⁵⁵ which is based on the combined axis method introduced by Woods.⁵⁶ The analysis rendered the experimental rotational and centrifugal distortion constants and the internal rotation parameters shown in Table 3. By direct comparison between the experimental values and those predicted theoretically it is straightforward to conclude that the observed species is the 5^γ(^γ6)–7 folded conformation. In spite that both sets of rotational constants are very similar, the agreement between the experimental one with those for 5^γ(^γ6)–7 folded is excellent. The structural information reported by the analysis of the internal rotation splitting allows us to confirm this assignment. V_3 , the value of the three-fold methyl internal rotation barrier height, gives an idea of the electronic and steric environment around the methyl top, which is also reflected in the F_0 value, the rotational constant of the methyl top. On the other hand, δ and ϵ are the angles between the principal axis of the molecule and the internal rotation axis (δ) and the projection of the internal rotation axis onto xy -plane (ϵ), respectively. Hence, they are directly related to the relative position and orientation of the methyl rotor within

the molecule structure and the angles between the methyl top and the principal inertial axes can be derived from them.

Table 3 shows the structural parameters obtained for the folded and the extended forms of Ac-Cys-NH₂ obtained at the B3LYP-D3(BJ)/def2-TZVP level of theory. It is interesting to notice that the RI-B97-D3(BJ)-abc method used for optical spectroscopy provide an equivalent set of rotational constants (Table 3). Complementary calculations (Table 3 and Table S4 of the ESI†) show that neither the RI option or the three-body option of the B97D3 method does affect significantly the results. Very comparable structural features are also obtained (Table S5, ESI†). Only moderate energetic changes (stabilization of the extended form by up to 2 kJ mol⁻¹) are noticed with the RI-B97-D3(BJ)-abc.

Examination of Table 3 shows that the calculated data are significantly different between the two lowest conformations of Ac-Cys-NH₂, and that those obtained for the most stable 5^γ(^γ6)–7 folded species agree very well with all the experimental data collected, in particular, beyond the rotational constants, the height of the barrier to the methyl rotation and the orientation of the methyl rotor relative to the principal axes, eventually providing an unambiguous assignment.

The non detection of the 5–^γ6 extended species can be rationalized in terms of gas phase abundances. It is assumed that the supersonic expansion brings all conformers to their lowest vibrational state and that the intensity of the rotational transition is proportional to the number density of each conformer and the square value of the corresponding component of the dipole moment, $N_i \cdot \mu_i^2$. N_i corresponds to the molecules populating all the rotational states of a given conformer, and μ_i is the corresponding electric dipole moment component. Therefore, if we consider the ΔG values and the dipole moment components from Table 3, the estimated line intensity of 5–^γ6 relative to that of 5^γ(^γ6)–7 is in the 3–15% range depending upon the theoretical method used. Owing to the S/N ratio achieved for the 5^γ(^γ6)–7 species (Fig. 4), the signal level does not appear to be high enough to allow the observation of 5–^γ6 under our experimental conditions.

6. Discussion

6.1. Assessing the theoretical methods

The best levels of theory used (B3LYP-D3(BJ)/def2-TZVPPD and RI-B97-D3(BJ)-abc/def2-TZVPPD) lead to very comparable structures, as testified by the intramolecular H-bond distances and the covalent bond lengths, (see Table S5, ESI†), which are found to differ by less than 2 and 1 pm respectively, as well as by the relevant backbone and side chain structural dihedral angles, which differ at most by 2 degrees.

The convergent values obtained provide a relatively narrow range of values (a few pm), independently from the quality achieved on rotational constants, namely about:

199 and 212 pm for the 7 and 5 NH \cdots O bonds resp.

258 and 250 pm for the 6^γ and 5^γ NH \cdots S H-bonds resp., and

238 pm for the ^γ6 SH \cdots O H-bond.

Table 3 Comparison between the rotational experimental data and calculated spectroscopic parameters and energetics at several levels of theory for the two lowest energy conformations of Ac-Cys-NH₂, showing that the best fit is provided by the 5^γ(^γ6)–7 conformation

	Experimental	B3LYP-D3(BJ)/ def2-TZVP	B3LYP-D3(BJ)/ def2-TZVPPD	RI-B97-D3(BJ)-abc/ def2-TZVPPD	B3LYP-D3(BJ)/ def2-TZVP	B3LYP-D3(BJ)/ def2-TZVPPD	RI-B97-D3(BJ)-abc/ def2-TZVPPD
		5 ^γ (^γ 6)–7			5– ^γ 6		
A ^a	1482.6098(35) ^b	1475	1475	1453	1432	1431	1423
B	832.7096(12)	833	833	828	781	781	771
C	602.5441(11)	601	600	596	538	538	534
Δ _J	0.0279(59)	—	—	—	—	—	—
μ _a	—	1.6	1.7	1.5	1.3	1.3	1.3
μ _b	—	1.4	1.5	1.4	0.8	0.8	0.8
μ _c	—	0.3	0.2	0.2	0.3	0.3	0.3
V ₃	100.32(18)	109.6	—	—	72.3	—	—
F ₀	151.81(27)	156.174	—	—	155.212	—	—
δ	2.132411(72)	2.076	—	—	0.119	—	—
ε	2.53656(16)	2.548	—	—	2.044	—	—
∠(i, a)	122.1782(42)	118.96	—	—	6.82	—	—
∠(i, b)	61.2220(88)	60.69	—	—	93.10	—	—
∠(i, c)	134.1188 (87)	136.48	—	—	83.93	—	—
N	95	—	—	—	—	—	—
σ	45.2	—	—	—	—	—	—
ΔH (0 K)	—	0	0	0	6.3	5.8	5.0
ΔG (300 K)	—	0	0	0	3.8	3.3	1.7

^a A, B, and C represent the rotational constants (in MHz); Δ_J is the quartic centrifugal distortion constant (in kHz); μ_a, μ_b and μ_c are the electric dipole moment components (in D); V₃ is the barrier height of the threefold methyl internal rotation (in cm⁻¹); F₀ is the rotational constant of the methyl top (in GHz); δ is the angle between the principal axis of the molecule and the internal rotation axis (in radians); ε is the angle between the principal axis of the molecule and the projection of the internal rotation axis onto xy-plane (in radians); ∠(i, x) are the angles between the internal rotor axis i and the principal axis a, b, c (in degrees); N is the number of A/E components included in the fit; σ is the rms error of the fit (in kHz); ΔH and ΔG are the relative energies and Gibbs free energies (in kJ mol⁻¹) to that of 5^γ(^γ6)–7 species. ^b Values in parentheses denote 1σ errors, applied to the last digit.

In some respects the differences in terms of rotational constants found between both methods (Table 3) can be considered as an assessment of the sensitivity of the rotational constants to the structure, which remains small: the difference reaches up to 1.5% for the A constant of the folded form and for the B constant of the extended species.

If one compares the theoretical data of the folded form with experiment (Table 3), the maximum discrepancy is reached for the A constant, since the theoretical values tend to be too small, by -0.6% with B3LYP and -2% for B97D, suggesting a general feature of these levels of theory for slightly too loose structures, inasmuch as the theoretical values presented are not corrected for vibrational averaging, which should tend to further decrease the rotational constants.

6.2. H-Bonding content

The competition between folded and extended backbones, a backbone feature already documented experimentally in dipeptide analogues of apolar residues,^{9,51} is still valid in the case of Cys, even in presence of secondary main-chain/side-chain NH-S H-bonds, indicating that the intrinsic backbone features, namely the folded and extended backbones, with both their H-bonding and their specific hyperconjugative effects, still have a leading role in controlling the qualitative shape of the conformational landscape.

The quantum chemistry structures, validated by the comparison to the high resolution data, have then been used to shed light on the H-bonding, relying on both the structures and the theoretical analysis of the electron density in the framework

of the NBO theory, in particular on the ΣE_{HB} NBO H-bonding indicators.³⁵ The interatomic HB distances of the 5 and 7 H-bonds (Table 1) are comparable to those encountered with similar secondary structures with non-polar side-chains (e.g. Ala; see Table S2 ESI†). The 7 H-bond of the folded form is found to be much stronger than the weak, non-linear 5 bond, as revealed by the ΣE_{HB} indicators calculated (Table 1). The NH···S H-bonds also exhibit very different strengths as illustrated by their NBO indicators. Beyond the strong stabilization provided by the 7 H-bond (~6.5 kcal mol⁻¹), the folded form also benefits from a substantial contribution of the 5^γ H-bond (~2.6 kcal mol⁻¹), complemented by a weaker ^γ6 interaction (1.3 kcal mol⁻¹). Accordingly, this ^γ6 bond is found to be much more elongated (239 pm) than its OH···O equivalent interaction in the serine dipeptide²² (212 pm), where, in this latter case, the ^γ6 interaction plays the major stabilizing role. In contrast, the extended form benefits from a rather strong 6^γ MC → SC H-bond (~6 kcal mol⁻¹). It results from the intrinsic stability of the backbone extended form, which stems essentially from strong SC-MC hyperconjugation interactions, involving the n_N lone pair NBO of the first amide and the σ*_{CαH} and σ*_{CαCβ} NBOs,³⁵ and to a lesser extent on the weak 5 H-bond. The Cys side chain has to accommodate this strong structure and the N-Cα-Cβ-S χ₁ dihedral tends to adopt the usual antiperiplanar arrangement (χ₁ = 180°). However the van der Waals radius of the S atom being large, the NH···S H-bond is constrained resulting in a relatively short and strong H-bond, and a significant deviation from the side chain antiperiplanar arrangement (χ₁ = -162°).

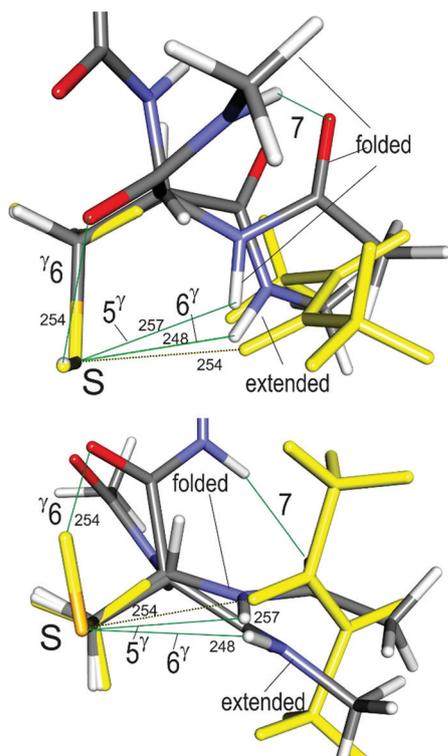


Fig. 5 Geometrical approach of the amide NH group to the Cys side chain S atom in both the folded and extended forms of the Ac-Cys-NHMe model molecule, compared to that in the methylacetamide-methanethiol intermolecular model (in yellow and dots), at the B97-D3 level of theory. For the sake of comparison the H(S), S and C^β atoms of the three structures have been overlapped in the views provided, either along the S-H axis (top panel) or along the S-C^β axis (bottom panel). H-bonds are identified by their labels (see text); H-bonding distances are given in pm.

These two NH...S interactions observed with the Cys side chain are characterized by different approaches of the NH donor to the S atom acceptor (Fig. 5), namely, for the 6^γ H-bond, a better linearity (with a larger N-H-S angle), and a shorter NH...S HB distance than its 5^γ counterpart (248 vs. 257 pm) with a strength varying accordingly (Table 1). It is then useful to compare these H-bonds to a natural reference, the NH...S interaction in the H-bonded intermolecular methylacetamide-methanethiol (see ESI[†]), whose NBO-based strength amounts to 6.3 kcal mol⁻¹ with an intermolecular distance of 254 pm. Thus the 5^γ bond appears to exhibit a H-bond distance comparable to that of the complex, for a strength being at least twice weaker. This illustrates the geometrical frustration of the 5^γ interaction, due to intrabackbone interactions, which forbids a linear H-bonding (Fig. 5). In contrast, the 6^γ H-bond is shorter than in the intermolecular bond, for a comparable strength (~6 kcal mol⁻¹), epitomizing the fact that a geometry fairly different from that of the complex can nevertheless lead to strong H-bonds in peptides, providing that a roughly linear, favorable NH...S approach can be achieved by a covalently constrained backbone. Such a conclusion was already drawn from a previous study in the context of amide-amide H-bonding.³⁵ This illustrates the crucial role of the peptide backbone, by

selecting the type of geometrical approaches that the NH...S interaction can adopt.

6.3. Spectral signatures and H-bond strength

The H-bonding strength discussed above is mainly rooted in theoretical data. IR spectroscopy confirms this view:

– Amide A (NH stretch) region. The experimental frequency associated to the 6^γ bond (3390 cm⁻¹) is red shifted relative to the 5^γ (3421 cm⁻¹). Despite the comparison is not straightforward since the NH group of the 6^γ bond belongs to a -CONH₂ carboxamide group, the vibrational coupling between both NH bonds has to be taken into account.⁵⁷ Calculations for both Ac-Cys-NH₂ and Ac-Cys-NHMe (Table S3, ESI[†]) show that the difference, for these particular H-bonded NHs, remains smaller than 3 cm⁻¹, making the 3390 cm⁻¹ frequency observed for the 6^γ bond a fair reference for comparison with the 5^γ (3421 cm⁻¹) bond.

Whether these frequency shifts can be taken as experimental indicators, in order to rank the strength of these interactions, can be assessed by comparing them to the NBO H-bond strength indicators. A previous study,³⁵ devoted to amide-amide H-bond spectroscopy, showed that such a correlation seems to exist, despite hyperconjugation effects linked to the backbone conformation tend to blur the monotonic dependence of the NH stretch frequency with the NBO indicator ΣE_{HB} . As a matter of fact the frequency pattern $7 < 6^{\gamma} < 5^{\gamma} < 5$ (Table 1) matches well the ΣE_{HB} pattern for these interactions ($6.5 < 5.7 < 2.6 < 1.9$ kcal mol⁻¹, resp.), demonstrating that the stretch frequency of H-bonded NHs remains a qualitative indicator of the H-bond strength, which can be used to compare H-bonds of different natures, even having different acceptors such as NH...O_{amide} and NH...S bonds.

– Amide I and II regions. A close examination of Table 2 shows that these regions can successfully distinguish extended and folded forms. First, the CO stretches couple differently in these two types of backbone: barely in the folded forms and strongly in the extended one, ending up with a pattern different enough for an assignment to be proposed. Second, the amide II region confirms this point, especially the NH bending region around 1500 cm⁻¹, with spectral positions which differ by more than by 10 cm⁻¹ (a value considered as significant in the amide I and II regions), and contribute to ascertain the assignment. These regions, which are not so currently used, demonstrate here their capability on a case study, where the assignment from the amide A region alone could be regarded as tentative.

– SH stretch spectroscopy. By comparison with the OH stretches of serine, one could expect invaluable assignment hints from SH stretch frequencies. However, the ^γ6 bond is a very weak SH...O interaction, in line with the long H...O distance (239 pm) compared to the 220 pm of the OH...O distance in the serine analogue, folded molecule.

As far as a direct spectral evidence is concerned, the corresponding spectral range (2600 cm⁻¹) remains seldom exploited so far, partly because of the low energy of OPOs in this region and the weak absorptions of the SH bonds.²⁵ In the present

case, indeed, the γ_6 bond is predicted at 2637 cm^{-1} (unscaled harmonic value), with a weak intensity (11 km mol^{-1}). This is to be compared with the free SH of the extended form predicted at 2634 cm^{-1} (unscaled harmonic value), and at a still weaker level of absorption (0.37 km mol^{-1}), making it very difficult to measure. Moreover, one should mention that the small, and barely significant, difference in frequency between bound and free SHs, does not seem to qualify the SH frequency shift as a clear marker of H-bonded SHs. This is in line with the expectations from a good H-bond acceptor and poor donor like the weakly polarized but strongly polarizable $-\text{SH}$ group.

6.4. Conformational landscape and side chain/main chain interactions

So far, as a part of a protein chain model, Cys was investigated in capped dipeptides.¹⁹ Owing to the relatively weak nature of the MC/SC interactions, the backbone essentially forced the dipeptide structure (7–7 or 10 backbone), and only those MC/SC interactions compatible with it could be observed, namely essentially 5^γ and weak γ_6 H-bonds. In the present case, the same basic trends apply but the natural backbone forms are the 7 and 5 structures, resulting in the observation of the 5^γ and γ_6 H-bonds as before, but also, and for the first time, of the 6^γ H-bond, compatible with the 5 extended backbone. Interestingly, and in line with this finding, recent data from Saclay group show that substitution of the sulfhydryl H atom by a methyl group or the inclusion of the S atom in a thietane ring in a γ -position leads to the exclusive observation of the extended form, stabilized by the same γ_6 H-bonding.⁵⁸

7. Conclusion

The present gas phase study combines microwave and optical spectroscopies in order to characterize the intrinsic bridging features of a cysteine side chain to its local backbone environment. Optical spectroscopy is used to sort out the stable conformations present in a supersonic expansion according to their H-bonding content. Microwave spectroscopy combined with theoretical structures enables us to go one step further in terms of structural properties, by validating the H-bonding distances provided by the quantum chemistry. It also provides a stringent test for the theoretical methods used. The present report illustrates the capacity of the weakly polar $-\text{CH}_2-\text{SH}$ side chain of Cys to adapt itself to the intrinsic local preferences of the peptide backbone, *i.e.*, γ -turn folded or β -sheet extended secondary structure, by establishing a local H-bonding, where the S atom of the Cys side chain, which essentially plays the role of a proton acceptor, is either weakly bound to the NH donor site of the same residue (5^γ bond with a folded backbone) or more strongly bound to the NH of the next residue along the chain (γ_6 bond in an extended backbone).

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

Support from the French National Research Agency (ANR; Grant ANR-17-CE29-0008 “TUNIFOLD-S”) and from the “Investissements d’Avenir” Funding program (LabEx PALM; grant ANR-10-LABX-0039-PALM; DIRCOS) are acknowledged. This work was granted access to the HPC facility of [TGCC/CINES/IDRIS] under the Grant 2019-A0050807540 awarded by GENCI (Grand Equipement National de Calcul Intensif) and to the CCRT High Performance Computing (HPC) facility at CEA under the Grant CCRT2019-p606bren. We also acknowledge the use of the computing facility cluster MésolUM of the LUMAT federation (FR LUMAT 2764). The financial fundings from Ministerio de Ciencia e Innovación (CTQ2016-76393-P), Junta de Castilla y León (Grant VA077U16) and the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013)/ERC-2013-SyG, Grant Agreement no. 610256 NANOCOSMOS, are gratefully acknowledged.

References

- 1 T. S. Zwier, *J. Phys. Chem. A*, 2001, **105**, 8827–8839.
- 2 M. S. de Vries and P. Hobza, *Annu. Rev. Phys. Chem.*, 2007, **58**, 585–612.
- 3 E. Gloaguen and M. Mons, in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, ed. A. M. Rijs and J. Oomens, 2015, vol. 364, pp. 225–270.
- 4 K. Schwing and M. Gerhards, *Int. Rev. Phys. Chem.*, 2016, **35**, 569–677.
- 5 A. M. Rijs and J. Oomens, in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, ed. A. M. Rijs and J. Oomens, 2015, vol. 364, pp. 1–42.
- 6 J. L. Alonso and J. C. Lopez, *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, 2015, vol. 364, pp. 335–401.
- 7 S. Kumar, S. K. Singh, C. Calabrese, A. Maris, S. Melandri and A. Das, *Phys. Chem. Chem. Phys.*, 2014, **16**, 17163–17171.
- 8 R. J. Lavrich, D. F. Plusquellic, R. D. Suenram, G. T. Fraser, A. R. H. Walker and M. J. Tubergen, *J. Chem. Phys.*, 2003, **118**, 1253–1265.
- 9 C. Cabezas, M. Varela, V. Cortijo, A. I. Jimenez, I. Pena, A. M. Daly, J. C. Lopez, C. Cativiela and J. L. Alonso, *Phys. Chem. Chem. Phys.*, 2013, **15**, 2580–2585.
- 10 E. Gloaguen, B. Tardivel and M. Mons, *Struct. Chem.*, 2016, **27**, 225–230.
- 11 C. Cabezas, M. Varela and J. L. Alonso, *ChemPhysChem*, 2013, **14**, 2539–2543.
- 12 I. Compagnon, J. Oomens, J. Bakker, G. Meijer and G. von Helden, *Phys. Chem. Chem. Phys.*, 2005, **7**, 13–15.
- 13 K. N. Blodgett, J. L. Fischer, J. Lee, S. H. Choi and T. S. Zwier, *J. Phys. Chem. A*, 2018, **122**, 8762–8775.
- 14 S. Habka, W. Y. Sohn, V. Vaquero-Vara, M. Geleoc, B. Tardivel, V. Brenner, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3411–3423.
- 15 P. S. Walsh, K. N. Blodgett, C. McBurney, S. H. Gellman and T. S. Zwier, *Angew. Chem. Int. Ed.*, 2016, **55**, 14618–14622.

- 16 P. S. Walsh, J. C. Dean, C. McBurney, H. Kang, S. H. Gellman and T. S. Zwier, *Phys. Chem. Chem. Phys.*, 2016, **18**, 11306–11322.
- 17 W. Y. Sohn, S. Habka, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2017, **19**, 17128–17142.
- 18 B. Yan, S. Jaecq, W. J. van der Zande and A. M. Rijs, *Phys. Chem. Chem. Phys.*, 2014, **16**, 10770–10778.
- 19 M. Alauddin, H. S. Biswal, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2169–2178.
- 20 H. S. Biswal, E. Gloaguen, Y. Loquais, B. Tardivel and M. Mons, *J. Chem. Phys. Lett.*, 2012, **3**, 755–759.
- 21 G. Goldsztejn, V. R. Mundlapati, J. Donon, B. Tardivel, E. Gloaguen, V. Brenner and M. Mons, *Phys. Chem. Chem. Phys.*, 2020, DOI: 10.1039/D0CP02825H.
- 22 C. Cabezas, M. A. T. Robben, A. M. Rijs, I. Pena and J. L. Alonso, *Phys. Chem. Chem. Phys.*, 2015, **17**, 20274–20280.
- 23 I. Leon, E. R. Alonso, S. Mata, C. Cabezas, M. A. Rodriguez, J. U. Grabow and J. L. Alonso, *Phys. Chem. Chem. Phys.*, 2017, **19**, 24985–24990.
- 24 E. Gloaguen, Y. Loquais, J. A. Thomas, D. W. Pratt and M. Mons, *J. Chem. Phys. B*, 2013, **117**, 4945–4955.
- 25 I. A. Lobo, P. A. Robertson, L. Villani, D. J. D. Wilson and E. G. Robertson, *J. Phys. Chem. A*, 2018, **122**, 7171–7180.
- 26 S. Mata, I. Pena, C. Cabezas, J. C. Lopez and J. L. Alonso, *J. Mol. Spectrosc.*, 2012, **280**, 91–96.
- 27 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
- 28 TURBOMOLE V7.2 2017, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.
- 29 D. Rappoport and F. Furche, *J. Chem. Phys.*, 2010, **133**, 134105.
- 30 A. Schafer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829–5835.
- 31 M. Sierka, A. Hogekamp and R. Ahlrichs, *J. Chem. Phys.*, 2003, **118**, 9136–9148.
- 32 K. Eichkorn, O. Treutler, H. Ohm, M. Haser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **242**, 652–660.
- 33 K. Eichkorn, O. Treutler, H. Ohm, M. Haser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **240**, 283–289.
- 34 K. Eichkorn, F. Weigend, O. Treutler and R. Ahlrichs, *Theor. Chem. Acc.*, 1997, **97**, 119–124.
- 35 V. Brenner, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, 2019, **21**, 24601–24619.
- 36 Y. Bouteiller, J.-C. Gillet, G. Grégoire and J. P. Schermann, *J. Phys. Chem. A*, 2008, **112**, 11656–11660.
- 37 H. S. Biswal, Y. Loquais, B. Tardivel, E. Gloaguen and M. Mons, *J. Am. Chem. Soc.*, 2011, **133**, 3931–3942.
- 38 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. R. Gomperts, B. 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, N. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, J. M. 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 and D. J. Fox, *Gaussian 16, Revision A.03*, 2016.
- 39 M. J. Frisch, J. A. Pople and J. S. Binkley, *J. Chem. Phys.*, 1984, **80**, 3265–3269.
- 40 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.
- 41 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 42 Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 289–300.
- 43 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 44 I. V. Alabugin, G. dos Pasos Gomes and M. A. Abdo, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2019, **9**, e1389.
- 45 E. D. Glendening, C. R. Landis and F. Weinhold, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 1–42.
- 46 A. E. Reed, L. A. Curtiss and F. Weinhold, *Chem. Rev.*, 1988, **88**, 899–926.
- 47 F. Weinhold, A. E. Reed, J. E. Carpenter and E. D. Glendening, NBO version 3.1., 2003.
- 48 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, J. J. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Rev. D01*, 2009.
- 49 E. Gloaguen, V. Brenner, M. Alauddin, B. Tardivel, M. Mons, A. Zehnacker-Rentien, V. Declerck and D. J. Aitken, *Angew. Chem., Int. Ed.*, 2014, **53**, 13756–13759.
- 50 U. Adhikari and S. Scheiner, *J. Phys. Chem. A*, 2013, **117**, 489–496.
- 51 C. M. Leavitt, K. B. Moore, P. L. Raston, J. Agarwal, G. H. Moody, C. C. Shirley, H. F. Schaefer and G. E. Doublerly, *J. Phys. Chem. A*, 2014, **118**, 9692–9700.
- 52 J. Gonzalez, R. Martinez, J. A. Fernandez and J. Millan, *Eur. Phys. J. D*, 2017, **71**, 203.

- 53 Y. Loquais, E. Gloaguen, S. Habka, V. Vaquero-Vara, V. Brenner, B. Tardivel and M. Mons, *J. Phys. Chem. A*, 2015, **119**, 5932–5941.
- 54 W. Gordy and R. L. Cook, *Microwave Molecular Spectra*, John Wiley Sons, New-York, 1984.
- 55 H. Hartwig and H. Dreizler, *Zeitschrift Fur Naturforschung Section a–a, J. Phys. Sci.*, 1996, **51**, 923–932.
- 56 R. C. Woods, *J. Mol. Spectrosc.*, 1966, **21**, 4–24.
- 57 W. Chin, F. Piuzzi, I. Dimicoli and M. Mons, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1033–1048.
- 58 Z. Imani, V. R. Mundlapati, G. Goldsztejn, V. Brenner, E. Gloaguen, R. Guillot, J.-P. Baltaze, K. Le Barbu-Debus, S. Robin, A. Zehnacker, M. Mons and D. J. Aitken, *Chem. Sci.*, 2020, DOI: 10.1039/D0SC03339A.