

Vibrational spectroscopy of deprotonated peptides containing an acidic side chain

Edith Nicol, Carine Clavaguéra, Gilles Ohanessian

▶ **To cite this version:**

Edith Nicol, Carine Clavaguéra, Gilles Ohanessian. Vibrational spectroscopy of deprotonated peptides containing an acidic side chain. International Journal of Mass Spectrometry, 2019, 435 (8), pp.42 - 50. 10.1016/j.ijms.2018.10.006 . hal-02161605

HAL Id: hal-02161605
https://hal.science/hal-02161605

Submitted on 21 Jun 2019

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.

Vibrational spectroscopy of deprotonated peptides containing an acidic side chain.

Edith Nicol,[1] Carine Clavaguéra,[2] Gilles Ohanessian[1,*]

[1] LCM, CNRS, Ecole Polytechnique, 91128 Palaiseau Cedex, France

[2] Laboratoire de Chimie Physique, CNRS, Université Paris Sud, Université Paris Saclay, 91405 Orsay Cedex, France

Abstract

The vibrational properties of four deprotonated dipeptides and three tripeptides containing an acidic residue, either Asp or Glu, have been studied using InfraRed Multiple Photon Dissociation (IRMPD) action spectroscopy in a mass spectrometer. It is found that these spectra fall into two broad categories, including clearly different spectra for sequence-reversed pairs of dipeptides. The observed bands were assigned using Density Functional Theory (DFT) calculations. Aside from the bands pertaining to the C=O stretching modes which all lie in the expected 1550-1750 cm^{-1} range, photon absorption is strongly influenced by the hydrogen bonding patterns, which differ according to whether the acidic residue is located at the C-terminus or not. This leads to two distinct frequency ranges of absorption in between 1250 and 1550 cm^{-1}. In addition, unexpectedly wide absorption around 1600 cm^{-1} may be attributed to non-classical proton sharing between the two carboxylates when Asp is at the C-terminus.

Footnote: dedicated to Helmut Schwarz in recognition of his immense contribution to chemical science in general and to mass spectrometry in particular.

Keywords: IRMPD spectroscopy, DFT calculations, deprotonated peptides, proton sharing, aspartate, glutamate

* Corresponding author
E-mail address: gilles.ohanessian@polytechnique.edu

I. Introduction

Isolated peptides in the gas phase have been much studied, especially with regard to the potential of mass spectrometry to generate both precise identification and extensive sequence data for peptides and proteins. Yet the majority of studies have been carried out on protonated (or multi-protonated) species. When multiple acidic sites are present, working in negative ion mode can however be especially efficient. It has been shown that deprotonated peptides can also be used to obtain useful sequence data, which are often complementary to those gleaned from positive ions [1-4]. A prominent feature of these results is that acidic residues, i. e. Asp and Glu, generate specific fragmentation schemes which can be integrated in sequencing strategies.

The above studies have all used collisional activation to generate fragments, including MS^3 experiments. Photodissociation has also proven useful, using either UV photons - UVPD [5] or IR photons – IRPD at low temperature or IRMPD at room temperature [6, 7a]. While UVPD has been used to provide sequence information (with rich radical chemistry induced by photon absorption), IRPD and IRMPD can provide valuable information about the 3D structures after confrontation to computed spectra. Initial studies by Oomens and Steill [8] have established reference vibrational frequencies for gaseous acetate, propionate and other carboxylates, and related them to their analogues in the liquid phase. Noticeably, low-temperature IRPD spectra have recently been obtained for these species, yielding slightly larger frequencies for both the symmetric and antisymmetric carboxylate stretching motions [9]. In further studies, IRMPD spectra have been obtained for a number of deprotonated amino acids [10], and for deprotonated trialanine and its a_3 fragment ion [6]. Following this successful attempt, IRMPD spectroscopy has been extensively used to characterize the structure of a-, b- and c-type fragments of deprotonated peptides [11]. The structuration of model peptides around a C-terminus carboxylate has also been studied in an IRPD study of deprotonated glycine oligomers, up to the tetramer [7a], as well as their Cu(II) complexes [7b].

The IRMPD spectra reported for amino acids [10] have shown that deprotonated Asp and Glu, $[Asp-H]^-$ and $[Glu-H]^-$, display absorption features that are completely different from those of all others. This has been interpreted as arising from non-classical proton sharing between two carboxylates. This can be either due to the proton being trapped between the two basic sites (instead of one covalent and one hydrogen bond) or to fast proton flipping between the two sites. Static quantum chemistry calculations of IR spectra such as those commonly carried out to interpret IR spectra, cannot provide clues to such behavior. We have shown that these spectra can be reproduced computationally when proton dynamics is properly taken into account [12]. Analogous perturbation was not observed for deprotonated Cys, [10] although previous work based on other experimental techniques [13-15] did suggest that a proton could be shared between the thiolate and the carboxylate. $[Cys-H]^-$ has been further characterized by low temperature IRPD [16], showing that non-classical proton sharing is likely to occur at least in cryogenic conditions. A similar trend was noted [17] for $[CysSO_2-H]^-$ with a shared-proton, as opposed to $[CysSO-H]^-$ and $[CysSO_3-H]^-$ for which it is localized either on the oxidized side chain or on the carboxylate. This non-classical behavior prompted us to undertake IRMPD measurements on deprotonated Asp [12] and Glu, and on a series of deprotonated carboxylic diacids [18]. We found that in most cases, a relatively large band centered near 1600 cm^{-1}, together with the absence of usual bands attributable to carboxylate and carboxylic acid moieties, is a strong indication of non-classical proton sharing.

In order to assess the modes of local interaction of a carboxylate when a carboxylic acid lies in its vicinity, a set of model peptides was studied by room-temperature IRMPD spectroscopy and quantum chemical calculations. The deprotonated dipeptides $[AspGly-H]^-$, named "DG" below, $[GlyAsp-H]^-$ (named GD), $[GluGly-H]^-$ (EG) and $[GlyGlu-H]^-$ (GE) are simple systems, yet allowing structural diversity in hydrogen-bonding opportunities. Slightly larger cases were also investigated, namely $[AspGlyGly-H]^-$ (DGG), $[GlyGlyAsp-H]^-$ (GGD) and $[GluGlyGly-H]^-$ (EGG), in which higher flexibility may lead to either stronger or more numerous hydrogen bonds or both. Among the tripeptides, calculations were done on DGG and GGD only.

II. Experimental & computational details

IRMPD spectra were recorded at the CLIO (Centre Laser Infrarouge d'Orsay, France) facility [19] using a free-electron laser (FEL) coupled to a 7T FT-ICR mass spectrometer. Samples were dissolved in H_2O/CH_3CN 50:50 mixtures, at a concentration of 10^{-5} M. The ions were formed by ESI negative mode in standard conditions and transferred to the ICR cell at room temperature. Mass-selected ions were stored in the cell and submitted to irradiation by IR photons during 0.15–0.5 s depending upon laser power, most often in the 0.5–1 W range. Attenuators were sometimes used to avoid absorption saturation. Absorption-induced fragmentation was monitored as a function of laser wavelength in the 1800–1250 cm^{-1} range.

The most abundant fragmentation is water elimination in all cases except GGD. For GD, DG and DGG, consecutive loss of ammonia is also observed, while secondary fragmentations are elimination of Gly from GE and of carbon dioxide from EG. Finally, elimination of Asp is seen for GD and is the exclusive fragmentation of GGD. This loss of Asp when it is the C-terminus residue has been observed and discussed before [2, 4b]. Comparison of these fragmentations with those induced by collision-induced dissociation (CID) [1,2,4] suggests that in the present IRMPD conditions, the amount of energy deposited in the ions before fragmentation is at the low end of those attained in the above CID experiments.

The IR spectra were obtained by representing the fragmentation efficiency, F_{eff}, as a function of wavelength. F_{eff} is defined as $F_{eff} = -\ln[I_p/(I_p+\Sigma I_{frag})]$, where I_p and I_{frag} are the parent and fragment ion intensities, respectively. The experimentally recorded IRMPD spectra are shown in Figure 1 in the 1150-1850 cm^{-1} range. In a few cases, spectra were recorded with a slightly more extended frequency window, however without any significant additional absorption.

Quantum chemical calculations were carried out on manually selected structures, with two issues in mind: deprotonation may occur on different sites, generating several isomers, and for each isomer, rotation around single bonds is expected to generate a manifold of conformers. Deprotonation was effected on either the C-terminus or on the Asp or Glu side chain. In four out of the six peptides on which computations were done, low energy structures were found for both deprotonated sites, C-terminus and side chain. Several attempts were also made at deprotonating the amide in DG, generating amidates with various hydrogen-bonding patterns for stabilization. In agreement with Oomens and Steill on $[A_3-H]^-$ [6], we have found that such isomers are much less stable, by ca. 20 $kJ.mol^{-1}$ or more. Thus they will not be described below. Hydrogen bonding interactions between the carboxylic and carboxylate groups were expected to bring about a large stabilization, thus they were present in a majority (ca. two thirds) of the structures investigated. A number of structures in which the carboxylate interacts with the terminal amine or with a peptidic N-H bond were also considered. It turned out that structures lying in the lowest 10 $kJ.mol^{-1}$ window have a carboxylic-carboxylate interaction, for all

peptides. Proton transfer to the other carboxylate was found to be unfavorable in all cases studied and no secondary minimum could be located, because the resulting carboxylate would be stabilized by only one non-covalent interaction instead of two. This can be seen easily for the lowest energy structures in Figures 2-7. As an example of the results, the six structures obtained in a 20 $kJ.mol^{-1}$ window for DG are shown in Figure S1 and their relative energies are listed in Table S1. All relative energies mentioned include the zero-point vibrational energy correction.

IR spectra were computed in the harmonic approximation, at the Density Functional Theory (DFT) level using the ωB97X-D density functional and the aug-cc-pVTZ basis set. Test computations were carried out on the low-energy subset of DG structures using three other functionals: B3LYP, cam-B3LYP and M11, with the same aug-cc-pVTZ basis set. There was no qualitative difference between the relative energies of isomers/conformers with the various functionals, i. e. the energetic ordering was the same for all four functionals, although significant differences in relative energies exist for some structures. In particular, all functionals yielded the same lowest energy structure(s) which is used to assign the experimentally observed bands in the next section. The results of this exploratory study are gathered in Table S1 and Figure S1 for the six lowest energy structures of DG. The ωB97X-D/aug-cc-pVTZ computed frequencies were scaled by 0.95, following the recommendation of Truhlar et al. for fundamental frequencies [20]. Band intensities were convoluted by a Lorentzian profile of 10 cm^{-1} FWHM. It is likely that a larger FWHM would be necessary to better account for band broadening due to intrinsic linewidth of the CLIO light source, room temperature and possible mixtures of conformations, however this would make it difficult to identify some of the bands described in the text when they have similar computed intensities. The spectra computed for the lowest energy structure of DG, with the same series of four density functionals and the wB97X-D basis set, are shown in Figure S2. They display rather limited differences in band positions and intensities. All DFT computations were done using the Gaussian09 software package [21].

III. Results

IRMPD spectra acquired in the conditions described above are all shown in Figure 1. It is immediately clear that all seven peptides absorb IR photons in the 1550-1750 cm^{-1} range, although their spectra display somewhat different features. This is expected as the C=O stretching frequency of an isolated aliphatic acid [22], that of a free C=O in a peptide bond (amide I) and that of the antisymmetric stretching frequency of a free carboxylate [8] lie near 1750, 1680 and 1600 cm^{-1}, respectively. Analysis of the absorption frequencies for each case may provide clues as to the structural features. This is strongly helped by comparison with literature results on similar cases ; in this case, deprotonated oligoglycines [7a] and deprotonated dimers of amino acids [23]. A second conclusion from inspection of Figure 1 is that the peptides fall into two groups, those having strong absorption in the 1450-1550 cm^{-1} range (blue box in Figure 1) and little at smaller frequencies (orange box in Figure 1), i.e. GD, GE and GGD, and those for which significant to strong absorption is seen in the 1250-1450 cm^{-1} range, with little or none in the 1450-1550 cm^{-1} range (DG, EG, DGG and EGG). Significant sequence effects are seen since e.g. DG and GD have clearly different spectra. Since symmetric stretch of a free carboxylate has been observed at 1300-1340 cm^{-1} and amide II bands are expected near 1500 cm^{-1}, it must be that intramolecular

interactions lead to significant frequency shifts. The latter should thus be useful clues about peptide structures.

We next describe the IRMPD spectrum for each peptide and assign the observed bands, based on DFT computed IR spectra. Out of the several structures obtained for each peptide, we only describe below and in Figures 2-7, the one of lowest energy and the second lowest if its relative energy is very small and its computed spectrum significantly different than that of the lowest structure. This leads to a single structure for GD, GE and EG, and two for GGD, DG and DGG. A more extended sample of structures is shown for DG in Figure S1. In several cases, low energy structures were found with computed spectra very similar to that of the lowest energy structure. Others probably exist because of non-exhaustive conformational search. Non negligible population of such structures at room temperature is expected to lead to IR band broadening. Hence computations help identifying the favorable families of conformations, not necessarily the most favorable individual conformers.

GD

The lowest energy structure of GD (GD-1, see Figure 2), has a direct interaction between the deprotonated C-terminus and the side chain acid in a nearly coplanar way. It has the N-terminus amine NH bonds pointing towards the peptidic C=O. The next lowest energy structure is 11.5 kJ.mol^{-1} higher in energy but only differs by the orientation of the N-terminus, which points towards the peptidic N-H, yielding a very similar IR spectrum. Thus only the lowest energy structure will be discussed. The large experimental absorption between 1550 and 1720 cm^{-1} is moderately well reproduced by three bands at 1725 (acid C=O stretch), 1667 (peptidic C=O stretch) and 1628 cm^{-1} (asymmetric carboxylate stretch). The experimental width is insufficiently covered by these three bands, suggesting either large anharmonicity effects or even non classical proton sharing, as has been previously observed for deprotonated Asp [10, 12]. The carboxylate environment in GD-1 is indeed very similar to that in D, with only an amine N-H replaced by a peptidic N-H. The intense band centered near 1500 cm^{-1} is attributable to the COH bend at its blue side (1535 cm^{-1}), and to the amide II mode of the peptidic bond at its red side (1466 cm^{-1}). The symmetric stretching motion of the carboxylate generates a very weak band at 1355 cm^{-1}.

GE

The experimental spectrum of GE (GE-1, see Figure 3) is very similar to that of GD, with the only difference that the experimental absorption in the 1600-1720 cm^{-1} range is much narrower. The lowest two energy structures obtained for GE are very similar to those described above for GD, thus only the lowest one will be briefly described. The three bands associated with C=O stretches are now in much better agreement with experiment. This is likely due to the different relative orientation of the acid and carboxylate moieties in GE, where they are no longer coplanar because of the larger flexibility of the Glu side chain. This apparently restores a classical proton sharing between the two carboxylates. The band centered near 1500 cm^{-1} is assigned in the same way as for GD, except that the weaker interaction of the acid O-H with the carboxylate red-shifts its frequency to 1500 cm^{-1} while the amide II motion is nearly unchanged.

GGD

The two most stable structures GGD-1 and GGD-2 (see Figure 4), are nearly degenerate with an energy difference of only 1.0 kJ.mol^{-1} at the ωB97X-D/aug-cc-pVTZ level (-6.4 kJ.mol^{-1} for the relative free energies at 298 K). They illustrate in a particularly acute way the competition that may occur between folded and unfolded structure stability. GGD–1 is strongly folded with direct interaction of the side chain carboxylate with both peptide termini and one of the peptidic N-H, the second being relatively close as well. The peptide chain bears a γ-turn or C_7 interaction holding the structure tight. On the opposite, GGD-2 is as extended as it can be, with only local interactions at each terminus. This expectedly leads to some significant differences in their IR spectra, even though the overall absorption regions are very similar.

As expected, the computed spectrum of GGD-2 (see Figure 4) is very close to those of GD-1 and GE-1 discussed above, with the C-terminus being deprotonated and very similar local interactions. The presence of two peptidic linkages gives rise to two amide I bands at 1679 and 1665 cm^{-1} and, more visibly, to two strongly coupled amide II modes yielding one intense combination at 1480 cm^{-1} and a weak one at 1501 cm^{-1}.

As discussed above, the structure of GGD-1 is largely different with strong folding and several interactions of the deprotonated side chain carboxylate receiving hydrogen bonds from the C-terminus acid, the N-terminus amine and one of the peptidic N-H's. Because of this sharing, the acid-carboxylate interaction is slightly weaker in GGD-1 than it is in GGD-2. This leads to a slightly higher stretching frequency for the acid C=O at 1741 cm^{-1}. The two amide I frequencies are slightly different (1671 and 1698 cm^{-1}) because one of the C=O's is free while the other has a weak interaction with an N-terminus N-H bond. An intense band computed at 1604 cm^{-1}, accompanied by a weaker one at 1597 cm^{-1}, corresponds to the couplings of the carboxylate antisymmetric stretching and NH_2 scissoring motions. It is the favorable relative orientation of these two groups which creates intensity enhancement. Another doublet of bands is found at 1529 and 1520 cm^{-1}, from combination of the NH_2 scissoring and amide II of the G-G linkage. Because of their orientation, their in-phase coupling leads to significant dipole moment change. The second amide II mode appears at 1481 cm^{-1}, essentially uncoupled and as such, of intensity that is in-between those of the previous two bands. Finally, the carboxylate symmetric stretch gives rise to a small band at 1356 cm^{-1}.

The simultaneous presence of these two nearly degenerate conformers is consistent with the experimental spectrum.

DG

Two nearly degenerate structures were found, within 1 kJ.mol^{-1} of each other (8.0 kJ.mol^{-1} for the relative free energies at 298 K). Their global structures (DG-1, DG-2, see Figure 5) are very similar, with a direct interaction of the deprotonated side chain of Asp with both the acid at the C-terminus of Gly in its trans conformation, and the N-terminus amine. In DG-1, both termini donate a hydrogen bond to the same carboxylate oxygen, leaving the second carboxylate oxygen free. In DG-2 the two hydrogen-bond

donors interact each with one of the carboxylate oxygens. Expectedly, the bands in the 1680-1720 cm^{-1} range correspond to the C=O stretches of the C-terminus acid and of the peptidic amide and are very close in both computed spectra. In marked contrast, bands in the 1550-1650 cm^{-1} range are rather different. For DG-1, an intense band at 1634 cm^{-1} corresponds mostly to the free C-O stretch of the carboxylate, while the NH_2 bending motion is only weakly excited at 1546 cm^{-1}. For DG-2, the two relatively intense bands at 1610 and 1563 cm^{-1} correspond to combinations of the NH_2 bending and carboxylate out-of-phase stretching motions. In each case, the most intense bands correspond to those with favorable relative orientations of the functional groups, i. e. leading to larger dipole moment changes upon motion. In DG-1, the carboxylate and acid groups are nearly coplanar, while they are roughly perpendicular in DG-2. Furthermore in DG-2, the N-terminus NH_2 and the carboxylate are in nearly parallel planes.

In both structures, the amide II bands arise near 1500 cm^{-1}, consistent with similar environments involving an interaction with the N-terminus. In the 1250-1350 cm^{-1} region, the COH bend is strongly coupled to the C-OH stretch and the symmetric carboxylate stretch for DG-2. The latter two also appear in this region for DG-1, however the COH bend is not coupled, thus appearing farther to the red, at 1174 cm^{-1}, where experimental absorption is very weak. Again this is due to the acid and carboxylate groups being nearly coplanar in DG-1, while they lie in nearly perpendicular planes in DG-2.

Clearly better overall agreement between experiments and calculations is met for DG-2, yet the presence of some amount of DG-1 cannot be ruled out.

EG

The most stable structure of EG (EG-1, see Figure 6) has non-bonded interactions very similar to those in DG-2, with the N- and C-termini donating hydrogen bonds to different carboxylate oxygens. Thus the computed spectra are rather similar (see Figures 5-bottom and 6) with only small differences for a number of bands. As an example, bands just above and below 1600 cm^{-1} are combinations of the carboxylate antisymmetric stretch and a N-terminus mode, scissoring in EG-1 and bending in DG-2. This difference, due to the relative orientations of the two groups, leads to a slightly different pair of frequencies (1563/1610 cm^{-1} in DG-2 vs. 1582/1612 cm^{-1} in EG-1). It is however only in the 1200-1300 cm^{-1} region that clear frequency differences can be seen between the spectra of EG-1 and DG-2. In particular at 1250-1260 cm^{-1} the spectrum of EG-1 displays three bands, accounting together for the large intensity visible in Figure 6. The most intense is dominated by the COH bend, red-shifted as it is stabilized by interaction with the nearby peptidic N-H. The other two have large components on the out of plane C-OH torsion coupled to the amide II band, as these two deformations can occur in parallel directions. Thus it is the presence of the peptidic N-H bond near the acid-carboxylate part which leads to an overall red-shift of this absorption. The convolution used to account roughly for temperature effects leads to an artificially large intensity at 1250-1260 cm^{-1} in the computed spectrum in Figure 6 while there are actually three low intensity bands computed in this range.

DGG

Absorption occurs in the same frequency regions as for DG and EG. There are now two amide I and two amide II bands, leading to more composite features in the 1500-1750 cm^{-1} range. The other significant difference is a larger intensity in the 1250-1400 cm^{-1} region, probably due to a drop in laser power in the higher frequency part of the spectrum. However, laser power was not monitored often enough to enable correction for peak intensity ratios.

Two low energy structures were obtained (DGG-1 and DGG-2, see Figure 7). Since the acidic side chain and C-terminus are now distant in the sequence, both structures are strongly folded in order to allow for several ionic hydrogen bonds to the carboxylate. Both structures are deprotonated at the Asp side chain with the carboxylate stabilized by strong interactions with the C-terminus, N-terminus and the peptidic N-H bond next to the C-terminus. The main difference, as described above for DG, is that in DGG-1, all three interactions are with the same carboxylate oxygen while both are involved in DGG-2.

As can be expected from the structures, the acid and amide I C=O stretching frequencies are very similar in both spectra. The two amide I frequencies are close and appear as a single band in both convoluted spectra of Figure 7 (at 1670 and 1680 cm^{-1}). The somewhat different arrangements of hydrogen bonds to the carboxylate group and relative orientations of polar groups generate some frequency shifts as for the previously described peptides. The stretching motions of the two carboxylate C-O groups are largely uncoupled in DGG-1 because one of them is free, while the usual coupling is found in DGG-2. The C-O stretch is coupled to the amide II mode next to the C-terminus in DGG-1, while it is coupled to the NH_2 scissoring in DGG-2. The resulting doublets are at 1648 and 1557 cm^{-1} in DGG-1, while significantly closer at 1622 and 1571 cm^{-1} in DGG-2. These coupling differences carry on to the amide II and NH_2 scissoring frequencies in the 1530-1570 cm^{-1} range. The lower frequency part of the spectrum undergoes some perturbations as well, with the partly uncoupled C-O stretching appearing at 1323 cm^{-1} in DGG-1, while the symmetrically coupled stretch is at 1356 cm^{-1} in DGG-2. The recovery of the experimental absorption range is satisfactory for both structures in this region.

IV. Discussion

The comparison of experimental IRMPD and computed IR spectra in the previous section showed that for most bands, reasonably good agreement is obtained and band assignments could thus be proposed with confidence. Based on these results, it is now possible to review the main variations in band frequencies and intensities, in order to understand why the spectra in Figure 1 fall into two clearly distinct groups. This discussion can be aided by comparison to frequencies of analogous vibrational modes in simple model systems.

Use of simple computed models

Reference values have been computed for simple models, at the same level of theory as for the peptides described above and using the same frequency scaling factor. For free acetate $CH_3CO_2^-$, the symmetric and antisymmetric stretching frequencies are computed to be 1321 and 1602 cm^{-1}, respectively, to be compared to the low temperature IRPD band maxima of 1335 and 1608 cm^{-1} [9] and to the room

temperature band maxima of IR electron detachment spectroscopy of 1305 and 1590 cm^{-1} [8]. For free acetic acid in its most stable cis conformation, the C=O stretch, C-OH stretch and COH bend frequencies are computed to be 1761, 1286 and 1164 cm^{-1}, to be compared to room temperature FTIR band maxima of 1788, 1280 and 1178 cm^{-1}, respectively [22]. The good agreement between experimental and computed values allows us to use analogous simple models, acetic acid in its trans conformation, both free and hydrogen-bonded to an acetate anion, to provide reference values to which the frequencies described previously can be compared for discussion. Based on the discussion in the previous section, we focus below on those vibrational motions whose frequencies are strongly sensitive to their environments. The acid and amide CO stretching frequencies appear to be relatively constant over the whole series of spectra and as such, not carrying very useful information about peptide structure.

COH bend

For free acetic acid in its trans conformation, the COH bending mode is strongly coupled with the C-OH stretching mode, with a frequency of 1255 cm^{-1}. It is blue-shifted to 1443 cm^{-1} when the trans acid is a hydrogen-bond donor to an acetate, because this motion weakens the hydrogen bond. This intermolecular interaction resembles the one within DG, EG and DGG, which is generated by main chain folding, thus a blue shift may be also expected in these cases. However, in these peptides, an additional interaction exists since COH bending moves the O-H bond towards the peptidic nitrogen lone pair, leading to significant stabilization and limiting the blue shift of the bending frequency. The bands with dominant COH bend contribution are at 1333 and 1255 cm^{-1} in DG-2 and EG-1, respectively. This motion then falls in the same frequency region as the symmetric carboxylate stretch and CCH bend motions, generating a number of coupled motions and thus much stronger absorption in the 1250-1450 cm^{-1} range than for the other series of peptides. The blue shift is stronger in DGG-1 and DGG-2 where the peptidic bond is less favorably oriented because of its direct interaction with the carboxylate (see Figure 7). As a result, the COH bending frequency is near 1400 cm^{-1}, on the blue side of this absorption feature.

In GD-1, GE-1, GGD-1 and GGD-2, where the trans acid interacts with the carboxylate within the same residue, the other carboxylate C-O bond is pointing away from the acid and there is no other functional group in the vicinity of the O-H bond. Without a stabilizing interaction to counter the weakening of a strong, ionic hydrogen bond, the blue shift of the COH bend frequency is larger, to 1483, 1523 and 1535 cm^{-1} in GE-1, GGD-1 and GD-1, respectively. As a consequence, this band lies in the middle (for GE) or even on the blue side (for GD and GGD) of the 1450-1550 cm^{-1} feature in these cases.

Carboxylate stretches

As mentioned above, the free acetate C-O stretching frequencies are near 1320 and 1600 cm^{-1}, both experimentally and computationally. Limited blue shifts of the antisymmetric mode to 1625-1660 cm^{-1}, and of the symmetric mode to 1330-1359 cm^{-1}, have been reported previously for deprotonated oligoglycines [7a], with the symmetric mode intensity being much smaller and even hardly discernible for $[G_4\text{-}H]^-$. A value of 1320 cm^{-1} was reported for deprotonated trialanine [6]. Somewhat larger shifts of the antisymmetric mode to 1580-1610 cm^{-1} have been reported for deprotonated dimers of amino acids [23]. Much larger perturbations have been reported when acetate is attached to a divalent cation [9], however such a situation is less relevant to the present study.

The out-of-phase stretching frequency changes computed in the present work carry some information about changes in the carboxylate environment. However these bands always lie in a region of strong absorption and are always rather intense, thus not distinctive between the two series of spectra. The in-phase stretching frequency changes are not larger, however changes in the chemical environment lead to large intensity changes which make them much more revealing.

In GD-1, GE-1, GGD-1 and GGD-2, the carboxylate in-phase stretching generates a very weak band between 1354 and 1367 cm^{-1}. Very little absorption is observed in this region, in reasonably good agreement with calculations. The frequencies cover a slightly wider range in DGG-1 (1323 cm^{-1}), DG-1 (1340 cm^{-1}), DGG-2 (1356 cm^{-1}), DG-2 (1368 cm^{-1}) and EG-1 (1378 cm^{-1}). It is however their increased intensities which lead to significant differences between the two series of spectra. In DG-1 and DG-2, the in-phase motion is strongly coupled to C-OH stretch and COH bend in the acid while in EG-1, DGG-1 and DGG-2, it is coupled to the COH bend only. The relative orientations of the acid and of the carboxylate allow for significant dipole moment changes during such coupled motions. The orientations are less favorable in the other series of peptides but more importantly, the COH bending motion lies in a different frequency range as described above, near 1500 cm^{-1} or larger. Thus coupling of these local modes cannot be efficient and small intensities are measured.

Amide II

Previous work on deprotonated oligoglycines by Garand et al. [7a] provides reference values for this band. The frequencies for the peptidic linkage next to the C-terminus were found to be significantly red-shifted because of the ionic interaction, between 1487 and 1503 cm^{-1}. Other peptidic N-H bonds were engaged into weak, neutral hydrogen bonds, leading to smaller shifts in the 1519-1555 cm^{-1} range. A relatively large band centered slightly below 1500 cm^{-1} was reported for deprotonated trialanine [6]. In GD-1, GE-1 and GGD-2, the peptidic N-H (the one next to the C-terminus in GGD-2) interacts with one of the carboxylate oxygens as in deprotonated polyglycines. Red shifts are also observed, with bands in the 1466-1480 cm^{-1} range. These shifts can be attributed mainly to an important CNH bending component of the amide II mode, moving the hydrogen closer to / away from the negatively charged oxygen. The amide II mode on the N-terminus side of GGD-2, devoid of such interaction, has a higher frequency (1529 cm^{-1}) as expected. In GGD-1, diverse interactions in the folded region lead to frequencies of 1481 and 1529 cm^{-1}.

In DG, EG and DGG, the peptidic N-H bond (the one next to the N-terminus in DGG) interacts with the amine at the N-terminus. This interaction being weaker than that to the carboxylate, the red-shift of the amide II band is smaller, to 1505 cm^{-1} in DG-1, 1498 cm^{-1} in DG-2, 1527 in EG-1, (1489, 1555 cm^{-1}) in DGG-1 and (1491,1533 cm^{-1}) in DGG-2.

As a result, the amide II band is more red-shifted in GD, GE and GGD (near 1470 cm^{-1}), accounting for the reddish part of the intense band centered near 1500 cm^{-1}, while it is close to or larger than 1500 cm^{-1} in DG, EG and DGG (and partly in GGD), on the red side of the large 1550-1750 cm^{-1} group of bands (and, in fact, too much to the red in the present computations).

Strong hydrogen bond or proton sharing?

It has been seen above that for GD and GGD, which have a carboxylic-carboxylate interaction very close to that in D, the width and lack of clear structure of the absorption feature in the 1550-1750 cm^{-1} range suggests that as for D (see Figure S3), non-classical proton sharing takes place [10, 12]. On the other hand, this does not appear to occur in GE, even though the carboxylate environment is nearly identical to that in E, for which non-classical proton sharing has been suggested [10], see Figure S3.

The crucial factors governing the mode of proton sharing are not yet completely clear. Localized structures either equivalent by symmetry or nearly degenerate with a very low barrier to proton transfer would appear to be particularly favorable. This is suggested by a recent study of the very low temperature IR spectrum of the deprotonated formic acid dimer trapped in He nanodroplets, which indicates a symmetrically shared proton [24]. In this case, inclusion of zero-point vibrational energy is enough to reverse the energy ordering of symmetrical and hydrogen-bonded structures. The symmetrical structure being more stable with the proton trapped in a very shallow potential may be at the origin of the particularly low frequency associated with proton transfer, ca. 600 cm^{-1}. It was shown that coupling proton transfer to vibrational modes of the formate anions is crucial in order to model this rather unusual behaviour.

Symmetry or near degeneracy of the anionic sites does not appear to be a sufficient condition, however, as it has been shown by low temperature IRPD spectroscopy with Ar-tagging that symmetrical cases involving two equivalent carboxylates have different behaviors: deprotonated dodecanoic acid involves a strong but classical hydrogen bond [25], as does oxalate [26]. In the latter case, strong coupling of proton transfer with skeletal vibrational modes has also been invoked to explain its band lying in a less unusual frequency range than for the deprotonated formic acid dimer, however broadened over hundreds of wavenumbers [26]. It may be that the rather different types of modes to which proton transfer is coupled in the formate dimer, oxalate and deprotonated dodecanoic acid are partly responsible for the large differences observed in their IR spectra.

To make the picture even more complex, species in which a proton is shared by chemically different sites may [16, 17] or may not [17,27,28] lead to IR signatures of unusual proton sharing at room temperature, similar to those described above for GD and GGD. Deprotonated cysteines with different oxidization states were shown to switch from hydrogen-bound oxidized side chain in $[CysSO\text{-}H]^-$ to a shared-proton in $[CysSO_2\text{-}H]^-$ to a hydrogen-bound carboxylate in $[CysSO_3\text{-}H]^-$ [17], underlining the importance of the relative proton affinities of the two anions. Still, nearly degenerate cases do not all generate unusual vibrational features [27,28].

The lack of clear trend overall is due in part to different experimental techniques being used, even if focus is restricted to vibrational spectroscopy of gaseous ions. In the realm of low temperature devices, spectra recorded in He droplets use a free electron laser with photon energies ranging from 400 to 1800 cm^{-1} [24], while Ar tagging spectroscopy is done using an OPO/OPA laser extending from 600 to 3600 cm^{-1} [25]. Based on recent results, both the 400-600 and 2000-3600 cm^{-1} ranges appear to be revealing, however comparison of the two experiments could not yet be made in the full spectral range. At the other end of the temperature range, many experiments described to date in the literature were carried out at room temperature. It may well be that different situations are being probed when the proton is held by a very soft potential. At room temperature, proton switching rapidly between the two oxygens may occur, i.e. swapping between the covalent and hydrogen bonds [12,18], instead of being trapped in between the two oxygens as appears to be the case at low temperature, either in a classical hydrogen bonding pattern [25] or with two non-classical O---H “bonds” neither covalent nor hydrogen bonds [24].

More work appears to be needed to better characterize proton sharing in gas phase anions in which two carboxylates share a proton. On the computational side, a dynamical treatment appears to be required to obtain a more complete and realistic picture at room temperature. Assessing the possible role of quantum nuclear effects, such as proton tunneling, would also be desirable.

V. Conclusions

IRMPD spectra for seven small peptides bearing one acidic side chain have been recorded in the negative ion mode to investigate their hydrogen bonding patterns. The bands observed in these spectra have been assigned from IR spectra computed by harmonic DFT calculations using the ωB97X-D functional. Even though room temperature spectra have intrinsically limited resolution, the absorption patterns observed indicate large structural differences. In particular, vibrational modes which often generate bands of relatively minor intensity, such as acidic COH bend or symmetric carboxylate stretch, undergo significant shifts which, in certain cases, lead to mode coupling and intensity enhancement. This leads to frequency regions which are largely silent for a series of peptides and absorption-active for others. When Asp or Glu is at the C-terminus, a large feature at 1450-1550 cm^{-1} arises from amide II and COH bending motions. When the acidic residue is away from the C-terminus, a group of bands arises from combinations of COH bend, C-OH stretch, symmetric carboxylate stretch, CCH and HCH bends and amide II motions. Structure-dependent combination of such motions leads to enhanced intensities.
In addition, strong band broadening in the 1550-1750 cm^{-1} range for two of these peptides, GD and GDD, may be a consequence of non-classical proton sharing. This appears to bring additional evidence of such behavior when two carboxylate groups share a proton, a feature which has also been invoked recently in a variety of other systems where the proton is shared by two identical or different anionic groups. With recent accumulation of evidence, this phenomenon clearly deserves further study, both experimental and theoretical. In the latter case, a dynamical treatment appears to be necessary to model room temperature spectra, including quantum nuclear effects.

Acknowledgements

CNRS and Ecole polytechnique are acknowledged for funding. Financial support from the National FT-ICR network (FR 3624 CNRS) for conducting the research is gratefully acknowledged. The authors thank the CLIO team for excellent technical support to IRMPD experiments.

Supplementary material

Supplementary data associated with this article are available: free relative energies and structures of DG in a 20 $kJ.mol^{-1}$ energy window ; computed IR spectra for the lowest-energy structure of DG with 4 density functionals ; experimental IRMPD spectra of the deprotonated amino acids Asp and Glu.

References

[1] A. G. Harrison, J. Mass Spectrom. 2004, 39, 136-144.

[2] A. G. Harrison, A. B. Young, Int. J. Mass Spectrom. 2006, 255-256, 111-122.

[3] (a) J. H. Bowie, C. S. Brinkworth, S. Dua, Mass Spectrom. Rev. 2002, 21, 87-107. (b) J. H. Bowie, D. Bilusich, Mass Spectrom. Rev. 2009, 28, 20-34. (c) T. Wang T, T. T. T. Nha, H. J. Andreazza, D. Bilusich, C. S. Brinkworth, J. H Bowie, Mass Spectrom. Rev. 2018, 37, 3-21.

[4] (a) N. L. Clipston, J. Jai-Nkuknan, C. J. Cassady, Int. J. Mass Spectrom. 2003, 222, 363. (b) Z. Li, T. Yalcin, C. J. Cassady, J. Mass Spectrom. 2006, 41, 939-949. (c) A. S. McNeill, C. Cui, S. R. Miller, M. L. Stover, M. Burns, C. J. Cassady, D. A.Dixon, Int. J. Mass Spectrom. 2018, 429, 212-226.

[5] M. A. Halim, M. Girod, L. MacAleese, J. Lemoine, R. Antoine, P. Dugourd, J. Am. Soc. Mass Spectrom. 2016, 27, 474-486.

[6] J. Oomens, J. D. Steill, J. Am. Soc. Mass Spectrom. 2010, 21, 698-706.

[7] B. M. Marsh, E. M. Duffy, M. T. Soukup, J. Zhou, E. Garand, J. Phys. Chem. A 2014, 118, 3906-3912. (b) B. M. Marsh, J. Zhou, E. Garand, RSC Adv. 2015, 5, 1790-1795.

[8] (a) J. Oomens, J. D. Steill, J. Phys. Chem. A 2008, 112, 3281-3283. (b) J. Oomens, J. D. Steill, J. Phys. Chem. A 2009, 113, 4941-4946.

[9] J. W. DePalma, P. J. Kelleher, L. C. Tavares, M. A. Johnson, J. Phys. Chem. Lett. 2017, 8, 484-488.

[10] J. Oomens, J. D. Steill, B. Redlich, J. Am. Chem. Soc. 2009, 131, 4310-4319.

[11] (a) J. Grzetic, J. Oomens, Int. J. Mass Spectrom. 2013, 354, 70-77. (b) J. Grzetic, J. Oomens, Int. J. Mass Spectrom. 2012, 316, 216-226. (c) J. Grzetic, J. Oomens, J. Am. Soc. Mass Spectrom. 2012, 23, 290-300. (d) A. Martin-Somer, J. Martens, J. Grzetic, W. L. Hase, J. Oomens, R. Spezia, J. Phys. Chem. A 2018, 122, 2612-2625. (e) A. L. Patrick, N. C. Polfer, Topics Curr. Chem. 2015, 364, 153-181.

[12] F. Thaunay, F. Calvo, G. Ohanessian, C. Clavaguéra, *Empirical valence-bond models based on polarizable force fields for infrared spectroscopy* (in Theory and Applications of the Empirical Valence Bond Approach: From Physical Chemistry to Chemical Biology*)* pp. 121−144, Ed. F. Duarte, S. C. L. Kamerlin, Wiley, 2017.

[13] H. K. Woo, K. C. Lau, X. B. Wang, L. S. Wang, J. Phys. Chem. A 2006, 110, 12603-12606.

[14] Z. X. Tian, A. Pawlow, J. C. Poutsma, S. R. Kass, J. Am. Chem. Soc. 2007, 129, 5403-5407.

[15] Y. Fang, F. Liu, R. Emre, J. Liu, J. Phys. Chem. B 2013, 117, 2878-2887.

[16] A. F. DeBlase, S. R. Kass, M. A. Johnson, Phys. Chem. Chem. Phys. 2014, 16, 4569-4575.

[17] D. Scuderi, E. Bodo, B. Chiavarino, S. Fornarini, M. E. Crestoni, Chem. Eur. J. 2016, 22, 17239-17250.

[18] F. Thaunay, F. Calvo, E. Nicol, G. Ohanessian, C. Clavaguéra, submitted.

[19] P. Maître, S. L. Caër, A. Simon, W. Jones, J. Lemaire, H. Mestdagh, M. Heninger, G. Mauclaire, P. Boissel, R. Prazeres, F. Glotin, J.-M. Ortega, Nucl. Instrum. Methods Phys. Res., Sect. A 2003, 507, 541.

[20] (a) D. G ; Truhlar et al. https://comp.chem.umn.edu/freqscale/version3b2.htm. (b) I. M. Alecu, J. Zheng, Y. Zhao, and D. G. Truhlar, J. Chem. Theory Comput. 2010, 6, 2872-2887.

[21] Gaussian09, rev. D.01, M. J. Frisch et al. Gaussian, Inc., Wallingford CT, 2013.

[22] Y. Maréchal, J. Chem. Phys. 1987, 87, 6344-6353.

[23] S. Heiles, G. Berden, J. Oomens, E. R. Williams, Phys. Chem. Chem. Phys. 2018, 20, 15641-15652

[24] D. A. Thomas, M. Marianski, E. Mucha, G. Meijer, M. A. Johnson, G. von Helden, Angew. Chem. Int. Ed. Engl. 2018, 57, 10615-10619.

[25] M. Z. Kamrath, R. A. Relph, T. L. Guasco, C. M. Leavitt, M. A. Johnson, Int. J. Mass Spectrom. 2011, 300, 91–98.

[26] C. T. Wolke, A. F. DeBlase, C. M. Leavitt, A. B. McCoy, M. A. Johnson, J. Phys. Chem. A 2015, 119, 13018-13024.
[27] D. Scuderi, C. F. Correia, O. P. Balaj, G. Ohanessian, J. Lemaire, P. Maitre, ChemPhysChem 2009, 10, 1631-1640.
[28] A. Cimas, P. Maitre, G. Ohanessian, M. P. Gaigeot, J. Chem. Theory Comput. 2009, 5, 2388-2400.

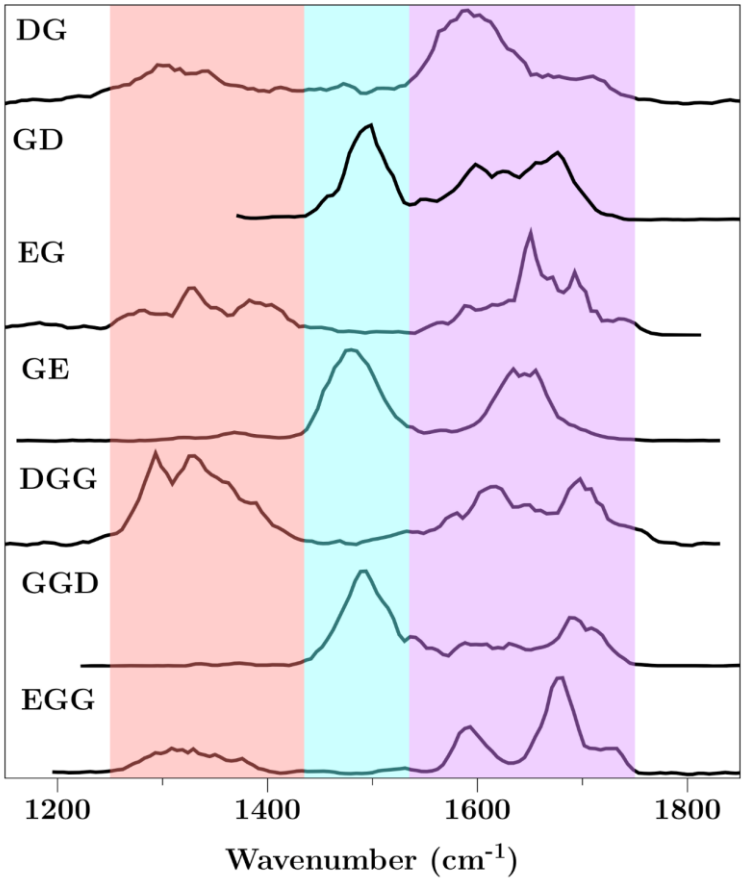


Figure 1: IRMPD spectra of the deprotonated peptides in the 1150-1850 cm^{-1} range, recorded at room temperature at the CLIO facility (Orsay, France). The coloured boxes are visual guides to define the main frequency ranges discussed in the text.

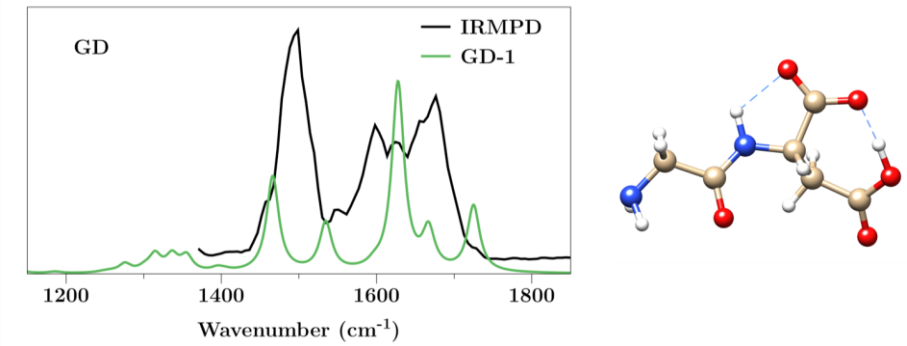


Figure 2: The IRMPD experimental spectrum for GD is represented in black and superimposed with the IR absorption spectrum (in green) computed for the GD-1 structure (right part).

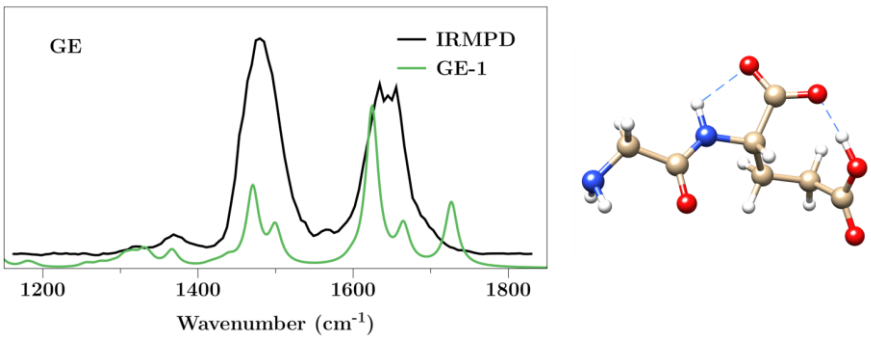


Figure 3: The IRMPD experimental spectrum for GE is represented in black and superimposed with the IR absorption spectrum (in green) computed for the GE-1 structure (right part).

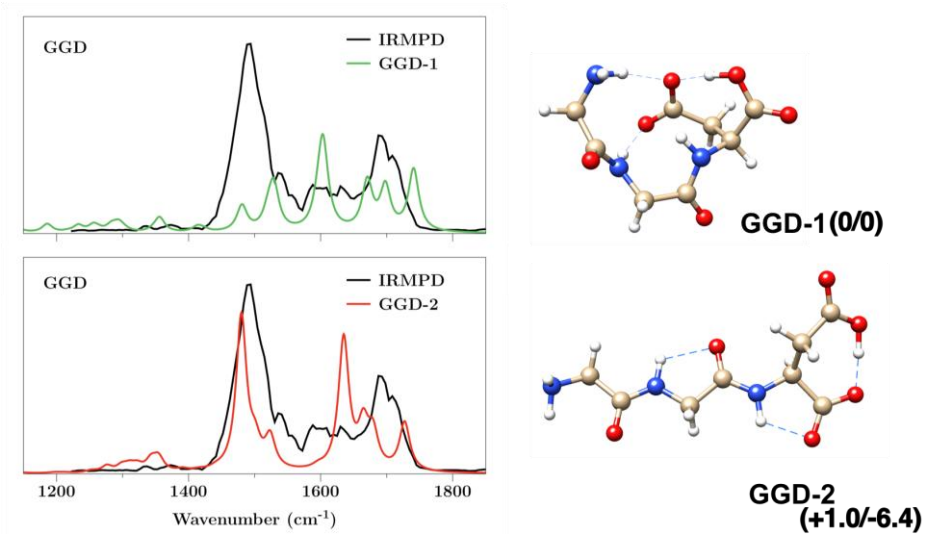


Figure 4: The IRMPD experimental spectrum for GGD is represented in black and superimposed with the IR absorption spectra (in green and red) computed for the GGD-1 and GGD-2 structures (right part), respectively. Relative free energies at 0/298 K in kJ.mol^{-1} are provided in parentheses.

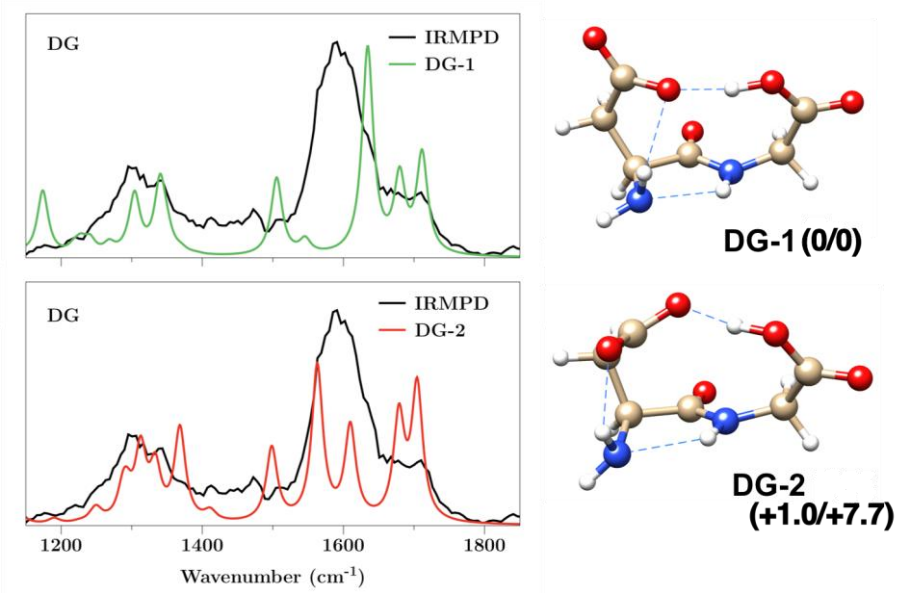


Figure 5: The IRMPD experimental spectrum for DG is represented in black and superimposed with the IR absorption spectra (in green and red) computed for the DG-1 and DG-2 structures (right part), respectively. Relative free energies at 0/298 K in kJ.mol^{-1} are provided in parentheses.

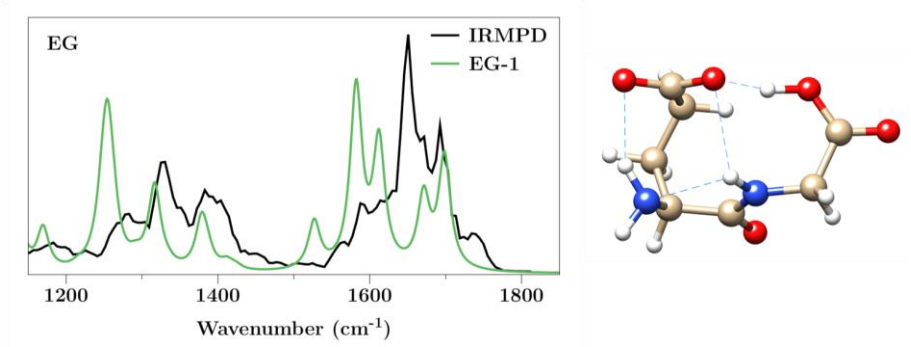


Figure 6: The IRMPD experimental spectrum for EG is represented in black and superimposed with the IR absorption spectrum (in green) computed for the EG-1 structure (right part).

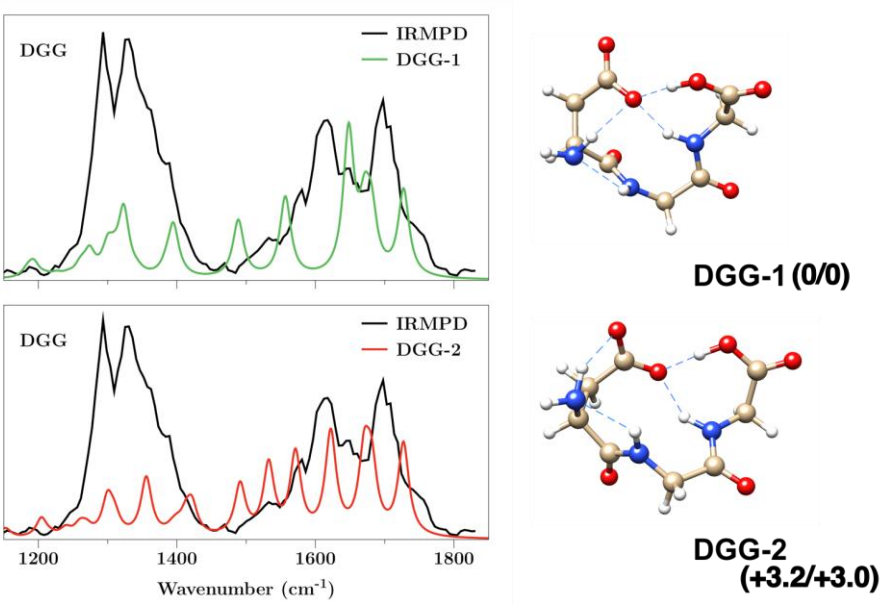


Figure 7: The IRMPD experimental spectrum for DGG is represented in black and superimposed with the IR absorption spectra (in green and red) computed for the DGG-1 and DGG-2 structures (right part), respectively. Relative free energies at 0/298 K in kJ.mol^{-1} are provided in parentheses.