# CC2 Benchmark for Models of Phenylalanine Protein Chains: 0–0 Transition Energies and IR Signatures of the $\pi\pi^*$ Excited State

Mi-Song Dupuy, Eric Gloaguen, Benjamin Tardivel, Michel Mons, Valérie Brenner

# CC2 Benchmark for models of phenylalanine protein chains: 0-0 transition energies and IR signatures of the ππ* excited state.

*Mi-Song Dupuy,† Eric Gloaguen, Benjamin Tardivel, Michel Mons and Valérie Brenner**

LIDYL, CEA, CNRS, Université Paris-Saclay, 91191 Gif-sur-Yvette, France.

KEYWORDS: Coupled-Cluster theory, Linear-response method, model CC2, Excited electronic states, Excitation energies, Excited states properties, Vibrational spectroscopy, Photochemistry and photophysics of biomolecules.

ABSTRACT: Extensive benchmarking calculations are presented to assess the accuracy of the standard approximate coupled cluster singles and doubles method (CC2) in studying ππ* excited states properties of model protein chains containing a phenylalanine residue, namely capped peptides, whose ground state conformers adopt the prototypical secondary structural features of proteins. First, the dependence with the basis-set of the CC2 excitation energies, CC2 geometry optimizations and *amide A* region frequencies of the lowest ππ* excited state in a reference system, the N-acetyl-phenylalaninylamide, are investigated and the results are compared with experimental data. Second, at the best level of theory determined, the CC2/aug(N,O,π)-cc-pVDZ//CC2/cc-

pVDZ level, a series of capped peptides of increasing size and containing residues of different nature are investigated. Along the series, compared to the experimental values, a mean absolute error of 0.10 eV is achieved for the 0-0 transition energies with a systematic overestimation. In addition, mode-dependent linear scaling functions for the calculated frequencies of the *amide A* region have been determined from the set of 95 experimental frequencies available; they lead to a quantitative simulation of the observed shifts of the *amide A* region frequencies upon ππ* excitation (root mean square deviation of 5 $cm^{-1}$). These results confirm the reliability of the CC2 method to characterize the lowest ππ* excited state of such medium-sized systems, emphasizing this class of theoretical approaches as a relevant spectroscopic tool, including for tasks as difficult as conformational assignment.

## 1. INTRODUCTION

Elucidating the mechanisms of excited-state deactivation following UV absorption is of key importance in understanding the photoinduced chemical dynamics of molecular systems, especially for those of paramount biological importance, like DNA bases or proteins.[1] Such studies on protein models shed light on the interplay between dynamics and biological function.[2-4] Deactivation processes are controlled by the energy and the nature of the electronic excited states of the chromophores, by their couplings and the resulting electron dynamics: UV light absorption populates excited states, which dissipate the electronic energy, either through a relatively slow radiative deactivation process, *i.e.* photon emission, or, more often and more efficiently by radiationless transitions, *e.g.* internal conversion or intersystem crossing. The latter, called nonadiabatic transfers, often involve ultrafast energy transfers through regions of the potential energy surfaces (PES) corresponding to avoided or surface crossings, of conical nature or not.[5-10]

In this context, we have developed an original innovative computational strategy in order to document the conformer-selective dynamics of capped peptide building blocks serving as models of proteins. Developed on small capped peptides,[11-13] this multi-step multi-level computational strategy allows us both to characterize the PES and the dynamics of their low-lying excited states using, first, nonadiabatic dynamic simulations based on time-dependent density functional theory (NA-TDDFT) to provide hints about the critical motions that drive the deactivation, which is then refined at two better levels of theory: i) the standard approximate coupled cluster singles and doubles method (CC2)[14-18] and ii) a multireference configuration interaction (MRCI) method.[19-21] This work is fully in line with these previous works[11-13] and focuses on the establishment of benchmark of the CC2 method on a series of capped peptides of increasing size and containing residues of different nature.

Nowadays, even if multireference methods, such as the complete active space perturbation theory of second order, CASPT2, or MRCI, are capable of accurately describing various types of excited states for a wide range of chemical systems, their application is still limited to small systems (no more than 10 atoms), since they are extremely time-consuming, especially if one wishes to investigate the spectroscopy or the dynamic of the excited states. In the case of larger systems, if the electronic structure of the excited states is dominated by single excitations out of the ground state, single reference method can be used. Among correlated single reference method, the CC2 method is one of the best choice due to its good compromise between computation times and accuracy. The validity of the CC2 method to treat low-lying excited states of medium size systems (more than 10 atoms and less than a few hundred atoms) has been already demonstrated by comparison with MRCI calculations for systems such as a retinal-chromophore[22] model or 9H-adenine[23] or even very recently for a small capped peptide.[13] Our objective is now to assess the accuracy of the CC2 method by comparison with experimental data. Benchmarks of the CC2 method against experimental data remain seldom for medium-sized systems. Moreover, they focus generally only on the adiabatic excitation energies, and evaluate the accuracy of both ground and excited states energies and gradients. Two of the most recent studies concern sets of the lowest excited state of different types of medium-sized systems, either focusing on aromatic organic molecules exhibiting different conformers (46 molecules and 66 singlet states: mean absolute error (MAE) on the adiabatic excitation energies of 0.07 eV and mean signed error or ME of +0.00 eV),[24] or covering both organic (polyenes, carbonyl compounds, aromatic hydrocarbons and heterocycle aromatic compounds) and inorganic (main-group and transition metal-compounds) systems (79 molecules, 84 singlet and 12 triplet states: MAE of 0.19 eV and ME of +0.11 eV).[25] The key questions that we presently address are then: (i) To which extent the good performances

obtained in these former studies[24-25] or on small systems[26] (11 molecules - 9 singlets and 4 triplet states – with MAE of 0.10 eV and a ME of -0.03 eV) apply to model protein chains, in particular capped peptides of increasing size, whose conformers in the ground state adopt the prototypical secondary structural features of proteins? and (ii) How does the CC2 method perform, not only in calculating energies and gradients of the excited states, but also in calculating their Hessian, and hence their vibrational frequencies?

The computational method, its practical details and the benchmark set composition are described in the Computational Details section. The basis-set convergence of the CC2 calculations is performed on four conformers of a reference system, the N-acetyl-phenylalaninylamide (Fa) and is discussed in the Results and Discussion section, where both the adiabatic excitation energies and the shifts of the vibrational frequencies between the ground and lowest ππ* excited states in the *amide A* region (the 3225-3580 $cm^{-1}$ region mainly due to the NH, $NH_{2sym.}$ and $NH_{2anti.}$ stretches) are compared to experimental data. This latter section also includes the discussion of the results related to the benchmarking of the CC2 method on both a derivative of Fa and several other relevant capped peptides, which contain two residues with different side chains, the objective being to determine the behavior of the CC2 method when increasing the system size and changing the nature of the residue.

## 2. COMPUTATIONAL DETAILS

### 2.1 Benchmark set composition:

The benchmark set consists of five capped peptides and their conformers, *i.e.* 18 ground states and 20 excited states. This set contains (i) two peptides with one residue, the phenylalanine (Phe), and two amide groups: Ac-Phe-$NH_2$ (four conformers A-D) and Ac-Phe-NH-$CH_3$ (three

conformers A-C), referred to in short as Fa and Fm respectively; (ii) two peptides with two residues and three amide groups: Ac-Gly-Phe-$NH_2$ (five conformers: A, A', B, B' and C), which contains one glycine (Gly) and one phenylalanine and Ac-Phe-Phe-$NH_2$ (three conformers: A-C), which contains two phenylalanines, referred to, in short, as GFa and FFa respectively; and (iii) one peptide with two residues but four amide groups, Ac-Gln-Phe-$NH_2$ (three conformers: A-C), which contains one phenylalanine and one glutamine (Gln) residue, whose side chain also bears an amide group, referred to in short as QFa. These systems (Figure 1) were selected since conformer-selective experimental data such as the 0-0 transition energies of the lowest ππ* excited states or the *amide A* region frequencies of both the ground and ππ* excited states were available and measured in our group. The ground state conformational landscape of each capped peptide has been characterized through their vibrational signature by combining conformer-selective double resonance IR/UV experiment and dispersion-corrected density functional theory methods (B97-D2/TZVPP, denoted hereafter DFT-D) in order to assign the conformers observed.[27] Then, when the intensity of the signal and the excited state lifetime (> 10 ns) were sufficient, three-color UV/IR/UV experiments were carried out in order to measure excited state IR (ESIR) spectra.[28-29] All these data are summarized in the following for each system as well as reported in more details in the hereafter cited references.

**Fa**: In the ground state, four conformers with two different types of folding backbone were observed and assigned to conformations lying in an energy range of 0-6.5 kJ/mol:[30] one β-strand extended conformation (A ($\beta_L$(a)) and three γ-turn folded conformations differing by the orientation of the phenyl side chain (B ($\gamma_L$(g+), C($\gamma_L$(g-) and D($\gamma_L$(a)[29]). The label L refers to the preferential orientation of the γ-turn feature (so-called inverse γ-turn) adopted by the natural amino acids of L configuration, as opposed to direct γ-turns, preferred by D aminoacids. Labels a, g+ or

g− refer to the anti, gauche+ or gauche- orientation of the phenyl side chain relative to the backbone, the $\chi^1$ dihedral angle N−$C_\alpha$-$C_\beta$-$C_\gamma$ being close to +180° for a, +60° for g+ and −60° for g-. These γ-turn and β-strand conformations are illustrative of the corresponding prototypical secondary structural features of proteins stabilized by $C_7$ H-bonds and $C_5$ H-bonds respectively, the index n of the label $C_n$ indicating the number of atoms in the ring formed by the H-bond. Noteworthy, two structures present an NH…π bond, the β-strand conformation A, and one of the γ-turns, B. The corresponding 0-0 transition energy to the lowest ππ* excited state was experimentally determined for each conformer[30] but only ESIR spectra of A and C could be measured.[28-29]

**Fm:** Fm corresponds to Fa in which the *cis*-position hydrogen of the C-terminal amide group has been replaced by a methyl group. In the ground state, Fm exhibits three conformers, structurally analogous to those of Fa (A-C), lying in an energy range of 0-4.5 kJ/mol.[12 and references therein] The corresponding 0-0 transition energy to the lowest ππ* excited state was determined for each conformer but only ESIR spectra of A and B could be measured.[28]

**GFa**: In the ground state, five conformers with three different types of folding backbone were observed and assigned to conformations lying in an energy range of 0-6 kJ/mol:[31] two 7-7 (double γ-turn) extended conformations, differing by the chirality L or D of the turns (A ($7_L$-$7_L$(g-)) and A' ($7_D$-$7_L$(g-))), two β-turn folded conformations of types I and II' (B (π-10I(g+)) and B'(π-10II'(g+))) and one β-strand (C (5-5-π(a))) extended conformation. All these families of backbone folding correspond to secondary structural features of proteins: the $2_7$ ribbon, β-strand and β-turn secondary structures, respectively stabilized by successive $C_7$ H-bonds (γ-turns), successive $C_5$ H-bond, and $C_{10}$ H-bonds. In addition, three structures present an NH…π bond, the two β-turns and

the β-strand. The corresponding 0-0 transition energy to the lowest ππ* excited state was determined for each conformer[31] but only ESIR spectra of A, B' and C could be measured.[28-29, 31]

**FFa:** In the ground state, three conformers were observed in an energetic range of 0-15 kJ/mol and assigned to conformations having both different types of backbone folding and different relative orientations of the two aromatic rings:[32 and references therein] one β-turn type I conformation (a $\alpha_L$-$\gamma_L$ structure, A (π-π-10I(g+,g+))), where the aromatic rings interact according to a T-shape arrangement, one unusual $\delta_L$ structure combined to a $\gamma_L$-turn (B (π-π-$7_L$(g+,g-))), i.e. a $\delta_L$-$\gamma_L$ conformation, where the aromatic rings interact according to a V-shaped arrangement, and one $\beta_L$-$\gamma_L$ conformation (C (5-π-$7_L$(a,g+))) where the aromatic rings interact according to a face-to-face arrangement. A is characterized by a $C_{10}$ H-bond, B by a $C_7$ H-bond and C by a usual combination of $C_5$ and $C_7^{eq}$ H-bonds. In addition, all these conformations present at least one NH-π bond. In the near UV spectrum, both conformers A and B lead to two intense transitions. The different ESIR spectra obtained after exciting these two bands, enable us to exclude any excitation delocalization in the excited state and thus to assign them to the origin transition of each excited chromophores. Comparison with comparable one-chromophore systems already studied even allows us to assign each origin transition to a chromophore ($A_1$, $A_2$ and $B_1$, $B_2$). For conformer C, only one weak band could be detected in the spectral region investigated, close to the 0-0 transition of toluene. This small band seems to correspond to a vibronic band of the origin transition of the conformer. This is confirmed by comparison with the conformer D of the FHa system, a capped peptide containing one phenylalanine and one histidine,[33] which exhibits a similar structure and where the phenyl chromophore experiences a similar environment. This similarity has been used to estimate the origin transition energy of the FFa C conformer. The corresponding 0-0 transition energies to the

lowest ππ* excited states were determined for each conformer ($A_1$-$A_2$, $B_1$-$B_2$ and C),[32] but only ESIR spectra of $A_1$ and $A_2$ were measured.[28]

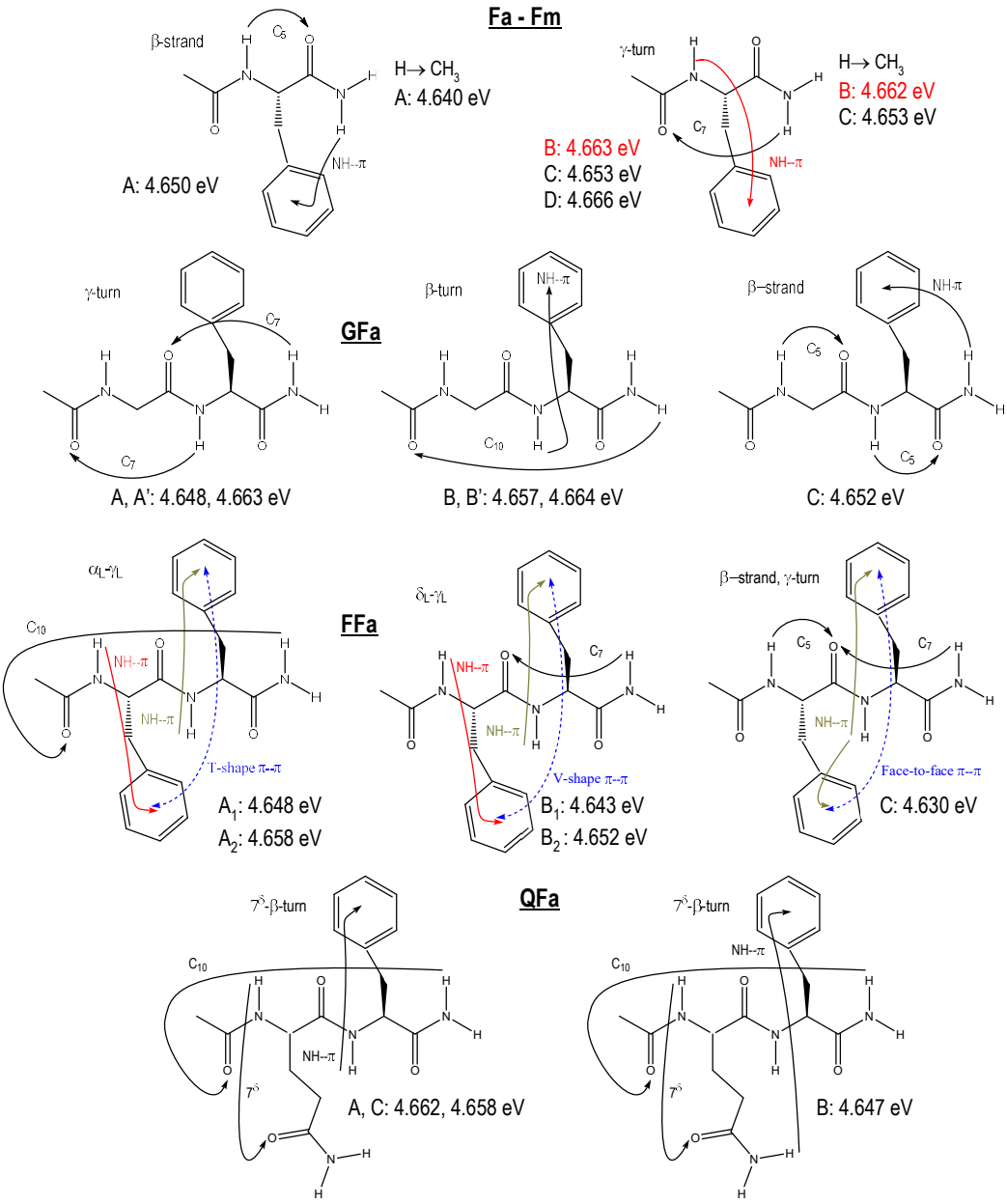



Figure 1: Ac-Phe-$NH_2$ (Fa), Ac-Phe-NH-$CH_3$ (Fm) and Ac-XX-Phe-$NH_2$ with XX=Gly, Phe, Gln (GFa, FFa and QFa) conformers and their experimental 0-0 transition energies (eV). In the Fm system, one hydrogen atom of the $NH_2$ group of the Fa system is replaced by a methyl group. The arrows indicate between which atoms the non-covalent intramolecular bonds (Hydrogen bonds, NH-π bonds and aromatic ring-ring arrangements) take place. The letters A, B, C refers to the notation adopted in the experimental reports to designate the several conformers observed, which

correspond to different types of backbone folding and non-covalent bonds networks (see the text for a detailed description). In case of several $S_0$ conformers which present both similar backbone folding and non-covalent networks, a single quotation mark is put to distinguish them such as A and A'. In case of several $S_1$ conformers whose the global conformation corresponds to that of the same $S_0$ conformer, a number is put in subscript to distinguish them such as $A_1$ and $A_2$."

**QFa:** In the ground state, three conformers were observed in an energetic range of 0-7.5 kJ/mol and assigned to conformations having the same type of folding backbone:[34 and Figure S11] all of them correspond to a type I β-turn backbone, stabilized by a $C_{10}$ H-bond combined to a side chain/main chain $C_7$ H-bond bridging the NH site of the first peptide bond to the oxygen atom of the Gln residue side chain CO-$NH_2$ group and labelled $7^{\delta}$. Each of them also present a NH…π bond. They differ by the hydrogen atom implied in the NH…π bond, by the orientation of phenylalanine residue or by the orientation of the CO-$NH_2$ group, *i.e.* the conformation of the Gln residue side chain. In A ($7^{\delta}$-π-10I(g+)) and C ($7^{\delta}$-π-10I(g+)), the NH…π bond involves the hydrogen atom of the second peptide bond but the orientation of the Gln residue side chain CO-$NH_2$ group differ. In B ($7^{\delta}$-π-10I(g-)), it is one of the terminal hydrogens of the Gln residue side chain CO-$NH_2$ which is involved in the NH…π bond. The corresponding 0-0 transition energy to the lowest ππ* excited state was determined for each conformer.[34] No ESIR spectra were measured for this system.

2.2 Method and basis sets:

Benchmark CC2 calculations[14-18] were carried out with the TURBOMOLE package.[35-36] All the CC2 calculations were performed by using the resolution-of-identity (RI) approximation for the electron repulsion integrals used in the correlation treatment and the description of the excitation processes. The CC2 model is based on coupled cluster (CC) theory and is designed as an approximation to the coupled cluster singles and doubles (CCSD) with an $O(N^5)$ scaling of the computational costs with system size N instead of $O(N^6)$: the singles amplitudes equations are kept

unchanged while the doubles amplitudes equations may be simplified by only retaining terms to lowest nonvanishing order. The attractive accuracy and convergence properties associated with CC methods are then transferred over to the calculation of electronic excitation energies and properties through the CC response theory framework by performing the derivation of the coupled cluster responses functions. Moreover, combining this model with the RI approximation leads to an acceleration of the calculations by one to two orders of magnitude, depending on the basis set, as well as to a decrease of memory demands to $O(N^2)$ and of disc storage demands to $O(N^3)$. The cc-pVXZ (X= D, T and Q) Dunning's correlation consistent basis sets[37] were employed in connection with optimized auxiliary basis sets for the RI approximation.[38] Furthermore, additional calculations were performed by adding to the cc-pVXZ (X= D and T) diffuse functions taken from the aug-cc-pVXZ (X= D and T) basis set for each oxygen and nitrogen atom and only for one in two carbon atoms of the phenyl ring. This kind of basis sets denoted hereafter aug(N,O,π)-cc-pVXZ (X=D and T) allows to avoid redundancy problems that arise if diffuse functions are supplied on all atoms[39-40] and their effectiveness has been recently demonstrated on Fa by comparison with MRCI calculations.[13] Frozen core for the 1s electrons were employed, and all calculations were carried out in the $C_1$ point-group symmetry. Ten singlet states were considered, and $D_1$, $D_2$ diagnostics, and % $\langle \overline{E_1} | E_1 \rangle$ biorthogonal norm were calculated in order to evaluate the capability of the CC2 method to properly describe the ground and excited states of such systems.[18, 41-42] Indeed, the $D_1$ and $D_2$ diagnostics, computed from the single and double substitution amplitudes in the CC2 wave function, are reliable indicators when static or dynamic correlation effects are not adequately treated at the CC2 level: their magnitude is correlated with the performances of the CC2 method. The initially recommended values for $D_1/D_2$ for ground state minima are 0.04(0.05)/0.17(0.18) in the case of MP2(CCSD)[41-42] but Köhn and Hättig have

extended these $D_1/D_2$ limit values up to 0.15/0.25 in particular from the evaluation of excited states of a set of small-sized molecules computed with CC2.[18] In addition, the biorthogonal norm $\langle\overline{E_1}|E_1\rangle$ gives a measure of the weight of the single excitation contributions to an excited state and should be larger than 85%.[18] Indeed, in order to be well described at the CC2 level, an excited state must be dominated by single excitations out of the ground state wave function.

The convergence criterion used in single point energy calculation is $10^{-8}$ on the density for the HF calculation, is $10^{-9}$ for the RI-CC2 ground state energy for the iterative coupled-cluster methods and $10^{-6}$ for the convergence threshold for norm of residual vectors in eigenvalue problems for the RI-CC2 excited states calculations. Finally, in the geometry optimization of both ground and lowest ππ* excited states of each conformer, the convergence criterion used corresponds to a norm of the Cartesian gradient lower than $10^{-4}$ au. For both optimized ground and excited states, orbital-relaxed first-order properties are determined, in particular the density. In the special FFa case, for which two close-lying ππ* excited states have been identified in the experiment for at least two conformers, each one being localized on a different phenyl ring, an exploration of the potential energy surface of the lowest ππ* excited states was performed in order to localize the minima corresponding to the two lowest ones. Whereas one of these ππ* excited states was easily identified by simply optimizing the first root starting from the ground state geometry, i.e., the protocol used to determine the lowest ππ* excited state of the capped peptides with one phenylalanine residue, for the second ππ* excited state, several geometrical deformations had to be investigated before achieving its geometry optimization. Indeed, the optimization of the second root starting from both the ground and the first ππ* excited state geometries systematically failed to converge to the second ππ* excited state.

The harmonic frequencies were calculated by numerical differentiation of the analytic gradients using central differences and a step length of 0.02 au. This allows to both verify that the optimized geometries correspond to true minima and calculate for each conformer and state the zero-point vibrational energy (ZPVE). The adiabatic ZPVE-corrected excitation energy of each lowest ππ* excited state were then calculated. Moreover, the IR data relevant for a comparison with experimental shifts were taken as the differences between the harmonic vibrational frequencies of the optimized ground state and the optimized ππ* excited state. In this way, we assumed that the basis set and method errors as well as anharmonicity effects are similar in both ground and excited states.

## 3. RESULTS AND DISCUSSION

In the next two subsections, the results of two benchmarking studies are presented. First, the basis set convergence of the CC2 method for energy, gradient (i.e. first derivatives of the energy) and hessian (i.e. second derivatives of the energy) calculations is tested on the four Fa conformers (A-D) by comparison between the results obtained with different basis sets, followed by a comparison with experimental data. The objective of this first benchmarking study is to determine the size and quality of the basis set required to obtain a good compromise between computation time and accuracy, keeping in mind that this reference system is the smallest system that we plan to investigate. Second, the selected basis sets are applied to the determination of the spectroscopic properties of a series of capped peptides of increasing size and containing different residues.

3.1 Basis set convergence of the CC2 calculations on the reference system, Fa:

Whatever the basis set, the ground state calculations of the four conformers exhibit $D_1/D_2$ values in the 0.075-0.085/0.17-0.27 ranges respectively while the lowest ππ* excited state calculations exhibit a $D_2$ value equal to 0.25 for all the conformers with a biorthogonal norm $\%\langle\overline{E_1}|E_1\rangle \geq 89\%$. These values confirm the reliability of the CC2 calculations on these systems even if some of them correspond to the upper limit of the recommended values.[18] Moreover, in the lowest ππ* excited states, the contributions of the canonical occupied ($\pi_{cycle}$)-unoccupied ($\pi^*_{cycle}$) HF orbitals to the total wave function change are all larger or equal to 94% for the four conformers as illustrated in Figure 2.

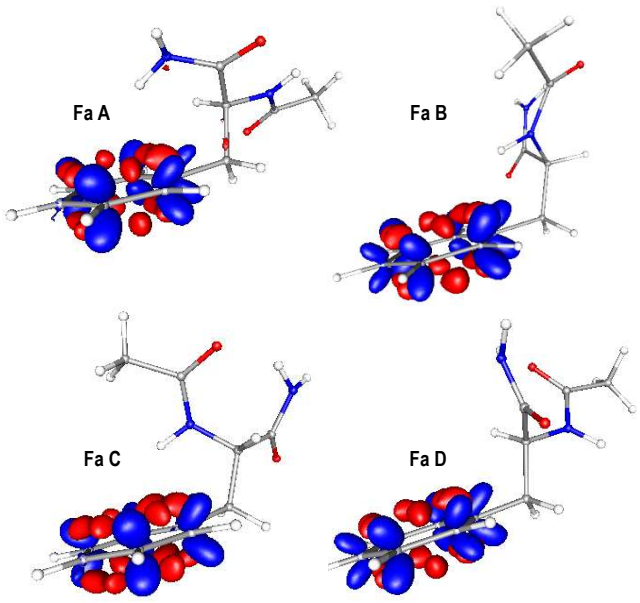



Figure 2: Contours (± 0.0019 au) of the difference between the CC2 density of the ππ* excited state and that of the ground state, for the four Fa conformers, calculated at the CC2/cc-pVDZ optimized geometry of the lowest ππ* excited state. A density increase (decrease) is indicated in blue (red).

| | | | Dihedral angles (°)[a] | | | Intramolecular distances (Å) | |
|---|---|---|---|---|---|---|---|
| | | | Φ | Ψ | $\chi^1$ | $d_{NH...O}$ | $d_{NH...\pi}$[b] |
| $S_0$ | | A | -159 | 168 | 201 | 2.15 | 2.45 (2.54, 3.10) |
| | cc-pVDZ | B | -82 | 53 | 44 | 1.96 | 2.37 (3.16, 2.49) |
| | | C | -88 | 80 | -51 | 1.98 | 2.56 (3.46, 2.70) |
| | | D | -80 | 85 | 192 | 2.06 | |
| | | A | -161 | 164 | 194 | 2.15 | 2.47 (2.69, 2.90) |
| | cc-pVTZ | B | -82 | 52 | 43 | 1.93 | 2.34 (3.13, 2.44) |
| | | C | -86 | 73 | -52 | 1.96 | 2.64 (3.60, 2.72) |
| | | D | -82 | 83 | 192 | 2.07 | |
| $S_1$ | | A | -164 | 155 | 182 | 2.22 | 2.51 (2.92, 2.63) |
| | cc-pVDZ | B | -83 | 52 | 45 | 1.95 | 2.37 (3.30, 2.35) |
| | | C | -88 | 79 | -50 | 1.99 | 2.52 (3.48, 2.53) |
| | | D | -81 | 84 | 192 | 2.06 | |
| | | A | -164 | 158 | 184 | 2.20 | 2.50 (2.93, 2.60) |
| | cc-pVTZ | B | -82 | 52 | 44 | 1.92 | 2.33 (3.25, 2.30) |
| | | C | -86 | 72 | -50 | 1.95 | 2.53 (3.51, 2.51) |
| | | D | -83 | 81 | 192 | 2.06 | |

Table 1: Characteristic geometrical parameters of CC2/cc-pVXZ (X=D and T) optimized geometry for both the ground ($S_0$) and lowest ππ* excited ($S_1$) states of the four Fa conformers.

[a] For the definition of the dihedral angles, see the Supporting Information (Figure S1).

[b] The NH…π bond is characterized by three distances: the distance of the $NH_{Phe}$ (A conformer) or $NH_2$ (B conformer and C conformer) hydrogen atom with the $C_\gamma$ carbon atom of the phenylalanine residue and, given in parentheses, the two distances with the two $C_\delta$ carbon atoms ($C_\delta^{C\text{-term}}$, $C_\delta^{N\text{-term}}$) of the phenylalanine residue.

The characteristic geometrical parameters of both the ground ($S_0$) and the lowest ππ* excited ($S_1$) state optimized geometries of the four Fa conformers are reported in Table 1 for the two basis set investigated, the cc-pVDZ and the cc-pVTZ basis sets. All the discrepancies for the $S_0$ optimized geometries fall into a range of [-7/+2]° for the dihedral angles with a mean absolute deviation (MAD) of 2.5° and into a range of [-0.03/+0.08] Å for the intramolecular distances with a MAD of 0.03 Å. Similar small discrepancies are obtained for the characteristic geometrical parameters of the $S_1$ optimized geometries: a range of [-7/+3]° for the dihedral angles with a MAD of 2° and a range of [-0.04/+0.02] Å for the intramolecular distances with a MAD of 0.02 Å.

Furthermore, no discrepancies larger than 0.01 Å are obtained for the covalent bond distances. These very small deviations in the geometries of ground and ππ* excited state between the two basis sets results are illustrated on the Figures S3-1&2, which highlight the superimposition of the two geometries of both states for all the conformers.

| $\Delta\nu_{S1/S0}$ ($cm^{-1}$) | cc-pVDZ | cc-pVTZ | Experiment |
|---|---|---|---|
| Fa A | | | |
| $NH_{Phe}$ | +11 | +2 | -1 |
| $NH_{2\ sym.}$ | -15 | -19 | -9 |
| $NH_{2\ anti.}$ | -18 | -14 | -6 |
| Fa B | | | |
| $NH_{Phe}$ | -44 | -53 | |
| $NH_{2\ sym.}$ | -5 | -6 | |
| $NH_{2\ anti.}$ | -1 | -1 | |
| Fa C | | | |
| $NH_{Phe}$ | -44 | -54 | -27 |
| $NH_{2\ sym.}$ | -2 | -7 | -1 |
| $NH_{2\ anti.}$ | -1 | -1 | -1 |
| Fa D | | | |
| $NH_{Phe}$ | 0 | -1 | |
| $NH_{2\ sym.}$ | -2 | -3 | |
| $NH_{2\ anti.}$ | 0 | +1 | |

Table 2: Theoretical shifts of the *amide A* region frequencies (CC2/cc-pVXD (X = D and T) harmonic) of the lowest ππ* excited state optimized geometry ($S_1$) relative to the ground state optimized geometry ($S_0$), for the four Fa conformers, together with the corresponding available experimental data ($cm^{-1}$).

In order to go further in the comparison, the vibrational frequencies, in particular the three NH stretch frequencies ($NH_{Phe}$, $NH_{2sym.}$ and $NH_{2anti.}$) of the *amide A* region and their shifts between the $S_1$ and the $S_0$ states, obtained for the two basis sets are determined and reported for the four Fa conformers in Table S4 and Table 2. Again, small discrepancies are observed with a MAD per $NH_{Phe}$, $NH_{2sym.}$ and $NH_{2anti.}$ stretch frequency of 11, 7 and 4 $cm^{-1}$ resp. for the $S_0$ state and of 10, 8 and 5 $cm^{-1}$ resp. for the $S_1$ state. This corresponds to a MAD for the three stretch frequencies of

7.5 $cm^{-1}$ for the two states, i.e. a deviation of only around 0.21% for this frequency range. In addition, the frequency shifts between $S_1$ and $S_0$ states also present similarly small discrepancies between the two basis sets: a MAD per stretch frequency of 7, 3 and 1 $cm^{-1}$ with a mean signed deviation (MD) of -3, -3 and 1 $cm^{-1}$.

As far as the adiabatic ZPVE-corrected excitation energy is concerned, the two basis sets give very similar results (see Table 3 and Table S5) for all the four conformers (a MAD of 0.02 eV for the ZPVE-corrected excitation energy, the MAD for the ZPVE being equal to 2-3 meV) confirming that the cc-pVDZ basis set is good enough for the geometry optimization. On the contrary, the test to circumvent the ground state geometry optimization by taking the DFT-D optimized geometry shows a significant effect on the adiabatic ZPVE-corrected excitation energies with a MAD of 0.08 eV (Table 3, first line). This effect is certainly due to a significant change in the intramolecular geometrical parameters obtained between the two levels of theory, a MAD of 0.11 Å on the intramolecular distances (Table S2) and this level of theory, DFT-D level for the ground state geometry optimization, was therefore not considered later. On the other hand, the effect of diffuse functions on the energetics seems to be relatively important (0.04 – 0.06 eV) and not negligible, as it will be confirmed below by the comparison with experiment. Basis sets with diffuse functions were therefore considered in the following for the calculation of the excitation energies. The level of theory considered in the following was then the CC2/cc-pVDZ level for the geometry optimization of both the ground and excited states and the CC2/aug(N,O,π)-cc-pVDZ level for the excitation energies calculation, leading to adiabatic ZPVE-corrected excitation energies calculated at the CC2/aug(N,O,π)-cc-pVDZ// CC2/cc-pVDZ level.

| $\Delta E_{adia}$ (eV) | Fa A | Fa B | Fa C | Fa D |
|---|---|---|---|---|
| *cc-pVDZ*[a] | 4.876 | 4.896 | 4.896 | 4.917 |
| cc-pVDZ | 4.807 | 4.823 | 4.815 | 4.835 |
| cc-pVTZ | 4.816 | 4.836 | 4.828 | 4.850 |
| cc-pVTZ//cc-pVDZ | 4.816 | 4.834 | 4.826 | 4.850 |
| aug(N,O,π)-cc-pVDZ//cc-pVDZ | 4.754 | 4.770 | 4.767 | 4.791 |
| aug(N,O,π)-cc-pVTZ//cc-pVDZ | 4.789 | 4.807 | 4.802 | 4.824 |
| Experiment[b] | 4.650 | 4.663 | 4.653 | 4.666 |

Table 3: Adiabatic ZPVE-corrected excitation energies of the lowest ππ* excited state ($S_1$) of the four Fa conformers calculated at the CC2/basis set2//CC2/basis set1 level, together with the experimental 0-0 transition energies.
[a] Here the geometry of the ground state ($S_0$) is optimized at the DFT-D/TZVPP level (see Table S2) whereas the geometry of the lowest ππ* excited state ($S_1$) is optimized at the CC2/cc-pVDZ level.
[b] The experimental values are presented with the number of significant digits obtained in the experiment and the theoretical values are given with the same number of digits.

Finally, comparing calculated and experimental 0-0 transition energies, the CC2 method at the CC2/aug(N,O,π)-cc-pVDZ//CC2/cc-pVDZ level proves to be very reliable. Indeed, the MAE, 0.11 eV, is equal to the ME with deviations in the [+0.11-+0.13] eV range (Table 3). Moreover, the calculated shifts of the *amide A* region frequencies between the $S_1$ and $S_0$ states are in a relatively good agreement with the experimental ones (Table 2) considering that they correspond to direct shifts obtained from harmonic frequencies, i.e. considering that basis set and method errors as well as anharmonicity effects are similar in both ground and excited states. Both the order of magnitude and sign of the experimental shifts have been well reproduced, especially the large shifts corresponding to significant geometrical changes of intramolecular parameters sensitive to the non-covalent interactions involving the phenyl ring where the excitation is localized. Upon excitation, the π system tends to extend farther from the ring $C_6$ axis, with a density increase above the C atoms and on the ring edge, and a decrease above the C-C bonds (Figure 2). In terms of

geometry, this electronic change leads to a strong shortening of the intramolecular distances $d_{NH...\pi}$ as illustrated in Figure 3 and Figure S6 for Fa B (-0.14 Å) and C (-0.17 Å), and consequently to a large calculated and measured shift of the *amide A* region frequency (Table 2). On the contrary, when the intramolecular distances $d_{NH...\pi}$ do not strongly vary or do not correspond to a significant intramolecular interaction such as in Fa A ($d_{NH...\pi}$ ($S_0$) = 2.45 Å and +0.06 Å in $S_1$) or Fa D, as shown in Figure 3 and Figure S6, no large shift between the $S_1$ and $S_0$ states has been calculated or measured (Table 2). Concerning the valence parameters such as covalent bonds, the valence angles, dihedral angles that do not involve the orientation of the backbone relative to the phenyl ring or other parameters which concern atoms not belonging to the π system such as the intramolecular distances $d_{NH...O}$, they are only very weakly changed upon ππ* excitation in all the Fa conformers: a MAD of 4° for the dihedral backbone angles, a MAD of 0.02 Å for the intramolecular distances $d_{NH...O}$ and a root mean square deviation (RMSD) change of the covalent bonds of only 0.01-0.02 Å (Table 1, Figure 3 and S6).

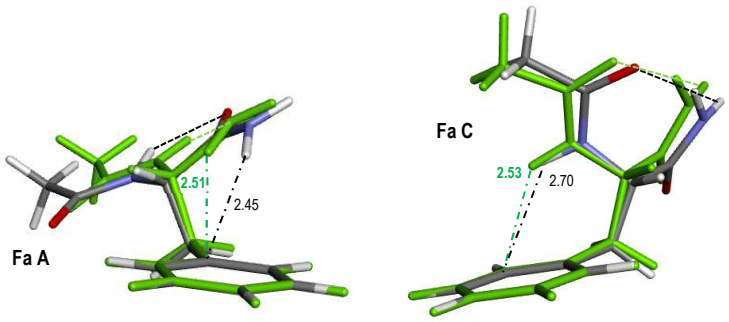



Figure 3: Comparison of the CC2/cc-pVDZ optimized geometries of the $S_0$ (atom-based colors) and $S_1$ (green) state for Fa A and C. For each conformer, the phenylalanine residues have been overlapped. Only distances (dash-dot) that vary significantly ($|d| > 0.01$ Å) between the ground and the excited state (see Table 1) are mentioned.

3.2 Adiabatic transition energies of the first ππ* excited states of the benchmark set:

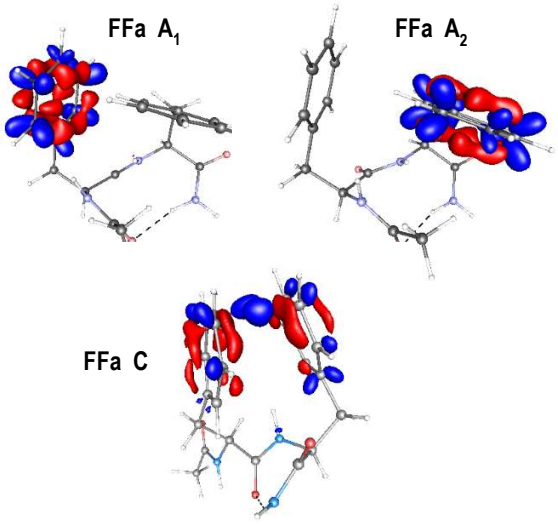



Figure 4: Contours (± 0.0019 au) of the difference between the CC2/cc-pVDZ density of each of the two ππ* excited states ($A_1$ and $A_2$) and that of the ground state, in FFa A (top panel) and between excited and ground states in FFa C (bottom panel) calculated at the CC2/cc-pVDZ optimized geometry of the ππ* excited states. A density increase (decrease) is indicated in blue (red).

Whatever the capped peptides considered, the optimized ground state of the conformers exhibits $D_1/D_2$ values in a range similar to that obtain for the Fa conformers while their lowest ππ* excited states exhibit $D_2$ and biorthogonal norm $\%\langle\overline{E_1}|E_1\rangle$ values similar to those obtained for Fa optimized ππ* excited states. Among the series of capped peptides, the lowest ππ* excited state optimized geometries exhibit an excited state localized on the phenyl ring when this one is unique in the molecule. The same observation holds for the FFa A and FFa B conformers where their two low-lying ππ* excitations are localized on either of the phenyl rings (Figure 4). The only exception is the FFa C conformer for which the lowest ππ* excited state is delocalized on the two phenyl

rings (Figure 4). Moreover, for all the ππ* excited states of the series, the contributions of the canonical occupied ($\pi_{cycle}$)-unoccupied ($\pi^*_{cycle}$) HF orbitals to the total wave function change are at least 95%. Finally, the non-covalent interaction network of the several conformers (intramolecular H-bonds, NH…π bonds and various relative orientations of the two aromatic rings) is preserved in the lowest optimized ππ* excited states as it will be discussed and illustrated in the following section.

Along the series, the trend observed for the transition energies (Table 3) is similar to that observed in Fa: the calculated adiabatic transition energies are overestimated compared to experimental ones. The only exception concerns the FFa C conformer for which the 0-0 transitions was not directly measured but extrapolated by comparison with an analogous system and their vibronic progression: the theoretical adiabatic transition energy underestimates the 0-0 transition by 0.15 eV. Even if the analogous system presents a conformation similar to FFa C, it contains only one phenylalanine carrying the ππ* excitation as opposed to FFa C, which may explain differences in comparison with experiment. A MAE equal to 0.10 eV with a very similar ME of +0.08 eV and positives deviations in the [0.05-0.13] eV range are obtained for the series. Then, it seems that the error is systematic along the series without no significant size effect since the MAE obtained for the capped monopeptide, 0.11 eV, is equivalent to that obtained for the capped dipeptides, i.e. 0.09 eV. Finally, this error is equivalent to that obtained for small systems[26] or others molecular systems,[24-25] within the expected error margins of more sophisticated methods such as the MS-CASPT2 method, i.e. ± 0.2 eV.

| Capped peptides | $\Delta E_{adia}$ | Experiment[a] |
|---|---|---|
| Fm A | 4.728 | 4.640 |
| Fm B | 4.769 | 4.662 |
| Fm C | 4.773 | 4.653 |
| GFa A | 4.754 | 4.648 |
| GFa A' | 4.776 | 4.663 |
| GFa B | 4.743 | 4.657 |
| GFa B' | 4.771 | 4.664 |
| GFa C | 4.766 | 4.652 |
| FFa A1 | 4.714 | 4.648 |
| FFa A2 | 4.729 | 4.658 |
| FFa B1 | 4.730 | 4.643 |
| FFa B2 | 4.705 | 4.652 |
| FFa C | 4.480 | *4.630* |
| QFa A | 4.736 | 4.662 |
| QFa B | 4.753 | 4.647 |
| QFa C | 4.725 | 4.658 |

Table 3: Adiabatic ZPVE-corrected excitation energies (eV) of the lowest ππ* excited state ($S_1$) of Fm, GFa, FFa and QFa conformers calculated at the CC2/aug(N,O,π)-cc-pVDZ //CC2/cc-pVDZ level, together with the experimental 0-0 transition energies (eV).
[a] The experimental values are presented with the number of significant digits obtained in the experiment and the theoretical values are given with the same number of digits. The value in italics correspond to an estimation by analogy with FHa system (see text).

3.3 IR signature of the *amide A* region vibrations:

As for the reference system, Fa, the valence parameters such as covalent bonds, valence angles, dihedral angles that do not involve the orientation of the backbone relative to the phenyl ring, or any other parameters, which concern atoms not belonging to the π system, such as the intramolecular distances $d_{NH...O}$, are only very weakly changed upon ππ* excitation (Table S7-1-4 and Figures 5, S8-1-4): (i) a range of [-6/+2]° for the Fm conformers dihedral angles with a MAD of 1°, respectively a range of [-11/+5]° for GFa and a range of [-4/+10]° for FFa with a MAD of

2° and a range of [-13, +9]° for QFa with a MAD of 5°; (ii) an almost zero MAD for the intramolecular distances $d_{NH...O}$ of Fm conformers, respectively 0.01 Å for GFa and 0.02 Å for both FFa and QFa; (iii) finally, a RMSD change of the covalent bonds of 0.01-0.02 Å only. On the contrary, upon excitation, the intramolecular distances $d_{NH...\pi}$ can be strongly shortened in the Fm conformers (A: -0.26 Å, B: -0.17 Å and -0.19 Å), in GFa A (-0.17 Å), FFa C (-0.14 and -0.25 Å) and QFa B (-0.39 Å), moderately shortened in GFa B (-0.08 Å) and B'(-0.11 Å), FFa $A_1$, $A_2$, $B_2$ (-0.07 Å) and QFa A (-0.12 Å), or weakly shortened in GFa C (-0.04 Å with a change of the π carbon atom involved in the interaction), FFa $B_1$ (-0.04 Å) and QFa C (-0.04 Å). The shortening of the intramolecular distance in these NH…π bonds, $d_{NH...\pi}$, upon ππ* excitation seems to be reflected in the calculated frequency shifts of the stretch vibration of the corresponding NH between the ground and the ππ* excited states (Table 4, unscaled shifts): the strongest shortenings lead to large calculated shifts (> 30 $cm^{-1}$) whereas moderate ones lead to smaller shifts (< 25 $cm^{-1}$) and the weaker ones to weak shifts (< 17 $cm^{-1}$). A careful examination of data, however, suggests that the dependence might be not so simple: the strength of the NH…π interactions has to be also considered in order to rationalize the value of the frequency shifts in the *amide A* region. For instance, for similar moderate shortenings, the conformers in which the intramolecular distance $d_{NH...\pi}$ becomes lower than 2.50 Å in the excited state present a shift larger than 30 $cm^{-1}$, such as in GFa B, FFa $A_1$ and $A_2$, a shift usually characteristic of a strong shortening such as in the Fm conformers, in FFa C or in QFa B. In the same spirit, for similar strong shortenings, the conformers in which $d_{NH...\pi}$ remains larger than 2.50 Å in the excited state present shifts lower than 25 $cm^{-1}$ such as in GFa A, a shift characteristic of a moderate shortening such as in GFa B' and in FFa $B_2$ (shift of the $NH_{Phe2}$ stretch frequency). In the special case of FFa $B_2$, the large shift of the $NH_{Phe1}$ stretch frequency (33 $cm^{-1}$) results from a strong shortening of $d_{NH...\pi}$ between the $NH_{Phe1}$ and the

π system of the Phe2 aromatic ring (-0.19 Å and a distance value which becomes lower than 2.50 Å in the excited state). Finally, when there is no NH…π interaction such as in GFa A', the calculated shift of the *amide A* region frequency is close to zero. It should be noticed that the shifts of NH stretch frequencies calculated for the NH…O interactions follow a similar trend. As $d_{NH...O}$ are in majority very weakly changed, all the calculated shifts of *amide A* region frequencies are weak (< ± 15 $cm^{-1}$) except when a shortening of the NH…O H-bond distance occurs leading to $d_{NH...O}$ values shorter than ~2 Å in the excited state, as in FFa C (-0.03 Å with a value of 1.96 Å in the excited state) and in QFa B (-0.03 Å with a value of 1.84 Å in the excited state) for which large shifts are calculated (> 30 $cm^{-1}$).

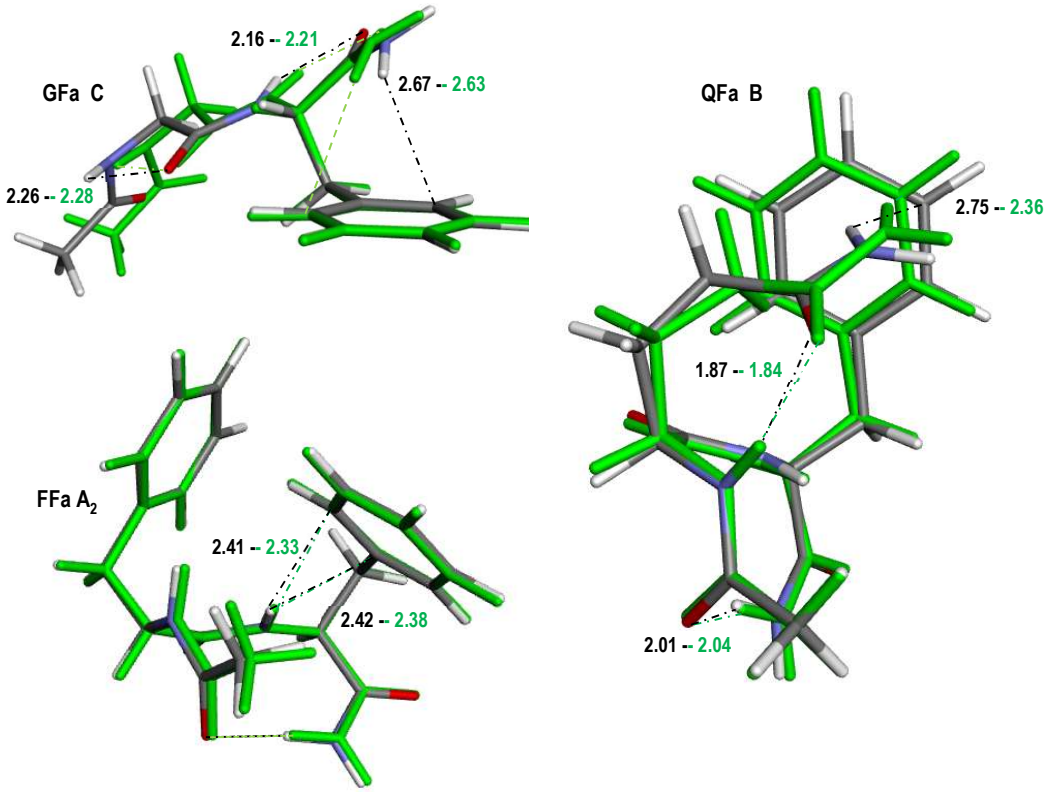



Figure 5: Comparison of the CC2/cc-pVDZ optimized geometries of the $S_0$ (atom-based colors) and $S_1$ (green) state for GFa C, FFa $A_1$ and QFa B. For GFa C and FFa $A_1$, the phenylalanine residues are overlapped whereas for QFa B, the backbones are overlapped. Only distances (dash-dot) that vary significantly (|d| > 0.01 Å) between the ground and the excited states (see Table S7.2-4) are mentioned.

| Capped peptides | $\Delta\nu_{S1/S0}$ ($cm^{-1}$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Theory[a] | | | | Experiment | | | |
| | $NH_{Phe}$ | | $NH_{C\text{-}term}$ | | $NH_{Phe}$ | | $NH_{C\text{-}term}$ | |
| Fm A | -5/-4 | | **-44/-37** | | 0 | | -27 | |
| Fm B | **-47/-40** | | -4/-3 | | -32 | | -4 | |
| Fm C | **-47/-40** | | -4/-3 | | | | | |
| | $NH_{Gly}$ | $NH_{Phe}$ | $NH_{2sym.}$ | $NH_{2anti.}$ | $NH_{Gly}$ | $NH_{Phe}$ | $NH_{2sym.}$ | $NH_{2anti.}$ |
| GFa A | -7/-6 | ***-18/-15*** | -5/-3 | 0/0 | -2 | -18 | +3 | -9 |
| GFa A' | -2/-2 | -5/-4 | -2/-1 | -2/-1 | | | | |
| GFa B | -1/-1 | **-37/-31** | -3/-2 | -1/-1 | | | | |
| GFa B' | 0/0 | ***-21/-18*** | +1/+1 | 0/0 | +1 | -18 | +2 | +1 |
| GFa C | +5/+4 | -2/-2 | *-15/-10* | *-14/-7* | +1 | -3 | -9 | -6 |
| | $NH_{Phe1}$ | $NH_{Phe2}$ | $NH_{2sym.}$ | $NH_{2anti.}$ | $NH_{Phe1}$ | $NH_{Phe2}$ | $NH_{2sym.}$ | $NH_{2anti.}$ |
| FFa A1 | **-41/-35** | -9/-8 | -4/-3 | -2/-1 | **-33** | 0 | -1 | 0 |
| FFa A2 | -5/-4 | **-34/-29** | -2/-1 | -1/-1 | -1 | **-24** | -1 | 0 |
| FFa B1 | *-11/-9* | -4/-3 | -6/-4 | -2/-1 | | | | |
| FFa B2 | **-33/-28** | ***-19/-16*** | 0/0 | 0/0 | | | | |
| FFa C | -12/-10 | **-74/-63** | **-30/-21** | -11/-6 | | | | |
| | Theoretical $\Delta\nu_{S1/S0}$ ($cm^{-1}$) | | | | | | | |
| | $NH_{Gln}$ | $NH_{Phe}$ | $NH_{2}$ sym./C-term | $NH_{2}$ anti./C-term | $NH_{2}$ sym./Chain | | $NH_{2}$ anti./Chain | |
| QFa A | -2/-2 | -12/-10 | +5/+3 | +8/+4 | -2/-1 | | -2/-1 | |
| QFa B | **-43/-37** | +6/+5 | -3/-2 | +1/+1 | **-45/-31** | | **-27/-14** | |
| QFa C | -9/-8 | *-16/-14* | -2/-1 | 0/0 | -1/-1 | | +1/+1 | |

Table 4: Theoretical (CC2/cc-pVDZ harmonic and mode-dependent corrected CC2/cc-pVDZ, i.e. x/y) frequency shifts between the lowest ππ* excited state ($S_1$) and the ground state ($S_0$) *amide A* stretches, ($cm^{-1}$) for the Fm, GFa, FFa and QFa conformers, together with the available experimental shifts. The large shifts of the frequencies of NH involved in NH…π bonds are in bold, the moderate ones in italic-bold and the weak in italics (see text). The large shifts of the frequencies of NH involved in NH…O bonds are in bold-underlined.

[a] The mode dependent corrected frequency shifts are calculated from the CC2/cc-pVDZ harmonic frequencies which are corrected by the optimal harmonic frequency mode-dependent (NH, $NH_{2sym}$ and $NH_{2anti}$) linear scaling functions determined from the whole set of 95 ($S_0$ plus $S_1$) experimental *amide A* region stretch frequencies available (see text).

As far as comparison with experiment is concerned, theory tends to overestimate the frequency shifts in the *amide A* region for all the capped peptides including the reference system, Fa, especially for large red shifts whereas moderate and weak shifts seem to be correctly reproduced (Table 4). A MAE of 6 $cm^{-1}$ is obtained together with a ME of -4 $cm^{-1}$ for the whole set of shifts whereas a MAE of 13 $cm^{-1}$ and a ME of -13 $cm^{-1}$ are obtained considering only the set

of large shifts (> 30 $cm^{-1}$). This may be due to differential anharmonicity effects between the ground and $\pi\pi^*$ excited states or to a differential intrinsic error of the CC2 method between the ground and excited state properties (geometry and frequencies).

A method error, which would arise from neglecting the intramolecular BSSE differential effects in the geometry optimization between the ground and excited states, can be immediately excluded in view of the previous results obtained for the reference system Fa. Indeed, calculations on Fa do not exhibit any significant increase of the intramolecular distances between the cc-pVDZ and the cc-pVTZ optimized geometries of both the $S_0$ and $S_1$ states: a MAD of 0.03 Å for $S_0$ with a mean signed deviation (MD) of +0.004 Å and a MAD of 0.02 Å for $S_1$ with a MD of -0.01 Å respectively. Furthermore, a MAD per $NH_{Phe}$, $NH_{2sym.}$ and $NH_{2anti.}$ stretch frequency of 11, 7 and 4 $cm^{-1}$ resp. for the $S_0$ states and of 10, 8 and 5 $cm^{-1}$ resp. for the $S_1$ states were determined previously between the cc-pVDZ and cc-PVTZ frequencies of Fa.

In order to go further in our analysis, we have taken advantage of the large amount of experimental data available to determine optimal harmonic frequency mode-dependent (NH, $NH_{2sym}$ and $NH_{2anti}$) linear scaling functions ($\nu_{exp.} = a\nu_{theo.} + b$) in order to correct the calculated $S_0$ and $S_1$ harmonic frequencies, and then the corresponding shifts, for method and basis set errors as well as anharmonicity effects. First, the linear scaling functions were determined from both the 67 experimental *amide A* region frequencies available for $S_0$ (Figure S10-1) and the 28 experimental *amide A* region frequencies available for $S_1$ (Figure S10-2) and second from the whole set of the 95 ($S_0$ plus $S_1$) experimental *amide A* region frequencies available (Figure 6). Except for the $NH_{2anti}$ stretch frequencies, the optimal harmonic frequency mode-dependent (NH, $NH_{2sym}$ and $NH_{2anti}$) linear scaling functions determined separately for the $S_0$ and $S_1$ states and those determined with

the whole set exhibit very similar behavior and parameters with a satisfactory correlation. This similar trend indicates that the errors in the calculation of the CC2/cc-pVDZ harmonic frequencies (method, basis set and anharmonicity) do not depend specifically on the ground or ππ* excited state. The most important disagreement obtained for the $S_1$ mode-dependent $NH_{2anti.}$ linear scaling function certainly results from the low number of experimental data available in this case. Once the $S_1$-$S_0$ shifts of the amide A frequencies are corrected using the linear scaling functions optimized with the whole set, a MAE of 3 $cm^{-1}$ is obtained together with a ME of -1.5 $cm^{-1}$ (Table 4). The values obtained considering only large shifts (> 30 $cm^{-1}$), a MAE of 7 $cm^{-1}$ and a ME of -7 $cm^{-1}$, still illustrate the presence of a systematic error for large shifts, as if the CC2 method would underestimate the intramolecular distances in case of strong non-covalent interactions but this effect on the shifts is greatly reduced using the mode-dependent corrected frequencies. Finally, these results demonstrate that a refined strategy, relying on mode-specific linear scaling functions, enables us to overcome the systematic disagreement between experimental and calculated harmonic frequencies and provides reliable predictions of the $S_1$-$S_0$ frequency shifts in the *amide A* region for such systems with a RMSD of 5 $cm^{-1}$.

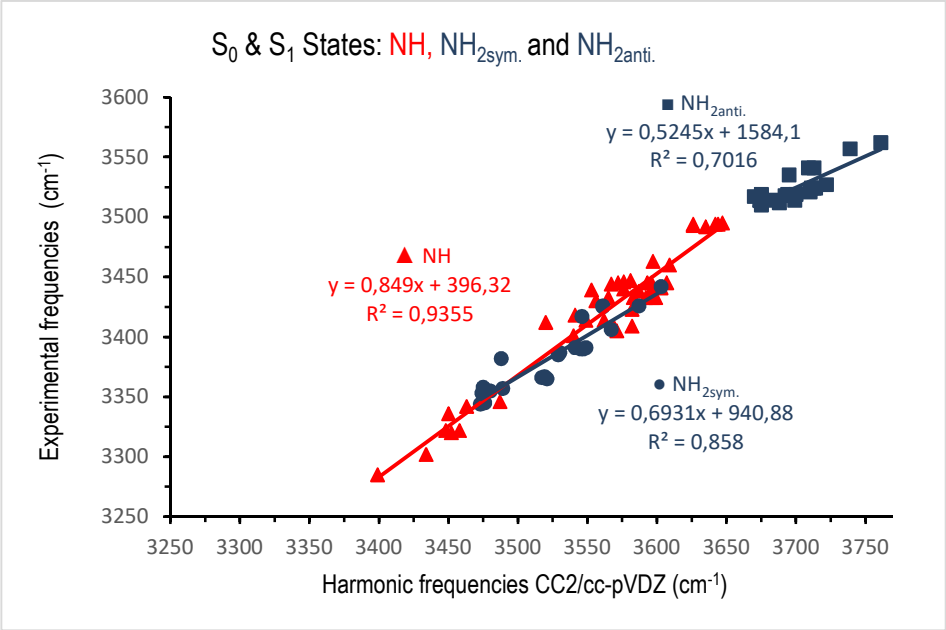



Figure 6: Experimental vs CC2/cc-pVDZ calculated harmonic *amide A* region frequencies of both the ground and the ππ* excited states of the series of capped peptides and the corresponding mode-dependent linear ($\nu_{exp.} = a\nu_{theo.} + b$) scaling functions.

## 4. CONCLUSIONS

Experimental data on a series of medium-sized capped peptides (20 conformers) were used to benchmark the CC2 method in its ability to reproduce the observed 0-0 transition energies as well as the frequency shifts in the *amide A* region upon ππ* excitation. We have demonstrated that the CC2/aug(N,O,π)-cc-pVDZ//CC2/cc-pVDZ level proves to be very reliable, giving a MAE of 0.10 eV for the 0-0 transition energies with a systematic overestimation compared to the experimental data (*i.e.* a very similar ME of +0.08 eV). Finally, mode-dependent linear ($\nu_{exp.} = a\nu_{harm.} + b$) scaling functions for the frequencies of the *amide A* region (NH, $NH_{2sym.}$ and $NH_{2anti.}$) have been determined from the whole set of the 95 experimental *amide A* region frequencies available (67 for $S_0$ and 28 for $S_1$). This leads to a quantitative simulation of the observed frequencies shifts of

the *amide A* region upon ππ* excitation (RMSD of 5 $cm^{-1}$) and allows an interpretation of this shift in terms of specific geometry changes. Being focused onto models of phenylalanine protein chains, the present work provides a state-of-the-art reference for detailed computational approaches addressing photochemistry and photobiology issues.

ASSOCIATED CONTENT

**Supporting Information**.

Electronic Supplementary Information (ESI) available:

S1: Definition of the characteristic angles of the backbone of capped peptides. S2: Characteristic geometrical parameters of the DFT-D optimized geometry of the ground state ($S_0$) of the four Fa conformers. S3: Comparison of the CC2/cc-pVXZ (X=D and T) optimized geometries of both the $S_0$ and $S_1$ states for the four Fa conformers. S4: *Amide A* region frequencies of both the ground ($S_0$) and ππ* excited ($S_1$) of the four Fa conformers. S5: ZPVE of both the ground ($S_0$) and ππ* excited ($S_1$) states according to the basis set for the four Fa conformers. S6: Comparison of the CC2/cc-pVDZ optimized geometry of the $S_0$ and $S_1$ states of Fa B and D. S7: Characteristic geometrical parameters of CC2/cc-pVDZ optimized geometry of both the ground ($S_0$) and lowest ππ* excited ($S_1$) states of the Fm, GFa, FFa and QFa conformers. S8: Comparison of the CC2/cc-pVDZ optimized geometry of the $S_0$ and $S_1$ states of the Fm, GFa, FFa and QFa conformers. S9: *Amide A* region frequencies of both the ground ($S_0$) and ππ* excited ($S_1$) states of the Fm, GFa, FFa and QFa conformers. S10: Experimental vs CC2/cc-pVDZ calculated harmonic *amide A* region frequencies of both the $S_0$ and $S_1$ states of the series of capped peptides and the corresponding mode-dependent linear ($\nu_{exp.} = a\nu_{theo.} + b$) scaling functions.

AUTHOR INFORMATION

**Corresponding Author**

*E-mail: valerie.brenner@cea.fr

ORCID

Valérie Brenner: 0000-0002-8004-1157

**Present Addresses**

†Present address: Technische Universität München, Zentrum Mathematik - M7, Boltzmannstraße 3, 85747 Garching.

**Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Funding Sources**

This work received a financial support from the Agence Nationale de la Recherche (ANR); Grant ANR-14-CE06-0019-01 - ESBODYR. This work was granted access to the HPC facility of [TGCC/CINES/IDRIS] under the Grant 2014- t2014087074, the Grant 2015- t2015087412&t2015087372, and the Grant 2016- t2016087540 awarded by GENCI (Grand Equipement National de Calcul Intensif), to the CCRT High Performance Computing (HPC) facility at CEA under the Grant CCRT2014/CCRT2015/CCRT2016-p606bren.

**Notes**

The authors declare no competing financial interest.

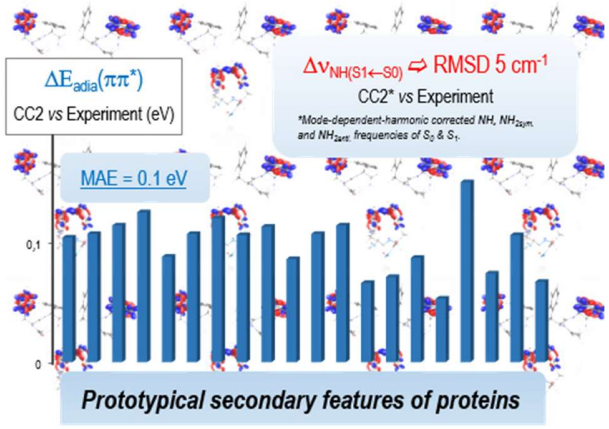
REFERENCES


1. Sobolewski, A. L.; Domcke, W., Molecular mechanisms of the photostability of life. *Phys. Chem. Chem. Phys.* **2010,** *12*, 4897-4898.
2. Chen, Y.; Barkley, M. D., Toward understanding tryptophan fluorescence in proteins. *Biochemistry* **1998,** *37*, 9976-9982.
3. Callis, P. R.; Liu, T. Q., Quantitative prediction of fluorescence quantum yields for tryptophan in proteins. *J. Phys. Chem. B* **2004,** *108*, 4248-4259.
4. Marazzi, M.; Sancho, U.; Castano, O.; Domcke, W.; Frutos, L. M., Photoinduced Proton Transfer as a Possible Mechanism for Highly Efficient Excited-State Deactivation in Proteins. *J. Phys. Chem. Lett.* **2010,** *1*, 425-428.
5. Roob, M. A.; Olivucci, M.; Bernadi, F., In *Encyclopedia of computational chemistry*, Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; J., G.; Kollman, P. A.; Schaefer, H. F.; Schreiner, P. R., Eds. John Wiley & Sons: Ltd Chichester, UK, 2002.
6. Worth, G. A.; Cederbaum, L. S., Beyond Born-Oppenheimer: Molecular dynamics through a conical intersection. In *Annual Review of Physical Chemistry*, 2004; Vol. 55, pp 127-158.
7. Levine, B. G.; Martinez, T. J., Isomerization through conical intersections. In *Annual Review of Physical Chemistry*, 2007; Vol. 58, pp 613-634.
8. Domcke, W.; Yarkony, D. R.; Koppel, H., *Conical Intersections: Electronic Structure, Dynamics & Spectroscopy*. World Scientific: River Edge, NJ, 2004.
9. Domcke, W.; Yarkony, D. R., Role of Conical Intersections in Molecular Spectroscopy and Photoinduced Chemical Dynamics. In *Annual Review of Physical Chemistry, Vol 63*, Johnson, M. A.; Martinez, T. J., Eds. 2012; Vol. 63, pp 325-352.
10. Yarkony, D. R., Nonadiabatic Quantum Chemistry-Past, Present, and Future. *Chem. Rev.* **2012,** *112*, 481-498.
11. Malis, M.; Loquais, Y.; Gloaguen, E.; Biswal, H. S.; Piuzzi, F.; Tardivel, B.; Brenner, V.; Broquier, M.; Jouvet, C.; Mons, M.; Doslic, N.; Ljubic, I., Unraveling the Mechanisms of Nonradiative Deactivation in Model Peptides Following Photoexcitation of a Phenylalanine Residue. *J. Am. Chem. Soc.* **2012,** *134*, 20340-20351.
12. Malis, M.; Loquais, Y.; Gloaguen, E.; Jouvet, C.; Brenner, V.; Mons, M.; Ljubic, I.; Doslic, N., Non-radiative relaxation of UV photoexcited phenylalanine residues: probing the role of conical intersections by chemical substitution. *Phys. Chem. Chem. Phys.* **2014,** *16*, 2285-2288.
13. Ben Amor, N.; Hoyau, S.; Maynau, D.; Brenner, V., Low-Lying excited states of model proteins: Performances of the CC2 method versus muitlreference methods. *J. Chem. Phys.* **2018,** *148*, 184105.
14. Christiansen, O.; Koch, H.; Jorgensen, P., The 2nd-order approximate coupled-cluster singles and doubles model CC2. *Chem. Phys. lett.* **1995,** *243*, 409-418.
15. Hattig, C.; Weigend, F., CC2 excitation energy calculations on large molecules using the resolution of the identity approximation. *J. Chem. Phys.* **2000,** *113*, 5154-5161.
16. Hattig, C.; Kohn, A., Transition moments and excited-state first-order properties in the coupled-cluster model CC2 using the resolution-of-the-identity approximation. *J. Chem. Phys.* **2002,** *117*, 6939-6951.
17. Hattig, C., Geometry optimizations with the coupled-cluster model CC2 using the resolution-of-the-identity approximation. *J. Chem. Phys.* **2003,** *118*, 7751-7761.

18. Kohn, A.; Hattig, C., Analytic gradients for excited states in the coupled-cluster model CC2 employing the resolution-of-the-identity approximation. *J. Chem. Phys.* **2003,** *119*, 5021-5036.
19. Bories, B.; Maynau, D.; Bonnet, M. L., Selected excitation for CAS-SDCI calculations. *J. Comput. Chem.* **2007,** *28*, 632-643.
20. Ben Amor, N.; Bessac, F.; Hoyau, S.; Maynau, D., Direct selected multireference configuration interaction calculations for large systems using localized orbitals. *J. Chem. Phys.* **2011,** *135*, 014101.
21. Chang, C.; Calzado, C. J.; Ben Amor, N.; Marin, J. S.; Maynau, D., Multi-scale multireference configuration interaction calculations for large systems using localized orbitals: Partition in zones. *J. Chem. Phys.* **2012,** *137*, 104102.
22. Tuna, D.; Lefrancois, D.; Wolanski, L.; Gozem, S.; Schapiro, I.; Andruniow, T.; Dreuw, A.; Olivucci, M., Assessment of Approximate Coupled-Cluster and Algebraic-Diagrammatic-Construction Methods for Ground- and Excited-State Reaction Paths and the Conical-Intersection Seam of a Retinal-Chromophore Model. *J. Chem. Theory Comput.* **2015,** *11*, 5758-5781.
23. Plasser, F.; Crespo-Otero, R.; Pederzoli, M.; Pittner, J.; Lischka, H.; Barbatti, M., Surface Hopping Dynamics with Correlated Single-Reference Methods: 9H-Adenine as a Case Study. *J. Chem. Theory Comput.* **2014,** *10*, 1395-1405.
24. Winter, N. O. C.; Graf, N. K.; Leutwyler, S.; Hattig, C., Benchmarks for 0-0 transitions of aromatic organic molecules: DFT/B3LYP, ADC(2), CC2, SOS-CC2 and SCS-CC2 compared to high-resolution gas-phase data. *Phys. Chem. Chem. Phys.* **2013,** *15*, 6623-6630.
25. Fang, C. F.; Oruganti, B.; Durbeej, B., How Method-Dependent Are Calculated Differences between Vertical, Adiabatic, and 0-0 Excitation Energies? *J. Phys. Chem. A* **2014,** *118*, 4157-4171.
26. Hattig, C., Structure optimizations for excited states with correlated second-order methods: CC2 and ADC(2). In *Advances in Quantum Chemistry, Vol 50: A Tribute to Jan Linderberg and Poul Jorgensen*, Sabin, J. R.; Brandas, E., Eds. 2005; Vol. 50, pp 37-60.
27. Gloaguen, E.; Mons, M., Isolated Neutral Peptides. In *Gas-Phase Ir Spectroscopy and Structure of Biological Molecules*, Rijs, A. M.; Oomens, J., Eds. 2015; Vol. 364, pp 225-270.
28. Sohn, W. Y.; Brenner, V.; Gloaguen, E.; Mons, M., Local NH-pi interactions involving aromatic residues of proteins: influence of backbone conformation and pi pi* excitation on the pi H-bond strength, as revealed from studies of isolated model peptides. *Phys. Chem. Chem. Phys.* **2016,** *18*, 29969-29978.
29. Alaudin, M.; Vaquero-Vara, V.; Habka, S.; Tardivel, B.; Gloaguen, E.; Mons, M., unpublished resutls.
30. Chin, W.; Mons, M.; Dognon, J. P.; Piuzzi, F.; Tardivel, B.; Dimicoli, I., Competition between local conformational preferences and secondary structures in gas-phase model tripeptides as revealed by laser spectroscopy and theoretical chemistry. *Phys. Chem. Chem. Phys.* **2004,** *6*, 2700-2709.
31. Loquais, Y.; Gloaguen, E.; Habka, S.; Vaquero-Vara, V.; Brenner, V.; Tardivel, B.; Mons, M., Secondary Structures in Phe-Containing Isolated Dipeptide Chains: Laser Spectroscopy vs Quantum Chemistry. *J. Phys. Chem. A* **2015,** *119*, 5932-5941.
32. Gloaguen, E.; Loquais, Y.; Thomas, J. A.; Pratt, D. W.; Mons, M., Spontaneous Formation of Hydrophobic Domains in Isolated Peptides. *J. Phys. Chem. B* **2013,** *117*, 4945-4955.

33. Sohn, W. Y.; Habka, S.; Gloaguen, E.; Mons, M., Unifying the microscopic picture of His-containing turns: from gas phase model peptides to crystallized proteins. *Phys. Chem. Chem. Phys.* **2017,** *19*, 17128-17142.
34. Vaquero-Vara, V.; Sohn, W. Y.; Tardivel, B.; Brenner, V.; Gloaguen, E.; Mons, M., in preparation.
35. TURBOMOLE v6.4-2012, v7.0-2015 & 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.
36. Furche, F.; Ahlrichs, R.; Hattig, C.; Klopper, W.; Sierka, M.; Weigend, F., Turbomole. *WIREs Comput. Mol. Sci.* **2014,** *4*, 91-100.
37. Dunning, T. H., Gaussian-basis sets for use in correlated molecular calculations.1. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989,** *90*, 1007-1023.
38. Weigend, F.; Kohn, A.; Hattig, C., Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002,** *116*, 3175-3183.
39. Granucci, G.; Hynes, J. T.; Millie, P.; Tran-Thi, T. H., A theoretical investigation of excited-state acidity of phenol and cyanophenols. *J. Am. Chem. Soc.* **2000,** *122*, 12243-12253.
40. Papajak, E.; Zheng, J. J.; Xu, X. F.; Leverentz, H. R.; Truhlar, D. G., Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.* **2011,** *7*, 3027-3034.
41. Janssen, C. L.; Nielsen, I. M. B., New diagnostics for coupled-cluster and Moller-Plesset perturbation theory. *Chem. Phys. lett.* **1998,** *290*, 423-430.
42. Nielsen, I. M. B.; Janssen, C. L., Double-substitution-based diagnostics for coupled-cluster and Moller-Plesset perturbation theory. *Chem. Phys. lett.* **1999,** *310*, 568-576.

For Table of Contents Only



## Supporting Information

# CC2 Benchmark for models of phenylalanine protein chains: 0-0 transition energies and IR signatures of the ππ* excited state.

*Mi-Song Dupuy,† Eric Gloaguen, Benjamin Tardivel, Michel Mons and Valérie Brenner**
LIDYL, CEA, CNRS, Université Paris-Saclay, 91191 Gif-sur-Yvette, France.

Appendix S1: Definition of the characteristic angles of the backbone of capped peptides.

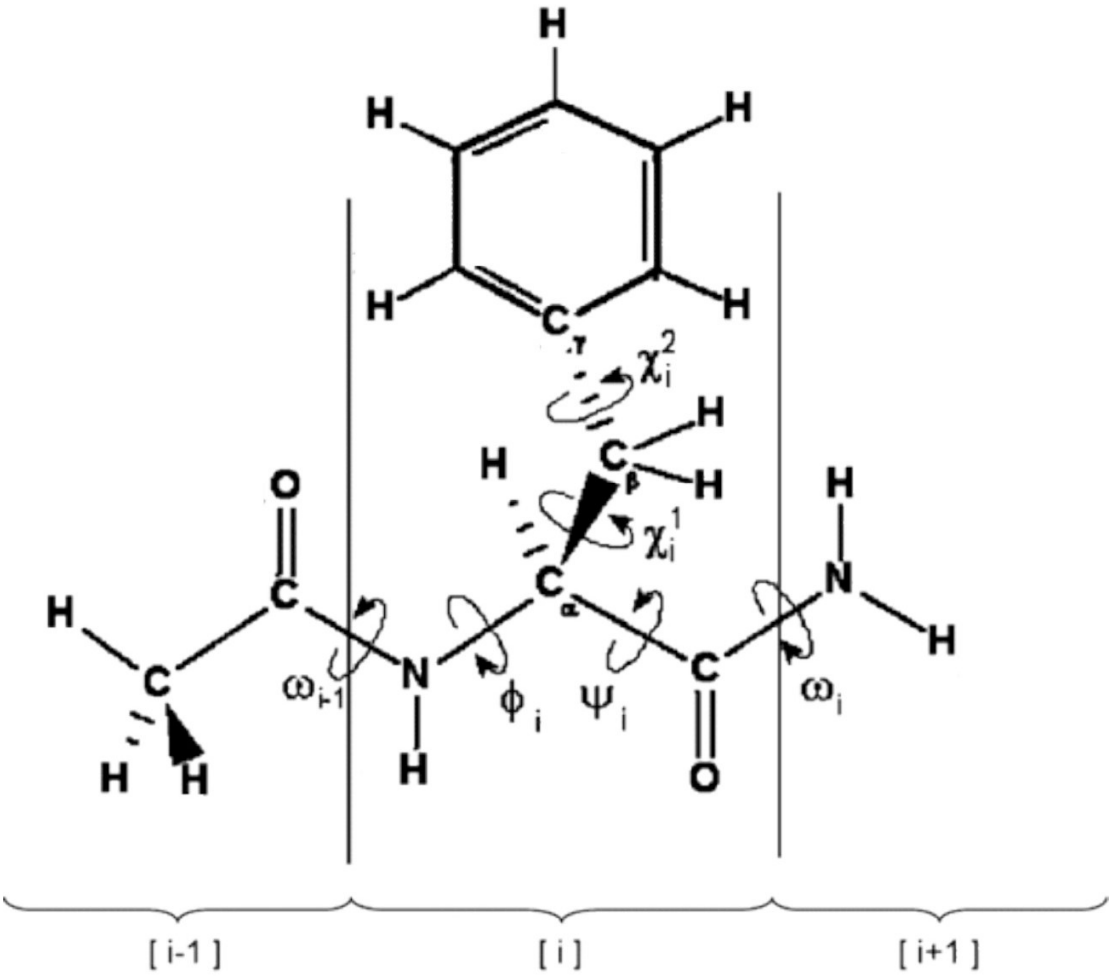



Figure S1: Definition of the characteristic dihedral angles of the backbone of capped peptides: Example of N-Ac-Phe-$NH_2$ (Fa) from the N-terminus (left-most [i-1] module) through the central Phe (central [i] module) to C-terminal $NH_2$ protecting group (right-most [i+1] module).

Appendix S2: Characteristic geometrical parameters of the DFT-D optimized geometry of the ground state ($S_0$) of the four Fa conformers.

| | | Dihedral angles (°)[a] | | | Intermolecular distances (Å) | |
|---|---|---|---|---|---|---|
| | | Φ | Ψ | $\chi^1$ | $d_{NH...O}$ | $d_{NH...\pi}$[b] |
| $S_0$ | A | -160 | 159 | 192 | 2.28 | 2.56 (3.37, 2.84) |
| | B | -83 | 55 | 44 | 2.02 | 2.44 (3.23, 2.54) |
| | C | -85 | 72 | -55 | 2.03 | 2.77 (3.76, 2.85) |
| | D | -83 | 84 | 193 | 2.24 | |

Table S2: Characteristic geometrical parameters of the ground state ($S_0$) of the four Fa conformers optimized at the DFT-D level.[1-2]

[a] For the definition of the dihedral angles, see the Supporting Information (Figure S1).

[b] The NH…π bond is characterized by three distances: the distance of the $NH_{Phe}$ (A conformer) or $NH_2$ (B conformer and C conformer) hydrogen atom with the $C_\gamma$ carbon atom of the phenylalanine residue and given in parentheses by the two distances with the two $C_\delta$ carbon atoms ($C_\delta^{C\text{-term}}$, $C_\delta^{N\text{-term}}$) of the phenylalanine residue.

Appendix S3: Comparison of the CC2/cc-pVXZ (X=D and T) optimized geometries of both the $S_0$ and $S_1$ states for the four Fa conformers.

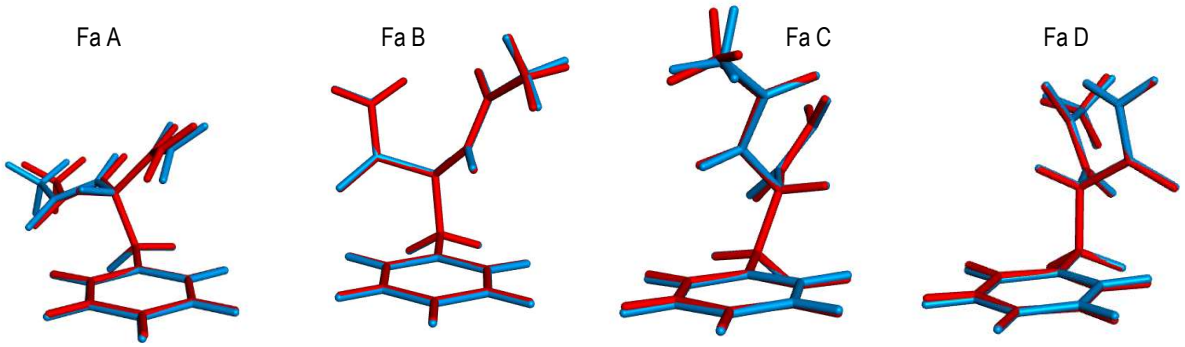



Figure S3-1: Comparison of the CC2/cc-pVDZ (blue) and CC2/cc-VTZ (red) optimized geometries of the $S_0$ state for the four Fa conformers. For each conformer, the phenyl rings have been overlapped.

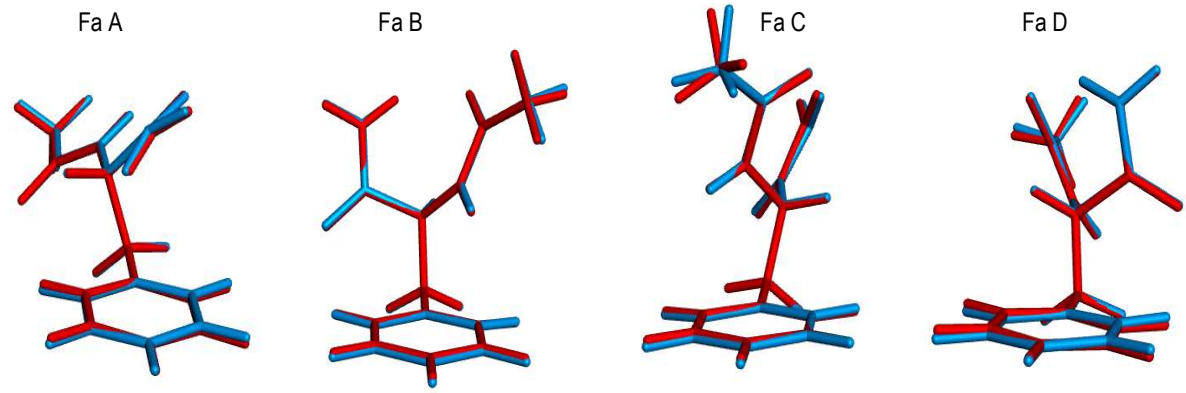



Figure S3-2: Comparison of the CC2/cc-pVDZ (blue) and CC2/cc-VTZ (red) optimized geometries of the $S_1$ state for the four Fa conformers. For each conformer, the phenyl rings have been overlapped.

Appendix S4: *Amide A* region frequencies of both the ground ($S_0$) and ππ* excited ($S_1$) states of the four Fa conformers.

| Conformer/State | | $NH_{phe}$ | $NH_{2\ sym.}$ | $NH_{2\ anti.}$ |
|---|---|---|---|---|
| **cc-pVDZ** | | | | |
| Fa A | $S_1$ | 3594 | 3546 | 3695 |
| | $S_0$ | 3583 | 3561 | 3713 |
| Fa B | $S_1$ | 3545 | 3469 | 3689 |
| | $S_0$ | 3589 | 3474 | 3690 |
| Fa C | $S_1$ | 3553 | 3473 | 3674 |
| | $S_0$ | 3597 | 3476 | 3675 |
| Fa D | $S_1$ | 3618 | 3498 | 3677 |
| | $S_0$ | 3618 | 3500 | 3677 |
| **Experiment** | | | | |
| Fa A | $S_1$ | 3434 | 3417 | 3535 |
| | $S_0$ | 3433 | 3426 | 3541 |
| Fa C | $S_1$ | 3439 | 3344 | 3514 |
| | $S_0$ | 3463 | 3345 | 3515 |
| **cc-pVTZ** | | | | |
| Fa A | $S_1$ | 3590 | 3549 | 3698 |
| | $S_0$ | 3588 | 3569 | 3712 |
| Fa B | $S_1$ | 3524 | 3454 | 3686 |
| | $S_0$ | 3577 | 3460 | 3687 |
| Fa C | $S_1$ | 3560 | 3463.9 | 3688 |
| | $S_0$ | 3614 | 3471 | 3683 |
| Fa D | $S_1$ | 3627 | 3501 | 3682 |
| | $S_0$ | 3627 | 3504 | 3681 |

Table S4: CC2/cc-pVXZ (X=D and T) *amide A* region frequencies ($cm^{-1}$) of both the ground ($S_0$) and ππ* excited ($S_1$) states of the four Fa conformers, together with the available experimental ones.

Appendix S5: ZPVE of both the ground ($S_0$) and ππ* excited ($S_1$) states according to the basis set for the four Fa conformers.

| | | | ZPVE (au) |
|---|---|---|---|
| $S_0$ | | A | 0.238368 |
| | cc-pVDZ | B | 0.239126 |
| | | C | 0.238832 |
| | | D | 0.239007 |
| | | A | 0.238510 |
| | cc-pVTZ | B | 0.239284 |
| | | C | 0.238696 |
| | | D | 0.239019 |
| $S_1$ | | A | 0.232277 |
| | cc-pVDZ | B | 0.232950 |
| | | C | 0.232845 |
| | | D | 0.232960 |
| | | A | 0.232336 |
| | cc-pVTZ | B | 0.233100 |
| | | C | 0.232730 |
| | | D | 0.232907 |

Table S5: ZPVE (au) of the optimized geometry of both the ground ($S_0$) and ππ* excited ($S_1$) states obtained at the CC2/cc-pVXD (X=D and T) levels for the four Fa conformers. The values for $S_0$ state at the DFT-D/TZVPP level are 0.233370 (A), 0.233924 (B), 0.233337 (C) and 0.233567 (D).

Appendix S6: Comparison of the CC2/cc-pVDZ optimized geometry of the $S_0$ and $S_1$ states of Fa B and D.

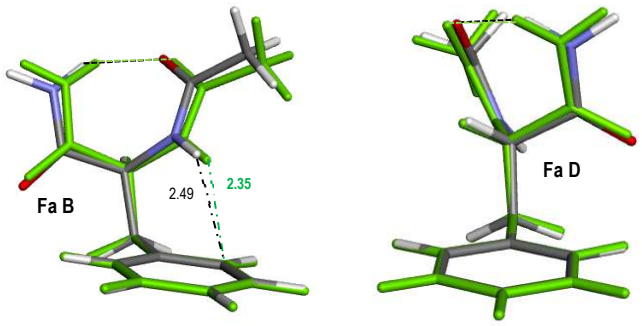



Figure S6: Comparison of the CC2/cc-pVDZ optimized geometries of the $S_0$ (atom-based colors) and $S_1$ (green) states for Fa B and D. For each conformer, the phenyl rings have been overlapped. Only distances (dash-dot) that vary significantly ($|d| > 0.01$ Å) between the ground and the excited state (see Table 1) are mentioned.

Appendix S7: Characteristic geometrical parameters of CC2/cc-pVDZ optimized geometry of both the ground ($S_0$) and lowest ππ* excited ($S_1$) states of the Fm, GFA, FFa and QFa conformers.

| Fm | | Dihedral angles (°)[a] | | | Intramolecular distances (Å) | |
|---|---|---|---|---|---|---|
| | | Φ | Ψ | $\chi^1$ | $d_{NH...O}$ | $d_{NH...\pi}$[b] |
| $S_0$ | A | -163 | 152 | 183 | 2.22 | 2.59 (2.94, 2.80) |
| | B | -83 | 60 | 41 | 1.92 | 2.34 (2.50, 3.10) |
| | C | -89 | 75 | -51 | 1.96 | 2.58 (3.48, 2.72) |
| $S_1$ | A | -165 | 153 | 177 | 2.23 | 2.54 (3.02, 2.54) |
| | B | -84 | 60 | 43 | 1.92 | 2.35 (2.33, 3.28) |
| | C | -89 | 75 | -50 | 1.96 | 2.52 (3.49, 2.53) |

Table S7-1: Characteristic geometrical parameters of CC2/cc-pVDZ optimized geometry of both the ground ($S_0$) and lowest ππ* excited ($S_1$) states of the Fm conformers.
[a] For the definition of the dihedral angles, see the Supporting Information (Figure S1).
[b] The NH…π bond is characterized by three distances: the distance of the $NH_{phe}$ (A conformer) or $NH_{C\text{-}term}$ (B and C conformer) hydrogen atom with the $C_\gamma$ carbon atom of the phenyl residue and given in parentheses by the two distances with the two $C_\delta$ carbon atoms ($C_\delta^{\text{to C-term}}$, $C_\delta^{\text{to N-term}}$) of the phenyl residue.

| GFa | | Dihedral angles (°)[a] | | | Dihedral angles (°) | | | Intramolecular distances (Å) | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\Phi_1$ | $\Psi_1$ | $\chi_1^1$ | $\Phi_2$ | $\Psi_2$ | $\chi_2^1$ | $d_{NH...O}$[b] | $d_{NH...\pi}$[c] |
| $S_0$ | A | -81 | 67 | | -84 | 79 | -59 | 1.96 - 1.99 | 2.87 (3.86, 3.03) |
| | A’ | -279 | 295 | | -81 | 80 | -57 | 1.92 - 1.99 | |
| | B | -71 | 346 | | -91 | 9 | 53 | 1.92 | 2.45 (3.29, 2.48) |
| | B’ | -304 | 220 | | -100 | 14 | 55 | 2.01 | 2.64 (3.47, 2.67) |
| | C | -115 | 166 | | -161 | 161 | 193 | 2.26 – 2.16 | 2.48 (2.67, 2.95) |
| $S_1$ | A | -80 | 71 | | -86 | 78 | -55 | 1.95 - 1.98 | 2.70 (3.54, 3.06) |
| | A’ | -279 | 293 | | -80 | 80 | -56 | 1.93 - 1.99 | |
| | B | -70 | 345 | | -91 | 9 | 53 | 1.91 | 2.44 (3.33, 2.40) |
| | B’ | -304 | 220 | | -95 | 10 | 57 | 2.01 | 2.60 (3.50, 2.56) |
| | C | -113 | 166 | | -165 | 155 | 181 | 2.28 – 2.21 | 2.52 (2.92, 2.63) |

Table S7-2: Characteristic geometrical parameters of CC2/cc-pVDZ optimized geometry of both the ground ($S_0$) and lowest ππ* excited ($S_1$) states of the GFa conformers.
[a] For the definition of the dihedral angles, see the Supporting Information (Figure S1). The residue 1 correspond to the first residue from the N terminal cap, the N-term.
[b] A and A’ conformers: the two distances correspond to the two $C_7$ hydrogen bond distances. B and B’ conformers, the distances is that of the $C_{10}$ H-bond. C conformer: the two distances correspond to the two $C_5$ H-bond distances.
[c] The NH…π bond is characterized by three distances: the distance of $NH_{phe}$ (B and B’ conformer) or $NH_2$ (C conformer) hydrogen atom with the $C_\gamma$ carbon atom of the phenylalanine residue and

given in parentheses by the two distances with the two $C_\delta$ carbon atoms ($C_\delta^{\text{to C-term}}$, $C_\delta^{\text{to N-term}}$) of the phenylalanine residue.

| FFa | | Dihedral angles (°)[a] | | | Dihedral angles (°) | | | Intramolecular distances (Å) | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\Phi_1$ | $\Psi_1$ | $\chi_1^1$ | $\Phi_2$ | $\Psi_2$ | $\chi_2^1$ | $d_{NH...O}$[b] | $d_{NH...\pi}$[c] |
| $S_0$ | A | -69 | 353 | 62 | -110 | 14 | 51 | 2.15 | [2.62, 2.54] -[2.42, 2.41] |
| | B | -153 | 27 | 51 | -95 | 75 | -41 | 2.05 | [2.70, 2.85] -[2.45, 3.00] |
| | C | -163 | 151 | 181 | -80 | 70 | 45 | 1.99 - 2.24 | 2.60 (2.94, 2.74) - 2.49 (3.41, 2.50) |
| $S_1$ | $A_1$ | -68 | 349 | 61 | -103 | 12 | 52 | 2.10 | [2.59, 2.47] - [2.43, 2.41] |
| | $A_2$ | -69 | 352 | 62 | -111 | 12 | 50 | 2.17 | [2.63, 2.56] - [2.38, 2.33] |
| | $B_1$ | -151 | 24 | 52 | -95 | 75 | -40 | 2.03 | [2.66, 2.78] - [2.43, 2.99] |
| | $B_2$ | -152 | 27 | 51 | -95 | 75 | -40 | 2.05 | [2.68, 2.87] - [2.42, 2.94] |
| | C | -160 | 161 | 182 | -73 | 67 | 41 | 1.96 - 2.21 | 2.46 (2.62, 3.15) - 2.24 (3.23, 2.17) |

Table S7-3: Characteristic geometrical parameters of CC2/cc-pVDZ optimized geometry of both the ground ($S_0$) and lowest ππ* excited ($S_1$) states of the FFa (Ac-Phe1-Phe2-$NH_2$) conformers.

[a] For the definition of the dihedral angles, see the Supporting Information (Figure S1). The residue 1 correspond to the first residue from the N-term.

[b] A, $A_1$ and $A_2$ conformers: the distance corresponds to the $C_{10}$ hydrogen bond distances. B, $B_1$ and $B_2$ conformers, the distance is that of the $C_7$ H-bond. C conformer: the distances correspond to the $C_7$ and the $C_5$ H-bond distances.

[c] The NH…π bond is characterized by two group of distances. A and B conformers: the distances of the $NH_{Phe1}$ hydrogen atom with the $C_\gamma$ and the $C_\delta^{\text{to N-term}}$ carbon atoms of Phe1 and those of the $NH_{phe2}$ hydrogen atom with the $C_\gamma$ and the $C_\delta^{\text{to N-term}}$ carbon atoms of Phe2. C conformer: the distances of the $NH_{Phe2}$ hydrogen atom with the $C_\gamma$ and the two $C_\delta$ carbon atoms ($C_\delta^{\text{to C-term}}$ and $C_\delta^{\text{to N-term}}$) of Phe1 and those of the same hydrogen atom with the $C_\gamma$ and the two $C_\delta$ carbon atoms ($C_\delta^{\text{to C-term}}$ and $C_\delta^{\text{to N-term}}$) carbon atoms of Phe2.

| QFa | | Dihedral angles (°)[a] | | | Dihedral angles (°) | | | Intramolecular distances (Å) | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\Phi_1$ | $\Psi_1$ | $\chi_1^1$ | $\Phi_2$ | $\Psi_2$ | $\chi_2^1$ | $d_{NH...O}$[b] | $d_{NH...\pi}$[c] |
| $S_0$ | A | -74 | 349 | -63 | -93 | 10 | 53 | 1.82 - 2.00 | 2.33 (3.25, |
| | B | -76 | 348 | 79 | -106 | 13 | -58 | | 2.47) |
| | C | -69 | 343 | 71 | -96 | 14 | 51 | 1.87 - 2.01 | 2.86 (2.63, |
| | | | | | | | | | 2.75) |
| | | | | | | | | 1.85 - 1.99 | 2.43 (3.19, |
| | | | | | | | | | 2.43) |
| $S_1$ | A | -65 | 340 | -58 | -106 | 15 | 47 | 1.83 - 2.06 | 2.37 (3.07, |
| | B | -69 | 341 | 67 | -112 | 16 | -50 | | 2.35) |
| | C | -69 | 342 | 69 | -97 | 14 | 48 | 1.84 - 2.04 | 2.92 (2.81, |
| | | | | | | | | | 2.36) |
| | | | | | | | | 1.85 - 2.00 | 2.39 (3.11, |
| | | | | | | | | | 2.39) |

Table S7-4: Characteristic geometrical parameters of CC2/cc-pVDZ optimized geometry of both the ground ($S_0$) and lowest ππ* excited ($S_1$) states of the QFa conformers.

[a] For the definition of the dihedral angles, see the Supporting Information (Figure S1). The residue 1 correspond to the first residue from the N-term.

[b] The two distances correspond to the $C_7$ and $C_{10}$ H-bond distances.

[c] The NH…π bond is characterized by three distances: the distance of the $NH_{Phe}$ (A and C conformer) or $NH_{2,chain}$ group (B conformer) hydrogen atom with the $C_\gamma$ (A and C conformer) or $C_\xi$ (B conformer) carbon atom of the phenylalanine residue and in parentheses, the two distances with the two $C_\delta$ carbon atoms ($C_\delta^{\text{to C-term}}$, $C_\delta^{\text{to N-term}}$) of the phenylalanine residue (A and C conformers) or with the $C_\delta^{\text{to C-term}}$ and $C\varepsilon^{\text{to C-term}}$ carbon atoms (B conformer).

Appendix S8: Comparison of the CC2/cc-pVDZ optimized geometry of the $S_0$ and $S_1$ states of the Fm, GFa, FFa and QFa conformers.

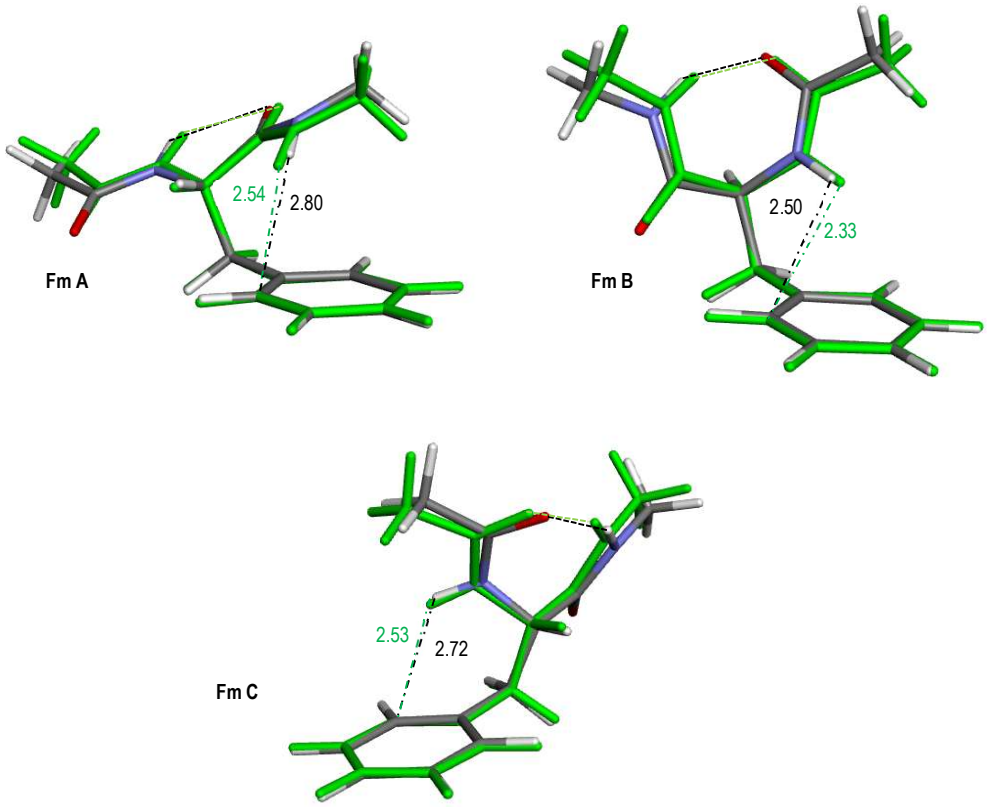



Figure S8-1: Comparison of the CC2/cc-pVDZ optimized geometry of the $S_0$ (atom-based colors) and $S_1$ states (green) of the Fm conformers. For each conformer, the phenyl rings have been overlapped. Only distances (dash-dot) that vary significantly ($|d| > 0.01$ Å) between the ground and the excited state (see Table S7.1) are mentioned.

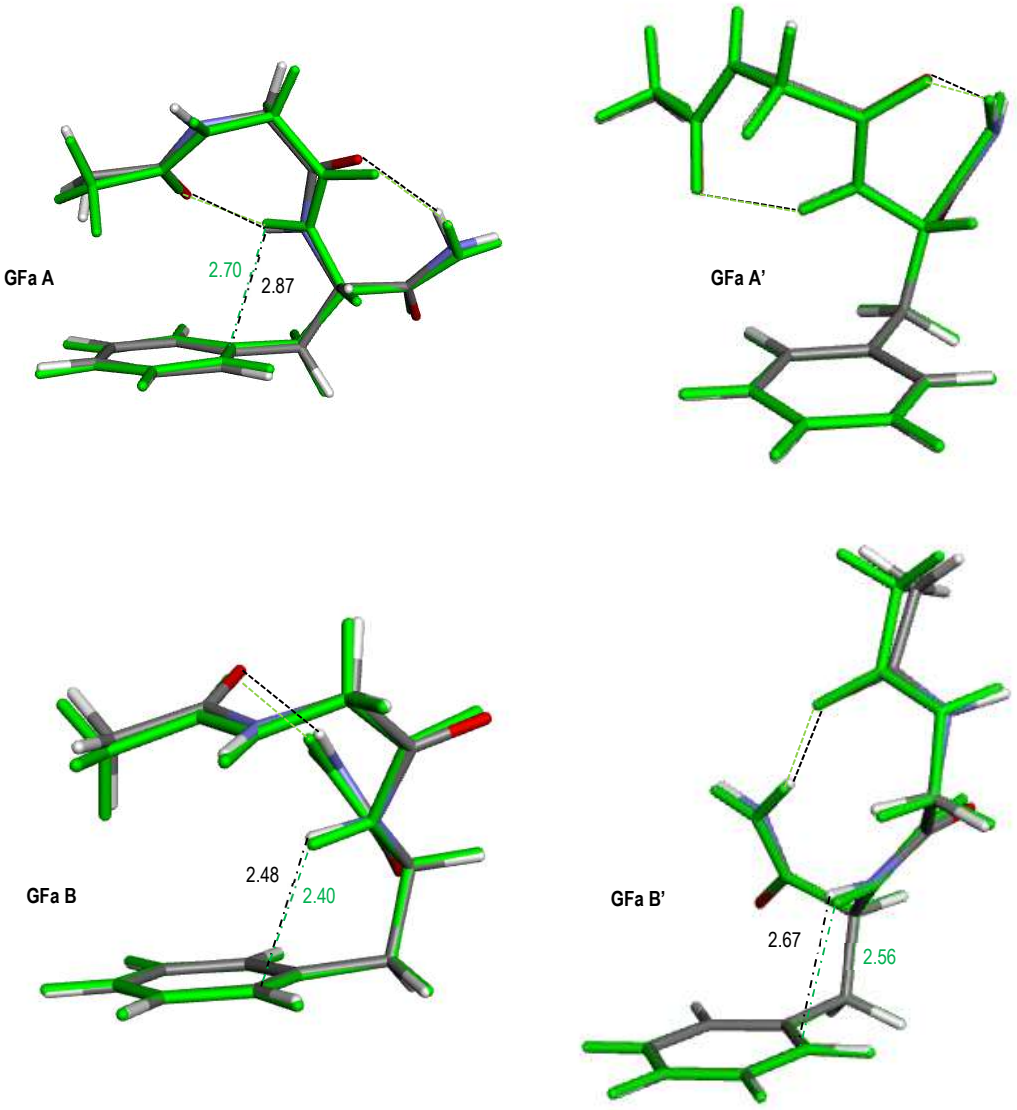



Figure S8-2: Comparison of the CC2/cc-pVDZ optimized geometry of the $S_0$ (atom-based colors) and $S_1$ states (green) of GFa A, A', B and B'. For each conformer, the phenyl rings have been overlapped. Only distances (dash-dot) that vary significantly ($|d| > 0.01$ Å) between the ground and the excited state (see Table S7.2) are mentioned.

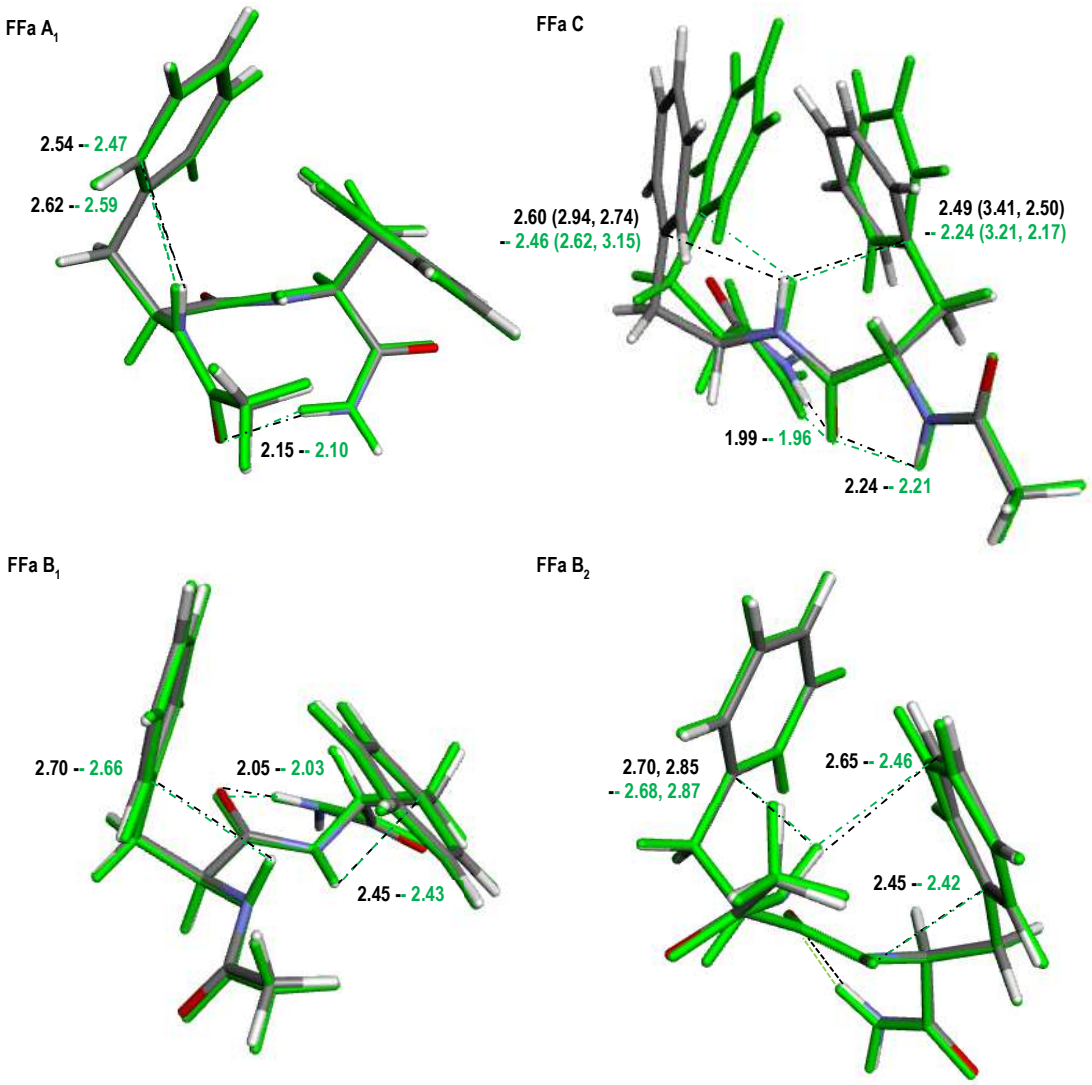



Figure S8-3: Comparison of the CC2/cc-pVDZ optimized geometry of the $S_0$ (atom-based colors) and $S_1$ states (green) of FFa $A_1$, $B_1$, $B_2$ and C. The phenyl rings have been overlapped for all conformers except FFa C for which it is not possible. In this latter case, the backbones until Phe1 are overlapped. Only distances (dash-dot) that vary significantly ($|d| > 0.01$ Å) between the ground and the excited state (see Table S7.3) are mentioned. In the case of FFa $B_2$, the distances of the $NH_{Phe1}$ hydrogen atom with the $C\varepsilon^{\text{to C-term}}$ carbon atoms of Phe2 are added.

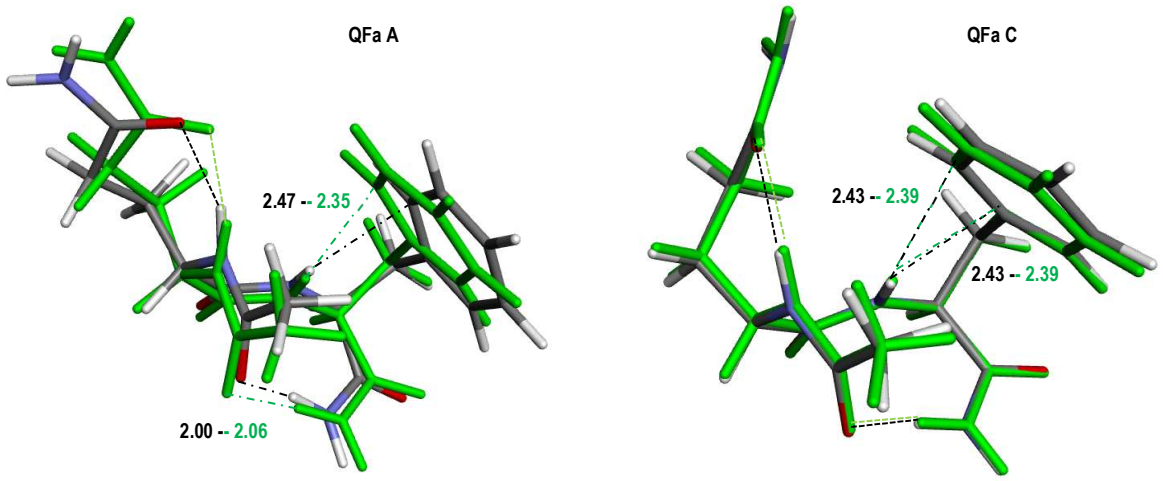



Figure S8-4: Comparison of the CC2/cc-pVDZ optimized geometry of the $S_0$ (atom-based colors) and $S_1$ states (green) of QFa A and C. For each conformer, the backbones have been overlapped. Only distances (dash-dot) that vary significantly ($|d| > 0.01$ Å) between the ground and the excited state (see Table S7.4) are mentioned.

Appendix S9: *Amide A* region frequencies of both the ground ($S_0$) and ππ* excited ($S_1$) states of the Fm, GFa, FFa and QFa conformers.

| Conformer/State | | $NH_{Phe}$ | $NH_{C\text{-}term}$ |
|---|---|---|---|
| Fm A | $S_1$ | 3592 | 3565 |
| | $S_0$ | 3597 | 3609 |
| Fm B | $S_1$ | 3540 | 3463 |
| | $S_0$ | 3587 | 3467 |
| Fm C | $S_1$ | 3552 | 3483 |
| | $S_0$ | 3599 | 3487 |
| **Experiment** | | | |
| Fm A | $S_1$ | 3433 | 3433 |
| | $S_0$ | 3433 | 3460 |
| Fm B | $S_1$ | 3401 | 3342 |
| | $S_0$ | 3433 | 3346 |

Table S9-1: CC2/cc-pVDZ *amide A* region frequencies ($cm^{-1}$) of both the ground ($S_0$) and ππ* excited ($S_1$) states of the Fm conformers, together with the available IR experimental ones ($cm^{-1}$).

| Conformer/State | | $NH_{Gln}$ | $NH_{Phe}$ | $NH_2$ sym./C-term | $NH_2$ anti./C-term | $NH_2$ sym./Chain | $NH_2$ anti./Chain |
|---|---|---|---|---|---|---|---|
| QFa A | $S_1$ | 3397 | 3570 | 3529 | 3705 | 3601 | 3759 |
| | $S_0$ | 3399 | 3582 | 3521 | 3700 | 3603 | 3761 |
| QFa B | $S_1$ | 3415 | 3599 | 3514 | 3689 | 3522 | 3695 |
| | $S_0$ | 3458 | 3593 | 3517 | 3688 | 3567 | 3722 |
| QFa C | $S_1$ | 3441 | 3560 | 3517 | 3699 | 3588 | 3740 |
| | $S_0$ | 3450 | 3576 | 3519 | 3699 | 3587 | 3739 |
| **Experiment** | | | | | | | |
| QFa A | $S_0$ | 3285 | 3409 | 3365 | 3519 | 3442 | 3562 |
| QFa B | $S_0$ | 3322 | 3445 | 3366 | 3512 | 3406 | 3527 |
| QFa C | $S_0$ | 3336 | 3440 | 3367 | 3514 | 3426 | 3557 |

Table S9-2: CC2/cc-pVDZ *amide A* region frequencies ($cm^{-1}$) of both the ground ($S_0$) and ππ* excited ($S_1$) states of the QFa conformers, together with the available IR experimental ones ($cm^{-1}$).

| Conformer/State | | $NH_{Gly}$ | $NH_{Phe}$ | $NH_{2sym.}$ | $NH_{2anti.}$ |
|---|---|---|---|---|---|
| GFa A | $S_1$ | 3635 | 3434 | 3475 | 3675 |
| | $S_0$ | 3642 | 3452 | 3480 | 3675 |
| GFa A' | $S_1$ | 3645 | 3443 | 3472 | 3668 |
| | $S_0$ | 3647 | 3448 | 3474 | 3670 |
| GFa B | $S_1$ | 3643 | 3559 | 3538 | 3709 |
| | $S_0$ | 3644 | 3596 | 3541 | 3710 |
| GFa B' | $S_1$ | 3626 | 3582 | 3530 | 3694 |
| | $S_0$ | 3626 | 3603 | 3529 | 3695 |
| GFa C | $S_1$ | 3572 | 3569 | 3546 | 3695 |
| | $S_0$ | 3567 | 3571 | 3561 | 3709 |
| **Experiment** | | | | | |
| GFa A | $S_1$ | 3492 | 3302 | 3358 | 3510 |
| | $S_0$ | 3494 | 3320 | 3355 | 3519 |
| GFa A' | $S_0$ | 3495 | 3322 | 3353 | 3517 |
| GFa B | $S_0$ | 3494 | 3445 | 3391 | 3521 |

| | | | | | |
|---|---|---|---|---|---|
| GFa B' | $S_1$ | 3494 | 3423 | 3387 | 3519 |
| | $S_0$ | 3493 | 3441 | 3385 | 3518 |
| GFa C | $S_1$ | 3445 | 3408 | 3416 | 3535 |
| | $S_0$ | 3444 | 3405 | 3425 | 3541 |

Table S9-3: CC2/cc-pVDZ *amide A* region frequencies ($cm^{-1}$) of both the ground ($S_0$) and ππ* excited ($S_1$) states of the GFa conformers, together with the IR available experimental ones ($cm^{-1}$).

| Conformer/State | | $NH_{Phe1}$ | $NH_{Phe2}$ | $NH_{2sym.}$ | $NH_{2anti.}$ |
|---|---|---|---|---|---|
| FFa $A_1$ | $S_1$ | 3540 | 3587 | 3545 | 3711 |
| FFa $A_2$ | $S_1$ | 3576 | 3562 | 3547 | 3713 |
| FFa A | $S_0$ | 3581 | 3596 | 3549 | 3714 |
| FFa $B_1$ | $S_1$ | 3509 | 3552 | 3483 | 3681 |
| FFa $B_2$ | $S_1$ | 3487 | 3537 | 3489 | 3683 |
| FFa B | $S_0$ | 3520 | 3556 | 3489 | 3683 |
| FFa C | $S_1$ | 3595 | 3467 | 3458 | 3681 |
| | $S_0$ | 3607 | 3541 | 3488 | 3692 |
| **Experiment** | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| FFa $A_1$ | $S_1$ | 3414 | 3438 | 3390 | 3524 |
| FFa $A_2$ | $S_1$ | 3446 | 3414 | 3390 | 3524 |
| FFa A | $S_0$ | 3447 | 3438 | 3391 | 3524 |
| FFa B | $S_0$ | 3412 | 3430 | 3357 | 3514 |
| FFa C | $S_0$ | 3445 | 3418 | 3382 | 3518 |

Table S9-4: CC2/cc-pVDZ *amide A* region frequencies ($cm^{-1}$) of both the ground ($S_0$) and $\pi\pi^*$ excited ($S_1$) states of the FFa conformers, together with the IR available experimental ones ($cm^{-1}$).

Appendix S10: Experimental vs CC2/cc-pVDZ calculated harmonic *amide A* region frequencies of both the $S_0$ and $S_1$ states of the series of capped peptides and the corresponding mode-dependent linear ($\nu_{exp.} = a\nu_{theo.} + b$) scaling functions.

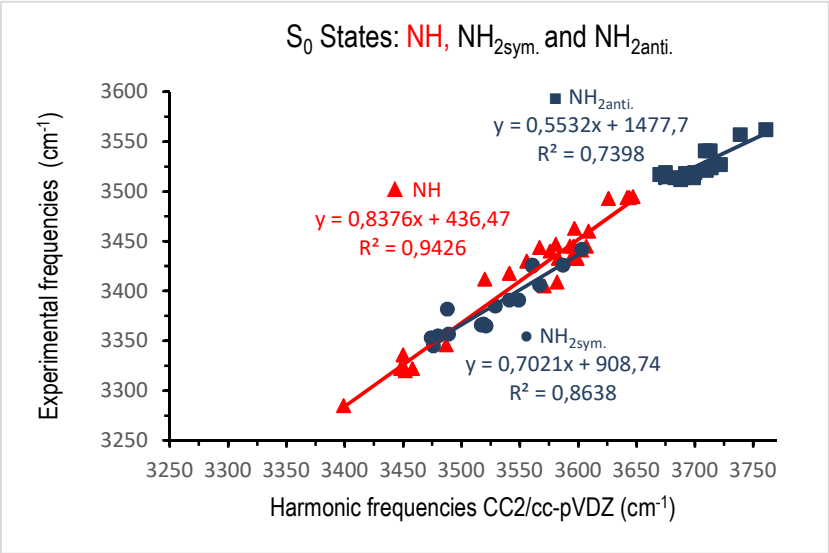



Figure S10-1: Experimental vs CC2/cc-pVDZ calculated harmonic *amide A* region frequencies of the $S_0$ states of the series of capped peptides and the corresponding mode-dependent linear ($\nu_{exp.} = a\nu_{theo.} + b$) scaling functions.

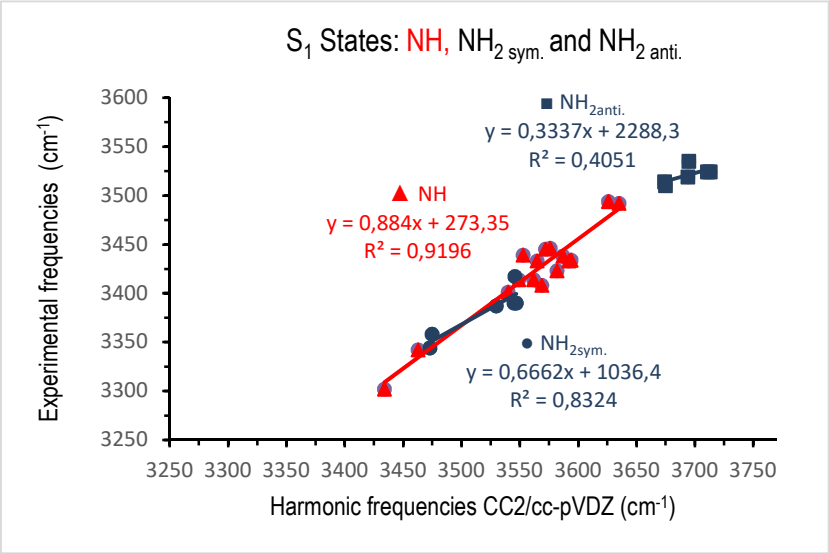


Figure S10-2: Experimental vs CC2/cc-pVDZ calculated harmonic *amide A* region frequencies of the $S_1$ states of the series of capped peptides and the corresponding mode-dependent linear ($\nu_{exp.} = a\nu_{theo.} + b$) scaling functions.

S11: DFT-D structures of the ground state of QFa A, B and C

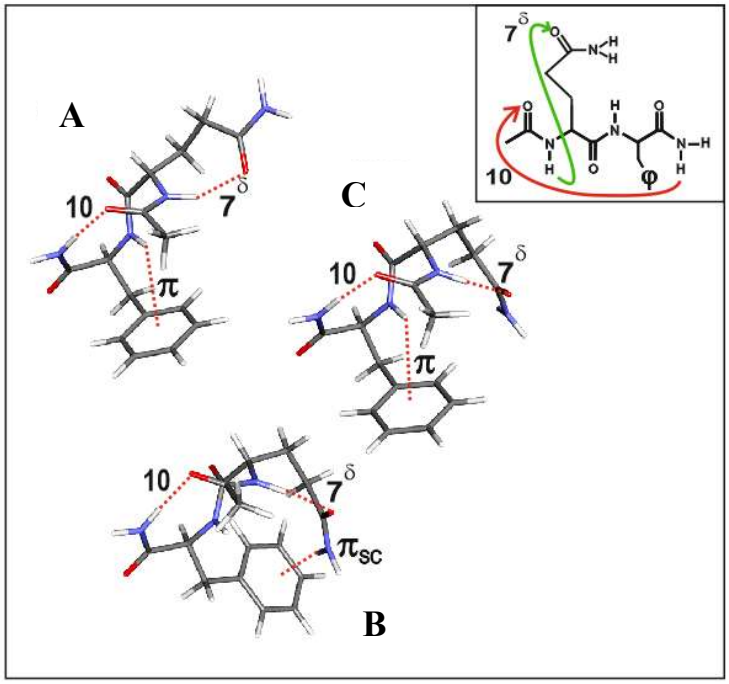



Figure S11: B97-D3 structures of the three most stable forms of QFa, which account for the conformer-selective IR spectra recorded (shown in Fig. 4). In these three forms the peptide backbone exhibits a β-turn structure stabilized by C10 H-bond and by a main chain/side chain H-bond, labelled $7^{\delta}$ (see molecule sketch in the insert). The 3 conformers differ by the arrangement of the Gly side chain relative to the backbone.

References:

1. Chin, W.; Mons, M.; Dognon, J. P.; Piuzzi, F.; Tardivel, B.; Dimicoli, I. Competition between local conformational preferences and secondary structures in gas-phase model tripeptides as revealed by laser spectroscopy and theoretical chemistry. *Physical Chemistry Chemical Physics* **2004,** *6* (10), 2700-2709.
2. Alaudin, M.; Vaquero-Vara, V.; Habka, S.; Tardivel, B.; Gloaguen, E.; Mons, M., unpublished resutls.