Structure of isolated tyrosyl-glycyl-glycine tripeptide. A comparative conformational study with peptides containing an aromatic ring

Research Article

Exequiel E. Barrera Guisasola, Marcelo F. Masman, Ricardo D. Enriz, Ana M. Rodríguez*

Department of Chemistry. Faculty of Chemistry, Biochemistry and Pharmacy, National University of San Luis, Chacabuco 917 - 5700, San Luis, Argentina

Abstract: The potential energy surface (PES) of tyrosyl-glycyl-glycine (YGG) tripeptide in solution was explored using EDMC (Electrostatically Driven Monte Carlo) and in the gas-phase by means of ab initio quantum chemical calculations. The theoretical computational analysis revealed that this tripeptide possesses a significant molecular flexibility. A C_7 backbone conformation was the most energetically preferred for the central Gly residue, using both methodologies. Some new stable conformers that have not been previously reported were identified in the gas phase as well. This study points out the interplay of backbone and side-chain contributions in determining the relative stabilities of energy minima. In addition, the peptide backbone of YGG was compared with other small peptides containing aromatic side-chains (Phe-Gly-Gly and Trp-Gly-Gly). The comparison with experimental X-ray results was also satisfactory.

Keywords: *Ab initio calculations • EDMC calculations • YGG • gas-phase chemistry*
© Versita Sp. z o.o.

1. Introduction

The formation of protein secondary and tertiary structure depends on the chemical environment as well as on chemical and physical constraints imposed by the individual properties of the protein building blocks, that is, the twenty amino acids and their sequence in the polypeptide chain [1-7]. A fundamental question that has not been satisfactorily answered yet is how side-chains and backbones interact in peptides. Side-chain folding is not only interesting, but also important because side-chain orientation can influence backbone folding *via* side-chain/backbone and/or side-chain/side-chain interactions. In spite of considerable effort to elucidate the nature of the forces that determine the conformational states of amino acids in a particular context of a protein or short peptide, the issue remains unsolved [8].

The study of peptide chains in the gas-phase are useful because the information obtained can be helpful for understanding the protein-folding process. Evaluation of non-covalent interactions in such extended systems like peptides is difficult and represents one of the most challenging tasks in computational chemistry

today [9]. The main reason being that peptides are very flexible systems showing an extensive conformational landscape.

Tripeptides can represent efficient building blocks for protein-structure prediction. A tripeptide constitutes a minimal model containing all the important forces resulting in distinct conformational states of the participating amino acids. A particular tripeptide contains all the necessary factors that influence the behaviour of the rotamer, maintain the interaction of side-chain with the backbone, and take into account φ and ψ preferences of allowed regions in protein structures. Recent findings show that there is relative structural rigidity between α and β atoms in some tripeptides, possibly due to an intramolecular stabilizing interaction [10]. In addition to their biological importance as peptides and proteins building blocks, these molecules are also of interest from a purely chemical point of view. They are able to form typical multiconformer systems with numerous local minimum structures associated with different conformational arrangements of the main chain and the side-chain.

Tripeptides containing at least one aromatic ring are of special interest due to a possible strong interaction

* E-mail: amrodri@unsl.edu.ar

between delocalized π electrons of the aromatic ring with peptide bonds. The strength of the weak polar interaction between the side-chain aromatic ring of an amino acid and an amide backbone of a polypeptide (Ar-HN interaction) can be as high as 16 kJ mol[-1] [11]. This is comparable with the strength (8-29 kJ mol[-1]) of a conventional hydrogen bond. The conformation of polypeptide fragments containing Ar-HN interactions can depend on factors including the amino acid sequence, the hydrophobicity or hydrophilicity of the local environment, the degree of solvation, and the structural flexibility of the polypeptide fragment [12-16]. There are four aromatic amino acids found in proteins, namely, phenylalanine (F), tyrosine (Y), tryptophan (W), and histidine (H). The structure of these amino acids and of some aromatic small peptides such as phenylalanyl-glycyl-glycine (FGG), tryptophyl-glycine (WG) and tryptophyl-glycyl-glycine (WGG), have already been reported by means of molecular dynamic simulations combined with high-level correlated *ab initio* quantum chemical calculations [17-19]. The structure of the tyrosyl-glycyl-glycine (YGG) peptide has recently been investigated using a combination of different strategies that employ a hierarchy of electronic structure theory [20]. Although three different methodologies to explore the conformational energy landscape of YGG were used in this study, the authors stated that the large changes in the relative stability of the obtained conformers indicate that B3LYP/6-31+G(d) as well as MP2/6-31+G(d) may not give correct structures and energetics for molecules containing interactions with π-electron clouds. Also, B3LYP/6-31+G(d) and MP2/6-31+G(d) calculations predicted markedly different structures for the tyrosil-glycine (YG) dipeptide [21]. The same conformational behaviour discrepancy between these two methods was observed for a number of other molecules containing an aromatic ring in addition to a flexible side chain, WG [17-19] and WGG [18,19]. Further investigation suggested that the very different structures obtained by B3LYP and MP2 are likely caused by dispersion (a true physical effect, underestimated by B3LYP) as well as large basis set superposition errors (BSSE, an artificial attraction) in the MP2 calculations [22]. Surprisingly, Hartree-Fock (HF) calculations – which do not describe dispersion interaction either – yielded an unambiguous triple-well potential for the conformers of YG dipeptide [21].

These previous considerations led us to reinvestigate the potential energy surface (PES) of the YGG tripeptide using a different strategy, such as the multidimensional conformational analysis (MDCA) [23,24]. This methodology has been used in previous works and has been shown to reduce the risk of missing important low-lying conformers [25-28]. In the present paper, we explore the full conformational space of YGG using two different procedures with modest computational requirements, namely, empirical potential and quantum-chemical methods. Such a reduced treatment is not expected to explain the entire behaviour of YGG in *vacuo* and in solution. Our aim in this study is less ambitious, we wish to predict by austere theoretical methods the structure and relative stability of the YGG tripeptide, and more importantly, to compare these results with others aromatic ring-related peptides such as FGG and WGG. This systematic structural study of aromatic ring-related peptides might permit elucidation of their structural features relationship. Finally, we compare our results with previously reported experimental data obtained from X-ray diffraction studies [29].

2. Computational Methods

2.1 Stochastic Conformational Search. EDMC calculations

The conformational space was explored using the method previously employed by Liwo *et al.* [30] that included the electrostatically driven Monte Carlo (EDMC) method [31,32] implemented in the ECEPPAK [33] package. Conformational energy was evaluated using the ECEPP/3 force field [34]. Hydration energy was evaluated using a hydration-shell model with a solvent sphere radius of 1.4 Å and atomic hydration parameters that have been optimized using non-peptide data (SRFOPT) [35,36]. In order to explore the conformational space extensively, 10 different runs were carried out, each of them with a different random number. Therefore, a total of 5,000 accepted conformations were collected. Each EDMC run was terminated after 500 energy-minimized conformations had been accepted. The parameters controlling the runs were the following: a temperature of 298.15 K for the simulations, a temperature jump of 50,000 K, and the maximum number of allowed repetitions of the same minimum was 50. Further paramenters were the maximum number of electrostatically predicted conformations per iteration was 400. the maximum number of random-generated conformations per iteration was 100, and the fraction of random/electrostically predicted conformations was 0.30. Also the maximum number of steps at one increased temperature was 20 with the maximum number of rejected conformations until a temperature jump was executed was 100. Finally, only trans peptide bonds ($\omega \cong 180°$) were considered. All accepted conformations were then clustered into families using the program ANALYZE [33] by applying the minimal-tree clustering algorithm for separation, using

all heavy atoms, energy threshold of 30 kcal mol^{-1}, and RMSD of 0.75 Å as separation criteria. This clustering step allows a substantial reduction of the number of conformations and the elimination of repetitions. A more detailed description of the procedure used here is given in section 4.4 Computational Methods of [37].

2.2 Systematic Conformational Search. Ab initio quantum chemical calculations

All gas-phase computations were carried out using the Gaussian 03 program package [38]. Each structure was optimized using the *ab initio* [39] restricted Hartree-Fock (RHF) [40] method with the split valence 6-31G(d) basis set [41,42]. The RHF/6-31G(d) geometry optimized structural parameters were then used as the input of single-point calculations at RHF/cc-pVTZ//RHF/6-31G(d) and RHF/aug-cc-pVTZ//RHF/6-31G(d) levels of theory [43,44] in order to obtain more reliable stability data. Additionally, each stable conformer was subjected to frequency calculations at the RHF/6-31G(d) level of theory to confirm their identities as being true minima.

2.3 Nomenclature of the structures

The used nomenclature describes both the order of the conformers according to their energies and their peptide backbone geometry described in terms of its principal torsional angles (see Fig. 1). The pattern followed for the peptide name is yggNN[$\psi_1 \cdot \varphi_2 \cdot \psi_2 \cdot \varphi_3 \cdot \psi_3$]. NN stands for the energetic position of the conformer according to the energy scale. ψ_1, φ_2, ψ_2, φ_3 and ψ_3 are the principal backbone torsional angles.
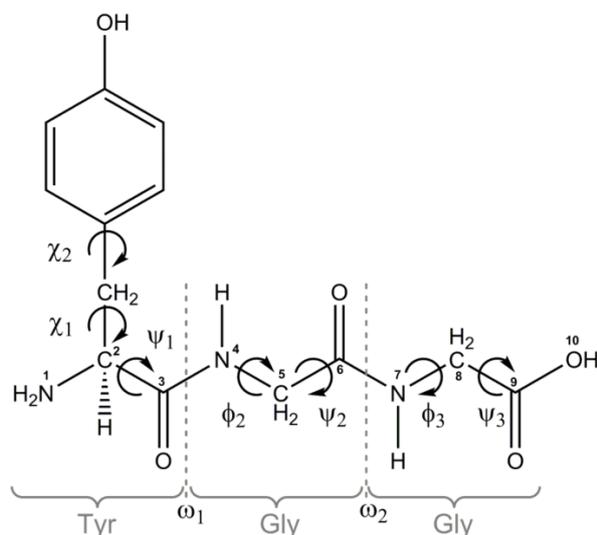
Figure 1. Numbering system employed for YGG with definition of backbone and side-chain dihedral angles.

3. Results and Discussion

3.1 Conformational study of YGG in gas-phase

The overall expression of the PES for the backbone of the YGG peptide is a function of five variables E = E (ψ_1, φ_2, ψ_2, φ_3, ψ_3) considering only the *trans*-peptide bonds (*i.e.*, $\omega \approx 180°$) (Fig. 1). Also, there are two torsional angles in the side-chain labelled χ_1 and χ_2. The X-ray diffraction experimental values ($\chi_1 = 74.2°$ and $\chi_2 = -91.6°$) [26] were used as starting values for these torsional angles. As three minima ($g+$, a, $g-$) are expected for each variable according to MDCA, this may lead to the existence of $3^5 = 243$ conformers. To truly characterize the full ensemble of the molecular conformers, *syn* and *anti* conformations of the carboxyl group (O=C-O-H) were also taken into account (0° and 180°, respectively).

Using 243 MDCA-predicted geometries with *syn* conformation of the carboxylic group as input, a total of 83 conformers were located on the PES at the RHF/6-31G(d) level of theory, instead of the 243 expected structures. These conformations may be grouped into two families which differ in the orientation of the torsional angle ψ_3 (approximately 0° and 180°). The conformations with $\psi_3 \cong 180°$ were energetically preferred in all the cases. The results of full optimization of the YGG peptide (*syn* isomer) at the RHF/6-31G(d) level of theory including geometric parameters, total energies, and relative energies are given in Supplemental Table 1. of the Supplementary Data. The geometry optimizations of the conformers with *anti* conformation of the carboxylic group resulted in 66 structures (Supplemental Table 2. in the Supplementary Data). In this case, the torsional angle ψ_3 adopted $g+$ and $g-$ as preferred spatial orientations (namely $\psi_3 \cong 60°$ and $\psi_3 \cong -60°$, respectively). All the structures presented in Supplemental Tables 1 and 2. lie within an interval of energy of about 14 kcal mol^{-1}, indicating that this tripeptide possesses a significant molecular flexibility in the gas-phase.

Table 1 shows the geometric parameters (five backbone torsional angles), total energies, and relative energies for the most 13 stable minima localized in the PES of the YGG tripeptide by means of RHF/6-31G(d) calculations considering an energy window of 3 kcal mol^{-1}. These 13 structures are depicted in Fig. 2. φ_2 and ψ_2 torsional angles determine the position of the carboxyl group and therefore, the possible formation of an OH$_{carb}$···O=C$_6$ intramolecular hydrogen bonding (H-bond). The OH$_{carb}$···O=C$_6$ hydrogen bond interaction involves two different families of conformers: those with a more bent (*via* an OH$_{carb}$···O=C$_6$ intramolecular H-bond) peptide backbone (*i.e.*, ygg*01* and ygg*02*, see Fig. 2)

Table 1. Torsional angles and total energy values for the 13 most stable conformers of YGG optimized at RHF/6-31G(d) level of theory. Calculated relative energies (DE$_{rel.}$)[a] are also shown. Total energies were given in hartrees, and the relative energies in kilocalories per mole.

Conformers	ψ_1	ϕ_2	ψ_2	ϕ_3	ψ_3	Energy (hartree)	$\Delta E_{rel.}$ (kcal mol^{-1})
ygg01	-15.48	119.46	-36.23	-80.05	65.71	-1039.9045	0.00
ygg02	-114.54	-116.08	5.15	-78.41	64.48	-1039.9043	0.14
ygg03	32.07	86.18	-69.98	-73.80	179.23	-1039.9039	0.35
ygg04	3.27	118.52	-14.45	84.08	-172.59	-1039.9026	1.18
ygg05	105.30	-89.63	59.31	76.97	-59.31	-1039.9024	1.31
ygg06	10.79	83.00	-71.65	-77.42	63.71	-1039.9009	2.27
ygg07	-10.08	-83.73	75.53	74.35	-177.56	-1039.9005	2.52
ygg08	4.56	112.36	-13.12	72.14	20.02	-1039.9003	2.60
ygg09	160.86	80.16	-80.23	-78.08	177.32	-1039.9003	2.64
ygg10	130.47	-84.42	66.84	77.91	-58.53	-1039.9002	2.68
ygg11	-37.91	85.36	-63.69	-78.12	56.10	-1039.9001	2.73
ygg12	5.71	89.78	3.59	77.38	-63.20	-1039.8999	2.88
ygg13	0.25	-179.93	178.61	-179.81	-179.87	-1039.8998	2.90

[a] The global minimum corresponds to the final geometry 01 (see Supplemental Table 2) having a total energy of -1039.9044926 hartree. This value is taken as reference value, corresponding to a relative energy 0.00 kcal mol^{-1}
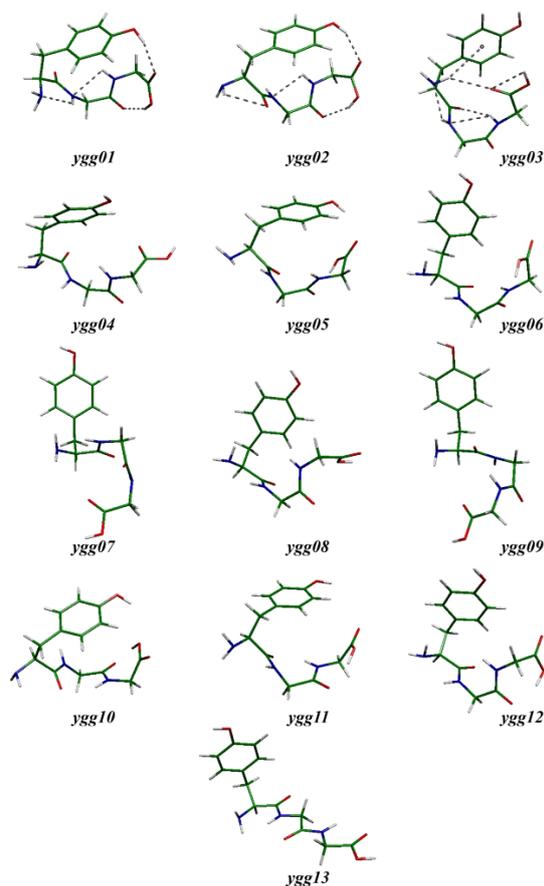
Figure 2. RHF/6-31G(d) geometries for the 13 most stable structures of YGG tripeptide. Intramolecular H-bonds are also shown for the three most stable structures.

and those with a fairly stretched or less bent (without an OH$_{carb}$…O=C$_6$ intramolecular H-bond) peptide backbone (*i.e.*, ygg03, see Fig. 2). It is also interesting that the usually preferred *syn* conformation of the carboxyl group (O=C-O-H = 0°) is lost when the OH$_{carb}$…O=C$_6$ intramolecular H-bond is formed, and instead, an *anti* configuration (O=C-O-H = 180°) is adopted (compare ygg01 and ygg03). The usually preferred *syn* configuration of the carboxyl group may be sacrificed in favor of the stability gained from the intramolecular H-bonds formed. RHF/6-31G(d) calculations predict folded conformations as the highly preferred forms for the backbone of the YGG tripeptide, showing a strong C$_3$=O…HN$_7$ intramolecular H-bond (the so-called C$_{eq}$7 or C$_{ax}$7) stabilizing the majority of the conformations. Only one conformation, ygg13, adopted an extended backbone form with a N$_4$H…C=O$_6$ intramolecular H-bond (the so-called C$_5$), and this fact could explain the lowest stability of this conformer.

The first three structures presented in Table 1 lie within a short interval of energy of 0.35 kcal mol^{-1}. The global minimum conformation (ygg01) is stabilized by four intramolecular H-bonds, namely: OH$_{carb}$…O=C6, N7H…N(H$_2$), N$_4$H…N(H$_2$) and OH$_{Tyr}$…O=C$_9$ side-chain/backbone H-bond (see Fig. 2). The first local minimum, ygg02, is only 0.14 kcal mol^{-1} less stable than the global minimum (see the ΔE_{rel} column in Table 1). Both structures essentially differ in the orientation of the tyrosyl residue and the first peptide bond (see Fig. 2). More specifically, comparing ygg01 and ygg02, ψ_1 and ϕ_2 angles are different, and consequently, the H-bond

pattern of both structures is not the same. For the first local minimum, ygg*02*, the $OH_{car}b…O=C_6$, $N_7H…N(H_2)$ and $OH_{Tyr}…O=C_9$ H-bond interactions remain, but in this case no $N_4H…N(H_2)$ intramolecular H-bond is formed. Instead, the carbonyl of the first peptide bond is interacting with one of the hydrogen atoms of the amino group (*i.e.*, $C_3=O… N(H_2)$ interaction). The conformer ygg*03*, corresponding to the second local minimum, shows a different H-bond pattern than ygg*01* and ygg*02* (Fig. 2). This conformation adopts the maximum number of H-bonds: $C_3=O…HN_7$ (typical $C_{eq}7$, $\varphi_2 = 86.18$ and $\psi_2 = -69.98$), $C_9=O…N(H_2)$, $OH_{carb}…O=C_9$ (*syn* isomer), $N_7H…N_4H$ and $N_4H…N(H_2)…\pi$ (cooperative hydrogen bonded "daisy chain" interaction). This hydrogen bond cooperativity has been already observed in some other amino acids [18, 45].

The 13 structures which lie within 3 kcal mol⁻¹ of relative energy were considered for the subsequent single point calculations with Dunning's correlation-consistent (cc-pVTZ) and augmented correlation-consistent (aug-cc-pVTZ) basis sets (see Table 2).

The order of structures, as well as the energetic differences among structures, remains almost the same at all levels of theory. Concerning the RHF/cc-pVTZ//RHF/6-31G(d) and RHF/aug-cc-pVTZ//RHF/6-31G(d) calculations (see Table 2), only one notable discrepancy in the order of conformers was obtained. The RHF/6-31G(d) last local minimum (ygg*13*) migrates to the fourth local minimum and to the second local minimum at RHF/cc-pVTZ//RHF/6-31G(d) and RHF/aug-cc-pVTZ//RHF/6-31G(d) levels of theory, respectively. For the rest of the conformers it might seem, at first sight, that the structures are placed in different positions. However, these

discrepancies are not significant since the energetic differences among the structures are very small (*e.g.*, the first three minima case). It should be noted that the energy intervals change when passing from the first (2.90 kcal mol⁻¹) to the last column (2.23 kcal mol⁻¹). Thus, a higher energetic cutoff has to be considered at the inferior level of theory to select the same number of conformers. From the analysis of Table 2, it seems that the RHF/6-31G(d) level of theory gives similar information compared with the higher level calculations at a lower computational cost.

Toroz *et al*. [20] found the 20 most stable conformers for the YGG tripeptide (based on the MP2 single-point energy calculations with inclusion of B3LYP zero-points energies). It is noteworthy that the global minimum found with B3LYP and MP2 remains the global minimum from our study with HF calculations, providing some confidence that this conformer may indeed be the most stable one. Also 8 of the 13 most stable conformers identified in the present work were included on the 20 most stable conformers according to the MP2 single point calculations. In addition, among the 20 most stable structures based on MP2 calculations, 19 conformers were found by using MDCA. It is interesting to note that the conformers ygg*02*, ygg*04*, ygg*07* and ygg*09* are missing minima in the B3LYP/MP2 landscape explored by Toroz and co-workers [20]. Among them, two of these new conformations should be distinguished because they showed low relative energies. Therefore, ygg*02* is our first local minimum (0.14 kcal mol⁻¹ less stable than the global minimum), while the structure of ygg*04* shows high correlation with the first local minimum on the PES of the aromatic ring-related peptides FGG and WGG reported by Valdés *et al*. [18]. On the other

Table 2. Relative energies for the most stable conformers of the YGG tripeptide evaluated with various basis sets (structures are labeled according to Fig. 2). The relative energies were given in kilocalories per mole.

Conformers	RHF/6-31G(d) (kcal mol⁻¹)	Conformers	RHF/cc-pVTZ (kcal mol⁻¹)	Conformers	RHF/aug-cc-pVTZ (kcal mol⁻¹)
ygg01	0.00	ygg02	0.00	ygg02	0.00
ygg02	0.14	ygg03	0.20	ygg13	0.20
ygg03	0.35	ygg01	0.54	ygg03	0.42
ygg04	1.18	ygg13	0.55	ygg04	0.63
ygg05	1.31	ygg04	0.58	ygg01	0.81
ygg06	2.27	ygg05	1.02	ygg05	0.95
ygg07	2.52	ygg07	1.14	ygg07	1.09
ygg08	2.60	ygg06	1.63	ygg06	1.58
ygg09	2.64	ygg09	1.73	ygg09	1.70
ygg10	2.68	ygg11	2.00	ygg11	1.85
ygg11	2.73	ygg08	2.03	ygg08	2.09
ygg12	2.88	ygg10	2.16	ygg10	2.12
ygg13	2.90	ygg12	2.27	ygg12	2.23
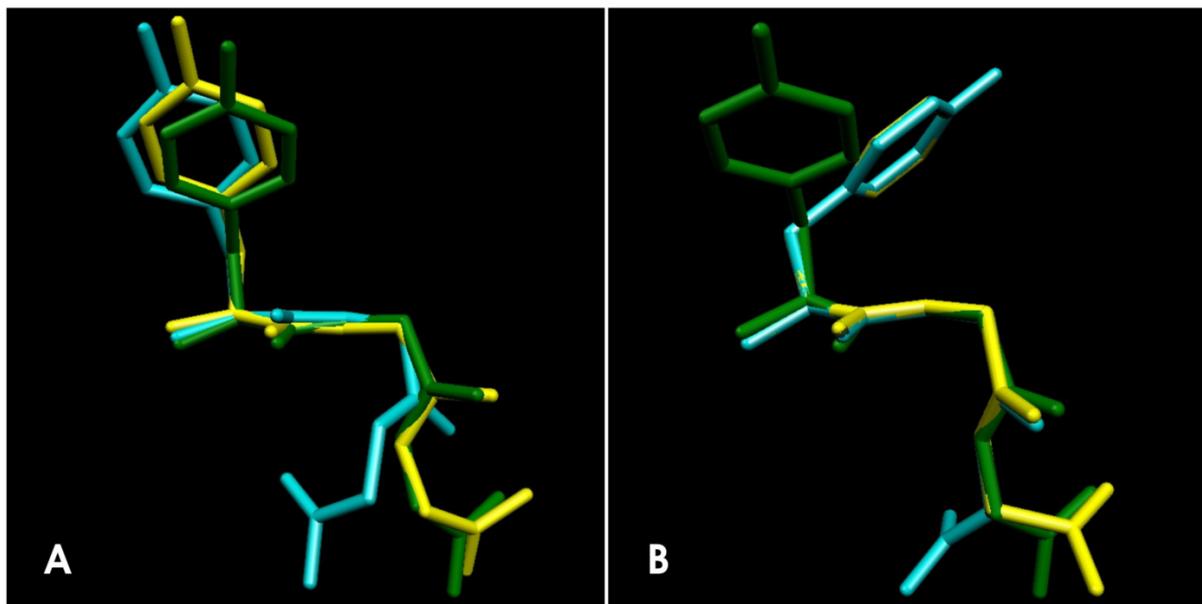
Figure 3. Overlapped geometries of the structure reported by Pichon-Pesme et al. [29] (in green) and; A) two YGG conformers obtained at RHF/6-31G(d) level of theory (ygg09 conformation in light blue and geometry number 18 in yellow) B) two YGG conformers obtained at EDMC/SRFOPT/ECCEP/3 level of theory. (Family 11 in light blue and family 12 in yellow). All hydrogen atoms are omitted for clarity.

hand, whereas there is a significant discrepancy in the relative energies resulting from the B3LYP and MP2 calculations, the energetic differences among structures based on HF calculations remain almost the same at all levels of theory.

Valdés *et al.* found a systematic structural behaviour for the tripeptides containing two glycyl residues (FGG and WGG) [18]. They argued that it is expected that the most stable conformers of other similar peptides may present equivalent geometries. Thus, the 13 most stable conformers on the PES of the YGG tripeptide have been compared with those most stable conformers on the PES of the FGG and WGG tripeptides [18] obtained using *ab initio* SCC-DFTB-D (self-consistent charge, density functional tight binding method with empirical dispersion energy) molecular dynamic simulations combined with high-level correlated *ab initio* quantum calculations (RI-MP2/cc-pVXZ, X = D, T). The geometries of the peptide backbones and side-chains are in general the same, but ygg*01* (global minimum) and ygg*02* (the first local minimum) conformations are not stable conformers in the PES of the FGG and WGG tripeptides. This difference may be caused by the presence of the *p*-OH substitution in the benzene ring of residue Tyr in the YGG tripeptide which may affect, explicitly or implicitly, its molecular structure. Such an effect may be related to the formation of the OH_{Tyr}…O=C9 side-chain/backbone H-bond (see Fig. 2). It is tempting to explain the higher stability of ygg*01* and ygg*02* conformations in terms of the presence of this intramolecular interaction. It is worth mentioning that the second local minimum obtained for

the YGG tripeptide is analogous to the global minimum conformation obtained for the FGG and WGG tripeptides. Besides, few geometrical similarities might be found between the three most stable minima conformations of the YGG, FGG, and WGG tripeptides. First, the peptide backbone is folded. Second, *g+* rotameric states for the side-chain ($\chi_1 \cong 60°$) are preferred. Third, the carboxyl group is in a preferential *anti* configuration.

At this point, the examination of the relative energy differences obtained for the conformations of the YGG tripeptide allows a comparison between theoretical calculations reported here and previously reported experimental data obtained from X-ray studies [29]. Single-crystals of tyrosyl-glycyl-glycine monohydrate were obtained by slow evaporation from 50% acetic solution containing 10 mg mL⁻¹ of the tripeptide equilibrated *via* the vapour phase against absolute 2-propanol. High-resolution X-ray diffraction data was collected at 123 K (detailed information is available, if required, in [29]). Pleasingly our theoretical calculations are in good agreement with the experimental data. The backbone torsional angles from this crystal structure are: ψ_1 = 164.73, φ_2 = 80.50, ψ_2 = 11.08, φ_3 = -103.41 and ψ_3 = -152.32. In Fig. 3A the structure reported by Pichon-Pesme *et al*. [29] (in green) was overlapped with two conformations obtained for the YGG tripeptide at RHF/6-31G(d) level of theory in vacuo. In this figure it can be appreciated that the ygg*09* conformation (in light blue) shows a good overlapping correlation, not only with the two first torsional angles of the backbone but also with the side-

chain (Tyr residue). In the case of the other conformer (in yellow), a complete overlap was obtained, but this form displays an energy gap of 8.85 kcal mol^{-1} above the global minimum (see Supplemental Table 2, final geometry number 18).

3.2 Conformational study of YGG in solution

EDMC results are summarized in Table 3 and Fig. 4 and more details are given in Supplemental Table 3 in the Supplementary Data. Calculations yielded a large set of conformational families for the YGG tripeptide. The total number of conformations generated was 24,575, and the number of those accepted was 50,000. In the clustering procedure, a RMSD (root mean square deviation) of 0.75 Å and a cutoff of 50 kcal mol^{-1} were used. The number of families after clustering was 121. The total number of families accepted with a relative population higher than or equal to 1.0% was 21, whose populations sum up to *ca* 90% of the total population. It is important to highlight at this point that none of these 21 families has an energy gap higher than 2.0 kcal mol^{-1} with respect to the global minimum.

All low-energy conformers of the studied peptide
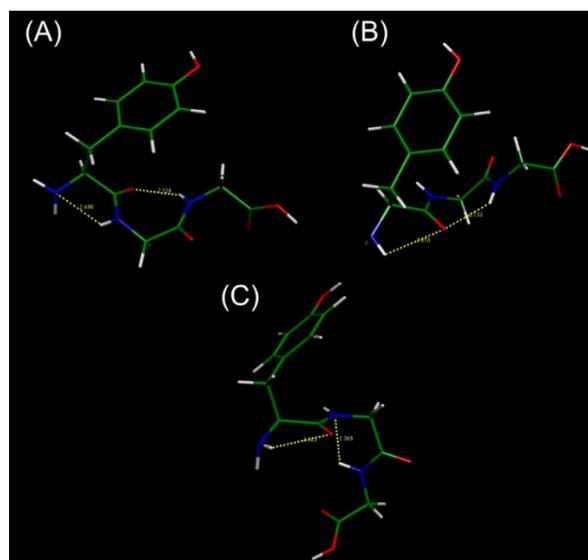
Figure 4. Stereoview of the three most populated families of YGG optimized at EDMC/SRFOPT/ECCEP/3 level of theory. A) Family 1 with a relative energy (DE) of 0.00 kcal mol^{-1}; B) Family 2 with a DE = 0.83 kcal mol^{-1} and C) Family 3 with a DE = 1.71 kcal mol^{-1}. The observed H-bonds are also shown.

were then compared to each other. The comparison involved the spatial arrangements, relative energy, and populations. It is interesting to note that none of

Table 3. Percent relative populations, relative energy values and torsional angles for the backbone conformers of YGG optimized at the EDMC/SRFOPT/ECEPP/3 level of theory. All conformational families shown here have relative population (%Pop) higher than 1.0% and a relative energy (DE) lower than 2.0 kcal mol^{-1}

Family	%Pop	ΔE	ψ_1	ϕ_2	ψ_2	ϕ_3	ψ_3
1	24.26	0.00	-40.65	80.74	-73.83	71.93	-145.86
2	19.20	0.83	122.06	-82.92	70.71	-72.25	146.31
3	5.68	1.71	154.74	75.83	29.63	161.62	-123.43
4	4.12	1.60	154.25	81.55	-75.28	70.88	-145.61
5	3.98	1.74	112.19	-79.39	-35.74	68.85	-144.06
6	3.04	1.78	-41.19	-74.55	-30.33	71.21	-145.45
7	2.74	1.88	-37.45	-72.91	-32.23	-171.29	107.40
8	2.70	1.99	-28.22	-74.53	-31.32	-170.43	104.97
9	2.68	1.60	154.20	81.56	-75.23	70.91	-145.62
10	2.58	1.88	-37.85	-72.94	-32.28	-171.29	107.33
11	2.50	1.63	154.76	75.83	29.66	161.67	-123.53
12	2.26	1.83	154.53	75.53	30.13	-71.37	145.38
13	2.10	1.76	-41.23	-74.45	-30.31	71.18	-145.34
14	1.88	1.82	112.24	-79.44	-35.83	68.92	-144.17
15	1.48	1.85	154.56	75.52	30.13	-71.35	145.35
16	1.46	1.73	-40.78	79.01	42.30	-69.05	143.89
17	1.16	1.97	-28.25	-74.59	-31.23	-170.48	105.00
18	1.16	1.98	-33.07	-81.67	74.62	-71.02	145.40
19	1.12	1.72	-42.71	-80.99	75.26	-70.95	145.40
20	1.08	1.75	-40.74	79.05	42.34	-69.14	143.88
21	1.04	1.72	-42.76	-80.92	75.19	-70.96	145.42

the energetically preferred families possesses a fully extended or a fully folded structure. Thus, the most populated family (24.26%, see Table 3), which is also the global minimum, adopts a semi-folded structure which is the most representative form of this molecule. This conformation is characterized by a stabilizing H-bond (r = 2.12 Å) between the carbonylic oxygen of residue *i* (Tyr) and the NH group of residue *i+1* (Gly) (Fig. 4A), thus it forms a $C_{eq}7$ interaction characterized by the values of ϕ and ψ (ϕ_2 = 80.74° and ψ_2 = -73.83°). It also shows a week C5 hydrogen bond type (r = 2.496 Å) encompassing the NH_2 group of residue *i* (as proton acceptor) and the NH group of residue *i+1* (as proton donor). The second most populated family (19.20%) corresponds to a structure also possessing a C7 interaction but in this case, it is a $C_{ax}7$ (r = 2.13 Å), which is characterized by the values of ϕ and ψ (ϕ_2 = -82.92° and ψ_2 = 70.71°) (Fig. 4B). Moreover, a C5 H-bond interaction (2.610 Å) between the NH_2 group of residue *i* and the carbonylic oxygen of the same residue (Tyr) is also observed. It is interesting to note that this family has an energy gap of 0.83 kcal mol⁻¹ with respect to the global minimum (see Table 3). On the other hand, the third most populated family (5.68%) has not shown a C7 interaction but a C5 one encompassing the NH group (as acceptor of the proton) of residue *i+1* and the NH group (as donor of the proton) of residue *i+2* (r = 2.37 Å). The NH group of residue *i+2* is also co-participating in a hydrogen bond with the carbonylic oxygen of residue *i+2* (r = 2.58 Å) (Fig. 4C). All families show a preference for a semi-extended side-chain conformation of residue Tyr ($\chi_1 \cong$ 180° and $\chi_2 \cong$ -60° or 60°).

According to the results of EDMC calculations summarized in Table 3 and Fig. 4, it is clear that the conformational preferences for the YGG tripeptide are different from those obtained in *vacuo*. It is noteworthy that both methods favoured C7 backbone conformations for the most stable conformers. Therefore, some interaction features were commonly observed by using both methodologies. Nevertheless, the comparison of the most populated conformations obtained in solution with the most stable conformers optimized at RHF/6-31G(d), showed some differences for the backbone torsional angles. The main discrepancy is obtained for φ_3 torsional angle, which determines the position of the second glycyl residue. This torsional angle adopts a g^- rotameric state ($\varphi_3 \cong$ -80°) in ygg*01* and ygg*11* conformations (see Table 1), while in family 1 it changes to a g^+ rotameric state ($\varphi_3 \cong$ 70°, see Table 3). A similar result was obtained comparing ygg*05* and ygg*10* conformers ($\varphi_3 \cong$ 70°) with family 2 ($\varphi_3 \cong$ -70°). Finally, χ_1 torsional angle determines the rotameric state of the side-chain of the Tyr residue. In case of the 13 most

stable conformers obtained from *ab initio* calculations, the g^+ rotameric state is systematically preferred, whereas EDMC calculations show a preference for a *trans* rotameric state ($\chi_1 \cong$ 180°). Regarding the terminal carboxyl group, it is worth mentioning that the stochastic conformational search in solution, which was carried out by using EDMC, showed a pronounced preference for *endo* (*syn*) conformers stabilized by an internal carboxyl H-bond (OHcarb…O=C9). The presence of the surrounding solvent might interfere in the formation of an H-bond between the carboxyl hydrogen and the C6=O, which have been observed in the gas phase for the *exo* (*anti*) conformers. On the other hand, the systematic conformational search in vacuo, which was carried out by using Gaussian, showed the presence of both conformers (*endo* and *exo*) but it showed a energetic preference for the *exo* (*anti*) conformers.

It is interesting to note that the EDMC calculations are also in good agreement with the experimental data, particularly comparing the family 12 (1.83 kcal mol⁻¹ less stable than the global minimum) and the structure reported by Pichon-Pesme *et al*. [29]. Fig. 3B shows the overlapping of the structure obtained from X-ray studies (in green) with two conformations obtained for the YGG tripeptide in solution (family 11 in light blue and family 12 in yellow). Thus, it can be appreciated as having good structural correlation, particularly with the backbone.

It is of great importance to mention that in general, empirical potentials provide inaccurate results for the study of isolated peptides containing aromatic side-chains. One of the most important reasons for this inaccuracy might be the fact that they work with effective average charges (over all existing structures), which may be far from the charges that properly describe each individual structure. It has recently been proven that the scan of the PES of FGG [17] and GFA [46] by using the AMBER empirical force field fails mainly due to a large variety of atomic charges for individual conformers. These calculations have demonstrated that introducing an average charge brings some uncertainty that results in incorrect structural predictions when an empirical potential is used. Despite this, the use of empirical potential is one of the most reliable techniques to study and predict the conformational behaviour of complex protein systems.

4. Conclusions

A previous work on the YGG tripeptide has shown that molecules containing aromatic rings are sensitive to intramolecular dispersion and basis set superposition error, the latter rendering MP2 calculations with small

to medium-sized basis set unsuitable for describing this type of molecules. B3LYP and small-basis MP2 may not be sufficiently accurate to unequivocally identify the most stable structures of aromatic peptides. The large number of possible conformers and the need to perform high-level quantum chemical methods accounting for dispersion make it nearly impossible to find all preferred conformers of aromatic peptides without experimental guidance. As experimental gas-phase data on the YGG tripeptide are not yet available, in the current study we reinvestigated the conformational preference of the YGG tripeptide using both a stochastic and a systematic conformational search (EDMC and *ab initio* quantum chemical calculations, respectively). Comparing the results obtained from both methodologies, a proper scan of the PES requires the use of a non-empirical method. EDMC calculations fail mainly to predict the position of the second glycyl residue (φ_3 torsional angle) and the spatial orientation of the terminal carboxyl group.

The most stable structures obtained from RHF/cc-pVTZ//RHF/6-31G(d) and RHF/aug-cc-pVTZ//RHF/6-31G(d) levels of theory are stabilized by C7 conformations for the central Gly residue and an intramolecular H-bond encompassing the *p*-OH of the aromatic side-chain (Tyr) and the carbonylic oxygen of residue *i+2*. Notwithstanding the three most energetically preferred structures lie within a small energy gap (0.35 kcal mol^{-1}), interestingly, they displayed different spatial orderings. It is worth mentioning that the methodology used in this study was able to predict additional stable conformers that have not been identified in a previous work. Comparing the structures and relative energies of the lowest energy conformers obtained for the YGG tripeptide with those of the FGG and WGG tripeptides, common geometrical features were observed for these sequences. The theoretical results reported here showed correspondence with those obtained by X-ray diffraction techniques.

An overall conclusion from the present work is that RHF/6-31G(d) calculations on the YGG tripeptide provide comparable results with those obtained on small related peptides by means of more sophisticated *ab initio* quantum chemical calculations (*i.e.*, RI-MP2/cc-pVXZ) at a considerably lower computational cost [18]. As it is well known, MP2 method allows the study of different systems at a high computer time but also with a high accuracy to the one obtained using the HF method. It is our experience that for these types of tripeptides the CPU time requirement increases, at least, in a threefold fashion when the level of theory is increased from HF to *post-ab inito* levels. Although it would be preferable to select the conformers using electronic structure methods that account for dispersion effects, our results indicate that this methodology can be used as a preliminary step to obtain conformations with reliable geometry and energy description, which can be used later as input for further calculations at higher levels of theory if needed.

Supplementary data

Supplementary data associated with this article can be found in the online version.

Acknowledgements

References

[1] S.S. Zimmerman, M.S. Pottle, G. Némethy, H.A. Scheraga, Macromolecules 10, 1 (1977)

[2] K.T. O'Neil, W.F. DeGrado, Science 250, 646 (1990)

[3] M.J. Rooman, J.P.A. Kocher, S.J. Wodak, Biochemistry 31, 10226 (1992)

[4] M. Blaber, X.J. Zhang, B.W. Matthews, Science 260, 1637 (1993)

[5] C. Brooks, D.A. Case, Chem. Rev. 93, 2487 (1993)

[6] A.G. Street, S.L. Mayo, Proc. Natl. Acad. Sci. U.S.A. 96, 9074 (1999)

[7] P. Koehl, V. Levitt, Proc. Natl. Acad. Sci. U.S.A. 96, 12524 (1999)

[8] G. Chasse et al., J. Mol. Struct. (THEOCHEM) 537, 319 (2001)

[9] H. Valdés, K. Pluháčková, M. Pitonák, J. Rezác, P. Hobza. Phys. Chem. Chem. Phys. 10, 2747, (2008)

[10] S. Anishetty, G. Pennathur, R. Anishetty. BMC Struct. Biol. 2, 9 (2002)

[11] G. Duan, V.H. Smith, Jr., D. Weaver, Chem. Phys. Lett. 310, 323 (1999)

[12] J.B.O. Mitchell, C.L. Nandi, I.K. McDonald, J.M. Thornton, J. Mol. Biol. 239, 315 (1994)

[13] M.M. Flocco, S.L. Mowbray, J. Mol. Biol. 235, 709 (1994)

[14] F. Nardi, G.A. Worth, R.C. Wade, Folding and

Design 2, S62 (1997)

[15] J.B.O. Mitchell et al., Nature 366, 413 (1993)

[16] G. Tóth, C.R. Watts, R.F. Murphy, S. Lovas, Proteins: Struct. Funct. Genet. 43, 373 (2001)

[17] D. Reha et al., Chem. Eur. J. 11, 6803 (2005)

[18] H. Valdés, D. Reha, P. Hobza, J. Phys. Chem. B 110, 6385 (2006)

[19] J. Cerný, P. Jurecka, P. Hobza, H. Valdés, J. Phys. Chem. A 111, 1146 (2007)

[20] D. Toroz, T. van Mourik, Mol. Phys. 105, 209 (2007)

[21] L.F. Holroyd, T. van Mourik, Chem. Phys. Lett. 442, 42 (2007)

[22] T. van Mourik, P.G. Karamertzanis, S.L. Price, J. Phys. Chem. A 110, 8 (2006)

[23] M.R. Peterson, I.G. Csizmadia, J. Am. Chem. Soc. 100, 6911 (1978)

[24] M.R. Peterson, I.G. Csizmadia, Prog. Theor. Org. Chem. 3, 190 (1982)

[25] M.F. Masman, S. Lovas, R.M. Murphy, R.D. Enriz, A.M. Rodríguez, J. Phys. Chem. A. 111, 10682 (2007)

[26] A.M. Rodríguez, J.C.P. Koo, D. Rojas, N. Peruchena, R.D. Enriz, Inter. J. Quant. Chem. 106, 1580 (2006)

[27] M.W. Klipfel et al., J. Phys. Chem. A. 107, 5079 (2003)

[28] M.F. Masman et al., Eur. Phys. J. D. 20, 531

[29] V. Pichon-Pesme, H. Lachekar, M. Souhassou, C. Lecomte, Acta Cryst. B56, 728 (2000)

[30] A. Liwo et al., Biopolymers 38, 157 (1996)

[31] D.R. Ripoll, H.A. Scheraga, Biopolymers 27, 1283 (1988)

[32] D.R. Ripoll, H.A. Scheraga, Biopolymers 30, 165 (1990)

[33] H.A. Scheraga, D.R. Ripoll, A. Liwo, C. Czaplewski, User Guide ECEPPAK and ANALYZE Programs.

[34] G. Némethy et al., J. Phys. Chem. 96, 6472 (1992)

[35] J. Vila, R.L. Williams, M. Vásquez, H.A. Scheraga, Proteins: Struct. Funct. Genet. 10, 199 (1991)

[36] R.L. Williams, J. Vila, G. Perrot, H.A. Scheraga, Proteins: Struct. Funct. Genet. 14, 110 (1992)

[37] M.F. Masman et al., Eur. J. Med. Chem. 44, 212 (2009)

[38] M.J. Frisch et al., Gaussian 03, Revision B.05, Gaussian Inc. (Pittsburgh PA, 2003)

[39] W.J. Hehre, L. Radom, P.V.R. Schleyer, J.A. Pople, Ab Initio Molecular Theory (John Wiley & Sons, New York, 1986)

[40] C.C.J. Roothaan, Rev. Mod. Phys. 23, 69 (1951)

[41] W.J. Hehre, R. Ditchfield, J.A. Pople, J. Chem. Phys. 56, 2257 (1972)

[42] P.C. Hariharan, J.A. Pople, Mol. Phys. 27, 209 (1974)

(2002)