C. R. Chimie 20 (2017) 500–507

Contents lists available at ScienceDirect

Comptes Rendus Chimie

www.sciencedirect.com

Full paper/Mémoire

# An exhaustive conformational analysis of *N*-formyl-L-tyrosinamide using a genetic algorithm for multimodal search

## *Une analyse conformationnelle exhaustive de laN-formyl-L-tyrosinamide par l'utilisation d'un algorithme génétique pour la recherche multi-modale*

Anouar El Guerdaoui [a, *], Brahim El Merbouh [a], Rachida Tijar [a], Malika Bourjila [a], Rachid Drissi El Bouzaidi [a, b], Abderrahman El Gridani [a], Mohamed El Mouhtadi [a]

[a] *Laboratoire de chimie physique, Faculté des Sciences, B.P. 8106, Université Ibn Zohr, 80000, Agadir, Morocco*
[b] *Centre Régional des Métiers de L'Education et de la Formation (CRMEF), Souss Massa Daraa, Inezgane, Morocco*

ARTICLE INFO

*Article history:*
Received 12 July 2016
Accepted 24 November 2016
Available online 31 December 2016

*Keywords:*
Genetic algorithm
Potential energy surface
Conformational analysis
L-Tyrosine
X-ray crystallography

ABSTRACT

This work reports a detailed conformational study on *N*-formyl-L-tyrosinamide diamide system through genetic algorithm based on multiniche crowding technique. We tried to evaluate the effect of the OH hydroxyl substitution of the benzene ring on the adopted folds by comparing them with those found previously for the *N*-formyl-L-phenylalaninamide. Among the 26 and 28 conformations detected for both systems *N*-formyl-L-tyrosinamide and *N*-formyl-L-phenylalaninamide successively, 19 conformations have similar geometrical structures (identical backbones and side-chain orientations). The comparison of 19 common conformation geometries of both systems revealed nearly perfect linear adjustments with $R^2$ values of 0.9997, 0.9996, 0.9998, and 0.9996 for the dihedral angles $\varphi$, $\psi$, $\chi_1$, and $\chi_2$ successively. This work also includes a comparison between theoretical calculations and experimental results of X-ray crystallography and nuclear magnetic resonance spectroscopy extracted from protein data bank.



*Motsclés:*
Algorithme génétique
Surface d'énergie potentielle
Analyse conformationnelle
L-Tyrosine
Cristallographie aux rayons X

RESUMÉ

Le présent travail rapporte une étude conformationnelle détaillée du système diamide *N*-formyl-L-tyrosinamide par le biais de l'algorithme génétique basé sur la technique Multi-Niche Crowding (MNC). Nous avons essayé, à travers cet article, d'évaluer l'effet de la substitution hydroxyle −OH du cycle benzénique de ce système sur les repliements adoptés, en les comparants avec ceux trouvées antérieurement pour le système *N*-formyl-L-phenylalaninamide. Parmi les 26 et 28 conformations détectées pour les deux systèmes *N*-formyl-L-tyrosinamide et *N*-formyl-L-phenylalaninamide successivement, 19 conformations ont des structures géométriques semblables (chaînes pricipales et orientation des

* Corresponding author.
*E-mail address:* anouarelguerdaoui@gmail.com (A. El Guerdaoui).

http://dx.doi.org/10.1016/j.crci.2016.11.008
1631-0748/

chaînes latérales identiques). La comparaison des géométries des 19 conformations communes des deux systèmes a révélé des ajustements linéaires presque parfaits avec des valeurs de $R^2$ de 0,9997, 0,9996, 0,9998 et 0,9996 pour les angles dièdres $\varphi$, $\psi$, $\chi_1$ et $\chi_2$, successivement. Ce travail inclut aussi une étude comparative entre les calculs théoriques et les résultats expérimentaux de la cristallographie aux rayons X et de la spectroscopie de résonance magnétique nucléaire (RMN) extraits de la *Protein Data Bank* (PDB).



## 1. Introduction

During the past three decades, many studies have been conducted on amino acids and peptides including those consisting of amino acids that have aromatic side chains. They have received a great experimental and theoretical attention because of their particular importance as chromophores to study the interactions within proteins. Their large faces $\pi$ can be bound to metal ions such as $Na^+$, $Mg^+$, and $Al^+$ [1,2], and they often take action to cause repulsive and attractive electrostatic forces, which lead to structural changes in their environments [3]. The presence of a UV chromophore in these biomolecules such as phenylalanine, tyrosine, tryptophan, and their peptides, respectively, has allowed the use of the different methods of spectroscopic analysis to determine their most stable structures in the gas phase. The identification of these structures experimentally detected requires the comparison with computed data on the energetics and vibrational frequencies to convert the observed spectra into structural assignments. This symbiotic relationship between experiment and theory has been used to study a wide range of biological molecules in the gas phase, including tryptamine neurotransmitters and serotonin [4–6], adrenergic neurotransmitters and their analogues [7–10], and more recently, small peptide chains [11–16]. Moreover, the experimental procedures performed on these molecules revealed a decay in the amount of conformations detected as the size of these systems increases [11,15,17]. In stark contrast with this small number of peptide conformations experimentally detected, there are a large number of conformations with low energies obtained through theoretical calculations. This poses extra requirements on computational techniques, in which the high-level quantum chemistry methods (e.g., which take into account a part of the electron correlation) will be necessary to accurately describe the intramolecular interactions within peptide, and in particular, the dispersion interactions involving the aromatic residues. Duan et al. [18] showed that the interaction between the aromatic cycle and the peptide backbone is surprisingly large (11 kJ $mol^{-1}$). This indicates the methods that describe the adequate dispersion interactions will be necessary to obtain a correct energy order of conformers. However, the identification of the global minimum by investigating all possible conformers to a high level of theory (theory of perturbations Møller-Plesset, MP2, or beyond) quickly becomes an impossible task for peptides that consist of more than two or three amino acid residues.

A solution to the conformational problem of larger-size molecules (long peptide chains) consists in using a less rigorous method of calculation to generate the most stable structures to be used than as starting structures for an ab initio calculation. In this sense, a new mathematical optimization road knows a growing development. The stochastic methods gradually take precedence over the conventional deterministic technical: on the one hand, they help in locating the optimum of a function in the parameters' space without using the derivatives of the function with respect to these parameters, and on the other hand, they do not get trapped by a local optimum and usually manage to determine the global optimum of the function in question.

This work is part of the development framework of a class of evolutionary methods called “multiniche crowding” (MNC) to locate the different typical secondary folds of the protected amino acid *N*-formyl-L-tyrosinamide (HCO-L-Tyr-$NH_2$). The combination of this technique, the basic operation of which is relevant to our desired objective to seek potential minima (global and local minimum) of molecular structures with the tools of semiempirical quantum calculation, should help afterward, on the one hand, to understand the conformational preference of this residue in the long polypeptide chains, and on the other hand, to provide vital data on the stacks problems of aromatic rings.

## 2. Calculation details

### 2.1. Numbering and abbreviations

Tyrosine can be derived from alanine amino acid by replacing one of its three hydrogen atoms of the methyl group by a phenol group constituting the side chain whose orientation is regulated by the three torsion angles: $\chi_1$ around $-C_4-C_{12}-$, $\chi_2$ around $-C_{12}-C_{14}-$, and $\chi_3$ around $-C_{17}-O_{24}-$. The backbone of HCO-L-Tyr-$NH_2$ has four torsion angles $\omega_0$, $\Phi$, $\psi$, and $\omega_1$ and represent the rotations around the bonds $C_2-N_3$, $N_3-C_4$, $C_4-C_5$, and $C_5-N_6$, respectively, as shown in Fig. 1a, where $\omega_i$ angles take 180° and correspond to *trans*-peptide bonds. The numbering of the different atoms of the compound HCO-L-Tyr-$NH_2$ is given in Fig. 1b.

The multidimensional conformational analysis (MDCA) [19] predicted the existence of nine possible conformations for the backbone of any amino acid defined by two torsion angles $\Phi$ and $\psi$ associated with the peptide bond. These nine conformations marked in Greek letters attached to both letters L or D ($\alpha_D$, $\alpha_L$, $\gamma_D$, $\gamma_L$, $\beta_L$, $\delta_D$, $\delta_L$, $\varepsilon_D$, and $\varepsilon_L$) are often represented on a map called Ramachandran map $E = f(\Phi, \psi)$ (Fig. S1a). The side chain of tyrosine is defined by three angles of $\chi_1$, $\chi_2$, and $\chi_3$, whose two angles $\chi_1$ and $\chi_2$

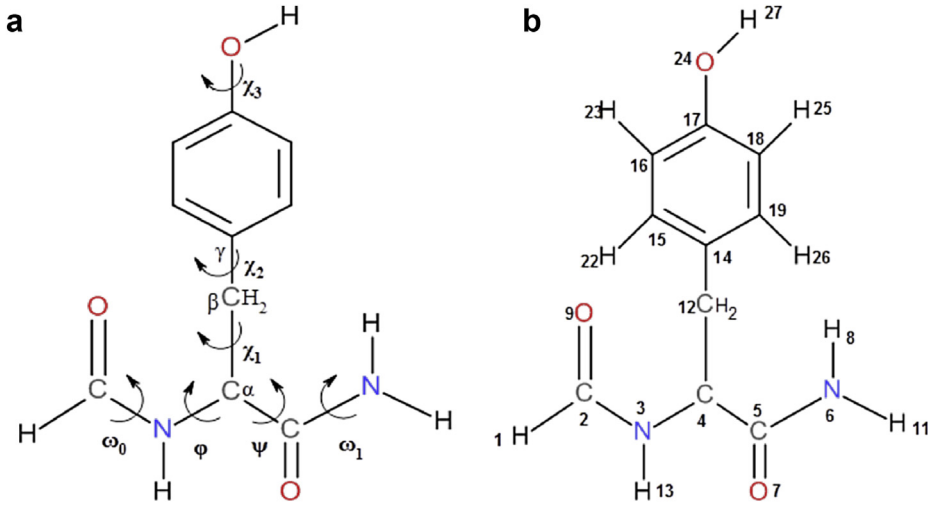


**Fig. 1.** (a) Numbering of atoms and (b) defining dihedral angles for HCO-L-Tyr-$NH_2$.

can take three possible orientations, *gauche*$^+$ ($g^+ = +60$), *anti* ($a = 180$), and *gauche*$^-$ ($g^- = -60$) leading in total to $3 \times 3 = 9$ possible orientations for this chain (Fig. S1b). Torsional angle $\chi_3$ may generally take both *cys* (s) and *trans* (a) positions corresponding to the values 0° and 180°, respectively.

In accordance with the IUPAC–IUB recommendation [20], dihedral or torsional angles were specified [21] within −180° and 180° for both backbone ($\Phi$, $\psi$) and side chain ($\chi_1$, $\chi_2$) conformations (Fig. S1).

### 2.2. Molecular computations

The adopted strategy for the scanning of the potential energy surfaces (PESs) of the peptide systems is to vary systematically and simultaneously both torsional angles $\Phi_i$ and $\psi_i$ of the *i*th polypeptide residue while maintaining other residues unchangeable attributing them to previously well-defined folds [22,23], a strategy that may show a great weakness in the exploration of the research space because it may neglect some conformations with low energies.

Our research team [24] published an approach that allows to explore the PES of an isolated molecular system and which may be of high molecular weight. The stochastic approach "genetic algorithm" based on the mechanisms of natural selection and genetic recombination, is used to create population generations from population point candidates called individuals or chromosomes that randomly product to optimize an objective function or fitness. Each chromosome is composed of a set of elements or characteristics called genes. In the study of the problem of conformational space of a given molecule, individuals correspond to conformations, the genes to dihedral angles, and the fitness function to the total energy of the system. This approach, which is coded into computer language and driven by scripts, automates the scan of all minima of molecular PES.

MNC method, proposed by Cedeno et al. [25] based on the concept of crowding, is used not only during the insertion of offspring in the population but also during the selection of the individuals that are going to mate. In MNC method, selection by fitness proportionate reproduction used in simple genetic algorithm (SGA) is replaced by what the author calls "crowding selection". For each individual $I_i$ of the population, its mate $I_j$ is selected from a group of individuals of crowding selection size (Cs), picked at random from the population. The mate $I_j$, thus chosen, must be the most similar one to $I_i$. During the replacement step, the algorithm uses a replacement policy called worst among most similar, as illustrated in Fig. 2. It consists first of all in choosing randomly Cf groups of *s* individuals per group from the population. These groups are called "crowding factor groups". Second, one individual from each group who is the most similar to the offspring is identified. Finally, among these Cf individuals who are candidates for replacement, the less fitted one is replaced by the offspring.

The result is an algorithm that (1) maintains stable subpopulations within different niches, (2) maintains diversity throughout the search, and (3) converges to different local minima. Another major advantage of this method is that it requires no prior knowledge of the research space. Therefore, it is adopted in this study to locate the global minimum of a given molecular system on the one hand, and to find all local minima on its PES on the other hand.

The genetic algorithm based on the MNC technique is implemented in a package of program interfaced with MOPAC [26] (version 6.0) to evaluate the quality of the individual to insert into the population in each iteration. The evaluation criterion is the energy of the molecule (the heat of formation in our case). The semiempirical method AM1 is used separately to accomplish this task. The data file has been designed so that a constraint is imposed to the values of the dihedral angles defining the freedom degrees of the conformation randomly generated beforehand. This

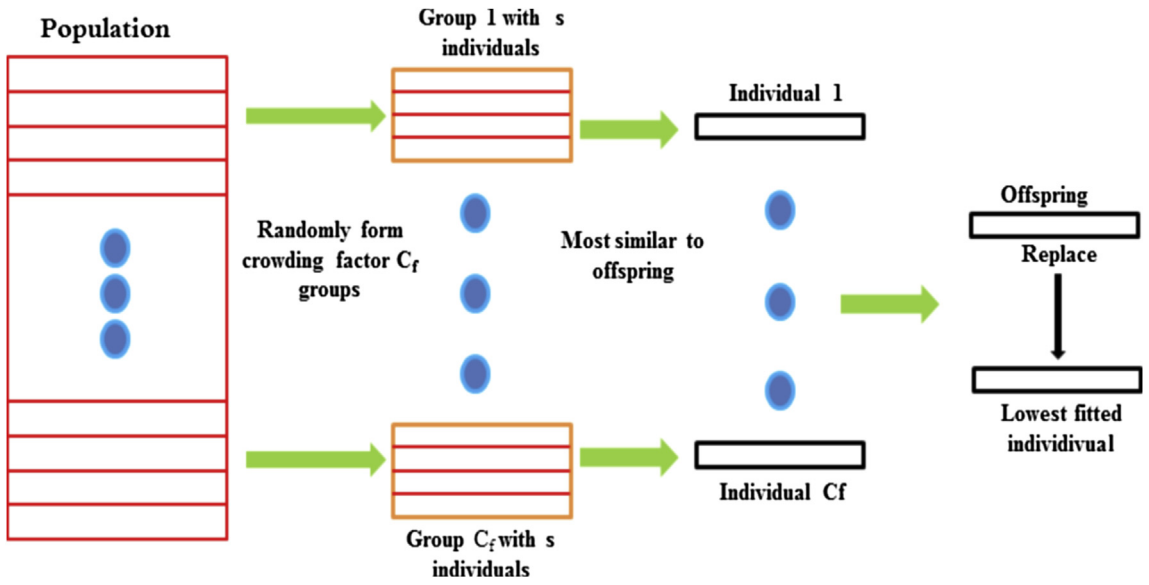


**Fig. 2.** Schematic illustration of worst among most similar technique.

constraint permits indeed to calculate exactly the energy of the conformation generated during the construction of the initial population and after the application of the crossover and mutation operators eventually. Once the algorithm converges after the fixed maximum number of generation, an optimization without constraint is performed to release the structure so that the individuals of the same niche converge to the corresponding minimum. It has been discussed elsewhere [27] that real encoding is ideally suited to handle problems in a continuous research space. The encoding scheme is so that the genome of each individual (conformation) is composed of $n$ dihedral angles ($\varphi_1$, $\varphi_2$,…, $\varphi_n$), in the case of a molecule of $n$ degrees of freedom, whose values represent a location on the PES. The crossover operator used is called interval crossover [28]. In interval crossover, only one offspring is generated. For each pair of parent genes $\varphi_1$ and $\varphi_2$, the offspring's gene is selected at random from the interval $[\varphi_1 - \varepsilon/2, \varphi_2 + \varepsilon/2]$, assuming without loss of generality that $\varphi_1 < \varphi_2$, if we use a real encoding as in our case. The parameter $\varepsilon$ will be called in the following the parameter of interval crossover. The crossover operation is thus performed so that it generates an offspring close to the parents. The mutation is applied to the offspring, generated by the crossover operation, with a $P_m$ probability (see Table 1), whereas permuting a couple of genes of the offspring selected at random.

**Table 1**
Control parameters of the MNC genetic algorithm.

| Parameters | Values |
|---|---|
| Population size | 500 |
| Crowding selection size (Cs) | 25 |
| Crowding factor size (Cf) | 3 |
| Crowding size ($s$) | 25 |
| Interval crossover parameter ($\varepsilon$) | 10 |
| Crossover probability ($P_c$) | 1.0 |
| Mutation probability ($P_m$) | 0.06 |

With this approach, a large number of conformations are treated. The exploration of the PES system is conducted intelligently and according to its promising and productive areas. The conformations of the first population are generated randomly. The values of the torsion angles, freedom degrees of the PES that define a molecular conformation are produced by mutation. The data files used by MOPAC are built automatically and one file in dynamic access gathers conformations forming the population of the current generation.

## 3. Results and discussion

### 3.1. Conformational analysis

Gordon et al. [29] have conducted the first systematic mapping of the PES of the formyl-L-alaninamide (HCO-L-Ala-$NH_2$) at Restricted Hartree-Fock (RHF)/3-21G and RHF/6-31+G* level of theory with a step size of 30° for $\varphi$ and $\psi$ angles successively, they found only seven minimum among the nine predicted by the MDCA. Therefore, the conformational building units of the right-handed helical structure $\alpha_L$ and that of polyproline II ($\varepsilon_L$) did not turn out be stationary points on the PES of HCO-L-Ala-$NH_2$.

Subsequently, the PES of the other amino acids, including glycine [30], phenylalanine [31], valine [32], and glutamine [33], have been explored and showed similar topologies to that of alanine. On the basis of these results, Császár and Perczel [34] suggested that the structural polyproline II ($\varepsilon_L$) element cannot be obtained for any of single amino acids using the diamide approximation. However, subsequent studies have reported the existence of $\varepsilon_L$ for Trp [35], Ile [36], and Cys [37].

The calculations carried out using the MNC genetic algorithm coupled with the AM1 method on the diamide system HCO-L-Tyr-$NH_2$ have revealed the existence of 26

different conformations (Table S1) spread over seven zones ($\alpha_L$, $\alpha_D$, $\beta_L$, $\gamma_L$, $\gamma_D$, $\delta_D$, and $\varepsilon_D$) among the nine legitimate predefined areas on the Ramachandran map, where the fold $\alpha_L$ appears this time as a potential minimum, a fold which usually presents as a minimum on the PESs of diamide models whose side chains comprise the hydroxyl grouping –OH [38,39], whereas no structure corresponding to the polyproline II $\varepsilon_L$ ($\varphi = -75 \pm 30°$, $\psi = 160 \pm 30°$) was detected. The number of different side-chain orientations ($g^+$ $g^+$, … …,$g^-$ $g^-$) varies according to each of the seven types of backbone folds found ($\alpha_D$, $\alpha_L$, $\gamma_D$, $\gamma_L$, $\beta_L$, $\varepsilon_D$, and $\delta_D$), as shown in Table 2.

The graphical representation of relative heat of formation $\Delta(\Delta H_f)$ for the 26 located conformations of HCO-L-Tyr-$NH_2$ diamide system (Fig. S2) shows that conformers with D subscript ($\alpha_D$, $\gamma_D$, $\varepsilon_D$, and $\delta_D$) have energy values (formation heat) relatively higher, which makes them structurally less stable. This result is consistent with results obtained for other N- and C-protected L-amino acids containing *trans*-peptide bonds ($\omega_i = 180°$) treated at an ab initio level as glycine [30], phenylalanine [31], serine [39], proline [40], and aspartic acid [41].

Note that the relative heat of formation $\Delta(\Delta H_f)_X$ of each conformation *X*, is calculated with regard to the global minimum $\gamma_L(g^- g^+)$ according to Eq. 1:

$$\Delta(\Delta H_f)_x = (\Delta H_f(X) - \Delta H_f(\gamma_L(g^- g^+))) \ (\text{kcal mol}^{-1}) \qquad (1)$$

where $\Delta H_f(X)$ and $\Delta H_f(\gamma_L(g^- g^+))$ are the heats of formation of conformers *X* and $\gamma_L(g^- g^+)$, respectively.

We can easily notice that the first four most stable conformers adopt the inverse gamma turn folding $\gamma_L$ with a low energy gap ($\approx$0.16 kcal), this shows that the orientation of the side chain $-CH_2$-($C_6H_5$) phenol has only a very modest effect on the backbone, which is quickly found by comparing the values of the angles ($\varphi$, $\psi$) that appear similar [$\varphi = -83.6 \pm 1°$, $\psi = 60.1 \pm 5.2°$]. This rigidity of the backbone is because of the strong hydrogen bond $CO_9 \cdots HN_6$ forming a ring with seven atoms (Table S1).

### 3.2. Molecular interactions

The studies conducted on N- and C-protected L-amino acids (HCO-X-$NH_2$ and $CH_3CO$-X-$NHCH_3$), which carry polar or nonpolar side chains, showed that the latter causes attractive electrostatic forces that result from hydrogen bondings, van der walls interactions, and other repulsive caused by the steric hindrance leading to structural changes in their environment, namely at the level of adjacent main chains, thus the length and nature of these side chains in addition to the interactions within the backbone itself play a key role in the adopted folds. The different types of hydrogen bonds, which may occur in the various conformers of HCO-L-Tyr-$NH_2$, are shown in Fig. 3.

The corresponding distances to these potential hydrogen bonds within each conformation of HCO-L-Tyr-$NH_2$ detected by the MNC/AM1 approach are reported in Table 3.

**Table 2**
Number of detected conformations of HCO-L-Tyr-$NH_2$ according to the different areas of the Ramachandran map.

| Configuration of the backbone conformers | Number of orientations of the side chain |
|---|---|
| $\alpha_D$ | 1 |
| $\beta_L$ | 4 |
| $\gamma_L$ | 6 |
| $\gamma_D$ | 5 |
| $\delta_D$ | 5 |
| $\varepsilon_D$ | 3 |
| $\alpha_L$ | 2 |

### 3.3. Comparison with HCO-L-phenylalanine-$NH_2$

#### 3.3.1. Structural analysis

The amino acid tyrosine differs from that of phenylalanine per hydroxyl substituent on its benzene ring. A recent study carried out on *N*-formyl-L-phenylalaninamide (HCO-L-Phe-$NH_2$) using genetic algorithm based on the MNC technique and coupled with the semiempirical method AM1 [42] allowed us to locate 28 different conformations for this diamide system (Table S2). The results obtained from this study will serve to make a comparison on the structural changes between these diamide systems. Then, if the –OH hydroxyl substitution in the benzene ring of the diamide system HCO-L-Tyr-$NH_2$ has no effect on the foldings adopted, then all stereochemical and energetic characteristics observed from tyrosyl residue must be identical to those of the phenylalanyl residue. However, comparing the two diamide systems revealed remarkable differences:

- The number of folds ($\alpha_D$, $\gamma_D$, $\gamma_L$, …) detected for the two systems is different. In the case of the protected tyrosine, two conformations whose dihedral angles are: ($\varphi_1 = -102.5°$; $\psi_1 = -0.1°$) and ($\varphi_2 = -103.7°$; $\psi_2 = -1.2°$) corresponding to the right-handed helical structure $\alpha_L$ are detected. However, no conformation that corresponds to this fold was located by either the stochastic approach MNC [42] or by ab initio calculations for the protected phenylalanine [31,43].
- Among the 26 and 28 conformations detected for both systems HCO-L-Tyr-$NH_2$ and HCO-L-Phe-$NH_2$, respectively, only 19 have similar geometric structures (identical backbones and side-chain orientations). The differences between the dihedral angles of 19 common structures of the two diamide systems HCO-L-Tyr-$NH_2$ and HCO-L-Phe-$NH_2$ (Table S3) are calculated by using Eq. 2:

$$\Delta(\text{Dihedral angle}) = \text{dihedral angle(Tyrosine)} - \text{dihedral angle(Phenylalanine)} \qquad (2)$$

These differences observed between the dihedral angles can be explained by the influence of the tyrosyl residue ($C_6H_4OH$) on some intramolecular interactions, which are responsible for the structures' rigidity in question.

The dihedral angles $\varphi$, $\psi$, $\chi_1$, and $\chi_2$ of HCO-L-Tyr-$NH_2$ can be represented in function of those of HCO-L-Phe-$NH_2$ to assess the concordance quality between the common conformers of the two systems (Figs. S3–S6).

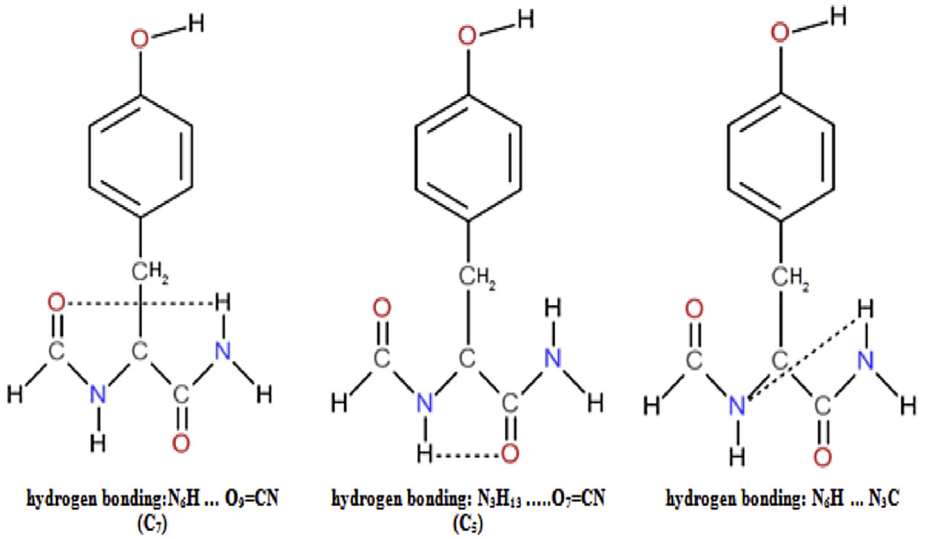


**Fig. 3.** Schematic representation of different types of hydrogen bonds for HCO-L-Tyr-$NH_2$.

**Table 3**
Relative distances of potential hydrogen bonds and electrostatic interactions within the 26 conformations of the HCO-L-Tyr-$NH_2$ obtained by MNC algorithm.

| Conformers | $N_6H\cdots N_3C$ | $N_6H\cdots O_9{=}CN$ ($C_7$) | $N_3H_{13}\cdots O_7{=}CN$ ($C_5$) |
|---|---|---|---|
| $\alpha_D(g^+\,g^-)$ | 2.64 | – | – |
| $\beta_L(g^+\,g^+)$ | – | – | 2.84 |
| $\beta_L(a\,g^+)$ | – | – | 2.45 |
| $\beta_L(a\,g^-)$ | – | – | 2.46 |
| $\beta_L(g^-\,g^-)$ | – | – | 2.77 |
| $\gamma_L(g^+\,g^+)$ | – | 1.81 | – |
| $\gamma_L(g^+\,g^-)$ | – | 1.87 | – |
| $\gamma_L(a\,g^+)$ | – | 1.74 | – |
| $\gamma_L(a\,g^-)$ | – | 1.80 | – |
| $\gamma_L(g^-\,g^+)$ | – | 1.81 | – |
| $\gamma_L(g^-\,g^-)$ | – | 1.85 | – |
| $\gamma_D(a\,g^+)$ | – | 1.54 | – |
| $\gamma_D(a\,a)$ | – | 1.64 | – |
| $\gamma_D(a\,g^-)$ | – | 1.65 | – |
| $\gamma_D(g^-\,g^+)$ | – | 1.55 | – |
| $\gamma_D(g^-\,g^-)$ | – | 1.54 | – |
| $\delta_D(g^+\,g^+)$ | – | – | – |
| $\delta_D(g^+\,g^-)$ | – | – | – |
| $\delta_D(a\,g^+)$ | – | – | – |
| $\delta_D(a\,g^-)$ | – | – | – |
| $\delta_D(g^-\,g^-)$ | – | – | – |
| $\varepsilon_D(a\,g^+)$ | – | – | – |
| $\varepsilon_D(a\,a)$ | – | – | – |
| $\varepsilon_D(g^-\,g^+)$ | – | – | – |
| $\alpha_L(g^-\,g^+)$ | 2.20 | – | – |
| $\alpha_L(g^-\,g^-)$ | 2.20 | – | – |

The dihedral angles $\varphi$ and $\chi_1$ for the conformations of the two diamide systems show a very good correlation ($R^2 = 0.9997$ and $R^2 = 0.9998$, respectively), which may show that the intervention of the grouping –OH is limited in a manner that did not disrupt certain intramolecular bonds responsible for the rigidity of these structures (Figs. S3 and S5). However, this correlation decreases slightly for both dihedral angles $\psi$ and $\chi_2$ ($R^2 = 0.9996$ and $R^2 = 0.9996$, respectively) (Figs. S4 and S6).

We can observe (from the spread of points in the figures) that the dihedral angles $\psi$ and $\chi_2$ are less ideal in terms of the population of their respective conformational space. These can be contrasted to the dihedral angles $\varphi$ and $\chi_1$, which have most of the stable conformers populating ideal states predicted by MDCA. The points representing the values of dihedral angles $\varphi$ and $\chi_1$ are intensively grouped in the probable areas predicted by MDCA ($g^+$, $a$, $g^-$) (Figs. S3 and S5). However, the points representing both dihedral angles $\psi$ and $\chi_2$ are distributed in a random manner, which indicates the flexible nature of each of those angles (Figs. S4 and S6).

### 3.3.2. Energetic analysis

The calculations carried out on the HCO-L-Tyr-$NH_2$ system (Table S1) predicted the folded conformation $\gamma_L(g^-\,g^+)$ as being the global minimum (see Fig. 4). The relative energy between the latter and the relative next higher minimum in energy $\gamma_L(g^-\,g^-)$ amounts to 0.01 kcal $mol^{-1}$, whereas the third and the fourth minimums are the folds $\gamma_L(g^+\,g^-)$ and $\gamma_L(g^+\,g^+)$ with relative energies of 0.15 and 0.16 kcal $mol^{-1}$, respectively, above the global minimum $\gamma_L(g^-\,g^+)$.

It is worth noting that these four most stable conformations of the HCO-L-Tyr-$NH_2$ system are also the preferred ones for HCO-L-Phe-$NH_2$ and in the same stability order but with a different energy gap (Fig. S7(a–e)).

Depicted 4D Ramachandran map based on four dihedral angles $\varphi$, $\psi$, and $\chi_1$, $\chi_2$ for located conformations of both HCO-L-Phe-$NH_2$ and HCO-L-Tyr-$NH_2$ diamide systems through the MNC/AM1 approach (Fig. S8) allows us a panoramic and simultaneously vision of adopted folds, relative orientations of the side chains, and relative heats of formation.

We easily notice that the conformations $\gamma_D(g^+\,g^+)$, $\beta_L(a\,g^+)$, $\alpha_D(g^+\,g^+)$, $\alpha_D(a\,g^+)$, and $\alpha_D(g^-\,g^-)$ are presented as potential minima in the PES of HCO-L-Phe-$NH_2$ when they are absent from that of HCO-L-Tyr-$NH_2$. A probable reason for this loss can be attributed to their relatively high energy

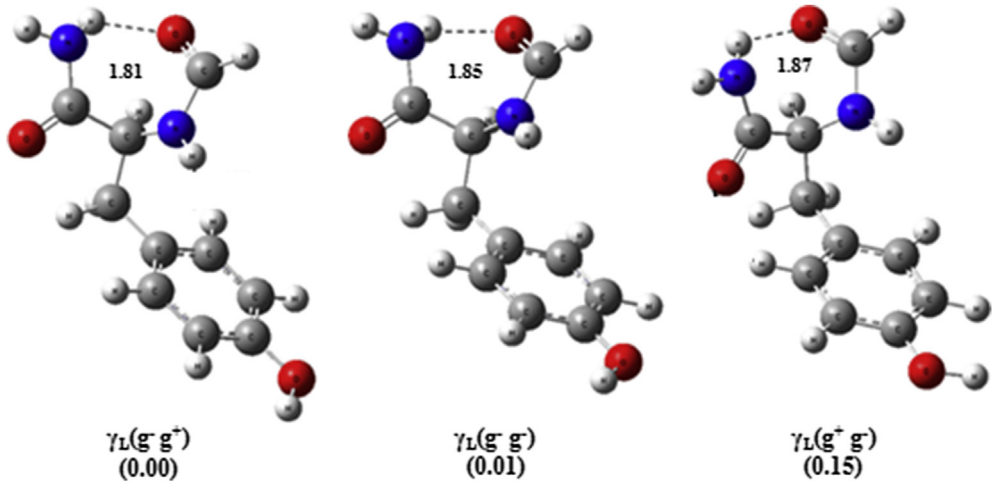


**Fig. 4.** Geometric representation of the three most stable conformers of the HCO-L-Tyr-$NH_2$ obtained by MNC/AM1 (bond lengths are in Å).

values (heats of formation) (Table S1), which makes them structurally less stable. Consequently, the intervention of the hydroxyl –OH grouping caused some structural changes (dihedral angles, hydrogen bonds, and so forth) that have led to the migration of these conformations to other more stable.

### 3.4. Correlation theory experience

The validity of the calculations presented in this work can be assessed by comparing, at the structural level, the conformations of the tyrosine predicted by the MNC approach with those derived experimentally either by X-ray crystallography, which will provide the technology needed to solve the three-dimensional structure of the proteins, or by nuclear magnetic resonance (NMR) spectroscopy.

The backbones of amino acids and consequently of proteins can be described by the two torsion angles $\varphi$ and $\Psi$ defining any secondary refolding. In this sense, a map showing the distribution of pairs ($\varphi$, $\Psi$) of 410 tyrosine residues, collected from 131 proteins was generated (Fig. 5a) by exploiting recent experimental protein data (May 2016) from protein data bank [44]. To conduct a comparison between the secondary folds observed experimentally and those predicted by the MNC approach, an additional map was drawn on the basis of the obtained results showing the pairs ($\varphi$, $\Psi$) corresponding to 500 conformations predicted by our calculations (Fig. 5b).

The comparison of these data shows a promising similarity. The map representing the experimental data (X-ray and NMR) has five zones, two of which are highly populated. The first zone (I) corresponds to the $\beta_L$ (extended β-strand) and $\varepsilon_L$ (polyproline II) areas; the second zone (II)

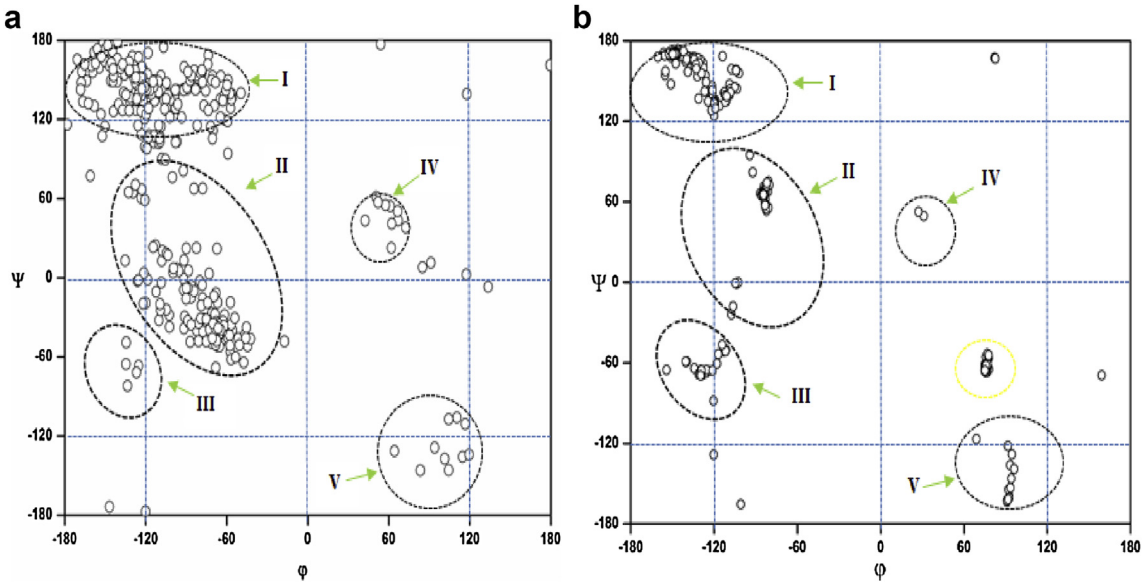


**Fig. 5.** (a) Couple distribution ($\varphi$, $\Psi$) of 410 tyrosine residues, collected from 131 proteins. (b) Couple distribution ($\varphi$, $\Psi$) corresponding to 500 HCO-L-Tyr-$NH_2$ conformations predicted by the MNC algorithm.

represents the $\gamma_L$ (inverse gamma turn), $\delta_L$, and $\alpha_L$ (right-hand α-helix) areas; the third zone (III) corresponds to the $\delta_D$; the fourth zone (IV) is the $\alpha_D$, which corresponds to the left-hand α-helix; and the fifth zone (V) corresponds to $\varepsilon_D$ (inverse polyproline II). The five zones are shown by dotted lines in Fig. 5a.

However, the map representing the folds theoretically detected through the MNC algorithm (Fig. 5b) also showed five zones but with the following remarkable differences:

- The absence of the conformation $\varepsilon_L$ (polyproline II). The same result was already mentioned by Chass et al. [45] by using the protected diamide system *N*-acetyl-tyrosyl-*N*-methylamide ($CH_3CO$-L-Tyrosine-$NHCH_3$) at B3LYP/6-31G(d) level of theory.
- The area designated by a dotted yellow line Fig. 5b corresponding to the folding $\gamma_D$ (right gamma turn) does not appear in the experimental map.

The zones III, IV, and V are less populated and experimentally correspond to the $\delta_D$, $\alpha_D$, and $\varepsilon_D$ folds, respectively. These folds are less favorites energetically by calculations carried out using the MNC genetic algorithm—as detailed in Section 3.1—and therefore far from being the preferred conformations of the compound HCO-L-Tyr-$NH_2$.

Such a correlation permits us to assume that if the diamide model is relevant to the description of main-chain folding of proteins, then the most stable conformers should have the lowest energy.

## 4. Conclusions

The exploration of the PES of HCO-L-Tyr-$NH_2$ through the genetic algorithm based on the MNC technique revealed the existence of 26 different conformations. The lowest energy structure corresponds to a $\gamma_L$ backbone conformation with $(g^-, g^+)$ orientations in the side chain. A comparative study made between HCO-L-Tyr-$NH_2$ and HCO-L-Phe-$NH_2$ systems suggests that the −OH substitution effect is indeed present. From the 19 common structures, dihedral angles $\varphi$ and $\chi_1$ showed perfect linear fits with $R^2$ values of 0.9997 and 0.9998, respectively. This correlation decreases slightly for dihedral angles $\psi$ and $\chi_2$ with $R^2$ values of 0.9996 for both angles.

The comparative study among theoretical calculations and experimental (NMR and X-ray) results revealed a promising similarity, which reflects the ability of the MNC genetic algorithm used to locate the different minima on the PES of HCO-L-Tyr-$NH_2$.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.crci.2016.11.008.

## References


[1] D.A. Dougherty, Science 271 (1996) 163.
[2] R.C. Dunbar, J. Phys. Chem. A 102 (1998) 8946.
[3] L.C. Snoek, E.G. Robertson, R.T. Kroemer, J.P. Simons, Chem. Phys. Lett. 321 (2000) 49.
[4] J.R. Carney, T.S. Zwier, J. Phys. Chem. A 104 (2000) 8677.
[5] J.R. Carney, T.S. Zwier, Chem. Phys. Lett. 341 (2001) 77.
[6] T.V. Mourik, L.E.V. Emson, Phys. Chem. Chem. Phys. 4 (2002) 5863.
[7] L.C. Snoek, T.V. Mourik, J.P. Simons, Mol. Phys. 101 (2003) 1239.
[8] L.C. Snoek, T. Van-Mourik, P. Çarçabal, J.P. Simons, Phys. Chem. Chem. Phys. 5 (2003) 4519.
[9] P. Çarçabal, L.C. Snoek, T. Van-Mourik, Mol. Phys. 103 (2005) 1633.
[10] R.G. Graham, R.T. Kroemer, M. Mons, E.G. Robertson, L.C. Snoek, J.P. Simons, J. Phys. Chem. A 103 (1999) 9706.
[11] I. Hunig, K. Kleinermanns, Phys. Chem. Chem. Phys. 6 (2004) 2650.
[12] W. Chin, M. Mons, J.P. Dognon, F. Piuzzi, B. Tardivel, I. Dimicoli, Phys. Chem. Chem. Phys. 6 (2004) 2700.
[13] W. Chin, J.P. Dognon, F. Piuzzi, B. Tardivel, I. Dimicoli, M. Mons, J. Am. Chem. Soc. 127 (2005) 707.
[14] W. Chin, J.P. Dognon, C. Canuel, F. Piuzzi, I. Dimicoli, M. Mons, I. Compagnon, G. Meijer, J. Chem. Phys. 122 (2005) 054317.
[15] D. Reha, H. Valdés, J. Vondrásk, P. Hobza, A. Abu-Riziq, B. Crews, M.S. Vries, Chem. Eur. J. 11 (2005) 6803.
[16] A. Abo-Riziq, L. Grace, B. Crews, M.P. Callahan, T. Van-Mourik, M.S. Vries, J. Phys. Chem. A 115 (2011) 6077.
[17] J.M. Bakker, C. Plützer, I. Hünig, T. Häber, I. Compagnon, G. Von-Helden, G. Meijer, K. Kleinermanns, ChemPhysChem 6 (2005) 120.
[18] G. Duan, V.H. Smith Jr., D.F. Weaver, Int. J. Quantum Chem. 90 (2002) 669.
[19] I.G. Csizmadia, in: J. Bertran (Ed.), Multidimensional Stereochemistry and Conformational Potential Energy Surface Topology, Reidel, Dordrecht, The Netherlands, 1989, pp. I–31.
[20] IUPAC-IUB, Biochemistry 9 (1970) 3471.
[21] Ö. Farkas, A. Perczel, J.F. Marcoccia, M. Hollósi, I.G. Csizmadia, J. Mol. Struct. 331 (1995) 27.
[22] J.C.C. Liao, J.C. Chua, G.A. Chass, A. Perczel, A. Varro, J.G. Papp, J. Mol. Struct. 621 (2003) 163.
[23] A. Mehdizadeh, G.A. Chass, Ö. Farkas, A. Perczel, L.L. Torday, A. Varro, J.G. Papp, J. Mol. Struct. 588 (2002) 187.
[24] B. El Merbouh, M. Bourjila, R. Tijar, R.D. El Bouzaidi, A. El Gridani, M. El Mouhtadi, J. Theor Comput. Chem. 13 (2014) 1450067.
[25] W. Cedeno, V.R. Vemuri, T. Slezak, Evol. Comput. 2 (1995) 321.
[26] J.J.P. Stewart, J. Comput. Chem. 10 (1989) 209.
[27] K. Deb, Multi-objective Optimization Using Evolutionary Algorithms, 1st ed., Wiley, Chichester, New York, USA, 2001.
[28] W. Cedeno, Ph.D. thesis, University of California Davis, CA, USA, 1995.
[29] T.H. Gordon, M.H. Gordon, M.J. Frish, J. Am. Chem. Soc. 113 (1991) 5989.
[30] A. Perczel, J.G. Angyan, M. Kajtar, W. Viviani, J.L. Rivail, J.F. Marcoccia, I.G. Csizmadia, J. Am. Chem. Soc. 113 (1991) 6256.
[31] I. Jàkli, A. Perczel, Ö. Farkas, M. Hollósi, I.G. Csizmadia, J. Mol. Struct. 455 (1998) 314.
[32] M. McAllister, A. Perczel, P. Császár, W. Viviani, J.L. Rivail, I.G. Csizmadia, J. Mol. Struct. 288 (1993) 161.
[33] M.A. Zamora, H.A. Baldoni, J.A. Bombasaro, M.L. Mak, A. Perczel, O. Farkas, R.D. Enriz, J. Mol. Struct. 540 (2001) 271.
[34] A.G. Császár, A. Perczel, Prog. Biophys. Mol. Biol. 71 (1999) 243.
[35] M.L. Ceci, M.A. López, S.S. Vallcaneras, J.A. Bombasaro, A.M. Rodrıguez, B. Penke, R.D. Enriz, J. Mol. Struct. 631 (2003) 277.
[36] F.C. Calaza, M.V. Rigo, A.N. Rinaldoni, M.F. Masman, J.C.P. Koo, A.M. Rodrıguez, R.D. Enriz, J. Mol. Struct. 634 (2003) 201.
[37] J.A. Bombasaro, M.A. Zamora, H.A. Baldoni, R.D. Enriz, J. Phys. Chem. A 109 (2005) 874.
[38] M.A. Sahai, S.S. Motiwala, G.A. Chass, E.F. Pai, B. Penke, I.G. Csizmadia, J. Mol. Struct. 666–667 (2003) 251.
[39] I. Jàkli, A. Perczel, Ö. Farkas, A.G. Császár, C. Sosa, I.G. Csizmadia, J. Comput. Chem. 21 (2000) 626.
[40] H.A. Baldoni, A.M. Rodríguez, M.A. Zamora, G. Zamarbide, R.D. Enriz, Ö. Farkas, L.L. Torday, C.P. Sosa, I. Jàkli, A. Perzel, J.G. Papp, M. Hollosi, I.G. Csizmadia, J. Mol. Struct. 465 (1999) 79.
[41] S.J. Salpietro, A. Perczel, Ö. Farkas, R.D. Enriz, I.G. Csizmadia, J. Mol. Struct. 497 (2000) 39.
[42] A. El Guerdaoui, R. Tijar, B. El Merbouh, M. Bourjila, R.D. El Bouzaidi, A. El Gridani, M. El Mouhtadi, Chem. Int. 2 (2016) 279.
[43] W. Chin, M. Mons, J.P. Dognon, R. Mirasol, G.A. Chass, I. Dimicoli, F. Piuzzi, P. Butz, B. Tardivel, I. Compagnon, G. Von-Helden, G. Meijer, J. Phys. Chem. A 109 (2005) 5281.
[44] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235.
[45] G.A. Chass, S. Lovas, R.F. Murphy, I.G. Csizmadia, Eur. Phys. J. D. 20 (2002) 481.