Method dependence of proline ring flexibility in the poly-L-proline type II polymer

On the Method Dependence of Proline Ring Flexibility in the Poly-L-Proline Type II Polymer

Michele Cutini, Marta Corno and Piero Ugliengo*

University of Torino, Department of Chemistry and NIS (Nanostructured Interfaces and Surfaces) Center, Via P. Giuria 7, 10125 Turin – ITALY

** e-mail: piero.ugliengo@unito.it*

Abstract

We studied the sensitivity of the energetic and geometrical features of the proline ring (pyrrolidine) to the quantum mechanical computational approach by adopting the proline monomer, trimer and polymer, as simplified collagen protein models. Within the Density Functional Theory (DFT) approach, we tested the ability of different functionals (GGA PBE and the hybrid B3LYP), added with *a posteriori* empirical dispersion corrections (D), to predict the conformational potential energy surface of the five-membered heterocycle pyrrolidine ring for the above models, dictating the collagen main features. We also compared the DFT-D results with those from the recently proposed cost-effective HF-3c method and our variant HF-3c-027, both based on Hartree-Fock Hamiltonian and Gaussian minimal basis set properly corrected for basis set superposition error, structure deficiencies and dispersion interactions. We found that dispersion interactions are essential to destabilize specific conformers. While the HF-3c and its HF-3c-027 variant are unreliable to predict accurately the energy of the ring conformers, structures are accurate. Indeed, the cost-effective DFT-D//HF-3c-027 approach in which the energetic is from the accurate DFT-D method on HF-3c-027 structures, provides energetic in line with that derived by the costly DFT-D//DFT-D approach, paving the way to simulate more realistic collagen models of much larger size. The adoption of either PBE or B3LYP functional for the electronic part of the DFT-D method gives very similar results, recommending the first as the most cost-effective method for simulating large collagen models. The predicted most stable conformation computed for the periodic poly-proline (type II) model allows for a higher flexibility, in agreement with experimental studies on collagen protein. The present results open the way to large-scale calculations of collagen/hydroxyapatite system, crucial for understanding the atomistic details in bones and teeth.

Keywords: DFT-D, dispersion interactions, collagen, poly-proline, pyrrolidine, ring puckering, HF-3c

Introduction

Collagen is one of the most abundant protein in mammals. It is the building block of complex hierarchical structures such as bones and tendons.[1] Its structural peculiarity is the geometrical motif in which three parallel polypeptides strands, in a poly-proline II type (PPII) fashion, coil about each other to form a triple helix.[1] The primary structure of collagen is mainly restricted to a triplet repeated sequence, which occurs in all types of collagens.[2] In each triplet, glycine (Gly) always occupies the first position, while proline (Pro) and its derivatives, e.g. Hydroxy-Proline (Hyp), are the most common amino acids in the second and the third positions, respectively. These positions of the collagen triplet are usually named as Xaa, the second, and Yaa, the third. Within all collagens, the Gly-Pro-Hyp (GPH) triplet appears most frequently (10.5%),[3] and Pro and Hyp represent the ~22% of all residues. The high content of imino acids gives stability to the triple helix and the pucker of the pyrrolidine ring plays an important role in this regard.[3]

There are several experimental proofs, provided by NMR studies, of the existence and interconversion between two proline ring puckers.[4–8] Pro envisages a five-membered heterocycle pyrrolidine ring as a side chain characterized by four carbon atoms usually labelled C_α, C_β, C_γ and C_δ. In turn, each carbon bond defines a torsion angle labelled as χ^1, χ^2, χ^3 and χ^4 (*vide infra*). London demonstrated that there are two stable conformational states for pyrrolidine rings in proline containing peptides, corresponding to a C_β-C_γ half-chairs geometry with C_γ more largely displaced than C_β.[4] Sone *et al.* focused on the side chain conformations of solid state PPII by ^{13}C NMR, proving the co-existence of two proline ring puckers also in the polymer.[5] Raman and Raman optical activity techniques were applied to the PPII polymer to characterize the side chain conformations at RT in water. The two ring conformations (hereafter indicated as DOWN (D) and UP (U)) of the proline ring puckers were found equally populated.[6] NMR studies on the dynamics of the pyrrolidine ring have demonstrated that the proline ring motion is uncorrelated to the backbone dynamics. Indeed, the motion of the backbone conformation is too slow to be affected by the proline ring conformation fluctuations, whose relaxation time is estimated to be about 1-30 ps.[4,7,8] Moreover, ^2H NMR study on the Pro crystal estimated the apparent activation energy of Pro ring flipping as 5.4 $kJ \cdot mol^{-1}$.[8] A very recent work on the rigidity of the two PPI and PPII helical forms, by means of both experimental (terahertz time-domain spectroscopy and X-ray diffraction) and theoretical methods (solid-state density functional theory), have highlighted the flexibility of these helices and found a great difference among the two poly-proline types in terms of Young's moduli.[9]

Several theoretical works have investigated the conformations and puckering of Pro as a single amino acid,[10,11] in small peptides,[12,13] and in proline oligomers.[14] The full description of the pyrrolidine ring

conformations is attained by computing the Adiabatic Potential Energy Surface (APES) on pseudorotational coordinates, e.g. phase angle and ring amplitude. In a pioneering work dated 1977, DeTar *et al.* studied the APES of the Pro side chain through molecular mechanics simulations.[15] More recently, Kapitán *et al.* computed the APES of Pro zwitterion in water by means of hybrid DFT functional.[16] The computationally demanding APES calculation simplifies by scanning the five-membered-ring potential energy using a single dihedral ring angle. Kang *et al.* used the χ^1 angle for the pyrrolidine ring.[17,18] As an alternative, few years ago, Wu proved that the χ^2 dihedral angle was a better descriptor of the puckering amplitude for a single proline aminoacid.[19]

As above-mentioned, the pyrrolidine ring has an important role in stabilizing the collagen triple helix. Many experimental as well as theoretical works have demonstrated that inducing the U conformation of Pro derivatives in the Yaa position (by the means of appropriate C^{γ} substitutions) increases the stability of collagen-like peptides.[20–22] Furthermore, inducing the D conformation of Pro in Xaa position stabilizes the triple helix.[23] These findings are related to the pre-organization of the D and U Pro to the Xaa and Yaa geometry of an ideal 7_2 helix, respectively.[2] Indeed, Pro and its derivatives show a strong interplay between side-chain pucker and main-chain conformation.[24] Interestingly, in the Gly-Pro-Pro (GPP) collagen-like peptide, the Pro puckering in the Xaa position prefers a D conformation. Instead, in GPH, the Pro in Xaa is almost uniformly distributed between the D/U states. Due to the interplay between the pucker and the main chain, the side chain mobility induces structural flexibility to the collagen protein. Recently, Chow et al.,[25] by means of NMR spectroscopy and force field calculations, have related the free ring flipping of Pro in GPH peptides to the presence of Hyp in Yaa position.

The focus of the present work is to establish an optimal computational procedure for characterizing the very delicate conformations of the pyrrolidine ring embedded in the collagen protein. We have considered three highly simplified models with increasing complexity, e.g. the proline molecule, the proline trimer and the polyproline type II polymer. For these three cases, we have computed the APES of the pyrrolidine ring both by pseudorotational coordinates and through a simplified procedure proposed by us as described in the section "APES on the χ^2 dihedral angle" of the Results and Discussion paragraph (*vide infra*). Among the various electronic structure methods for predicting periodic systems features,[26] such as polymers,[27] we focused on standard DFT, adopting the PBE[28] and B3LYP[29–31] functionals, with polarized valence triple-zeta (VTZP) quality basis set. Among the available London dispersion corrections,[32] we employed the Grimme's a posteriori correction to the DFT energy[33] using three different schemes, *i.e.* D2, D3 and $D3^{ABC}$.[34] All the DFT-based methods can become expensive for large systems, like a realistic triple collagen helix. Grimme and coworkers[35] have recently proposed a fast alternative to DFT, *i.e.* HF-3c, meant to reduce the

computational burden for large organic molecules. We have recently tested HF-3c to study the structure and energetic of a large set of molecular crystals, proposing a modified version of the method (hereafter named HF-3c-027)[36] which improves over the original HF-3c for the studied data set. Further work is needed to assess whether the HF-3c-027 improvement will also apply to a broader class of materials, inclusive of inorganic crystals. For an accurate estimation of energetic, we showed that running a single point B3LYP-D3^{ABC} on the HF-3c-027 optimum geometry (B3LYP-D3^{ABC}//HF-3c-027) is the proper procedure. In this way, we achieved a great time saving in the time-consuming geometry optimization, while keeping the accuracy in the energetic to the same level of a plain B3LYP-D3^{ABC}. We showed that the time saving with the above procedure is close to a factor 30. Here we extend that methodological comparison to establish a reliable reference for investigating more complex collagen models, *e.g.* a realistic full triple helix models, as well as to study its interactions with hydroxyapatite surfaces involved in bone mineralization processes. For all these cases, the need of a fast and reliable method like B3LYP-D3^{ABC}//HF-3c-027 is clearly mandatory.

Computation Details

All the HF-3c and DFT calculations were performed with a development version of the CRYSTAL14 code.[35,37] HF-3c is based on the Hartree-Fock method, with minimal MINI-1 basis set, supplemented by three correction terms curing the drawbacks of the plain HF/MINI-1 approach: i) a short range correction to cope with deficiencies of HF/MINI-1 in geometry prediction; ii) a basis set deficiencies correction term to reduce the large basis set superposition error associated with the MINI-1 basis set; iii) a London type term of the D3^{ABC} kind to cope with dispersion interactions. Interestingly, ii) and iii) corrections have been applied also to hybrid DFT based method.[38] We also employed the revised HF-3c-027 form of the original HF-3c, in which the s_8 term of the D3 scheme is scaled by a factor 0.27, as detailed in Ref [36]. For the DFT calculations, we used both the B3LYP hybrid functional,[29–31] and the PBE one.[28] The polarized valence triple-zeta (VTZP) basis set devised by Ahlrichs and co-workers was employed for all atoms,[39] with the exception of hydrogen in the BS1 computational setup (vide infra). Moreover, we have assessed the effect of different dispersion schemes: the D2 scheme[33] and its revised version[40] D* were applied for the PBE and B3LYP methods, respectively. The D3 correction,[34] including the Axilrod–Teller–Muto (ATM)-three-body-term (D3^{ABC}),[41,42] was employed for both PBE and B3LYP functionals.

For the 2D APES calculations, the χ^1 dihedral angle (see Fig. 1) was incremented by ~2.5° within 50° and -50°, and for each value of χ^1, the χ^2 was scanned with same step length and boundaries. For each fixed couple of dihedrals, all the internal coordinates and the cell vector were fully relaxed. In

the 1D APES, we only scanned one dihedral and keeping its value fixed during the optimization. The step length and boundaries were the same of the 2D case. When fixing one or more coordinates, the geometry optimization was run in redundant internal coordinates. The back-transformations from internal to Cartesian coordinates followed a procedure based on the conjugate gradient.[43] Atomic positions and cell vectors optimization adopted the analytical gradient method. The Hessian was upgraded with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm.[44–46] We set tolerances for the convergence of the maximum allowed gradient and the maximum atomic displacement to the default values (0.00045 Ha·bohr^{-1} and 0.00030 Bohr, respectively). The Γ-centered k-point grid was generated via the Monkhorst−Pack scheme[47] with a system dependent number of k points. The shrink factor for all polymer calculations was set to 4. The eigenvalue level-shifting technique was used to lock the system in a non-conducting state,[43] with level shifter set to 0.6 Ha. To help convergence of the SCF, the Fock/KS matrix at a cycle was mixed with 30% of the one of the previous cycle.[43] The electron density and its gradient were integrated over a pruned grid consisting of 974 angular and 75 radial points generated through the Gauss–Legendre quadrature and Lebedev schemes.[48]

We employed two computational setups, different in terms of tolerance values controlling the Coulomb and exchange series in periodic systems[43] and of employed basis sets, namely BS1 and BS2. Tolerances are set to 10^{-6} Hartree for Coulomb overlap, Coulomb penetration, exchange overlap and exchange pseudo-overlap in the direct space, and 10^{-14} for exchange pseudo overlap in the reciprocal space for the BS1 setup, while to 10^{-7} and 10^{-18} for the BS2. In both BS1 and BS2 cases, valence triple-zeta with polarization (VTZP) basis sets were adopted for carbon, nitrogen and oxygen atoms,[39] while hydrogen was described at the same VTZP quality level in BS2, but with a 3-1G(p) basis set[49] in the BS1 case (see SI Basis Details section for details). We run all the calculations with the BS1 setup, with the exception of the 2D-APES calculation for the pyrrolidine ring of the proline molecule, for which we employed the more accurate BS2 setup.

As discussed in the introduction, the adoption of the optimized structures at HF-3c (or HF-3c-027) for single point energy evaluation at DFT-D/VTZP//HF-3c brought about a 30 times speed up of the calculations compared to the full B3LYP-D3^{ABC}/VTZP//B3LYP-D3^{ABC}/VTZP level on the PPII system. When adopting PBE-D3^{ABC}/VTZP//PBE-D3^{ABC}/VTZP method we got a further speed up of 2.2 compared to the B3LYP-D3^{ABC}/VTZP//B3LYP-D3^{ABC}/VTZP level.

The graphical visualization and structural manipulation of structures was performed with MOLDRAW version 2.0.[50] Images were rendered with POV-Ray version 3.7.[51]

Results and discussion

Molecular Models

In this work we have considered three simplified models of proline within the collagen protein: *i)* the proline molecule (PRO); ii) the proline trimer, HN-$(Pro)_3$-COOH, (PRO3) and iii) the poly-proline type II polymer (PPII). The PRO model represents the L-proline molecule *in vacuum* with N-H in *cis* position with respect to C^{α}-H and an intramolecular H bond (N---HO). PRO resulted as the most stable conformer for gaseous proline at the CCSD(T)/CBS//B3LYP/6-311G(d,p) level of theory.[11]
Both the PRO3 and the PPII are in a PPII conformation, which is a 3_1 left handed helix with *trans* peptide bonds. Regarding the PRO3 terminal groups, both the N-H and C=O groups are in *trans* with respect to C^{α}-H. The PPII polymer is a one-dimension periodic system whose unit cell contains three proline molecules.
Since the ii) and iii) models have more than one Pro residue and in order to define a reference nomenclature for the side chain conformations, the Pro ring has been considered as 'DOWN puckered (D)' or 'UP puckered (U)' for negative/positive values of the χ^2 angle (see Figure 1 for its definition), respectively. The χ^2 suits as a marker for the proline pucker, as already proposed by DeTar and Ho.[15,24] For multiple Pro containing systems D and U represent a simpler way to describe the ring conformation than by using the pseudorotational coordinates, P and Q.
In this work, we considered the PRO3 system in the DDD and DUD states. Conversely, we investigated the PPII for all the side chain conformations, e.g. DDD, UDD, UUD and UUU. It is worth mentioning that the polymeric nature of PPII, imposes UDD≡DUD≡DDU, as well as UUD≡UDU≡DUU.

Pseudorotation of a Five-Membered ring

The side chain of Pro and derivatives consists of the five-membered heterocycle pyrrolidine. Usually, this ring puckers either in an UP or DOWN fashion, where four out of five ring atoms, namely N, C_{α}, C_{β}, and C_{δ}, lie on a common plane and Cγ carbon is relaxed out of the plane. Figure 1 shows the Pro atoms and dihedrals definitions. The proline is UP puckered (U) when the Cγ atom and the carbonyl group lie on the same side of the above-mentioned plane. Conversely, if the Cγ atom and the carbonyl group are at the opposite sides with respect to that plane, the proline is DOWN puckered (D).

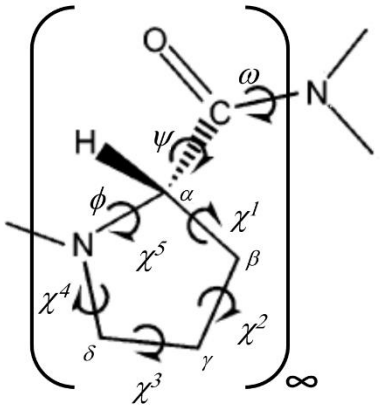


Figure 1. Definition of the atoms and torsional angles of a Pro within a polymeric structure.

For an extensive analysis of the Pro side chain conformations, we have applied the pseudorotation concept, firstly adopted to characterize cyclopentane,[52] but quickly employed to analyze heterocycles, such as furanose[53] and pyrrolidine rings.[15,54] It consists in only two variables defining the conformations of a five-membered heterocycle, i.e. the phase angle (P) and the ring amplitude (Q).[15] Following the definition in Ref [53]:

$$\tan(P) = \frac{\chi^4+\chi^1-\chi^3-\chi^5}{2\cdot\chi^2\left[\sin\left(\frac{\pi}{5}\right)+\sin\left(\frac{2\pi}{5}\right)\right]} \quad (1)$$

$$Q = \frac{\chi^2}{\cos P} \quad (2)$$

If $\chi^2<0$ then 180° must be added to P.[19] The five torsional angles can be calculated back from P and Q as follows

$$\chi^j = Q\cos\left(P + \frac{4\pi(j-2)}{5}\right) \quad (3)$$

with *j*=1, 2, 3, 4, 5.

Physically, Q represents the maximum value assumed by χ^j, and P identifies the type of ring puckering. If P assumes values that are even multiples of 18° (0, 36°, 72°, …), the ring is puckered in the half chair conformation (H), while for odd multiples of 18° (18, 54°, …), it adopts an envelope conformation (E). Both conformations are degenerate in cyclopentane, but when hetero-atoms are introduced in the ring, the degeneration breaks down, and one or more minima come up. Although 1), 2), and 3) are approximate equations, they correlate torsional angles within ±1° over the ranges of low energy conformations.

In Figure 2, we report cyclopentane in the E and H conformations. In the E conformation, four atoms lie on the same plane and the remaining atom, which is out of plane, is labelled as C. In the H conformation, A and B are defined as the atoms not directly bonded to F, that is the atom crossed by the C_2 axis. Moreover, in Figure 2, we report the pseudorotation pathway of a pyrrolidine ring: for

the odd (red line) and even (black line) multiples of 18° (e.g. E and H conformations) C, A and B atoms are reported, respectively. The A, B and C positions with respect to the C=O group are in superscript.

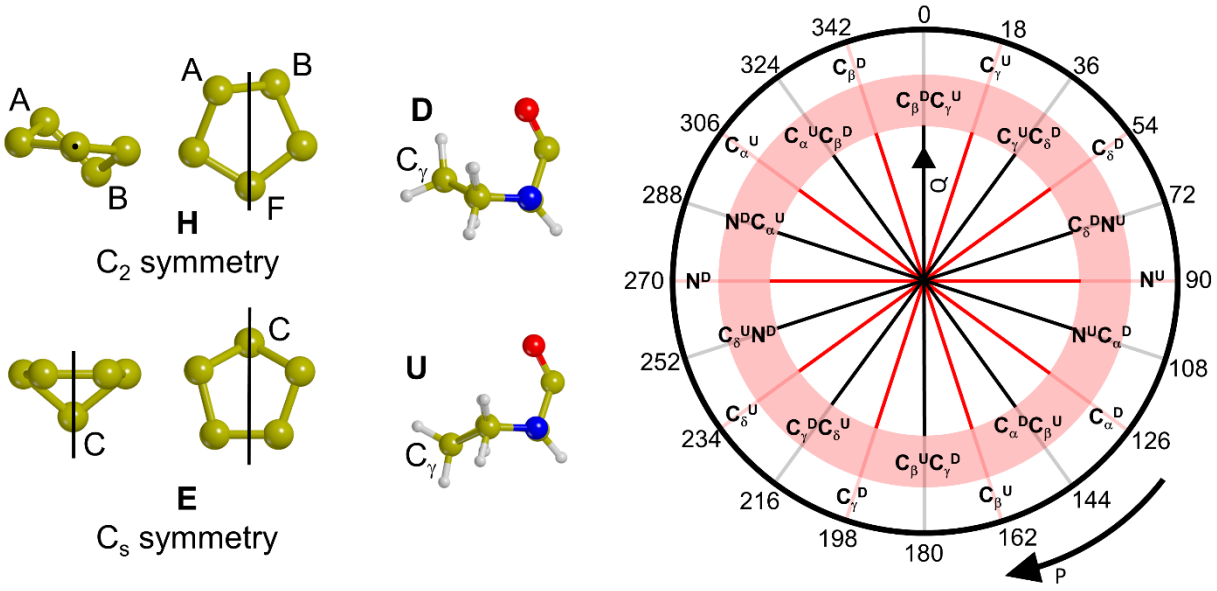


Figure 2. On the left, five-membered homocycles in the H and E conformations with symmetry axis and mirror plane. On the centre, Pro molecule extracted from a PPII polymer, in an E conformation with D and U C_γ atom, respectively. On the right, the pseudorotation pathway of a pyrrolidine ring (Q is radial and P is angular).

The Pro Ring Pucker

Initially, we have tested the performance of the HF-3c and HF-3c-027 as well as of the hybrid DFT-D/VTZP//HF-3c-027 approach in computing the 2D-APES for the side chain of PRO *in vacuum*, see Figure 3. The results for the HF-3c show the coexistence of three stable ring conformations, i. e. A, B and C, Figure 3. The adoption of HF-3c-027 method does not change the minima positions only slightly affecting their relative stability. When SP-B3LYP-D*≡B3LYP-D*/VTZP//HF-3c-027 is computed, the shape of the energy surface changes radically. A new large and shallower minimum (D) assimilates the B and C minima leading to a 2D-APES with two wells only, as expected from CCSD(T)/CBS calculations.[11] We compute the H C_γ^U/C_δ^D geometry (D state) 2.3 kJ·mol^{-1} higher in energy than the E C_γ^D (A state), see Figure 2. The SP-B3LYP-D* results are in good agreement with the wave-function reference method, e.g. CCSD(T)/CBS, see Table 1. The validity of the hybrid SP-B3LYP-D* approach is confirmed also by the full B3LYP-D3^{ABC}/BS2 method, that is here considered as a reference method for DFT approaches. The latter method combines both energy and geometry relaxation at the B3LYP level inclusive of the recent D3^{ABC} dispersion scheme, with the

most BS2 accurate computational setup. Interestingly, the reference method maintains the same shape of the SP-B3LYP-D* energy surface, only slightly affecting the conformer relative stability.

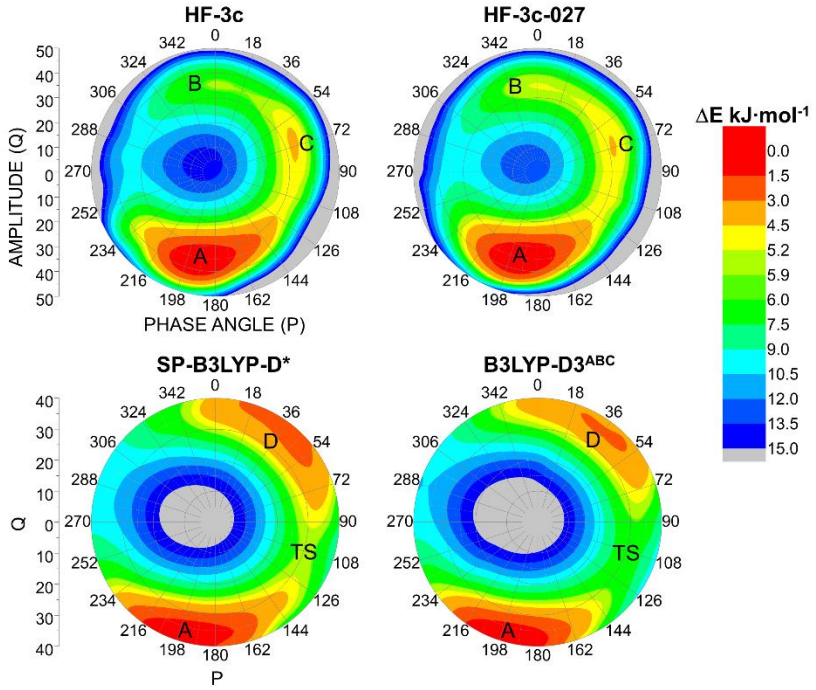


Figure 3. 2D-APES for the pyrrolidine ring of PRO computed with different methods. For all the computational methods, the down pucker conformation is found in the same APES region, namely A. Other stable conformers found in different position on the APES are named as B, C and D. The transition state between the A and D state is named as TS. See text for further details.

Table 1. Relative stability, ΔE (kJ·mol^{-1}), amplitude and phase angles, Q and P respectively (in degrees), for the fully optimized, A, B, C and D, PRO conformers. The transition state (TS) energy between the A and D state is also reported.

	ΔE	**Q**	**P**
HF-3c			
A	0	36.5	189.1
B	5.8	35.1	3.4
C	4.4	33.4	72.9
HF-3c-027			
A	0	35.8	190.4
B	5.1	34.5	2.1
C	4.5	32.5	69.5
[a]SP-B3LYP-D*			
A	0	39.6	195.5
D	2.3	36.2	35.9
TS	5.4	30.0	103.0
B3LYP-D3^{ABC}/BS2			
A	0	40.0	195.6
D	2.6	37.6	26.3
[a]TS	6.6	31.0	107.0
[b]CCSD(T)/CBS//B3LYP/6-311G(d,p)[11]			
A	0	//	[c]198.0
D	2.2	//	[c]18.0

[a] *The results are taken from the 2D-APES diagram.*
[b] *The results include the core correlation energy and the relativistic effects.*
[c] *See Ref* [11] *for further details.*

The effect of the Chain Length

To evaluate the effect of increasing the length of the Pro chain on the pyrrolidine ring mobility, we have computed the 2D-APES of the trimer and the polymer, both in the PPII conformation. We have investigated the DDD→DUD process at the HF-3c-027 and SP-B3LYP-D* levels of theory. From the SP-B3LYP-D*(see Figure 4), we deduced that increasing the length of the Pro chain lowers the pyrrolidine ring mobility. The TS energy for the ring flipping increases, i.e. from 5.3 kJ·mol^{-1} to 7.6 kJ·mol^{-1} up to 8.0 kJ·mol^{-1} for PRO, PRO3 and PPII cases, respectively, despite a moderate change in geometry (see Figure S1). Indeed, the TS for both PRO3 and PPII is in the E N^{UP} conformation,

while for the PRO case is in the $N^{UP}/C_{\delta}{}^{D}$ conformation, see Figure 2. The energy difference between D/U conformers increases, e.g. from 2.3 kJ·mol^{-1} to 4.3 kJ·mol^{-1} up to 6.4 kJ·mol^{-1} for PRO, PRO3 and PPII, respectively (Figure S1). Moreover, the low energy area on the 2D-APES decreases, as the Pro chain length increases. The strong relation between side and main chain in iminoacids may explain this effect. The conformational freedom of the main chain is, not surprisingly, lower in a protein than in a free residue. Therefore, by increasing the length of the system, the main chain stiffens, and so does the pyrrolidine ring.

The stable geometries of the side chain ring are very similar for both the trimer and the polymer at SP-B3LYP-D*. The ring geometry in the U polymer is strongly dependent on the dispersion scheme adopted to supplement the pure DFT method (vide *infra*). We noticed no pucker changes in the neighbouring Pro during the ring flipping.

It has to be underlined that, even if the HF-3c-027 method gives reliable results for the PRO3 molecule, it computes only one stable minimum for the PPII polymer.

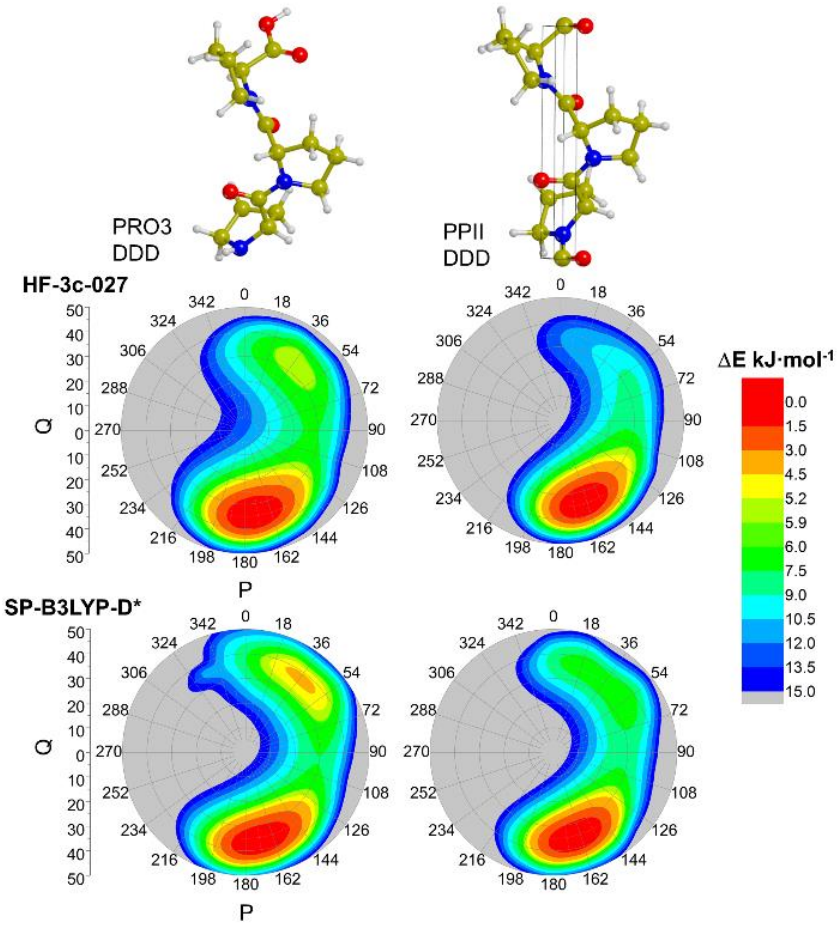


Figure 4. Comparison between the HF-3c-027 and the SP-B3LYP-D* methods 2D-APES for the DDD→DUD process for PRO3 and PPII.

APES on the χ^2 Dihedral Angle

As anticipated in the Computational Details, we have reduced the computational burden associated to the 2D-APES through the 1D energy scan only of the χ^2 dihedral angle. We first tested this approach for the PRO case at HF-3c level. Figure 5 shows the projection of the 1D path on the 2D-APES, which closely matches the full APES features. The error on the energy calculated at the A, TS1, B, TS2, and C points is less than 0.01 kJ·mol^{-1}. As expected, the results for the 1D energy scan on the χ^1 dihedral angle only, show hysteresis due to the high-energy path followed on the 2D-APES. The same agreement is between 1D and 2D scans for the DDD→UDD transition in PPII at SP-B3LYP-D* (see Figure S2). For the polymeric case, we calculated a deviation within 0.1 kJ·mol^{-1} on the energy between the 1D and 2D approaches.

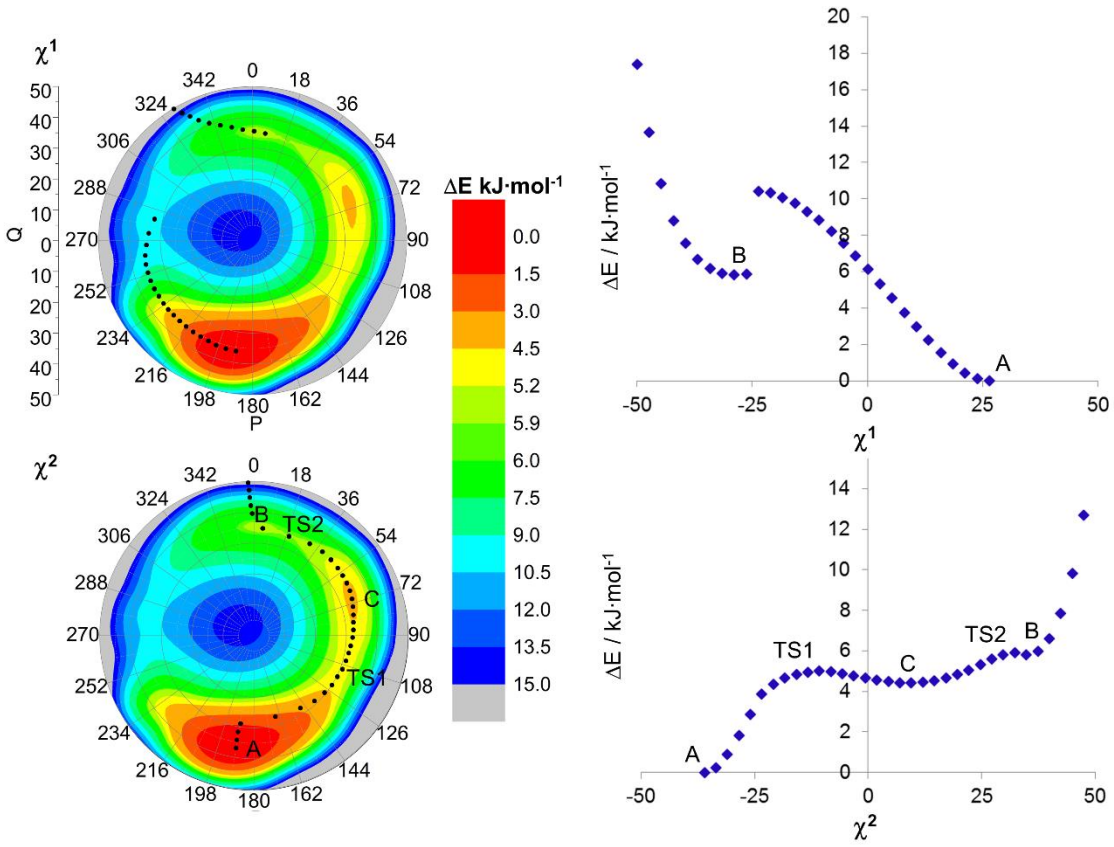


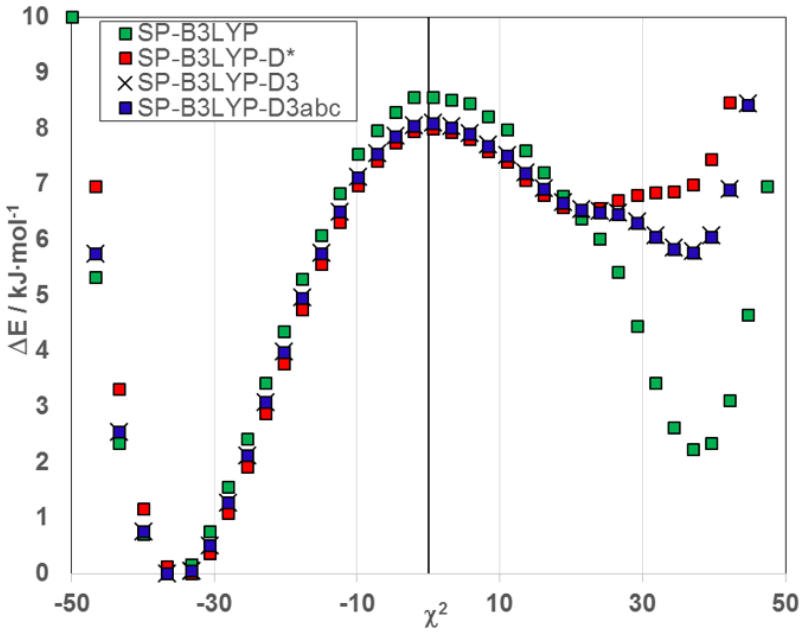
Figure 5. APES for the PRO side chain at the HF-3c level of theory. Dots represent the energy path on the 2D APES computed through χ^1 and χ^2 scans.

The Effect of Dispersion Scheme and Functional

With the simplified 1D approach, we have compared the effect of different dispersion schemes for the DDD→UDD process in the PPII polymer. On the HF-3c-027 1D-APES optimized geometries, we have calculated single point energies at the B3LYP level and at B3LYP-D*, B3LYP-D3 and B3LYP-D3^{ABC} level, see Figure 6. The dispersion free B3LYP energies exhibit a very deep minimum for the UDD conformer with a ΔE (E_{UDD}-E_{DDD}) of 2.2 kJ·mol^{-1}. The inclusion of dispersion contribution to the energy clearly destabilizes the UDD conformation, filling the well of the pure electronic calculation and leading to a rather flat energy profile in the region of the UDD state. The ΔE is 6.5 kJ·mol^{-1} and 5.8 kJ·mol^{-1} for the SP-B3LYP-D* and SP-B3LYP-D3 methods, respectively. The contribution of the ABC correction to the D3 energy is irrelevant. The destabilization effect due to the dispersion energy contribution on the U pucker is unrelated to the periodicity of the model. In PRO case, the dispersion energy flattens and destabilizes the U state region of the APES (Figure S3). The same destabilization effect appears in the PRO3 system (see Table S2). The flattening of the APES on the UDD region causes the variation of the minimum position. The SP-B3YLP-D* predicts a minimum in χ^2 = ~21° with P = 50° while the SP-B3LYP-D3 methods gives χ^2 = ~37° with P = 13°. Interestingly, the B3LYP calculation confirms the position of the UDD SP-B3LYP-D3 minimum. From high resolution X-ray structure analysis on the Protein DataBank,[15,24] we expected two minima on the APES, with values of P for U Pro of 12° and a χ^2 of ~37°. To investigate on this point, also the PBE functional coupled with the D2 dispersion scheme is tested. It computes the UDD region of the APES either with a minimum at χ^2= ~20° or with two almost degenerate minima (see Figure S5). Overall, the D3 scheme excels in accounting for dispersion interactions in PPII polymers compared to D2 and D* schemes. It is worth noting, that this effect is due only to the adopted dispersion scheme and not on the geometry optimization method (see Figure S5).

Moreover, we have investigated the effect of the chosen DFT functional on the energy. The SP-B3LYP-D3^{ABC} and SP-PBE-D3^{ABC} 1D-APES exhibit similar energy profiles, see Figure S6. The TS and minima occur at the same position, and the energetic only differs by no more than 1 kJ·mol^{-1}. Full geometry optimization confirms the agreement between B3LYP-D3^{ABC} and PBE-D3^{ABC}. The geometries on the DDD→UDD path show agreement between the HF-3c-027 and B3LYP-D3^{ABC} methods. In details, the main torsional angles on the minima positions change by 3.3° on average between the HF-3c-027 and B3LYP-D3^{ABC} methods. The deviation between B3LYP-D3^{ABC} and PBE-D3^{ABC} is 0.6° on average (see Figures S4 and S6). The effectiveness of the simplified 1D-APES χ^2 scan is confirmed for different methods by virtue of the projections on the P/Q surface (see Figure S2).

a)



b)

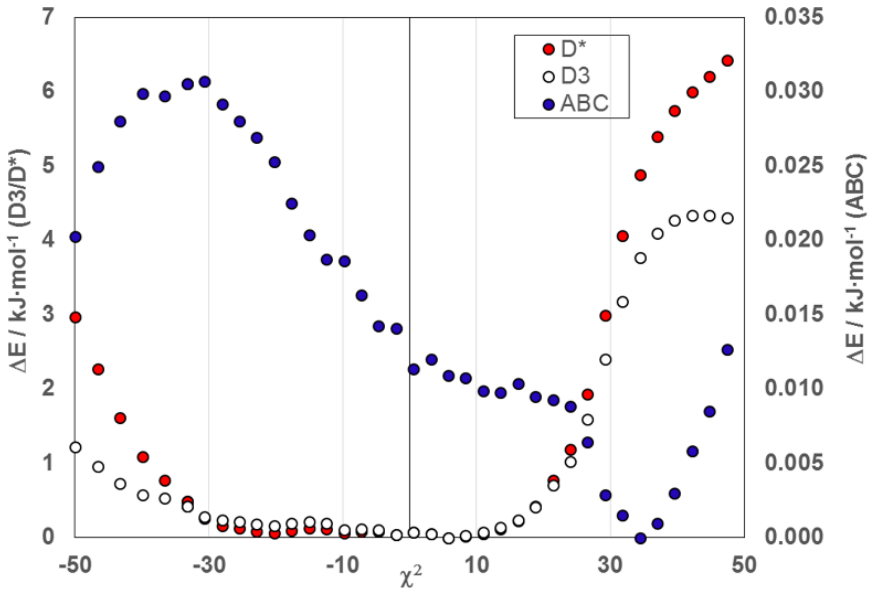


Figure 6. DDD →UDD APES on the χ^2 variable: a) energy differences for the SP-B3LYP-X level with three different dispersion schemes, i.e. X= D*, D3 or D3^{ABC}; b) dispersion energy contributions only for X=D*/D3 of the a) SP-B3LYP-X curves. In b), the left y-axis is for the D3/D* reference energy, the right y-axis is for the contribution due to the three-body dispersion ABC term.

Energy Path from DDD to UUU

We have explored all the possible conformations of the PPII polymer by computing the energy path, e.g. 1D-APES on χ^2, from DDD towards UUU conformer, Figure 7. Figure 8 shows the conformational state stabilities and TS energies at the SP-PBE-D3^{ABC}, SP-B3LYP-D3^{ABC}, PBE-D3^{ABC}, B3LYP-D3^{ABC}, HF-3c and HF-3c-027 levels. Our results demonstrate that by increasing the U Pro content in the polymer chain, the overall stability decreases (Table S3). The only exception is the UUU state, which is slightly more stable than the UUD state at the PBE-D3^{ABC} level only. The trend in the relative stabilities agrees between single point and full geometry optimization calculations, with an energy up-shift within 2 kJ·mol^{-1}. In contrast with the good results for SP-DFT-D methods, pure HF-3c methods performs poorly. Indeed, the UDD conformers is predicted as unstable, and the UUD and UUU states are at least 10 kJ·mol^{-1} less stable with respect to the DFT-D approach. Even though HF-3c-027 is unreliable on the energies, we found a very good agreement between SP-B3LYP-D3^{ABC} and B3LYP-D3^{ABC} geometries for all considered PPII conformers. In particular, the Pro stable puckering states are fully consistent between the full DFT-D optimization and SP methods, see Table S4, and largely agree with the experimental results on Pro ring pucker in biological systems.[4] Regarding the main chain torsional angles, as expected, the results show a progressive reduction of the ϕ from -77° to -56° and of the ψ angle from 160° to 132° moving from DDD to UUU state, respectively. The ψ trend parallels the results for the B3LYP/6-31G* relaxed Pro_6 geometries, as reported in Ref [55]. In the molecule, the reported average ψ values are 153° and 126° for the DDD and UUU cases, respectively. Conversely, the ϕ angle for the B3LYP relaxed hexamer moves from -72° to -66° on average. This trend is slightly different with respect to our results and this may be due to the periodic nature of our model. During the DDD→UUU transition ω oscillates of ±5° around 174°, which is close to the ideal *trans* peptide bond of 180°, see Figure S10. Although the small hysteresis on the UUD→UUU path, Figure S9, for all ring flipping, the TS states are always in an E N^{UP} conformation (Table S4).

Interestingly, the length of the unit cell shortens as the Pro ring puckers in the U fashion. The B3LYP-D3^{ABC} method computes a length of the unit cell of 9.12 Å, 8.96 Å, 8.77 Å and 8.61 Å for the DDD, UDD, UUD and UUU conformers, respectively. The PBE-D3^{ABC} gives similar trend with unit cell length slightly longer, e.g. 9.16 Å, 8.98 Å, 8.79 Å and 8.62 Å, respectively. At room temperature, the all DOWN state is the most populated for both B3LYP-D3^{ABC} and PBE-D3^{ABC} methods with a Boltzmann distribution of the 80% and 60%, respectively. If we keep a Pro ring fixed in the U

puckering, the population of the UDD, UUD and UUU states at RT will be almost equal for PBE-D3^{ABC}. Moving from the UDD to the UUU states will result in almost no energy change and a reduction of the unit cell length of ~4%. Therefore, the above data shows that the U pucker of a Pro in PPII system enhances the flexibility of the polymer chain at RT.

Assuming a first order reaction for the ring conformation interconversion, we calculated the half-life between puckering states, $t_{1/2}$. In the equations (3) and (4), we used the electronic energy E instead of the rigorous free Gibbs energy, G:

$$t_{1/2} = \frac{0.69}{k(T)} \tag{3}$$

$$k(T) = \frac{k_b T}{h} e^{-(E_{TS}-E_R)/RT} \tag{4}$$

Our results show a $t_{1/2}$ for a D→U pucker change of 3-10 ps, depending on the process, within the range of the experimental observed values of 1-30 ps.

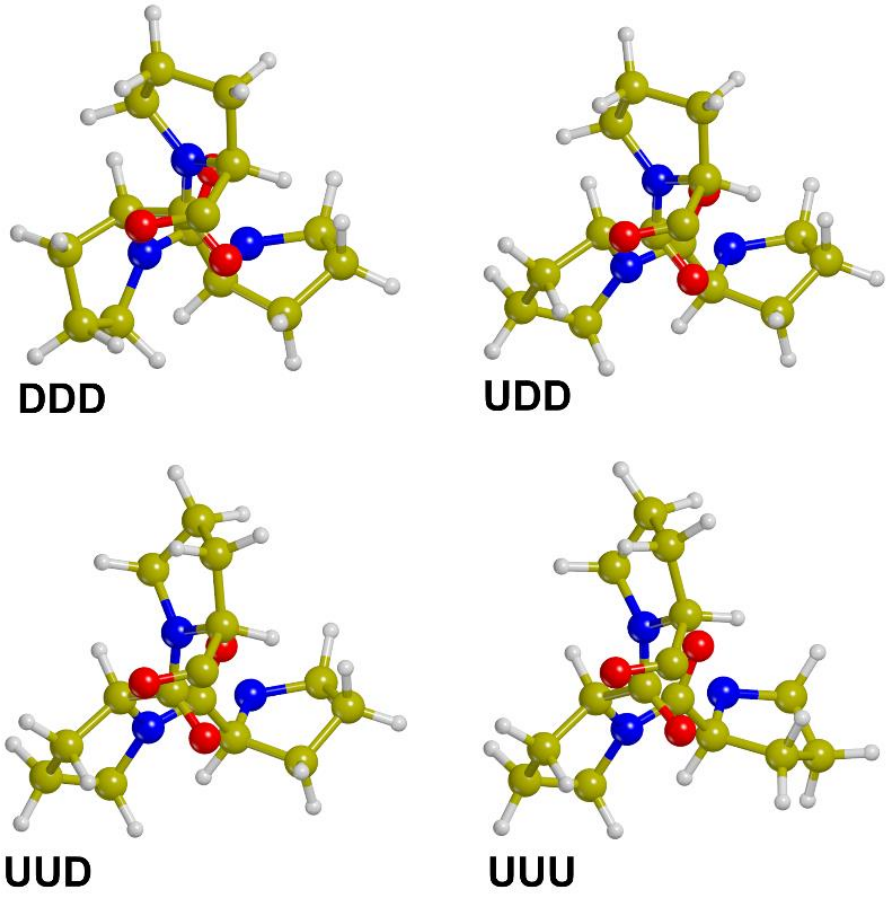


Figure 7. Views along the polymer chain of the B3LYP-D3^{ABC} relaxed structures of the PPII conformers.

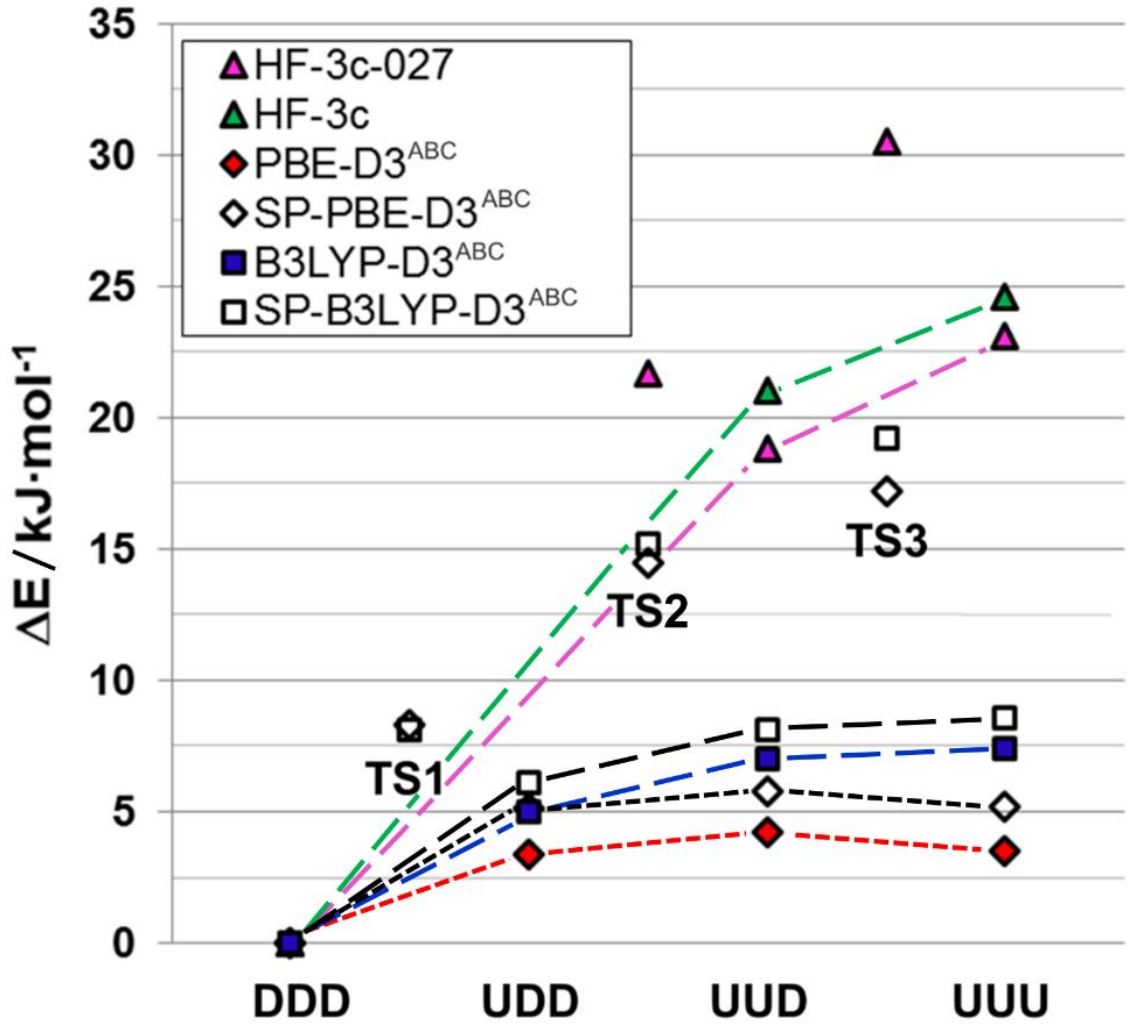


Figure 8. Energy ranking for the PPII polymer conformations. Lines are included to guide the reader eyes only. Transition state energies (TS) for all the process are reported. The UDD conformer is unstable at the HF-3c and HF-3c-027 levels (missing symbol).

Conclusions

In this work, we have investigated the sensitivity to the computational approach of the energetic and geometrical features of the proline ring (pyrrolidine) belonging to simplified models of the collagen protein, *i.e*. the proline monomer, trimer and polymer. The focus of the work is on the accurate prediction of the conformational flexibility of the five-membered heterocycle pyrrolidine ring for the above models. Following the standard nomenclature, the ring usually puckers either in an UP or DOWN fashion, where four out of five ring atoms lie on a common plane and Cγ carbon is relaxed out of the plane. The ring is UP puckered (U) when the Cγ atom and the carbonyl group lie on the

same side of the N, C_α, C_β, and C_δ plane while it is DOWN puckered (D) when they are at the opposite side. Within the DFT approach, we tested the ability of GGA PBE and the hybrid B3LYP functionals, supplemented with the D2 and D3 empirical dispersion corrections to predict the adiabatic potential energy surface of the pyrrolidine ring for each collagen model. Moreover, the results obtained by the application of the recently proposed cost-effective HF-3c method and our variant HF-3c-027, which improves over HF-3c for the structure prediction of a rather extended set of molecular crystals, were contrasted with the DFT results.

The main findings of our work are the following:

i) Dispersion interactions are of fundamental importance in determining the relative stabilities between different proline pucker conformers. We found the UP state systematically destabilized only when dispersion interactions are explicitly included in the calculations. This is an important point, as many computational works in the literature lack to account for dispersion interactions, which we proved crucial for the correct conformational behaviour.

ii) The HF-3c method is unreliable in computing the puckering for proline containing systems. In most of the cases, the shape of the HF-3c adiabatic potential energy surface as well as the conformer stability ranking is significantly different from the reference DFT-D values.

iii) Despite the poor results for the plain HF-3c method, the B3LYP-D*/VTZP//HF-3c-027 approach has proven to be a cost-effective method whose results are in full agreement with those of the B3LYP-D*/VTZP//B3LYP-D*/VTZP calculations, with a significant reduction of the computational cost.

iv) Calculated energies of the different conformers at the PBE and B3LYP levels were within few $kJ \cdot mol^{-1}$ to each other, showing the moderate effect of the hybrid functional in this topic. Therefore, considering that GGA functionals in CRYSTAL14 are about 2-3 times faster than hybrid functionals, we recommend PBE as the method of choice to deal with large collagen models. As for the empirical dispersion schemes, we suggest to use the D3 scheme, which showed to be more consistent and reliable than the D2 or D* schemes. The role of three-body as in $D3^{ABC}$ proved to be inessential.

v) In agreement with previous work,[19] we confirmed the adoption of one single ring dihedral angle (χ^2) as a fast alternative to scan the ring adiabatic potential energy surface.

vi) We showed that the effect of the UP puckering within the PPII polymer is to increase its flexibility. This is in agreement with the recent study on the collagen protein by Chow et

al.[25] in which they showed that a Pro followed by a Hyp, which is always UP puckered, has a frustrated puckering conformation leading to an enhanced protein flexibility. Therefore we demonstrated that the Pro frustration in collagen may be related only to the UP pucker geometry of Hyp and not to γ hydroxylation.

From a purely computational perspective, the proven reliability of the cost-effective DFT-D//HF-3c-027 approach paves the way to more challenging investigations, already under study in our laboratory. One is ab-initio simulation of the hybrid organic/inorganic PPII/HAP interface as a model of the collagen/hydroxyapatite interface, crucial for bone mineralization processes; the other deals with the design of more realistic collagen models, such as a Gly-Pro-Hyp triple helix polymers. Despite the promising results obtained at HF-3c-027 level, limitations are also apparent. For instance, the Gaussian nature of the HF-3c-027 model chemistry does not allow molecular dynamics calculations to be efficiently performed, like for the plane-wave DFT standard code incarnation. Even for DFT Hamiltonian, transition state structures are hardly characterized for constrained search variable, as in the case of the pyrrolidine ring. We, however, believe that the development of smart and computationally cheap methods like HF-3c, grounded on the rigour of quantum mechanics, will in the near future be competitive with classical force fields, especially for the study of the bio-organic molecules/inorganic surfaces interaction, which are critical for the present classical force fields.

Supporting Information

Basis set details, Details for the calculation of the energy path from the DDD to the UUU conformer, models energetic at the SP-B3LYP-D* level (Figure S1), projection of the fully optimized 1D-APES (Figure 2 and Table S1), effect of dispersion on PRO (Figure S3) and PRO3 APES (Table S2), geometry and energy of the DDD→ UDD reaction at different level of theory (Figure S4-7), geometry and energy of the DDD→ UUU reaction at the SP-DFT level (Figure S8 and S10, Table S3 and S4), projection of the 1D-APES on χ^2 dihedral angle onto the 2D for the DDD→UUU reaction paths (Figure S9).

This information is available free of charge via the Internet at http://pubs.acs.org/.

References

(1) *Collagen. Structure and Mechanics*; Fratzl, P., Ed.; Springer Berlin/Heidelberg: Postdam, Germany, 2008.
(2) Shoulders, M. D.; Raines, R. T. *Annu. Rev. Biochem.* **2009**, *78*, 929-958.
(3) Ramshaw, J. A. M.; Shah, N. K.; Brodsky, B. *J. Struct. Biol.* **1998**, *122* (1–2), 86-91.
(4) London, R. E. *J. Am. Chem. Soc.* **1978**, *100* (9), 2678-2685.
(5) Sone, M.; Yoshimizua, H.; Kurosu, H.; Ando, I. *J. Molec. Struct.* **1994**, *317*, 111-118.
(6) Kapitán, J.; Baumruk, V.; Bouř, P. *J. Am. Chem. Soc.* **2006**, *128* (7), 2438-2443.
(7) Mádi, Z. L.; Griesinger, C.; Ernst, R. R. *J. Am. Chem. Soc.* **1990**, *112* (21), 2908-2914.
(8) Sarkar, S. K.; Young, P. E.; Torchia, D. A. *J. Am. Chem. Soc.* **1986**, *108* (21), 6459-6464.
(9) Ruggiero, M. T.; Sibik, J.; Orlando, R.; Zeitler, J. A.; Korter, T. M. *Angew. Chemie Int. Ed.* **2016**, *55* (24), 6877-6881.
(10) Ramek, M.; Kelterer, A.; Nikolic, S. *Int. J. Quantum Chem.* **1997**, *65*, 1033-1045.
(11) Czinki, E.; Császár, A. G. *Chem. - A Eur. J.* **2003**, *9* (4), 1008-1019.
(12) Kang, Y. K.; Byun, B. J. *J. Comput. Chem.* **2010**, *31* (16), 2915-2923.
(13) Kang, Y. K.; Park, H. S. *Chem. Phys. Lett.* **2014**, *600*, 112-117.
(14) Zhong, H.; Heather, C. A. *J. Chem. Theory Comput.* **2006**, *2*, 342-353.
(15) Detar, D. F.; Luthra, N. P. *J. Am. Chem. Soc.* **1977**, *99* (4), 1232-1244.
(16) Kapitán, J.; Baumruk, V.; Kopecký, V.; Pohl, R.; Bouř, P. *J. Am. Chem. Soc.* **2006**, *128*, 13451-13462.
(17) Kang, Y. K.; Park, H. S. *J. Mol. Struct. THEOCHEM* **2005**, *718* (1–3), 17-21.
(18) Song, I. K.; Kang, Y. K. *J. Phys. Chem. B* **2005**, *109* (35), 16982-16987.
(19) Wu, D. *AIP Adv.* **2013**, *3* (3), 032141.
(20) Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. *Chem. Biol.* **1999**, *6*, 63-70.
(21) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392* (6677), 666-667.
(22) Shoulders, M. D.; Hodges, J. A.; Raines, R. T. *J. Am. Chem. Soc.* **2006**, *128* (25), 8112-8113.
(23) Shoulders, M. D.; Guzei, I. A.; Raines, R. T. *Biopolymers* **2008**, *89* (5), 443-454.
(24) Ho, B. K.; Coutsias, E. A.; Seok, C.; Dill, K. A. *Protein Sci.* **2005**, *14* (4), 1011-1018.
(25) Chow, W. Ying; Bihan, D.; Forman, C. J.; Slatter, D. A.; Reid, D. G.; Wales, D. J.; Farndale, R. W.; Duer, M. J. *Sci. Rep.* **2015**, *5* (February), 12556.
(26) Beran, G. J. O. *Chem. Rev.* **2016**, *116* (9), 5567-5613.
(27) Pham, T. H.; Ramprasad, R.; Nguyen, H. V. *J. Chem. Phys.* **2016**, *144*, 214905.
(28) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (18), 3865-3868.
(29) Becke, A. D. *J. Chem. Phys.* **1993**, *98* (7), 5648.
(30) Becke, A. D. *Phys. Rev. A* **1988**, *38* (6), 3098-3100.
(31) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37* (2), 785-789.
(32) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. *Chem. Rev.* **2016**, *116*, 5105-5154.
(33) Grimme, S. *J. Comput. Chem.* **2006**, *27* (15), 1787-1799.
(34) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
(35) Sure, R.; Grimme, S. *J. Comput. Chem.* **2013**, *34*, 1672-1685.
(36) Cutini, M.; Civalleri, B.; Corno, M.; Orlando, R.; Brandenburg, J. G.; Maschio, L.; Ugliengo, P. *J. Chem. Theory Comput.* **2016**, *12* (7), 3340-3352.
(37) Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco, P.; Noël, Y.; Causà, M.; Rérat, M.; Kirtman, B. *Int. J. Quantum Chem.* **2014**, *114* (19), 1287-1317.
(38) Brandenburg, J. G.; Caldeweyher, E.; Grimme, S. *Phys. Chem. Chem. Phys.* **2016**, *18* (23), 15519-15523.
(39) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.

(40) Civalleri, B.; Zicovich-Wilson, C. M.; Valenzano, L.; Ugliengo, P. *CrystEngComm* **2008**, *10*, 405-410.
(41) Muto, Y. *Proc. Phys. Math. Soc. Japan* **1943**, *17*, 629.
(42) Axilrod, B. M.; Teller, E. *J. Chem. Phys.* **1943**, *11*, 299-300.
(43) Dovesi, R.; Saunders, V. R.; Roetti, C.; Orlando, R.; Zicovich-Wilson, C. M.; Pascale, F.; Civalleri, B.; Doll, K.; Harrison, N. M.; Bush, I. J.; D'Arco, P.; Llunell, M.; Causà, M.; Noël, Y. *CRYSTAL14, User's Manual*; Università di Torino: Torino, Italy, 2014.
(44) Broyden, C. G. *IMA J. Appl. Math.* **1970**, *6* (1), 76-90.
(45) Fletcher, R. A. *Comput. J.* **1970**, *13*, 317-322.
(46) Shanno, D. F.; Kettler, P. C. *Math. Comput.* **1970**, *24* (111), 657-664.
(47) Monkhorst, H. J.; Pack, J. D. *Phys. Rev. B.* **1976**, *8*, 5188-5192.
(48) Prencipe, M.; Pascale, F.; Zicovich-Wilson, C. M.; Saunders, V. R.; Orlando, R.; Dovesi, R. *Phys. Chem. Miner.* **2004**, *31*, 559-564.
(49) Corno, M.; Busco, C.; Civalleri, B.; Ugliengo, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2464-2472.
(50) Ugliengo, P.; Viterbo, D.; Chiari, G. *Z. Krist.* **1993**, *207*, 9-23.
(51) Persistence of Vision Raytracer. http://www.povray.org (accessed November 26, 2016).
(52) Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. *J. Am. Chem. Soc.* **1947**, *69* (10), 2483-2488.
(53) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94* (23), 8205-8212.
(54) Pfafferott, G.; Oberhammer, H.; Boggs, J. E.; Caminati, W. *J. Am. Chem. Soc.* **1985**, *107* (8), 2305-2309.
(55) Zhong, H.; Carlson, H. A. *J. Chem. Theory Comput.* **2006**, *2*, 342-353.

Table of Content Graphic