Elucidating the Nature of Interactions in Collagen Triple-Helix Wrapping

Biophysical Chemistry, Biomolecules, and Biomaterials; Surfactants and Membranes

Elucidating the Nature of Interactions in Collagen Triple Helix Wrapping

Michele Cutini, Stefano Pantaleone, and Piero Ugliengo

J. Phys. Chem. Lett., **Just Accepted Manuscript** • DOI: 10.1021/acs.jpclett.9b03125 • Publication Date (Web): 18 Nov 2019

Elucidating the Nature of Interactions in Collagen Triple Helix Wrapping

Michele Cutini,[a] Stefano Pantaleone,[b] and Piero Ugliengo[a]**

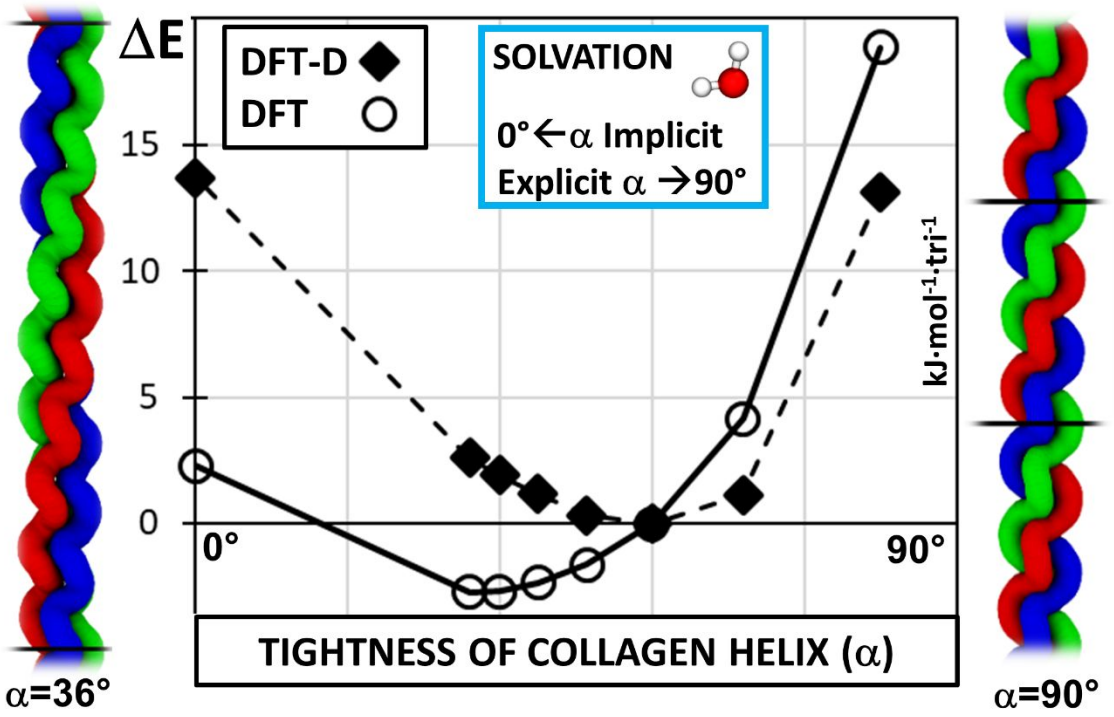
[a]University of Torino, Department of Chemistry and NIS (Nanostructured Interfaces and Surfaces) Center, Via P. Giuria 7, 10125 Turin - Italy

[b]Grenoble Alps University, CNRS, Institute of Planetary Sciences and Astrophysics, Grenoble (IPAG), 38000 Grenoble - France

*Corresponding Authors: michele.cutini@unito.it, piero.ugliengo@unito.it

Collagen is the most abundant protein family in the animal kingdom. Its structural motif envisages three polypeptide chains coiled in the so-called collagen triple helix. Depending on the triplet amino acidic sequence of the chains, collagen has different helical arrangements. Such atomic-scale structural variations have large impact on the large-scale structure of collagen. In this letter, we elucidate the interactions responsible of a specific helical pattern of the collagen protein by means of DFT-D based computer simulations. We demonstrate that inter-chains interactions and solvation effects stabilize compact helices over elongated ones. Conversely, elongated helices are stabilized by less geometrical strain and entropic factors. Our computational procedure predicts the collagen helical pattern in agreement with the experimental evidences.

TOC GRAPHICS



KEYWORDS: collagen, triple helix, DFT, micro-solvation, dispersion interactions

Protein structure prediction is a broadly studied problem since long ago.[1–4] In particular, due to the structure-property one-to-one correspondence, its understanding at atomistic level is very useful in the protein engineering field. Many factors contribute to the protein structure: amino acidic sequence, temperature, pressure, pH, presence and nature of the solvent, intra- and inter-molecular interactions.[1,5,6] Due to the large size of proteins, the multiple weak interactions (H-bonds and dispersive forces) both within the protein itself, and with the solvent, play a delicate role in determining the protein geometry.[7] The experimental approach to elucidate the weight of each component of these interactions is hindered by their intermingled nature. In this note, we show that modern density functional theory based on accurate hybrid functionals and large variational basis set is capable to cope with this problem.

Among all proteins, collagen is the main structural protein in mammals. It is found in skin, tendons, cartilage and bones, with the role of providing mechanical support to tissues. Collagen has a triple helical structure, in which three parallel polypeptide strands wrap together.[8–10] This geometrical organization imposes severe condition on the protein amino acidic content. Indeed, the amino acidic sequence rests on a specific triplet pattern

characterized by the presence of Glycine (Gly) every three amino acids (G-X-Y).[11,12] Proline (Pro) and (2S,4R)-4-Hydroxyl-Proline (Hyp) are the most common residues and occupy positions X and Y of the G-X-Y triplet, respectively. Therefore, Gly-Pro-Hyp is the most frequent amino acidic triplet in collagen.[13]

Collagen can organize in different helical geometries depending on the amino acidic sequence.[11] The helicity-composition relationship has been a matter of debate since long time ago.[14] A high amount of experimental evidences have proven that, for high content of Pros and derivatives, collagen exhibits a 7/2 helicity.[15,16] Conversely, in Pro-free zone, collagen seems to prefer a 10/3 helix.[17] The former case is a tight helix, in which 7 residues fit into two helical turns, thus there are 3.5 residues per turn, with amino acidic triplet rotation angle α = 51.4°. The latter is a loose helix, in which 10 residues fit into three helical turns (3.33 residues per turn with α = 36.0°). Even if the differences between these helices seem to be small at the atomic scale, they affect significantly the long triple helical domains.[11]

The purpose of the present letter is to clarify the interactions responsible to bring collagen in a specific helical geometry. This is of interest from both fundamental and applied points of view. Indeed, the present work also aims to provide useful insights for collagen protein

engineering.[17–26] To achieve our goals we rely on computer simulations, based on DFT and HF theories (see computational method section and SI for further details). We have analyzed seven collagen helices, which differ for the amino acidic triplet rotational angle (α), i.e. α = 90°, 72°, 60°, 51.4°, 40°, 36° and 0°. The higher the α value, the tighter the collagen helix is (see Figure 1). As shown in Figure 1, we have used a periodic model, in which the chosen motif is repeated infinitely. The advantages and details of this particular choice to represent collagen compared to finite models has been recently discussed in Ref. 27.

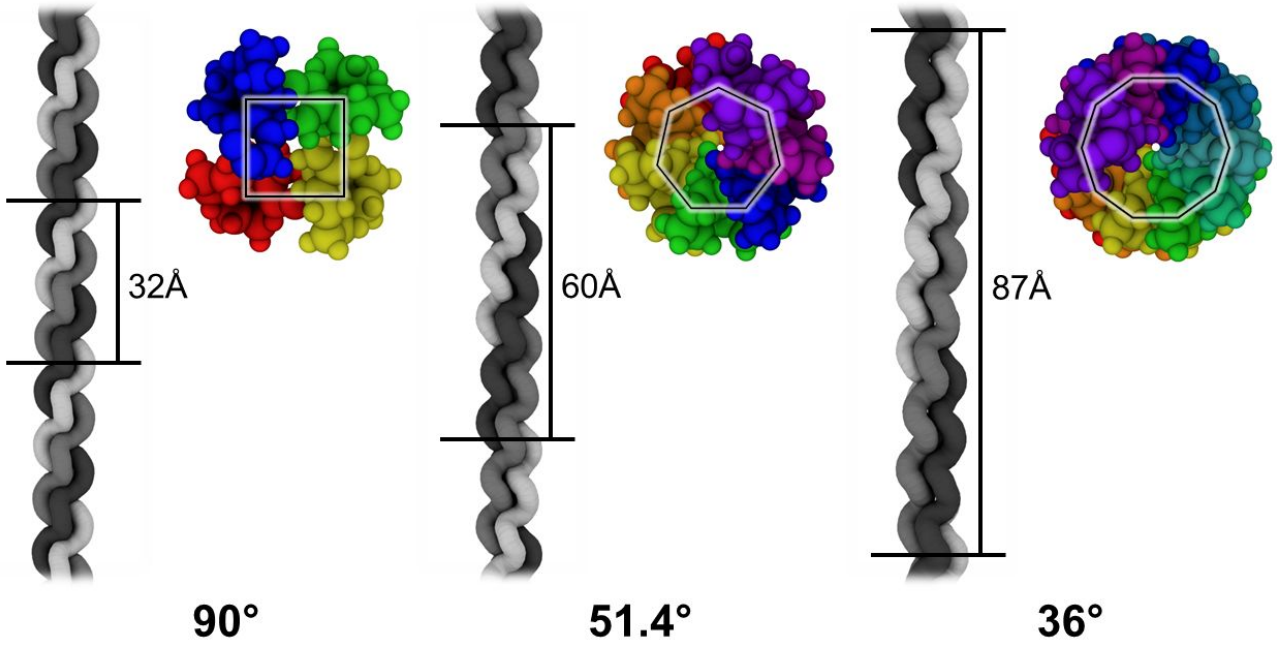


Figure 1. Collagen models with α = 90°, 51.4° and 36°. *Lateral view*: tube representation in different grey levels for each collagen single strand. The black segments delimitate the unit cell of the model. *Top view:* a single protein strand is reported with all amino acidic triplet

within the unit cell depicted in different colors. The geometrical shapes follow the triplets wrapping.

For the α = 0° case, the three collagen strands are no longer wrapped together, as they stand parallel to each other, while keeping the peculiar collagen inter-strand H-bond pattern. Regarding the composition, we have simulated homo-trimeric collagens, in which each strand is made of the repetition of the Gly-Pro-Pro triplet. We chose this composition because it resembles the most diffuse one in collagens, i.e. Gly-Pro-Hyp, but it is simpler to handle, see Ref. 27 for further details and a precise description of the models. As we want to compare the reasons of the relative energy stability, $\Delta E(COL)^C$, between different collagen wrappings, ultimately controlled by the α values, we define the following relationship:

$$\Delta E(COL)^C = E(COL_\alpha)^C - E(COL_{60.0°})^C = \Delta E(ss\text{-}COL) - \Delta BE^{*C} \qquad (1)$$

Where $E(COL_\alpha)^C$ is the BSSE-corrected ("C" apex) energy of a fully relaxed collagen triple helix (COL_α) with triplet rotation equal to α minus the energy $E(COL_{60.0°})^C$ of the COL with α = 60.0° set as a reference state. The $COL_{60.0°}$ structure is, indeed, the most stable wrapped conformation with the whole set of the quantum mechanical methods here adopted (see

points i-ii) in the computational method section). As there are no conflicting results, we do not discuss here the details, which can be found in the SI (see Figure S2 and S3 and Table S4). Therefore, from now on, the discussion will be focused only on the results obtained at the B3LYP-D3^{ABC}/VTZP levels, with some comparison with B3LYP/VTZP to assess the role of dispersion.

As anticipated, the $\Delta E(COL)^C$ has a minimum value for α = 60°, see Figure 2A, in good agreement with the experimentally reported value of α = 51.4°.[11] The discrepancy between the two cases (α = 60° and 51.4°) are fully justifiable, as the $\Delta E(COL)^C$ minimum is very shallow, with energy difference between the two structures of only 0.04 kJ mol^{-1} triplet^{-1}.

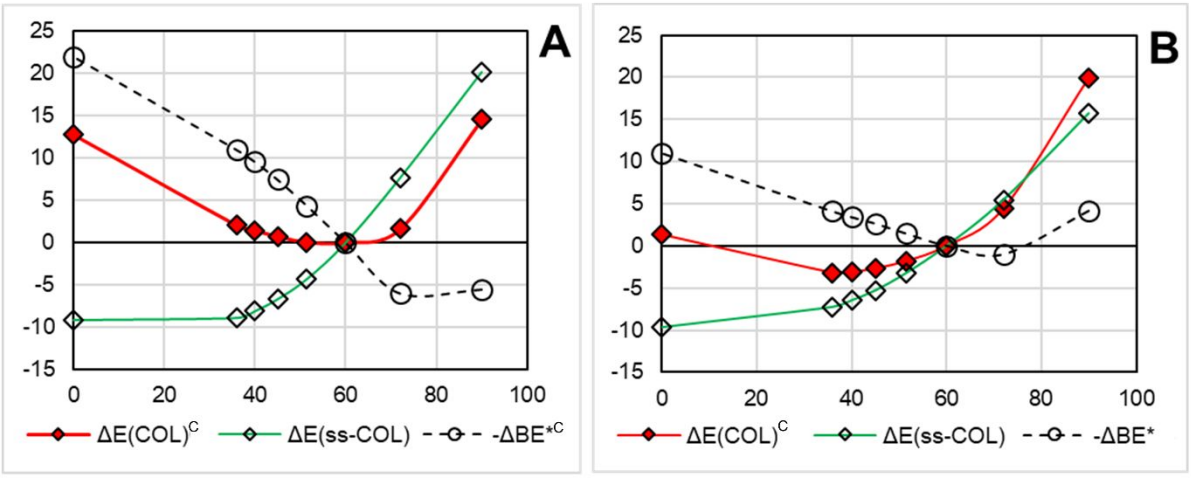


Figure 2. Energy decomposition of $\Delta E(COL)^C$ (vertical axis, kJ mol^{-1} triplet^{-1}) vs α (horizontal axes, degrees). **A**: Results at B3LYP-D3^{ABC}/VTZP level. **B**: Results at B3LYP/VTZP level. Lines reported to guide the readers eye only.

$\Delta E(COL)^C$ can also be dissected following the second term of equation 1), in which $\Delta E(ss\text{-}COL) = E(ss\text{-}COL_\alpha//COL_\alpha) - E(ss\text{-}COL_{60.0}//COL_{60.0})$ is the energy difference of a single collagen strand, fixed at the geometry assumed in the triple helix structure for an angle α, and the corresponding value for the collagen structure at α = 60.0° (the minimum). $\Delta E(ss\text{-}COL)$ is, therefore, a measure of the geometrical distortions of a single collagen strand due to a different angle wrapping compared to the most stable one. Its nature is obviously dominated by covalent intra-strand contributions (bond distances, angles and torsions changes). The term ΔBE^{*C} is the difference between $BE_\alpha{}^{*C} = E(ss\text{-}COL_\alpha//COL_\alpha) - E(COL_\alpha//COL_\alpha)$ and the corresponding $BE_{60.0}{}^{*C}$ one. It measures the BSSE-corrected inter-strands interaction energy within a triple helix with wrapping angle α with respect to the energy of the most stable one (wrapping angle equal to 60.0°). Interestingly, within the $\Delta E(COL)^C$ these two terms act against each other: indeed, $\Delta E(ss\text{-}COL)$ favors less packed

structures with less twisted geometries at variance with ΔBE^{*C} which favors more packed ones (see Fig. 2).

We also split the pure DFT contribution (electrostatic, exchange, polarization and charge transfer) from the dispersive (D) contributions from ΔE(ss-COL) and ΔBE^{*C} terms. The results can be summarized as:

1) the ΔBE^{*C} term is dominated by the dispersion interactions (Figure S4) in good correlation with the compactness of the helix (Figure S5). The pure DFT contribution to BE^{*C} comes from the interactions of N-H and C-H groups with the C=O group of Pro in X position within the protein core (Figure S6).
2) the ΔE(ss-COL) term is dominated by the DFT contribution (Figure S4), mainly arising from the dihedrals, angles and bond distances deformation occurring in the collagen helix.

Therefore, if we exclude the D term from our computational set-up, *i.e.* running the calculations with the bare B3LYP method (or any other pure GGA/hybrid functional), the ΔE(ss-COL) term will dominate the ΔE(COL) expression, leading to an artificial stabilization of more elongated helical geometries. This is clearly shown in Figure 2B, in which the

B3LYP/VTZP simulation, leads to a shift of the minimum region from α = 60° to 0° $< \alpha <$ 36°. Therefore, lowering inter-strand interactions loosens the collagen triple helix structure and *vice versa.*

To better mimic the collagen environment, we should take into account dynamic and solvent effects. Within the quantum mechanical framework, they can be rigorously included with *Ab-Initio* Molecular Dynamics (AIMD) simulations in explicit solvent. Unfortunately, this is undoable at present, due to the large size of the protein models. A possible and popular alternative, is to rely on classical Force-Field Molecular Dynamics (FFMD) simulations. Unfortunately, the very tiny energy differences characterizing the various helix conformations (see Figure 2A-B) are at the limits of the FFMD accuracy.

We choose a more pragmatic and simpler approach to keep the cost of the calculations reasonable while enforcing good accuracy: i) dynamical effects are accounted for by classical statistical thermodynamic based on the harmonic approximation of the vibrational contributions, which define the entropic contribution to the Gibbs free energy. As the calculation of the whole set of frequencies is time consuming for large structures, we relied on the fast HF-3c method,[28] ii) solvent effects are estimated through the Polarizable

Continuum Model (PCM) which brings to the free energy G(COL)sol definition, computed at room T and in water solvent:

$$G(COL)sol = E(COL)^C + HYD + ZPE + H - TS \quad (2)$$

Here, the $E(COL)^C$ is the BSSE-free total energy of collagen (*vide supra*), the HYD term is the interaction energy of the protein with the continuum medium along with the cavitation energy, the ZPE term is the protein vibrational zero-point energy computed within the harmonic approximation, the H term is the protein thermal vibrational energy correction at room T and S is the protein vibrational entropic contribution to the Gibbs free energy. The vibrational quantities are computed correcting the HF-3c harmonic frequencies with a 0.86 scaling factor, as suggested in the original paper.[28] The relative collagen free energy, $\Delta G(COL)sol$, is defined along the line of eq. (1) as:

$$\Delta G(COL)sol = G(COL_{\alpha})sol - G(COL_{60.0^{\circ}})sol \quad (3)$$

and following eq. (2), $\Delta G(COL)sol$ can be decomposed in its components:

$$\Delta G(COL)sol = \Delta E(COL)^C + \Delta HYD + \Delta ZPE + \Delta H - T\Delta S \quad (4)$$

The terms of eq. (4) are shown as a function of α in Figure 3.

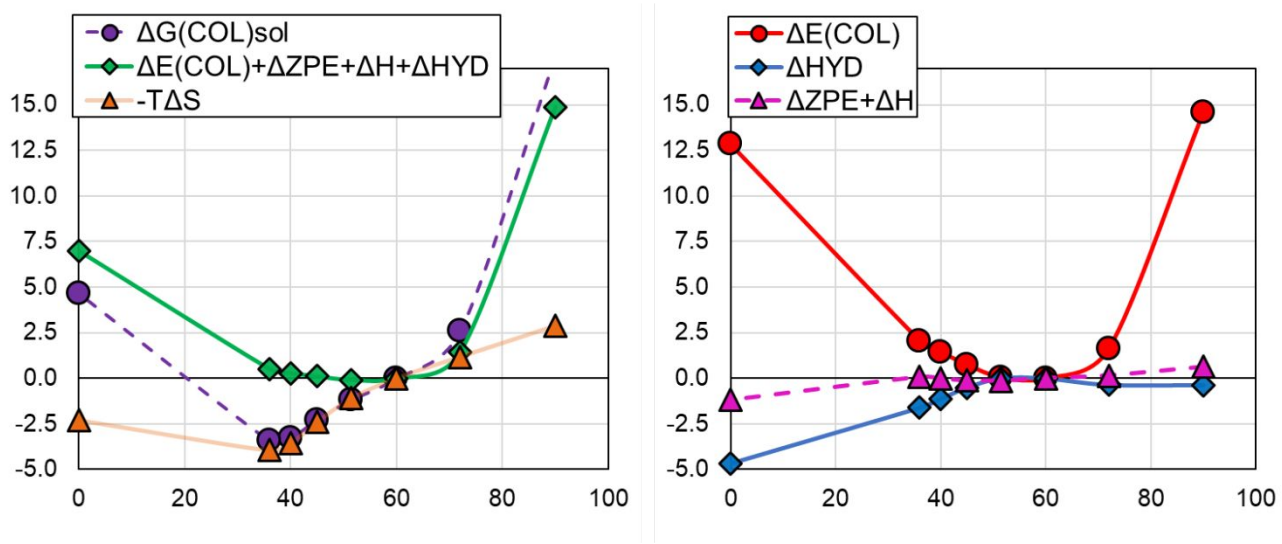


Figure 3. Decomposition of ΔG(COL)sol computed at room T vs α. Energy (vertical axis) in kJ mol^{-1} triplet^{-1}) and α (horizontal axis) in degrees.

Remarkably, by including the dynamic and solvent effects, the minimum of the energy curve moves to lower α = 36° values, a clear indication of the triple helix unwrapping. The change is mainly due to entropic effects, which stabilize more elongated structures. The continuum solvation model gives solvent-protein interaction similar for most of the cases. Only for the most elongated structures (small α values) hydration energy HYD is higher. We believe this is due to the more exposed Gly and Pro (Y) C=O groups, which maximize the protein-continuum solvent interaction. Zero-point and thermal vibrational energies do not vary notably with α, thus we have grouped them together in Figure 3.

The over-stabilization of the elongated collagen helices (small α values), due to hydration and entropic effects, leads to results in disagreement with the experimental evidences, for which α = 51.4°. We believe the problem is due to: i) the accuracy of the HF-3c computed vibration frequencies; ii) the implicit solvation approach, not accurate enough when specific hydrogen bond interactions are important as in the present case. As for point i), we computed, for few cases only, the vibrational frequencies at the more accurate B3LYP-D3^{ABC}/VTZP level. The vibrational corrections, reported in Table S3, indicate that the HF-3c level is capable of computing the right trend, but, indeed, over-stabilizes the elongated

structures. As for point ii), we improved the solvation model via explicit micro-solvation. In micro-solvation, few critically important water molecules are explicitly added to the models, with the purpose of representing the most important H-bonds between explicit water molecules and hydrophilic groups of the collagen polymer. On top of the micro-solvated model we run the PCM approach to account for long range bulk solvation effects. Unfortunately, describing the solvent through the micro-solvation approach increases the cost of the simulations notably. Therefore, we have applied it only to the two most representative cases, i.e. collagens with α = 36° and 51.4°. We have analyzed three levels of micro-solvation with 3, 4 and 5 water molecules per amino acidic triplet, named as 3w, 4w and 5w, in the following. The models are reported graphically in Figure 4.

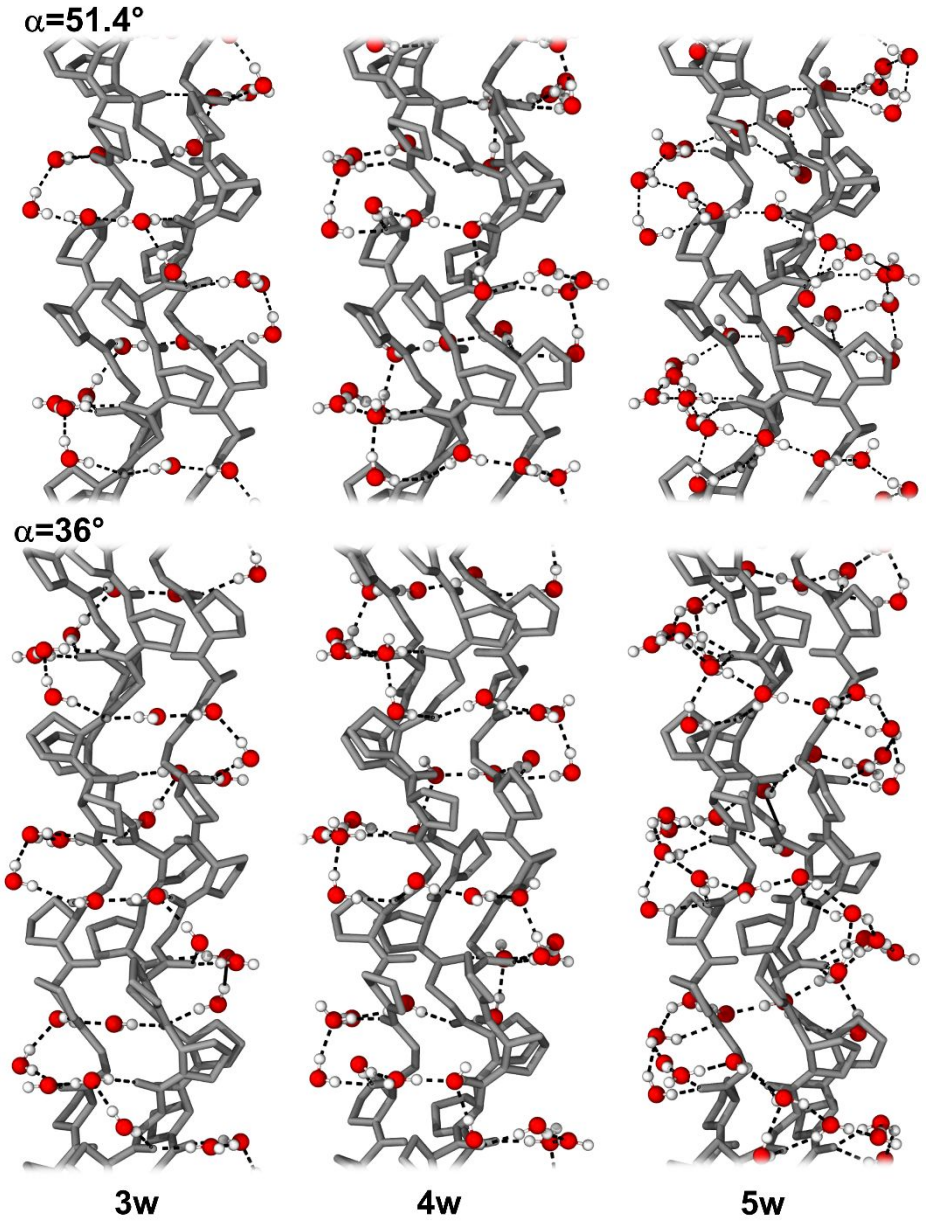


Figure 4. Micro-solvated collagen triple helices (full unit cell shown) with α = 51.4° (TOP) and α = 36° (BOTTOM).

We have computed the G(COL)sol for all models as defined in eq. (3). In this case, the BSSE-corrected total energy of the micro solvated collagens ($E(COL)^C$) is decomposed as follows:

$$E(COL)^C = - BE^{*C} + E(ss\text{-}COL) - BE^{*C}_{hyd} - BE^{*C}_{wat} + E(W) \quad (5)$$

Where the BE^{*C} (i), BE^{*C}_{hyd} (ii), and BE^{*C}_{wat} (iii) are the BSSE corrected interaction energies between: i) collagen strands; ii) the solvated waters and the protein; iii) water molecules. The E(ss-COL) term, as defined above, is the energy of an isolated single collagen strand and the E(W) term is the sum of the energies of each isolate water molecule, both energy terms considered at the relaxed micro solvated collagen geometry. Substituting eq (5) into eq (3), and then subtracting the corresponding free energies between helices, we obtain the ΔG(COL)sol expression for the micro solvated case:

$$\Delta G(COL)sol = G(COL_{36^\circ})sol - G(COL_{51.4^\circ})sol =$$

$$-\Delta BE^{*C} + \Delta E(ss\text{-}COL) - \Delta BE^{*C}_{hyd} - \Delta BE^{*C}_{wat} + \Delta E(W) + \Delta HYD + \Delta ZPE + \Delta H - T\Delta S \quad (6)$$

We reported in Figure 5 the above-mentioned terms contributing to the ΔG(COL)sol definition for micro solvated (3w, 4w and 5w) implicit solvated (PCM) collagens.

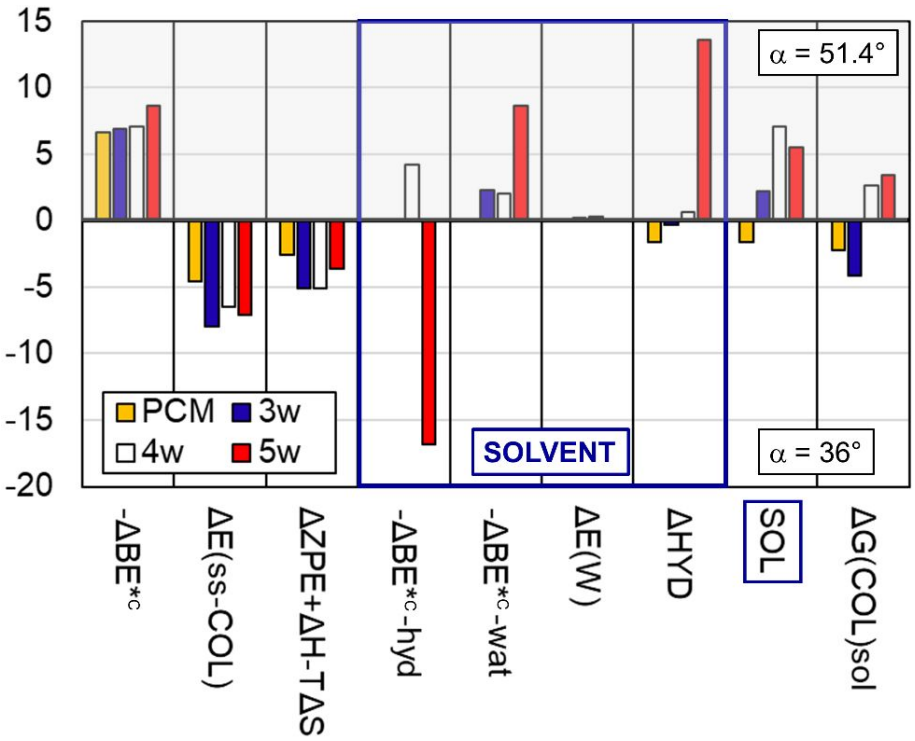


Figure 5. ΔG(COL)sol (vertical axis, kJ mol^{-1} triplet^{-1}) contributions (see Equation 6) for micro solvated (3w, 4w and 5w) and implicit solvated (PCM) collagens. Positive/negative ΔG(COL)sol values stabilize/destabilize the α = 51.4° wrapping.

As in gas-phase, also in micro solvation, the $-\Delta BE^{*C}$ term stabilizes tight collagens (α = 51.4°), as shown by the bars in the positive region of Figure 5; on the contrary, both the ΔE(ss-COL) and the vibrational/entropic contributions counterbalance the $-\Delta BE^{*C}$ term (bars in the negative region of Figure 5). Differently from gas-phase cases, when explicit solvation is taken into account the $-\Delta BE^{*C}$ term has a balanced contribution from DFT and D components due to the shortening of the inter-chains N-H---O=C H-bond (Table S10-11).

The $-\Delta BE^{*C}_{hyd}$, $-\Delta BE^{*C}_{wat}$, $\Delta E(W)$ and ΔHYD terms, highlighted in a blue frame in Figure 5, describe the solvent contributions to the free energy difference. All solvent effects contributions are grouped in the SOL term of Figure 5. Each single component, as well as the overall value, vary notably depending on the level and geometry of solvation. Notable is the 5w case; in this case the α = 51.4° collagen has a low value for the water-protein interaction ($-\Delta BE^{*C}_{hyd}$) which is balanced by a high inter-waters ($-\Delta BE^{*C}_{wat}$) and water-bulk (ΔHYD) interaction. The comprehensive SOL term amounts to -1.6, 2.2, 7.1 and 5.5 kJ mol^{-1} triplet^{-1} for the PCM, 3w, 4w and 5w cases. As expected, the sole PCM stabilizes the elongated collagen and it is inaccurate to simulate the water-collagen case.

The computed interaction energy felt by each water molecules, which comes from the bulk and the inter-waters interaction, increases with the increasing number of water molecules. This is a consequence of the shortening of the inter-water H-bond length, as shown in Figure S10, due to enhanced H-bond cooperativity. An exception to this trend is the 5w case with α = 36°, in which an addition water-protein H-bond reduces the inter-waters and water-bulk interaction with respect to the other cases. Interestingly, the inter-waters cooperative effects also affect the H-bond strength with the protein, which increases along with the inter-waters

and water-bulk interaction, see Figure 6. This is due to a shortening of the water-protein H-bond length as reported in Figure S10.

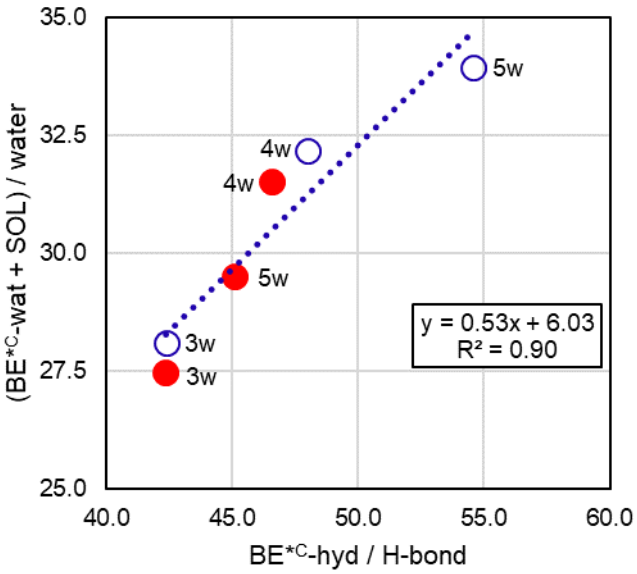


Figure 6. Correlation between average water-protein (vertical axis) and water-water binding energy (horizontal axis), normalized by the number of water-protein H-bonds and the number of water molecules, respectively. Empty blue circle for α=51.4°, red filled ones for α=36°. Energy in kJ mol^{-1} triplet^{-1}.

Overall, the computed ΔG(COL)sol is -2.2, -4.1, 2.6 and 3.4 kJ mol^{-1} triplet^{-1} for the PCM, 3w, 4w and 5w cases, respectively (see Fig. 5). Only for the collagens with high solvation, i.e. 4w and 5w case, the α = 51.4° helix conformation is favored, in agreement with the

experiments. This indicate that the minimum water number to solvate accurately Gly-Pro-Pro collagen models is 4 water per aminoacidic triplet.

In summary, we have built up a robust, controllable, reproducible and relatively cheap computational procedure capable of reproducing the experimental evidences about collagen helical wrapping. The fine physico-chemical features driving collagen to specific helical wrappings are captured by accurate hybrid DFT simulations on reliable collagen protein polymer-like models. The models included also explicit water micro solvation supplemented by PCM solvation to mimic bulk water effects. Through these simulations we have dissected the energetic terms and discovered those favoring more compact/elongated collagen helices. We demonstrated that a tight collagen helices wrapping is favored by: i) a stronger inter-chains interaction, not only coming from the dispersion interaction, but also from inter-chains electrostatic/polarization/charge-transfer terms; ii) a favorable water-collagen interaction. On the contrary, a loose collagen helices wrapping is stabilized by: i) a more stable polymer structure, resembling closely the expected geometry of a free Pro rich peptide (poly-Pro type II); ii) entropic factors. The delicate balance of these factors induces

more compact helices to be more stable for the more realistic micro solvated collagens. In these cases, the helical features are in agreement with the experimental evidences.

The approach presented here is tested on the Gly-Pro-Pro collagen composition, which is well known in literature, but it can be extended to any other collagen composition. Preliminary results on collagens with Gly-Leu-Hyp and Gly-Phe-Hyp compositions indicate a preference for more elongated helices with respect to Gly-Pro-Pro, in agreement with the experimental evidences.[17] The proven predictivity of this approach can be useful for the design of collagen like peptides with specific helical features.

Computational Methods

We relaxed the geometry for all models using the CRYSTAL17 suit,[29] relying on several *ab-initio* methodologies. Within the DFT framework we employed the B3LYP-D,[30–32] and PBE-D,[33] functionals with a large Gaussian VTZP quality basis set.[34] The adoption of localized Gaussian function implies that many quantities are affected by the basis set superposition error (BSSE) which has, therefore, taken into proper account. Due to the very large size of the systems we run frequencies calculations with the cost-effective HF-3c method.[28] Furthermore, we run single point energy calculations to assess the effect of: i) the basis set quality, ii) the dispersion forces description, and iii) the water solvation. To do so we also employed the VASP code.[35–38] Further details on the computational approach are reported in the SI.

Associated Content

Supporting information content: Computational Methods; Collagen Models Description; Definition of the Computed Quantities; Table S1-2: Basis sets employed; Table S3: Vibrational corrections to the energy for gas phase collagens at the HF-3c and DFT-D levels; Table S4: Relative total energies between helices in function of the method; Table S5: BE*, ΔE(ss-COL) and ΔE(COL) and all their contributions for DFT and DFT-D results; Table S6: Computing BSSE with the counter poise method and a plane waves basis sets; Figure S2-3: Dispersion scheme and geometry relaxation effect of energy ranking; Figure S4: Energy decomposition of ΔE(ss-COL) and ΔBE* terms; Figure S5: Height per triplet and dispersion component of BE correlation; Figure S6: Average inter-strands electrostatic contacts lengths in function of α; Figure S7: Correlation between ΔE(ss-COL) and $-\Delta BE^{*C}$ terms with α; Figure S8: Main torsional angles at the HF-3c level of theory in function of α; Figure S9: Correlation of TΔS, ΔET and ΔE0 contributions with the axial rise per residue; Energy Decomposition in Micro-Solvated Collagens (Table S7-10); Table S11: Geometrical analysis of the COL-W models, Figure S10: Correlation energetic and geometry within micro solvated collagen models.

Author Information

Corresponding authors: Dr. Michele Cutini and Prof. Piero Ugliengo

NAME: ORCID

Dr. Michele Cutini: 0000-0001-6896-7005

Dr. Stefano Pantaleone: 0000-0002-2457-1065

Prof. Piero Ugliengo: 0000-0001-8886-9832

Acknowledgments

SP acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, for the Project “The Dawn of Organic Chemistry” (DOC), grant agreement No 741002. MC acknowledges Massimo Bocus, M.Sc. and Irene Bechis, M.Sc for the efforts devoted to support the “Collagen Project”. PU and MC acknowledge the University of Torino for funding (grant agreement No CHI.2019.21/XXII) and the CRYSTAL team for continuous support in the usage of the CRYSTAL program.

References

(1) Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, *426*, 884.

(2) Dill, K. A.; Maccallum, J. L.; Folding, P. The Protein-Folding Problem , 50 Years On. **2012**, No. NOVEMBER, 1042–1047.

(3) Onuchic, J. N.; Wolynes, P. G. Theory of Protein Folding. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 70–75.

(4) Dill, K. A.; Ozkan, B. S.; Schell, S. M.; Thomas, W. R. The Protein Folding Problem. *Annu. Rev. Biophys.* **2008**, *37*, 289–316.

(5) Gething MJ; J, S. Protein Folding in the Cell. *Nature* **1992**, *355*, 33–45.

(6) Ulrich Hartl, F. Molecular Chaperones in Cellular Protein Folding. *Nature* **2010**, *3* (9), 571–580.

(7) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29* (31), 7133–7155.

(8) Rich, A. and Crick, F. H. C. The Structure of Collagen. *Nature* **1955**, *176*, 915–916.

(9) Ramachandran, G. N.; Kartha, G. Structure of Collagen. *Nature* **1955**, *176*, 593.

(10) Cowan, P. M.; Mcgavin, S.; North, A. C. T. The Polypeptide Chain Configuration of Collagen. *Nature* **1955**, *176* (4492), 1062–1064.

(11) Bella, J. Collagen Structure: New Tricks from a Very Old Dog. *Biochem. J.* **2016**, *473* (8), 1001–1025.

(12) Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78* (1), 929–958.

(13) Ramshaw, J. A. M.; Shah, N. K.; Brodsky, B. Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host-Guest Triple-Helical Peptides. *J. Struct. Biol.* **1998**, *122* (1–2), 86–91.

(14) Okuyama, K.; Takayanagi, M.; Ashida, T.; Kakudo, M. A New Structural Model for Collagen. *Polym. J.* **1977**, *3*, 341–343.

(15) Okuyama, K.; Miyama, K.; Mizuno, K.; Bächinger, H. P. Crystal Structure of (Gly-Pro-Hyp)9: Implications for the Collagen Molecular Model. *Biopolymers* **2012**, *97* (8), 607–616.

(16) Okuyama, K. Revisiting the Molecular Structure of Collagen. *Connect. Tissue Res.* **2008**, *49* (5), 299–310.

(17) Bella, J. A New Method for Describing the Helical Conformation of Collagen: Dependence of the Triple Helical Twist on Amino Acid Sequence. *J. Struct. Biol.* **2010**, *170* (2), 377–391.

(18) Luo, T.; Kiick, K. L. Collagen-like Peptides and Peptide-Polymer Conjugates in the Design of Assembled Materials. *Eur. Polym. J.* **2013**, *49*, 2998–3009.

(19) Fields, G. B. Synthesis and Biological Applications of Collagen-Model Triple-Helical Peptides. *Org. Biomol. Chem.* **2010**, *8* (6), 1237.

(20) Fallas, J. A.; O'Leary, L. E. R.; Hartgerink, J. D. Synthetic Collagen Mimics: Self-Assembly of Homotrimers, Heterotrimers and Higher Order Structures. *Chem. Soc. Rev.* **2010**, *39* (9), 3510–3527.

(21) Rele, S.; Song, Y.; Apkarian, R. P.; Qu, Z.; Conticello, V. P.; Chaikof, E. L. D-Periodic Collagen-Mimetic Microfibers. *J. Am. Chem. Soc.* **2007**, *129* (47), 14780–14787.

(22) Kusebauch, U.; Cadamuro, S. A.; Musiol, H. J.; Lenz, M. O.; Wachtveitl, J.; Moroder,

L.; Renner, C. Photocontrolled Folding and Unfolding of a Collagen Triple Helix. *Angew. Chemie - Int. Ed.* **2006**, *45* (42), 7015–7018.

(23) Wang, A. Y.; Mo, X.; Chen, C. S.; Yu, S. M. Facile Modification of Collagen Directed by Collagen Mimetic Peptides. *J. Am. Chem. Soc.* **2005**, *127* (12), 4130–4131.

(24) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. Solution Structure of an ABC Collagen Heterotrimer Reveals a Single-Register Helix Stabilized by Electrostatic Interactions. *J. Biol. Chem.* **2009**, *284* (39), 26851–26859.

(25) Newberry, R. W.; VanVeller, B.; Raines, R. T. Thioamides in the Collagen Triple Helix. *Chem. Commun. (Camb).* **2015**, *51* (47), 9624–9627.

(26) Egli, J.; Siebler, C.; Köhler, M.; Zenobi, R.; Wennemers, H. Hydrophobic Moieties Bestow Fast-Folding and Hyperstability on Collagen Triple Helices. *J. Am. Chem. Soc.* **2019**, jacs.8b13871.

(27) Cutini, M.; Bocus, M.; Ugliengo, P. Decoding Collagen Triple Helix Stability by Means of Hybrid DFT Simulations. *J. Phys. Chem. B*.

(28) Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large

Systems. *J. Comput. Chem.* **2013**, *34* (19), 1672–1685.

(29) Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalleri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S.; et al. Quantum-Mechanical Condensed Matter Simulations with CRYSTAL. *WIREs Comput Mol Sci.* 2018, pp 1–36.

(30) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100.

(31) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.

(32) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789.

(33) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868.

(34) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets

of Triple Zeta Valence Quality for Atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829.

(35) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamcis for Liquid Metals. *Phys. Rev. B* **1993**, *47* (1), 558.

(36) Kresse, G.; Furthmüller, J.; Hafner, J. Ab Initio Molecular-Dynamics Simulation of the Liquid-Metal–Amorphous-Semiconductor Transition in Germanium. *Phys. Rev. B* **1996**, *6* (1), 558–561.

(37) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B - Condens. Matter Mater. Phys.* **1996**, *54* (16), 11169–11186.

(38) Kresse, G.; Furthmiiller, J. Water News Roundup. *J. / Am. Water Work. Assoc.* **2004**, *96* (10), 14–20.