Multidiszciplináris Tudományok, 14. kötet, 1. sz. (2024) pp. 71–82. **https://doi.org/10.35925/j.multi.2024.1.7**

MOLECULAR DYNAMICS SIMULATIONS OF THE PROLINE AND HYDROXYPROLINE OF COLLAGEN

Nesreen Alkanakri
PhD student, Institute of Chemistry, University of Miskolc
3515 Miskolc-Egyetemváros, e-mail: nesreen.alkanakri@student.uni-miskolc.hu

Michael C. Owen
Higher Education and Industrial Cooperation Centre, University of Miskolc,
Institute of Chemistry, University of Miskolc
3515 Miskolc-Egyetemváros, e-mail: michael.christopher.owen@uni-miskolc.hu

Abstract

Collagen is an important natural, bioactive, and abundant material in living systems. Developing collagen materials that resolve practical issues in tissue engineering is the focus of significant research. However, its properties and behavior are not sufficiently understood, not in the least because proteins undergo significant conformational changes while performing their function. Moreover, it is difficult to determine the solvent impact on the structure and interatomic bonding at the atomistic level. Molecular Dynamics (MD) simulation is a technique that can be used successfully to understand macromolecular structure-to-function relationships. This work investigates the influence of hydroxyproline and proline on hexamer and heptamer collagen structures using the GROMACS software. We applied the Amber99sb force field to conduct molecular dynamics simulations in triplicate of the collagen fragments over a trajectory of 200 ns. We studied the root mean square (RMS) distribution, hydrogen bonds, and solvent accessible surface area (SASA). The results showed proline and hydroxyproline helped to stabilize the 3-helix of collagen; hydroxyproline did so more extensively than proline did. Hydroxyproline is responsible for the formation of intermolecular hydrogen bonds. It increases the stability of the triple helical, while proline promotes the formation of the intramolecular hydrogen bonds and makes the overall structure less stable than hydroxyproline. The solvent-accessible surface area (SASA) indicates that collagen is a lipophilic polymer.

Keywords: *Collagen, tissue engineering, Molecular Dynamics Simulation (MD), hexamer, heptamer*

1. Introduction

The word collagen is derived from the Greek words “kola” and “gen”, which means “gum” and “producing”, respectively. (Silvipriya et al., 2015) It is the most abundant protein in the animal kingdom, but it isn't present in unicellular organisms and plants. Collagen makes up 25% of the protein content of the whole body, especially in mammals. It is found in bones, cartilage, and dentin of teeth. (Sangeetha et al., 2020) There are different types of collagens, and about 29 types of collagens have been identified. (Wang, 2021) Type I is the most prevalent collagen type, representing 90% of the total collagen. (Nurubhasha et al., 2019) The collagen molecule is composed of three α chains; in the main structure of an α-chain can a repeating Gly-Xxx-Yyy triplet be found; where Xxx and Yyy can be any amino acid, but generally (Xxx) are often proline and (Yyy) is hydroxyproline. (Johansson, 2013) A high abundance

of glycine and proline induces the formation of left-handed α-helices in the α- chains. When three left-handed α-helices come together, they form a 300 nm long righthanded superhelix called collagen monomer or tropocollagen. (Streeter and De Leeuw, 2011) The collagen fibrils comprise microfibrils, clusters of tropocollagens. Many fibrils together form larger fibers *(Figure 1)*. Except for the helical region, the triple-helical domain has two nonhelical ends known as telopeptides. Telopeptides include about 20 amino acid residues and don't contain the repeating -(Gly-Xxx-Yyy)-motif. (Hulmes, 2008)

The distribution of hydrogen bonds is the distinctive feature of the collagen triple helix, and hydrogen bonds are responsible for the 3_{10}-helix's stability. (Silvipriya et al., 2015) There are different types of hydrogen bonds. (Brodsky, 1999) One of them is direct hydrogen bonds between carbonyl and amino groups of neighboring chains, carbonyl of the Xaa residue with a glycine N–H from a parallel chain (N–H (Gly) ••• O=C Xxx), (Shoulders and Raines, 2009) or between a hydroxyproline hydroxyl group in one tropocollagen and a glycine carbonyl group in its neighbor (O-H (Hyp in Yyy) ••• O=C). (Zhang et al., 2020) The second type contains a bridging water molecule, which links two adjacent tropocollagens by forming hydrogen bonds to each of them. (Rýglová et al., 2017)

Collagen has been prepared in various forms, such as films, powders, sponges, gels, and fibers. It has been widely used in several forms and applications in the biomedical industry. (Antonio et al., 2021) Collagen is used in drug delivery, skin replacement, bone regeneration scaffolds, ophthalmology, and wound healing. (Xie et al., 2018)

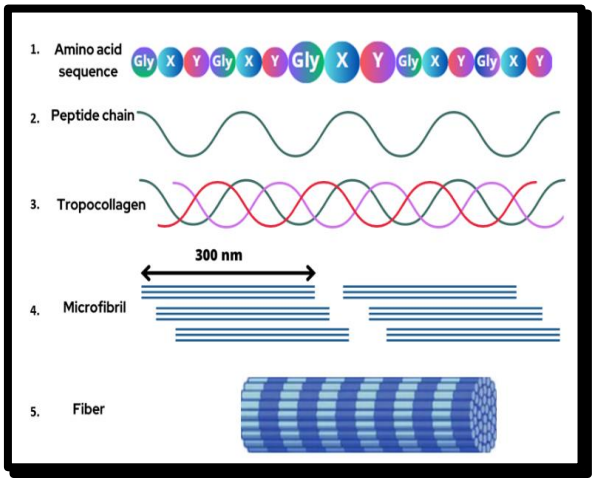

Figure 1. *The hierarchical structural levels of collagen are shown. 1. The primary structure with repeating Gly-Xxx-Yyy residues. 2. A left-handed helix is formed in the secondary structure. 3. Three helices form together a 300 nm long super-helix, called tropocollagen. 4. Several tropocollagens build up microfibrils. 5. Many fibrils together form larger fibers.*

Because collagen is a molecule with a complex structure, scientists have tried throughout the 20th century to use methods that help understand its molecular structure. In recent years, one of the most common methods for understanding molecular structure and interactions is known as molecular dynamics (MD) simulations; MD simulations mimic the changes in the structures of biological molecules over a given period of time, giving us atomistic insight into the structural changes. (Hospital et al., 2015) Bodian et al. investigated the structure and kinetics of the native collagen; they ran a 10 ns molecular dynamics simulation of the heterotrimeric, triple helical domain of human type I collagen. The simulated structures show heterogeneity in the triple helical domain, which is consistent with the

results of experiments but at a higher resolution. (Bodian et al., n.d.) Streeter et al. studied the interprotein interactions that are present within a collagen fibril. The interactions studied include direct interprotein hydrogen bonds, water-mediated interprotein hydrogen bonds, and lipophilic interactions. The simulations are used to calculate the number of interprotein interactions, to determine which functional groups influence the interactions most, and to observe the spatial distribution of interprotein interactions throughout the fibrils. (Streeter & De Leeuw, 2011) Leo et al. studied how the Type I collagen fragments from rat tail sequence self-assemble. The results indicate that collagen fibrillogenesis is driven by the loss of water molecules from monomer surfaces. (Leo et al., 2019) Venkatram et al. used the partition coefficient log P to quantitatively measure lipophilicity, which is an important factor when considering small molecules and single chemical moieties (like polymer end groups). However, studies have found that this characteristic loses its predictive power when applied to larger polymer systems. (Venkatram et al., 2019) In this research, we investigated the influence of hydroxyproline and proline on hexamer and heptamer collagen structures by using the GROMACS software. We studied the hydroxyproline hexamer structure and compared it with the proline hexamer and hydroxyproline heptamer.

2. Methods

2.1. Software and molecular mode

Molecular dynamics simulations were performed with the GROMACS 2020 program package. (Van Der Spoel, 2005) We studied and compared the hydroxyproline hexamer structure with two structures *(Figure 2)*. In the first structure, we converted the hydroxyproline residues to proline residues to form the proline hexamer, and in the second, we added a triple helix and got hydroxyproline heptamer.

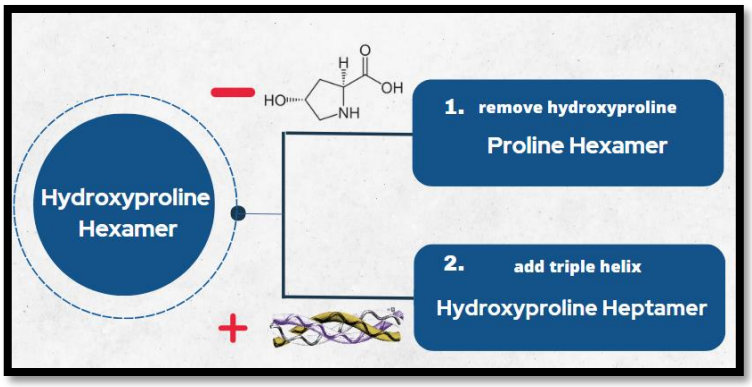

Figure 2. *The systems were studied using GROMACS software*

The force field describes all energetic parameters for intermolecular and intramolecular interactions. We applied the Amber99sb force field, which has been used extensively in the study of collagen and other polymers, and the results have been reasonable compared to experiments when such comparisons are possible. AMBER force field parameters were developed particularly for use with proteins. (Wang, 2000) We are using the TIP3P water model, which is a generic equilibrated 3-point solvent model. (Price, 2004)

2.2. Molecular dynamics simulation protocol

MD Simulations are performed in four primary steps. *(Figure 3)* showed the general MD simulation flowchart used for collagen simulation.

Model Selection: we generate the PDB. A PDB file (Protein Data Bank file) is a typical format for describing atoms' three-dimensional coordinates in a molecular structure. The PDB file is commonly used in computational chemistry and molecular dynamics simulations to input initial atomic coordinates for a simulation; it contains atomic coordinates and atom types. topology file for the three systems centered the protein in the cubic box and placed it at least 1.0 nm from the box edge. We fill it with solvent (water).

- Energy Minimization: the structure is relaxed using the steepest descent algorithm. We must check that the system has no steric clashes or inappropriate geometry. The algorithm stopped when the maximum force on an atom was less than 1000 kJ/mol/nm.
- Equilibration: equilibration is often carried out in two steps. The first phase is conducted under an NVT ensemble (constant Number of particles, Volume, and Temperature). The second is conducted under an NPT ensemble (pressure and the system's density are constant).
- Production Run and Analysis: in order to obtain the output trajectories, a production run is carried out for the appropriate amount of time.

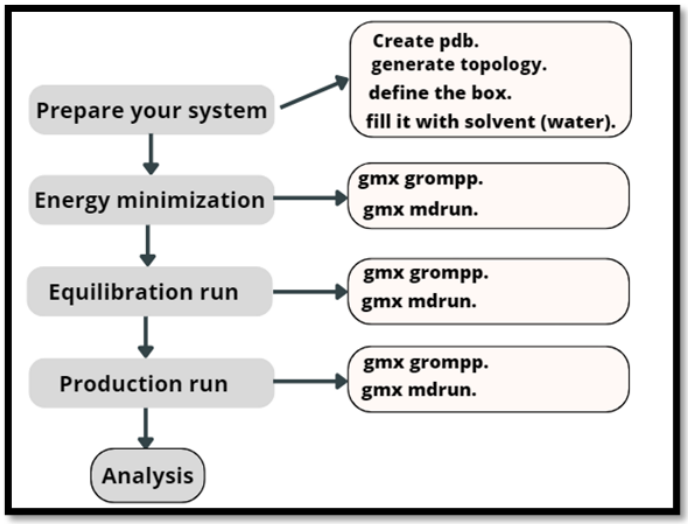

Figure 3. *The general MD simulation flowchart is used for collagen stimulation*

Each system was equilibrated for 100 ns, and Because of the large number of conformations that MD simulations produce, which makes analysis is difficult in practice. Therefore, clustering algorithms have been applied to MD results in order to divide protein ensembles into groups of structures with comparable physicochemical and structural characteristics. This approach is useful because it makes it possible to characterize a conformational ensemble created by MD and directs the analysis to concentrate on the most important alterations. So, we chose the three most prevalent structures, made the second equilibration, and ran for 200 ns.

Table 1. *Details of the simulated systems*

System	Protein atoms	Protein chains	Total atoms	water mol.	T (k)	P (bar)	box (nm^3)
Hyp_Hexamer	6390	18	171510	55040	300	1.0	1749.67
Pro_Hexamer	6228	18	173250	55674	300	1.0	1773.58
Hyp_Heptamer	7455	21	171510	54685	300	1.0	1749.67
Pro_Heptamer	7266	21	173241	55325	300	1.0	1773.58

3. Results and discussion

We presented the average values of the three runs for the three systems. To begin with, we used VMD to render each structure. For example, we have included the structure of hydroxyproline hexamer before and after simulations. In *Figure 4.a*, it can be observed that each helix was aligned along the Z-axis, and all helices were parallel to each other. However, after stimulation in *Figure 4.b*, the effect of the solvent on the structure is evident, as the helices have changed their direction.

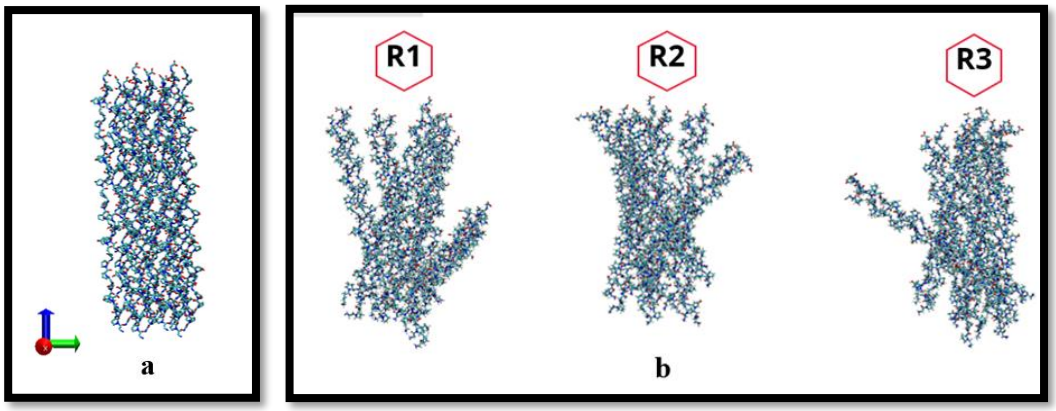

Figure 4. *The structure of hydroxyproline hexamer: a. before simulation, b. after simulation*

3.1. RMS distribution

The RMS (Root Mean Square) distribution in GROMACS generally pertains to the dispersion of root mean square deviations (RMSD) or root mean square fluctuations (RMSF) in the atomic positions observed throughout a molecular dynamics (MD) simulation trajectory. The RMS distribution of the collagen backbone was calculated to quantitatively measure the change in collagen structure during the simulation and to evaluate the stability of the native state of the system. The RMS distribution values of a protein can be compared to determine changes in protein molecular dynamics. *(Figure 5)* the root mean squared (RMS) distribution for the three systems is shown.

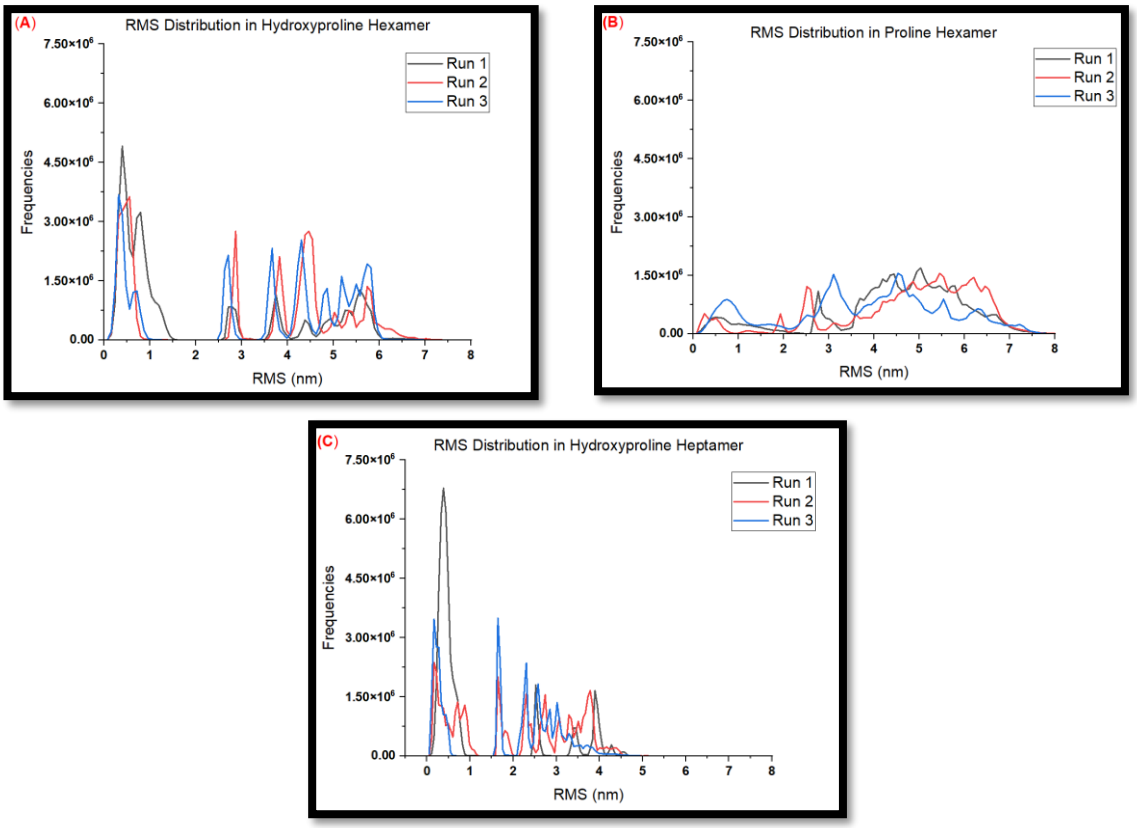

Figure 5. *RMS Distribution; (A) Hydroxyproline Hexamer, (B) Proline Hexamer, (C) Hydroxyproline Heptamer*

The low deviations mean that the structure is more stable. The results indicate that the proline hexamer is less stable than hydroxyproline hexamer and hydroxyproline heptamer, and the deviation of the helices within it is larger than in the other systems. Also, hydroxyproline heptamer is more stable than hydroxyproline hexamer, which means the system becomes more stable with an increased number of tropocollagen in the structure.

3.2. Hydrogen bonds

Hydrogen bonds are a non-covalent interaction between a hydrogen atom and a more electronegative atom, typically oxygen, nitrogen, or fluorine. These bonds play an important role in the structure and properties of many biological molecules, such as proteins, DNA, and water molecules. Hydrogen bonds play an essential role in maintaining protein stability. Also, the hydrogen bond is important in providing a stable foundation for biological systems. Therefore, it is possible to compute the hydrogen bonds in the protein structure using the MD simulation trajectories. (Ghahremanian et al., 2022) We studied intramolecular hydrogen bonds and intermolecular hydrogen bonds for three systems. We took the average of the values in each system.

3.2.1. Hydrogen bonds within triple helices

The hydrogen bonds within triple helices are shown in *Figure 6*, and the average of these values is shown in *Table 2*.

Table 2. *The average number of hydrogen bonds within triple helices in three systems*

Hydrogen Bonds		
System Name	**Average**	**Standard Error**
Hydroxyproline Hexamer	23.77	0.06
Proline Hexamer	24.21	0.05
Hydroxyproline Heptamer	23.53	0.06

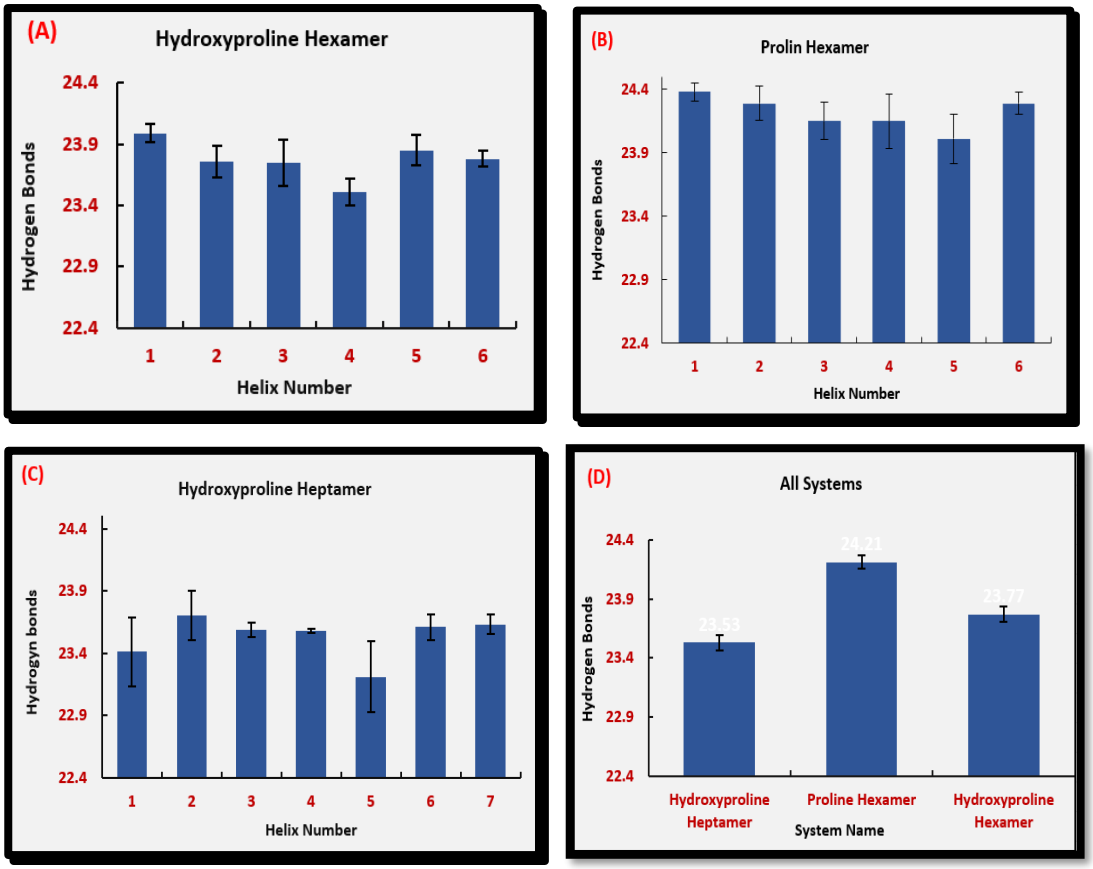

Figure 5. *Average intramolecular hydrogen bonds in (A) Hydroxyproline Hexamer (B) Proline Hexamer, (C) Hydroxyproline Heptamer, (D) the average of all systems*

The average number of hydrogen bonds in proline hexamer was higher than in hydroxyproline hexamer or heptamer, meaning the interactions within triple helices seem strongest in proline hexamer. The average number of hydrogen bonds in hydroxyproline hexamer is (23.53) almost similar to that in

hydroxyproline heptamer (23.77). However, hydroxyproline forms intramolecular hydrogen bonds through the hydroxyl group with a glycine carbonyl group. However, the number of intramolecular hydrogen bonds through proline was more than those in the case of hydroxyproline.

3.2.2. Hydrogen bonds between triple helices

The hydrogen bonds between triple helices are shown in *Figure 7*, and the average of these values is shown in *Table 3*.

Table 3. *The average number of hydrogen bonds between triple helices in three systems*

Hydrogen Bonds		
System Name	**Average**	**Standard Error**
Hydroxyproline Hexamer	3.4	0.41
Proline Hexamer	1.49	0.24
Hydroxyproline Heptamer	4.15	0.38

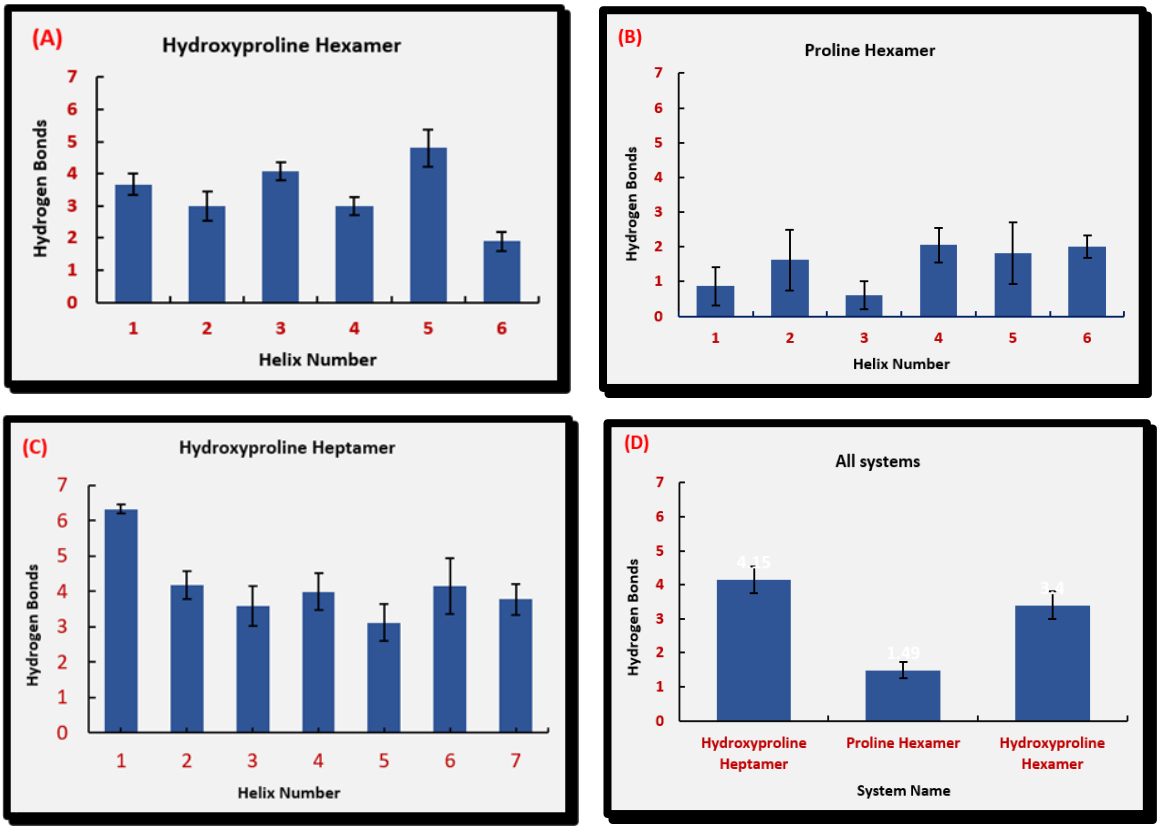

Figure 7. *Average intermolecular hydrogen bond in (A) Hydroxyproline Hexamer (B) Proline Hexamer (C) Hydroxyproline Heptamer (D) the average of all systems*

The average number of hydrogen bonds in hydroxyproline hexamer and heptamer was more than in proline hexamer, which had the lowest number of intermolecular hydrogen bonds (1.49); that means the interactions between triple helices were much weaker compared with those in the hydroxyproline systems. The average number of hydrogen bonds in hydroxyproline heptamer was (4.15) more than in hydroxyproline hexamer, which was (3.4). So we noticed adding hydroxyproline caused an increase in the number of intermolecular hydrogen bonds, and the presence of hydroxyproline leads to making the interactions between triple helices stronger and the structure more stable.

3.3. Solvent accessible surface area (SASA)

Protein solvent accessible surface area (SASA) has been regarded as an essential element in protein folding and stability. It is defined as a surface area of protein interacting with its solvent molecules. Protein atoms with charges between –0.2 and 0.2 are considered lipophilic, whereas those with charges greater than 0.2 and less than –0.2 are considered hydrophilic. The solvent-accessible surface area of each group of atoms is computed, along with the total surface area as shown in *(Figure 8)*, and its average value for helices is shown in *Table 4*, and for total in *Table 5*.

Table 4. *The average value of Solvent Accessible Surface Area (SASA) for Helices*

Solvent Accessible Surface Area (SASA) (nm^2)			
System Name	**Helices**	lipophilic	**Hydrophilic**
Hyp Hexamer	60.03	41.26	18.76
Pro Hexamer	58.54	49.28	9.25
Hyp Heptamer	59.99	41.22	18.78

Table 5. *The average value of Solvent Accessible Surface Area (SASA) for total protein*

Solvent Accessible Surface Area (SASA) (nm^2)			
System Name	**Total**	lipophilic	**Hydrophilic**
Hyp Hexamer	275.13	187.90	87.22
Pro Hexamer	283.18	237.73	45.45
Hyp Heptamer	296.09	201.98	94.11

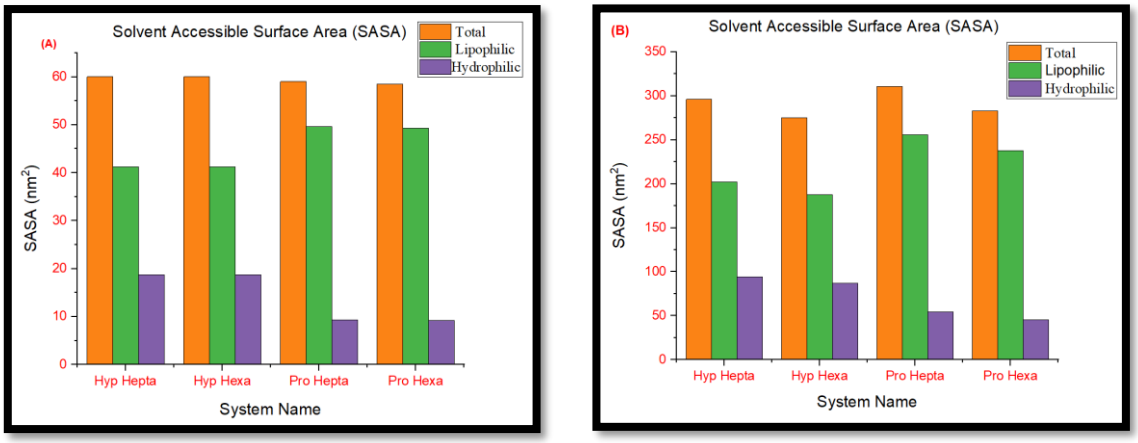

Figure 8. *The solvent Accessible Surface Area (SASA): (A) Helices, (B) Total*

The results showed the average value of lipophilic helices was in proline hexamer (49.28 **nm²**) higher than in the hydroxyproline hexamer and heptamer, which were (41.26 **nm²**), and (41.22 **nm²**) respectively. Furthermore, the average value of hydrophilic helices was lower in proline hexamer (9.25) than in hydroxyproline hexamer and heptamer, which were (18.76 nm\S2\N), (18.78 nm\S2\N), respectively. The same thing happened to the average value of lipophilic for total protein, which was also the highest in proline hexamer (237.73 nm\S2\N) and the lowest for the average value of hydrophilic (45.45 **nm²**). In general, hydrophobic interactions took the bigger surface area from the total SASA in the three systems, this means collagen is a lipophilic polymer. Particularly, it has been demonstrated that the lipophilicity of polymers controls cell adhesion, migration, and survival within biological systems. The vast majority of the research points to the importance of achieving the ideal balance of lipophilicity and hydrophilicity in biomaterial polymers for achieving the best functionality in biological systems, but it is often a fine balance to increase advantageous lipophilic interactions and reduce detrimental ones. (Pearce and O'Reilly, 2021)

4. Summary

Collagen, an abundant extracellular matrix protein, has been used numerous times in pharmaceuticals, medicine, food, and cosmetics. Increased knowledge of collagen structure and properties in the last decades has helped develop more collagen-based products and tissue engineering biomaterials. A systematic understanding of collagen's structure can promote an understanding of collagen's biological functions. In this research, we studied the influence of hydroxyproline and proline on hexamer and heptamer collagen structures using the GROMACS software. Each proline and hydroxyproline residue helps to stabilize the 310-helix of collagen; hydroxyproline is more responsible for the formation of intermolecular hydrogen bonds and increases the stability of the triple helical. While proline promotes the formation of intramolecular hydrogen bonds. So hydroxyproline makes the structure more effective than proline, which is confirmed by RMS distribution; we found the RMS distribution for proline hexamer less stable than hydroxyproline hexamer or heptamer, and the tropocollagen in it was destroyed. SASA results indicate that collagen is a lipophilic polymer. The largest lipophilic SASA was in proline hexamer, meaning the polymer's aggregate tendency would be stronger in it than in other systems.

Studying the effect of both proline and hydroxyproline on collagen structure is very important because both have a significant role in stabilizing the structure.

5. Acknowledgments

The research reported here was carried out as part of the EFOP-3.6.1-16-2016-00011 "Younger and Renewing University – Innovative Knowledge City – Institutional development of the University of Miskolc aiming at intelligent specialization" project implemented in the framework of the Szechenyi 2020 program. The realization of this project is supported by the European Union, co-financed by the European Social Fund.

References

[1] Silvipriya, K. S., Krishna Kumar, K., Bhat, A. R., Dinesh Kumar, B., John, A., Lakshmanan, P. (2015). Collagen: Animal sources and biomedical application. *Journal of Applied Pharmaceutical Science*, 5 (3), pp. 123–127, **https://doi.org/10.7324/JAPS.2015.50322**

[2] Sangeetha, K., Jisha Kumari, A. V., Radha, E., Sudha, P. N. (2020). Pharmaceutical applications of collagen. In A. Kumar Nayak, Md S. Hasnai, D. Pal (eds.). *Natural Polymers for Pharmaceutical Applications* (pp. 61–92). Apple Academic Press. **https://doi.org/10.1201/9780429328350-3**

[3] Wang, H. (2021). A review of the effects of collagen treatment in clinical studies. *Polymers*, 13 (22), p. 3868, MDPI, **https://doi.org/10.3390/polym13223868**

[4] Nurubhasha, R., Sampath Kumar, N. S., Thirumalasetti, S. K., Simhachalam, G., Dirisala, V. R. (2019). Extraction and characterization of collagen from the skin of Pterygoplichthys pardalis and its potential application in food industries. *Food Science and Biotechnology*, 28 (6), pp. 1811–1817, **https://doi.org/10.1007/s10068-019-00601-z**

[5] Johansson, P. (2013). *Characterization of Protein Surface Interaction: Collagen and Osteocalcin*. Master's Thesis, Linköping University. **https://doi.org/10.13140/RG.2.1.4080.3688**

[6] Streeter, I., De Leeuw, N. H. (2011). A molecular dynamics study of the interprotein interactions in collagen fibrils. *Soft Matter*, 7 (7), pp. 3373–3382. **https://doi.org/10.1039/c0sm01192d**

[7] Hulmes, D. J. S. (n.d.). Collagen Diversity, Synthesis and Assembly.

[8] Brodsky, B. (1999). Hydrogen bonding in the triple-helix. *Proceedings of the Indian Academy of Sciences (Chem. Sci.)*, 111, pp. 13–18, **https://doi.org/10.1007/BF02869892**

[9] Shoulders, M. D., Raines, R. T. (2009). Collagen structure and stability. *Annual Review of Biochemistry*, 78, pp. 929–958, **https://doi.org/10.1146/annurev.biochem.77.032207.120833**

[10] Zhang, X., Xu, S., Shen, L., Li, G. (2020). Factors affecting thermal stability of collagen from the aspects of extraction, processing and modification. *Journal of Leather Science and Engineering*, 2 (1), Springer, **https://doi.org/10.1186/s42825-020-00033-0**

[11] Rýglová, Š., Braun, M., Suchý, T. (2017). Collagen and its modifications - Crucial aspects with concern to its processing and analysis. *Macromolecular Materials and Engineering*, 302 (6), Wiley-VCH Verlag, **https://doi.org/10.1002/mame.201600460**

[12] Antonio, J. D. S., Jacenko, O., Fertala, A., Orgel, J. P. R. O. (2021). Collagen structure-function mapping informs applications for regenerative medicine. *Bioengineering*, 8 (1), pp. 1–23, MDPI AG, **https://doi.org/10.3390/bioengineering8010003**

[13] Xie, H., Chen, X., Shen, X., He, Y., Chen, W., Luo, Q., Ge, W., Yuan, W., Tang, X., Hou, D., Jiang, D., Wang, Q., Liu, Y., Liu, Q., Li, K. (2018). Preparation of chitosan-collagen-alginate composite dressing and its promoting effects on wound healing. *International Journal of Biological Macromolecules*, 107 (PartA), pp. 93–104. **https://doi.org/10.1016/j.ijbiomac.2017.08.142**

[14] Hospital, A., Goñi, J. R., Orozco, M., Gelpí, J. L. (2015). Molecular dynamics simulations: Advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, 8 (1), 37–47, Dove Medical Press Ltd., **https://doi.org/10.2147/AABC.S70333**

[15] Bodian, D. L., Radmer, R. J., Holbert, S., & Klein, T. E.: *Molecular dynamics simulations of the full triple helical region of collagen type I provide an atomic scale view of the protein's regional heterogeneity*, 2011 Pacific Symposium on Biocomputing, pp. 193–204. **https://doi.org/10.1142/9789814335058_0021**

[16] Leo, L., Bridelli, M. G., Polverini, E. (2019). Insight on collagen self-assembly mechanisms by coupling molecular dynamics and UV spectroscopy techniques. *Biophysical Chemistry*, 253, 106224, **https://doi.org/10.1016/j.bpc.2019.106224**

[17] Venkatram, S., Kim, C., Chandrasekaran, A., Ramprasad, R. (2019). Critical assessment of the Hildebrand and Hansen solubility parameters for polymers. *Journal of Chemical Information and Modeling*, 59 (10), pp. 4188–4194, **https://doi.org/10.1021/acs.jcim.9b00656**

[18] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26 (16), pp. 1701–1718. **https://doi.org/10.1002/jcc.20291**

[19] Wang, J., Cieplak, P. and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21 (12), pp. 1049–1074. **https://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.3.CO;2-6**

[20] Price, D. J., Brooks, C. L., 3rd (2004). A modified TIP3P water potential for simulation with Ewald summation. *The Journal of Chemical Physics*, 121 (20), pp. 10096–10103. **https://doi.org/10.1063/1.1808117**

[21] Ghahremanian, S., Rashidi, M. M., Raeisi, K., & Toghraie, D. (2022). Molecular dynamics simulation approach for discovering potential inhibitors against SARS-CoV-2: A structural review. *Journal of Molecular Liquids*, 354. Elsevier B.V. **https://doi.org/10.1016/j.molliq.2022.118901**

[22] Pearce, A. K., O'Reilly, R. K. (2021). Polymers for biomedical applications: The importance of hydrophobicity in directing biological interactions and application efficacy. *Biomacromolecules*, 22 (11), pp. 4459–4469, **https://doi.org/10.1021/acs.biomac.1c00434**