**A Computational Study of Tropocollagen Proline and Hydroxyproline**

**Nesreen Alkanakri\_1**

*PhD student, Institute of Chemistry, University of Miskolc*

*Address: 3515 Miskolc, Miskolc-Egyetemváros, e-mail:* [*nesreen.alkanakri@student.uni-miskolc.hu*](mailto:nesreen.alkanakri@student.uni-miskolc.hu)

**Michael C. Owen\_2**

*Higher Education and Industrial Cooperation Centre, University of Miskolc, Institute of Chemistry, University of Miskolc*

*Address: 3515 Miskolc, Miskolc-Egyetemváros, e-mail: michael.christopher.owen@uni-miskolc.hu*

**Abstract**

Collagen is an important structural component of bones, cartilage, and teeth and the most abundant protein in mammals; however, it is also widely used as a synthetic biomaterial. Tropocollagen is the main substructural building block of collagen fibrils. [It consists of three twisted polypeptide chains in a unique triple-helix structure](https://www.biologyonline.com/dictionary/tropocollagen). While they retain unique properties in isolation, their main function is determined when several tropocollagens come together to form collagen fibers. The fibers' properties depend on the primary structure of each of the protein chains within, which are rich in proline and hydroxyproline residues, and the interactions between tropocollagen strand. These interactions are maintained through cross-linking, but hydrogen bonding and other non-covalent interactions have also been shown to play a crucial role. To understand how protein primary structure affects the interactions between tropocollagen clusters, we employed MD simulations to investigate the structure of model hexameric and heptameric tropocollagen strands that are rich in either proline or hydroxyproline residues. The results indicate that the cross-linking of tropocollagen is thought to enhance its stability not by improving the inherent stability of the triple helix, but through the properties of cross-linking and intermolecular interactions. Hexamer systems have a higher total interaction energy than heptamer systems due to more efficient packing and stronger intermolecular interactions, especially those involving hydroxyproline, while heptamers face steric hindrance.

**Keywords:** Collagen, Gromacs, tropocollagen, Hydroxyproline, Proline.

# Introduction

Collagen is an essential extracellular matrix protein that provides structural and mechanical support to tissues such as tendons, skin, and cartilage and is the most abundant protein in humans and other mammals.1,2 In addition to its natural function within organisms, collagen has many industrial applications. It is often used as a food supplement to support bone and joint health, as a cosmetic supplement to promote healthy skin and a youthful appearance, as a medical device in tissue regeneration, and also in drug delivery systems.3,4 Collagen is a highly-organized coiled aggregate of singular tropocollagen strands, each of which is a triple helix comprised of three individual polypeptide chains.2 Each protein chain is in an extended polyproline II-like helix coiled around a common axis.5 Tropocollagen is stabilized by hydrogen bonds within each polyproline II peptide chain and between the peptide chains of each tropocollagen strand, stabilizing it and enabling it to form a structural collagen subunit.6 This intricate structure gives tissues the strength and flexibility to maintain structural integrity under different conditions and between living organisms. Understanding tropocollagen’s unique structure offers promising opportunities for advancement in the field of materials science.7

Regarding the protein strands within tropocollagen, each polyproline II triple helix consists of repeated sequences of three amino acids (Gly-Xxx-Yyy), with proline (Pro) and 4-hydroxyproline (Hyp) commonly found in the Xxx and Yyy positions.8 This sequence is widely observed in collagen and has been extensively examined due to its remarkable stability, structure, and dynamics.9 Glycine and proline are pivotal building blocks of collagen.10 Glycine, the smallest amino acid used for proteins, facilitates the close packing of the three polypeptide chains in the helix by fitting into confined spaces within each tropocollagen strand.2 Substituting glycine with larger amino acids can disrupt this tight arrangement, resulting in structural abnormalities and potential diseases such as osteogenesis imperfecta, which occurs with glycine residues are substituted for arginine residues.11 10 Proline plays a critical role in the synthesis of hydroxyproline through a process known as hydroxylation, a transformation crucial for upholding the stability of the collagen triple helix structure.12 Hydroxyproline's participation in hydrogen bonding and possession of a cyclic structure contribute to the rigidity and stability of the helix, albeit at the expense of reduced flexibility.13,14 It has been suggested that hydroxyproline íimproves the thermal stability of collagen by increasing the number of hydrogen bonds between collagen chains, which helps. This is particularly important for sustain the triple helix structure.15 Hydroxyproline also improves hydration and prevents the dysfunctional assembly of collagen.16 Further studies could further delineate how amino acid primary structure contributes to the stability of tropocollagen and bulk properties of collagen..

There are around 29 different types of collagens, and their classification is based on various factors.17These include their protein primary structure, the length of the triple-helical domain, the charge profile of the helix, interruptions within the triple helix, and the size and shape of the terminal domain.18 These structural differences are essential in defining the specific functional properties of each type of collagen. For example, fibril-forming collagens such as types I, II, III, V, and XI are elongated rod-like molecules about 300 nm long. They align parallel to each other, allowing them to provide tensile strength and structural integrity to various tissues.19,20 On the other hand, collagen types IV, VIII, and X form intricate networks, with type IV collagen molecules measuring around 400 nm in length. Types VIII and X collagens, known as "short-chain" collagens, form networks and play crucial roles in creating the structural scaffolding within tissues.21 Furthermore, some collagens do not form homotypic fibers or networks. For example, type IX collagen covers the surface of type II collagen fibers, presenting a unique structural interaction.22 More can be understood about how these different collagen types evolved.

A diagram of chemical structures

Description automatically generated

***Figure 1.*****The hierarchical levels of the collagen structure are illustrated as follows: A. The chemical structure of glycine, proline, and hydroxyproline is presented. B. The helix, which consists of an amino acid sequence Gly-Xxx-Yyy repeat, with Xxx and Yyy often being proline and hydroxyproline residues. C. Three helices combine to form a 300 nm long super-helix known as tropocollagen. D. Multiple tropocollagen molecules collectively form microfibrils.**

To understand how collagen works, one can examine how individual tropocollagenstrands behave. One key question is: What characteristics of the tropocollagen triple helix contribute to the properties of the collagen superstructures? Most research has focused on understanding the stability of tropocollagen.23,24 It's been found that hydroxyproline residues are crucial for linking collagen molecules within fibrils. They typically form strong covalent bonds involving the carbon (C) and alpha-carbon (Cα) atoms that stabilize the fibrillar structure and improve its mechanical strength.25 Gautieri *et al*. used molecular dynamics simulations to study the thermal stability of collagen triple helices. They found that water molecules and the specific amino acid sequence in glycine, proline, and hydroxyproline presence significantly impact stability.26 Water stabilizes collagen by forming hydrogen bonds between the collagen molecule and the surrounding water molecules, which increases its stability in a hydrated environment. Bailey *et al*. investigated how different types of cross-links affect the thermal stability of the triple helix through simulations. They concluded that covalent cross-linking enhances collagen stability by strengthening pre-existing hydrogen bond networks.27 Similarly, Madhavi *et al.* studied how temperature affects the movement of a hydrated small tropocollagen fragment. They revealed that the strength of the hydrogen bonds between the chains remained relatively stable, but the occupancy of hydrogen bonds slightly increased when the temperature decreased.28 Matamoros *et al*. discussed the importance of lipophilicity, or hydrophobicity, in proline and its relation to the pharmacological, toxicological, and biochemical factors of significant importance.29

Previous research underscores the vital role of amino acid composition and sequence, hydrogen bonding, and cross-links in preserving the stability and assembly of tropocollagen into collagen fibers. In order to further understand the role of collagen within living organisms and to improve our use of exogenous collagen in the aforementioned applications safely and effectively, it is of the utmost importance to understand the properties of collagen at the molecular and submolecular levels as other biomolecules are, including how it interacts with itself and other compounds and substances under various physiologically-relevant conditions. It is not clearly known as to what number of tropocollagen strands from the most stable collagen fibrils and how proline and hydroxyproline residues influence these interactions. This study will use molecular dynamics (MD) simulations to characterize the interactions between tropocollagen hexamers and heptamers, using model peptides comprised mainly of proline or hydroxyproline residues. We will explore how differences in the quantity and arrangement of collagen polypeptide chains within hexamer and heptamer tropocollagen complexes contribute to their structural diversity and biological significance in tissue biology.

# Methods

## Simulation Systems

In this study, we utilized molecular dynamics (MD) simulations to explore the structural characteristics of hexameric and heptameric tropocollagen strands, both with and without hydroxyproline residues. Each system was simulated in water, and no ions were added as the systems were electrically neutral. The simulations were carried out in a cubic box, and the number of molecules in each system is detailed in ***Table 1***.

***Table 1.* The molecular composition of the four systems (hydroxyproline- rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer) includes the number of protein atoms, protein chains, total atoms, water molecules, and simulation box volume (nm³).**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tropocollagen Systems** | **Protein Atoms** | **Protein Chains** | **Total Atoms** | **Water Molecules** | **Box Volume**  **(nm3)** |
| hydroxyproline-rich hexamer | 6390 | 18 | 171510 | 55040 | 1749.67 |
| proline-rich hexamer | 6228 | 18 | 173250 | 55674 | 1773.58 |
| hydroxyproline-rich heptamer | 7455 | 21 | 171510 | 54685 | 1749.67 |
| proline-rich heptamer | 7266 | 21 | 173241 | 55325 | 1773.58 |

## Force Field

The force field parameters for all molecules used in this study are compatible with the Amber99sb force field.30,31 The TIP3P model was utilized for water.32

## Model Systems

In our study, a model tropocollagen structures were employed either glycine-proline-hydroxyproline (Gly-Pro-Hyp) triplets in the hydroxyproline-rich hexamer and heptamer systems or glycine-proline-proline (Gly-Pro-Pro) triplets in the proline-rich hexameric and heptameric tropocollagen systems.

## Simulation Protocol

The molecular dynamics simulation employed the LINCS algorithm to constrain bond lengths, utilizing a time step of 2 fs.33 A constant pressure of 1 bar was maintained using the Parrinello-Rahman barostat to scale the box volume isotropically.34 The simulation temperature was set to 300 K and controlled with the V-rescale thermostat, independently regulating polymer and aolvent groups.35 Non-bonded atom pairs were updated every ten simulation steps (nstlist = 10), and Lennard-Jones interactions were truncated at 1.0 nm (rvdw = 1.0), with energy and pressure dispersion corrections applied to mitigate cutoff length effects and ensure compatibility with the Amber99sb force field. Periodic boundary conditions were enforced in all dimensions (xyz). Electrostatic interactions were computed using particle-mesh Ewald (PME) summation,36 employing an interpolation order of 6, a direct sum tolerance of 10-5, and a real-space cutoff of 1.0 nm. These simulations were conducted using the GROMACS 2020 software package.37

## Analysis

We used VMD to render the primary structure of hexamer and heptamer tropocollagen, as shown in ***Figure 2***, respectively.

A close-up of a blue and red molecule

Description automatically generated

***Fig. 2.* The structures of tropocollagen: a) a hexamer containing six tropocollagen strands, b) a heptamer containing seven tropocollagen strands.**

The study commenced with a preliminary 100 ns simulation. Initial configurations were obtained using the g\_cluster program, employing the grooms method. This method is the most used clustering algorithm by GROMACS for clustering trajectories based on their RMSD deviation values.38 In this approach, the protein backbone atoms of all structure pairs (extracted from the simulation frames) were calculated and grouped based on a cutoff distance of 1.0 nm. The structure with the highest number of atoms was designated as the central structure, and all structures within this cutoff distance were assigned to the same cluster. The algorithm removed this cluster from the remaining structures and repeated the clustering process to generate groups of non-overlapping clusters, each with a central structure. Following this selection, each system underwent a 200 ns production run using the configuration of the central structure from each of the three largest clusters. Consequently, we analyzed four systems: hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer, with three runs conducted for each system. Data from these three 200 ns runs were averaged and are presented as such in the results section.

## Tropocollagen Interactions

All analysis programs mentioned in this section are part of the standard release of the GROMACS 2020 program package. The *gmx mindist* tool determined the number of contacts between each triple helix and the other helices within the system. The propensity for hydrogen bonding was determined by counting the occurrences of hydrogen bond formations between atoms that donate and accept hydrogen bonds within (intramolecular) and between (intermolecular) triple helices. A hydrogen bond was recorded when the angle between the hydrogen bond donor, the hydrogen-bonded hydrogen, and the hydrogen bond acceptor was between 150° and 180°, and the distance between the donor and acceptor atoms was less than 0.35 nm**.**39The *gmx sasa* program was utilized to calculate the solvent-accessible surface area (SASA). The analysis categorized atoms into two groups: "Hydrophobic," including those in the system with charges ranging from -0.2 to 0.2, and "Hydrophilic," comprising those outside the -0.2 to 0.2 range. These along with the total solvent-accessible surface area are presented for each system. The *gmx rmsf* tool calculates each system's root mean square fluctuation (RMSF) of C-alpha positions.40 The *gmx msd* program was used to calculate each helix's mean square displacement (MSD), This provides an easy way to compute the diffusion constant using the Einstein relation. The *gmx energy* program was used to calculate the Coulombic and Lennard-Jones interaction energies between each triple helix and the remaining helices in the system.

# Results

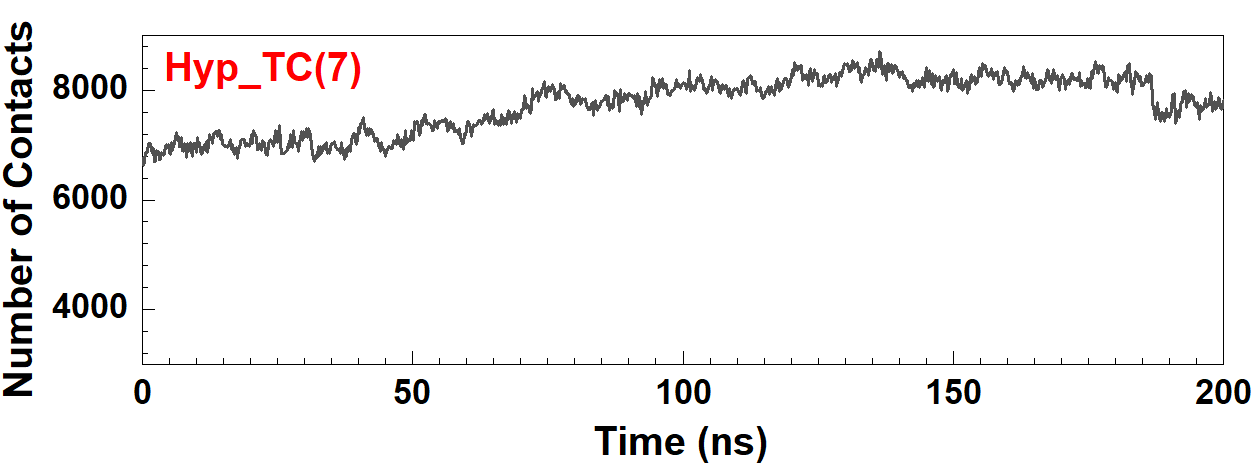
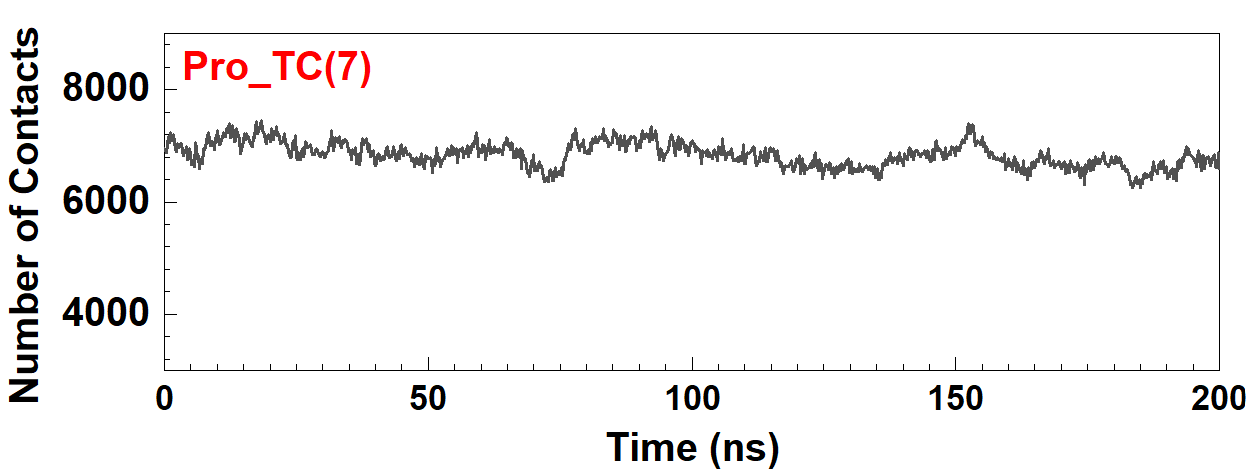
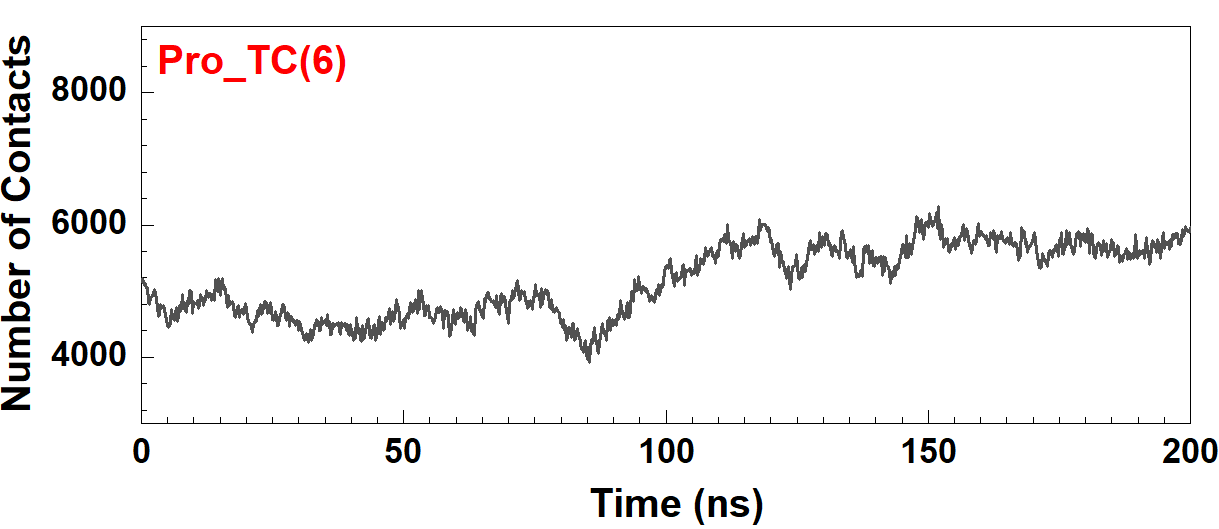
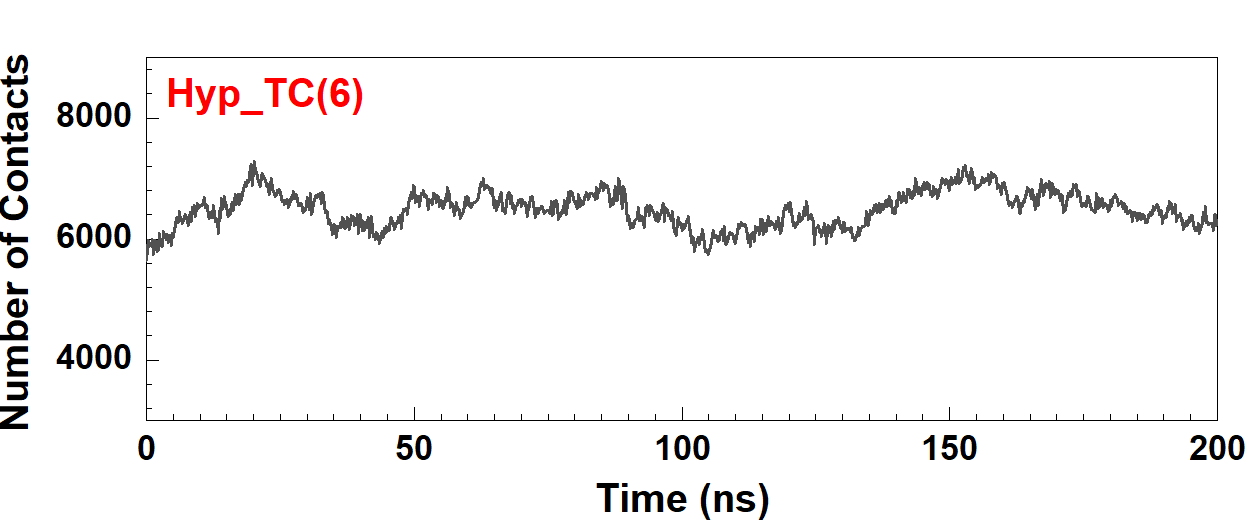
## Number of Contacts Between Tropocollagen Strands

***Table 2***, Shows the average number of contacts with the number of contacts per tropocollagen in the proline- and hydroxyproline-rich tropocollagen strands in both hexameric and heptameric forms. The hydroxyproline-rich hexamer had an average of 6491 contacts, with 1081 contacts per tropocollagen strand. Meanwhile, the hydroxyproline-rich heptamer had an average of 7743 contacts, with 1106 contacts per tropocollagen strand. The proline-rich hexamer had an average of 5156 contacts, with 859 contacts per tropocollagen strand, while the proline-rich heptamer had an average of 6841 contacts, with 977 contacts per tropocollagen strand.

***Table 2****.* **The average number of contacts with a number of contacts per tropocollagen was observed in three runs for each tropocollagen system: hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tropocollagen Systems** | **Number of Contacts** | | |
| **Avg** | **SE** | **Contacts per TC** |
| hydroxyproline-rich hexamer | 6491 | 9.18 | 1081 |
| proline-rich hexamer | 5156 | 17 | 859 |
| hydroxyproline-rich heptamer | 7743 | 16.1 | 1106 |
| proline-rich heptamer | 6841 | 6.90 | 977 |

***Figure 3***, shows that the hydroxyproline-rich systems had more contacts between the strands than the proline-rich systems. Specifically, the hexamers and heptamers of the hydroxyproline-rich systems had a maximum of 7284 and 8704 contacts, respectively. On the other hand, the proline-rich tropocollagen strands had 6279 and 7450 contacts in the hexamer and heptamer, respectively. Notably, the hexameric tropocollagen systems had fewer contacts between the tropocollagen strands than the heptameric ones. The minimum values for each system followed similar patterns to the respective maxima. The hydroxyproline-rich hexamers and heptamers had a minimum of 5645 and 6632 contacts, respectively, while the proline-rich systems had 3923 and 6261 contacts, respectively, between each of the tropocollagens within each system.



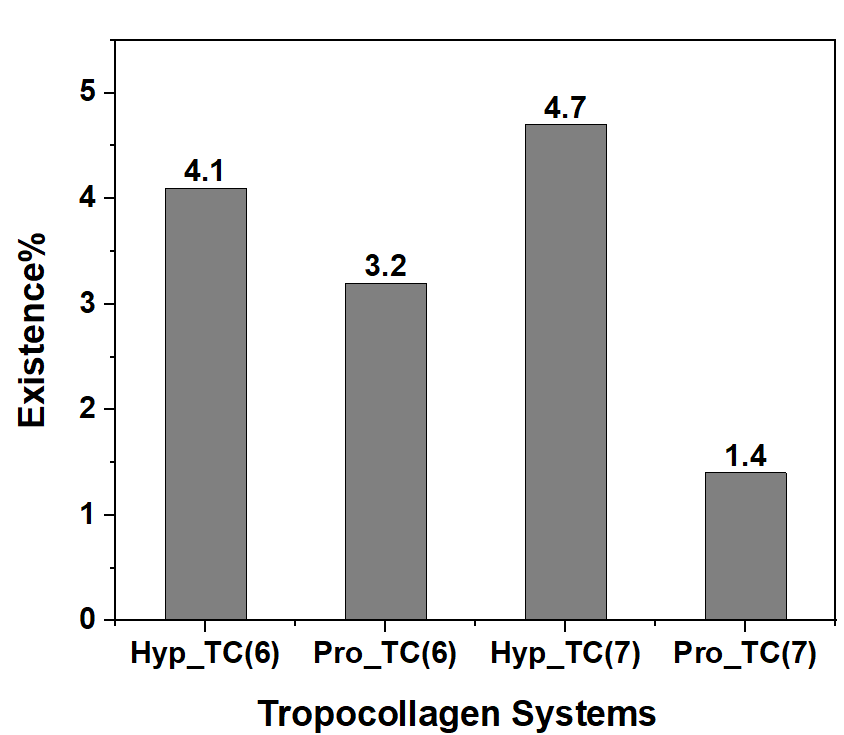
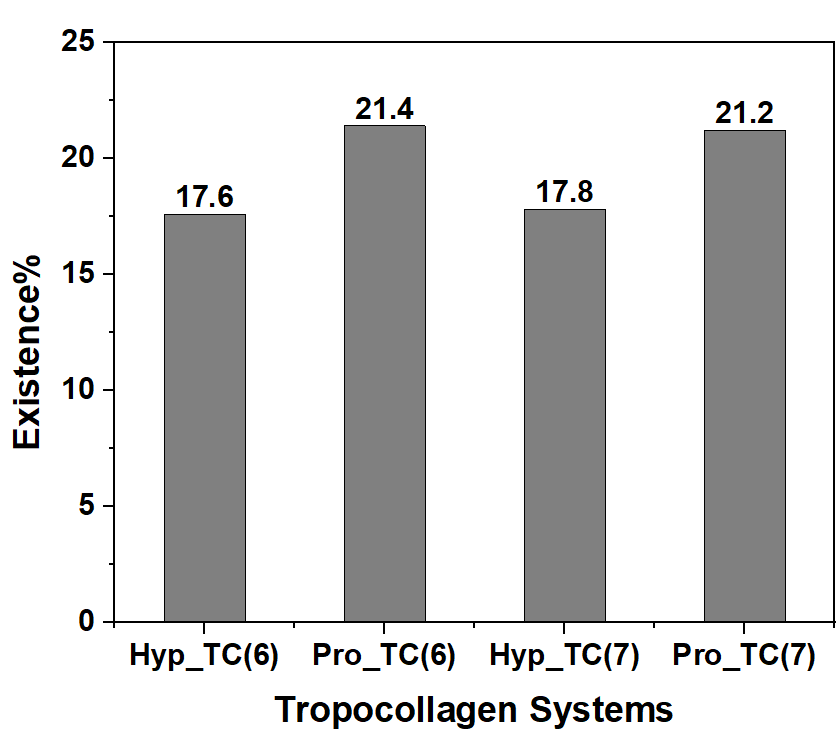
***Figure 3.*****shows the average number of contacts for three runs in the four tropocollagen systems: Hyp\_TC(6), Pro\_TC(6), Hyp\_TC(7), and Pro\_TC(7) over a 200 ns simulation period.**

## Hydrogen Bonds Both Within and Between Tropocollagen Strands

***Table 3*** provides the average number of hydrogen bonds with the number of hydrogen bonds per tropocollagen observed in three runs of the hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer. The number of hydrogen bonds within the tropocollagen strands was higher in the proline-rich systems, with 24.2 in the hexamer and 24.0 in the heptamer, and the number of hydrogen bonds per tropocollagen was 4 and 3.4, respectively***.*** Meanwhile, the average number of hydrogen bonds within hydroxyproline-rich hexamer was 23.7, with 3.9 bonds per tropocollagen strand, and the hydroxyproline-rich heptamer was 23.5, with 3.3 hydrogen bonds per tropocollagen strand. Conversely, the number of hydrogen bonds between the tropocollagen strands was higher in those containing hydroxyproline, with an average of 3.4 hydrogen bonds with 0.5 bonds per tropocollagen observed between the tropocollagens in the hydroxyproline-rich hexamer and an average of 4.2 with 0.5 bonds per tropocollagen in the hydroxyproline-rich heptamer. In the case of the proline-rich tropocollagen hexamer, an average of 1.5 hydrogen bonds formed between tropocollagens in the hexamer with 0.2 bonds per tropocollagen, and practically none (0.03) were found in the proline-rich heptamer. These results are similar to those shown in ***Figure 4***, which presents the average distribution ofhydrogen bonds within and between the tropocollagens in each system. Most hydrogen bonds within tropocollagens formed between the amide nitrogen of a glycine residue (acting as a hydrogen bond donor) and the amide oxygen of a proline residue, acting as a hydrogen bond acceptor. Additionally, hydroxyproline is essential in forming hydrogen bonds between the tropocollagens. Specifically, most of the hydrogen bonds were formed between the oxygen atom in the hydroxyl group (-OH) bonded to the γ-carbon of the pyrrolidine ring of a hydroxyproline residue, acting as a hydrogen bond donor, and the oxygen (O) of a glycine or hydroxyproline residue, acting as a hydrogen bond acceptor.

***Table 3.* The average number of bond interactions with the number of bonds per tropocollagen was observed in three runs for each tropocollagen system: hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tropocollagen Systems** | **Hydrogen Bonds** | | | | | |
| **Within Tropocollagen Strands** | | | **Between Tropocollagen Strands** | | |
| **Avg** | **SE** | **Avg/TC** | **Avg** | **SE** | **Avg/TC** |
| hydroxyproline-rich hexamer | 23.7 | 0.02 | 3.9 | 3.4 | 0.28 | 0.5 |
| proline-rich hexamer | 24.2 | 0.09 | 4 | 1.5 | 0.45 | 0.2 |
| hydroxyproline-rich heptamer | 23.5 | 0.01 | 3.3 | 4.1 | 0.38 | 0.5 |
| proline-rich heptamer | 24.0 | 0.04 | 3.4 | 0.03 | 0.01 | 0.0 |



**(A)**

**(B)**

***Figure 4.*****shows the average distribution of hydrogen bonds: (A) within, (B) between tropocollagen strands for each system's three runs: hydroxyproline-rich tropocollagen hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer.**

## Total, Hydrophobic, and Hydrophilic Solvent-Accessible Surface Area (SASA) of Tropocollagen Fragments

***Figure 5*** displays the average solvent-accessible surface area observed in three runs in each of the four systems: hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer. In the hydroxyproline-rich tropocollagen hexamer, the total SASA varied from approximately 273.1 to 279.9 nm², the hydrophobic SASA ranged from 185.9 to 187.9 nm², and the hydrophilic SASA ranged from 86.2 to 89.0 nm². The proline-rich tropocollagen hexamer had a total SASA between 272.0 and 295.1 nm², with hydrophobic SASA values between 229.6 and 246.5 nm² and hydrophilic SASA values ranging from 42.3 to 48.6 nm². The total SASA of the hydroxyproline-rich tropocollagen heptamer system, the total SASA values varied from 285.2 to 305.8 nm², with hydrophobic and hydrophilic SASA values ranging from 194.0 to 209.1 nm² and 91.2 to 96.7 nm², respectively. Finally, the proline-rich heptamer system exhibited total SASA values from 304.0 to 321.7 nm², with hydrophobic SASA ranging from 250.3 to 265.4 nm² and hydrophilic SASA ranging from 53.6 to 56.3 nm².

A graph of different types of hydrophobic

Description automatically generated

***Figure 5.* The average SASA of each system's three runs as a time function** **for Hyp\_TC(6), Pro\_TC(6), Hyp\_TC(7), and Pro\_TC(7). The black line represents the total SASA, the red line indicates the hydrophobic region and the blue line represents the hydrophilic region.**

The average SASA values in nm² for the total surface in four systems, categorized by hydrophobic and hydrophilic regions, are shown in ***Table 4***. The surface areas of the proline and hydroxyproline-rich hexamer and heptamer differed. The proline-rich hexamer and heptamer had 283.1 nm² and 310.5 nm² more surface area, respectively, compared to the hydroxyproline-rich hexamer and heptamer, which had 275.1 nm² and 296.0 nm². When comparing hydroxyproline and proline, it was discovered that the hydroxyproline-rich hexamer and heptamer have more significant areas of hydrophilic surfaces, making up 31.7% and 31.8% of their total surface area, respectively. The proline-rich hexamer and heptamer were 16% and 17.6% hydrophilic, respectively. On the other hand, the proline-rich systems show more extensive regions of hydrophobic surfaces that are accessible to the solvent, accounting for 84.1% and 82.3% of the total surface area for the hexamer and heptamer, respectively, and in hydroxyproline-rich hexamer and heptamer were 68.3% and 68.2% hydrophobic, respectively.

***Table 4****.* **The average SASA values in nm² for the total surface in four systems (hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer) are categorized by hydrophobic and hydrophilic regions. Includes percentages of hydrophobic and hydrophilic areas.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tropocollagen Systems** | **Solvent Accessible Surface Area (SASA) (nm2)** | | | | | | | |
| **Total** | | **Hydrophobic** | | | **Hydrophilic** | | |
| **Avg** | **SE** | **Avg** | **SE** | **%** | **Avg** | **SE** | **%** |
| hydroxyproline-rich hexamer | 275.1 | 2.43 | 187.9 | 1.52 | 68.3 | 87.2 | 0.91 | 31.7 |
| proline-rich hexamer | 283.1 | 6.68 | 237.7 | 4.89 | 84.1 | 45.4 | 1.81 | 16.0 |
| hydroxyproline-rich heptamer | 296.0 | 5.97 | 201.9 | 4.37 | 68.2 | 94.1 | 1.59 | 31.8 |
| proline-rich heptamer | 310.5 | 5.61 | 255.8 | 4.80 | 82.3 | 54.6 | 0.84 | 17.6 |

## Root Mean Square Fluctuation (RMSF) of α-carbons within the tropocollagens

***Figure 6*** presents the average RMSF values of α-carbons of the amino acid residues within each tropocollagen fragment the hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer. Comparing the RMSF of hydroxyproline and proline positions revealed that hydroxyproline-rich hexamer and heptamer generally fluctuate less than those of proline. For instance, the maximum RMSF of the α-carbons of hydroxyproline in the hydroxyproline-rich hexamer was 4.09 nm; the heptamer was 2.86 nm. In contrast, the maximum RMSF value in the proline-rich hexamer was 5.57 nm, and in the heptamer, it was 4.62 nm. In short, heptamers (hydroxyproline and proline) also exhibited lower RMSF values than hexamers.

A group of graphs showing different numbers

Description automatically generated with medium confidence

***Figure 6.*****shows the average root-mean-squared fluctuation (RMSF) values of c-alpha atoms for three runs in the four systems: Hyp\_TC(6), Pro\_TC(6), Hyp\_TC(7), and Pro\_TC(7).**

## Mean-Squared Displacements (MSD)

In our study, we delved into the MSD analysis to track the translation of atoms within the hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer. ***Table 5*** presents the average diffusion constants (D) for each system. We observed higher diffusion constants within the proline-rich systems for the hexamer 0.20nm2/s and the heptamer 0.29nm2/s, resulting in elevated MSD. This is further illustrated in ***Figure 7***, depicting the average MSD for each system over three runs. Specifically, the MSD values were 212.5 nm² at 200 ns for the proline-rich hexamer and 332.9 nm² for the proline-rich heptamer. Conversely, we noted lower diffusion constants within the hydroxyproline-rich systems for the hexamer 0.18 nm2/s and heptamer 0.16 nm2/s, indicative of reduced MSD. At 200 ns, the MSD values were 236.6 nm² for the hydroxyproline-rich hexamer and 209.4 nm² for the hydroxyproline-rich heptamer. Interestingly, the MSD was observed to be higher in the hydroxyproline-rich hexamer compared to the heptamer, while a lower MSD was evident in the proline-rich hexamer in contrast to the proline-rich heptamer.

***Table 5.*** **The average diffusion constant (D) for three runs in the four systems: hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer.**

|  |  |  |
| --- | --- | --- |
| **Systems** | **Diffusion Constant (D) (nm2/s)** | |
| **Avg** | **SE** |
| hydroxyproline-rich hexamer | 0.18 | 0.08 |
| proline-rich hexamer | 0.20 | 0.18 |
| hydroxyproline-rich heptamer | 0.16 | 0.10 |
| proline-rich heptamer | 0.29 | 0.21 |

A graph of different colored lines

Description automatically generated

***Figure 7.*****The average mean square displacements for three runs of the four systems (hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, proline-rich heptamer) during the 200 ns simulation period.**

## Interaction energies

We examined the interactions between contacts in three runs in each of the four systems: hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer to determine if they were driven by electrostatic or Lennard-Jones (LJ) interactions. ***Table 6*** shows the total interaction energies, including the average Coulombic and Lennard-Jones energies for each system and for each tropocollagen. The total interaction energy was higher in hexamer systems compared to heptamer systems. In the hydroxyproline-rich hexamer, the interaction energy was 115 kJ·mol⁻¹, while in the proline-rich hexamer, it was 113.5 kJ·mol⁻¹. Per the tropocollagen unit, the interaction energy was 19.2 kJ·mol⁻¹ in the hydroxyproline-rich hexamer and 18.9 kJ·mol⁻¹ in the proline-rich hexamer. For the heptamer systems, the hydroxyproline-rich heptamer exhibited an interaction energy of 111.1 kJ·mol⁻¹, with 15.8 kJ·mol⁻¹ per tropocollagen. In the proline-rich heptamer, the total interaction energy was 82.2 kJ·mol⁻¹, with 11.7 kJ·mol⁻¹ per tropocollagen.

***Table 6.* The total interaction energies, including the average of Coulombic and Lennard-Jones energies for three runs in each of the four systems: hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer and for each tropocollagen.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tropocollagen Systems** | **Interaction Energy (kJ·mol-1)** | | | | | | | |
| **Total** | | **Coulombic** | | | **LJ** | | |
| **Sum** | **Sum/TC** | **Avg** | **SE** | **Avg/TC** | **Avg** | **SE** | **Avg/TC** |
| hydroxyproline-rich hexamer | 115.0 | 19.2 | -26.9 | 0.9 | -4.4 | -88.1 | 4.5 | -14.6 |
| proline-rich hexamer | 113.5 | 18.9 | -45.6 | 11.5 | -7.6 | -67.9 | 6.2 | -11.3 |
| hydroxyproline-rich heptamer | 111.1 | 15.8 | -24.5 | 1.1 | -3.5 | -86.6 | 3.8 | -12.3 |
| proline-rich heptamer | 82.2 | 11.7 | -6.1 | 0.2 | -0.8 | -76.1 | 5.2 | -10.8 |

***Figure 8*** shows that the strongest LJ interactions were observed in hydroxyproline-rich systems. In the hydroxyproline-rich hexamer, the average LJ interactions were -88.1 kJ·mol⁻¹, with -14.6 kJ·mol⁻¹ per tropocollagen. For the hydroxyproline-rich heptamer, the average LJ interactions were -86.6 kJ·mol⁻¹, and

-12.3 kJ·mol⁻¹ per tropocollagen. In contrast, the average LJ interactions in the proline-rich hexamer were

-67.9 kJ·mol⁻¹, with -11.3 kJ·mol⁻¹ per tropocollagen. In the proline-rich heptamer, the average LJ interactions were -76.1 kJ·mol⁻¹, and -10.8 kJ·mol⁻¹ per tropocollagen. The strongest contributions to the Coulombic interaction energy were found in the hexamer systems. In the proline-rich hexamer, the average Coulombic energy was -45.6 kJ·mol⁻¹ in total and -7.6 kJ·mol⁻¹ per tropocollagen, while in the hydroxyproline-rich hexamer, the average energy was -26.9 kJ·mol⁻¹ in total, and -4.4 kJ·mol⁻¹ per tropocollagen. Compared to the heptamer systems, the average Coulombic interaction in the hydroxyproline-rich heptamer was -24.5 kJ·mol⁻¹ and -3.5 kJ·mol⁻¹ per tropocollagen. The proline-rich heptamer was -6.1 kJ·mol⁻¹ in total and -0.8 kJ·mol⁻¹ per tropocollagen.

A graph of a number of green and orange bars

Description automatically generated

***Figure 8.*****The average interaction energies, including the Coulombic and Lennard-Jones energies for three runs in each of the four systems: hydroxyproline-rich hexamer, proline-rich hexamer, hydroxyproline-rich heptamer, and proline-rich heptamer.**

# Discussion

The collagen models used in this study are denoted as [(Gly-Pro-Hyp)30]3 and [(Gly-Pro-Pro)30]3, and contain 90 amino acid residues per tropocollagen strand. This truncation was necessary due to computational limitations, as the full-length collagen molecule, which stretches 300 nm long, is too large for atomistic simulations. However, our use of shorter fragments still provides atomistic insight into the interactions between tropopollagen strands.

**4.1 More Contacts Shown in Hyp-Containing Tropocollagens Indicate Tighter Packing**

The comparison of the number of contacts in the four systems indicates that hydroxyproline contributes to a more stable molecular structure. Atomic contacts are defined as pairs of atoms positioned closer to each other than a set cutoff distance.41 The cutoff of x.x was chosen in this study to indicate a possible interaction between points within the system without trying to characterize any specific interaction type. These contacts show where the tropocollagen strands maintained more contacts with each other, which suggests that the strands have a greater number of interactions that support tighter packing. This is despite the fact that the Hyp-containing tropocollagen strands contain an extra OH functional group. The extra space provided required for the OH group induces tighter packing between the strands even without the formation of covalent cross-links. The tighter packing afforded by Hyp could be the reason why hydroxyproline enhances the structural stability of collagen more effectively than proline does.15 42 This phenomenon is observed both in the heptamer and hexamer. Moreover, the tighter packing is also reflected in the solvent-accessible surface area results, where in spite of the extra OH group, the hydrxyproline-containing tropocollagen strands have a smaller accessible surface area the the proline-rich ones do, which is also indicative of a more tightly-packed structure. The additional triple helix in the heptamer leads to more interactions and binding with surrounding molecules, resulting in higher stability. Therefore, collagen stability is influenced by residue-residue contacts, impacting its overall mechanical properties. This emphasizes the need to maintain these contacts to prevent the structural dissociation into individual tropocollagen strands.43

Inter-strand hydrogen bonds may well be the driving force for the tighter packing of Hyp-containing tropocollagen. Investigating the intra- and intermolecular hydrogen bonds in the Hyp and Pro-rich tropocollagen strands (Table 3) revealed the proline-rich tropocollagens displayed slightly more intramolecular hydrogen bonds than the hydroxyproline-rich systems did; however, the Hyp-rich tropocollagens had many more hydrogen bonds between tropocollagen strands than the proline-rich systems did. Conclusively, our analysis indicates that proline facilitates the formation of intramolecular hydrogen bonds more than hydroxyproline does. While these intramolecular bonds contribute to the stability of tropocollagen, their influence is relatively minor when compared to the contributions of glycine, with its small size and proline, with its fixed ϕ angle. Glycine, the smallest amino acid, perfectly fits the tight spaces within the triple helix structure, and its presence in every third residue is critical in compactly packing the triple helix and vital for maintaining its stability. Studies show that the hydrogen bonds between the NH groups of glycine and the CO groups of adjacent amino acids are essential for holding the triple helix structure together at a molecular level.44The fixed phi angle is crucial for properly aligning collagen chains, enabling the formation of the triple helical structure. Various studies have observed this, where alterations in proline or glycine residues led to helix destabilization.45 Adding a triple helix increases the number of these intermolecular bonds, as illustrated in ***Figure 4***. The heptamer system has a higher number than the hexamer system. The cross-linking of tropocollagen is theorized to enhance its stability, not by improving the inherent stability of the triple helix, but through the properties of cross-linking and intermolecular interactions. Various studies support this perspective, emphasizing the role of cross-linking in improving collagen's thermal and mechanical properties.46 Therefore **Hyp increases the hydrogen bonds between strands and allows tighter packing in the absence of covalent cross links**.

**4.2 Solvent-Accessible Surface Area remains constant, while Hyp doubles the Hydrophillic Surface Area**

The surface area of a polymer that is available for interacting with neighboring molecules and solvent particles is known as the solvent-accessible surface area.47 This factor is crucial in determining the ability to interact with other entities in the environment, including other polymers, proteins, carbohydrates, therapeutics and other small molecules.48 These accessible regions have been divided into the hydrophilic and hydrophobic regions, which can further predict how the molecule will partition between hydrophobic and hydrophilic media, such as between the surface of or buried within a membrane or between other hydrophobic or hydrophilic polymers or drugs. ***Figure 5*** shows that throughout 200 ns, the average SASA values remained stable across all the systems, indicating minimal fluctuations in the accessible surface areas for both the total, hydrophobic, and hydrophilic regions. This can be expected in the fixed aqueous environment, since changes to the surface area are most likely to be induced by large changes to the system or solvent. As noted in the data in ***Table 4***, comparing the structures of proline-rich heptamer and hydroxyproline-rich heptamer with proline and hydroxyproline-rich hexamer reveals. The average total solvent-accessible surface area values for the heptamers are higher than the hexamers. This increase in SASA is attributed to an additional tropocollagen strand, resulting in a larger molecular surface area exposed to solvent than that of its respective hexamer. This is important for protein interactions and stability, especially since mutations linked to diseases often occur in regions with limited solvent accessibility. This highlights the significance of SASA in protein functionality. When comparing the hydroxyproline-rich tropocollagens to the proline-rich ones, it is observed that hydroxyproline-rich systems have larger hydrophilic surface areas due to the additional hydroxyl group in hydroxyproline. This feature encourages greater interactions with water and other molecules, enabling these polymers to interact more with hydrophilic molecules and surfaces.50 Simply converting the Hyp residues to Pro reduced the hydrophilic SASA by nearly half, and the magnitude of this change was greater than expected. These results suggest that this conversion could alter the location of collagen from hydrophilic regions to hydrophobic ones, similarly alter the type of molecules that interact with collagen, and significantly alter its interactions with small molecules such as drugs and other therapeutics both in molecule type as well as kinetics and dissociation rates. Although proline-rich tropocollagen is not entirely hydrophobic, it is interesting to see the degree to which segments rich in proline and glycine contribute to its largely hydrophobic surface area.51

4.3 **Root-mean-squared fluctuation and mean-squared displacement indicates a less dynamic structure for Hyp-enriched tropocollagen**

The RMSF is a measure commonly used to indicate the displacement of a specific atom or group of atoms from a reference structure.52 The RMSF analysis offers insights into the dynamic regions of proteins and polymers, particularly flexible loops. Additionally, it can highlight the importance of specific regions in conformational changes and polymer interactions.40 The analysis revealed that the heptamer and hexamer hydroxyproline-rich systems exhibited lower RMSF than the proline-rich systems, as shown in ***Figure 6***. This finding suggests that the hydroxyproline-rich tropocollagen structures may have higher resistance to motion in an aqueous environment than the proline-rich structures. The hydroxyl group in hydroxyproline is expected to promote additional hydrogen bonding and other polar interactions with solvent, small molecules, and surfaces in its environment, ultimately contributing to increased overall rigidity and stability of the structure.4653 Furthermore, our comparison of the tropocollagen heptamers and hexamers revealed a higher RMSF in the hexameric systems than in the heptameric ones, both in the case of the hydroxyproline and proline-rich tropoocollagens.. This higher RMSF indicates greater fluctuation. One possible explanation for this difference could be that the higher number of chains in heptamers may lead to a more stable structure, as discussed in the analysis of the number of contacts, consequently reducing atomic displacement and flexibility.

The RMSF outcomes align with mean-squared displacement, a parameter commonly utilized to characterize atomic movement.54,55 A comparison between proline and hydroxyproline reveals disparities in their interactions and structural organizations.56 ***Figure 7*** shows that the systems rich in proline demonstrate a higher MSD compared to those rich in hydroxyproline. As described in Section 4.1, the hydroxyl group can form hydrogen bonds with adjacent molecules, creating a more organized and less dynamic structure. This structured environment leads to stronger intermolecular interactions and restricts the movement of hydroxyproline molecules, resulting in lower diffusion constants, as seen in ***Table 5***.The lower diffusion constants associated with hydroxyproline suggest they form more stable and less dynamic structures, contributing to collagen fibers' rigidity and resistance to motion.57 In contrast, proline lacks this hydroxyl group, which results in fewer hydrogen bonds and polar interactions and allows for less restricted movement and faster diffusion. Proline’s higher mobility might facilitate more dynamic interactions within the cellular environment, impacting processes like protein folding and intracellular transport.58 Therefore, the stability provided by hydroxyproline in collagen is essential for the mechanical properties of tissues such as skin, tendons, and bones, as it helps maintain their integrity and resilience under stress.14 Proline’s role can influence different tissues' flexibility and dynamic behavior, affecting cell signaling and movement processes.59

***4.4 Quantified energetics between Tropocollagen Strands***

***Table 5*** presents compelling evidence that the total interaction energy is significantly higher in hexamer systems than in heptamer systems. This difference can be attributed to the distinct structural characteristics of hexamers and heptamers, which greatly impact their interaction energies due to variations in packing efficiency, hydrogen bonding, and overall stability. Hexamers, comprising six units, can achieve a more compact and optimal packing arrangement compared to heptamers, which consist of seven units. This efficient packing in hexamers allows for closer intermolecular interactions, resulting in a reduced distance between units and enhanced attractive forces, ultimately leading to higher interaction energies.10 Conversely, the additional unit in heptamers can introduce steric hindrance, disrupting optimal arrangements and leading to weaker interactions. ***Figure 8*** shows that LJ interactions are more robust in hydroxyproline-rich systems compared to proline-rich systems. As previously mentioned, the strong intermolecular interactions of hydroxyproline play a significant role in LJ interactions. The hydroxyl group in hydroxyproline enhances polarity, resulting in stronger van der Waals forces and dipole-dipole interactions, which contribute to the LJ potential.60 Conversely, Coulombic interactions are reduced due to the presence of the hydroxyl group in hydroxyproline altering electron distribution, creating a region with higher electron density. Consequently, hydrogen bonding with water molecules may weaken the Coulombic interactions compared to proline. The hydroxyl group in hydroxyproline makes it more hydrophilic, leading to stronger interactions with water molecules, shielding its charges, and reducing Coulombic interactions.10 In contrast, proline is less hydrophilic, resulting in stronger Coulombic interactions as it does not attract as much water to shield its charges.

**4.5 Both Hyp and Pro-rich tropocollagen are more stable as a hexamer than as a heptamer.**

Both the heptamers and hexamers showed similar trends when the properties of the Hyp and Pro-enriched tropocollagens. This demonstrates that the effect of the Pro and Hyp residues in prevalent in multiple tropocollagen configurations. The aim of computing both hexameric and heptameric Hyp and Pro-enriched tropocollagens is to also compare the relative stability of each configuration. To facilitate this comparison, in the case of the number of contacts, hydrogen bonds, interaction energies were computed per tropocollagen strand, in order to make the values of the two configurations comparable. Adding an additional tropocollagen strand to both the Hyp and Pro-enriched system slightly increased the number of contacts per strand, slightly decreased the number of hydrogen bonds per strand in the case of Pro only, while remaining constant in Hyp. Most tellingly, the total interaction energy per strand decreased from -19.2 to -15.8 kJ·mol-1 by adding a tropocollagen strand to the Hyp-enriched system, and from -18.9 to -11.7 kJ·mol-1 when adding a tropocollagen strand to the Pro-rich system. With stronger per-strand interactions present in the hexamer, it seems that in both the Hyp and Pro-enriched systems the hexameric form is more stable. The difference between is not particularly large, and in both cases the stability and rigidity of the Hyp-rich system, and the flexibility and fluidity of the Pro-rich system is preserved.

# Conclusion

The findings significantly improve our understanding of the structural and functional properties of collagen. The study highlights the differences between hydroxyproline-rich and proline-rich tropocollagen systems, revealing how specific amino acid compositions influence stability, flexibility, and environmental interaction. We found that the hydroxyproline-rich systems have more atomic contacts than proline. In particular, the proline-rich systems have a slightly higher number of intramolecular hydrogen bonds, while hydroxyproline is more significant in forming intermolecular hydrogen bonds. In our analysis of solvent-accessible surface areas, we discovered that heptamer systems have a more expansive molecular surface area exposed to the solvent due to their higher solvent-accessible surface area. Furthermore, we observed hydroxyproline-rich systems possess larger hydrophilic surface areas than proline-rich systems. On the contrary, proline-rich systems exhibit higher hydrophobic surface areas. These observations emphasize the substantial influence of amino acid composition on the solvent accessibility and the hydrophobic/hydrophilic balance of protein structures. The analysis of RMSF reveals that the structures containing hydroxyproline display lower flexibility. The mean square displacement analysis showed reduced MSD and diffusion constants in hydroxyproline-rich systems. On the other hand, in Proline-rich systems, there is higher MSD, which indicates faster diffusion and smoother particle movement. Hydroxyproline-rich systems show stronger LJ interactions and reduced Coulombic interactions than proline-rich systems, attributed to the hydroxyl group’s influence on polarity and hydrophilicity. This insight is crucial for the development of collagen-based materials with tailored properties for biomedical applications, such as tissue engineering and drug delivery systems.

**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. Buehler MJ. Nature Designs Tough Collagen: Explaining the Nanostructure of Collagen Fibrils.; 2006. www.pnas.orgcgidoi10.1073pnas.0603216103

2. Yu Z, An B, Ramshaw JAM, Brodsky B. Bacterial collagen-like proteins that form triple-helical structures. *J Struct Biol*. 2014;186(3):451-461. doi:10.1016/j.jsb.2014.01.003

3. Nuñez SM, Guzmán F, Valencia P, Almonacid S, Cárdenas C. Collagen as a source of bioactive peptides: A bioinformatics approach. *Electronic Journal of Biotechnology*. 2020;48:101-108. doi:10.1016/j.ejbt.2020.09.009

4. Sorushanova A, Delgado LM, Wu Z, et al. The Collagen Suprafamily: From Biosynthesis to Advanced Biomaterial Development. *Advanced Materials*. 2019;31(1). doi:10.1002/adma.201801651

5. Kar K, Ibrar S, Nanda V, Getz TM, Kunapuli SP, Brodsky B. Aromatic interactions promote self-association of collagen triple-helical peptides to higher-order structures. *Biochemistry*. 2009;48(33):7959-7968. doi:10.1021/bi900496m

6. Buehler MJ. Defining nascent bone by the molecular nanomechanics of mineralized collagen fibrils. In: *ASME International Mechanical Engineering Congress and Exposition, Proceedings*. Vol 12. American Society of Mechanical Engineers (ASME); 2010:795-798. doi:10.1115/IMECE2009-12137

7. Khare E, Yu CH, Gonzalez Obeso C, et al. Discovering design principles of collagen molecular stability using a genetic algorithm, deep learning, and experimental validation. 2022;119:2209524119. doi:10.1073/pnas

8. Jenkins CL, Bretscher LE, Guzei IA, Raines RT. Effect of 3-hydroxyproline residues on collagen stability. *J Am Chem Soc*. 2003;125(21):6422-6427. doi:10.1021/ja034015j

9. Walker KT, Nan R, Wright DW, et al. Non-linearity of the collagen triple helix in solution and implications for collagen function. *Biochemical Journal*. 2017;474(13):2203-2217. doi:10.1042/BCJ20170217

10. Shoulders MD, Raines RT. Collagen structure and stability. *Annu Rev Biochem*. 2009;78:929-958. doi:10.1146/annurev.biochem.77.032207.120833

11. Zelenovskii PS, Vasileva DS, Vasilev SG, Kopyl S, Kholkin A. Ferroelectricity in glycine: A mini-review. *Front Mater*. 2022;9. doi:10.3389/fmats.2022.918890

12. Phang JM. The regulatory mechanisms of proline and hydroxyproline metabolism: Recent advances in perspective. *Front Oncol*. 2023;12. doi:10.3389/fonc.2022.1118675

13. Cotuá J, Cotes S. In silico Study of Solvent Effects on the Intramolecular Hydrogen Bond of Hydroxy Proline. *Oriental Journal Of Chemistry*. 2022;38(3):676-680. doi:10.13005/ojc/380318

14. Hu S, He W, Wu G. Hydroxyproline in animal metabolism, nutrition, and cell signaling. *Amino Acids*. 2022;54(4):513-528. doi:10.1007/s00726-021-03056-x

15. Xu S, Gu M, Wu K, Li G. Unraveling the Role of Hydroxyproline in Maintaining the Thermal Stability of the Collagen Triple Helix Structure Using Simulation. *Journal of Physical Chemistry B*. 2019;123(36):7754-7763. doi:10.1021/acs.jpcb.9b05006

16. Chow WY, Bihan D, Forman CJ, et al. Hydroxyproline Ring Pucker Causes Frustration of Helix Parameters in the Collagen Triple Helix. *Sci Rep*. 2015;5. doi:10.1038/srep12556

17. Wang H. A review of the effects of collagen treatment in clinical studies. *Polymers (Basel)*. 2021;13(22). doi:10.3390/polym13223868

18. Haq F, Ahmed N, Qasim M. Comparative genomic analysis of collagen gene diversity. *3 Biotech*. 2019;9(3). doi:10.1007/s13205-019-1616-9

19. Richards AJ, Snead MP. Molecular Basis of Pathogenic Variants in the Fibrillar Collagens. *Genes (Basel)*. 2022;13(7). doi:10.3390/genes13071199

20. Tsujita Y, Kondo T. A building block of collagen fibrils demonstrated by sequential aqueous counter collision process. *Journal of Fiber Science and Technology*. 2019;75(9):112-118. doi:10.2115/fiberst.2019-0014

21. Hulmes DJS. Building collagen molecules, fibrils, and suprafibrillar structures. In: *Journal of Structural Biology*. Vol 137. Academic Press Inc.; 2002:2-10. doi:10.1006/jsbi.2002.4450

22. Gordon MK, Gerecke DR, Olsen BR. *Type XII Collagen: Distinct Extracellular Matrix Component Discovered by CDNA Cloning (Recombinant DNA/Nucleotide Sequencing/Collagen Genes)*. Vol 84.; 1987. https://www.pnas.org

23. Cutini M, Bocus M, Ugliengo P. Decoding Collagen Triple Helix Stability by Means of Hybrid DFT Simulations. *Journal of Physical Chemistry B*. Published online 2019:7354-7364. doi:10.1021/acs.jpcb.9b05222

24. Kirkness MWH, Lehmann K, Forde NR. *Mechanics and Structural Stability of the Collagen Triple Helix*.

25. Rowe J, Rö K. Chemical bonds in collagen rupture selectively under tensile stress †. *Phys Chem Chem Phys*. 2023;25:2331. doi:10.5281/zenodo.7107608

26. Gautieri A, Vesentini S, Redaelli A, Buehler MJ. Viscoelastic properties of model segments of collagen molecules. *Matrix Biology*. 2012;31(2):141-149. doi:10.1016/j.matbio.2011.11.005

27. Bailey AJ, Paul RG, Knott L. *Mechanisms of Maturation and Ageing of Collagen*. Vol 106.; 1998.

28. Madhavi WAM, Weerasinghe S, Fullerton GD, Momot KI. Structure and Dynamics of Collagen Hydration Water from Molecular Dynamics Simulations: Implications of Temperature and Pressure. *Journal of Physical Chemistry B*. 2019;123(23):4901-4914. doi:10.1021/acs.jpcb.9b03078

29. Matamoros P, Pinheiro S, Viayna A, Zamora W. Towards an Understanding of the Lipophilicity of Non-Coded Amino Acids: Computational Simulations of Proline Analogs. In: *2022 IEEE 4th International Conference on BioInspired Processing, BIP 2022*. Institute of Electrical and Electronics Engineers Inc.; 2022. doi:10.1109/BIP56202.2022.10032480

30. Aduri R, Psciuk BT, Saro P, Taniga H, Schlegel HB, SantaLucia J. AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J Chem Theory Comput*. 2007;3(4):1464-1475. doi:10.1021/ct600329w

31. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. 2015;11(8):3696-3713. doi:10.1021/acs.jctc.5b00255

32. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926-935. doi:10.1063/1.445869

33. Bekker H, Berendsen HJC. LINCS: A Linear Constraint Solver for molecular simulations. *Article in Journal of Computational Chemistry*. Published online 1998. doi:10.1002/(SICI)1096-987X(199709)18:123.0.CO;2-H

34. Nosé S. A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys*. 1984;81(1):511-519. doi:10.1063/1.447334

35. Ke Q, Gong X, Liao S, Duan C, Li L. Effects of thermostats/barostats on physical properties of liquids by molecular dynamics simulations. *J Mol Liq*. 2022;365. doi:10.1016/j.molliq.2022.120116

36. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103(19):8577-8593. doi:10.1063/1.470117

37. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 2008;4(3):435-447. doi:10.1021/ct700301q

38. Mukhopadhyay A. On Clustering using Gromos Algorithm on biased and unbiased trajectories followed by reweighting. On Clustering using Gromos Algorithm on biased and unbiased trajectories followed by reweighting. Published online 2022. doi:10.13140/RG.2.2.10132.48007

39. Gorelov S V, Titov AI, Shvetsov A V. Determination of hydrogen bonds in Gromacs: new implementation to overcome the limitation. doi:10.1101/2023.09.01.555860

40. Yekeen AA, Durojaye OA, Idris MO, Muritala HF, Arise RO. CHAPERONg: A tool for automated GROMACS-based molecular dynamics simulations and trajectory analyses. *Comput Struct Biotechnol J*. 2023;21:4849-4858. doi:10.1016/j.csbj.2023.09.024

41. Cheng S, Robbins MO. Defining Contact at the Atomic Scale. Published online April 7, 2010. doi:10.1007/s11249-010-9682-5

42. Bhatnagar RS, Pattabiraman N, Sorensen KR, Langridge R, Macelroy RD, Renugopalakrishnan V. Inter-chain proline: Proline contacts contribute to the stability of the triple helical conformation. *J Biomol Struct Dyn*. 1988;6(2):223-233. doi:10.1080/07391102.1988.10507709

43. Hamacher K. Efficient quantification of the importance of contacts for the dynamical stability of proteins. *J Comput Chem*. 2011;32(5):810-815. doi:10.1002/jcc.21659

44. Qi Y, Zhou D, Kessler JL, et al. Terminal repeats impact collagen triple-helix stability through hydrogen bonding. *Chem Sci*. 2022;13(42):12567-12576. doi:10.1039/d2sc03666e

45. Ratnatilaka Na Bhuket P, Eastwood JR, Qin Z, et al. Structural and Positional Effects of Peptoid Residues on Stability of the Collagen Triple Helix 1 2. doi:10.26434/chemrxiv-2024-pz51s

46. Wu X, Liu Y, Liu A, Wang W. Improved thermal-stability and mechanical properties of type I collagen by crosslinking with casein, keratin and soy protein isolate using transglutaminase. *Int J Biol Macromol*. 2017;98:292-301. doi:10.1016/j.ijbiomac.2017.01.127

47. Mitternacht S. FreeSASA: An open source C library for solvent accessible surface area calculations. Published online January 25, 2016. doi:10.12688/f1000research.7931.1

48. Raghunathan S. Solvent accessible surface area-assessed molecular basis of osmolyte-induced protein stability. *RSC Adv*. 2024;14(34):25031-25041. doi:10.1039/D4RA02576H

49. Savojardo C, Manfredi M, Martelli PL, Casadio R. Solvent Accessibility of Residues Undergoing Pathogenic Variations in Humans: From Protein Structures to Protein Sequences. *Front Mol Biosci*. 2021;7. doi:10.3389/fmolb.2020.626363

50. Watanabe S, Tanimoto Y, Nishiwaki H, Watanabe Y. Identification and characterization of bifunctional proline racemase/hydroxyproline epimerase from archaea: Discrimination of substrates and molecular evolution. *PLoS One*. 2015;10(3). doi:10.1371/journal.pone.0120349

51. Schobert B, Tschesche H. *UNUSUAL SOLUTION PROPERTIES OF PROLINE AND ITS INTERACTION WITH PROTEINS*. Vol 541.; 1978.

52. Martínez L. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS One*. 2015;10(3). doi:10.1371/journal.pone.0119264

53. Chen YS, Chen CC, Horng JC. Thermodynamic and kinetic consequences of substituting glycine at different positions in a Pro-Hyp-Gly repeat collagen model peptide. *Biopolymers*. 2011;96(1):60-68. doi:10.1002/bip.21470

54. Riahi MK, Qattan IA, Hassan J, Homouz D. Identifying short- and long-time modes of the mean-square displacement: An improved nonlinear fitting approach. *AIP Adv*. 2019;9(5). doi:10.1063/1.5098051

55. Li T, Shen J, Wu L, et al. Atomic-Scale Observation of Carbon Distribution in High-Performance Carbon-Doped Ge2Sb2Te5 and Its Influence on Crystallization Behavior. *Journal of Physical Chemistry C*. 2019;123(21):13377-13384. doi:10.1021/acs.jpcc.9b02098

56. Mangaud E, Rotenberg B. Sampling mobility profiles of confined fluids with equilibrium molecular dynamics simulations. Published online May 7, 2020. doi:10.1063/5.0013952

57. Rawal A, Rhinehardt KL, Mohan R V., Pendse M. Influence of Hydroxyproline on Mechanical Behavior of Collagen Mimetic Proteins Under Fraying Deformation-Molecular Dynamics Investigations. *J Biomech Eng*. 2021;143(8). doi:10.1115/1.4050648

58. Świątek A, Kuczera K, Szoszkiewicz R. Effects of Proline on Internal Friction in Simulated Folding Dynamics of Several Alanine-Based α-Helical Peptides. *Journal of Physical Chemistry B*. 2024;128(16):3856-3869. doi:10.1021/acs.jpcb.4c00623

59. Umumararungu T, Gahamanyi N, Mukiza J, et al. Proline, a unique amino acid whose polymer, polyproline II helix, and its analogues are involved in many biological processes: a review. *Amino Acids*. 2024;56(1). doi:10.1007/s00726-024-03410-9

60. Improta R, Berisio R, Vitagliano L. Contribution of dipole–dipole interactions to the stability of the collagen triple helix. *Protein Science*. 2008;17(5):955-961. doi:10.1110/ps.073301908