**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

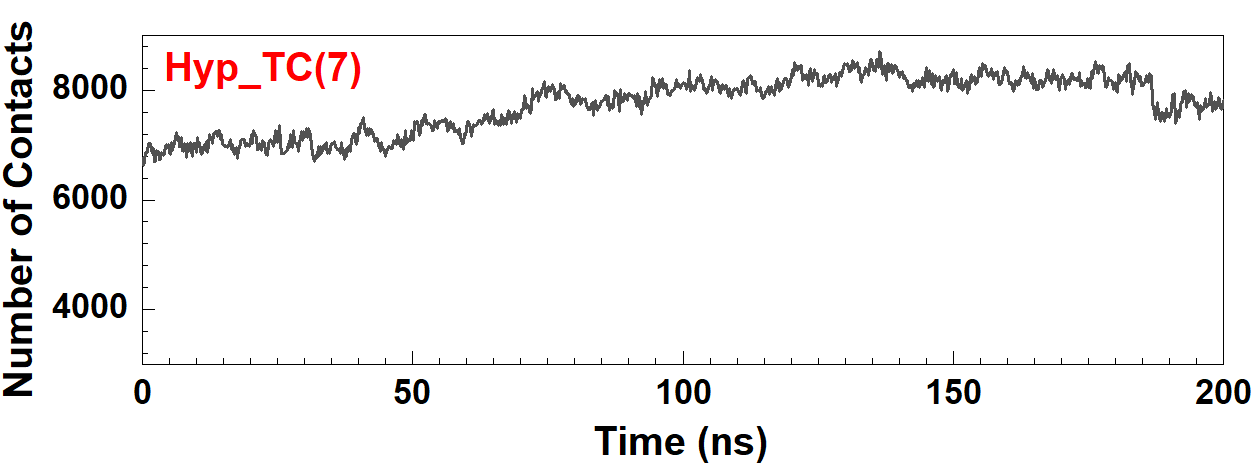
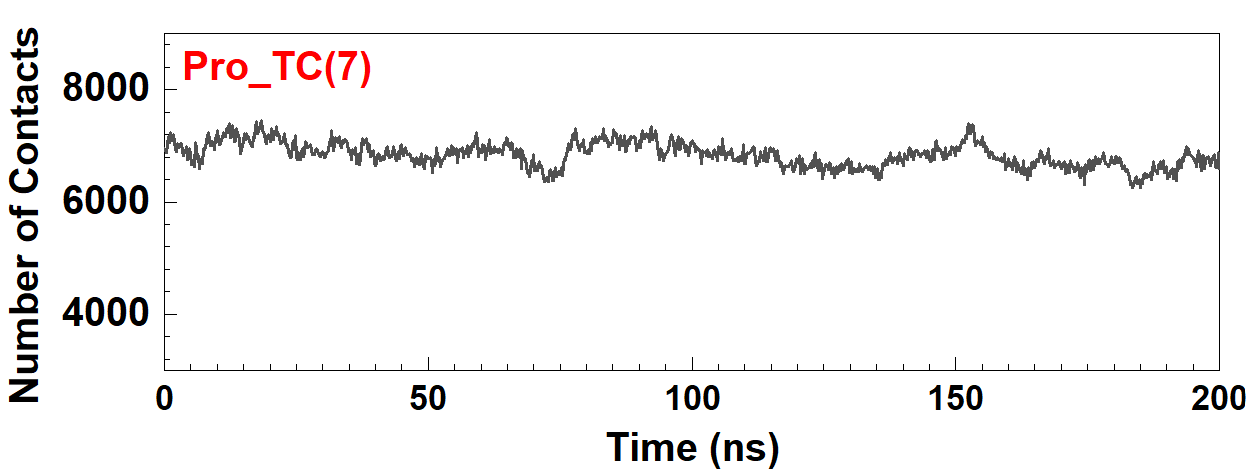
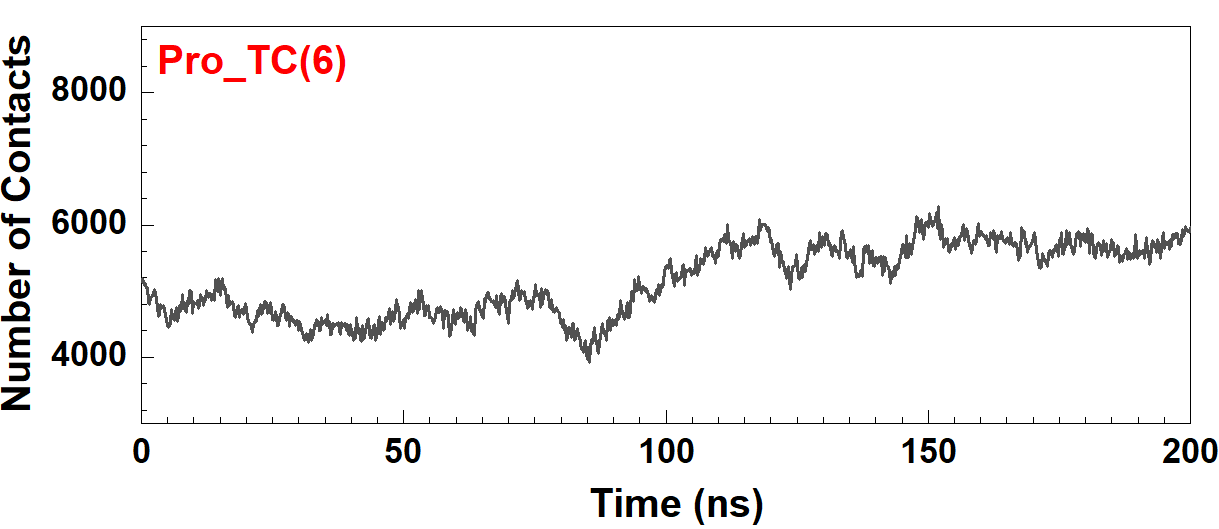
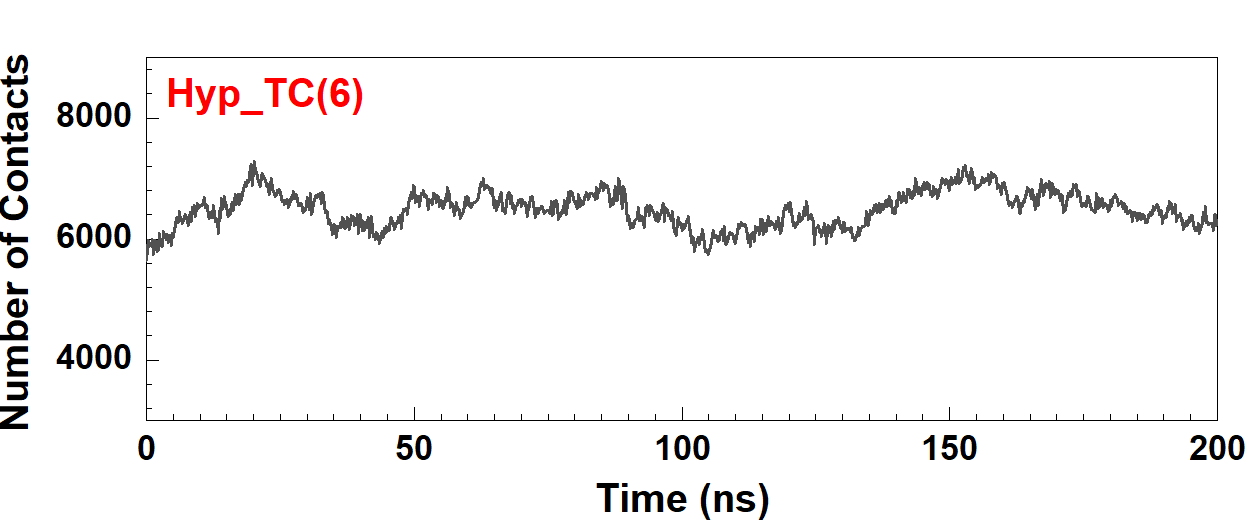
## Number of Contacts Between Tropocollagen Strands

***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.



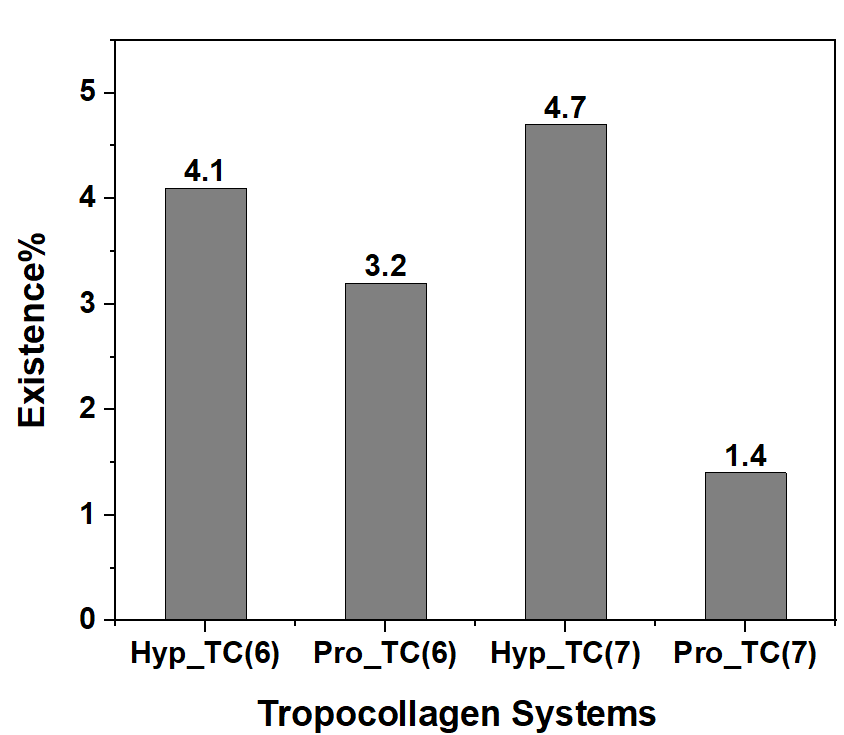
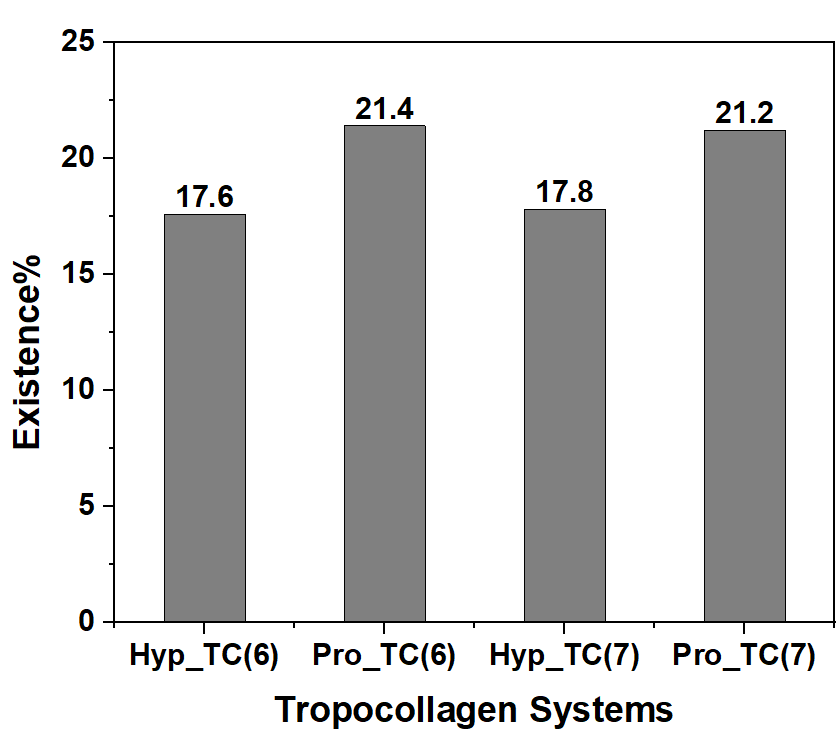
***Figure 3.*****shows the average number of contacts for three runs in the four tropocollagen systems: Hyp\_TC(6), Pro\_TC(6), Hyp\_TC(7), and Pro\_TC(7) over a 200 ns simulation period.**

## 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.