Hungarian Materials and Chemical Sciences and Engineering, Volume 47, No. 1 (2023), pp. 118–127.
http://doi.org/10.32974/mse.2022.012

INVESTIGATING COLLAGEN AS A BIO-MATERIAL BY MOLECULAR DYNAMICS SIMULATIONS

DIMAH ZAKARAIA[1] – DALAL K. THBAYH[2] –
BÉLA FISER[3] – MICHAEL C. OWEN[4]

Abstract: In this work, molecular dynamics simulation is used to describe and analyze the behavior of model collagen polymer $(Pro\text{-}Pro\text{-}Gly)_9$. This project aims to highlight the important role of molecular dynamic simulation in determining the structural stability of collagen, and establishing collagen as a hydrophobic or hydrophilic protein under different temperatures. The system was simulated at four different temperatures (300, 310, 320, and 330 K). The results indicate that the average number of hydrogen bonds within the protein and the protein backbone was similar at each temperature. The solvent-accessible surface area of hydrophobic and hydrophilic atoms for the four temperatures indicates that the collagen model peptide is mostly hydrophobic. All the results show that the structure of the studied polymer was the least stable at 320 K and the most stable at lower temperatures (300 K). The average effect across the first 100 ns was investigated. The dominant states obtained within this time interval will be explored in following studies. Researchers can use the results of this work to develop collagen with the appropriate thermal stability for biological applications.

Keywords: Collagen, polymer, hydrogen bonds, hydrophilic, hydrophobic

INTRODUCTION

Collagen is a structural protein found in abundance in all mammals. It makes up one-third of the total protein in humans, accounts for three-quarters of the dry weight of skin, and is the most widespread component of the extracellular matrix [1]. Collagen is a significant fibrous material responsible for the structural integrity of many body parts like skin, bone, tendons, and teeth. Moreover, it is a scaffold for cell cohesion as an extracellular matrix in many organs [2]. In vertebrates, there are 28 different forms of collagen, each with at least forty-six different polypeptide chains, and many other proteins have collagenous domains [3], [4]. Its

[1] Institute of Chemistry, University of Miskolc
H-3515 Miskolc-Egyetemváros, Hungary
zakaraia.dimah@uni-miskolc.hu

[2] Institute of Chemistry, University of Miskolc
Higher Education and Industrial Cooperation Centre, University of Miskolc
H-3515 Miskolc-Egyetemváros, Hungary
Polymer Research Center, University of Basrah, Basrah, Iraq
dalal.thebayh@uobasrah.edu.iq

[3] Institute of Chemistry, University of Miskolc
Higher Education and Industrial Cooperation Centre, University of Miskolc
H-3515 Miskolc-Egyetemváros, Hungary
Ferenc Rakoczi II Transcarpathian Hungarian College of Higher Education, 90200 Beregszász, Transcarpathia, Ukraine
kemfiser@uni-miskolc.hu

[4] Institute of Chemistry, University of Miskolc
Higher Education and Industrial Cooperation Centre, University of Miskolc
H-3515 Miskolc-Egyetemváros, Hungary
kemowen@uni-miskolc.hu

amino acid sequence is generally $(Gly\text{-}Xxx\text{-}Yyy)_n$, with strict sequence constraints of a glycine residue (Gly) at every third position, and the Xxx and Yyy positions occupied by proline (Pro) and hydroxyproline (HYP) residues, respectively [2]. The collagen molecule has a triple-helical structure due to the above characteristics of its amino acid sequence. This enables the smallest amino acid residue, Gly, to be positioned at the interface of the triple-helix, while the amino acids at the X and Y positions form the exterior. A microfibril is formed when triple-helical molecules combine [1]. Collagen model peptide $(Pro\text{-}Pro\text{-}Gly)_9$ has been used for this work. Molecular dynamics (MD) simulation is a rapidly growing science branch that has proven to be a reliable method for understanding the dynamic behavior of biomolecules [5]. MD simulations analyze the movement of molecules by using time-dependent numerical integration of Newton's equations of motion, which, when supplemented with experimental studies, can become effective instruments for understanding complex systems that exist in conformational ensembles [6]. There are various programs for performing molecular dynamic simulations, such as GROMACS, Amber, CHARMM, NAMD, Open MM, and LAMMPS. Researcher Leikina and his colleagues experimentally determined that at room temperature, collagen type I exhibits thermal instability [7]. Other researchers used thermal kinetic analysis to evaluate the thermal stability of fish collagen and found that fish collagen is not stable at temperature 21 °C [8]. Through the MD simulations applied by researchers Streeter and Leeuw on collagen fibrils, they proved that by MD it is possible to determine the total number of interprotein interactions, the functional groups that contribute most to these interactions, and the spatial distribution of these interactions throughout the D period of the fibril [9]. Researchers Ebrahimi and his colleagues also reported the efficiency of the MD simulation in studying the mechanical properties of collagen [10]. This work hypothesizes that the appropriate collagen environment could be predicted by MD simulation by controlling the system's temperature and understanding the effect of different temperatures on the thermal stability of collagen. In this study, GROMACS (GROningen MAchine for Chemical Simulations) was used to simulate the structural ensembles of collagen model peptide $(Pro\text{-}Pro\text{-}Gly)_9$ at four different temperatures (300, 310, 320, and 330 K). the influences of temperatures on the behavior, structure, and stability of collagen were investigated to gain insight about the effectiveness of MD simulation in understanding the dynamic behavior and thermal stability of polymers in general. The particular structure is relevant for ongoing study of proline-containing collagen.

1. METHODS

1.1. Molecular Dynamics Simulation

All simulations were carried out using the GROMACS 2020 program package [11].

1.1.1. Simulated Systems

In the present work, MD simulations were used to study a specific collagen model peptide $(Pro\text{-}Pro\text{-}Gly)_9$ at four different temperatures (300, 310, 320, and 330 K). In each case, the model system consisted of six peptide chains and water as solvent. The number of solvent molecules was 36950 at all four temperatures. The model system in each simulation was placed in a cubic box at least 1.0 nm from the edges.

1.1.2. Force Field and Water Models

Optimized Potentials for Liquid Simulations for all-atom parameters (OPLS-AA/L) force field was used to represent the protein in our system [12]. The parameters of the simple point charge (SPC216) model were used for water. The SPC216 water model is a 3-site

rigid water molecule with individual charges and Lennard-Jones parameters for each of its three atoms [13].

1.1.3. Simulation Protocol for Equilibration and Production

The leap-frog algorithm with a 2 ps time step was used for the time propagation. The minimization stopped after the maximum force on any single atom in the system was less than 1,000.0 kJ/mol/nm. The system was equilibrated to remove steric atomic clashes and simulated at four different temperatures (300, 310, 320, 330 K). The system was first equilibrated for 100 ps under constant Number of particles, Volume, and Temperature (NVT) and constant Number of particles, Pressure, and Temperature (NPT) conditions. The system was then simulated for 100 ns under NPT conditions, and this trajectory was subsequently analyzed [14]. A thermostat and barostat are required to simulate the NPT and NVT ensembles. Parrinello-Rahman barostat and Berendsen thermostat were used in this work.

1.2. Analysis of MD Simulations

The GROMACS 2020 program package was used to do all the simulation analyses. The protein structures were rendered using the 1.9.3 version of the Visual Molecular Dynamics (VMD) program [15]. The hydrogen bonds were analyzed with the *gmx hbond* program. In the analysis, the maximum hydrogen bond length and bond angle were set to 0.36 nm and 30°, respectively. The entire protein and the protein backbone were used as two subsets to compute the hydrogen bonds. The Origin 2018 program was used to smoothen the graphs of the hydrogen bonds within the protein and the protein backbone.

The solvent-accessible surface area (SASA) was computed using the *gmx sasa* program, which utilizes the double cubic lattice method of Eisenhaber et al. [16]. Hydrophobic atoms with charges (–0.2 to 0.2), hydrophilic atoms with charges greater than 0.2 and less than –0.2, and an all-atom group were used as subsets of protein atoms to compute the SASA. Representative protein structures from each trajectory were obtained with the *gmx cluster* program using the method developed by Daura et al. [17]. The algorithm was a gromos method for clustering, and the root-mean-square deviation (RMSD) cutoff was 0.35 nm for the clustered groups.

The RMSD was computed using the *gmx rms* program to compare the structures. RMSD value for collagen backbone was calculated for 100 ns simulation to check the stability of the structure of the protein. To check the structures at different times, VMD was used.

2. RESULTS

The standard error for averages of the number of hydrogen bonds, solvent-accessible surface area, and the root-mean-square deviation was computed and was found to be around 0.3.

2.1. Hydrogen Bonding

For four temperatures (300, 310, 320, and 330 K), the number of hydrogen bonds within the collagen and the collagen backbone was calculated as a function of time, as shown in *Figure 1*. The average of these values is shown in *Table 1*. The results indicated that the average number of hydrogen bonds within the protein and the protein backbone had very similar values at four different temperatures. Thus, it was noticed that the average number of hydrogen bonds within the protein and the protein backbone increased from 20.4 to 23.1 when the temperature was increased from 300 K to 320 K and then decreased to 21.7 at 330 K. The biggest average

number of hydrogen bonds was recorded at 320 K, which means the interactions within this model polymer seemed to be strongest at 320 K, and the lowest number of hydrogen bonds was at 300 K, which means the interactions seemed to be weakest at lower temperatures (300 K).

Table 1

The average number of hydrogen bonds within the protein and the protein backbone

Temperature (K)	300	310	320	330
Average number of hydrogen bonds within the protein	20.4	21.6	23.1	21.7
Average number of hydrogen bonds within the protein backbone	20.4	21.6	23.1	21.7

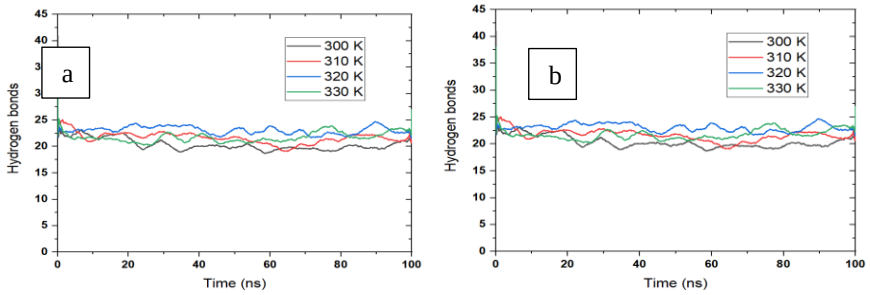

Figure 1

Number of hydrogen bonds (a) within the protein and (b) the protein backbone as a function of time for the studied collagen model peptide at 300, 310, 320, and 330 K

2.2. Solvent Accessible Surface Area

The solvent-accessible surface area (SASA) of the studied collagen is the surface that can be accessed by water in this specific case. The SASA of the studied polymer as a function of time at four different temperatures (300, 310, 320, and 330 K) is shown in *Figure 2*, and the average values are shown in *Table 2*. The results showed that the average value of SASA of all atoms at all four different temperatures (300, 310, 320, and 330 K) were between 88.9 nm^2 and 90.3 nm^2. Furthermore, the average SASA values of hydrophobic atoms were between 73.2 nm^2 and 74.5 nm^2, while the average SASA values of hydrophilic atoms were between 15.7 nm^2 and 16.6 nm^2 at all four different temperatures (300, 310, 320 and 330 K). Thus, collagen is a hydrophobic polymer because it has a higher hydrophobic SASA. The SASA of all atoms decreased slightly by increasing the temperatures (320 K and 330 K). The hydrophobic SASA was the highest at 300 K (74.5 nm^2).

Table 2

The average values of solvent-accessible surface area (SASA) of the studied polymer at four different temperatures (300, 310, 320, and 330 K)

Temperature (K)	The average SASA values of all atoms (nm^2)	The average SASA values of hydrophobic atoms (nm^2)	The average SASA values of hydrophilic atoms (nm^2)
300	90.3	74.5	15.8
310	90.3	73.7	16.6
320	88.9	73.2	15.7
330	89.4	73.4	16.1

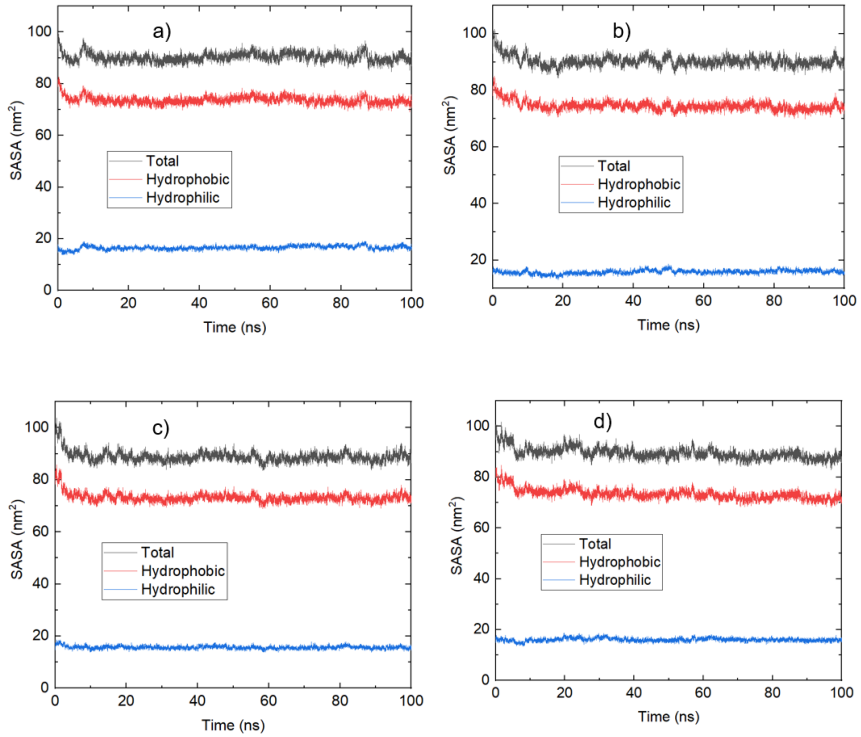

Figure 2

The solvent-accessible surface area (SASA) of hydrophobic (red), hydrophilic (blue), and all atoms (black) as a function of time for the studied collagen structure at four different temperatures (a) 300 K, (b) 310 K, (c) 320 K, and (d) 330 K

2.3. Central structures of the largest clusters

The size of the three largest clusters was analyzed at all temperatures (300, 310, 320, and 330 K) for the studied collagen model peptide *(Table 3)*. The cluster analysis showed that the percentages of all structures at 300 K for the largest clusters were 78.9%, 2.7%, and 2.2%, respectively. The percentages of all structures for the other temperatures were relatively similar, meaning the largest cluster was the first one because it had the biggest number of structures. Because of that, only the central structures of the first cluster for four temperatures were rendered. The biggest number of structures in the largest cluster was at 300 K compared with the other temperatures, and the smallest number of structures in the largest cluster was at 320 K. The central structures of the largest clusters obtained at temperatures 300, 310, 320, and 330 K for the studied collagen are shown in *Figure 3*. It can be noticed from *Figure 3* that the protein folds into a well-defined three-dimensional structure, and the type of collagen with 3D structures is fibrous. This type of protein forms long fibers and mostly consist of repeated sequences of amino acids which are insoluble in water.

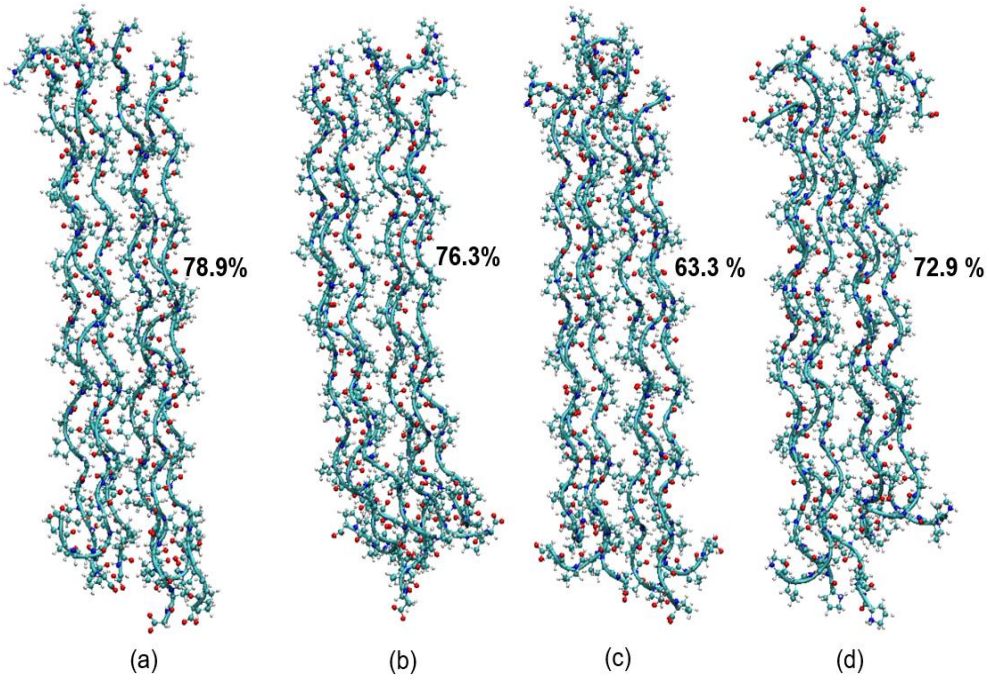

Figure 3

The central structures of the largest clusters of the studied collagen model obtained at different temperatures (a) 300 K, (b) 310 K, (c) 320 K, and (d) 330 K. 78.9% of the structures were in the largest cluster at 300 K, 76.3% of the structures were in the largest cluster at 310 K, 63.3% of the structures were in the largest cluster at 320 K, and 72.9% of the structures were in the largest cluster at 330 K.

Table 3

The sizes of the largest clusters of the simulated collagen model peptide at different temperatures, determined by cluster analysis

Number of clusters	**Number of structures in each cluster**				**% of all structures**			
	300 K	310 K	320 K	330 K	300 K	310 K	320 K	330 K
1	7,897	7,627	6,326	7,295	78.9%	76.3%	63.3%	72.9%
2	270	290	607	548	2.7%	2.9%	6.1%	5.5%
3	221	239	277	216	2.2%	2.4%	2.8%	2.2%

2.4. The root-mean-square deviation

The root-mean-square deviation (RMSD**)** obtained at 300, 310, 320, and 330 K for collagen is shown in *Figure 4*. The average RMSD at 300, 310, 320, and 330 K for the studied collagen is shown in *Table 4*. The results indicated that the average RMSD at 320 K was higher than the average RMSD at other temperatures. From the RMSD plot (a) at temperature 300 K, it can be observed that the structures before 30 ns moved to 0.5 nm then, between 30 ns and 80 ns,

jumped to between 4 nm and 6 nm and then, between 80 and 100 ns changed back to 0.5 nm. RMSD results indicated larger changes and distributions in protein structure at the other temperatures.

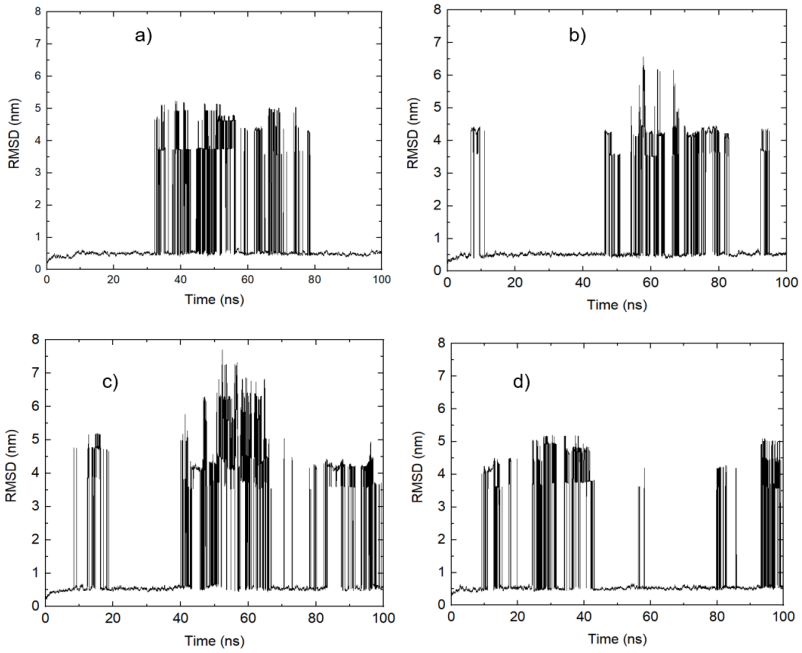

Figure 4
The root-mean-square deviation (RMSD) of the studied collagen obtained at temperatures (a) 300 K, (b) 310 K, (c) 320 K, and (d) 330 K

Table 4
The average number of RMSD at temperatures (300, 310, 320, and 330 K) for the studied collagen

Temperature (K)	300	310	320	330
RMSD (nm)	1.2	1.3	1.8	1.4

DISCUSSION

The time step of our simulation was two fs in length, whereas the duration of the simulation was 100 ns. The evaluation of the full configurational space was not mentioned in this work, but rather the focus was on the early denaturation process from an experimentally determined structure. The limitations of MD simulations required large memory allocation and time for output files. So, in the future we will try to use longer time steps to see many changes in its properties. Its strengths are the atomic resolution and capability to generate a configurational and dynamic ensemble.

The effect of the temperature on the hydrogen bonds within the collagen and its backbone was studied. From the results, it can be seen that by increasing the temperature from 300 K to 330 K, the average number of hydrogen bonds within the protein had the same value as

the average number of hydrogen bonds within the protein backbone for all the studied temperatures. The biggest average number of hydrogen bonds was at 320 K, meaning the interactions within polymers seemed to be strongest at 320 K, and the polymer tends to be more rigid at this temperature. The lowest number of hydrogen bonds was at 300 K, meaning the interactions within polymers seemed the weakest at lower temperatures (300 K), and the polymer tends to be more flexible at low temperatures (300 K).

Based on the extent to which the hydrophobic and hydrophilic surface area comprised the total surface area, hydrophobic and hydrophilic interactions occurred between parts of the studied structure. Still, hydrophobic interactions took the bigger surface area from the total SASA, therefore collagen can be considered a hydrophobic polymer. The larger the hydrophobic SASA is, the stronger the aggregation tendency of the polymer would be at lower temperatures (e.g., 300 K).

The distribution of structures between the groups in the cluster analysis did not differ at the four temperatures; most of the structures were distributed in the first cluster for all temperatures, making this cluster the biggest one. From here, the largest clusters, the first one at each temperature, were used to render the central structure of the studied polymer to evaluate how effectively these structures represent our protein structure at each temperature. Based on the size of the largest cluster, at temperature 300 K, the size of the largest cluster was more than the others at other temperatures, and from here, at temperature 300 K, the protein structure was the most stable. On the other side, at temperature 320 K, the size of the largest cluster was less than the others at other temperatures, and from here, at temperature 320 K, the protein structure was the least stable.

The results indicate no differences between the rendered images at the studied four temperatures, and the structures had not been affected by increasing the temperature from 300 K to 330 K. These results suggest that, generally, the structure of collagen within this range of temperatures remains the same. Collagen folds into a well-defined three-dimensional structure, and the type of this polymer with 3D structures is fibrous.

The difference between the backbones of a protein from its initial structural conformation to its final position is measured by using RMSD. The deviations produced during the simulation can be used to estimate the stability of the protein in relation to its structure. Visualization results indicated that the six peptide chains come together for the low value of RMSD, and the structure of the polymer was separate for the high value of RMSD, which means that the smaller the deviations are, the more stable the protein structure is. The average number of RMSD at 300 K was less than the average number of RMSD at other temperatures, ensuring that the structure of the studied polymer was the most stable at lower temperatures (300 K). In comparison, the average number of RMSD at 320 K was higher than the average number of RMSD at other temperatures, ensuring that the structure of the studied polymer at 320 K was the least stable.

In conclusion, at lower temperatures (300 K), the interactions within collagen seemed the weakest, the tendency of the collagen to aggregate was stronger, and the protein structure was the most stable. At higher temperatures (320 K), the interactions within collagen seemed to be the strongest, the tendency of collagen to aggregate was weaker, and the protein structure was the least stable.

CONCLUSION

Collagen model peptide $(Pro\text{-}Pro\text{-}Gly)_9$ had been studied using molecular dynamics simulations and OPLS-AA/L all-atoms as a force field to know the properties and behavior of this protein and its interactions. SASA results indicated that collagen is a hydrophobic polymer. The largest

hydrophobic SASA was at 300 K, which means the tendency for the polymer to aggregate would be stronger at lower temperatures (300 K). The biggest size of the largest cluster was at 300 K, which means that at lower temperatures (300 K), the structure of the protein was the most stable. RMSD at 300 K was lower than RMSD at the rest temperatures, ensuring that the structure of the studied polymer at this temperature was the most stable. Studying the behavior of collagen by MD methods at different temperatures gives us a good step to investigate the influences of temperature on the behavior, structure, and stability of the macromolecule and, thus, to gain insight into the stability of polymers in general.

ACKNOWLEDGEMENTS

This research is supported by the European Union and the Hungarian State, co-financed by the European Regional Development Fund in the framework of the GINOP-2.3.4-15-2016-00004 project, which aims to promote cooperation between higher education and the industry. Further support was provided by the National Research, Development, and Innovation Fund (Hungary) within the TKP2021-NVA-14 project. The GITDA (Governmental Information-Technology Development Agency, Hungary) is gratefully acknowledged for allocating computing resources used in this work.

REFERENCES

[1] M. D. Shoulders and R. T. Raines: Collagen structure and stability. Annu. Rev. Biochem., Vol. 78, pp. 929–958, 2009. https://doi.org/10.1146/annurev.biochem.77.032207.120833

[2] C. Hongo, K. Noguchi, K. Okuyama, Y. Tanaka, and N. Nishino: Repetitive interactions observed in the crystal structure of a collagen-model peptide, [(Pro-Pro-Gly)$_9$]$_3$. *J. Biochem.*, Vol. 138, No. 2, pp. 135–144, Aug. 2005. https://doi.org/10.1093/jb/mvi108

[3] J. Brinckmann: Collagens at a glance. *Top. Curr. Chem.*, Vol. 247, pp. 1–6, 2005. https://doi.org/10.1007/b103817

[4] G. Veit, B. Kobbe, D. R. Keene, M. Paulsson, M. Koch, and R. Wagener: Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. *J. Biol. Chem.*, Vol. 281, No. 6, pp. 3494–3504, 2006, https://doi.org/10.1074/jbc.M509333200.

[5] T. Hansson, C. Oostenbrink, and W. F. Van Gunsteren: Molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, Vol. 12, No. 2, pp. 190–196, 2002. https://doi.org/10.1016/S0959-440X(02)00308-1

[6] M. C. Owen, A. Karner, R. Šachl, J. Preiner, M. Amaro, and R. Vácha: Force Field Comparison of GM1 in a DOPC Bilayer Validated with AFM and FRET Experiments. *J. Phys. Chem. B*, Vol. 123, No. 35, pp. 7504–7517, Sep. 2019. https://doi.org/10.1021/acs.jpcb.9b05095

[7] E. Leikina, M. V. Mertts, N. Kuznetsova, and S. Leikin: Type I collagen is thermally unstable at body temperature. *Proc. Natl. Acad. Sci. U.S.A.*, Vol. 99, No. 3, pp. 1314–1318, 2002, https://doi.org/10.1073/pnas.032307099.

[8] W. Liu, Z. Tian, C. Li, and G. Li: Thermal denaturation of fish collagen in solution: A calorimetric and kinetic analysis. *Thermochim. Acta*, Vol. 581, pp. 32–40, 2014. https://doi.org/10.1016/j.tca.2014.02.012

[9] I. Streeter and N. H. De Leeuw: A molecular dynamics study of the interprotein interactions in collagen fibrils. *Soft Matter*, Vol. 7, No. 7, pp. 3373–3382, 2011. https://doi.org/10.1039/c0sm01192d

[10] S. Ebrahimi, A. Montazeri, and H. Rafii-Tabar: Molecular dynamics study of the interfacial mechanical properties of the graphene-collagen biological nanocomposite. *Comput. Mater. Sci.*, Vol. 69, pp. 29–39, 2013. https://doi.org/10.1016/j.commatsci.2012.11.030

[11] P. S. Chauhan et al.: Gromacs 5.0.7. *Proteins Struct. Funct. Genet.*, Vol. 12, No. 3, pp. 169–200, 2011.

[12] E. Monvall: Statsbudgeten: de stora reformerna gäller arbetslivet. *Tidskr. Sver. Sjukskot.*, Vol. 43, No. 2, pp. 54–58, 1976.

[13] P. Mark and L. Nilsson: Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A*, Vol. 105, No. 43, pp. 9954–9960, 2001. https://doi.org/10.1021/jp003020w

[14] J. Wong-ekkabut and M. Karttunen: The good, the bad and the user in soft matter simulations. *Biochim. Biophys. Acta – Biomembr.*, Vol. 1858, No. 10, pp. 2529–2538, 2016, https://doi.org/10.1016/j.bbamem.2016.02.004.

[15] W. Humphrey, A. Dalke, and K. Schulten: VMD: Visual Molecular Dynamics. *J. Mol. Graph.*, Vol. 14, No. 1, pp. 33-38, 1996. https://doi.org/10.1016/0263-7855(96)00018-5

[16] J. Harris: Early Language Development: Implications for Clinical and Educational Practice. *Early Lang. Dev. Implic. Clin. Educ. Pract.*, Vol. 16, No. 3, pp. 1–265, 2020. https://doi.org/10.4324/9781003028253

[17] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. Van Gunsteren, and A. E. Mark: Peptide folding: When simulation meets experiment. *Angew. Chemie – Int. Ed.*, Vol. 38, No. 1–2, pp. 236–240, 1999. https://doi.org/10.1002/(sici)1521-3773(19990115)38:1/2<236::aid-anie236>3.0.co;2-m