Hungarian Materials and Chemical Sciences and Engineering, Volume 47, No. 1 (2023), pp. 109–117.
http://doi.org/10.32974/mse.2022.011

ANALYSIS OF MOLECULAR DYNAMICS (MD) SIMULATION OF CARBONIC ANHYDRASE

SAEED TALEI[1] – RACHID HADJADJ[2] –
PÉTER MIZSEY[3] – MICHAEL C. OWEN[4]

Abstract: Molecular Dynamics (MD) simulation is a computational method for analyzing the physical movements of atoms and molecules allowed to interact for a fixed duration of time. In this study, the thermal stability of carbonic anhydrase, which catalyzes the reaction of water and carbon dioxide, was investigated. Our simulations were performed in a box of water at four different temperatures, 300 K, 310 K, 320 K, and 330 K. The duration of each simulation was 100 ns, and thereafter the hydrogen bonds, Solvent Accessible Surface Area (SASA), as well as Root Mean Square Deviation (RMSD) were analyzed. Moreover, cluster analysis was done to identify representative structures at each temperature. The results showed that changing the temperature did not significantly impact the number of hydrogen bonds. The SASA had more fluctuation when the temperature increased. Moreover, the higher the temperature of the simulation was, the more clusters were obtained. The higher number of clusters indicates higher conformational flexibility and less-stable conformers forming during the simulation.

Keywords: Carbon anhydrase, MD simulation, RMSD, SASA, carbon dioxide

INTRODUCTION

Carbonic anhydrase is an enzyme found in red blood cells, gastric mucosa, pancreatic cells, and renal tubules. This enzyme catalyzes the interconversion of carbon dioxide (CO_2) and carbonic acid (H_2CO_3) [1]. This is a reaction that produces carbonic acid by capturing CO_2 into water.

$$CO_2 + H_2O \xrightarrow{\text{Carbonic Anhydrase}} H_2CO_3 \quad (1)$$

Carbonic anhydrase is a widespread enzyme in nature having a key role in the respiratory process of animals by facilitating CO_2 transportation. Carbonic anhydrase is utilized in the photosynthetic fixation of CO_2 in plants. The presence of carbonic anhydrase in plants was confirmed by Bradfield in 1947 [2, 3]. About 86% to 95% of the total carbonic anhydrase is found in chloroplasts [4], while the rest is completely restricted to the cytosol of mesophyll

[1] Institute of Chemistry, University of Miskolc
H-3515 Miskolc-Egyetemváros, Hungary
talei.saeed@uni-miskolc.hu
[2] Institute of Chemistry, University of Miskolc
H-3515 Miskolc-Egyetemváros, Hungary
hrachid.chemeng@gmail.com
[3] Institute of Chemistry, University of Miskolc
H-3515 Miskolc-Egyetemváros, Hungary
kemizsey@uni-miskolc.hu
[4] Higher Education and Industrial Cooperation Centre, University of Miskolc
H-3515 Miskolc-Egyetemváros, Hungary
kemowen@uni-miskolc.hu

cells [5]. A correlation was found by Khan [6] between carbonic anhydrase activity and photosynthetic rate suggesting that the enzyme serves as a biochemical marker for productivity, as this helps in carbon sequestration. Carbonic anhydrase activity is kept up by light and CO_2 concentration [7, 10].

Carbonic anhydrase is one of the fastest enzymes. It has the potential to catalyze over a million CO_2 hydration reactions per second. According to its capability for capturing carbon dioxide through CO_2 hydration reaction, however, its commercial implementation is limited by its thermal and chemical instability. Several researchers consider it a novel bio-inspired carbon capture and storage approach [11]. Carbonic anhydrase recently received a lot of attention to be used as an additive for solvent-based carbon capture process, which can improve the liquid side mass transfer coefficient [12]. As solvent-based capture absorption and desorption are thermal processes, thereby the thermal stability of this enzyme should be considered. In this study, the effect of temperature variation on the thermal stability and structure of carbonic anhydrase was investigated.

1. METHODS

1.1. Molecular Dynamics Simulation

Molecular dynamic (MD) simulation was carried out using the GROMACS 2020 program package. It is a widely used simulation used in different applications within chemistry, biology, and physics [13].

1.1.1. Simulated System

In this work, the system contained a carbonic anhydrase that was noted as '1dmy' in the Protein Data Bank (PDB). It was simulated in a rectangular box of water with the dimensions of 7.99 by 7.99 by 7.99 nm. Since the net electrostatic charge of carbonic anhydrase is negative two, two Na^+ were added to neutralize the system. One simulation was done at four different temperatures, 300 K, 310 K, 320 K, and 330 K. The molecular composition of the system is shown in *Table 1*.

Table 1

Molecular composition of the simulated system. The composition of the system was the same for all four temperatures

Molecule	Number
Carbon Anhydrase	1
Water	15807
Na^+	2

1.1.2. Force Fields and Water Models

The OPLS-AA force field was used for the simulations in this study, using the parameters developed for this force field. The TIP3P model was used for water [14].

1.1.3. Simulation Protocol

The system was equilibrated in two phases. The first one is under an NVT (constant number of particles, volume, and temperature) and the other is under NPT (constant number of particles, pressure, and temperature). The duration of both equilibration steps was 100 ps. MD simulations were then carried out using the Verlet algorithm under NPT conditions for 100 ns.

1.2. Analysis of MD Simulations

All simulation analyses in this work were conducted with the GROMACS 2020 program package. The Visual Molecular Dynamics (VMD) program was used to visualize and render the protein structures. To analyze the hydrogen bonds, the 'gmx h-bond' program was used. The maximum hydrogen bond length for the simulation was 0.36 nm and the bond angle used in the analysis was 30 degrees. The hydrogen bonds in two subsets of atoms were calculated, Hydrogen bonds within the entire protein and the protein main-chain atoms.

The solvent-accessible surface area was simulated by the 'gmx sasa' program. Three different subsets of atoms with our protein were considered, namely, the hydrophobic atoms, which included those with charges between –0.2 and 0.2 coulombs, the hydrophilic group with charges less than –0.2 and greater than 0.2 coulombs, and the third with all the atoms of the protein.

The cluster analysis was done by the 'gmx cluster' program with the RMSD cutoff of 0.1 nm and the algorithm described by Daura et al. [15] was used.

The 'gmx rms' program was used for RMSD. It is typically used as a quantitative measure of similarity between two or more protein structures reporting how well a submitted structure matches the known, target structure.

2. RESULT AND DISCUSSION

In this section the impact of temperature on the hydrogen bonds, solvent accessible surface area, root mean square deviation, and clustering of carbonic anhydrase was examined. Moreover, the average value of hydrogen bonds, SASA, and RMSD was calculated. For the hydrogen bonds and SASA, the standard error was significant, but for the last one standard error was insignificant.

2.1. Hydrogen Bonds within the Protein

In *Figure 1*, the hydrogen bonds within the protein main-chain atoms and within the entire protein are shown. According to *Figure 1(a)*, the number of hydrogen bonds for all temperatures fluctuated between 80 and 95 over the 100 ns trajectory. While the lowest number of bonds was obtained at 330 K, the highest number of hydrogen bonds were observed at 310 K and 320 K. Although at the beginning of the trajectory the number of hydrogen bonds was the same for all temperatures (since they start at the same structure), the figure at 330 K decreased during the 100 ns trajectory and stayed on the lowest amount between all the temperatures.

Regarding *Figure 1(b)*, there were also fluctuations at each temperature. The fewest number of hydrogen bonds in the total time interval was 153, which was observed at 330 K, while the highest number of hydrogen bonds was 180, and was observed at 310 K.

According to *Figure 1*, the number of hydrogen bonds in the entire protein was almost double of those only observed within the protein backbone. This suggests that half of the hydrogen bonds were within the protein side chain or between the side chain and backbone. *Table 2* represents the average number of hydrogen bonds for both the main chain atoms and entire protein alike. While the former slightly decreased by increasing the temperature, in the latter the highest value was at 310 K. Since the backbone H bonds are affected more by temperature variation and the backbone best defines the overall folding of the protein, there is a higher dependence on the backbone H bonds on temperature compared to the side-chain hydrogen bonds. Generally, hydrogen bonds contribute favorably to protein stability, thus, by increasing the number of hydrogen bonds increases the protein stability.

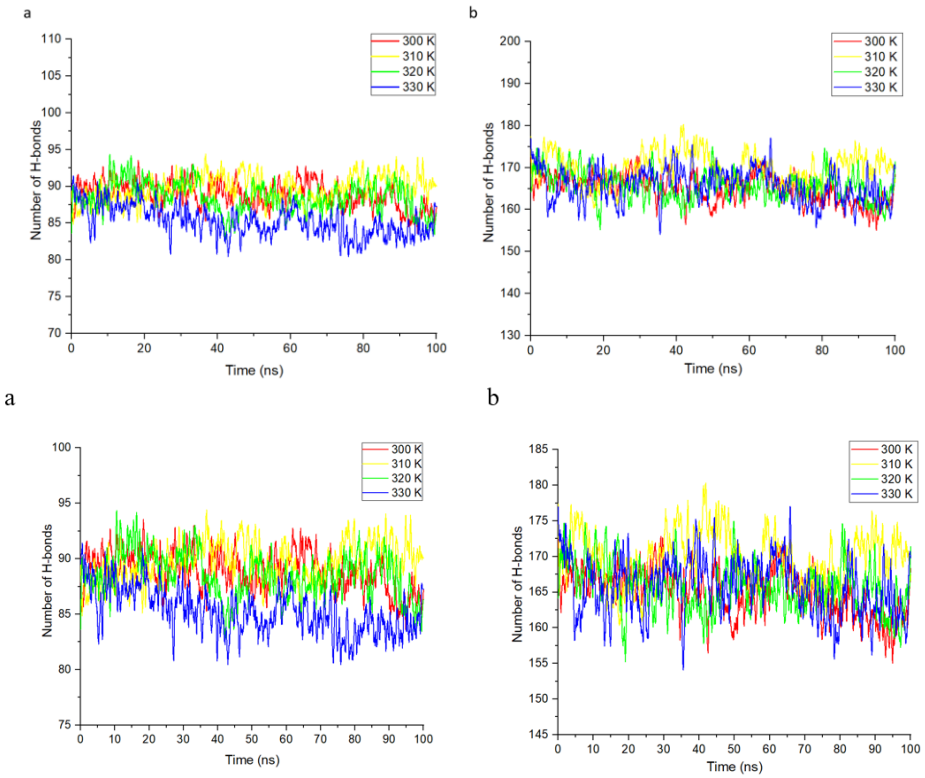

Figure 1

Variation of hydrogen bonds within the protein main-chain atoms (a) and within the entire protein and (b) over the 100 ns trajectory for each simulated temperature

***Table** 2*

The average number of hydrogen bonds within the main chain atoms and the entire protein at the considered temperatures in the simulation

Temperature (K)	Hydrogen bond within the main chain atoms	Hydrogen bond within the entire protein
300	88.9 ±0.1	164.9 ±0.1
310	88.7 ±0.1	169.9 ±0.1
320	88.5 ±0.1	165.6 ±0.1
330	85.2 ±0.1	165.7 ±0.1

2.2. Solvent Accessible Surface Area (SASA)

The SASA of proteins has always been considered a decisive factor in protein folding and stability studies [16, 17]. It is defined as the surface characterized around a protein by a hypothetical center of a solvent sphere with the van der Waals contact surface of the molecule [18].

The hydrophilic, hydrophobic, and total SASA analysis are depicted in *Figure 2* for all temperatures. The SASA of hydrophobic and hydrophilic groups fluctuated between 50 and

60 nm^2. *Figure 2a* and *Figure 2b* indicate the surface area of hydrophilic and hydrophobic protein, respectively. As driving force of folding process comes from the hydrophobic effect, the fluctuations in the surface area of hydrophobic protein compared to hydrophilic.

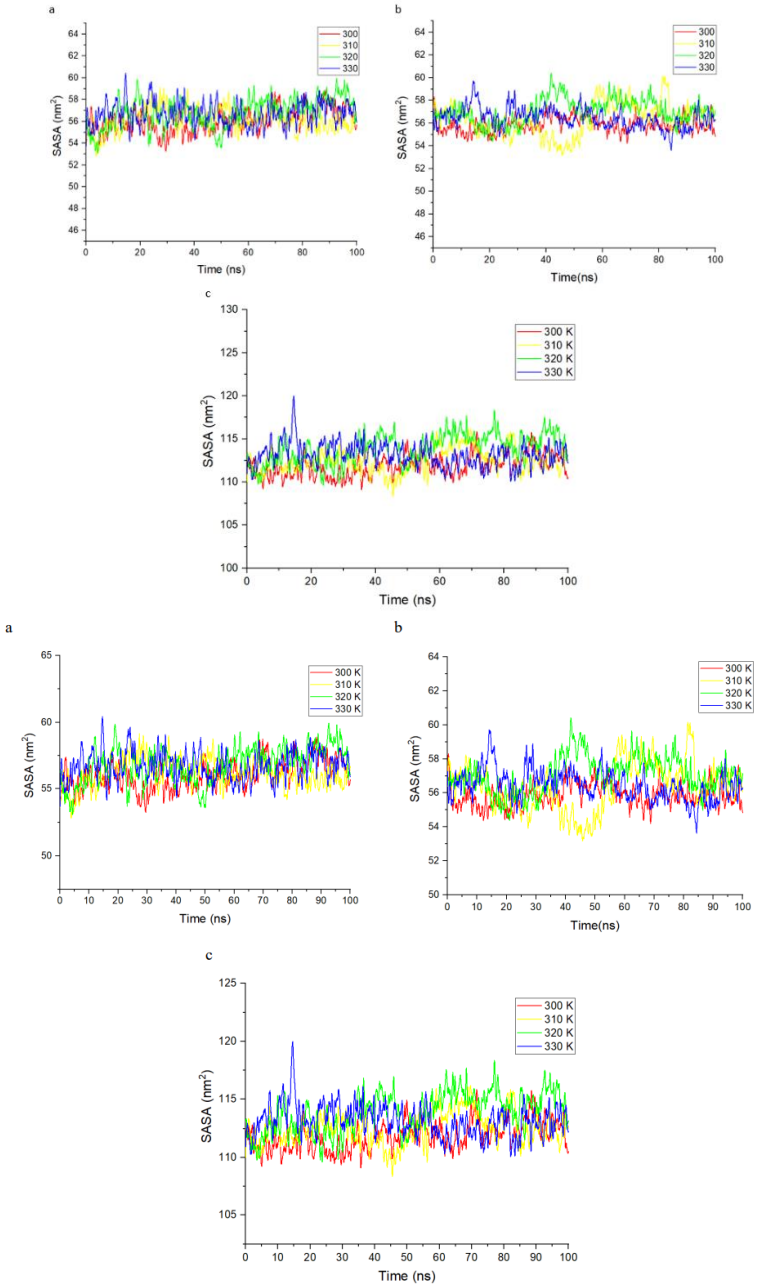

Figure 2

Solvent Accessible Surface Area for hydrophilic group (a), hydrophobic group (b), and the total (c) over the 100 ns trajectory for each simulated temperature

Figure 2c demonstrates the total surface area of protein which is calculated as a sum of the hydrophobic and hydrophilic surface area. The maximum value for total solvent accessible surface area of protein was 120 nm^2 and was observed at 330 K. *Table 3* represents the average value for SASA of hydrophobic, hydrophilic, and total groups at the considered temperature during the simulation. As is shown in the table, there was a burgeoning trend in the values from 300 K to 320 K, followed by a decrease in 330 K, the reason for this is that higher temperature may have facilitated the sampling of a more compact configuration. The SASA of proteins has always been considered a decisive factor in protein folding and stability studies [16, 17]. It is defined as the surface characterized around a protein by a hypothetical center of a solvent sphere with the van der Waals contact surface of the molecule [18].

Table 3
The average SASA value of hydrophobic, hydrophilic, and total groups at the considered temperatures in the simulation

Temperature (K)	Hydrophilic	Hydrophobic	Total
300	56.02 ± 0.01	55.83 ±0.01	111.85 ±0.01
310	56.20 ± 0.01	56.36 ±0.01	112.56 ±0.01
320	56.81 ± 0.01	57.08 ±0.01	113.89 ±0.01
330	56.73 ± 0.01	56.45 ±0.01	113.19 ±0.01

2.3. Clusters in protein structures

The size and number of the structural clusters formed at each temperature can indicate the stability of the protein. Fewer clusters that contain more structures can indicate a greater protein stability under a particular set of conditions [19]. The number of clusters that exists for each temperature was 9, 22, 39, and 52 for the temperatures of 300 K, 310 K, 320 K, and 330 K, respectively. As the number of clusters increases by temperature increment, the stability of protein decreases.

There is a central member in each cluster which is the cluster centroid. Since cluster number 1 is the most popular cluster in each temperature and the most representative frame, the central member of this cluster shows the main configuration of protein. *Figure 3* depicts the configuration for each temperature.

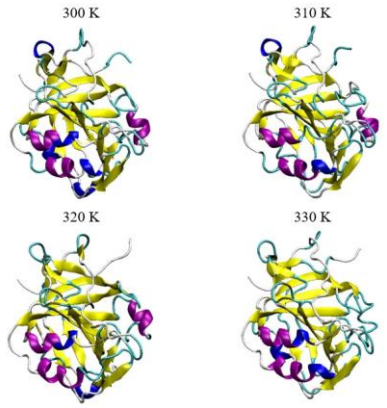

Figure 3
Different configuration of carbonic anhydrase according to main cluster at the considered temperatures. Each color shows the secondary structure of the protein. Yellow: β-sheet, white: random coil, cyan: β-turn, blue: α-helix, and purple: 3_{10}-helix.

Table 4 shows the first population of the three largest clusters and the percent of the total population of each cluster at each temperature. In 300 K there were 9,394 members in cluster number 1 which is almost 94% of the total structures that exist. This number for other temperatures was 6045, 3524, and 4238 at 310 K, 320 K, and 330 K, respectively.

Table 4
Variation in the number of members in each cluster for the first five clusters at the considered temperatures

Cluster	Number of members in each cluster (%)			
	300 K	310 K	320 K	330 K
1	9,394 (93.9)	6,045 (60.4)	3,524 (35.2)	4,238 (42.3)
2	286 (2.8)	1,765 (17.6)	1,722 (17.2)	996 (9.9)
3	118 (1.1)	903 (9.0)	1,147 (11.1)	871 .7)

2.4. Root Mean Square Deviation (RMSD)

The RMSD of carbonic anhydrase at 300 K, 310 K, 320 K, and 330 K is shown in *Figure 4*. At the starting point of the simulation, the RMSD increased to 0.11 nm for all temperatures. Thereafter, at 300 K and 330 K, the value fluctuated and reached the same values at the end of the interval as the beginning value. In contrast, for 310 K and 320 K, these figures had a burgeoning trend. In addition to the RMSD calculation, the standard error was calculated for these results. This figure for all the temperatures was around 0.0001 nm. According to *Table 5* that conntains the average RMSD, the highest value was for 320 K which had the highest value in the substantial proportion of the period.

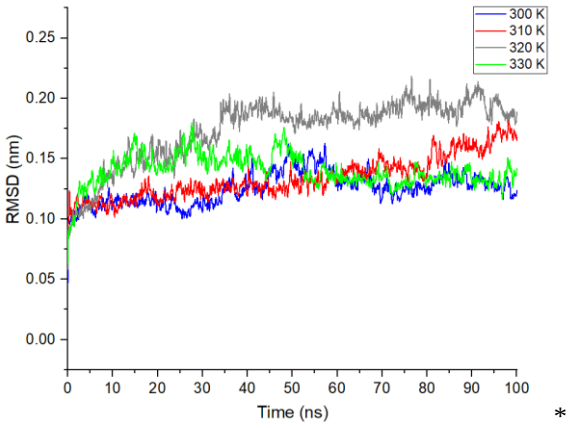

Figure 4
RMSD analysis for carbonic anhydrase at the four different considered temperatures

Table 5
Average amount of RMSD at the considered temperatures in MD simulation

Temperature (K)	300	310	320	330
Average RMSD (nm)	0.1250	0.1340	0.1740	0.1402

CONCLUSION

In this report, the thermal stability of carbonic anhydrase over the 100 ns trajectory was investigated by the MD simulation. The results of the simulation illustrated the higher dependence on the backbone H bonds on temperature compared to the side-chain hydrogen bond. The thermal stability of protein increased because of an increase in the hydrogen bond. The SASA calculation demonstrated the same fluctuation for both hydrophobic and hydrophilic groups at four different considered temperatures. Though the maximum value for the total solvent accessible surface area of protein was observed at 330 K at 120 nm^2. Moreover, the number of clusters increased by the temperature increment which is represented by the lower stability in higher temperatures. Regarding RMSD calculation, there were fluctuations at two temperatures, 320 K and 330 K, while this figure for two other temperatures increased over the time trajectory.

REFERENCES

[1] D. Keilin and T. Mann: Carbonic anhydrase. Purification and nature of the enzyme. *Biochemical Journal*, Vol. 34, No. 8–9, p. 1163, 1940.

[2] J. R. G. Bradfield: Plant carbonic anhydrase. *Nature*, Vol. 159, No. 4040, pp. 467–468, 1947.

[3] K. Okabe, S. Y. Yang, M. Tsuzuki, and S. Miyachi: Carbonic anhydrase: Its content in spinach leaves and its taxonomic diversity studied with anti-spinach leaf carbonic anhydrase antibody. *Plant Sci. Lett.*, Vol. 33, No. 2, 1984. https://doi.org/10.1016/0304-4211(84)90004-X

[4] M. Tsuzuki, S. Miyachi, and G. E. Edwards: Localization of carbonic anhydrase in mesophyll cells of terrestrial C3 plants in relation to CO2 assimilation. *Plant Cell Physiol.*, Vol. 26, No. 5, 1985, https://doi.org/10.1093/oxfordjournals.pcp.a076983.

[5] R. J. DiMario, J. C. Quebedeaux, D. J. Longstreth, M. Dassanayake, M. M. Hartman, and J. v Moroney: The cytoplasmic carbonic anhydrases β CA2 and β CA4 are required for optimal plant growth at low CO2. *Plant Physiol.*, Vol. 171, No. 1, pp. 280–293, 2016, https://doi.org/10.1104/pp.15.01990.

[6] N. A. Khan: Variation in carbonic anhydrase activity and its relationship with photosynthesis and dry mass of mustard. *Photosynthetica*, Vol. 30, No. 2, 1994, pp. 145–147, https://doi.org/10.1007/BF02879650.

[7] N. L. Reed: Carbonic anhydrase in plants: distribution, properties and possible physiological roles. 1981. In Progress in Phytochemistry, L. Reinhold, J.B. Harborne, and T. Swain, eds. (Oxford, U.K.: Pergammon Press), pp. 47–94.

[8] L. Bao and M. C. Trachtenberg: Facilitated transport of CO2 across a liquid membrane: comparing enzyme, amine, and alkaline. *J. Memb. Sci.*, Vol. 280, No. 1–2, pp. 330–334, 2006, https://doi.org/10.1016/j.memsci.2006.01.036.

[9] A. S. Bhown and B. C. Freeman: Analysis and status of post-combustion carbon dioxide capture technologies. *Environ. Sci. Technol.*, Vol. 45, No. 20, pp. 8624–8632, 2011, https://doi.org/10.1021/es104291d.

[10] G. M. Bond, J. Stringer, D. K. Brandvold, F. A. Simsek, M.-G. Medina, and G. Egeland: Development of integrated system for biomimetic CO_2 sequestration using the enzyme carbonic anhydrase. *Energy & Fuels*, Vol. 15, No. 2, pp. 309–316, 2001. https://doi.org/10.1021/ef000246p

[11] J. F. Domsic and R. McKenna: Sequestration of carbon dioxide by the hydrophobic pocket of the carbonic anhydrases. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, Vol. 1804, No. 2, pp. 326–331, 2010. https://doi.org/10.1016/j.bbapap.2009.07.025

[12] A. Gladis, M. T. Gundersen, P. L. Fosbøl, J. M. Woodley, and N. von Solms: Influence of temperature and solvent concentration on the kinetics of the enzyme carbonic anhydrase in carbon capture technology. *Chemical Engineering Journal*, Vol. 309, pp. 772–786, 2017, https://doi.org/10.1016/j.cej.2016.10.056.

[13] M. J. Abraham et al.: GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, Vol. 1, pp. 19–25, 2015, https://doi.org/10.1016/j.softx.2015.06.001.

[14] Y. Mao and Y. Zhang: Nonequilibrium molecular dynamics simulation of nanobubble growth and annihilation in liquid water. *Nanoscale and Microscale Thermophysical Engineering*, Vol. 17, No. 2, pp. 79–91, 2013. https://doi.org/10.1080/15567265.2012.760692

[15] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark: Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, Vol. 38, No. 1–2, pp. 236–240, 1999.

[16] S. Ausaf Ali, I. Hassan, A. Islam, and F. Ahmad: A review of methods available to estimate solvent-accessible surface areas of soluble proteins in the folded and unfolded states. *Curr. Protein Pept. Sci.*, Vol. 15, No. 5, pp. 456–476, 2014. https://doi.org/10.2174/1389203715666140327114232

[17] A. M. Lesk and C. Chothia: Solvent accessibility, protein surfaces, and protein folding. *Biophys. J.*, Vol. 32, No. 1, pp. 35–47, 1980. https://doi.org/10.1016/S0006-3495(80)84914-9

[18] S. Lu and A. S. Wagaman: On methods for determining solvent accessible surface area for proteins in their unfolded state. *BMC Res. Notes*, Vol. 7, No. 1, pp. 1–7, 2014. https://doi.org/10.1186/1756-0500-7-602

[19] J. Arunachalam and N. Gautham: Hydrophobic clusters in protein structures. *Proteins: Structure, Function, and Bioinformatics*, Vol. 71, No. 4, pp. 2012–2025, 2008, https://doi.org/10.1002/prot.21881.