

Shape measurement for cylindrical structures formed by tyrosine molecules

Cite as: AIP Conference Proceedings **2183**, 080003 (2019); <https://doi.org/10.1063/1.5136196>
Published Online: 06 December 2019

M. Gokhan Habiboglu, and Sahin Uyaver



View Online



Export Citation

ARTICLES YOU MAY BE INTERESTED IN

Sentiment analysis of Turkish Twitter data

AIP Conference Proceedings **2183**, 080004 (2019); <https://doi.org/10.1063/1.5136197>

Preface to “Computer Science and Technology”

AIP Conference Proceedings **2183**, 080001 (2019); <https://doi.org/10.1063/1.5136194>



Your Qubits. Measured.

Meet the next generation of quantum analyzers

- Readout for up to 64 qubits
- Operation at up to 8.5 GHz, mixer-calibration-free
- Signal optimization with minimal latency

Find out more

 Zurich Instruments

Shape Measurement for Cylindrical Structures Formed by Tyrosine Molecules

M. Gokhan Habiboglu^{1, a)} and Sahin Uyaver^{2, b)}

¹⁾Turkish-German University, Faculty of Engineering, Sahinkaya Cad. No: 108 34820 Beykoz Istanbul Turkey

²⁾Turkish-German University, Faculty of Science, Sahinkaya Cad. No: 108 34820 Beykoz Istanbul Turkey

^{a)}Corresponding author: habiboglu@tau.edu.tr

^{b)}Corresponding author: uyaver@tau.edu.tr

Abstract.

Tyrosine is one of the important aromatic amino acids, the wrong metabolization of which usually results in severe mental diseases. Based on the simulation results it was observed that tyrosine molecules form fibril-like structures at high temperatures. In order to have a better understanding of tyrosine self-assembly, we developed a quantitative measure to analyze fibril-like shapes formed by tyrosine molecules. This was applied to 4 different temperatures and then was compared with the analysis from simulation data. As expected, tyrosine molecules indeed exhibit fibril-like structures at 350 K at a fast rate, which is perfectly in agreement with the analysis in literature.

Keywords: Self-assembly, molecular dynamics, amino acids, clustering.

INTRODUCTION

A four-fold assembly is observed at high temperatures in molecular dynamics simulations of tyrosine [1]. Fibril-like structures formed in these simulations are important due to the role of tyrosine in brain-related diseases. In this current study we apply a mathematical method to measure the shapes of the structures obtained by molecular dynamics simulations of tyrosine molecules. That is, we know from the previous simulation studies [1, 2, 3] that tyrosine molecules form a well-formed four-fold structures at especially high temperatures like phenylalanine molecules. These four-fold structures are in the shape of a cylinder (Fig. 1). We develop a measure to estimate the cylinder-likeness of the structures formed by tyrosine molecules. To do this, we identify the structures quantitatively, using the produced data from [1].

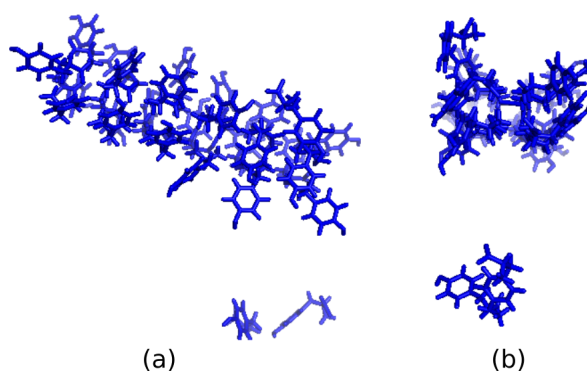


FIGURE 1. (a) 27 tyrosine molecules at 350K form a tube like structure at 350 Kelvin. (b) Profile view at right shows that it is a tubular structure.

Analyzing the data from simulations is mostly an important task. In the simulations of proteins and amino acids, a clustering algorithm is often used. In the previous studies [2, 3] the researchers have implemented a home-made script to analyze the structures, where they mainly followed the similar steps of Hoshen-Kopelman's algorithm [4]. In the literature the clustering works appear very often: Harder et al. [5] has implemented a method of clustering using Gauss integrals, which is based on the root mean square deviation. In this method the well-known K-means clustering is performed afterwards. It is noted that this method is significantly faster in time and also can cluster a big number of structures. Another method, called Local-Global Alignment, was designed to facilitate to find the similarities between

protein structures [6]. This method also enables us to cluster similar fragments of protein structures. In this method one needs an extensive analysis of regions of local similarities, so that our system does not require such an expensive analysis. Another code development is by de Hoon et al. [7]. They develop a clustering software library from other known algorithms and the software is available as open-source. Since our consideration focuses on fibril structures, this software bundle may need a special tuning. Another software package as being a new solution for protein 3D structure prediction is MUFOLD by Zhang et al. [8]. The authors of this paper developed a number of novel methods and combined them in a software package. Another clustering algorithm, which is a fundamental element of our application, is called Density-Based Spatial Clustering of applications with noise (DBSCAN) [9, 10]. This method is a density-based clustering method, where initially a set of points in some space is given, then closely packed points are grouped together, finally the points outside of the group are said to be present in low-density regions.

As we see, clustering of structures is a widely used method for different purposes. In our method, which is dedicated to the specific structures obtained by the simulations of tyrosine molecules, we focus on the occurring of such structures, and also the level of quality of these structures.

METHOD

Since tyrosine molecules form fibril-like structures, we need to describe the cylindrical shape quantitatively. In order to have an analytical insight for the likeliness of the self-assembly in tyrosine to a cylinder-like shape, we developed a geometric measure, which estimates the resemblance of the molecular formed structure to a perfect cylinder is. In our previous simulations 27 tyrosine molecules were simulated at 360 mM concentrations and the temperatures 275 K, 300 K, 325 K and 350 K were studied. We have chosen C- α atoms of tyrosine molecules as the center of masses instead of all atoms (Fig. 2(a)).

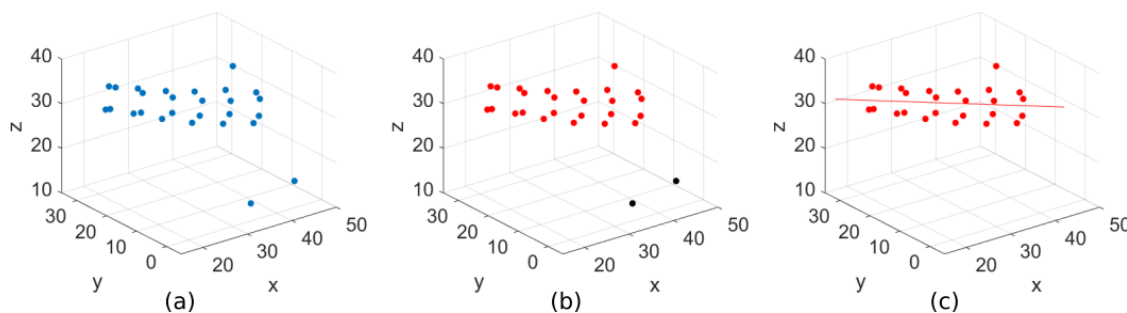


FIGURE 2. (a) Formation of C- α atoms at 350 K and at $t=300$ ns. (b) DBSCAN algorithm identifies largest cluster (red) (c) Line of best fit is estimated for largest cluster.

The details of the method are as follows: We first identify the largest cluster in the molecular conformation at each time step. For this purpose, we use the well-known DBSCAN algorithm as described in [9, 10] with the parameters $Eps = 8$ and $MinPts = 4$. These parameters are determined by trial and error and selected accordingly because these values are comparable with our clustering observations from the simulation data. If the number of molecules in the largest cluster (major cluster) is below 4, then the self-assembly at that particular time is considered insufficient and hence is not taken into consideration (Fig. 2(b)). For the sufficiently large major clusters, we first estimate the 3 dimensional line of best fit for all molecules (Fig. 2(c)). This line corresponds to the axis of a hypothetical cylinder covering all the molecules in the cluster. Then the standard deviation of distances from each molecule to this line is calculated. Note that this distance would be equal for each molecule in a perfect cylinder, accordingly its standard deviation becomes zero. For this reason, it provides us an analytical sense of understanding for the similarity of the shape formed by self-assembly to a perfect cylinder. The higher standard deviation is, the less likely the self-assembly forms a cylinder-like shape. Finally, this measure is divided with the ratio of molecules in the largest cluster to the total number of molecules. When the cluster size increases, this ratio approaches to one, which is the case where all molecules are included in the cylinder shaped structure, and the division equates to the standard deviation. Thus, this division particularly provides us indirectly to take the presence of isolated molecules into account.

Mathematical representation of this measure, named as cylinder likelihood measure, is as follows:

$$\phi = \frac{\sigma(d_i)}{R}, \quad (1)$$

where σ is standard deviation, d_i is the distance of i^{th} molecule to the line of best fit and R is the ratio of the number of molecules in largest cluster, n_c to the number of all molecules, n_{all} , $R = n_c/n_{\text{all}}$.

RESULTS

The evolution of ϕ with respect to time at 4 different temperatures is given in Fig. 3. In order to have a better understanding of the shape dynamics, the moving average of ϕ is also depicted. This figure clearly provides us that tyrosine molecules continuously tend to assemble into a cylinder-like shape at 350 K (Fig. 3(d)), because the moving average decreases over time. This result is completely in agreement with the literature [1] stating that tyrosine molecules form four-fold tube-like structures at $T = 350$ K. We also observe that the cylinder-like shape poorly exists in other temperatures, therefore ϕ does not reach down to low enough values which we obtained at $T = 350$ K. One of the reasons of this result can be attributed to the relatively high value of R at $T = 350$ K. The largest cluster formed through simulations at this temperature consists of 25 molecules. Additionally, the average of R for all frames at $T = 350$ K is found as 19.36, which is considerably higher than the average of R in other temperatures. Hence, we can state, even without considering the shape of formation, that the self-organization of tyrosine at $T = 350$ K is stronger as compared to other temperatures. Averages of ϕ between certain time intervals are given in Table I.

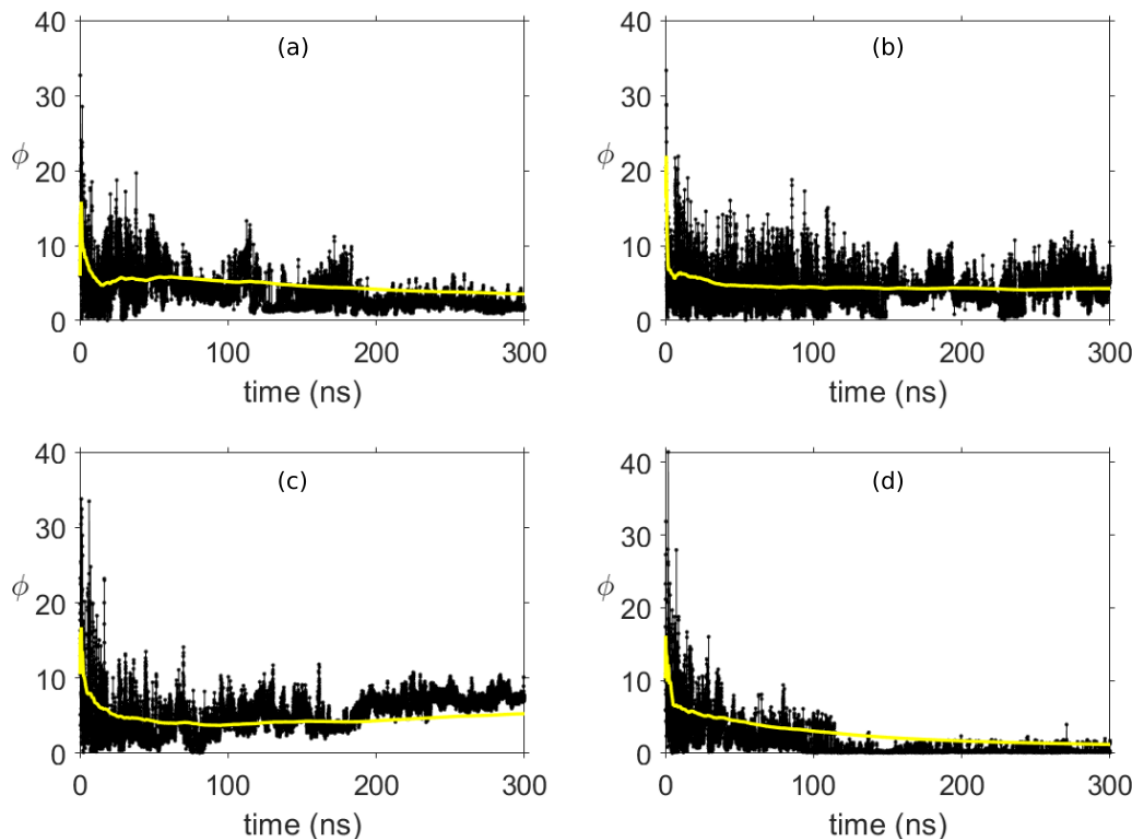


FIGURE 3. Evolution of cylinder likeliness measure and its moving average (yellow line) with respect to time for temperatures (a) 275 K, (b) 300 K, (c) 325 K and (d) 350 K.

Two important results are drawn from Table I, namely at which temperature tyrosine molecules forms a tube like structure and how fast they assemble into that structure. At $T = 350$ K, ϕ average after 150 ns has a significantly low

TABLE I. Time interval averages of ϕ at 4 different temperatures.

Time Interval (ns)	275 K	300 K	325 K	350 K
0-100	5.2011	4.4856	3.7888	3.0478
100-200	3.1816	4.1941	4.8701	0.3889
200-250	2.4117	3.2944	6.9573	0.2920
250-300	2.1198	5.0402	7.2646	0.2810

value as compared to other temperatures and even approaches to zero, which clearly shows that a cylindrical formation appears. Furthermore, as shown in Fig. 5 in [1] the self-assembly process occurs fastest at 350 K, which is also in agreement with our calculations, since the change in ϕ average is noticeably small after 100 ns at this temperature. In addition, we observe that ϕ values increase at 325 K. This phenomenon can be explained by looking at simulation results which indicate that a crystal-like shape is formed [1].

CONCLUSION

We conclude that our method can identify the cylindrical structures in principle. In our study we have studied the self-assembly of tyrosine molecules, which lead to the formation of cylindrical structures at some temperatures. The estimation of the formed structures are of high interest, therefore one can use an automated method to get information on the structures. We see that our method works especially well with distinguishing of cylindrical structures. Furthermore one gets an insight to the level of formation. This approach may lead us to understand the formation of the structures better, which is especially important both in health science and bio-technology. If the formed structures are far from being a cylindrical one, then the measurement result will be relatively larger than zero, so that one can conclude that different structures like crystal-like, sphere-like or dispersed one are formed.

REFERENCES

1. S. Uyaver, H.W. Hernandez and M.G. Habiboglu, "Self-assembly of aromatic amino acids: a molecular dynamics study", *Phys. Chem. Chem. Phys.*, 20, 30525-30536 (2018).
2. H.W. German, S. Uyaver and U.H.E. Hansmann, "Self-Assembly of Phenylalanine-Based Molecules", *Journal of Physical Chemistry A*, 119 (9), 1609-1615 (2015).
3. H.W. German, M. Bhavaraju, S. Uyaver and U. H. E. Hansmann, "Computational Insights Into the Self-Assembly of Phenylalanine-Based Molecules", *Task Quarterly* 18(4), 357-363 (2015).
4. J. Hoshen, R. Kopelman, "Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm", *Phys. Rev. B*, 14 (8), 3438-3445 (1976).
5. T. Harder, M. Borg., W. Boomsma, P. Rogen and T. Hamelryck, "Fast large-scale clustering of protein structures using Gauss integrals", *Bioinformatics*, 28 (4), 510-515 (2012).
6. A. Zemla, "LGA: a method for finding 3D similarities in protein structures", *Nucleic Acids Research*, 31 (13), 3370-3374 (2003).
7. M.J.L. de Hoon, S. Imoto, J. Nolan and S. Miyano, "Open source clustering software", *Bioinformatics*, 20 (9), 1453-1454 (2004).
8. J. Zhang, Q. Wand, B. Barz, Z. He, I. Kostin, Y. Shang and D. Zu, "MUFOLD: A new solution for protein 3D structure prediction", *Proteins*, 78 (5), 1137-1152 (2010).
9. M. Ester, H.P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226-231 (1996).
10. E. Schubert, J. Sander, M. Ester, H.P. Kriegel and X. Xu, "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN", *ACM Trans. Database Syst.* 42 (3), 1-21 (2017).