Proceedings of the 36th European Peptide Symposium
Michal Lebl (Editor)
European Peptide Society, 2022 *https://doi.org/10.17952/36EPS.2022.283*

Protein Structure Alignment and RMSD Calculations

Fatima I. Sapundzhi and Metodi S. Popstoilov
South-West University "Neofit Rilski", 2700 Blagoevgrad, Bulgaria

Introduction

In recent years, the alignment of protein structures is a fundamental problem for computational chemistry, computational structure biology and bioinformatics research. Alignment of protein structures can be performed as pairwise structure alignment (PSA) and multiple structure alignment (MSA). This task can be considered as an NP-hard problem. More information and reviews about this problem could be found in [1-5]. Initially, it is important to choose the way in which the protein structure will be represented - for example, it can be represented as a C-alpha-based distance map (C- map), secondary structure elements (SSEs) and other methods [6-7]. The next step is to consider the protein as rigid or flexible and choose a structure alignment method, which may consist of a scoring function to measure protein similarity and a search algorithm to optimize the function [8-9]. For more details of scoring function see Sapundzhi et al [10-11]. An important task for researchers is to evaluate the similarity of two protein structures. To compare the degree of similarity of two three-dimensional (3D) structures, the root means square deviation (RMSD) measure or the sum of squared "errors" (the squared distance between the corresponding points) is used. We use the RMSD function

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\delta_i^2} \quad (1)$$

where δ_i is the distance between atom i and either a reference structure or the mean position of the n equivalent atoms [12]. The purpose of this study is to present a simple procedure to calculate the RMSD between pairs of 3D structures and to find the minimum RMSD value. The implementation for the program is realized in the C# programming language.

Results and Discussion

In the current study, we present a program implemented in the C# programming language that can calculate the root mean square distance (RMSD) between two 3D structures [13,14]. The developed tool calculates minimal RMSD between two Cartesian coordinates in either. xyz or .pdb format (Protein Data Bank files). Both files (structures) must contain the same number of atoms in a similar order for the program to run correctly. The implementation of the program is based on the Kabsch algorithm [15,16] and Quaternion algorithm [17] for rotation. First, the structures are translated to the center of the coordinate system, then the covariance matrix and the optimal rotation matrix are calculated. The developed tool shows the RMSD calculated in three ways (Figure 1).

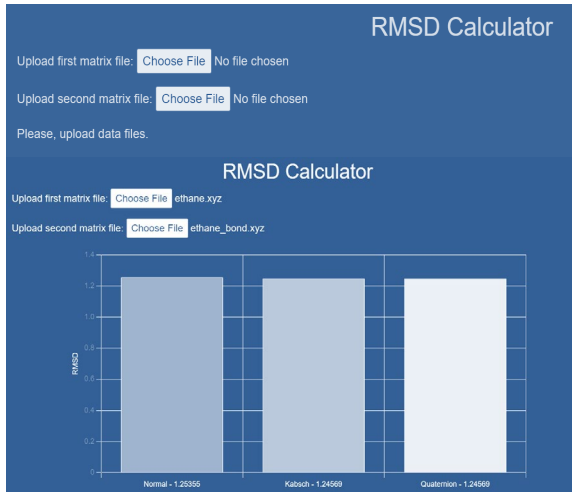

Fig.1. RMSD calculation between two structures. The tool accepts as arguments two files: .pdb and .xyz.

The results of the program implementation with the given example from https://github.com/ are shown as follows: Normal RMSD: 1.25, Translated RMSD: 1.24; Rotated RMSD: 1.245. The obtained results are compared with the similar tools developed in Python [18] and Perl [19] and show that for these implementations the algorithm gives almost identical results. The developed tool

has an easy-to-use graphical user interface that will be uploaded to a server and can be freely used by researchers in the field of bioinformatics research.

Acknowledgments

This paper is partially supported by SWU "Neofit Rilski" Project RPY-B4/19; RP-B7/20; Project, BNSF H27/36; National Scientific Program "Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICTinSES)", financed by the Ministry of Education and Science.

References

1. Eidhammer, I., Jonassen, I., Taylor, W.R. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* **7(5)**, 685-716 (2000), https://doi.org/10.1089/106652701446152
2. Koehl, P. *Reviews in Computational Chemistry* **22**, 1 (2006), https://doi.org/10.1002/0471780367.ch1
3. Brylinski, M., Skolnick, J. *Proceedings of the National Academy of Sciences* **105**(1), 129-134 (2008)
4. Ilinkin, I., Ye, J., Janardan, R. *BMC Bioinformatics* **11**, 71 (2010), https://doi.org/10.1186/1471-2105-11-71
5. Irving, J.A., Whisstock, J.C., Lesk, A.M. *Proteins: Structure, Function, and Bioinformatics* **42(3)**, 378-382 (2001), https://doi.org/10.1017/S0033583503003901
6. Jianzhu, Ma, Wang, S. *Advances in Protein Chemistry and Structural Biology* **94**, 121-75 (2014), https://doi.org/10.1016/B978-0-12-800168-4.00005-6
7. Holm, L., Rosenstrom, P. *Nucleic Acids Research* **38** (Suppl. 2), W545-W549 (2010), https://doi.org/10.1093/nar/gkq366
8. Taylor, W.R. *Comparing Secondary Structure 'Stick' Models of Proteins Using Graph Matching with Double Dynamic Programming* (Mewes, H, Seidel, H., Weiss, B. Eds) Bioinformatics and Genome Analysis. Ernst Schering Research Foundation Workshop, **38**. Springer, Berlin, Heidelberg, 2002, https://doi.org/10.1007/978-3-662-04747-7_7
9. Hasegawa, H., Holm, L. *Current Opinion in Structural Biology* **19 (**3) 341-348 (2009), https://doi.org/10.1016/j.sbi.2009.04.003
10. Sapundzhi, F., Prodanova, K. and Lazarova, M., *AIP Conference Proceedings* **2172**, 1, 100008 (2019), https://doi.org/10.1063/1.5133601
11. Sapundzhi, F. *International Journal of Online and Biomedical Engineering (iJOE)* **15(11)**, 139-145 (2019), https://doi.org/10.3991/ijoe.v15i11.10893
12. Topalska, R., Sapundzhi, F. *Journal of Chemical Technology and Metallurgy* **55(4)**, 714-718 (2020)
13. Coutsias, E., Seok, C., Dill, K. *J Comput Chem*. **25** (15), 1849-1857, (2004)
14. Armougom, F., Moretti, S., Keduas, V., Notredame, C. *Bioinformatics* **22** (14) e35-39 (2006), https://doi.org/10.1093/bioinformatics/btl218
15. W. Kabsch, *Acta Crystallographica Section A* **32**(5), 922-923, (1976), https://doi.org/10.1107/S0567739476001873
16. W. Kabsch, *Acta Crystallographica Section A* **34** (5), 827-828, (1978), https://doi.org/10.1107/S0567739478001680
17. Walker, M., Shao, L., Volz, R., *CVGIP: Image Understanding* **54 (3)**, 358-367 (1991), https://doi.org/10.1016/1049-9660(91)90036-O
18. http://github.com/charnley/rmsd
19. Sapundzhi, F., Slavov, V. *Journal of Chemical Technology and Metallurgy* **55(4)**, 935-938 (2020)