Hypothesis

A Local Spherical Coordinate System Approach to Protein 3D Structure Description

Wei Li [1] ID *

[1]   Institute of Special Environmental Medicine, Nantong University, No. 9, Seyuan Road, Nantong City, Jiangsu Province, P. R. China; wli148@aucklanduni.ac.nz

*   Correspondence: wli148@aucklanduni.ac.nz;

Abstract: To date, Cartesian (x, y, z) coordinate system (CCS) has been the default approach to geometrically specify atomic spatial positions in protein structures since the launch of Protein Data Bank (PDB) in 1971. To this end, this paper proposes a local spherical coordinate system (SCS) approach as an alternative to the default approach and a previously reported global SCS approach. The local SCS approach here requires only two parameters (ϑ and φ), instead of x, y and z as required by the default CCS approach. Essentially, CCS and SCS are like the two sides of one coin, i.e., geometric coordinate system for three-dimensional position specification. Therefore, this paper furthermore argues that it is time to flip the coin over, and have a look at the other side of the coin, e.g., the local SCS approach, which possesses an intrinsically lower degree of descriptional complexity than that of the default CCS approach, and constitutes a potentially useful alternative perspective for all protein structural research field.

1. Introduction

Protein Data Bank (PDB) is the primary database for experimentally determined biomolecular structures [1]. In PDB-format files, Cartesian coordinate system (CCS) is currently the default system (x, y, z) used to specify atomic positions in protein structures since the launch of PDB in 1971 [2,3]. In 2011, it was proposed for the first time (refereed to as the 2011 approach below) that protein structures be represented in spherical coordinates (ρ, ϑ and φ), aiming at expressing all PDB-deposited protein structures in spherical coordinates [4]. Essentially, this 2011 approach is a global spherical coordinate system (SCS) one, where the protein geometric centroid is taken as the unique original starting point for all atoms within a protein molecule. This 2011 approach results in two applications, i.e., the separation of the protein outer layer from its inner core, and the identification of protrusions and invaginations on the protein surface [4]. Similarly, this 2011 global SCS approach requires three parameters (ρ, ϑ and φ), as does the default 48-year-old CCS approach.

2. Theory

Since CCS and SCS are two geometrically interconvertible coordinate systems, they are like the two sides of one coin. This paper therefore proposes a local SCS approach, where the spatial position of an atom is specified theoretically by three parameters: ρ, ϑ and φ, representing radial distance (Å), polar angle (°), and azimuthal angle (°), respectively.

In practice, a standard model for the geometry of a molecule is constructed by representing each atom with a solid ball in three-dimensional space with the van der Waal radius of the corresponding atom. While this model is easy to generate, it has an obvious limitation: the van der Waal radius

34 of a given atom depends on its chemical environment and is not a universal constant. Furthermore,
35 conventional mechanical molecular model considers atoms as spheres and bonds as springs, where
36 the process of spring deformation can be linked to the ability of atomic bond to stretch, bend and twist.
37 Thus, the inter-nuclear distance at which the system energy minimum occurs defines the equilibrium
38 atomic bond length [5,6], which is taken as the value for ρ for the local SCS approach here.

3. An alternative local SCS approach

40 The local SCS approach here leads to a dimensionality shift (from three to two) in the number of
41 geometric parameters for protein structure description, i.e., boiling its geometric description down to
42 two parameters (ϑ and φ), instead of *x, y and z* as required by the default CCS approach, and instead of
43 ρ, ϑ and φ in the global SCS approach.
44 Take protein main chain structure as an example. Each amino acid residue's main chain are
45 constituted by N_N (amide nitrogen), C_α and the carboxyl carbon (with a double-bonded oxygen atom),
46 i.e., N-C_α-C=O. To simplify protein structure description, removing all side chains leaves only the
47 backbone atoms, i.e., three main chain atoms and another three backbone atoms, H_α, H_N (amide
48 hydrogen) and O=C (the carboxyl oxygen with a double-bonded carbon). Thus, the protein main chain
49 geometry can be abstracted as a linear sequence, -[N_i-$C_{\alpha i}$-CO_i-N_{i+1}-$C_{\alpha i+1}$-CO_{i+1}]-, where $1 < i < n$, n
50 represents the total number of amino acid residues in a protein. In this linear atom sequence, with an
51 arbitrary position of the starting atom (j) specified in the local SCS approach, the next ($j + 1$) atom's
52 position can be specified with three parameters (ρ, ϑ and φ), where ρ is the equilibrium atomic bond
53 length of atom (j) and atom ($j + 1$). Under this geometric arrangement, atom (j) and atom ($j + 1$)
54 constitute a vector in a three-dimensional space from atom (j) as the beginning point to atom ($j + 1$)
55 as the ending point, where $1 < j < m$, m represents the number of atoms in a protein main chain.
56 Similarly, the rest of the protein main chain structure can be specified in a way that each and every bit
57 of the three-dimensional structural information is encoded in the two local SCS parameters (ϑ and φ).
58 Given that each atom is covalently bonded to at least one neighbouring atom within a protein
59 structure, the local SCS approach here is also applicable to a protein structure as a whole, such that
60 two parameters (ϑ and φ) will be sufficient to geometrically define the three-dimensional structure
61 of a protein. Moreover, all PDB-deposited protein structures are able to be converted to the local
62 SCS format (ϑ and φ) with the current default CCS format (*x, y and z*), since CCS and SCS are two
63 geometrically interconvertible coordinate systems.

4. Summary

65 This article proposes an local SCS approach to establish a two-parameter description of protein
66 three-dimensional structure, as an alternative to the global SCS approach and the 48-year-old default
67 CCS approach [2,3]. This local SCS approach presents an alternative angular perspective for the
68 description of protein conformational flexibility, both main chain and side chain. Furthermore, this
69 local SCS approach also constitutes a potentially useful alternative perspective for protein structure
70 tree-decomposition, which can be of use for protein structural/domain classification, protein threading
71 in both backbone and side chain structure [7], and functional prediction of the *proteios* building block
72 of life [8–10].

5. Discussion

74 While protein structures change during evolution in response to mutations, they are three to
75 ten times more conserved than their amino acid sequences [11]. While some genetic mutations are
76 probably more perturbative of protein structure than others and are able to cause major structural
77 rearrangements of protein structures, it is highly likely that the net structural consequence of a genetic
78 mutation is the instalment of a different side chain (a side chain placement problem [12,13]) at the
79 same amino acid residue position, while the overall architecture of the protein remains unchanged.
80 In light of the fact that a side chain placement problem [12,13] is essentially an assignment of a set of

angles for each amino acid residue, only the local SCS approach exhibits its uniqueness here, neither the default CCS approach, nor the global SCS approach, because only the local SCS approach takes the equilibrium atomic bond length as the value for ρ in protein structure description. In other words, the degree of conservation of protein molecules (either sequence or structure) is linked to the simplicity of the way we approach them, either as a set of points/atoms/nuclei (*x, y and z*, i.e., the default CCS approach), or as a globular object with one geometric centroid taken as the unique original starting point for all atoms (the global SCS approach), or the local SCS approach here with a dimensionality shift from three to two in the number of geometric parameters necessary.

Finally, it is reasonable to not rule out the possibility that this local SCS approach will find its application in fields such as protein structure alignment [14], comparison [15,16], clustering [17–19], and structural classification and functional prediction from an evolutionary point of view [20–24].

Author Contributions: Cconceptualization, W.L.; methodology, W.L.; software, W.L.; validation, W.L.; formal analysis, W.L.; investigation, W.L.; resources, W.L.; data curation, W.L.; writing–original draft preparation, W.L.; writing–review and editing, W.L.; visualization, W.L.; supervision, W.L.; project administration, W.L.; funding acquisition, not applicable.

Conflicts of Interest: The author declares no conflict of interest.

1. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* **2003**, *10*, 980–980. doi:10.1038/nsb1203-980.
2. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. In *Methods in Molecular Biology*; Springer New York, 2017; pp. 627–641. doi:10.1007/978-1-4939-7000-1_26.
3. Crystallography: Protein Data Bank. *Nature New Biology* **1971**, *233*, 223–223. doi:10.1038/newbio233223b0.
4. Reyes, V.M. Representation of protein 3D structures in spherical (ρ, φ, ϑ) coordinates and two of its potential applications. *Interdisciplinary Sciences: Computational Life Sciences* **2011**, *3*, 161–174. doi:10.1007/s12539-011-0099-0.
5. Batsanov, S. Calculation of van der Waals radii of atoms from bond distances. *Journal of Molecular Structure: THEOCHEM* **1999**, *468*, 151 – 159. doi:https://doi.org/10.1016/S0166-1280(99)00077-9.
6. Caine, B.A.; Bronzato, M.; Popelier, P.L.A. Experiment stands corrected: accurate prediction of the aqueous pKa values of sulfonamide drugs using equilibrium bond lengths. *Chemical Science* **2019**, *10*, 6368–6381. doi:10.1039/c9sc01818b.
7. Xu, J.; Jiao, F.; Berger, B. A tree-decomposition approach to protein structure prediction. 2005 IEEE Computational Systems Bioinformatics Conference (CSB'05). IEEE, 2005. doi:10.1109/csb.2005.9.
8. Dawson, N.; Sillitoe, I.; Marsden, R.L.; Orengo, C.A. The Classification of Protein Domains. In *Methods in Molecular Biology*; Springer New York, 2016; pp. 137–164. doi:10.1007/978-1-4939-6622-6_7.
9. Sillitoe, I.; Dawson, N.; Thornton, J.; Orengo, C. The history of the CATH structural classification of protein domains. *Biochimie* **2015**, *119*, 209–217. doi:10.1016/j.biochi.2015.08.004.
10. Das, S.; Dawson, N.L.; Orengo, C.A. Diversity in protein domain superfamilies. *Current Opinion in Genetics & Development* **2015**, *35*, 40–49. doi:10.1016/j.gde.2015.09.005.
11. Illergård, K.; Ardell, D.H.; Elofsson, A. Structure is three to ten times more conserved than sequence-A study of structural response in protein cores. *Proteins: Structure, Function, and Bioinformatics* **2009**, *77*, 499–508. doi:10.1002/prot.22458.
12. Santana, R.; Larrañaga, P.; Lozano, J.A. Side chain placement using estimation of distribution algorithms. *Artificial Intelligence in Medicine* **2007**, *39*, 49–63. doi:10.1016/j.artmed.2006.04.004.
13. Fahmy, A.; Wagner, G. Optimization of van der Waals Energy for Protein Side-Chain Placement and Design. *Biophysical Journal* **2011**, *101*, 1690–1698. doi:10.1016/j.bpj.2011.07.052.
14. Zhang, Y. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research* **2005**, *33*, 2302–2309. doi:10.1093/nar/gki524.

15. Betancourt, M.R.; Skolnick, J. Universal similarity measure for comparing protein structures. *Biopolymers* **2001**, *59*, 305–309. doi:10.1002/1097-0282(20011015)59:5$<$305::aid-bip1027$>$3.0.co;2-6.

16. Ortiz, A.R.; Strauss, C.E.; Olmea, O. MAMMOTH (Matching molecular models obtained from theory): An automated method for model comparison. *Protein Science* **2009**, *11*, 2606–2621. doi:10.1110/ps.0215902.

17. Suryanto, C.H.; Saigo, H.; Fukui, K. Structural Class Classification of 3D Protein Structure Based on Multi-View 2D Images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2018**, *15*, 286–299. doi:10.1109/tcbb.2016.2603987.

18. Baran, D.; Pszolla, M.G.; Lapidoth, G.D.; Norn, C.; Dym, O.; Unger, T.; Albeck, S.; Tyka, M.D.; Fleishman, S.J. Principles for computational design of binding antibodies. *Proceedings of the National Academy of Sciences* **2017**, p. 201707171. doi:10.1073/pnas.1707171114.

19. Holmes, I. Faculty of 1000 evaluation for End-to-end differentiable learning of protein structure., 2018. doi:10.3410/f.732717449.793543371.

20. Figueroa, M.; Sleutel, M.; Vandevenne, M.; Parvizi, G.; Attout, S.; Jacquin, O.; Vandenameele, J.; Fischer, A.W.; Damblon, C.; Goormaghtigh, E.; Valerio-Lepiniec, M.; Urvoas, A.; Durand, D.; Pardon, E.; Steyaert, J.; Minard, P.; Maes, D.; Meiler, J.; Matagne, A.; Martial, J.A.; de Weerdt, C.V. The unexpected structure of the designed protein Octarellin V.1 forms a challenge for protein structure prediction tools. *Journal of Structural Biology* **2016**, *195*, 19–30. doi:10.1016/j.jsb.2016.05.004.

21. Spencer-Smith, R.; O'Bryan, J.P. Direct inhibition of RAS: Quest for the Holy Grail? *Seminars in Cancer Biology* **2019**, *54*, 138–148. doi:10.1016/j.semcancer.2017.12.005.

22. Houk, K.N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Accounts of Chemical Research* **2017**, *50*, 539–543. doi:10.1021/acs.accounts.6b00532.

23. Murakami, Y.; Jones, S. SHARP2: protein-protein interaction predictions using patch analysis. *Bioinformatics* **2006**, *22*, 1794–1795. doi:10.1093/bioinformatics/btl171.

24. Deng, H.; Jia, Y.; Zhang, Y. Protein structure prediction. *International Journal of Modern Physics B* **2018**, *32*, 1840009. doi:10.1142/s021797921840009x.