# Protein Structure Description with ρ, θ and ϕ: A Case Study with Caenopore-5

Wei Li

Contrebola Institute of Computational Interstructural Biophysics,
No. 88, Renaissance East Road, Nantong City, 226000,
Jiangsu Province, People's Republic of China

* Corresponding email: wli148@aucklanduni.ac.nz

## Abstract

Since its establishment in 1971, the Protein Data Bank (PDB) has been using Cartesian coordinate system (CCS) as the standard framework for protein structure description with $x$, $y$, $z$. Despite the interconvertibility of CCS and spherical coordinate systems (SCS, ρ, θ and ϕ), CCS remains to date the default and the only framework for protein structure description in PDB. Recent advances in protein structure prediction (e.g., AlphaFold) revolutionized the field by integrating deep learning algorithms with experimental structural data, achieving unprecedented accuracy of protein structure prediction and relying on Cartesian representation of protein structures to extract geometric features, such as inter-residue distances and dihedral angles, which are essential for building and visualizing protein structures with $x$, $y$, $z$. To this end, questions remain about the potential for further improvement in the performance of protein structure prediction, i.e., what is going to drive the next stage of its continued development with improved accuracy and efficiency. Therefore, this article introduces an alternative coordinate system for protein structure description and feature extraction. Using Caenopore-5 as an example, this article redefines protein backbone structures using atomic bonding networks (ABN) within the SCS framework (ABN-SCS), leading to the extraction of a set of spherical parameters (ρ, θ and ϕ) from the NMR ensemble of Caenopore-5, encompassing 477 covalent bonds and 80 peptide bonds within its backbone for each structural model in its NMR ensemble. Overall, this work offers an alternative spherical framework (ABN-SCS), i.e., a new language (ρ, θ, ϕ), for protein structure description and ABN-SCS-based extraction of spherical geometric features for protein structure prediction, aiming to contribute a little bit to its performance and application in drug discovery and design.

## Keywords

Protein Data Bank; Spherical Coordinate System; Atomic Bonding Network-based Spherical Coordinate System; Protein structure prediction; Feature extraction.

# Introduction

## [$x$, $y$, $z$] and [ρ, θ, ϕ] for protein structure description

In 1971, Protein Data Bank (PDB) was established as the single global repository for experimentally determined biomolecular 3D structures [1–4]. Since then, PDB has been acting as a catalyst [5] for structural biology, i.e., understanding [6–8], determination [9–12], description [13–18] and prediction [19–24] of the structure of protein, the essential building block of life. To this end, protein structure determination and description are like the two sides of a coin, both essential for understanding its function, and uncovering disease mechanisms caused by structural abnormalities, which drive advancements in protein engineering and structural modeling for drug discovery and design [25–27].

Since 1971 [1, 2], Cartesian coordinate system (CCS) has been the default framework (Figure 1) to describe protein 3D structure, i.e., the specification of spatial atomic positions with Cartesian coordinates [$x$, $y$, $z$]. In biochemistry, the primary structure of protein is described with an alphabet of twenty letters, each letter representing one standard amino acid, and the secondary structure of protein a mini alphabet of three letters for α-helix, β-strand and random coil, and the tertiary and the quaternary structures of protein a set of [$x$, $y$, $z$]. That is, for over half a century [28], [$x$, $y$, $z$] has been used as the standard language (i.e, the default Cartesian coordinate system) for protein structure description, reflecting an intuitive simplification and conceptualization of protein 3D structure as a set of discretely distributed points in 3D space, where each point corresponds to an atom in the protein, and its position is defined by the [$x$, $y$, $z$]. In this intuitive simplification and conceptualization with [$x$, $y$, $z$], nonetheless, one key aspect is not taken into account, i.e., the fact that structurally, protein is a set of chemically bonded points in 3D space, consisting of an intrinsic inter-atomic bonding network, instead of a set of discretely distributed points.

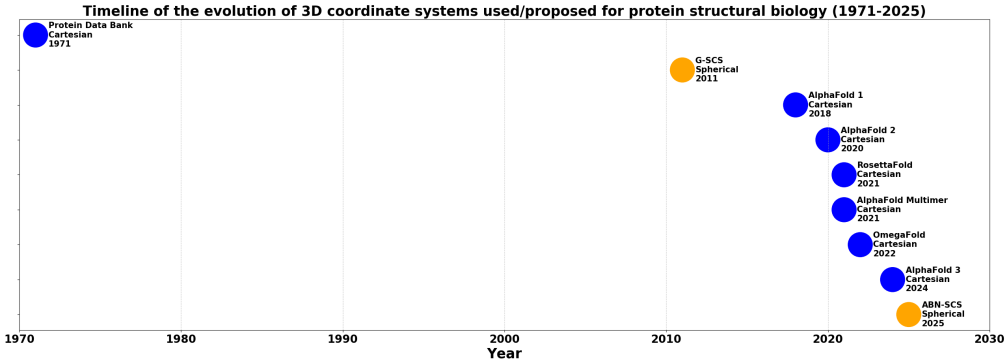


**Fig 1.** Over half-a-century burial of [ρ, θ, ϕ] in the Protein Data Bank [2, 3, 5].

In geometry, CCS ([$x$, $y$, $z$]) and spherical coordinate system (SCS, [ρ, θ, ϕ]) are interconvertible, both applicable in the specification of atomic positions [17, 18, 28]. In 2011, for instance, it was proposed for the first time that protein structure be represented in spherical coordinates [14, 15], where a global spherical coordinate system (G-SCS, Figure 1) was proposed and used for the separation of the protein outer layer from its inner core, and for the identification of protrusions and invaginations on the protein surface [14, 15]. By definition, G-SCS (Figure 1) requires three parameters ($R$, θ and ϕ), and takes the geometric centroid (an imaginary point) of the protein structure as origin ([0, 0, 0]), where $R$ is used to signify the distance between the imaginary centroid and any atom of the protein [14, 15].

## A Cartesian Protein Data Bank for protein structure prediction

As of January 8, 2025, more than 229,000 biomolecular structures have been determined and deposited in PDB [29, 30]. These structures were predominantly resolved using experimental techniques such as X-ray crystallography [9, 12], NMR spectroscopy [10, 12], cryo-electron microscopy [11, 12, 31], etc [12, 32]. Nonetheless, 229,000 represents only a small fraction of the billions of known protein sequences [33]. To address this disparity, computational tools become indispensable for large-scale structural bioinformatics, driving advances in protein structure prediction in the past few years [23, 34–39]. For instance, AlphaFold (Figure 1) [34] and RoseTTAFold (Figure 1) [35] transformed protein structure prediction through the increasing adoption of deep learning methodologies [33, 40–42], which use neural networks to replace traditional energy models and sampling procedures, resulting in significant improvements in prediction accuracy [23, 24, 43–46].

As described above, the Protein Data Bank that we have now is essentially a Cartesian Protein Data Bank due to CCS's long-standing role as the default system for specifying atomic positions with [$x$, $y$, $z$] [1, 2]. As the primary repository of protein structure data, PDB [1, 2] serves as the foundation for the training and the validation of most protein structure prediction tools [23, 34–39]. Take AlphaFold (Figure 1) [34] for instance, which takes Cartesian-format PDB files as input for the extraction of geometric features, such as inter-residue distances and dihedral angles. Subsequently, along with the protein structure data in PDB, these structural features are plugged into deep learning algorithm(s) for the prediction of discrete probability distributions of inter-residue distances, resulting in an $L \times L$ distance map for the protein of interest [33, 40, 47]). This $L \times L$ distance map serves as the foundation for constructing Cartesian format ([$x$, $y$, $z$]) structural models, subsequent visualization with PyMol [48] or VMD [49], or further analysis and use in computational drug discovery and design [25–27].

# Aim

Traditional protein structural modeling in the 1980s is based on known structural templates [50–52], followed by advancements in molecular dynamics and energy minimization techniques [53]. Recent years saw a paradigm shift with the advent of AlphaFold, which leverages deep learning to achieve unprecedented accuracy in predicting protein structures [20–22, 34, 54, 55]. However, questions remain about the potential for further improvement in the performance (e.g., accuracy, efficiency) of protein structure prediction, i.e., what is going to drive the next stage of the development in this field. This article, therefore, reports an atomic bonding network-based spherical coordinate system (ABN-SCS) [16–18] for protein structure description with [$\rho$, $\theta$, $\phi$], instead of [$x$, $y$, $z$]. Aiming to provide an alternative framework rooted in the intrinsic geometry of atomic bonding network of protein structure, this article builds on this ABN-SCS framework, using Caenopore-5 [56–58] as an example to demonstrate the use of ABN-SCS in redefining protein structure ([$\rho$, $\theta$, $\phi$]) and extracting spherical structural features, in addition to those Cartesian structural features used by current protein structure prediction tools [59–65].

# Materials and Methods

## Two geometrically inter-convertabile coordinate systems

This section is described in the first and the second sections of supplemtary file **S2.pdf**.

## An ABN-SCS description of the structures of twenty stand-alone natural amino acids

This section is the same as the third section of supplemtary file **S2.pdf**, and is described in supplemtary file **S1.pdf**.

## An ABN-SCS description of protein backbone structure: a case study with a model peptide with five amino acids

This section is described in the fourth section of supplemtary file **S2.pdf**.

## An ABN-SCS description of protein peptide bond structure: a case study with a model peptide with five amino acids

This section is described in the fifth section of supplemtary file **S2.pdf**.

## An ABN-SCS description of protein backbone and peptide bond structures: Caenopore-5 as an example

To illustrate the redefinition of protein backbone structure with ρ, θ and ϕ, this article takes Caenopore-5 (Figure 1) as an example, which adopts a helical conformation, forming oligomeric pores upon interaction with lipid bilayers, with five helices stabilized by three disulfide bridges (Figure 3).

**Fig 2.** Amino acid sequence and secondary structure of Caenopore-5 (81 amino acids) [56–58]. This figure was prepared by the PSIPRED server [66].

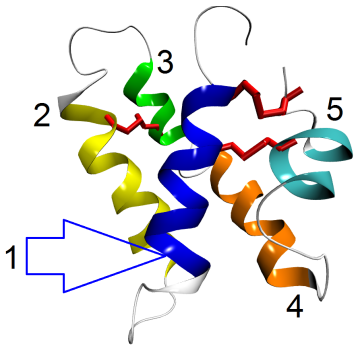


**Fig 3.** Three-dimensional structure of Caenopore-5 (PDB ID: 2JSA) stabilized by three disulfide bonds (red sticks) [56–58]. This figure is prepared by VMD [49].

As the structure of Caenopore-5 was determined by liquid-state NMR spectroscopy [56, 57], the structural data deposited in PDB [2] is an NMR ensemble of 15, instead of 1, structural models calculated with restraints from experimental NMR data. As a result, the redefinition of the backbone structure of Caenopore-5 with [ρ, θ, ϕ] is for all 15 structural models [56–58]. Specifically, the backbone structure of Caenopore-5 is described with [ρ, θ, ϕ] according to the fourth section of supplemtary file **S2.pdf**, and the peptide bond structure of Caenopore-5 is described with [ρ, θ, ϕ] according to the fifth section of supplemtary file **S2.pdf**.

## Results

This article used Caenopore-5 [56–58] as an example to define an alternative framework, i.e., ABN-SCS, for protein structure description with [ρ, θ, ϕ] [67] and the extraction of spherical structural features in addition to those Cartesian structural features used by current protein structure prediction tools [59–65]. In short, for Caenopore-5 [56–58], two sets of spherical parameters (ρ, θ and ϕ) were extracted from its NMR ensemble (PDB ID: 2JSA) and summarized in supplemtary file **S4.pdf**. Specifically, for the backbone structure of Caenopore-5 [56–58], a set of spherical parameters (a $3 \times 477$ matrix) were extracted from its NMR ensemble (PDB ID: 2JSA) and visualized with Figures 1, 2 and 3 in supplemtary file **S3.pdf** for ρ, θ and ϕ, respectively; And for the peptide bond structures of Caenopore-5 [56–58], a set of spherical parameters (a $3 \times 80$ matrix) were extracted from its NMR ensemble (PDB ID: 2JSA) and visualized with Figures 4, 5 and 6 in supplemtary file **S3.pdf** for ρ, θ and ϕ, respectively. Further details of the results and analysis of them are included in supplementary file **S3.pdf**.

## Conclusion

To sum up, this article used Caenopore-5 [56–58] as an example for protein 3D structure description with a spherical coordinate system based on the intrinsic inter-atomic bonding (ABN-SCS) network of protein. Geometrically, this spherical ABN-SCS framework is an uncharted territory, different from the default Cartesian approach since 1971 [1–4], in the sense that ABN-SCS (ρ, θ and ϕ) not only is a new language for protein structure description [13–18] and understanding [6–8], but also paves a new way for the extraction of protein structure features beyond conventional Cartesian feature extraction (e.g., inter-atomic and inter-residue distances) by computational tools such as AlphaFold [34] and RoseTTAFold [35].

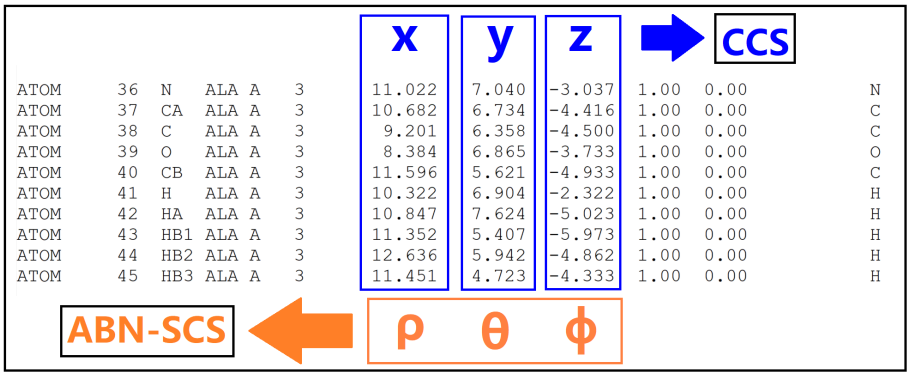


**Fig 4.** Towards a spherical Protein Data Bank in addition to its Cartesian counterpart [1, 2]. This figure represents a simplified example of the PDB-format ATOM record with Cartesian coordinates ([$x$, $y$, $z$]) replaceable by ABN-SCS coordinates [ρ, θ, ϕ].

Essentially, this spherical ABN-SCS approach is more of a geometric algorithm, through which spherical structural features (ρ, θ and ϕ) of protein can be extracted from the entire collection of experimentally determined protein structures inside PDB [1–4]. Therefore, this article here proposes a spherical Protein Data Bank on top (Figure 4) of its 54-year-old Cartesian counterpart [1–4], such that additional spherical structural features are extracted and can be incorporated into deep learning algorithms for continued improvement of the performance of protein structure prediction tools [19–22].

# Discussion

## [ρ, θ, ϕ]: a new language for protein structure description

As mentioned above, since 1971 [1,2], [$x$, $y$, $z$] has been used as the standard language (i.e, the default Cartesian coordinate system) for protein structure description (Figure 1), reflecting an intuitive simplification and conceptualization of protein 3D structure as a set of discretely distributed points in 3D space, where each point corresponds to an atom in the protein, and its position is defined by the [$x$, $y$, $z$]. From a linguistic point of view, language shapes and constrains thought, according to the Sapir-Whorf Hypothesis [68]. In protein structural biology, thus, there is an interdependence between protein structure understanding [6–8] and protein structure description [13–18], which is mirrored in the relationship between signifier (description) and signified (understanding). Therefore, the tools and terminologies used to describe protein structures shape how scientists conceptualize molecular configurations with coordinate systems (e.g., CCS [1], G-SCS [14,15], ABN-SCS [16–18]). These descriptive frameworks not only represent our understanding but actively influence it by focusing our attention on certain structural aspects while potentially obscuring others.

For instance, here, as the standard Cartesian language, [$x$, $y$, $z$] offers a Cartesian perspective on how we understand and describe protein structure. In addition to this Cartesian perspective, nonetheless, [$x$, $y$, $z$] constrains our understanding and description of protein structure with an intuitive Cartesian simplification and conceptualization, where one key aspect is not taken into account, i.e., the fact that structurally, protein is a set of chemically bonded points in 3D space, consisting of an intrinsic inter-atomic bonding network, instead of a set of discretely distributed points, which is one of the reasons why this article aims to redefine protein structure with an atomic bonding networks (ABN) within the SCS framework (ABN-SCS). Moreover, protein structural feature extraction [69,70], particularly through spherical feature extraction ([ρ, θ, ϕ]) [67], offers a new paradigm for the integration of this ABN-SCS framework with existing protein structure prediction tools, improving both accuracy and computational efficiency [17].

## An ABN-SCS definition of covalent bond length with ρ

In the past few years, while AlphaFold has significantly advanced protein structure prediction with homologous single-chain sequences [34], it has limitations in its performance while handling complex systems and dynamic interactions, which underscore the need for continued development in this field. Here, this article proposes an ABN-SCS-based protein structure description, which incorporates chemically accurate constraints to avoid unrealistic bond stretching or compression in protein structure prediction, because ABN-SCS encodes protein structures with spherical coordinates [ρ, θ, ϕ], where ρ represents equilibrium bond length, i.e., the inter-nuclear distances at which system energy is minimized [71,72].

By constraining the value of ρ to the equilibrium bond lengths in a spherical coordinate system, this ABN-SCS framework enhances the realism of the predicted protein structure by maintaining chemically accurate distances between covelently bonded atoms. To this end, the ABN-SCS framework ensures that the system is aligned with chemical bonding principles, making sure that inter-atomic distances reflect the energetically favored configurations of atoms in protein structure, which could be particularly relevant in the context of force-field-based energy minimization in NMR data-driven structural calculation or existing protein structure prediction algorithms [69,70,73,74].

## Covalent Bond Length in ABN-SCS: is ρ a real constant?

By definition, Cartesian Coordinate System (CCS) requires three parameters, i.e., [$x$, $y$, $z$], to describe protein structures. Here, in this article, the atomic bonding network-based spherical coordinate system (ABN-SCS) also requires three parameters, i.e., [ρ, θ, ϕ], for protein structure description. In theory, ρ is a physical constant, since it is defined as the equilibrium bond length—a physical constant representing inter-nuclear distances that minimize system energy [71, 72]. Thus, the ABN-SCS framework proposed here facilitates a dimensionality reduction by transforming a three-dimensional matrix ([ρ, θ, ϕ]) into a two-dimensional spherical vector comprising polar (θ) and azimuthal (ϕ) angles. As a result, despite using only two spherical parameters, ABN-SCS is able to theoretically preserve the structural information encoded by [$x$, $y$, $z$] in the Cartesian framework [1–4].

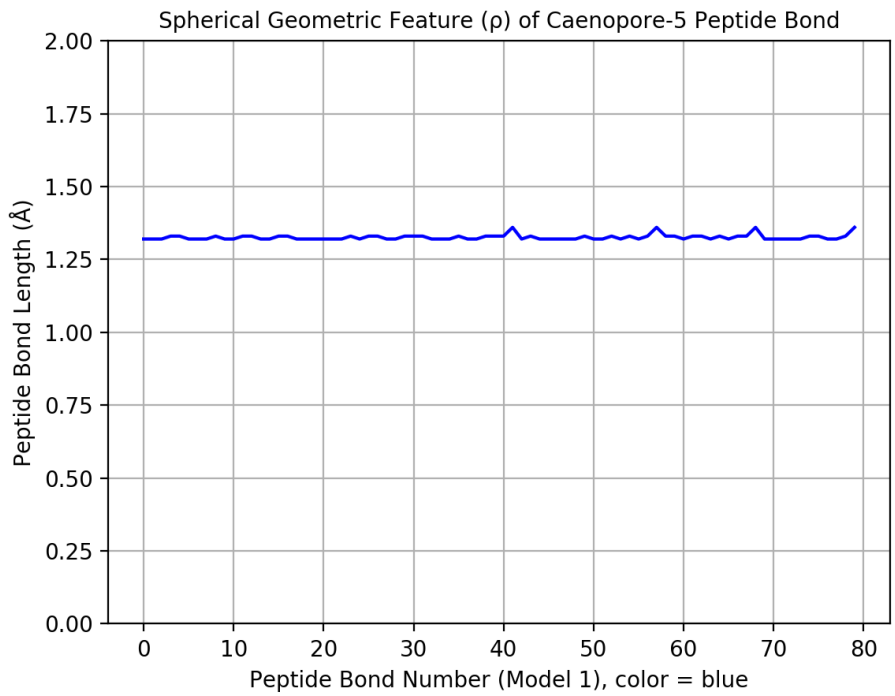


**Fig 5.** Distribution of the lengths (ρ) of 80 peptide bonds in Caenopore-5 [56–58]. This data corresponds to the first structural model in the NMR ensemble (PDB ID: 2JSA), depicted as the top-left subfigure in Supplementary File **S3.pdf**.

In practice, however, ρ is not a truly fixed value [71, 72]. By definition in the ABN-SCS framework, ρ is the equilibrium bond length, i.e., the distance at which the attractive and repulsive forces between two nuclei are balanced, resulting in the lowest energy configuration [71, 72]. In fact, the value of the equilibrium bond length is highly sensitive to environmental conditions, in a way similar to how atomic nucleus chemical shift (i.e., Larmor or precessional frequency) is sensitive to chemical environmental conditions. For example, factors such as temperature [75, 76] can influence atomic vibrations or bond stretching, while pH variations [58, 77–80] can alter protonation states and electrostatic interactions within protein structure. Moreover, the composition of the buffer and the ionic strength [81] of the surrounding medium can affect the local environment, influencing the forces between atoms and slightly altering the bond

lengths. Taken together, ρ varies dynamically depending on these environmental conditions, rather than remaining constant under all circumstances.

| Model | Mean (Å) | Std (Å) | Min (Å) | Max (Å) | Std/Mean |
|---|---|---|---|---|---|
| 1 | 1.3260 | 0.0092 | 1.3200 | 1.3600 | 0.0070 |
| 2 | 1.3266 | 0.0091 | 1.3200 | 1.3600 | 0.0069 |
| 3 | 1.3265 | 0.0092 | 1.3200 | 1.3600 | 0.0069 |
| 4 | 1.3261 | 0.0092 | 1.3200 | 1.3600 | 0.0069 |
| 5 | 1.3272 | 0.0090 | 1.3200 | 1.3600 | 0.0068 |
| 6 | 1.3261 | 0.0092 | 1.3200 | 1.3600 | 0.0069 |
| 7 | 1.3262 | 0.0092 | 1.3200 | 1.3600 | 0.0069 |
| 8 | 1.3262 | 0.0092 | 1.3200 | 1.3600 | 0.0069 |
| 9 | 1.3274 | 0.0090 | 1.3200 | 1.3600 | 0.0068 |
| 10 | 1.3267 | 0.0091 | 1.3200 | 1.3600 | 0.0069 |
| 11 | 1.3265 | 0.0092 | 1.3200 | 1.3600 | 0.0069 |
| 12 | 1.3272 | 0.0090 | 1.3200 | 1.3600 | 0.0068 |
| 13 | 1.3270 | 0.0091 | 1.3200 | 1.3600 | 0.0068 |
| 14 | 1.3275 | 0.0089 | 1.3200 | 1.3600 | 0.0067 |
| 15 | 1.3267 | 0.0091 | 1.3200 | 1.3600 | 0.0069 |

**Table 1.** A summary of the statistics for the lengths (ρ) of 80 peptide bonds in Caenopore-5 (PDB ID: 2JSA). Each row represents one structural model from the NMR ensemble, with Mean, Std, Min, and Max indicating average bond length, standard deviation, minimum, and maximum values of the lengths (ρ) of 80 peptide bonds, respectively.

Here, with Caenopore-5 [56–58] an example, its 80 peptide bond lengths (ρ) appear rather stable (Figure 5) but also exhibit variations of certain degree across its 81 amino acid residues, as shown in Figure 5 and fifteen subfigures of Figure 4 in supplementary file **S3.pdf**. These variations are represented by the distribution pattern as shown in Figure 5 and quantitatively described by a statistics summary for the lengths (ρ) of 80 peptide bonds in Caenopore-5 (Table 1). Specifically, the 80 bond lengths (ρ) possess an average value of approximately 1.326 Å (Figure 5) and a standard deviation of ∼0.009 Å, yielding a ratio of **Std/Mean** of ∼0.7% (Table 1). Assume that the distribution of ρ is a standard normal one and its **Mean** value equals its equilibrium bond length in theory, its **Mean** ± **3** × **Std** is able to represent 99.74% of the entire collection of the real situation for the lengths (ρ) of 80 peptide bonds in Caenopore-5. Finally, given that [ρ, θ, ϕ] is a three-dimension description of protein 3D structure, ρ represents one degree of dimensionality, the variation of ρ represents 3 × **Std/Mean** (Table 1), i.e., 3 × 0.7% = 2.1%. Since the equilibrium bond length in theory is by definition a physical constant, for Caenopore-5 (PDB ID: 2JSA), its structure description does not require a three-dimension description with [$x$, $y$, $z$]. Instead, its structure description does not require a 2.021-dimension description with [Std, θ, ϕ].

Taken together, since ρ is not a truly fixed value [71, 72], the ABN-SCS framework proposed here could not induce a dimensionality reduction by transforming a three-dimensional matrix ([ρ, θ, ϕ]) into a two-dimensional spherical vector comprising polar (θ) and azimuthal (ϕ) angles. Yet, the ABN-SCS framework induces a dimensionality reduction from 3 to 2.021 in the case of the 80 peptide bonds of Caenopore-5 [56–58], and this article argues that this dimensionality reduction is applicable not just to the entire 3D structure of Caenopore-5 [56–58], but also to the entire collection of protein structures deposited in the Protein Data Bank [1–4], and that this dimensionality reduction is able to enhance computational efficiency in fields such as molecular dynamics simulations [82].

## Supporting information

**S1 File.** **S1.pdf** ABN-SCS-based descriptions of the structures of twenty stand-alone natural amino acids.

**S2 File.** **S2.pdf** Supplementary Materials and Methods of this manuscript.

**S3 File.** **S3.pdf** Supplementary Results of this manuscript.

**S4 File.** **S4.pdf** Two sets of spherical parameters (ρ, θ and ϕ) extracted from the NMR ensemble (PDB ID: 2JSA) of Caenopore-5 [56–58].

**S5 File.** **spherical.py** Python script for ABN-SCS-based extraction of spherical geometric features, i.e., [ρ, θ, ϕ].

## Acknowledgments

I am grateful to the community of structural biology, for the continued accumulation of knowledge and data of the biomolecular space.

## References

1. Crystallography: Protein Data Bank. Nature New Biology. 1971;233(42):223–223.
2. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nature Structural & Molecular Biology. 2003;10(12):980–980.
3. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S. In: Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. Springer New York; 2017. p. 627–641. Available from: `http://dx.doi.org/10.1007/978-1-4939-7000-1_26`.
4. Velankar S, Burley SK, Kurisu G, Hoch JC, Markley JL. In: The Protein Data Bank Archive. Springer US; 2021. p. 3–21. Available from: `http://dx.doi.org/10.1007/978-1-0716-1406-8_1`.
5. Bonvin AMJJ. 50 years of PDB: a catalyst in structural biology. Nature Methods. 2021;18(5):448–449.
6. Wang T, Zhang W. Understanding Protein Functions in the Biological Context. Protein & Peptide Letters. 2023;30(6):449–458. doi:10.2174/0929866530666230507212638.
7. Deeds EJ, Shakhnovich EI. In: A Structure-Centric View of Protein Evolution, Design, and Adaptation. John Wiley & Sons, Inc.; 2010. p. 133–191. Available from: `http://dx.doi.org/10.1002/9780471224464.ch2`.
8. Lamb AL, Kappock TJ, Silvaggi NR. You are lost without a map: Navigating the sea of protein structures. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2015;1854(4):258–268. doi:10.1016/j.bbapap.2014.12.021.
9. Liu H, Lee W. The XFEL Protein Crystallography: Developments and Perspectives. International Journal of Molecular Sciences. 2019;20(14):3421. doi:10.3390/ijms20143421.

10. Luchinat E, Barbieri L, Cremonini M, Banci L. Protein in-cell NMR spectroscopy at 1.2 GHz. Journal of Biomolecular NMR. 2021;75(2–3):97–107. doi:10.1007/s10858-021-00358-w.

11. García-Nafría J, Tate CG. Structure determination of GPCRs: cryo-EM compared with X-ray crystallography. Biochemical Society Transactions. 2021;49(5):2345–2355. doi:10.1042/bst20210431.

12. Steven AC, Baumeister W. The future is hybrid. Journal of Structural Biology. 2008;163(3):186–195.

13. Koch I, Schäfer T. Protein super-secondary structure and quaternary structure topology: theoretical description and application. Current Opinion in Structural Biology. 2018;50:134–143. doi:10.1016/j.sbi.2018.02.005.

14. Reyes VM. Implementation of the Spherical Coordinate Representation of Protein 3D Structures and its Applications Using FORTRAN 77/90 Language; 2015. Available from: `https://arxiv.org/abs/1512.00424`.

15. Reyes VM. Representation of protein 3D structures in spherical ($\rho$, $\phi$, $\theta$) coordinates and two of its potential applications. Interdisciplinary Sciences: Computational Life Sciences. 2011;3(3).

16. Li W. Half-a-century Burial of $\rho$, $\theta$, and $\varphi$ in PDB. 2021;.

17. Li W. A Local Spherical Coordinate System Approach to Protein 3D Structure Description. 2020;doi:10.20944/preprints202001.0133.v1.

18. Li W. A Reversible Spherical Geometric Conversion of Protein Backbone Structure Coordinate Matrix to Three Independent Vectors of $\rho$, $\theta$ and $\phi$. 2024;doi:10.20944/preprints202404.0576.v1.

19. Pinheiro F, Santos J, Ventura S. AlphaFold and the amyloid landscape. Journal of Molecular Biology. 2021; p. 167059.

20. Ruff KM, Pappu RV. AlphaFold and Implications for Intrinsically Disordered Proteins. Journal of Molecular Biology. 2021; p. 167208.

21. Higgins MK. Can We AlphaFold Our Way Out of the Next Pandemic? Journal of Molecular Biology. 2021; p. 167093.

22. Fersht AR. AlphaFold – A Personal Perspective on the Impact of Machine Learning. Journal of Molecular Biology. 2021; p. 167088.

23. Zweckstetter M. NMR hawk-eyed view of AlphaFold2 structures. Protein Science. 2021;.

24. Strodel B. Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins. Journal of Molecular Biology. 2021; p. 167182.

25. Pathak RK, Singh DB, Sagar M, Baunthiyal M, Kumar A. In: Computational Approaches in Drug Discovery and Design. Springer Singapore; 2020. p. 1–21. Available from: `http://dx.doi.org/10.1007/978-981-15-6815-2_1`.

26. Li W, Vottevor G. Towards a Truly General Intermolecular Binding Affinity Calculator for Drug Discovery & Design. 2023;.

27. Xu W. Current Status of Computational Approaches for Small Molecule Drug Discovery. Journal of Medicinal Chemistry. 2024;67(21):18633–18636. doi:10.1021/acs.jmedchem.4c02462.

28. Li W. Half-a-century Burial of ρ, θ and φ in PDB. 2021;.

29. Berman HM. The Protein Data Bank. Nucleic Acids Research. 2000;28(1):235–242. doi:10.1093/nar/28.1.235.

30. Burley SK, Berman HM, Duarte JM, Feng Z, Flatt JW, Hudson BP, et al. Protein Data Bank: A Comprehensive Review of 3D Structure Holdings and Worldwide Utilization by Researchers, Educators, and Students. Biomolecules. 2022;12(10):1425. doi:10.3390/biom12101425.

31. Hao Y, Yu X, Bai Y, McBride HJ, Huang X. Cryo-EM Structure of HER2-trastuzumab-pertuzumab complex. PLOS ONE. 2019;14(5):e0216095.

32. Turk M, Baumeister W. The promise and the challenges of cryo-electron tomography. FEBS Letters. 2020;594(20):3243–3261. doi:10.1002/1873-3468.13948.

33. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–589.

34. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–589.

35. Krishna R, Wang J, Ahern W, Sturmfels P, Venkatesh P, Kalvet I, et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science. 2024;384(6693). doi:10.1126/science.adl2528.

36. Aubel M, Eicholt L, Bornberg-Bauer E. Assessing structure and disorder prediction tools for de novo emerged proteins in the age of machine learning. F1000Research. 2023;12:347. doi:10.12688/f1000research.130443.1.

37. Wu R, Ding F, Wang R, Shen R, Zhang X, Luo S, et al. High-resolutionde novostructure prediction from primary sequence. 2022;doi:10.1101/2022.07.21.500999.

38. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023;379(6637):1123–1130. doi:10.1126/science.ade2574.

39. Ahdritz G, Bouatta N, Floristean C, Kadyan S, Xia Q, Gerecke W, et al. OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. Nature Methods. 2024;21(8):1514–1524. doi:10.1038/s41592-024-02272-z.

40. AlQuraishi M. Machine learning in protein structure prediction. Current Opinion in Chemical Biology. 2021;65:1–8.

41. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373(6557):871–876.

42. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. Proceedings of the National Academy of Sciences. 2020;117(3):1496–1503.

43. AlQuraishi M. A watershed moment for protein structure prediction. Nature. 2020;577(7792):627–628.

44. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020;577(7792):706–710.

45. Hey T, Butler K, Jackson S, Thiyagalingam J. Machine learning and big scientific data. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2020;378(2166):20190054.

46. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). Proteins: Structure, Function, and Bioinformatics. 2019;87(12):1141–1148.

47. AlQuraishi M. AlphaFold at CASP13. Bioinformatics. 2019;35(22):4862–4865.

48. DeLano WL. Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography. 2002;40:82–92.

49. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of Molecular Graphics. 1996;14:33–38.

50. Jalily Hasani H, Barakat K. Homology Modeling: an Overview of Fundamentals and Tools. International Review on Modelling and Simulations (IREMOS). 2017;10(2):129. doi:10.15866/iremos.v10i2.11412.

51. Webb B, Sali A. Protein Structure Modeling with MODELLER. In: Methods in Molecular Biology. Springer US; 2020. p. 239–255.

52. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research. 2018;46(W1):W296–W303.

53. Du X, Li Y, Xia Y, Ai J, Wu Y. Molecular dynamics simulations and free energy calculations of protein–ligand interactions: recent advances and future perspectives. Current pharmaceutical design. 2017;23(30):4436–4450.

54. Cramer P. AlphaFold2 and the future of structural biology. Nature Structural & Molecular Biology. 2021;28(9):704–705.

55. Read RJ, Baker EN, Bond CS, Garman EF, van Raaij MJ. AlphaFold and the future of structural biology. Acta Crystallographica Section D Structural Biology. 2023;79(7):556–558.

56. Mysliwy J, Dingley AJ, Stanisak M, Jung S, Lorenzen I, Roeder T, et al. Caenopore-5: The three-dimensional structure of an antimicrobial protein from *Caenorhabditis elegans*. Dev Comp Immunol. 2010;34:323–330.

57. Roeder T, Stanisak M, Gelhaus C, Bruchhaus I, Grötzinger J, Leippe M. Caenopores are antimicrobial peptides in the nematode *Caenorhabditis elegans* instrumental in nutrition and immunity. Dev Comp Immunol. 2010;34:203–209.

58. Li W. Characterising the interaction between caenopore-5 and model membranes by NMR spectroscopy and molecular dynamics simulations [PhD thesis]. University of Auckland; 2016.

59. Ayoub R, Lee Y. Protein structure search to support the development of protein structure prediction methods. Proteins: Structure, Function, and Bioinformatics. 2021;89(6):648–658.

60. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021;596(7873):590–596.

61. Simpkin AJ, Rodríguez FS, Mesdaghi S, Kryshtafovych A, Rigden DJ. Evaluation of model refinement in CASP14. Proteins: Structure, Function, and Bioinformatics. 2021;.

62. Diwan GD, Gonzalez-Sanchez JC, Apic G, Russell RB. Next Generation Protein Structure Predictions and Genetic Variant Interpretation. Journal of Molecular Biology. 2021; p. 167180.

63. Abriata LA, Peraro MD. State-of-the-art web services for de novo protein structure prediction. Briefings in Bioinformatics. 2020;22(3).

64. Billings WM, Morris CJ, Corte DD. The whole is greater than its parts: ensembling improves protein contact prediction. Scientific Reports. 2021;11(1).

65. Layer G, Reichelt J, Jahn D, Heinz DW. Structure and function of enzymes in heme biosynthesis. Protein Science. 2010;19(6):1137–1161.

66. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices 1 1Edited by G. Von Heijne. Journal of Molecular Biology. 1999;292(2):195–202. doi:10.1006/jmbi.1999.3091.

67. Li W. A Reversible Spherical Geometric Conversion of Protein Backbone Structure Coordinate Matrix to Three Independent Vectors of ρ, θ and ϕ. 2024;doi:10.20944/preprints202404.0576.v1.

68. Beeman WO. In: Linguistics and Anthropology. Elsevier; 2012. p. 531–551. Available from: `http://dx.doi.org/10.1016/B978-0-444-51747-0.50016-0`.

69. Tyagi M, de Brevern AG, Srinivasan N, Offmann B. Protein structure mining using a structural alphabet. Proteins: Structure, Function, and Bioinformatics. 2007;71(2):920–937. doi:10.1002/prot.21776.

70. Potrzebowski W, Trewhella J, Andre I. Bayesian inference of protein conformational ensembles from limited structural data. PLOS Computational Biology. 2018;14(12):e1006641. doi:10.1371/journal.pcbi.1006641.

71. Batsanov SS. Calculation of van der Waals radii of atoms from bond distances. Journal of Molecular Structure: THEOCHEM. 1999;468(1-2):151–159.

72. Caine BA, Bronzato M, Popelier PLA. Experiment stands corrected: accurate prediction of the aqueous pKa values of sulfonamide drugs using equilibrium bond lengths. Chemical Science. 2019;10(25):6368–6381.

73. Cao L, Coventry B, Goreshnik I, Huang B, Sheffler W, Park JS, et al. Design of protein-binding proteins from the target structure alone. Nature. 2022;605(7910):551–560. doi:10.1038/s41586-022-04654-9.

74. Pillai A, Idris A, Philomin A, Weidle C, Skotheim R, Leung PJY, et al. De novo design of allosterically switchable protein assemblies. Nature. 2024;632(8026):911–920. doi:10.1038/s41586-024-07813-2.

75. Becka LN, Cruickshank DWJ. The variation of apparent bond lengths with temperature in molecular crystals. Acta Crystallographica. 1961;14(10):1092–1092. doi:10.1107/s0365110x6100317x.

76. Steiner T, Saenger W. Distribution of observed C–H bond lengths in neutron crystal structures and temperature dependence of the mean values. Acta Crystallographica Section A Foundations of Crystallography. 1993;49(3):379–384. doi:10.1107/s0108767392010249.

77. Webb H, Tynan-Connolly BM, Lee GM, Farrell D, O'Meara F, Søndergaard CR, et al. Remeasuring HEWL pKa values by NMR spectroscopy: Methods, analysis, accuracy, and implications for theoretical pKa calculations. Proteins: Structure, Function, and Bioinformatics. 2010;79(3):685–702.

78. Li W. Gravity-driven pH adjustment for site-specific protein $pK_a$ measurement by solution-state NMR. Measurement Science and Technology. 2017;28(12):127002.

79. Yang AS, Honig B. On the pH Dependence of Protein Stability. Journal of Molecular Biology. 1993;231(2):459–474.

80. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical $pK_a$ Predictions. Journal of Chemical Theory and Computation. 2011;7(2):525–537.

81. Ziolkowski J. New relation between ionic radii, bond length, and bond strength. Journal of Solid State Chemistry. 1985;57(3):269–290. doi:10.1016/0022-4596(85)90152-5.

82. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nature structural biology. 2002;9(9):646–652.