

Peer Review

Review of: "Over Half a Century of Burial of ρ , θ and ϕ in PDB"

Patrick Kunzmann¹

1. Independent researcher

The article motivates the usage of spherical coordinates as input features to protein structure prediction models. Indeed, using Cartesian coordinates directly has the major drawback that the model is not invariant to translations and rotations of the same structure. For this reason, structure prediction models like AlphaFold use features like torsion angles which are invariant to these transformations. However, spherical coordinates as proposed in your article are not rotation/translation invariant: As Fig 2. depicts, θ lies in the xy-plane, thus the ϕ - θ -pair is still dependent on the rotation of the structure, limiting the usability as input features to structure prediction models.

Second, I am concerned about floating point errors when using these coordinates: When Cartesian coordinates should be reconstructed from spherical coordinates (e.g., for some computation that requires them, or a molecular visualization) by iterating through the bonding network, I wonder if numerical errors accumulate. I think a reference implementation and an example 'spherical structure file' would be helpful here to clarify such details. These would also support the explanation in the 'definition, principles and methodology' section, which is a bit vague in my opinion.

Finally, you propose to create a complementary PDB describing structures in spherical coordinates. This would be a tremendous effort, where creating the database itself is arguably the easiest part. However, the existing tooling that reads structures from CIF/PDB files and works on them is also based on Cartesian coordinates. To motivate the community to adopt the 'spherical perspective' across a multitude of software packages would require some fundamental advantage of this alternative coordinate system, which I unfortunately do not see here. Even if the spherical representation benefits structure prediction models, the desired values can still be computed quickly on-the-fly, which is probably easier than switching to an alternative data representation.

Declarations

Potential competing interests: No potential competing interests to declare.