Extraction of Continuous Protein Dynamics from CryoEM aided by Coefficient Conversion between Normal Modes and 3D Zernike Polynomials

James Krieger, David Herreros, Carlos Oscar Sanchez Sorzano, Jose Maria Carazo

Biocomputing Unit, Centro Nacional de Biotecnología, CSIC, Madrid 28049, Spain

Continuous dynamics and conformational transitions are critical for understanding the mechanisms of biological molecular machines. CryoEM provides many 2D projection images of these machines rapidly frozen in all possible conformational states, potentially allowing us to reconstruct these dynamics. This analysis remains very challenging.

Several methods show some success, but there is still room for improvement. Our recent method uses 3D Zernike polynomials to calculate possible motions from CryoEM maps and optimises their coefficients to describe transitions to other maps, images, and atomic structures. This deformation is broadly applicable, even for making new structures.

However, the coefficient optimisation is not always successful and may induce unrealistic deformations. To overcome this, we have developed a way of converting projection coefficients of normal modes to 3D Zernike coefficients and *vice versa*, providing a better connection to physically reasonable motions.

Introduction

- Proteins have intrinsic global dynamics encoded by their structure and important for their function **[1]**
- **Normal mode analysis (NMA)** with coarse grained elastic network models (ENMs) is a good method for describing such dynamics **[1]**
- Traditional structural biology methods, including CryoEM with discrete maximum likelihood classification, provide preferred conformational states with differences between them that are in line with NMA
- This information can be harnessed to extract continuous dynamics and landscapes using new CryoEM image processing methods
- We demonstrate the beginnings of such a method, building upon our recent **Zernikes3D** method **[2]**

Theory

- Conformational changes between two structures can be described by deformation fields comprised of a basis set and corresponding coefficients
- **Zernikes3D [2]** approximates a volume $V'(r)$ as the deformation of a reference volume $V(r)$ as a linear combination of the 3D Zernike polynomials $Z_{lnm}(r)$ at each voxel r with coefficients α_{lnm}
- These polynomials can also be calculated for atoms
- NMA eigenvectors can also be projected onto a deformation vector to approximate conformational change as a linear combination of mode vectors $\mathbf{u}_k$ with scaling coefficients g_k
- Hence, there's an equivalence between the two sets when considering structures with P atoms

$$\begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_P \end{pmatrix} \mathbf{g} = (I_P \otimes A) \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_P \end{pmatrix} \quad \text{or} \quad (I_P \otimes \mathbf{g}^T) \begin{pmatrix} U_1^T \\ U_2^T \\ \vdots \\ U_P^T \end{pmatrix} = \begin{pmatrix} Z_1^T \\ Z_2^T \\ \vdots \\ Z_P^T \end{pmatrix} A^T$$

and we can solve it by least squares:

$$\mathbf{g} = (U^T U)^{-1} U^T (I_P \otimes A) Z$$

and

$$A^T = (Z^T Z)^{-1} Z^T (I_P \otimes \mathbf{g}^T) U^T$$

Results

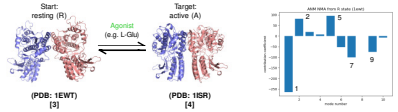
1. NMA Projection Coefficients capture Transitions

- The simplest approximation of the transition between 2 structures of P atoms comes from subtraction of atom positions and is called the deformation vector

$$\Delta \boldsymbol{q} = [\Delta x_1 \quad y_1 \quad z_1 \quad \Delta x_2 \quad \dots \quad \dots \quad \dots \quad \dots \quad \Delta z_P]^T$$

- (Linear interpolation along this vector gives a morph)
- Normal mode vectors have the same form with x, y and z movement components for each atom
- A dot product between normal modes (NMs) and the deformation vector yields projection coefficients g_k

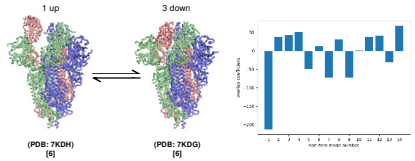
Example 1: metabotropic glutamate receptor (mGluR)



- Activation steps for Venus fly trap (VFT) dimer upon agonist binding **[3, 4]** are captured by global modes **[5]**:
  - Mode 1: inter-subunit rotation
  - Mode 2: inter-subunit closure
  - Mode 5: closing of clamshell B
  - Mode 7: closing of clamshell A
  - Mode 9: symmetric clamshell tilting

- With 10 NMs, RMSD to target drops from 11.2 Å → 3.4 Å
- With 100 NMs, RMSD to target drops from 11.2 Å → 1.2 Å

Example 2: SARS-CoV-2 Spike



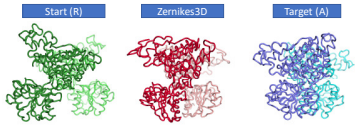
- Receptor-binding domains (RBDs) alternate between an "up" state to engage host cells and a "down" state that evades the immune system, and can be impacted by mutations including D614G **[6-8]**
- A single mode captures the RBD coming down
- RMSD to target drops from 5.9 Å → 3.0 Å

2. Conversion of Coefficients Helps Zernikes3D

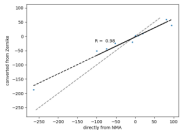
- Conversions can be applied in either direction
- These are illustrated with the same two examples

Example 1: metabotropic glutamate receptor (mGluR)

- The Zernikes3D atoms and volumes programs can capture the VFT dimer transition reasonably well



- Converting to NMA removes unphysical deformations by not having modes that include them
- It recovers the same modes as the projection with similar weights but smaller amplitudes
- RMSD improves relative to Zernikes3D alone although still worse than NMA projection

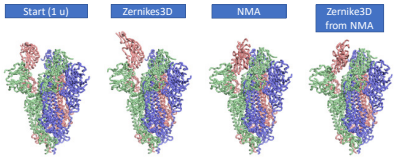


Deformation	RMSD to target (Å)
None (starting RMSD)	11.23
Zernikes3D direct	6.57
10 modes (converted from Zernikes3D)	5.21
10 modes (projection)	3.43

- This could be useful for selecting modes to save efficiency in hybrid electron microscopy normal mode analysis (HEMNMA) continuous heterogeneity analysis and for flexible fitting and other map-based simulations

Example 2: SARS-CoV-2 Spike

- Zernikes3D has difficulty with RBD up/down motion, even using higher order polynomials
- Converting from NMA to Zernikes3D really helps



- The improved 3D Zernike polynomials could be used as priors for continuous heterogeneity analysis and improving reconstructions of less populated states.

Methods

- Atomic structures were parsed and handled, and normal modes, deformation vectors, and projections were calculated with ProDy 2
- Zernikes3D command line programs in Xmipp were used directly, including atoms program for mGluR VFT and volumes program for the SARS-CoV-2 Spike
- Volumes of the Spike were simulated by converting atomic structures with Xmipp programs using Scipion
- Original volumes from the EMDB were also tried with similar results
- Prototype code was developed in Python
- Key steps include:
  - Parsing of NMA vectors and Zernike coefficients
  - Calculation of NMA coefficients and 3D Zernike polynomials
  - Least squares conversion
  - Application of deformations
  - Analysis and output of new coefficients and deformed structures

Conclusions

- Normal modes and 3D Zernike polynomials can both work well for describing conformational changes of biological macromolecules
- A conversion between them using least squares enables their use in more circumstances and will facilitate further methods development
- This conversion is particularly helpful for the SARS-CoV-2 Spike, which causes difficulties for the Zernikes3D method

Acknowledgments

We gratefully acknowledge the Marie Sklodowska-Curie Actions Individual Fellowship grant EnLaCES (H2020-MSCA-IF-2020, Proposal: 101024130) to JMK, as well as HighResCells (ERC - 2018 - SyG, Proposal: 810057), and "ERDF: A way of making Europe" from the European Union and Horizon 2020, and PID2019-104757RB-I00 funded by MCIN/AEI/10.13039/501100011033

References

[1] Zhang *et al.* (2020). *Curr. Opin. Struct. Biol.* **62**, 14-21.

[2] Herreros *et al.* (2021). *IUCrJ* **8**, 992-1005.

[3] Kunishima *et al.* (2000). *Nature* **407**, 971-977.

[4] Tsuchiya *et al.* (2002). *PNAS* **99**, 2660-2665.

[5] Krieger *et al.* (2015). *Biophys. J.* **109**, 1136-1148.

[6] Gobeil *et al.* (2021). *Cell Rep.* **34**, 108630

[7] Queirós-Reis *et al.* (2021). *Int. J. Mol. Sci.* **22**, 10836

[8] Ginex *et al. bioRXiv.* **2021.12.05.471263**; *PLOS Pathogens* (under revision)