

¹ HeliXplore: A Python package for analyzing multi-strand helix deformations

³ **Ronnie Mondal**  ^{1,2} and **Valerie Vaissier Welborn**  ^{1,2} ¶

⁴ 1 Department of Chemistry, Virginia Tech, Blacksburg, VA 24061 2 Macromolecules Innovation Institute,
⁵ Virginia Tech, Blacksburg, VA 24061 ¶ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Open Journals](#) ↗

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

⁶ Corresponding author email: vwelborn@vt.edu

⁷ Summary

⁸ Multi-stranded helices are a recurring structural motif in biomolecular systems, most prominently found in DNA and collagen ([Ramachandran & Kartha, 1954](#); [Watson & Crick, 1953](#)). They also appear in synthetic polymers and macromolecules ([Yashima et al., 2009](#)). Helix properties depend on strand length, strand composition and the environment, which can be modeled using classical Molecular Dynamics (MD). However, multi-stranded helices can undergo local and global deformations that directly impact their function. MD lacks a tool to systematically quantify deformation across systems and conditions. In the past, people have used polysaccharide atom distances ([Khatami et al., 2021](#)), collagen cross-sectional triangles ([Ravikumar & Hwang, 2008](#)), pitch, or principal axis measures ([Koneru et al., 2019](#); [Zhang et al., 2006](#)). Although adequate for the system under investigation, these metrics are not generalizable and miss key local or collective distortions. Without local or inter-strand descriptors, comparisons across systems remain largely qualitative.

¹⁴ We present HeliXplore, an open-source Python package for the systematic and quantitative analysis of multi-strand helix deformation. Originally inspired by collagen, HeliXplore is ¹⁵ generalizable to any helical bundle, including single-stranded helices. HeliXplore measures ¹⁶ how helices deviate from their ideal geometry using user-defined backbone atoms. HeliXplore ¹⁷ runs calculations in three sections: Section 1 for intra-strand deformations (rise, radius, twist ¹⁸ and windowed deviations) per atom, or group of atoms and per time frame and helical ¹⁹ regularity per strand; Section 2 for inter-strand deformations (axial shifts, axis angle deviations, ²⁰ axis distance deviations and averaged distance deviations); and Section 3 for triple-helix ²¹ deformations, using the area and shape of the cross-sectional triangle.

²² In practice, one only needs HeliXplore.py, the MD trajectory file (in TINKER .arc format ²³ or the standard RCSB .pdb format) and the number of strands to be able to run the code. ²⁴ Users can also input the atom names or atom types (for TINKER .arc format) to mark the ²⁵ backbone. `read_tinker_arc()` and `read_traj_pdb()` functions can be replaced to cater to ²⁶ other trajectory file formats. HeliXplore checks for the four required Python dependencies ²⁷ (`numpy`, `scipy`, `pandas` and `matplotlib`) before running the main code. No other installations ²⁸ are required. For a detailed description of the inputs and examples, see the README file on ²⁹ GitHub. A shorter description is provided with `python HeliXplore.py --help`.

³⁰ Statement of Need

³¹ HeliXplore provides the first open-source, Python-based implementation of a quantitative ³² framework for analyzing multi-helix deformations. By resolving intra- and inter-strand ³³ deformations, HeliXplore enables atomic resolution of structural distortions that have ³⁴ previously eluded MD studies. More broadly, HeliXplore establishes a transferable ³⁵

42 methodology for analyzing helices, positioning it as a foundational tool for systematic
 43 comparisons across systems, conditions and force fields. Users of HeliXplore are free to
 44 modify the code and underlying mathematical formulations to adapt to their specific research
 45 needs.

46 Mathematics

47 Section I:

48 For a helix of length N , with $\mathbf{p}^{i,m}(t)$ the coordinates of atom i in strand m at time t , the
 49 **helical axis vector** $\mathbf{v}_1^m(t)$ is determined via principal component analysis:

$$\left[\frac{1}{N} \sum_{i=1}^N (\mathbf{p}^{i,m}(t) - \mathbf{c}^m(t))(\mathbf{p}^{i,m}(t) - \mathbf{c}^m(t))^T \right] \mathbf{v}_1^m(t) = \lambda_1^m(t) \mathbf{v}_1^m(t),$$

50 where $\mathbf{c}^m(t)$ is the average coordinates of strand m at time t :

$$\mathbf{c}^m(t) = \frac{1}{N} \sum_{i=1}^N \mathbf{p}^{i,m}(t),$$

51 and λ_1^m indicates the principal eigenvalue. The unit vector of $\mathbf{v}_1^m(t)$ is hereafter referred to as
 52 $\hat{\mathbf{v}}_1^m(t)$.

53 **Deviations in rise** for atom i at strand m are calculated as:

$$\delta_{\text{Rise}}^{i,m}(t) = \frac{\overbrace{(\mathbf{p}^{i+1,m}(t) - \mathbf{p}^{i,m}(t)) \cdot \mathbf{v}_1^m(t)}^{\text{Rise}^{i,m}(t)} - \text{Rise}^{i,m}(0)}{\text{Rise}^{i,m}(0)}.$$

54 **Deviations in radius** are calculated using the circumradius $R^{i,m}$ as:

$$\delta_{\text{Radius}}^{i,m}(t) = \frac{R^{i,m}(t) - R^{i,m}(0)}{R^{i,m}(0)},$$

55 where:

$$R^{i,m}(t) = \frac{\|\mathbf{p}^{i,m}(t) - \mathbf{p}^{i-1,m}(t)\| \|\mathbf{p}^{i+1,m}(t) - \mathbf{p}^{i-1,m}(t)\| \|\mathbf{p}^{i,m}(t) - \mathbf{p}^{i+1,m}(t)\|}{4 \times \frac{1}{2} \|(\mathbf{p}^{i,m}(t) - \mathbf{p}^{i-1,m}(t)) \times (\mathbf{p}^{i+1,m}(t) - \mathbf{p}^{i-1,m}(t))\|}.$$

56 **Deviations in twist** for atom i on strand m at time t are calculated as the angle between the
 57 normals of the planes defined by consecutive atoms $(i-1, i, i+1)$ ($\mathbf{n}_a^{i,m}(t)$) and $(i, i+1, i+2)$
 58 ($\mathbf{n}_b^{i,m}(t)$) as:

$$\delta_{\text{Twist}}^{i,m}(t) = \frac{\overbrace{\arccos \left(\frac{\mathbf{n}_a^{i,m}(t) \cdot \mathbf{n}_b^{i,m}(t)}{\|\mathbf{n}_a^{i,m}(t)\| \|\mathbf{n}_b^{i,m}(t)\|} \right)}^{\text{Twist}^{i,m}(t)} - \text{Twist}^{i,m}(0)}{\text{Twist}^{i,m}(0)}.$$

59 **Windowed deviations** are calculated after superposition using a window of 5 atoms centered
 60 around atom i with the Kabsch algorithm (Kabsch, 1976) as:

$$d_{\text{windowed}}^{i,m}(t) = \|\mathbf{q}^{i,m}(t) - \mathbf{p}^{i,m}(0)\|,$$

61 where $\mathbf{q}^{i,m}(t)$ are the aligned coordinates of atom i . $\mathbf{q}^{i,m}(t)$ is calculated from the coordinates
 62 of the atoms within the window $\mathbf{p}^{j,m}(t)$, where $j \in [i-2, i+2]$ as:

$$\mathbf{q}^{j,m}(t) = \mathbf{R}^{j,m}(t) (\mathbf{p}^{j,m}(t) - \langle \mathbf{p}^{j,m}(t) \rangle_j) + \langle \mathbf{p}^{j,m}(0) \rangle_j,$$

63 where the optimal rotation matrix $\mathbf{R}^{j,m}(t)$ is obtained by singular value decomposition (using
 64 NUMPY.LINALG.SVD) of the cross-covariance matrix $\mathbf{H}^{j,m}(t)$:

$$\mathbf{H}^{j,m}(t) = \sum_j (\mathbf{p}^{j,m}(t) - \langle \mathbf{p}^{j,m}(t) \rangle_j) (\mathbf{p}^{j,m}(0) - \langle \mathbf{p}^{j,m}(0) \rangle_j)^T.$$

65 **Helical regularity** for a strand m is calculated as:

$$\mathcal{R}^m(t) = \frac{\sigma_{\text{Rise}}(t)}{\langle |\text{Rise}^{i,m}(t)| \rangle_i} + \frac{\sigma_{\text{Radius}}(t)}{\langle |\text{Radius}^{i,m}(t)| \rangle_i} + \frac{\sigma_{\text{Twist}}(t)}{\langle |\text{Twist}^{i,m}(t)| \rangle_i},$$

66 where σ is the standard deviation of the metric over i .

67 Section II:

68 **Axial shift** of strand m with respect to strand n along the central axis is calculated as:

$$s^{mn}(t) = |\langle \tilde{\mathbf{v}}_1^m(t) \rangle_m \cdot (\mathbf{c}^m(t) - \mathbf{c}^n(t))|, \quad \text{where } \tilde{\mathbf{v}}_1^n(t) = \begin{cases} -\hat{\mathbf{v}}_1^n, & \text{when } \hat{\mathbf{v}}_1^m \cdot \hat{\mathbf{v}}_1^n < 0 \\ \hat{\mathbf{v}}_1^n, & \text{otherwise.} \end{cases}$$

69 **Deviations in axis angles** are calculated between the unit vectors of the principal axes of the
 70 two strands m and n as:

$$\delta_\theta^{mn}(t) = \frac{\theta^{mn}(t)}{|\arccos(\hat{\mathbf{v}}_1^m(t) \cdot \hat{\mathbf{v}}_1^n(t)) - \theta^{mn}(0)|}.$$

71 **Deviations in axis distances** are calculated using the perpendicular distance between the axes
 72 of strands m and n as:

$$\delta_d^{mn}(t) = \frac{\frac{d^{mn}(t)}{|\langle (\mathbf{c}^m(t) - \mathbf{c}^n(t)) \cdot (\hat{\mathbf{v}}_1^m(t) \times \hat{\mathbf{v}}_1^n(t)) |} - d^{mn}(0)}{d^{mn}(0)}.$$

73 **Deviations in averaged distances** are calculated using the distance between the average
 74 coordinates of strands m and n as:

$$\delta_c^{mn}(t) = \frac{\frac{c^{mn}(t)}{\|\mathbf{c}^m(t) - \mathbf{c}^n(t)\|} - c^{mn}(0)}{c^{mn}(0)}.$$

75 **Section III:**

- 76 For triple helices, one atom i on each strand is taken to form a triangular cross-section.
 77 **Deviations in area** are calculated from the area of the triangle as:

$$\delta_{\text{Area}}^i(t) = \frac{\overbrace{\frac{1}{2} \|(\mathbf{p}^{i,2}(t) - \mathbf{p}^{i,1}(t)) \times (\mathbf{p}^{i,3}(t) - \mathbf{p}^{i,1}(t))\|}^{\text{Area}^i(t)} - \text{Area}^i(0)}{\text{Area}^i(0)}.$$

- 78 **Deviations in shape** are calculated from the normalized isoperimetric ratio (IP) as:

$$\delta_{\text{Shape}}^i(t) = \frac{\overbrace{\frac{\text{IP}^i(t)}{4\pi \text{Area}^i(t)}}^{\text{IP}^i(t)} - \text{IP}^i(0)}{\text{IP}^i(0)},$$

- 79 where $P^i(t)$ is the corresponding perimeter.

- 80 The first frame is taken as the reference. The reference can be changed by appending a new
 81 reference frame to the beginning of the input MD trajectory. All mathematical details are also
 82 outlined as comments in the code.

83 **Acknowledgements**

- 84 The authors thank the National Institute of Health, National Institute of General Medical
 85 Sciences, under award number R35-GM150409 and Advanced Research Computing at Virginia
 86 Tech for providing computational resources and technical support.

87 **References**

- 88 Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Foundations
 89 of Crystallography*, 32(5), 922–923. <https://doi.org/10.1107/S0567739476001873>
- 90 Khatami, M. H., Barber, W., & Haan, H. W. de. (2021). Using geometric criteria to study
 91 helix-like structures produced in molecular dynamics simulations of single amylose chains
 92 in water. *RSC Advances*, 11(20), 11992–12002. <https://doi.org/10.1039/DRA00071C>
- 93 Koneru, J. K., Zhu, X., & Mondal, J. (2019). Quantitative assessment of the conformational
 94 heterogeneity in amylose across force fields. *Journal of Chemical Theory and Computation*,
 95 15(11), 6203–6212. <https://doi.org/10.1021/acs.jctc.9b00630>
- 96 Ramachandran, G. N., & Kartha, G. (1954). Structure of collagen. *Nature*, 174(4423),
 97 269–270. <https://doi.org/10.1038/174269c0>
- 98 Ravikumar, K. M., & Hwang, W. (2008). Region-specific role of water in collagen unwinding
 99 and assembly. *Proteins: Structure, Function, and Bioinformatics*, 72(4), 1320–1332.
 100 <https://doi.org/10.1002/prot.22026>
- 101 Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: A structure for
 102 deoxyribose nucleic acid. *Nature*, 171(4356), 737–738. <https://doi.org/10.1038/171737a0>
- 103 Yashima, E., Maeda, K., Iida, H., Furusho, Y., & Nagai, K. (2009). Helical polymers:
 104 Synthesis, structures, and functions. *Chemical Reviews*, 109(11), 6102–6211. <https://doi.org/10.1021/cr900162q>

- ¹⁰⁶ Zhang, Q., Lu, Z., Hu, H., Yang, W., & Marszalek, P. E. (2006). Direct detection of the
¹⁰⁷ formation of ν -amylose helix by single molecule force spectroscopy. *Journal of the American
¹⁰⁸ Chemical Society*, 128(29), 9387–9393. <https://doi.org/10.1021/ja057693+>

DRAFT