Writhe to Drive: Using entanglement to guide protein structure predictions.

Christopher Prior[1], Robert Rambo[2], Arron Bale[1,3]

[1]Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, United Kingdom
[2]Diamond Light Source Ltd, Harwell Science Innovation Campus, Didcot, OX11 0DE, United Kingdom
[3]PhD Candidate for MoSMed CDT

Introduction

The absolute writhe is a measure of entanglement of a curve. Intuitively, one can think of the writhe as the average number of crossings seen by observers sat around the curve, as shown in Fig.1. Due to the strict constraints on the local geometry, proteins have a preferred amount of complexity as measured by the absolute writhe. When making predictions to fit BioSAXS data[1], we use the absolute writhe as a quick and easy to calculate diagnostic tool to assess if the predicted structure is realistic.

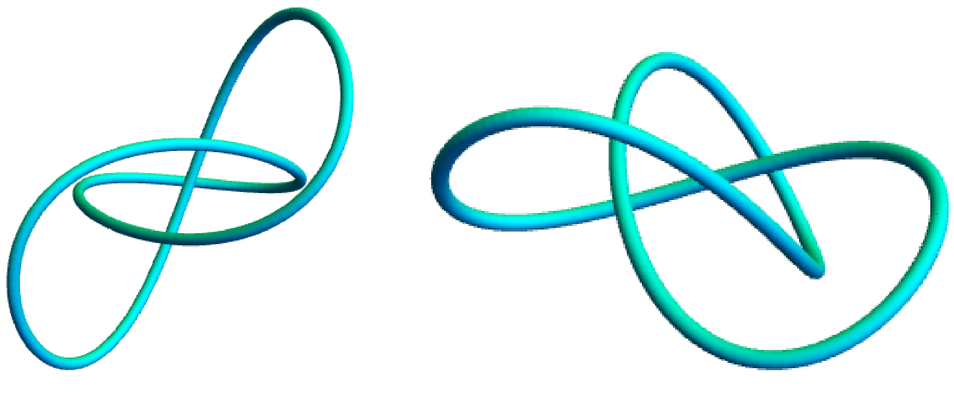

Fig. 1: Two different projections of the trefoil knot. On the left, we can see 3 crossings, whereas on the right we can see 4 crossings. Averaged over all projections then, the trefoil has a writhe of just over 3.

Bounding the Absolute Writhe

- We consider a protein backbone as a discrete curve whose points represent the alpha-carbon residues.
- We replace secondary structure elements with straight edges to ”smooth out” the backbone, as seen on the left of Fig.2.
- When smoothing in this way, we are measuring entanglement on the tertiary structure scale, not the local entanglement of say, the α-helices. See Fig.3 for a visual example of this.
- We can then plot the absolute writhe of this smoothed curve against its length.
- We repeat this process for a sample of over 2000 protein structures from the PDB to produce the scatter plot in Fig.2, and fit bounds to this distribution.
- We then use these two bounding curves, especially the lower bound, to penalise unrealistic predictions in the fitting algorithm.

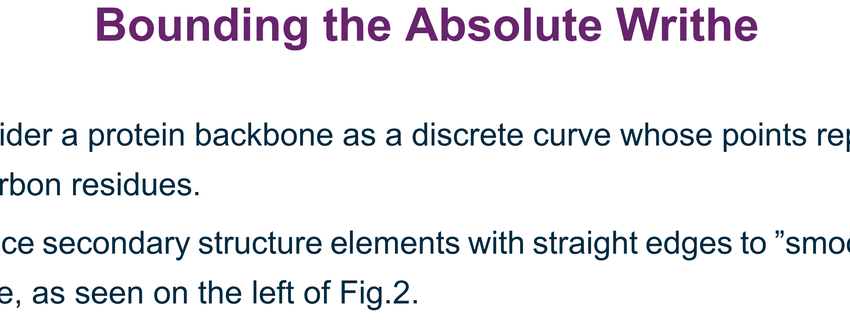

Fig. 2: Left: A visual example of smoothing the secondary structure for the protein CC Chemokine CCL14 (PDB: 2Q8R). Right: A plot of the absolute writhe against length for a representative sample of over 2000 proteins from the PDB.

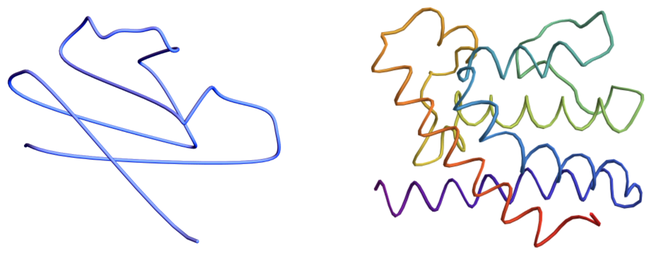

Fig. 3: Left: The smoothed backbone of the gene product Af1862 from Archaeoglobus fulgidus (PDB: 2oeb). Right: The full structure of 2oeb.

The MicA Problem

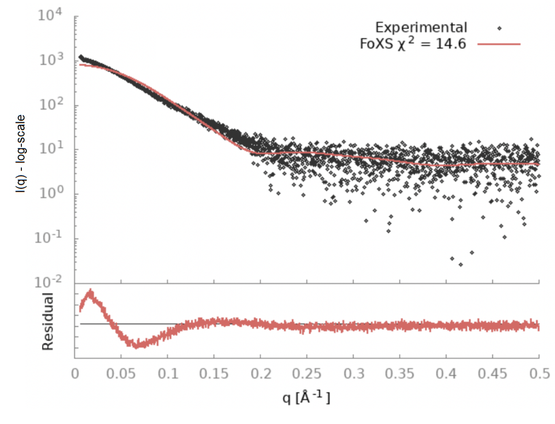

Fig. 4: In black, we see the experimental BioSAXS data for MicA in solution. In red, we see the scattering profile from FoXS[2].

In Fig.4 we see the FoXS[2] scattering profile of the crystallographic (and AlphaFold[3] predicted) structure of MicA. This is a poor fit, especially at the lower values of q where instead of the flat sloping profile, indicative of an larger opened out structure, the red curve has a hill-like peak.

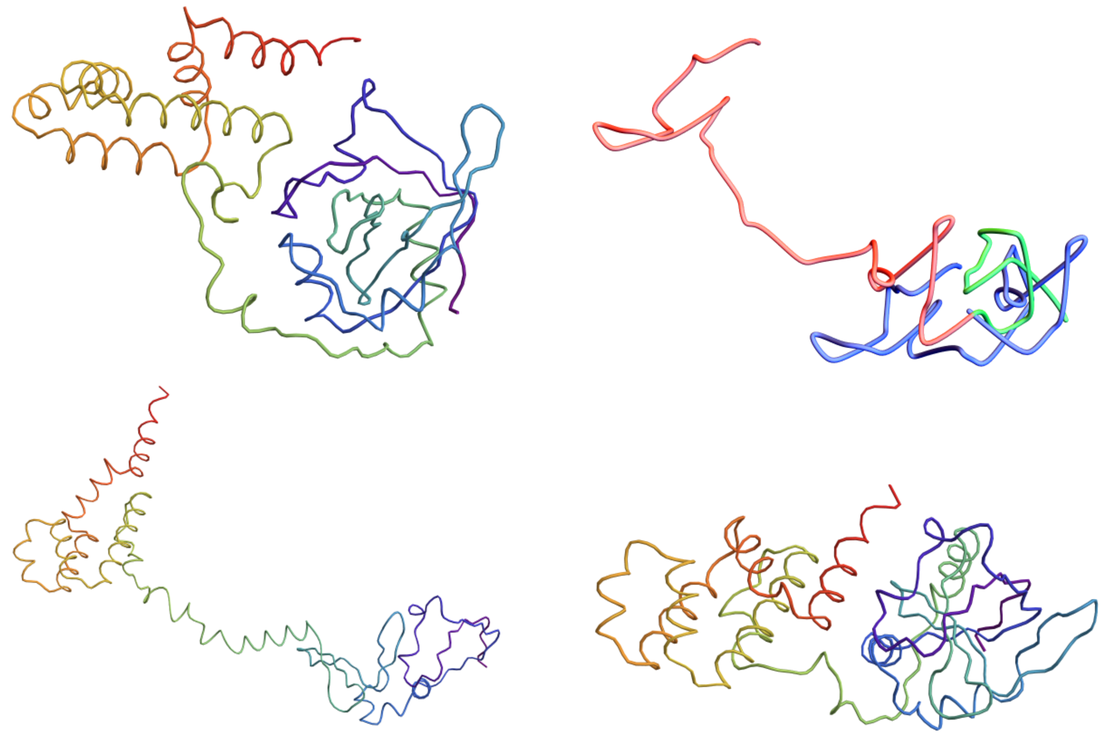

Fig. 5: Top left: The crystallographic structure of MicA. Top Right: Two good fits to the experimental SAXS data for MicA, in our smoothed representation. In green, the regions where the predictions agree. In red/blue, the α-helix heavy sub-unit whose relative position varies between the two predictions. Bottom Left: The full structure of the opened out prediction (red top right). Bottom Right: The full structure of the compact prediction (blue top right).

Results and Discussion

In Fig.5 we see two examples of good fits to the SAXS data for MicA. The red structure has opened out, leading to an absolute writhe that falls below the empirical bound. The blue structure on the other hand is much more compactly entangled, with an absolute writhe that falls within the realistic region. This may initially seem to contradict the SAXS data, which is indicative of a large opened out structure. In fact, there has been some dimerization in solution, and our predictions agree with this. In essence, our red structure in Fig.5 is unfolding to match the size of the dimer.

References

[1] Prior, C, et al.: *J. Chem. Theory Comput.*, journal 2020, **16**(3):1985, pMID: 32023061.

[2] Schneidman-Duhovny, D, et al.: *Nucleic Acids Res.*, journal .

[3] Jumper, J, et al.: *Nature*, journal 2021, **596**(7873):583.