Describing protein structure geometry to aid in functional understanding

Charlotte M. Deane[1], James Dunbar[1], Angelika Fuchs[2], Kanti V. Mardia[1,3], Jiye Shi[4] and Henry R. Wilman[1]

[1] Department of Statistics, University of Oxford
[2] Roche Diagnostics
[3] Department of Statistics, University of Leeds
[4]UCB Pharma

1 Introduction

The shape of a protein is known to be fundamental in its ability to perform a specific function. Here we consider two case examples, the variation of angles between the two variable chains of antibodies and the annotation of kinks in membrane helices.

2 ABangle: Characterising the VH-VL orientation in antibodies

Antibodies are a class of proteins indispensable for the vertebrate immune system. The general architecture of all antibodies is very similar, but they are able to bind to many different molecules or antigens. This binding malleability means that antibodies are an increasingly important category of biopharmaceuticals and biomarkers (Kuroda *et al.* (2012)).

The binding site of antibodies is situated between the two variable domains, VH and VL, of the antibody's antigen binding fragment. The relative orientation between these domains contributes to determining the topology of the binding site which in turn influences antigen specificity and affinity. Subsequently, understanding and predicting the VH-VL orientation is important for antibody modelling, docking and as well as for studying the mechanisms of antigen specificity and affinity (Dunbar *et al.* (2013)) .

We have developed a method to characterise the VH-VL orientation in an absolute sense and implemented it in the computational tool, ABangle (Dunbar *et al.* (2013)). This allows us to calculate differences in orientation between structures in a consistent fashion and determine how they compare to other structures globally.

By fully characterising the VH-VL orientation in an absolute sense we are able to quantify not just the relative changes between pairs of structures, but how this change relates to variation observed in all structures. An analysis of conservation of VH-VL orientation in sequence-identical structures revealed that antibodies specific for protein antigens were more flexible than those specific for hapten antigens in the unbound form. However, in the bound form, both types of antibody structures were found to have a similar level of orientation conservation.

The characterisation has also allowed us to identify preferences in orientation space for different VH-VL pairings and how the presence of certain amino-acids at particular positions affects pose. We have used this information to investigate how better knowledge about the VH-VL orientation can improve antibody structure prediction.

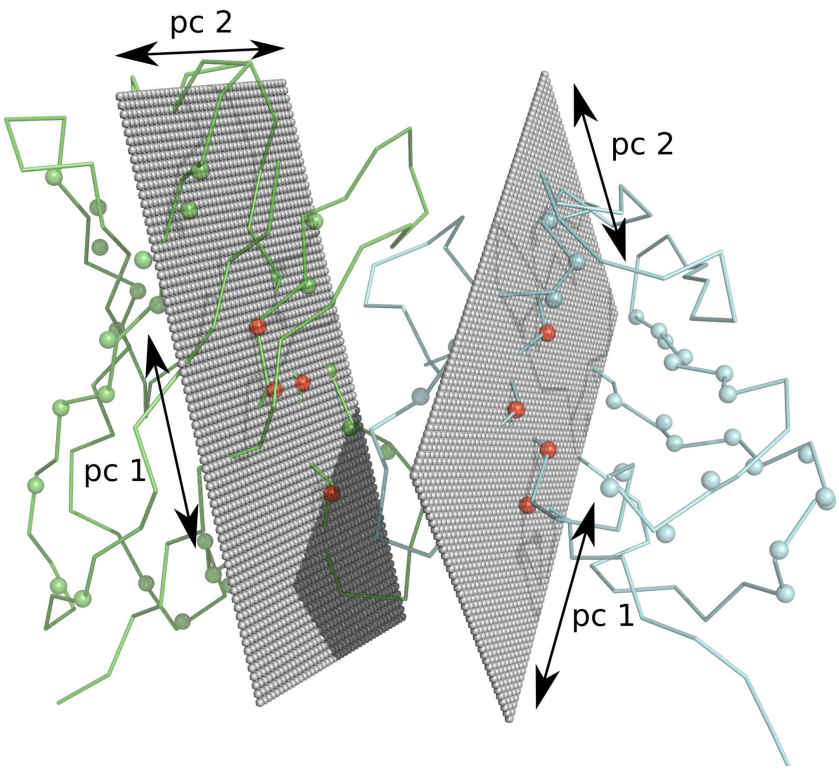

Figure 1: The heavy (green) and light (cyan) variable domains of an antibody (ribbon pdb 1b4j). Planes are mapped onto the domains by performing a structural superposition of the VH and the VL consensus structures (spheres) onto the coreset positions. Those positions used to fit the planes for the VH and the VL domains are coloured red.

3 Statistical approaches to kinked helices

In the second case we consider kinks alpha helices, such kinks (deviations from the ideal helix structure) are thought to be functionally important. There are a number of methods available to identify helix kinks but they disagree as to what constitutes a kink (Sansom and Weinstein (2000)). Kink Finder is a standard method it takes 6 residue segments of the helix, and fits an axis to each segment using a cylinder fit. Calculating the angle between the axis of adjoining sections gives an angle for each residue. The largest angle is selected, and if it is above a threshold, the helix is designated as kinked, otherwise it is not kinked . However, at what angle is a helix kinked, and at what angle is it not? Given there is no consensus definition of a kinked helix, how can an appropriate cut-off be chosen? In a standard method, a string of heuristic decisions are made, for example, the number of atoms used to calculate a local helix axis, or the order of fit to use. Our new methodology uses a simple statistical parameter and a statistical model to classify helices. This removes many of these heuristic decisions.

Here we describe a method that makes this decision in a more statistically robust manner. A log-likelihood ratio based statistic is calculated for each helix, which depends only on the second and third eigenvalues from a principle component analysis of the C-alpha coordinates of all the helix residues (Mardia *et al.* (1979)). A mixture model is used for this statistic to identify the distributions of the two separate types of helix – kinked and not kinked. For a query helix, applying discriminate analysis to this model provides a log odds score to classify between the

two types of helix.

A second challenge facing all kink classifiers is the error associated with the estimate of the angle (or other statistic). Effective quantification of this allows us to more accurately describe helices as kinked or not kinked. For example, given two helices, both with maximum angle estimates of 25, one of these may be based on good fits, whilst the other is based on poorer fits. The first we are more confident is kinked, whilst the other may or may not be kinked.

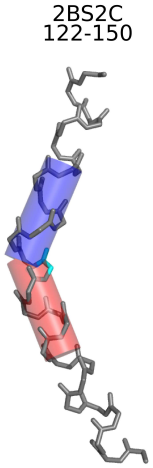

Figure 2: A kinked helix with Kink Finder cylinders shown and the Kink Residue identified. The helix is labelled with its 4 letter PDB code and single letter chain identifier and the number of the first and last residue in the helix.

4 Conclusion

These two examples show how accurately capturing and describing the geometry of proteins can help our functional understanding.

References

Kuroda, D., Shirai, H., Jacobson, M.P. and Nakamura, H. (2012) *Protein Eng. Des. Sel.*, 25, 507–522

Dunbar, J., Fuchs, A., Shi, J., and Deane, C.M.(2013) *Protein Eng. Des. Sel.*, doi: 10.1093/protein/gzt020 .

Sansom, M. S. and Weinstein, H. (2000).*Trends in Pharmacological Sciences*, 21(11), 445–51.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) *Multivariate Analysis. Academic Press (London).* pp 235-236.