New directional statistics for patterns in protein and RNA structure

Zhengzheng Zhang[1*], Kanti V. Mardia[1], Charles C. Taylor[1], John T. Kent[1], and Thomas Hamelryck[2]

[1] Department of Statistics, University of Leeds
[2] Bioinformatics Centre, Department of Biology, University of Copenhagen

There are many statistical challenging problems in protein and RNA Bioinformatics. We can regard the backbone of a typical protein as an articulated object in three dimensions with fixed bond lengths between successive amino acids. Hence it can be viewed as a long time series (with hundreds or thousands of amino acids), where all the information lies in the angles between successive bonds. There are two types of angles: bond or planar angles, analogous to colatitude, which are nearly constant here; and dihedral angles, analogous to longitude, which contain all the information. Thus the basic protein description is reduced to a circular time series. There is also further angular information coming from side chains.

The aim is to find patterns in such data (see for example, Boomsma et al, 2008). This poster will give new methods and visualisation tools, including circular principal component analysis and clustering (cf: Mu et al, 2005; Altis et al, 2007), together with examples. A similar analysis can also be carried for the structure of RNA molecules. Our method proposes an alternative approach to Frellsen et al (2009).

References

Boomsma, W., Mardia, K.V., Taylor, C.C., Ferkinghoff-Borg, J., Krogh, A., and Hamelryck, T. (2008). A generative, probabilistic model of local protein structure. *Proceedings of the National Academy of Sciences*, **105**, 8932-8937.

Frellsen, J., Moltke, I., Thiim, M., Mardia, K., Ferkinghoff-Borg, J., and Hamelryck, T. (2009). A probabilistic model of RNA conformational space. *PLoS computational Biology, in Press.*

Mu, Y., Nguyen, P.H., and Stock, G. (2005). Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *PROTEINS: Structure, Function, and Bioinformatics*, **58**, 45–52.

Altis, A., Nguyen, P.H., Hegger, R., and Stock., G. (2007). Dihedral angle principal component analysis of molecular dynamics simulations. *The Journal of Chemical Physics*, **126**, 244111.