Poster I-24

How parallel are the helices of your favourite protein?

Authors:
Peter Røgen (*Department of Mathematics, Technical University of Denmark*)
Per W. Karlsson (*Department of Mathematics, Technical University of Denmark*)

Short Abstract: How parallel and anti-parallel are the helices of your favourite protein? - And which other proteins have a similar parallelity of their helices? To answer such questions, we introduce a notion of Local Self Parallelity of protein secondary structures, and use it for automatic classification of protein structures.

Long Abstract:
How parallel and anti-parallel are the helices of your favourite protein? - And which other proteins have a similar parallelity of their helices?

We quantify the contents and the relative positions of the secondary structure elements of a protein structure in a way that is continuous and smooth under deformation of the protein structure. To do this we introduce:

1) a notion of Local Self Parallelity of a space curve and
2) a real-valued smooth (alpha,beta,coil)-coloring of the backbone giving
3) the alpha-alpha, the alpha-beta, the alpha-coil, the beta-alpha etc.
Local Self Parallelity of a protein chain.

Hereby we can quantify how parallel the helices and the sheets of a protein are, or e.g. how parallel the sheets and the "coil"-regions of the protein are, etc. Using the N-to-C direction of the backbone, we furthermore distinguish between the helices lying upstream and downstream relatively to the sheets. That is: To the alpha-beta Local Self Parallelity only the helices lying upstream relatively to the sheets contribute and to the beta-alpha Local Self Parallelity only the helices lying downstream relatively to the sheets contribute.

The algorithm starts with the carbon alpha curve given by a pdb-file. This curve is then smoothened by a local linear filter [1,2] and the (alpha,beta,coil)-coloring is given as a spline-function of neighbour and next-neighbour distances on the smoothened backbone. Using the (alpha,beta,coil)-coloring as weights the measures of Local Self Parallelity are linear maps from the distance matrix of the smoothened backbone. The calculation time is thus very short (> 1000 domains per minute including reading the pdb-files).

As functions of the distance matrix, these structural measures are blind to chirality. It is thus a bit surprising that a Jack knife test shows that they can automatically classify 96% of 18861 connected CATH-domains correctly which is the same rate of success as the highly chirality sensitive Gauss integrals was reported to have in [3]. However the error rate of 1% here is a lot higher than the Gauss integrals 0.02% error rate reflecting that eg. four helix bundles come in a right-handed and in a left-handed tertiary structure which is indistinguishable by

these distance matrix based structural measures.

[1] P. Røgen, Evaluating protein structure descriptors and tuning Gauss integral based descriptors J. Phys.: Condens. Matter 17 (2005) S1523-S1538928.

[2] K. Lindorff-Larsen, P. Røgen, E. Paci, M Vendrusscolo & C.M. Dobson, Protein folding and the organization of the protein topology universe, Trends in Biochemical Sciences, 30(1), 13-19. 2005.

[3] P. Røgen & B. Fain, Automatic classification of protein structure by using Gauss integrals, PNAS, 100(1), 119-124, 2003.