New insights into coiled coil formation by means of support vector machines

Ingrid G. Abfalter[1*], Carsten C. Mahrenholz[2*], Ulrich Bodenhofer[1] and Sepp Hochreiter[1]

[1]Institut für Bioinformatik, Johannes Keppler Universität, Altenbergerstrasse 69, 4040 Linz
[2]Institut für Medizinische Immunologie, Charité Universitätsmedizin Berlin, Hessische Straße 3-4, 10115 Berlin

* These authors contributed equally to this work.

Coiled coil proteins consist of two up to seven α-helices that are wrapped around each other. They can associate as either homomeric or heteromeric structures and bind in parallel or antiparallel topologies. [1]. One of the important functions of these proteins is playing an integral role in the infectious mechanism of viruses like bird flu, swine flu or HIV, as they enter host cells using coiled coil fusion proteins. Such fusion proteins are highly conserved, hence, they are considered a very attractive target for the development of anti-viral drugs [2].

While structure and occurrence of coiled coils are well known, the rules for oligomeric formation, the key to biological function and a prerequisite for rational drug design, are poorly understood. In order to be able to comprehend and manipulate coiled-coil formation, we determine specific properties that influence the oligomerization of coiled coil proteins, i.e. how many α-helices form the coiled coil or which heptad positions are essential for each structural type.

To find rules that determine oligomerization, we first apply support vector machines and statistical methods to classify dimers and trimers on the basis of their amino acid sequences. The data set for this classification task was collected by searching the entire RCSB Protein Data Base (PDB) for coiled coil structures, extracting the according amino acid sequences and sorting them into types of oligomers based on properties of their 3D structures. We then extract important features like specific amino acids at certain key positions or amino acid patterns that are characteristic for each type of oligomer. Amino acid patterns are retrieved from traditional string kernels and from a new kernel detecting amino acid co-occurrences. The relevance of the selected patterns is ensured by statistical tests and the excellent classification results measured by cross-validation.

The rules derived of these patterns will then be combined and used for designing coiled coil amino acid sequences with desired stoichiometry. We determine the actual oligomerization state of our newly constructed sequences by testing them experimentally through SPOT technique, peptide synthesis and biophysical verification [3].

References

1. A.N. Lupas, M. Gruber. *Adv Protein Chem*. **710**, 37-78. (2005).
2. H.M. Strauss , S. Keller. *Handb Exp Pharmacol.* **186**, 461-82. (2008).
3. C.C. Mahrenholz*, Z. Fidan*, M. Portwich, R. Volkmer. *Chem Today*. **26**, 22-25. (2008).