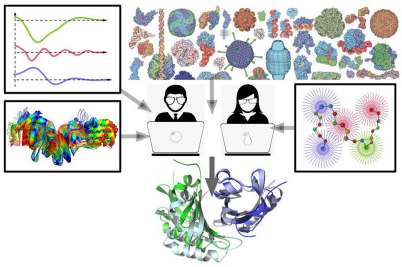

Contribution ID: **20** Type: **Oral**

How to combine sequence and structure information in multiple protein alignements?

Multiple protein sequence alignments are used daily in bioinformatics to annotate and predict the characteristics of currently mass produced sequences. The quality of their results have been assessed many times and have recahed a plateau. Proteins fold into stable three-dimensional structures with a topology much more conserved than sequence. Consequently, it should be advantageous to use this other source of information to align the sequences, in order to find the homologous positions. Several programs have been developped to align proteins according to their structure or to their sequence and their structure. In this study, we wanted to assess the added value of structural information in multiple alignments and compared the results of these programs to the results of the sequence alignment programs.
We compared the multiple alignments resulting from 25 programs either based on sequence, structure, or both, to reference align-ments deposited in five databases (BALIBASE 2 and 3, HOMSTRAD, OXBENCH and SISYPHUS). On the whole, the structure-based methods compute more reliable alignments than the sequence-based ones, and even than the sequence+structure-based programs whatever the databases. Two programs lead, MAMMOTH and MATRAS, nevertheless the perfor- mances of MUSTANG, MATT, 3DCOMB, TCOFFEE+TM ALIGN and TCOFFEE+SAP are better for some alignments. The advantage of structure-based methods increases at low levels of sequence identity, or for residues in regular secondary structures or buried ones. Concerning gap management, sequence-based programs set less gaps than structure-based programs. Concerning the databases, the alignments of the manually built databases are more challenging for the programs.

Primary author: CARPENTIER, Mathilde (ISYEB - MNHN - SU - CNRS - EPHE)

Presenter: CARPENTIER, Mathilde (ISYEB - MNHN - SU - CNRS - EPHE)

Session Classification: Session 4