Protein structure comparison in 3D based on secondary structure matching (SSM) followed by C_α alignment, scored by a new structural similarity function.

E. Krissinel and K. Henrick

European Bioinformatics Institute, Genome Campus, Hinxton,
Cambridge CB10 1SD, United Kingdom

SSM, a new tool for protein structure comparison and recognition in 3D, developed in EBI, is available at http://www.ebi.ac.uk/msd-srv/ssm . The structure comparison is done in two stages. In the first stage, an original graph-matching algorithm is applied to the mapping of protein secondary structure elements, with their positions and orientations in 3D taken into account. In the second stage, the mapped SSEs are used as a starting point for the precise iterative 3D alignment of the protein backbone C_α atoms. The alignment algorithm maximises an empirical structural similarity function Q, which controls the balance between alignment length N_a and RMSD achieved.

Comparison of SSM with similar resources available in the Net (DALI, VAST, CE) shows a good agreement to the degree of difference between all of them. While the difference in N_a and RMSD often increases significantly with decreasing structural similarity, the quality function Q shows a remarkable agreement. These results imply that simple measures like alignment length and RMSD do not give a sufficiently good indication of structural similarity.

SSM performance allows for structural studies on the scale of the whole PDB archive (currently about 20,000 structures with more than 40,000 chains). We used SSM for studying the relationship between structure and sequence similarities in the course of cross matching of all PDB chains. It appears that 80% of protein backbone atoms superpose with RMSD of 2.5 Å or better at sequence identity S as low as 20%. Combined analysis of N_a , RMSD, Q and statistical significance of matches suggests that S<20% is a solid indication of structural dissimilarity. The structural similarity function Q shows a better correlation with S than either N_a or RMSD and also was found to be a better measure of structural similarity. Our results show that reliable automation of protein structure recognition is achievable only with an appropriate score function, which essentially expresses the accepted definition of structural similarity.

SSM was developed to run on UNIX platforms, either as CGI or standalone application, optionally using many-CPU clusters for parallel processing. A typical query (comparison of a 200-300 residue protein with the whole PDB or SCOP archives) normally takes less than a minute. SSM is available for download and in-house installation under both academic (free) and commercial licenses.

References

1. E. Krissinel, K. Henrick, Secondary Structure Matching (SSM), a tool for protein structure alignment in 3D. Submitted to *Proteins: Structure, Function and Genomics*.
2. E. Krissinel, K. Henrick. Common subgraph isomorphism detection by backtracking search. Submitted to *Software – Practice and Experience*.