

# Dimensional reduction and clustering of class A G-protein-coupled receptors

Julien Pelé\* and Marie Chabbert

UMR CNRS 6214 - INSERM U771, Faculté de médecine d'Angers, 49045 ANGERS, FRANCE

\* Contact author: julien.pele@etud.univ-angers.fr

**Keywords:** Dimensional reduction, Clustering, Biological objects

In protein families, amino acid sequences reflect the evolutionary history that led to biological diversity. This available information remains to be unveiled from multiple sequence alignments. Clustering sequences into groups of similar features can provide clues to the evolutionary relationship between them. A solution to cluster objects is to generate a space by dimensional reduction. Objects correspond to points whose mutual distances depend on the metric.

To carry out our project, we used the programming language Perl and the statistical environment of R. We analyzed class A G-protein-coupled receptors from five species. Distance matrices were generated from multiple sequence alignments by different similarity measures. They were analyzed by two statistic techniques: metric multidimensional scaling (`cmdscale` in R package **stats**) and principal component analysis (`dudi.pca` in R package **ade4**). Biological objects were grouped by k-means (`Kmeans` in R package **amap**). Groups were validated with the **ciValid** functions from the R package. The clustering analysis was bootstrapped in order to assess the grouping robustness. Best parameters were determined in relation with biological meaningfulness. Our approach indicates that receptors from different genomes share a similar clustering pattern that might relate to major evolutionary determinants of class A G-protein-coupled receptors.

## References

Grishin VN, Grishin NV. (2002). Euclidian space and grouping of biological objects. *Bioinformatics*, 18, 1523-34.

Casari G, Sander C, Valencia A. (1995). A method to predict functional residues in proteins. *Nat Struct Biol.*, 2, 171-8.