

Parallel tiered clustering for large data sets using a modified Taylor's algorithm

J. MacCuish¹

N. MacCuish¹, M. Chapman¹

¹*Mesa Analytics & Computing, Inc., Santa Fe, New Mexico, USA*

Clustering large sets has many applications in drug discovery, among them compound acquisition decisions and combinatorial library diversification. Molecular fingerprints (2D) and molecular shape conformers (3D) from PubChem are the basic descriptors comprising the large sets utilized in this study. A parallel tiered clustering algorithm, implementing a modified Taylor's algorithm, will be described as an efficient method for analyzing datasets of such large scale. Results will be presented in SAESAR (Shape And Electrostatics Structure Activity Relationships).