Protein structures may share the geometry of local folds in cases of distant homology or even if totally unrelated. BORGES is a software that builds customized libraries of composite secondary structure fragments extracting them from all possible real structures deposited in the PDB. Description of the folds is independent of the primary sequence, but relies on a novel algorithm based on distribution of characteristic vectors computed over the mainchain coordinates atoms. Filtering out models with no significant differences expected in terms of phasing reduces the size of the library without loss of generality. BORGES libraries are used within a novel procedure combining rotation clustering, rotation refinement, and model refinement (BORGES) with the model location (PHASER[1]), density modification and autotracing (SHELXE[4]), extending the method underlying the previous ARCIMBOLDO[2,3,5], which exploited enforcing secondary structure to constrain phasing. BORGES exploits cloud computing technologies (in particular middleware CONDOR[6] and SGE[7]) to parallelize jobs, building up a dynamic multisolution frame in which decisions are taken in running time to lead towards the final structure of the protein. Usability is considered providing the user an easy graphical input/output interface, this also includes the possibility to remote access to external supercomputing center for cloud computing. With this new approach three novel structures were so far solved, where both the classical ARCIMBOLDO and other software had failed, and many test cases of both alpha and beta structures.

[7] https://arc.liv.ac.uk/trac/SGE

Keywords: Model Libraries, Phasing methods, Bioinformatics