3D Retrieval and Classification for Structural Bioinformatics with implications to Biological Image Databases

Afzal Godil[1] and Talapady N. Bhat[2]

[1]Information Access Division, ITL, [2]Biochemical Science Division, CSTL, National Institute of Standards and Technology, Gaithersburg, MD 20899.

It is widely believed that the 3D shapes of macro molecules and their active sites provide a discriminating role in bio-molecular recognition and function. Geometrical shapes determine their ability to bind to their targets. Characterization of geometrical shape may thus provide information to classify and retrieve related and functionally relevant macro-molecules for purposes such as drug targeting. There are over 46,000 protein structures in the Protein Data Bank (PDB). These 3D structural databases pose challenges for storing, indexing, searching, clustering, retrieval of shape based structural information. Techniques used in text based retrieval of structural information may not be easily extended to shape based 3D or 2D searches that require surface-based and volume-based descriptors to effectively characterize the shape, semantics and geometric topology. Hence there is a need for an automated rule-based 3D retrieval and classification system to efficiently manage Structural Bioinformatics Databases. Some of these requirements may be met by the existing technology used in 3D shape searching for instance in Computer Aided Design (CAD) as illustrated here using a Web interface (fig 1) that facilitates query on images. We have developed a shape based retrieval and classification method for a few of the structures taken from the PDB. The method involves developing 3D shape descriptors to describe the 3D shape of each structure. The shape descriptors that we have developed are based on histograms of distances between atoms, moments and Spherical harmonics of the surface of the molecules. We have used this method to develop proximity measures for structures that may be used for assessing their similarity.

Though we have not tested this method for 2D images of biological objects such as those from live cells, we believe that the technique may be applied to them as well. For instance, one may use the above mentioned shape descriptors to identify cellular objects and then use their similarity matrix to compare them. This similarity matrix may be used to develop shape based taxonomy (fig 1) to be used in a Web browser that displays and facilitates query on cellular images and on the features derived from them. Our efforts towards his effort also will be presented.
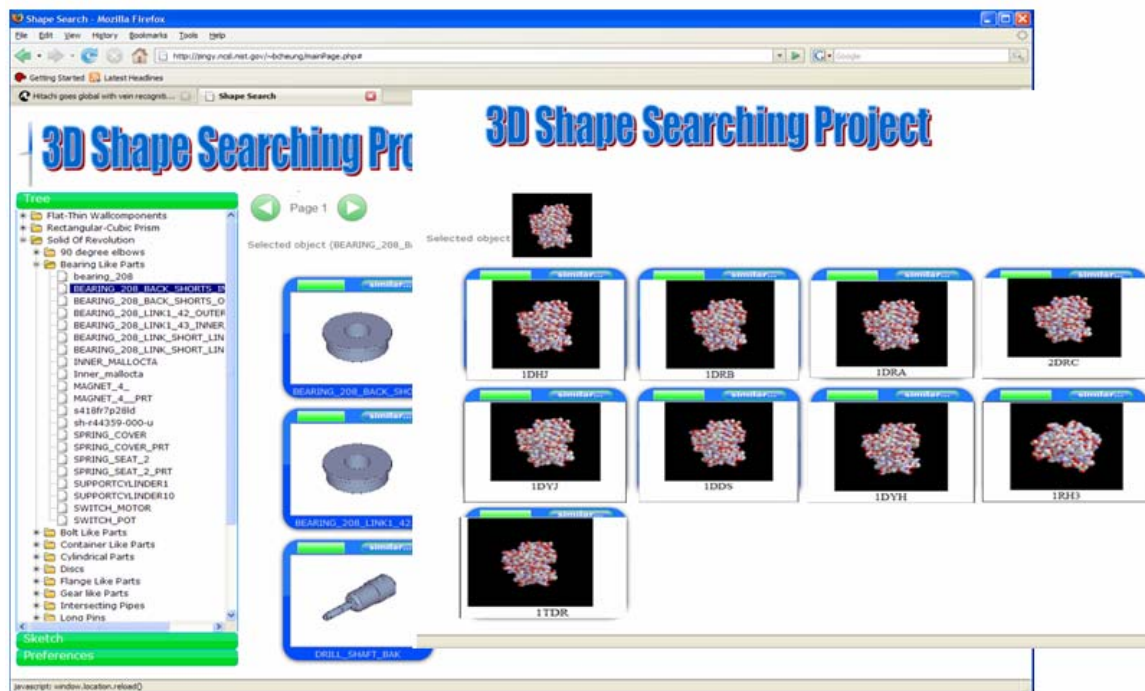
Figure 1. 3D Shape Searching from CAD and Structural Bioinformatics Database