Example of use (Use case)

The use of Geomfinder begins with the selection of the protein structures. The user can set the PDBid of the Protein Data Bank database and the query protein will be downloaded automatically. Alternatively, the user can upload a protein structure generated by *in silico* comparative modeling. Once this process is finished, Geomfinder carries out a PDB validation process. If the files are correct, all chains of the two protein structures are captured (Fig. 1).

Fig. 1: Selecting the protein structures to compare with Geomfinder

When the processing of each PDB files is completed, the user must specify some parameters, such as: the chain (on each protein), the near and far threshold, and the filters, among others, as is shown in Fig. 2.

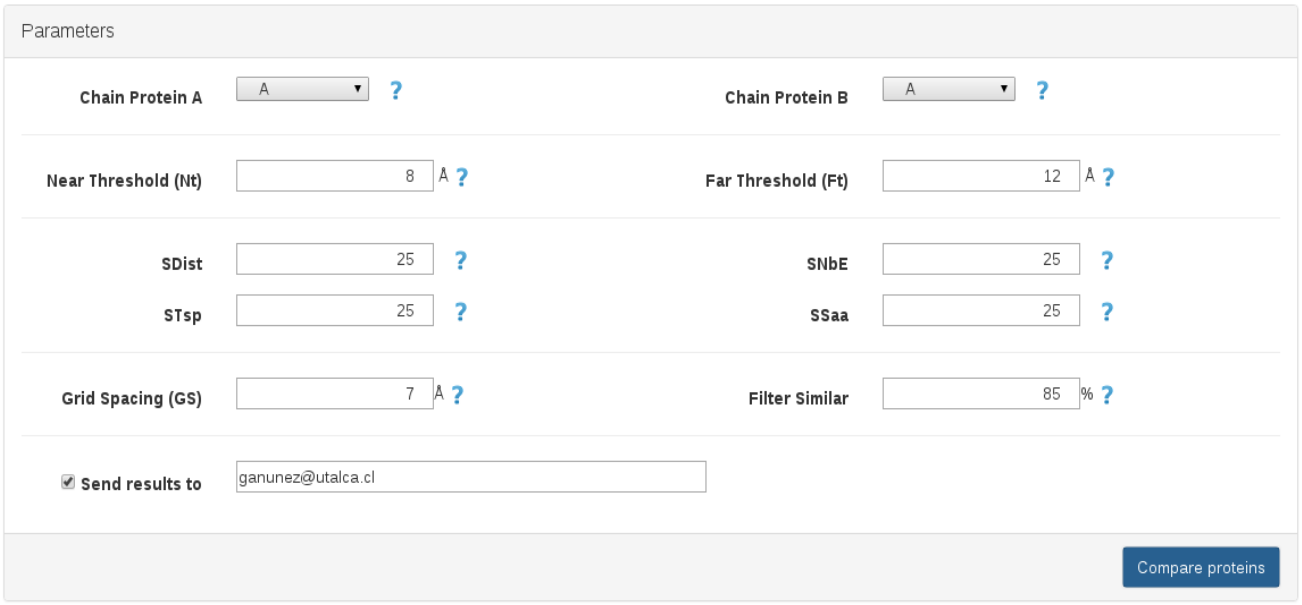

Fig. 2: Setting the parameters of Geomfinder

As described in the paper, the parameters used relate to a research question. In this example, we used the recently resolved X-ray structure of the *Drosophila melanogaster* dopamine transporter (DAT) (PDBid: 4M48). The second protein selected was the structure of human monoamine oxidase B (MAO-B) (PDBid: 1GOS). Both proteins are considerably different and while DAT is a transmembrane protein, MAO-B is a mitochondrial outermembrane flavoenzyme. Nevertheless, these proteins interact with the same endogenous ligand (dopamine), so it is reasonable to conclude that some structural features of their dopamine binding sites would be shared.

In this example, we are only interested in highly similar 3D patterns, so the "Filter Similarities" parameter was set to 85% (only the pairs of 3D patterns with a Gscore higher or equal than 85% are shown). To calculate the final similarity score (GScore), the contributions of the partial similarities of the distance, non-bonded energy, perimeter and the sequence component were set equally as 25%.

The parameters used are shown in the Fig. 3.

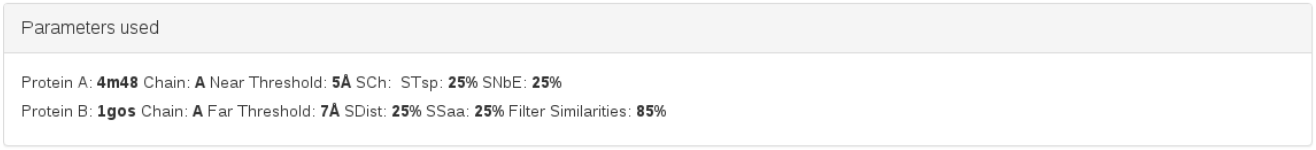

Fig. 3: Parameters used

Here we can analyze each pair of 3D pattern similarities scores (yellow box). The dynamic feature of the data tables allow us to sort, filter and search the results (red boxes). By default, all the pairs of 3D patterns are sorted in descending order based on their GScore.

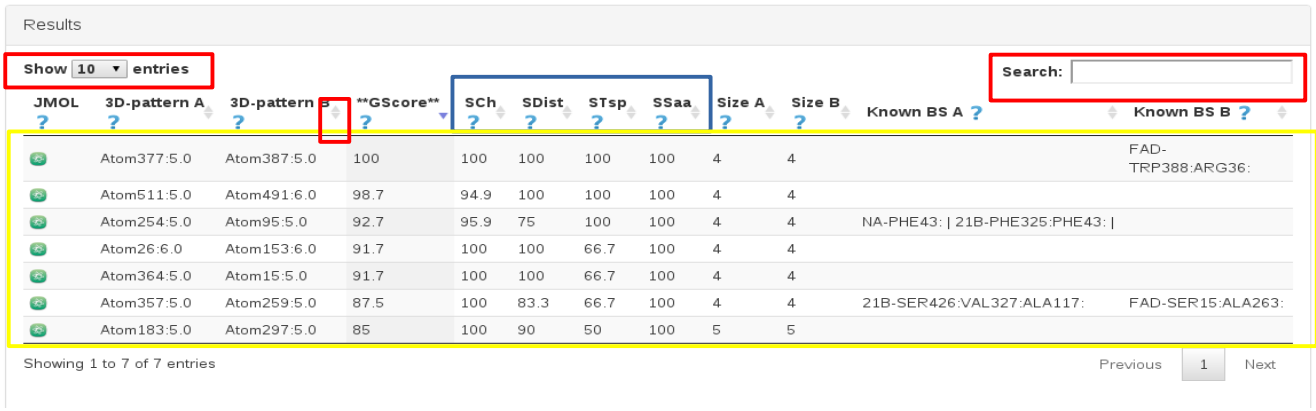

Fig. 4: Main results page

In this example, Fig. 4 shows seven pairs of 3D patterns identified with a GScore higher than 85%. Additionally, it is possible to observe the partial similarities of Non-Bonded energy, distance, perimeter and sequence component. The most similar pair of 3D patterns found was that denoted by the atom377 label in the protein A (4M48) and the atom387 label in the protein B (1GOS). Here, the GScore was 100%. Additionally, we can observe the size (number of residues) of each of the 3D patterns.

If we are interested in a particular pair of 3D patterns, we can click on the green button displayed in Fig. 4. Four 3D visualization windows will be shown below the table results. The first two frames (Fig. 5) show the two 3D patterns that have been identified and selected. The second two frames (Fig. 6) show a secondary structure representation of the query proteins. The green spheres show the location of the residues forming the 3D pattern that has been identified. All these files can be downloaded as PDB formatted files (green boxes in Fig. 5).

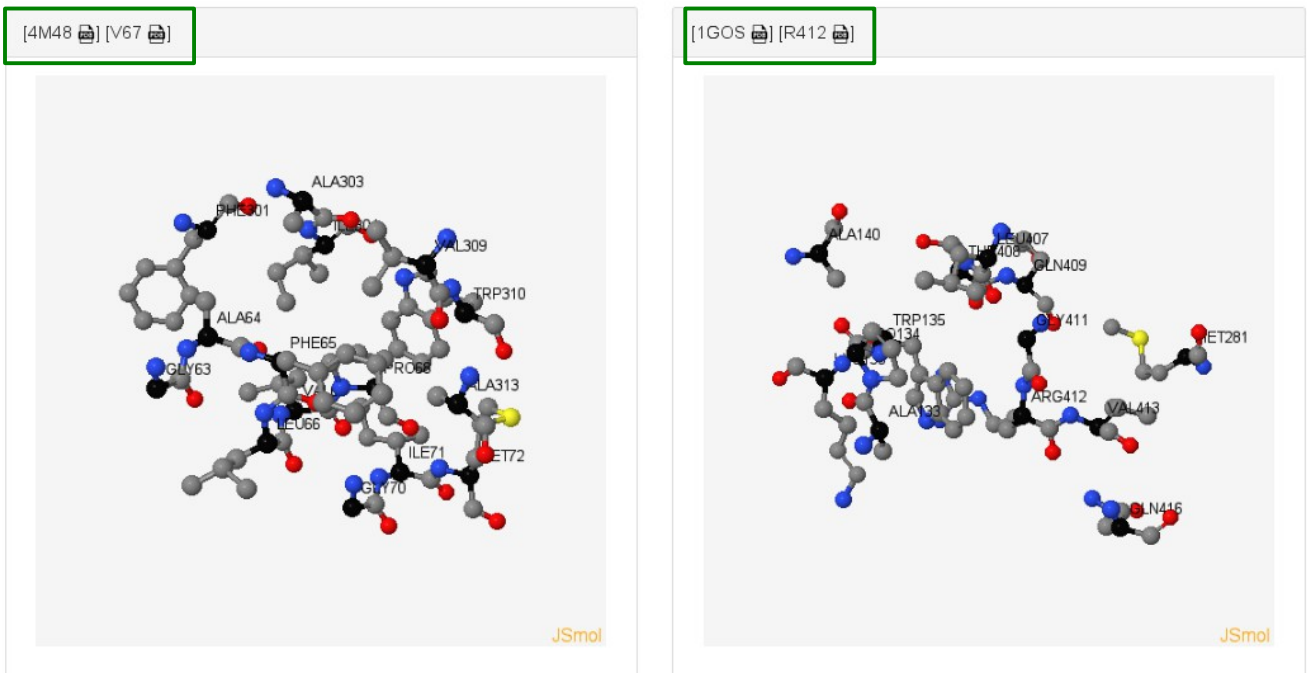

Fig. 5: Residues forming a pair of similar 3D pattern identified

Fig. 6: Residues forming a pair of similar 3D pattern identified located on each source protein