Geomfinder: A multi-feature identifier of similar three-dimensional protein patterns: A ligand-independent approach

Núñez-Vivanco G.

Valdés-Jiménez A.

Besoaín F.

Reyes-Parada M.

Background: Since the structure of proteins is more conserved than the sequence, the identification of conserved three-dimensional (3D) patterns among a set of proteins, can be important for protein function prediction, protein clustering, drug discovery and the establishment of evolutionary relationships. Thus, several computational applications to identify, describe and compare 3D patterns (or motifs) have been developed. Often, these tools consider a 3D pattern as that described by the residues surrounding co-crystallized/docked ligands available from X-ray crystal structures or homology models. Nevertheless, many of the protein structures stored in public databases do not provide information about the location and characteristics of ligand binding sites and/or other important 3D patterns such as allosteric sites, enzyme-cofactor interaction motifs, etc. This makes necessary the development of new ligand-independent methods to search and compare 3D patterns in all available protein structures. Results: Here we introduce Geomfinder, an intuitive, flexible, alignment-free and ligand-independent web server for detailed estimation of similarities between all pairs of 3D patterns detected in any two given protein structures. We used around 1100 protein structures to form pairs of proteins which were assessed with Geomfinder. In these analyses each protein was considered in only one pair (e.g. in a subset of 100 different proteins, 50 pairs of proteins can be defined). Thus: (a) Geomfinder detected identical pairs of 3D patterns in a series of monoamine oxidase-B structures, which corresponded to the effectively similar ligand binding sites at these proteins; (b) we identified structural similarities among pairs of protein structures which are targets of compounds such as acarbose, benzamidine, adenosine triphosphate and pyridoxal phosphate; these similar 3D patterns are not detected using sequence-based methods; (c) the

detailed evaluation of three specific cases showed the versatility of Geomfinder, which was able to discriminate between similar and different 3D patterns related to binding sites of common substrates in a range of diverse proteins. Conclusions: Geomfinder allows detecting similar 3D patterns between any two pair of protein structures, regardless of the divergency among their amino acids sequences. Although the software is not intended for simultaneous multiple comparisons in a large number of proteins, it can be particularly useful in cases such as the structure-based design of multitarget drugs, where a detailed analysis of 3D patterns similarities between a few selected protein targets is essential.

Pattern

Protein-structure

Similarity