Similarity search for local protein structures at atomic resolution by exploiting a database management system

A method to search for local structural similarities in proteins at atomic resolution is presented. It is demonstrated that a huge amount of structural data can be handled within a reasonable CPU time by using a conventional relational database management system with appropriate indexing of geometric data. This method, which we call geometric indexing, can enumerate ligand binding sites that are structurally similar to sub-structures of a query protein among more than 160,000 possible candidates within a few hours of CPU time on an ordinary desktop computer. After detecting a set of high scoring ligand binding sites by the geometric indexing search, structural alignments at atomic resolution are constructed by iteratively applying the Hungarian algorithm, and the statistical significance of the final score is estimated from an empirical model based on a gamma distribution. Applications of this method to several protein structures clearly shows that significant similarities can be detected between local structures of non-homologous as well as homologous proteins.

存档文本