

Computational Screening of Novel Enzymes Using OCR-based Protein Fingerprint

이선구[†], Amit Goyal, 황규석

부산대학교

(sungulee@pusan.ac.kr[†])

Design of an efficient fingerprint that detects homologous proteins at distant sequence identity has been a great challenge. We have developed a strategy to extract an ideal-like fingerprint with high specificity and sensitivity from a group of sequences related to a fold. The approach is devised based on the assumptions that the critical residues for a protein fold may be conserved in three aspects, i.e. sequence, structure, and intramolecular interaction, and embedded in secondary structures. The fingerprint was generated by selecting the overlapped conserved residues (OCR) from the conserved residues of homologous proteins obtained using independent three alignment methods, i.e. multiple sequence alignment, structure-based alignment, and alignment based on the interstrand hydrogen-bonds. The OCR-based fingerprints were tested on protein folds of various classes with discrete sequence similarities, and showed more than 90% detection efficiency for all the folds tested. Finally, the approach was employed in the computational screening of novel enzymes such as Cupredoxin proteins and Cytochrome P450 proteins from protein sequence database.