A statistical model for measuring binding site similarity

John R. Davies*1, Richard M. Jackson2, Charles C. Taylor1, & Kanti V. Mardia1

1 Department of Statistics, University of Leeds
2 School of Molecular Biology and Biochemistry, University of Leeds

We examine the problem of measuring the similarity of protein-ligand binding sites using a newly developed score, the Poisson Index (PI). PI is based upon a distribution of the random matching of two configurations of points in three dimensional space derived in Green & Mardia (2006). The model simulates a set of points taken from the superpopulation that are distributed as a homogeneous Poisson process. Using this model the PI examines the probability that two sites (size m and n with \( m \leq n \)) will match the observed \( L \) matching atoms or better by chance.

PI requires only three parameters (\( m, n \) and \( L \)) to be calculated and is a reasonably quick and efficient method of capturing binding site structural similarity. We investigate the ability of PI to measure similarity in protein-ligand binding site matching.

Reference