Exploring protein binding site similarity

Richard M. Jackson

Institute of Mollecular and Cellular Biology, University of Leeds

The large-scale comparison of protein-ligand binding sites is problematic, in that measures of structural similarity are difficult to quantify and are not easily understood in terms of statistical similarity that can ultimately be related to structure and function. We present a binding site matching score the Poisson Index (PI) based upon a well-defined statistical model. PI requires only the number of matching atoms between two sites and the size of the two sites-the same information used by the Tanimoto Index (TI), a comparable and widely used measure for molecular similarity. Despite the difficulty of determining a biological 'ground truth' for binding site similarity we conclude that PI is a suitable measure of binding site similarity and could form the basis for a binding site classification scheme.

We have recently undertaken a large-scale comparison of protein kinase ATP-binding sites. This has allowed us to discover binding site similarity in different sub-families of protein kinase that are not evident from sequence similarity alone. We propose a relevant classification of the protein kinase family based on the similarity of their binding sites. Not only does this classification highlight features that are important for the potency and selectivity of kinase inhibitors, but it also allows us to rationalise cross-reactivity among the protein kinases.