PocketPicker: Introduction of a Pocket Prediction Method for Ligand Binding Site Analysis with Shape-Descriptors

M. Weisel1,2, E. Proschak1, J. Kriegl2, G. Schneider1
martin.weisel@bioinformatik.uni-frankfurt.de

Introduction
We developed a new grid-based pocket prediction method PocketPicker1,2 for the identification of potential protein binding sites. Predicted pockets are visualized as sets of grid points colored by their respective accessibility. Furthermore we developed a new shape-descriptor that encodes the size and shape of a potential binding site with respect to its accessibility. This descriptor was suited for automated comparative binding site analysis in aldose reductase.

Prediction of Binding Pockets
• The pocket detection routine of PocketPicker employs a sophisticated scanning process and comprises the calculation of “buriedness indices” for probes installed in the grid.
• Buriedness is calculated for probes installed in areas close to the protein surface. Clustering is applied for adjoining probes (Fig. 1).

Clustering of Grid Probes
• A three-dimensional Flood-fill algorithm was applied to identify neighboring grid probes and group them into disjoint clusters.
• Geometric neighborhood-search was used to prevent the grouping of spacially close concavities (Fig. 2).

Evaluation of Pocket Prediction
• Performance of binding site detection was tested on a data set of 48 ligand-receptor complexes and their 48 unbound apo-forms[3].
• Correct predictions were termed “TOP1-hits”, predictions with the respective ligand residing within the three largest pockets were termed “TOP3-hits”.
• Prediction results match performances of common methods (Fig. 3).

Design of a Shape-Descriptor regarding Pocket Buriedness
• A descriptor was designed to describe the shape of a pocket with respect to the buriedness of the site.
• Grid probes were grouped into six categories A, B, C, D, E, F holding grid point coordinates with ascending buriedness values (Fig. 4).

Clustering of Grid Probes
• The shape-descriptor was designed to record the appearance of distances between these categories. Distances are stagecd in 20 distance bins covering ranges up to 20 Å (Fig. 5).

Shape-similarity is expressed as Euclidean Distance d between the shape-descriptors r and s of two pockets:

\[ d = \sqrt{\sum_{i=1}^{n} (r_i - s_i)^2} \]

References

Supported by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften