Modelling receptor binding sites

Andrea Schafferhans, Gerhard Klebe

Department of Pharmaceutical Chemistry, Philipps-University Marburg,
Marbacher Weg 6, D-35032 Marburg, Germany
(andrea.schafferhans@mailer.uni-marburg.de)

In structure based drug design protein modelling and QSAR (Quantitative Structure Activity Relationships) are currently used as two completely independent methods to approach the same problem: gaining information about an unknown receptor structure in order to allow the design of suitable ligands. Protein modelling approaches the problem by extracting common features from homologous representatives of the protein class the target protein belongs to and maps the sequence of the protein under consideration onto these common features. QSAR in contrast uses ligand-information to extract 3D-features responsible for binding affinity. Thus, homology modelling mainly contributes general information about the protein conformation and binding site geometry that is rather unspecific for a particular receptor, whereas QSAR implicitly contains this specificity information through the ligand binding data, however it knows nothing about the surrounding protein structures. In this study we explore ways of combining the information obtained by both methods in order to generate a more detailed and reliable model.

The first crucial step in combining ligand and protein information is the alignment of individual ligands or a pharmacophore model (e.g. one used and validated in QSAR) onto a preliminary protein model. Since structural information is still quite diffuse in the incipient stage, one may not rely too strongly on directional interactions (hydrogen bonds) as most docking tools do. In addition, the representation of ligands and protein should reflect and cope with the still given uncertainties of the model-built protein structure. Therefore, we represent structural information in terms of Gaussian functions associated with volume[1], lipophilic[3], hydrogen bond donor and acceptor properties. This representation has the advantage that it allows the merging of several protein models into a combined one using geometric or arithmetic averages over the different Gaussian-type representations in order to consider conformational alternatives simultaneously.

The search for the individual ligand orientations matched onto this binding site representation is performed by a modified version of the program SEAL[4,5]. Since SEAL always suggests several possible solutions for each ligand, criteria to select the most likely solutions have to be defined. To select the most convincing orientation from these multiple solutions, a similarity measure among the considered ligands in the various computed orientations is calculated. The most consistent and best scored alignment among the ligands is finally extracted. The resulting ligand alignment is then used to calculate a QSAR model; its results can subsequently be fed back into the protein modelling process (e.g. considering volume constraints).

We will present results from the evaluation of the described method using homology-built thrombin structures and a set of thrombin inhibitors as a test case.

References:

[1] Points representing positive volume properties are possible interaction sites inside the binding cavity (e.g. as calculated by LUDI[2]), while the protein atoms are included as points associated with a negative volume.
[2] H.-J. Böhm, *J. of Comput.-Aided Mol. Design* **6**, 61-78 (1992)
[3] Lipophilic properties are located at lipophilic interaction sites (also calculated by LUDI).
[4] G. Klebe, T. Mietzner, F. Weber, *J. of Comput.-Aided Mol. Design* **8**, 751-778 (1994)
[5] S. K. Kearsly, G. M. Smith, *Tetrahedron Comput. Methodol.* **3**, 615 (1990)