[P4] Prediction of protein kinase-ligand interactions through 2.5D Proteochemometrics

Fabrice Carles[1], Nicolas Bosc[1], Pascal Bonnet[1], Christophe Meyer [2]

[1]Institut de Chimie Organique et Analytique, UMR CNRS-Université d'Orléans 7311, Université d'Orléans BP 6759, 45067 Orléans Cedex 2, France
[2]Janssen-Cilag, Centre de Recherche Pharma, CS10615 - Chaussée du Vexin, 27106 Val-de-Reuil, France

The human genome encodes 518 proteins from the kinase family[1]. They are all involved in protein phosphorylation, a regulation mechanism of many biological processes. In the event of a perturbation in their regulation pathway, or a mutation of their nucleic acid sequence, a variation of the protein kinase expression may lead to cancer, diabetes or inflammatory diseases. As a consequence, significant efforts have been made to uncover and design new potent and selective protein kinase inhibitors.

Because the ATP binding site is highly conserved throughout the kinase family, designing selective inhibitors remains a challenge for the pharmaceutical industry. Routine screening and profiling on broad kinase panels is producing an increasing amount of data. Proteochemometrics modelling[2] is a computational technique which combines both ligand and target information within a single predictive model of the biological activity. Experimental data serves to train machine learning algorithms and the resulting model is used to predict the activity of unknown compounds.

Here we present a 2.5D proteochemometrics (PCM) approach where the protein kinase structures are described by a novel 3D descriptor whereas the ligands are encoded by a 2D fingerprint. Using two examples, we demonstrated that the protein descriptor successfully classifies protein kinases based on their group membership and on the conformation of their DFG motif. We also compared the performance of our models with those obtained from a full 2D PCM modeling. In both cases, the internal validation of the models demonstrated good capabilities to distinguish "active" from "inactive" protein kinase – ligand pairs. However, the external validation performed on an independent dataset showed that both models tend to overestimate the number of "inactive" pairs. We presumed that the unbalanced training set could explain this effect and we therefore modified the ratio of active versus inactive pairs to tackle this issue.

Our results suggest that characterizing protein-ligand interactions at the 3D level may provide PCM models with a superior predictive power as compared to full 2D models. Particular care must be given to the distribution of active/inactive protein – ligand pairs composing the training set.

Bibliography:

[1]G. Manning; D. B. Whyte; R. Martinez; T. Hunter; S. Sudarsanam. Science. 298 (2002) 1912–1934.
[2]M. Lapinsh; P. Prusis; A. Gutcaits; T. Lundstedt; J. E. Wikberg., Biochim. Biophys. Acta. 1525 (2001) 180–190.