Application of interaction patterns to discriminate ligand preference to target/antitarget protein

Jagna Witek[a,b], Krzysztof Rataj[a,b], Sabina Smusz[a,b],Tomasz Kosciolek[a], Stefan Mordalski[a], Andrzej J. Bojarski[a]

[a] Department of Medicinal Chemistry, Institute of Pharmacology Polish Academy of Sciences, 12 Smętna Street, 31-343 Kraków, Poland
[b] Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, 7 Gronostajowa Street, 30-007 Kraków, Poland
e-mail: jagna.witek@uj.edu.pl

Introduction

SIFts (Structural Interaction Fingerprints) are precise and rapid tool for binding site description. In this research, fingerprint describing physical ligand-protein interactions consists of 9-bit fragments providing information with residues and type of interaction.
A collection of such fingerprints created for certain ligands group can be further used in machine learning procedure, to generate an interaction profile. Such approach enables us to evaluate whether a given compound possibly exhibits desired activity.

In this project, interaction patterns are generated for active (Ki < 10 nM) and inactive (Ki > 1000 nM) ligands docked into our target structures. Two varying protein families were designated as targets: G-protein coupled receptors (GPCRs) class A, and protein kinases. Afterwards, compounds were docked to their crystal structures in order to generate training SIFt sets.
SIFts for pairwise selective ligands were then compared with appropriate profiles in order to evaluate if their activity profile can be properly recognized. Basing on complementarity between particular SIFt fingerprint and two ML profiles, one can select compounds with high affinity to target and low affinity to antitarget. Such an approach enables rapid and accurate prediction of ligand preference to protein, and also convenient data analysis.

Fingerprint preparation

The fist step of our study, was generating fingerprints or crystal structures of Serine/Threonine protein kinase Aurora-A (PDB ID: 1MUO) and Cyclin Dependent Kinase 2 (PDB ID: 4AAA), representing target and antitarget. Active compounds, retrieved from ChEMBL database, were docked to all receptor structures; on a basis of interactions in ligand-protein complexes, SIFts were generated.

SIFts enabled recognition of residues involved in ligand binding and furthermore, types of interactions between specific residues. In this research nine bits were used to describe those associations: any contact, backbone, side chain, polar, hydrophobic, hydrogen bond donor/acceptor, aromatic and charged (Fig. 1). The fingerprints were produced for every ligand-receptor complex retrieved.

SIFt analysis

Crucial stage in interaction analysis was applying machine learning to create interaction profiles of compounds displaying certain activity. Training sets were composed of several active and inactive compounds, and further method evaluation was performed on test sets (Table 1) by means of a SMO (Sequential Minimal Optimization) algorithm. Its performance was evaluated by recall (R – fraction of positives selected from test set), precision (P – correctness of positive instances prediction; low values indicate a high rate of false positives), and F-measure (gives balanced measure of machine learning methods performance (Table 2)).

Results and conclusions

Application of machine learning in SIFt analysis enabled discriminating ligand's preference to protein structure. This method proved to be useful in distinguishing between active an inactive compounds for single target (Table 2). Hovewer, it was originally dedicated to ligands with selective activity to target/antitarget protein, as a multitargeting tool. Unfortunately, it is impossible to apply this methodology to ligands with opposite activities to two targets with satisfying accuracy.

Presented method may be useful in discriminating ligand preference to receptor in easy and automated way. However, further evaluation of method, would allow to investigate its capabilities and limitations.

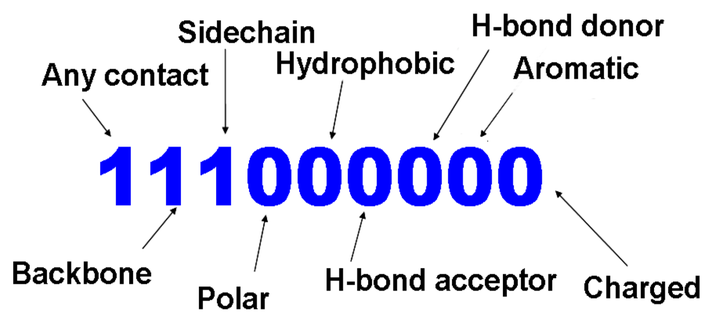

Figure 1. Fragment of SIFt describing bit positions for individual ligand-residue interactions.

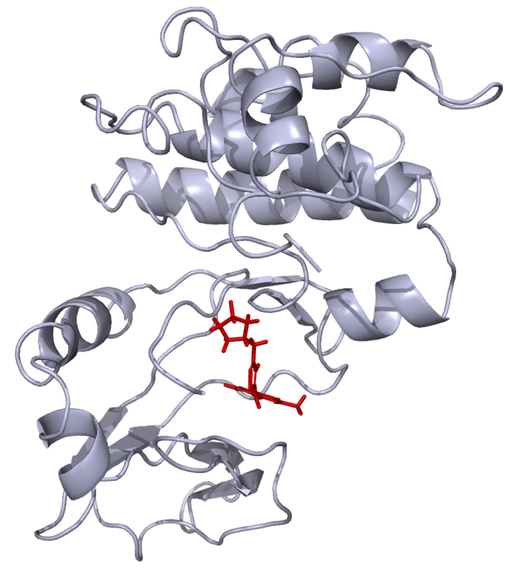

Figure 2. Crystal structure of Aurora-A kinase with active ligand docked.

$$R = \frac{TP}{TP + FN}$$

$$P = \frac{TP}{TP + FP}$$

$$F - measure = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Figure 3. Measures of machine learning performance.

R – recall

P – precision

TP – the number of true positives (correctly classified actives)

FP – the number of false positives (inactives wrongly classified as actives)

TN – the number of true negatives (correctly classified inactives)

FN – the number of false negatives (actives wrongly classified as inactives)

Target	Active compounds		Inactive compounds	
	Ligands № in training set	Ligands № in test set	Ligands № in training set	Ligands № in test set
Aurora-A	31	58	12	43
CDK2	31	124	18	77

Table 1. Composition of training and test datasets.

Target	Class	Recall	Precision	F-measure
Aurora-A	Active	0.966	0.918	0.941
	Inactive	0.884	0.95	0.916
CDK2	Active	0.984	0.961	0.972
	Inactive	0.935	0.973	0.954

Table 2. Evalutation of machine learning performance in predicting compounds activity.

Literature

(1)Deng Z, Chuaqui C, Singh J *„Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein-Ligand Binding Interactions”,* J. Med. Chem. 2004, 47, 337 -344
(2)Mordalski S, Kosciolek T, Kristiansen K, Sylte I, Bojarski A J *„Protein binding site analysis by means of Structural Interaction Fingerprint patterns”*

Acknowledgments

This study is supported by project UDA-POIG.01.03.01-12-100/08-00 co-financed by European Union from the European Fund of Regional Development (EFRD); http://www.prokog.pl

INNOVATIVE ECONOMY
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN FUND OF
REGIONAL DEVELOPMENT