# An application of ligand interaction profiles as a novel approach in virtual screening of GPCR ligands

**Jagna Witek, Sabina Smusz, Krzysztof Rataj, Stefan Mordalski, Dawid Warszycki, Andrzej J. Bojarski**

Department of Medicinal Chemistry, Institute of Pharmacology Polish Academy of Sciences, 12 Smętna Street, 31-343 Kraków, Poland
e-mail: jagna.witek@gmail.com

## Introduction

Cheminformatic methods, such as Virtual Screening (VS), constitute vital part of modern drug design. VS enables effective database mining, being particularly useful tool in search for ligands of desired activity. The scope of this study was to find an efficient methodology of automatic discrimination between active and inactive compounds for the given target, thus facilitating VS procedure.
In this research we present a workflow combining docking, application of interaction profiles for ligands created on the basis of structural interaction fingerprints (SIFt) and machine learning algorithms [1]. Due to the fact, that the method is strongly dependent on the target structure, it was examined against demanding conditions, being homology models of G-protein coupled receptors.

## Homology modeling

Homology models of beta1 and beta2 adrenergic receptors were built on crystal structures of class A GPCRs retrieved from the Protein Data Bank (**Table 1**). Sequences of modeled receptor and its template were aligned manually, assuring that the most conserved amino acid in each helix, and motifs characteristic for class A GPCRs are on matching positions. Ranges of the helices were determined on the basis of crystal structures. For each template, 100 models were built, and underwent validation by docking. Active ($K$i < 100 nM) and inactive ($K$i > 1000 nM) compounds towards each target, were retrieved from ChEMBL database. The compounds were clustered, and centroids of each cluster were selected for docking.

**Model selection**
For every model, an enrichment curve was calculated basing on the Glide Score values of docked compounds, including not docked actives and inactives as false negatives and true negatives, respectively. The final model quality was determined by the Area Under ROC curve (AUROC). Three models per template with the highest AUROC value were selected for further studies. To assess quality of homology models in this study, all the experiments were also performed for three crystal structures of beta1 and beta2 receptors.

In order to assess the ability of homology models and crystals to discriminate between active and inactive compounds, four sets of compounds were docked (**Table 2**). Active and inactive compounds were retrieved from ChEMBL database. The set of inactives was enriched by random selection of compounds from ZINC database, and generation of decoys according to the Directory of Useful Decoys approach [2].

## Fingerprint preparation

Structural Interaction Fingerprints (SIFts) enable recognition of amino acids involved in ligand binding and additionally, they provide information about types of interactions between specific residues [3]. In this research nine bits were used to describe following associations: any contact, backbone, sidechain, polar, hydrophobic, hydrogen bond donor/acceptor, aromatic and charged (**Figure 1**).
SIFts generated for each ligand docked into at least one of protein structures, were subsequently utilized to create SIFt profile [4]. It was performed by averaging all fingerprint strings obtained for each ligand into single profile, describing ligand's interaction pattern in simplified manner (**Figure 2**).

## Interaction profile analysis

Crucial stage of interaction examination, was application of machine learning algorithms to SIFt profiles.
Analysis was performed using Sequential Minimal Optimization algorithm in 10-fold cross validation experiment. Its performance was evaluated by MCC parameter, which provides a balanced measure of classification efficiency of machine learning methods (**Figure 3**).

## Results and discussion

Application of machine learning to SIFt analysis enabled discrimination between active and inactive compounds towards given target, when applied both to crystal structures and homology models (**Figure 4**).
Low values of MCC in distinguishing in case of true active and inactive compounds may result from insufficient representation of inactives in the set. Moreover, true inactive compounds are very difficult to discriminate due to the fact that they are extremely similar to actives. High MCC values in case of active and random or active and decoy compounds suggest that the method may be feasible for screening ligands in search of novel structures. What is more, in case of beta1AR, MCC values are higher for homology models than for crystal structures. It may imply that models are equally good or even more suitable for application of presented methodology.

Presented method may be useful in assessment of ligand's affinity towards target receptor structure, in case of paucity of experimental data. However, the most beneficial way to exploit this procedure would be determination of multitarget profile of ligand's interaction. Further evaluation may allow to investigate its capabilities and limitations.

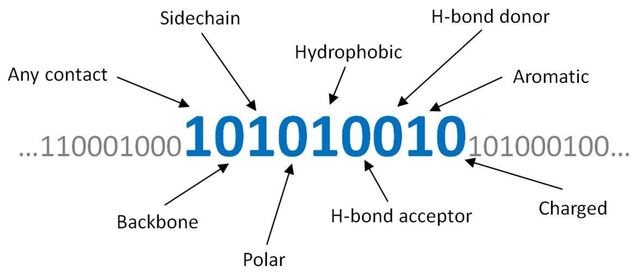


**Figure 1.** Fragment of SIFt describing bit positions for individual ligand-residue interactions.

**Table 1.** PDB IDs of crystal structures used in homology modeling and docking.

| Target | 5-HT1B | 5-HT2B | A2A | Beta1 | Beta2 | CXCR4 | D3 | H1 | M2 | M3 |
|---|---|---|---|---|---|---|---|---|---|---|
| PDB ID | 4IAR | 4IB4 | 3QAK | 2Y00*<br>2Y02<br>2VT4 | 3P0G*<br>3KJ6<br>2RH1 | 3OE0 | 3PBL | 3RZE | 3UON | 4DAJ |

*structures used in homology modeling

**Table 2.** Number of compounds used in docking to homology models and crystal target structures.

| Target | Active | Active (centroids) | Inactive | Inactive (centroids) | Random | Decoy |
|---|---|---|---|---|---|---|
| Beta1AR | 342 | 104 | 715 | 204 | 2557 | 2526 |
| Beta2AR | 550 | 103 | 601 | 173 | 2557 | 2619 |

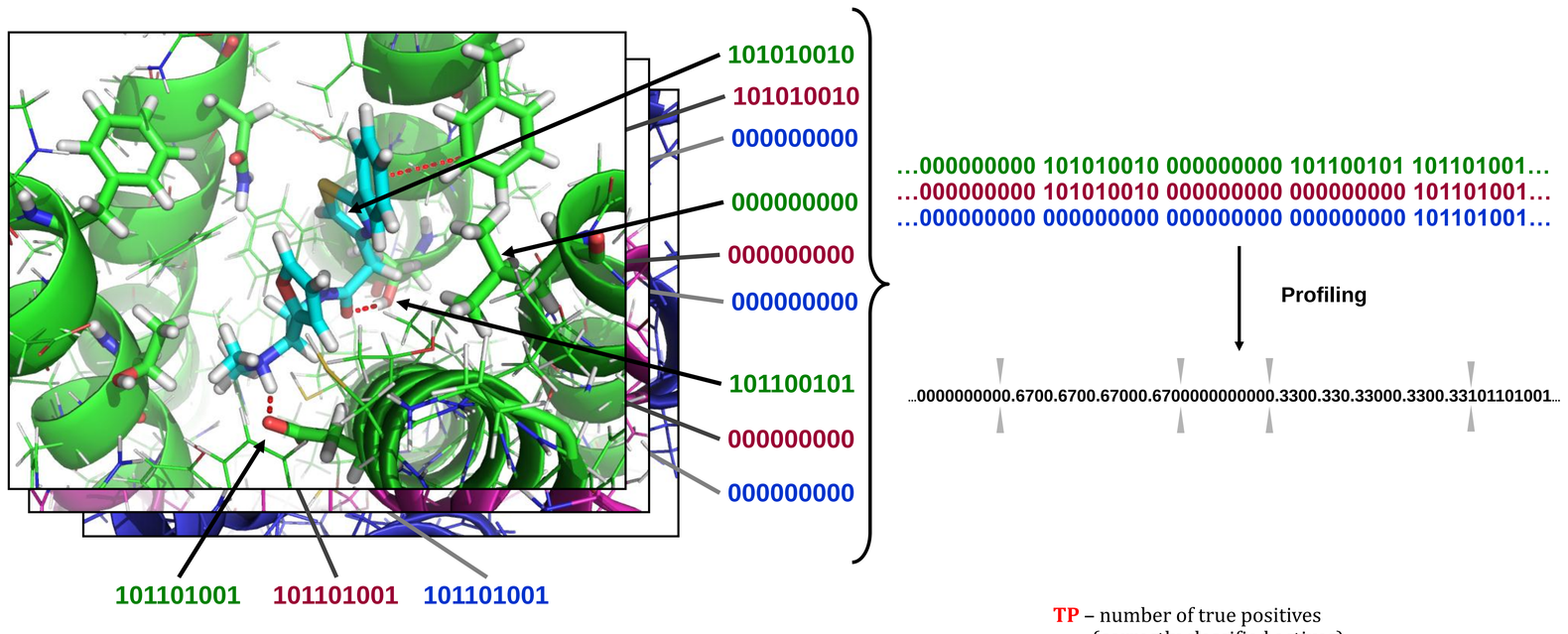


**Figure 2.** Scheme of SIFt profile construction.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

**Figure 3.** Measures of machine learning performance.

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

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

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

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

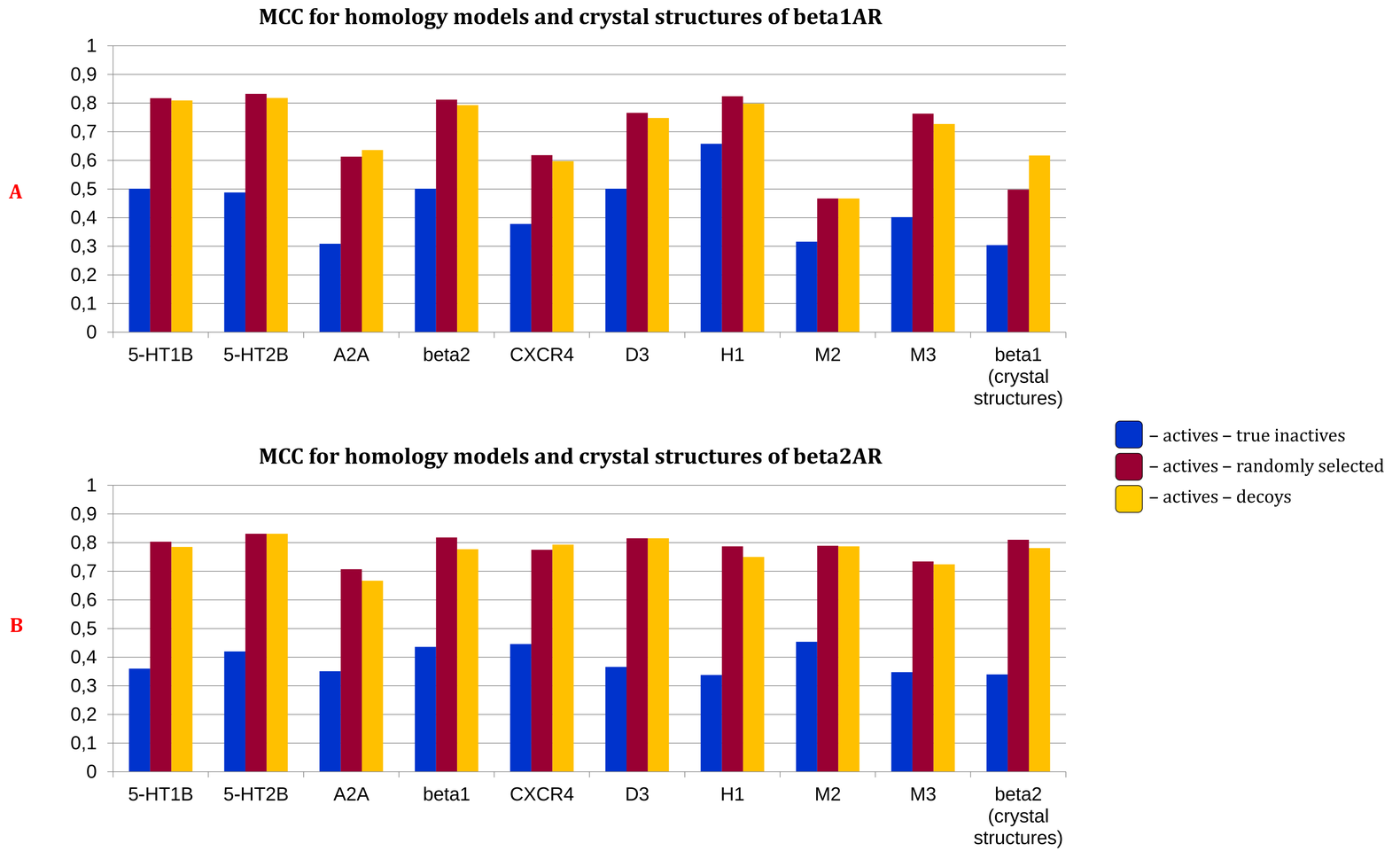


**Figure 4.** Evalutation of machine learning performance in predicting compounds activity towards homology models and crystal structures of beta1 (A) and beta2 (B) adrenergic receptors.

### Literature

1) Witek, J.; Smusz, S.; Rataj, K.; Mordalski, S.; Bojarski, A.J.: *An application of machine learning methods to structural interaction fingerprints – a case study of kinase inhibitors*, Bioorg Med Chem Lett, 2014, 24 (2), 580–585.
2) Huang, N.; Shoichet, B.K.; Irwin, J.J.: *Benchmarking sets for molecular docking*, J. Med. Chem., 2006, 49(23), 6789-6801.
3) 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.
4) Mordalski, S.; Kosciolek, T.; Kristiansen, K.; Sylte, I.; Bojarski, A. J.: *Protein binding site analysis by means of Structural Interaction Fingerprint patterns,* Bioorg Med Chem Lett, 2011, 21, 6816-6819.

### Acknowledgments

The study was partially supported by the Polish-Norwegian Research Programme operated by the National Centre for Research and Development under the Norwegian Financial Mechanism 2009-2014 in the frame of Project PLATFORMex (Pol-Nor/198887/73/2013).