Structural interaction profiles combination as a method for optimization of its application in docking results analysis – beta-2 adrenergic receptor case study

Sabina Smusz,[1,2] Jagna Witek[1] , Krzysztof Rataj[1], Stefan Mordalski[1], Andrzej J. Bojarski[1]

[1]Department of Medicinal Chemistry, Institute of Pharmacology Polish Academy of Sciences, Smętna 12 Street, 31-343 Kraków
[2]Faculty of Chemistry, Jagiellonian University, Ingardena 3 Street, 30-060 Kraków

e-mail: smusz@if-pan.krakow.pl

Introduction

Docking belongs to one of the most popular tools used in computer-aided drug design protocols. However, inferring about compounds activity on the basis of its results is still a very challenging task. There are some approaches that enable automation of the procedure of ligand-protein complexes analysis, among which there is a combination of Structural Interaction Fingerprints with machine learning algorithms.[1] Nevertheless, there still remains the problem of selection of proper set of models for docking studies that can be considered from both quantitative (the number of receptors that should be taken into account) and qualitative (the way the best models are selected) point of view. In the study, the number of receptors considered in post-docking analysis was optimized and the confrontation between the application of homology models and crystal structures was carried out.

Experimental

Beta-2 adrenergic receptors[2] were chosen as a case study due to presence of crystal structures and relatively high number of ligands. Homology models of beta-2 were constructed on nine templates (20 models per template were generated). Compounds with reported activity towards beta-2 were picked from the ChEMBL[3] database. Moreover, two sets of assumed inactives were generated: by random selection from ZINC and with the use of DUD approach. All the compounds were docked into the binding site of the constructed homology models and to 10 crystal structures of beta-2.

Ligand-receptor complexes obtained in the docking procedure were represented by the Structural Interaction Fingerprint.[4] Then, for each docked compound, the SIFt profile was calculated in a way that on each position in the string, the values were averaged over all receptors/crystal structures that were considered in the given situation. The number of receptors that were taken into account was varying from 3 to 20 (10 for the crystal structures).

Such representation constituted an input for machine learning experiments. They were performed with the use of the WEKA package with the Support Vector Machines[5] applied as a classification method in 10-fold cross-validation experiments.

The machine learning method effectiveness was evaluated by Matthews Correlation Coefficient (MCC).

Results

In general, increase in the number of receptors that were included in profiles calculation led to the improvement of the classification effectiveness. In the majority of cases, the optimal number of receptors was found to be 20. The number of receptors for which the MCC values were the lowest was equal to 3 in the majority of cases (A).

Changes occurring in actives/true inactives classification experiments were characterized by the highest rate of randomness – for each template, there were several number of receptors included in the profile that were not beneficial in comparison to the experiments with one less number of receptors (D). However, when actives were distinguished from DUDs (E) and ZINC (F) compounds, it was clear that the higher number of receptors included in the profile, the higher MCC values.

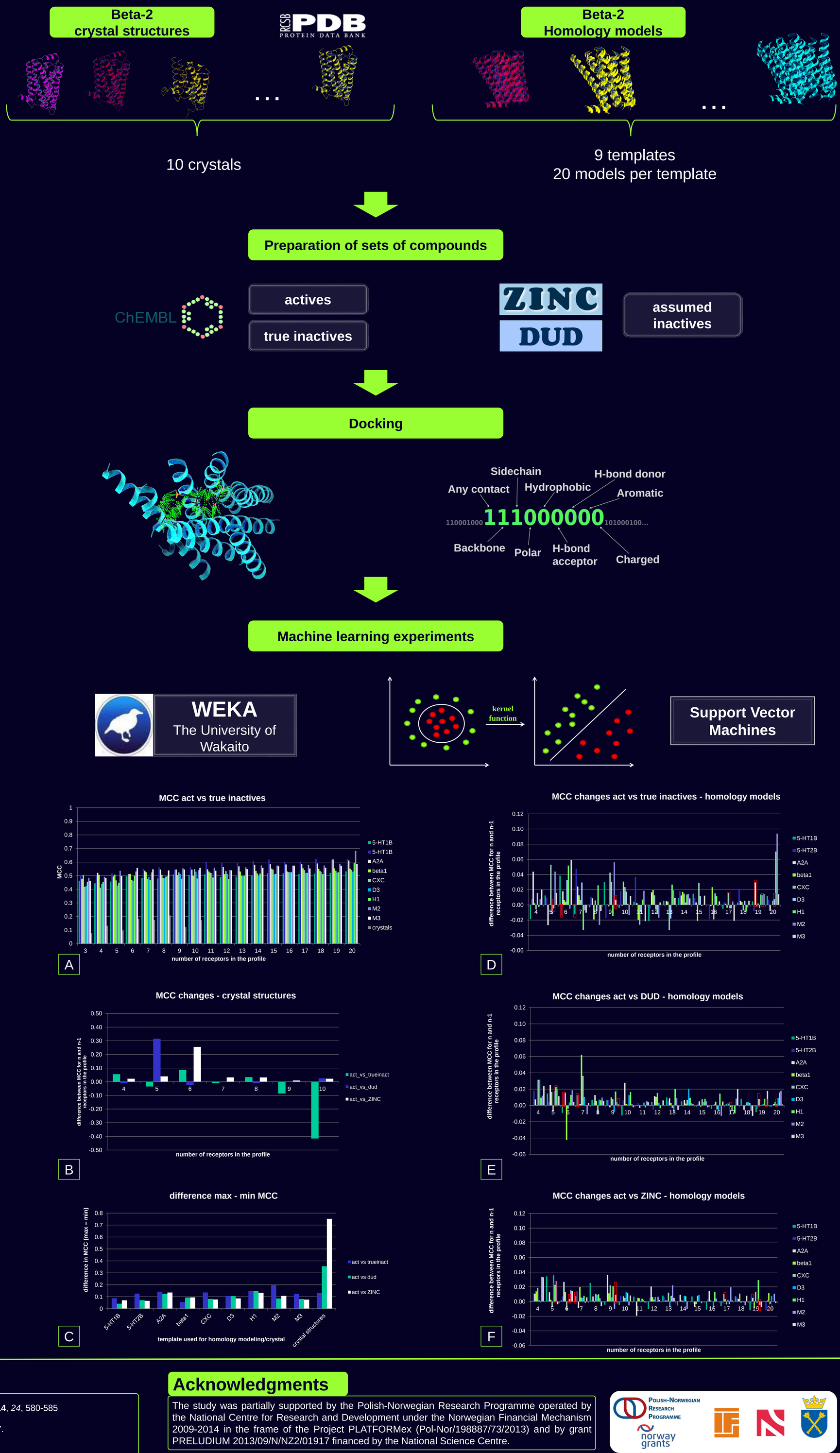
Changes in MCC for crystal structures were more significant than it was in the case of homology models (B). There were some experiments in which inclusion of additional crystal to the profile caused significant changes in MCC values – up to 0.3 improvement in act/DUD and act/ZINC experiments and up to 0.4 decrease in case of act/true inactives discrimination.

The analysis of the results show that the number of receptors included in the profiles section has significant influence on the obtained results. In case of homology models it was around 10-15% improvement in MCC when the optimal number of receptors considered was compared with the worst SIFt profile composition. The highest impact however occurred for experiments with crystal structures – in case of discrimination between actives from true inactives it was around 15% change in MCC, over 35% for actives/DUD and 75% for actives/ZINC experiments (C).

The best performing individual models (in terms of AUROC) are indicated with red frames. The lack of significant change in MCC values after adding the best model to the profile indicates that the model quality assessed in the previous step has rather no influence on the obtained results. The MCC values were more dependent on the number of models in the profile than on their quality.

Conclusions

It was proved that increasing the number of models improved the results in all types of experiments (actives/true inactives, actives/DUDs, actives/ZINC cmds discrimination). Although further addition of receptors would probably also cause further increase in MCC values, due to computational expenses arising from the necessity of performing the docking procedure, the set of considered receptors was not further extended. It was also proved that for beta-2 receptor, homology models are much more effective in terms of identification of active molecules than crystal structures and also that the number of receptors had more influence on the obtained results than their quality evaluated by AUROC. Therefore, despite the presence of crystallographic data for some proteins, for virtual screening purposes it is advisable to use also a set of homology models of a given protein to provide the diversity of receptor conformations.



References

1. Witek, J.; Smusz, S.; Rataj, K.; Mordalski, S.; Bojarski, A.J. Bio*org. Med. Chem. Lett.* **2014**, *24*, 580-585
2. Liggett, S. B. *J. Allergy Clin. Immunol.* **1999**, *104*, S42-6.
3. Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers et al. *Nucl. Acids Res.* **2011**, *40*, D1100-7.
4. Deng, Z.; Chuaqui, C.; Singh, J. *J. Med. Chem.* **2004**, *47*, 337-344.
5. Cortes, C.; Vapnik, V. *Mach. Learn.* **1995**, *20*, 273-297.

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 the Project PLATFORMex (Pol-Nor/198887/73/2013) and by grant PRELUDIUM 2013/09/N/NZ2/01917 financed by the National Science Centre.