Discovery of Potential Thioredoxin Reductase I Inhibitors by QSAR-Based Virtual Screening

Meryck F. B. Silva (IC), Marília N. Nascimento (PG), Marcelo N. Gomes (PG), Cleber C. Melo-Filho (PG), Rodolpho C. Braga (PQ), Bruno J. Neves (PG),* Carolina H. Andrade (PQ)**

***bruno.labmol@gmail.com ****carolina@ufg.br**

LabMol - Laboratory for Molecular Modeling and Drud Design, Faculdade de Farmácia, Universidade Federal de Goiás, Goiânia, Brazil

Keywords: antioxidant system, machine learning, fingerprints, virtual screening.

Abstract

We report the discovery of 30 potential thioredoxin reductase I (TrxR) inhibitors by QSAR-based virtual screening (VS).

Introduction

The TrxR plays a crucial role in the redox regulation of numerous cellular signaling pathways, involved in cell survival and proliferation. In recent years, accumulating evidence supports that TrxR is a promising target for development of novel anticancer agents as the thioredoxin system is often overexpressed in many tumors.[1] In this study, we developed predictive binary QSAR models for TrxR inhibition which is fully compliant to Organization for Economic Cooperation (OECD) guidance on development and validation of models. Then, we performed a QSAR-based VS and prioritized some hits for experimental validation.

Results and Discussion

We retrieved a large dataset containing 392,548 compounds with inhibition data for TrxR of *Rattus norvegicus* (PubChem Bioassay ID: 588453). The dataset was prepared according to following steps: inconclusive IC_{50} data were removed; a threshold value was defined to discriminate between inhibitors (≤10 μM) and non-inhibitors (>10 μM); chemical structures were standardized and duplicates identified. Importantly, we found an overall concordance of 99.9% between duplicates, considering multiple samples. Because the prepared dataset was highly unbalanced (1,327 inhibitors and 367,069 non-inhibitors), it is not recommended building binary QSAR models for the entire dataset. Thus, we decided to balance the dataset with ratios of 1:1, 1:2, and 1:3 (inhibitor:non-inhibitor). Unlike the traditional under-sampling methods which randomly balance the dataset, we developed a linear under-sampling strategy that retains most of the representative structures of the non-inhibitors set, thus ensuring as high as possible coverage of original chemical space. Then, QSAR models were developed combining three machine learning classifiers (Random Forest, SVM, and GBM) and five molecular fingerprints (MACCS keys, Avalon, FeatMorgan, AtomPair, Morgan). The robustness and the applicability domain (AD) of these models were estimated using the *5-fold* external cross-validation method and the Euclidean distances of *k*-NN, respectively. In general, all combinations generated using balanced dataset led to predictive models, with correct classification rate (CCR) values ranging between 0.81 – 0.88 and coverage of 0.61 – 0.67. The best individual model was built using the combination of Avalon/GBM (CCR = 0.88, SE = 0.87, and SP = 0.88). To assure that the accuracy of the models was not due to chance correlation, 10 rounds of Y-randomization were performed for each model, led to CCR values around 0.50, indicating that models built are statistically robust. The most predictive model was mechanistically interpreted using predicted probability maps (PPMs). Developed models were then used in the most conservative way (i.e., in consensus fashion and with the strictest AD criteria) for VS of our in house database. Next, we removed pan-assay interference compounds (PAINS), and compounds that were outside of Lipinski and Veber rules. As a result, 30 predicted hits were selected for experimental validation in enzymatic TrxR inhibition assay and against a panel of tumour cell lines. In addition, predicted hits showed new chemical scaffolds, structurally dissimilar to known inhibitors of TrxR. All steps of this study were performed in our in-house KSAR workflow (http://labmol.farmacia.ufg.br/ksar), a tool that integrate R, ChemAxon, Python and KNIME.

Conclusions

The developed QSAR models are robust and predictive, and fully compliant to OECD's guidance. The models were useful to rank 30 new VS hits with different chemical scaffolds that could be potentially TrxR inhibitors. The next step will be the experimental validation of compounds using TrxR inhibition assays and cytotoxicity against a panel of tumour cell lines.

Acknowledgements

CAPES, FAPEG, CNPq, and ChemAxon

[1] Mahmood, D. F. D.; Abderrazak, A.; El Hadri, K.; Simmet, T.; Rouis, M. Antioxid Redox Signal **2013**, *19*, 1266.