FDA drug repurposing – A machine learning prediction model as virtual screening tool for O-GlcNAc transferase (OGT) potential inhibitors discovery

Azevedo, P.H.R.deA.[1*]; Alves, T.F.[1]; Peçanha, B. R.de B. [1]; Flores-Junior, L.A.P.[1]; Lima, C.H.S.[2]; Dias, L.R.S.[1]; Muri, E.M.F.[1**]

1 Laboratório de Química Medicinal, Faculdade de Farmácia, Universidade Federal Fluminense, Niterói, RJ, Brazil
2 Universidade Federal do Rio de Janeiro, Instituto de Química, Rio de Janeiro, RJ, Brazil
**pedrohraa@id.uff.br **estelamuri@id.uff.br*

Introduction

O-linked N-acetylglucosamine (O-GlcNAc) is a ubiquitous, single N-acetyl-glucosamine sugar that cycles on and off serine or threonine residues in nuclear, cytoplasmic, and mitochondrial proteins. O-GlcNAc is added to proteins by a single enzyme, O-GlcNAc transferase (OGT), while the enzyme O-GlcNAcase (OGA) removes the modification. The synthesis of UDP-GlcNAc, the substrate for OGT, occurs through the hexosamine biosynthetic pathway. Furthermore, OGT activity is sensitive to a wide range of UDP-GlcNAc concentrations, easily changing the O-GlcNAcylation levels, with downstream effects on the modified proteins' function, location, or stability[1]. Due to the O-GlcNAcylation importance in metabolic diseases as cancer and diabetes, the number of studies focused on unveiling OGT catalysis increased[2]. Although studies allowed the discovery of enzymatic inhibitors, most OGT inhibitors until now are limited mainly because of their lack of specificity, cytotoxicity, or solubility[3]. Considering the development of a new drug is a costly and slow process, repurposing of available drug can be an attractive tool to treat both common and rare diseases. Also, drug repurposing possesses lower costs and shorter development timelines once it can use lots of experimental data about the available drugs[4]. In this context, this work focuses on drug repurposing to OGT target through the molecular modeling approach as molecular docking and machine learning prediction to a virtual screening of FDA-approved drugs.

Method

The molecular docking study was realized with the X-ray crystal structure of hOGT (PDB:5NPS, Reso1ution = 0.68 Å) and performed by GOLD 5.2.2 with Goldscore function through the previous protocol described in the literature[5]. The OGT inhibitors with IC_{50} equal to or lower of 100µM were obtained on the ChemBL databank. Approximately 20 to 30 decoy (inactive) compounds were built from each inhibitor using DecoyFinder 2.0 from the ZINC databank. The parameter used to create the decoy dataset was Tanimoto threshold of 0.5 and 0.9 to active x decoy and decoy x decoy, respectively.

The FDA-approved drugs were obtained from the ZINC databank and filtered by Lipinski's Rule of 5. All 3D compounds of OGT inhibitor, decoys, and drugs structures had their geometry minimized at ground state with molecular mechanics by Merck Molecular Force Field using Spartan v10. The physicochemical properties were calculated by DataWarrior version 5.5.0 concatenated with docking results on Excel 2019. The machine learning model was developed from different algorithms like Logistic Regression (LR), Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Best-First Tree WEKA (BFT-WEKA), and Extreme Gradient Boosting Tree (XGBOOST) available in KNIME 4.3.4. The sample treatments were normalized with a Z-score, followed by correlation filter application with a threshold of 0.5, and partitioned into a training set (70%) and test set (30%) for validation. The imbalanced dataset was correct by upsampling using the Synthetic Minority Over-Sampling Technique (SMOTE). Statistical parameters were used to evaluate the models as: recall, precision, sensitivity, specificity, F-measure, overall accuracy and error and Cohen's kappa.

Results and Discussion

The ChemBL search from OGT inhibitors returned OGT inhibitors 28 compounds, with IC_{50} value below 100µM. For the 28 compounds selected, 830 decoys were generated assembling the machine learning base sample. For the screening FDA-approved drug bank assembling containing 2115 drugs, duplicates names were excluded returning 1430 drugs that using the Lipinski's Rule resulted a

total of 804 drugs, which just 95 of them were able to perform the docking against OGT enzyme. The OGT inhibitors, decoys, and the FDA-approved compounds were submitted to molecular docking protocol previously validated. The docking results to OGT inhibitors and decoys were concatenated with 13 of their physicochemical descriptors to create a machine learning model.

As the number of active compounds in the dataset bank was too small compared with the number of decoys, the imbalanced bank was upsampling using SMOTE allowing the generation of synthetic samples for the active minority category. This step is essential for improving the prediction parameters as sensibility, specificity, and accuracy. The statical parameters to different prediction algorithms indicated that NB and SVM possess the worse values to sensibility, specificity, and accuracy whereas, LR, DT, RF, BFT-WEKA, and XGBOOST showed statistical values over 90%. The FDA-approved drug dataset was evaluated by these 5 best prediction algorithms and for the prediction reliability evaluation, the train dataset application domain was calculated (APD = 4.063) and confronted against the FDA drug dataset where 46,3% (44/95) of the drugs were reliable into the APD train data. The 5 algorithms returned 32 different drugs predicted as active against OGT, of which 13 were within the application domain and 8 of them were also predicted as active by 2 or more algorithms. The celecoxib, a nonsteroidal anti-inflammatory drug, presented the prediction as OGT active for 4 of 5 algorithms, and its complex with OGT enzyme was analyzed for a better understanding of the interactions between both. The molecular docking study of celecoxib showed H-bond interactions with the C-Catalytic domain (Lys842, Lys898, Cys917, Thr922 Asp925), which form the catalytic binding site.

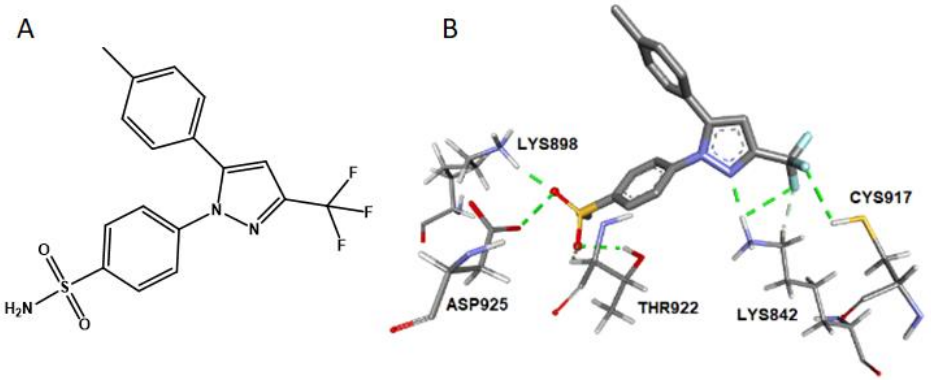

Figure 1 – A) Celecoxib structure. B) Celecoxib-OGT complex and their H-bonds interactions.

Conclusion

In conclusion, we developed a machine learning classification model using docking score and ligand descriptors to predict the activity of potential hOGT inhibitors. The consensus machine learning model showed excellent discrimination between active and inactive compounds with sensibility, specificity, and accuracy over 90%, respectively. We have carried out the FDA-approved drug repurposing by virtual screening of 95 drugs using a machine learning predict model identifying 8 drugs with OGT potential inhibitor. The celecoxib was identified as a potential inhibitor by 4 of 5 machine learning predict models. The intermolecular analysis reveals H-bond interaction with C-Catalytic domain residues. The identification of celecoxib as a potential OGT inhibitor allows us to consider its structure for future design of others OGT inhibitors and further enzymatic inhibition assays are essential to evaluate its real potential hOGT inhibition corroborating the OGT inhibitor machine learning predict model developed on this work.

Acknowledgments

We gratefully acknowledge the financial support from the National Council for Scientific and Technological Development (CNPq), Carlos Chagas Filho Foundation for Support of Research of the State of Rio de Janeiro (Edital FAPERJ 22/2016 –Emergency Support for Stricto Sensu Graduate Programs and Courses in the State of Rio de Janeiro Project E -26/200.930/2017) and Cancer Foundation.

Bibliographic References

[1] Hart, G. W.: Nutrient regulation of signaling and transcription, J. Biol. Chem., 2019, 294, (7), pp. 2211–2231.

[2] Vasconcelos-dos-Santos, A. et al.: Hyperglycemia and aberrant O-GlcNAcylation: contributions to tumor progression, J. Bioenerg. Biomembr., 2018, 50, (3), pp. 175–187.

[3] Makwana, V. et al.: Essential role of O-GlcNAcylation in stabilization of oncogenic factors, Biochim. Biophy.s Acta.

Gen. Subj., 2019, 1863, (8), pp. 1302–1317.
[4] Pushpakom, S. et al.: Drug repurposing: progress, challenges and recommendations, Nat. Rev. Drug Discov., 2019, 18, (1), pp. 41–58.
[5] Albuquerque, S. O. et al.: Biological evaluation and molecular modeling of peptidomimetic compounds as inhibitors for O-GlcNAc transferase (OGT), Eur. J. Pharm. Sci., 2020, 154, pp. 105510.