Targeted In Silico Screening of Small Molecule Databases against Breast Cancer

Lalehan Oktay [a,b], Umut Yücel [b], Alpsu Olkan [b] and Serdar Durdağı [b]*

[a]Bioengineering Program, Graduate School of Natural and Applied Sciences, Bahçeşehir University, Istanbul, Turkey

[b]Computational Biology and Molecular Simulations Laboratory, Department of Biophysics, School of Medicine, Bahçeşehir University, Istanbul, Turkey

Around 75% of all breast cancers are either estrogen receptor positive and/or progesterone receptor-positive, and in these cases, estrogen is the major stimulant responsible for the progression of these types of tumors. [1] Therapy approaches for estrogen receptor and progesterone receptor-positive tumors include usage of tamoxifen - an anti-estrogen - that stops estrogen from binding to its receptors. [1] Tamoxifen functions mainly as an antagonist, but multiple data sources imply that tamoxifen also has agonist activity and that, this is the reason for its loss of effectiveness against breast cancer and, therefore, the relapse of the disease. [1] An alternative strategy to achieve estrogen deprivation is to inhibit the aromatization of androgens into estrone or estradiol through the blockage of aromatase. [2] Currently, there are two generations of aromatase inhibitors (AIs). The goal during clinical development of aromatase inhibitors has been to achieve complete estrogen suppression at doses that have no significant toxicity and no nonspecific hormonal effects. [1] Keeping this in mind, a literature check was made, and 28 molecules with known IC_{50} data for MCF-7 cell lines were retrieved from the ChEMBL chemical database and utilized in the PHASE 3D-QSAR application. From this application, a common four-sited pharmacophore hypothesis, AHHH.565, was generated and further expanded into a 3D-QSAR model. This implies that for favorable binding to a pseudo receptor, a ligand should possess a hydrogen-bond acceptor site along with 3 hydrophobic sites. The model was validated using a set of gallic acid derivatives. Using this model, screening of chemical databases (OTAVA, NCI, Peptidomimetic, SPECS, Chembridge, and Enamine) and a total of over 1.28 million molecules was conducted. Top 1000 molecules from each database were chosen according to their fitness to the model, AHHH.565, and were used in the GLIDE/SP (standard precision) mode. Top 100 resulting poses were chosen for extra precision (XP) docking using GLIDE. 10 resulting poses were chosen for Induced Fit Docking (IFD). GOLD docking program was also used for identified hits to further validate the findings. Surviving molecules were tested against 26 toxicity QSAR models and against a universal cancer therapeutic activity QSAR model through the Metacore/Metadrug of Clarivate Analytics. Surviving molecules were put through molecular dynamic (MD) simulations with the aromatase enzyme, and results were compared with known clinically used AIs to validate our screening methodology for the discovery of novel, non-toxic AIs.

[1] S. R. Johnston, M. Dowsett, *NRC*, **(2003)**, *3*, 821 – 831.

[2] A. Carmen, M. J. Carlos, *Medicinal chemistry of anticancer drugs,* **(2008)**, 54 – 90.