Design of Selective Compounds for the Estrogen Receptors: A Molecular Docking and Machine Learning Study

Chiappori F, Ferrario MG, Gaiji N, Fantucci P

Department of Biotecnology and Bioscience, Università Milano - Bicocca,
Milano DELOS s.r.l, Bresso (MI)

Motivation

Estrogens are steroidal hormones that regulate reproductive functions and the bone mass maintenance in mammals. The molecular action of the estrogens is mediated by the nuclear receptors hERs, isoforms ERa and ERb, which show high structural similarity, but low sequence identity. In women, after menopause the production of estrogens is strongly reduced, it is partly recovered by the Hormone Replacement Therapy (HRT) that prevents some of the menopause symptoms, but enhances the risk of tumors of the reproductive tissues. The most promising model compounds for HRT are SERMs (Selective Estrogen Receptor Modulators) and phytoestrogens. SERMs act in a tissue-specific manner, but they are not sufficiently strong agonists in target tissues, while in other tissues they are antagonists; phytoestrogens are not tissue-selective, but they act mainly on the ERb, which seems to be responsible for the non-reproductive estrogen functions. Our research aims at identifying new compounds for HRT, using a Virtual High-Throughput Screening (VHTS) approach, based on generation of a large number of drug-like compounds and analysis of their interaction with the Ligand Binding Site (LBS) of the ERa and ERb. The 3D structures of the two proteins have been obtained as described in another contribution to BITS (1). All the results reported have been obtained using the DELOS suite of programs (2) which has been recently developed by us. The LBSs found by DELOS correspond very closely to the binding cavity of the ERa and ERb, experimentally known (3).

Results

Two compounds libraries have been generated using as scaffolds the skeletons of estradiol and genestein, respectively, on which three substitution points have been defined for ten different substituents. The two libraries, named EST and GEN, respectively, include 1000 compounds each. The DELOS program has been used for optimization of the 3D structures of the ligands using a quantum chemical approach and for the evaluation of 346 molecular descriptors for each compound. Another section of DELOS allows the construction of the potential maps for the ligand-LBS interaction, to be used in rigid docking simulations. The rigid docking module of DELOS is based on a fast Genetic Algorithm (GA) coupled with an extrapolation scheme (GA-E) for the global optimization of the best roto-translation of the ligand within the LBS. On the basis of the differential docking energy to ERa and ERb (minimum with ERa and maximum with ERb) 65 and 57 molecules have been selected and considered as "good molecules" from EST and GEN libraries, respectively. All other elements of the libraries have been considered as "bad molecules". This coarse classification is carried out in order to attempt a two-way classification of the library elements, using the molecular descriptors as features and the docking behaviour as class variable. We submitted to the classifier the table of the all 346 descriptors, after that, we tried different feature selection methods: · the descriptors with non-zero variance (264 for EST and 290 for GEN) · the descriptors with correlation index above 0.4 for the docking energy to ERα (68 for EST and 111 for GEN) and to ERβ (84 for EST and 94 for GEN), and for the difference of docking energy from ERα to ERβ (12 for EST and 40 for GEN). The results of the classifier trained by the EST library are good in the case of the descriptors with non-zero variance (74% true positives, 77% true negatives) and the descriptors correlated with the docking energy for ERβ (80% true positives, 61% true negatives). In the case of GEN library, the best results are those provided by the classifier trained by the descriptors with non-zero variance (90% true positives, 65% true negative) and the descriptors correlated to the docking energy for ERβ (88% true positives, 52% true negatives). The final selection performed on GEN and EST libraries was carried out using a combined criterion

which incorporate both docking energy and a series of ADME properties. The final subset of molecules includes 12 and 8 molecules for EST and GEN library, respectively. Concerning the research of molecules for the HRT, we have individuated some characteristics that may play a major role for finding new ERβ-selective compounds, i.e. the absence of bulk side chains, the presence of bulk substituents with polar groups at the ends. The present investigation shows how our VHTS procedure is efficient and relatively fast. The bioinformatics platform DELOS, which allows to follow all the important steps along the rational drug design process, proved to be a reliable and very promising tool for massive, extended studies in this field.

Contact email: noura.gaiji@unimib.it

References

1. Molecular dynamic study of the ligand binding domain of estrogen receptor alpha and beta. Chiappori F., Fantucci P., Ferrario MG, Gaiji N BITS2005
2. DELOS is an integrated platform H/S developed by the academic spin-off DELOS.srl
3. Brzozowski et all Nature vol 389, 753-758 (1997)