

Development and validation of a virtual high-throughput screening protocol for estrogen receptor antagonists (Drug discovery acceleration & Lead Optimization)

Andrew S Knox,¹ Mary J Meegan,¹ Dermot Frost,³ and James C Sexton²

¹Department of Pharmaceutical Chemistry, Trinity College, Dublin, ²Department of Mathematics, Trinity College Dublin, ³Trinity Centre for High Performance Computing

Introduction

R&D production of new drugs has remained constant over the last number of years with major pharmaceuticals each launching roughly one new drug per year(1)

In the pharmaceutical industry a continuous demand exists for novel molecules with specific physical or biological properties. Traditionally these molecules are found either by accidental observation of an interesting characteristic or by testing many molecules, from natural sources or manmade, for the desired properties. The dawn of the use of computational methods for drug discovery acceleration and lead optimization has provided a complementary strategy in finding such molecules.

In this respect, we have validated and developed a virtual High Throughput Screening protocol involving a suite of conformational and docking computational algorithms optimized to carry out screening of candidate compounds for estrogen receptor antagonist activity using X-ray crystal models of the Estrogen Receptor alpha (ER α). The ER α was chosen as our target protein for the study, as it is a model with therapeutic importance (ER α is expressed in 50% of all breast cancers (2)) for which there is available a significant amount of crystallographic data with distinct differences.

The objective of this was to assess the ability of our computational drug design suite in identifying potential modulators that would exhibit selectivity towards the estrogen receptor alpha (ER α). The enrichment and hit rate achieved by our protocol outperforms two commercially available programs FlexX (Tripos Inc.) and FRED (Openeye software).

Problems?

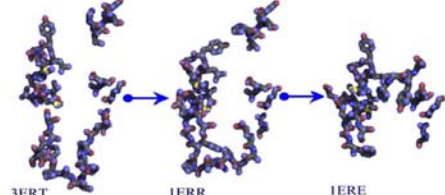
Typical problems associated with Virtual Screening and the identification of the bioactive conformation of a ligand are outlined below:

1. Accounting for target & ligand flexibility – When a ligand binds to a receptor, the receptor and ligand undergo a conformational change.
2. Obtaining a rapid execution time for the screening protocol.

Our approach....

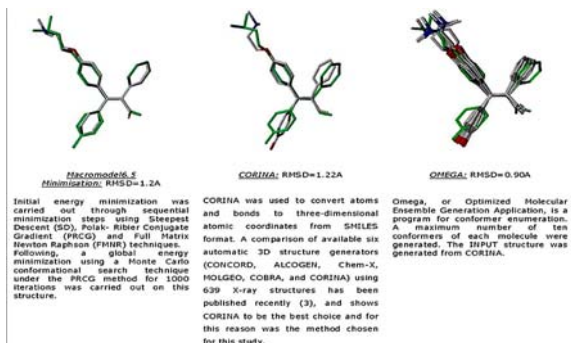
1. Our protocol involved screening the dataset through three available crystal structures with inherently different active site conformations (3ERT, 1UOM, 1ERE). Figure 1 shows that a significant range of conformational differences within a 5 Angstrom radius of the active site, are provided by these crystal structures of the ER α . This allows a broad scope of 'hits' to be identified in our screening process, thus taking into account target flexibility.

Fig 1: Conformational differences within 5 Angstrom radius of active site using POVRAY



There are several methods for taking ligand flexibility into account. The RMSD produced by each method compared with the bioactive conformation of Hydroxytamoxifen by superpositioning is outlined below. 2D-3D structure generation using CORINA (Molecular Networks GmbH), followed by conformer ensemble generation using OMEGA (Openeye software) outperforms the other methods with the lowest RMSD of 0.9A.

Fig 2: Comparison of conformer generation of Hydroxytamoxifen using several techniques

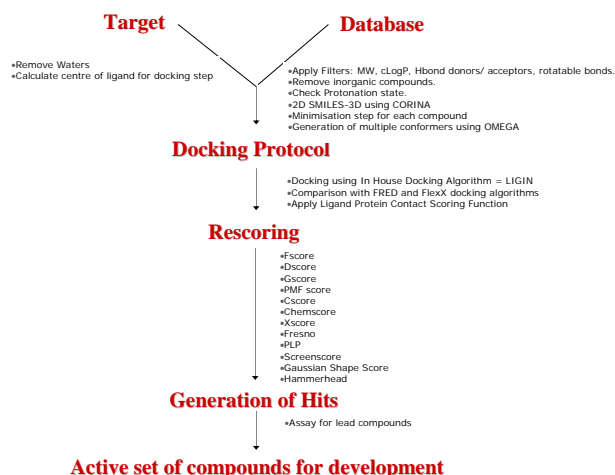


2. The code was written in C and is trivially parallelisable and so could be run on as many CPUs as available. At present we have a new cluster with 160 processors available. The execution time to run the protocol outlined in the validation study would be 30 minutes showing the benefits of this approach.

References:

1. Smith, A. Screening for Drug Discovery-The leading question. *Nature* **2002**, *418*, 353-463.
2. J. A. Scott and W. L. McGuire, in *Endocrine-Dependent Tumors*, K. D. Voight and C. Knabbe, Eds. (Raven, New York, **1991**), 179-196.
3. Sobolev V., Wade R.C., Vriend G. and Edelman M. *Proteins* **1996**, *25*, 120-129.

Summary of Virtual Screening Protocol



Validation Study

A haystack built from 1000 'drug-like' compounds from Maybridge HTS collection (Maybridge plc HITS Kits) was seeded a set of known actives (35 Estrogen Receptor Modulators taken from literature.

Physicochemical properties such as cLogP, numbers of H-bond donors/acceptors, numbers of rotatable bonds calculated and 2D filters were applied to remove inorganics, and compounds that were not drug-like. The filtered dataset was converted from 2-D to 3-D using CORINA and conformational ensembles generated subsequently using OMEGA, to allow sequential docking of each ligand in the dataset, using our 'in house' rigid docking algorithm (LIGIN(3)).

3ERT (Hydroxytamoxifen bound), 1ERR (Raloxifene bound), 1UOM (Isoquinoline bound) pdb entries were downloaded from the crystallographic database and modified to remove crystallographic waters and solvent remaining. A scoring function based on ligand protein contacts (LPC) calculates the shape complementarity of each compound with the ER α and allows rejection of inappropriately bonded ligands. The dataset is rescored using a number of appropriate scoring functions in order to assess and retrieve a set of known actives and also possible lead compounds. The performance measures of our screening protocol were assessed by enrichment value using two commercially available docking and scoring programs FRED (Openeye software), FlexX (Tripos Inc):

$$\text{Enrichment} = \text{Hit Rate in Subset} / \text{Hit Rate in Database}$$

Fig 3: Hits retrieved using several known scoring functions to re-score the database of ligand, and Enrichment rate for protocols.

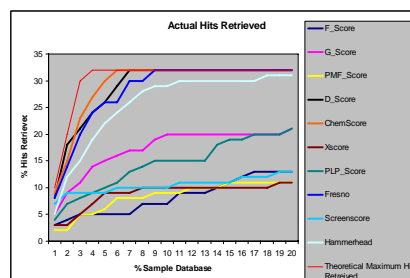


Figure 3 shows the hits retrieved using each scoring function to re-score the database of ligands. Applying D_score or ChemScore as our rescoring system provides us with the best-hit rate.

The enrichment rate is the increase in the proportion of hits found in any given sample of compounds, compared with the proportion expected from a random sample. It is clear from Figure 3 also that our protocol outperforms both FRED and FlexX.

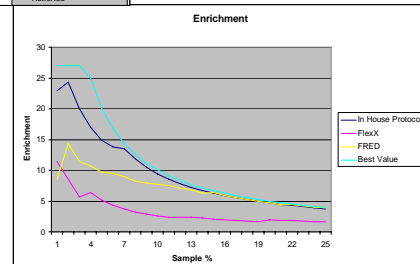


Fig 4: Potential modulators retrieved from top 1% of database.

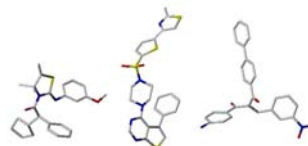


Figure 4 shows a selection of three compounds retrieved within the top 1% of the database among the seeded antagonists using our protocol. This demonstrates the ability of the system to screen for potential modulators also, as they all have the same structural motif needed to bind within the active site of the ER