Drug Design Teaching

Answers to Exercise 7 and to questions of Session3 (Ligand-based Virtual Screening)

Exercise 7. Enrichment of antiviral molecules in FDA-approved (and related) drugs

This exercise aims to show how virtual screening is an enrichment procedure. Virtual screening allows not to test an entire chemical collection, yet a reduced selection of molecules. Instead of random picking, virtual screening helps choosing the most interesting ones to be assayed experimentally. In the given example at the beginning of Session 3, we rely on similarity of molecules in 2D, with the hypothesis that compounds most similar chemically to Saquinavir have a higher probability to be also protease inhibitors and having thus an antiviral activity.
In any virtual screening procedure, you'll monitor whether your criteria and setup make sense. This can be quantified by counting how many "known actives" are top-ranked. In the context of this exercise the known actives are FDA-approved antiviral molecules by inhibition of the viral protease.

From the virtual screening of Saquinavir against DrugBank with FP2, calculate the enrichment factor (**EF**) at **top 24** (=0.2%)

- How many of these antiviral protease inhibitors can be found in the top 24 of your screening **(n_{screen})** ?

  Saquinavir, Telinavir, Atazanavir and Indinavir can be found among the 24 top-ranked molecules **($n_{screen} = 4$)**

- The rate at top 24: **$r_{screen} = 4 / 24 = 0.17$**

- The entire DrugBank database contains 10,575 molecules among which 18 molecules are antiviral protease inhibitors. This corresponds to a rate **$r_{db} = 18 / 10575 = 0.0017$**

- enrichment factor (*ER*) is the ratio of rates (**r_{screen} / r_{db}**)
  $EF_{0.2}$ = 0.17 / 0.0017 = **100**

This virtual screening produced an enrichment of 100-fold of anti-viral protease inhibitor drugs at 0.22% compared to random picking in the entire DrugBank collection. This is *huge* compared to 'real-life' virtual screening enrichment.

Answers to questions along Session 3 (EGFR inhibitors).

- With that results let's try to answer the following questions about compound CHEMBL461792:

  - *What is the similarity score and ranking of compound CHEMBL461792?*
    Score : 0.964
    Rank : #19

  - *What are the two structural differences between CHEMBL461792 and Erlotinib?*
    1. The cyano substituent of on the aromatic of erlotinib is replaced by a chloro substituent in compound CHEMBL461792.
    2. The ether "side chains" of erlotinib are linked together by an additional C-C bond in CHEMBL461792.

  - *Which of these chemical modifications makes CHEMBL461792 more rigid than Erlotinib?*
    The modification 2 rigidifies the long "side chains".

  - *Any clue about the potential benefit to test a more rigid ligand?*
    By limiting the flexibility of some parts of the ligand is one typical trick in drug design. If the geometry of the molecule is locked in its bioactive conformation (the geometry of the binding mode), the ligand doesn't need to freeze in the right geometry to complement the active side of the protein. This represents a gain of entropy, a component of the free energy of binding. The favorable impact on the free energy of binding is effective only if the rigid molecule has a geometry complementary with the binding site, not making clashes nor unfavorable interactions with the atoms of the protein.
    Another effect is that very flexible molecules are less prone to pass through biological barrier (such as gastrointestinal wall for absorption, or blood brain barrier to access CNS).