

THE EMERGENCE OF VIRTUAL SCREENING

High throughput screening of large collections of chemical compounds remains an important step in many drug discovery research projects. Jin Li and Timothy Lovell, Chemical Computing, AstraZeneca, review the complementary virtual screening methodologies.

In recent years, virtual screening – sometimes referred to as *in silico* high throughput screening (HTS) – has emerged as a complementary technology to HTS for lead discovery. It is also an attractive method for accessing new chemical compounds. Virtual screening methods have been developed in order to search large chemical databases in the computer and select a subset of candidate compounds for laboratory testing or synthesis. The subset of compounds can be selected by a number of virtual screening approaches, including, for example, 2D descriptors, 3D pharmacophores, shape matching, QSAR models and protein 3D structure directed docking.

One of the key concepts developed in assessing the effectiveness of virtual screening methods is the enrichment rate, illustrated in Figure 1. In essence, enrichment is a measure of the ability of a given virtual screening method to separate a small subset of active compounds for a given protein target from a large set of random compounds. If this ability to separate actives from inactives is over and above a random selection, the virtual screening method is said to have a certain enrichment rate. Most companies using virtual screening have invested considerable effort in establishing such a platform.

Docking and structure-based screening

Among all virtual screening methods, protein-structure-based docking has received considerable attention. This approach is a direct way to use the rapidly increasing number of protein 3D

structures that appear in the lead discovery process. Structure-based virtual screening typically involves fast docking of a (large) number of chemical compounds against a protein-binding or druggable site. The docked conformations are then scored, usually in the form of a relative rank order by a variety of scoring functions as a way to select a small subset of compounds for further analysis, purchase or testing. Compounds with good scores are supposedly indicative of potentially good binders. Virtual screening has been used in a range of discovery-research projects, all with varying degrees of success.

Structure-based virtual screening approaches are underpinned by the accuracy of the molecular-docking methods, which are, in turn, dependent on the computational algorithms for conformational sampling and scoring of different ligand-binding poses. There is a growing consensus that a small number of docking software packages provide users with reliable, and therefore, useful performance over a large range of receptor-ligand structural motifs. The speed and the accuracy, examined for a large number (approximately 200) of protein-ligand complexes, are of such quality and reliability that they can be used with a reassuring level of confidence (see Figure 2).

A key concept in assessing the effectiveness of virtual screening methods is the enrichment rate

Figure 1. Virtual screening is about enriching the subset of compounds with those likely to be active for a given biological target or to meet a certain calculated property profile

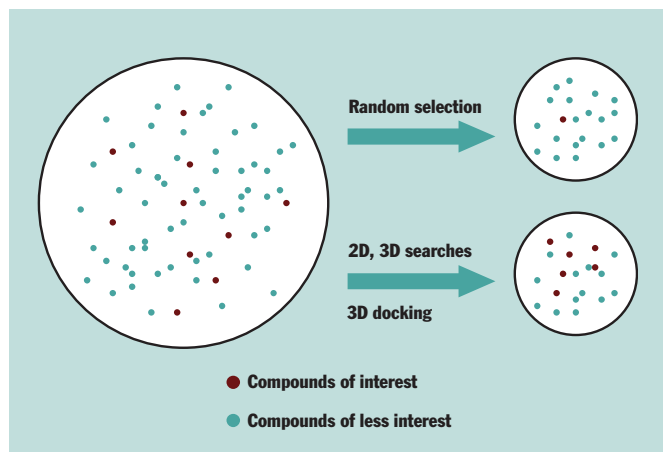
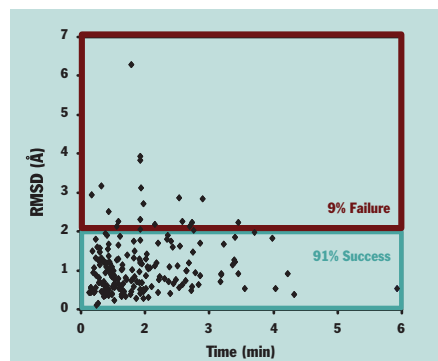


Figure 2. Docking accuracy, as assessed by the RMS difference of ligand conformation predicted by docking software (in this case, results from ICM) and that determined by x-ray crystallography, for 195 complexes. Typically, when the difference is less than 2 Å, the result is considered acceptable or successful.



From the viewpoint of applying virtual screening in a prospective manner, for instance, towards the discovery of new lead compounds, a key question is whether a given virtual screening method can generate new active leads, and what the likely benefits of such an approach in practice are. The answer to these questions can only be obtained by applying different docking methods in a wide range of

structure determination, targeted chemical synthesis and site-directed mutagenesis studies.

Generally, virtual screening has found more applications in the lead-identification or lead generation stage rather than lead optimisation stage of a drug discovery programme, although there are exceptions. But, in spite of the wealth of studies reporting successful examples of virtual screening applied in lead discovery projects, there are a number of areas in structure-based virtual screening that present opportunities for improvement. The development of a truly predictive and reliable scoring function is welcome, as all of the current scoring functions display variable performance and are protein target and chemical series dependent.

In the difficult cases, elaborate filtering methods or target-specific scoring techniques are often needed to enrich the selected subset to an acceptable level. More challenging still is that a protein is an inherently dynamic and flexible entity that can respond to the binding of a small molecule and other external stimuli. During virtual screening, the ability of the protein side chains and backbone atoms of the binding site to alter geometrically to changes in the coordinates and conformation of the ligand, the so-called induced fit, is generally not handled well. As a result, opportunities may be missed to explore novel leads, which otherwise would score well and be ranked highly. Currently, activities in this area of developmental research are generating promising results.

For in-house compound collections, the principle use of virtual screening is to make subsets of those compounds, preferably new ones, for further processing. Virtual chemical libraries perhaps represent the greatest potential application of virtual screening in lead discovery. Mining large, untapped reserves of new chemically-accessible virtual compounds is an underexploited avenue of research, partly due to speed and accuracy concerns, along with the inability to assess 100sM or more compounds in a focused and timely manner. But, given that the number of chemically accessible compounds in the universe is around 10^{63} , and arguably, a significant number of these are potentially drug-like in nature, the would-be rewards for exploring such massive chemical libraries are clearly manifold.

Despite considerable scope for further improvement, virtual screening has proven useful in a number of drug discovery projects. In combination with the well-established ligand-based methods, the reliable molecular docking of compounds and rapid screening of large compound libraries, both real and virtual, will yield many more success stories in the future. **END**

Authors

Jin Li and Timothy Lovell work at GSI Chemical Computing of AstraZeneca R&D

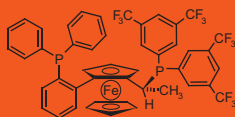
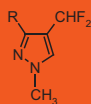
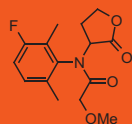
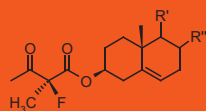
Acknowledgement

The authors would like to thank Dr Hongming Chen for providing the initial data quoted in Figure 2, and many other AstraZeneca colleagues in the effort to advance these methods and their applications in the discovery projects.

Advanced Fluorinations

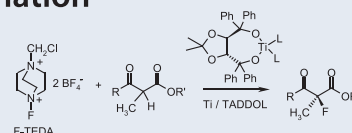
A Powerful Selection

solvias



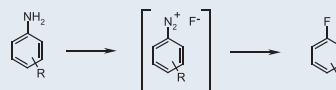
Enantioselective Catalytic Fluorination

- The first asymmetric fluorination
- Up to 90 % ee



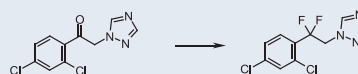
Modified Schiemann Reaction

- Broad scope
- Mild reaction conditions
- Scalable to production scale



Deoxofluorination with SF₄ and HF

- High chemoselectivity
- Also applicable for CF₃ introduction
- Viable method up to kg quantities



Solvias Services:

- Synthesis of customized molecules
- Process R&D

Solvias AG
P.O. Box
4002 Basel
Switzerland
Tel. +41 61 686 61 61
Fax +41 61 686 65 65
info@solvias.com
www.solvias.com

circle 18