Make Way for Virtual Screening

Virtual screening is being hailed by some as a superior method of evaluating compounds compared to high throughput screening. However, it’s still important to choose wisely between a ligand- or protein-based model template, as both can be prone to mishandling.

Over the last 20 years, high throughput screening (HTS) has been the standard method of discovering new active chemical matter. However, conducting a high throughput screen of a large compound collection against a given drug target is a complex, time-consuming and expensive process with no guarantee of success. By contrast, virtual screening is an effective and cheaper method of narrowing down a large compound collection to a smaller focused set, with a higher chance of hitting the target.

Virtual screening is a computational method for evaluating a large number of compounds in order to assess whether they are likely to be active at a biological target of interest. It is also an effective technique for discovering new hits, or for identifying a back-up series for a discovery project. Unlike an HTS campaign, a virtual screen requires some starting information before it can begin; this has slowed its uptake as the default method for starting discovery projects. But recent advances in protein crystallography have narrowed the gap, thus enabling virtual screening to take its place as a full replacement for HTS.

Traditionally, virtual screening would search a database of molecules comprising the structures of the millions of real compounds from a corporate collection, or from vendors. As virtual screening methods and throughput have improved, so the size of the database that can be screened has increased. Virtual screening is now being used to screen tens or hundreds of millions of virtual compounds, created by enumerating a combinatorial library or by decorating known scaffolds with novel side-chains, resulting in a knowledge-based selection of compounds to synthesise. Whatever the database that is screened, virtual screening results in a focused set of compounds for wet screening in high information content assays.

Quality for Less

One of the key drivers towards the adoption of virtual screening as the method of choice for starting discovery projects is the desire for high-quality biological data on all hit molecules at a fraction of the cost of HTS. Virtual screening provides a fast, effective and inexpensive way of triaging large databases of compounds to get a small number that can be assayed intensively, giving high-quality data.

Outsourcing a virtual screening project to a specialist company will typically cost £10,000-20,000. Compared to the cost of a high throughput screen, which could cost in the region of £2 million, this is remarkably cost-effective.

Choosing the Right Screening Method

There are two main approaches to virtual screening: ligand-based virtual screening and protein structure-based virtual screening. The key difference between the two methods is whether the protein or the ligand is used as a model template against which potential screening compounds can be virtually tested. If the interaction of a ligand with a protein is thought of as a lock and key mechanism, then protein-based virtual screening is the equivalent of taking the lock as the template, rather than the key.

There are many pros and cons to ligand- and protein-based virtual screening, and both can be prone to mishandling. Each approach has its strengths,
and the results obtained from either method can be comparable and, in many cases, highly complementary.

**Protein Structure-Based Screening**

Protein structure-based screening, often referred to as ‘docking’, involves taking the structure of the target protein and calculating the binding energy for a series of ligand structures in one or more poses. A pose is a given conformation of the ligand in a docked position within the protein’s active site.

Theoretically, docking should be the method of choice for all virtual screens. Docking can perform very well if conducted by an expert on an amenable protein target, but there are many situations where it is not as effective as might be desired. A number of challenges presented by this approach reduce its effectiveness, and can lead to the use of ligand-based methods in preference.

Firstly, it is most useful to consider the protein structure in a relevant activated state, but this is not always possible and where it could become a viable option, the protein may contain flexible loops that significantly alter the shape of the active site.

Secondly, there are a number of different approaches to the computationally intensive process of determining the poses to be scored. Each approach has certain benefits in some circumstances, but knowing which is best for the current target is difficult to determine.

Finally, there are a number of methods used to calculate a score for the generated poses. This is usually a pseudo-binding energy score, and relies heavily on the method that was used to parameterise the scoring function. No single scoring function performs well in all circumstances.

There are, therefore, many different docking methods; that is, combinations of pose generation and pose scoring. Each has its own strengths and weaknesses, and there is no clear ‘best’ docking method. This can leave the virtual screening practitioner in a quandary over which method to choose.

**Ligand-Based Screening**

Ligand-based virtual screening does not rely on the use of a protein structure to guide the process. The method depends on the assumption that an active ligand or ligands will define the required features for activity in general against the given target.

Methods are usually based either on the finding of compounds in the database with a similar atomic structure to known actives, or upon discovering compounds that can make similar interactions with the protein as the known actives. In either case, the drug targets do not need to have an elucidated structure. This makes ligand-based virtual screening ideal for ion channels and G protein-coupled receptors, which represent the majority of drug targets.

Using atomistic methods for ligand-based virtual screening, such as ‘2D similarity’ or ‘substructure’ searching, has long been used in conjunction with HTS to flesh out any new active molecule with structure-activity relationship information. The great advantage of these methods is the speed at which they can return compounds that have a high probability of being active. Unfortunately, the very nature of these methods means that it is difficult to find actives that are truly novel. By definition, any compounds found are structurally similar to the query molecule, and hence are likely to share any absorption, distribution, metabolism, excretion, toxicity (ADMET) or intellectual property issues it may have.

In contrast to atomistic methods, ligand-based virtual screening approaches that mimic the interactions of known actives with the protein in 3D, excel at generating novel actives with the freedom to operate. However, using a 3D-based method does slow down the rate at which compounds can be virtually screened. It also incurs an overhead in the generation of the database in a 3D format through conformation exploration. Nonetheless, these methods are quite capable of screening tens of millions of compounds in a short timeframe of 24 to 48 hours by deploying calculations to a Linux cluster.

Historically, using a 3D ligand-based screening method meant using pharmacophores that were defined as feature-based labels of the interactions of ligand and protein, to search for novel actives. However, a desire for more physical-based methods led to the use of the shape of the known active when bound to the protein, and a renaissance in 3D ligand-based screening. These simplistic physical models are now being superseded by more detailed models that combine the shape of actives with physically derived features, such as the electrostatic environment surrounding the active molecule, to give new, high information content pharmacophores.
Additionally, compounds with very similar shapes can show radically different biological properties, driven by differing electrostatic or hydrophobic properties. A demonstration of this can be seen with a simple series of substituted benzenes (see Figure 2).

**Electrostatics Mean Realistic Pharmacophores**

Developing a more realistic and descriptive pharmacophore requires using the electrostatic, shape and hydrophobic properties of a molecule in combination in order to build up a detailed ‘protein’s-eye view’ of the molecule. These molecular fields yield an immense amount of information about the properties of a ligand and the likely interactions it will make with a protein.

Using the full electrostatic environment around active ligands is computationally expensive. One approach is to distil the fields down to only the most intense regions in space. Termed ‘field points’, these can be rapidly calculated and compared between molecules. The field points represent a physically derived pharmacophore that is far more relevant to the biological activity of a molecule than 2D chemical structure, or simple 3D shape.

Compounds with very different 2D properties can display similar biological activity (see Figure 1).

Using the full electrostatic environment around active ligands is computationally expensive. One approach is to distil the fields down to only the most intense regions in space. Termed ‘field points’, these can be rapidly calculated and compared between molecules.
approaches will further increase the utility of virtual screening, reducing the time taken to run searches, or alternatively expanding the available search space for large virtual chemistry libraries.

**HTS Rescue**

Both protein structure-based and ligand-based virtual screening have a number of potential application areas, including HTS rescue and identifying robust backup series.

HTS rescue uses virtual screening – which is implemented as an insurance method – to remove false negatives from wet screening experiments and to find new active compounds that failed the HTS assay. In HTS, the biological result measured is usually a per cent inhibition of protein activity at a single fixed concentration of the ligand. This may be measured once, in duplicate or in triplicate and the results averaged. Often, even where the per cent inhibition has been measured more than once, there can be a relatively high rate of false negatives and positives in the assay.

Running a virtual screen of the compounds alongside the wet screening allows for the identification of potential false negatives and positives. It also enriches the compounds selected for follow-up hit confirmation in which a compound will be studied in more detail and its activity measured at multiple concentrations to give a more accurate activity, such as an IC50 value.

**Generating Backup Series**

As a drug discovery program progresses, it is important to have a good backup series in waiting, in case unexpected ADMET or other development issues affect the main series. It is usually desirable to have a backup series which is structurally highly distinct from the main series, to minimise the risk of it having the same liabilities. Ligand-based 3D virtual screening is often an excellent choice for finding such a backup. The wealth of ligand-based knowledge acquired can then guide the screening process, providing good hit rates, and searching by shape or electrostatic space provides hits with low structural similarity.

**Proven Technology**

Choosing the most appropriate virtual screening method depends on a number of factors, including the application of the method and the available structural and ligand data. The choice of ligand- or structure-based virtual screening is largely dependent on the quality of data available. The results of each method are largely comparable, and the choice of approach must be made on a per project basis. Whichever technique is used, expert knowledge is critical to the success of a virtual screening experiment.

Virtual screening is a proven technology that can replace large-scale HTS runs in order to create a more focused data set for a smaller-scale screen. It is a useful tool to be run as an adjunct to HTS in order to validate the results, and also has applications for chemotype switching and generating backup series. In these cases, the use of a well-validated 3D ligand-based virtual screening method is an efficient and cost-effective way to make the leap to new structures of interest.