Integrating computational methods:Layout 1 14/1/10 20:00 Page 61

Integrating computational methods and drug discovery – avoiding pitfalls in chemical space

The immensity of the task in discovering suitable molecules for development among the vast number which are theoretically available is well stated. With 90% of potentially efficacious compounds failing at latter stages, the earlier toxic candidates are dropped from the development process the greater the amount of time and money saved. We discuss the advantages and limitations of the integration of computational methods into the drug discovery process involving the use of a reliable *in silico* screening method.

By Alexandre Ismail and Professor Philippe Manivet

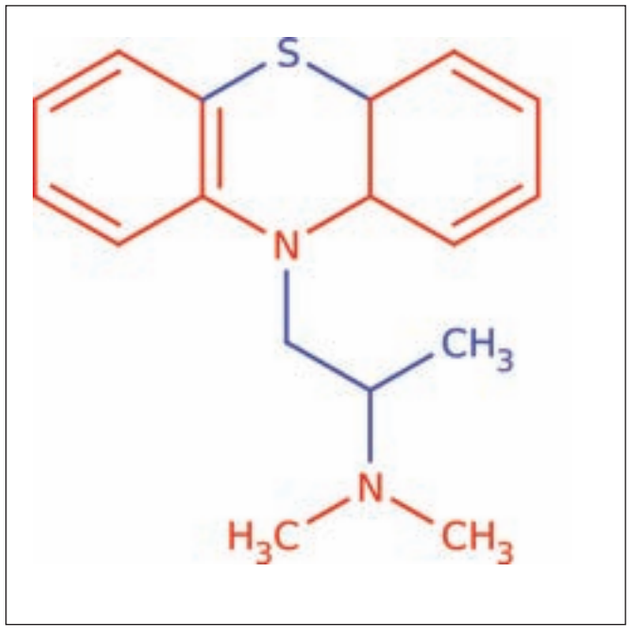
Chemical space is the set of all possible compounds, and it is vast – it has been estimated that there are 10^{60} possible small-molecule compounds in this space. This number exceeds the number of atoms in the universe required to construct them. While biological systems have explored and used only a small fraction of this enormous space, this reduced set is still very large. In addition to small molecules, biological systems use a myriad of proteins to accomplish vital tasks.

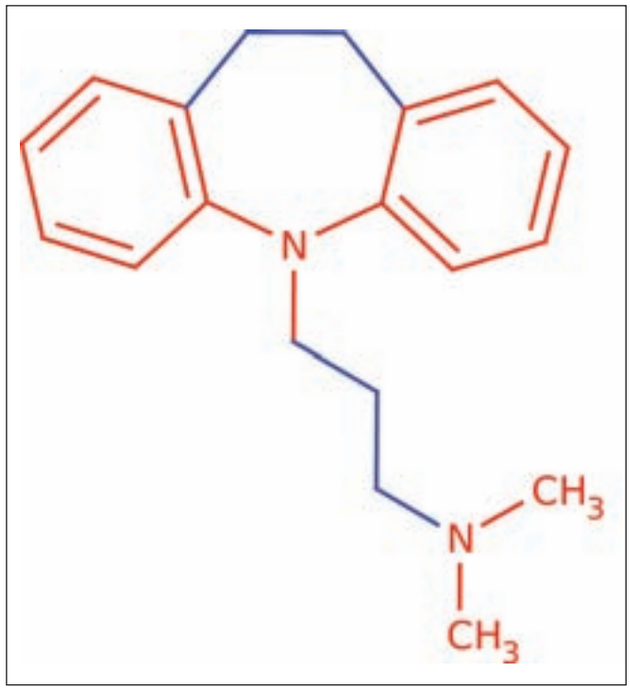
It is among this set of natural molecular structures that the drug discovery process seeks its targets. Conceiving drugs to accurately and precisely hit targets in this biologically relevant space is a problem of the needle-in-a-haystack type. It is further complicated by the spatial arrangement of these molecular types: targets are often sequestered behind membranes and other layers of molecular and cellular machinery. This presents the challenge for the drug molecule to penetrate these barriers and accurately bind to the target in order to be effective at a low concentration. Often, these intervening structures also contain molecules structurally related to the target, which may be affected by the drug. Such off-target reactions can cause toxic effects in the organism, despite the potency of the proposed drug. Toxicity is a major pitfall in drug development, and one that often remains undetected until it is too late.

Search methods

In industrial terms, the searching of biologically relevant chemical space often begins with an HTS campaign. This methodology has the strength of exploring chemical space deeply, but it tends to do so rather narrowly. Complementary approaches, such as fragment-based drug design, can search chemical space more broadly, revealing new chemistries capable of therapeutic effect. While these methods take advantage of molecular and cellular contexts for drug discovery, they are much less adept at revealing the toxicity of a given compound

Integrating computational methods:Layout 1 14/1/10 20:00 Page 62





Figures A1 (left) and B1 (right): 2D representations of promethazine and imipramine: structural similarities drawn in red

in a complete biological system. These approaches can generate valuable hits and lead series, but pruning these results by toxicity remains a necessity. This is an area where a large amount of data has already been collected, and where a system based on this knowledge can be used to accurately predict and eliminate toxic compounds in the discovery and development process.

Therefore, drug discovery consists not only of selecting a potent molecule from a large set, but also of understanding the mapping from potent molecules to therapeutically efficacious drugs. Having a reliable way to establish this mapping is essential, as it allows the searching of chemical space while avoiding the pitfalls of toxicity. Computational methods offer an excellent way of surveying chemical space for toxic 'dead-ends' in development.

Avoiding dead ends in chemical space

Late stage discovery of toxicity in a lead compound means the molecule must return to prior stages of development for optimisation. If it is not possible to 'design away' the toxicity (ie the toxicity results from a pharmacophoric feature of the scaffold), then considerable resources have been spent on a molecule that could never have been a viable drug. There is also the case where toxicity is only revealed after the drug is on the market, endangering patients, exposing the company responsible to major liabilities and resulting in market withdrawals.

Clearly, an effective method for predicting toxicity and characterising its mechanism is essential to avoiding these scenarios in drug development.

Key elements of a toxicology prediction system

A reliable *in silico* toxicity screening method must combine a high fidelity pharmacophore representation with an extensive database of compounds known to be toxic.

Many existing pharmacophore models rely on 2D fragment matching, which is demonstrably ineffective at recognising the biological activity of molecules. This stems mainly from the fact that such systems reduce the dimensionality of the molecular description to one or two dimensions. Two molecules can contain one or several common fragments in their chemical structure, yet not show the same activity and/or the same toxicity mechanism. A good example is promethazine (targeting histamine H1 receptors) and imipramine (targeting presynaptic serotonin and norepinephrine transporters). Both molecules have two similar fragments representing 60% of their structure (**Figures A1** and **B1**). Their respective targets, activities and side effects are, however, completely different. An additive approach considering that whole-molecule physical properties can be estimated by the summation of those 2D fragments is not a correct assumption (**Figures A2, A3** and **B2, B3**).

Integrating computational methods:Layout 1 14/1/10 20:00 Page 63

The biologically relevant conformational and electronic-structure properties of a molecule usually require three (and often more) dimensions for accurate representation and comparison. Once a high-fidelity pharmacophore has been established, it must be used to screen the novel compound against compounds known to be toxic. Databases of these compounds, such as the FDA's DSSTox, represent the knowledge collected from thousands of man-hours spent characterising toxicity. Matching this existing body of work to the questions posed by newly discovered compounds is the key to realising significant time and cost savings in discovery and development.

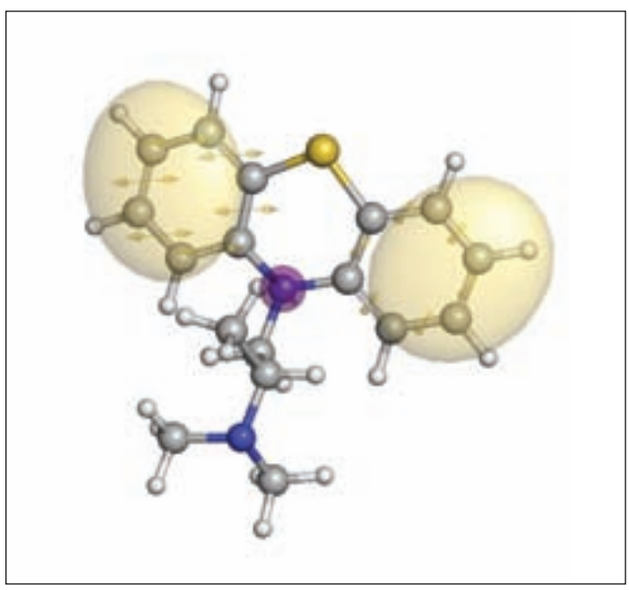
This approach has the advantage of directly analysing the existing data with a rational pharmacophore based on realistic molecular geometry and electronic structure. This replaces intermediary rules of thumb, such as Lipinski's rule-of-five and other criteria of 'drug-likeness'. Such criteria have proved modestly successful at selecting leads for development, but fail to predict the efficacy of biological signalling molecules such as peptides (which do not fit the rules of drug-likeness), as well as toxicity. Indeed, the brevity of such rules is itself an example of an attempt at reducing the dimensionality of the underlying system. However, this approach causes a critical loss of information when understanding molecular systems, which are inherently highly-dimensional. A high-fidelity description of molecular reality is necessary to move from metaphor to pharmacophore, and from simile to simulation.

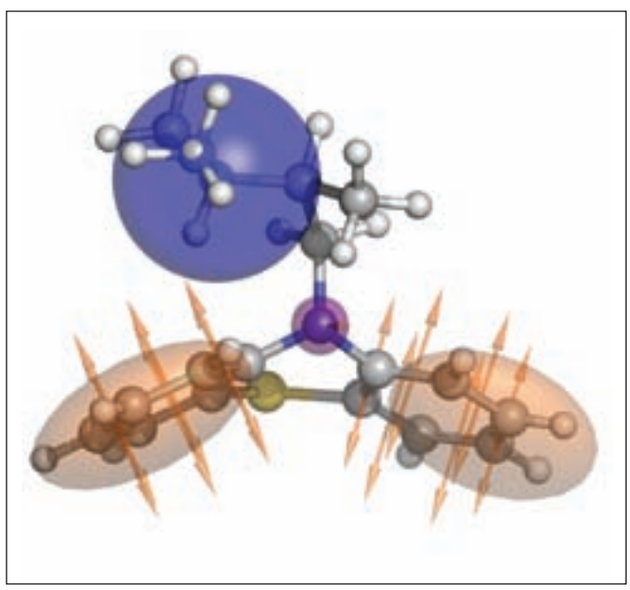
Key applications of toxicity screening in discovery and development

It is an oft-quoted statistic that more than 90% of compounds in development will not become successful drugs. While ballpark figures of this sort may be a direct and unavoidable consequence of the size of chemical space, this single figure encapsulates many variables that can be improved. The earlier toxic candidates are dropped from the development process, the greater the amount of time and money saved.

Virtual library design, pre-screening, and cross-screening

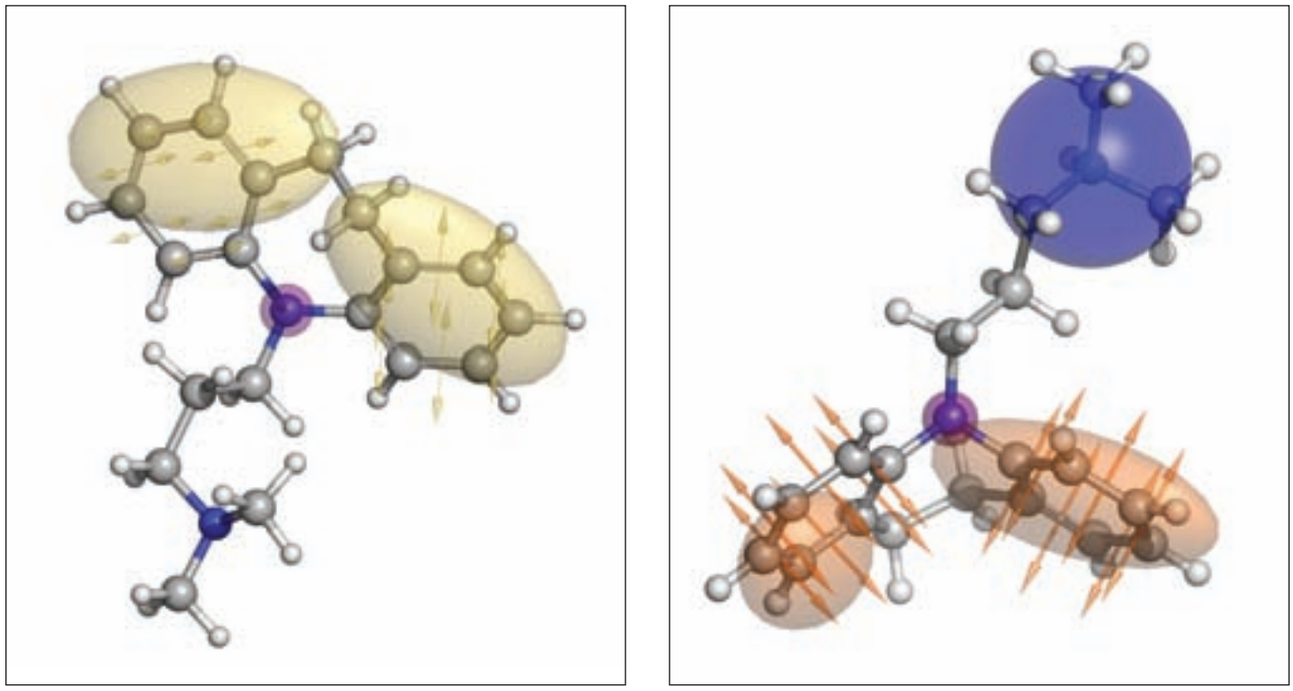
A high quality pharmacophore can improve key stages of discovery and development. When applied to the very beginning of the ligand-based discovery process, it can be the basis of a rationally designed virtual library. This can better orient an initial HTS/HCS search of chemical space. One of the characteristic limitations of HTS is the library it is based on. These libraries tend to be variations on a scaffold. While there may be a great number of unique compounds in such an HTS 'deck', their derivation from a small set of scaffolds means they are all rather closely related.





Figures A2 (left) and A3 (right): 3D biologically relevant conformer pharmacophore representations. Note the large deviation from planarity in imipramine's central ring, as well as the conformational difference of the tertiary amine chain. *Pharmacophore legend – orange arrows: aromatic centres; orange ellipses: hydrophobic centre; blue sphere: polar centre (omitted in A2 and B2 for clarity); purple sphere: H-bond donor/acceptor site; red arrow: H-bond donor/acceptor vector*

Integrating computational methods:Layout 1 14/1/10 20:00 Page 64



Figures B2 (left) and B3 (right): Another view of the same pair. Note the puckering of imipramine's central ring and the consequent re-orientation of the side rings. In three dimensions, apparently similar structures yield significantly different conformations and hydrogen bonding patterns. These yield definitive differences for interactions with biological targets. *Pharmacophore legend – orange arrows: aromatic centres; orange ellipses: hydrophobic centre; blue sphere: polar centre (omitted in A2 and B2 for clarity); purple sphere: H-bond donor/acceptor site; red arrow: H-bond donor/acceptor vector*

That is to say, HTS can explore chemical space deeply along the axis of a given scaffold, but it is this same underpinning of library construction which makes the search space narrow.

In addition, libraries of this sort are often not pre-screened for toxicity. The subsequent HTS search may reveal compounds with nanomolar activities at the level of individual cells that are toxic at the level of the whole organism. Thus, a high definition pharmacophore can be used to design focused virtual libraries and remove toxic hits from the search space before HTS resources are wasted on them.

After initial identification of hits, hit confirmation can also benefit from computational methods. Cross-screening assays can be more accurately and efficiently selected if the toxicophoric points of a hit have already been well characterised. Hit expansion can also be done in a more rationally directed way. This can save significant resources when applied to large scale HTS campaigns.

Virtual fragment linking

Fragment-based drug design is an approach that is inherently structurally based. Crystallography of fragments bound to targets and SAR-by-NMR are experimental techniques that yield highly relevant structural data, such as the binding pose of a given fragment. The next step of FBDD is to link these fragments into a larger therapeutic molecule. This is another stage where computational methods can be used to guide design and evaluate molecular properties.

Binding immediately reveals the importance of the ligand's conformational flexibility and electronic structure, and these are features that must be evaluated on the molecule as a whole. While FBDD by definition builds the atomic skeleton by fragments, it does so by adding linking moieties that do not participate directly in binding to join the binding moieties together. The resulting molecule is not simply the sum of its parts, but rather a new chemical entity, which may have unexpected properties. These constructs must be evaluated for other parameters relevant to a drug, such as stability, reactivity, solubility and toxicity. Optimisation of these properties can be performed at this key stage of development *in silico*. Medicinal chemistry resources need not be wasted on synthesising F2L (fragment-to-lead) candidates that can already be pruned out for toxicity or other pharmacophoric features.

Integrating computational methods:Layout 1 14/1/10 20:00 Page 65

Computational methods: guide and complement to existing techniques

Robust computational methods are uniquely adapted to dealing with the problems of chemical space. Experimental methods are essential for the generation of new structural information. However, on the scale of industrial drug discovery programmes, complementary computational techniques are necessary to focus experimental resources efficiently. Such programmes can benefit immediately from reliable early stage derisking made possible by robust computational methods, as well as increased efficiency in lead-optimisation. DDW

Professor Philippe Manivet *is the President of the Scientific Committee of BioQuanta. He is also director of BioQuanta's* in silico *department, where he has led the development of the MultiDIP, a next-generation molecular modelling platform for identifying and characterising the toxicity and biological activity of small molecules. He has pioneered the intensive use of* in silico *methods to greatly accelerate the molecular development cycle for external clients, as well as BioQuanta's own diagnostic reagents. He is currently a member of the ECVAM expert pool for* in silico *and* in vitro *alternative methods for predicting xenobiotic toxicity. Professor Manivet holds a PhD in theoretical chemistry and biochemistry from Ecole Polytechnique of Paris, France. He is a hospital practitioner in clinical biochemistry and haematology. He has co-authored more than 30 international peer-reviewed publications as well as several patents.*

Alexandre Ismail is a Master's student at supBiotech in Paris. He holds a BA in biochemistry from Hunter College, and has participated in computational chemistry research programmes at Hunter College, Rutgers University, Wolfram Research, Duquesne University, the University of Pittsburgh and the City College of New York. He has also worked as a molecular modelling consultant at Avatar Biotechnologies for the rational design of post-translational modifications to protein drugs.