Molecular Field Technology and its Applications in Drug Discovery

Translating molecular structure information into molecular fields helps to decode ligand-protein binding, and facilitates the discovery and design of truly novel, active compounds.

By Sally Rose and Andy Vinter at Cresset BioMolecular Discovery Limited

Dr Sally Rose is Director of Business Development at Cresset BioMolecular Discovery Ltd. Previously, she was a Founder and Director of Molecular Informatics at BioFocus and, prior to that, Head of Small Molecule Modelling at GlaxoWellcome. Dr Rose has a PhD in Molecular Diversity from Reading University, UK.

Dr Andy Vinter is Founder and Chief Scientific Officer at Cresset BioMolecular Discovery Ltd. From 1964 to 1990, he worked as an organic, physical and computational chemist for Wellcome and SKB. From 1990 to 2001, he returned to academic research at the University of Cambridge and consulted on molecular modelling for many companies involved in drug discovery. His academic period allowed him to develop new computational science which culminated in the formation of Cresset BioMolecular Discovery Ltd in 2001.

Many molecular modellers and medicinal chemists have become frustrated by the limitations of the current state-of-the-art of molecular modelling technology. Too often, modelling studies fail to explain the difference in activity between compounds. Docking methodologies for example can be especially poorly parameterised, in some cases with results being no better than simply comparing how closely the volumes of two molecules can overlay. A solution to this problem is to look deeper than structure by modelling compounds using their molecular fields. This approach has wide-ranging applications in drug discovery from hit-finding through to lead optimisation – but to be effective it requires a radical change in the way we think about molecular structure. At Cresset, we have successfully applied the technology to virtual screening and finding the bioactive conformation of ligands.

MOLECULAR FIELD TECHNOLOGY

Molecular fields provide a way of analysing the surface properties of molecules that allows us to understand how the atomic structure of a compound can be translated into biologically-relevant binding properties. Modelling in 'field' rather than 'structural' space facilitates more innovative discoveries, resulting in a diversification of the chemotypes of ligands which are active against a specific target. Longer term, this will result in a more diverse range of drugs against a target, providing improved choice for patients.

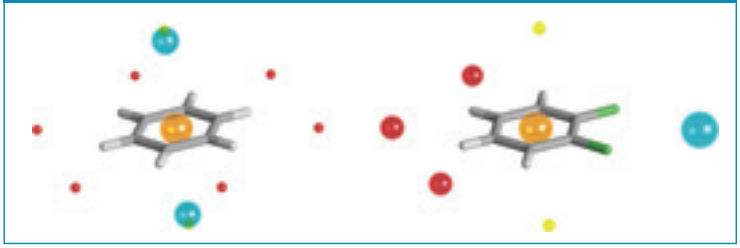
Traditionally, medicinal chemists have optimised molecules by making relatively small changes to structure based on experience. However, modifying fields in a predictable way allows the chemist to develop new skills and think about molecules in a different way.

This is demonstrated in the following simple example. A medicinal chemist is optimising a pendant benzyl group attached to a core structure by varying substitution around the phenyl ring. Addition of 2,3-di-F substituents is found to increase activity relative to the unsubstituted ring. As fluorine and hydrogen are similar in size, the classic logic would be to infer that the fluorines are electronegative compared to hydrogen, and so must be making a favourable interaction with a positive group on the protein. However, in field terms, addition of the di-fluoro group has three main effects:

- The fluorines create a negative field on their side of the phenyl ring
- The hydrogens on the opposite side of the phenyl ring become more positive
- The pi cloud of electrons above and below the phenyl ring disappears

These effects are shown in Figure 1, represented as field points (see description below); any one of these factors

Figure 1: Comparison of the fields of benzene (left) and di-fluoro benzene (right), represented as field points



could be important for increasing activity. (Further support for this theory can be found in reference 1.)

Cresset calculates four molecular fields to represent the binding properties of a ligand. These are:

- Positive electrostatic (coloured red)
- Negative electrostatic (coloured blue)
- Van der Waals attractive i.e. 'steric' (coloured yellow)
- Hydrophobic (coloured orange)

The fields are calculated by determining the interaction of a probe atom (carrying a +1, 0 or -1 charge) at the surface of a molecule. The fields form a continuous skin around the atomic skeleton and 'field points' are placed at the extreme values of the fields (local maxima). These are represented as spheres, colour-coded as shown in the list above and whose diameter represents the magnitude of the extrema. The hydrophobic field is calculated in a slightly different way and field points are located at the centre of hydrophobic groups such as phenyl, halogens and alkyl groups. The field points occur at regions around the molecule where binding interactions are likely to be strongest, and so they provide a summary of the entire field and its biologically important sites. Figure 2 shows an example of the relationship between field points and molecular structure for a hypothetical molecule.

The field point pattern is a sophisticated 'pharmacophore' which can be used to define a template for binding. Molecules can be overlaid using their fields, rather than structure, and the field similarity between two molecules can be quantified and converted to a similarity value – for example, for determining the similarity of a compound to a known active ligand in virtual screening. (For more detailed technical information see references 2 and 3.)

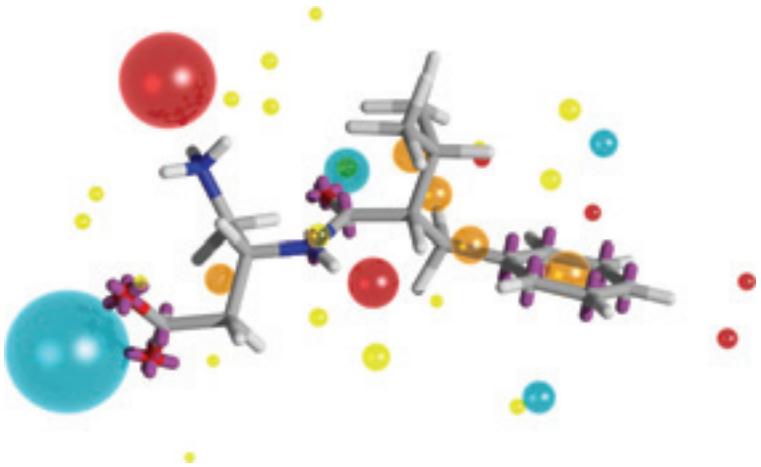


Figure 2: The field points around a hypothetical molecule. The field points occur at regions where binding to a protein is most likely to occur

THE XED FORCE FIELD

Accurate fields require accurate 3D molecular models with the correct electron distribution. However, most molecular mechanics force fields handle electron distribution poorly because of their reliance on atom-centred charge models that do not reflect the delocalisation found in pi systems. These are abundant in nature; for example, the side-chains of nine of the 20 naturally occurring L-amino acids contain pi systems, and the DNA helix is dominated by pi-stacking interactions. It is therefore crucial to model pi-electron interactions accurately to correctly interpret protein-ligand binding.

At Cresset, we have developed the eXtended Electron Distribution (XED) force field to address this problem. XED puts electron density where it should be – that is, away from the atom centres in pi systems (4, 5). The XED force field generates molecular fields which correctly reproduce experimental data. The force field, molecular fields and field points have been extensively validated using x-ray data from the Cambridge Crystallographic Data Centre (6).

APPLICATIONS OF FIELDS TO DRUG DISCOVERY

Ligand-Based Virtual Screening

Our first application of field technology to drug discovery was in ligand-based virtual screening. The software, called FieldScreen™, runs on a linux cluster from an easy-to-use web-based graphical user interface (GUI). The technology is based on the assumption that two molecules that bind to the same protein site would be expected to have the same field pattern in their bound conformation. For virtual screening, we simply take the field pattern of a known active ligand in its bioactive conformation, and use this to virtually screen a database of molecules in field format to find molecules with a similar field to the template. These molecules have a high probability of being active. As fields vary with molecular conformation, each molecule in the database is present up to 100 times in different conformations, each with its associated field.

The ideal template can be generated directly from the 3D structure of a ligand which has been co-crystallised with its target protein. Alternatively, a docked ligand or a 3D overlay model of active compounds may be used.

In practice, the field approach is especially useful for targets with no x-ray data (and hence docking is not an option), but where there are a few known active ligands that can be used to define a field model for activity. This is described in the following section. If target x-ray information is available, it can be used to define an excluded volume during the virtual screening process and so further refine the results.

FieldScreen™ is more flexible than the ligand-based pharmacophore methods as:

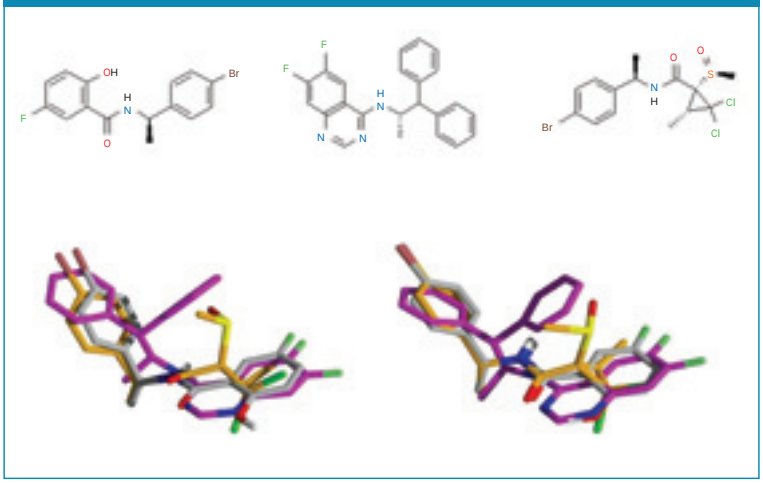
- There is no need to pre-define pharmacophoric groups, because any atom or group capable of creating an area of local electron deficiency or richness is automatically identified by the force field
- There are no problems with too many or few hits, because a list of the entire virtual database is returned, sorted by similarity to the template, so the user can select the desired number of compounds to test from the top of the list

Field-based virtual screening technology is ideal for major pharma companies with vast in-house compound collections who wish to rationalise their high-throughput screening. It is also readily accessible to small biotech companies who can access Cresset's database of circa 4.5 million commercially-available compounds (stored in field format), via consultancy projects, to select small sets of compounds for biological screening. This is a highly cost-effective way to kick-start a hit-finding project for small companies that lack an in-house compound collection of sufficient size to be suitable for HTS. Virtual screening projects generally take about two weeks. An analysis of past projects shows that our hit-finding 'success-rate' is >80%, where success is defined as a client finding hits they wish to follow up from screening tens to a few hundred compounds.

Modelling the Bioactive Conformation

A feature of fields is their high dependence on molecular conformation. Although at first this may appear to be a major complication, it can also be used to our advantage to further develop applications of field technology. One such application is to find the bioactive conformation of a set of ligands starting from 2D structures and using no 3D target information. This software (FieldTemplater™) runs under command-line mode in linux; a Windows GUI is under development. As previously mentioned, if two or more ligands can bind to the same protein target and elicit the same biological response, then the likelihood is that they will present the same field pattern

Figure 3: The FieldTemplater 'bioactive' conformation model (right) compared with the overlay seen by comparing the x-ray data of the ligand-protein complexes for three syctalone dehydratase ligands

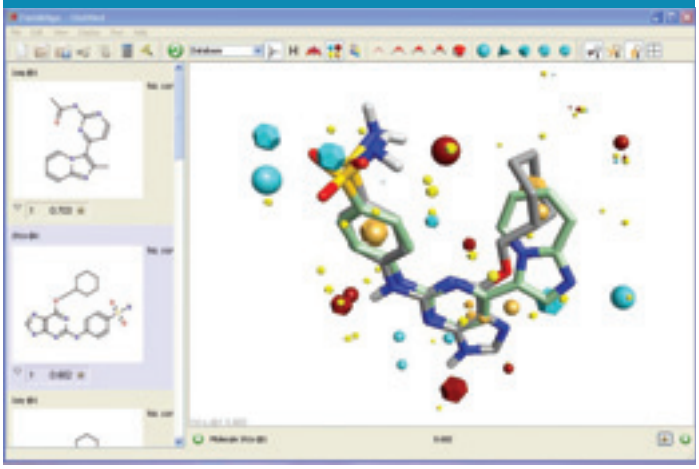


to the target in their bound form. Therefore, we can analyse the field patterns for two different ligands in a variety of conformations and identify those pairs which have similar field patterns. These pairs of conformations are hypotheses for the bioactive conformation. In practice, we generally compare sets of conformations for three or more diverse ligands to reduce the number of chance matches and produce more accurate results.

To date we have compared FieldTemplater™ alignments with x-ray data from ligand-protein complexes for seven targets (p38, PDE4, thrombin, HIV NNRTI, neuraminidase A, CDK-2, and scytalone dehydratase). FieldTemplater found the bioactive conformation (within 1 angstrom root mean square deviation of the x-ray conformation) in the majority of cases, and produced a result that was close enough to be useful in almost all the other instances. An example is shown in Figure 3 for scytalone dehydratase. However, it should be noted that the great value of the method is in defining bioactive

The field point pattern is a sophisticated 'pharmacophore' which can be used to define a template for binding. Molecules can be overlaid using their fields, rather than structure, and the field similarity between two molecules can be quantified and converted to a similarity value – for example, for determining the similarity of a compound to a known active ligand in virtual screening.

Figure 4: A screenshot from FieldAlign showing the 3D bioactive conformation (green) and the overlay obtained for a molecule entered in 2D (grey, second molecule down on left-hand 2D list)



conformation models for ligands of transmembrane protein targets, such as GPRCs and ion channels, where no x-ray data of the target are available and docking studies based on speculative homology models are characterised by large errors and uncertainty.

Molecular Field Overlays in Lead Optimisation

Once a field template has been generated, it can be used to align multiple molecules as part of the lead optimisation process. This software has recently been released (FieldAlign™) and is available to run under Windows (with an easy-to-use GUI) or linux (command-line mode or GUI). The Windows GUI was specifically developed for use by medicinal chemists. The template molecule is entered in its 3D bioactive conformation and any compound which the modeller or chemist wishes to align to the template is entered in 2D. The system provides a rapid, easy way to align compounds to a 3D molecule template using fields. A screenshot is shown in Figure 4.

We envisage a variety of applications for FieldAlign™, including:

- Interpretation of SAR (Structure-Activity Relationships) – with respect to field point patterns
- Translation of SAR between series – overlay series in field space and identify equivalent substitution sites
- Exploration of lead optimisation ideas – design molecules to test the importance of specific field points
- Identification of fragments which are isosteric in field space – invent novel bioisosteres; and Replace non-ideal functional groups
- Virtual library design – select scaffold and reagent (monomer) combinations with high field similarity to known actives
- Align compounds for 3D QSAR – prior to CoMFA; and for 3D property calculation

WHERE NEXT?

Cresset's field software has only been available for two years but has already seen significant uptake by both global pharma and small biotech companies. There are many rational and highly innovative ways we can develop the technology over the next few years. One of our main focuses will be to develop software which is suitable for use by both medicinal chemists and molecular modellers. We firmly believe that much potential innovation from medicinal chemists has been stifled by lack of access to good, easy-to-use modelling software that can support them in unravelling SAR and prioritising their synthetic ideas.

The authors can be contacted at s.rose@cresset-bmd.com

References

1. **Matsushima A., Fujita, T., Nose, T. and Shimohigashi, Y. 'Edge-to-face CH/_ interaction between ligand phe-phenyl and receptor aromatice group in the thrombin receptor activation', *J. Biochem.* (2000) 128, 225-232**

2. **Cheeseright, T., Mackey, M., Rose, S. and Vinter, A., 'Molecular field technology applied to virtual screening and finding the bioactive conformation', *Expert Opin. Drug Discov.* (2007) 2, 131-144**

3. **Cheeseright, T., Mackey, M., Rose, S. and Vinter, A., 'Molecular field extrema as descriptors of biological activity: definition and validation', *J. Chem. Inf. Model.* (2006) 46, 665-676**

4. **Vinter, J.G., 'Extended electron distributions applied to the molecular mechanics of intermolecular interactions', *J. Comp-Aid. Mol. Des.* (1994) 8, 653-668**

5. **Vinter, J.G., 'Extended electron distributions applied to the molecular mechanics of intermolecular interactions. II Organic complexes', J. *Comp-Aid. Mol. Des.* (1996) 10, 417-426**

6. **Cambridge Crystallographic Data Centre: www.ccdc.cam.ak.uk**