Finding Better Leads using Molecular Fields

Sally Rose, Tim Cheeseright, Cresset BioMolecular Discovery Ltd

2D drawings are a ubiquitous representation for molecular structures. Despite this, they provide only a limited insight into the activity and molecular properties of a compound. Cresset's Field based representations provide much more accurate descriptions of a compound's shape and binding interactions. Fields can be used throughout the drug discovery and development process from escaping the chemotype trap in lead optimization to defining optimal patent protection strategies. Example protocols and a validation case study based on lead optimization of CCK-2 antagonists are presented in the article below.

The Limitations of 2D Structures

While we all are trained to instantly recognise and think about compounds in 2D structure terms, we do not often consider how much insight they really give us into the properties and activities of our molecules. We use our experience to guide our decisions, but we are constantly aware that we cannot always accurately predict the impact that even small changes to a 2D structure will have on important molecular properties such as ADME, toxicity and activity.

In many ways, looking at the 2D structure is like looking at a human skull. While we might see the basic underpinnings of the structure, we cannot see the flesh and movement of the face that allows us to recognise the person and understand their personality. Chemists rely on their experience to make structural changes that they feel will have the desired effect, but in reality have very few tools that allow them to predict before synthesizing and testing the compounds that this is actually the case. This has several unfortunate consequences.

For medicinal chemists there are two main issues. Firstly, many structurally similar compounds that do not have the desired properties get synthesized and tested. This has a significant cost in time, money and lost opportunities to make more interesting molecules with the correct activity and properties. Secondly, and perhaps more importantly, chemists find it hard to make large leaps in structural terms whilst still retaining confidence that activity will be retained in the new series. This second problem, known as the chemotype trap, is seen in lead optimization programs when there is a need to change direction or when developing a backup series for the same activity. The inability to predictably identify other chemotypes that will exhibit similar activity causes programs to stall or advance slowly.

Other areas of the drug discovery and development process are impacted by the inadequacy of 2D structure to predict activity. One of the most valuable of these is Intellectual Property and patents. It is common practise to seek protection for an area of chemical space in a Composition of Matter patent by describing Markush expansions of 2D structures with combinatorial sets of prescribed R-groups around a common scaffold. The explicit implication of the use of Markush expansions in a patent is that 'a

group of constituents is considered equivalent for the purposes of the invention'[1] or in other words, that all the potential permutations will have the same biological activity.

All chemists know that this is both incorrect and incomplete. Not only will closely-related derivative structures show a range of activities and properties, some very similar and some quite different from a lead compound, but also and more disturbingly, entirely unrelated structures with radically different chemotypes can exhibit comparable or even better activity than a given lead compound. While seeking protection over a wider chemical space than necessary is unlikely to be a major concern, the potential for competitors to circumvent the protection offered by a Composition of Matter patent is a serious issue, as it undermines the potential value of the invention.

This problem also affects the in-licensing function, which is routinely asked to make very rapid decisions about the potential value of a compound that is offered to the company by a biotech company or other inventor. There are many challenges here and many questions to decide– does the compound work?, does it have side-effects?, can it be made cost-effectively?, how competitive is it?, and what is it worth? being high on the list. A much more subtle question however is how well protected is it? This question is obviously closely related to the patenting strategy employed for the Composition of Matter filings and the coverage of molecules active against the specified target included in the filing. If the chemotype range is very narrow, this coverage may be good. However, if many chemistries satisfy the binding requirements for a target then it is unlikely that a simple Markush based expansion on a single scaffold will provide effective coverage. It may be that many compounds might already be known that would, if tested, exhibit the specified activity. Knowing this would obviously have a significant impact on the potential value of the in-licensed compound.

These facets of the drug discovery and development process are not alone. Other critical areas that rely on understanding and prediction of activity and molecular properties including safety, trials design, marketing and regulatory submission. These are areas of the business where each decision made has the potential to cost or make millions of dollars. We therefore need a way of thinking about molecules that is capable not only of providing a convenient, easily manipulated representation, but also provides much more insight into the activity, properties and similarity of compounds.

Cresset's Molecular Fields

In order to gain more insight into the 'personality' of our molecules we need a means of thinking about chemical structure not in terms of the underlying skeleton, but in terms of its full range of interactions with its protein target. Proteins interact with the ligand's molecular surface, not with the underlying atomic structure. The complementarity of the molecular fields in terms of four properties, positive and negative electrostatics, van der Waals attraction and hydrophobicity, describe the binding properties of two molecules. These fields are continuous properties around a molecule and so are hard to compute and hard to visualize.

[1] "*Patent Law for the Nonlawyer*" (Burton A. Amernick; 2nd edition, 1991)

Cresset therefore developed a method to position 'Field Points' at the locations of the extrema of the field values to represent the regions that are most likely to contribute to important binding interactions. The field points are shown as spheres, sized according to the size of fields they represent. The explanatory power of Field Points is shown in Figure 1 below:

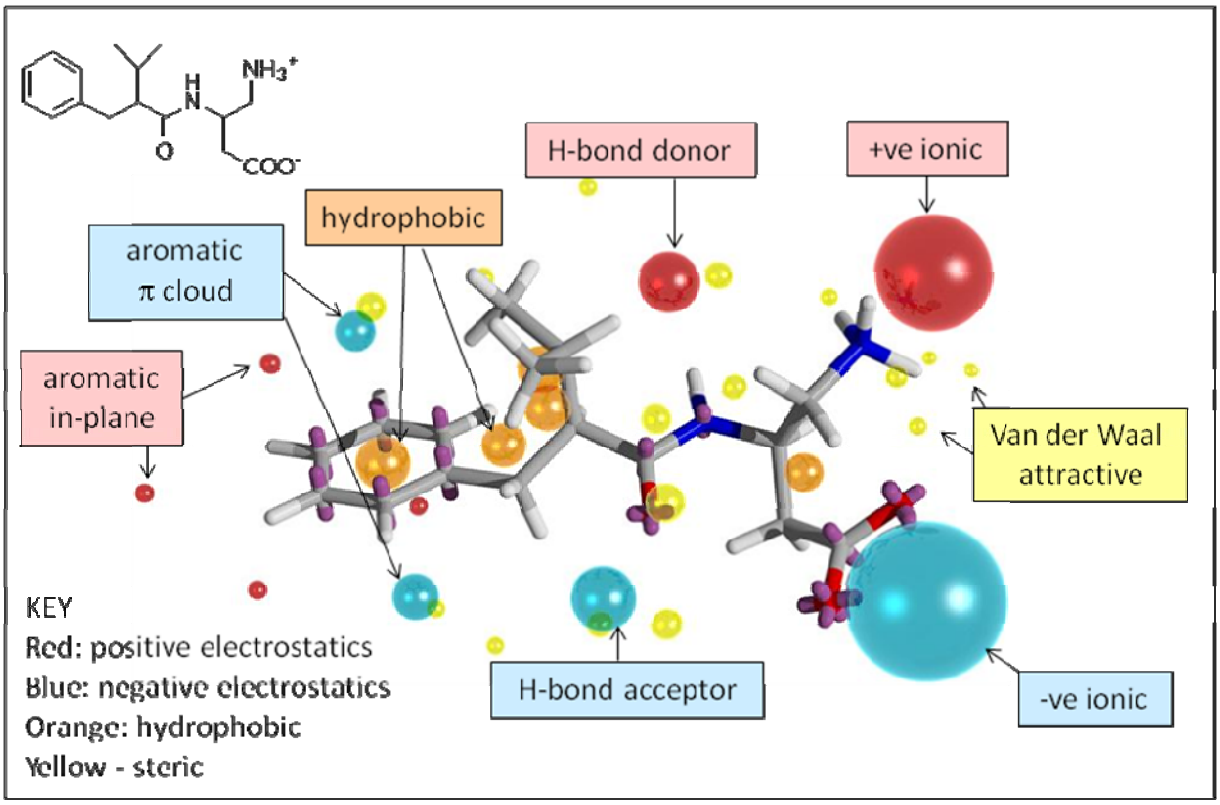

Figure 1. How Fields relate to structure, properties and activity

Cresset's Molecular Field Software

Cresset has developed a number of Field based products to model, analyze and compare the fields of molecules. These products, which are available as easy-to-use Windows PC packages, allow the user to align molecules using fields, shape or a combination of the two.

FieldTemplater™ - Creates a model of the 3D bioactive conformation of active molecules. This requires input of 2 or more active ligands drawn in 2D. It works by identifying multiple conformations for each molecule and searches for a common field pattern across the various conformations of the different molecules. Where it finds a common pattern, this is a hypothesis for the bioactive conformation because all the active ligands can present this same field to the active site. In practice, using 3 (or more) diverse ligands gives remarkably good solutions to this complex problem.

FieldAlign™ – Once you know the bioactive conformation of a single molecule or group of molecules, FieldAlign™ allows you to align any molecule drawn in 2D to this 3D template and calculate a field similarity value. In this way you can compare other actives or unknowns to your activity template and see how the field pattern changes and which field points are conserved in actives. A screenshot of FieldAlign™ is shown in Fig 2.

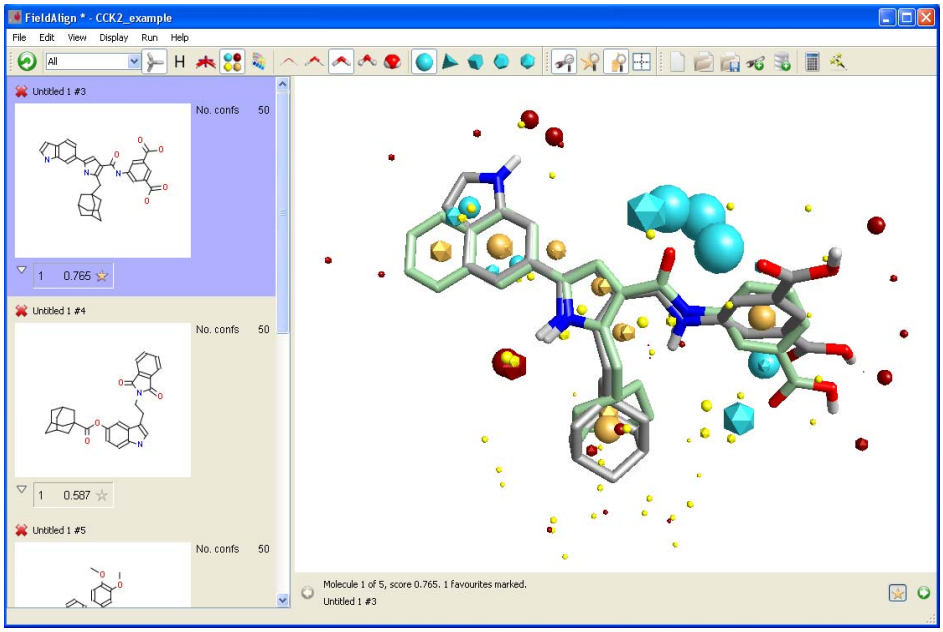

Figure 2. Screen view of FieldAlign

Examples of Using FieldAlign and FieldTemplater

FieldAlign and FieldTemplater can be used by a medicinal chemist on a desktop PC to guide a lead optimization program using a rational approach based around exploring and exploiting molecular field similarities. The software can be used in a straightforward way to help you prioritize your ideas for compounds to synthesize that increase your understanding of the field SAR.

1. Define a Field template for activity

A field template is the field of a single active molecule in its bioactive conformation, or the combined field created from more than 1 aligned active molecules. This defines the field characteristics required for activity. The creation of a robust field model for activity gives a head start to any optimization project as it allows interpretation of SAR within a series and comparison of substitution patterns across different series.

The template can be generated in a variety of ways:

a) The simplest approach, if available, is to use the 3D structure of an active ligand bound to its target protein and extracted from the x-ray complex. This provides the ligand's bioactive conformation and its field is the template for activity.
b) Use Cresset's FieldTemplater™ to create 3D models of the bound conformation from 3 or more 2D structures of active ligands. This is ideal in situations where an x-ray structure of the target protein is not available e.g. for GPCRs and ion channels.
c) If you don't have FieldTemplater™, then you could use a 3D structural model generated by another approach, such as docking or pharmacophore studies, that you had high confidence in.

Once you have a template for activity it can be used to guide scaffold hopping or lead optimization.

2. Scaffold hopping

As noted above, scaffold hopping is essential to escape the chemotype trap. An example is given in Fig 3. It can be achieved using the following practical approach.

a) Use your medicinal chemistry experience and creativity to design novel scaffolds
b) Draw these as final compounds in 2D using ChemDraw (or similar)
c) Copy and paste into FieldAlign™ and calculate the field similarity to your 3D template
d) Select the compounds with highest similarity to synthesize as these will be the most likely compounds to retain biological activity

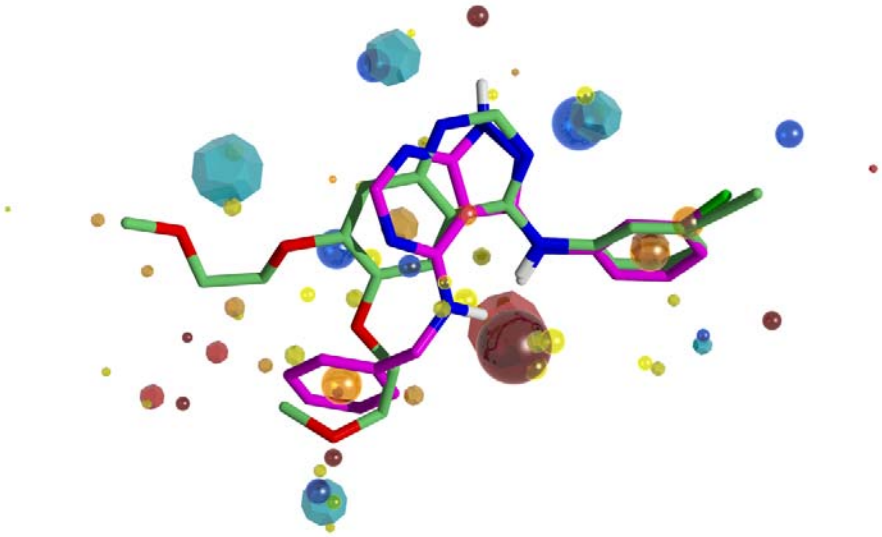

Figure 3. Scaffold hopping example for two egfr ligands – Same field, different structures.

3. Quantitative Field-Activity Relationships

FieldAlign can be used to retrieve the similarity values of your template across an active series of tens of molecules. In many cases, a quantitative correlation of the similarity value to activity can be plotted and used throughout a project to pre-check the likely activity of your next synthetic target.

4. Selecting analogues to synthesize

Once you have selected one or more series to follow-up, you can use FieldAlign™ to help explore and exploit field similarity and probe the importance of specific field points in a manual, qualitative manner.

a) Draw a variety of analogues in 2D with different substituents that are synthetically accessible or build a small library of up to 500 analogues
b) Copy and paste (or import an SDF file) into FieldAlign™ and calculate field similarity to your template
c) Identify any specific field points that you would like to investigate
d) Select a set of analogues that give specific variation of certain field points while retaining high overall similarity e.g. explore if a field point should be positive or negative, explore the importance of the magnitude of the field point etc.

The field template can be refined as activity data from new compounds is obtained.

Case Study – Lead Optimisation of CCK2 Antagonists

Molecular Field templates were used successfully by the James Black Foundation to design new chemotypes for cholecystokinin-2 (CCK2) receptor antagonists. CCK2 is a peptide-activated GPCR so the search for novel small molecule antagonists is hampered by the lack of detailed knowledge about the 3D structure of the receptor. As a result, Cresset's ligand- and Field-based approach was considered ideal.

An early lead, exemplified by **1** (pKi=8.8), contained an undesirable dibenzobicyclo[2.2.2]octane (BCO) core. To find replacement groups for this, a crude Field template was generated using a low energy conformation of **1** and new core structures were sought to mimic the field pattern around the BCO core. FieldAlign™ showed that the indole derivative, **2**, and naphthyl derivative, **3**, both fulfilled this requirement. They were subsequently synthesized and found to possess a similar pKi to **1** (pKi of **2** = 9.0; pKi of **3** = 7.9), so supporting the use of the Field template model.

The next step was to reduce the molecular weight and look at modifying the two amide-linked side chains which were thought to contribute to the poor half-life. A new field template was constructed using FieldTemplater™ from compounds **1**, **2** and **3** with their optimal sidechains. SAR showed the phenyl was not essential for activity, so could be removed. Novel analogues were then designed by the chemists and compared to the template using FieldAlign™. The template confirmed that the two amides could be replaced by a pyrrole, **4**, while still retaining the same field pattern, but the substructures are now linked in a totally different manner.

Unfortunately, **4** could not be synthesized, but close analogues were subsequently suggested and analysed using FieldAlign™. These included compounds in which the pyrrole was replaced by an imidazole and the indole was replaced by phenyl or phthalate. The compounds showed slightly lower

affinity for the CCK2 receptor than compounds **1** to **3**. However, the molecular weight was now significantly lower and subsequent small structural changes resulted in compounds with a better half-life. Compounds **5** (MW 467, pKi = 5.9) and **6** (MW 468, pKi = 6.1) were selected as lead development compounds.

Having successfully identified two development candidates, the James Black Foundation was further interested in using virtual screening to identify potential back-up series to **5** and **6** with completely different chemical structures. A new Field template model was built from compounds **2** and **7** (YF476, a competitor compound) using FieldTemplater™ methodology and a database of circa 600,000 commercially-available compounds was run through Cresset's Field-based virtual screening package FieldScreen™. The database was screened twice; first using the field of compound **2** and then that of **7** (both in their modeled bioactive conformation) resulting in two lists of compounds sorted by field similarity to their respective template. The top 500 compounds from each search were visually examined by the chemists and a smaller set was selected for purchase and testing. Compounds which appeared synthetically challenging were excluded. As novelty was important, compounds which were structurally very similar to known CCK2 compounds were also excluded.

Eighty-eight compounds were finally bought and tested. The results are summarized below:

- 1,000 compounds with high field similarity were 'eye-balled'
- 88 of the 1000 compounds were purchased and tested
- 27 had >20% binding at 10μM (30% hit rate)
- 4 had >20% binding at 1μM
- Full binding curves were measured for 12 compounds (n=2)
- Molecular weight range of actives 400-600

The structures of three of the active compounds (**8**, pKi 5.3; **9**, pKi 5.5 and **10**, pKi 5.1) found using molecule **7's** field hit-list are shown in Fig 4 and it is apparent that they are structurally distinct from **7,** so this approach provided the desired novelty. Other novel compounds found from this virtual screening project showed higher affinity and were considered more favorable hit structures, but cannot be disclosed here for commercial reasons.

This project demonstrates a practical example of how fields can be used to assist the medicinal chemists in their logical approach to designing and prioritizing novel scaffolds for series switching and also how a purely computational approach using virtual screening with fields can find diverse chemotypes for chemists to follow-up.

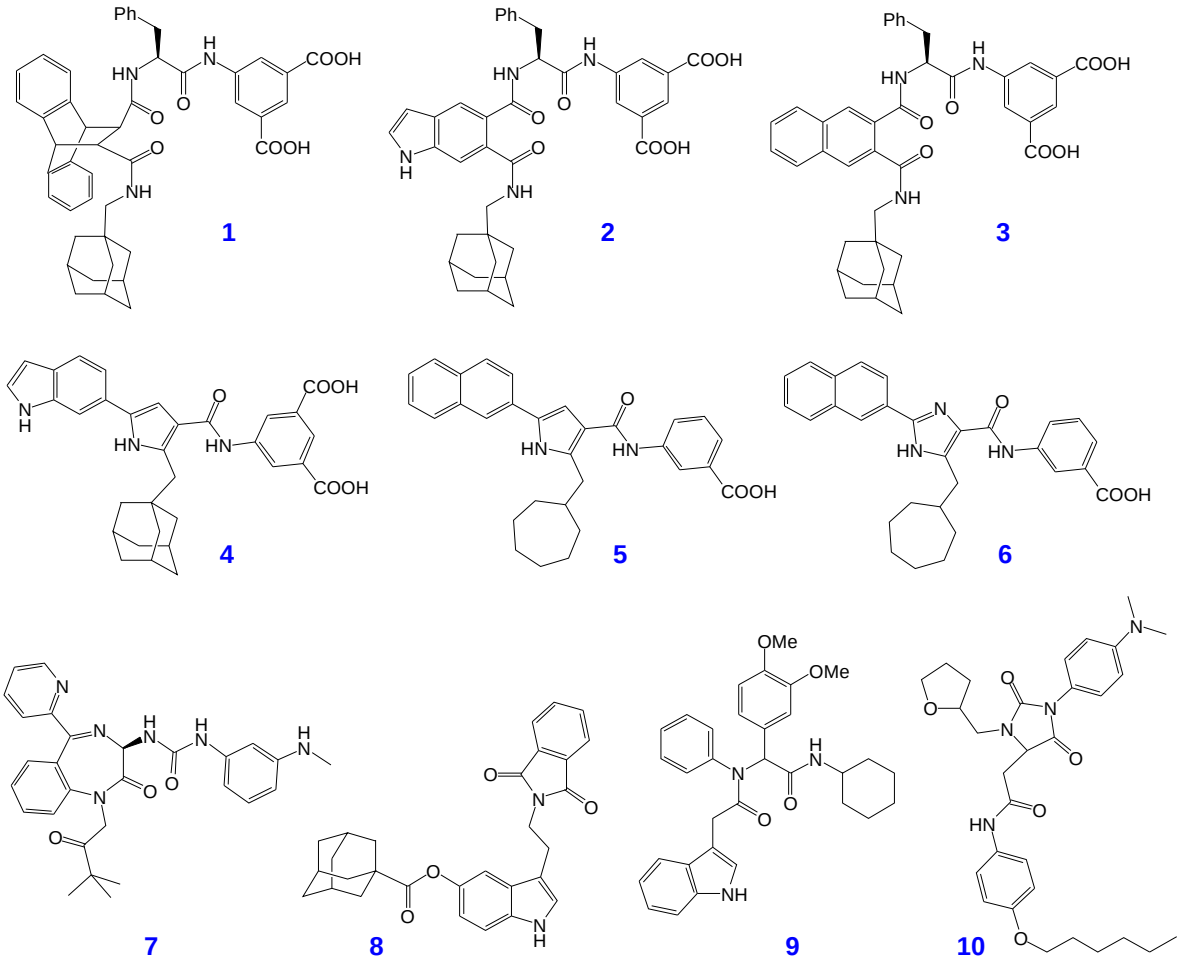

Figure 4. CCK2 Compounds 1 to 10

Further Information

FieldAlign™ and FieldTemplater™ are available from CambridgeSoft's SciStore
http://scistore.cambridgesoft.com/

More detailed information on Fields and products can be obtained from Cresset's website at
http://www.cresset-bmd.com