

Overcoming the limitations of chemical structure

The pharmaceutical business has been profoundly hampered by a ubiquitous and unexpected obstacle – the way it draws its chemical compounds. Scientists, patent agents and business decision makers from R&D, safety, and in-licensing are locked into a world of 2D ‘atom and bond’ and Markush representations of the molecular structure of compounds. Biological targets do not recognise a drug’s structural make-up. They respond instead to the properties around the drug that are generated by the structure. We can now generate and compare these property fields and show how they alone eliminate the largest impediments to drug discovery – the time and cost of evaluating activity and the serendipity involved in jumping from one chemical class to another for resolving patent issues, biological cross-reactions, ADMET and alternative clinical applications.

**By Dr Andy Vinter,
Dr Steve Gardner
and Dr Sally Rose**

In the pharmaceutical world, scientists routinely use 2D structural likeness to indicate similar biological action. 2D ‘atom and bond’ representation such as those shown in **Figure 1** are, after all, the *de facto* language of chemistry. So why question the authority of this representation after applying it for so long on all aspects of drug discovery from activity enhancement to toxicity and from patent application to safety issues?

The reason is simple – it has a serious limitation that inhibits our ability to discover, protect and bring to market new drugs. That limitation is familiar to most chemists. We have known for many years that very different structures can act at the same biological target to effect the same biological action. The obvious conclusion must be that chemical structure – atoms linked by bonds –

is not what is recognised by a biological target.

Molecule A and molecule B in **Figure 1** bear no structural relationship to one other and yet molecule B mimics molecule A – both are able to occupy the same site on a common target (see **Figure 4**). The corollary is also often true; very closely related derivative structures regularly exhibit very diverse activities at a given target.

This inability of 2D structure to describe a compound’s activity has profound consequences. Chemists are routinely required to make ‘chemotype jumps’ to move from one chemical class, all members of which have one or more common structural features, to another. Such chemotype jumps are vital if structures are to be novel and patentable, if pre-clinical ADME (absorption, distribution, metabolism, and excretion), safety and

toxicity problems are to be overcome and FDA requirements satisfied. It often takes chemists up to three years to make a single chemotype jump and the process is essentially serendipitous as the structure offers no clues as to the likely activity of the compounds. This obviously requires a huge investment in cost and time and there is always a risk that no such alternate chemotype with appropriate activity will be found.

Many chemotype jumps are required in the long haul from early hit to marketable drug. By the late 1980s, so costly and time-consuming had this process become that companies embraced the ideas of high throughput chemistry and screening in the hope that applying these parallel processes to tens of thousands of molecules, they would alleviate the 'chemotype trap'. In the event, it not only reduced the chances of chemical novelty but left them to drown in a sea of irresolvable data.

The alternative to chemical structure

Before discussing how molecules recognise each other, we must remind ourselves that molecules are three-dimensional and exist in a specific shape (or conformation). To make matters more complex, most molecules change their shape and volume in nanoseconds. In this context the impression of a static regular structure given by the normal 2D-representations used by patent lawyers, in-licensing executives and other non-chemical sectors is positively misleading. In reality, the shape and volume changes provide the protein target with an exquisite means of selectivity, allowing it to pick out only one specific complementary shape when it combines with a drug or hormone.

The most important question then is – what does a protein recognise about its natural hormone or a targeted drug if not its 2D structure? To answer this question, imagine a 3D chemical structure to be like a human skeleton. Rough size, sex and mobility can be judged from a person's bones but there is nothing to indicate either the visual characteristics or indeed the personality of the living individual. An experienced chemist can get a 'feel' for how a 2D structure may look and move in 3D just as a forensic scientist can build a speculative model of the physical appearance of a human face from a skull, but this basic 'feel' is not enough to predict in detail either the properties of a molecule nor the personality of the human. To do this we need to understand how a molecule will interact with its target (and potentially other molecules) through its behaviour. While human behaviour and personality arises from many genetic imprints subsequently coloured by environmental influences, we can be

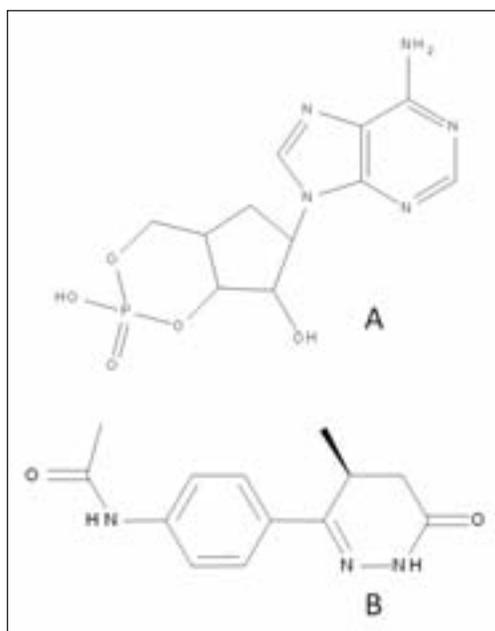


Figure 1
These two structurally diverse molecules act at the same biological site

thankful that, by changing its shape and surroundings, a molecule can exhibit a wide range of behaviour using just four basic 'genetic' components.

The first important trait is that all molecules and atoms are attracted to all others. If some surface contours of a protein complement those of a drug, the two molecules will stick together along that common contour. This is the root of the old 'lock and key' idea of drug/protein interaction. It is claimed that the convolutions on a gecko's feet present such a large surface area to a wall or window, that nothing other than this tendency for atoms to stick to each other are needed to stop him falling to the ground. However in reality, the matching surfaces have to be quite extensive for strong binding and hence high drug affinity to occur. Large antigen-antibody interactions are often driven by this surface 'stickiness'.

The next pair of properties that act to define how a molecule behaves, take us back to school physics where we learnt that like electrical and magnetic poles repel while unlike poles attract. Electrons are negatively charged and also define the outer 'skin' of a molecule's surface. Depending on its atomic make-up, electron poor and electron rich regions appear at its surface, stretching out into space with ever decreasing influence, sometimes over distances many times the molecule's own width (see **Box 1** for more detail). If this 'electrostatic field' is perceived by another molecule with regions of the opposite polarity, the two molecules will come together. These electrostatic attractions are the strongest of all inter-molecular

Molecular modelling

Box I: Molecular field technology

Molecular fields have been applied in computational chemistry for more than two decades to analyse the activity of small molecules¹⁻³. However, the problems of aligning diverse structures, the use of coarse grids around the molecule to define the field and poorly-validated force fields have been paramount in limiting the general utility of the methods. The early field methodology did not live up to its promises because of approximations buried in the molecular mechanics used to generate fields. Vinter⁴⁻⁶ has addressed these basic flaws.

The molecular field description of a compound comprises four fields: positive and negative electrostatic, van der Waals attractive and hydrophobic^{7,8}. The fields are derived from calculating the interaction energy of a 'probe' atom with the target molecule. Potential fields around a molecule are continuous functions which make intermolecular comparisons difficult. To overcome this computational intractability, we use molecular field extrema⁷ as a means of finding the most biologically relevant regions of a molecule, where interactions with a protein are most likely to occur. These extrema are visualised as coloured spheres ('field points') and their size reflects their magnitude. There is approximately the same number of field points as heavy atoms in the average 'drug-like' compound. **Figure 3** outlines the procedure for converting a structure to a field pattern.

Field points provide a convenient way of visualising and sampling the key information in the field and facilitating rapid molecular similarity comparisons. As fields and field points are conformation dependent, comparisons of molecular similarity are performed over a range of conformations (commonly up to 100) for both molecules. The highest similarity value obtained gives the best conformational overlay in terms of field similarity and this value is taken as the molecular similarity score.

If fields are added to the 3D conformers of molecules A and B in **Figure 1**, the field similarity of the two can clearly be seen (**Figure 4a and b**). To determine the similarity of the 2 molecules, A and B, the field points of A are used as sampling points into the field potentials of B. This score is asymmetric, so the calculation is repeated for the field points of B sampling into the actual field of A and the two values are then averaged and normalised to give the final optimum similarity score (**Figure 4c**).

Because the energies are analytically recalculated, the entire 'true' field is used in the calculation which overcomes problems associated with the relative widths and exact locations of potential energy wells. Sampling at the location of field points ensures that comparisons are biologically-relevant.

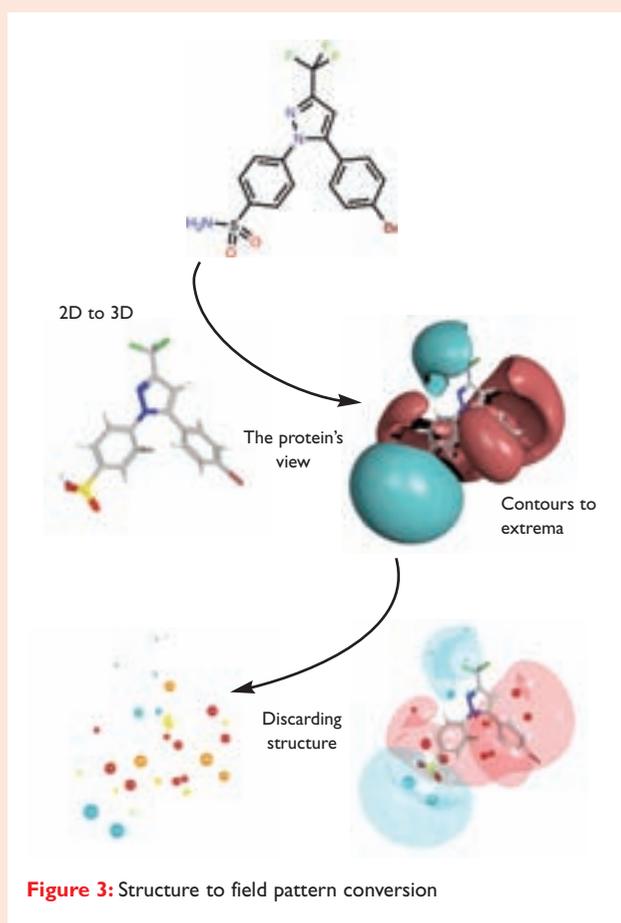


Figure 3: Structure to field pattern conversion

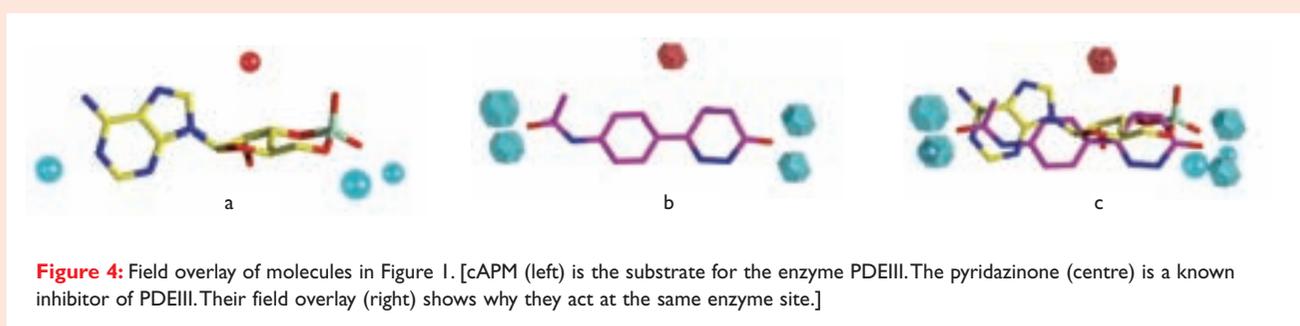


Figure 4: Field overlay of molecules in Figure 1. [cAPM (left) is the substrate for the enzyme PDEIII. The pyridazinone (centre) is a known inhibitor of PDEIII. Their field overlay (right) shows why they act at the same enzyme site.]

interactions and manifest themselves most strongly in hydrogen bonds and ion pair bridges. Conversely, like poles are equally vigorously repelled. The two electrostatic actions of repulsion and attraction make up two of our four recognition forces. These electrostatic fields are, most importantly, dynamic and change drastically as the drug changes shape. We can now begin to see why a protein only wants to entertain the one drug shape that complements its own fixed field.

The fourth influence is akin to the human tendency to 'belong'. It is called the hydrophobic effect and is complex to the extent that it still eludes precise definition. Simply speaking, molecules of an oily nature (eg hydrocarbons) that contain atoms with no superfluous electrons will club together. Polar molecules (eg water) with spare electrons (usually in lone pairs) will form another club. The two clubs may mix but only under duress. Measurements of properties such as logP are used to find out which club a molecule prefers to join.

In summary, the four properties of surface 'stickiness', electronic attraction, electronic repulsion and hydrophobicity can be calculated for a drug molecule in a given shape. The result is a complex skin over the skeleton of the molecule showing the most prevalent properties in each area. Unfortunately working with such complex patterns requires such enormous computer power that it is effectively impossible. In order to make the problem tractable, we therefore use a similar approach to facial recognition, by reducing the complex skin shape and property patterns down to a series of points that represent the size and character of the most extreme parts of the fields. These resulting 'field patterns' (Figure 2) represent all that is needed to define how that molecule will be perceived and bind with another molecule, be it a protein active site, a piece of DNA or any other molecule at all.

The application of fields

The field pattern reflects how the outside world perceives and interacts with the molecule and although the field pattern is dependent on the structure, it is not possible to reverse engineer the field pattern to definitively regain the original atom arrangement. Attempts to do so will result in a list of diverse structures all seemingly satisfying the starting field pattern. Far from being a problem, this is precisely the point of using fields over structure. Across all of chemical space, there must be many diverse structures that can generate a similar field pattern. It is worth noting that the larger the structure becomes, the more specific its field pat-

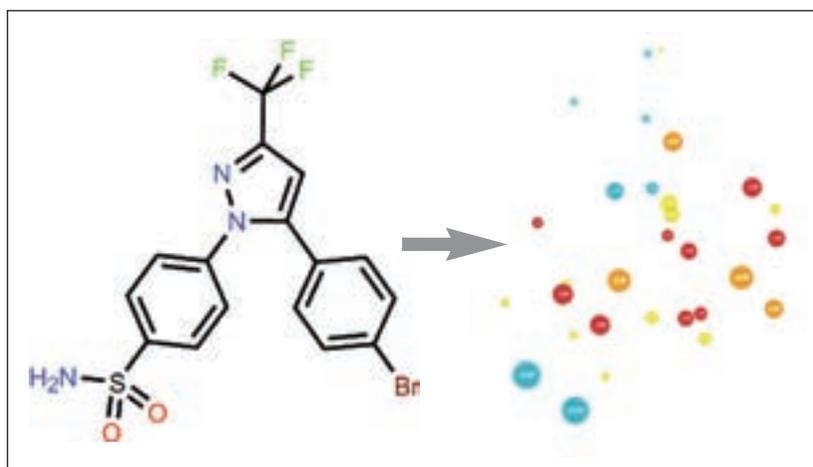


Figure 2
The field equivalent (right) of the structure (left)

tern to that structure. This may be one reason for nature to adopt large molecules for specific purposes such as locking the active sites of biological targets into large proteins.

The axiom therefore follows that; if two molecules from different classes can generate the same field pattern, they will be active at the same biological site. It is also certain that a structural 'metoo' that has been derived from an existing biologically active drug is not guaranteed to be active unless its field pattern closely matches the original. Examples of this phenomenon abound. Grafting a polar hydroxyl group on to benzene drastically changes its field pattern and its consequent behaviour, but a methyl group has little effect. This is well known chemistry; just compare the properties of benzene with phenol and toluene!

This ability to compare the fields generated by two or more molecules and hence determine the similarity of their biological activity and properties has many useful applications in the real world of pharmaceutical discovery and development.

Hit-finding (virtual field screening)

The field pattern of a known active compound can be used as a seed to screen an in-house or external database⁹, itself made up of field patterns rather than structures. We are looking for field pattern matches which will indicate similar biological activity to the seed. The returned structures will include examples across all chemotypes within the database. If the database contains commercially available molecules, they are rarely drug-like, usually small and are regarded as diverse hits suitable for further medicinal chemistry to progress to leads. If a more drug focused resource such as the World Drug Index is converted to a field database and searched with the seed, drug-like molecules are

Molecular modelling

References

1 Crucciani, G (Ed). Molecular Interaction Fields. vol 27 of Methods and Principles in Medicinal Chemistry, Wiley-VCH 2006.

2 Cramer III, RD, Patterson, DE and Bunce, JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110: 5959-5967, 1988.

3 Goodford, PJ. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J Med Chem* 28: 849-857, 1985.

returned. This is an excellent resource for repositioning drugs, suggesting other biological actions for the seed compound, checking contra-indications and taking advantage of pre-clinical data already available from the screened hits. Using virtual field screening to preselect a smaller number of compounds with a higher probability of showing activity prior to biological testing significantly reduces assay costs in terms of reagents, compound and post-assay analysis. GSK¹⁰ has shown that lower throughput assays are much less error-prone than HTC/HTS, resulting in fewer false negatives and positives, which greatly facilitates the selection of new chemotypes to follow-up.

Hit-to-lead

Hit-to-lead and lead optimisation projects invariably proceed faster and more efficiently if there is a robust model to help guide synthesis and promote creativity. This may be an X-ray of the protein active site, a robust QSAR model, a traditional pharmacophore or a molecular field pattern.

The medicinal and computational chemist can then design compounds to explore and test the limits of the model and improve its power of prediction.

A molecular field model gives a pattern for binding. It is created by overlaying three or more known active compounds to find the common field pattern exclusively recognised by the target active site. Alternatively, the model can be created directly from the field of a ligand extracted from a protein-ligand co-crystal. New compounds can be proposed that explore the importance of certain field points and help define shape constraints ('excluded and allowed volumes' in the field model). Combinatorial arrays can be designed by the chemist and similarity values for each compound to the model calculated. Compounds with high similarity to the model are more likely to show activity, allowing the model to direct the selection of interesting building blocks, increasing the probability of synthesising active compounds. This reduces the time and cost to reach the patent filing stage in a drug discovery project.

Continued on page 63

Box 2: Bioisostere replacement

The screenshot shows a set of bioisosteric replacements for the lactone moiety of the COX-2 compound Rofecoxib. They include known actives such as Valdecoxib and Parecoxib and more interesting analogues such as the thiazolotriazole compound¹¹. The field patterns are clearly highly similar.



Lead optimisation and ADME/toxicity

The field model can be further refined during lead optimisation. Importantly, ADME properties need to be considered during this stage of discovery. ADME and toxicity field models can be constructed for important proteins. A specific hERG field pattern has already been successfully used to filter unwanted chemotypes and guide further synthesis.

A key role for fields at this stage is in the search for bioisosteres. Frequently, certain groups are found which cause ADME or toxicity problems, but which appear essential for activity. They are often the cause for serious concern and can result in the termination of a project. Field patterns have been used to substitute the offending groups by finding bioisosteres that retain the same field pattern across the whole molecule while overcoming a specific ADME/toxicity issue, offering a potential solution to this impasse (Box 2).

The identification of a follow-up series and IP

Bioisosterism can be used to find new scaffolds which have the same field properties (and therefore activity) as the original lead series, yet have different chemotypes offering the possibility for a new patent. Field-based bioisosterism is a fast method to find completely new chemotypes with the desired activity, but potentially different ADME/toxicity profiles to the lead series facilitating the extension of a product's exclusivity and revenue stream.

Because a field pattern defines the key binding

features required for activity at a given target, it provides additional, complementary information to the Markush chemical structure, and can be used to evaluate and/or increase the protection afforded by a composition of matter (CoM) patent. A field model developed in-house can be used to explore the patent space of competitor compounds very rapidly and identify new bioisosteres of the known core scaffolds for synthesis and evaluation in-house. This offers the possibility of fast follow-up to a competitor's new blockbuster drug (Box 2). In a similar vein, a patent that described not only the 2D structure and its synthesis but also the key activity features in terms both of fields and a number of Markush structures covering all active chemotypes could provide a pharma company with greater IP protection against trivial structural changes ('patent busting').

Field pattern-based protection may be possible in the future. Indeed, a small biotechnology company patented a field pattern as a pharmacophore as long ago as 1994. Although this has since been dropped, the concept is certainly interesting and one that could increase the level of protection afforded by a CoM patent.

In-licensing and portfolio management

Using a molecular field model of known active compounds allows the in-licensing manager to assess more rapidly the value of a novel compound presented by an external company for out-sourcing or co-development. If it can present the same field to a protein as known actives then it is likely to bind at the same site. Alternatively, if the field is dissimilar to known actives then it is most likely to be working by a different mechanism and possibly of greater interest. Field analysis thus facilitates a deeper understanding of the types of compounds in a company's portfolio of drugs in development and their novelty compared to any presented compound or drugs under development in other companies. It provides unique, biologically-relevant information to assist the decision-making process.

Conclusions

Molecular fields offer a significant advance on chemical structure as a means of looking at the biologically relevant properties of drugs and hormones. The technology is valuably used by different departments throughout a pharma company to improve efficiency, effectiveness and innovation. R&D, IP and business development functions can all benefit from understanding and correctly applying the technology described above.

Limiting the chemistry world to 2D is passé and

constraining, yet it continues to be the *de facto* representation used throughout the industry. There are however significant competitive advantages throughout the pharma business to adopting a smarter way to compare the biological activity and properties of our molecules. **DDW**

Dr J G Vinter (known as Andy) is Chief Scientific Officer of Cresset BioMolecular Discovery, the company he founded in 2001. From 1964 Andy worked in the pharmaceutical industry as an organic and computational chemist and in 1990 moved into academic and consulting activities at the University of Cambridge, expanding on the software that resulted in the formation of Cresset. Andy is recognised as a pioneer in the application of molecular fields for drug discovery.

Dr Steve Gardner is Chief Operating Officer of Cresset. Steve has built a number of innovative informatics technologies, products and companies in life sciences and healthcare. Steve was previously CTO of BioWisdom and Astra's Director of Research Informatics. He consults and publishes widely and has 20 semantic data integration patents.

Dr Sally Rose was 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 and recently retired to live in France.

Continued from page 62

4 Vinter, JG. Extended electron distributions applied to the molecular mechanics of intermolecular interactions. *J Comp-Aid Mol Design* 8: 653-668, 1994.

5 Vinter, JG. Extended electron distributions applied to the molecular mechanics of intermolecular interactions. II-Organic Complexes. *J Comp-Aid Mol Design* 10: 417-426, 1996.

6 Chessari, G, Hunter, CA, Low, CMR, Packer, MJ, Vinter, JG, and Zonta, C. An Evaluation of Force Field Treatments of Aromatic Interactions. *Chem. Eur. J.*, 8: 2860-2867, 2002.

7 Cheeseright, T, Mackey, M, Rose, S and Vinter, A. Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *J. Chem. Inf. Model.* 46: 665-676, 2006.

8 Cheeseright, T, Mackey, M, Rose, S and Vinter, A. Molecular Field technology applied to virtual screening and finding the bioactive conformation. *Exp Opin Drug Disc.* 2: 131-144, 2007.

9 Cheeseright, T, Mackey, MD, Melville, JL and Vinter, JG. FieldScreen: Virtual screening using molecular fields. Application to the DUD dataset. *FieldScreen: Virtual Screening Using Molecular Fields. Application to the DUD Data Set. J. Chem. Inf. Model.* 48, No 11, 2108-2117, 2008.

10 Bellenie, BR, Barton, NP, Emmons, AJ, Heer, JP and Salvagno, C. Discovery and optimization of highly ligand-efficient oxytocin receptor antagonists using structure-based drug design. *Bioorg Med Chem Lett* 19: 990-994, 2009.

11 Cheeseright, T. The identification of bioisosteres as drug development candidates. *Innov Pharm Tech* April, 22-26, 2009.