

Peptides to non-peptides: leads from structureless virtual screening

Tim Cheeseright, Mark Mackey and Andy Vinter*, Cresset BioMolecular Discovery, The Spirella Building, Letchworth, UK SG6 4ET;

*e-mail: a.vinter@cresset-bmd.com

Cresset BioMolecular Discovery (BMD) was founded in 2001 as an *in silico* drug design company providing new leads for drug discovery programs. Cresset-BMD has developed a unique database of commercially available molecules described by their surface fields rather than their structural topology. For any known biologically active molecule a list of compounds with potentially the same activity and unprecedented structural diversity can be generated; this forms an ideal starting point for medicinal chemistry programs or can be a fast and economic fall-back reference in the event of development problems. The flexibility and power of this searching method is apparent from its ability to locate drug-like leads from traditionally difficult hits such as peptides, steroids and highly charged molecules.

This database is built on structure independent virtual screening technology that describes a molecule as a protein 'sees' it rather than by its atoms and bonds (Box 1). Most current computational methods describe molecular similarity in terms of structure- or shape-based models. However, it is known that two structurally different molecules can bind to the same region of a protein and, as such, they must have something in common. The common factor is the outer 'skin' of the molecules rather than what is shown by the more traditional structure-based views of medicinal chemistry.

Scientific basis

The idea of the 'common skin' of a molecule being the essential clue to its biological activity has been around for 20 years or more. Many workers have tried to define the

form and constituents of this skin but have failed to correlate the results to anything consistently useful. It transpires that the main source of error is the model of electrostatic charge within the molecule. Correct charge distribution is crucial to the *in silico* definition of these skins, and had to be researched and developed correctly before the skins could be properly generated. The research has resulted in a new molecular mechanics force field that uses extended electron distribution (XED) to redefine the electrostatic description of molecules [1,2]. XEDs reproduce experimental observations of structural shape and conformational energy difference better than any previous force field [3] and are able to create the accuracy necessary for generating molecular electrostatic potential (MEP) fields, which are used in the representation of these molecular skins. The difficulty of working with MEP fields has been solved by distilling them down to their potential extrema [4,5,6], which we call 'field points' (Box 1). Four different fields – positive, negative, shape and hydrophobic – are combined in a 'field pattern' that represents the ligand from the viewpoint of the protein [7].

The advantage of a field description over traditional pharmacophore methods is that field patterns are calculated directly from the structure of a conformation. They take into account all of the steric and electrostatic properties of the conformation in an analogue manner: the field points near a hydrogen-bonding group will vary depending on the steric restrictions around the group and on what electron-donating and electron-withdrawing substituents are attached to that part of the molecule. As a result, much

more information is encoded in the field pattern than in a simple set of 'hydrogen-bond donor', 'hydrogen-bond acceptor', 'aromatic' and 'hydrophobic' pharmacophore points. Moreover, owing to the accuracy of the XED force field, the potential for less common interactions such as C–H hydrogen bonding, aromatic–aromatic stacking and cation– π interactions [3,8,9] is fully represented in the field pattern.

It needs only the shortest leap of imagination to visualize a database of chemicals, each with its own set of accessible conformations stored as field-point fingerprints or FieldPrints™ rather than structures. A 'template' consisting of the FieldPrint™ of an active compound could be used to scan the database for similar FieldPrints™. Hits would not be selected on the basis of direct structural similarity, so the search would return a set of compounds broad enough to provide new and truly diverse structural leads. Cresset-BMD was formed to develop this technology and has already produced independent validation of the approach.

Applications

FieldPrint™ techniques can generate significant value for pharmaceutical companies who are looking to jump from a stalled lead to a structurally unrelated molecule while retaining biological activity. Cresset-BMD technology could be used to circumvent many common problems that arise in drug development programs, including:

- molecular toxicity, clearance or overall bioavailability;
- freedom to operate within a competitive patent area;

Box I. The principles behind structure-independent virtual screening

Proteins respond to the electron cloud around a molecule rather than to the 3D arrangement of its individual atoms. The electron cloud is created by the atomic charges acting in concert, and can be represented by a 3D molecular electrostatic potential (MEP) map around the molecule. For analytical ease, the electron clouds can be distilled to 3D 'field points' around the molecule, the sizes of which depict their relative strengths (Figure I).

Although MEPs have been used for many years, they depend on the quality of the atom charge model used. Electrical charge distribution on an atom is conventionally described using atom-centred charges (ACCs) *in silico*. This description is too inaccurate to reproduce MEPs and many observed phenomena such as aromatic interactions and hydrogen bonding. This technology overcomes the problem by distributing charges in a more realistic way. The new description is called extended electron distribution (XED). Using ACCs, the MEP is incorrectly expressed; XEDs rectify the shortcomings of ACCs and lead to a significant increase in the accuracy of the MEP maps (Figure II).

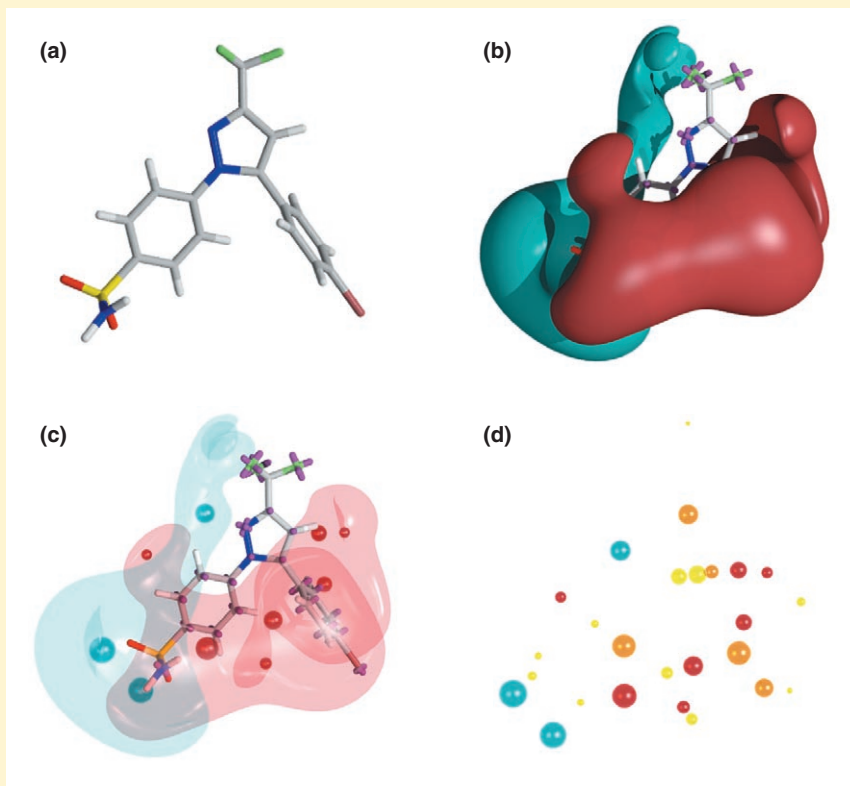


Figure I. Derivation of field points. (a) Traditional 3D view used by modellers. (b) Molecular electrostatic potential (MEP) view from a protein – simplified to show the electron-rich areas (blue) and the electron-deficient areas (red). (c) The complex electron cloud is distilled into field point extrema that describe the binding properties of the molecule. (d) Surface (yellow) and hydrophobic (orange) properties are added as field points.

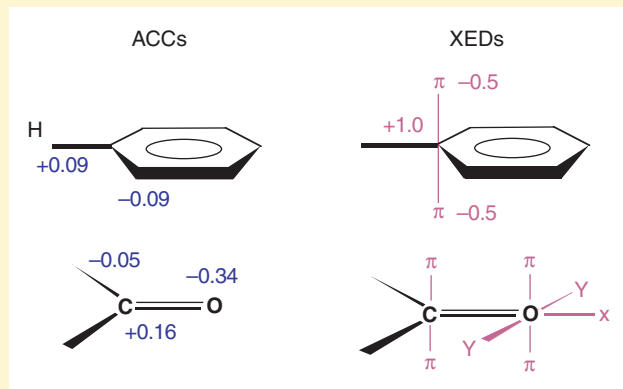


Figure II. The extended electron distribution (XED) description. Conventional modelling sets atom charges at the nucleus [atom centred charges (ACCs)]. The XED description enables the atomic nuclear charge to remain positive by extending its associated electrons out onto stalks (labelled X, Y or π).

- target intractability with standard screening technologies;
- switching from a peptide to a non-peptide small molecule lead.

All that is required for a FieldPrint™ search is a known active molecule and enough auxiliary information to be able to approximate the binding conformation of that molecule. In practice, two structurally diverse active molecules from the literature usually suffice (note that protein information is not required). Therefore, the

Cresset-BMD *in silico* database-searching algorithms can be used to rapidly kick-start a lead-finding program providing the following benefits:

- rapid and diverse hit-finding provides a better starting point for drug discovery;
- the ability to move rapidly from a natural ligand to active small molecule hits;
- increased speed of lead optimization, ensuring rapid entry into a full medicinal chemistry program and leads to associated cost savings;

- identification of novel, structurally distinct back-up series.

Use of the Cresset-BMD technology will help to lower the lead compound attrition rate in drug development as diverse structural classes can be used from the start. It also reduces the number of compounds required for development of candidates to drugs using medicinal chemistry and reduces the costs and delays of high throughput screening (Figure I). When working with an intractable target,

a drug company might take between 1–3 years to generate a lead suitable for further medicinal chemistry. Potentially, this new technology is able to reduce this period to 2–5 months with unprecedented cost savings. For example, by searching our database with the peptidic thrombin inhibitor D-Phe-Pro-Arg-chloromethyl ketone (PPACK), a non-peptidic spike that took Merck (<http://www.merck.com>) over three years to develop using at least ten full-time medicinal chemists (>30 man years) was retrieved in a few hours.

Competition

There are many potential competitors but none, to date, are able to go beyond the use of structural comparisons with any success. Comparative molecular field analysis (CoMFA) from Tripos (<http://www.tripos.com>) uses field analysis akin to the Cresset-BMD approach, but suffers from the electron-distribution approximation outlined previously. *In silico* design companies who offer contract services or sell software directly, and structural genomics companies who identify leads using X-ray analysis and large-scale computer docking, use only standard science and methodology in proprietary packages. Cresset-BMD uses new science that releases such structural constraints. Therefore, at present, there is no direct competitor.

Validation

New methods have to be proven, especially if they are offered commercially. The virtual screening method was tested by constructing a database of 600 000 commercially available compounds (Figure 1). When searched for inhibitors of thrombin with, for example, the PPACK FieldPrint™ (Figure 2), the database found 39 times more spikes in the top 1% of the database than would be expected by chance (Figure 1). One major advantage of using fields rather than structural fingerprints or pharmacophores is that searching the database using a peptide as the active search molecule produces a better retrieval of the spikes over chance even though the spike

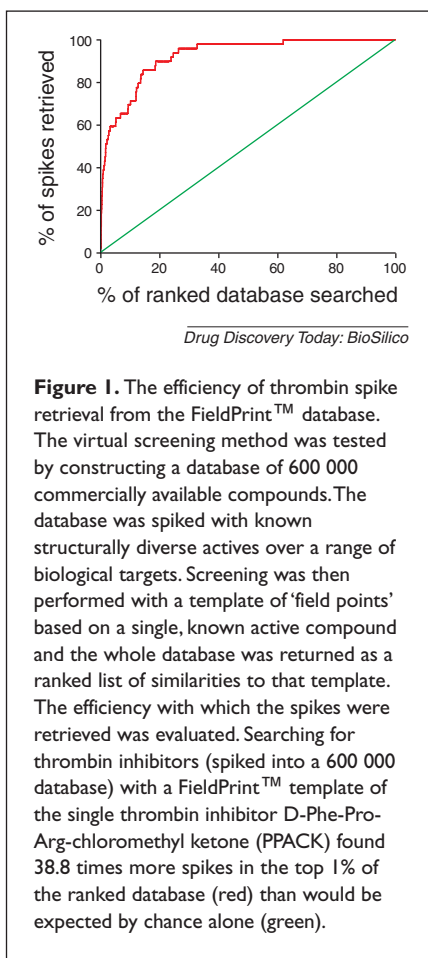


Figure 1. The efficiency of thrombin spike retrieval from the FieldPrint™ database. The virtual screening method was tested by constructing a database of 600 000 commercially available compounds. The database was spiked with known structurally diverse actives over a range of biological targets. Screening was then performed with a template of 'field points' based on a single, known active compound and the whole database was returned as a ranked list of similarities to that template. The efficiency with which the spikes were retrieved was evaluated. Searching for thrombin inhibitors (spiked into a 600 000 database) with a FieldPrint™ template of the single thrombin inhibitor D-Phe-Pro-Arg-chloromethyl ketone (PPACK) found 38.8 times more spikes in the top 1% of the ranked database (red) than would be expected by chance alone (green).

molecules were predominantly non-peptidic in nature. The FieldPrints™ encode the surface properties of a molecule, not its chemical structure. As a result, this method should be equally applicable to peptide and non-peptide agonists or substrates, and provides a truly novel way to progress from peptidic leads to drug-like molecules. For those problems that cause drug developers the most grief – patent cover, poor ADME (absorption, distribution, metabolism, excretion) and toxicology, intractable chemistry or just frustration at the lack of positive progress – the FieldPrint™ virtual screening technology offers a simple and effective solution.

The final test of any method is to apply it to a real situation. The FieldPrint™ virtual screening technique was validated in collaboration with the James Black Foundation (JBF), sponsored by Johnson and Johnson (<http://www.jnj.com>). The binding

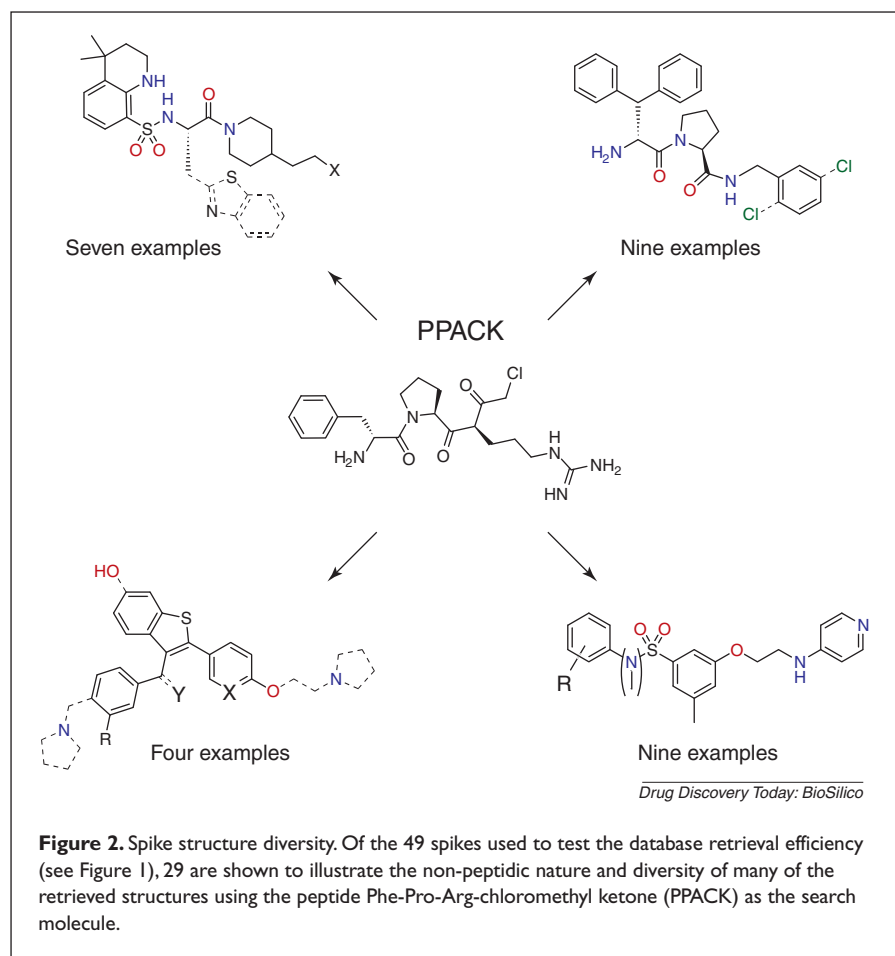
conformations for two of their gastrin peptide guanine nucleotide (G-protein)-coupled receptor (GPCR) antagonists were determined by field-point-based overlays and subsequently used to search the database. A list of 1 000 potential hits was gathered from the top of the similarity results and the most promising 88 were selected and purchased from commercial vendors. Using their in-house binding assay, JBF found 27 (30%) active at the pre-agreed threshold (10 μM) and four with activities >1 μM. All but two of the hits had no structural similarity to any known class of actives at the gastrin GPCR. These results vindicated the Cresset-BMD approach, and further projects that are expected to reinforce the conclusions from JBF are in progress.

Present limitations

Currently, Cresset-BMD is limited to using commercial compound databases that contain simple, easily made molecules. If searching with a particularly demanding template (e.g. highly charged, oddly shaped or conformationally extensive), as most drugs tend to be, there is a reasonable chance that the database will contain no compounds with field patterns that match that of the search molecule. Under these circumstances the search would obviously fail.

Although we have every scientific confidence in our electrostatic descriptors, the surface and hydrophobic descriptors might not be ideal because their basis on experimental observations means they are less rigorous. Ongoing research will improve the quality of surface and hydrophobic descriptors but we have no evidence, as yet, to show that they are inadequate for our present purposes.

Finally, a search is actually run in two stages. The first is a fast approximate method designed to cope with 100 million conformational entries for 2 million compounds in a reasonable computational timeframe; as a result, this preliminary protocol loses information such as chirality. The second is a refinement stage, which is accurate but takes considerable time; on average, 10% of the database is refined in



four days with a cluster 25 Linux processors. Ultimately, the approximate method will be dispensed with as our cluster grows and computing speed increases.

Conclusions

FieldPrints™ and field-based overlays have shown themselves to be a robust method

for locating new leads for medicinal chemistry programs. Unlike docking protocols, no X-ray protein information is required and only 2D structures are needed for two or more ligands known to act at the same site. Thus, the technique is applicable to many classes of target for which no acceptable lead-finding solution currently

exists because the technique is based on the surface properties of molecules. Lead hopping is now possible from peptide to non-peptide, toxic to non-toxic, patented to non-patented, or any time a new organic species is needed.

References

- Vinter, J.G. (1994) Extended electron distributions applied to the molecular mechanics of intermolecular interactions. *J. Comput. Aided Mol. Des.* 8, 653–668
- Vinter, J.G. (1996) Extended electron distributions applied to the molecular mechanics of intermolecular interactions. II-organic complexes. *J. Comput. Aided Mol. Des.* 10, 417–426
- Chessari, G. *et al.* (2002) An evaluation of force field treatments of aromatic interactions. *Chemistry* 8, 2860–2867
- Davis, A. *et al.* (1987) Extended electron distributions applied to the molecular mechanics of intermolecular interactions. II-Organic Complexes. *J. Comput. Aided Mol. Des.* 1, 97–120
- Vinter, J.G. and Saunders, M.R. (1991) In *Host-Guest Molecular Interactions: from Chemistry to Biology*, Ciba Foundation Symposium No 158. pp 249–261, Wiley
- Apaya, R.P. *et al.* (1995) The matching of electrostatic extrema: a useful method in drug design? A study of phosphodiesterase III inhibitors. *J. Comput. Aided Mol. Des.* 9, 33–43
- Vinter, J.G. and Trollope, K.I. (1995) Multi-conformational composite molecular fields in the analysis of drug design. (1) Methodology and first evaluation using 5HT and histamine action as examples. *J. Comput. Aided Mol. Des.* 9, 297–307
- Hunter, C.A. *et al.* (2002) Substituent effects on cation-p interactions: A quantitative study. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4873–4876
- Lozman, O.R. *et al.* (2001) Complementary polytopic interactions (CPI) as revealed by molecular modelling using the XED force field. *J. Chem. Soc. Perkin Trans.* 29, 1446–1452

Bio-Tools articles

Drug Discovery Today: BIOSILICO Bio-Tools articles provide an insight into the latest product developments in the informatics arena. Articles briefly describe the basic principles of a novel or developing technique and then discuss the current and potential applications of this technique, placing this new/emerging technology in the context of currently available tools. If you would like to contribute to this section, please submit your proposals to:

Dr Christopher Watson, e-mail: BIOSILICO@drugdiscoverytoday.com