The Virtual Elaboration of Fragment Ideas: Growing, Merging and Linking Fragments with Realistic Chemistry

PEER REVIEWED

Fragment-based drug discovery seeks to identify relatively simple and low molecular weight molecules which interact with protein targets, then grow or link them into active leads. Molecular design software can be used to create realistic chemistry between and around fragments. Two key factors for success are: using computational techniques that are suited to handling small molecule fragments; and prioritising the results by evaluating each new potential molecule as a whole.

Fragment-based drug design (FBDD), also referred to as fragment-based lead discovery (FBLD), has emerged as a successful new method in drug discovery for both lead identification and optimisation. Rather than starting out with screening compounds, which are typically drug-sized molecules, FBDD focuses on smaller molecules, referred to as fragments, that show some affinity to a protein target, then seeks to grow them into more drug-like and, therefore, larger active leads. One key advantage of FBDD is that it typically results in a lower molecular weight lead, which is likely to have a much better PK profile, including, for instance, higher oral bioavailability [1].

FBDD encompasses both experimental and computational methods. Fragments are relatively simple molecules with few chemical features. Their typical small size and low molecular weight mean that the affinity shown when binding to the target can be low in absolute terms. Hence expensive and sensitive biophysical techniques such as NMR, X-ray crystallography or surface plasmon resonance (SPR) are needed to detect the fragment interactions[2]. Experimental methods are, therefore, typically low throughput and high cost.

Computational drug design techniques complement experimental methods and help researchers focus expensive lab resources on the areas most likely to succeed. Early in the discovery process, computational techniques are used to pre-screen fragments in order to focus on those most likely to bind. This is typically done by screening a database of fragments using a docking method. Computational methods are also useful once initial hits have been detected for growing and merging fragments. Example uses are for exploring additional features within a protein active site and for suggesting suitable linker chemistry between two fragments that have been found to bind in different positions within the active site.

However, it is important that the computational chemistry software is suitable for use with fragments. Most drug design software has been designed to deal with drug-like molecules that have a molecular weight between 350 and 500 Daltons. Fragments are typically of the order of 200 to 350 Daltons. Therefore, algorithms that have been proven to work with larger molecules are not always successful when applied to fragments.

An example of this is where a docking method utilises a scoring function to evaluate the binding energy for a given pose of a molecule. The algorithms typically use a functional group contribution, also known as a pharmacophore approach. In other words, features are identified on the ligand and on the protein then an overall interaction energy is calculated by summing up pairwise ligand-protein feature scores. This works better for larger and more complex molecules as the approximations used tend to average out more. For smaller molecules, the errors in the approximations used do not get averaged so well, leading to less accurate results.

Scaffold Hopping Methods Demonstrate Successful Fragment Handling

The field of scaffold hopping has been developed over the last 15 years, coming from both ligand-based (LBDD) and structure-based (SBDD) approaches to the virtual screening of whole molecules. The techniques have been refined and modified to deal with the replacement of specific portions of a molecule.

Scaffold hopping tools work by replacing a specific portion of a molecule with structurally different chemistry that is predicted to retain similar biological properties. For scaffold hopping, the portion of the molecule to be replaced is the core of the reference molecule. For other applications, the replaced fragment may be an R-Group or a core of a leaf group or substituent (Figure 1).

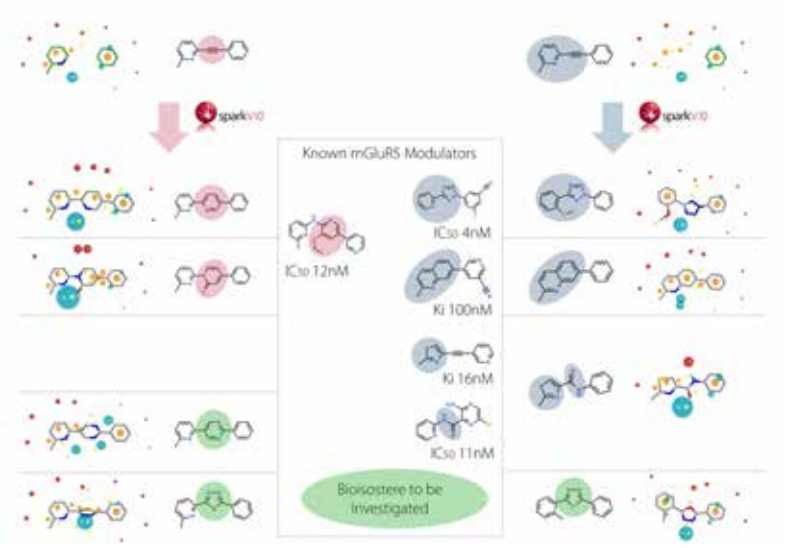

Figure 1: The results of the fragment swapping experiments performed on the mGluR5 modulator MPEP using Spark[5]. Spark automatically searches internal database fragments for fit to a selected region of a molecule, rebuilds it and scores by field similarity back to the reference molecule and ranks the output[6].

In essence, scaffold hopping tools are fragment replacement tools.

Scaffold hopping tools represent a proven method of working with fragments that are in wide use. They have been specifically tailored to handle fragments well and have enjoyed successful applications[3, 4].

Extending Fragment Swapping to Growing, Merging and Linking

Scaffold hopping techniques, based on the identification and replacement of small sub-sets of a given molecule, are readily extendable to fragment growing, merging, linking and optimisation.

Fragment growing is the process of building sensible chemistry around a fragment. Fragment growing typically starts with a single fragment hit. It proceeds by extending the molecule, typically in the context of an X-ray crystal structure, to explore further parts of the protein binding site in order to increase potency.

Fragment merges start with two overlapping reference molecules. The method takes the decoration of one molecule and substitutes it with the core of the other. The output fragments are, therefore, a combination of both.

For fragment linking, the starting point is two non-overlapping references. The first molecule is grown, then chemically linked to replacements that match the second. The challenge is to build successful bridging chemistry between the two fragments. Scaffold hopping methods have successfully been used to address the challenge of finding linker chemistry. This may be a case of merging adjacent pockets, or linking distant pockets. In either case, the chemistry has to be sensible, tractable and synthesisable, which requires an appropriate computational scoring of the entire proposed new molecule.

Using Scaffold Hopping Software to Link Fragments Bound to P38 Kinase

This P38 kinase example shows how scaffold hopping software can be used to link two separate fragment binding modes from X-ray crystal structures, combining them to form a series of drug-like molecules.

A scaffold hopping software tool was used to grow a fragment bound to P38 kinase. Two inhibitors that had been co-crystallised and deposited into the PDB as 3K3I and 3ROC were used as the starting points (Figure 2). The smaller inhibitor was grown to be 'like' the larger, DFG-in inhibitor.

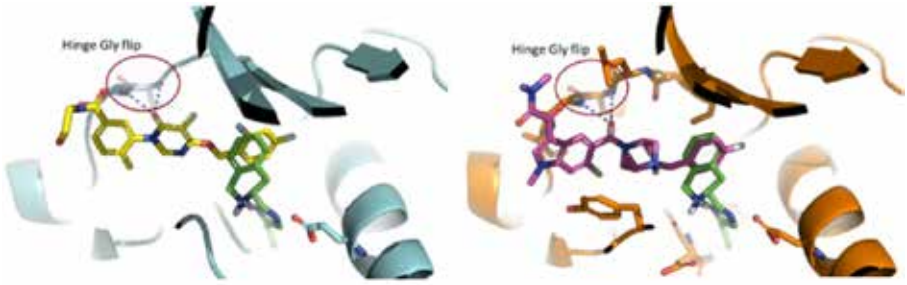

Figure 2. (a) P38 PDB: 3K3I DFG-out inhibitor (green) and (b) P38 PDB: 3ROC DFG-in, hinge-flipped inhibitor (yellow).

Initially the experiment was run starting from the 3K3I derived inhibitor as the fixed component and a dummy atom provided in the ortho position of the tetrahydro isoquinoline (TIC) moiety as the site for fragment replacement. The output field similarity score was weighted 80:20 between these two references derived from the two inhibitors – 3ROC : 3K3I.

A protein excluded volume was not used. The results were evaluated, and the examples shown (Figure 3) display a beautiful fit into the P38 protein. These results were found using a database derived from a commercial pool of compounds. Interesting output was also obtained using different reagent pools e.g., amines and alcohols which would involve the use of different chemical transformations but possibly via the same or a different starting chemical TIC intermediate.

This example shows how scaffold hopping software can be used effectively for fragment growth. The resultant candidate molecules in this example included interesting and sensible results, including highly selective p38 actives.

Key Factors for Success in Growing, Linking and Merging Fragments

There are three key factors that are important to have in place in order for software to deliver a successful elaboration of fragment ideas into proposed molecules with realistic chemistry:

- Firstly, the electrostatics and steric calculations have to be right for small molecules;
- secondly, you have to have a good database of subsets of existing molecules to use as building blocks for growing, merging and linking the fragments;
- finally, it is important to assess each result molecule as a whole in its fully energy-minimised form.

Key 1: Get the Electrostatics Right for Small Molecules

An accurate computational expression of the interaction potential of a fragment is vital for predicting the binding energy of the fragment with the protein target. This relies on a combination of calculations of electrostatic, hydrophobic and shape properties. It is important that the underlying algorithms used by any modelling software translate to realistic results.

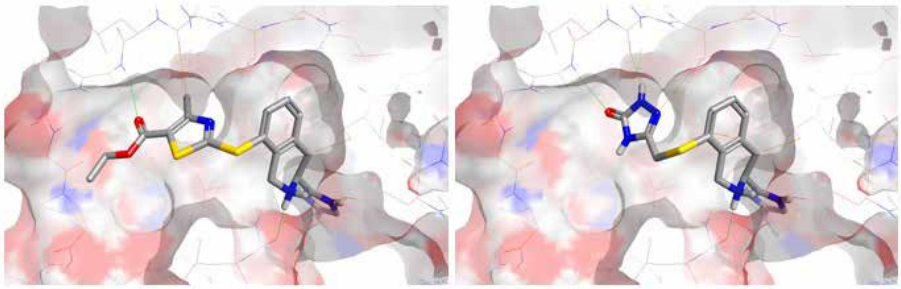

Figure 3. The results display a beautiful fit into the P38 protein. Left, Spark output example 8; right, Spark output example 84 from the thiol reagent pool.

Any errors in calculations will tend to be amplified when dealing with fragments. There is simply less margin for

error since there are fewer atoms from which to calculate properties. Figure 4 shows the types of important detail that need to be calculated accurately when dealing with fragments.

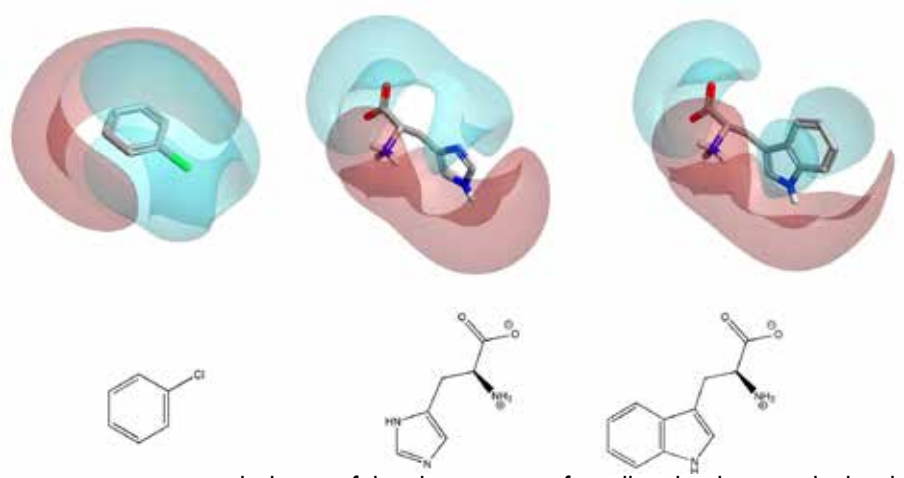

Figure 4: A correct calculation of the electrostatics of small molecules reveals details such as the sigma hole of chlorobenzene (left) and the complex electrostatics of amino acids histidine (centre) and tryptophan (right). These surfaces were calculated using the XED force field from Cresset[7].

A correct calculation of the electrostatics of atoms and molecules results in a significantly improved description of intermolecular interactions. For example, the interaction of aromatic groups, a common interaction in protein-ligand complexes, is correctly predicted by the XED model to prefer an edge-to-face arrangement (Figure 5, right) whereas another algorithm incorrectly suggests that a face-to-face arrangement would be preferred (Figure 5, left).

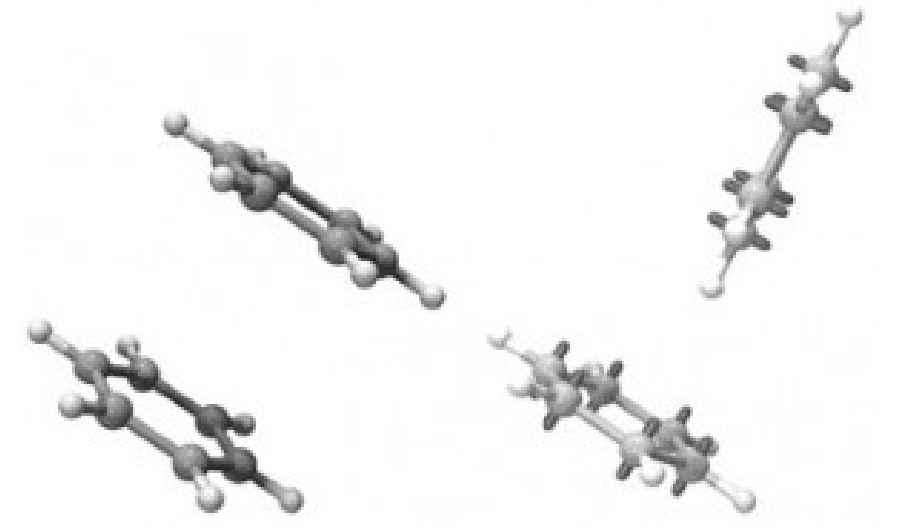

Figure 5. The difference in interaction calculations that can result from an incorrect consideration of the electrostatics. Left: Benzene-benzene interactions as predicted using calculations suited to larger molecules incorrectly suggest face-to-face interactions. Right: Benzene-benzene interactions correctly predicted as edge-to-face using the XED model.

Key 2: Use Subsets of Existing Molecules to Build Realistic Geometry and Realistic Chemistry

The second key factor for success in growing, merging and linking fragments is to have a wide range of alternative pieces of chemistry available from which to build on the fragment. Rather than inventing de novo chemistry to add to the fragment of interest, successful fragment handling software makes use of databases of small sections of existing molecules with realistic geometries and chemistry.

Essentially, such a database is developed by breaking up existing molecules into ever smaller pieces, using chemical rules to identify the sensible bonds to break at each step. If a molecule has been synthesised before then we can have a high degree of confidence that the links can be synthesised again. For example, Cresset's Spark application has a large database of molecular substructures that includes moieties derived from the CHEMBL and ZINC databases.

The use of databases of moieties derived from real compounds also gives a high degree of diversity combined with reasonable probability of chemical synthesis. Further control over the chemistry that is employed can also be gained by the manual selection of connected atoms that form the new bonds to the existing fragment hits.

Key 3: Assess Each Result Molecule in its Fully Energy-minimised Form

The final key factor for success is that each result molecule grown from a fragment must be assessed in its fully energy-minimised form. Once it has been minimised, it can be scored for similarity to the ideal target, or for other criteria to ensure that only linking fragments that can truly work are progressed into the final results.

The electrostatic and steric properties of each result molecule need to be considered in the chemical context in which it will be synthesised. In other words, it is really important to make sure that the added linker chemistry has not adversely affected the interactions from the parent fragments. The molecules need to be assessed in product space to ensure that the final molecule has retained the binding affinity for the target.

This assessment considers both the final shape and the electrostatics, but in this case the emphasis is on the electrostatics. Generally the overall geometry of the molecule doesn't change too much when linked to another moiety, but the electronic structure can be very different. If we take two active fragments bound to different areas of a binding site, then the linking chemistry is unlikely to radically alter the shape of the active fragments; however it could radically alter the electrostatics. The creation of result molecules that are fully energy-minimised before computational scoring

ensures that only linking fragments that can truly work are progressed into the final results. The output molecules can be ranked towards protein pockets or surfaces (excluded volume) and by ranking their 3D field similarity against other bound ligands or substrates.

Case Study - F-uracil as an Anti-cancer and Anti-viral Treatment[8]

Uracil DNA glycosylase (UDG) is a potentially interesting target for both cancer and anti-viral therapies. Recent efforts to produce synthetic inhibitors of this protein relied on an active fragment tethering approach, yielding some interesting bis-oxime linked active ligands. This case study describes an alternative method using molecular modelling software for the efficient elaboration of tractable fragment growing or linking chemistry.

To illustrate the utility of scaffold hopping software for the reconnection of distant fragments, the bis-oxime linker was excised from a 1.5MicM active (PDB: 3FCI, Figure 6).

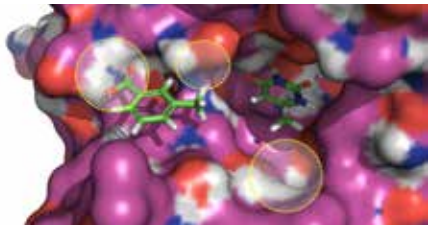

Figure 6. Active fragments originally tethered by a bis-oxime linker (disconnected for this experiment), shown embedded in UDG. In the modelling software, the protein is effectively employed as an excluded volume.

The two fragment atoms to be joined were selected (i.e., the methyl groups of m-benzoic acid and uracil) and suitable fragments with appropriate vectors, which were capable of re-joining them, were inserted. Fragments were sourced from the software tool's large internal databases, derived from ZINC and ChEMBL. The resultant molecules were automatically constructed in-situ, minimised and, in the simplest case, scored against the parent using a field-based 3D similarity metric[7].

The example results shown in Figure 7 show the diverse range of the output suggestions for new linking chemistry for the initial experiments. More importantly, each of the new fragments not only satisfies the geometry and length required for the reconnection of the fragments, but also possesses features which are consistent with important interactions within the protein. Figure 8 shows the electrostatic surfaces of 'Output 3' sited in the protein target.

Rank (BIF%)	Structure	Rank (BIF%)	Structure	Rank (BIF%)	Structure
3 (38)		16 (37)		32 (34)	
60 (34)		76 (33)		174 (29)	
196 (28)		203 (28)		860 (22)	

Figure 7: Example outputs from a fragment linking experiment using the scaffold hopping tool Spark.

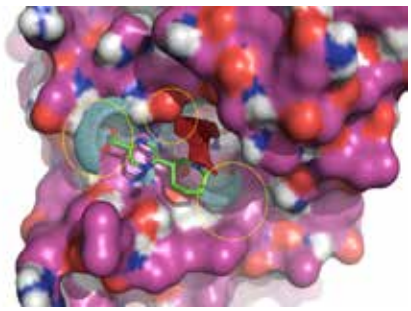

Figure 8. Spark output 'entry 3' with Cresset's XED force field derived negative (cyan mesh) and positive (red mesh) electrostatic iso-potential surface

References

1. Murray CW, Rees DC, Nature Chemistry 1, 187-192 (2009).
2. Kumar A, Voet A, Zhang KY, Fragment-based drug design: from experimental to computational approaches, Curr Med Chem. 2012; 19(30): 5128-47.
3. Bioisosteric Replacements for the Neurokinin 1 receptor (NK1R), pages 259-278 of Scaffold Hopping in Medicinal Chemistry, edited by Nathan Brown, published by Wiley: a retrospective study.
4. Bioisosteric Replacements for the Neurokinin 1 receptor (NK1R), Francesca Perruccio, Scaffold Hopping in Medicinal Chemistry, 259-278.
5. www.cresset-group.com/spark.
6. Cheeseright T, Finding Potential New IP with Novel Bioisosteres of mGluR5 Modulators, http://www.cresset-group.com/wp-content/uploads/2013/11/Finding-Potential-New-IP-with-Novel-Bioisosteres-of-mGluR5-Modulators.pdf.
7. Cheeseright T, Mackey M, Rose S, Vinter A, Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation, J. Chem. Inf. Model., 46, 665-676, 2006.
8. Slater M, Cheeseright T, Using Cresset's Spark to grow and link distant fragment hits with sensible chemistry, Presented at Fragments 2015, Cambridge, UK.

Dr Robert Scoffin is CEO of Cresset. He is an expert in the fields of molecular modeling and cheminformatics, and his previous roles include CEO of Amedis and VP, Europe at CambridgeSoft. His DPhil is in Chemistry from the University of Oxford.
Email: rob@cresset-group.com

Dr Martin Slater is Director of Consulting Services at Cresset. He studied medicinal chemistry at the Universities of Huddersfield and Leeds, and spent 14 years in the highly dynamic pharmaceutical 'fee for service' environment at BioFocus in which he supported medicinal chemistry and a diverse range of programs.
Email: martin@cresset-group.com