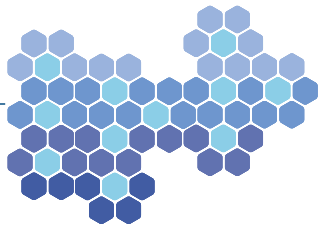


# Chemical Reaction

By Dr Giovanna Tedesco and Katriona Scoffin at Cresset

There is often a divide between computational results and synthetic accessibility in the identification process of alternative compounds. However, there is now a way to produce results more quickly and precisely, and to ultimately accelerate the research process

Bioisostere replacement is a popular computational method for lead discovery. Accurate and fast scoring functions make it possible to identify promising alternative compounds that are likely to preserve activity. However, if there is a significant disconnect between computational results and synthetic accessibility, the suggested compounds will only ever be of theoretical interest. This article explains how the intelligent use of reagent databases from available commercial sources can rank computational outcomes according to synthetic tractability. New ideas can be filtered to focus on results that can be synthesised from reagents, which are available to be ordered online and delivered to the lab within days.

## Scaffold Hopping

Bioisostere replacement – also known as scaffold hopping – is a useful technology for lead detection and optimisation. Projects often start with promising chemistry, but encounter challenges such as toxicity or patent issues. The path to a drug can lead to many dead ends as different scaffolds and R-groups are modified, replaced or discarded. The multi-parameter puzzle requires unwanted side effects to be minimised, while retaining or improving activity and potency. Switching to compounds that are biologically similar, but structurally different, is an efficient way to explore chemical space with a high likelihood of retaining activity.

Keywords

Reagent databases
Synthetic tractability
Field points
Fragment growing

The interaction of a ligand with a protein depends on three main considerations: the reciprocity of electrostatic patterns, the matching of 3D shapes and the matching of hydrophobicity patterns. To make a realistic and useful computational assessment of biological similarity, these three factors must be encoded in such a way that they can be compared over hundreds or thousands of fragments and compounds.

Field points (see Figure 1) are a computationally viable and chemically significant way of encoding the biological effects of chemical changes. They condense electrostatics, shape and hydrophobicity in such a way that it becomes possible to scan large databases of compounds in a reasonable amount of time in order to usefully assess their biological similarity, as demonstrated in the following case study.

## Case Study 1: Finding Bioisosteres of DPP-IV Inhibitors

A leading bioisostere replacement programme was employed to identify ideas for new dipeptidyl peptidase IV (DPP-IV) inhibitors, using known inhibitors as starting points (see Figure 1). DPP-IV inhibitors have demonstrated that they are important agents in treating type 2 diabetes.

Alogliptin (A), omarigliptin (B) and the experimental fluoro olefin compound (C, PDB: 3C45), represent some of the most ligand-efficient examples among known DPP-IV inhibitors. Two experiments were performed using alogliptin and omarigliptin in a simple scaffold hopping exercise. A final chemotype merging experiment was also carried out using a truncated omarigliptin, with fluoro olefin as a second template, to find molecules bridging the two series.

This workflow is shown in Figure 1, along with example outcomes that depict the diverse range of output suggestions provided for new chemistry. The results included examples which already have precedents in patents and literature, serving to validate the method (1).

## Fragment Reconnection

Bioisosteric replacement using fragment- and reagent-specific databases can be readily applied to the popular discovery approaches of fragment growing and linking.

The starting point is one or more fragments that are recognised as responsible for activity. Perhaps they have been isolated from other projects, or it is known that they link with important elements of the protein binding site. The challenge is to grow a compound from one fragment, or to identify linking chemistry for two or more fragments.

Uracil DNA glycosylase (UDG) is a potentially interesting target for both cancer and anti-viral therapies (2). Recent efforts to produce synthetic inhibitors of this protein relied on an active fragment tethering approach, yielding interesting bis-oxime linked active ligands. This example describes an alternative method using scaffold hopping software for the reconnection of distant fragments.

To mimic a fragment linking experiment with the software, the bis-oxime linker was excised from an active ligand (PDB: 3FCI) (see Figure 2). The two fragment atoms to be joined were selected and suitable fragments with appropriate distance and geometry, which were capable of re-joining them, were inserted. Fragments were sourced from databases such as ZINC and ChEMBL. The resulting molecules were automatically constructed *in situ*, minimised and then scored against the parent compound using a field-based 3D similarity metric (3).

The example results (see Figure 3) show the diverse range of the output suggestions for new linking chemistry (4). Significantly, each of the new fragments not only satisfies the geometry and length requirements for their reconnection, but also has features consistent with important interactions within the protein.

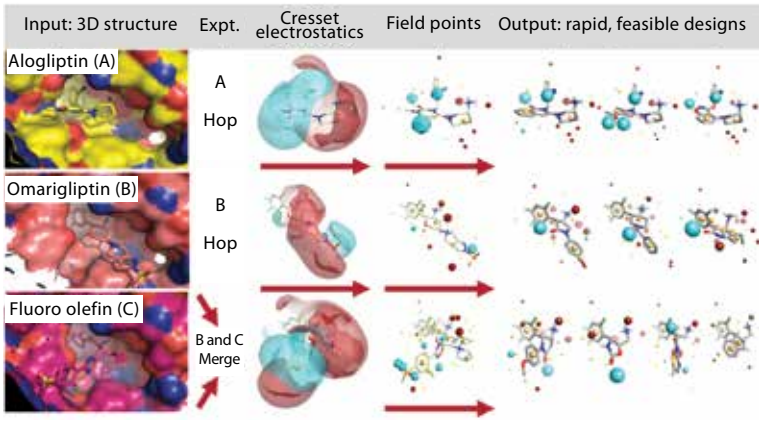



**Figure 1:** Using bioisosteric replacements to find new DPP-IV inhibitors. Cyan represents negative electrostatic field and field points; red represents positive field points; yellow represents steric field points; and orange represents hydrophobic field points

## Accessible Chemistry

*In silico*-generated ideas for R-group replacements become far more practical when they are tied to available reagents. Chemistry being chemistry, nothing can guarantee that a given compound can be readily synthesised – however, narrowing the searches to accessible chemistry is one way to ensure that the bioisosteric replacements have the highest likelihood of being synthesised in real life.

Many attempts have been made to encode all possible transformations and reagents in order to score their synthetic accessibility; however, the results have proven mixed, with several exceptions which render decision-making difficult. An alternative is to come up with a set of interpretable rules for chemistry space – a challenge that also proves to be quite hard. A more powerful approach is to encode

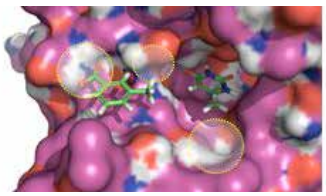


**Figure 2:** Active fragments originally tethered by a bis-oxime linker (disconnected for this experiment), shown embedded in UDG. In the modelling software, the protein is effectively employed as an excluded volume

the origins of an R-group. For example, an R-group could come from a reagent that has undergone multiple steps before it ends up as an R-group. This method makes it possible to encode chemistry space, so that the choices being made are realistic.

Focusing searches on R-group libraries that are classified by specific chemistry provides a systematic way to rapidly exploit a particular chemical reaction.

## Case Study 2: Identifying Optimisation Routes

The reagent databases (5) available with Spark (3) were used to identify alternative decorations

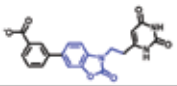
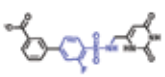
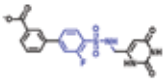
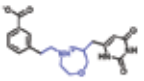
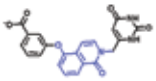
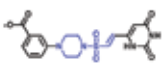
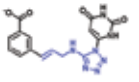
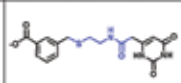
| Rank (BIF %) | Structure | Rank (BIF %) | Structure | Rank (BIF %) | Structure |
|---|---|---|---|---|---|
| 3 (38) |  | 16 (37) |  | 32 (34) |  |
| 60 (34) |  | 76 (33) |  | 174 (29) |  |
| 196 (28) |  | 203 (28) |  | 860 (22) |  |

**Figure 3:** Example outputs from a fragment linking experiment using scaffold hopping

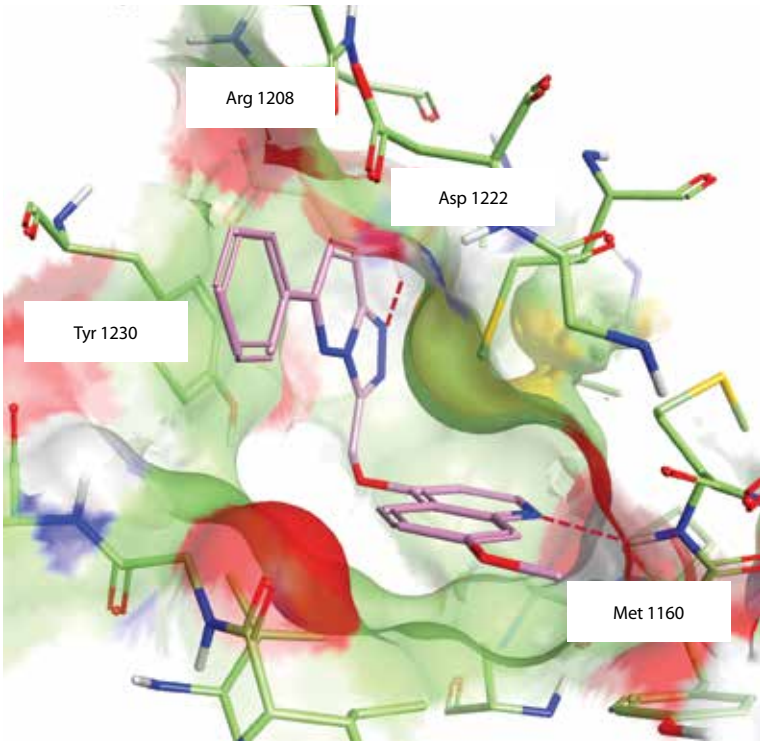



**Figure 4:** X-ray crystal structure of compound C in the active site of c-MET (PDB 3CD8), which elucidates four key interactions (6)

for a series of triazolopyridazine and 8-fluorotriazolopyridine selective inhibitors of the c-MET kinase, developed by Amgen as a potential treatment for cancer (6,7). The use of databases derived from available reagents ensured that the results could be tethered to molecules that were readily synthetically accessible.

The published X-ray crystal structure of an early lead compound D (6) bound to c-MET (PDB 3CD8) shows that this molecule adopts a 'U-shaped' binding mode into the active site. Four crucial interactions can be identified with key residues, which are highlighted in Figure 4.

Based on this experimental information, researchers at Amgen

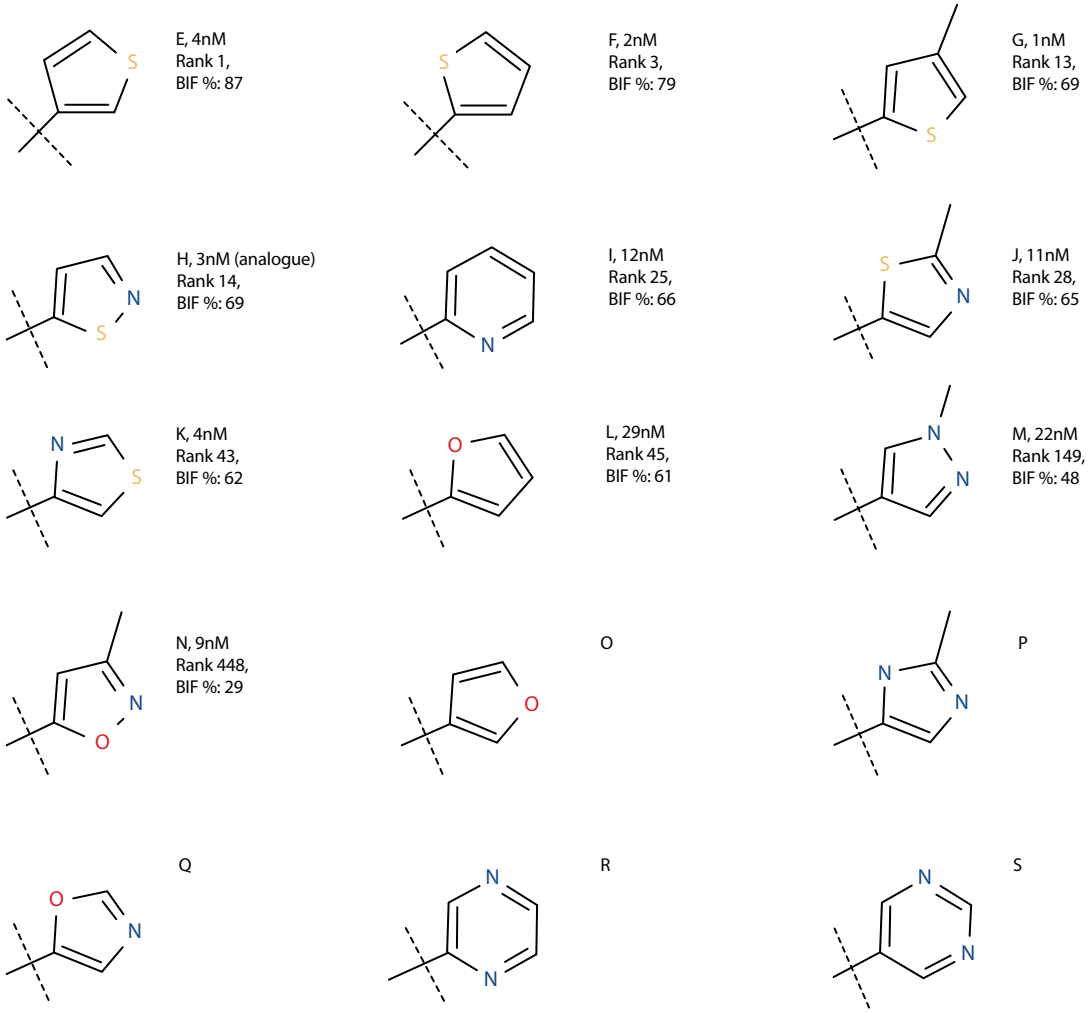



**Figure 5:** R-groups associated with known active inhibitors of c-MET found by Spark

speculated that modifications of the C-6 phenyl group on the triazolopyridazine core would modulate the π-stacking interaction with Tyr1230, allowing for increased potency. They therefore started a chemical exploration based on the synthesis of aryl and heteroaromatic analogues (6). The same strategy was applied to the exploration of 8-fluorotriazolopyridine compounds (7).

The 3D structure of compound D (6) was used in a Spark case study – in combination with reagent databases based on eMolecules building blocks (8) – to verify whether this bioisosteric replacement method could have facilitated the chemical exploration work at Amgen.

The initial experiment was run on a database of 9,500 aromatic boronic acids derived from eMolecules building blocks to closely replicate the chemistry utilised in the original publication (6,7). This search was able to identify the large majority of the monocyclic heterocycles used to explore the C-6 position of c-MET kinase inhibitors published in (6) (see Figure 5). In particular, 3-thienyl (E), 2-thienyl (F), 5-isothiazolyl (H, a close analogue of a 3nM active) and 4-methyl-2-thienyl (G) were correctly identified among the 15 top ranking Spark results.

Heterocycles used in subsequent iterations of the project to explore the 8-fluorotriazolopyridine scaffold (7) were also correctly retrieved. However, while 2-pyridyl (I), 4-thiazolyl (K) and 2-methyl-5-thiazolyl (J) ranked reasonably high in the list of results, 1-methyl-4-pyrazolyl (M) and 3-methyl-5-isoxazolyl (N) were found with a lower rank.

A final Spark search carried out with the 3-methyl-isothiazol-5-yl derivative as a starter molecule on an expanded set of reagent databases (boronic acids and aromatic halides), provided some interesting suggestions about alternative small heterocycles which could have been tried (O, P, Q, R, S).

## Available Synthetic Chemistry

It is possible to link the results of reagent searches to the eMolecules site to enable a check of real time availability information. The results table of a Spark search run on eMolecules reagent databases displays compound availability information. This is important for planning laboratory activity, taking into account realistic delivery timelines.

For example, for some fragments, shipment is to be expected within 1-5 days from order. Delivery times for other reagents are longer: 5-pyrimidinyl boronic acid can be shipped within four weeks (to make S) from order, while 5-oxazolinyl boronic acid (for Q) needs to be synthesised, which may take up to 12 weeks.

## Future Optimisation

In the second case study outlined above, a Spark R-group replacement experiment successfully identified the majority of active heterocycles used by Amgen in the discovery of new potent inhibitors of c-MET kinase.

Lead optimisation is an iterative process. Access to reagent availability information plays an important role in deciding which fragments should be included in each round of optimisation. Reagents with short delivery times should be preferred during the initial stages of the project to facilitate quick structure-activity relationship information gathering, which will enable a more informed choice of fragments to explore in the successive rounds of lead optimisation.

References

1. Slater M, Scaffold hopping into new DPP-IV protease inhibitors, 2015. Visit: www.cresset-group.com/wp-content/uploads/2015/05/Scaffold-hopping-into-new-DPP-IV-protease-inhibitors_Poster_APPROVED.pdf
2. Krosky, *Nucleic Acids Research* 34(20): pp5,872-5,879, 2006
3. Visit: www.cresset-group.com/products/spark
4. Slater M and Cheeseright T, Using Cresset's Spark to grow and link distant fragment hits with sensible chemistry, Presented at Fragments 2015, Cambridge, UK
5. Visit: www.cresset-group.com/products/spark/current-spark-databases
6. Albrecht BK *et al, J Med Chem* 51: pp2,879-2,882, 2008
7. Peterson EA *et al, J Med Chem* 58: pp2,417-2,430, 2015
8. Visit: www.emolecules.com/info/building-blocks

Dr Giovanna Tedesco is Cresset's Product Manager for computational chemists. Previously, she worked as a Senior Computational Chemist at Glaxo, where she supported a variety of drug discovery programmes in the antibacterial and central nervous system (CNS) areas and led target-to-lead CNS programmes.
Email: giovanna@cresset-group.com

Katriona Scoffin is a freelance science writer and marketing professional with extensive experience in the life sciences industry. She currently works for Cresset, an innovative company that uses software to help chemists discover, design and optimise the best small molecules for their project.
Email: katriona@cresset-group.com