Avoiding chemotype bias in virtual screening: diverse chemical space exploration using 3D electrostatic field and shape descriptors

Oliver Hills & Nathan Kidley

Cresset, Cambridgeshire, UK oliver.hills@cresset-group.com cresset-group.com

Highlight

Using a prevalent Malaria target, this presentation illustrates the advantages of unique 3D Cresset field technology over, commonly used, 2D fingerprint similarity methods; as well as exemplifies the diversity enrichment and reduction of the vast chemical space explored.

Background

Malaria is a tropical parasite infection spread by mosquitos, with a global infection frequency of approximately 219 million people, causing approximately 435,000 deaths annually. The causative parasite is *plasmodium falciparum*, transmitted to humans by female mosquitos.[1]

Inhibiting the tautomerase activity of the *P. falciparum* macrophage migration inhibitory factor protein (pfMIF; **Figure 1**) has been the central drug-design approach taken to design novel anti-malarials.[2]

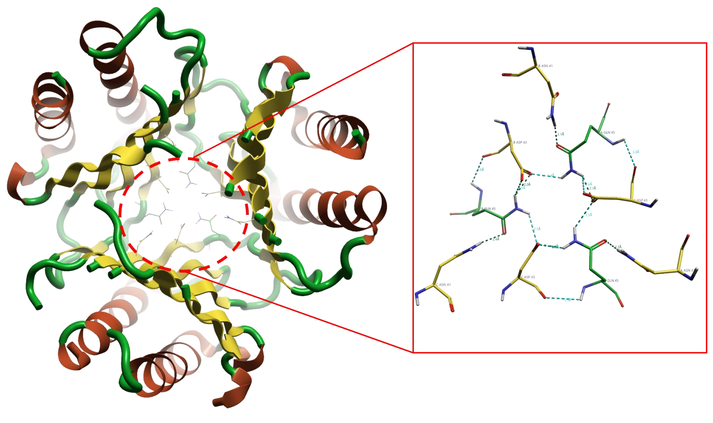

Figure 1: pfMIF functional trimer viewed down its 3-fold axis (PDB ID 4P7M). The amino acid residues implicated in stabilising the trimeric structure are exemplified.

Virtual Screening References

A previous experimental high-throughput screen identified two potent pfMIF inhibitors with nanomolar Ki's[3]:

- 3-[(2-methyl-6-phenylpyridin-4-yl)oxy]phenol (**Compound 1**)
- 4-(3-methoxy-5-methylphenoxy)-2-(4-methoxyphenyl)-6-methylpyridine (**Compound 2**)

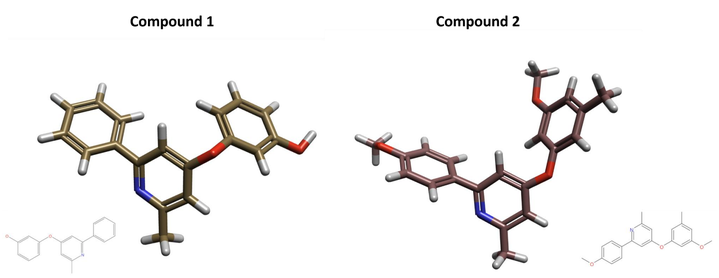

Figure 2: pfMIF inhibitor compounds 1 and 2 in their bioactive conformations.

The bioactive conformations of **1** and **2**, enriched by the 3D field points (**Figure 3**), were used as queries for the virtual screening.

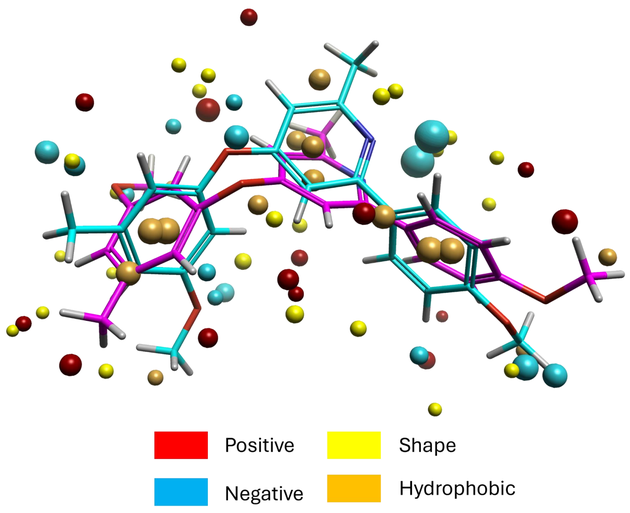

Figure 3: pfMIF inhibitor **compounds 1 & 2** in their bioactive conformations with their field patterns superposed.

Virtual Screening using Blaze™

Blaze is Cresset's advanced ligand-based virtual screening platform, which uses our unique 3D field and shape similarity technology.[4] The Blaze virtual screening workflow is as follows:

For the 3D bioactive input conformation of the query molecule:

- Blaze executes a multi-tiered search using 3D electrostatic and shape similarity over conformationally enriched databases of commercial ("off-the-shelf") compounds
- This returns a ranked list of chemically diverse lead-like hits, with a 3D pose available for every hit

Compounds 1 and **2** were supplied as query seeds and Blaze conducted a multi-tiered search across the full molecular weight range of all commercial compound databases (> 30M compounds). Hits were identified as compounds with high electrostatic and shape similarity (**Figure 4**), which are expected to map the bioactivity of the query molecule.

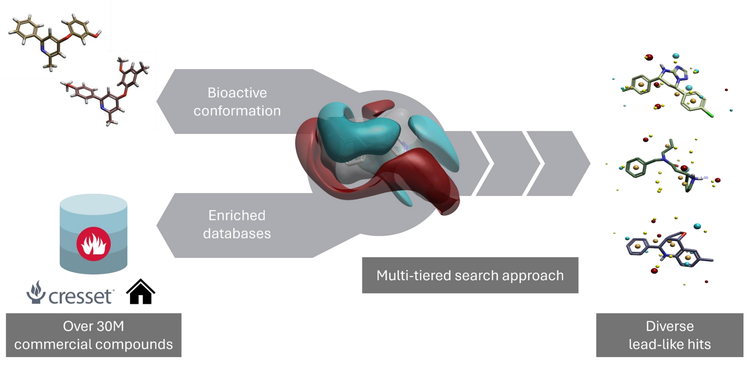

Figure 4: The Blaze virtual screening workflow.

2D vs. 3D Similarity

Analysis of the Blaze hits using 2D RDK5 fingerprint similarity clearly demonstrates that Blaze identifies leads with a wide chemical diversity (**Figure 5**).

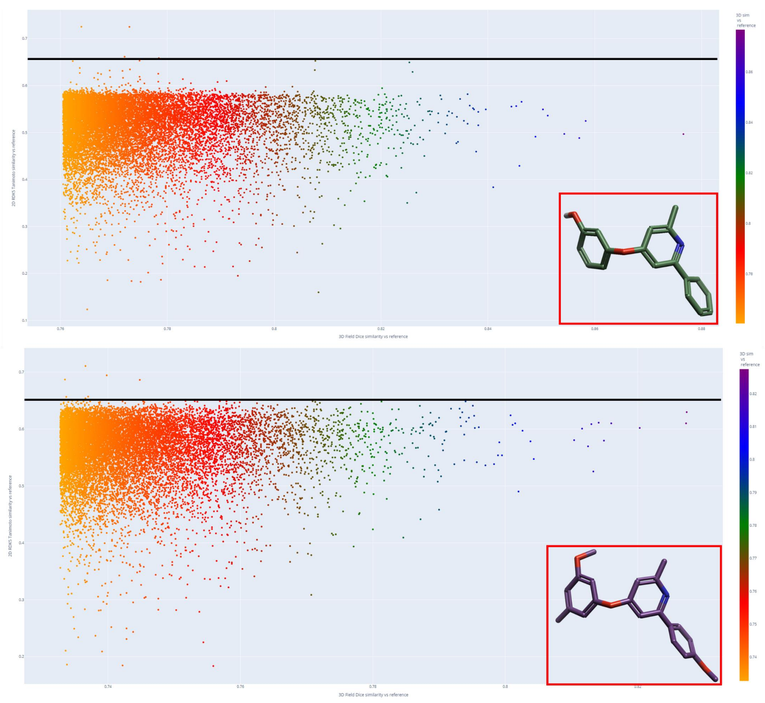

Figure 5: Scatter plots of 2D RDK5 fingerprint similarity vs. 3D Field similarity for **compounds 1** (top) and **2** (bottom). Data points are coloured by 3D Field similarity to the reference. **Compounds 1** and **2** are shown as insets and the conventional 0.65 2D similarity threshold shown as a black line.

Both scatter plots encompass ligands high in 3D Field similarity, but low in 2D similarity:

- Blaze enriches chemical diversity.
- Blaze identified new molecules well below the widely used 2D similarity threshold value to 0.65.[5]
- Screening in 3D Field descriptor space charts expansive regions of 2D dissimilar chemical space.

Chemical Space Visualizations

The expanse of the chemical space explored was visualised using t-Distributed Stochastic Neighbour Embedded (t-SNE) projections of 2D RDK5 fingerprint similarities (**Figure 6**).

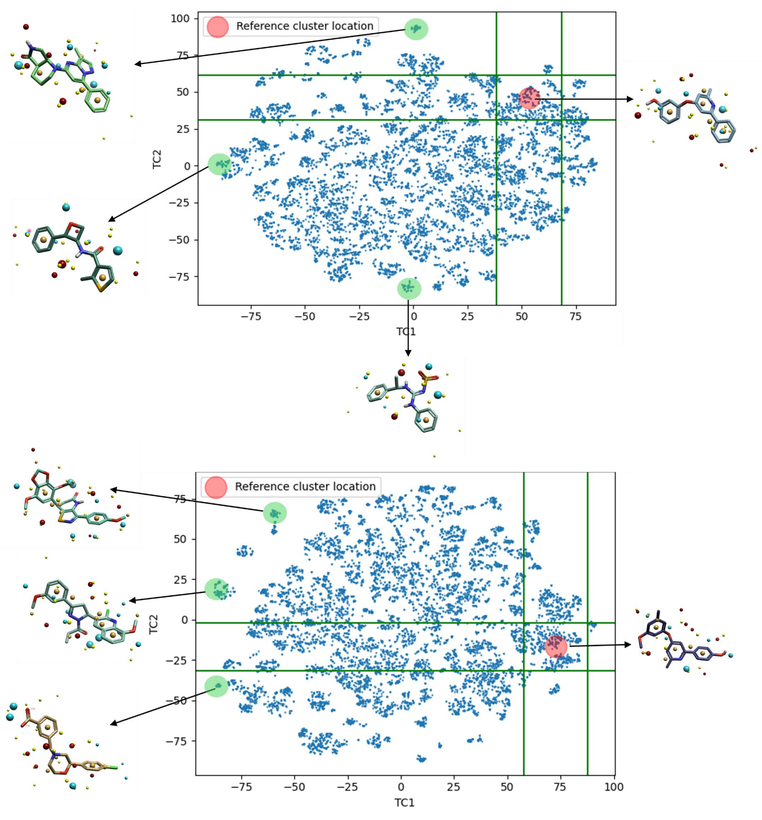

Figure 6: t-SNE plot for the Blaze hits using query **compounds 1** (top) and **2** (bottom). The ligands shown about the periphery of the plot are exemplar ligands from the fringe chemical clusters highlighted in green. The reference molecule is shown also alongside its cluster location highlighted in red.

Blaze identifies molecules far away from the query molecule in 2D chemical space:

- Compounds at the 'fringes' of the explored chemical space belong to very different, and non-obvious, chemical series
- Blaze hits are not intuitive design ideas.
- The 3D electrostatic and shape fields are conserved, despite low 2D similarity.
- Blaze hits can form similar interactions with the target protein, despite the diversity in their chemical structures.

Avoiding Chemotype Bias

Butina clustering of the Blaze hits exemplifies the ability for Blaze to overcome substructure and chemotype biases (**Figure 7**).

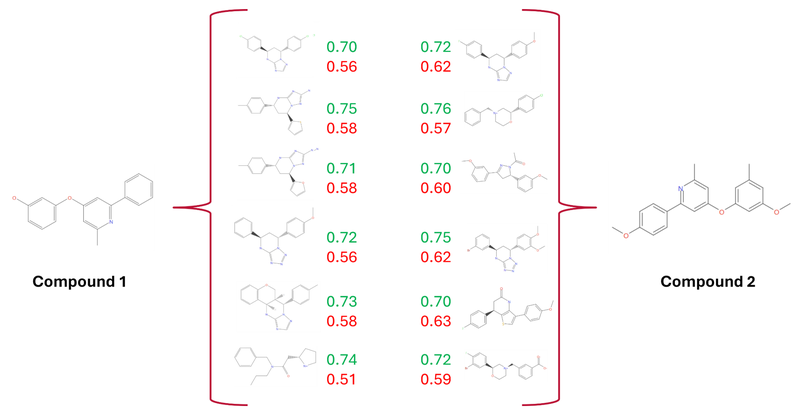

Figure 7: cluster exemplars from the first 12 Butina clusters for the Blaze hits using **compound 1** (left) and **2** (right) as query seeds. The 3D electrostatics and shape similarity score is shown in green, whereas the 2D similarity score is shown in red.

References

1. Cordery, D.V., Kishore, U., Kyes, S., Shafi, M.J., Watkins, K.R., Williams, T.N., Marsh, K., Urban, B.C., 2007. Characterisation of Plasmodium falciparum macrophage migration inhibitory factor homologue. J Infect Dis 195, 905–912. https://doi.org/10.1086/511309
2. 2. Dahlgren, M.K., Baeza Garcia, A., Hare, A.A., Tirado-Rives, J., Leng, L., Bucala, R., Jorgensen, W.L., 2012. Virtual Screening and Optimization Yield Low-Nanomolar Inhibitors of the Tautomerase Activity of Plasmodium falciparum Macrophage Migration Inhibitory Factor. J. Med. Chem. 55, 10148–10159. https://doi.org/10.1021/jm301269s
3. Pantouris, G., Rajasekaran, D., Baeza Garcia, A., Ruiz, V.G., Leng, L., Jorgensen, W.L., Bucala, R., Lolis, E.J., 2014. Crystallographic and Receptor Binding Characterization of Plasmodium falciparum Macrophage Migration Inhibitory Factor Complexed to Two Potent Inhibitors. J. Med. Chem. 57, 8652–8656. https://doi.org/10.1021/jm501168q
4. Cheeseright, T., Mackey PhD, M., Rose PhD, S., Vinter PhD, A., 2007. Molecular field technology applied to virtual screening and finding the bioactive conformation. Expert Opin. Drug Discov. 2, 131–144. https://doi.org/10.1517/17460441.2.1.131
5. Maggiora, G., Vogt, M., Stumpfe, D., Bajorath, J., 2014. Molecular similarity in medicinal chemistry. J Med Chem 57, 3186–3204. https://doi.org/10.1021/jm401411z