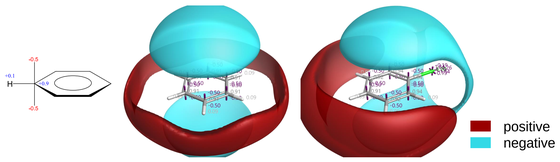
Adding pharmacophores to shape and electrostatics: too much of a good thing?

Martin J. Slater, Tim Cheeseright, Mark D. Mackey, Susana Tomásio

Cresset, Cambridgeshire, UK martin@cresset-group.com cresset-group.com

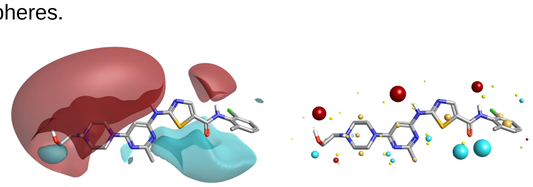
XED force field

Cresset's XED force field[1] provides a detailed description of molecular electrostatics through the use of off-atom center charges. Critical to the XED molecular mechanics approach is the ability to separate partial charges into π- and σ-components.



Calculating fields to assess molecular interactions

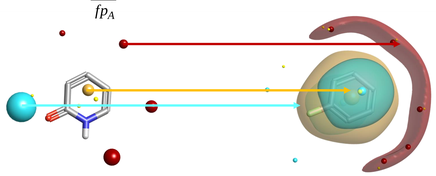
Cresset's 3D ligand similarity technology compares molecules in terms of their molecular electrostatic interaction potentials, aka 'fields'. The local extrema (maximum/minimum) of the fields are termed 'field points', and displayed as colored spheres with size determined by the magnitude of the field – stronger fields get larger spheres.



Similarity calculations

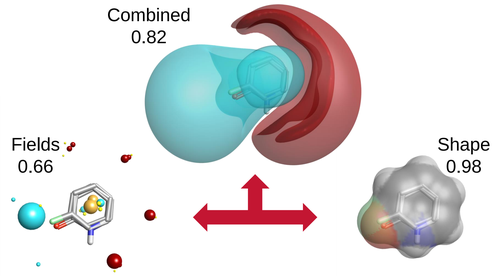
Field similarity[2] between two aligned conformations is based on their full electrostatic potential, sampled only in those places where one of the conformations has a field point. In this way, a raw score can be computed for conformation A into conformation B by determining the field potential for B at the places where the field points for A lie. The overall score can be made symmetric by computing the converse B-into-A score and averaging the two.

$$E_{A\to B} = \sum_{fp_A} size(fp_A) \times F_B(position(fp_A))$$



1. *J. Comp. Aided. Mol. Des*. **1994**, *8*, 653-668
2. *J. Chem. Inf. Model*. **2006**, *46*, 665-676
3. *J. Comput. Chem.*,**1996**, *17*, 1653-1666
4. https://www.cresset-group.com/products/blaze/
5. http://dud.docking.org/
6. US patent 2015/0038510

A shape similarity[3] calculation is also carried out for the two aligned conformations. Typically, field and shape similarity values are then combined 50/50 to get a combined similarity score: this ratio can be fine-tuned by the user according to the specific experiment needs.



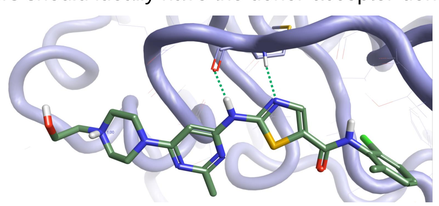
To align compounds in a virtual screening experiment, where the 3D conformation of the query molecule is known, each molecule to be searched is populated with conformations. Each generated conformation is then initially aligned to the query using a variant of color-coded clique matching, followed by least-squares fitting of the field point maps. These initial alignments are then submitted to a simplex optimizer, and the best-scoring alignment (in terms of field/shape similarity) is taken as the 'correct' alignment, and its score as the overall score, for that molecule.

Introducing pharmacophore constraints

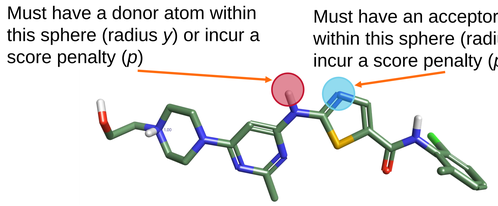
The combination of shape and electrostatics in Blaze™,[4] Cresset's platform for virtual screening of large chemical databases, has proved to be highly successful. However, the user often needs to prioritize particular interactions or molecular features in the query.

We have previously presented methods for doing this in electrostatics space, by adding a field constraint. These work by down-weighting any results that do not have a specific electrostatic or hydrophobic field at a user-specified location, and allow matches across chemical features. For example, a positive field constraint can be matched by both H-bond donors and other electropositive features such as aromatic C-Hs.

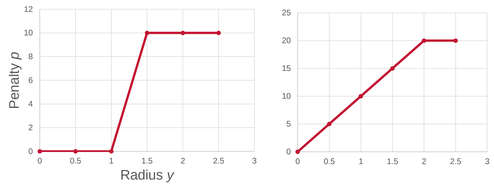
However, there are also cases where a more specific constraint is desired. A typical example are kinases, where all results retrieved from a virtual screen for new hinge binders should ideally have the donor-acceptor-donor motif.



Here we present an extension to Blaze that adds pharmacophore constraints to its electrostatics and shape-based similarity matching algorithm. These will match specific functional groups, such as metal binding functional groups ('metal binder'), electrophilic centers ('covalent'), or other standard pharmacophore types such as 'donor H', 'acceptor', 'cation', 'anion'. Alignments that do not feature a suitable atom on top or close to the constrained atom incur a score penalty.



The optimal set-up for the penalty function was assessed by using a selection of kinase targets taken from the DUD[5] dataset, after adding pharmacophore constraints to the hinge binding motif. Three parameters were evaluated: the radius *y* between the query atom and the hit atom, beyond which a penalty *p* starts to be applied; the maximum value of *p*; the steepness of the ramp function which relates *y* to *p*. Different combinations of these parameters generate different penalty functions, as shown in the graphs below.

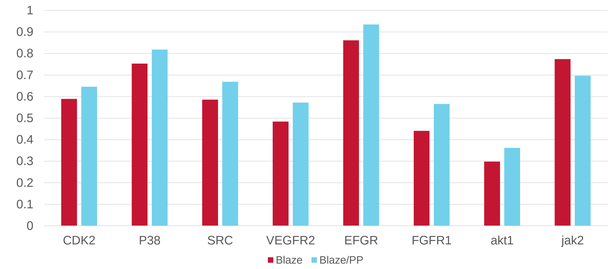


Results

The best overall virtual screening performance (expressed as ROC-AUC) is achieved with a penalty *p* being applied when the radius *y* between the query atom and the hit atom is larger than 0 Å, with *p* reaching a maximum value of 20 at *y* = 2 Å (graph on the right).

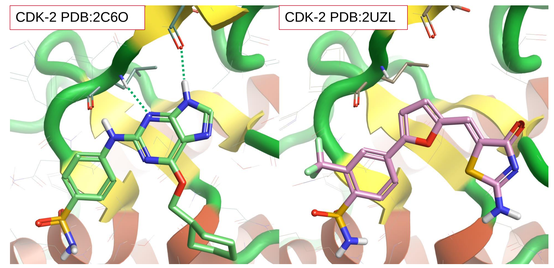
Target	PDB	Normal	*y* 1,2,10	*y* 0,2,10	*y* 1,2,20	*y* 1,2,30	*y* 1,3,10	*y* 1,3,30	*y* 1,2,5
CDK2	1CKP	0.589	0.677	0.646	0.684	**0.693**	0.659	0.689	0.644
P38	1OUK	0.753	0.726	**0.818**	0.517	0.421	0.710	0.417	0.753
SRC	1YOL	0.586	0.629	**0.669**	0.617	0.627	0.626	0.610	0.660
VEGFR2	-	0.484	0.548/ 0.506	**0.572**	0.541/ 0.450	0.546/ 0.413	0.542/	0.537/	
EGFR	1XKK	0.861	0.921	**0.935**	0.919	0.923	0.915	0.920	
FGFR1	-	0.441	0.556	0.566	0.596	**0.600**	0.550	-	
tysy	1syn		0.881	**0.886**	0.870				0.876
akt1	3cqw	0.298	0.353	0.362	0.375	**0.385**			0.350
jak2	3lpb	0.774		**0.697**	0.707	0.494	0.697	0.723	0.678

Overall, we found an average improvement of around 0.13 in ROC-AUC across the tested targets which represents a reasonable gain given the deficiencies in the dataset.



Conclusion

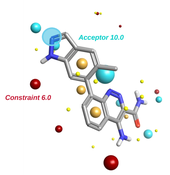
We have found that use of pharmacophore constraints with our electrostatic and shape based scoring method is generally favorable. However, we recommend caution in their use as they represent human bias on the search. For example, the CDK2 ligand from PDB 2UZL, although less active than some chemotypes, lacks the classic donor-acceptor-donor hinge binding motif and hence would not be retrieved by a query with constraints on these features.



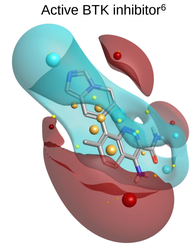
Pharmacophore constraints represent a significant extension to Blaze giving a comprehensive and fully customizable virtual screening system where users can:

- Score hits by electrostatic/shape similarity (customizable, 50/50 default)
- Score hits by electrostatic or shape similarity only
- Use Dice, Tanimoto, or Tversky similarity
- Use excluded volumes (the protein structure) to promote hits that fit in a protein cavity
- Manually intervene to promote specific features:
  - field constraints (electrostatic or hydrophobic matching)
  - pharmacophore constraints (feature matching)

Query molecule (PDB: 4Z3V)



Active BTK inhibitor[6]



Note that the aromatic hydrogens in the active BTK inhibitor match the field constraint but would not match a pharmacophore constraint on the indazole NH.