cresset

smarter chemistry | smarter decisions™

Bioisosteres in Accessible Chemistry Space

Tim Cheeseright

About Cresset

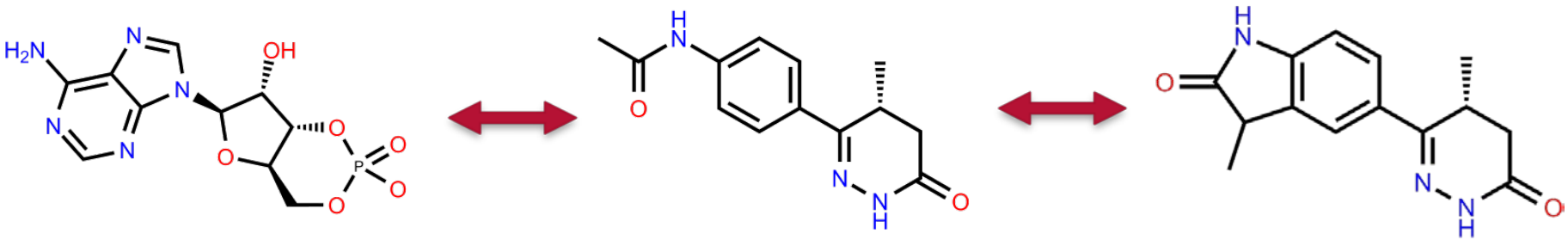
- Founded in 2002 by Dr. Andy Vinter – initial Wellcome Trust funding
- Located in Cambridge, UK
- Primary market pharmaceutical and biotech R&D
  - Other markets include flavour and fragrance, agrochemical and chemicals
- Software:
  - 14 of the top 20 pharmaceutical companies use Cresset technology in their research programmes
  - First software sale 2005
- Consultancy Services:
  - 160 collaborative projects delivered to global clients

cresset

What are Bioisosteres?

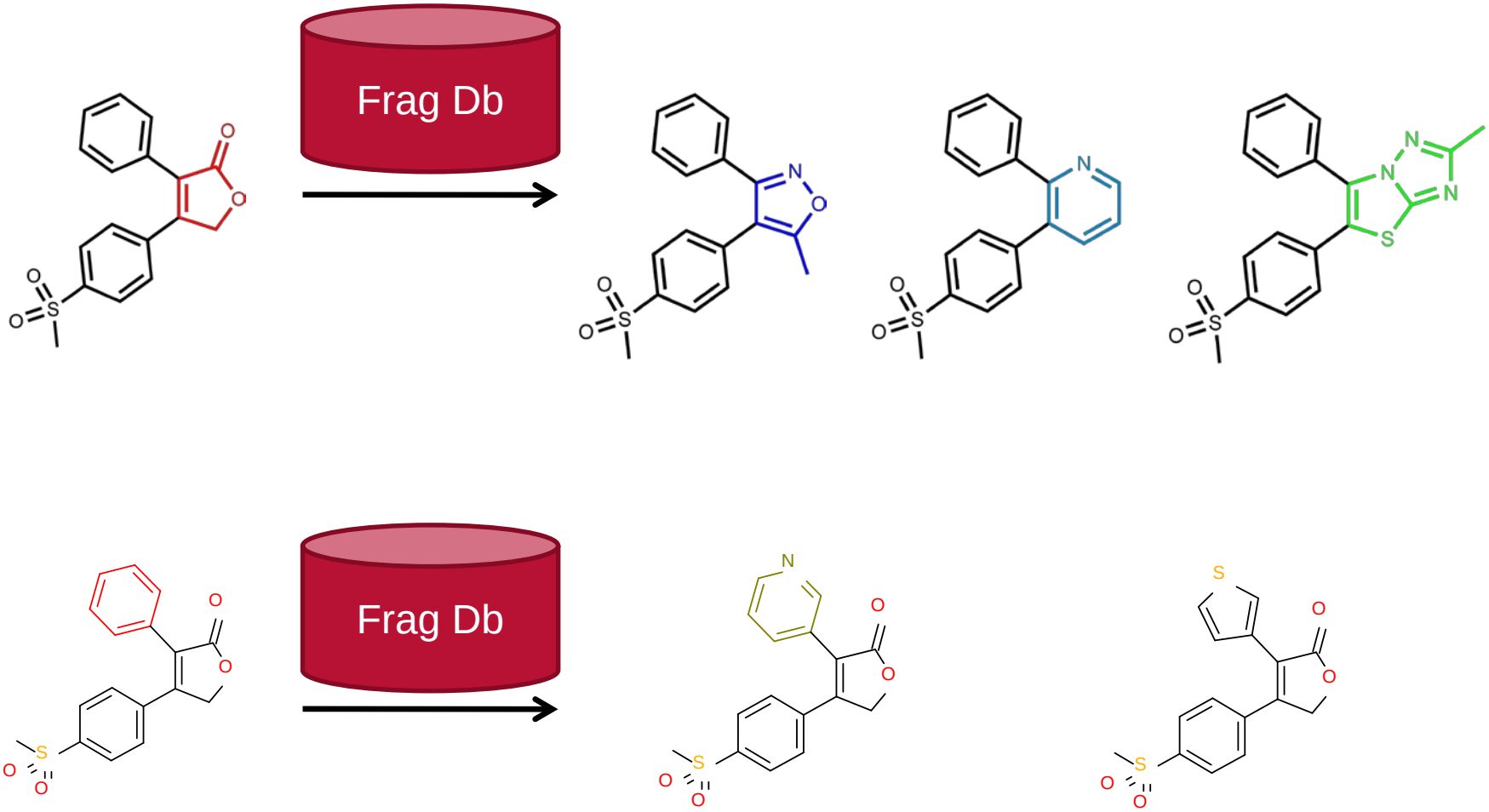
> Broadly:

Any two structures that show the same biological effect



Bioisosteres Bioisosteric groups

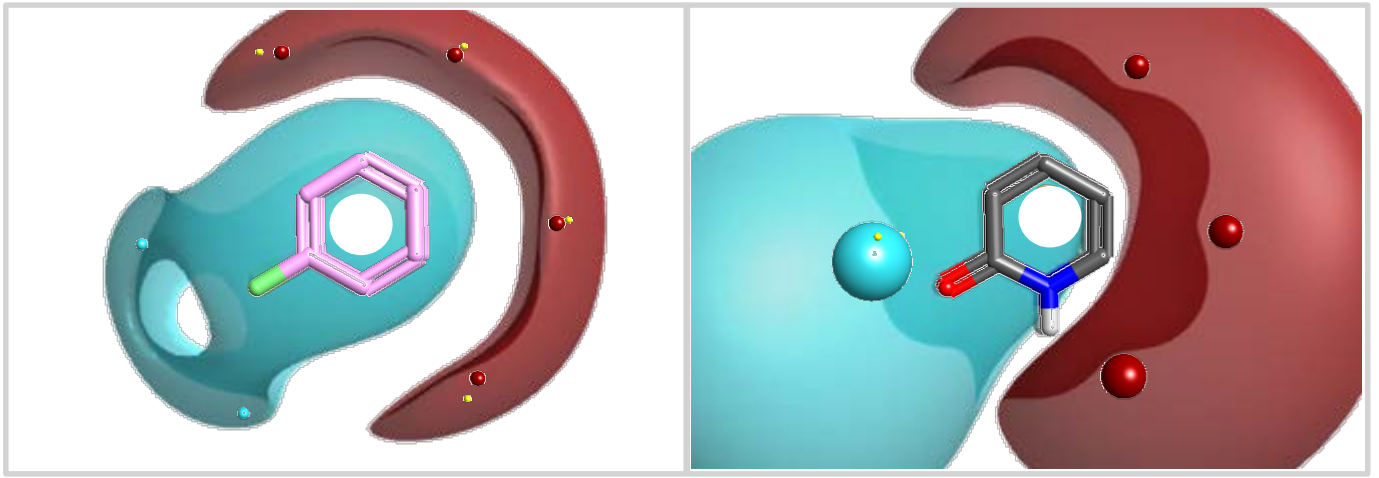
Searching for Bioisosteric Groups



Scoring Replacements

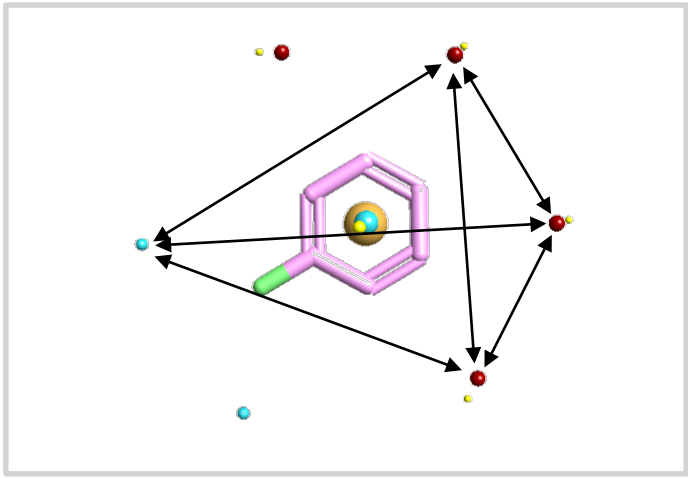
- Structure Based
  - Docking
- Ligand Based
  - 2D similarity
  - Pharmacophores
  - Electrostatic and Shape Similarity

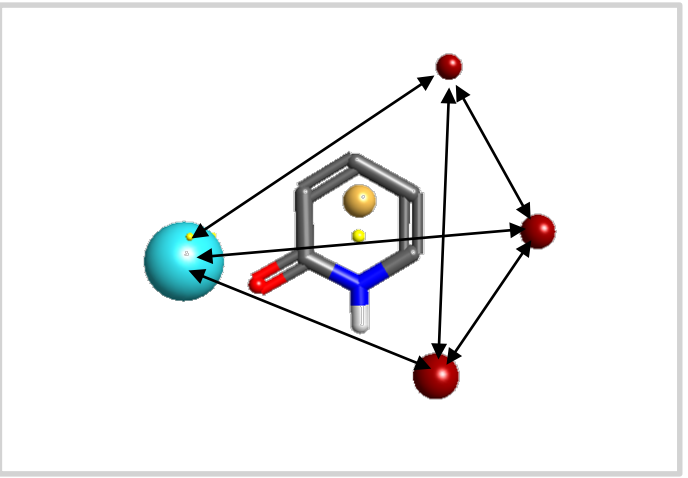
Alignment, Scoring and Comparisons

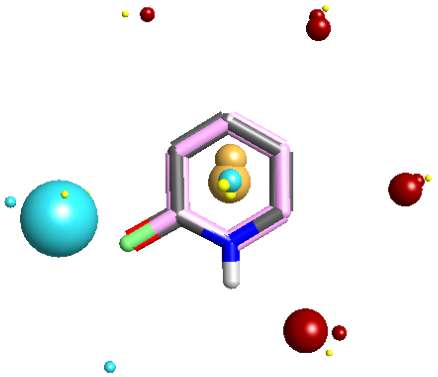


Alignment, Scoring and Comparisons

Clique based alignment

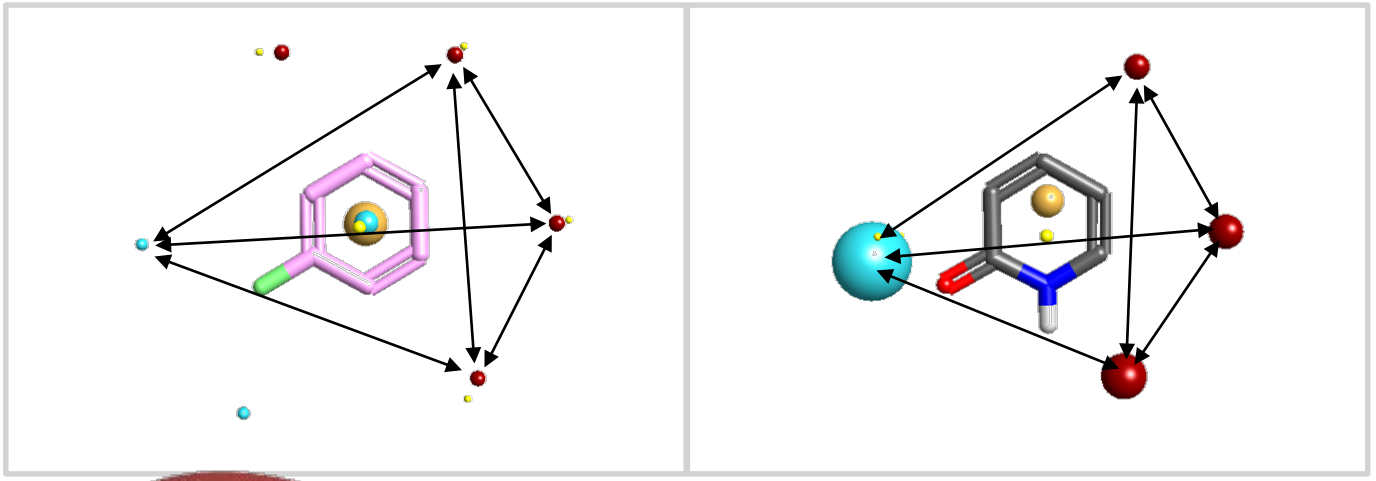


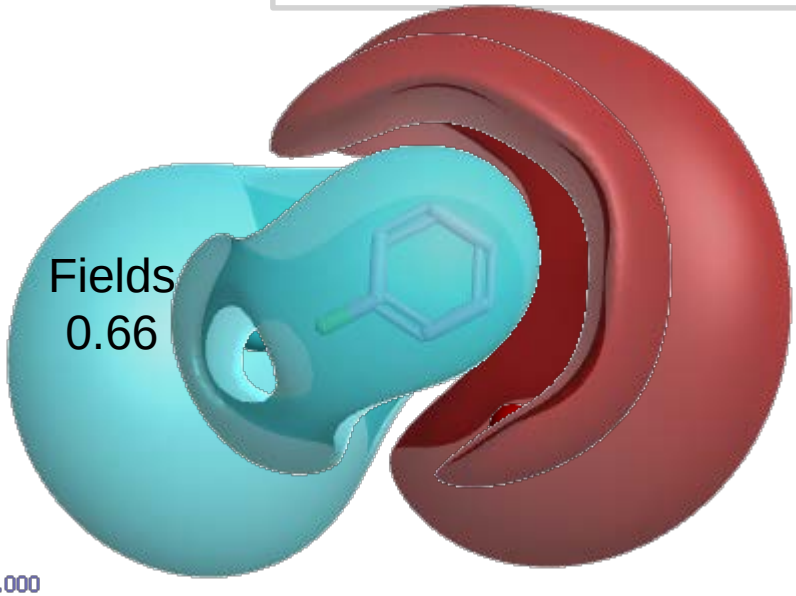




Alignment, Scoring and Comparisons

Clique based alignment

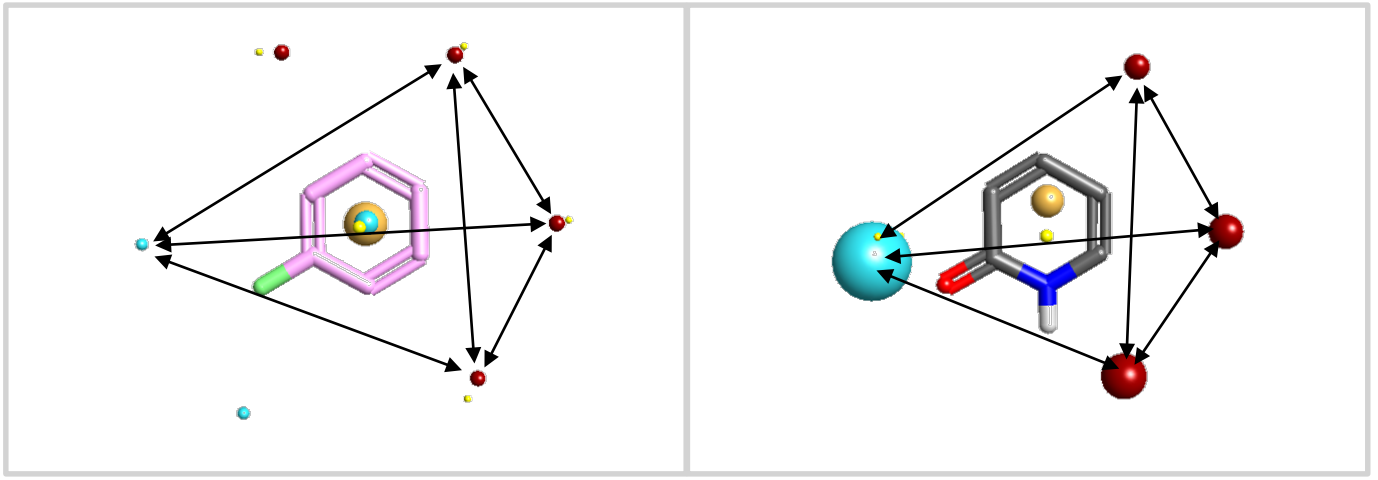




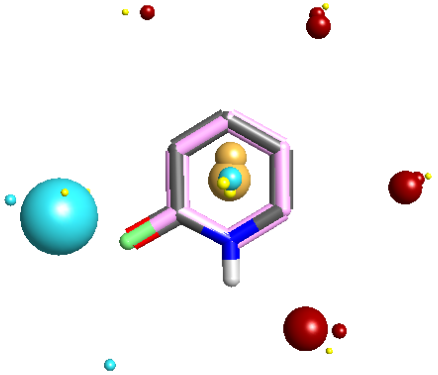
cresset

Alignment, Scoring and Comparisons

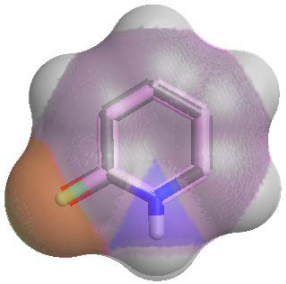
Clique based alignment



Fields
0.66



Cheeseright et al,
J. Chem Inf. Mod., 2006, 665



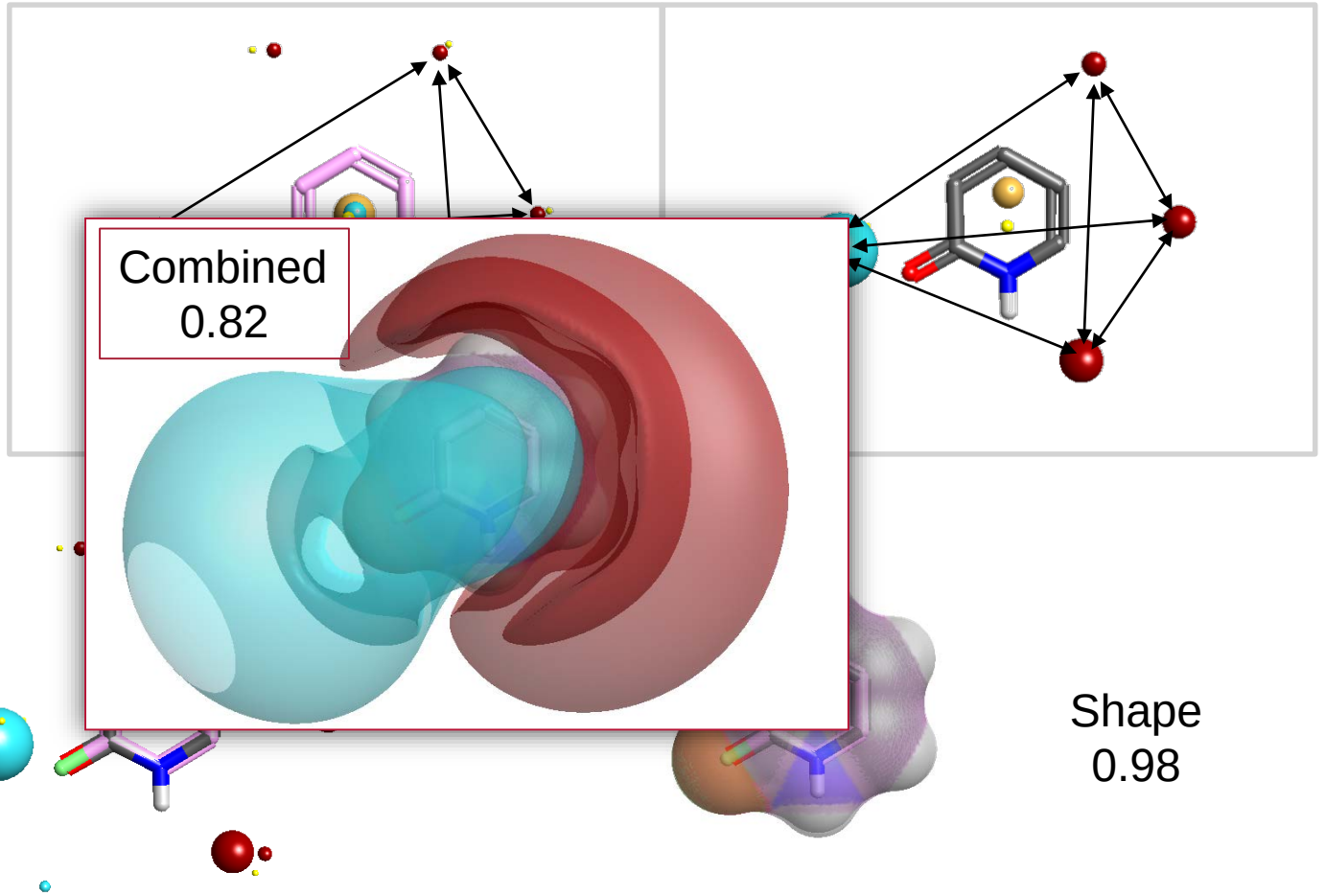
Shape
0.98

Grant, Gallardo, Pickup,
J. Comp. Chem., 1996, 1653

cresset

Alignment, Scoring and Comparisons

Clique based alignment



Fields
0.66

Shape
0.98

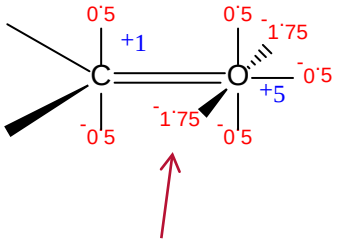
Cheeseright et al,
J. Chem Inf. Mod., 2006, 665

Grant, Gallardo, Pickup,
J. Comp. Chem., 1996, 1653

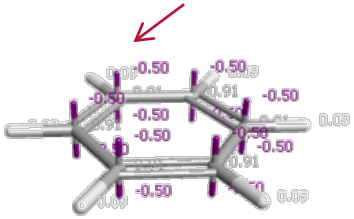
cresset

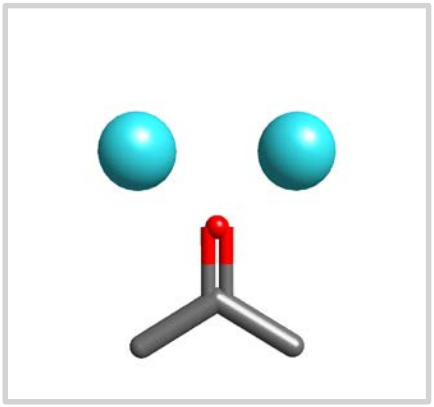
Detailed Electrostatics from XED

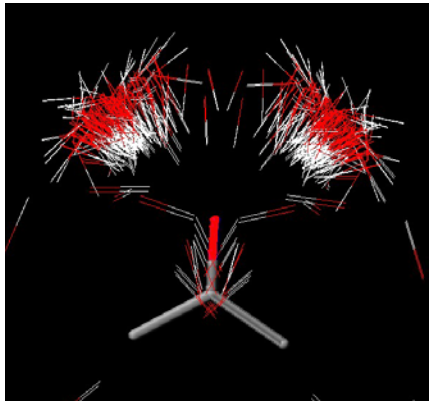
> eXtended Electron Distribution gives detailed electrostatic interaction patterns



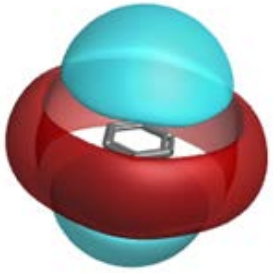
XED adds p-orbitals to get detailed representation of atoms



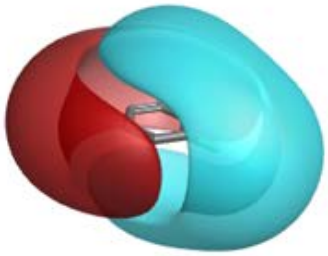




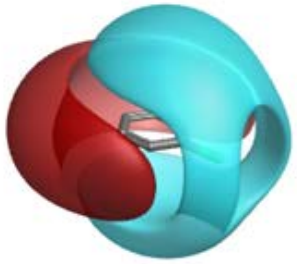
Interaction of Acetone and Any-OH from small molecule crystal structures



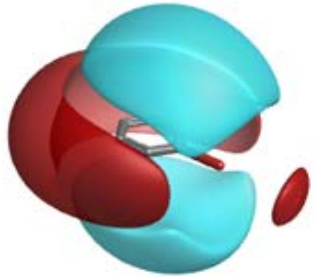
benzene



flurobenzene



chlorobenzene



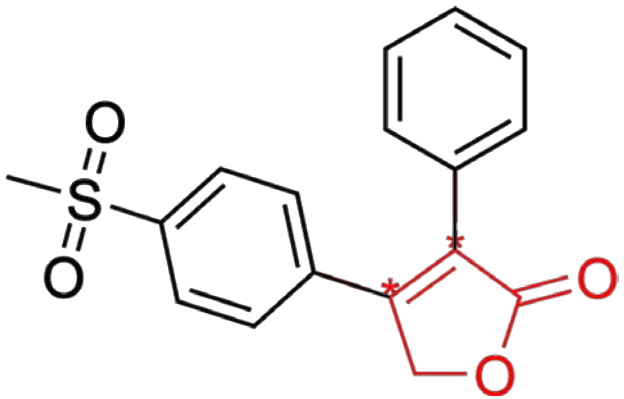
bromobenzene

Scoring with Shape and Electrostatics

- Conformation dependant
- Electrostatics are a property of the whole molecule
- Shape is a property of the whole molecule

Our Approach

1 Select a region to replace and remove these atoms

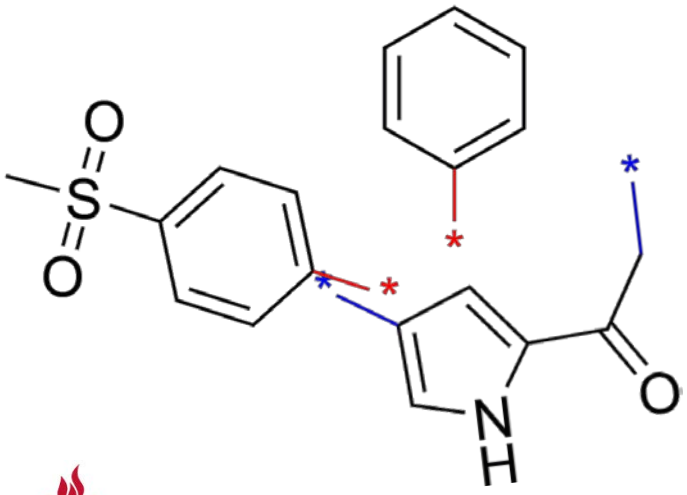


cresset

Our Approach

1 Select a region to replace and remove these atoms
2 Search database for matching fragments
(geometric search only)
(search runs on fragment conformations)

Wrong distance

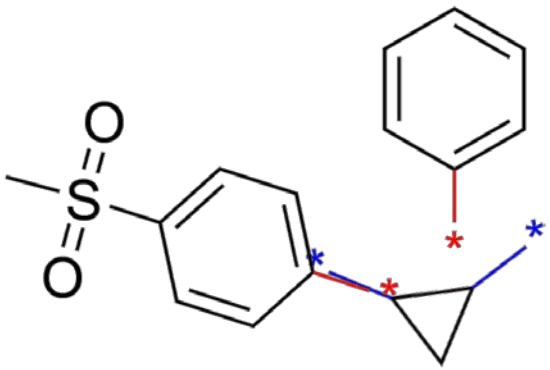


cresset

Our Approach

1 Select a region to replace and remove these atoms
2 Search database for matching fragments
(geometric search only)
(search runs on fragment conformations)

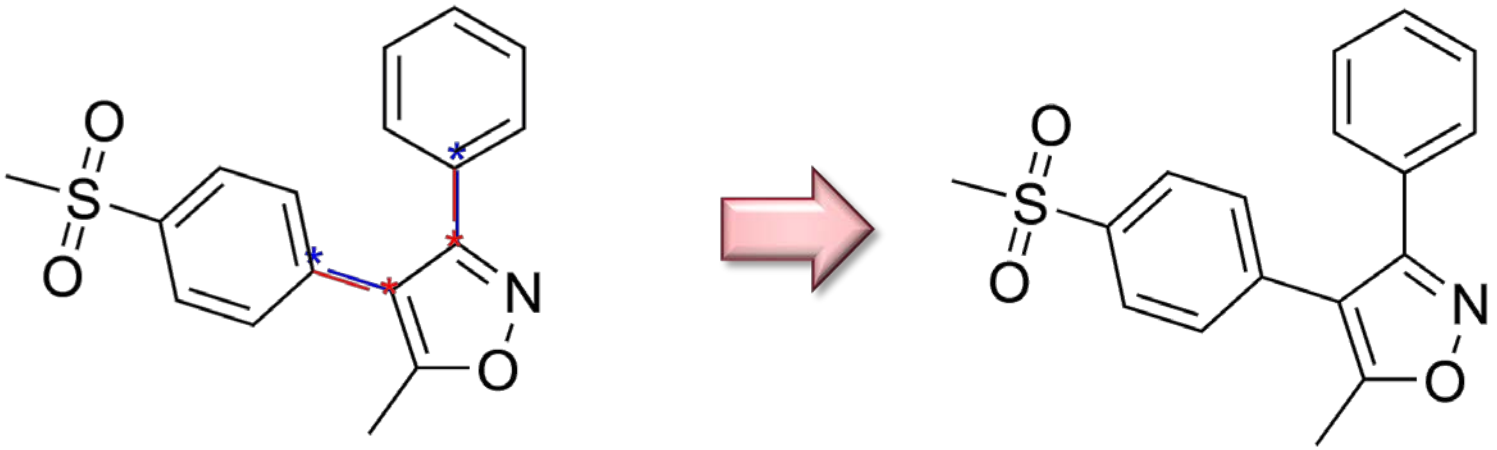
Wrong angle



Our Approach

1. Select a region to replace and remove these atoms
2. Search database for matching fragments
   (geometric search only)
   (search runs on fragment conformations)
3. Form Products
   (minimise and add Field Points)

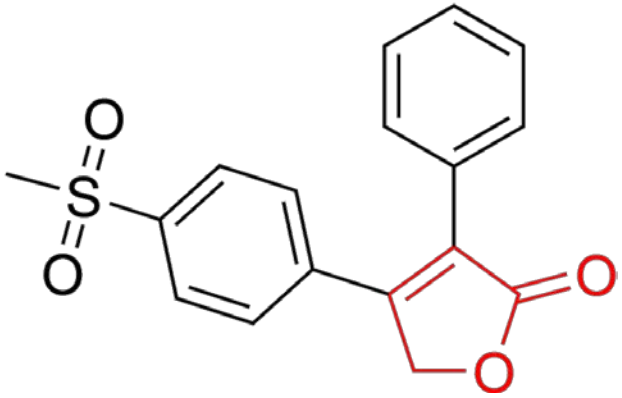
Good match

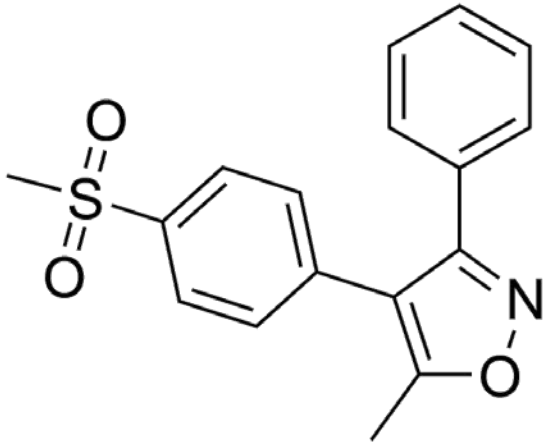


cresset

Our Approach

1. Select a region to replace and remove these atoms
2. Search database for matching fragments
   (geometric search only)
   (search runs on fragment conformations)
3. Form Products
   (minimise and add Field Points)
4. Score



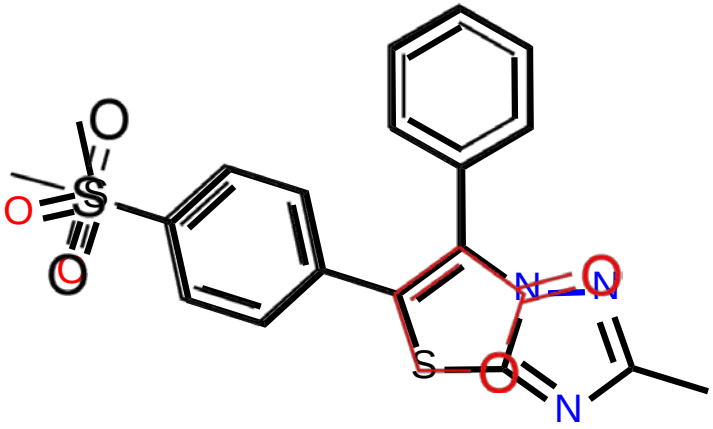


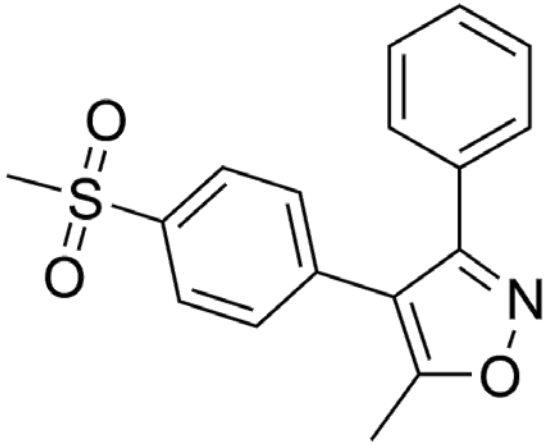
0.88

cresset

Cross scoring modification

1. Select a region to replace and remove these atoms
2. Search database for matching fragments
   (geometric search only)
   (search runs on fragment conformations)
3. Form Products
   (minimise and add Field Points)
4. Score

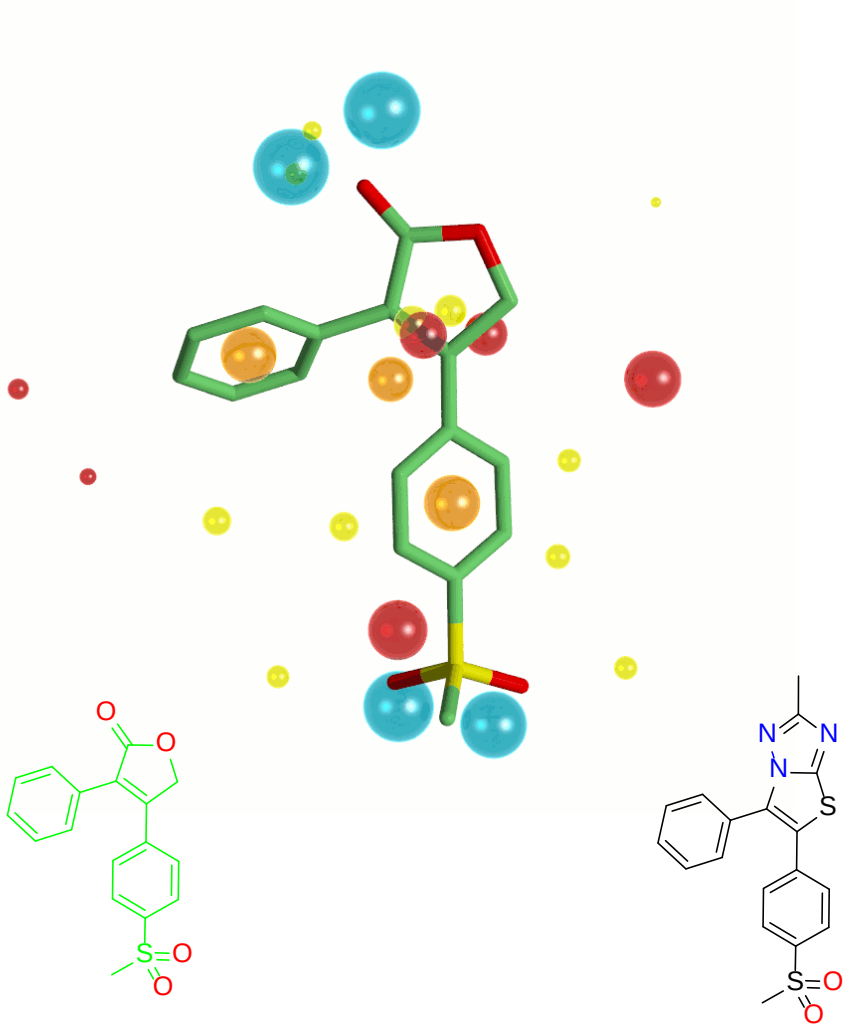




0.68

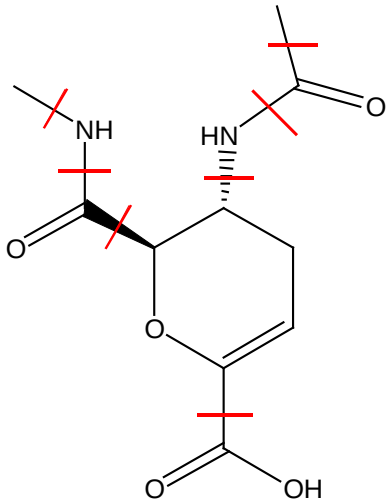
Whole-Molecule Scoring Advantages

- Produces more diverse, non-obvious bioisosteres
  - Avoids fragment scoring limitations
  - Allows for electronic influence of replacing a moiety on the rest of the molecule and vice versa
  - Allows for neighbouring group effects



Where do you get fragments from?

- Fragment existing molecules
- Fragment at
  - Heteroatoms
  - C=O and C=S
  - Bonds to rings
  - Not NO2, COOH, etc



- Generate all sets of connected pieces
  - Subject to MW, NH, and rotatable bond limits
- Sort fragments on frequency
  - Zinc derived → ‘VeryCommon’, ‘Common’, ‘LessCommon’ etc
  - ChEMBL derived → ‘Chembl_common’ etc

Core vs Terminal Replacements

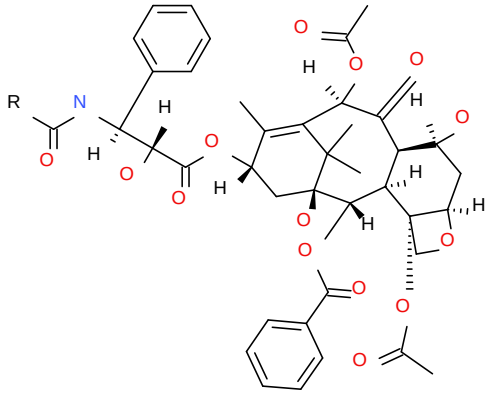
- Core replacement highly utilized in Cresset customers
- Algorithm supports both core and terminal replacements
  - Terminal fragments require rotation after joining to find optimal alignment
- Customer request to link replacements heavily to available chemistry space more tightly

→ Find terminal replacements from list of available reagents

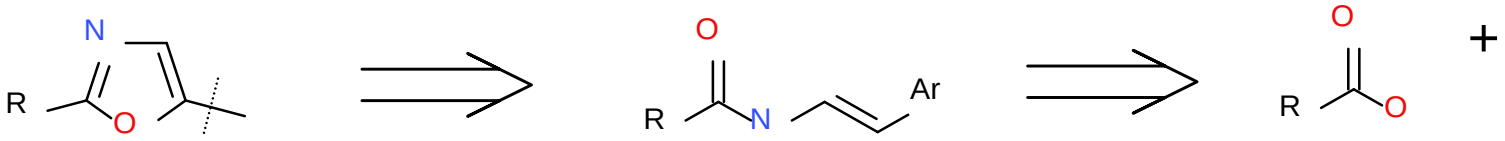
cresset

Processing Reagents into Databases

- Intended to aid synthetically accessible choices for a series
  - Not assessing global “Synthesizability”



- Want to capture the origins of an R group
  - Not trying to encode all chemical transformations
  - Transformation can represent multiple chemical steps



Atpat – Atom/pattern matching

- Cresset's Cheminformatics platform
- Derived from (and similar to) Merck's internal PCP program
- "AWK for Molecules"
- Used extensively within Cresset software to process and modify molecules
  - Fragmentation
  - Re-charging
  - Tautomerization
  - Calculations of TPSA, logP etc

More on atpat

- Pattern/action language
- Syntax:

PATTERN COMMAND

- Pattern can be smarts like atom and/or bond pattern or property that is held by atoms/bonds
- Commands or Actions can operate on
  - Atomic - change atom, charge, delete, add
  - Bond - change type etc
  - Residue
  - Molecule based

cresset

Advantage over SMARTS

- Flexibility
  - Assign properties to atoms and match these properties in patterns
- Multi-line processing enables simple understanding of complex patterns
- E.g. Definition of a protonatable amine
  - Recursive SMARTS (Wikipedia):

```
[$([NH2][CX4]),$([NH]([CX4])[CX4]),$([NX3]([CX4])([CX4])[CX4])]
```

  - ATPAT:

```
N&sp3~sp2|sp|N  > Ntri _ ;
N&sp3!Ntri      > basicN ;
```

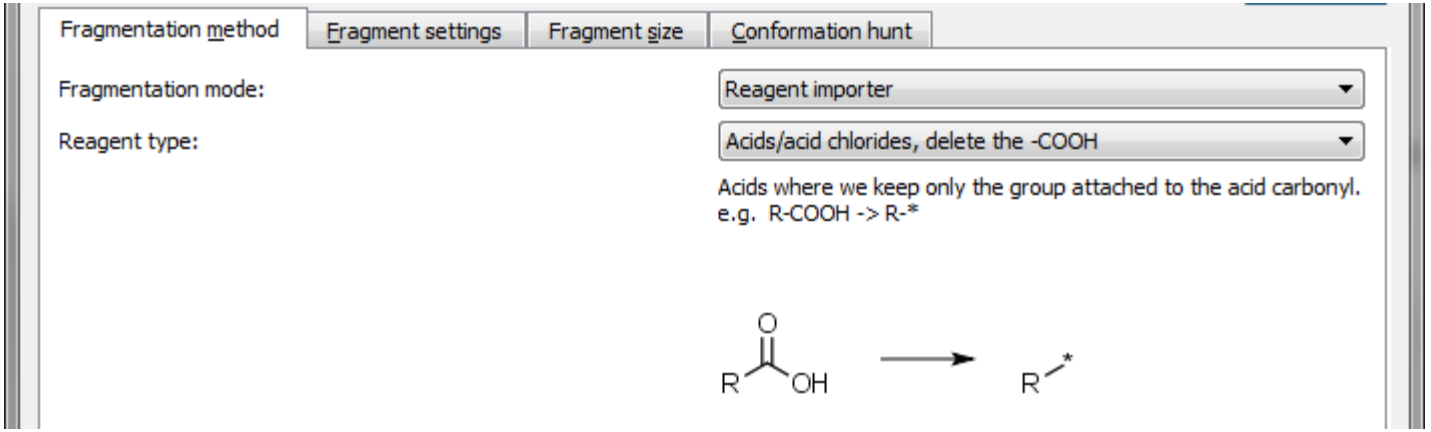
Disadvantages of ATPAT

> “What’s ATPAT?”

> Learning curve

> Odd syntax

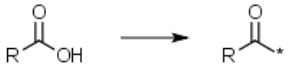
Reagent Processing using Atpat




```
{
TITLE Acids/acid chlorides, delete the –COOH
DESCRIPTION Acids where we keep only the group attached to
the acid carbonyl.DESCRIPTION e.g.  R-COOH -> R-*
}
   !H~C&c3&sp2(~O&c1)~O|F|Cl|Br|I&c1
$n _  52       _      _
>  _  _        del    " ;
del  > # ;
```

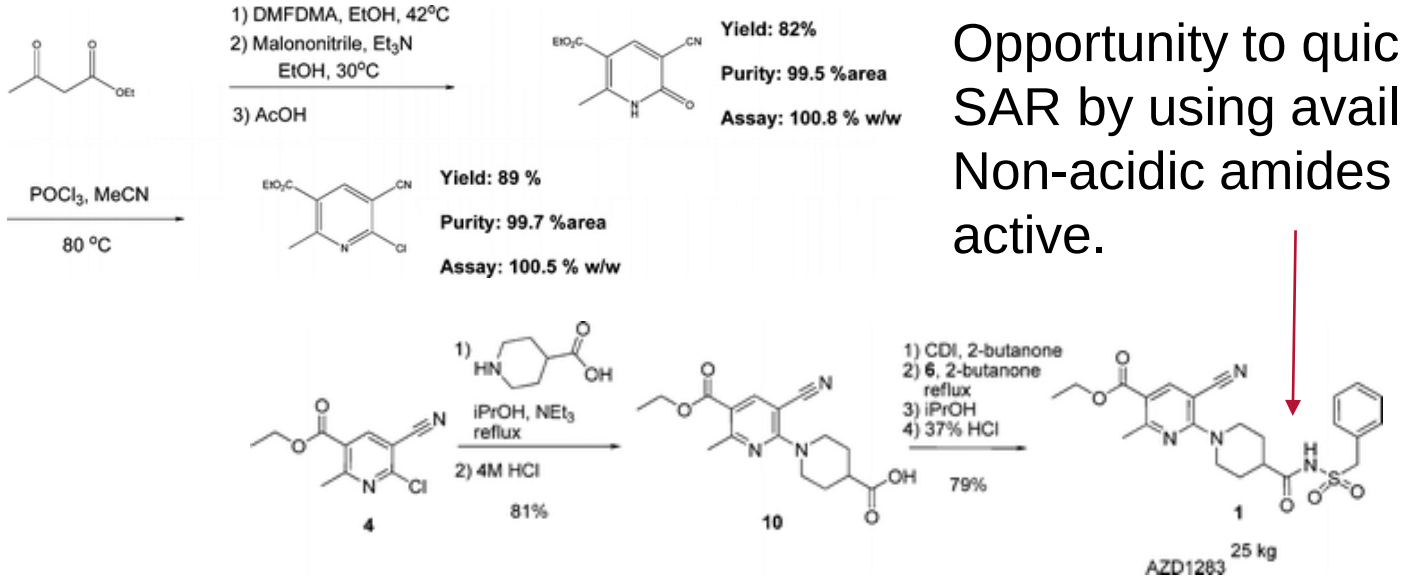
Current implementation

- 17 rules for conversion of R groups to fragments
- Only single attachment points
  - Leaf groups not cores
- http://cresset-group.com/sparkreagentdb/

Transformation	Title	Description
	Acids/acid chlorides, delete the -COOH	Acids where we keep only the group attached to the acid carbonyl. e.g. R-COOH -> R-*
	Acids/acid chlorides, keep the carbonyl	Acids where we attach through the carbonyl group (eg acylations) e.g. R-COOH -> R-C(=O)-*
	Aliphatic alcohols, delete the O	Alcohols used as alkylating agents where the O is deleted on addition e.g. R-OH -> R-*
	Alcohols, keep the O	Alcohols where the attachment is through the oxygen e.g. R-OH -> R-O-*
	Aliphatic halide	Primary/secondary/tertiary aliphatic halides (Cl,Br,I) e.g. R(1-3)C-Cl -> R(1-3)C-*
	Alkynes, delete the -C#C	Alkynes, keep only the attached group e.g. R-C#C -> R-*
	Amines, delete the N	Primary amines as an alkylating agent where the N is deleted on addition e.g. R-NH2 -> R-*
	Amines, keep the N	Primary and secondary amines where the N is the attachment point such as in reductive aminations e.g.

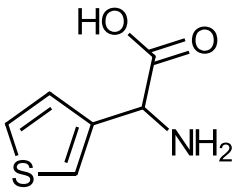
Case Study – Application to $P2Y_{12}$

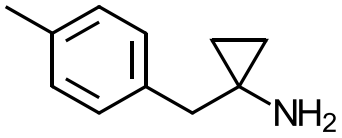
- $P2Y_{12}$ is a GPCR chemoreceptor located on platelet cell membranes: inhibition of $P2Y_{12}$ is an important antithrombotic drug target.
- Apply reagent databases to AZD1283

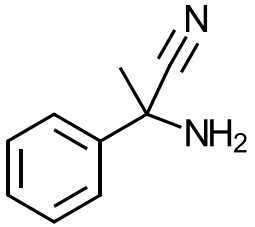


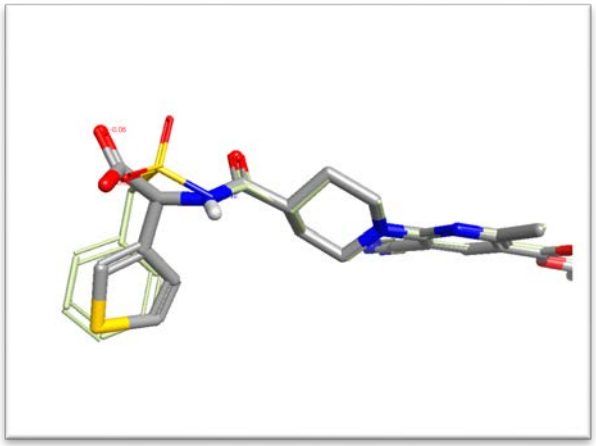
Opportunity to quickly explore SAR by using available amines. Non-acidic amides known to be active.

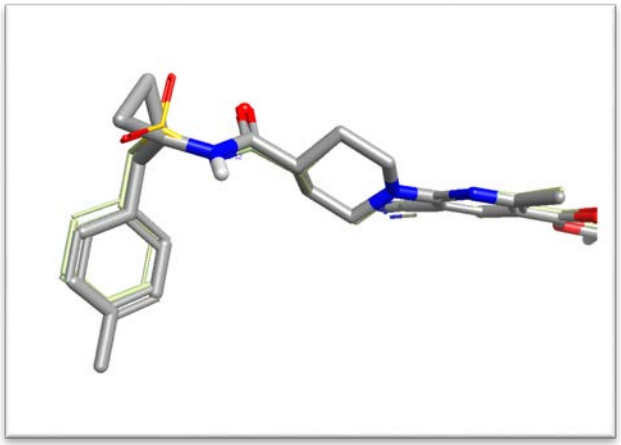
$P2Y_{12}$ Selected Results

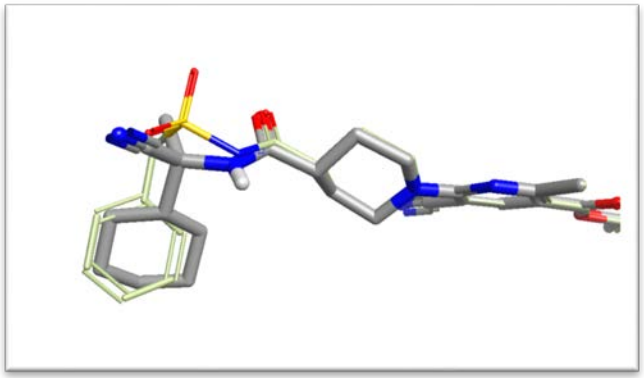












Conclusion

- Reagent based bioisostere databases good for
  - finding imaginative new molecules
  - Linking computational output to synthetic chemistry
- ATPAT
  - Great cheminformatics language
  - Steep learning curve
- Challenges
  - Keeping reagent databases up to date
    - KNIME/PP workflows

Acknowledgements

> Mark Mackey

> Martin Slater

> Rae Lawrence

Thank you

tim@cresset-group.com

cressetgroup

Slides available on request