
innovative science • intuitive software


# Diverse R Groups for Library Design

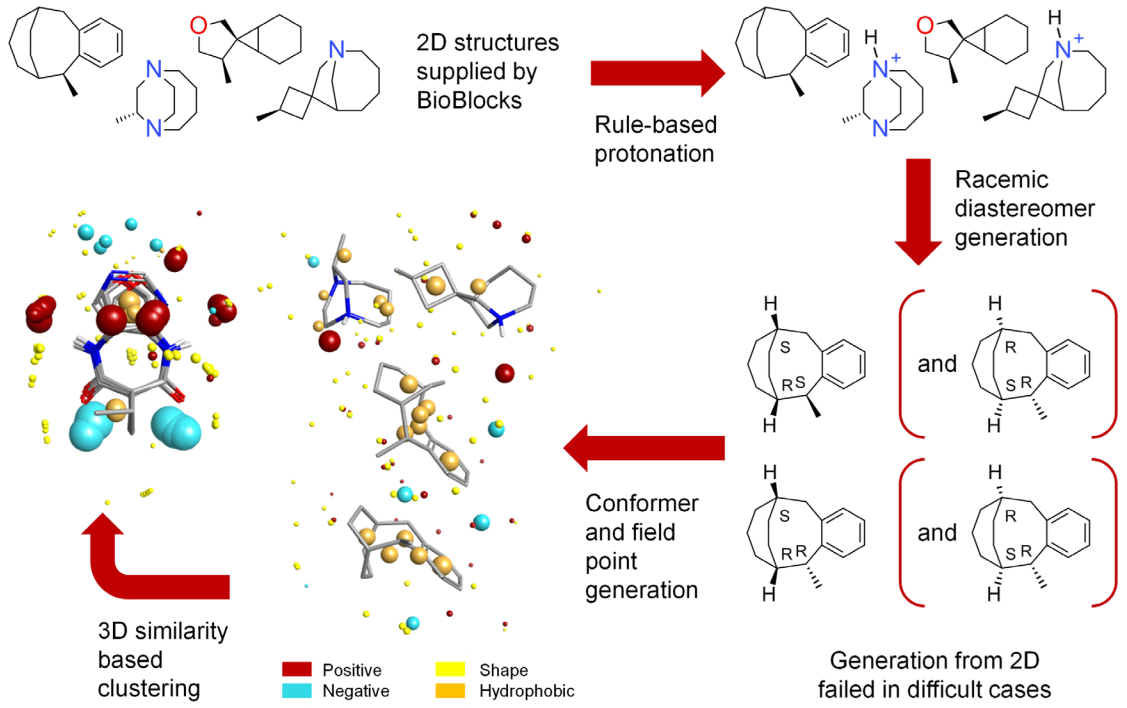
Paolo Tosco and Tim Cheeseright

# The inspiration for PickR™ – customer requests

> Customer request to assess electrostatic diversity in large collection of monomers

→ Collaborative project:

https://www.cresset-group.com/about/news/build-and-cluster-diverse-3d-libraries/



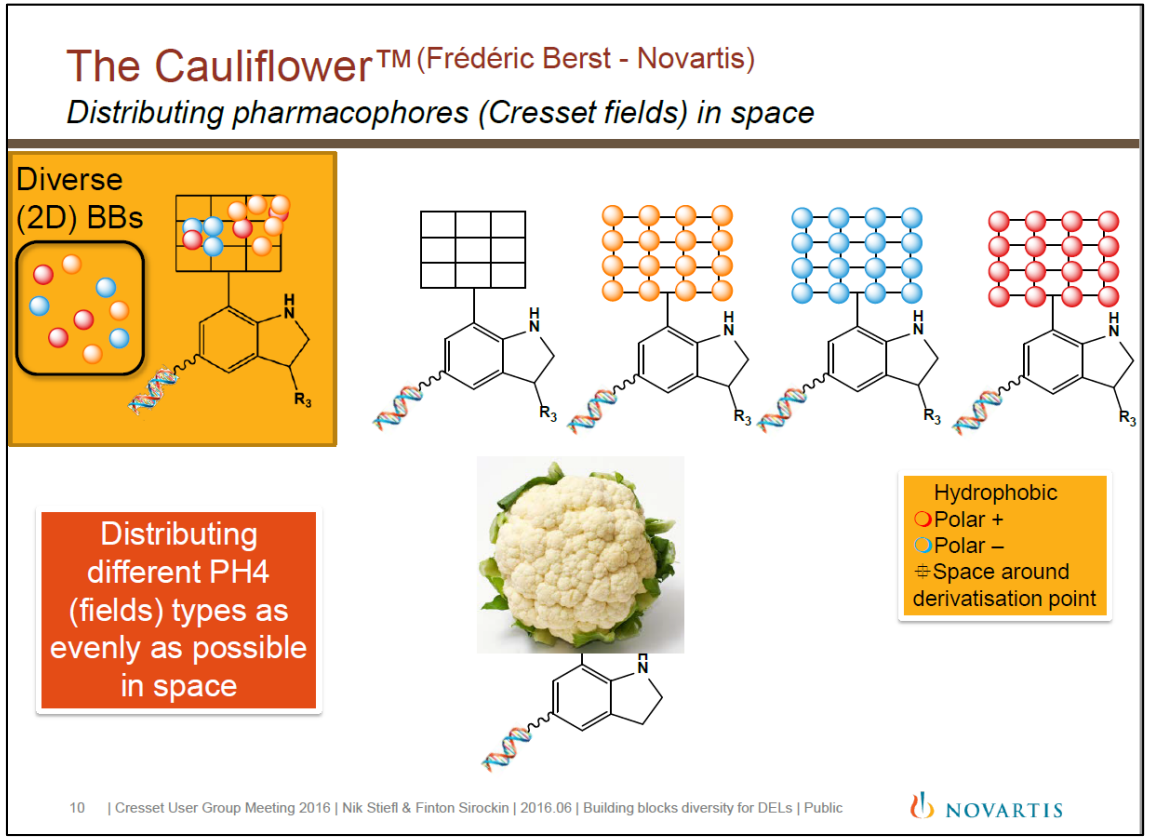
# The inspiration for PickR – customer requests

> Second request for similar workflow to be applied to DNA encoded libraries

→ Collaborative project and improved code:

https://www.cresset-group.com/media/uploads/files/1415_Nik_and_Finton_Novartis_Analysing_Building_Blocks_Diversity_for_DNA_Encoded_Library_Design.pdf

"…. has become the method of choice for many hit finding libraries."

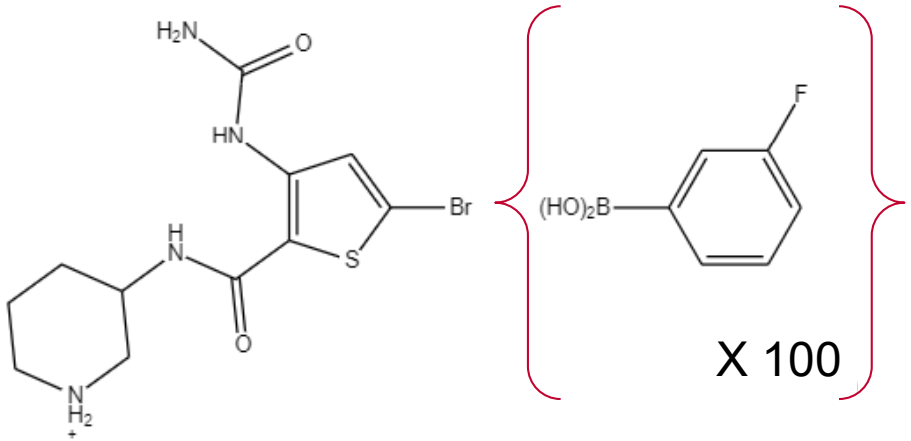


# What is PickR?

- PickR
  - is a monomer selection tool for library design
  - enables users to choose ‘diverse’ reagents from a larger set
  - uses shape and electrostatics to calculate diversity
  - picks R-groups from a list

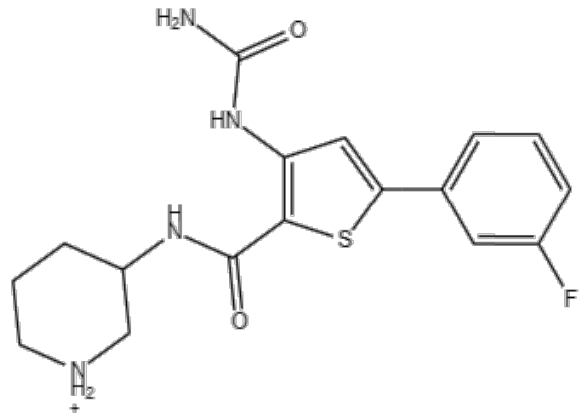
- It is
  - Command line
  - Not a library design tool
    - Library design tools look at the big picture – combining multiple monomers together and then designing the space

# PickR is complementary to Spark

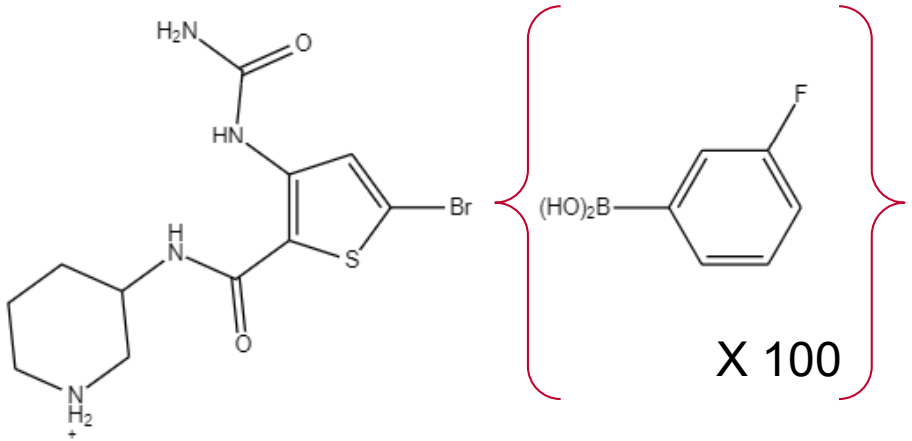



Which 100 boronic acids?





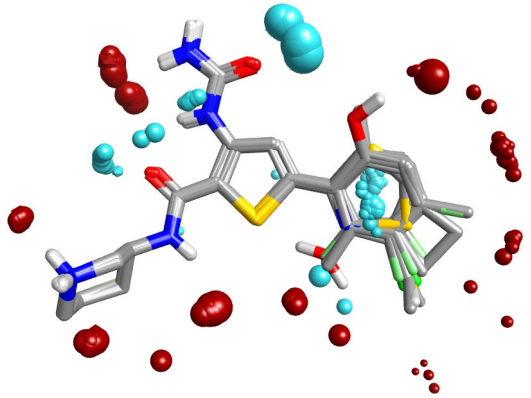
# PickR is complementary to Spark




Which 100 boronic acids?

Spark

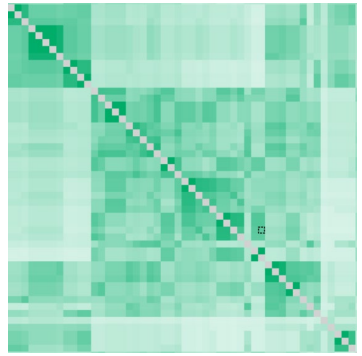
Available Boronic acids
Multi-stage process
Optimised for similarity
to starting molecule



All similar to m-F-Phenyl

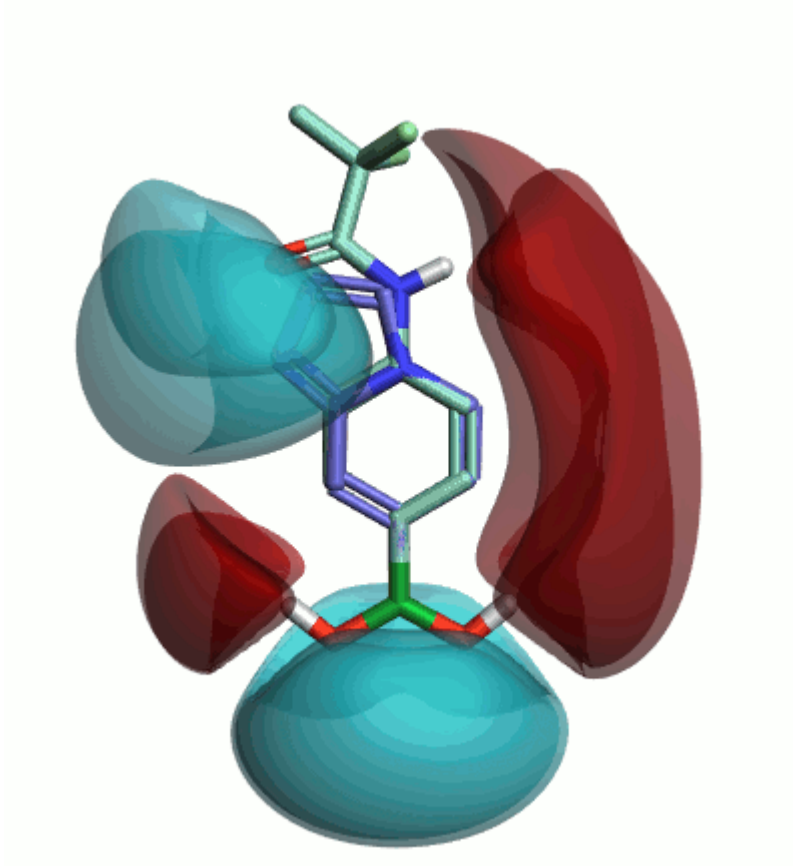
PickR

Available Boronic acids
Multi-stage process
All against all similarity



Diverse pick of R-groups

# Assessing reagent similarity in 3D

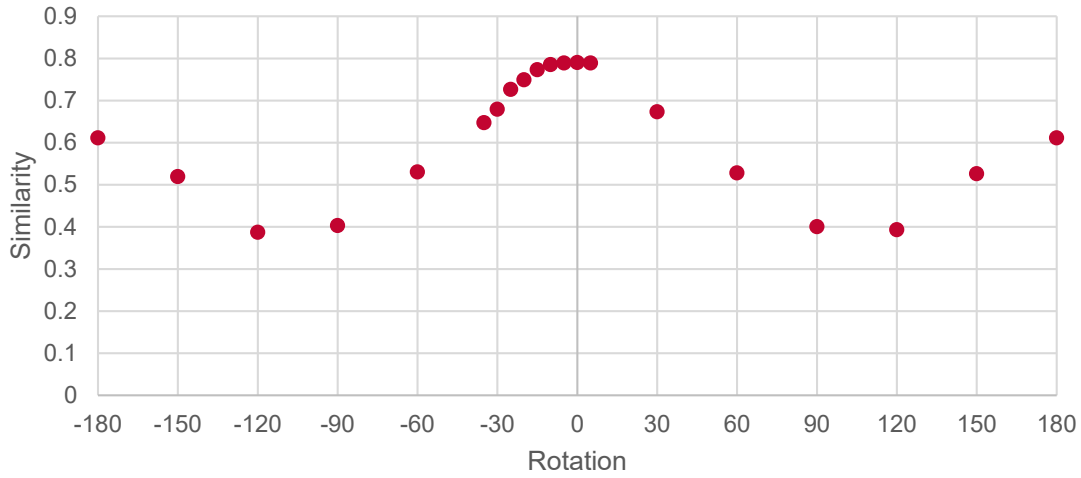


> Conformation on conformation similarity in a restricted context

→ Align on scissile bond

→ Twist around the bond to find the best

→ Repeat for all conformations of A and B



# Complexity from conformations

| Conformations A vs B | B1 | B2 | B3 | B4 |
|---|---|---|---|---|
| A1 | 0.76 | 0.65 | 0.68 | 0.77 |
| A2 | 0.55 | 0.88 | 0.86 | 0.87 |
| A3 | **0.91** | 0.66 | 0.75 | 0.79 |
| A4 | 0.89 | 0.78 | 0.77 | 0.80 |

| Conformations B vs C | C1 | C2 | C3 | C4 |
|---|---|---|---|---|
| B1 | 0.74 | 0.81 | 0.83 | **0.84** |
| B2 | 0.53 | 0.58 | 0.61 | 0.65 |
| B3 | 0.73 | 0.76 | 0.75 | 0.65 |
| B4 | 0.71 | 0.55 | 0.45 | 0.55 |

| Conformations A vs C | C1 | C2 | C3 | C4 |
|---|---|---|---|---|
| A1 | 0.72 | 0.85 | **0.88** | 0.87 |
| A2 | 0.55 | 0.68 | 0.66 | 0.67 |
| A3 | 0.81 | 0.86 | 0.85 | 0.79 |
| A4 | 0.82 | 0.68 | 0.75 | 0.70 |

Simmatrix

| Molecules | A | B | C |
|---|---|---|---|
| A | 1 | 0.91 | 0.88 |
| B | | 1 | 0.84 |
| C | | | 1 |

# Clustering the similarity matrix

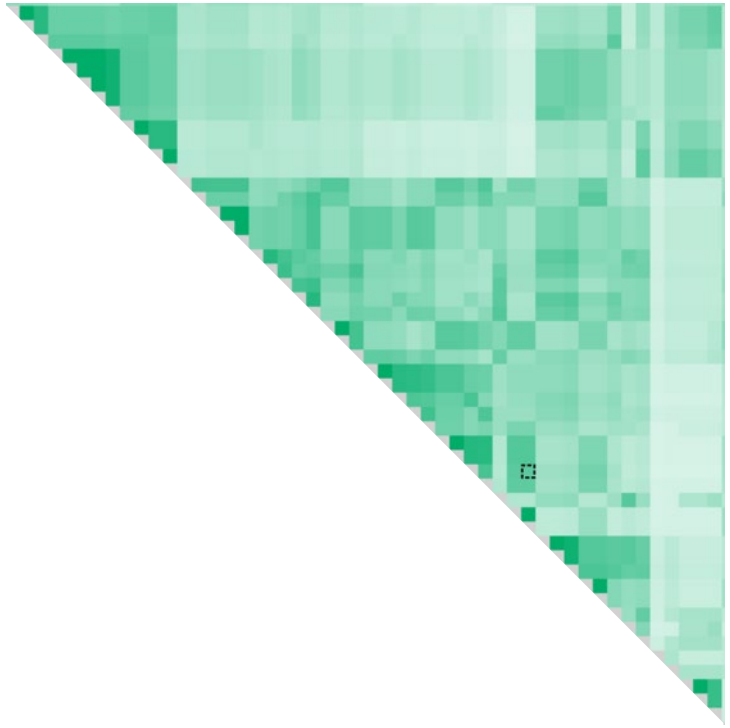
- K-Medoids clustering
  - Matrix available for KNIME

→ Usually desire a specific number of clusters

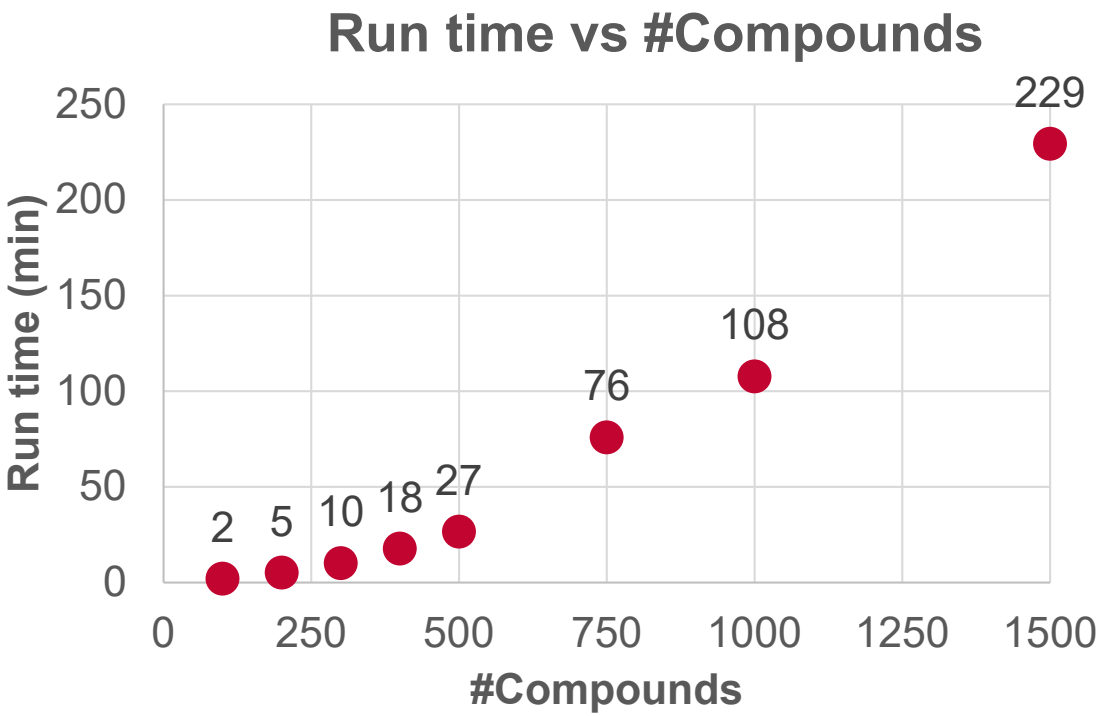
→ Obtain a specific compound as centre of each cluster

→ Clusters ordered by similarity to medoid

  → Select alternative easily



# Calculation time




- Calculation time for sets of boronic acids
- All calculations on AMD Ryzen
  - 8 core with HT
- 3D calculations significantly slower than 2D methods
- Easy integration to cluster facilitates larger calculations
  - Built in support for SGE, LSF, PBS, SLURM

# Usage

> Simple command line:

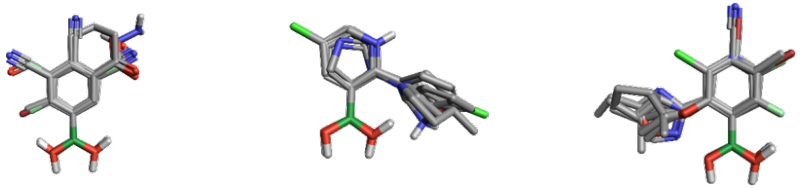
```
pickr  -s <SMARTS> <REAGENTS>.smi|sdf
```

```
pickr  -s <SMARTS> -v -Q sge <REAGENTS>.smi|sdf
```

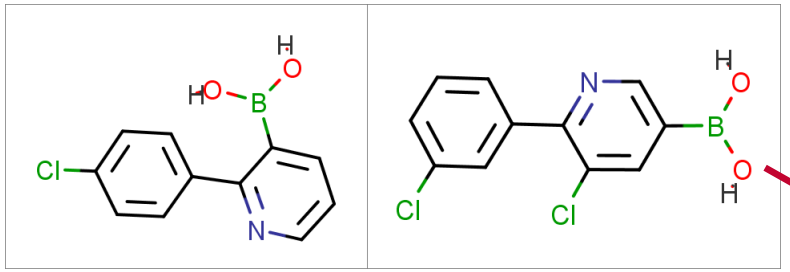
Verbose output

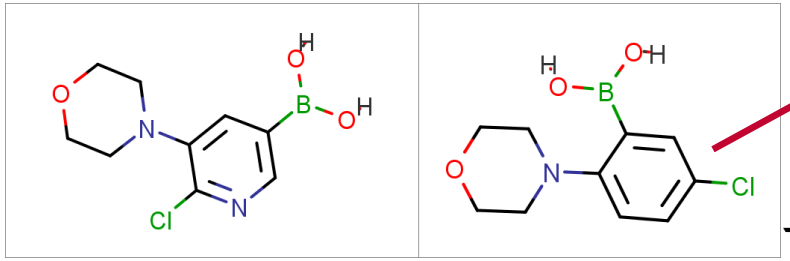
Queuing engine to use

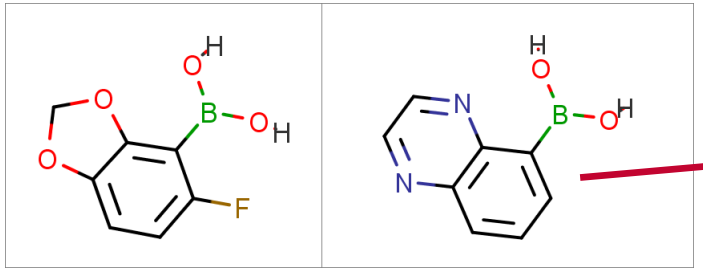
> Results presented in multiple formats
  > CSV
  > SDF single file
  > 2D SDF files of each cluster
  > 3D SDF files of each cluster containing best alignment for each molecule

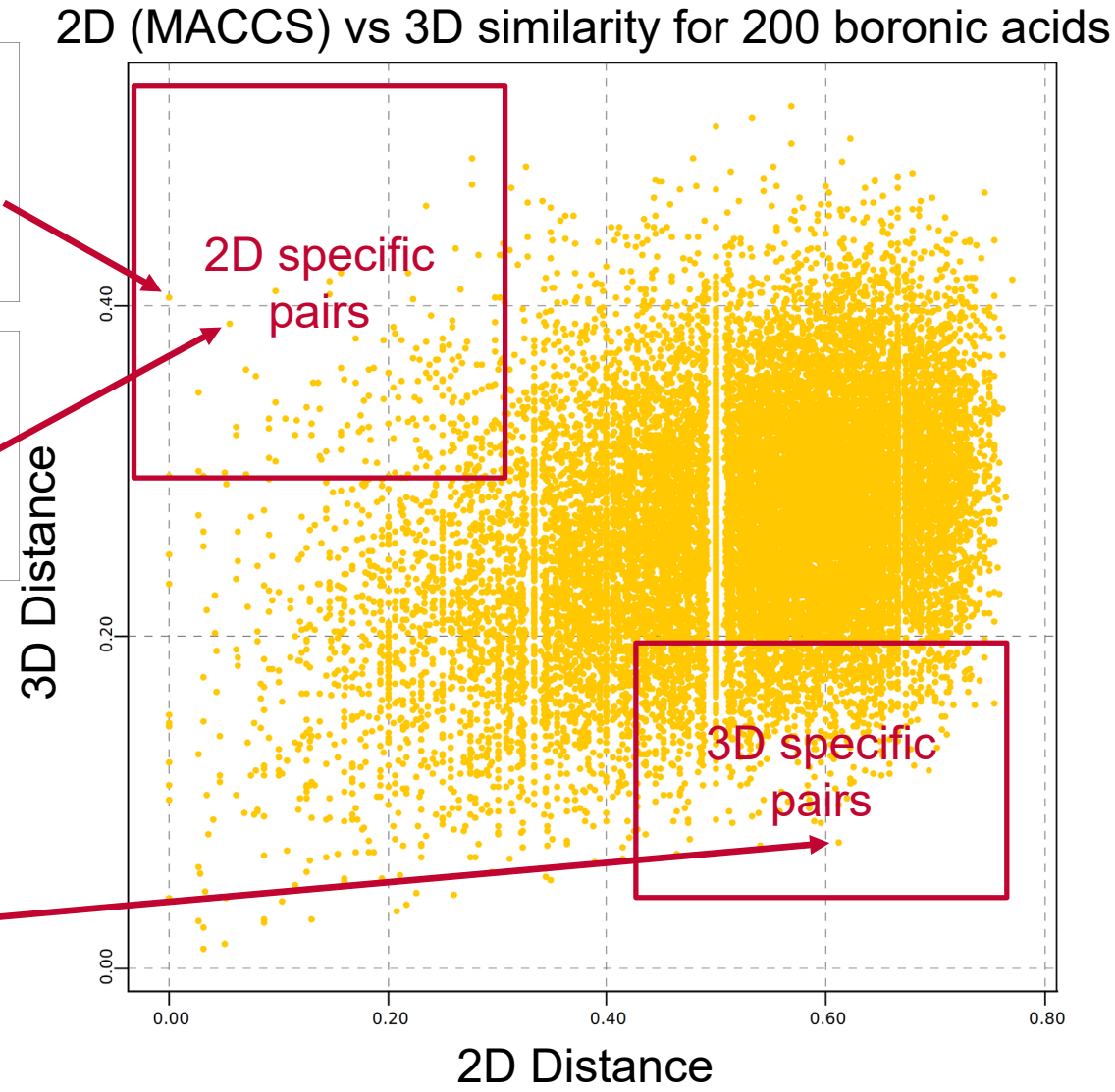


# Results – 3D descriptors generate different relationships

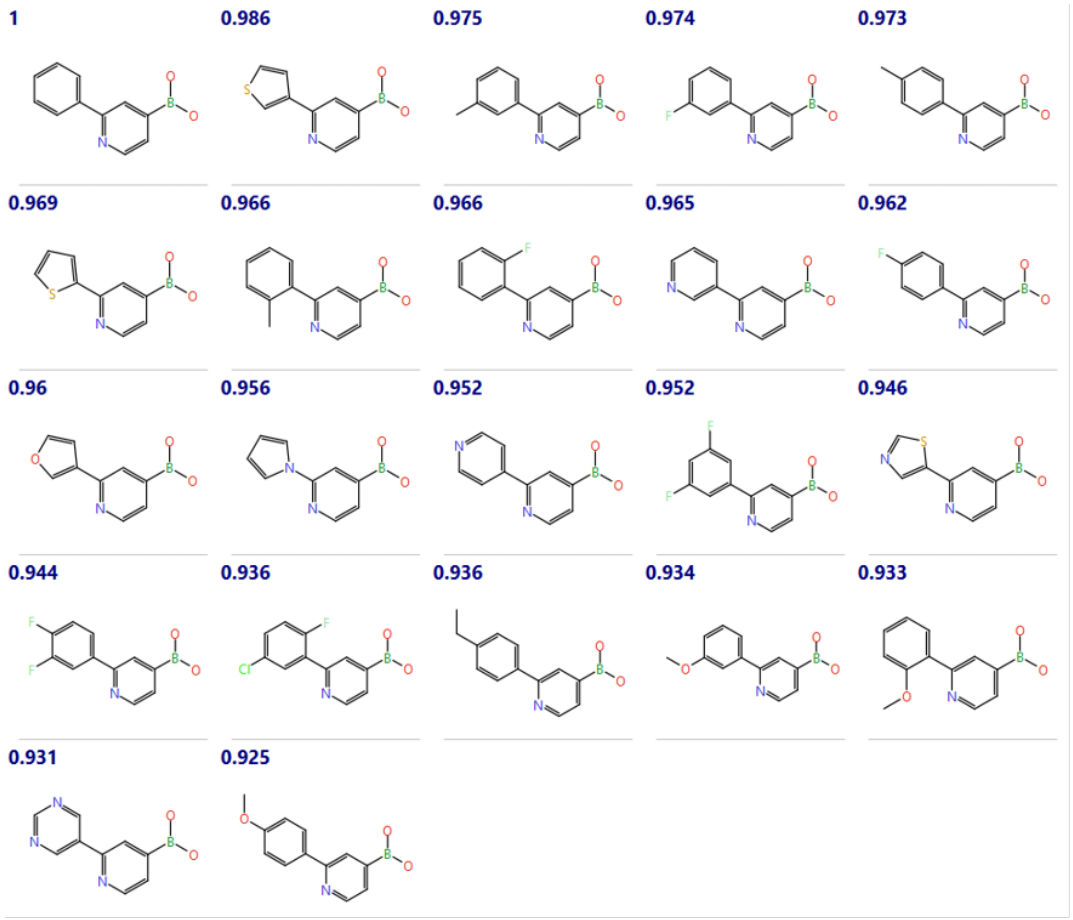




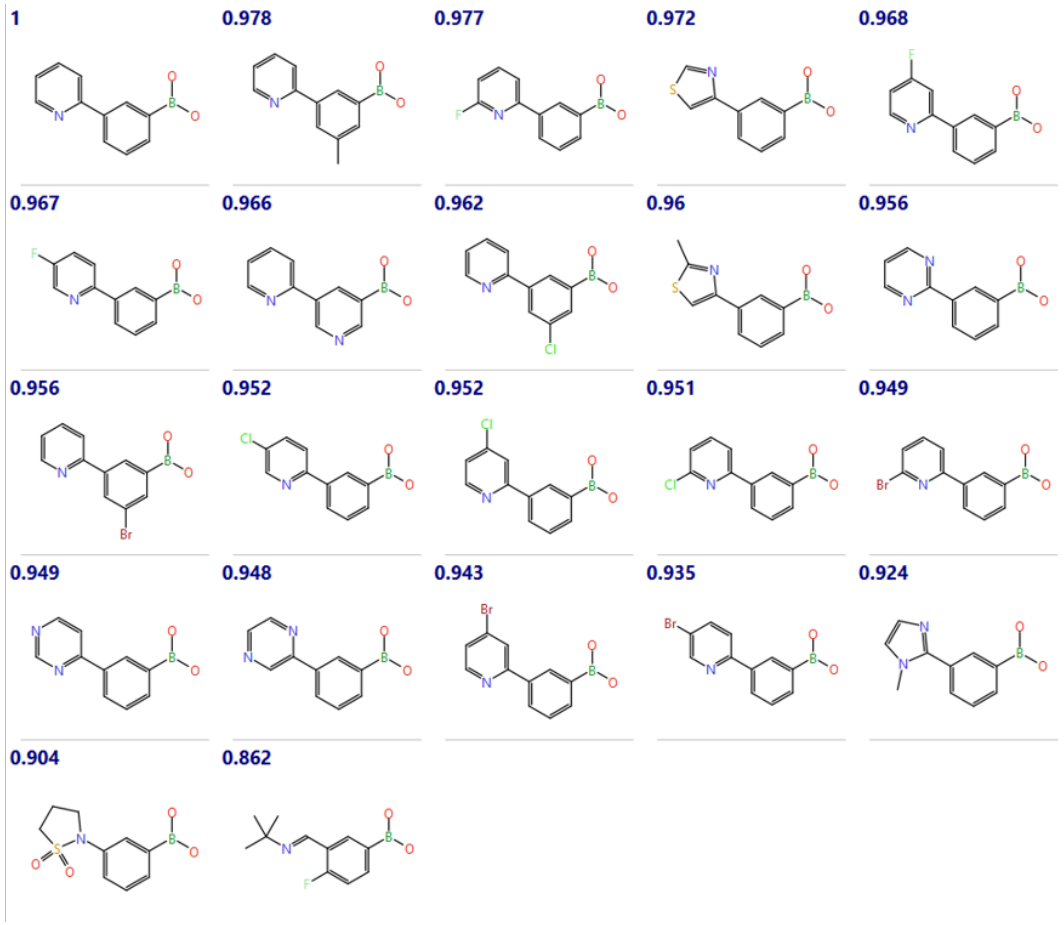




# Results fit with chemical intuition




2-Ar-4-pyridyl boronics




3-(2-H-bond acceptor-Ar)-phenyl boronics

# Case study

DOWNLOADED ALL NITRILE CONTAINING REAGENTS FROM ENAMINE

FILTER TO ENSURE THEY CONTAIN C-C#N

FILTERED FOR HEAVY ATOMS < 15

➔ 4500 MOLS

RUN PICKR

→ 450 CLUSTERS

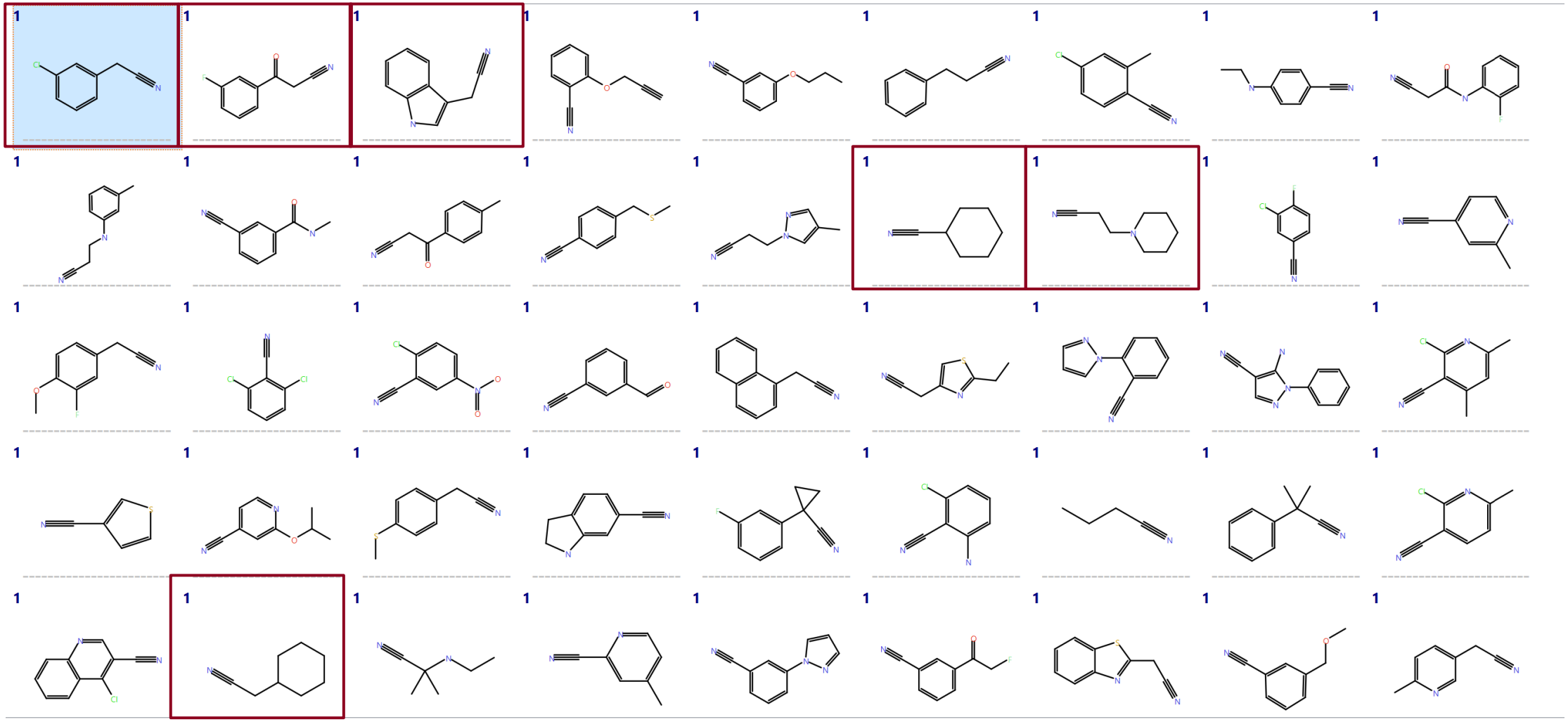
# Case Study – Enamine nitriles

```
/apps/cresset/PickR/bin/pickr -s '[N:1]#[#6:2][#6]' \
                              -v \
                              -Q sge \
                              -c \
                              -t ID \
                              Enamine_nitriles_filtered.sdf
```

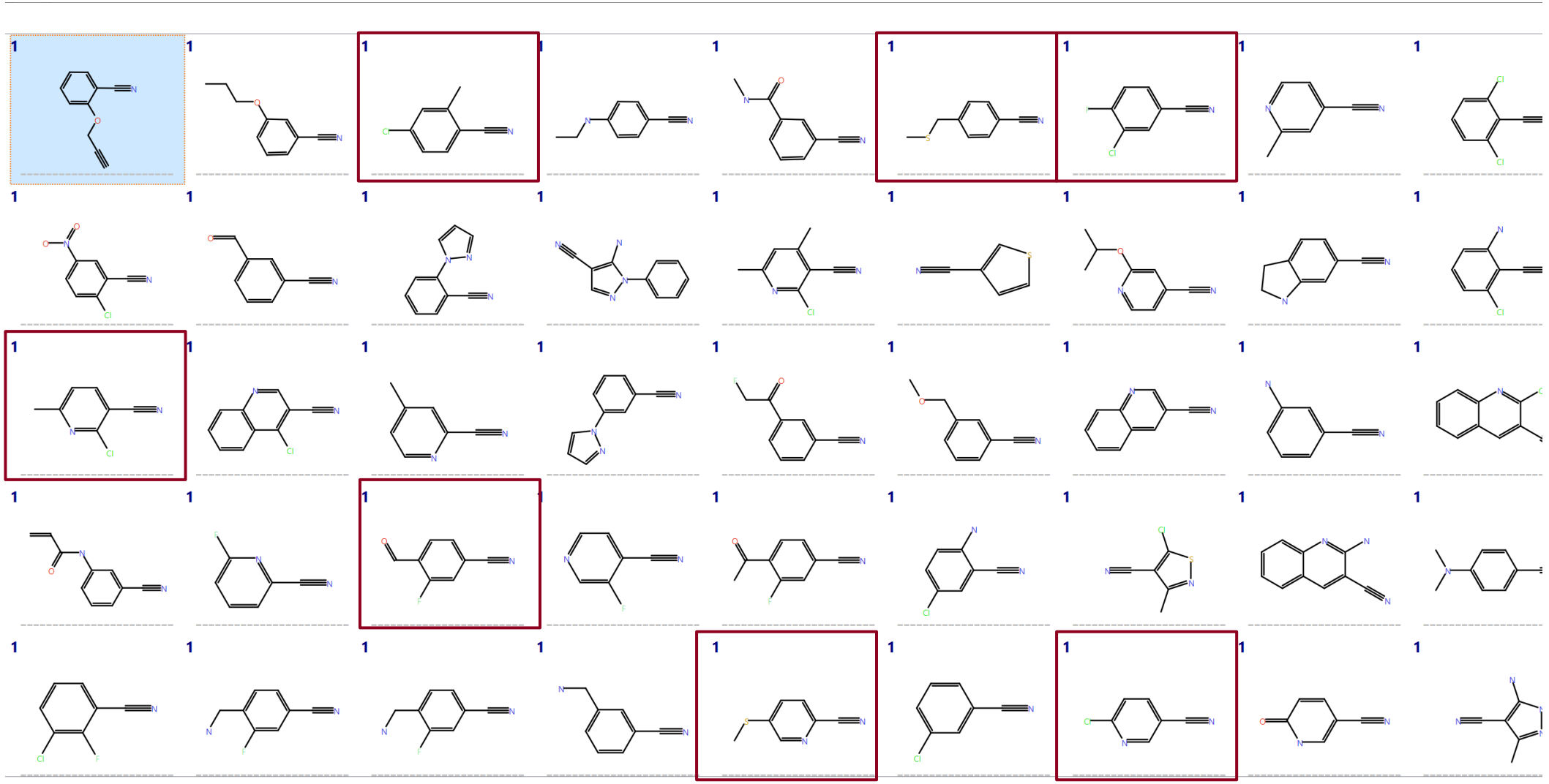
Use Cresset's charging rules

SDF tag for compound name

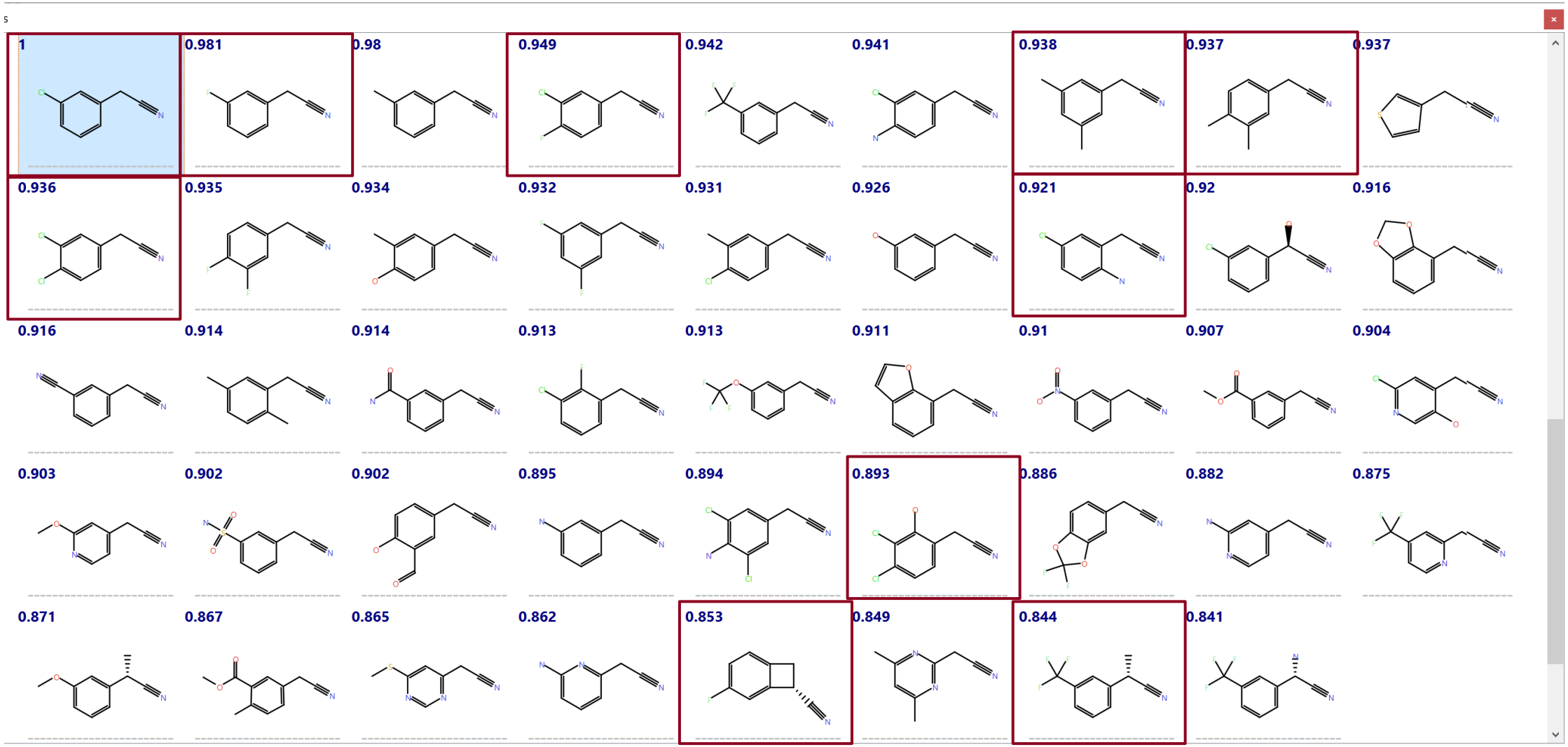
# Case study – Nitriles – medoids of largest 45 clusters



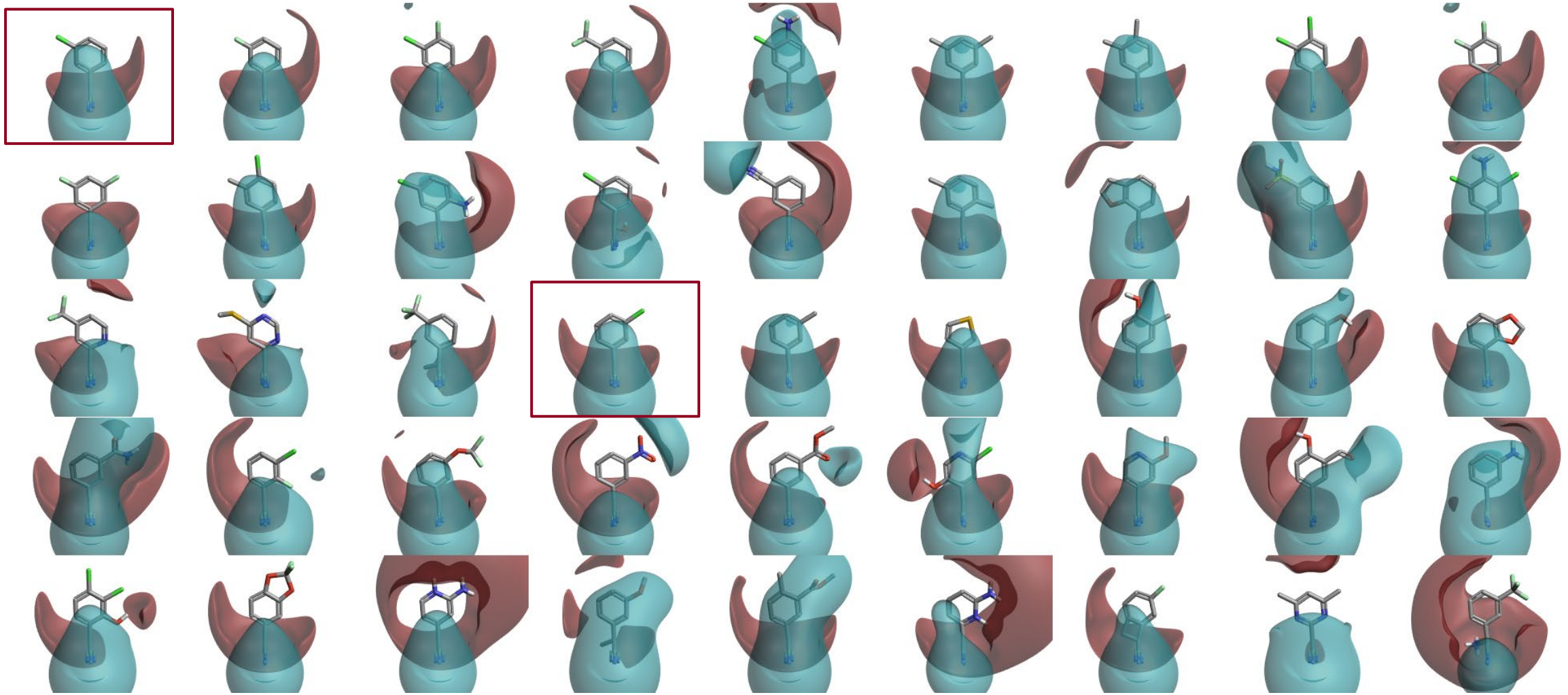
# Case study – Nitriles – 45 aromatic nitrile medoids



# Case study – Nitriles – largest cluster



# Case study – Nitriles – largest cluster – 3D



# Conclusion – Why use PickR

- Electrostatic diversity gives excellent division of functional groups
  - Fits well with chemical intuition
- Easy cluster integration offsets additional calculation times
- Combine 3D with 2D similarity for optimal space coverage
- Try it!