

# COMPOUND COLLECTION ENHANCEMENT STRATEGY

# Key steps of HTS library design

I. Design of scaffold library

II. Reshaping of the reagent library

III. Virtual coupling

IV. Filtering by physico-chemical/structural parameters

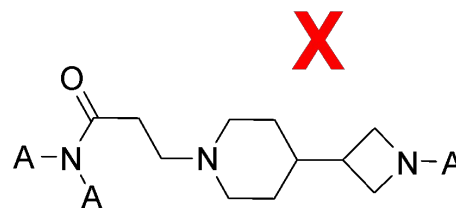
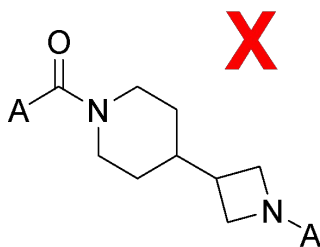
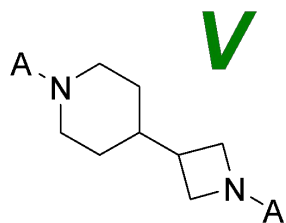
V. Checking for the sufficient novelty

VI. Control of  $F_{sp^3}$  and abundant chemotypes

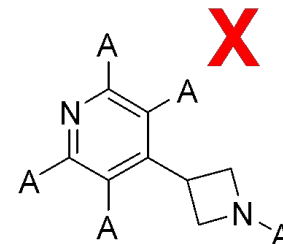
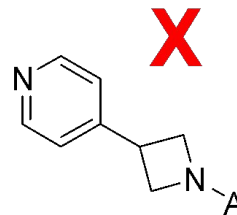
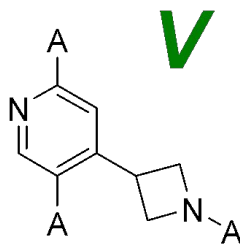
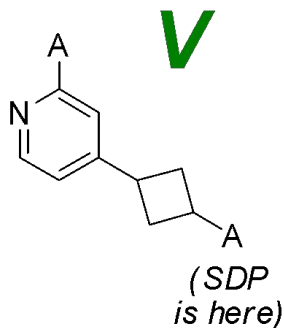
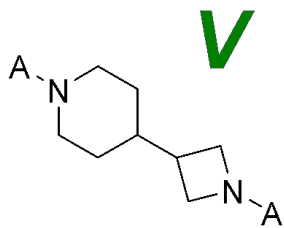
IV. Two-phase diversity selection

# I. Design of scaffold library

1. Scaffold selection is a **CRUCIAL** step in design of HTS library.
2. For a year program, ~200 scaffolds are selected (can be done in parts) – mostly proposed by chemists
3. A **chemoinformatic** definition of scaffold should be used (*i.e.* the scaffold is a combination of core ring systems, linkers combining them, and attachment points). Therefore, different side chains are assigned to the same scaffold.

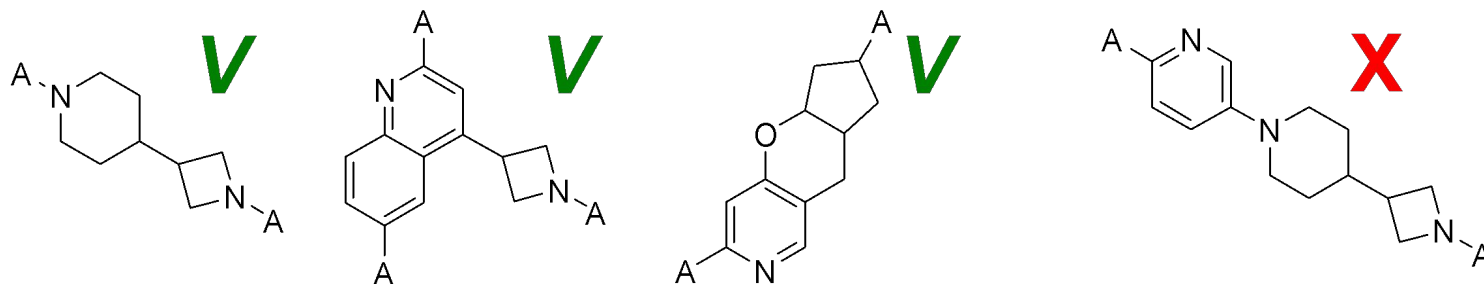


1. The scaffold should contain one strong and 1–2 weak diversity points.

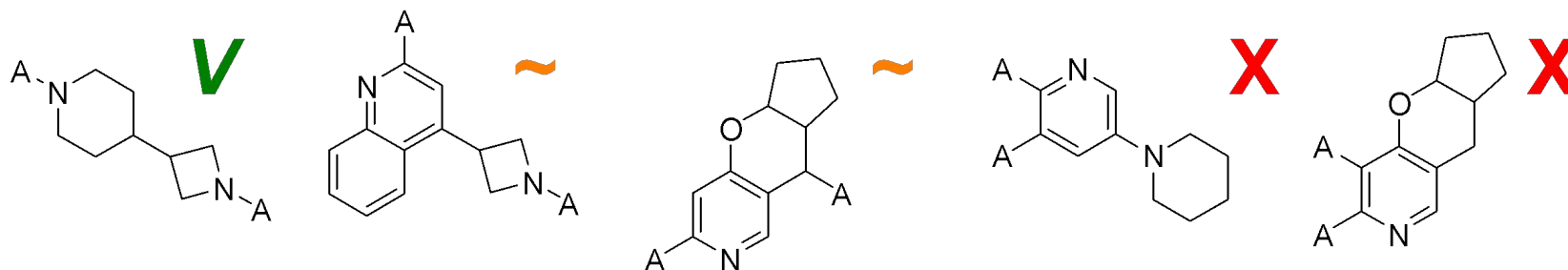


# I. Design of scaffold library

5. Scaffold should contain 2–4 (preferably 2–3) rings and 1–2 ring systems



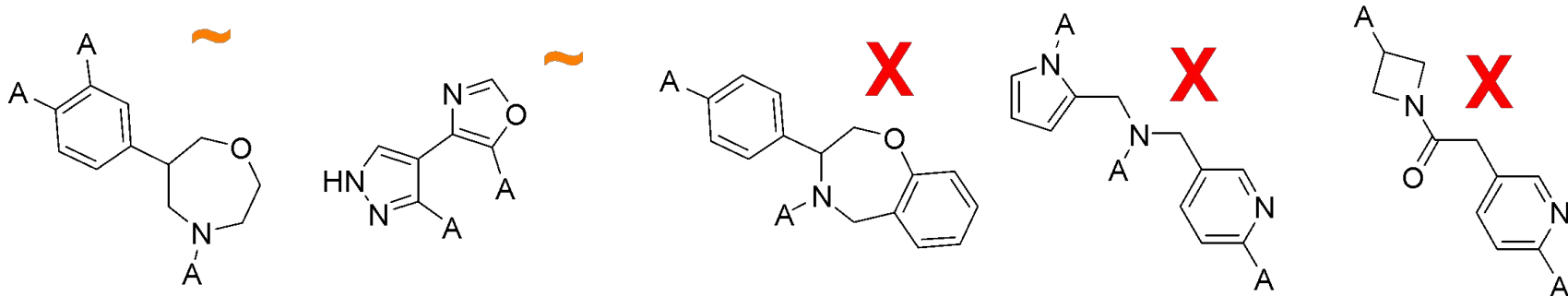
5. Diversity points should be present in every ring system, preferably in every ring. For bi- and polycyclic scaffolds with one ring system, diversity points should be present in at least two different rings, preferably distant ones.



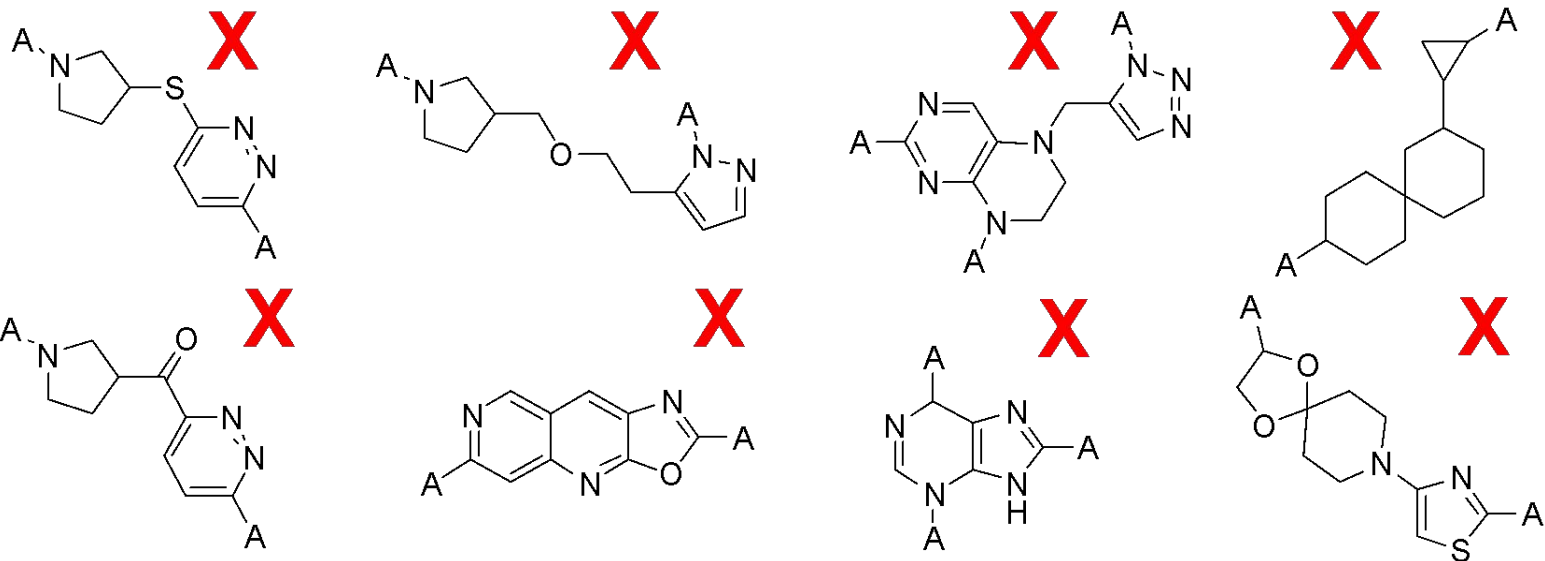
# I. Design of scaffold library

---

7. Scaffold should represent an attractive chemotype
- No bad groups, reactive groups, PAINS etc.
  - No isolated benzene rings (with rare exceptions)
  - No more than 1 benzene ring (even fused)
  - No chain amides/sulfamides/ureas/sulfides etc.
  - HAC < 20 (preferably <=15)
  - Rotatable bonds <=4 (preferably <=3)
  - Heteroatom count: 1–6 (preferably 2–5)
  - Preferably at least 1 aliphatic and 1 aromatic ring
  - Approved by MedChem visual expertize



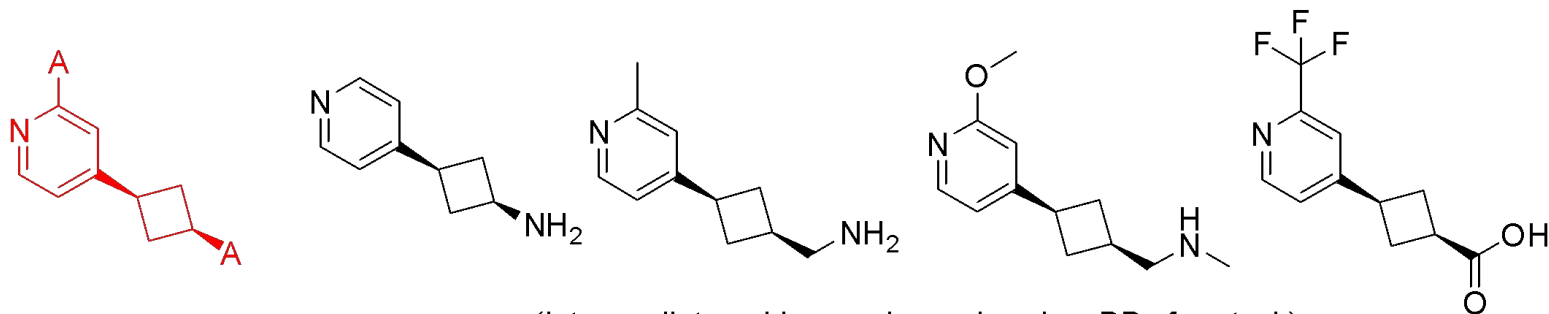
# I. Design of scaffold library



8. Scaffold should not be represented in *superior set* of LC stock (including scaffold-based drug-like compounds only) and opened orders ( $\leq 50$  cpds)
9. Scaffold should not be represented in E-molecules ( $\leq 50$  cpds,  $\leq 100$  cpds for especially attractive chemotypes – avoid “negative filtering”)
10. Scaffold should show  $< 85\%$  similarity to existing templates
11. BBs of scaffold should be prepared in 3–8 steps so that total salary of the BB synthesis per scaffold is 1200-2000 USD

# I. Enumeration of scaffold BBs

1. At least four variations are proposed for the scaffold's weak DPs (preferably 5–6, if it is possible, up to 10 variations can be included, with lower amounts of BBs ordered for synthesis).
2. Preferable substituent examples: H, Me, *i*-Pr, CF<sub>3</sub>, CF<sub>3</sub>CH<sub>2</sub>, *c*-Pr, *c*-Bu, *i*-Bu, *t*-Bu, OH, OMe, CH<sub>2</sub>OH, SO<sub>2</sub>Me and F (avoid arylation agents), C(O)NH<sub>2</sub>, THP, Py and other hetaryls, fused alicyclic rings (for 2 neighbor week DPs) etc. n-Alkyl groups should be avoided; limited use of benzene derivatives is allowed (*e.g.* fluorine-substituted).
3. If several types of BBs are possible for the same scaffold (*e.g.* amine and carboxylic acid), it is preferable to use their different variations.



## II. Reagent database

---

1. A database of at least ~800–1000 appropriate reagents is needed for such a project – at least 400 to start.
2. The following sets should be considered:
  - already available reagents (at least 25 g);
  - purchase from suppliers in China (<3\$/g; at least 100 g);
  - specially designed and synthesized (<10\$/g; at least 100 g)
3. 50% arylation / alkylation / epoxides / aldehydes; 20% acylation / sulfonylation / urea synthesis; 30% amines.
4. No more than 5 reagents derived from the same BMFL scaffold
5. The following rules should be applied:
  - Ro2 compliant;
  - <5% containing isolated benzene;
  - 1–2 rings (except the short list of 10–20 acyclic);
  - no bad groups / PAINS / reactive (after modification) etc.



# IV. Filtering by PhysChem / Structure

---

1. All available filters for bad, PAINS, reactive etc. (should not give any problems).
2. Lead-likeness PhysChem parameters:
  - MW 200...450
  - cLogP  $-1 \dots 4$
  - HDonors  $\leq 5$
  - Hacceptors  $\leq 8$
  - RotatableBonds  $\leq 6$
  - Molecular complexity by Wyeth  $> 20$
  - TPSA  $< 140$
  - LogS  $> -6$
  - No too saturated without H-donor
3. Less than 2 benzenes,  $\leq 2$  S,  $\leq 2$  Cl, Br,  $\leq 5$  F; no chains longer than 6 bonds; at least 2 heteroatoms.
4. Additional structural filters depending on particular set.

## V. Filtering by Novelty

---

1. 98% Tanimoto diversity to competitors (E-molecules).
2. 95% Tanimoto diversity to Advanced HTS collection and opened orders.

## VI. Chemotype control

1. Checking for  $F_{sp^3}$  distribution (perfect  $F_{sp^3} = 0.6-0.7$ ).
2. Checking for amide/sulfonamide/urea % (<30%).

## VII. Final diversity selection

1. <300–400 cpds per scaffold (the rest is removed by Tanimoto).
2. Check for BMFL scaffold distribution ( $\leq 30$  compounds)

# Final remarks

---

1. A key feature is separation of scaffold synthesis and combinatorial step.
2. ~100 g of building blocks per scaffold should be ordered for synthesis (+extra amount for stock), *e. g.* 3×40 g, 4×30 g, 5×25 g or 10×15 g – in the latter case, more compounds and more BBs can be ordered, so that budget for BBs can be somewhat increased.
3. Reactivity rules for virtual coupling should be developed.