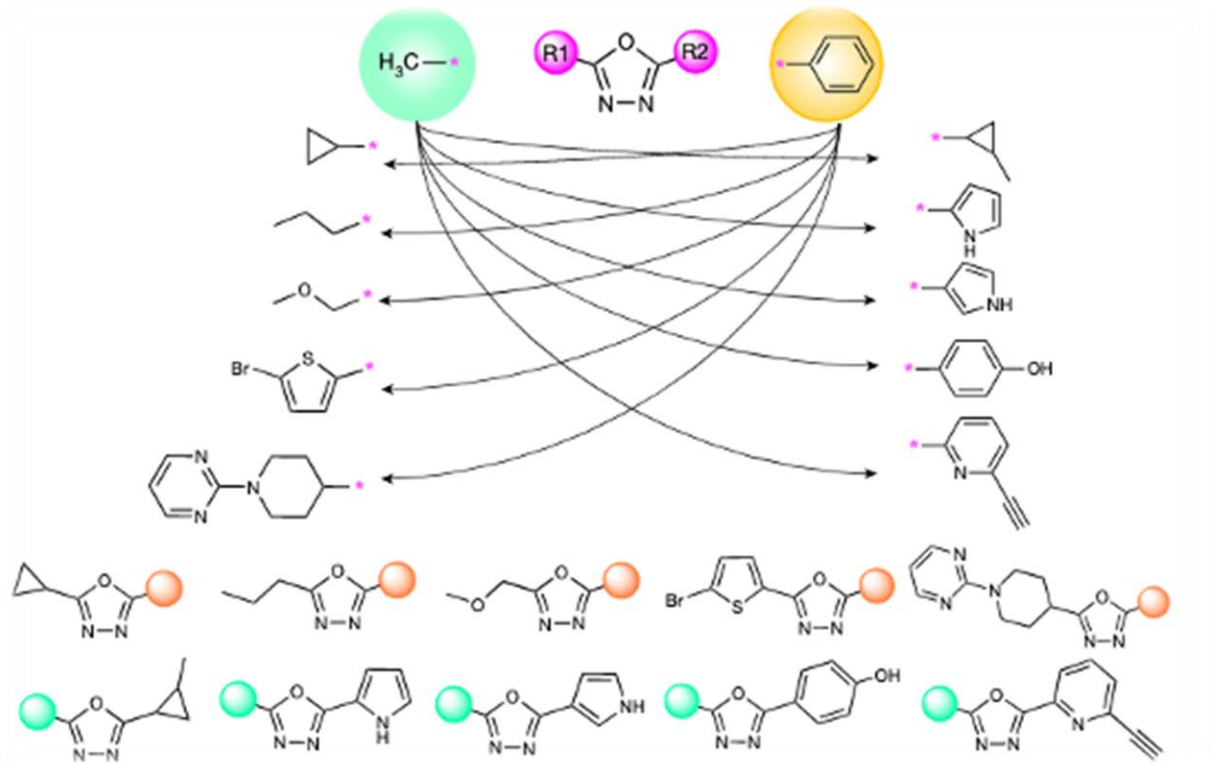
Expanding ChemDiv compound collections with Virtual Chemical Space technology

molsoft
molecules *in silico*

How we define the Chemical Space ™ : Parts and Rules

- The Chemical Space is defined by Molsoft as a compact representation of billions to trillions of compounds by parts and rules and without listing each of them explicitly.
- The Space is built upon ChemDiv's commercially available anchors, building blocks and cores using MolSoft's ICM cheminformatics tools.
- Molsoft developed tools to search or 3D-dock-screen the Chem Space to extract catalogs or target guided hit lists.
- The Space has a unique discovery potential
  - It ensures novelty
  - It may avoid known liabilities
  - Covers of trillions of synthesizable compounds

Generating Chemical Space in MolSoft ICM

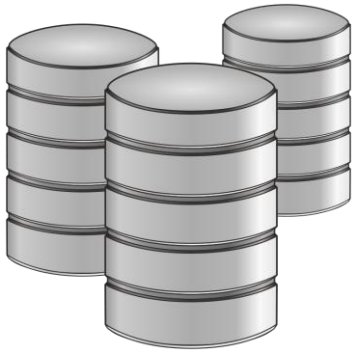


A library of fragment-like anchors representing all possible combinations for all reactions in the entire.

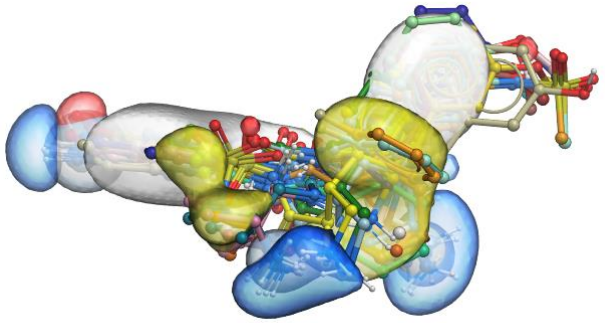
The chemical space was generated using:

- **Rule-based enumeration**: Combines multiple chemical reactions into simpler rules (e.g., Suzuki, amide coupling) with anchor libraries.
- **Markush combinatorial expansion**: Defined core structures and variable R-groups.
- **Custom Catalogs** Generate very large catalogs with specific properties from the Chemical Spaces.

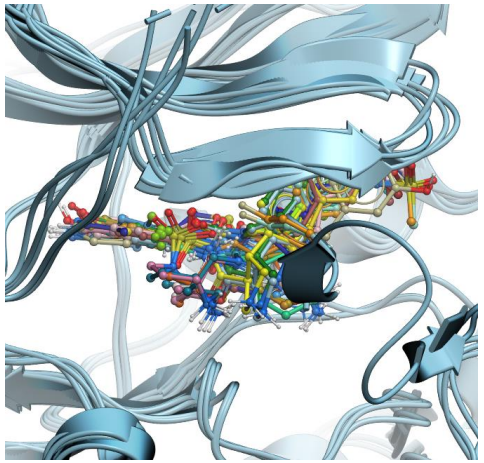
GPU and AI Enhanced Fast Highly Accurate Chemical Space Screening Methods in ICM



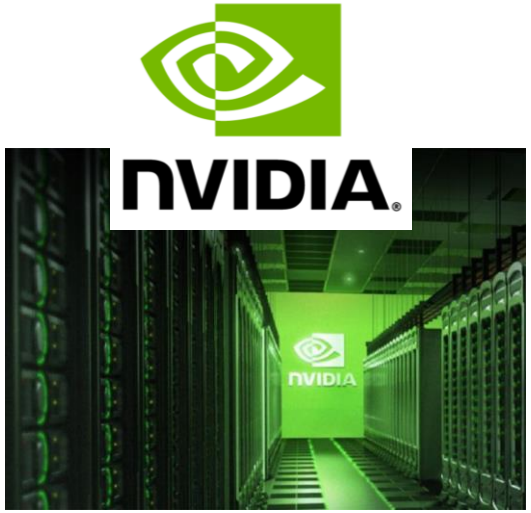
Chemical Space Data



Ligand-based: RIDE

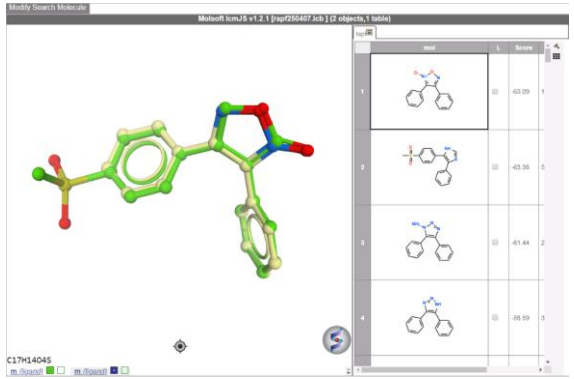


Structure-based: RIDGE and CombiRIDGE



GPU Enhanced Screening

Scoring and Evaluation



Lead Candidates – top scoring hitlist of 1-10K compounds from the Space.

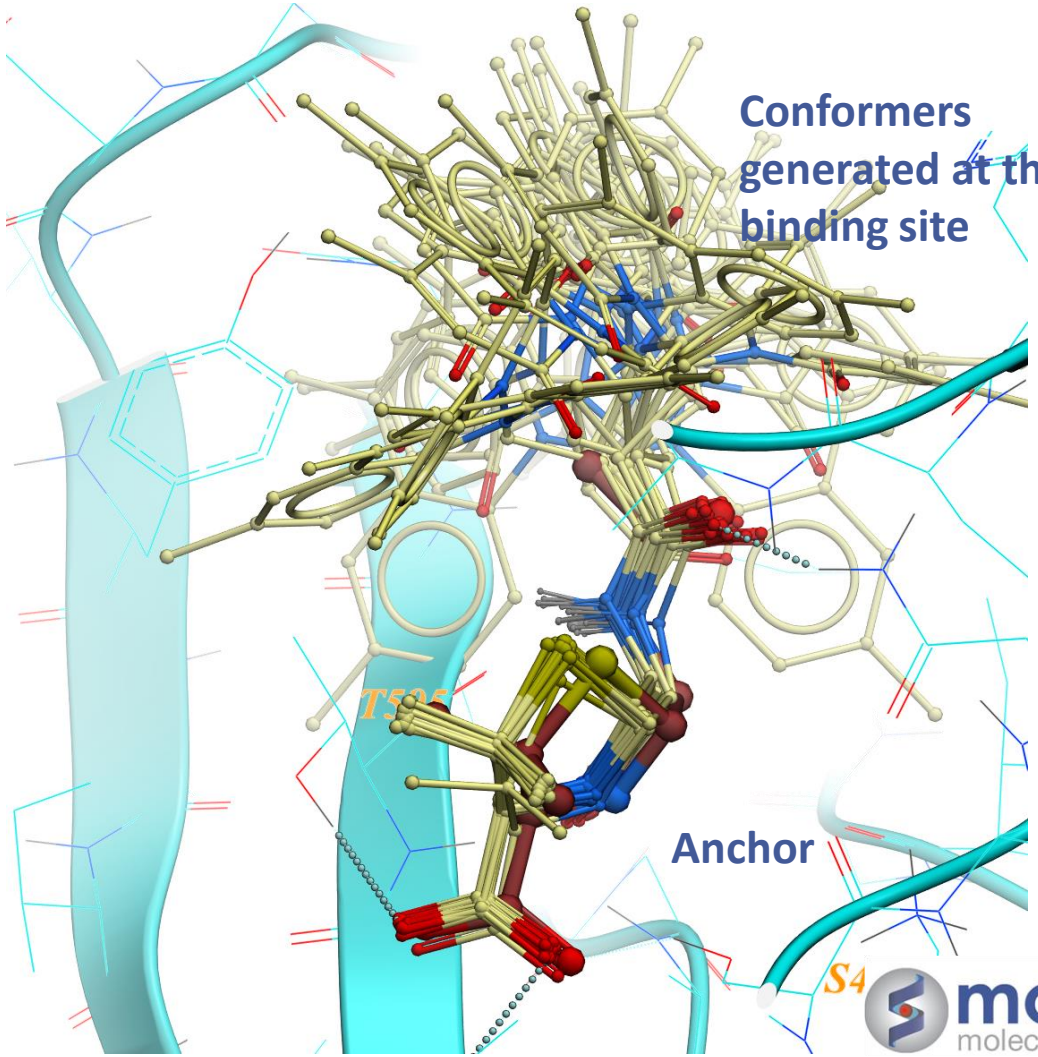
CombiRIDGE – Ultrafast GPU Accelerated 3D - Docking Compound Generation using the Space Parts and Rules

- *In-situ* GINGER - conformer generation directly at the binding site:
  - start with 3D anchor/core fragment + 2D R-group
  - GINGER generates full 3D conformers while keeping anchor/core in place
- GPU grid/MMFF Cartesian minimization refines poses
- RTCNN scoring

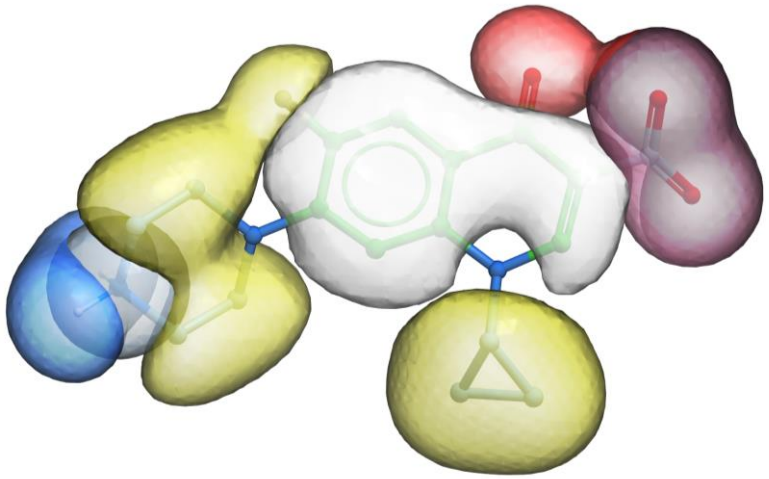
All stages GPU-accelerated

Applications:

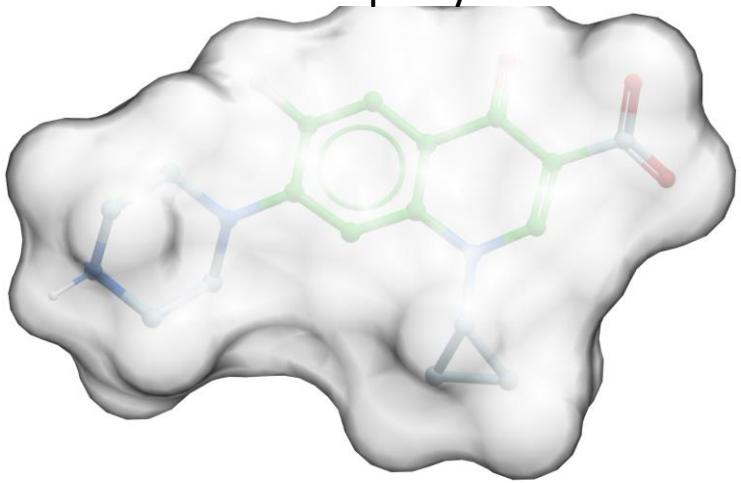
- R-group screening
- Covalent docking
- Combinatorial docking



Ligand Guided GPU Accelerated Screen through Ultra Large Libraries RIDE – Rapid Isostere Discovery Engine



Atomic Property Fields

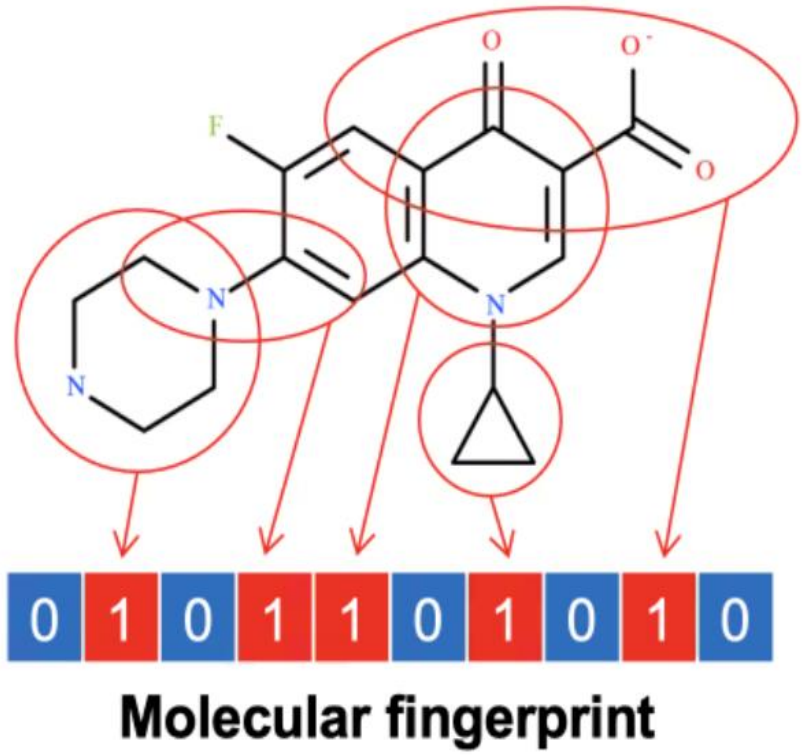


Shape Matching and Excluded Volume

- **RIDE** is a fast 3D molecular similarity search method based on APF. RIDE searches databases of compound conformers for molecules that are isosteric to the query.
- Applications include – virtual screening, scaffold hopping, hit follow up and core replacement.
- Atom Weighting - Contributions of different portions of the molecule can be modulated with per-atom weights to reflect relative importance of certain moieties.
- Excluded Volumes meet Shape Matching - An envelope penalty can be applied to the regions that surround all or part of the query molecule to prioritize hits without bulky extensions in constrained regions.

Screen 1.5M chemicals/sec/GPU on RTX 4090

2D Chemical Fingerprint and Substructure Search using MolSoft's MolCart GigaSearch



- MolSoft's **MolCart GigaSearch Engine** enables you to perform substructure and fingerprint search of the chemical space in seconds.
- The method adds fingerprint bit statistics to the MolCart search engine which allows extremely fast and efficient way of filtering out molecules based on the input chemical pattern.
- It also provides a new efficient way of storing chemical fingerprints to minimize the amount of data to be scanned on server side.