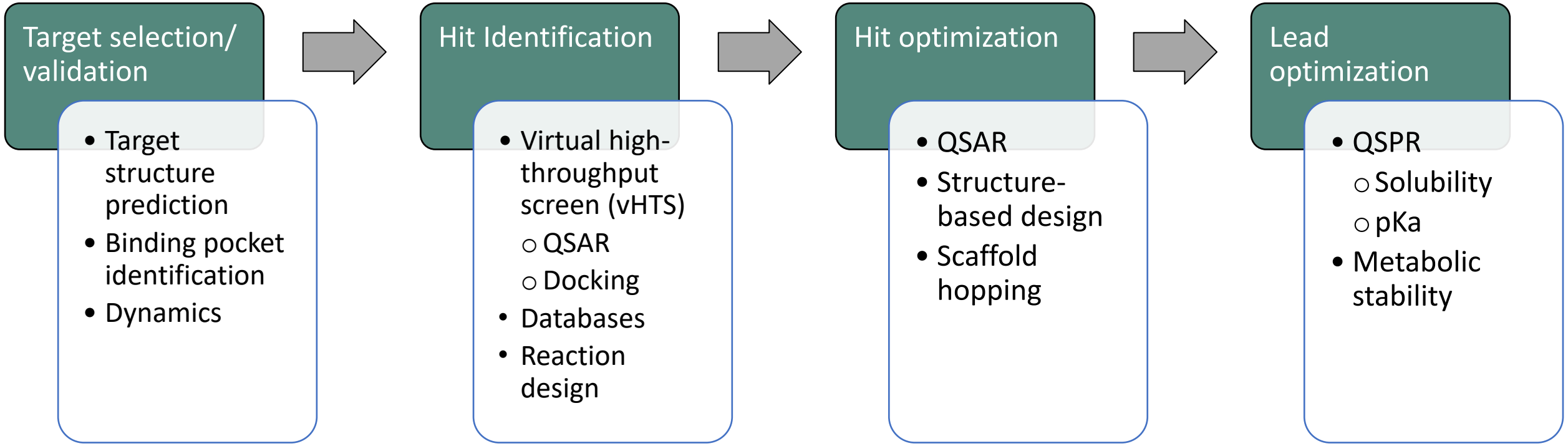
Small molecule processing, handling, and miscellaneous tasks

BCL-Rosetta Drug Design Workshop

January 2022

Computer-aided drug design can increase the efficiency and efficacy of multiple parts of the drug design process



In reality, this is often not linear. This is a cyclic, iterative process.

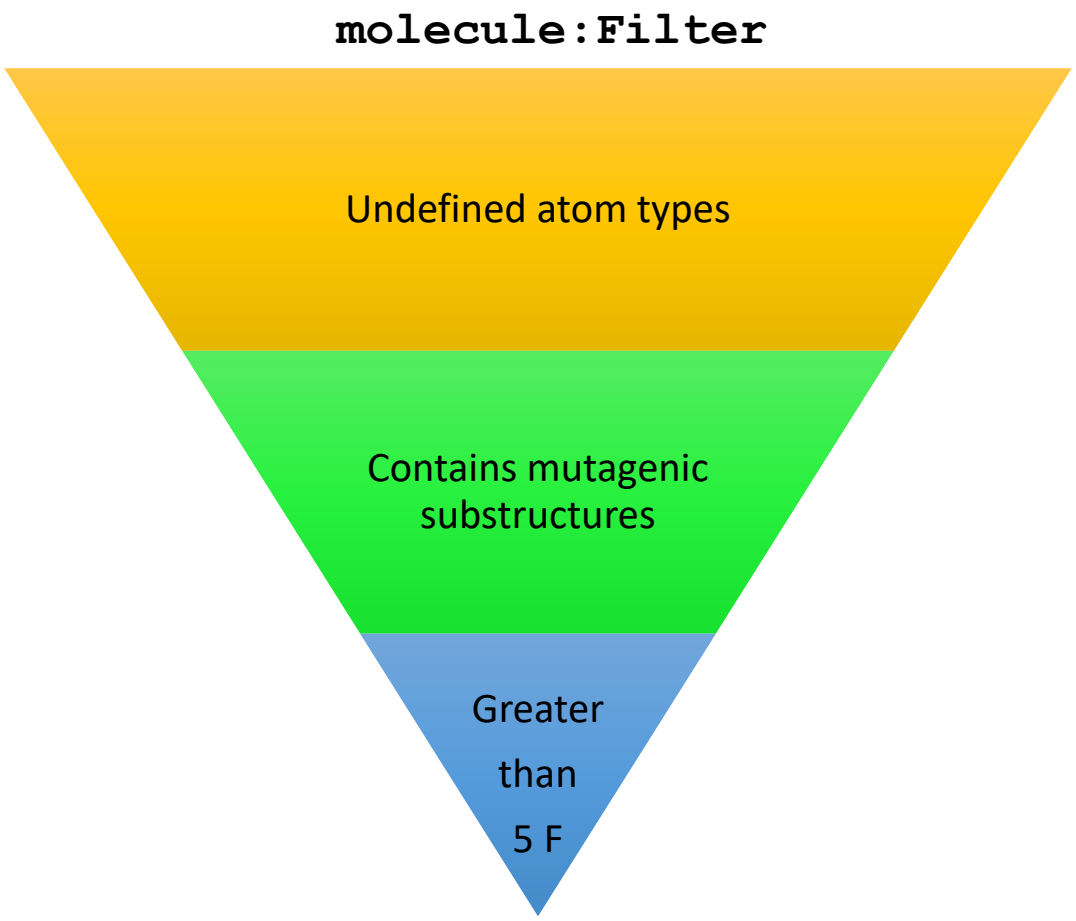
There are some practical considerations that need to made before one can build models for drug discovery

- Valid atom types / connectivities
- Impurities in molecule data files
- Filtering fragments / molecules that are not helpful
- High quality 3D structures
- Data labels
- Duplicates
- Orientation in 3D space
- Comparisons to other molecules in dataset
- Ability to compute / extract features for model development

Most of the tools that we need for day-to-day molecule handling are in the `molecule` application group

Application	Purpose
Filter	Select molecules by property (e.g., predicted activity > 5), substructures present, or filter out molecules that are chemically or conformationally invalid
Unique	Select unique molecules in an ensemble at varying levels of substructure resolution
Split	Split out parts from molecules – rings, chains, Murcko scaffold, etc. – from the input structure
Reorder	Sort input molecules or atoms within the molecules
Compare	Compare molecules by substructure, RMSD, properties, or other customizable metrics
Properties	Compute numeric/string properties from input molecules, e.g., QSAR predictions, partial charges, etc.

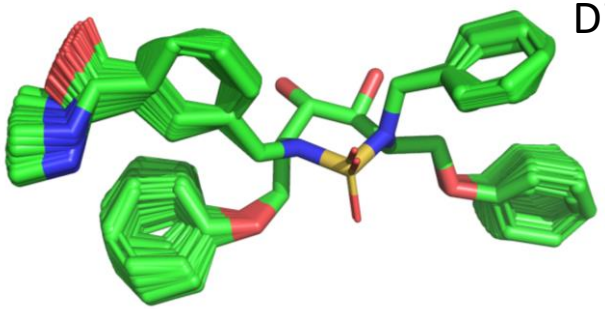
For example, we can remove molecules from our datasets that undesired properties or substructures



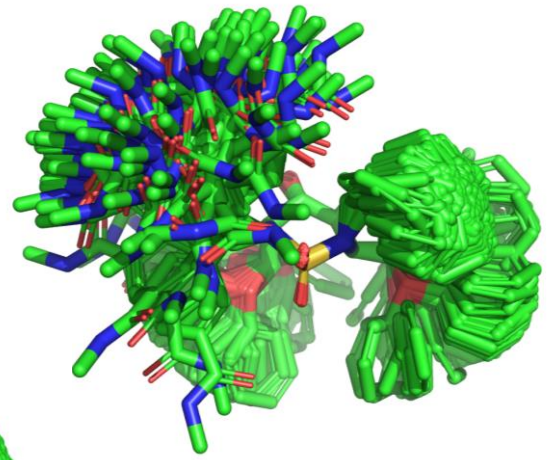
- ✓ Contains existing presets
- ✓ Allows user definitions
- ✓ Extensible

We can generate high-quality 3D conformers of small molecules for use with docking, alignment, and other methods

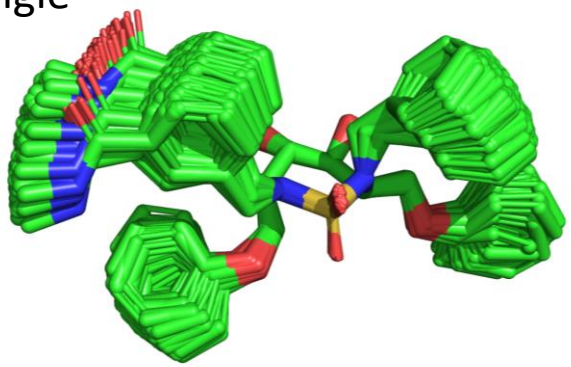
Wiggle only



Wiggle &
Bond Angle &
Dihedral Angle



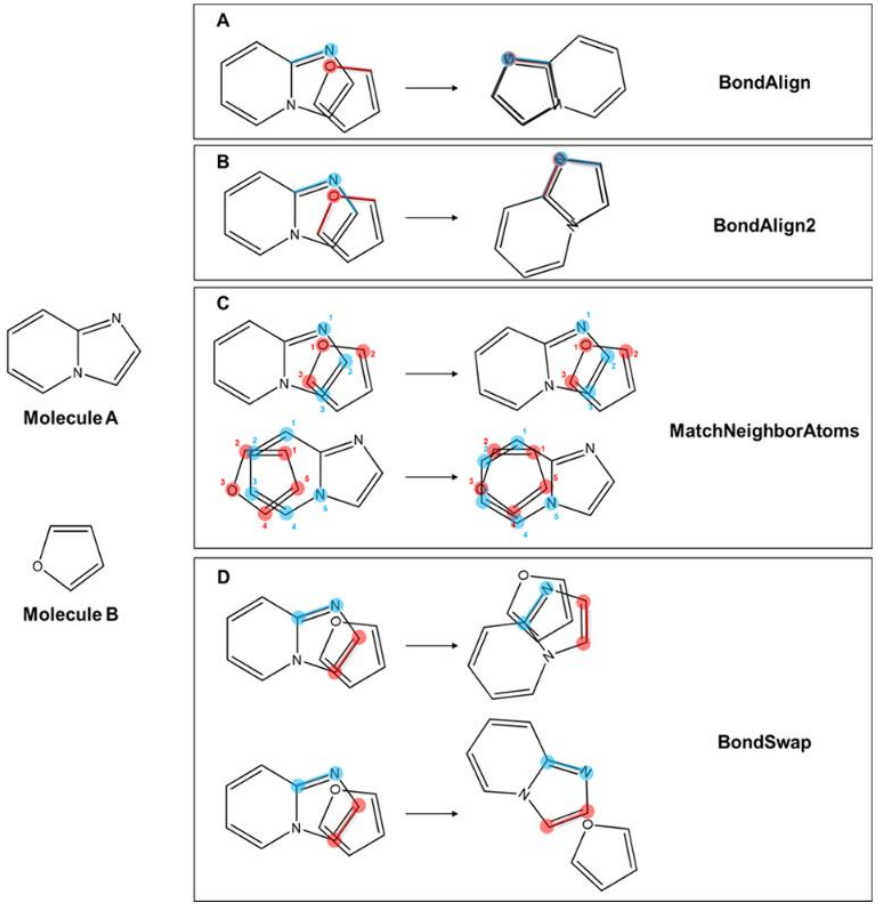
Wiggle &
Bond Angle

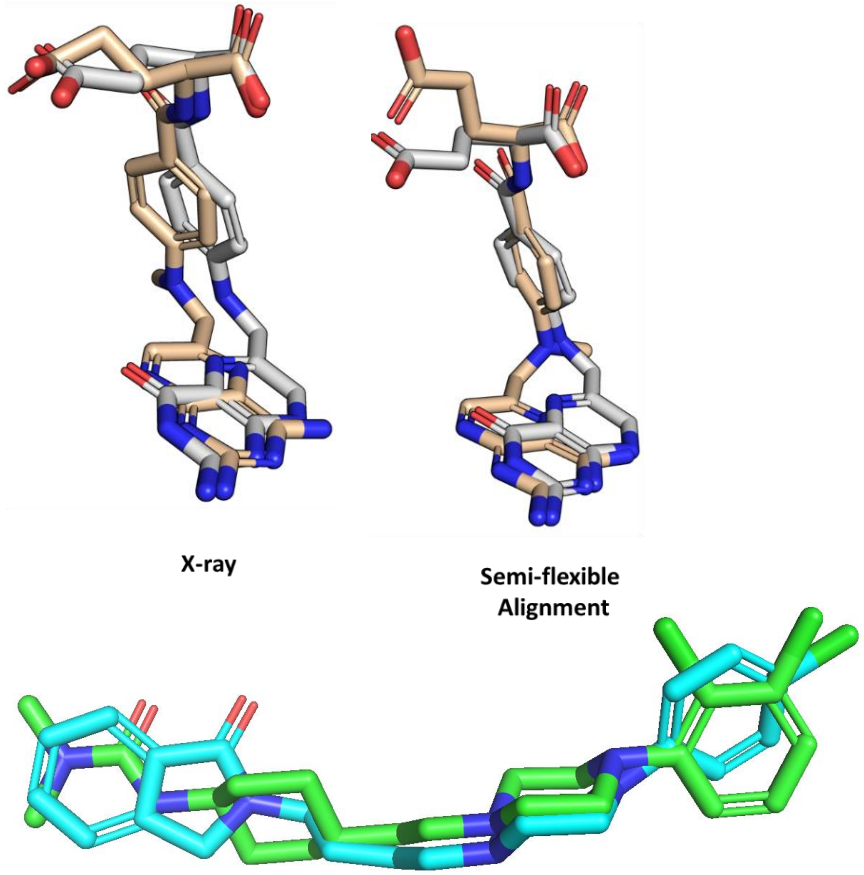


Maximum ensemble size	Mean 10	Median 10	Mean 50	Median 50	Mean 250	Median 250
Balloon DG	1.10	0.97	1.00	0.86	0.92	0.77
Balloon GA	1.22	1.10	0.90	0.80	0.72	0.63
RDKit DG	1.00	0.89	0.77	0.64	0.63	0.52
ETKDG	0.98	0.87	0.77	0.66	0.63	0.54
Confab [c]	0.81	0.70	0.72	0.61	0.65	0.53
Frog [c]	1.18	1.19	0.93	0.85	0.75	0.65
Multiconf-DOCK	0.99	0.89	0.84	0.72	0.80	0.69
Conformator	NA	NA	0.68	0.58	0.57	0.47
CSD [a,d]	0.85	0.69	0.65	0.52	0.53	0.44
ConfGenX OPSL3 forcefield [a]	NA	NA	0.63	0.52	0.54	0.44
Omega [b]	NA	NA	0.67	0.51	0.57	0.46
Excalc [a]	NA	NA	0.87	0.77	0.73	0.61
iCon [a]	NA	NA	0.72	0.53	0.60	0.47
MOE Stochastic w/clustering [a]	NA	NA	0.75	0.55	0.64	0.52
MOE LowModeMD w/clustering [a]	NA	NA	0.75	0.54	0.62	0.50
MOE Import [a]	NA	NA	0.90	0.79	0.65	0.56
BCL::Conf2016 [a,c,d]	0.96	0.87	0.76	0.68	0.68	0.60
BCL::Conf [b,d]	0.76	0.68	0.57	0.49	0.46	0.36

Mendenhall et al. 2020

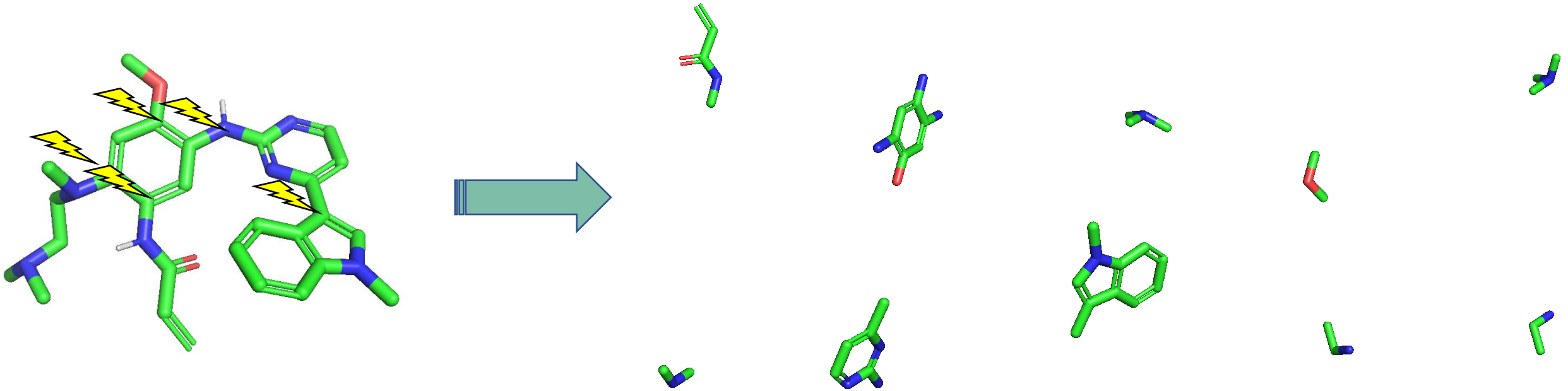
We can perform small molecule substructure- or property-based flexible alignments





Brown et al. 2019

We can split molecules into fragments that can subsequently be used for drug design, filtering, or other protocols.



This is a small sampling of tools available in the BCL

Get ready to learn more!