Atomwise

# A virtual high-throughput screening pipeline for covalent inhibitors

August 23rd, 2021 – ACS Fall 2021

Contributors: Srimukh Prasad Veccham, Adrian Morrison, Pawel Gniewek, Saulo de Oliveira, Venkatesh Mysore, Henry van den Bedem

**Srimukh Prasad Veccham**

Cheminformatics Scientist

Atomwise

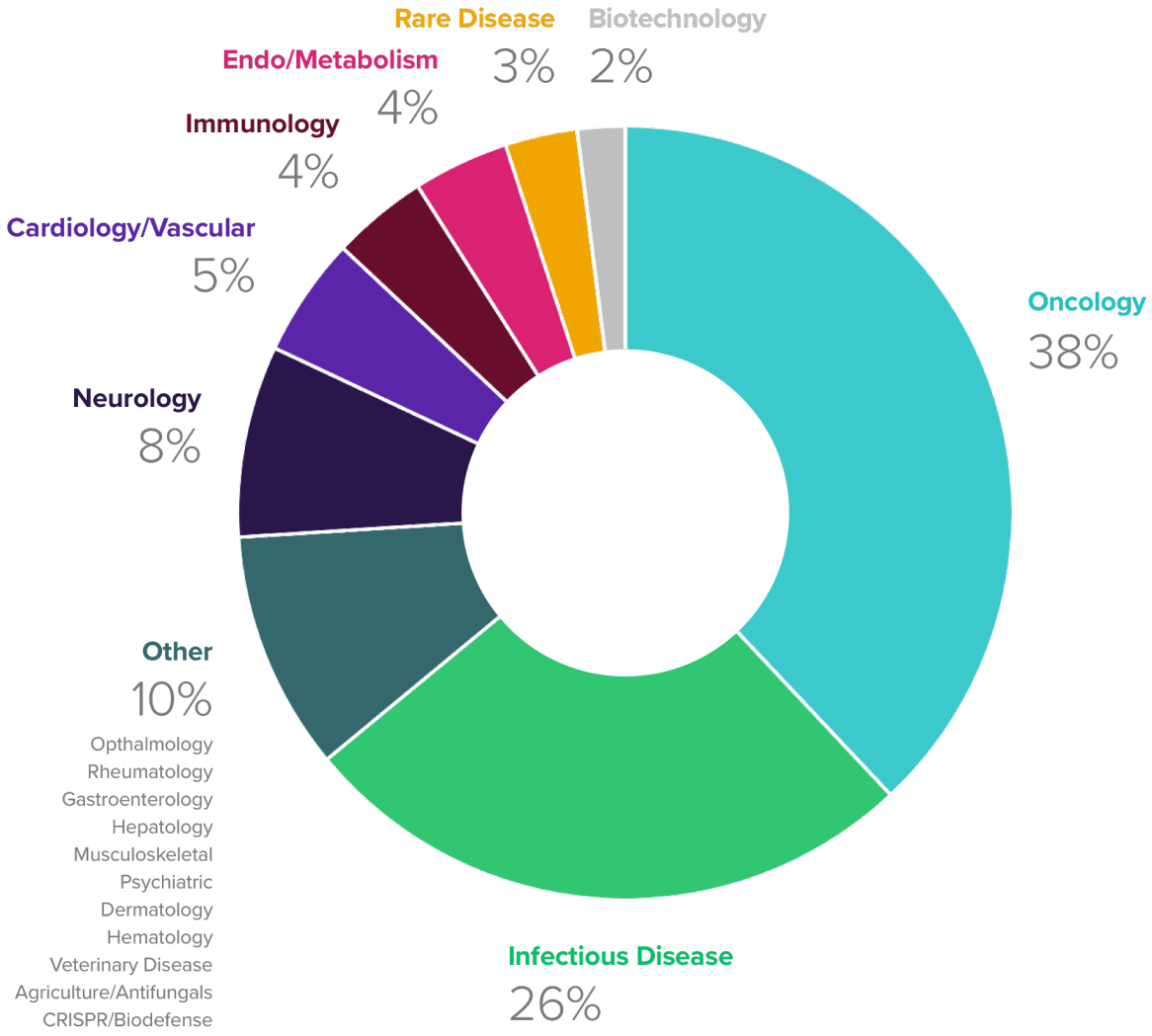
# Agenda

- Covalent inhibitors
- Screening with AtomNet® model
- Covalent adaptation of AtomNet® model training protocol

# Atomwise

"Drug The Undruggable"

- Pharmaceutical company using machine learning for drug discovery
- Wide and diverse portfolio of internal and joint venture assets
- Developing innovative strategies to target undruggable genome
- Covalent inhibition is one such strategy

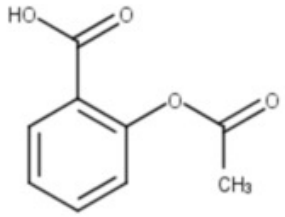


# Covalent Inhibitors

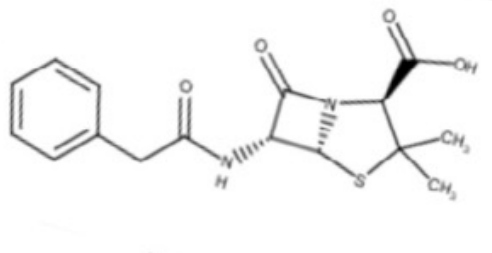
## Covalent Inhibitors form covalent bonds with targets

- Covalent inhibitors act by forming covalent bonds with targets
- Traditionally under-explored by pharmaceutical companies due to concerns surrounding off-target toxicity
- Historically covalent inhibitors were discovered serendipitously
- 14 covalent drugs approved by US FDA in the last decade

**Aspirin**



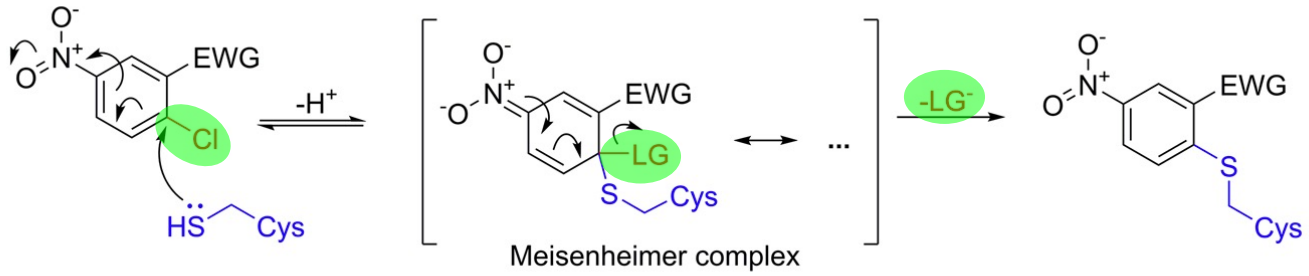

**Penicillin**

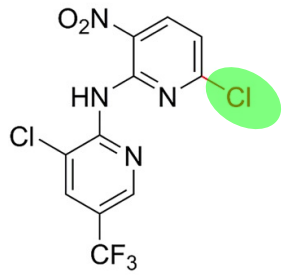


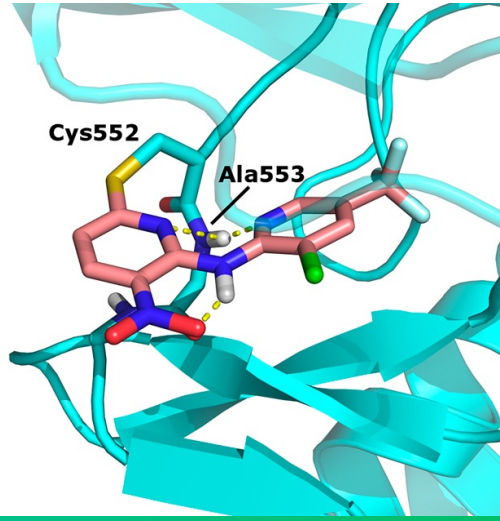
# Covalent Inhibitors

Warhead, an electrophilic group on inhibitor, reacts with nucleophilic residues

- Inhibitors usually have an electrophilic group with reacts with nucleophilic amino acid residues
- Electrophilic group on inhibitor is called a warhead
- Identity of warhead decides
  - Mechanism of reaction: reversible covalent vs. irreversible covalent
  - Reactivity and selectivity

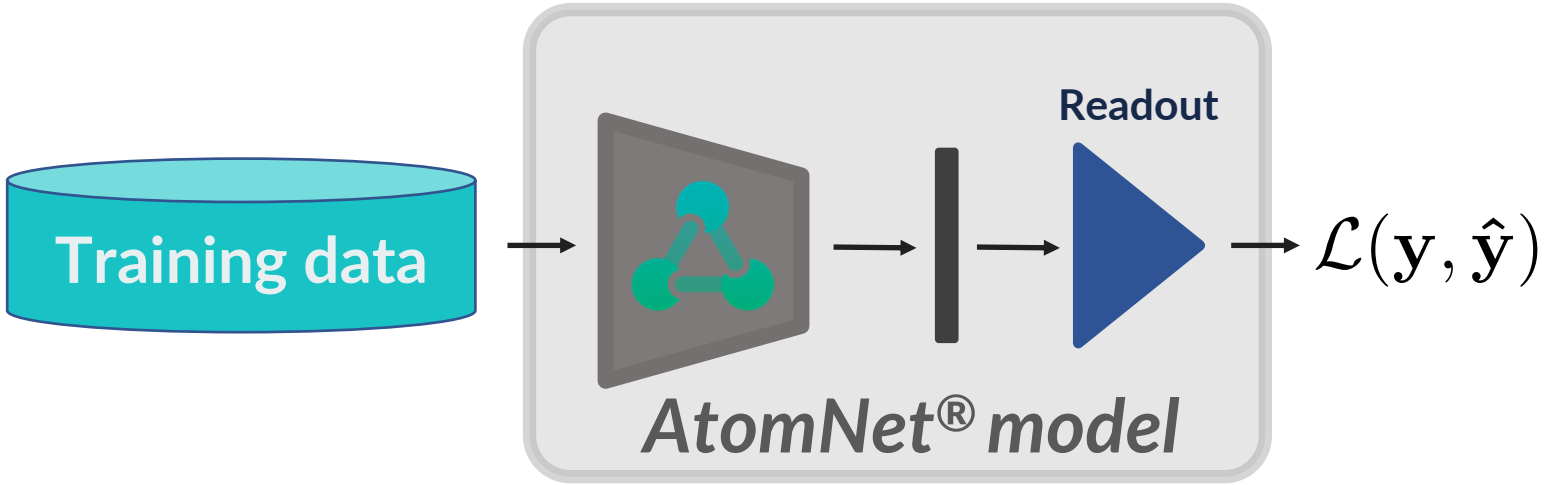



Picture credits: Gehringer, M. et al *J. Med. Chem* **2019**, *62*, 5673

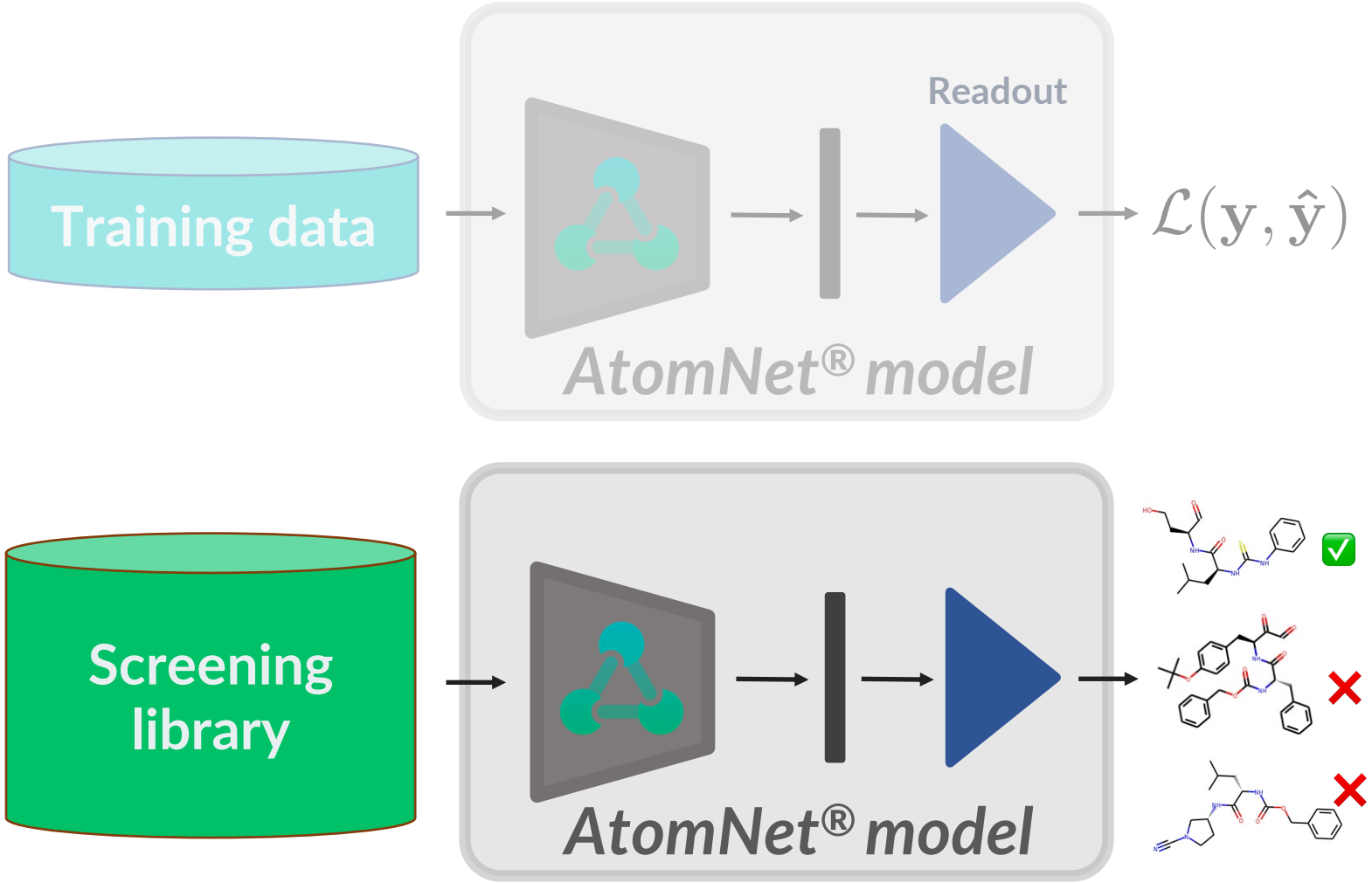
# Virtual screening using AtomNet® model




Training

- Bioactivity dataset from multiple sources with diverse targets and ligands
- Classifier and regressor models trained on bioactivity prediction task

# Virtual screening using AtomNet® model



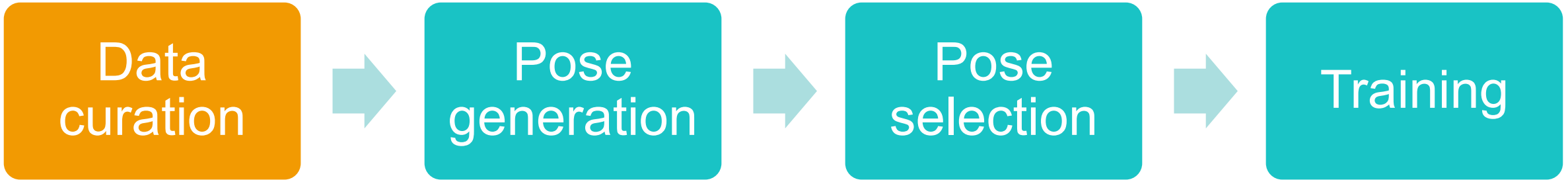

## Training

- Bioactivity dataset from multiple sources with diverse targets and ligands
- Classifier and regressor models trained on bioactivity prediction task

## Screening

- Library of compounds screened against target of interest
- Classification and ranking based on affinity prediction

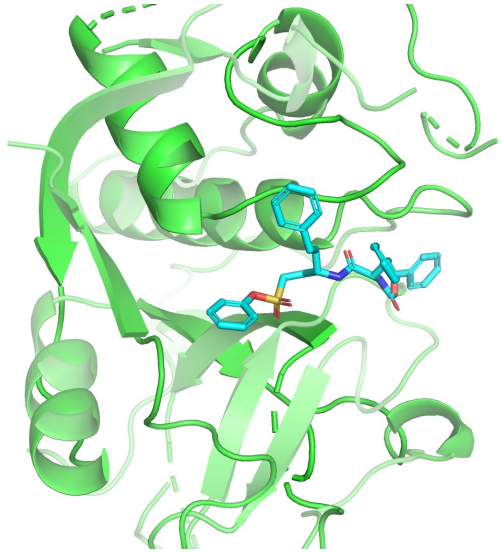
# AtomNet® model training protocol

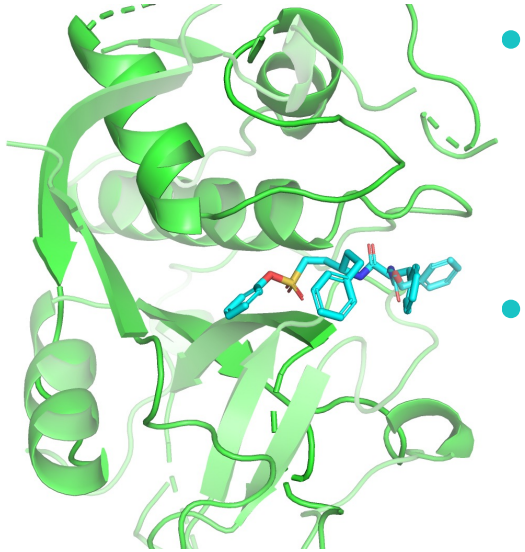



- Standardization of binding affinity data
- Filtering out potentially noisy and incorrect data

# AtomNet® model training protocol

Data curation → **Pose generation** → Pose selection → Training



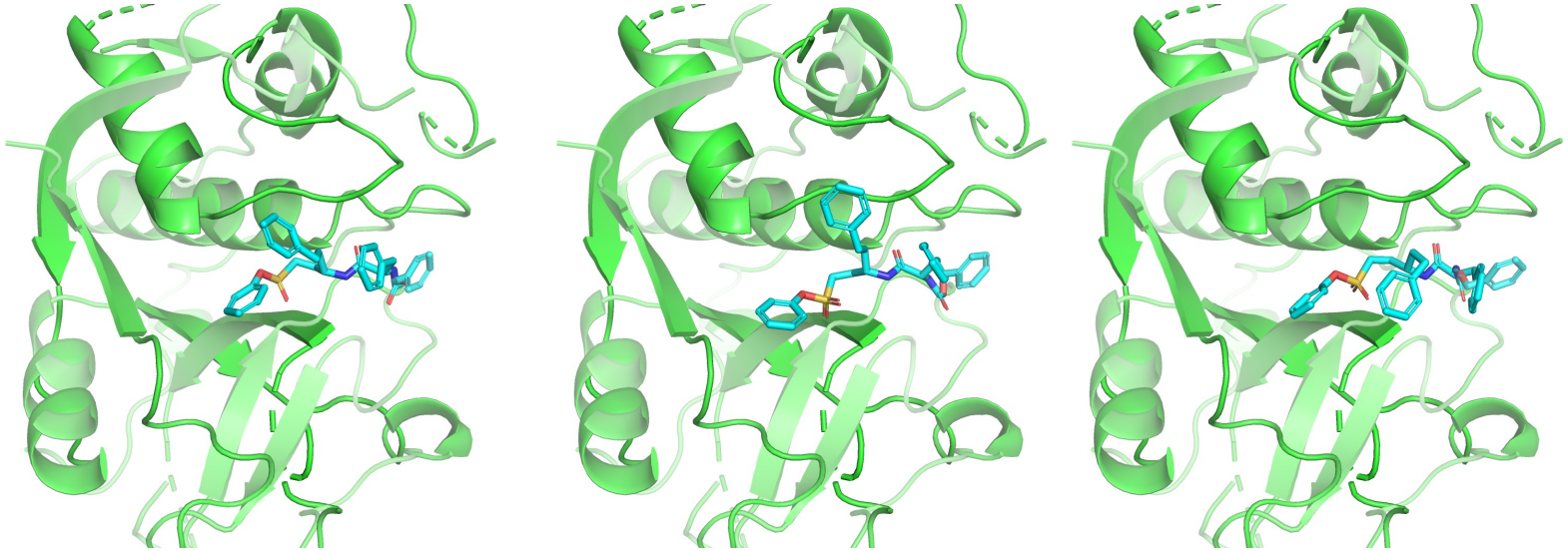


- Protein-ligand conformations (poses) generated using docking
- Cuina,[1] our in-house GPU implementation of the Vina scoring function, used

[1] https://blog.atomwise.com/efficient-gpu-implementation-of-autodock-vina

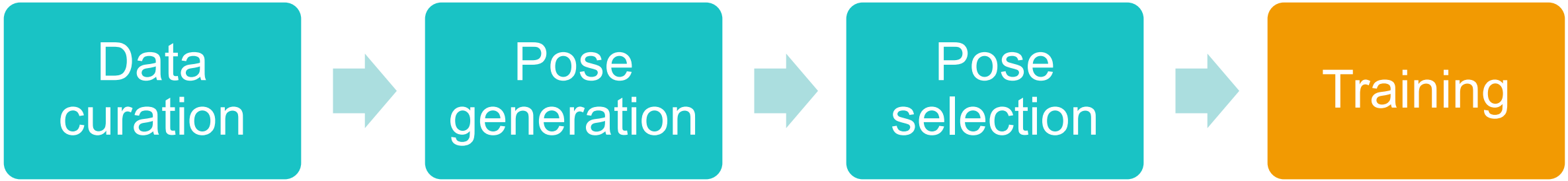
# AtomNet® model training protocol

Data curation → Pose generation → **Pose selection** → Training



- Selection of poses from ensemble
- Ranking of poses using AtomNet® pose ranker

# AtomNet® model training protocol

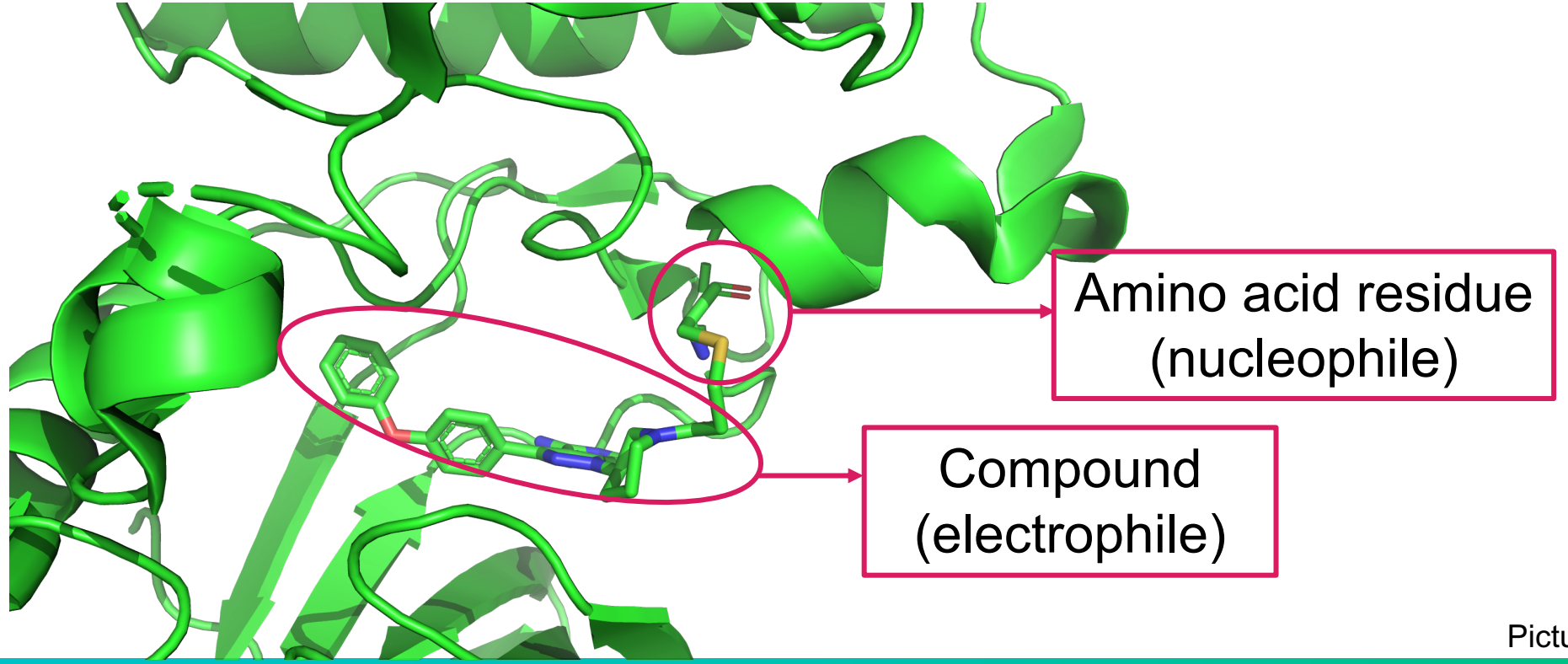



- Decoy binding affinity data generation
- Cross-validation splits

# Data curation for covalent inhibitors

## Identify electrophile and nucleophile required for reaction

- Covalent reaction requires a warhead-containing compound and nucleophilic residue in the binding pocket

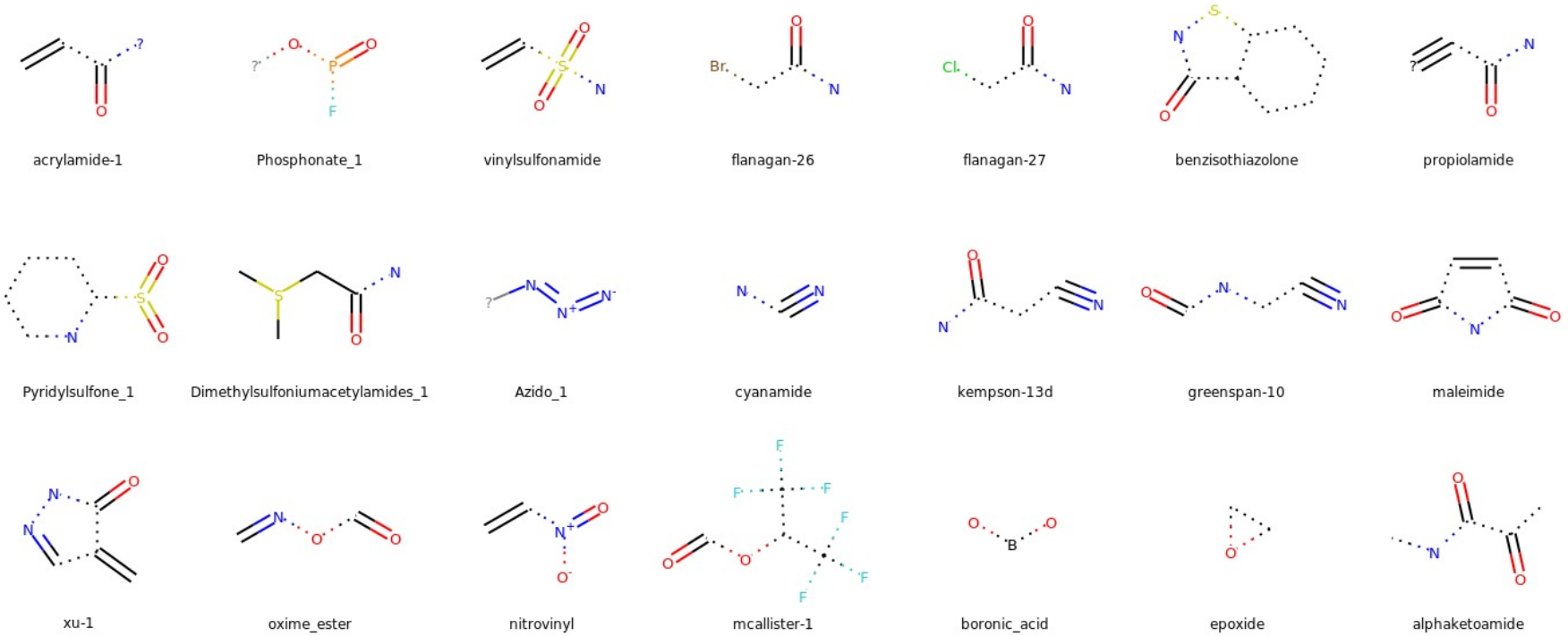



Picture credits: PDB ID 5P9J

# Data curation for covalent inhibitors

## Electrophile identified by SMARTS pattern matching

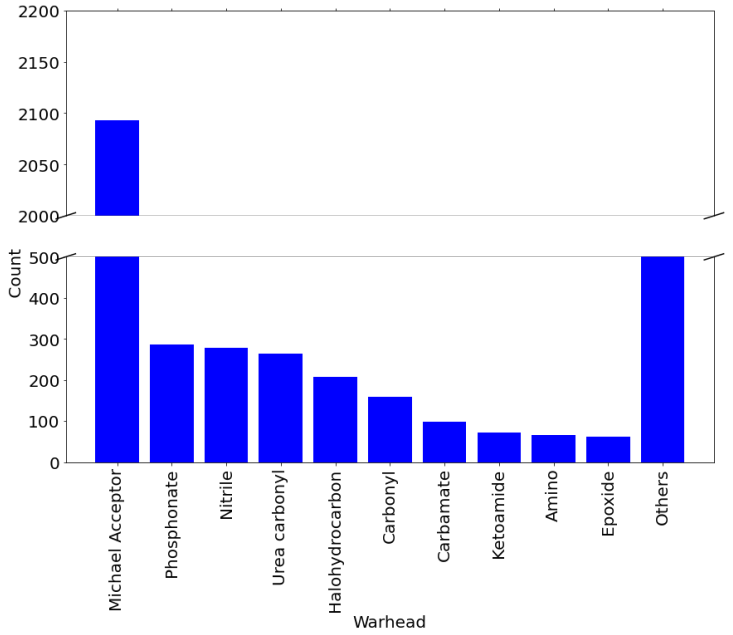
- Warheads identified by SMARTS pattern matching
- ~80 SMARTS patterns compiled

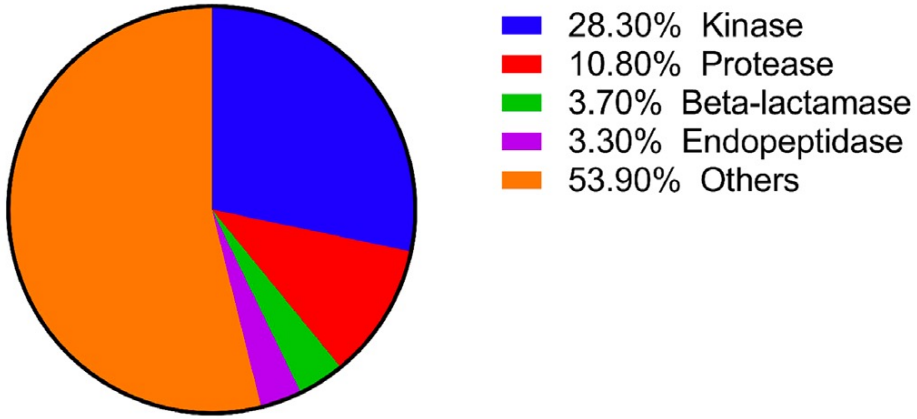


# Data curation for covalent inhibitors

## Electrophile identified by SMARTS pattern matching

- Warheads identified by SMARTS pattern matching
- ~80 SMARTS patterns compiled
- CovalentInDB[1] is a comprehensive database of covalent inhibitors
  - 4500 inhibitors
  - 280 protein targets

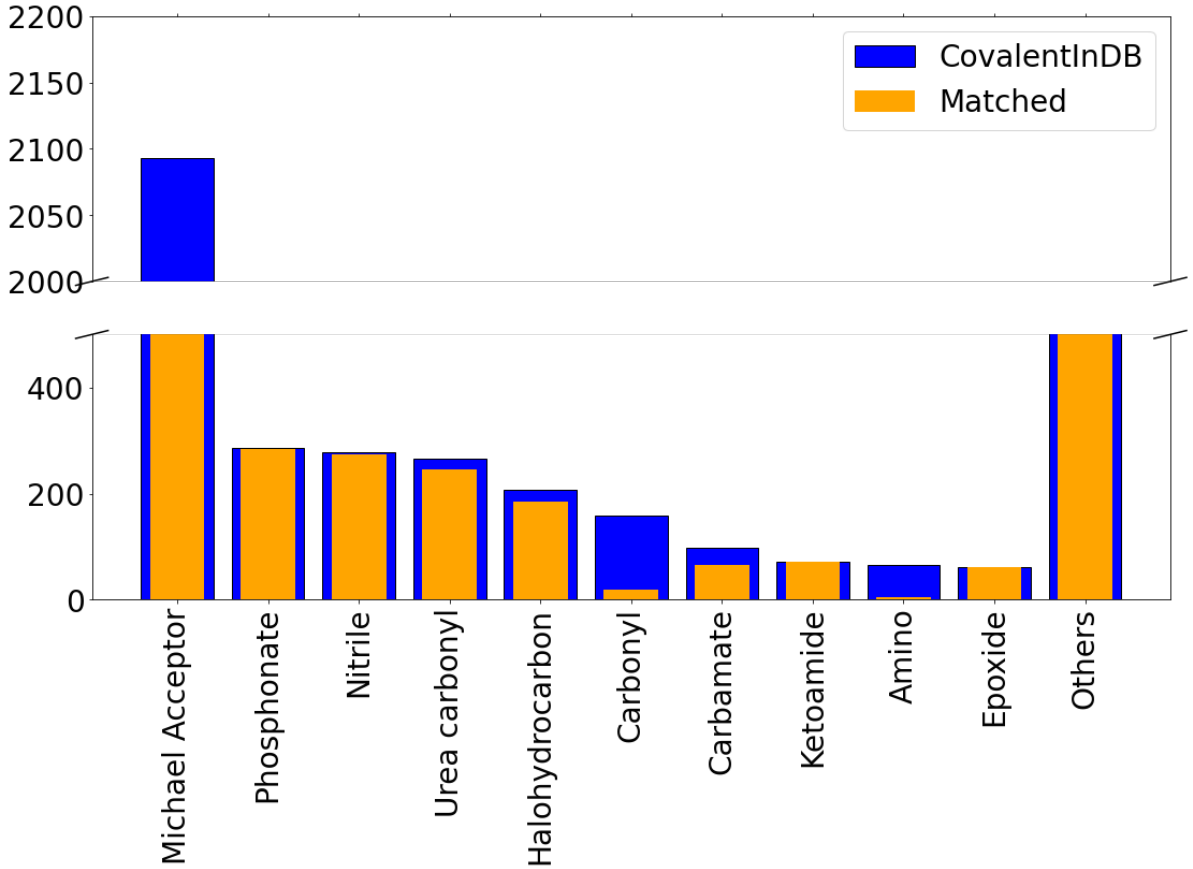



[1] Du, H. et al. *Nucleic Acids Res.* **2021**, *49*, D1122

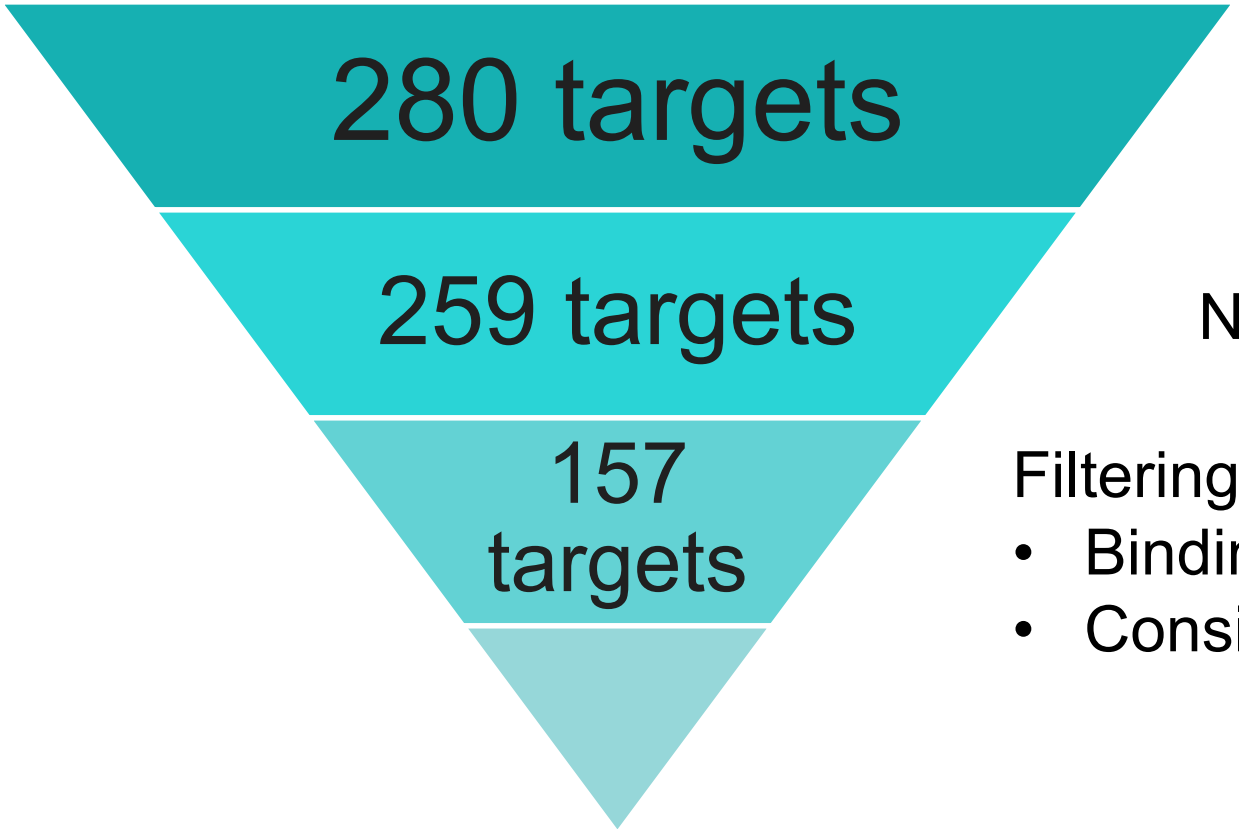
# Data curation for covalent inhibitors

- ~80 SMARTS patterns cover 87.6% of CovalentInDB



# Data curation for covalent inhibitors

Potential covalent targets and site of attachment identified from CovalentInDB



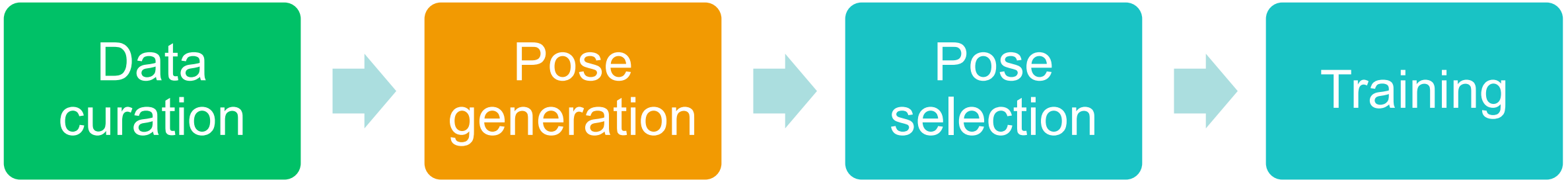

Number of unique targets CovalentInDB

Number of targets in Atomwise DB

Filtering

- Binding to amino acid residues and not cofactors
- Consistent nucleophilic residue identity

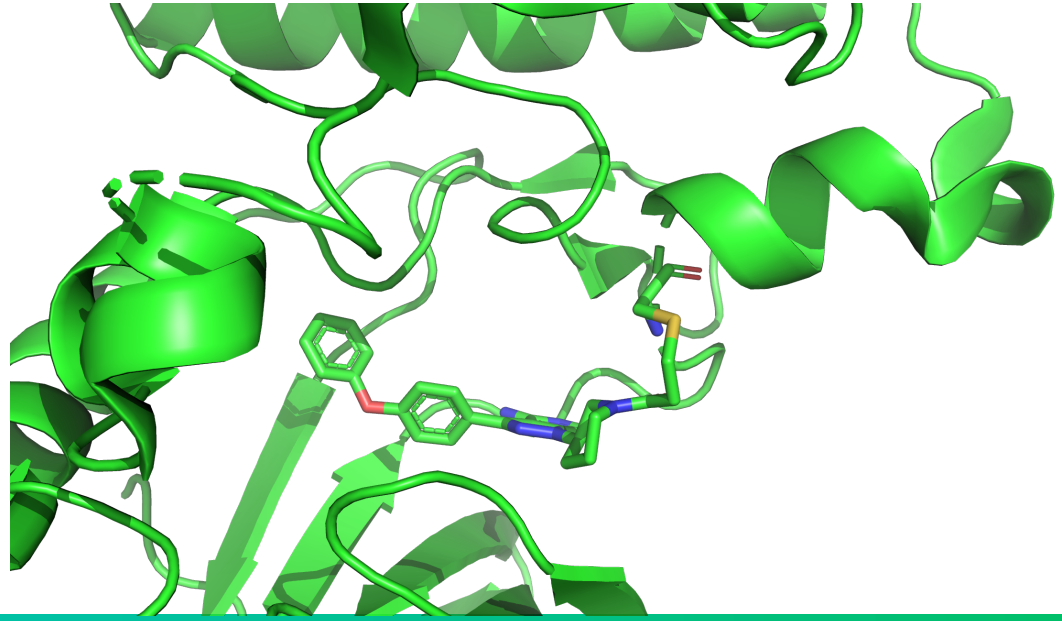
# AtomNet® model training protocol




- Covalent pose generation for covalent target and covalent inhibitor

# Pose generation

- Covalent poses generated using constrained docking
- Requirements
  1. Modification of amino acid residue for binding
  2. Modification of ligand
  3. Position and atom(s) to restrain

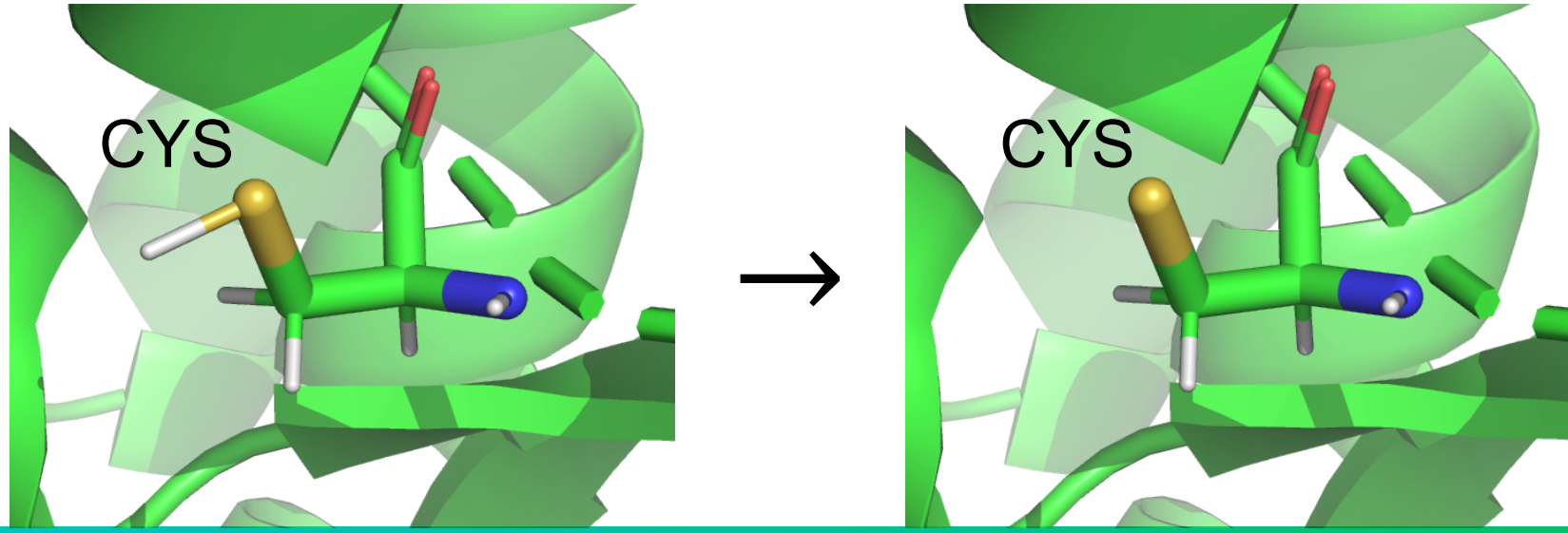


# Pose generation

## Target is modified by deleting hydrogen on nucleophilic atom

- Deletion of appropriate hydrogen on the covalent residue

| Residue | Modification |
|---|---|
| CYS | Delete HG |
| SER | Delete HG |
| LYS | Delete 1HZ/2HZ/3HZ |
| HIS | Delete 2HE |



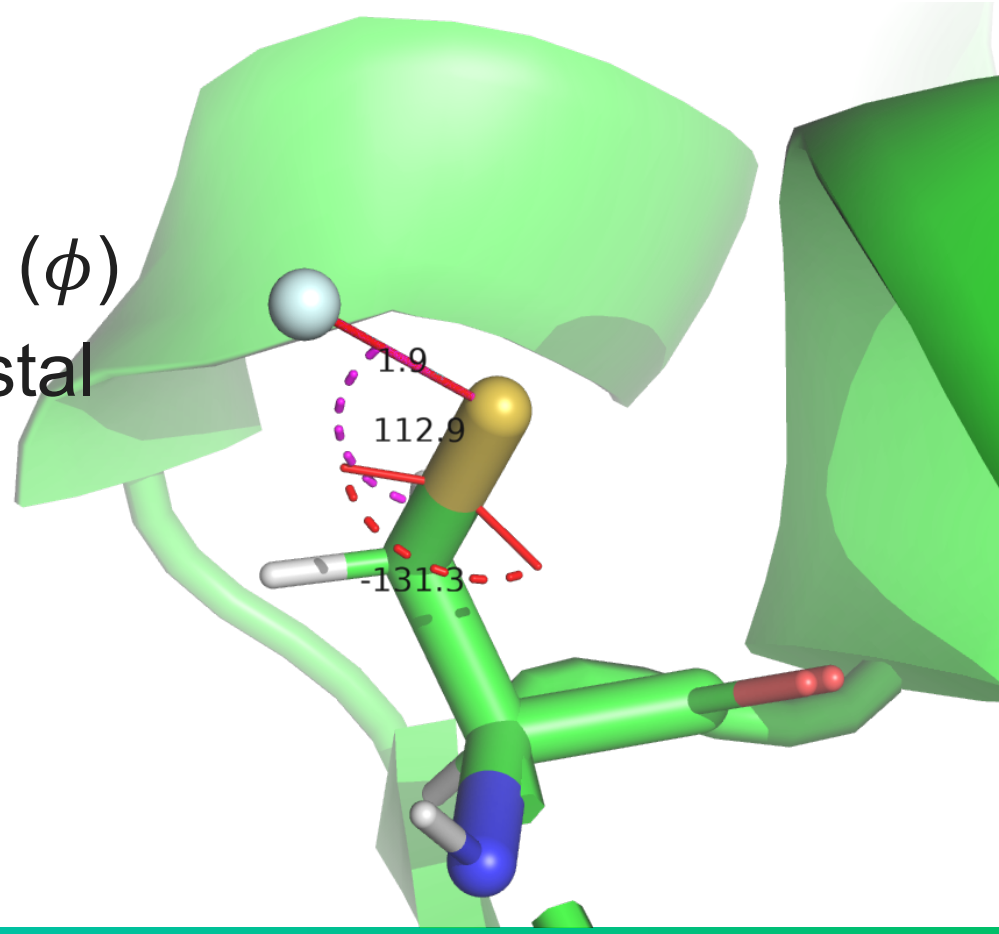


Picture credits: PDB ID 5P9J

# Pose generation

## Target is modified by deleting hydrogen on nucleophilic atom

- Deletion of appropriate hydrogen on the covalent residue
- Restraint point is deduced from the bond distance ($r$), bond angle ($\theta$) and dihedral ($\phi$)
- Parameter set ($r, \theta, \phi$) deduced from crystal structures



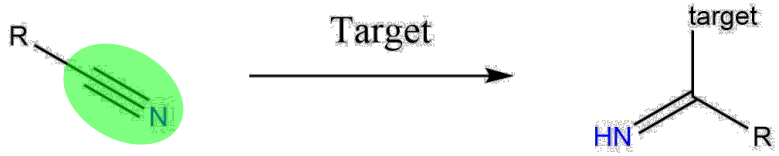
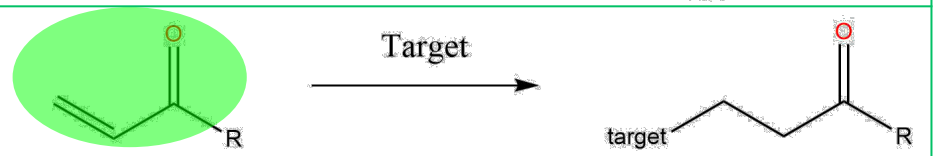


Picture credits: PDB ID 5P9J

# Pose generation

Ligand is modified to resemble the product after reacting covalently with target

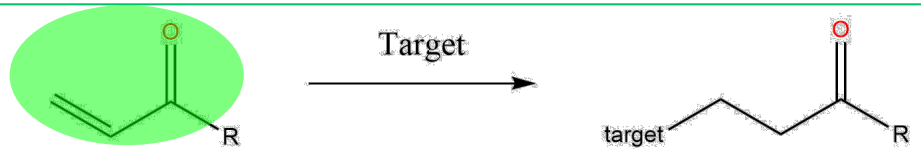
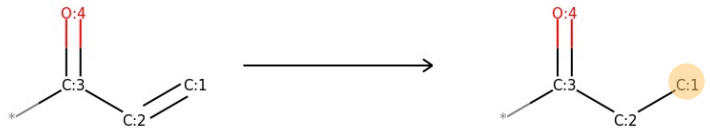
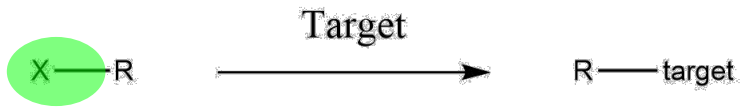
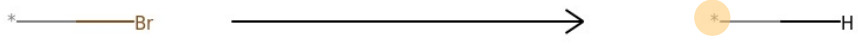
- Each warhead undergoes a certain type of reaction with residue
- Reaction is carried out programmatically using reaction SMARTS

| Warhead | Reaction mechanism |
|---|---|
| Nitrile |   |
| Michael acceptor |   |
| Halohydrocarbons |   |

# Pose generation

Ligand is modified to resemble the product after reacting covalently with target

- Each warhead undergoes a certain type of reaction with residue
- Reaction is carried out programmatically using reaction SMARTS
- Atom index number of atom to be restrained is saved

| Warhead | Reaction mechanism | Reaction SMARTS |
|---|---|---|
| Nitrile |  |  |
| Michael acceptor |  |  |
| Halohydrocarbons |  |  |

# Pose generation

## Covalent bond is modelled by restraining bonding atoms to be proximal

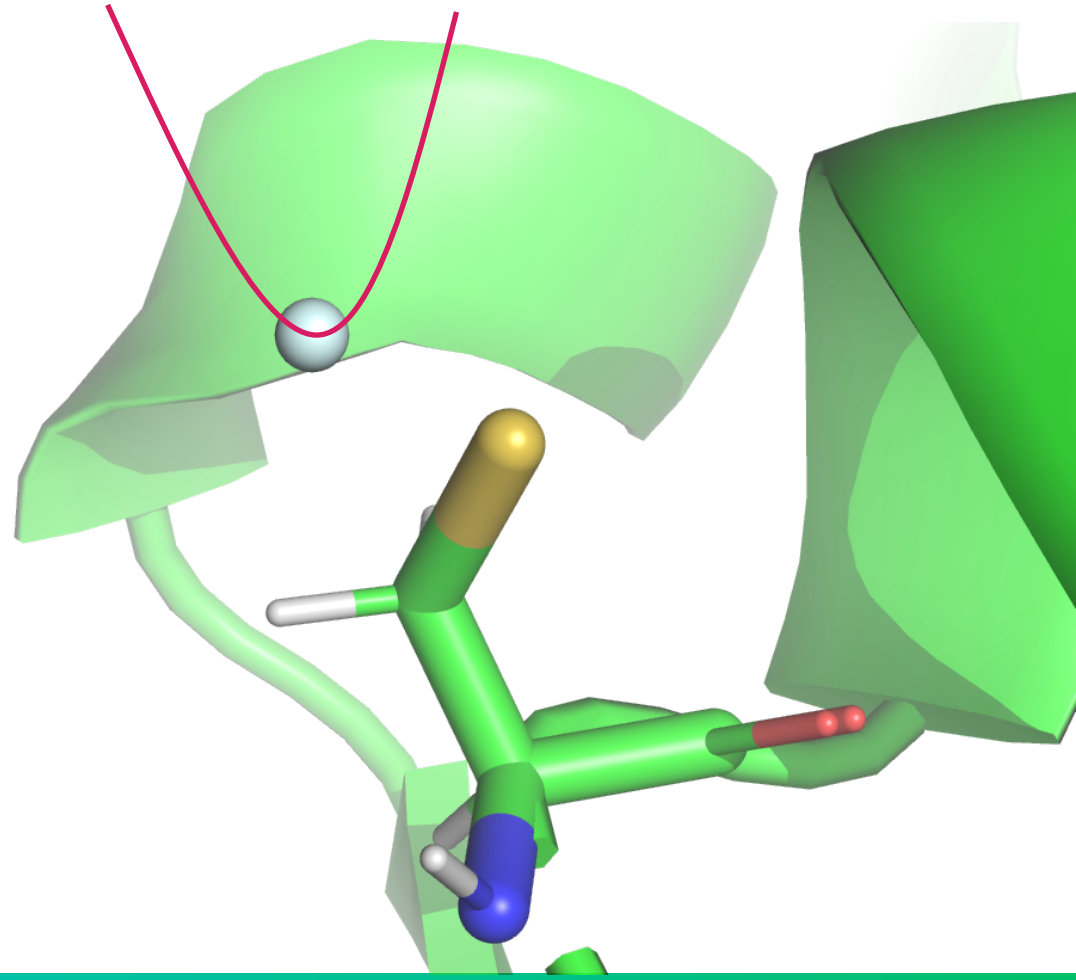
- Covalent bond is modeled by adding a penalty to the Vina scoring function
- Penalty is added on ligand atom that forms covalent bond
- Penalty form:

$$f(\vec{r}) = a_0(\vec{r} - \overrightarrow{r_0})^2$$

$\vec{r}$: Position of ligand atom

$\overrightarrow{r_0}$: Restraint point

$a_0$: Prefactor

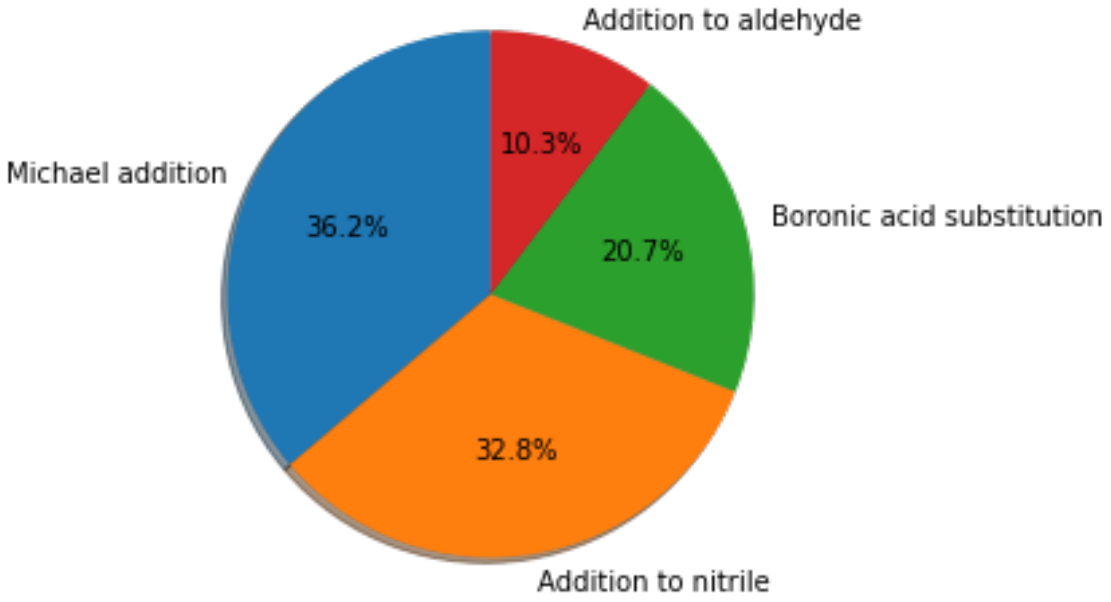



Picture credits: PDB ID 5P9J

# Dataset and protocol for docking evaluation

## Dataset

- 116 crystal structures
- 92 cysteines and 24 serines

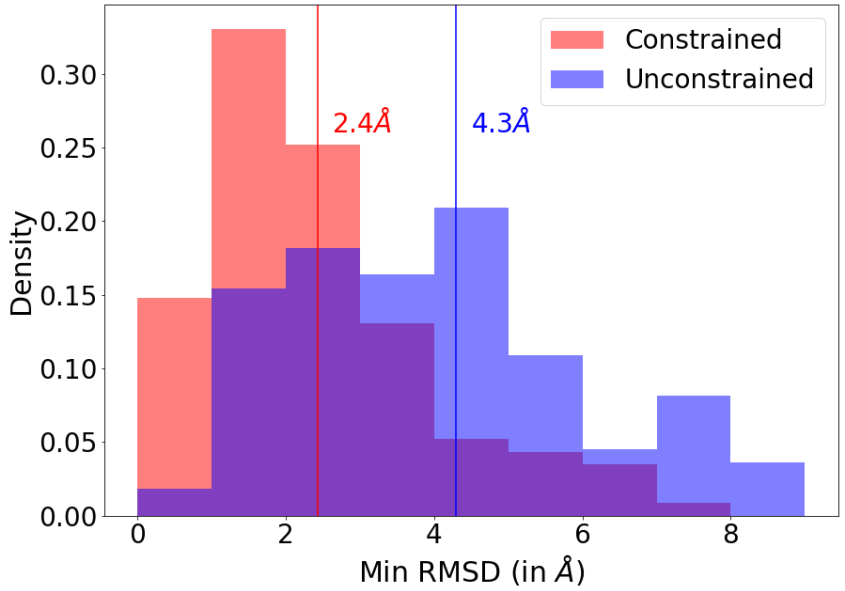



## Protocol

- Top 64 out of 2048 poses chosen
- 64 poses compared with crystal structure pose
- RMSD computed for heavy atoms of ligand
- Minimum of the RMSDs reported for each complex

# Pose generation

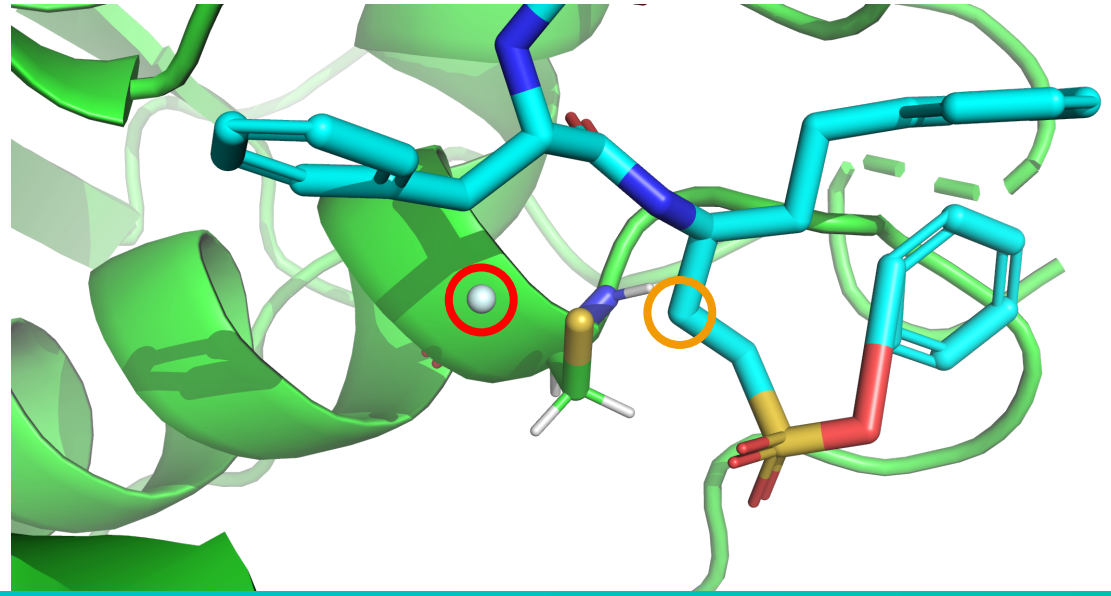
- Constrained docking can recapitulate crystal structure pose significantly better than unconstrained docking
- Restraint parameters set $(r, \theta, \phi)$ derived from crystal structure

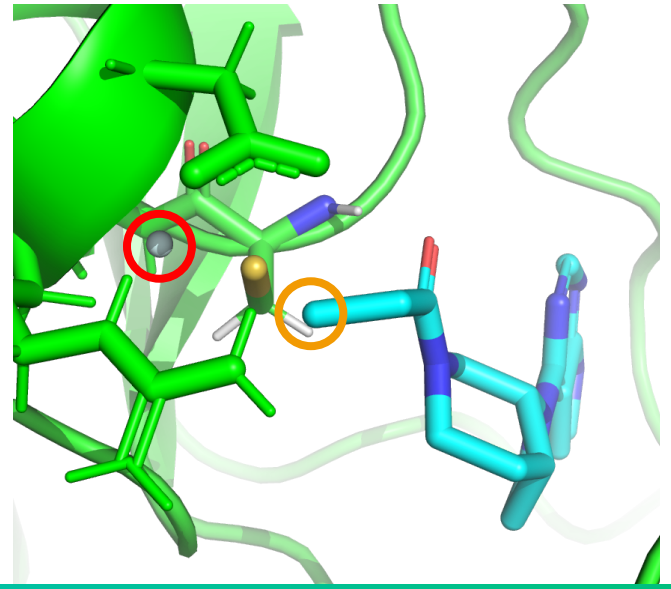


# Pose generation

## Correct location of restraint is essential to generate good poses

- Bond distance and angle are transferable but not dihedral

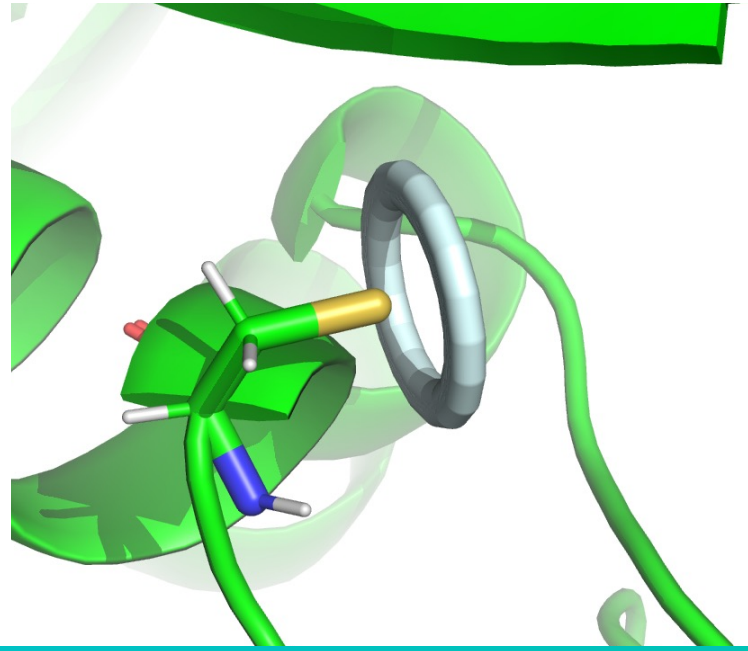


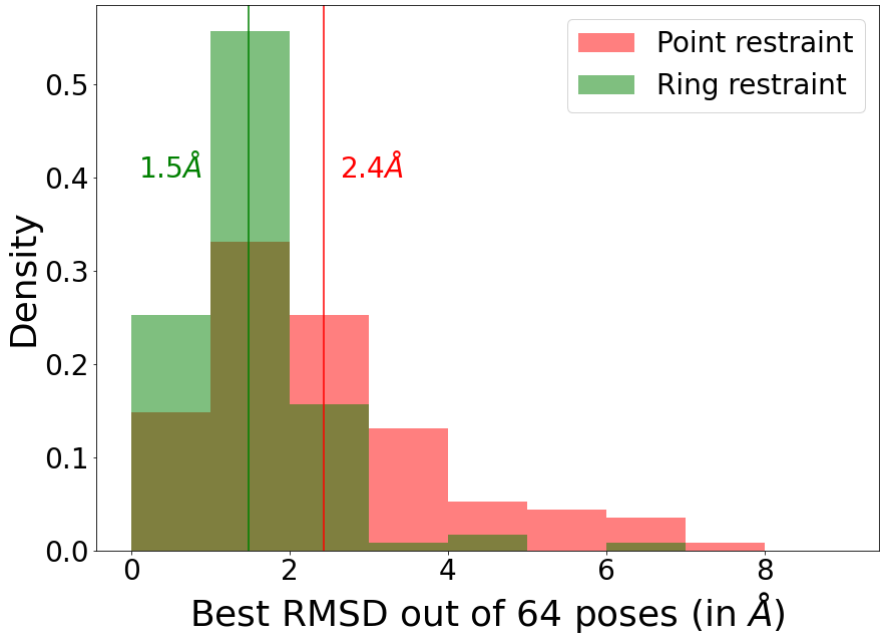


# Pose generation

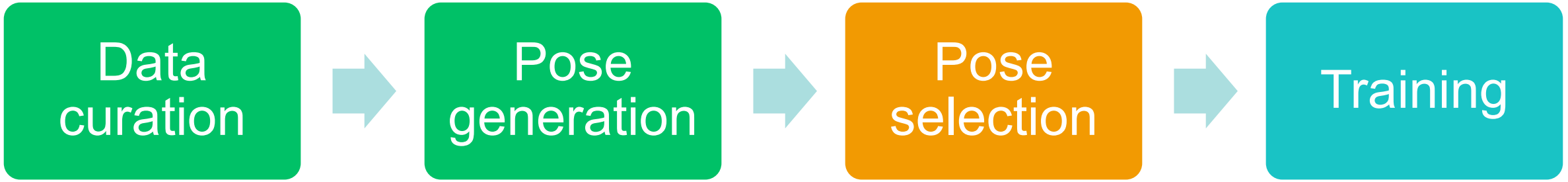
## Correct location of restraint is essential to generate good poses

- Bond distance and angle are transferable but not dihedral
- Scanning the restraint point by scanning dihedral with fixed $(r, \theta)$
- "Ring restraint" reduced RMSD of best pose by ~0.9 Å





# AtomNet® model training protocol



# Pose selection

## Good poses are important for binding affinity predictions

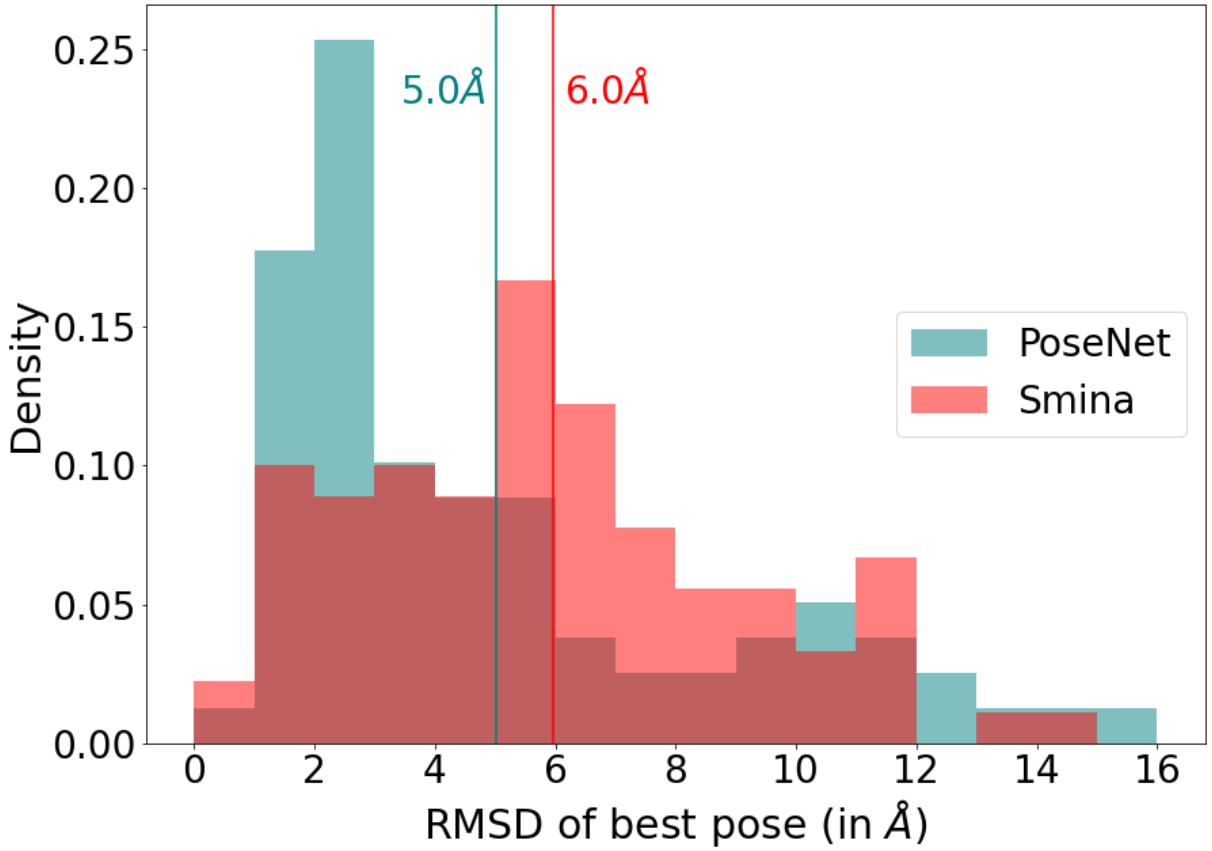
- Vina scoring function is good at generating good poses but not ranking them
- Better poses improve performance of binding activity prediction models
- AtomNet® pose ranker[1] is a neural network trained to rescore poses and identify the best ones
- AtomNet® pose ranker currently trained only on non-covalent interactions

[1] https://blog.atomwise.com/ligand-pose-ensembles-improves-affinity-prediction-in-structure-based-virtual-screening

# Pose selection

## AtomNet® pose ranker improves performance on covalent pose selection

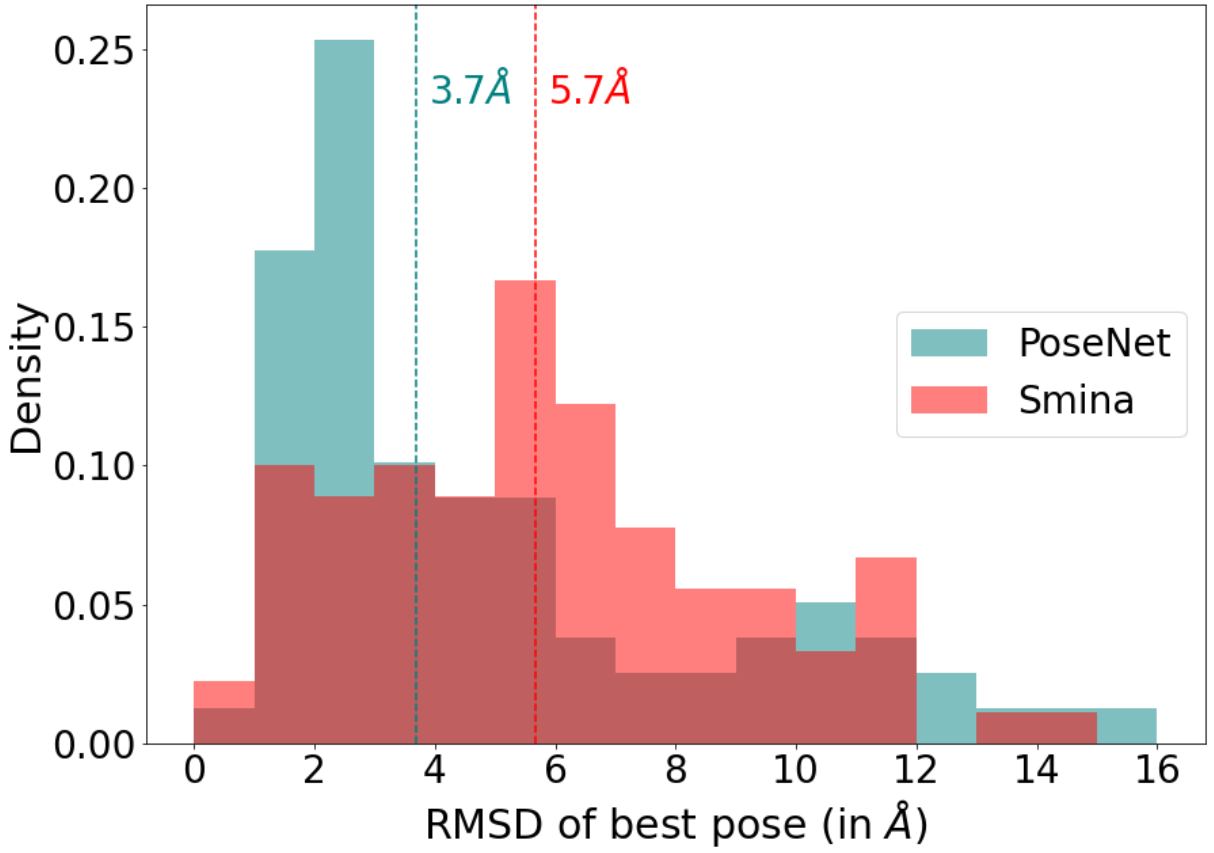
- AtomNet® pose ranker already improves performance on covalent poses
- Mean is contaminated by extreme values



# Pose selection

## AtomNet® pose ranker improves performance on covalent pose selection

- AtomNet® pose ranker already improves performance on covalent poses
- Mean is contaminated by extreme values
- Median performance improves by 2 Å
- Curation of larger dataset consisting of covalent poses needed for re-training/finetuning AtomNet® pose ranker



# Thank you


## Acknowledgements

Co-contributors
Jon Sorenson
Kate Stafford


## Conclusions

- Demonstrated protocol for identifying covalently binding protein-inhibitor pairs
- Developed a constrained docking protocol for covalent pose generation
- Constrained docking protocol can generate poses with RMSD of 1.5 Å

## Future directions

- Training covalent AtomNet® pose ranker
- Training and evaluation of models for bioactivity prediction based on good covalent poses

Atomwise

**Srimukh Prasad Veccham**

Cheminformatics Scientist

srimukh@atomwise.com

# Learn more about Atomwise

Atomwise at Fall ACS 2021:

https://info.atomwise.com/acs_fall2021

Join us!

https://www.atomwise.com/careers/