Atomwise

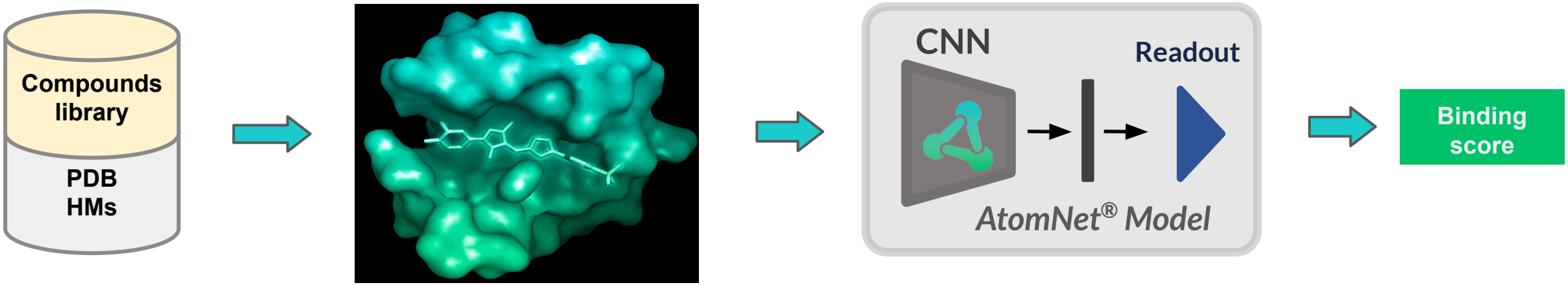
Structure-Based Drug Design with Multi-Task Learning and Data Augmentation

Training vHTS models that care about poses

Paweł Gniewek, PhD
Senior Scientist
Atomwise

Structure-based drug discovery

The Atomwise story



Data

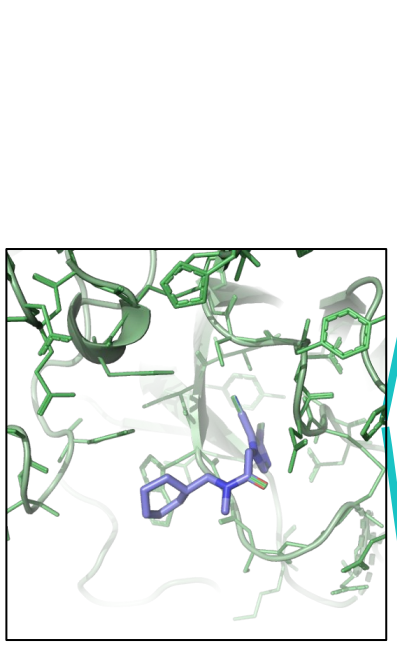
Docked Complex

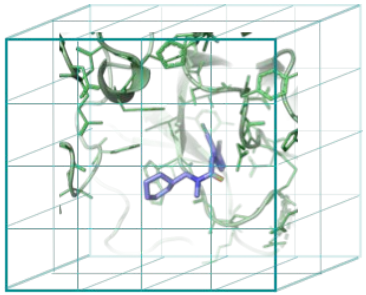
ML

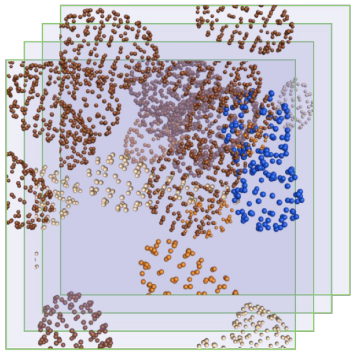
Prediction

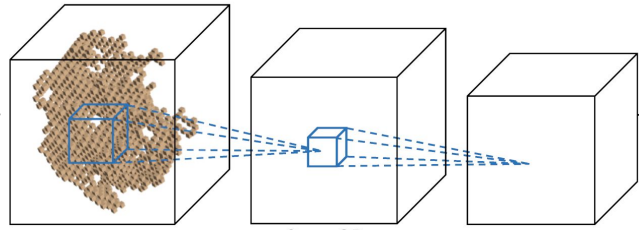
Deep Learning in Structure-Based Drug Discovery

3D grid-based and Graph-based neural networks

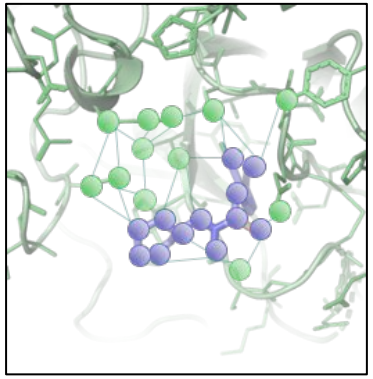


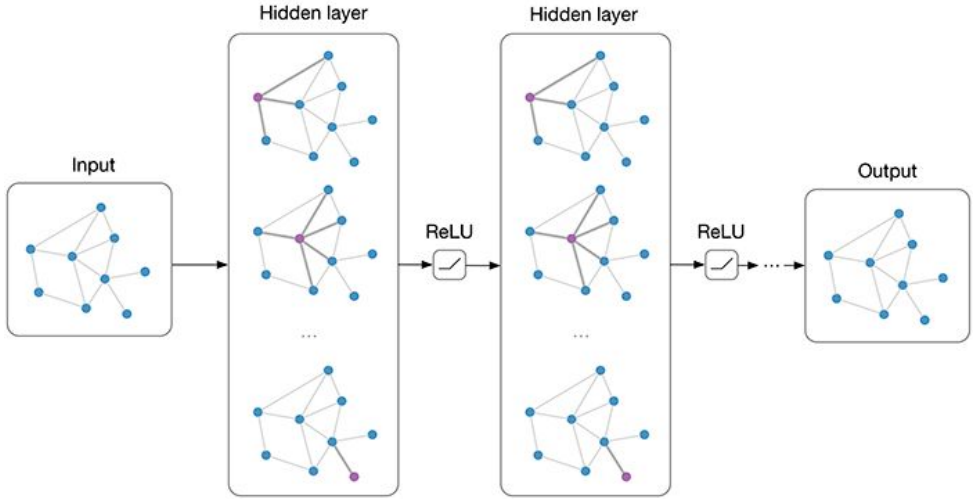






Binding score

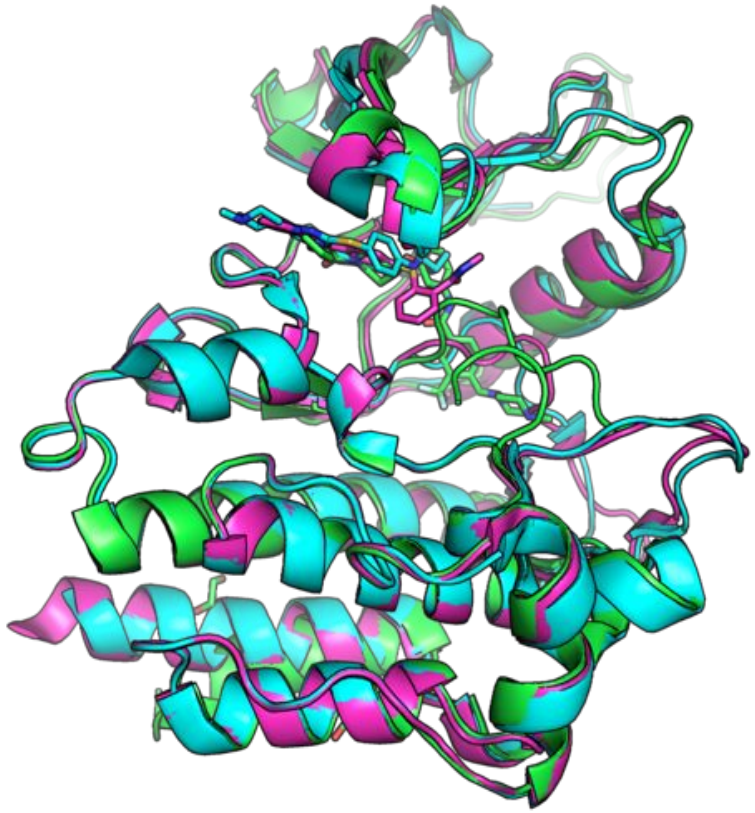




Binding score

Structure-based models

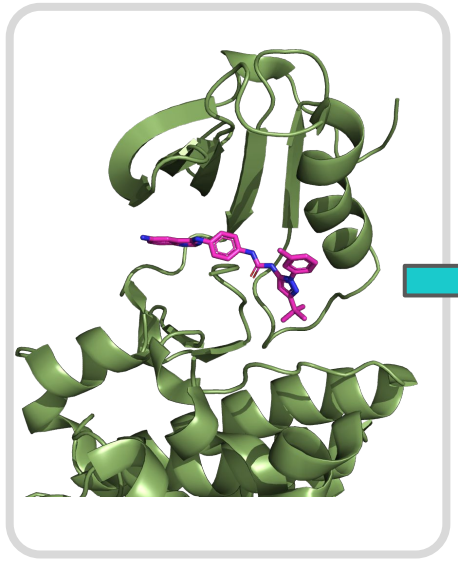
It is assumed that with structural data available, a model can learn ligand-receptor interactions and this knowledge will carry to new targets and novel chemical spaces.



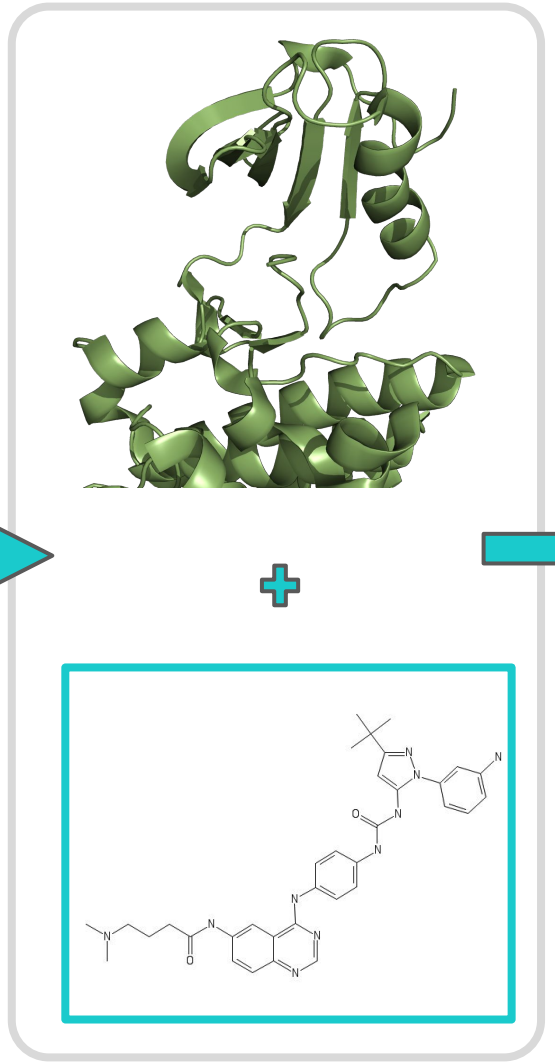
Pose-sensitivity enables:

1. Targeting novel and allosteric sites
2. Rational drug design and lead optimization
3. Visualize neural networks

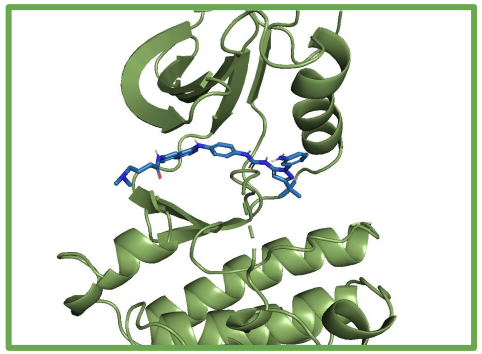
How can we make sure that the model is pose-sensitive?



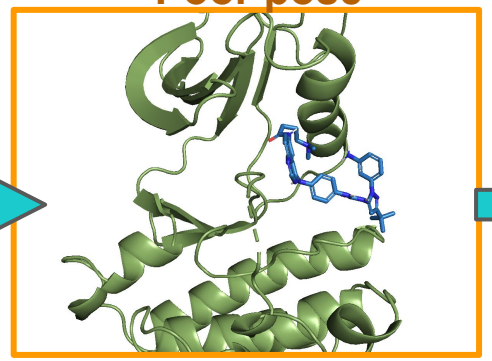
Protein kinase
PDB ID 3F3V



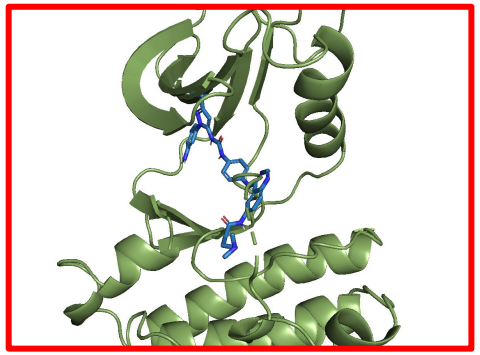
Good

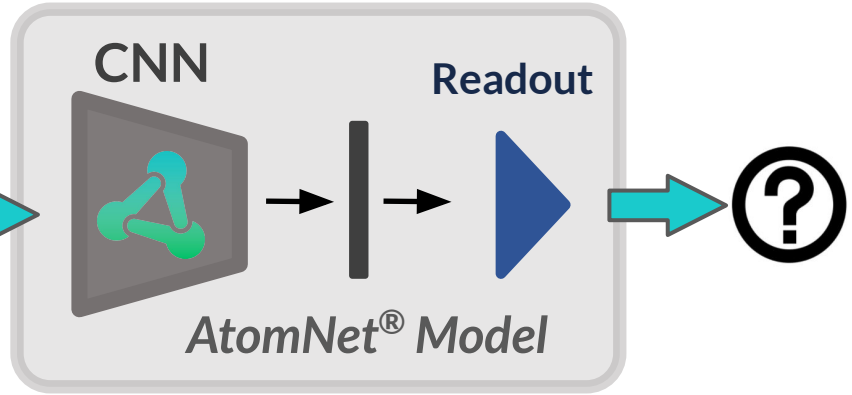


Poor pose



Physically implausible pose





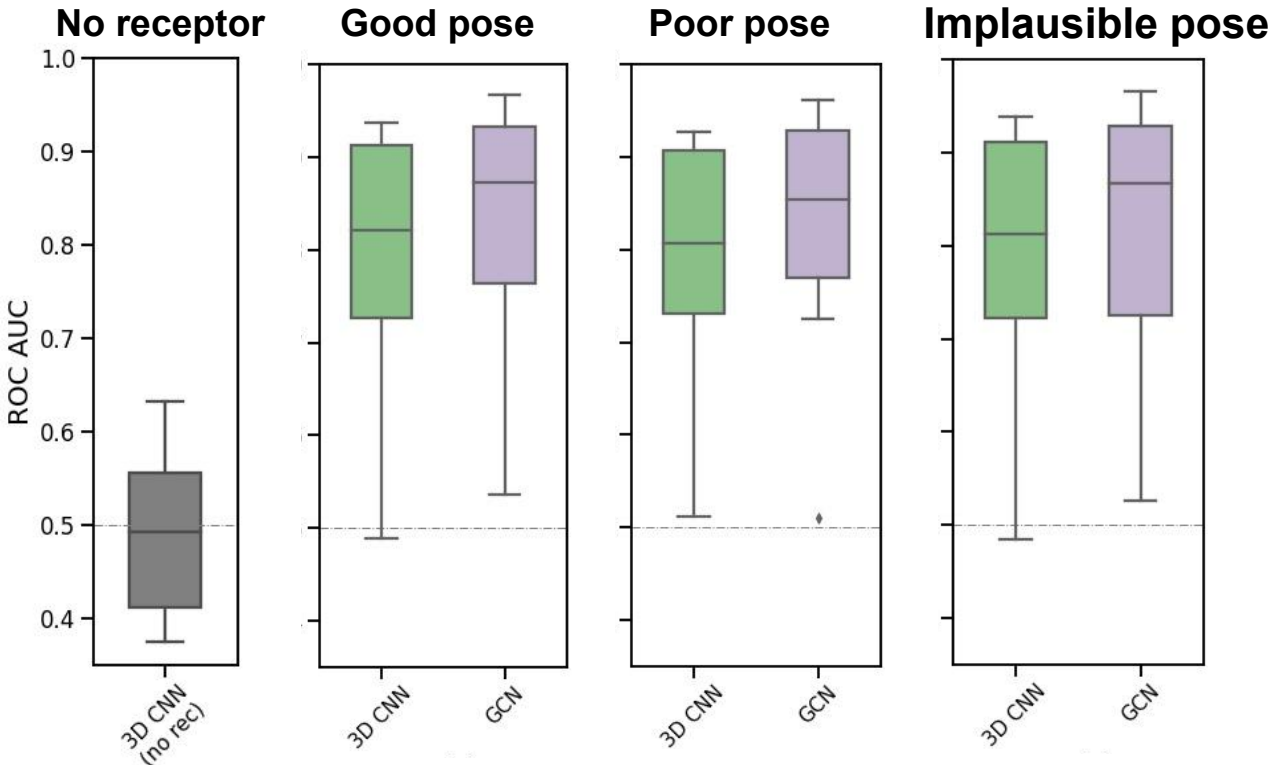
How can we make sure that the model is pose-sensitive?

Test set

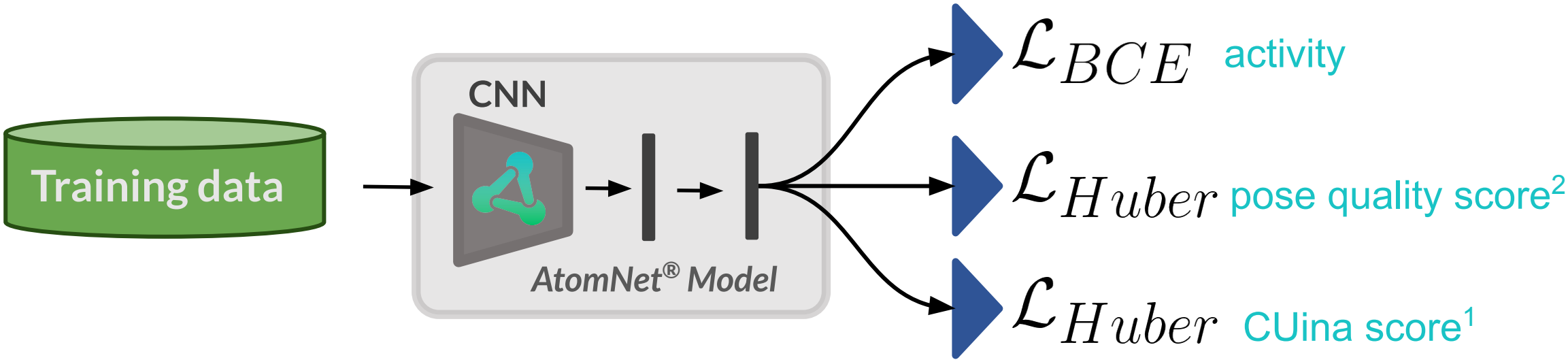
- A set of 12 target proteins spanning diverse chemical space and protein folds

Training set

- Atomwise internal bioactivity data with the test set targets excluded (including close homologs)



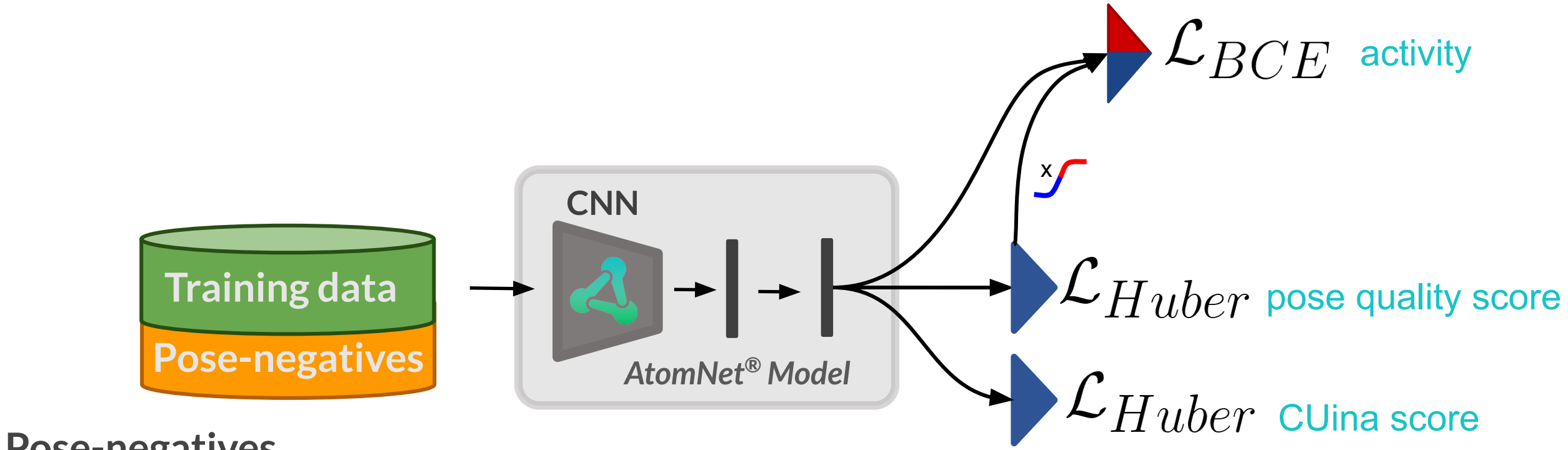
From single-task to multi-task model



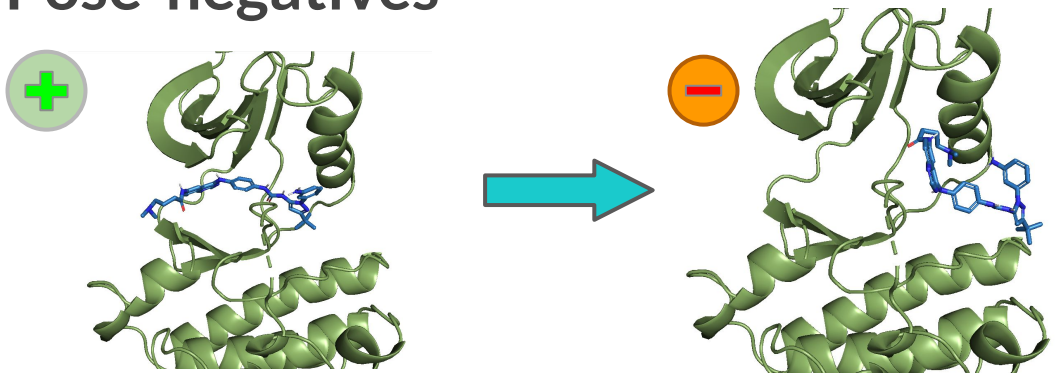
[1] Adrian Morrison, https://blog.atomwise.com/behind-the-ai-reconfiguring-key-code-to-supercharge-molecular-docking

[2] Kate Stafford *et al.*, “Modeling protein flexibility with conformational sampling improves ligand pose and bioactivity prediction”, ACS Fall 2019

Multi-task network with conditioned tasks and data augmentation

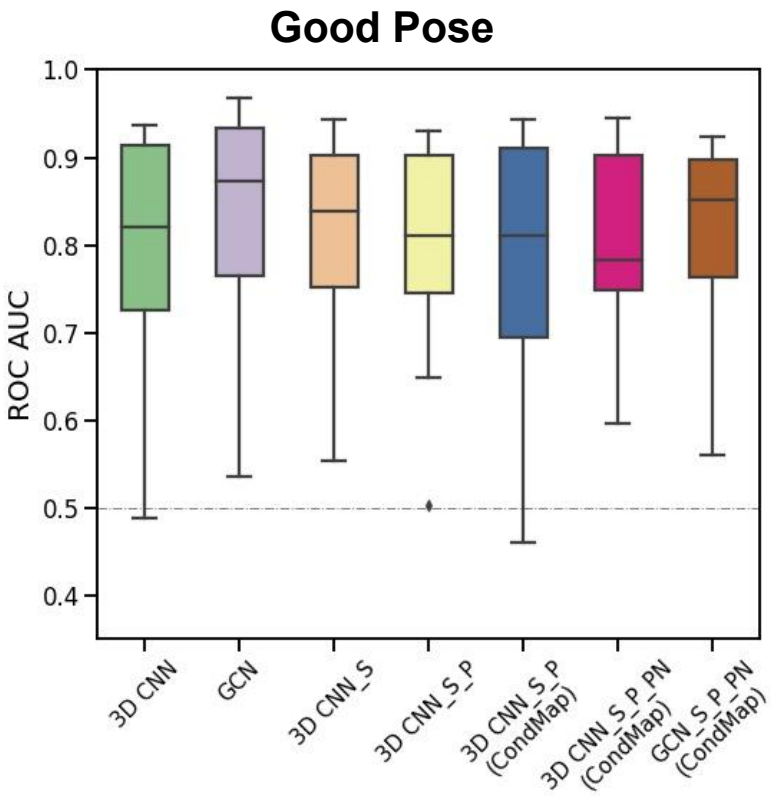


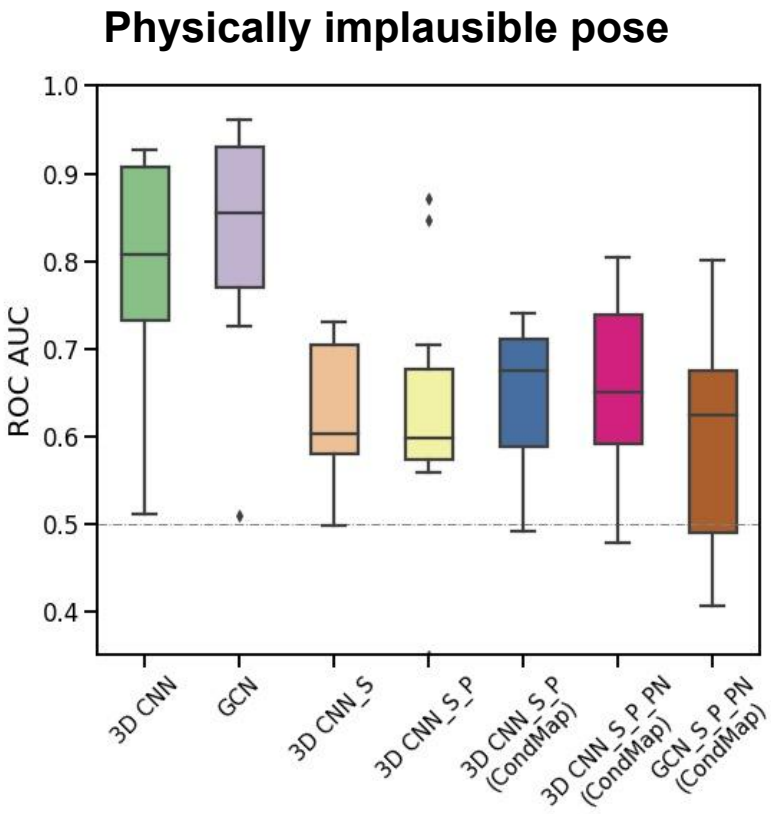
Pose-negatives

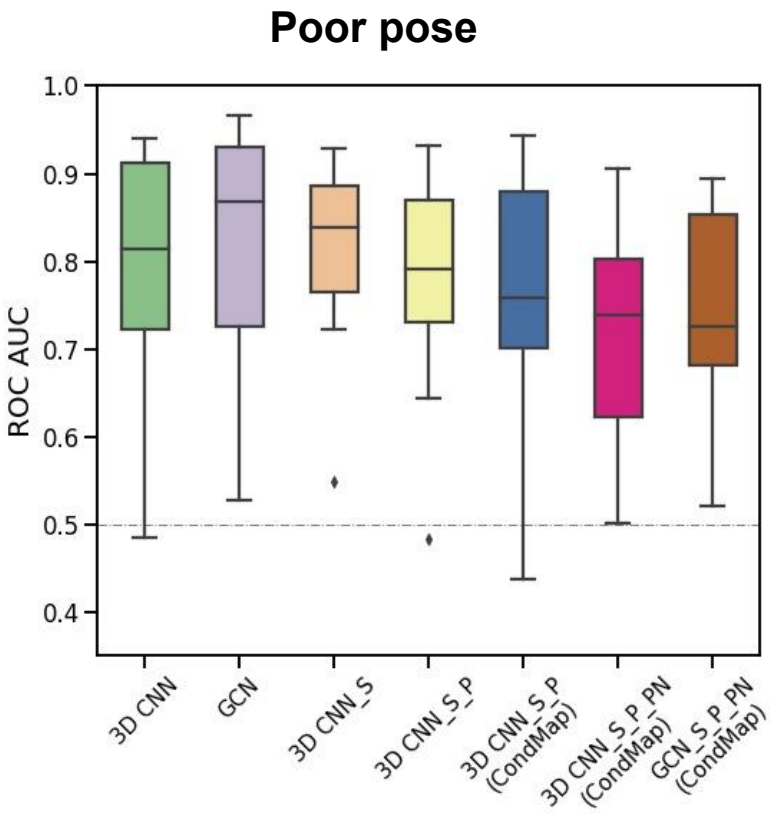


Long et al., *Conditional Adversarial Domain Adaptation*, NeurIPS 2018

Multi-task learning with data-augmentation lead to pose-sensitive DL models



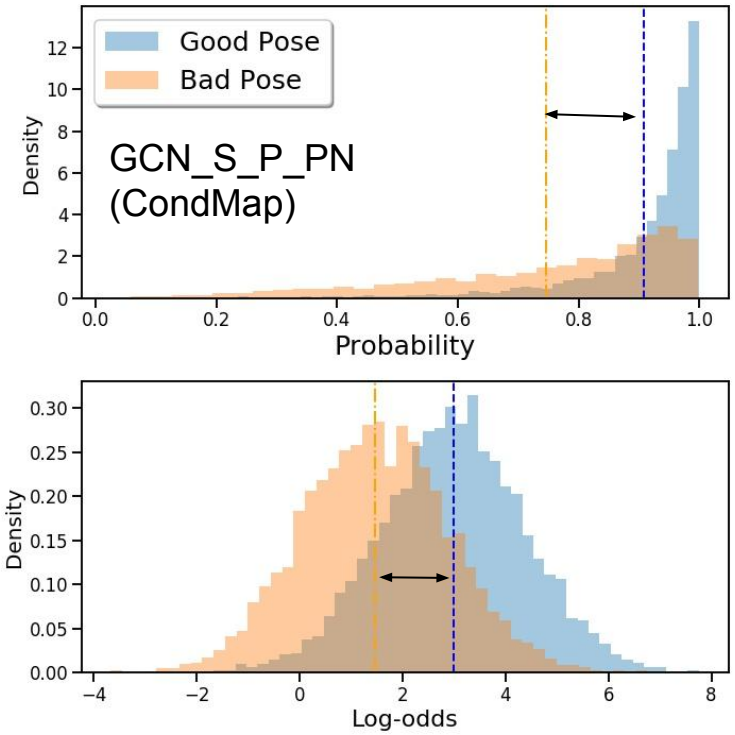




Physically implausible and poor poses are penalized

Models can grade the penalty based on how bad the pose is

Consider active compounds only

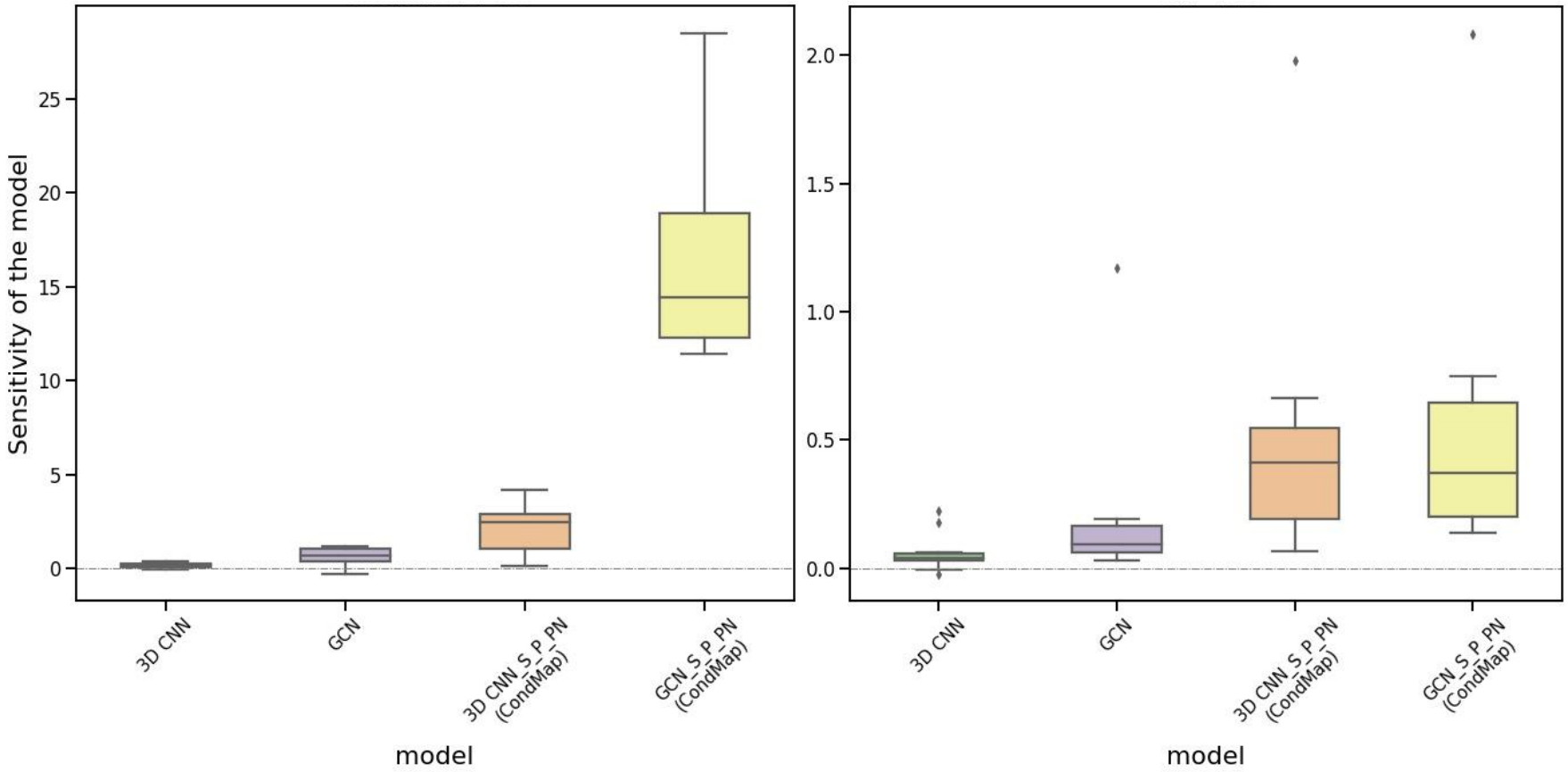


Sensitivity of the model = log-odds drop due to perturbation to the system

$$\log\text{-odds} = \log\left(\frac{p}{1-p}\right)$$

Physically implausible pose

Poor pose



Key takeaways

- Pose-sensitivity is essential for screening non-primary (allosteric) binding sites and rational compound optimization
- A combination of multi-task learning and data-augmentation enables training of pose-sensitive models
- Pose-sensitive models offer new possibilities towards better generalizability of the DL models

Acknowledgments

In collaboration with Brad Worley, Bastiaan Bergman and Atomwise Team!

Our Other Talks & Posters

https://info.atomwise.com/acs_spring2021

Join Us!

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