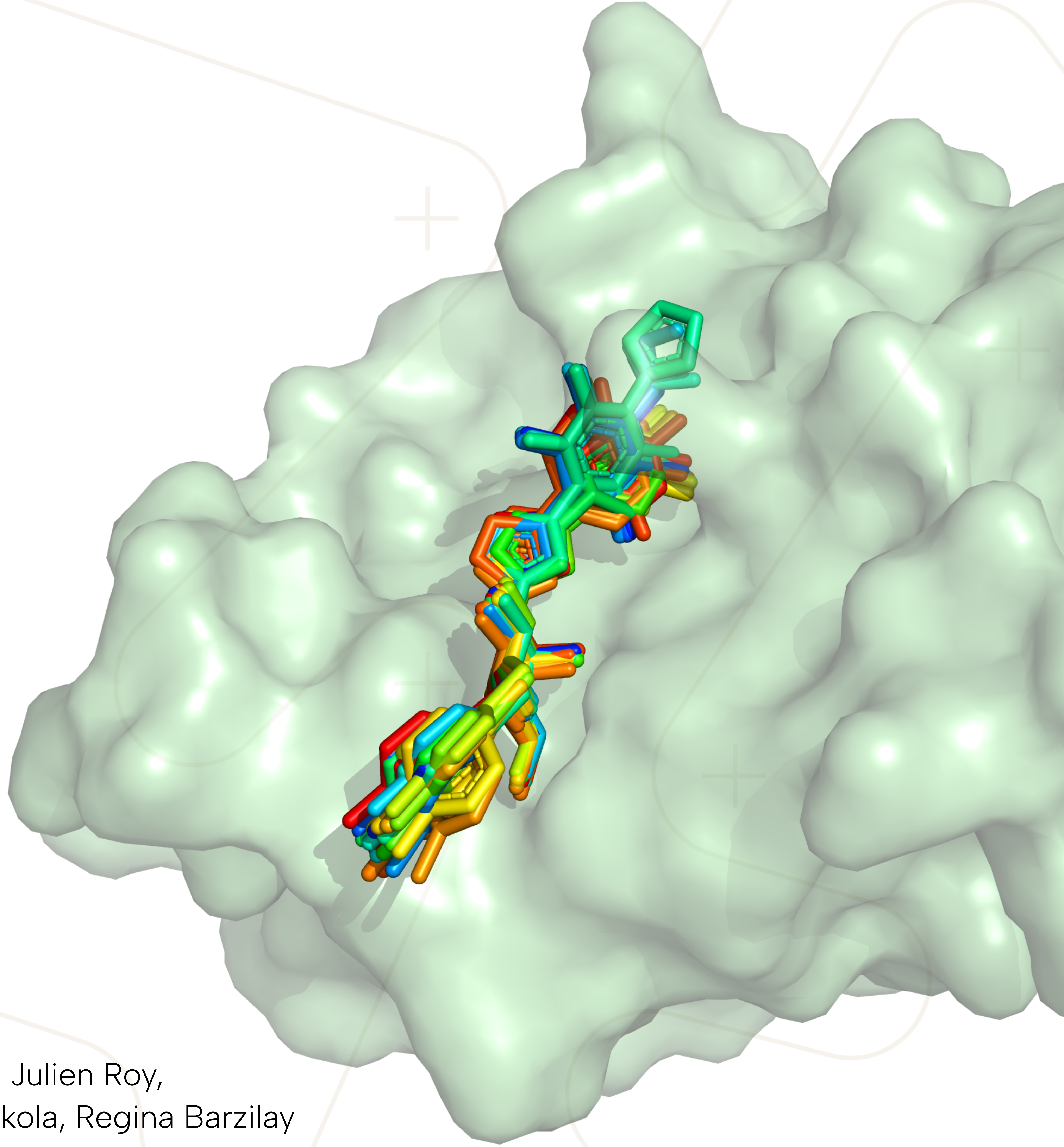
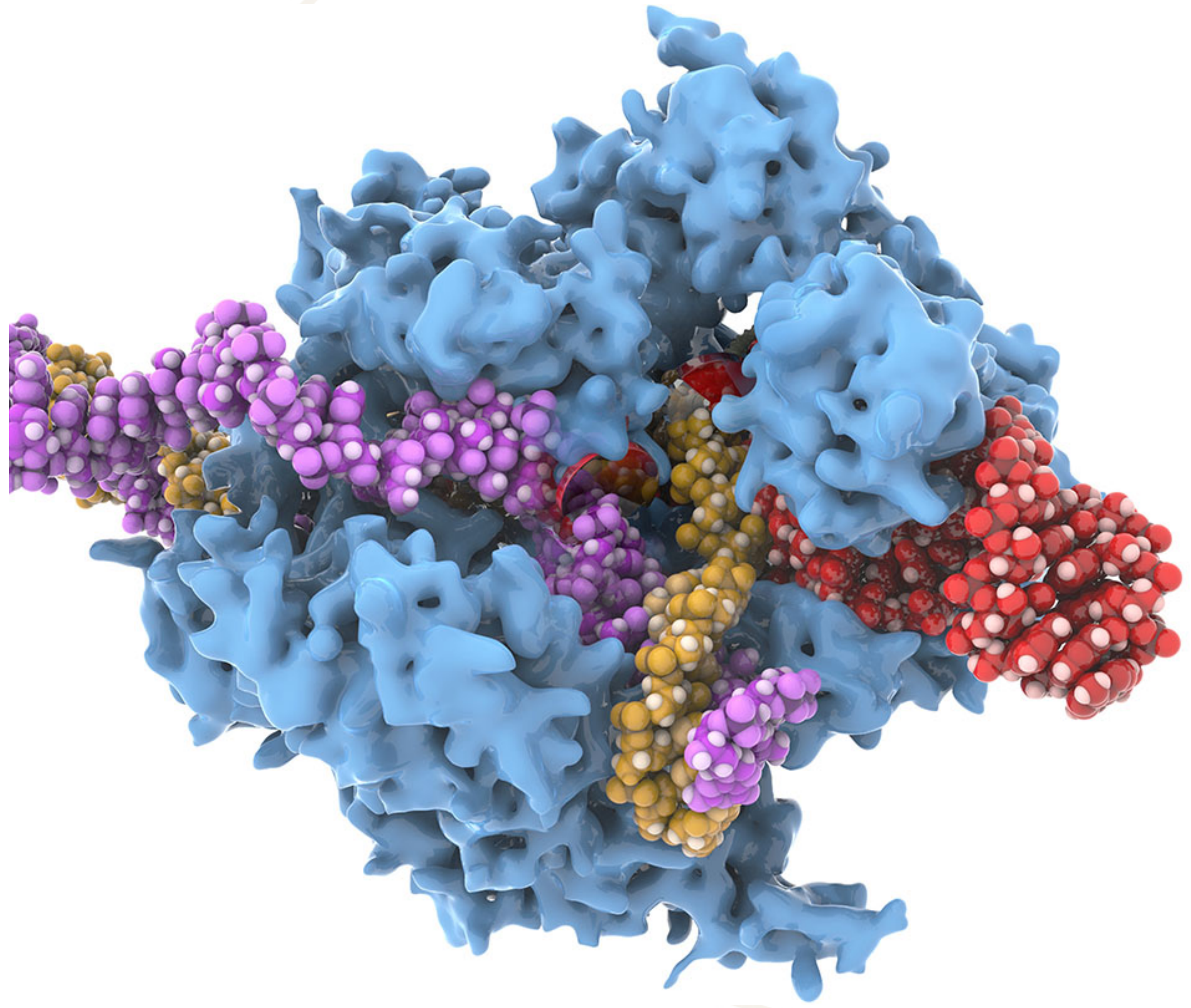
Boltz-2

Towards Accurate and Efficient Binding Affinity Prediction

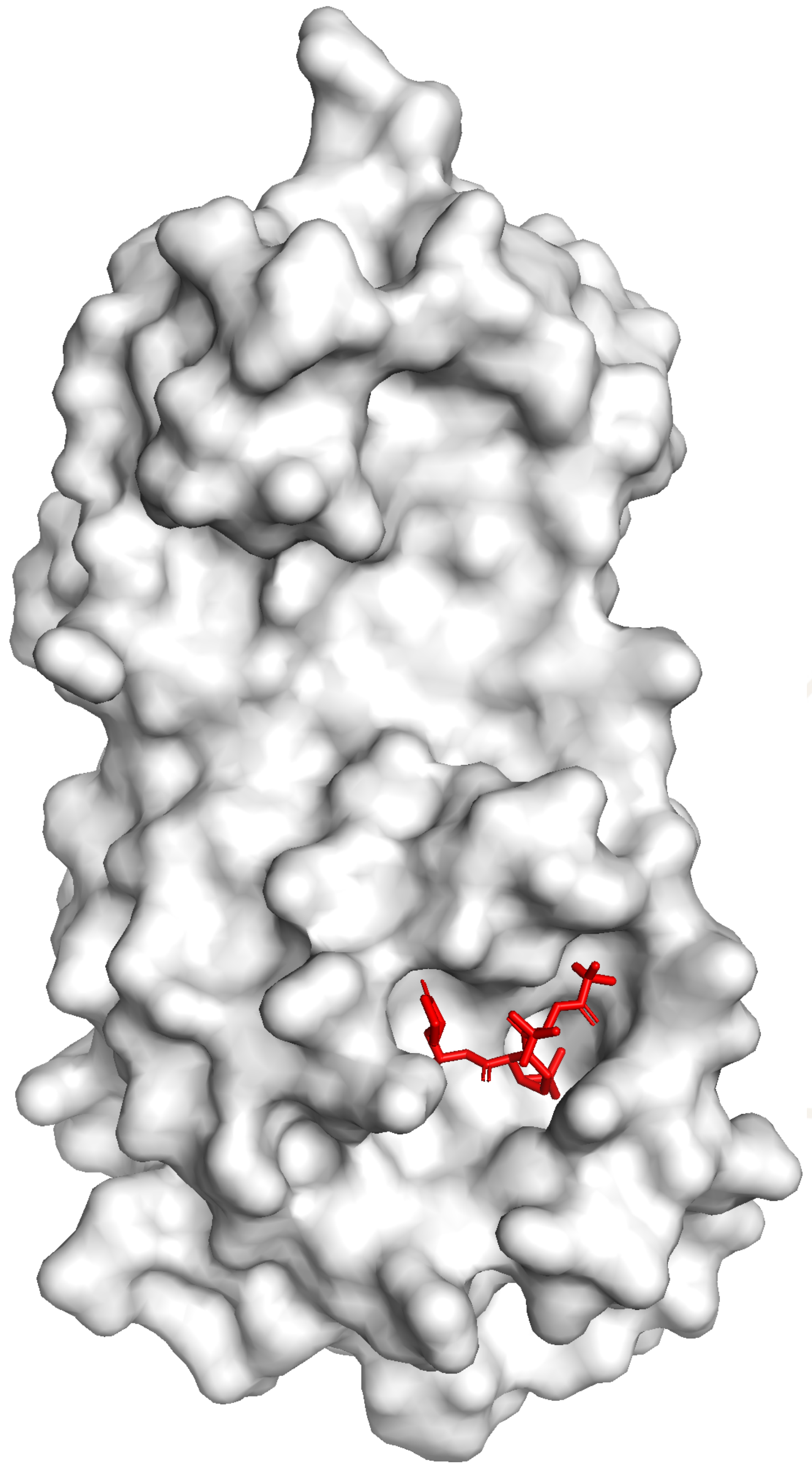


Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz,
Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy,
Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, Regina Barzilay

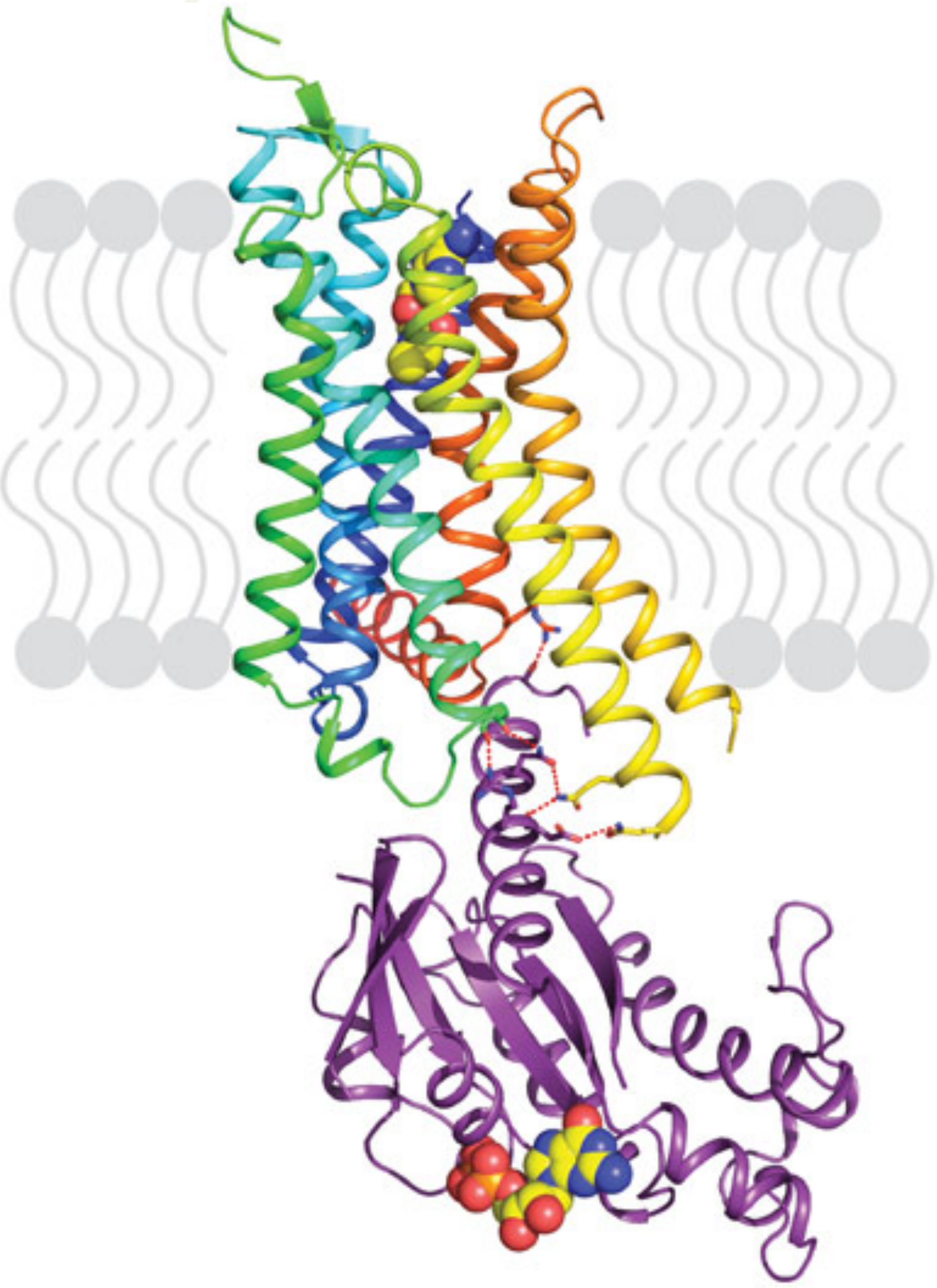
The importance of structure and affinity prediction



CRISPR-Cas9 interaction.
Image credit R. Andrade.

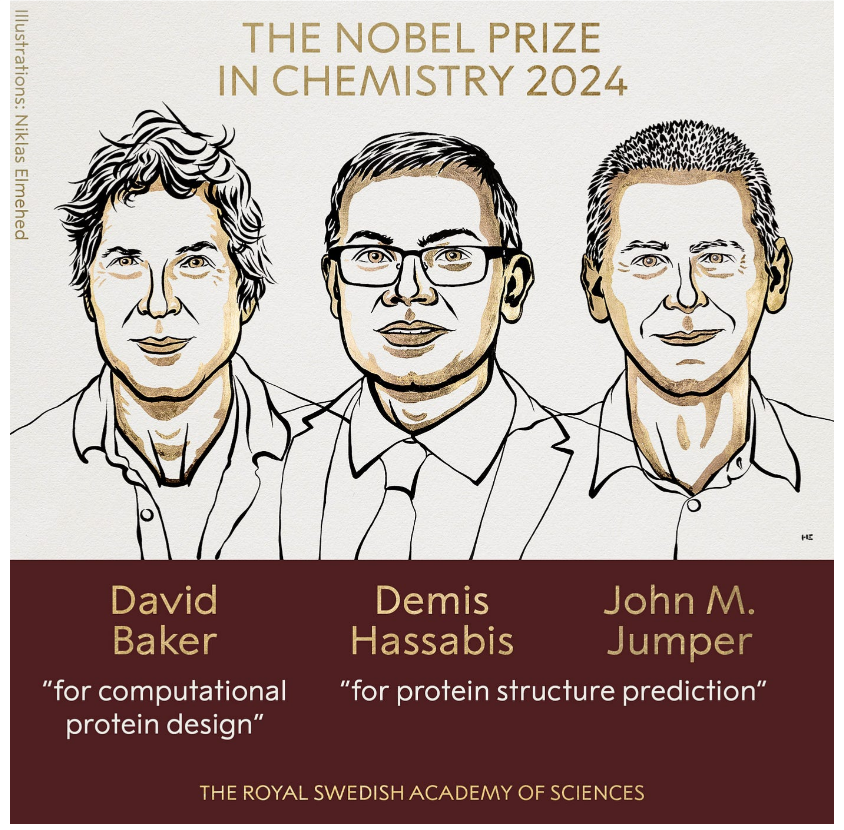


Interaction Paxlovid
(Pfizer) and SARS MPro



Engineered G protein.
Carpenter et al.

AlphaFold3 and Boltz-1



2020

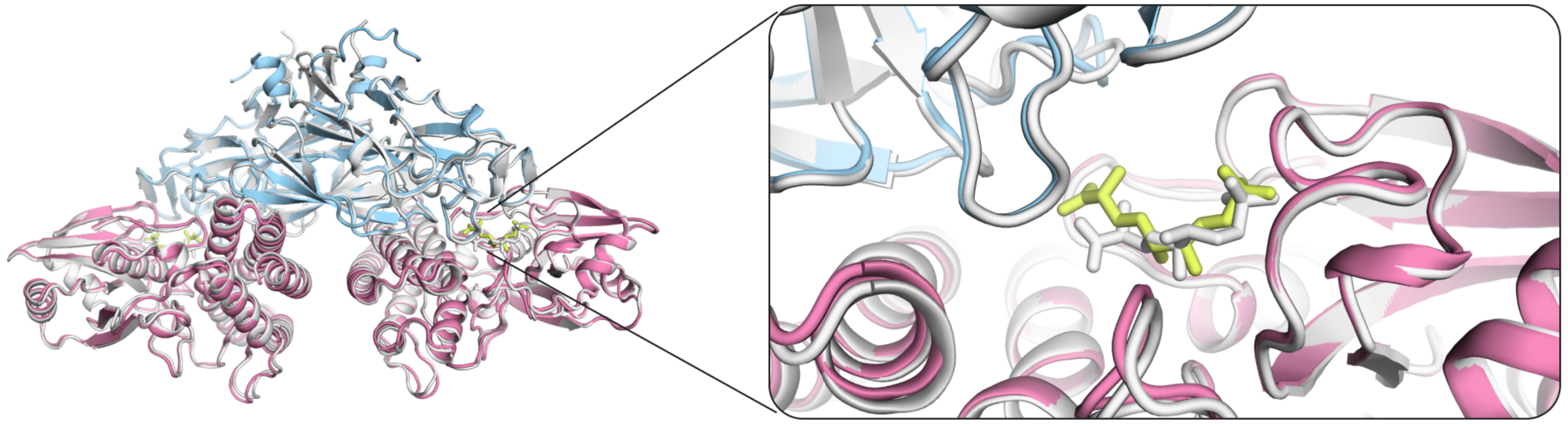
DeepMind announces AlphaFold2 for single chain protein structure prediction

May 2024

DeepMind publishes AlphaFold3 for arbitrary structure prediction without releasing the model

Nov 2024

Boltz-1 is the first fully open source model approaching AF3-level



Boltz adoption and feedback

2000+
members in the
community slack

2800+ stars
430+ forks
40+ contributors

Beyond structure: Affinity Prediction

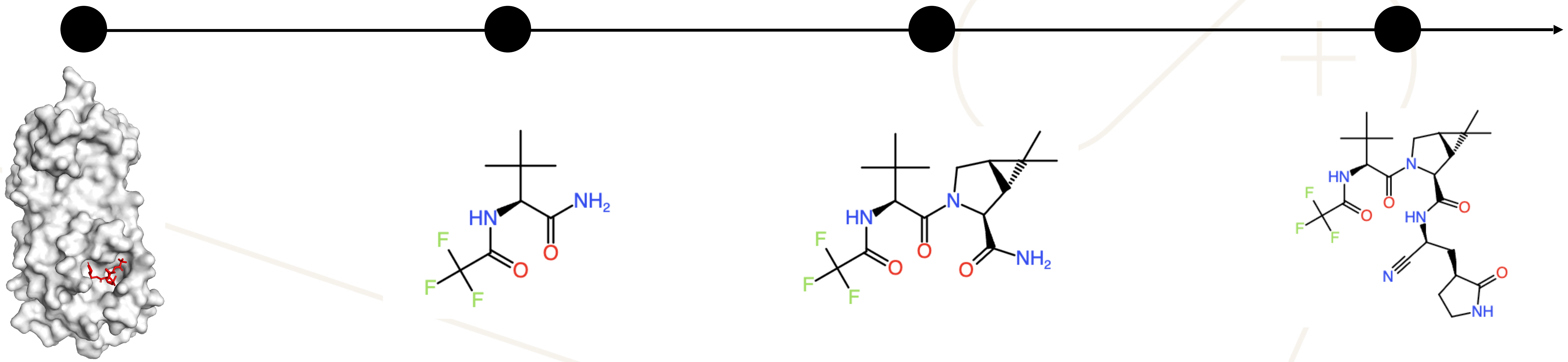
Accurate Binding Affinity calculation is likely the biggest bottleneck in preclinical drug development.

Target discovery
Find the right protein target for the disease

Hit-discovery
Find an initial small molecule that binds

Hit-to-lead
Maximize the affinity to the target

Lead-optimization
Minimize off-target interactions

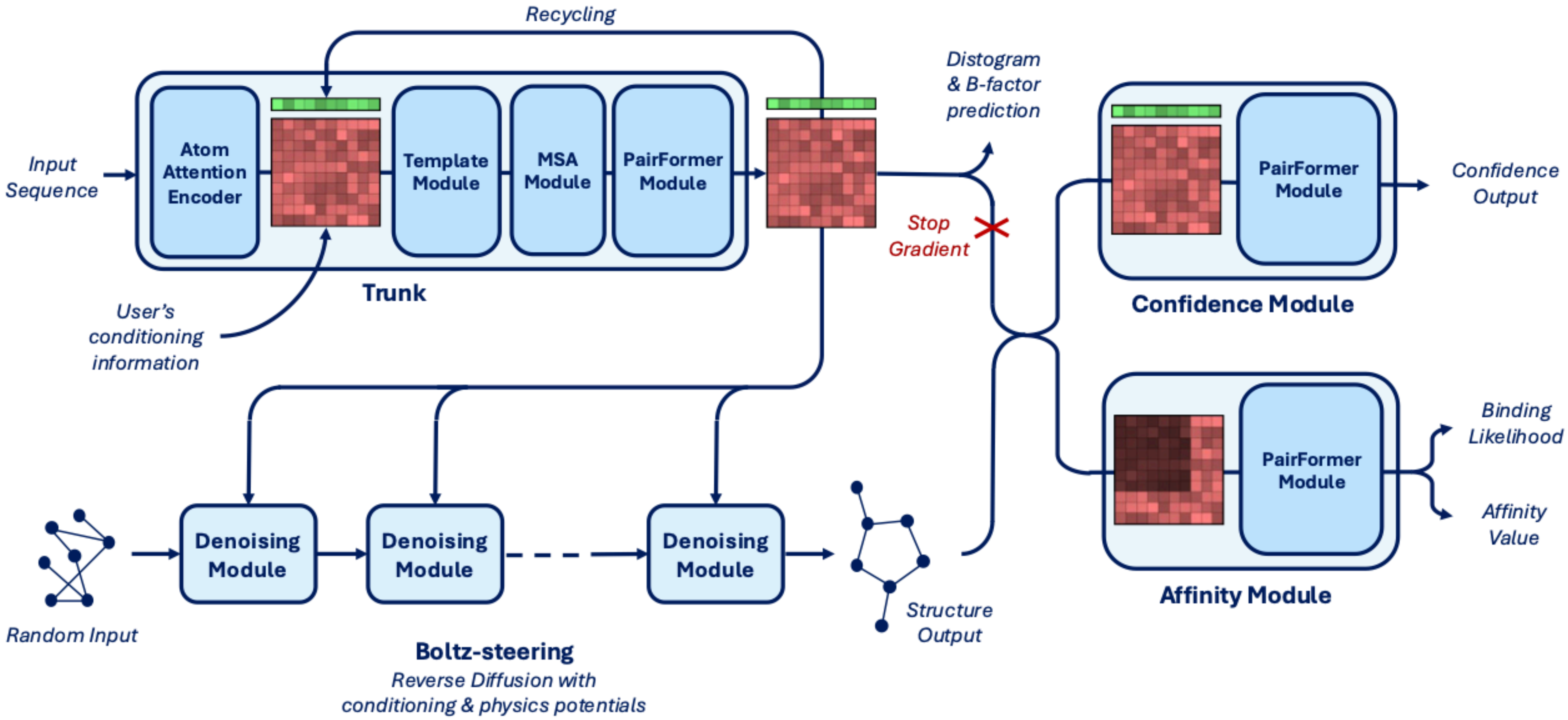


Invivo & clinical trials

While we got good at structure prediction, across all these stages the **cost & time** of precise **binding affinity measurements** limit the number of molecules & proteins that are tested.

Boltz-2 Structure Prediction

Boltz-2 Architecture



Boltz-1 Feedback

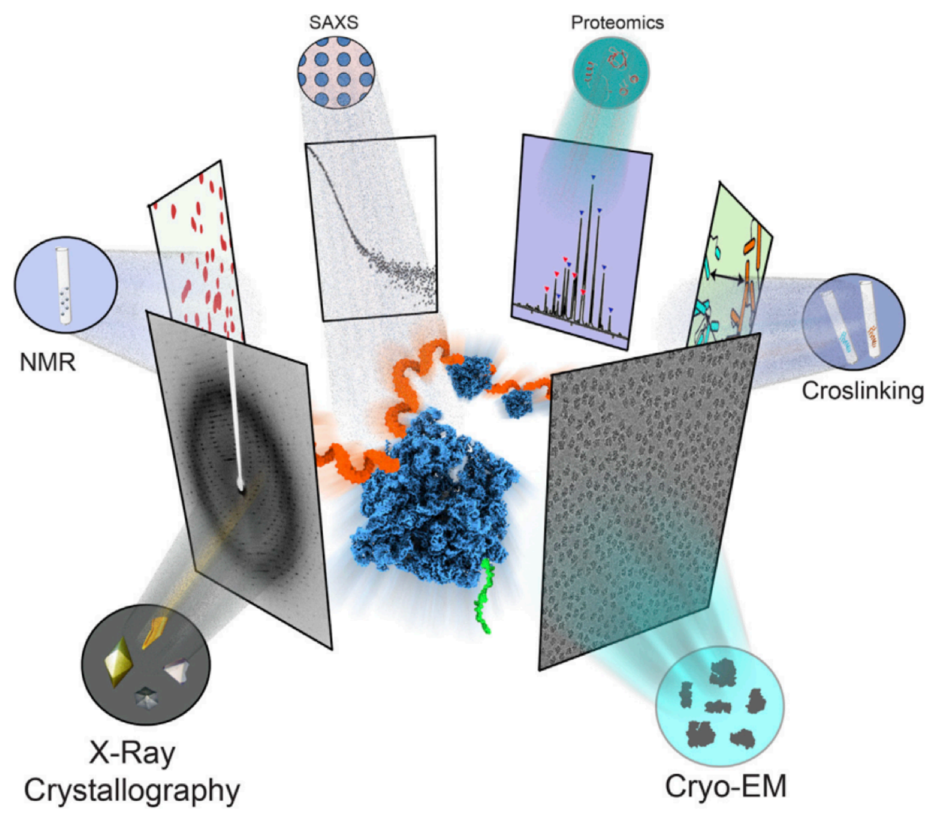
- User control in terms of templates, contacts, and other domain knowledge
- Modeling dynamics
- Improving the physical quality of poses
- Faster speed and lower memory consumption

Boltz-1 Feedback

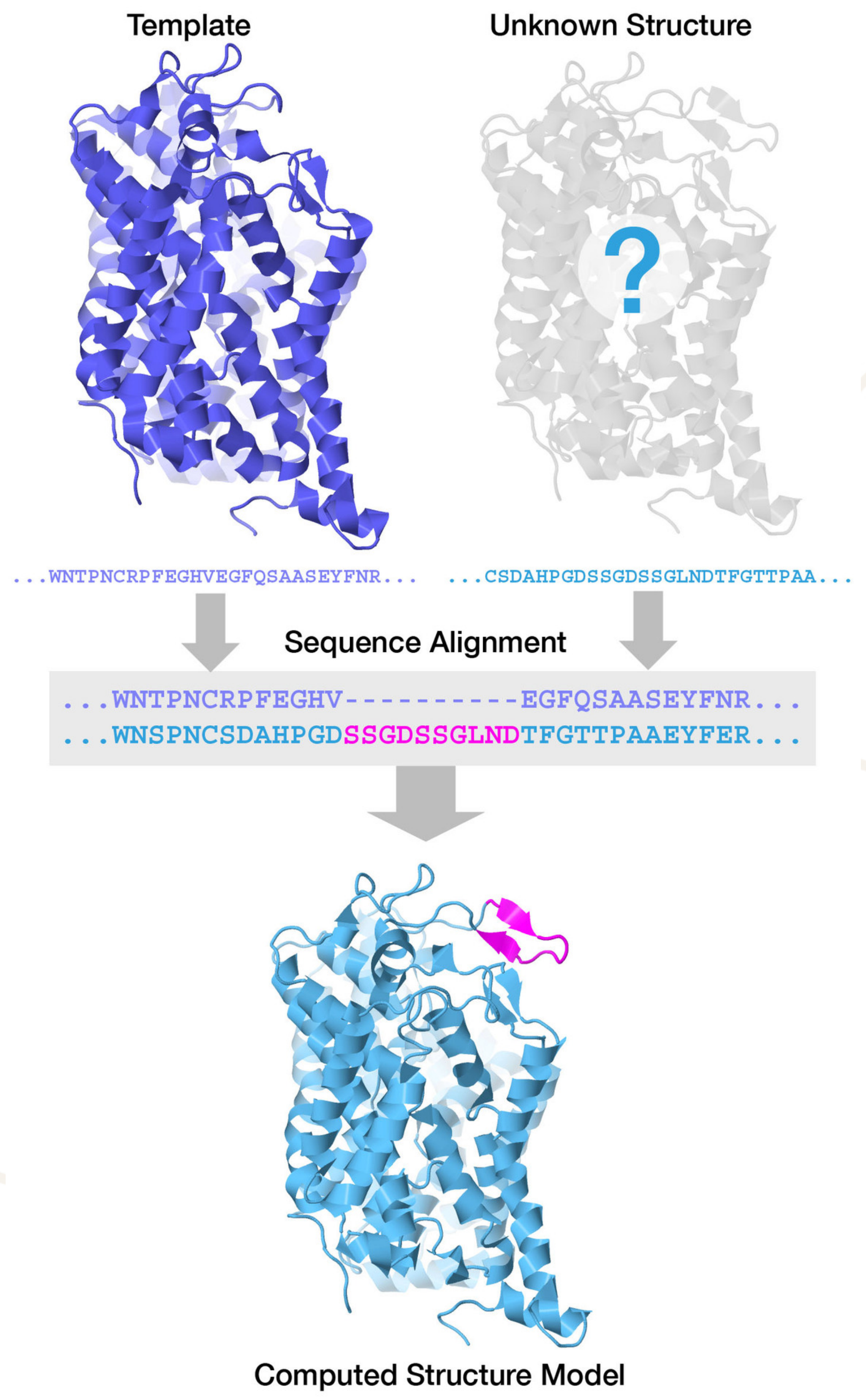
- **User control in terms of templates, contacts, and other domain knowledge**
- Modeling dynamics
- Improving the physical quality of poses
- Faster speed and lower memory consumption

New ways to condition the model

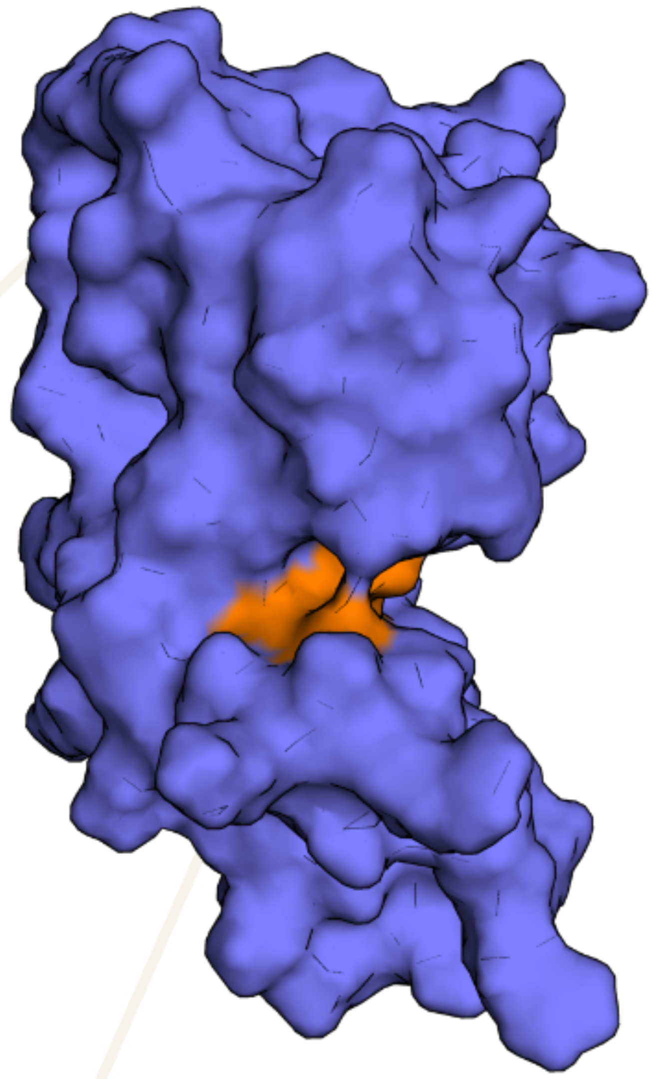
Method conditioning



Templates (+ Multimer)



Contacts or pocket



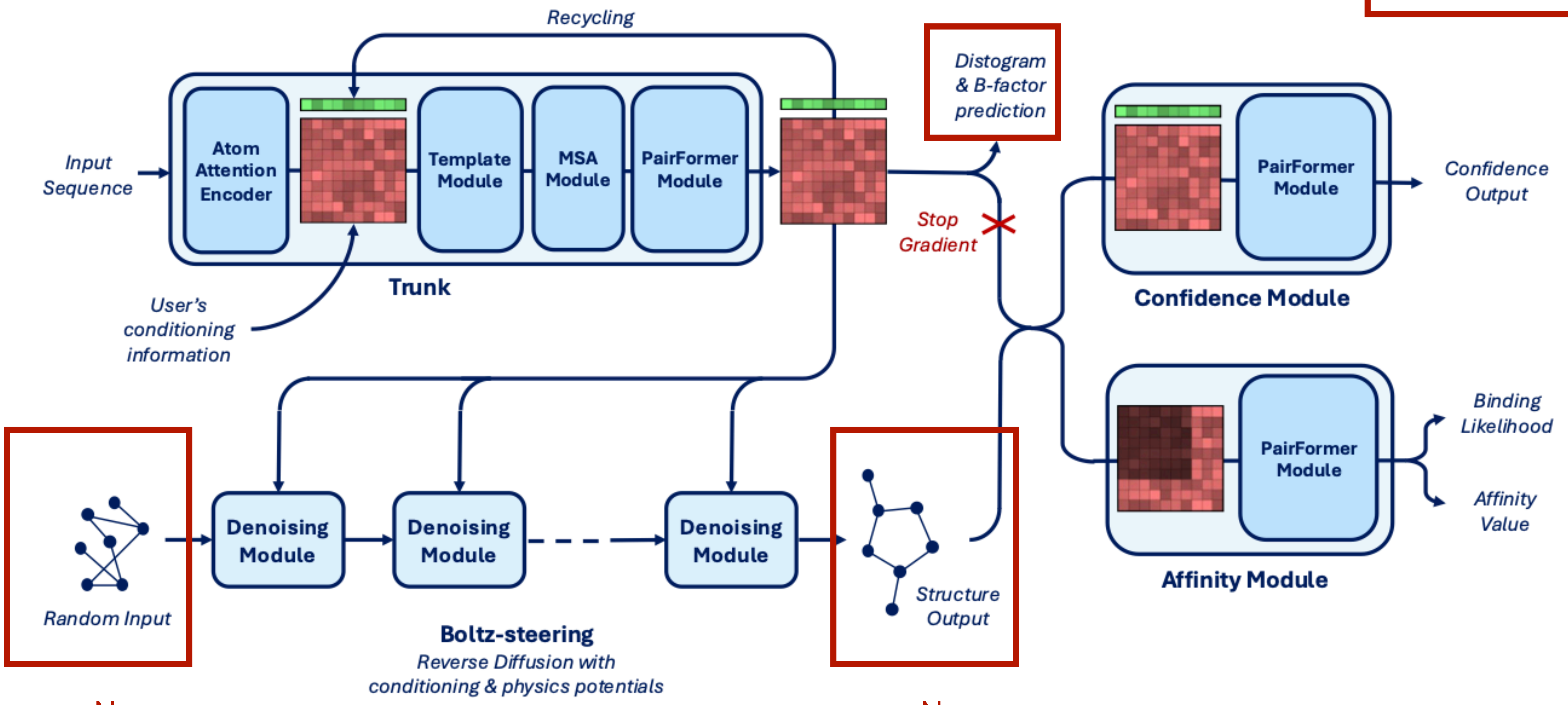
Boltz-1 Feedback

- User control in terms of templates, contacts, and other domain knowledge
- **Modeling dynamics**
- Improving the physical quality of poses
- Faster speed and lower memory consumption

Modeling Local Dynamics

MD datasets
MISATO
mdCATH
ATLAS

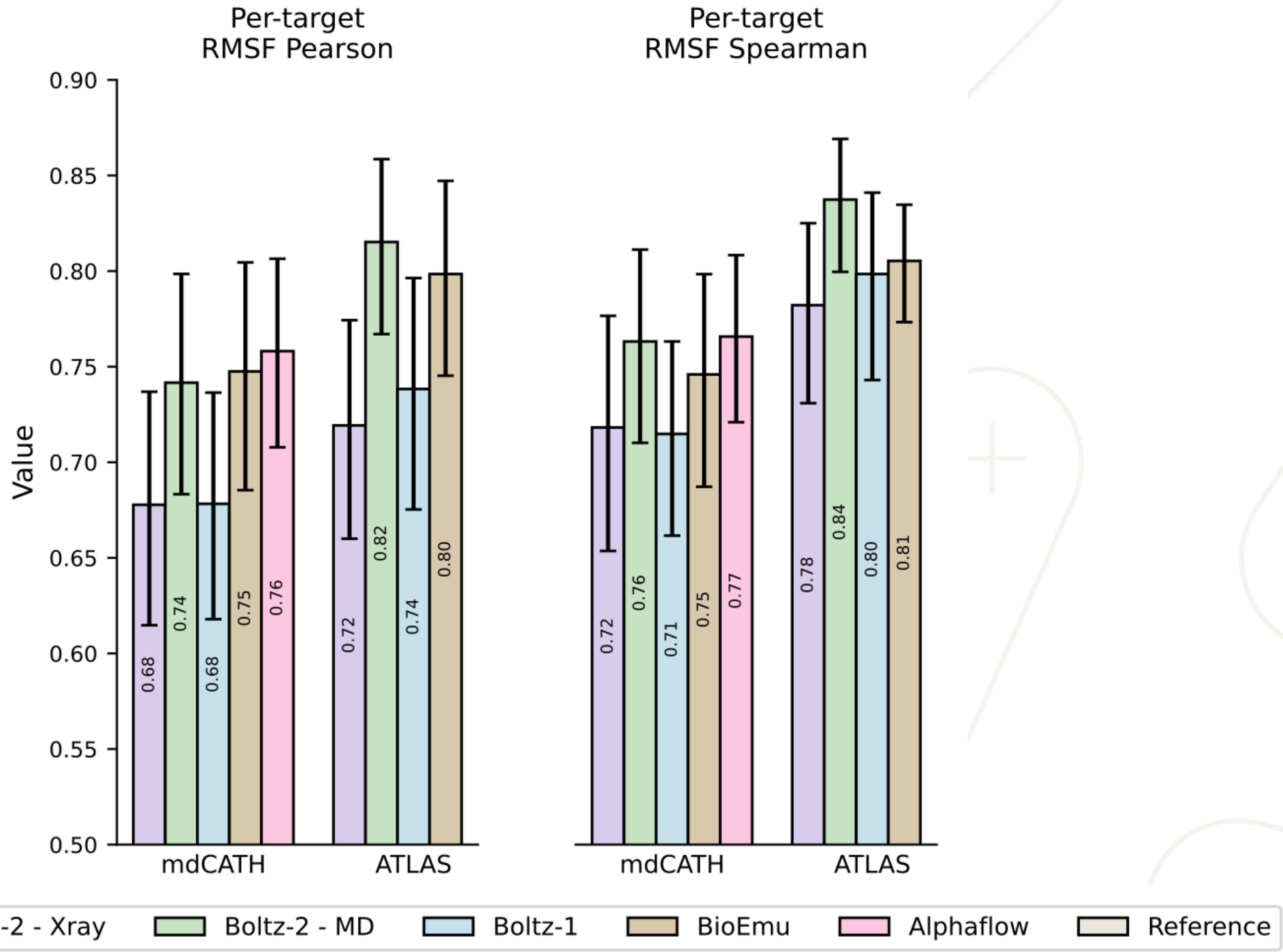
Distribution over N



x N

x N

Improvement on capturing local dynamics

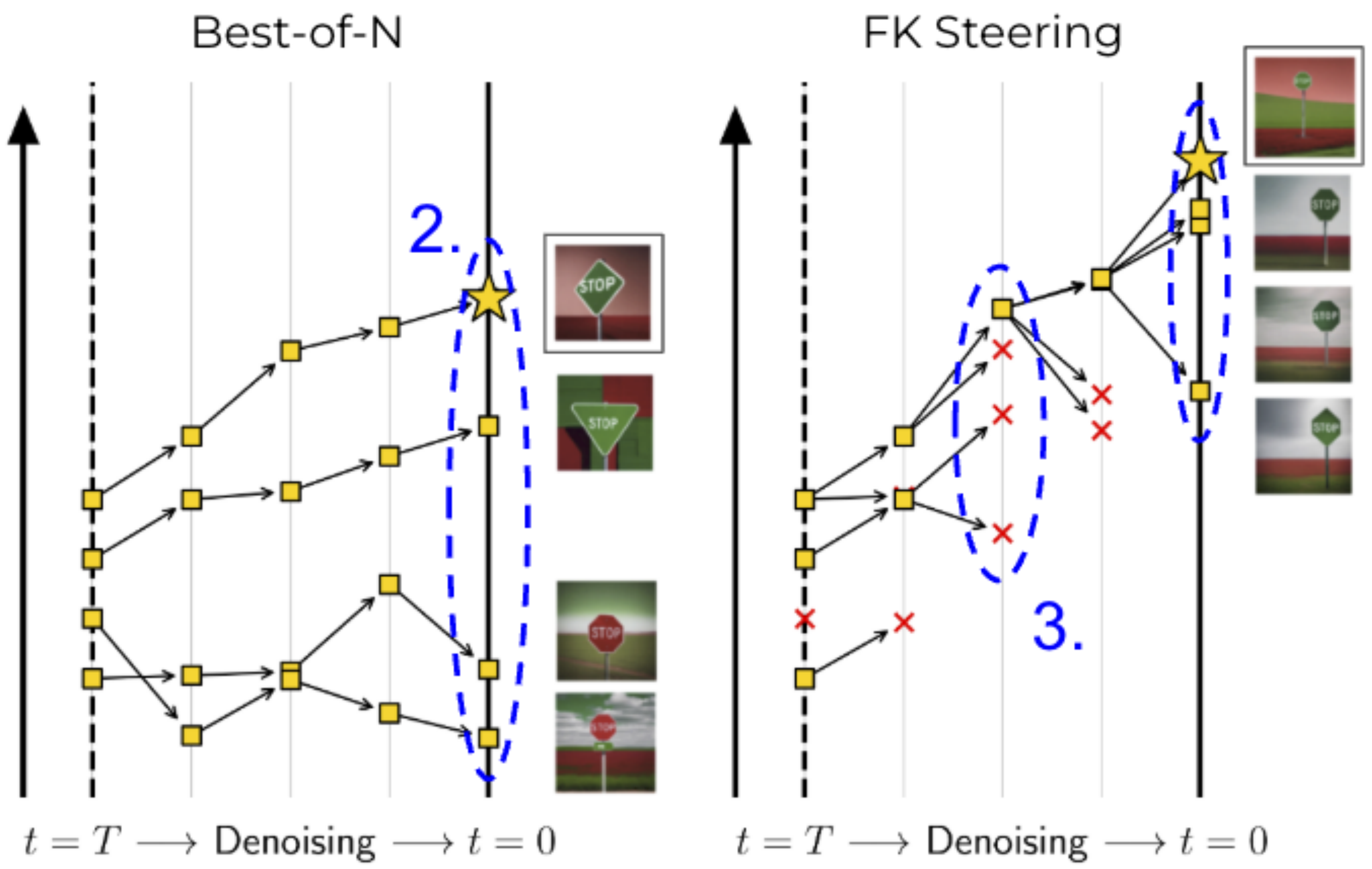


Boltz-1 Feedback

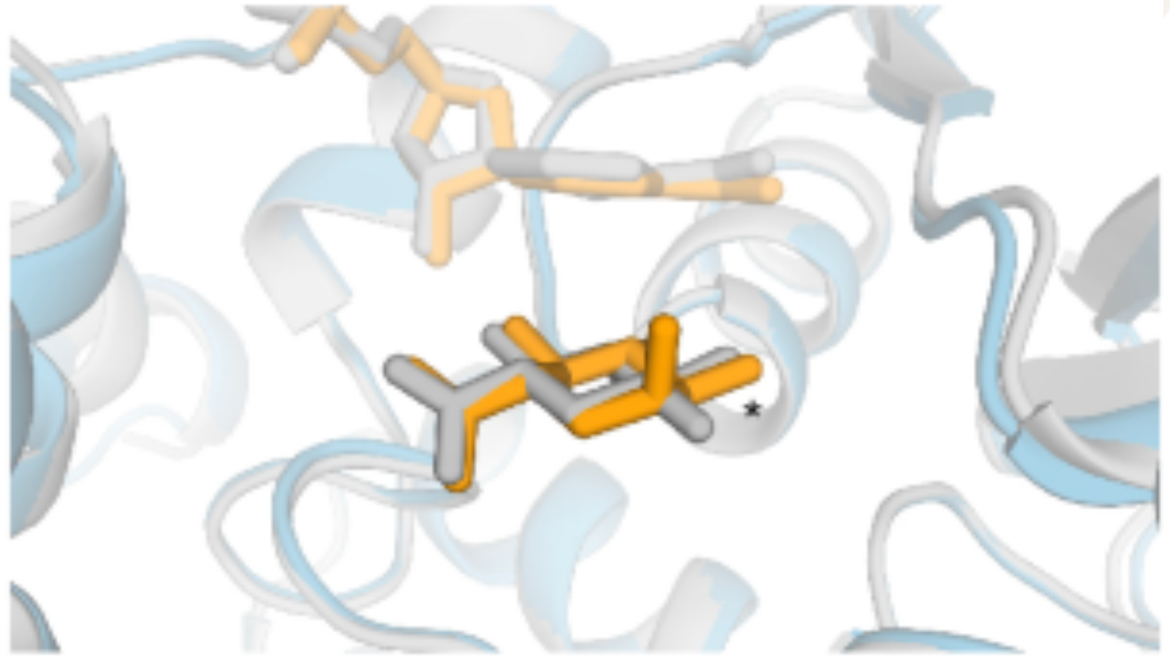
- User control in terms of templates, contacts, and other domain knowledge
- Modeling dynamics
- **Improving the physical quality of poses**
- Faster speed and lower memory consumption

Boltz-x Steering Potentials

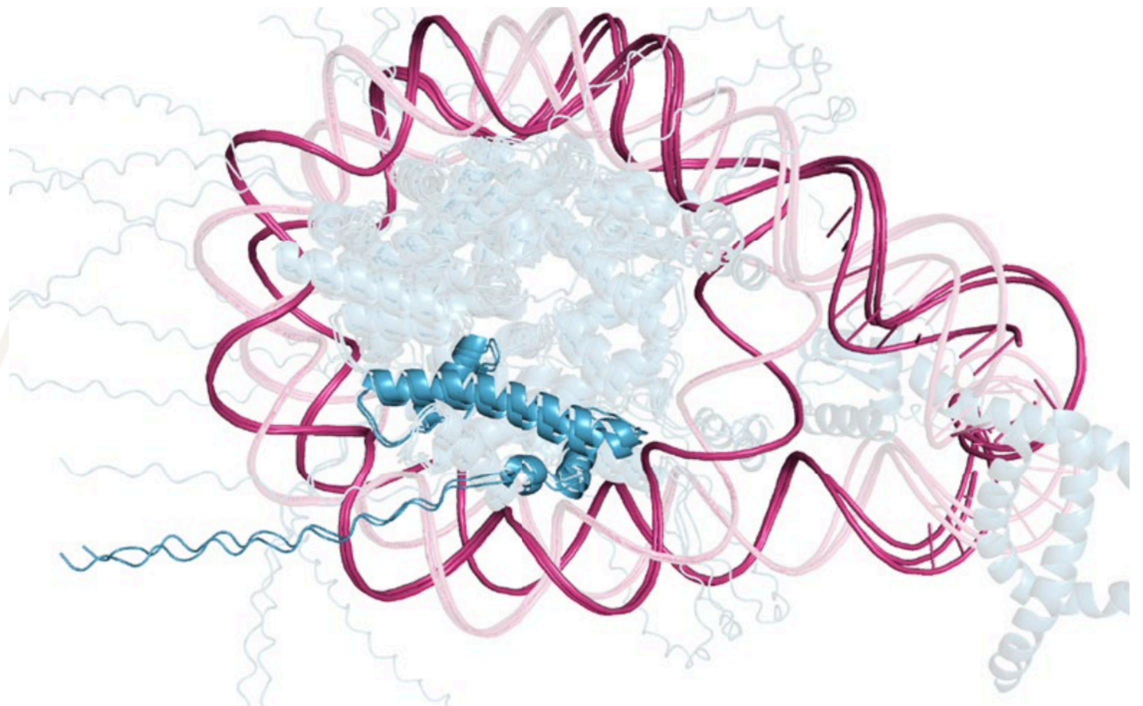
- Structure prediction models have so far had poor physical quality of the poses with issues with chirality (~5%), bond distances (~25%), stereochemistry, steric clashes and even overlapping chains.
- We addressed these by introducing a custom Feynman-Kac steering inference-time potentials



AlphaFold3 report

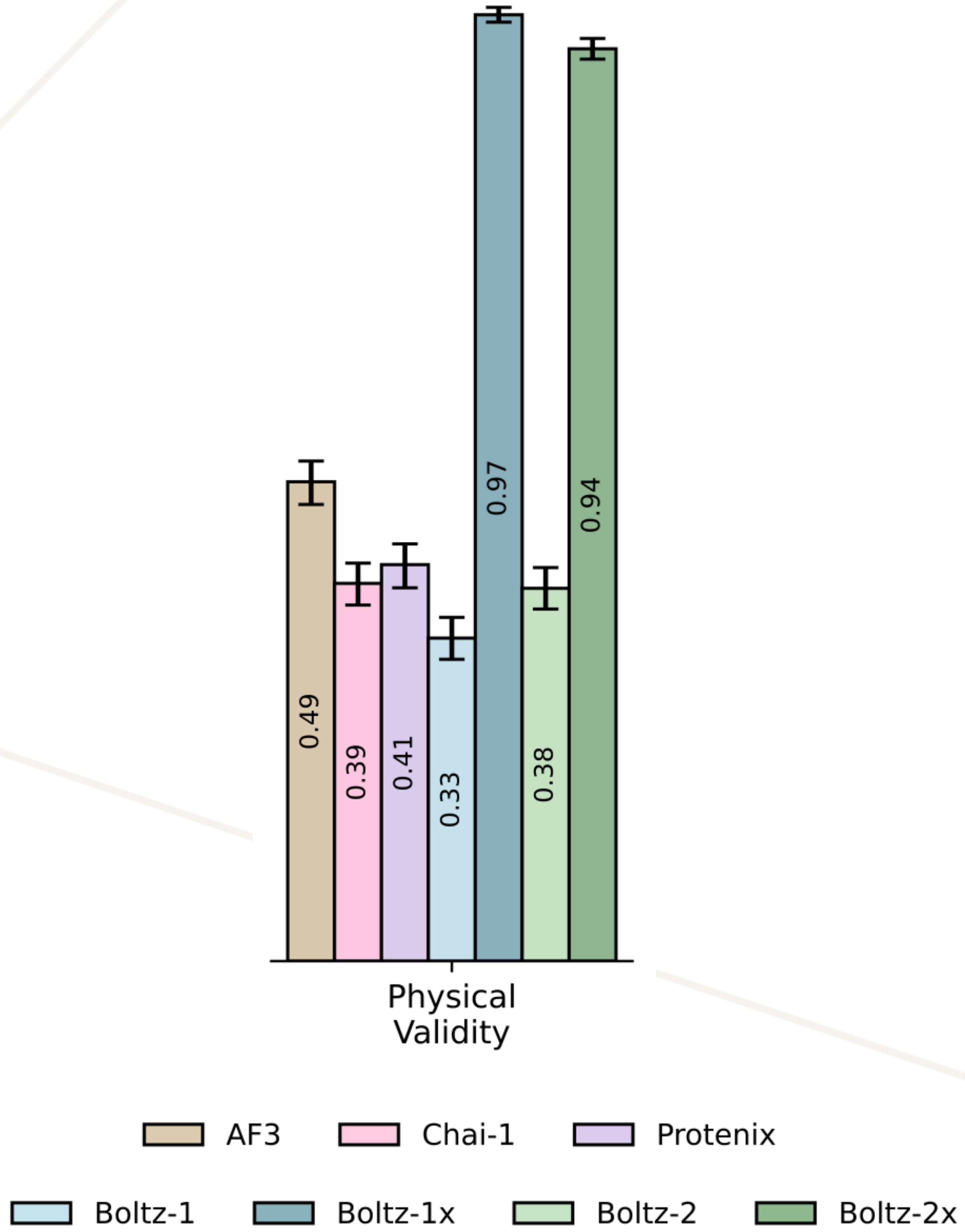


Chirality error

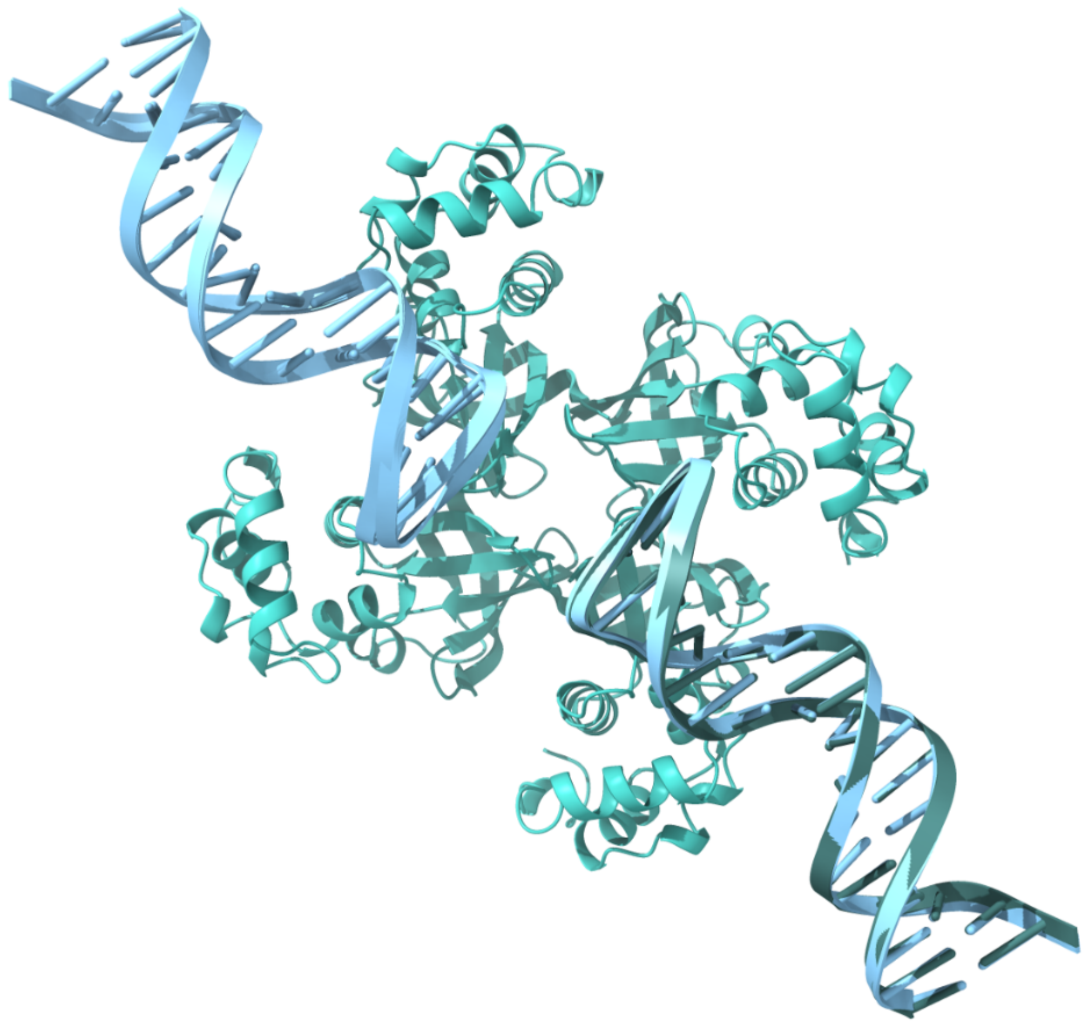


Clashing chains

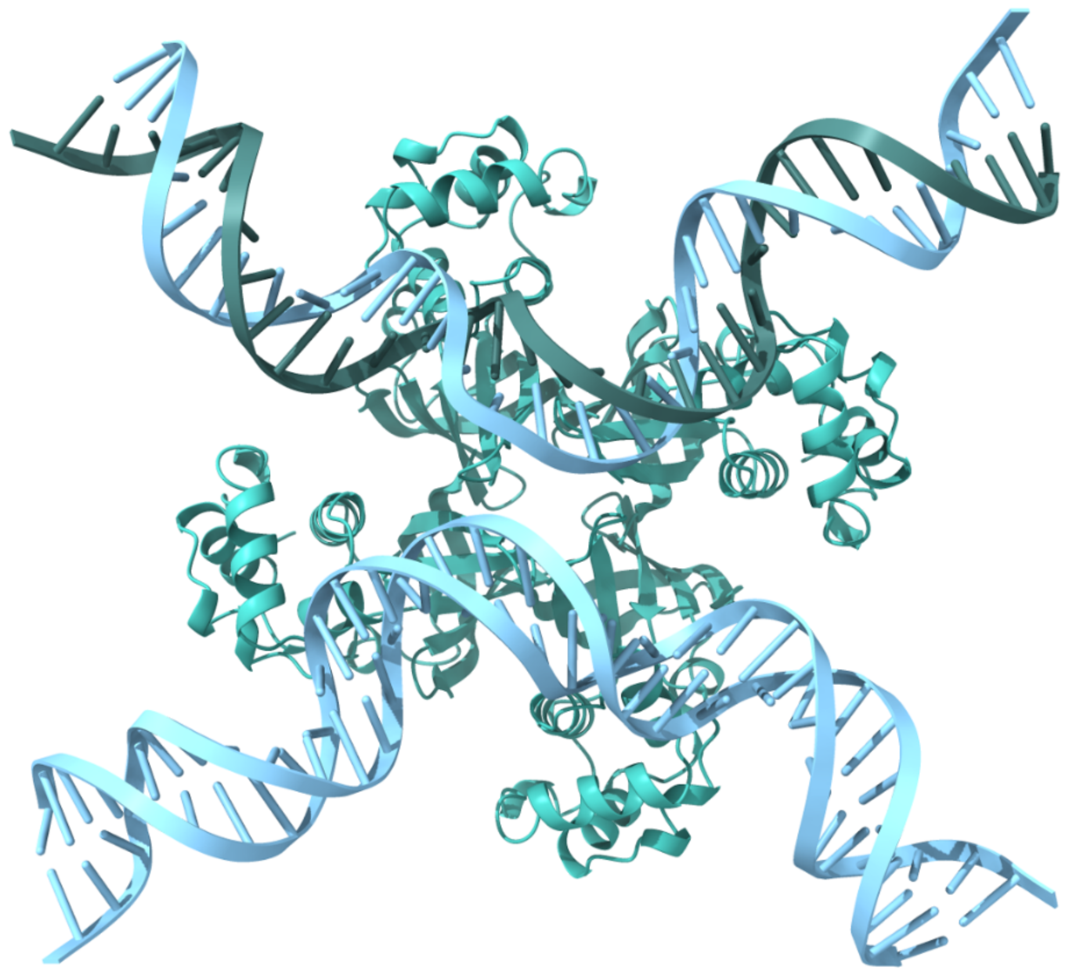
Boltz-1x & Boltz-2x for High Quality Poses



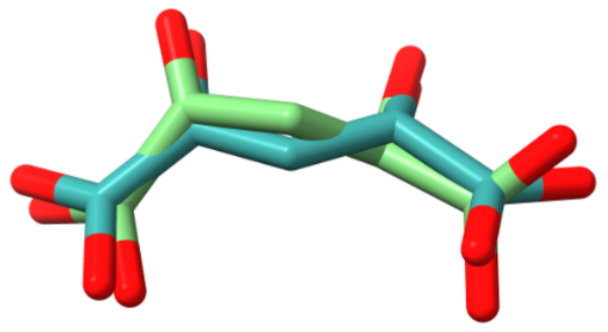
Boltz-1 PDB id: 8GH8
Overlapping DNA chains



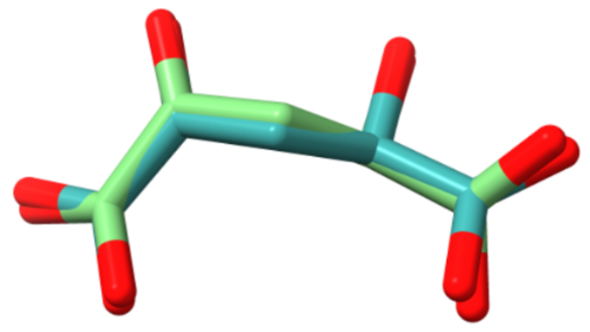
Boltz-1x PDB id: 8GH8
Physical DNA chain placement



Boltz-1 PDB id: 8SUT
Wrong small molecule chirality



Boltz-1x PDB id: 8SUT
Correct small molecule chirality

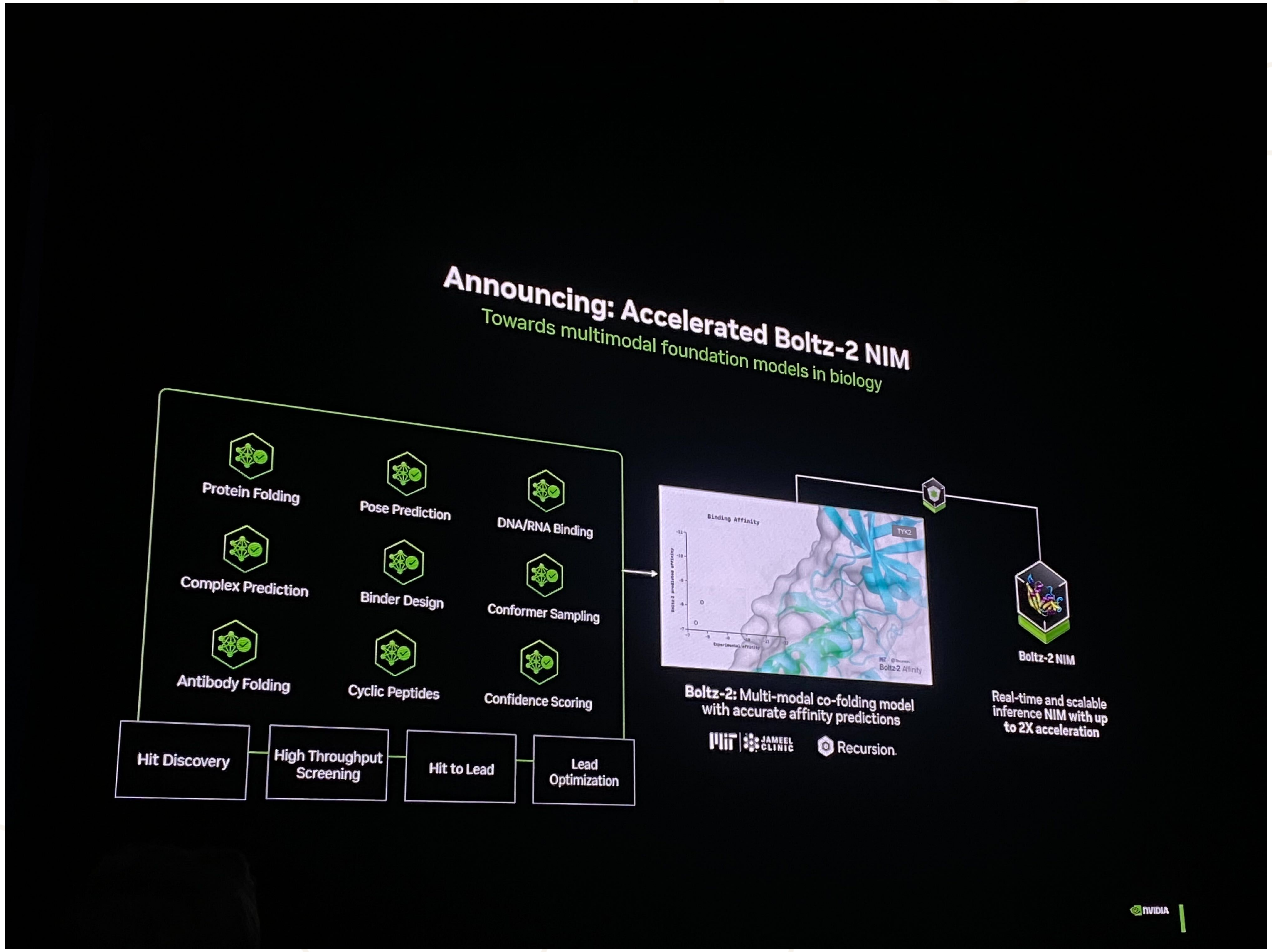


Boltz-1 Feedback

- User control in terms of templates, contacts, and other domain knowledge
- Modeling dynamics
- Improving the physical quality of poses
- **Faster speed and lower memory consumption**

Scalability

Just announced collaboration with NVIDIA to accelerate Boltz-1 & Boltz-2!



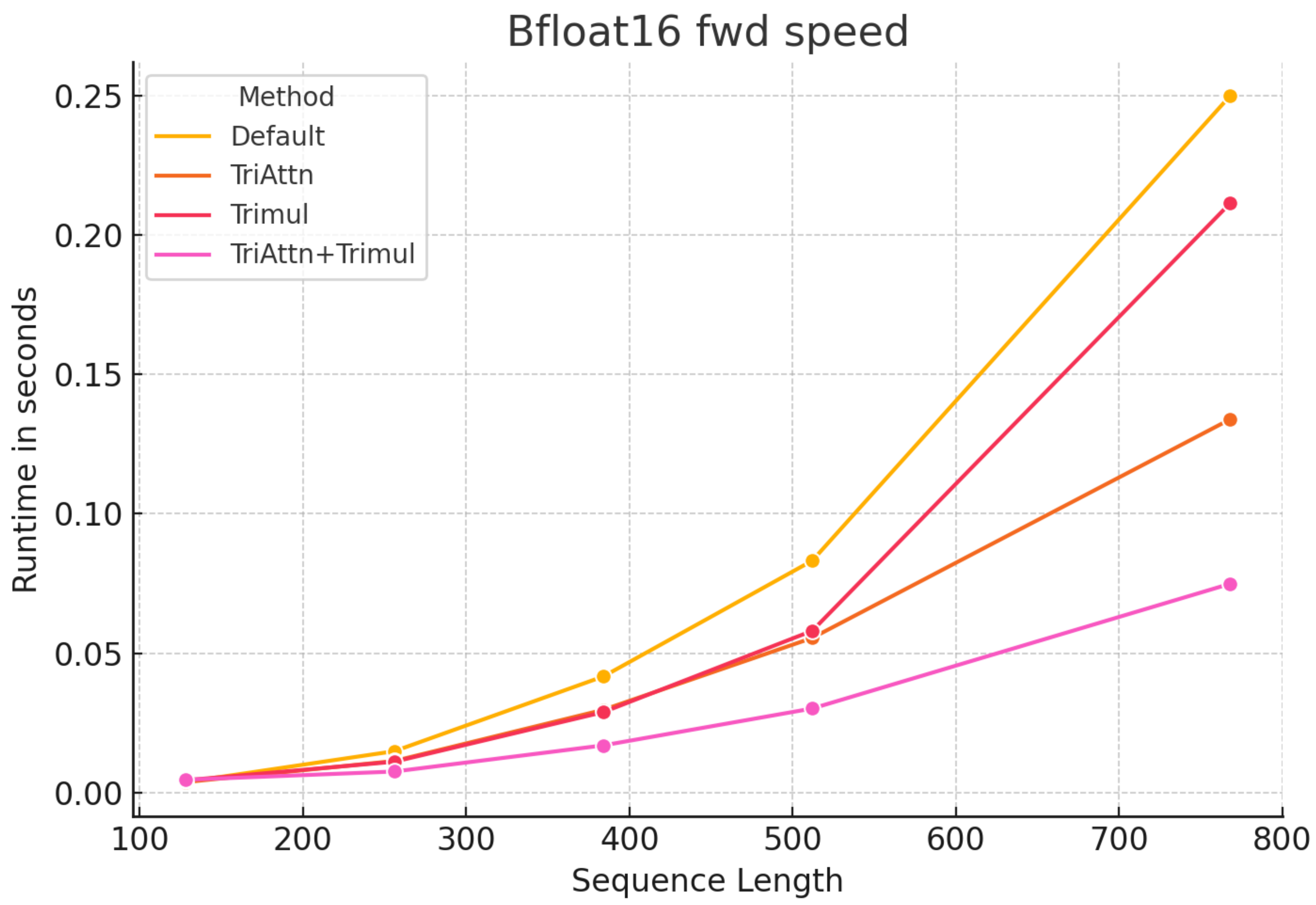
Announcing: Accelerated Boltz-2 NIM

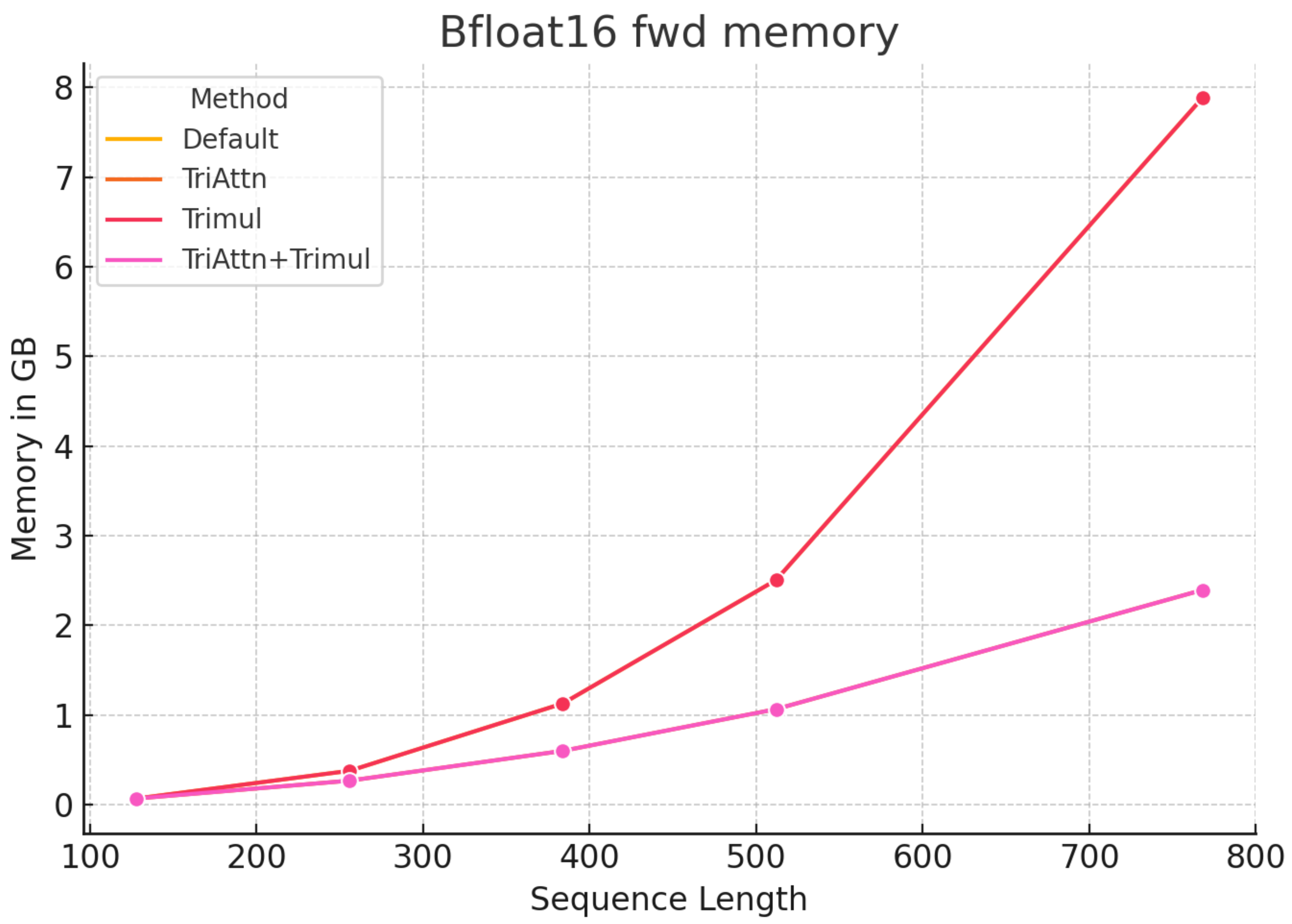
Towards multimodal foundation models in biology

Boltz-2: Multi-modal co-folding model with accurate affinity predictions

Real-time and scalable inference NIM with up to 2X acceleration

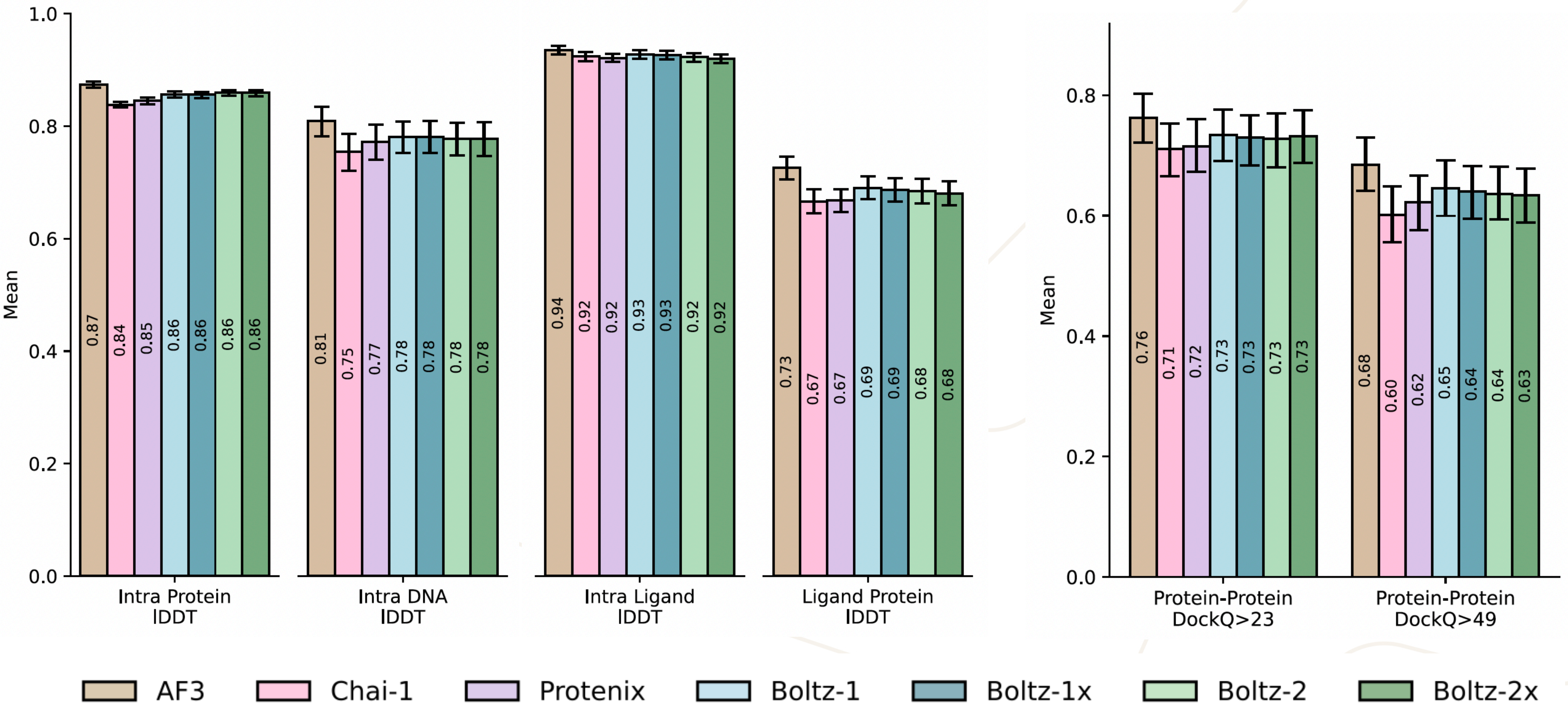
Scalability


Bfloat16 fwd speed

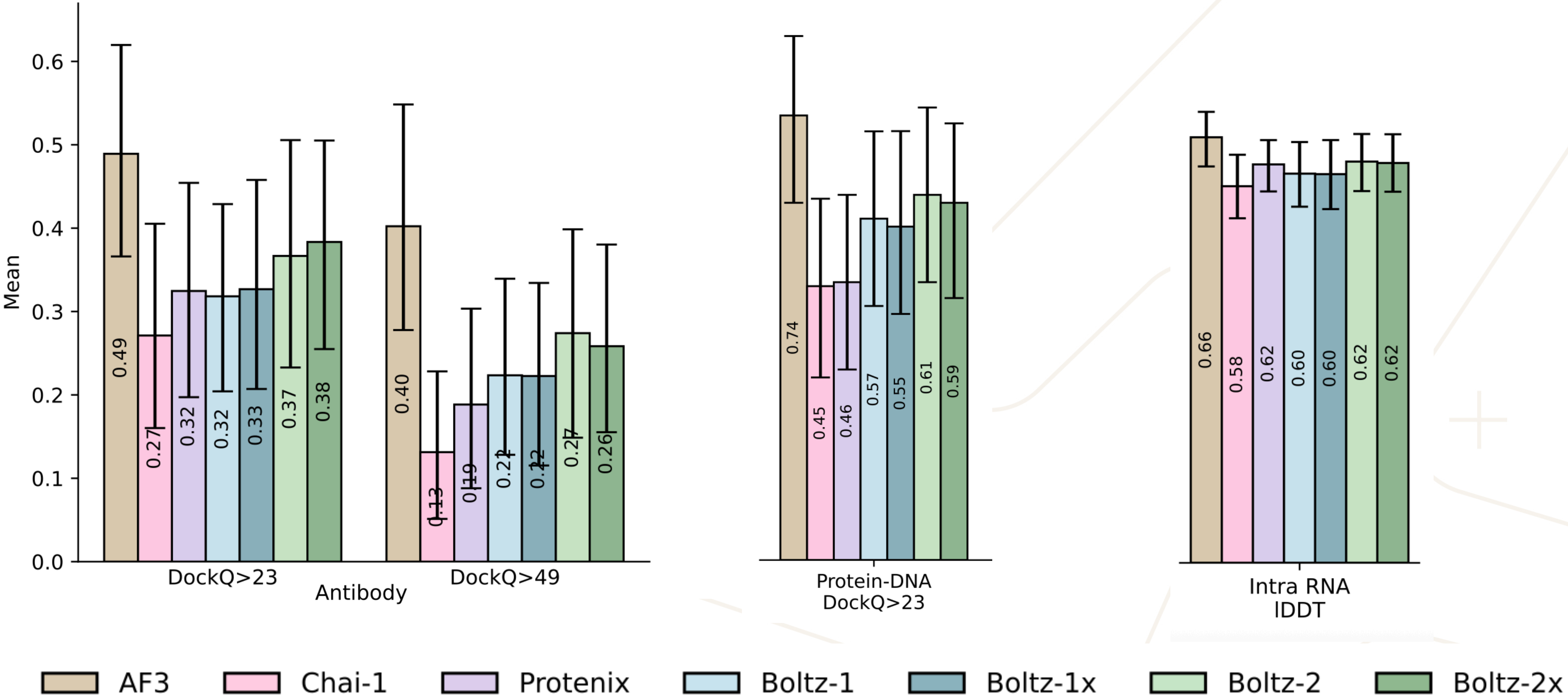

Bfloat16 fwd memory

Now live on the Boltz repo!

On “easier” modalities similar performance



Improvement in hard modalities

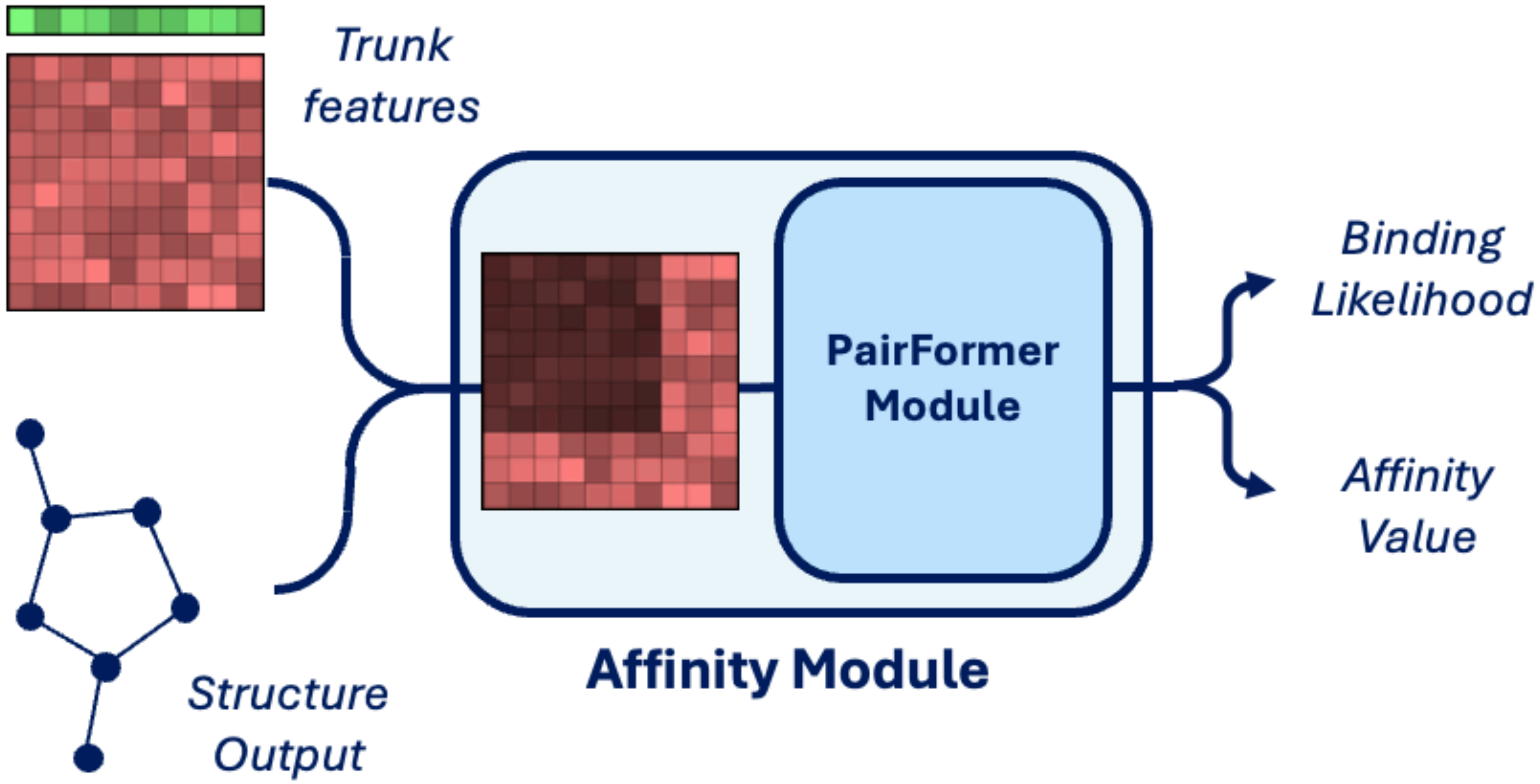


Boltz-2 Affinity Prediction

Key components for accurate affinity prediction

- Exploit deep representations learned by structural model.
- Curation of large amounts of good quality data.
- Designing supervision framework to extract signal from noise.
- Counter biases in the data with synthetic data and custom sampling.

Affinity module architecture



- PairFormer-based module
- Inputs:
  - Trunk **representation**
  - Predicted **coordinates**
- Output:
  - Binding Likelihood → **hit discovery**
  - Affinity Value (IC50) → **hit-to-lead** and **lead-optimisation**

Affinity data

Type	Affinity values	Binary data
Sources	PubChem, CheEMBL, BindingDB	PubChem, CheEMBL, BindingDB, CeMM, MIDAS
Total Size	**~1.5M datapoints** across **2.5k targets**	**~0.5M positives** and **~5M negatives** datapoints across **2.5k targets**

Critical data curation:

- **Keep highest quality data:** filter low confidence assays and ligands.
- **Avoid false positives**: retain HTS binders only with a dose-response measurement.
- **Reduce size bias**: augment the binary classification dataset with synthetic decoys.

Training Affinity: extracting signal over noise

- **Binary focal loss** → avoid overfitting to positives

Training Affinity: extracting signal over noise

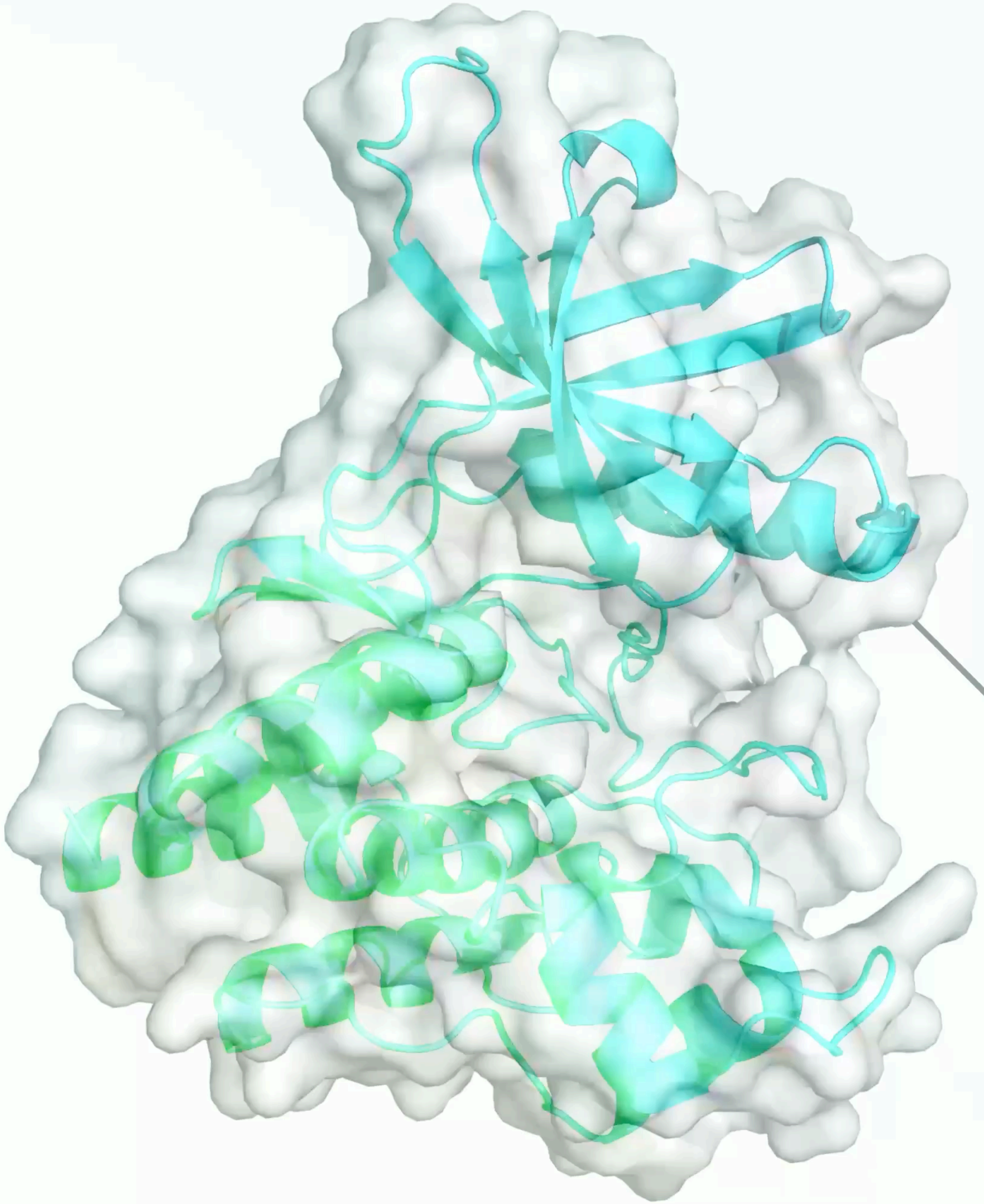
- **Binary focal loss** → avoid overfitting to positives
- **Structural confidence filter** (ipTM ≥ 0.75)
  → Avoid confounding the model with wrong structures

Training Affinity: extracting signal over noise

- **Binary focal loss** → avoid overfitting to positives
- **Structural confidence filter** (ipTM ≥ 0.75)
  - → Avoid confounding the model with wrong structures
- 90% weight on **relative difference** in affinity value loss:
  - → Reduce inter-lab experimental noise and merge multiple assay types

Training Affinity: extracting signal over noise

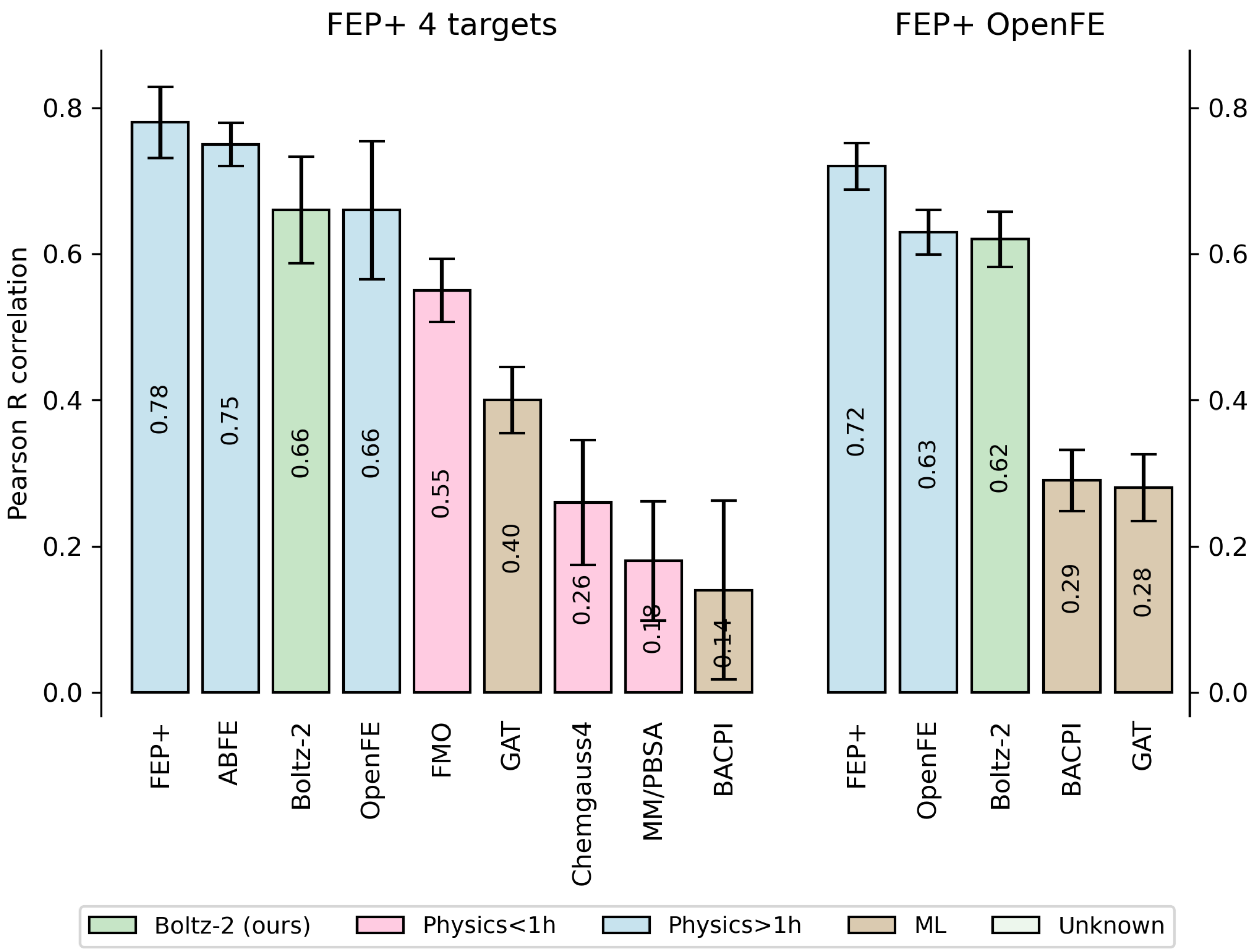
- **Binary focal loss** → avoid overfitting to positives
- **Structural confidence filter** (ipTM ≥ 0.75)
  - → Avoid confounding the model with wrong structures
- 90% weight on **relative difference** in affinity value loss:
  - → Reduce inter-lab experimental noise and merge multiple assay types
- **Activity Cliff Sampler**
  - → Focus on learning activity cliffs that are critical for ligand optimization

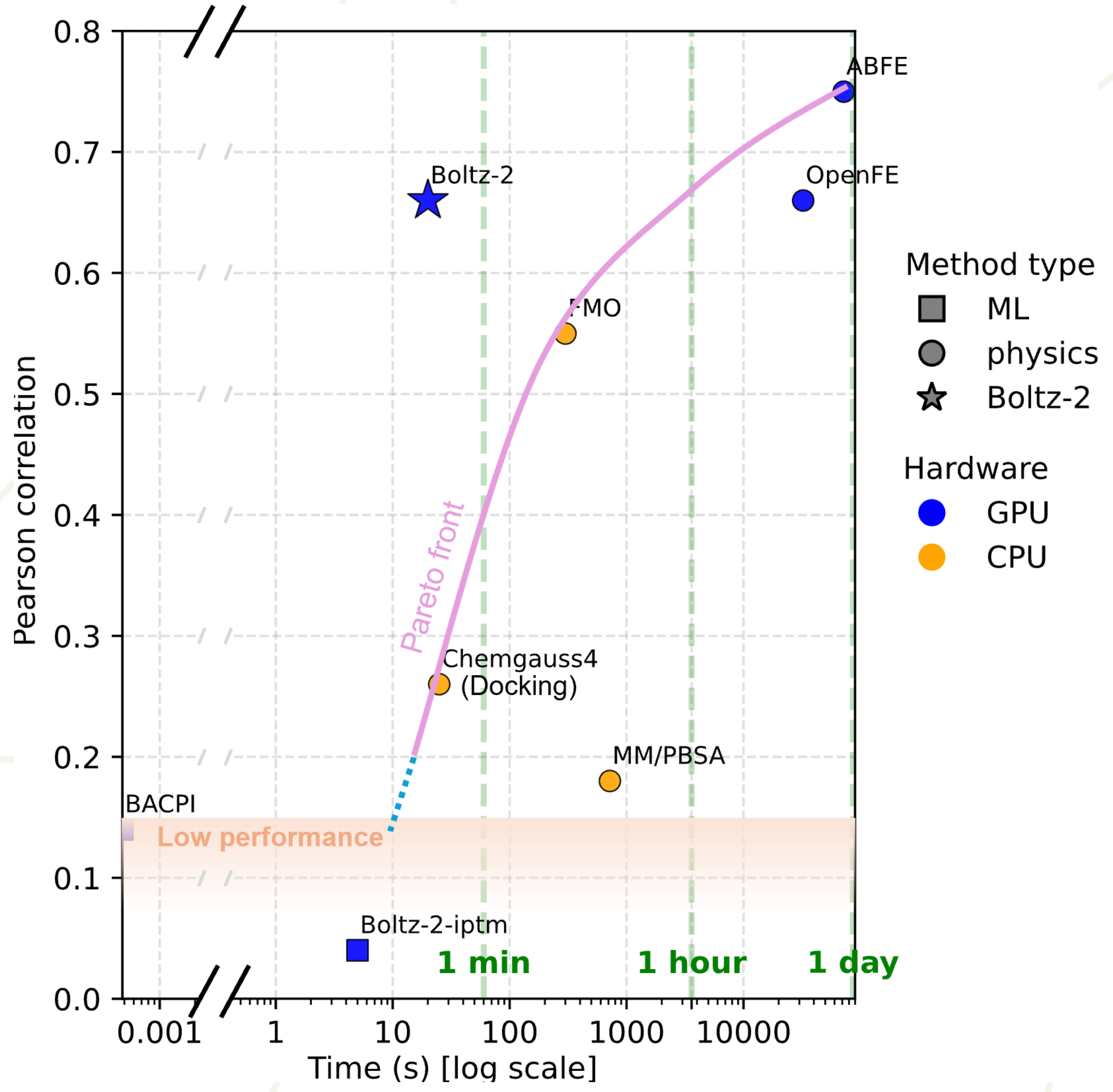


MIT × Recursion

Boltz-2 Affinity

Hit-to-lead results: FEP+ benchmark





Boltz-2 approaches FEP methods on the FEP+ benchmark at **1000x the speed**

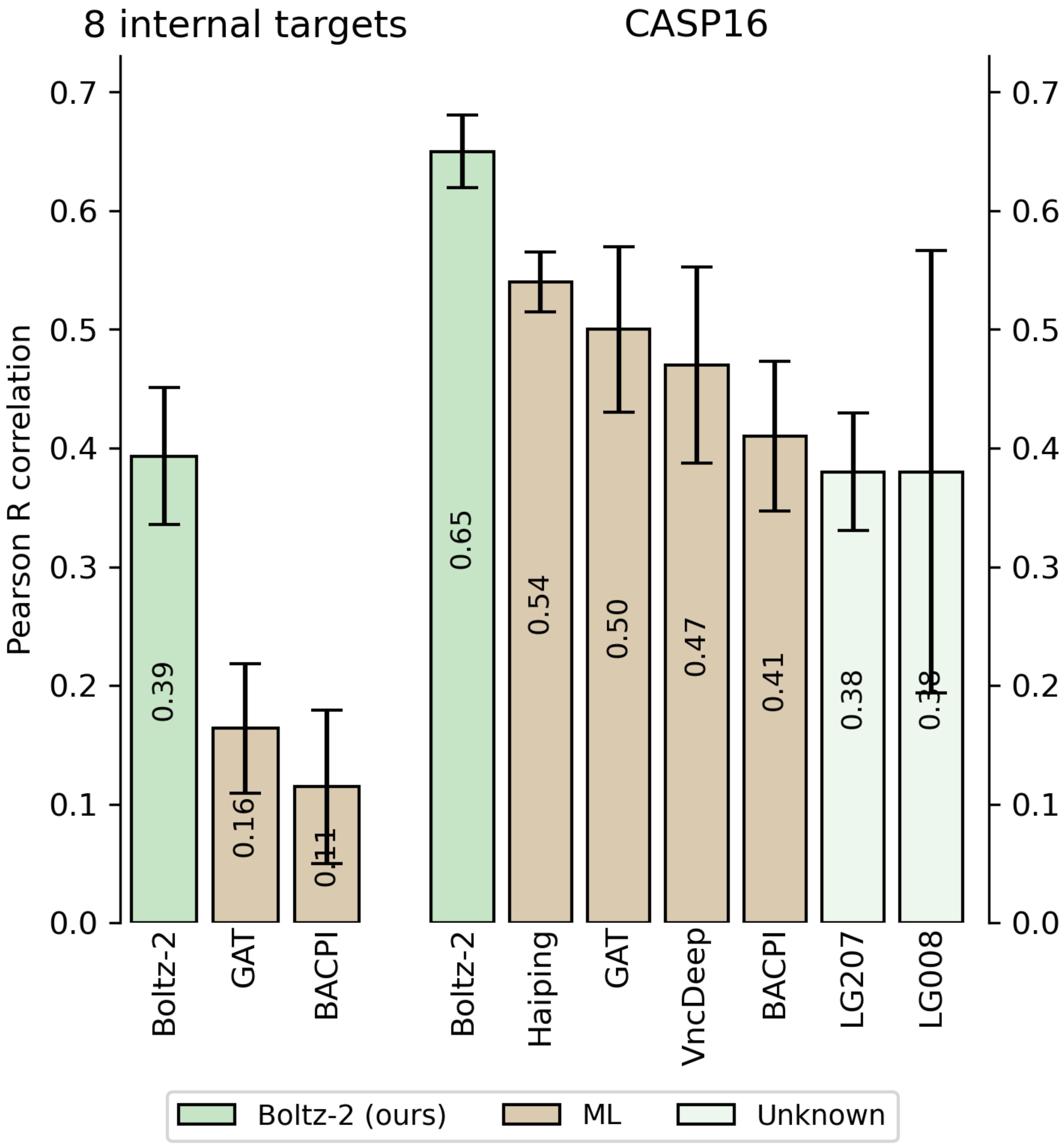
Hit-to-lead results: CASP16 and internal targets

CASP16 affinity challenge:

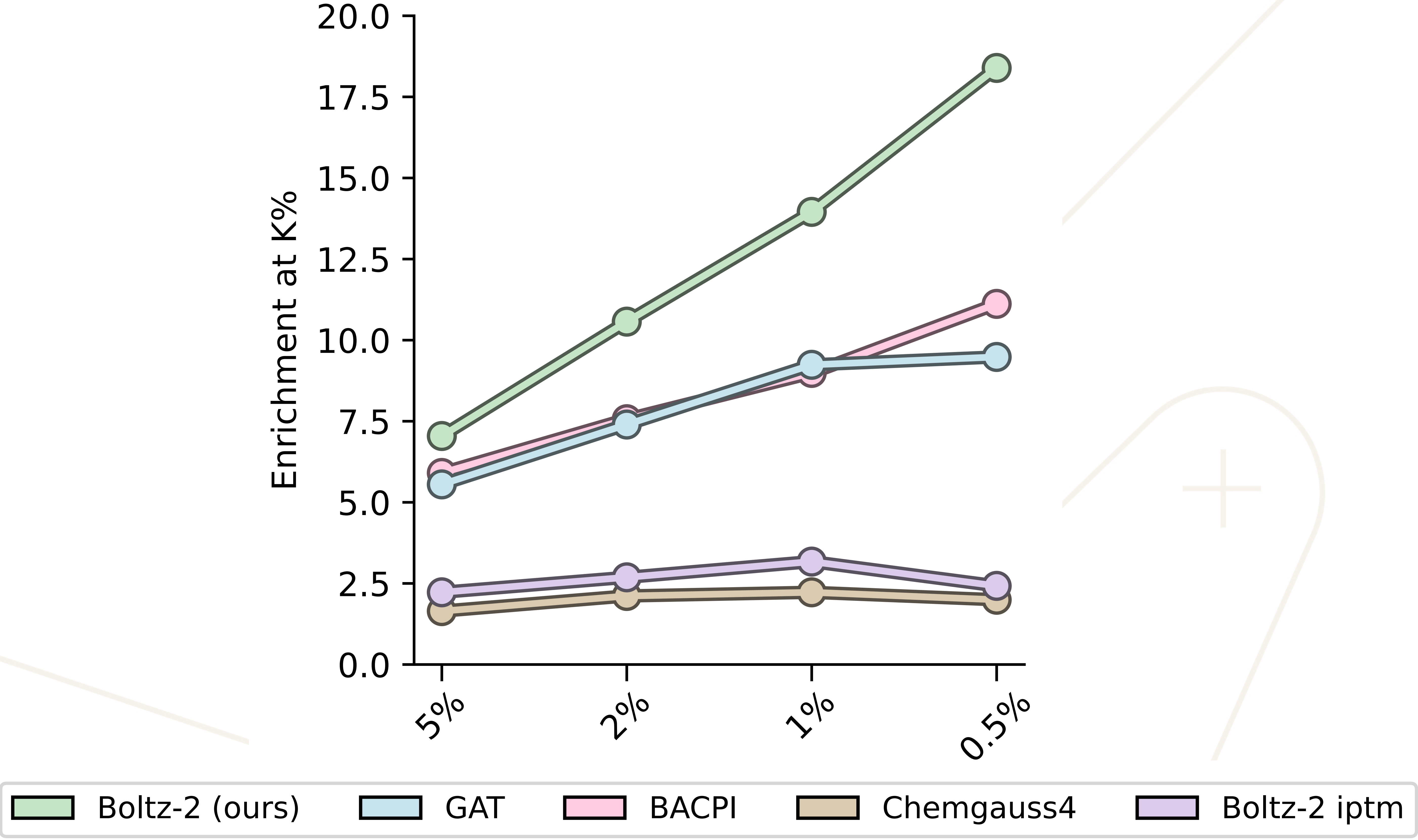
- 140 protein-ligand pairs across two targets from Roche
- 31 entries from top labs
- Boltz-2 retrospectively ranked 1st

Internal targets:

- Challenging real-world assays
- 8 hit-to-lead assays
- 3 out of 8 targets with >0.55 Pearson



Hit-discovery results: MF-PCBA benchmark

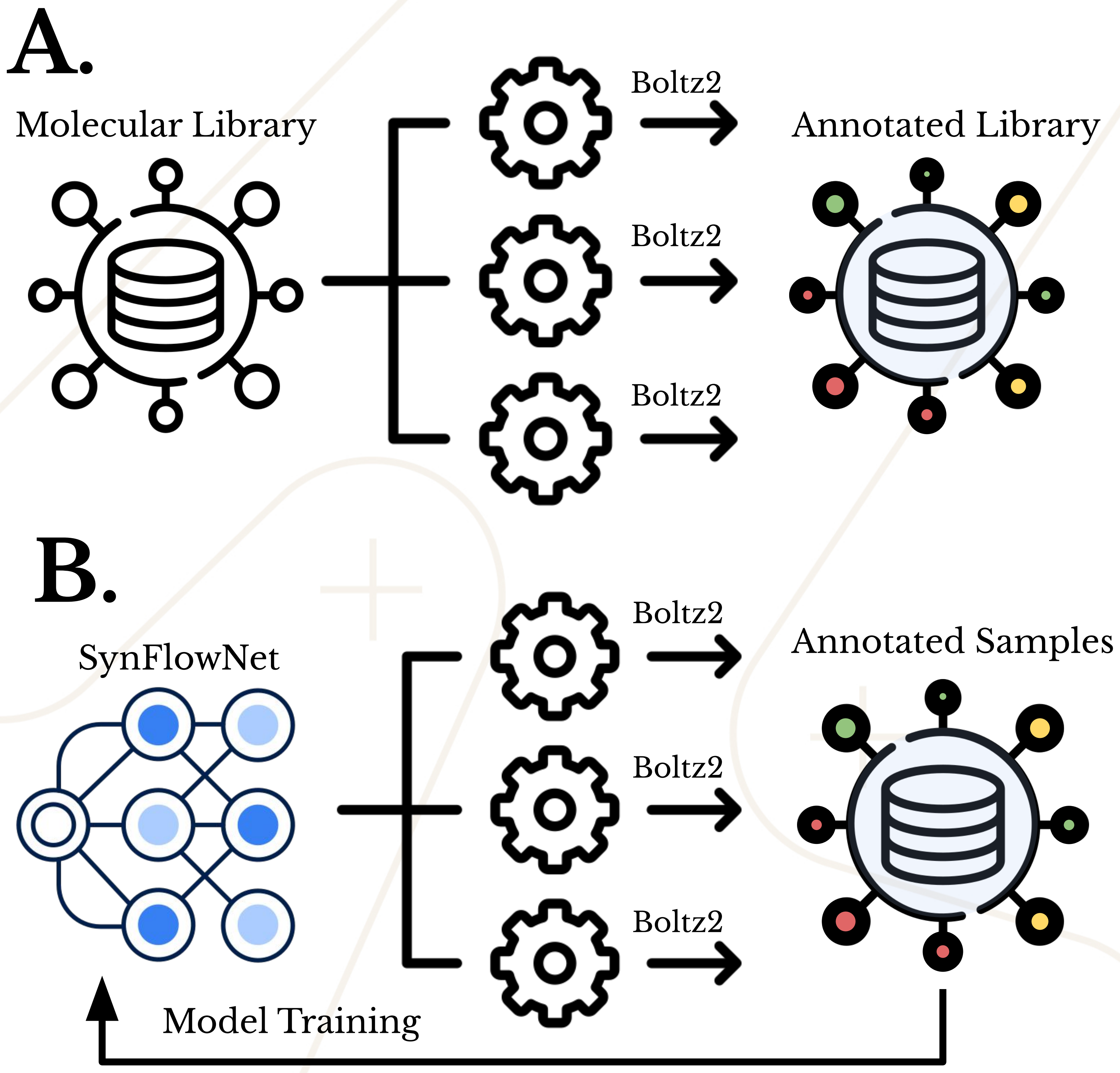


Boltz-2 achieves **2x enrichment** compared to **ML baselines and docking**

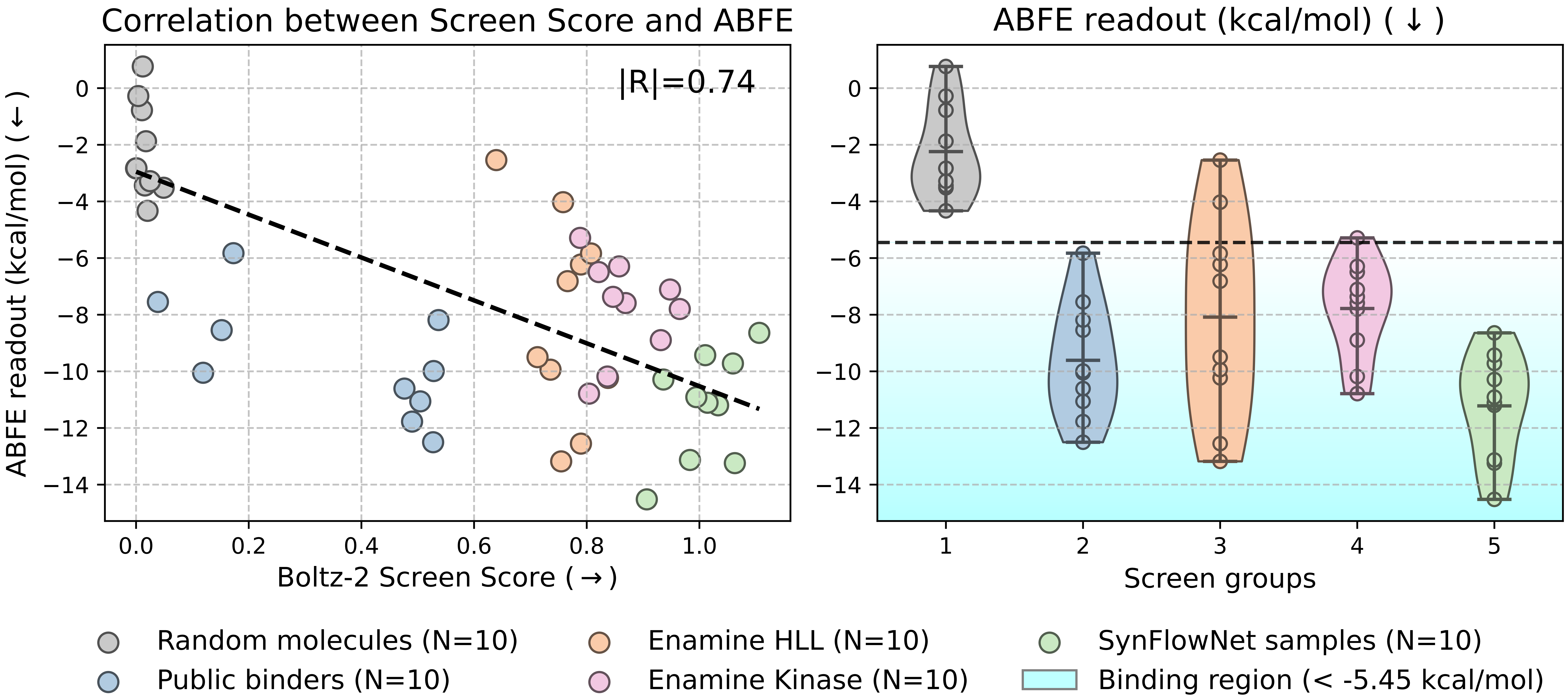
Prospective generative virtual screen

We use Boltz-2 to prospectively screen molecules on TYK2 using:

- Enamine Hit Locator Library (460k)
- Enamine Kinase Library (65k)
- SynFlowNet (117k)



Prospective generative virtual screen



Our **top-10 compounds** all bind and on average they **bind stronger than existing binders** according to ABFE validation.

Limitations

- Biomolecular structure prediction is not a solved problem: performance remains limited on many interaction modalities.
- Boltz-2 struggles to model large conformational changes.
- Good affinity prediction is dependent on good structure prediction.
- More work is needed to fully understand the applicability of affinity predictions.

Next: improve structure, affinity and more

On top of further improving Boltz-2 performance, we are working on further expanding Boltz's applicability! Stay tuned for:

- Protein-protein affinities
- Small-molecule and protein design pipelines
- ADME, toxicity and developability predictions
- Tools for fine-tuning and iterative design

Would you like to work together? Please reach out to us!

Thank You! We'd love to hear from you!

Manuscript
bit.ly/boltz2-pdf

Code and model
bit.ly/boltz-git

Slack channel
bit.ly/slack-boltz

saro00@csail.mit.edu
gcorso@csail.mit.edu
jwohlwend@csail.mit.edu

Jameel Clinic

Recursion