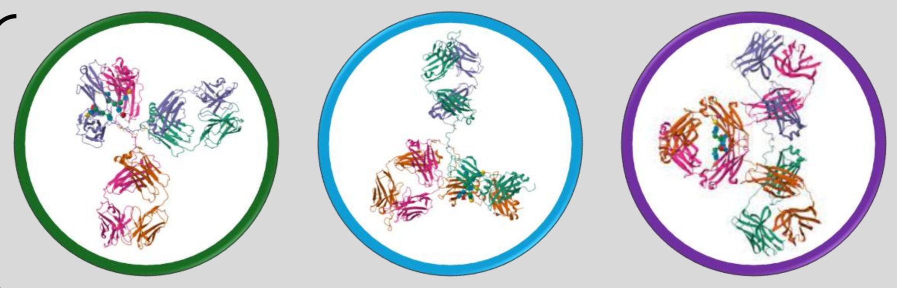
Antibody Design with Constrained Bayesian Optimization

Yimeng Zeng[1], Hunter Elliott[2], Phillip Maffettone[2], Peyton Greenside[2], Osbert Bastani[1], Jacob R. Gardner[1]

[1]University of Pennsylvania [2]BigHat Biosciences

Computational Antibody Design

As an optimization problem:

$\mathcal{A}:\{$  $\cdots\}$

$$\underset{x\in\mathcal{A}}{\text{maximize}}\ f(x)$$ ← Maximize binding affinity

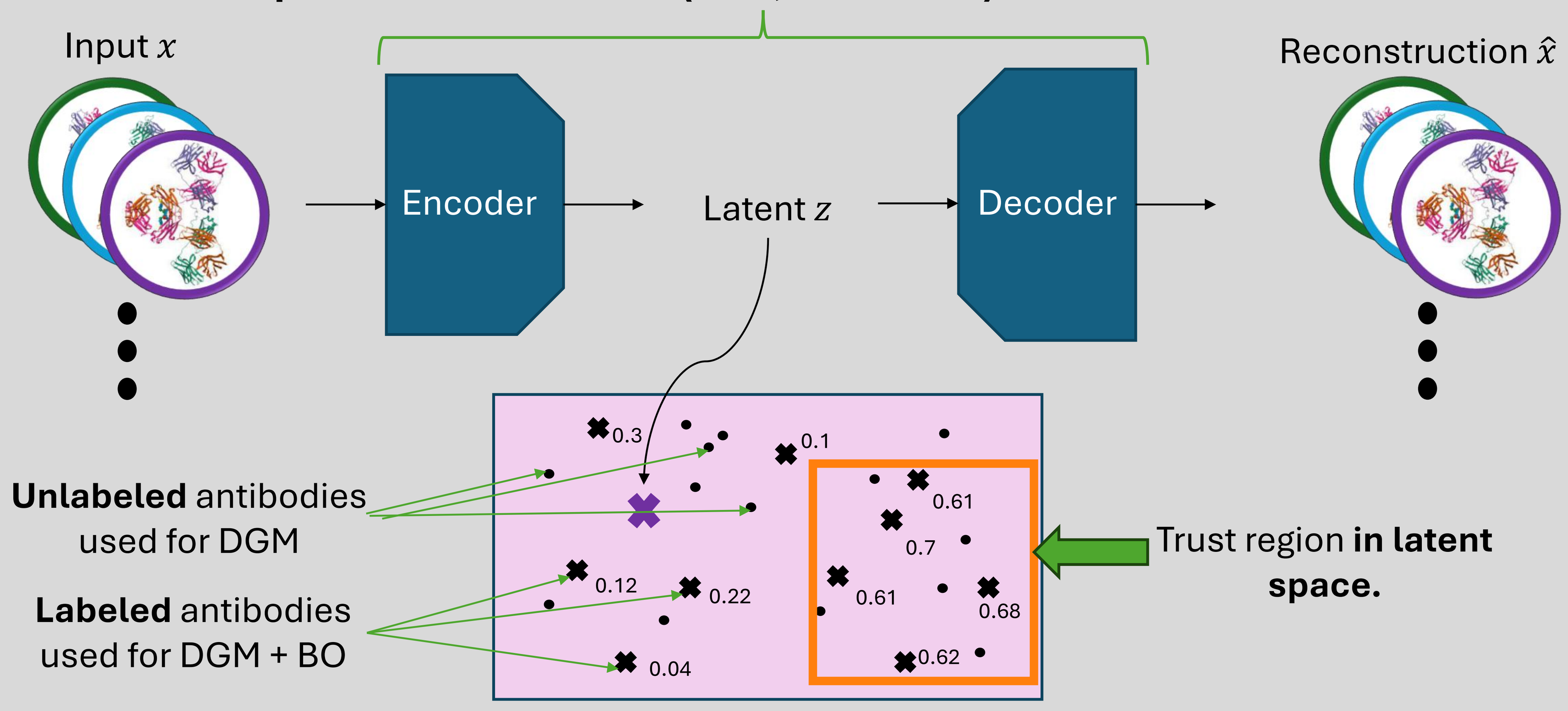
$$s.t.\ \vec{\boldsymbol{c}}(x) \geq \vec{\boldsymbol{\tau}}$$ ← Under design constraints

Constraints might include thermostability, diversity of multiple solutions, ease of synthesis...

Obvious challenge: optimizing over the space of all antibodies (or even just CDRs) is hard.

(Local) Deep Bayesian Optimization

Deep Generative Model (VAE, Diffusion) over Antibodies



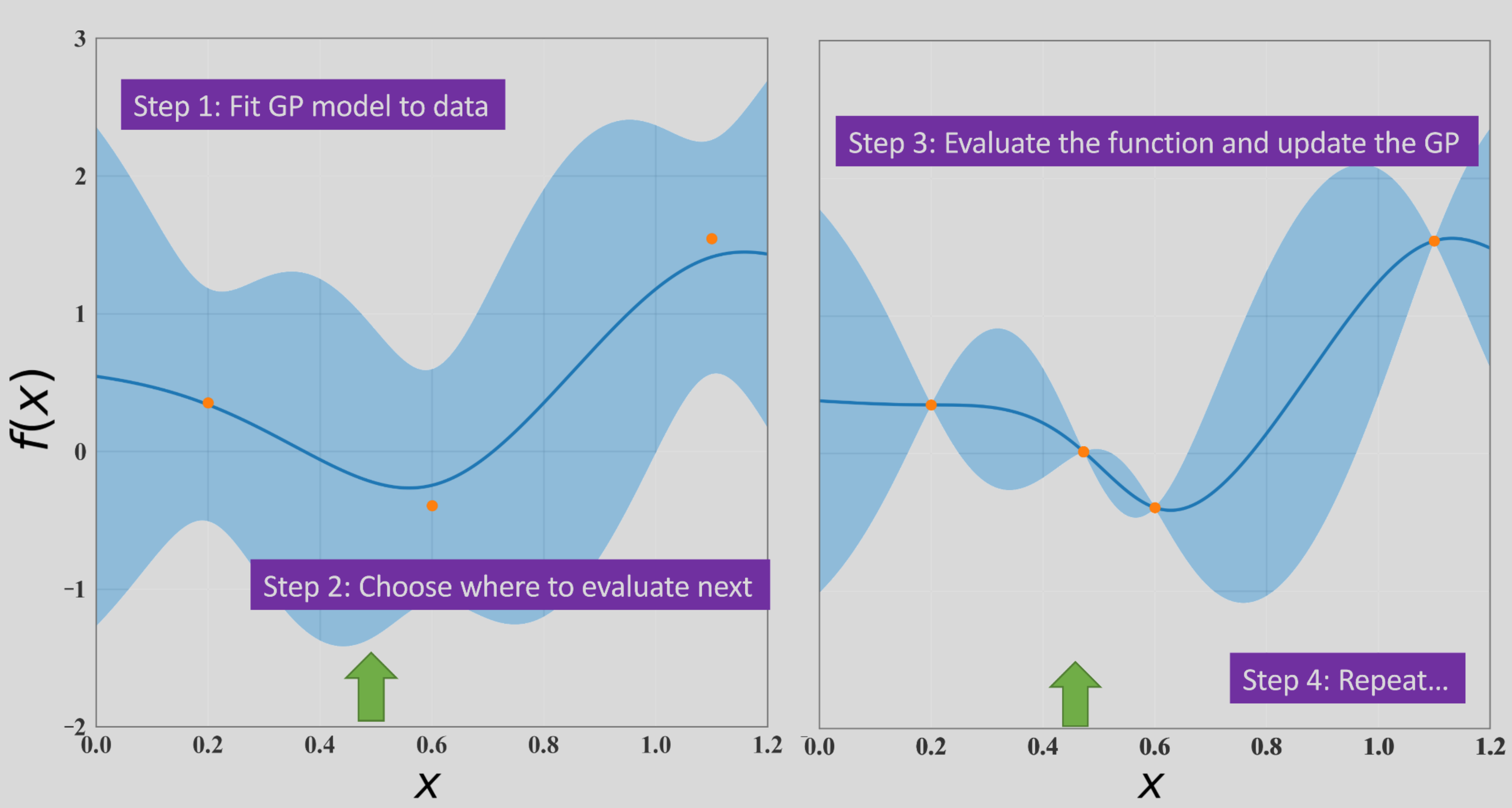
Key Idea: Apply Bayesian optimization over **continuous latent vectors** z instead of discrete x.

- **Objective function:** $\hat{f}(z) = f(\text{Decode}(z))$, where Decode(z) produces an antibody.
- Define constraints $\hat{c}(\cdot)$ similarly over z instead of over x.
- **Challenge 1:** Latent spaces are very high dimensional ⇒ **use trust regions.**
- **Challenge 2:** $\hat{f}(\cdot)$, $\hat{c}(\cdot)$ probably not smooth in z ⇒ **Train VAE and BO surrogate *jointly*.**

$$\mathcal{L}_{\text{joint}} = \underbrace{\mathbb{E}_{\text{Enc}(z|x)}\left[\mathcal{L}_{\text{svgp}}(\theta_{\text{GP}}, \theta_{\text{enc}}; \boldsymbol{y}, \boldsymbol{Z})\right]}_{\text{Expected supervised loss}} + \underbrace{\mathcal{L}_{\text{VAE}}(\theta_{\text{enc}}, \theta_{\text{dec}}; \boldsymbol{X})}_{\text{Typical VAE loss}}$$

Expected supervised loss (for data that has labels) — **Typical VAE loss** (for data that has no labels)

Bayesian Optimization

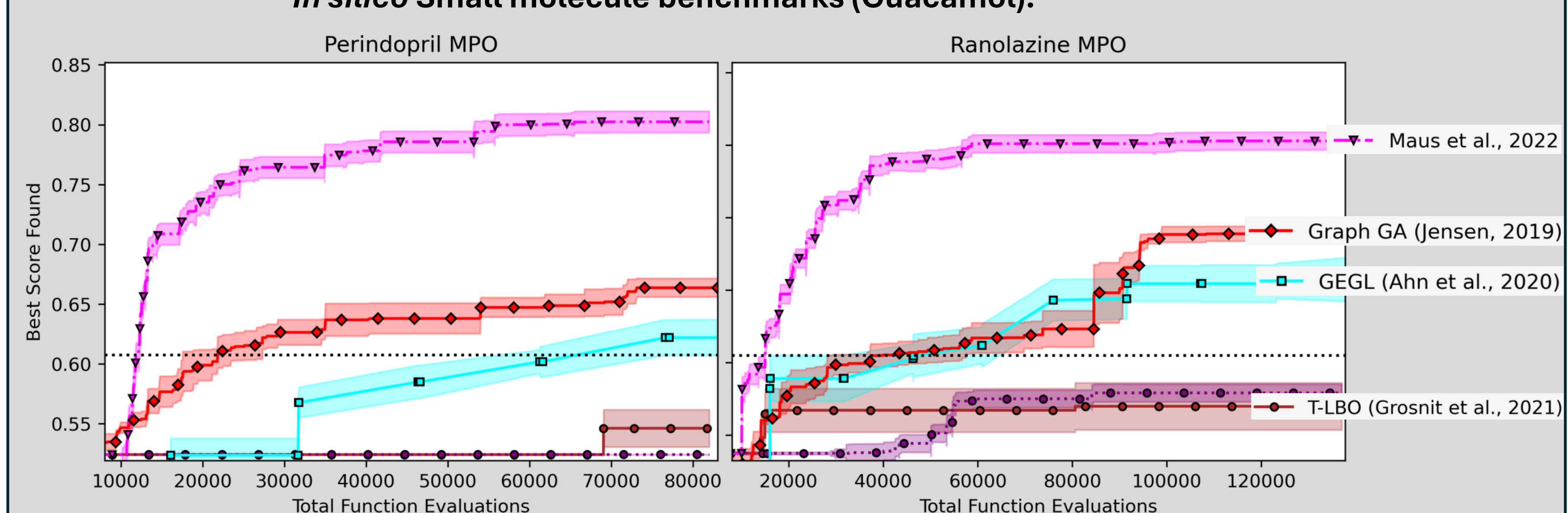


The literature supports: constraints, multi objective, diverse solutions, and more!

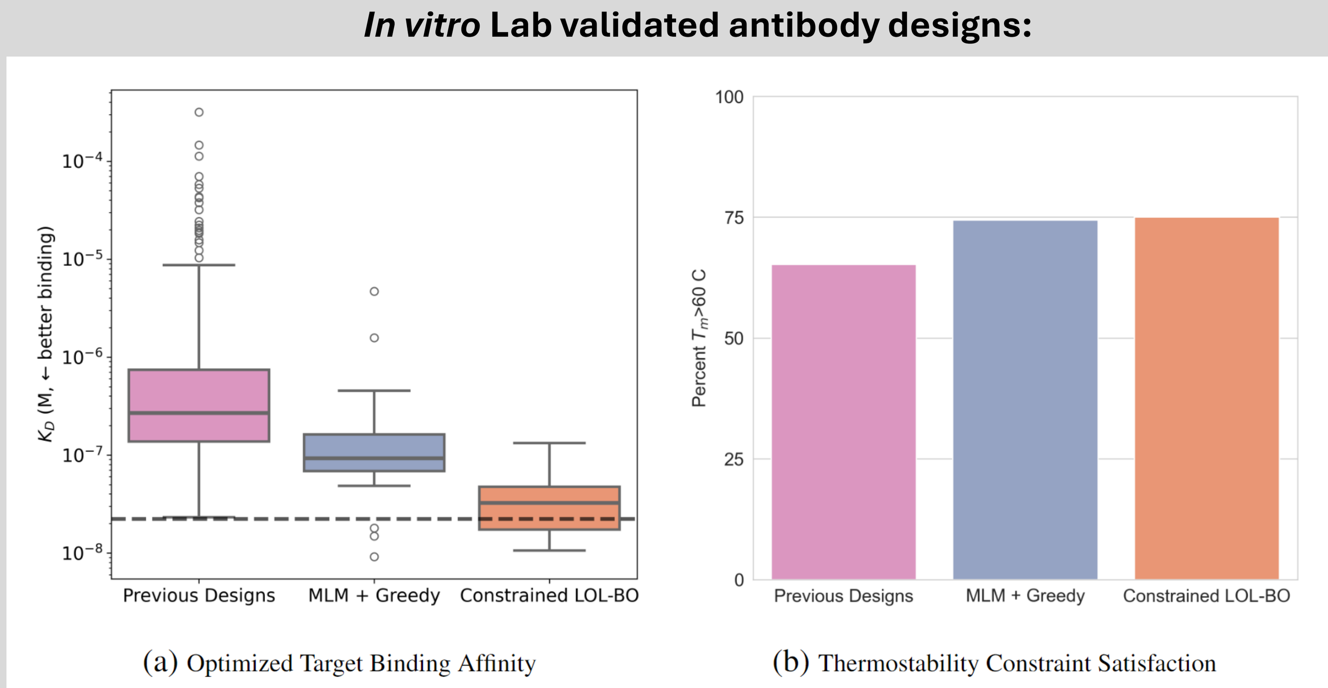
Historically: not very useful for high dimensional optimization, large data regime, discrete/structured optimization, ...

Results

In silico **Small molecule benchmarks (Guacamol):**

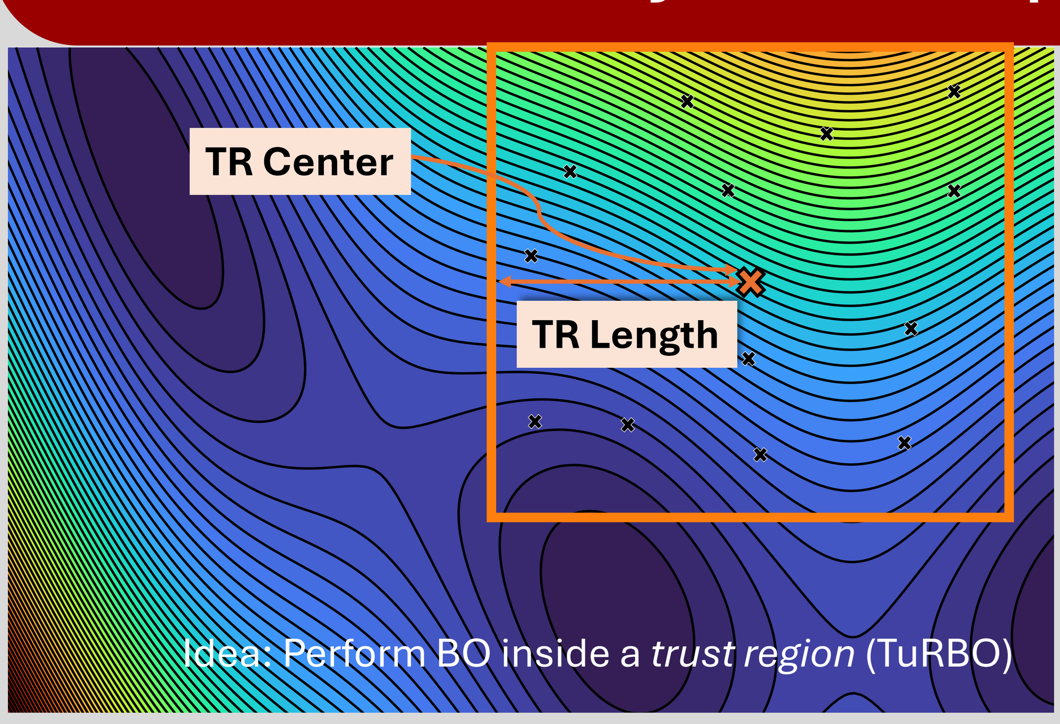


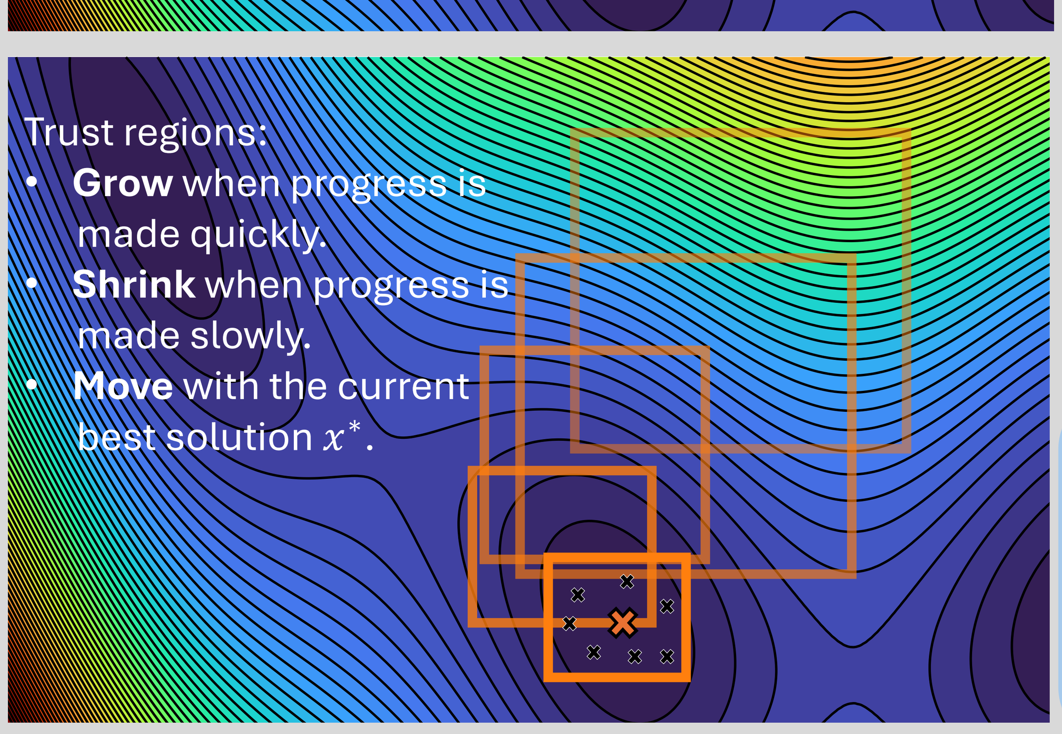
In vitro **Lab validated antibody designs:**



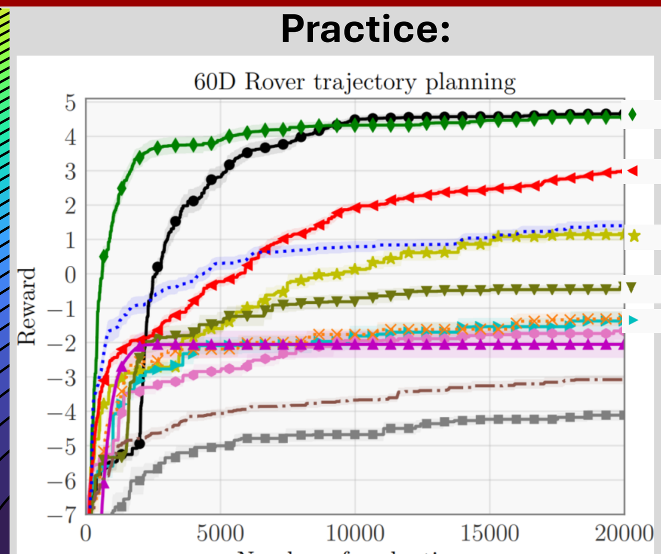
- Better binding affinity than masked language model and prior designs.
- Thermostability constraints satisfied in 75% of designs.

Local Bayesian Optimization





Practice:



Theory:
Under assumptions commonly used to analyze BO, convergence to a stationary point on noisy functions has rate $O\left(\frac{d^{1.25}}{T^{0.25}}\right)$.

Much better than exponential in *d*

State of the art for high dimensional black-box optimization, and not just for BO methods.

References

For local Bayesian optimization:

- David Eriksson, Michael Pearce, Jacob R. Gardner, Ryan Turner, Matthias Poloczek. **Scalable Global Optimization via Local Bayesian Optimization.** (NeurIPS 2019). ← TuRBO
- Kaiwen Wu, Kyurae Kim, Roman Garnett, Jacob R. Gardner. **The Behavior and Convergence of Local Bayesian Optimization.** (NeurIPS 2023). ← Theory
- David Eriksson, Matthias Poloczek. **Scalable Constrained Bayesian Optimization** (AISTATS 2021). ← Constraints

For local latent space Bayesian optimization:

- Natalie Maus, Haydn T. Jones, Juston Moore, Matthew J. Kusner, John Bradshaw, Jacob R. Gardner. **Local Latent Space Bayesian Optimization over Structured Inputs.** (NeurIPS 2022).
- Natalie Maus, Kaiwen Wu, David Eriksson, Jacob R. Gardner. **Discovering Many Diverse Solutions for Bayesian Optimization.** (AISTATS 2023). ← Diverse solutions

For fast approximate GP models:

- James Hensman, Nicolo Fusi, Neil D. Lawrence **Gaussian Processes for Big Data** (UAI 2013). ← Original
- Martin Jankowiak, Geoff Pleiss, Jacob R. Gardner. **Parametric Gaussian Process Regressors.** (ICML 2020). ← Used here