Bayesian optimization needs better deep learning

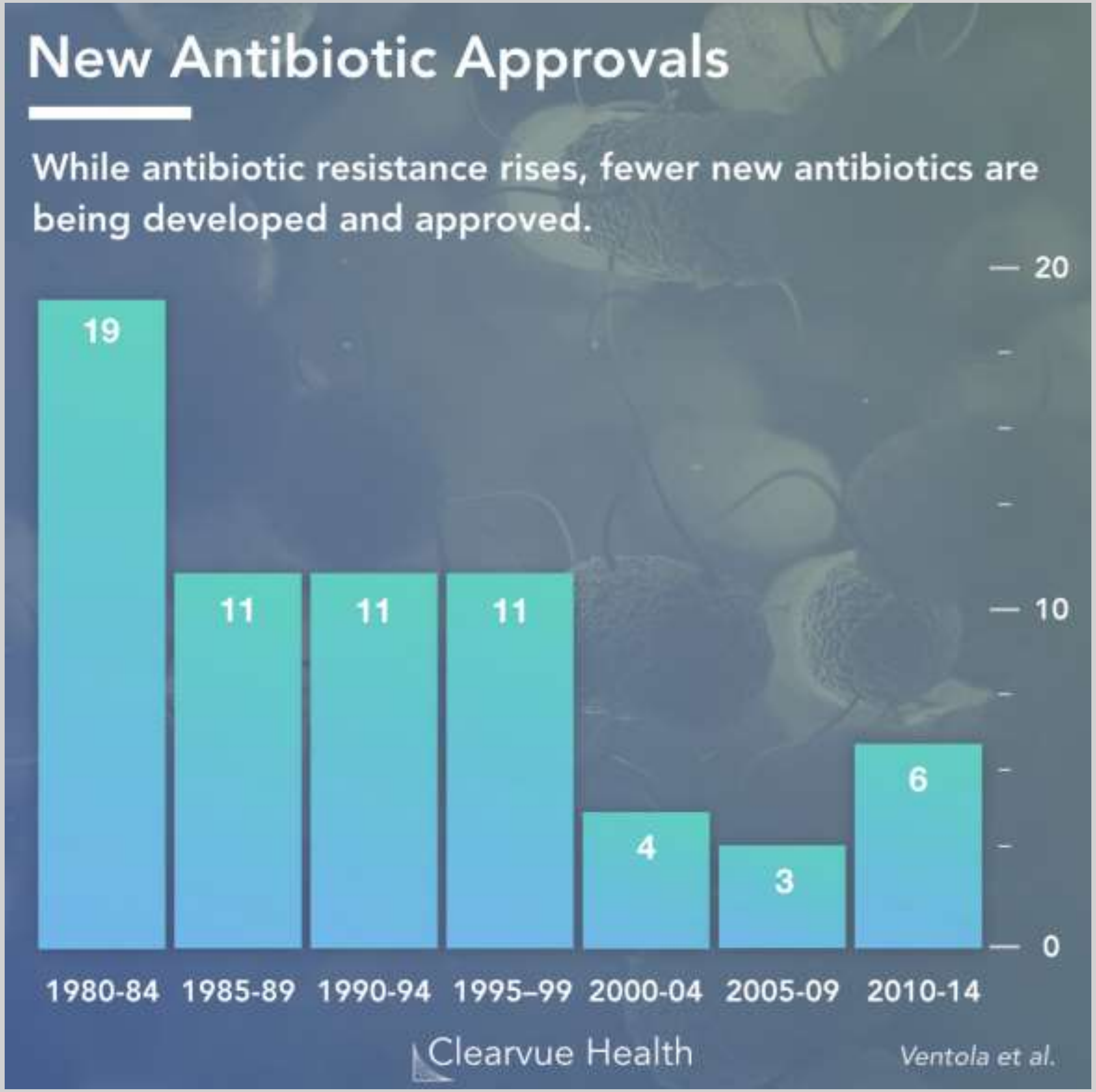
or

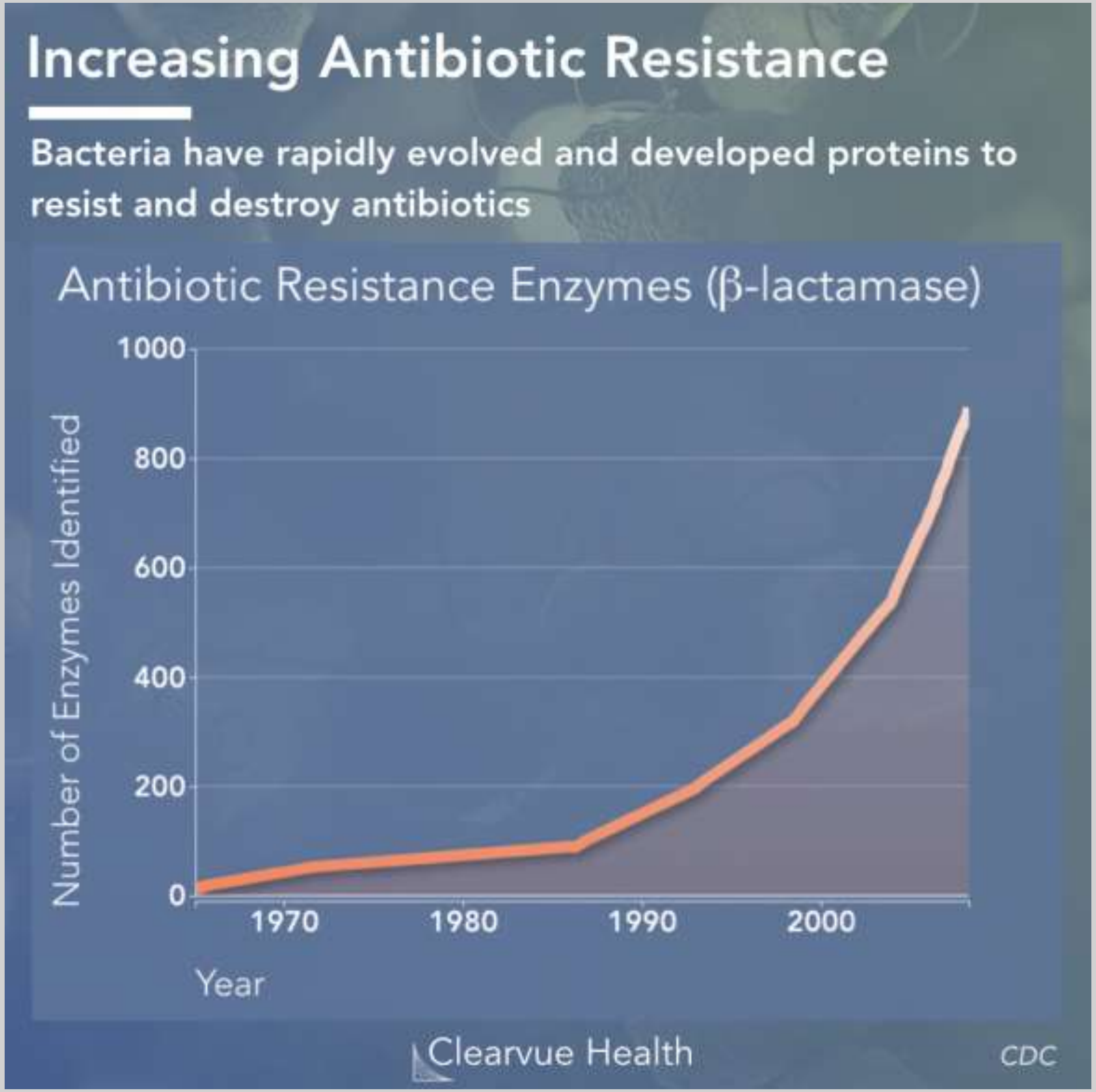
Some fun antibiotic design stuff and how it could be way better.

Jacob Gardner, University of Pennsylvani

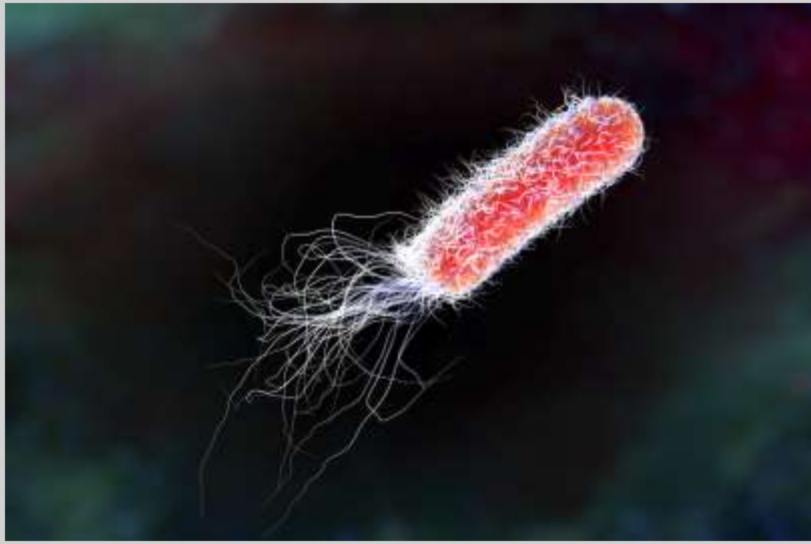
Penn
UNIVERSITY of PENNSYLVANIA

Motivating Example: Antimicrobial Pept



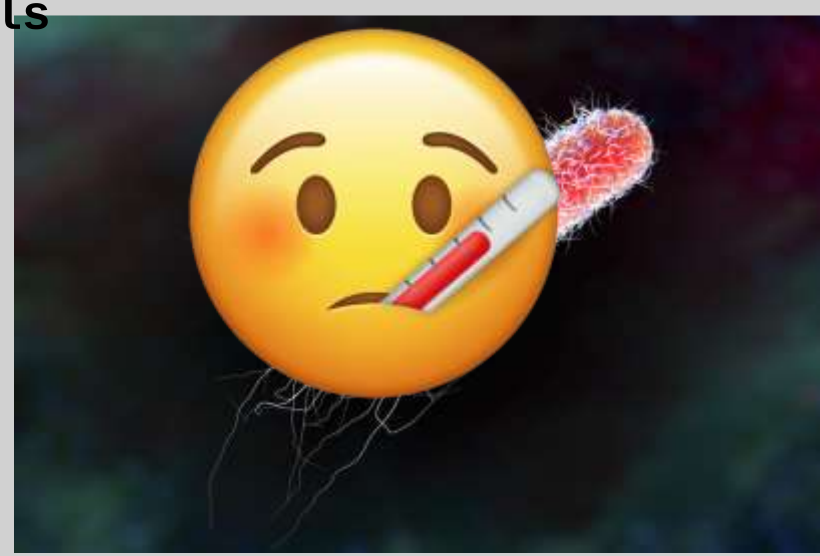


Motivating Example: Antimicrobial Pept



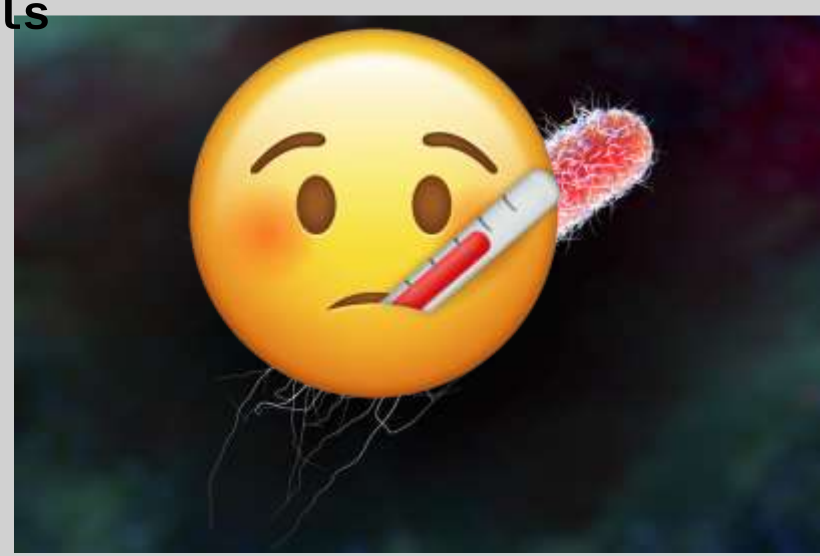
Motivating Example: Antimicrobial Pept

Mine AMPs from extinct mammals

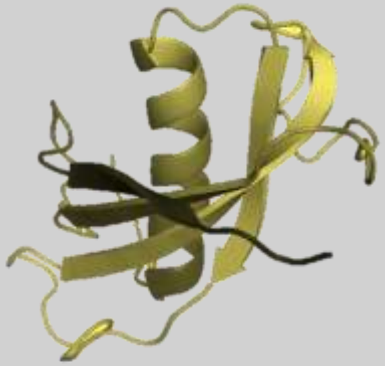


Motivating Example: Antimicrobial Pept

Mine AMPs from extinct mammals



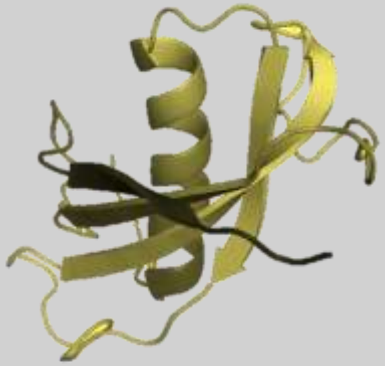
Bayesian optimization



Motivating Example: Antimicrobial Pept

Mine AMPs from extinct mammals

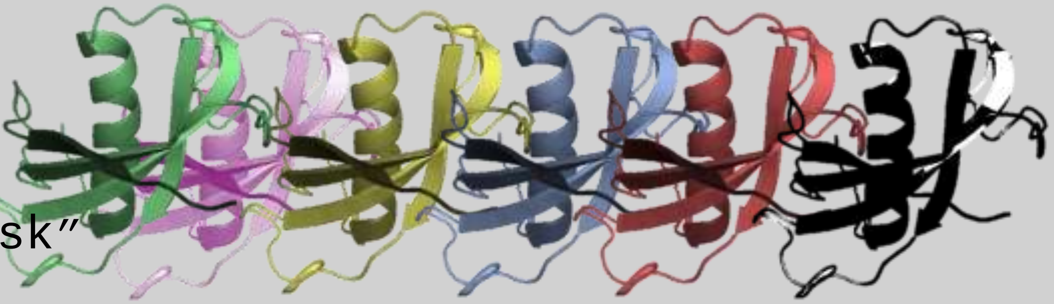
Bayesian optimization



Motivating Example: Antimicrobial Pept

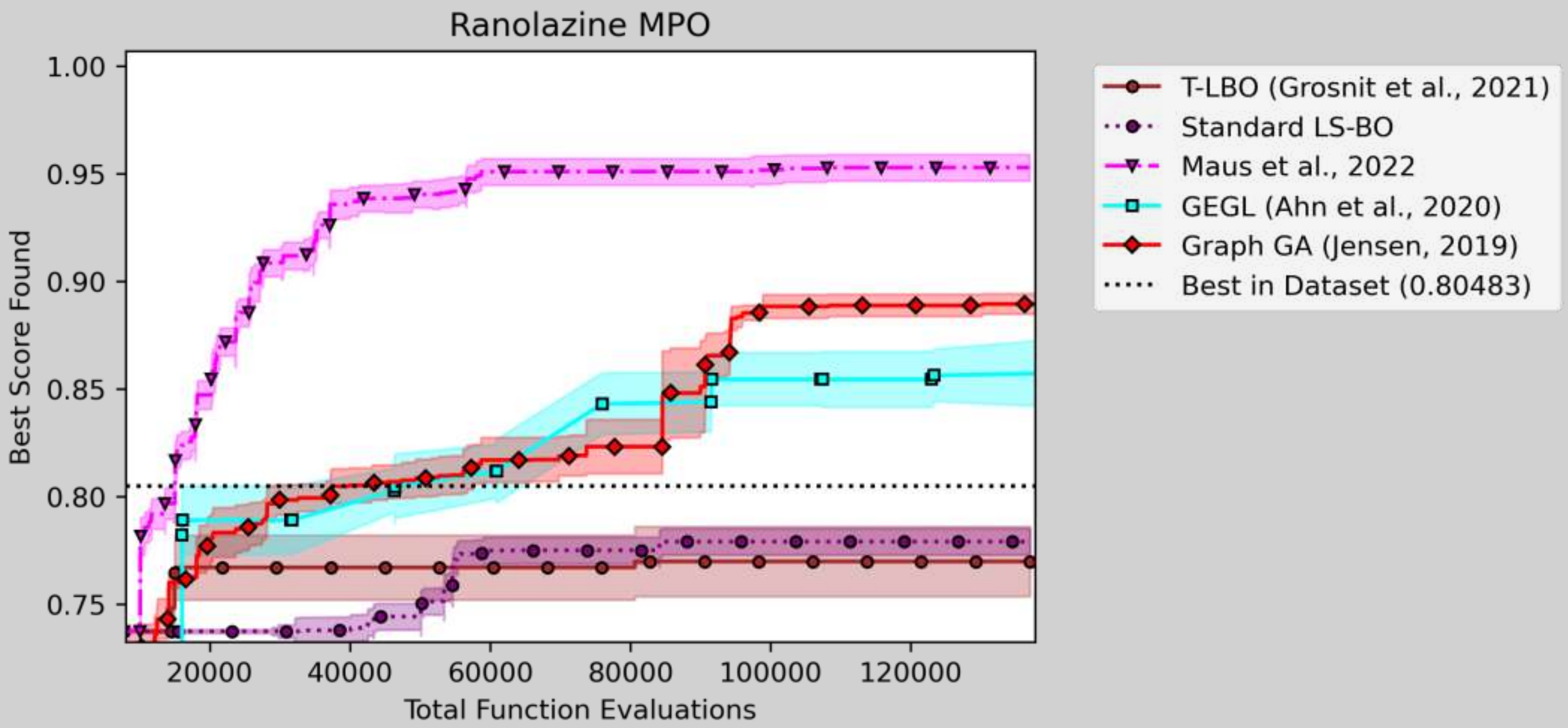
Mine AMPs from extinct mammals

Bayesian optimization

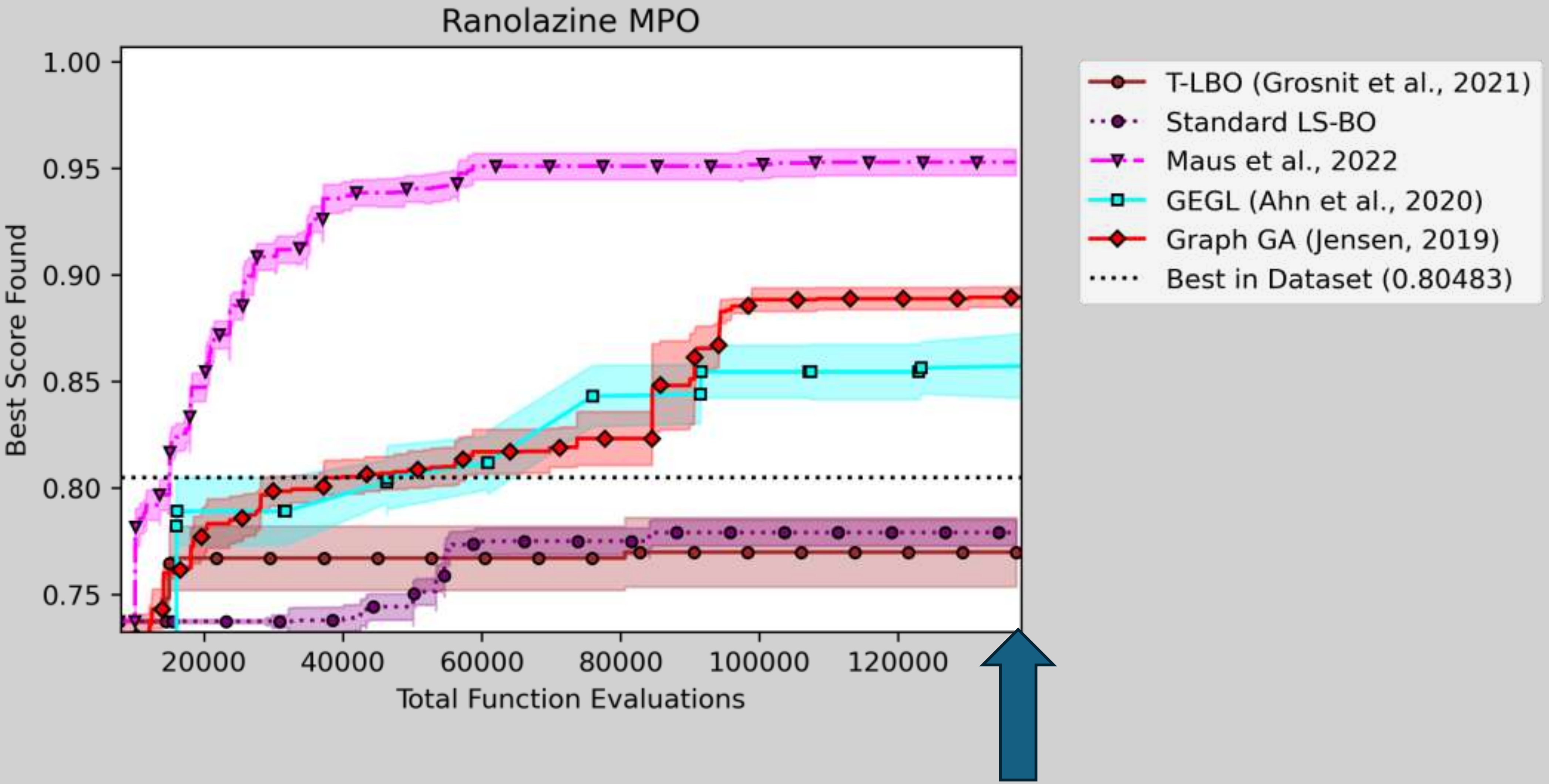


“Portfolio risk”

The Anatomy of a Result Plot in this S



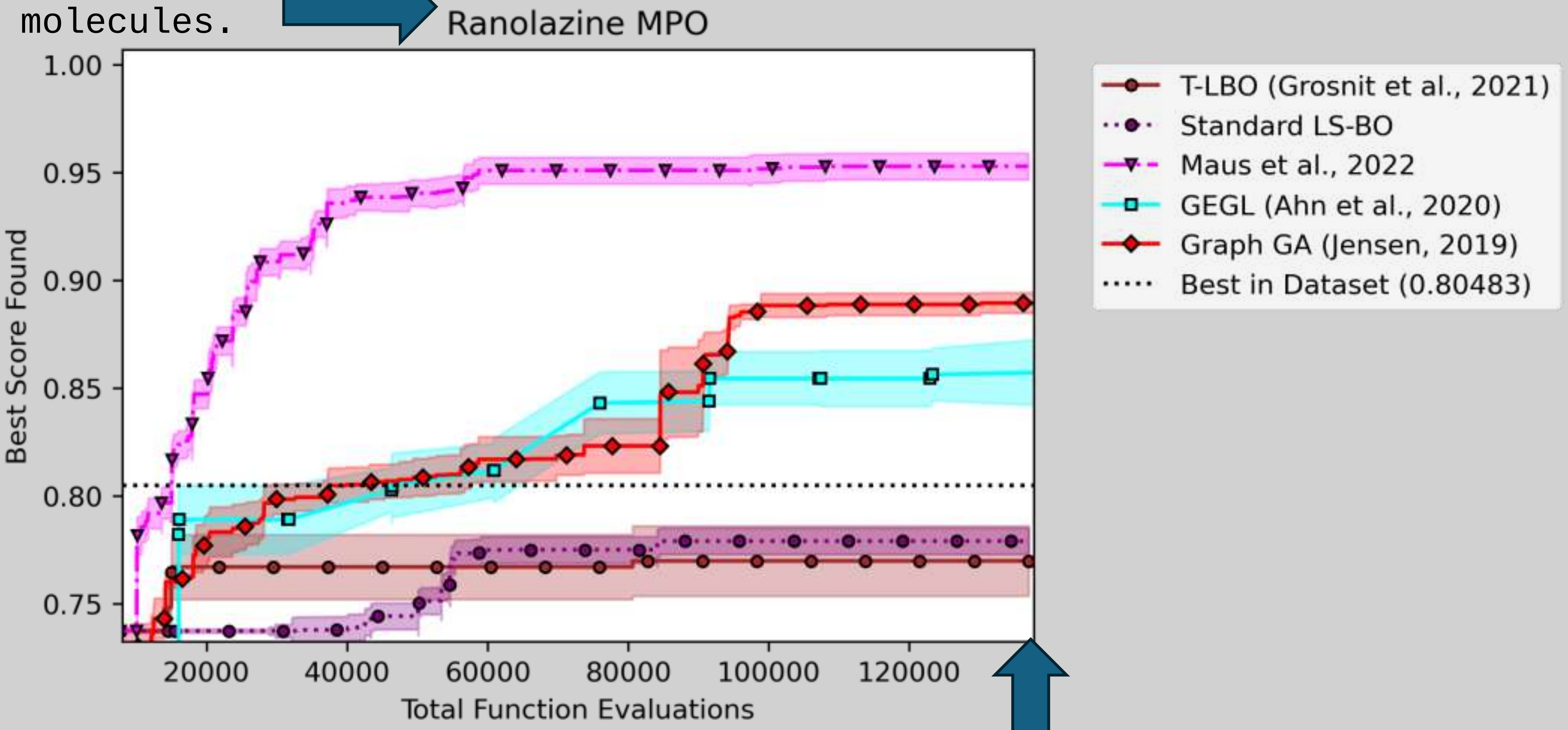
The Anatomy of a Result Plot in this S



Tons of evaluations,
large batch size.

The Anatomy of a Result Plot in this S

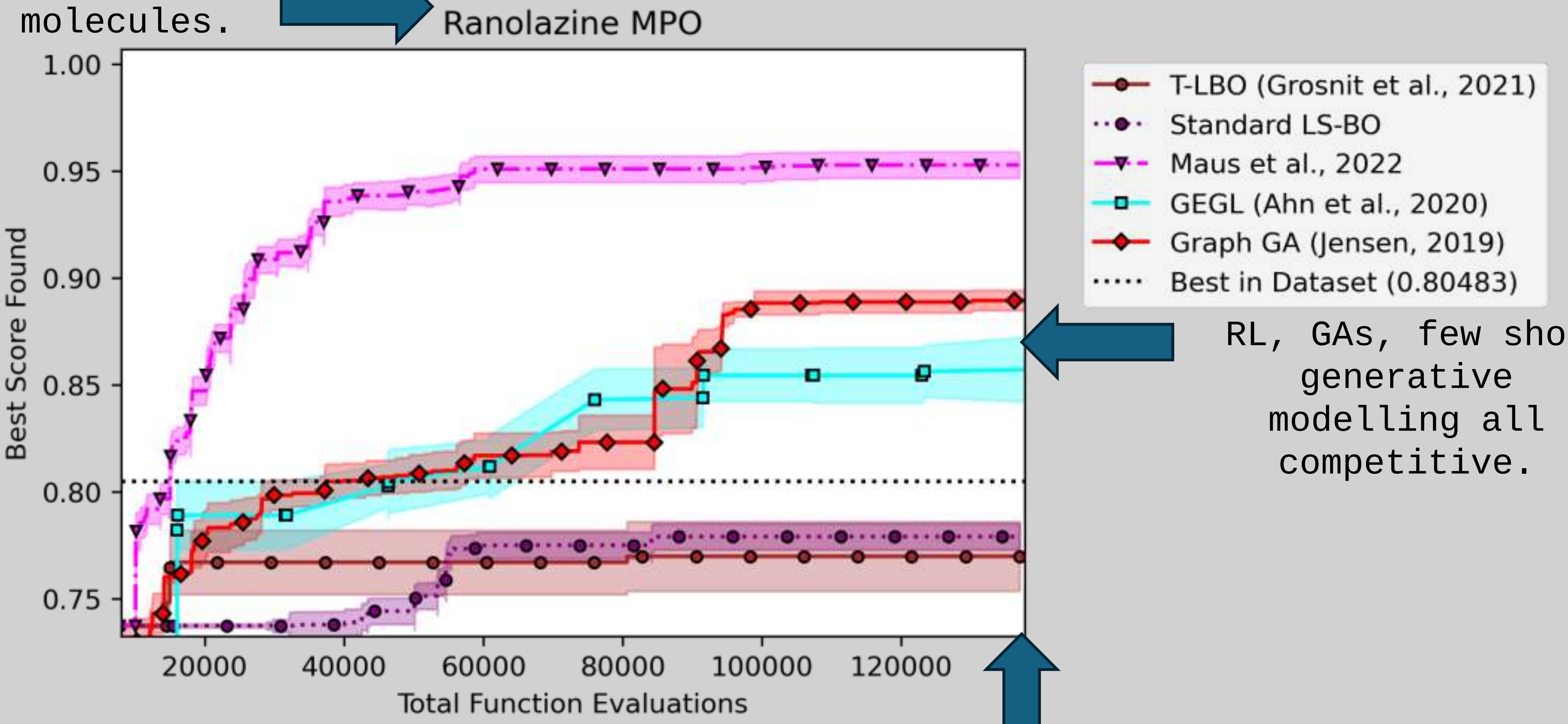
Objectives defined
over **discrete structures**
like molecules.



Tons of evaluations,
large batch size.

The Anatomy of a Result Plot in this S

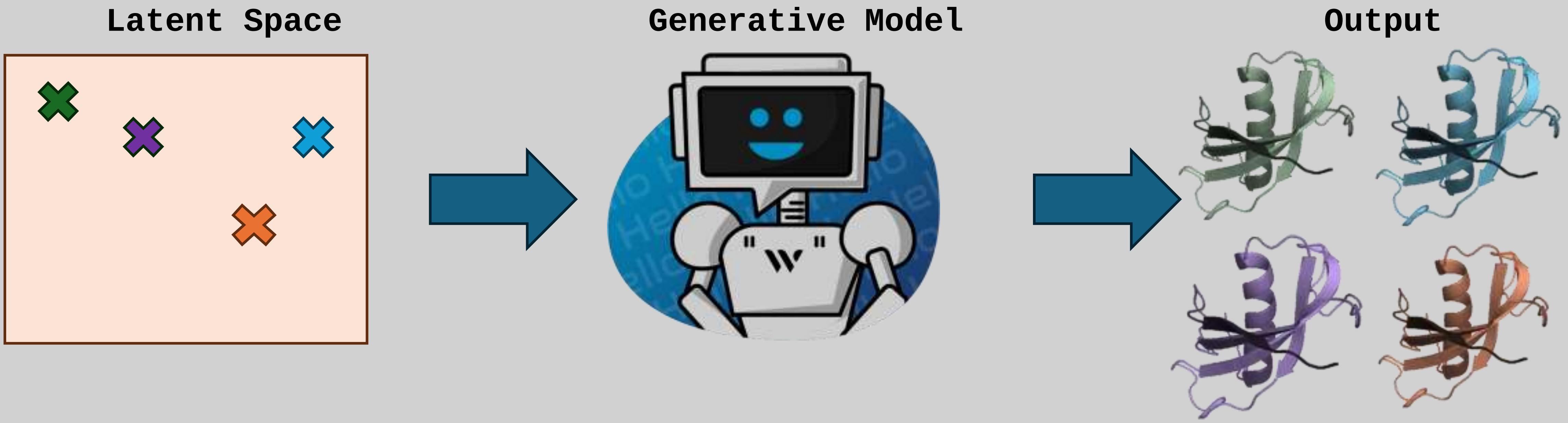
Objectives defined
over **discrete structures**
like molecules.



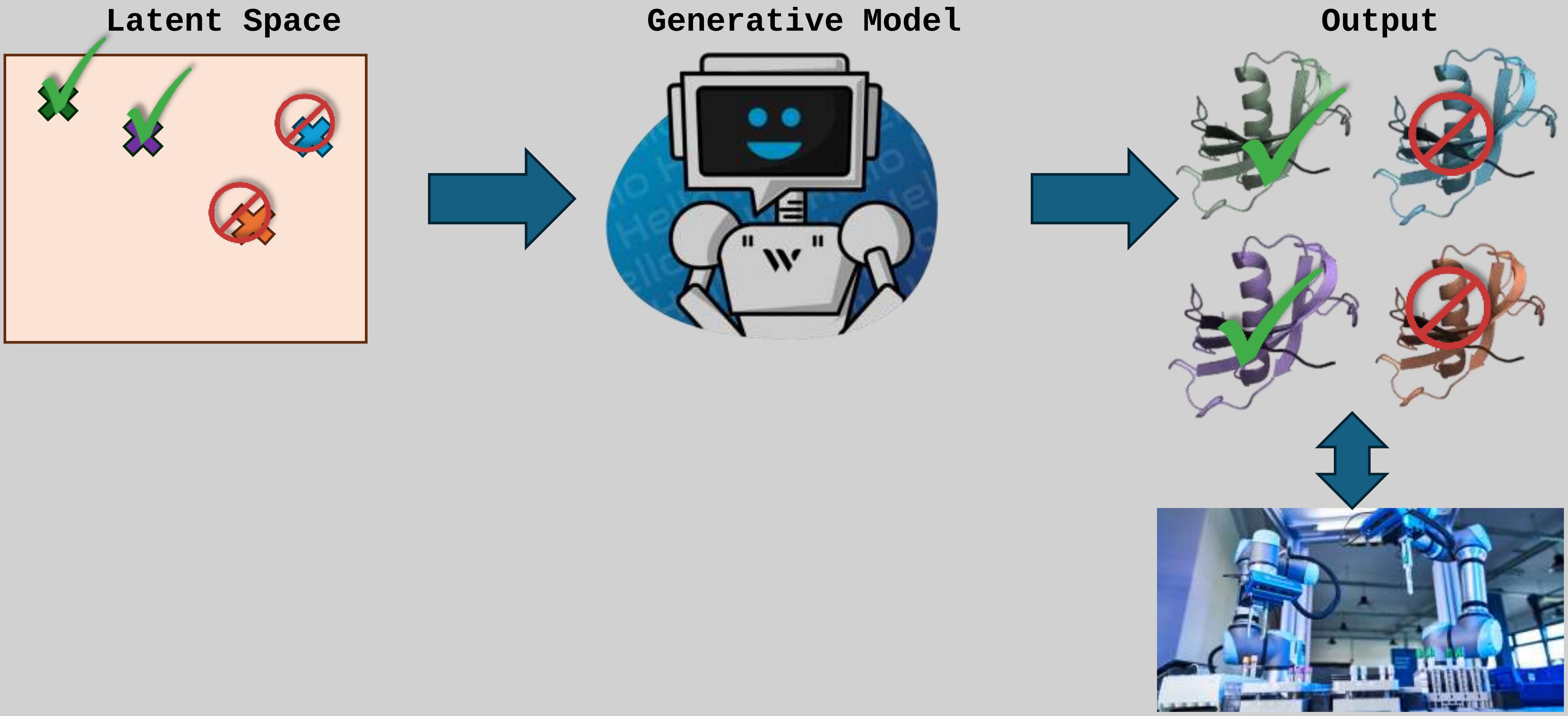
RL, GAs, few shot
generative
modelling all
competitive.

Tons of evaluations,
large batch size.

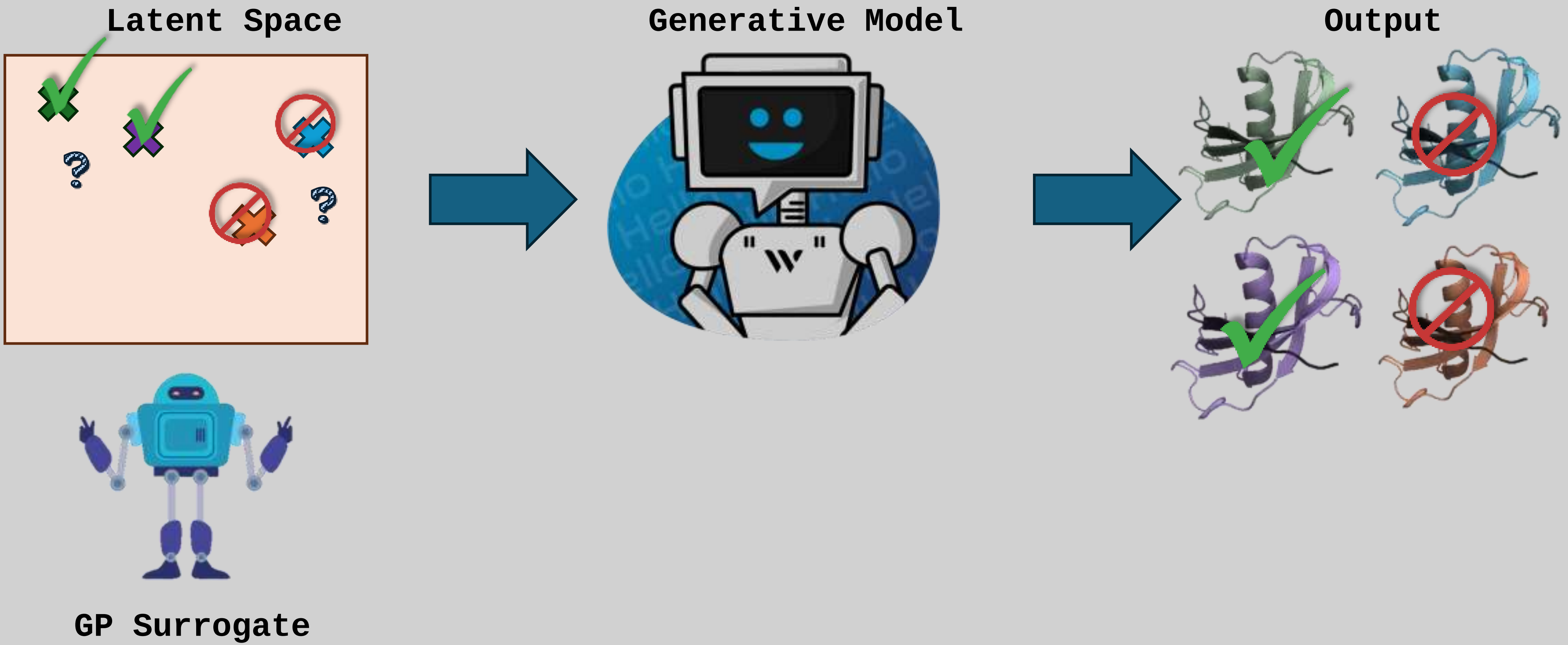
Challenge 1: Discrete, Structured



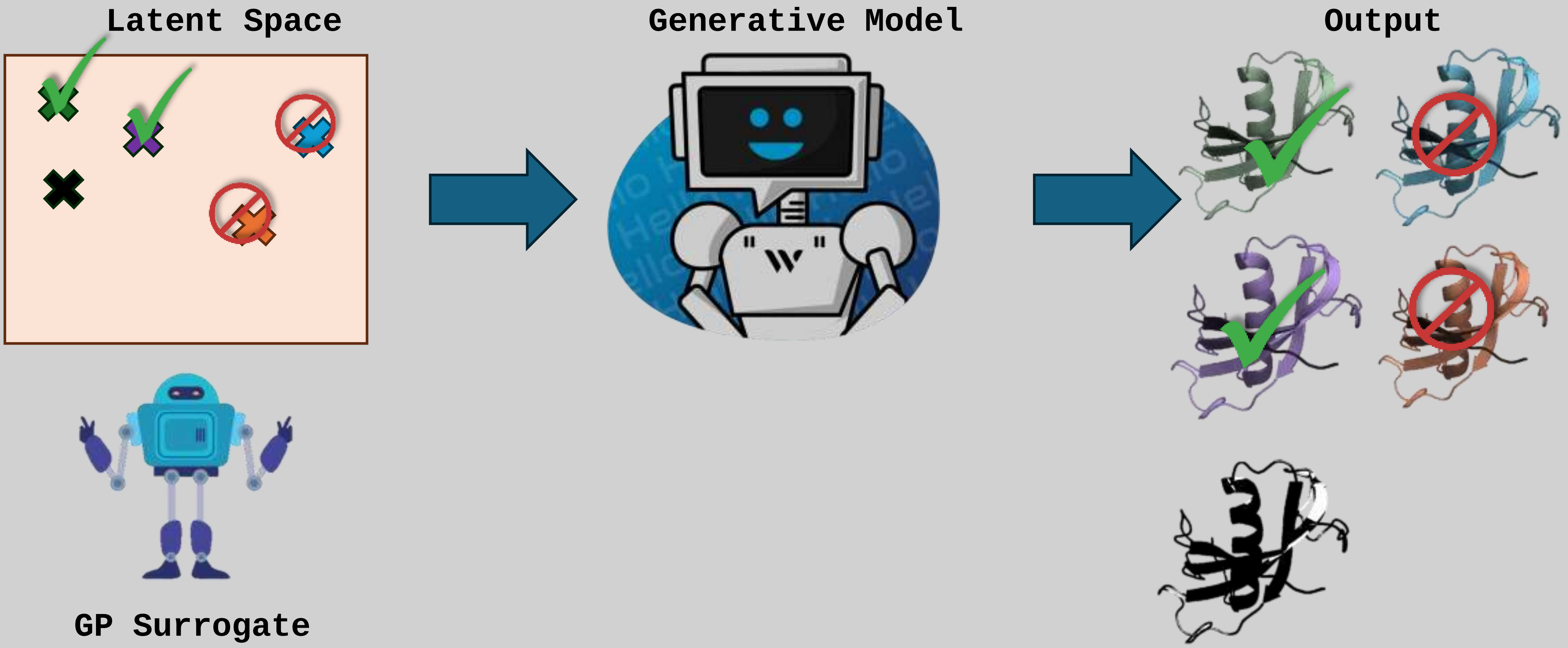
Challenge 1: Discrete, Structured



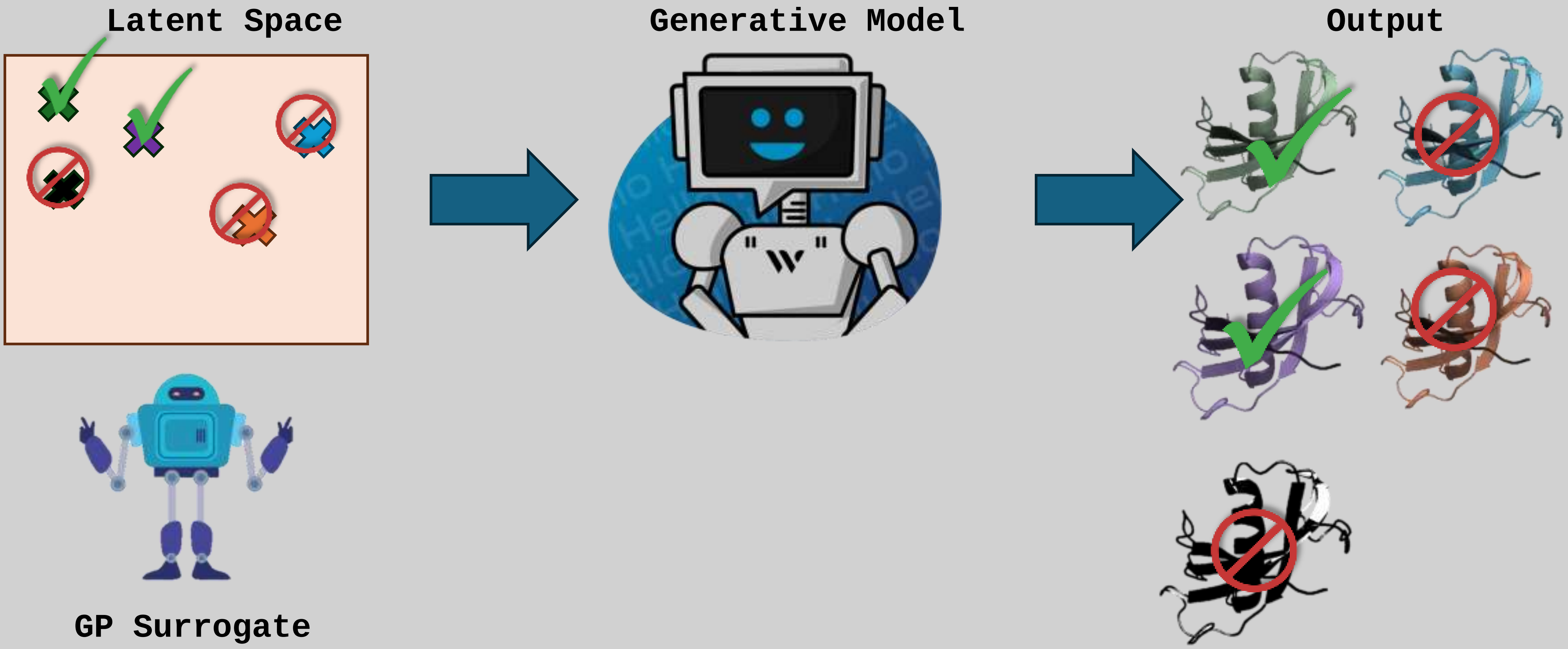
Challenge 1: Discrete, Structured



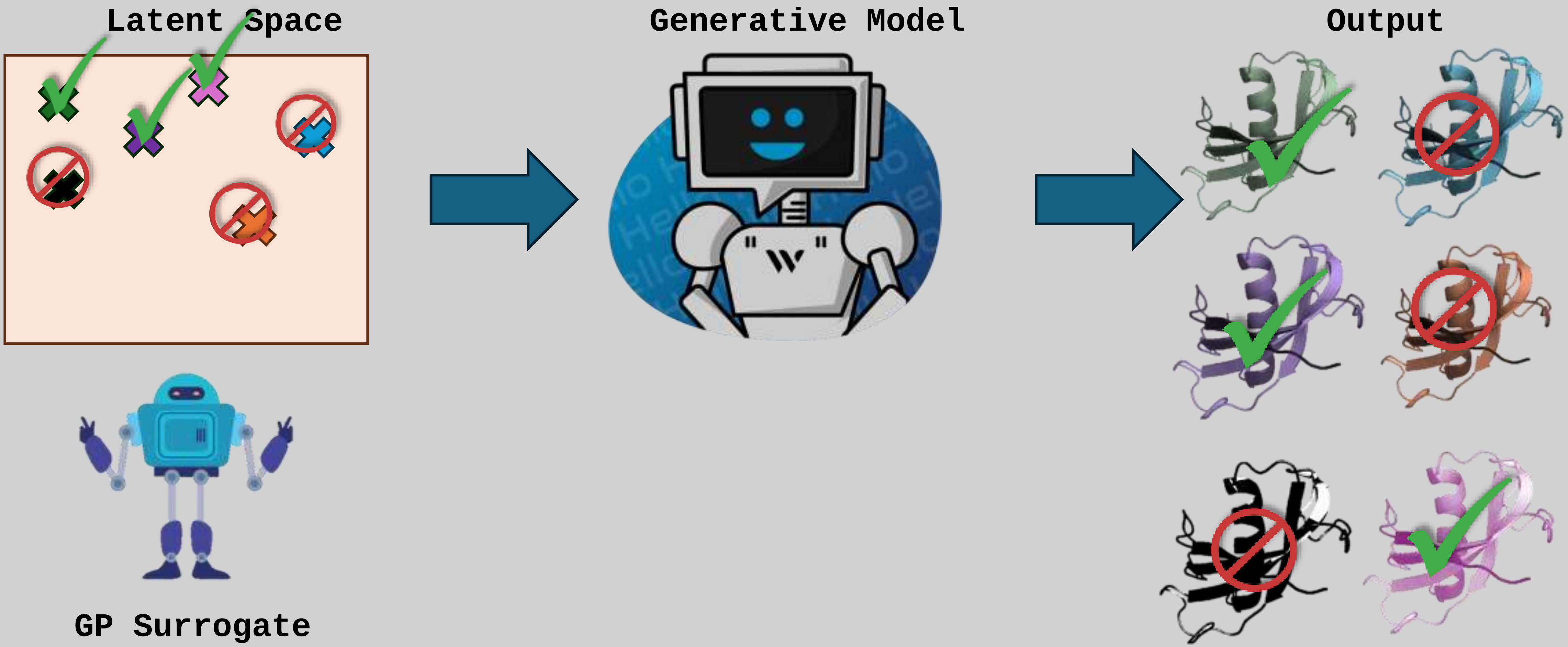
Challenge 1: Discrete, Structured



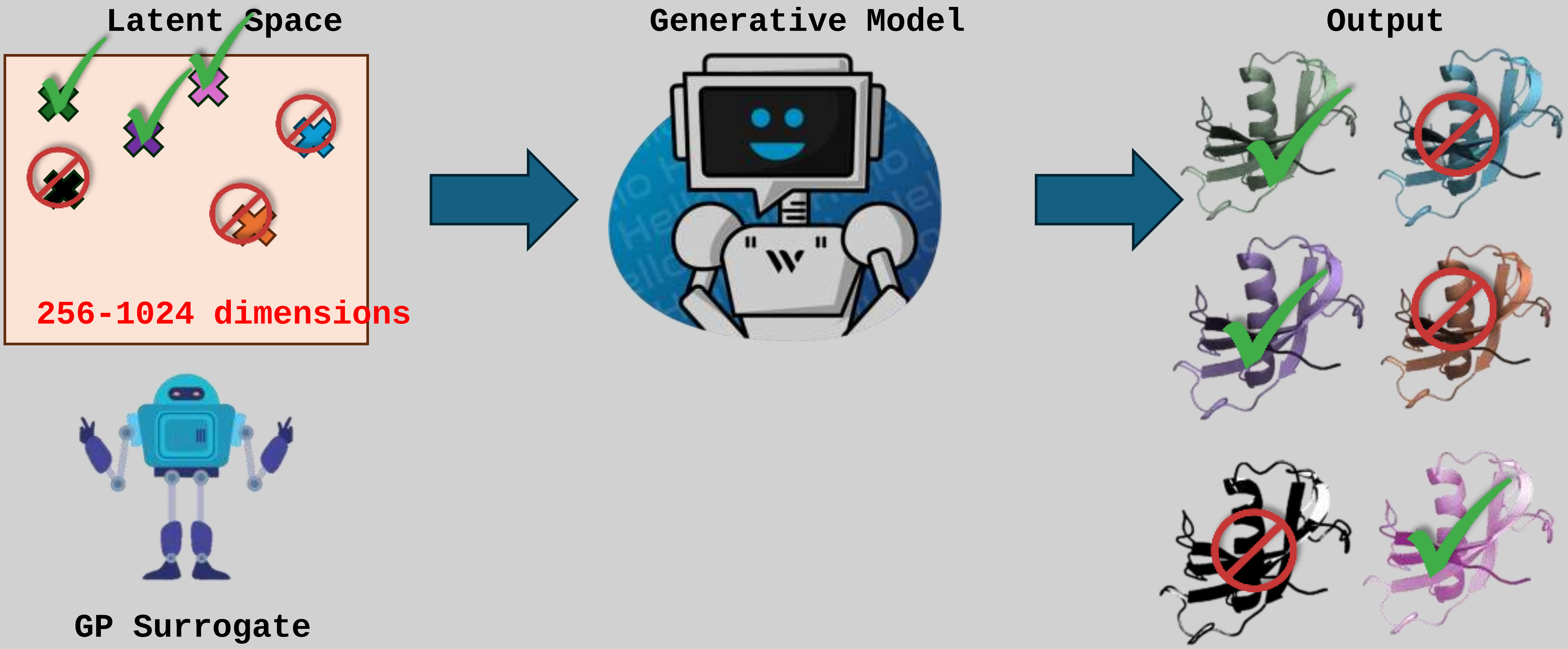
Challenge 1: Discrete, Structured



Challenge 1: Discrete, Structured



Challenge 1: Discrete, Structured

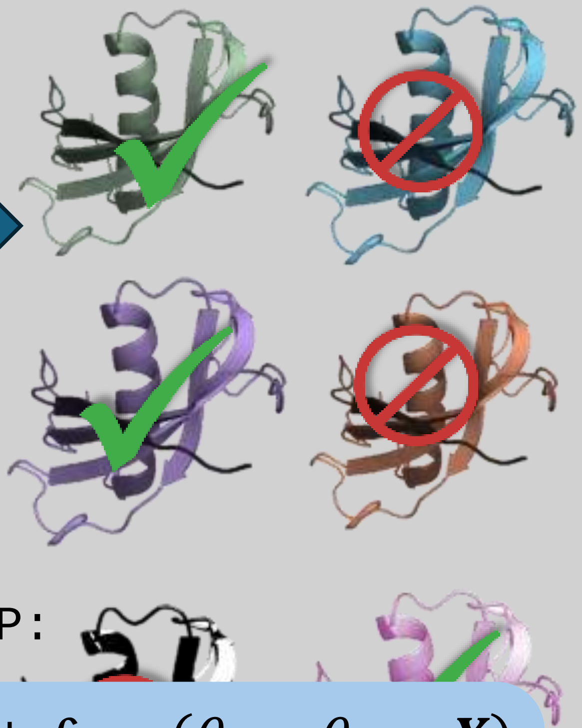


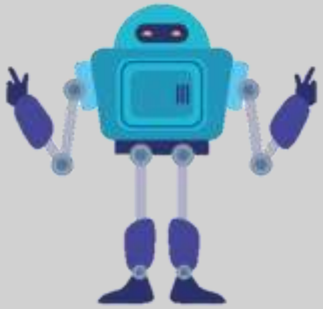
Challenge 1: Discrete, Structured

Latent Space

Generative Model

Output





GP Surrogate

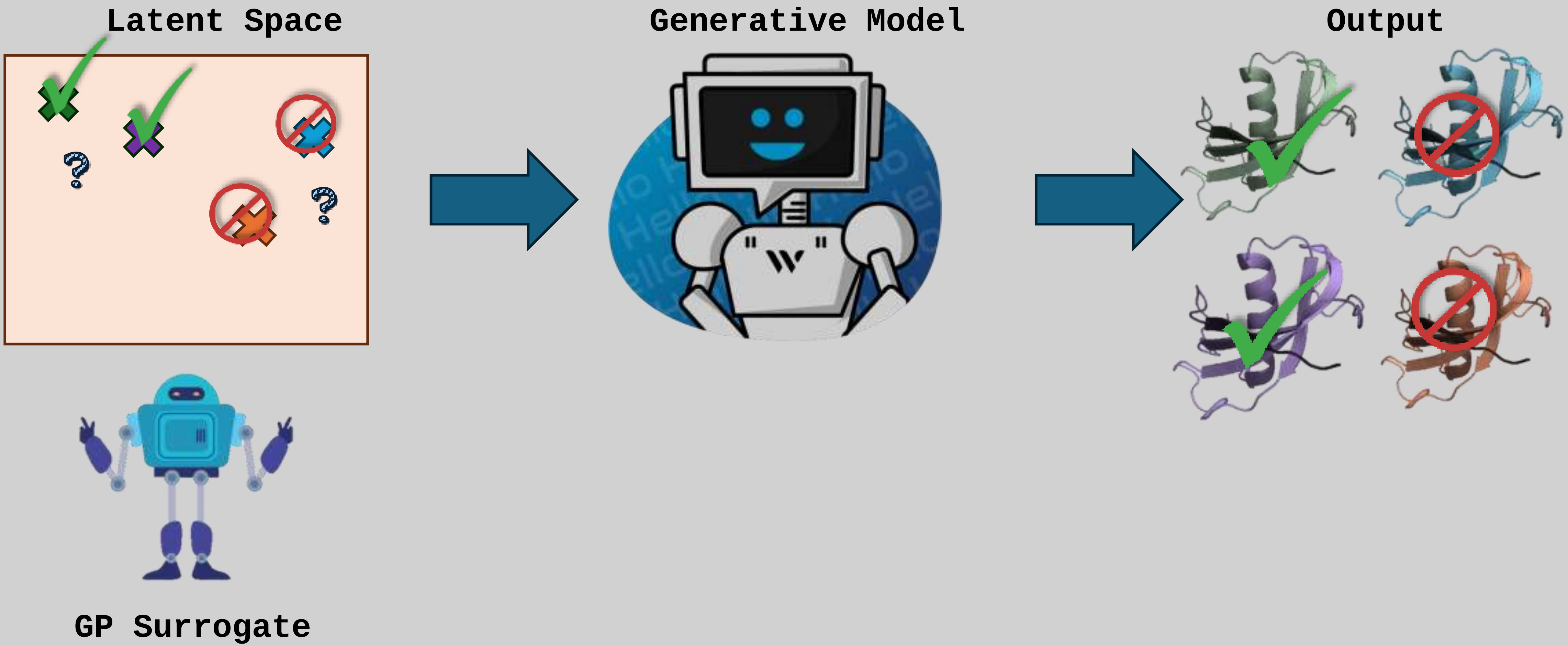
Our paper (NeurIPS 2022):

- Use high dim. BO (local BO)
- Joint ELBO over VAE and SVGP:

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

(for data that has labels) (for data that has no labels)

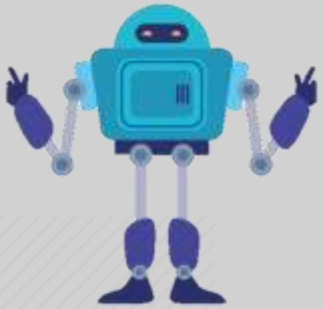
Challenge 2: High Throughput



Challenge 2: High Throughput

(NeurIPS 2024)

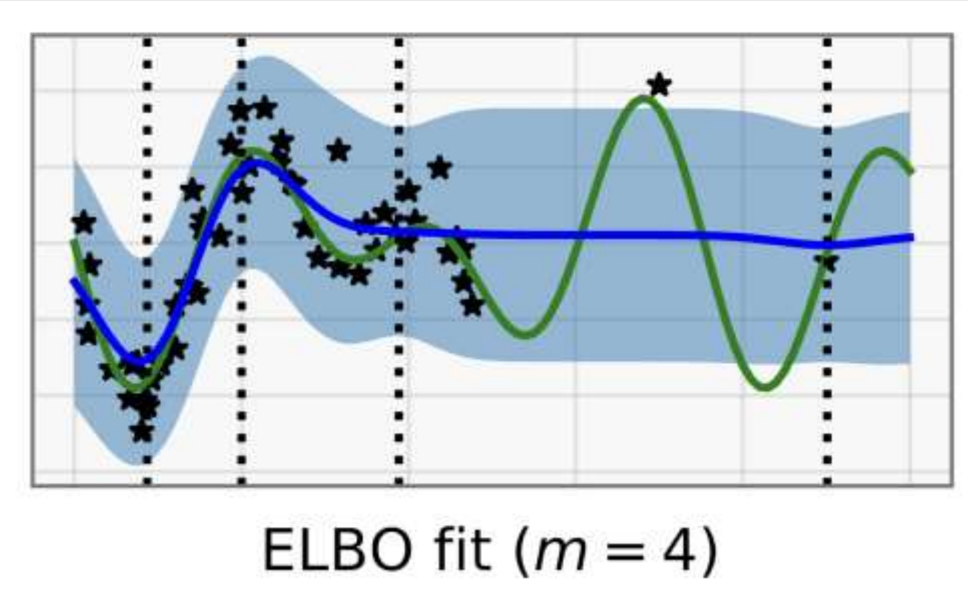
Latent Space **Generative Model** **Output**

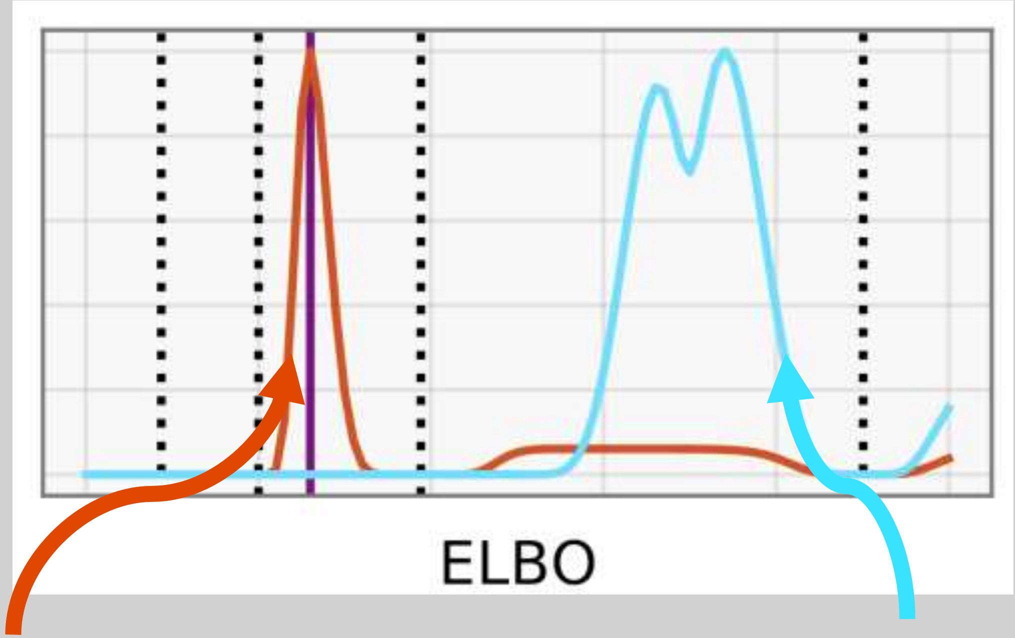


GP Surrogate

(Maddox et al., 2021; Vakili et al., 2021;
Maus et al., 2022; Stanton et al., 2022;
Moss et al., 2023)

Challenge 2: High Throughput

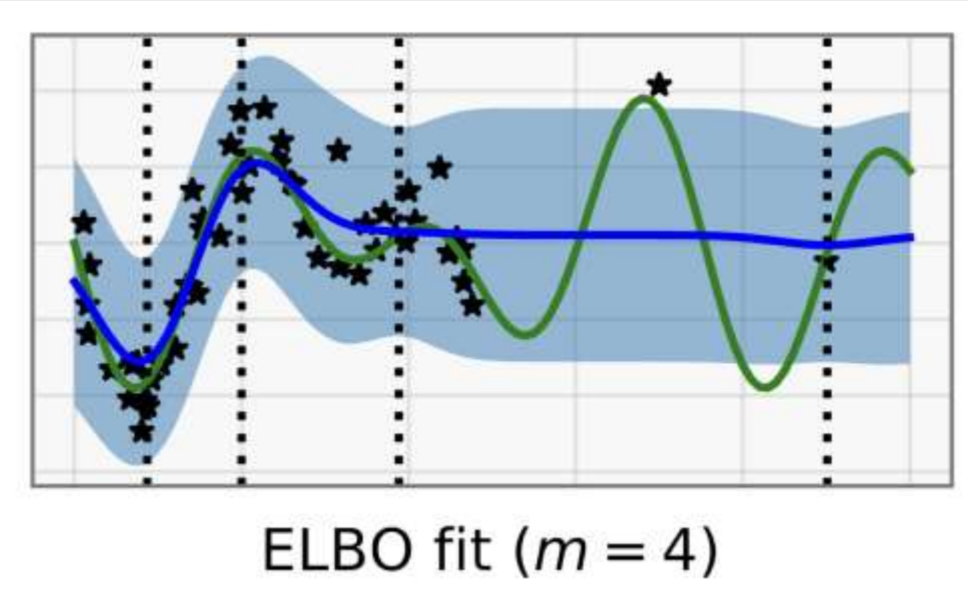


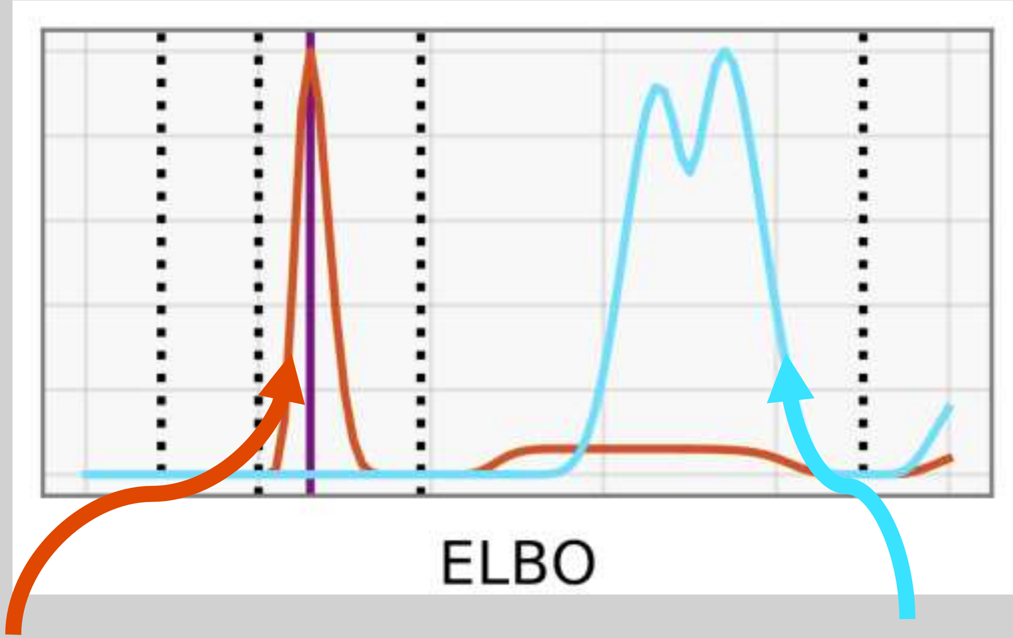


Train SVGP, compute EI

Exact EI

Challenge 2: High Throughput



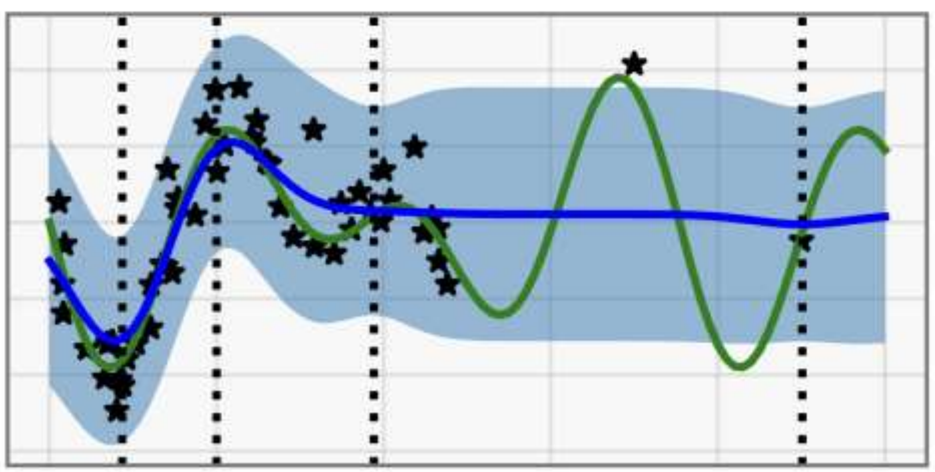


Train SVGP, compute EI

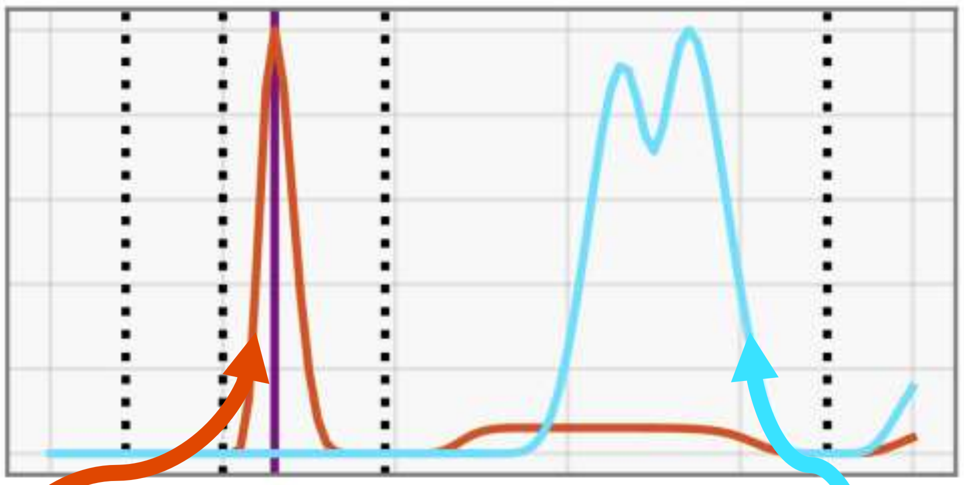
Exact EI

Suggestion: SVGPs are bad for BO because they aren't designed to be good for it.

Challenge 2: High Throughput



ELBO fit $(m = 4)$



ELBO

Train SVGP, compute EI **Exact EI**

Suggestion: SVGPs are bad for BO because they aren't designed to be good for it.

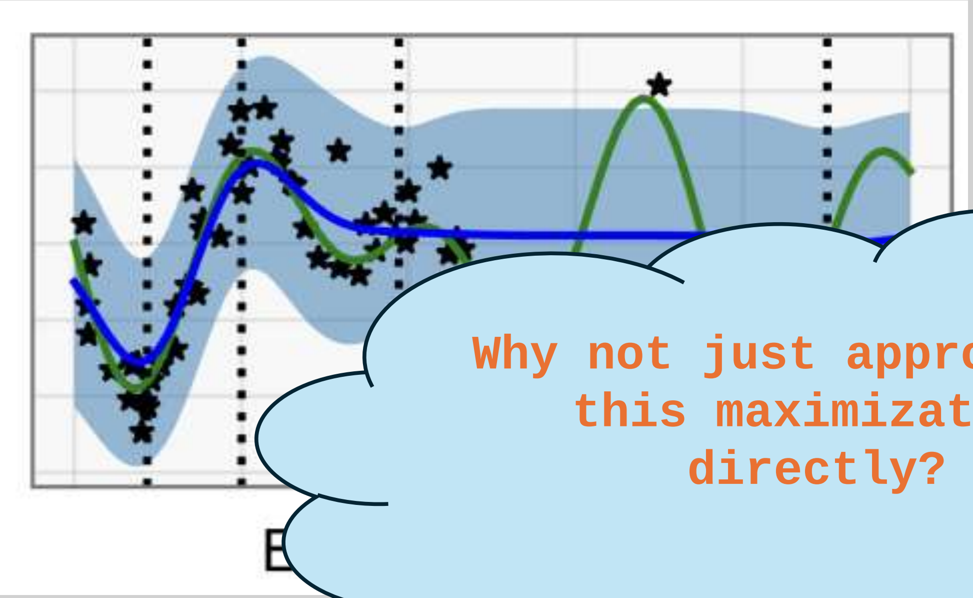
$$\underbrace{\arg\max \int u(\boldsymbol{x}, f; \mathcal{D}_t)\pi(f \mid \mathcal{D}_t)df}_{\text{Posterior expected utility}}$$

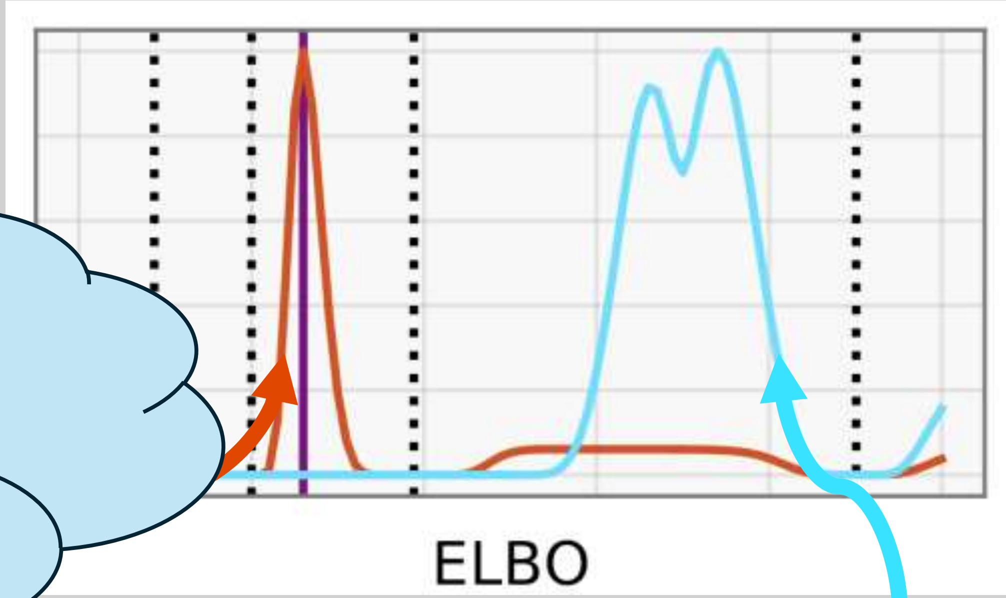
is not approximated by

1. $q_\lambda(f) \coloneqq \arg\max \mathcal{L}_{\text{ELBO}}(\lambda)$
2. $\arg\max \int u(\boldsymbol{x}, f; \mathcal{D}_t) q_\lambda(f) df$

"first approximate the posterior,
Then just plug that in."

Challenge 2: High Throughput





Why not just approximate this maximization directly?

ain SVGP, compute EI

Exact EI

Suggestion: SVGPs are bad for BO because they aren't designed to be good for it.

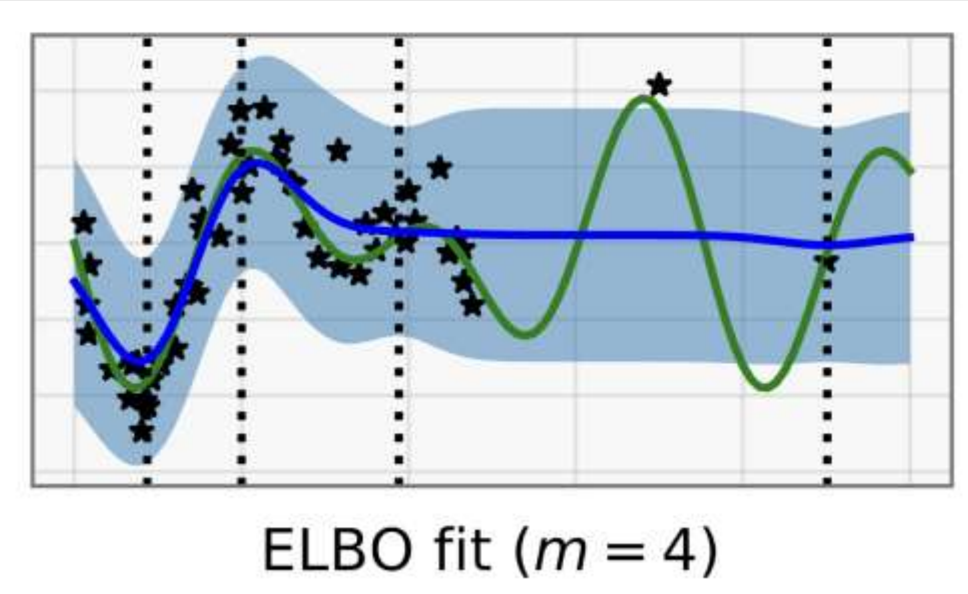
$$\underbrace{\arg\max \int u(\boldsymbol{x}, f; \mathcal{D}_t)\pi(f \mid \mathcal{D}_t)df}_{\text{Posterior expected utility}}$$

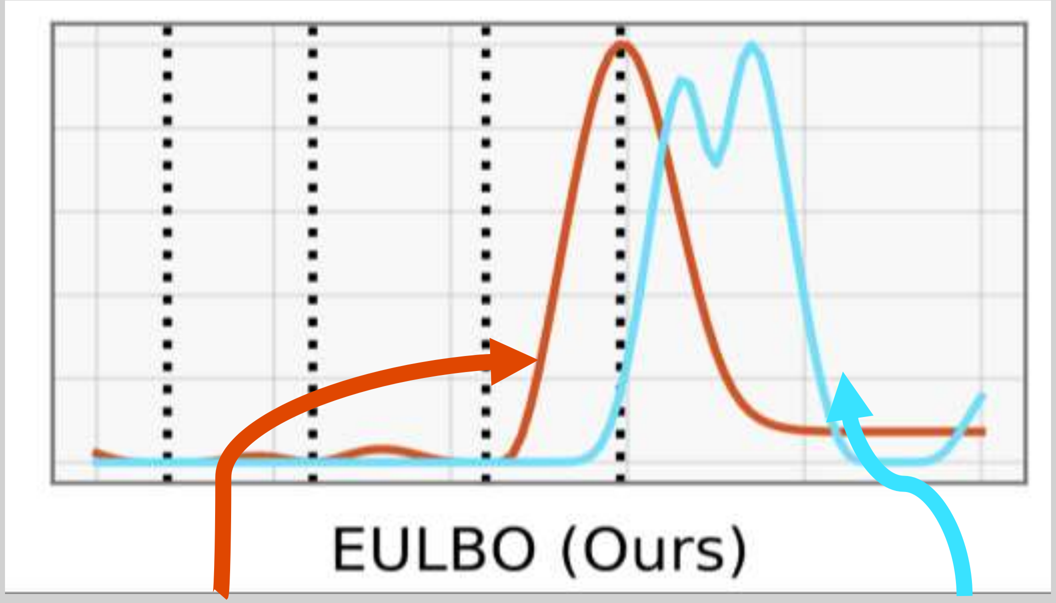
is not approximated by

1. $q_\lambda(f) := \arg\max \mathcal{L}_{\mathrm{ELBO}}(\lambda)$
2. $\arg\max \int u(\boldsymbol{x}, f; \mathcal{D}_t)q_\lambda(f)df$

"first approximate the posterior, Then just plug that in."

Challenge 2: High Throughput





Joint model selection + acquisition (EULBO)

Exact EI

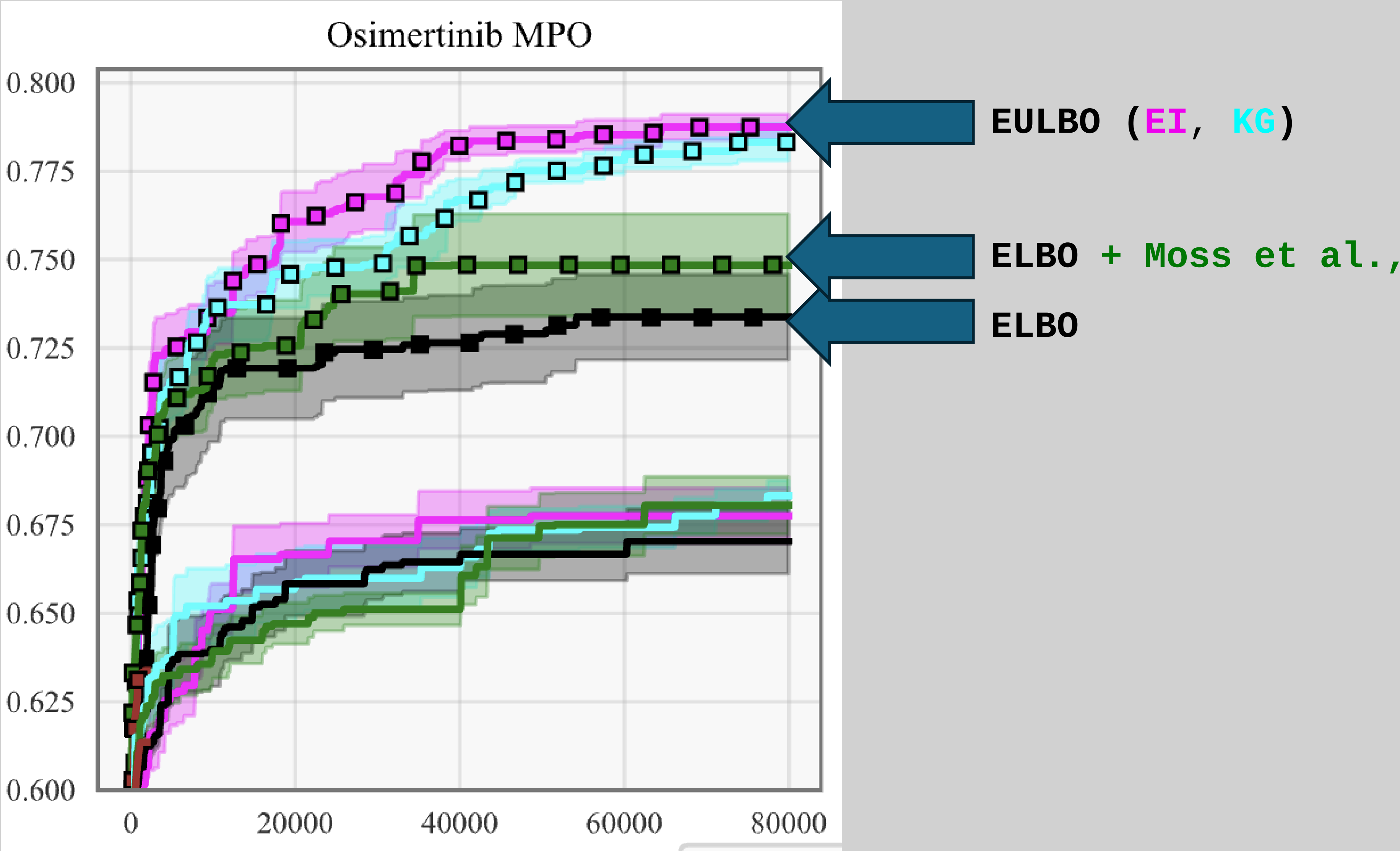
Idea: directly lower bound exact EI instead, **Expected Utility Lower Bound (EULBO).** Use for model selection and acquisition jointly.

After derivation:

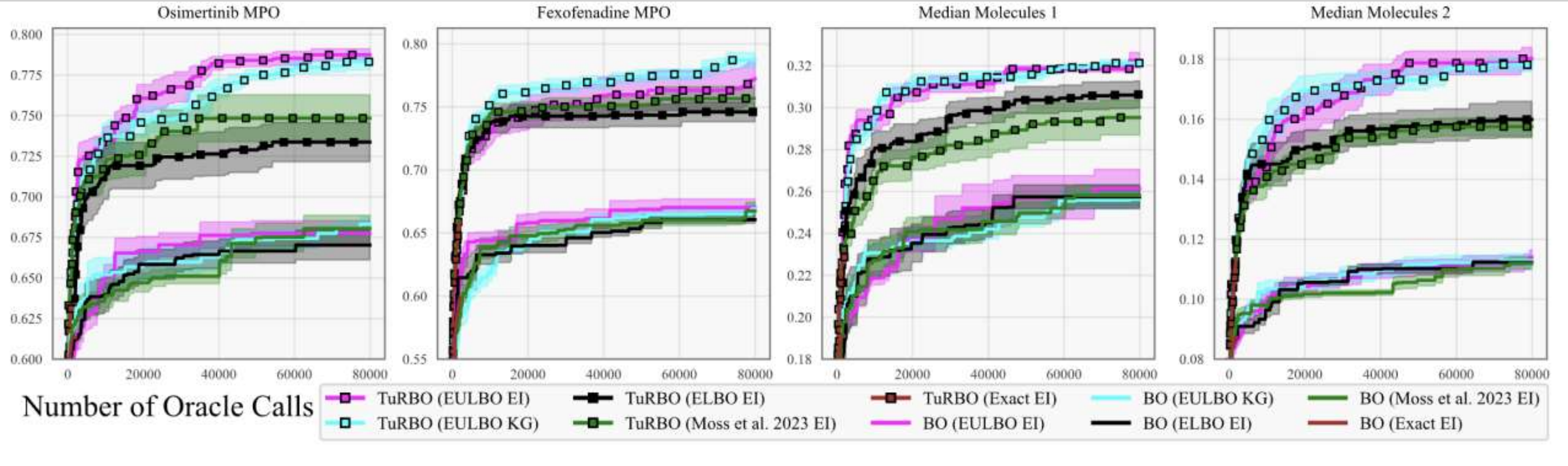
$$\mathcal{L}_{\text{EULBO}} = \mathcal{L}_{\text{ELBO}} + \mathbb{E}_{q_\lambda(f)}[\log u(\boldsymbol{x}, f; \mathcal{D}_t)]$$

(Flavor of loss calibrated VI; Lacoste-Julien et al., 2011)

Challenge 2: High Throughput



Challenge 2: High Throughput

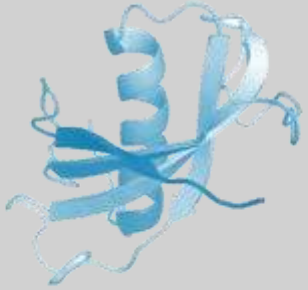
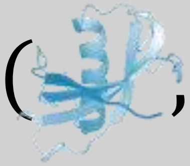


Challenge 3: Portfolio Risk

(AISTATS 2023)

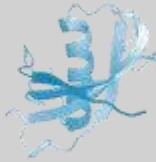
Idea: Find M solutions, all pairwise

 $= \arg\max f(x)$

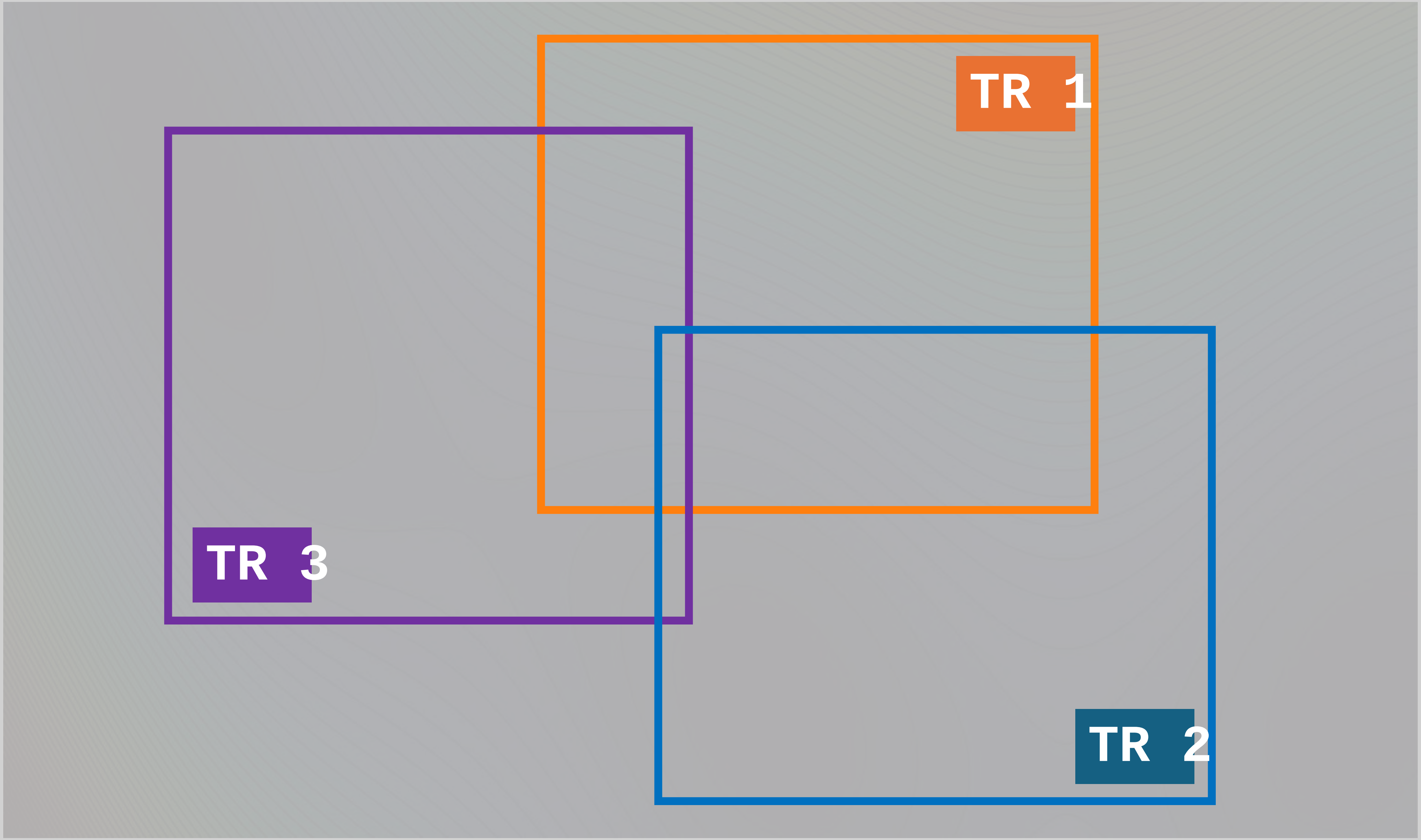
 $= \arg\max f(x)$ $\quad s.t. \quad$ $\delta($  ,  $) \geq \tau$

 $= \arg\max f(x)$ $\quad s.t. \quad$

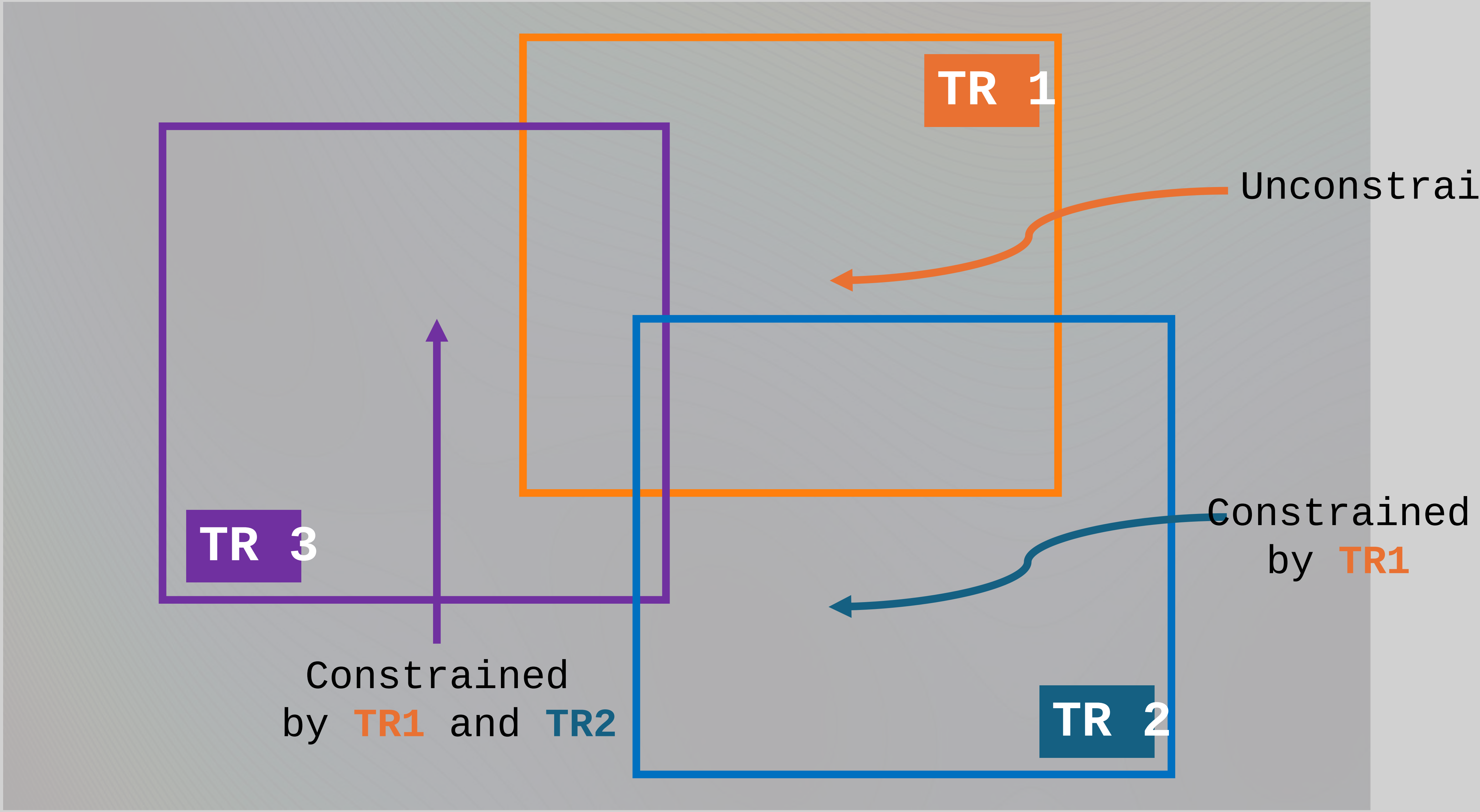
$\delta($  ,  $) \geq \tau$

$\delta($  ,  $) \geq \tau$

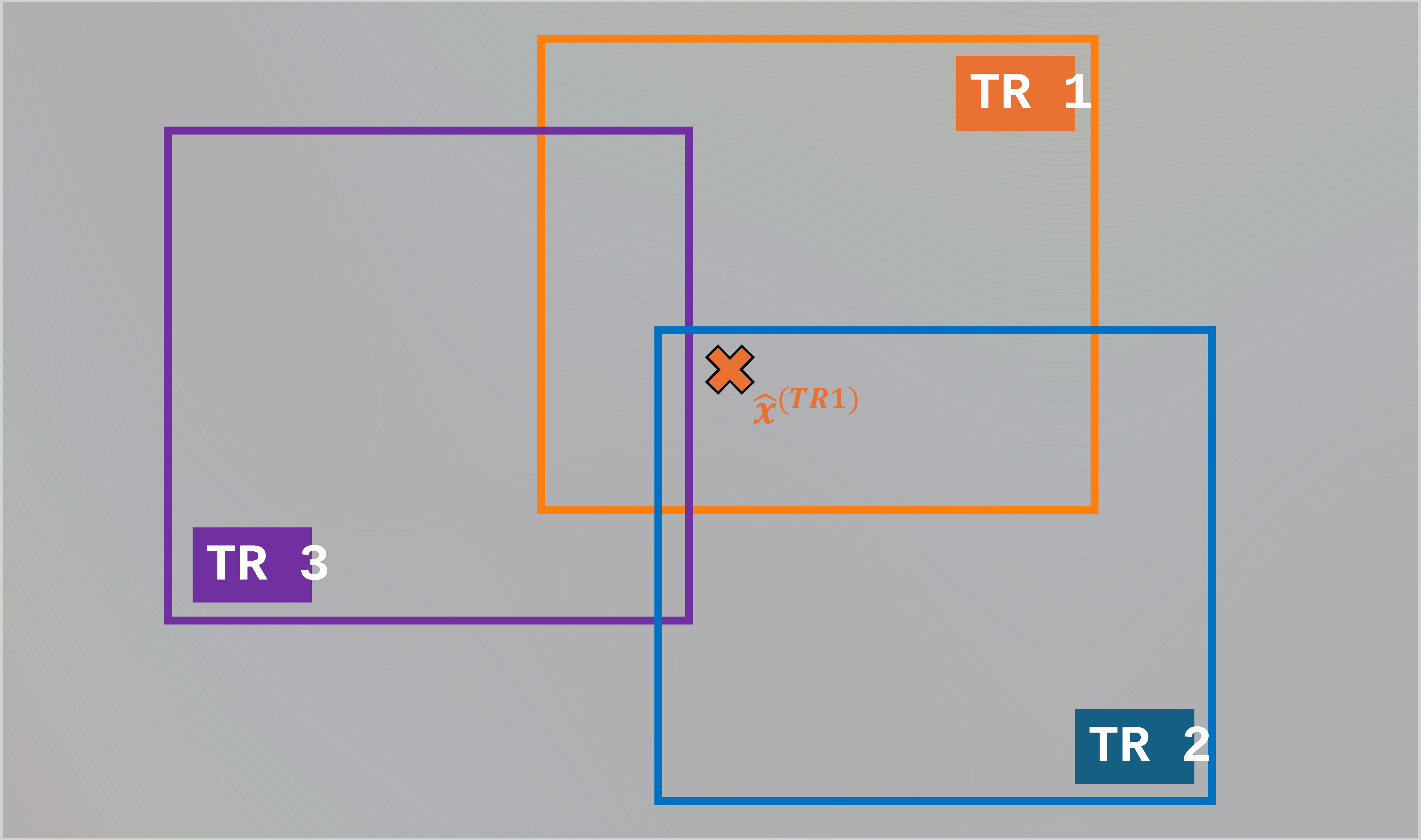
Finding Diverse Solutions



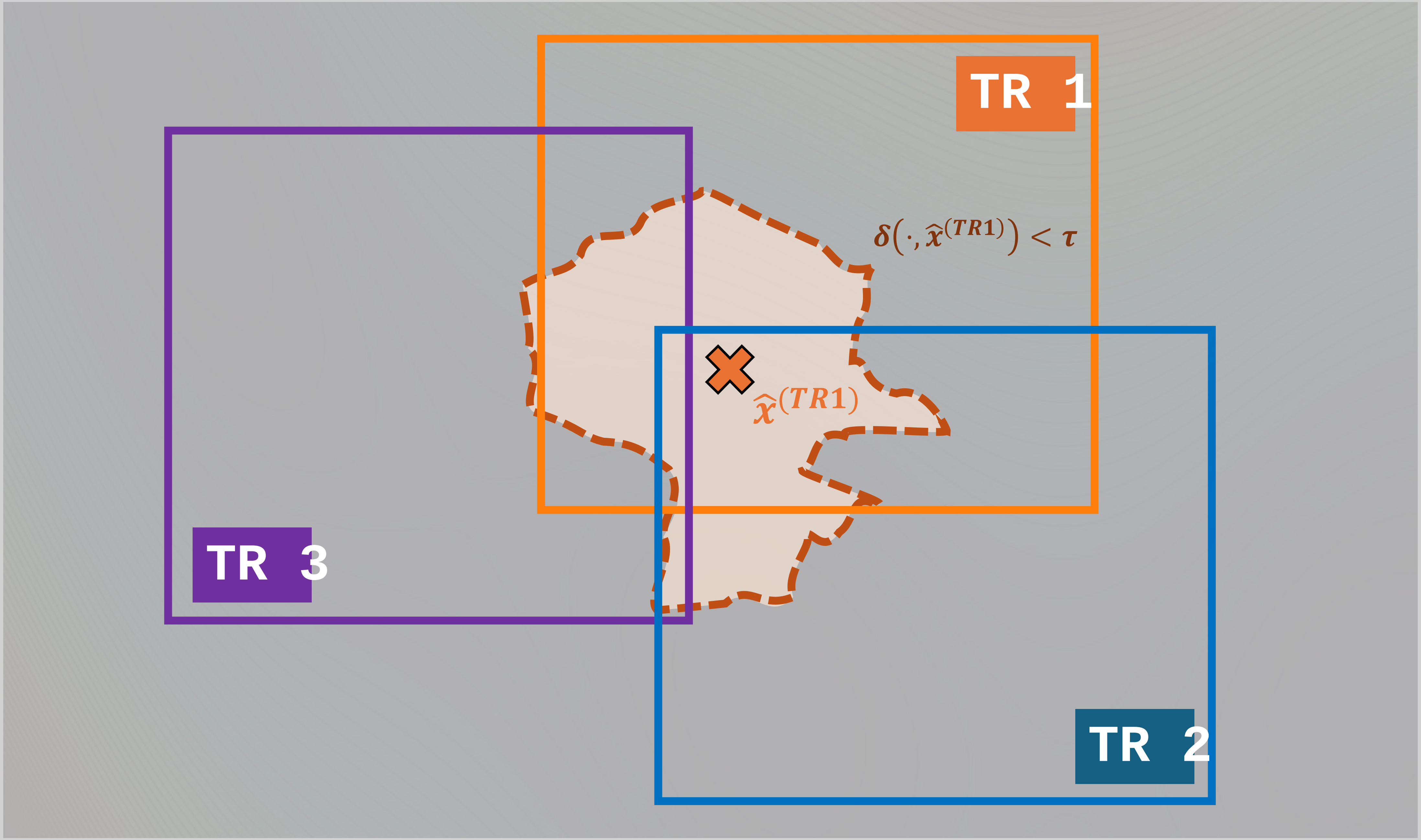
Finding Diverse Solutions



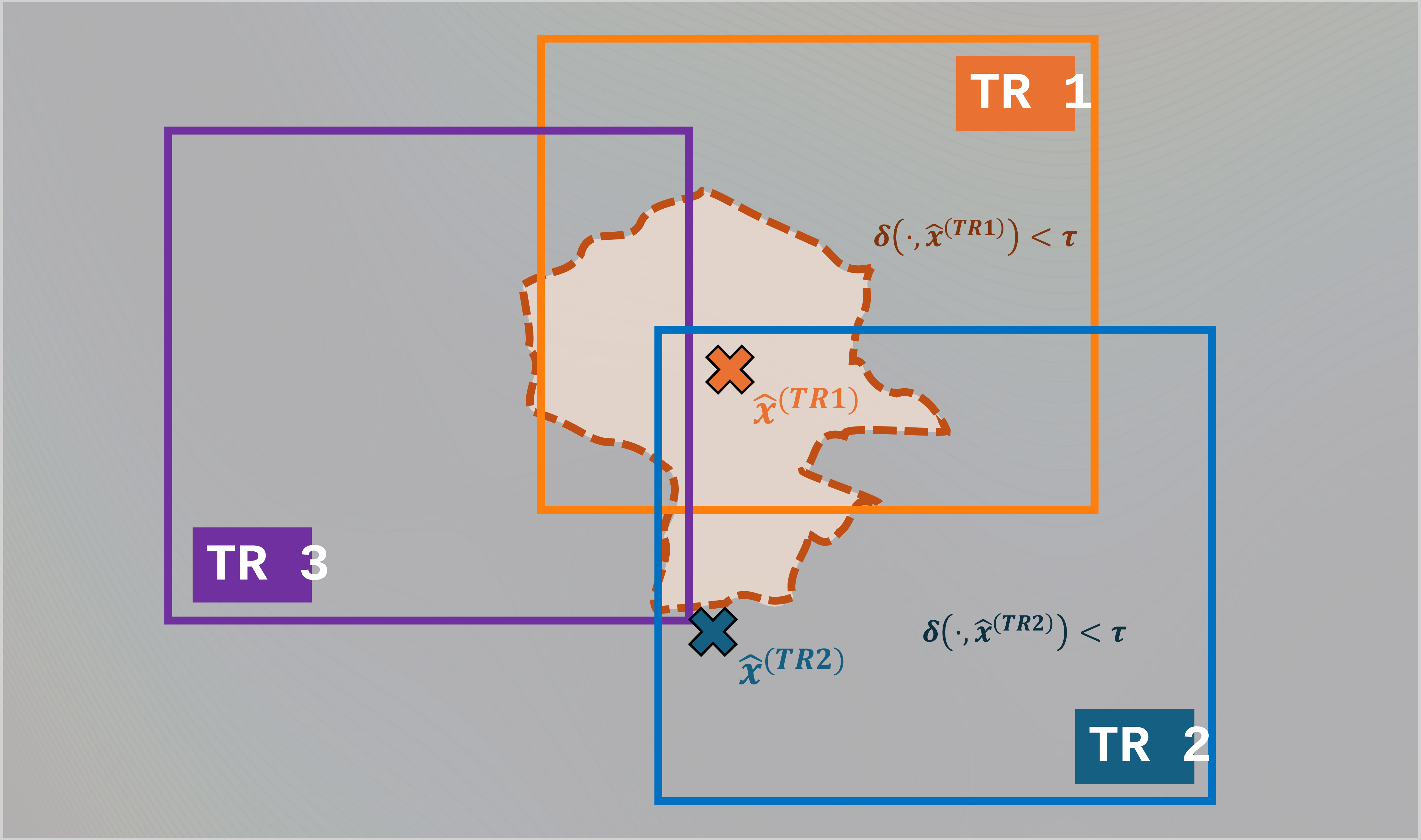
Finding Diverse Solutions



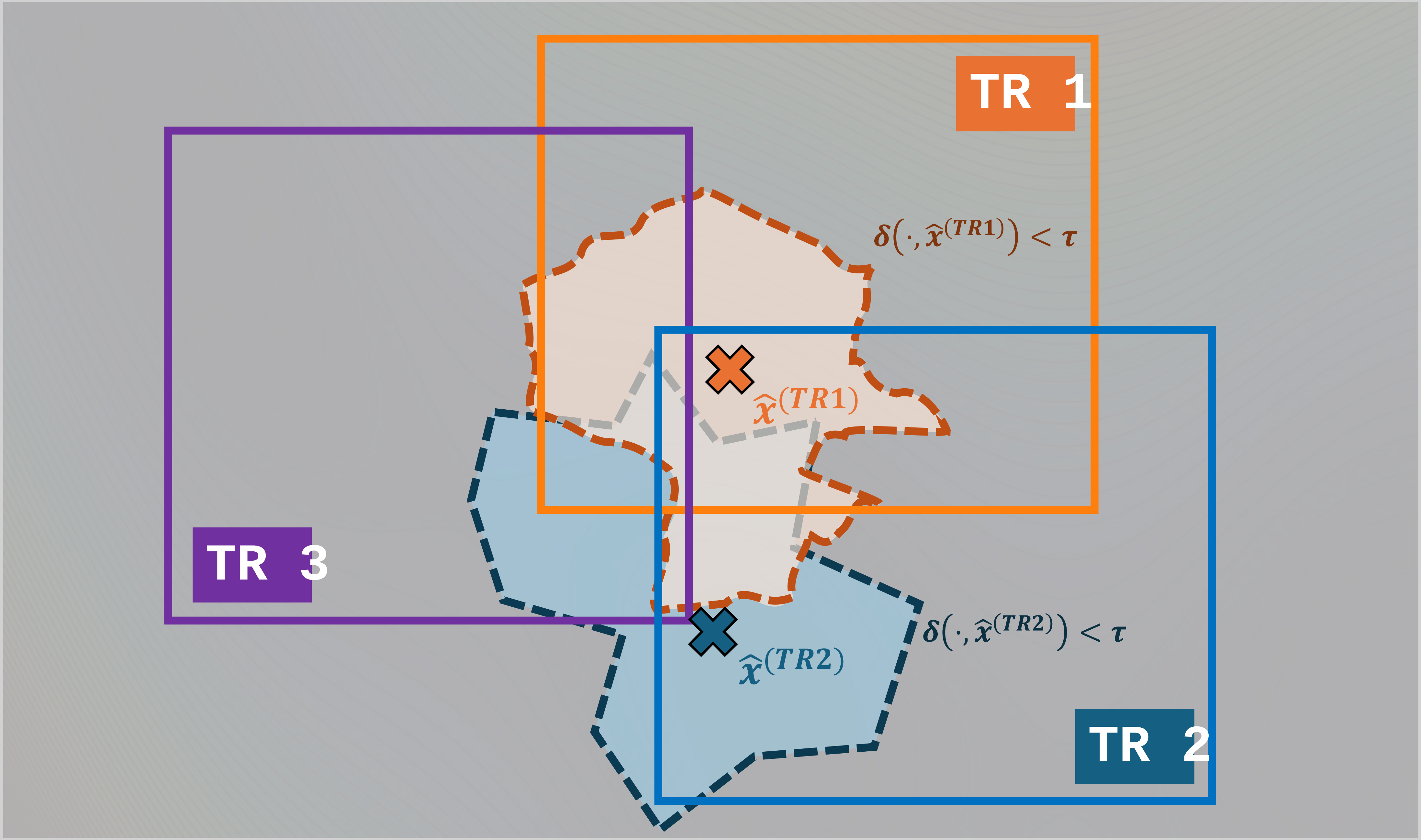
Finding Diverse Solutions



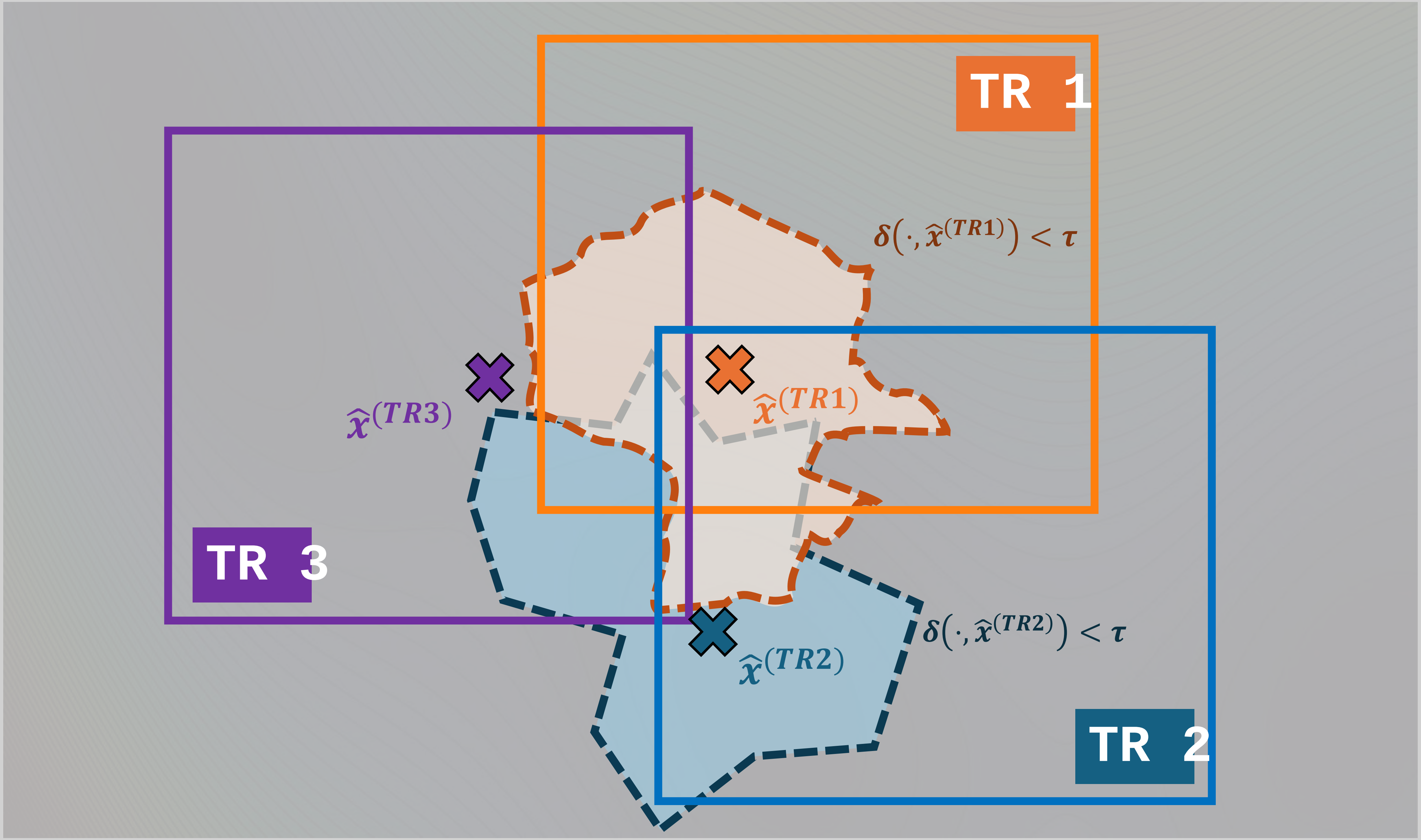
Finding Diverse Solutions



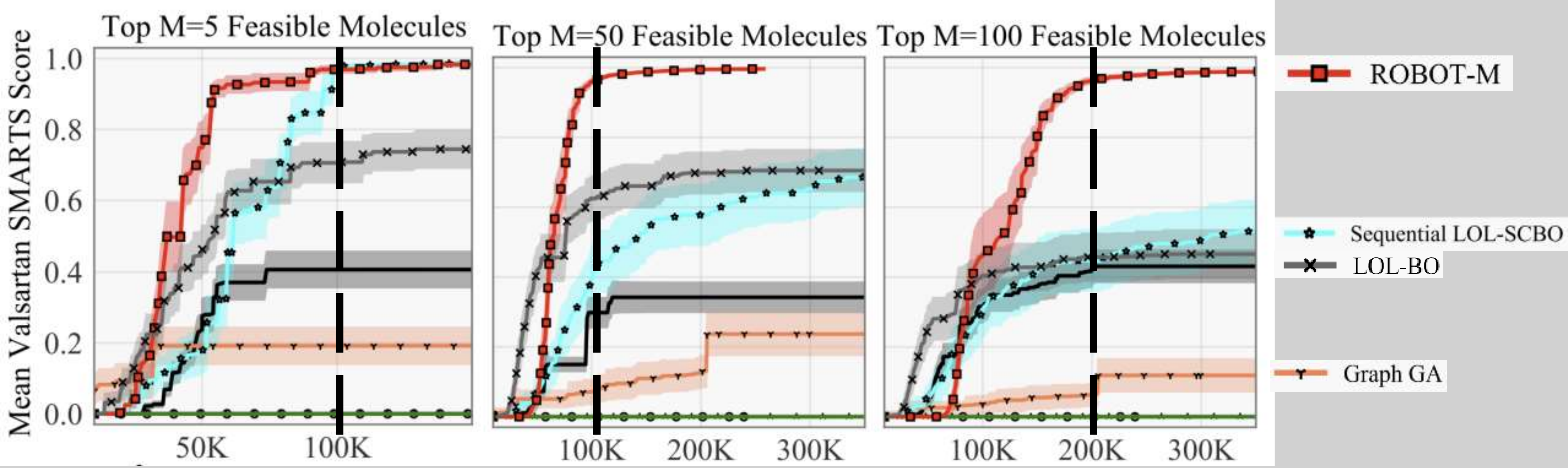
Finding Diverse Solutions



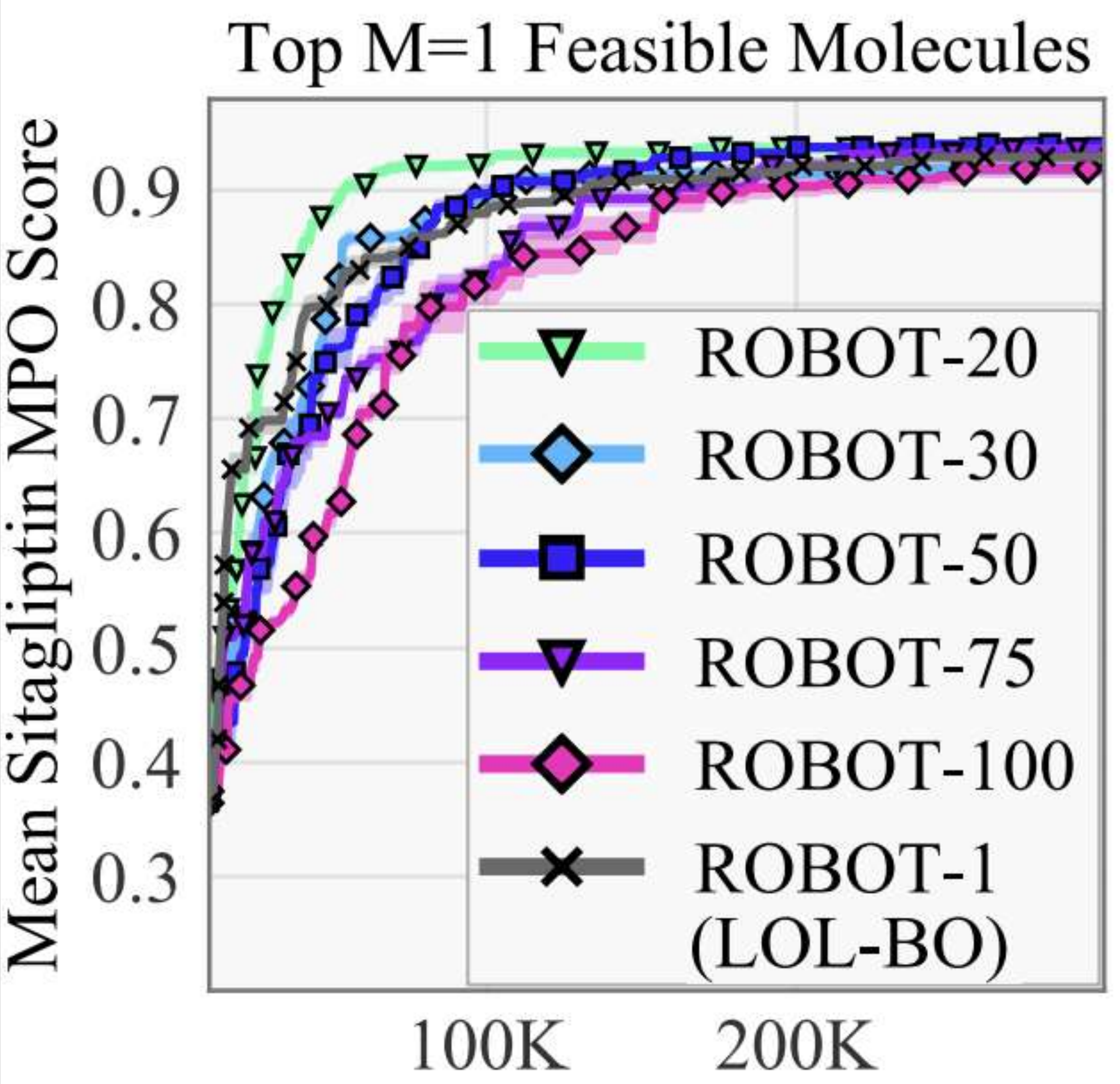
Finding Diverse
Solutions



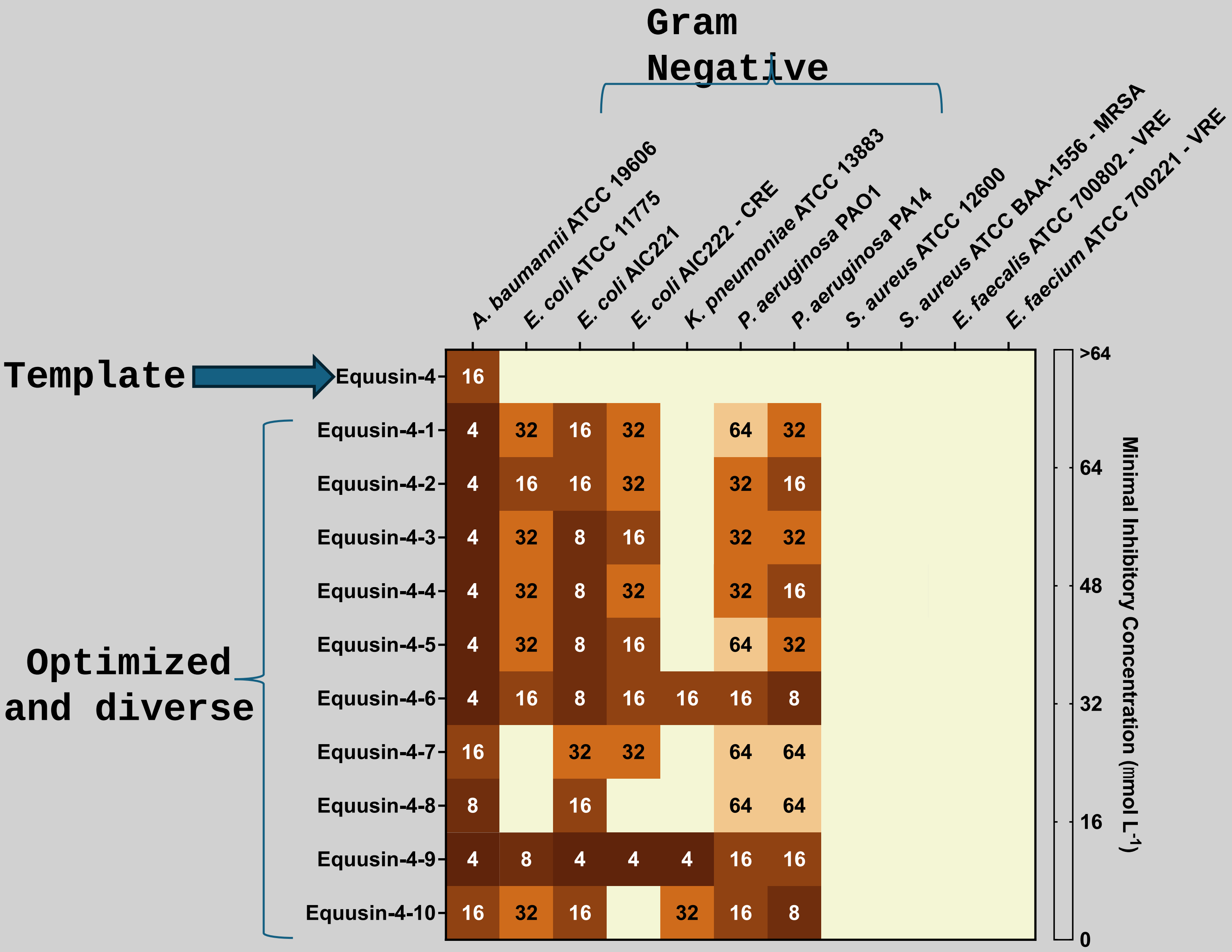
Finding Diverse Solutions



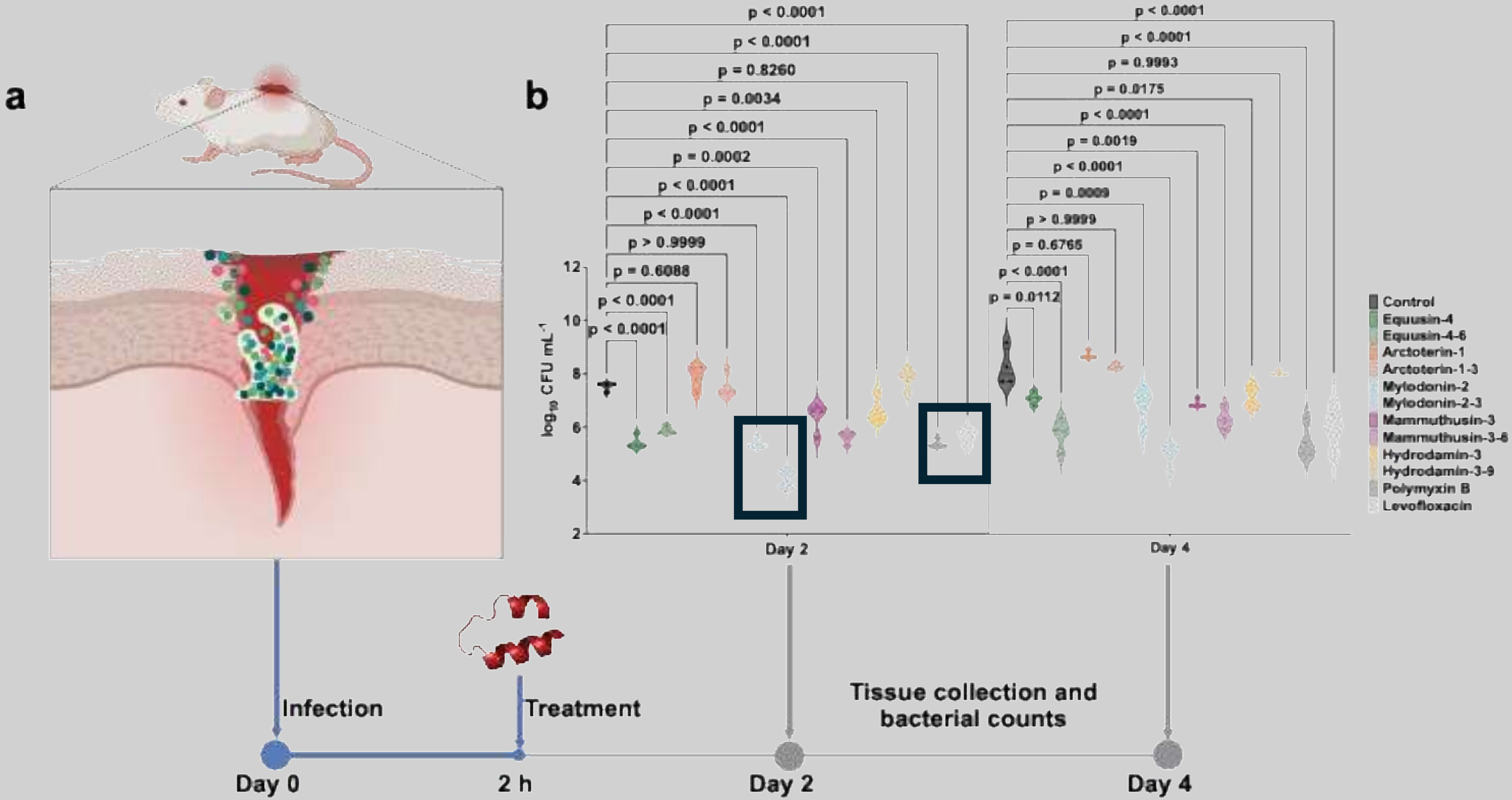
Finding Diverse Solutions



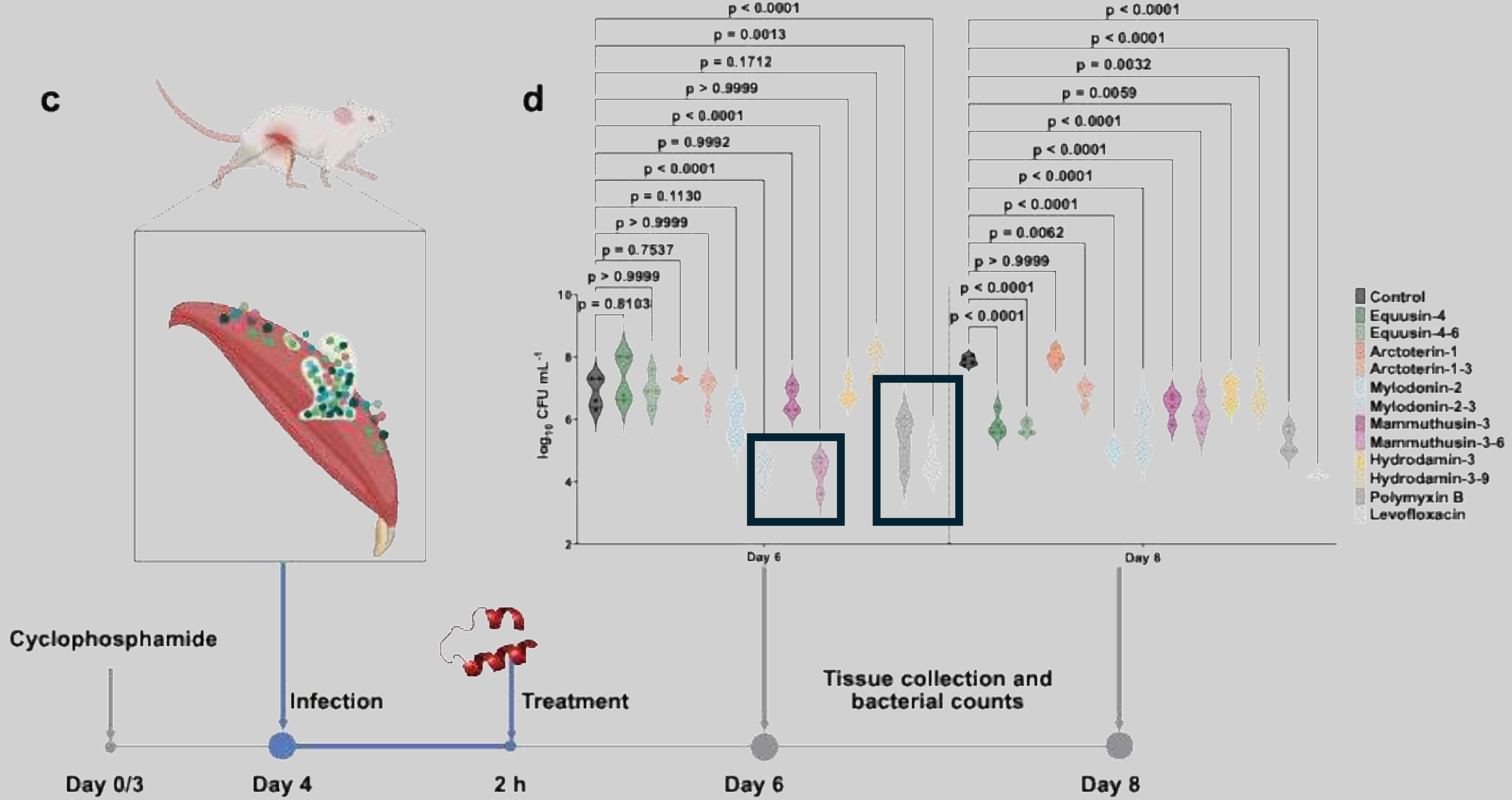
Antibiotics



Antibiotics



Antibiotics



Antibiotics — some stats

- 10 template peptides from 10 extinct organisms, optimize 10 diverse variants
  - **100** total peptides.
  - **68%** success rate in vitro, **6/8** tested working in vivo (although lower sam
  - **Two** were stronger than control antibiotics in mice.

- **Virtual screening** (e.g., Stokes et al., 2020; Liu et al., 2023): 3-5% success

ayesian optimization in a latent space is quite encouraging compared to virtual screening, generative approaches, ….

What can we be doing better?

What can we be doing better?

Generative Model

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C

Molecule in.

Molecule out.

What can we be doing better?

Generative Model

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

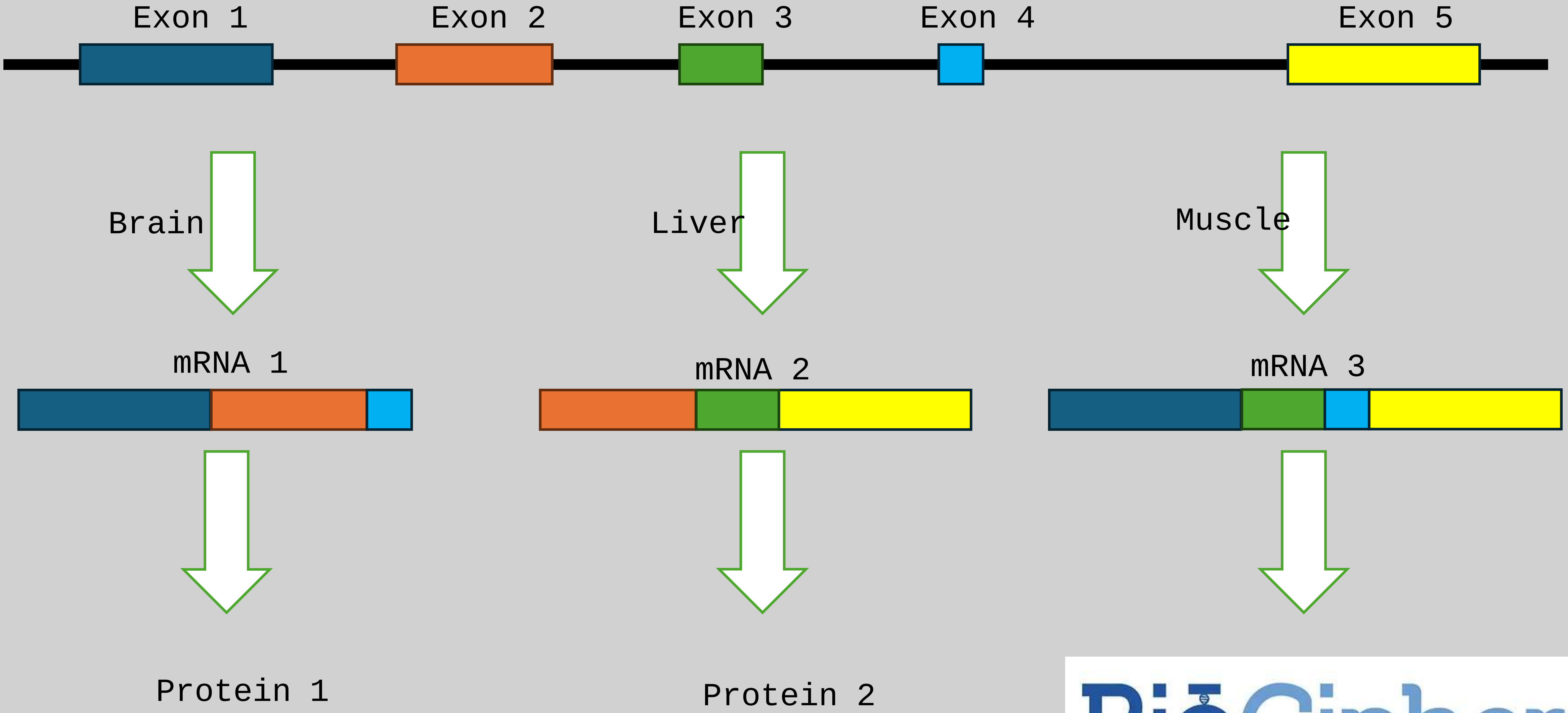
Molecule in.

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C

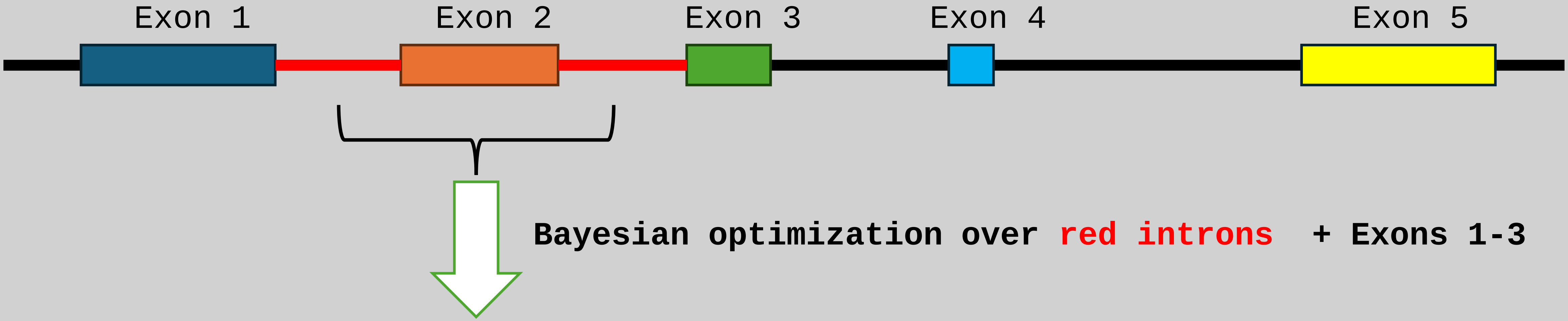
Molecule out.

Designed to **generate natural language text**, not **search for new molecules**

Example 1: Sparse BO in RNA



Example 1: Sparse BO in RNA

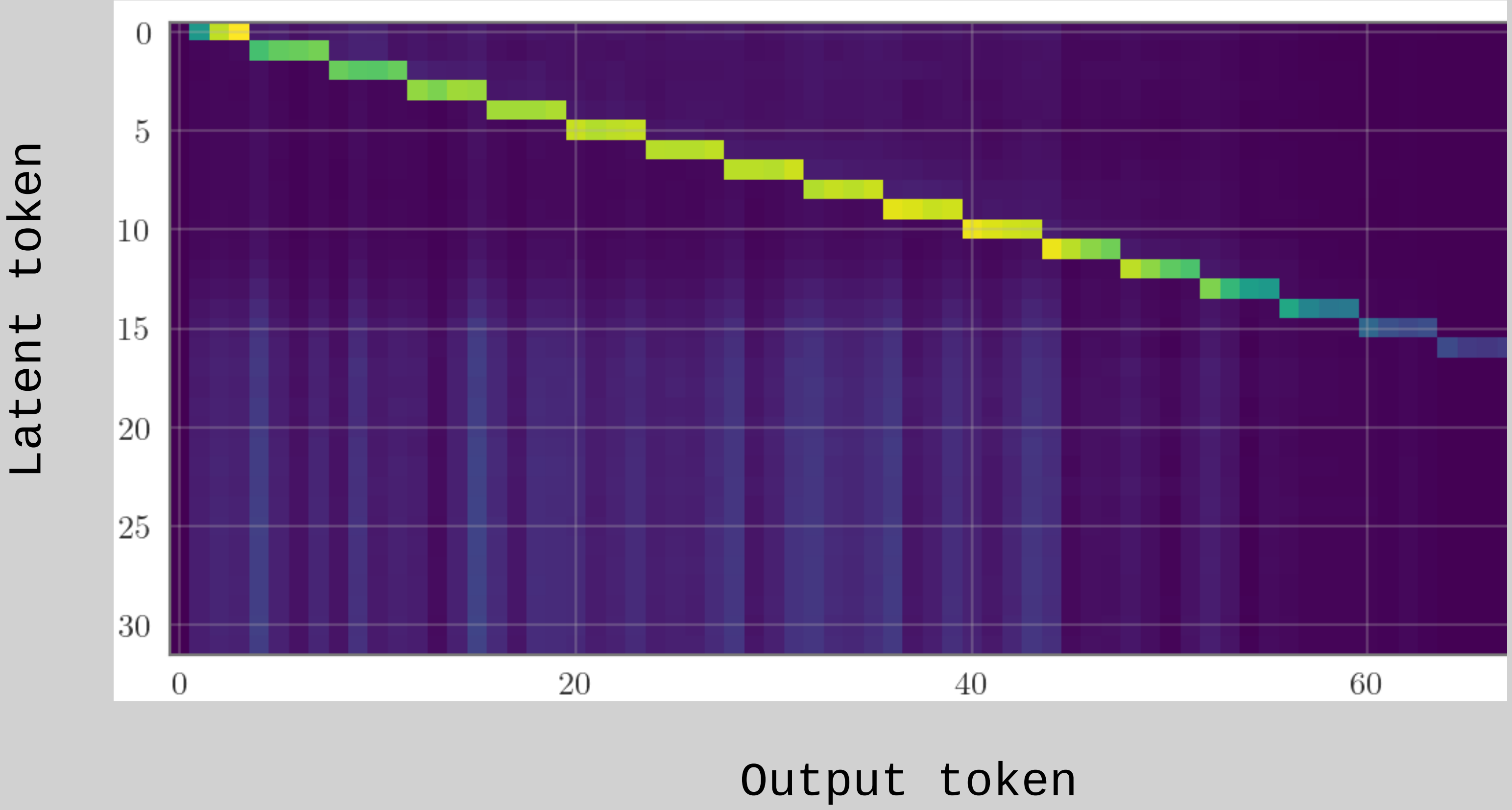


Maximize inclusion of **orange exon** in brain, minimize in all other tissues.

Experimentally: relatively few locations in sequence involved in splicing contro

Example 1: Sparse BO in RNA

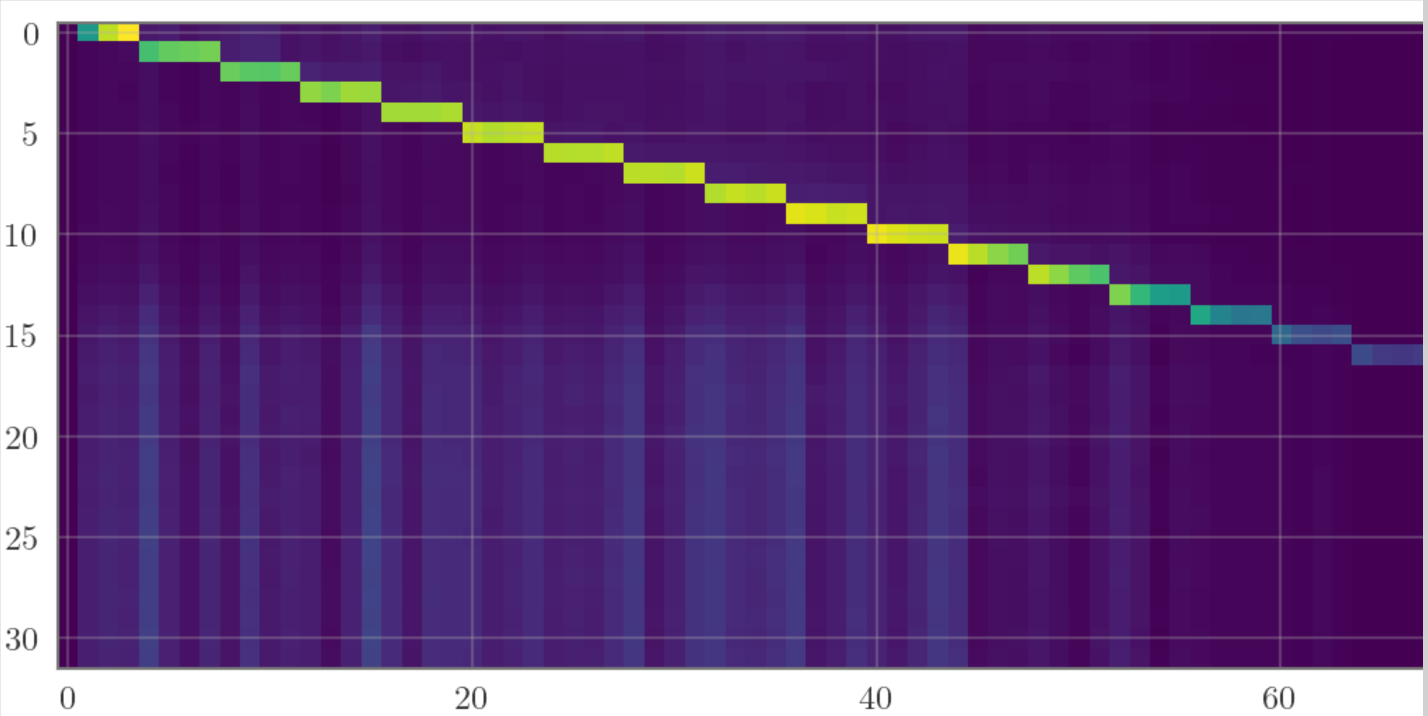
Step 1: Make **output tokens** sparse in **latent tokens.**



Example 1: Sparse BO in RNA

Step 1: Make **output tokens** sparse in **latent tokens.**

Latent token



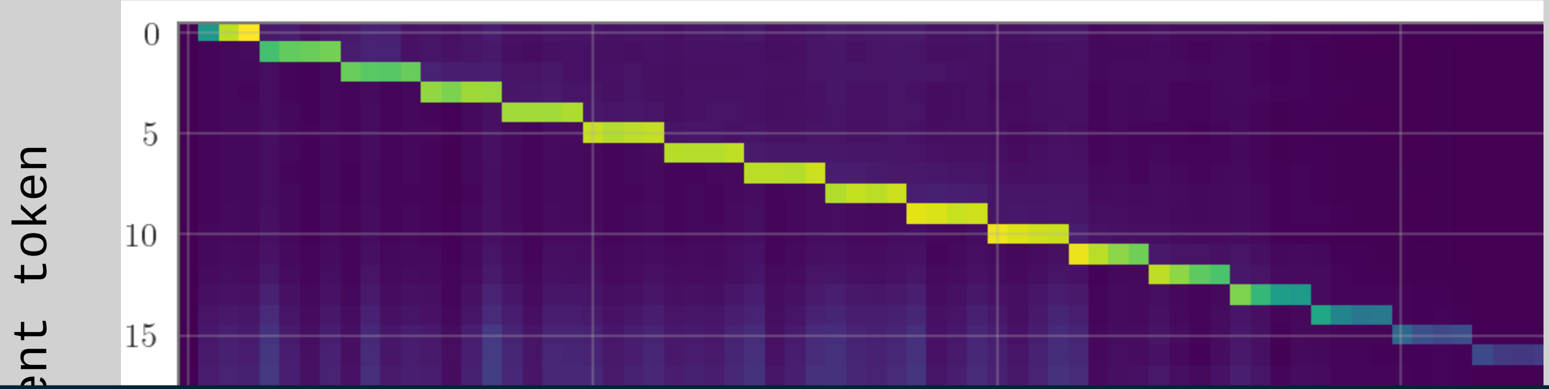
Output token

Step 2: **Just run e.g. SAASBO**

BioCiphers

Example 1: Sparse BO in RNA

Step 1: Make **output tokens** sparse in **latent tokens.**



If we want to use BO Algorithm X, the latent space needs to display the advantages that BO Algorithm X exploits.

Step 2: **Just run e.g. SAASBO**

BioCiphers

What can we be doing better?

Generative Model

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C

Molecule in.

Molecule out.

Designed to **generate natural language text**, not **search for new molecule**

What can we be doing better?

Generative Model

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C

Molecule in.

Molecule out.

Not where the knowledge is.

What can we be doing better?

Generative Model

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C

Molecule in.

Molecule out.

Not where the knowledge is.

Measurable, data exists.

✓ Make a version of this extinct peptide that $\underbrace{\text{kills A. baumannii}}$

Make a version of this extinct peptide that **kills A. baumannii**

What can we be doing better?

Generative Model

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C

Molecule in.

Molecule out.

Not where the knowledge is.

Measurable, data exists.

✓ Make a version of this extinct peptide that **kills A. baumannii**

🚫 Make a version of this peptide that kills A. baumannii **via membrane perm**

What can we be doing better?

Generative Model

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C

Molecule in.

Molecule out.

Not where the knowledge is.

Measurable, data exists.

✓ Make a version of this extinct peptide that **kills A. baumannii**

⊘ Make a version of this peptide that kills A. baumannii **via membrane depo**

What can we be doing better?

Generative Model

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO → CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C

Molecule in.

Molecule out.

Not where the knowledge is.

Measurable, data exists.

Make a version of this extinct peptide that **kills A. baumannii**

Make a version of this peptide that kills A. baumannii **via membrane depo**

Thermostability? Cytotoxicity?

Example 2: People talk about drugs.

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

Example 2: People talk about drugs.

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

Hydroxychloroquine is an anti-malarial drug commonly used in rheumatologic or dermatological conditions. Its use is partly due to its good safety profile**; however, it is estimated that up to 7.5% of long-term users may develop retinal toxicity.** [1] The mechanism of drug toxicity is still controversial. **Hydroxychloroquine tends to accumulate in pigmented tissues such as retinal pigment epithelium (RPE), inhibiting the activity of anionic transporter 1A2 (OATP1A2) . [2] Another theory suggests that the drug accumulates in the photoreceptors, leading to secondary degeneration of the outer nuclear layer and, subsequently, of RPE**. [3] Currently, it is hypothesized that hydroxychloroquine accumulates primarily at the level of the ganglion cell layer but its implications are still unknown. [4] Although usually asymptomatic, retinal toxicity may be associated with photosensitivity and presence of scotomas.'

Example 2. People talk about drugs.

C1[C@@H]([C@H](O[C@H]1N2C=C(C(=O)NC2=O)C(F)(F)F)CO)O

'Shitara and colleagues suggest that adverse effects caused by trifluridine/tipiracil might be easier to manage than those caused by immune checkpoint inhibitors, and underline this as a possible benefit of trifluridine/tipiracil. Although this hypothesis might be true, **grade 3 or worse adverse events occurred in 267 (80%) of 335 patients given trifluridine/tipiracil in this trial, compared with 137 (42%) of 330 patients given nivolumab in the ATTRACTION-2 trial**. 2 Compared with previous trials of trifluridine/tipiracil, the number of adverse events seems to be high in the TAGS study. For example, **in the RECOURSE-2 trial, 4 in a population with pretreated metastatic colorectal cancer, 69% of patients given trifluridine/tipiracil had grade 3 or worse adverse events.**'

Example 2: People talk about drugs.

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

13.1 Consolidated References

7,096 items　Download

Search　SORT BY Publication Date - Most Recent

LaCo0.95Mo0.05O3/CeO2 composite can promote the effective activation of peroxymonosulfate via Co3+/Co2+ cycle and realize the efficient degradation of hydroxychloroquine sulfate

Publication Name: Journal of Colloid and Interface Science
Publication Date: 2025-01-15
PMID: 39341146　**DOI:** 10.1016/j.jcis.2024.09.174

Inter-eye asymmetry of microvascular density in patients on hydroxychloroquine therapy by optical coherence tomography angiography

Publication Name: Microvascular Research
Publication Date: 2025-01
PMID: 39288847　**DOI:** 10.1016/j.mvr.2024.104747

A new three-dimensional modeling of fluorescence excitation-emission measurements to explore the interaction of hydroxychloroquine and DNA, and to quantify their binding constant

Publication Name: Colloids and surfaces. B, Biointerfaces
Publication Date: 2025-01
PMID: 39332057　**DOI:** 10.1016/j.colsurfb.2024.114266

Innovations in Cutaneous Lupus

Publication Name: Dermatologic Clinics

Example 2: People talk about drugs.

Tools · Download (9k) · Add · View: Cards Table · Share

P18615 · NELFE_HUMAN

Negative elongation factor E · **Gene:** NELFE (RD, RDBP) · Homo sapiens (Human) · 380 amino acids · Evidence at protein level · **Annotation score:** 5/5

#Repressor #RNA-binding #Transcription #Transcription regulation

1 domain · 21 PTMs · 3 isoforms · 6 interactions · 12 3D structures · 34 reviewed publications

P42285 · MTREX_HUMAN

Exosome RNA helicase MTR4 · **Gene:** MTREX (DOB1, KIAA0052, MTR4, SKIV2L2) · Homo sapiens (Human) · EC:3.6.4.13 · 1042 amino acids · Evidence at protein level · **Annotation score:** 5/5

#Helicase #Hydrolase #DNA damage #mRNA processing #mRNA splicing #rRNA processing

2 domains · 4 PTMs · 1 reviewed variant · 1 interaction · 11 3D structures · 32 reviewed publications

Q16270 · IBP7_HUMAN

Insulin-like growth factor-binding protein 7 · **Gene:** IGFBP7 (MAC25, PSF) · Homo sapiens (Human) · 282 amino acids · Evidence at protein level · **Annotation score:** 5/5

#Growth factor binding #Cell adhesion

3 domains · 1 PTM · 3 reviewed variants · 2 isoforms · 1 disease · 1 3D structure · 15 reviewed publications

Example 2: People talk about drugs.

Training

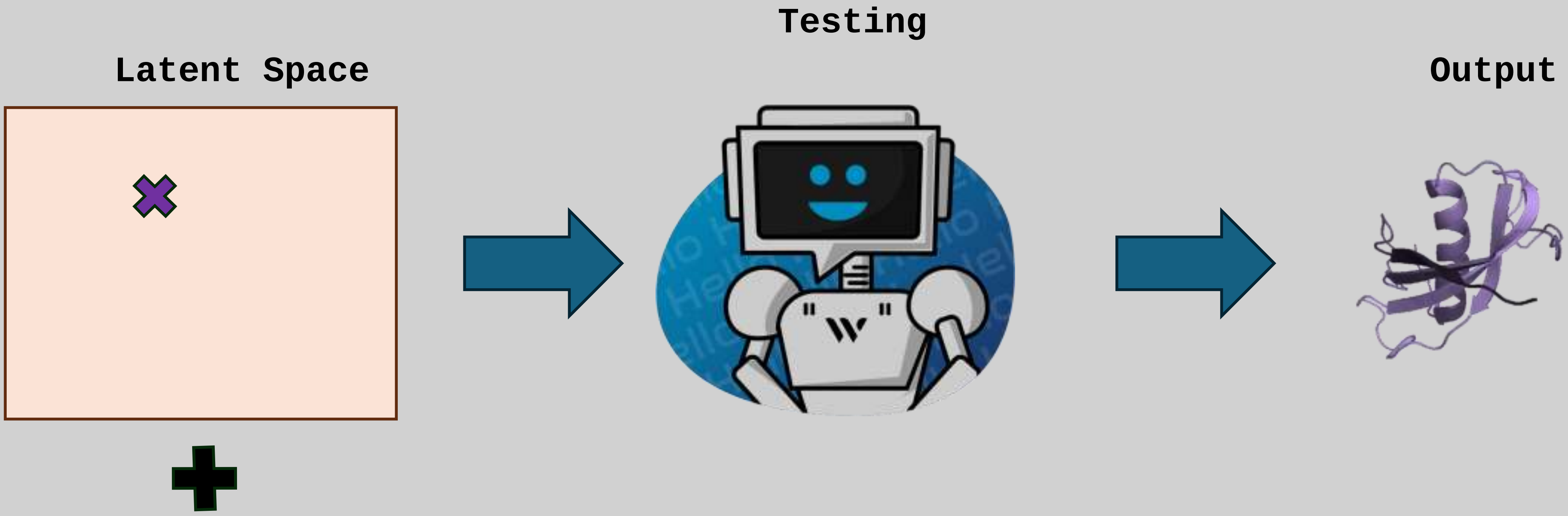
CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO

CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C

“Hydroxychloroquine tends to accumulate in pigmented tissues such as retinal pigment epithelium (RPE)”

+ e.g. CLIP objective.

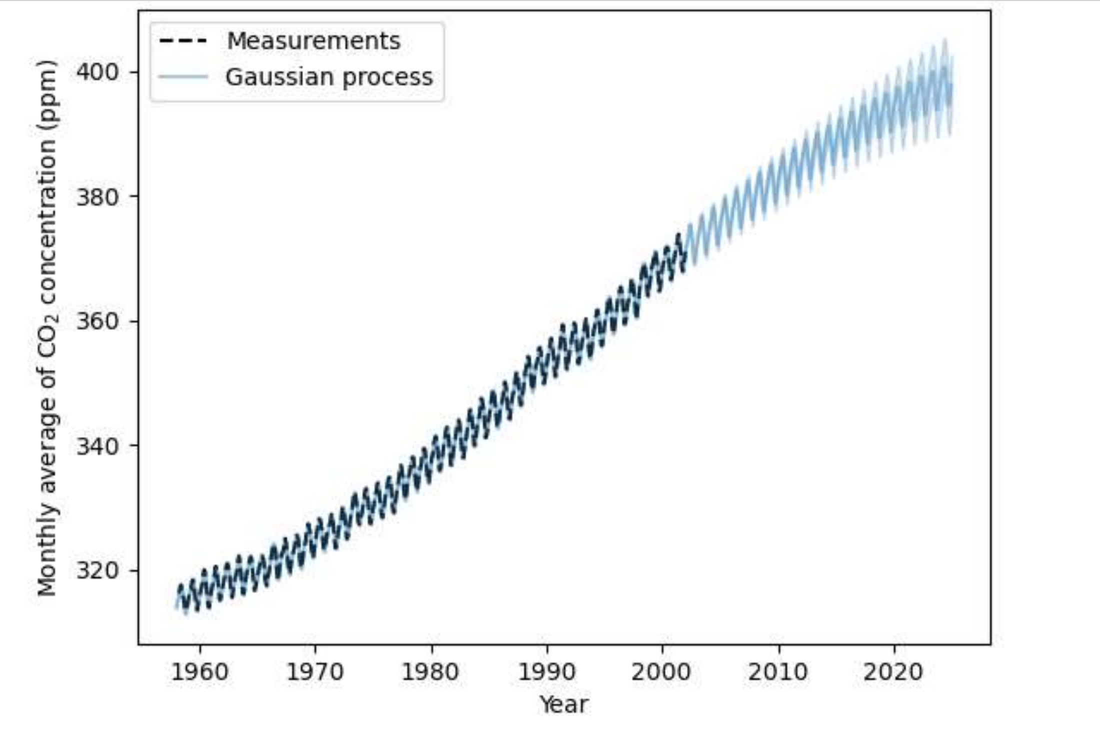
Example 2: People talk about drugs.



This peptide's primary mechanism of action is **membrane depolarization.** It melts at 65 °C.

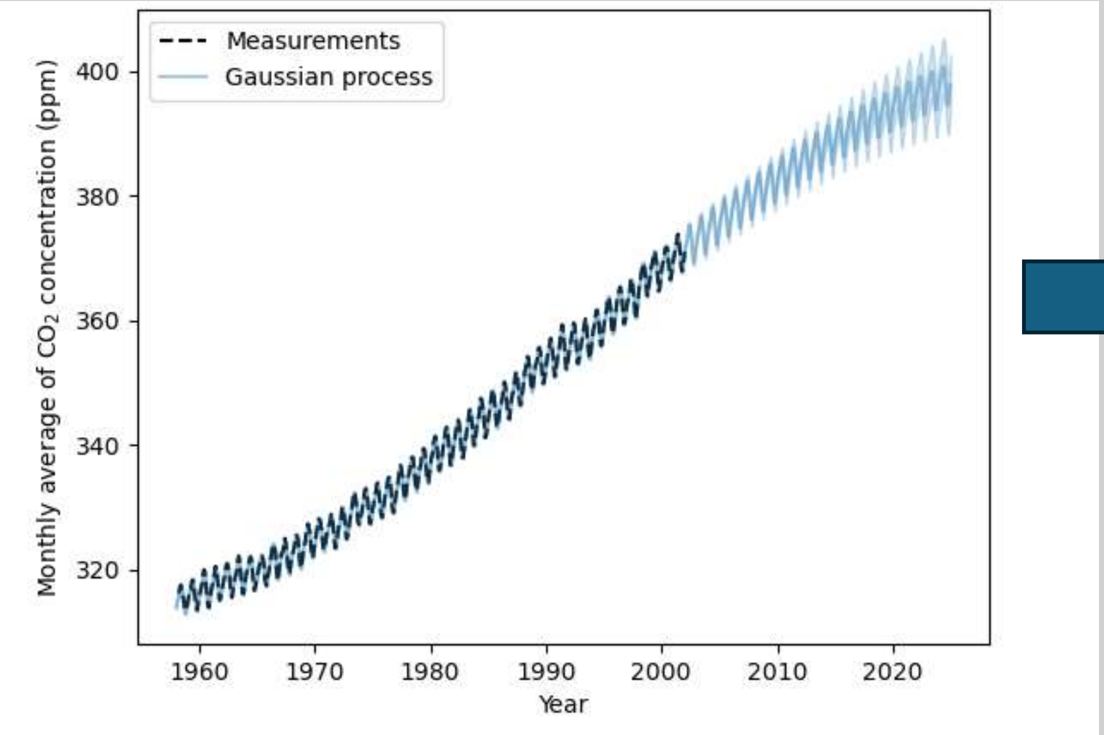
The betrayal slide

Prior knowledge means:

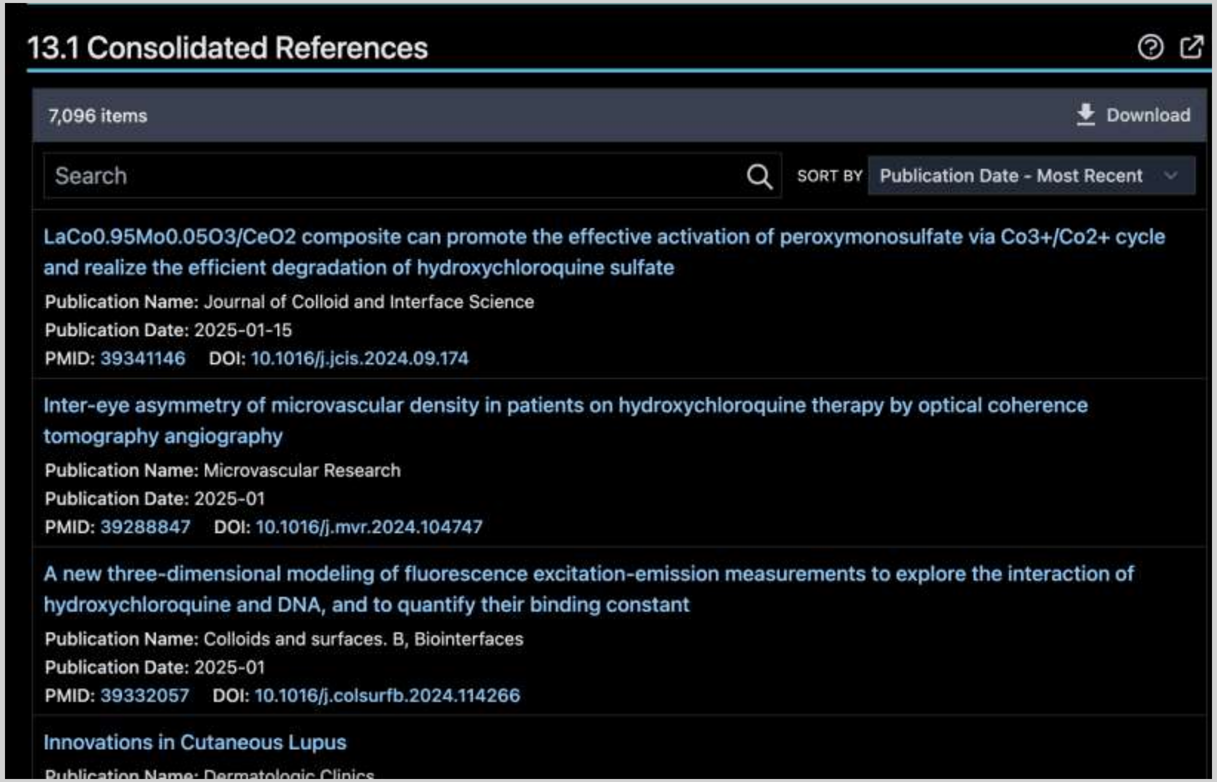


The betrayal slide

Prior knowledge means:



Prior knowledge means:



Conclusion

Bayesian optimization is great, quite possibly even state of the art for design pr

The representations of molecules that we’re searching through are **incredibly naïve**

The next big steps in this space might not be about the surrogate model, the acquisition function, or most of the other things you might find in my own methods sections.