

# DockTDesign: Many-objective Optimization with Deep Generative Models for *de novo* Drug Design

Matheus M. P. Silva<sup>1</sup>, Jaqueline S. Angelo<sup>2</sup>, Isabella A. Guedes<sup>1</sup>, Laurent E. Dardenne<sup>1</sup>

<sup>1</sup>Laboratório Nacional de Computação Científica (LNCC/MCTI)

<sup>2</sup>Fundação Oswaldo Cruz (Fiocruz)

Support:

GMMsB  
GRUPO DE MODELAGEM MOLÉCULAR DE SISTEMAS BIOLÓGICOS

LNCC  
Laboratório Nacional de  
Computação Científica

CAPES

FAPERJ  
Fundação Carlos Chagas Filho de Amparo  
& Pesquisas do Estado do Rio de Janeiro

dockthor

## Introduction

- *de novo* drug design is inherently a **many-objective optimization problem**, requiring simultaneous optimization of **multiple, often competing properties**—such as potency, selectivity, and synthetic accessibility [1].
- Despite its relevance, optimizing **four or more objectives** in generative drug discovery remains relatively **underexplored** in current scientific literature [1].
- We present a novel framework that combines a pre-trained variational autoencoder (**VAE**) with state-of-the-art many-objective evolutionary algorithms [2].
- The VAE enables chemically valid molecular generation in a **structured latent space**, while ManyEAs efficiently **navigate complex optimization landscapes**.

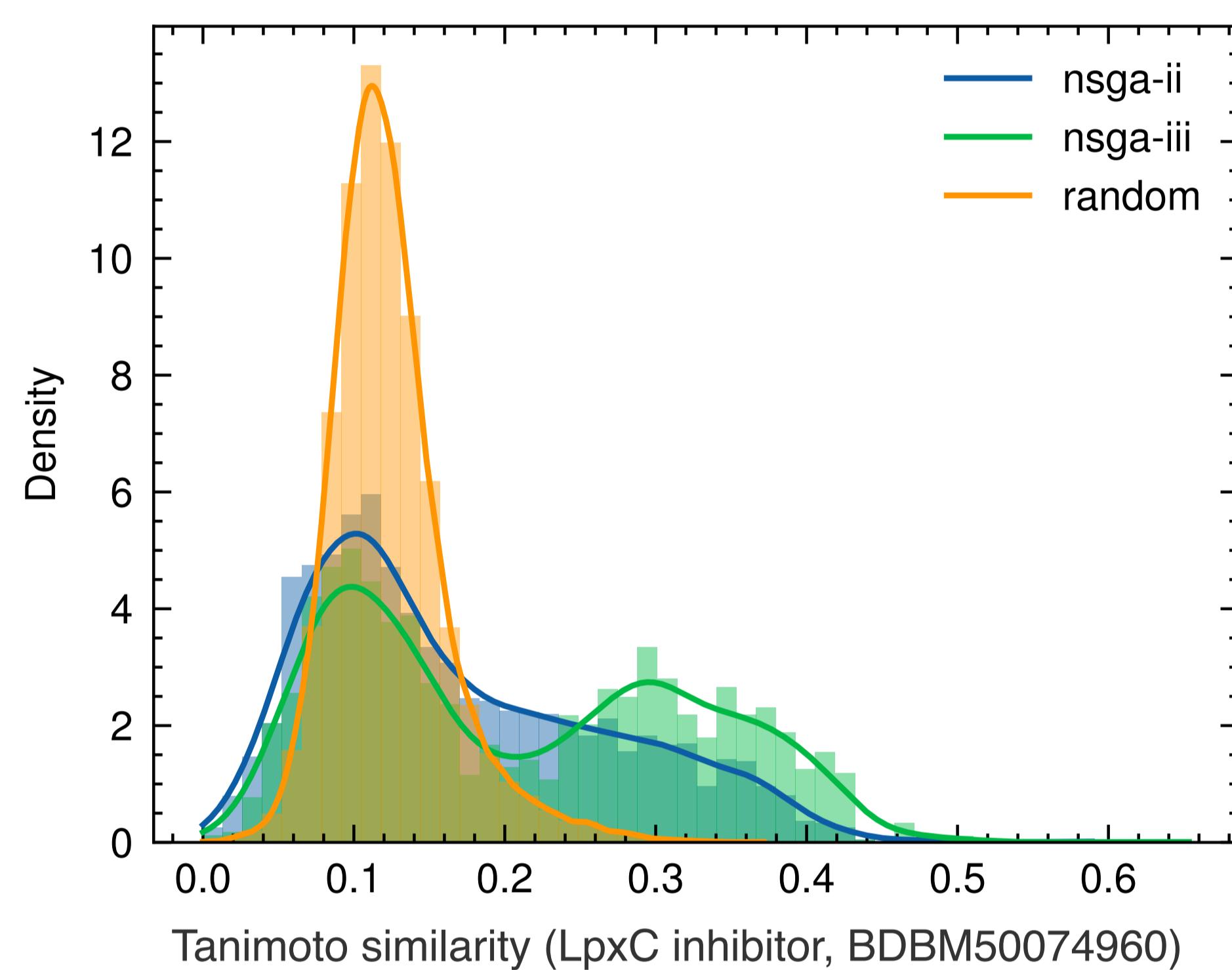


Fig. 1. Distribution of Tanimoto similarity to the reference LpxC inhibitor (BDBM50074960) for non-dominated solutions generated by NSGA-II (blue) and NSGA-III (green), compared to randomly generated molecules (orange). NSGA-III achieves higher similarity, reflecting better target alignment.

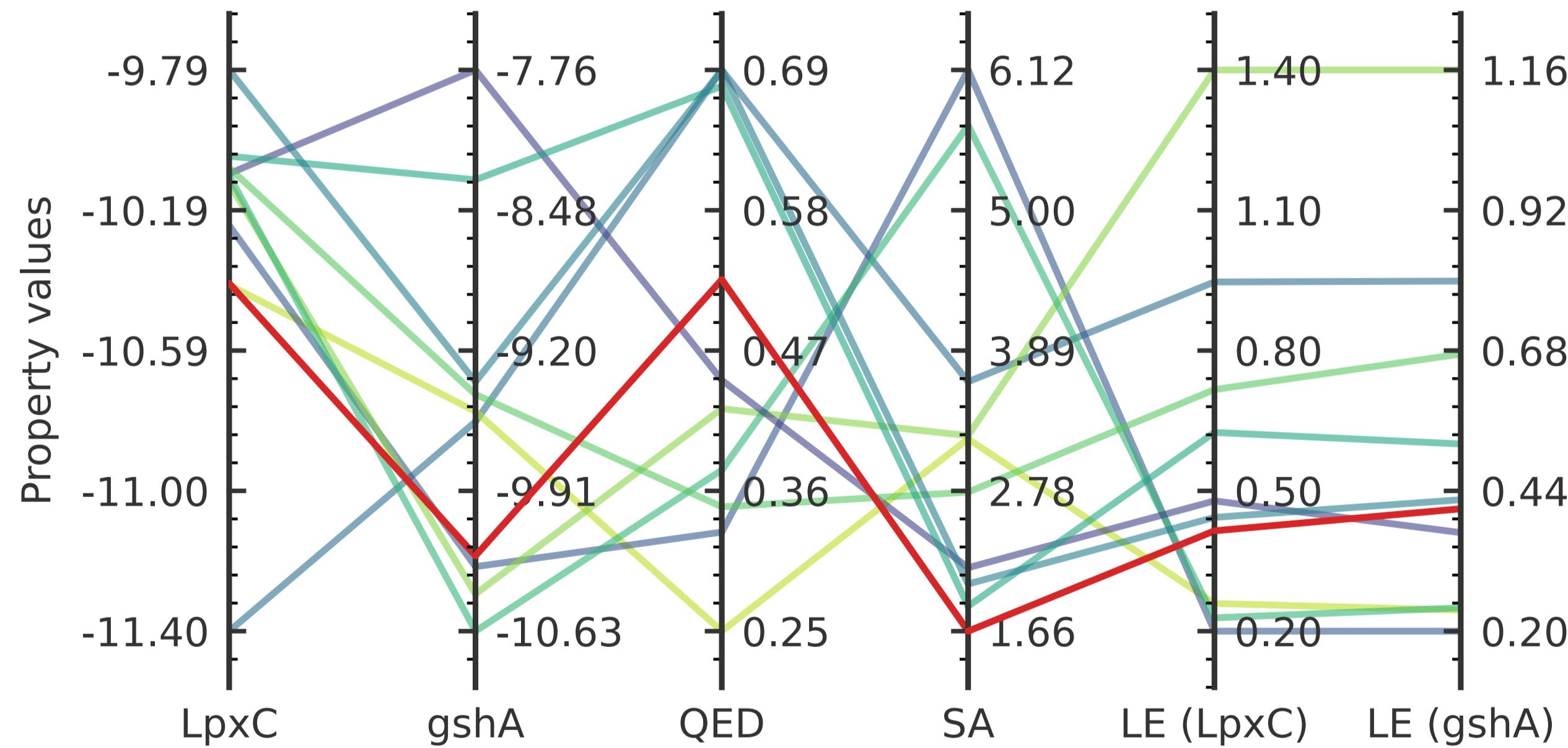


Fig. 3. Trade-offs among the first 10 unique NSGA-III Pareto-optimal solutions in the multi-target scenario. Properties shown:  $\Delta G_{bind}$  (LpxC, gshA), QED, SA, and ligand efficiency (LE) for both targets.

## Conclusions

This work demonstrates the potential of many-objective optimization to enable more selective and effective *de novo* drug discovery. Our framework successfully generates novel, Pareto-optimal molecules that simultaneously optimize up to six drug-relevant properties.



By integrating HierVAE with NSGA-III and docking based-objectives, we provide a flexible and customizable platform enabling medicinal chemists to define target-specific objectives, explore trade-offs, and prioritize candidates aligned with their discovery goals. Openly available at: [github.com/gmmbsb-lnc/docktdesign](https://github.com/gmmbsb-lnc/docktdesign).

## Methods

- **Evolutionary algorithms.** NSGA-II and NSGA-III, for multi- and many-objective optimization, respectively.
- **Generative chemistry.** Graph-based VAE pre-trained on ~1.8M ChEMBL molecules; generates 100% valid structures and enables smooth latent space exploration.
- **Generative framework.** Evolutionary search is performed in the VAE's latent space. Each latent vector is decoded into a molecular structure, which is then evaluated across multiple drug-relevant properties. The process iteratively evolves a population of candidates toward the Pareto front.
- **Selectivity scenario (ligand-based, 5 objectives).**  $\uparrow$  QED (drug-likeness),  $\downarrow$  SA (synthetic accessibility),  $\downarrow$  Cx (Böttcher molecular complexity),  $\uparrow$  Tanimoto similarity to a potent LpxC inhibitor (BDBM50074960,  $K_i = 0.053$  nM),  $\downarrow$  Tanimoto similarity to a known inhibitor of human MMP-9 (BDBM50478376,  $IC_{50} = 97$  nM; also binds LpxC).
- **Multi-target scenario (structure-based, 6 objectives).** Focused on dual activity against antibacterial targets LpxC and gshA. Binding affinities predicted via DockThor (docking) + DockTDeep (scoring) [3]. Objectives:  $\uparrow$  QED,  $\downarrow$  SA,  $\uparrow$  ligand efficiency ( $-\Delta G_{bind} / \#$  heavy atoms),  $\downarrow$   $\Delta G_{bind}$  (predicted binding affinity in kcal/mol) for both targets.

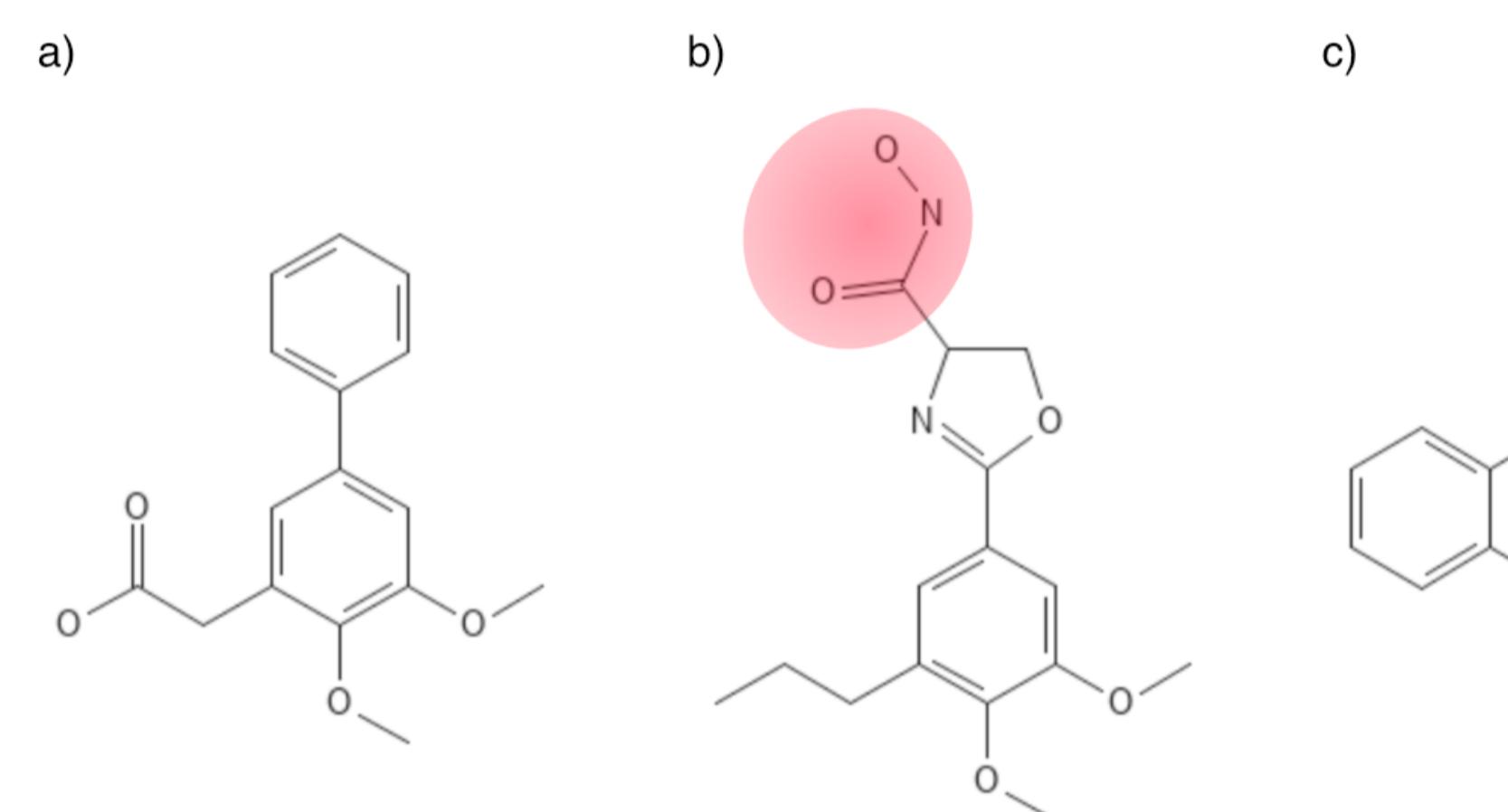


Fig. 2. (a) NSGA-III compromise solution, (b) similarity target BDBM50074960 (LpxC inhibitor), and (c) dissimilarity target BDBM50478376 (MMP-9/LpxC binder).

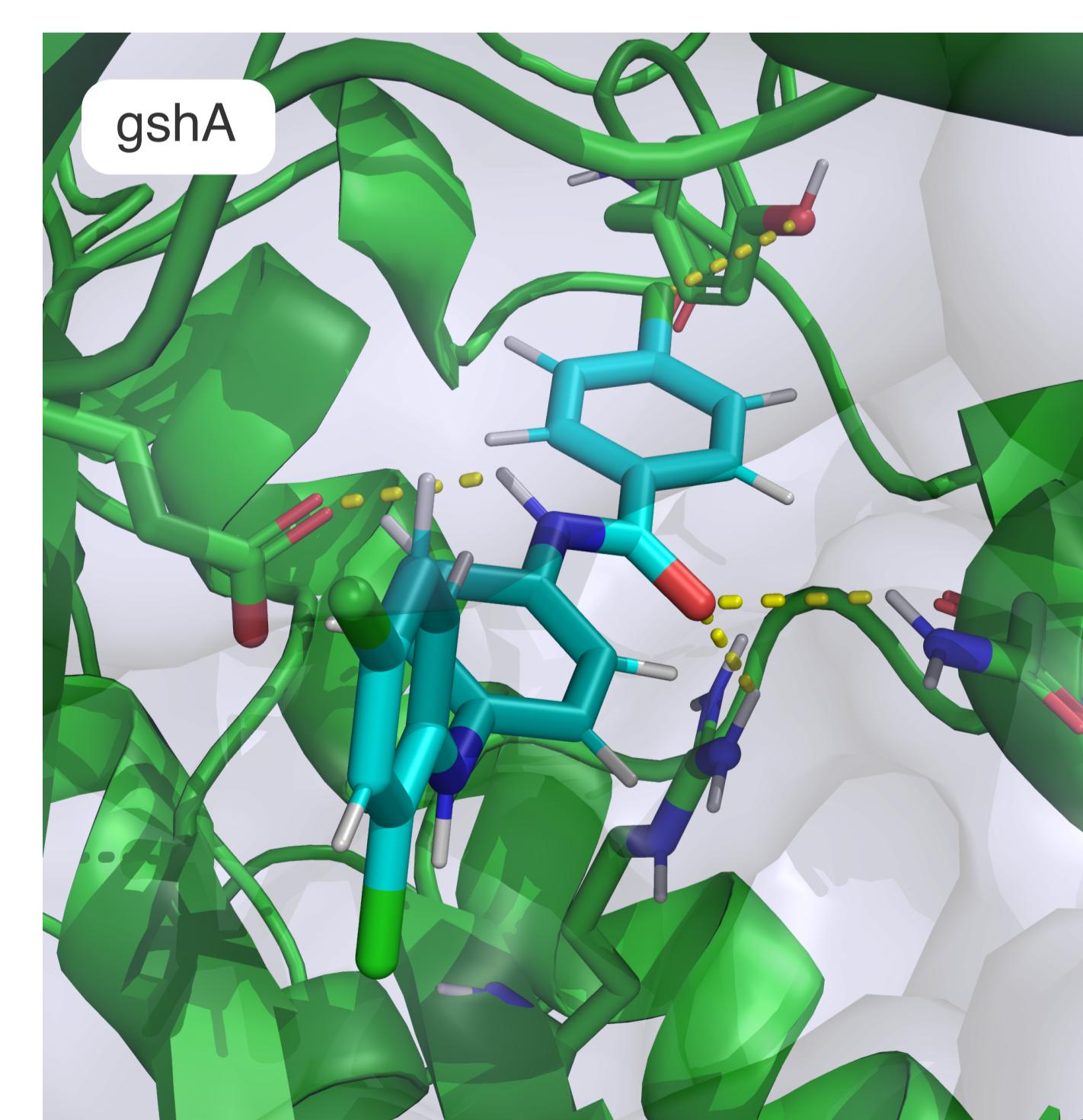
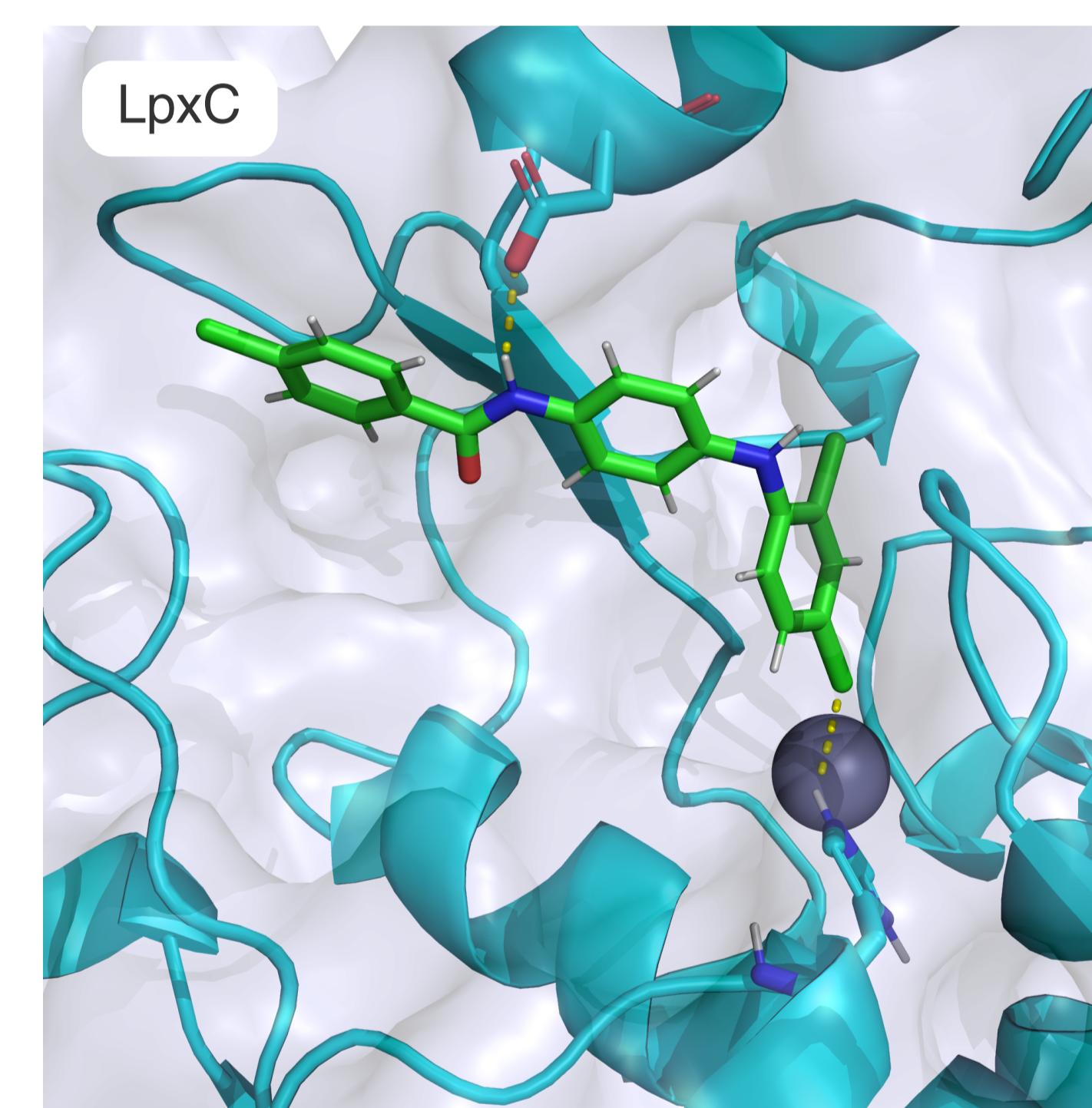


Fig. 4. Predicted binding modes of a dual-target compound (NSGA-III solution #411) with LpxC (top) and gshA (bottom) in the multi-target scenario. 3D poses (left) and 2D interaction maps (right) show favorable interactions for both targets.  $\Delta G_{bind} = -10.4$  (LpxC) /  $-10.25$  kcal/mol (gshA), QED = 0.523, SA = 1.66, LE  $\approx 0.4$  for both targets.

## References

- [1] Angelo, J. S., et al. "Multi-and many-objective optimization: present and future in *de novo* drug design." *Frontiers in Chemistry* 11 (2023): 1288626.
- [2] da Silva, Matheus M. P., et al. "A Generative Evolutionary Many-Objective Framework: A Case Study in Antimicrobial Agent Design." *Proceedings of the GECCO Companion*. 2024.
- [3] da Silva, Matheus M. P., et al. "Data-centric training enables meaningful interaction learning in protein–ligand binding affinity prediction." (2025). *ChemRxiv*.

## Funding

FAPERJ projects E-26/010.001415/2019, E-26/211.357/2021, E-26/200.608/2022, E-26/210.372/2022, and E-26/200.393/2023. CNPq project: 309744/2022-9.