



Instituto de
Inteligência Artificial



DockTDesign

Deep Generative Models for *de novo* Drug Design

Matheus M. P. da Silva

PhD in Computational Modeling
Laboratório Nacional de Computação Científica (LNCC/MCTI)

✉ matheusp@posgrad.lncc.br GitHub: github.com/mpds

November 6, 2025

Summary

1. Introduction

- 1.1 Generative modelling: variational autoencoders
- 1.2 Multi and *many*-objective optimization

2. Methodology

3. Results

4. Conclusions

Motivation

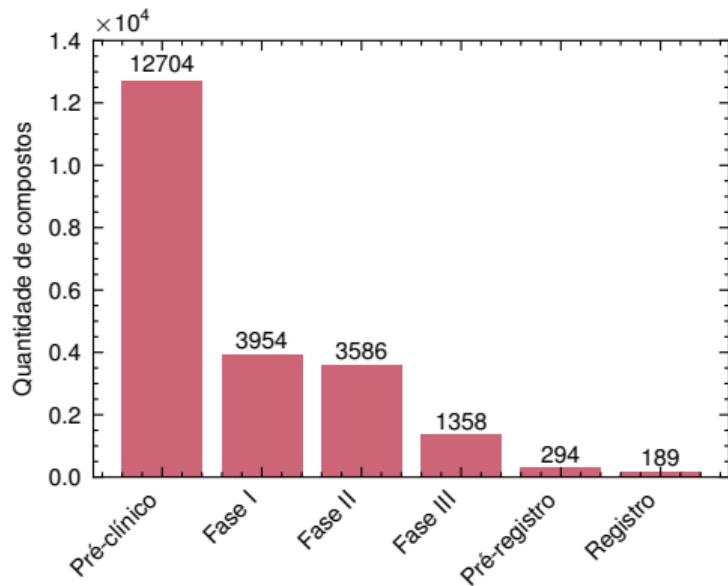


Figure: Number of compounds per development phase in 2025. Source: [Pharmaprojects](#).

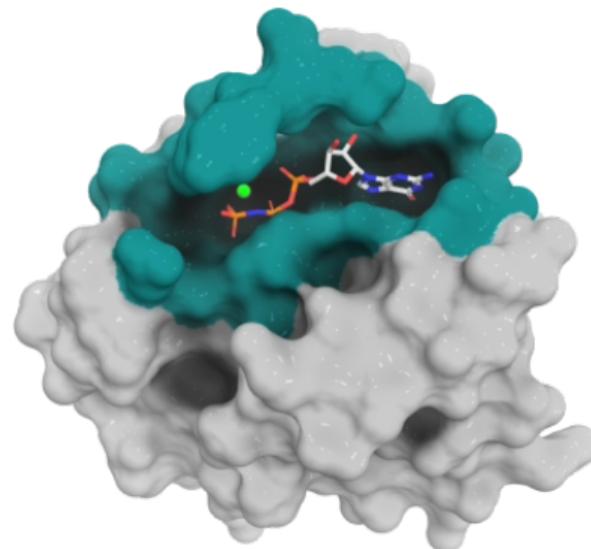
- Drug development: long, costly, and with low success rates ($\leq 5\%$ in the preclinical phase).
- Increasing the success rate in early phases can have a large economic and public-health impact.
- Computational methods are essential for the early identification of promising molecules (*hits*).

Objectives

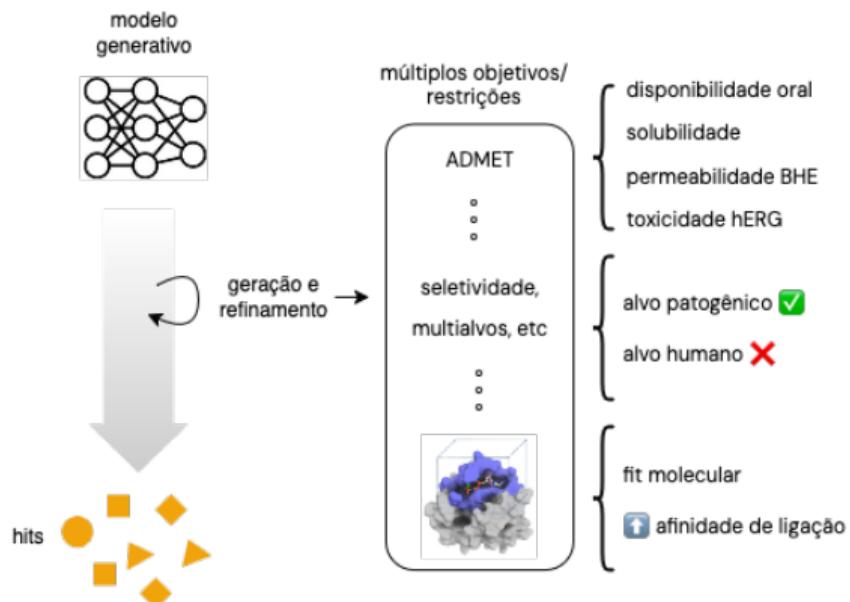
- Pose prediction
- Binding affinity prediction

Components

- Search algorithm
- Scoring function:
 - pose
 - binding affinity



- Generate molecules with desirable properties without the need for pre-defined structures.
- Essentially multi-objective problem.



Variational autoencoders (VAEs)

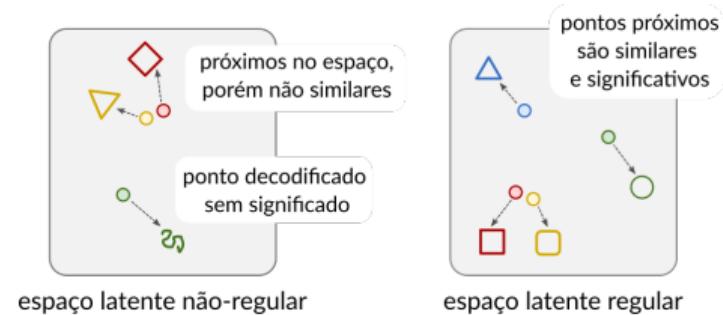
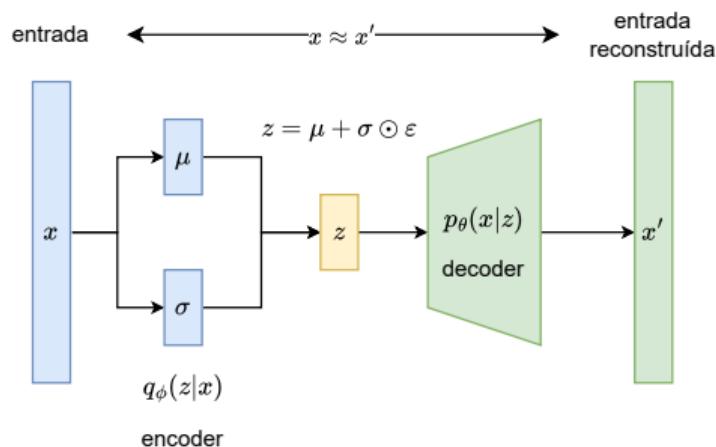


Figure: Properties of the latent space.

Figure: Variational autoencoder (VAE).

Multi-objective optimization

$$\begin{array}{ll}\text{minimize} & F(x) = (f_1(x), f_2(x), \dots, f_k(x))^T \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, 2, \dots, m \\ & h_j(x) = 0, \quad j = 1, 2, \dots, p\end{array}$$

$$f_i : \mathbb{R}^n \rightarrow \mathbb{R}$$

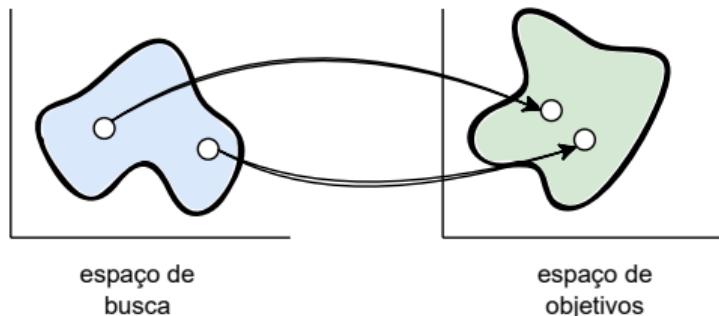


Figure: Different spaces in multi-objective optimization.

Pareto dominance: $x \succ y$ if $f_i(x) \leq f_i(y)$ for all i and $f_j(x) < f_j(y)$ for some j .

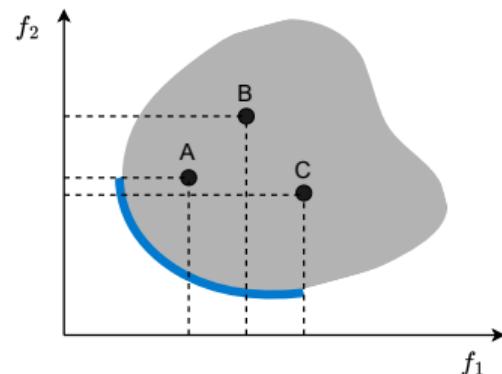


Figure: Dominance and Pareto front.

Literature review: *de novo* design

Three main approaches are employed for the *de novo* design of molecules using generative models:

- Distribution learning
- Conditional generation
- Objective-guided learning

Current limitations

Methodologies that handle $k \geq 4$ objectives are still little explored in the literature, with few studies considering more than three objectives, especially integrating appropriate techniques and generative models [?].

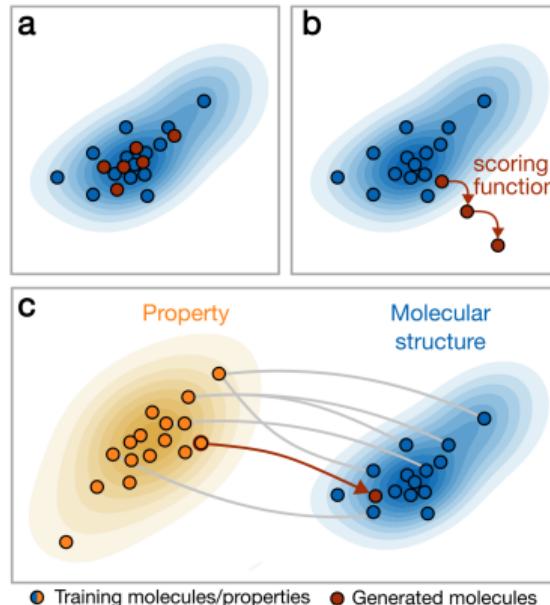


Figure: Generative approaches: (a) distribution learning, (b) objective-guided learning, and (c) conditional generation [?].

Proposed generative platform

DockTDesign

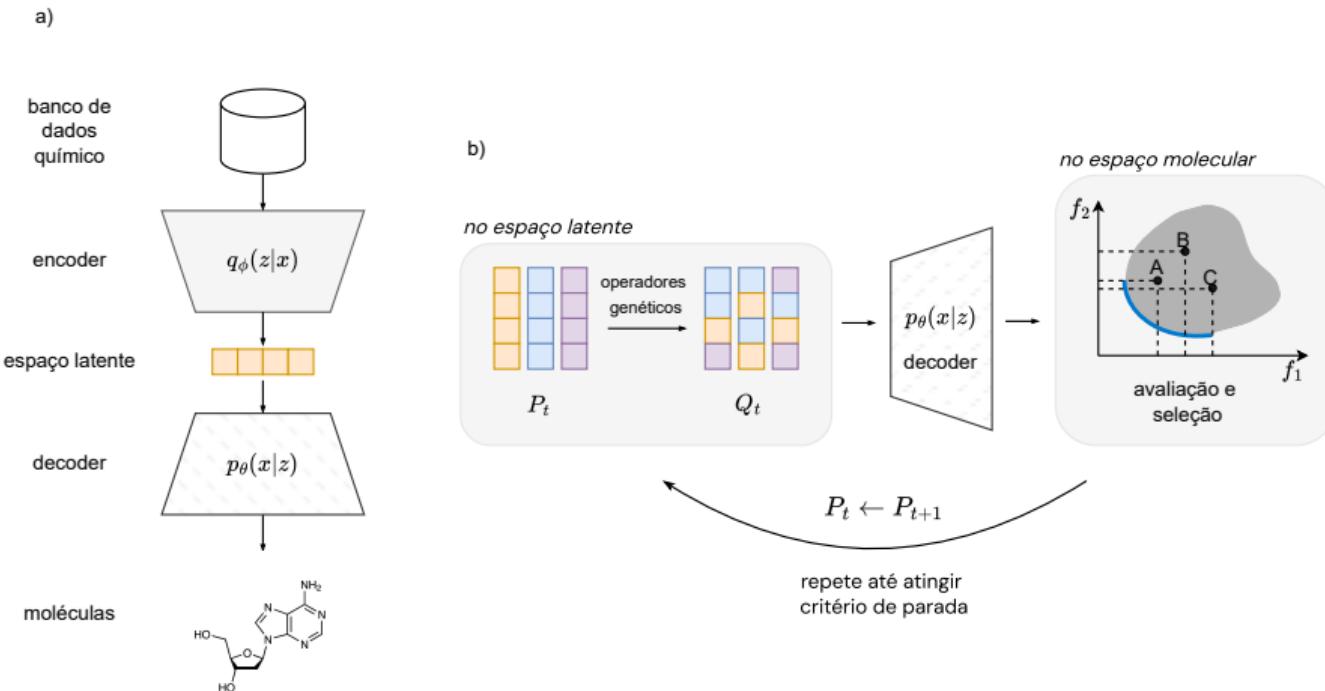


Figure: DockTDesign generative platform.

Pharmacological targets

DockTDesign

A **target** and an **anti-target** (*off-target*) were selected as a case study for generating selective inhibitors:

- LpxC^a (**target** antimicrobial involved in LPS synthesis).
- MMP-9^b (**anti-target** human important for extracellular matrix degradation).

The choice of targets was made based on the identification of a potent LpxC inhibitor that also inhibits MMP-9 (BDBM50478376).

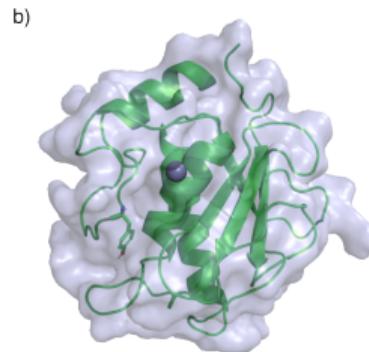
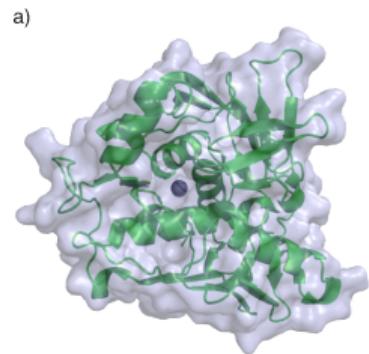


Figure: Receptors (a) LpxC and (b) MMP-9.

^aUDP-3-O-acyl-N-acetylglucosamine deacetylase

^bmatrix metalloproteinase 9

Experimental design

Algorithms: Single-objective GA, NSGA-II and NSGA-III.

Objectives:

1. maximize QED;
2. minimize SA;
3. minimize complexity;
4. maximize similarity against molecule BDBM50074960 (active LpxC);
5. minimize similarity against molecule BDBM50478376 (active LpxC, active MMP-9).

Work resulted in a publication in an international conference [?].

Results: molecular docking (protocol I)

DockTDesign

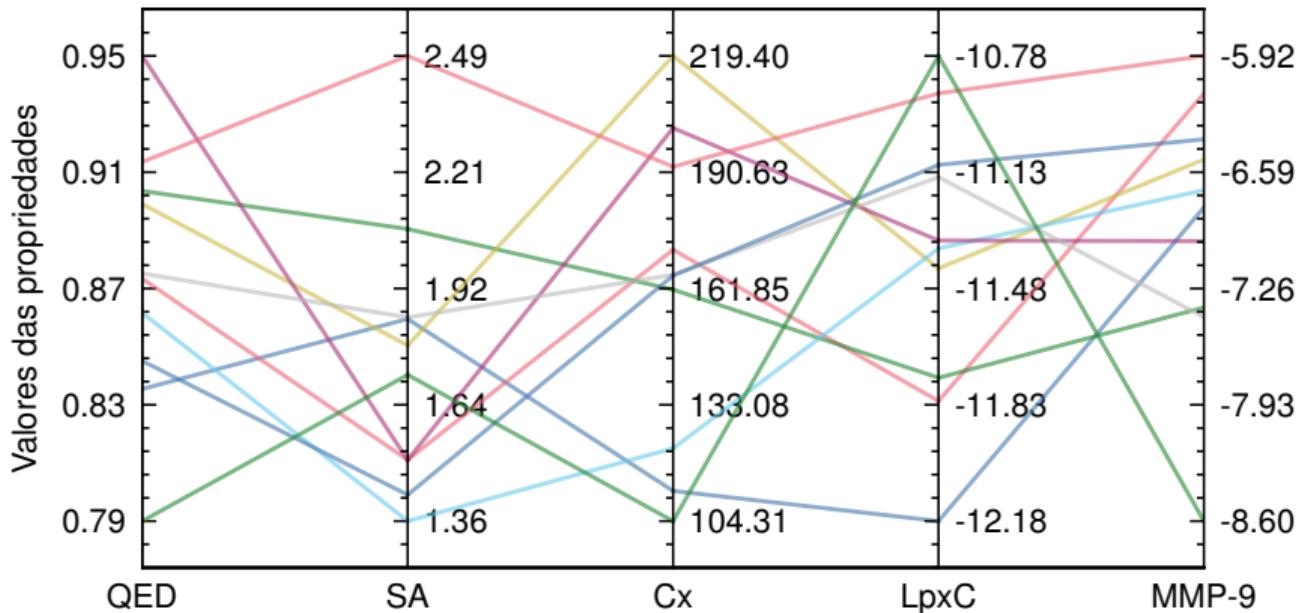
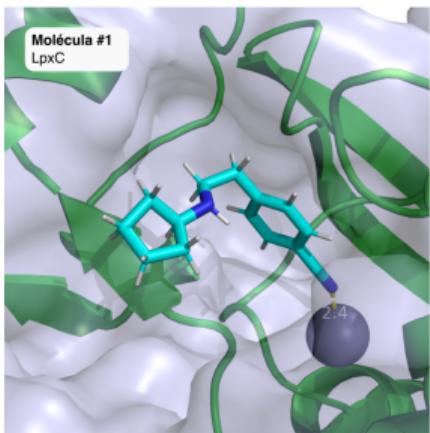


Figure: Parallel coordinates plot for the solutions obtained with protocol I.

Results: molecular docking (protocol I)

DockTDesign

a)



b)

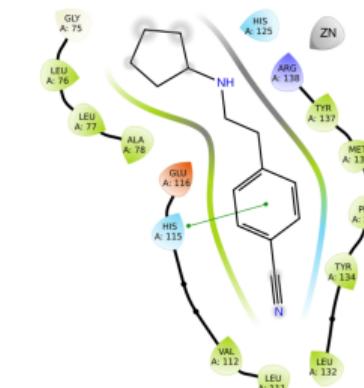
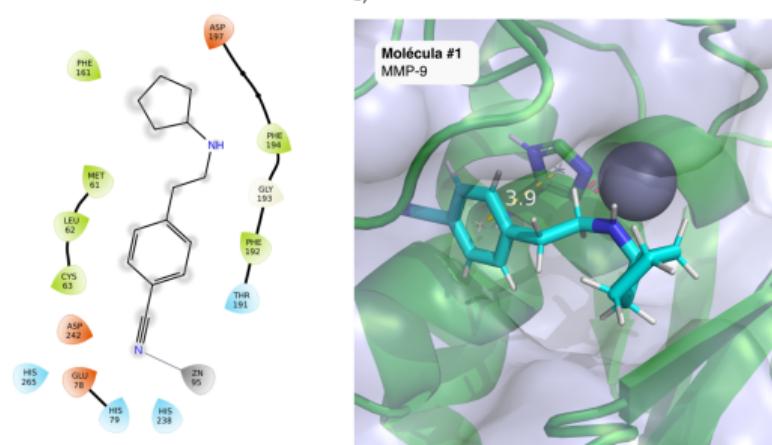


Figure: Compound 1 from protocol I in the active sites of (a) LpxC and (b) MMP-9. LpxC: -12,176 kcal/mol; MMP-9: -6,795 kcal/mol; QED: 0,836; SA: 1,85; Cx: 111,781.

- **DockTDesign**, together with **DockThor** and **DockTDeep**, presents itself as a promising and flexible platform for hit identification, capable of suggesting to the specialist a set of molecules that simultaneously meet multiple objectives.

Thank you!



*Instituto de
Inteligência Artificial*

