Universidade do Minho
Escola de Engenharia

CENTRE OF BIOLOGICAL ENGINEERING

Combining Evolutionary Algorithms with Reaction Rules Towards Focused Molecular Design

João Correia[1,2], Vítor Pereira[1,2], Miguel Rocha[1,2]

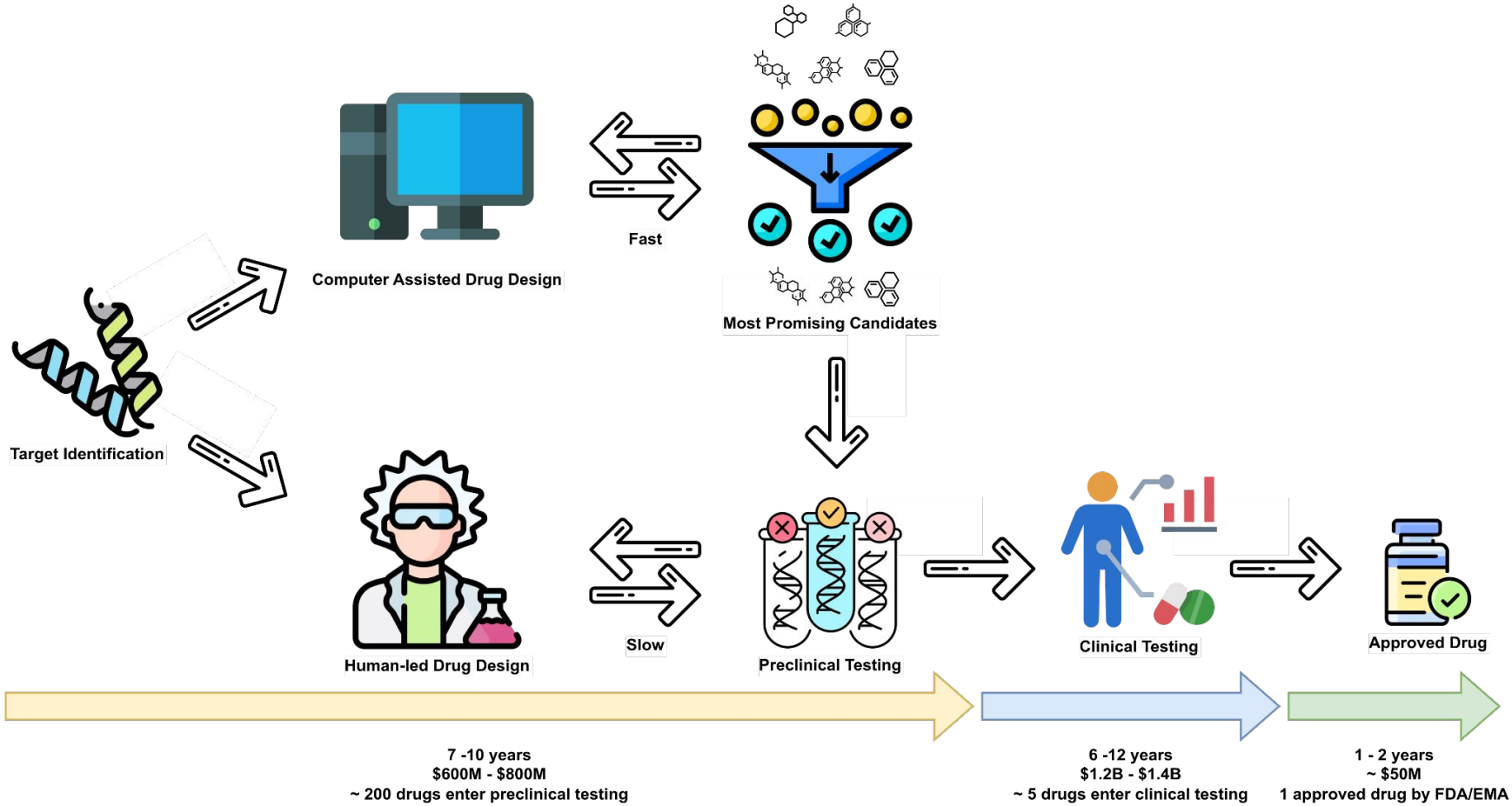
1 - CEB - Centre of Biological Engineering, University of Minho, Braga, Portugal

2 - LABBELS - Associate Laboratory, Braga/Guimarães, Portugal

jfscorreia95@gmail.com

18th July, 2023

Drug Design and Development



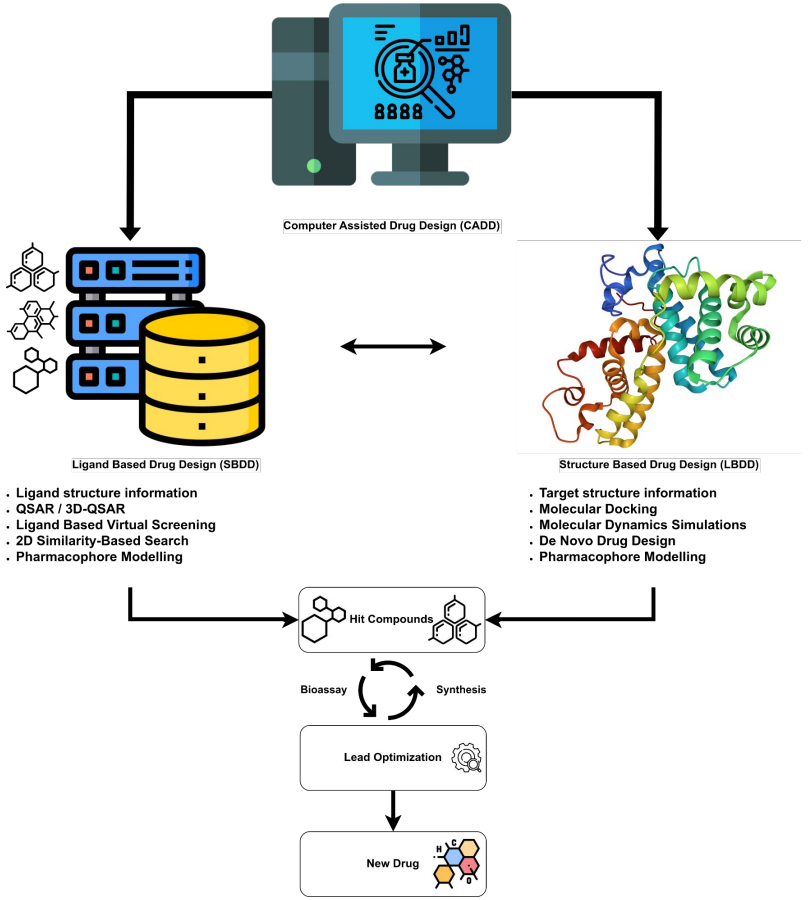
Computer Assisted Drug Design (CADD)

- **Ligand-Based Drug Design (LBDD):**

LBDD focuses on the interaction between a drug molecule and its target receptor or enzyme, using **known ligands** to design new compounds.

- **Structure-Based Drug Design (SBDD):**

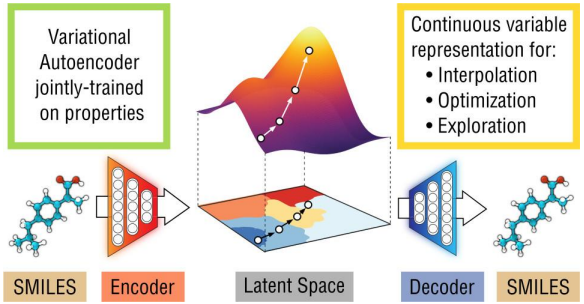
SBDD utilizes detailed **structural information of the target** receptor or enzyme to design drugs that interact optimally.



Gradient-based and Gradient-free Methods for Molecular Optimization

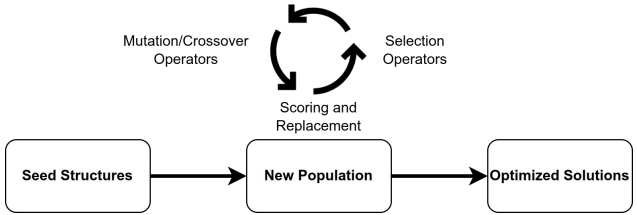
Optimization algorithms are instrumental in generating **optimal molecules** by leveraging a provided **molecular representation** and calculable **objective functions**.

- **Gradient-based Molecular Optimization - Deep Learning Models**



A diagram of an autoencoder used for molecular design. (https://pubs.acs.org/doi/10.1021/acscentsci.7b00572)

- **Gradient-free Molecular Optimization - Population-Based Stochastic Optimization Algorithms**



Overall flowchart of a genetic algorithm for molecular optimization.

Atom, Fragment and Reaction-Based Molecular Design

- Level of specificity of molecular structure generation: **atom-based, fragment-based, and reaction-based** methods.
- **Reaction-based gradient-free**: ensure **validity** and **synthetic feasibility** and provide potential **synthesis paths**.
- N**ovelty and diversity** of the generated molecules highly depend on the availability of **comprehensive reaction templates**.

Table 1: Examples of the different gradient-free atom, fragment, and reaction-based methods.

Method	Molecule construction method	Evolutionary technique
Kawai *et al.* [18]	Atom-based	Genetic Algorithm
iSyn [27]	Reaction-based	Genetic Algorithm
GB-GA [16]	Atom-based	Genetic Algorithm
MolFinder [21]	Atom-based	Conformational Space Annealing
AutoGrow4 [33]	Reaction-based	Genetic Algorithm
EvoMol [26]	Atom-based	Genetic Algorithm
LEADD [19]	Fragment-based	Genetic Algorithm
ChemGE [41]	Fragment-based	$(\mu + \lambda)$ Evolutionary Strategy
MSO [39]	Atom-based	Particle Swarm Optimization
MOARF [13]	Fragment-based	Multi-objective Evolutionary Algorithm
CReM [30]	Fragment-based	Stochastic exploration

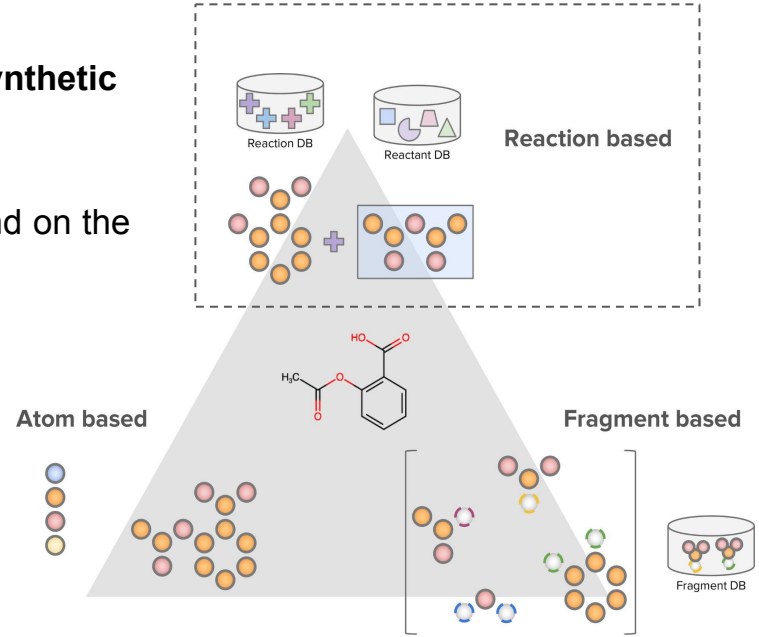


Illustration of the continuum between atom-based, fragment-based, and reaction-based molecular representation paradigms. 10.1016/j.drudis.2021.05.019

ReactEA: a Reaction-Based Gradient-Free Framework for Molecular Optimization

- Here, we introduce **ReactEA**, a new open-source evolutionary framework for computer-aided drug discovery that utilizes a comprehensive set of **22.949 biochemical reaction rules**.
- ReactEA can be used for **single or multi-objective optimization**.
- ReactEA can virtually **optimize any objective function** and **track potential synthetic routes** during the optimization process.

https://github.com/BioSystemsUM/ReactEA

Reaction Rules

- Reaction rules are **generic descriptions of reactions that encode the way reactants are converted into products**.

- A reaction rule can be applied to a compound if the compound **contains a particular substructure** that is encoded by the reaction rule.

- Reaction rules are represented as **SMARTS** (SMiles ARbitrary Target Specification) strings (https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html).

oxidation of alcohols to ketones: **[#6:1][O:2]>>[#6:1]=[O:2]**

- The compounds are represented as **SMILES** (Simplified Molecular Input Line Entry System) strings (https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html).

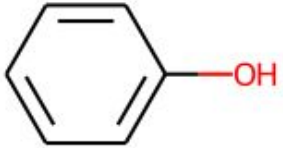
caffeine: **CN1C=NC2=C1C(=O)N(C(=O)N2C)C**

Reaction Rules

Let's look at a simple example: the oxidation of alcohols to ketones

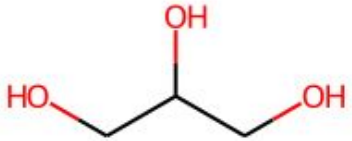
ethanol

CCO



phenol

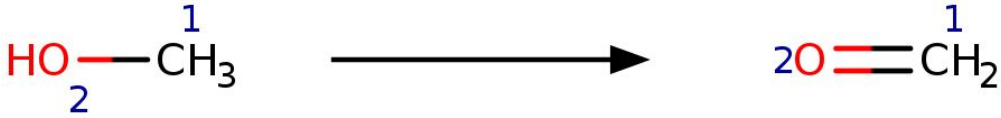
c1cccc(O)c1



glycerol

OCC(O)CO

Reaction Rule: **[#6:1][O:2]>>[#6:1]=[O:2]**



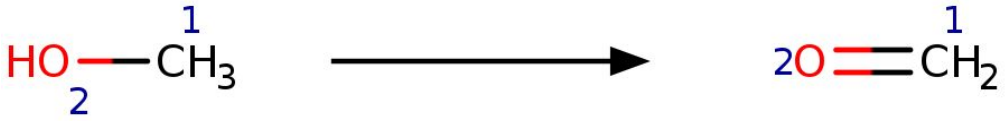
This definition says:

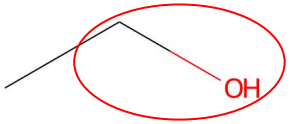
Take any carbon atom, which we'll label as atom #1, whether aliphatic or aromatic, that is **bound via a single bond to an aliphatic oxygen atom** (capital O), that we label as atom #2, and **transform the linkage between these atoms to be a double bond**.

Reaction Rules

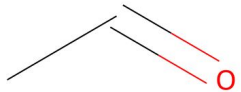
Let's apply this rule to our molecules:

Reaction Rule: **[#6:1][O:2]>>[#6:1]=[O:2]**

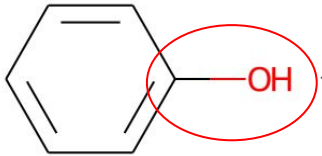




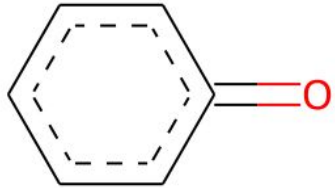
ethanol



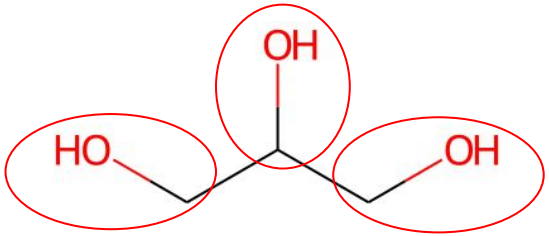
acetaldehyde



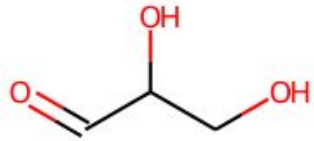
phenol



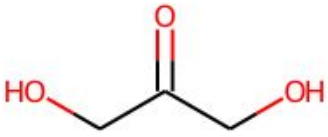
invalid: pentavalent carbon



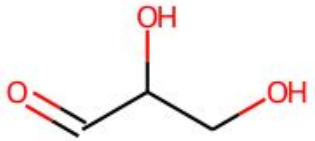
glycerol



Glyceraldehyde



Dihydroxyacetone



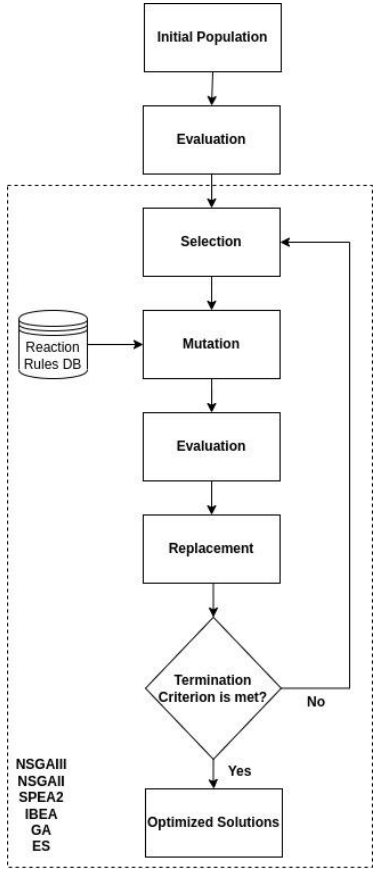
Glyceraldehyde

ReactEA: General Overview

1. **Initial population**: first set of candidate solutions (e.g. a set of **available precursors in a particular organism**, a **diverse set of molecular fragments** for a de novo design experiment, or **known ligands** for lead optimization);

2. **Evaluation:** determines the fitness of each molecule in the initial population using the defined objective functions;

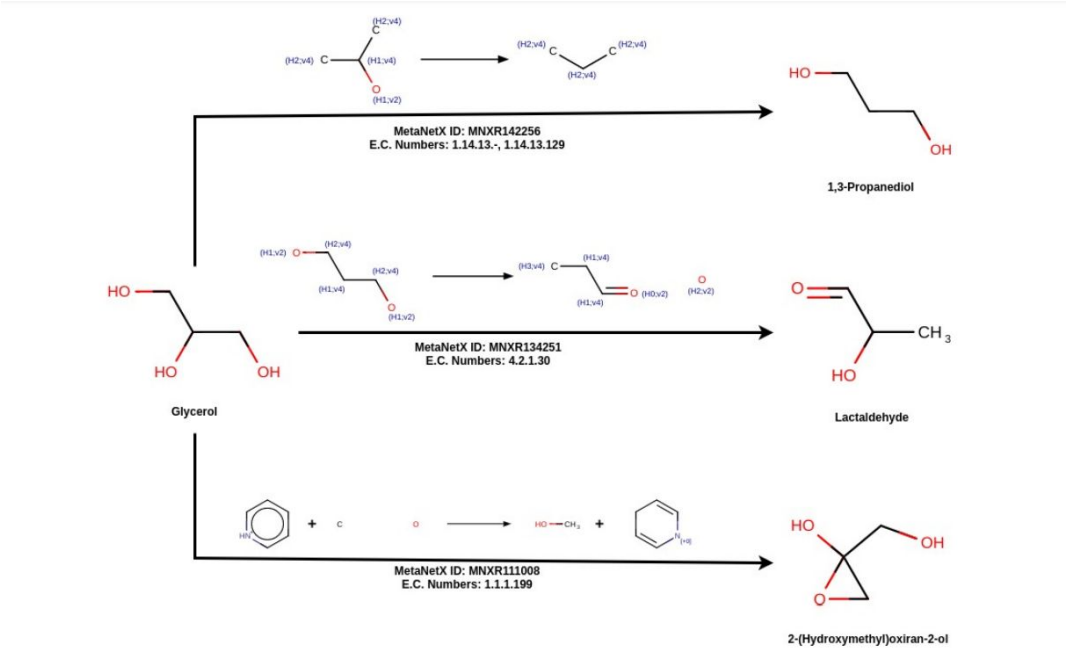
3. **Selection**: determines which solutions from the current population will be used to generate the next population;



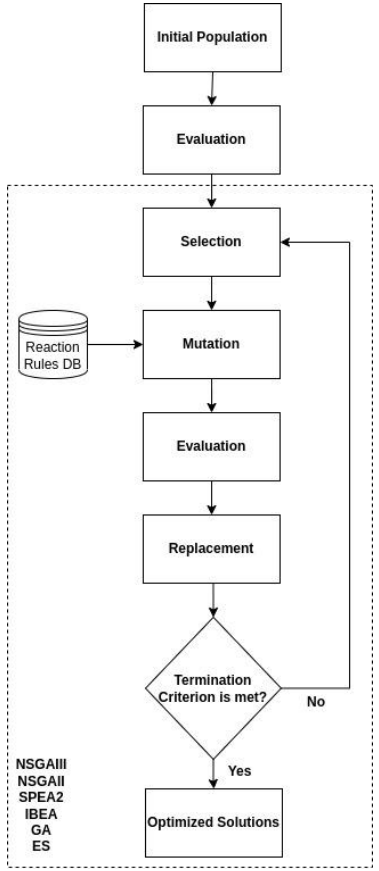
General overview of the ReactEA framework.

ReactEA: General Overview

4. **Mutation**: introduces new variations into the population. In ReactEA we apply reaction rules to the molecules (transform reactants into products);



Example of how the mutation operator works. In this example, three different reaction rules are used to generate three different products (mutants) from a single reactant (parent).



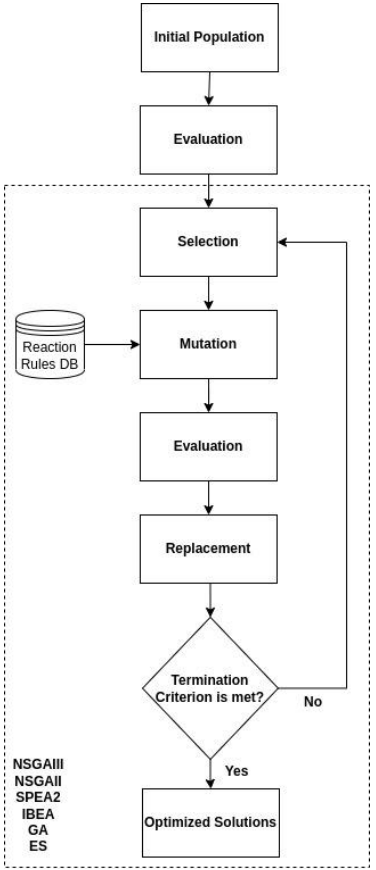
General overview of the ReactEA framework.

ReactEA: General Overview

5. **Evaluation**: determines the fitness of the newly generated molecules using the defined objective functions;

6. **Replacement**: determines which individuals from the current population will be replaced by offspring in the next generation. It plays a significant role in shaping the evolution of the population towards high-quality solutions.

7. **Termination Criterion**: determines when to stop the search process. It is essential to ensure that the algorithm has found a satisfactory solution within a reasonable time and computational budget.



General overview of the ReactEA framework.

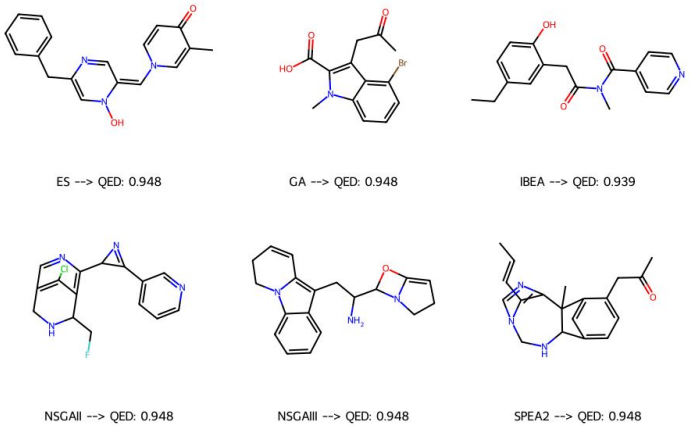
Results: Case 1 - Optimization of simple molecular properties

- Optimization of **simple objective functions** (drug-likeness (QED), synthetic accessibility score (SAS), etc.);
- Initial population: **648 precursors from the Escherichia coli iJO1366** metabolic model;
- **100 generations** or until no improvements were observed for 5 generations;
- The results showed **similar performance to state-of-the-art (SOTA) methods** under comparable conditions.

QED:

SOTA:
0.954

Ours:
0.948

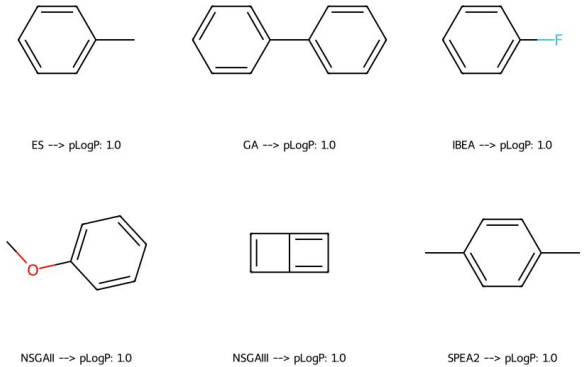


Best molecules, QED optimization.

SAS:

SOTA:
0.95

Ours:
1.0



Best molecules, SAS optimization.

Results: Case 2 - Similarity to Aspartame

- Main objectives:
  - Assess the ability of ReactEA in **finding paths to target molecules**;
  - Understand the **impact of the initial population** on the performance of ReactEA;

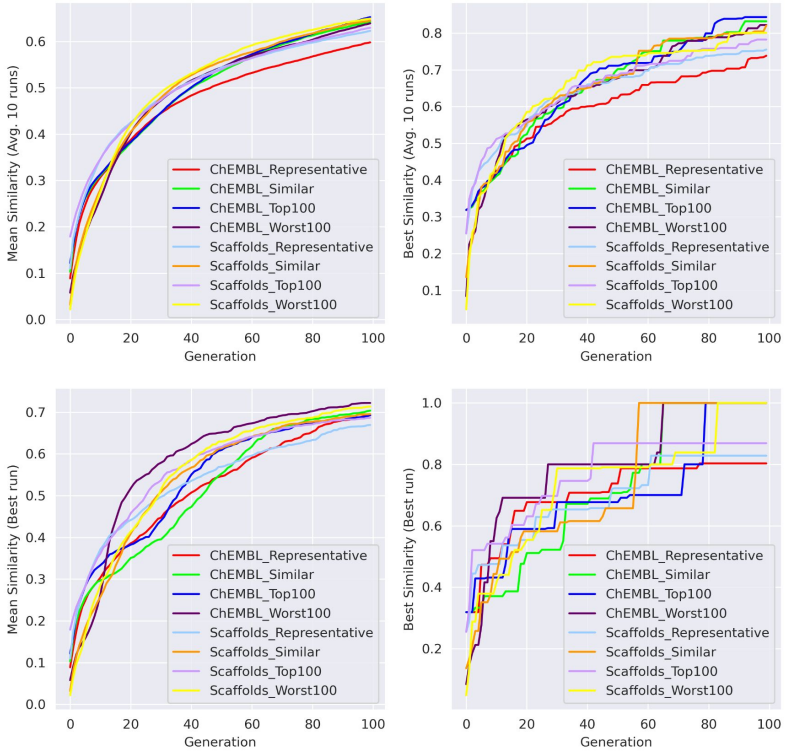
- We tested **8 different initial populations**, each consisting of 100 molecules, including **natural product (NP)-based approved drugs from ChEMBL** and **scaffolds from known NPs**;

- We were able to r**each the Aspartame** molecule in **5 out of the 8** initial populations.

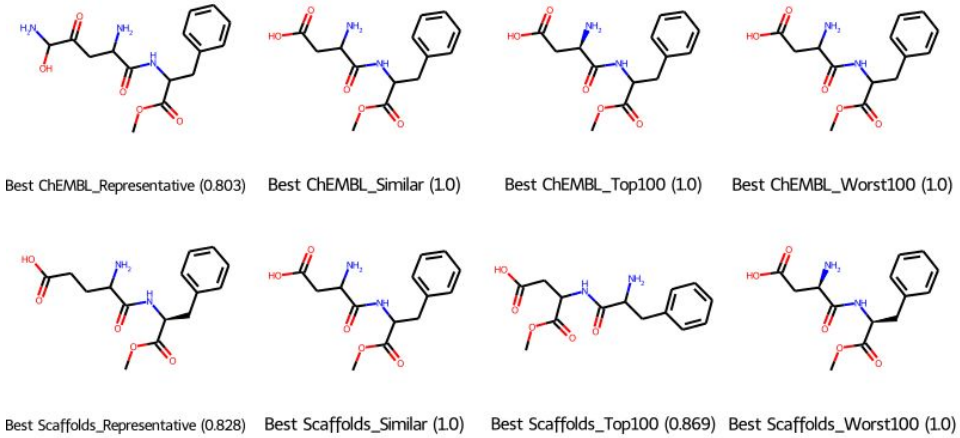
Table 2: Impact of different initial populations on the performance of ReactEA.

Set	Unique/Novel	Internal Sim.	Mean Sim. to Aspartame	Best Sim. to Aspartame	QED	SAS
ChEMBL Representative	100%	0.457	0.598	0.803	0.436	3.030
ChEMBL Similar	100%	0.500	0.642	1.000	0.452	2.953
ChEMBL Top100	100%	0.510	0.653	1.000	0.447	2.971
ChEMBL Worst100	100%	0.506	0.639	1.000	0.447	2.971
Scaffolds Representative	100%	0.487	0.623	0.828	0.445	3.004
Scaffolds Similar	100%	0.508	0.645	1.000	0.428	2.973
Scaffolds Top100	100%	0.476	0.630	0.869	0.450	2.991
Scaffolds Worst100	100%	0.519	0.649	1.000	0.447	3.015

Results: Case 2 - Similarity to Aspartame



Evolution of solutions across generations for the different initial populations. Top left represents the average similarity to aspartame of all solutions (average 10 runs). Top right the best solution (average 10 runs). Bottom left the mean similarity of the best run only. Bottom right the best solution of the best run.



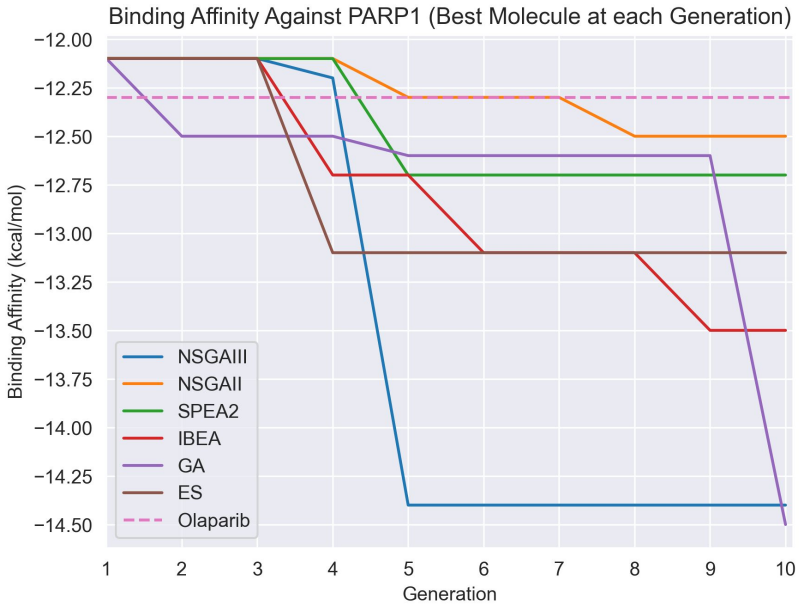
Best molecule according to its similarity to aspartame generated by each EA.

Results: Case 3 - Docking to PARP1

- To show how ReactEA can be used for **lead optimization**, we used a library of 94 seed molecules, including 11 **known PARP inhibitors** and 83 molecular fragments;
- We used the **DOCKSTRING** package (AutoDock Vina wrapper) to calculate the binding affinity against the PARP1 enzyme.

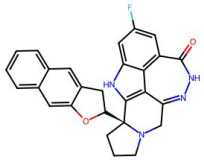
Table 3: Results of the Docking to PARP1 optimization.

EA	Worst	Best	Mean	Std. Dev.
Init. Pop.	50.0000	-12.1000	35.277660	25.8735
NSGAIII	-11.1000	-14.4000	-11.769149	0.5998
NSGAII	-10.6000	-12.5000	-11.198936	0.4759
SPEA2	-10.7000	-12.7000	-11.427660	0.5284
IBEA	-10.9000	-13.5000	-11.547872	0.5303
GA	-10.7000	-14.5000	-11.378723	0.6341
ES	-11.5000	-13.1000	-11.954255	0.3935

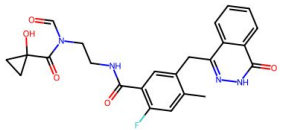


Binding affinity of the best molecule at each generation for the different EAs. The best inhibitor in the initial population, Olaparib, is shown for reference.

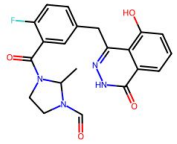
Results: Case 3 - Docking to PARP1



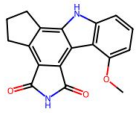
NSGAIII (Binding Affinity: -14.4)



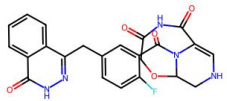
NSGAII (Binding Affinity: -12.5)



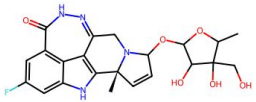
SPEA2 (Binding Affinity: -12.7)



IBEA (Binding Affinity: -13.5)

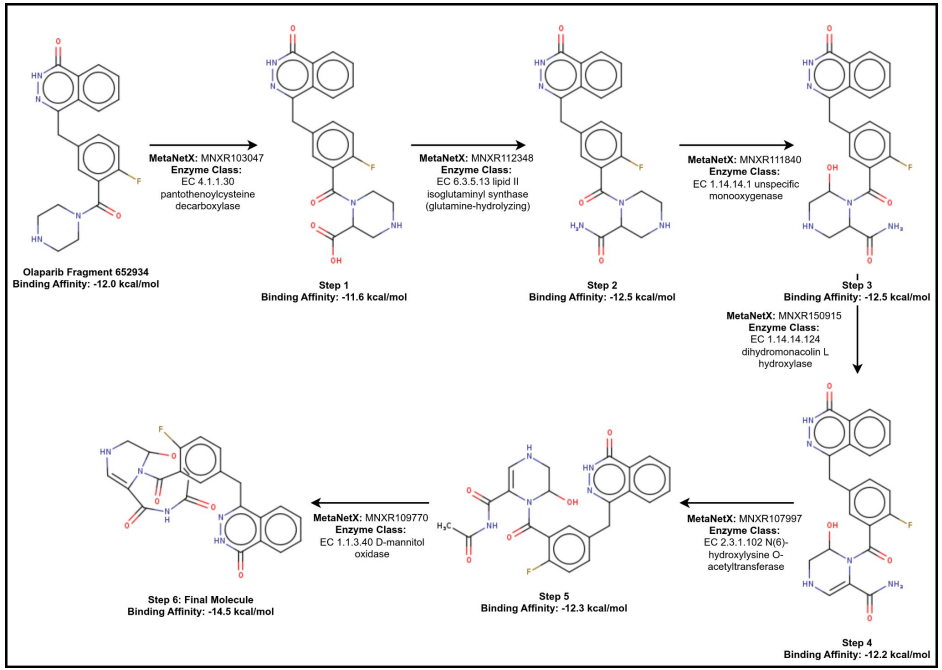


GA (Binding Affinity: -14.5)



ES (Binding Affinity: -13.1)

Best molecule and respective binding affinity of the best generated molecules by each EA.



Molecule with the best binding affinity and the reaction templates that were applied from the starting compound (Olaparib Fragment 652934) until the final molecule.

Conclusions

- We have introduced **ReactEA**, a new open-source reaction-based **EA framework for molecule generation**;
- ReactEA can be used in **single and multi-objective problems**, optimizing **any (set of) measurable objective functions** with its **large set of biochemical reaction rules**;
- The use of reaction templates allows for the generation of **valid structures** and assures **chemical feasibility**, with potential **synthesis paths** provided for the generated molecules;
- Our results demonstrate ReactEA's **high configurability and versatility**, achieving excellent results in optimizing **simple objectives like QED or SAS**, as well as more **complex tasks like similarity to a target molecule and docking to proteins**;
- One of the limitations of ReactEA and other reaction-based approaches is their dependence on the **available reaction rules** and their **coverage of the chemical space**;
- Additionally, the reaction rules are limited to **human knowledge**, which may limit the exploration of **new areas of the chemical space**.

Acknowledgements

Ph.D. Advisors: Vítor Pereira and Miguel Rocha

BiSBi Group - University of Minho

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Questions?

João Correia[1,2], Vítor Pereira[1,2], Miguel Rocha[1,2]

1 - CEB - Centre of Biological Engineering, University of Minho, Braga, Portugal

2 - LABBELS - Associate Laboratory, Braga/Guimarães, Portugal

jfscorreia95@gmail.com

18th July, 2023