# GALAPAGOS: Fragment-based Evolutionary Algorithm for Simultaneous Optimization of Drug-likeness and Affinity

**Hokyun Jeon[1], Hui Sun Lee[1], Insuk Joung[1]**

[1]Standigm Inc., Seoul, Republic of Korea

## Background

- In the hit and lead optimization process, advanced computational methods contribute to designing molecules with desired pharmaceutical properties.
- However, it is still challenging to generate the molecule that satisfies multiple properties in broad chemical space avoiding being trapped in local minima.

## Purpose

- We propose a seamless workflow for molecular optimization using a genetic algorithm consisting of a fragment-based next generation scaling up methods.
- We aimed to find molecules with optimized properties for multi-objective functions, and simultaneously, maximizing the efficiency of chemical space exploration to reach the global optima.

## Methods

### GALAPAGOS workflow

- The algorithm has a strategy to extensive search for optimized molecules in multiple distinct local minima **(Figure 1)**.

***Initialize***

- The initial pool, which is user defined data source, is divided into each seed pool and auxiliary pool. The seed pool is the ancestors for offspring pool.
- Clustering-based selection guarantees diversity of the seed pool.

***Generation***

- Crossover: exchange the sidechain fragment between the molecules with maximum common structure.
- Mutation: Moiety-based Neural Networks (MONET), which is another asset of the Standigm inc., modifies a randomly selected sidechain fragment.

***Selection***

- Objective function
  - Linear combination of multiple pharmaceutical property weighted scores.

$$S_{obj} = w_g S_{gold} + w_r S_{ROCS} - w_p(LogP + SA\ score + Ring\ penalty) + w_q QED$$

- The distance cutoff of the clustering decrease by episode for extensive maturation.

$$D_e = 0.97 D_{e-1}$$

- Selection rule facilitate to avoid the chemical space area of molecules from the previous round.

### Validation

- To demonstrate the feasibility of the algorithm, we conducted a pilot study to optimize the virtual hit molecules of Adenosine receptor A2 (AA2AR) target from the ChEMBL26 database.
- The initial pool was defined as the top 10,000 of ligands in the database ranked by Tanimoto similarity with crystal ligand (< 0.323).
- The crystal ligand and active molecules in DUD-E datasets are compared to the final generated pool of the GALAPAGOS.

## GALAPAGOS workflow

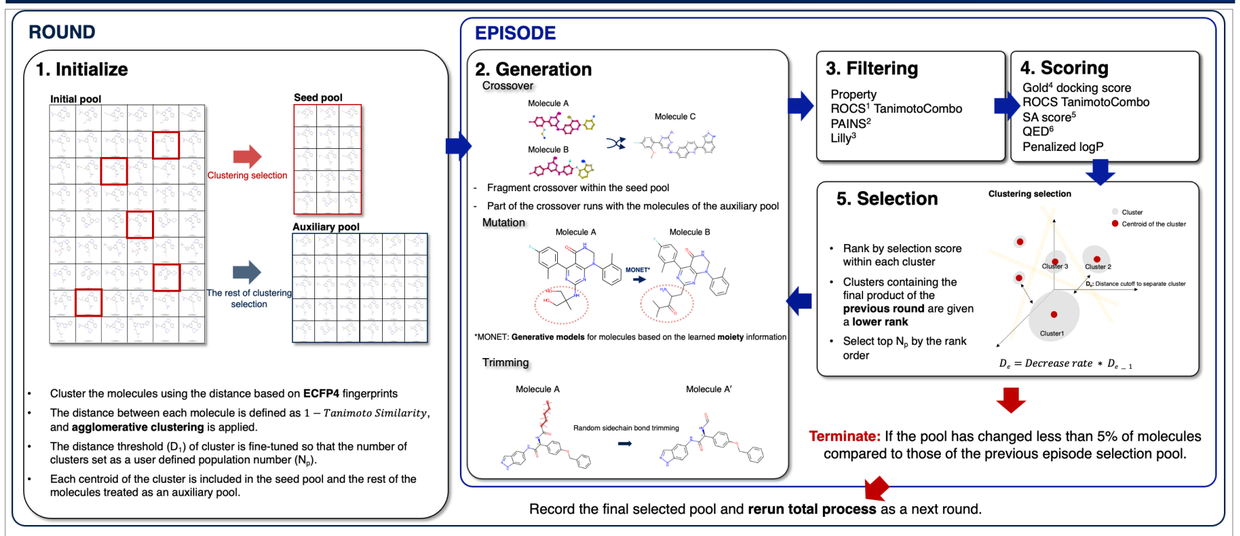


**Figure 1.** Overall workflow for hit and lead optimization in the GALAPAGOS.

## Result

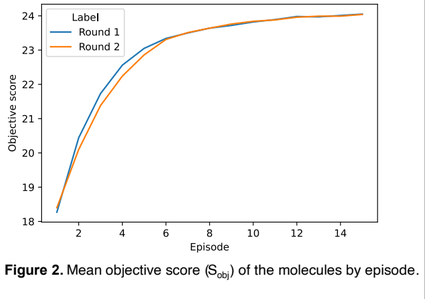


**Figure 2.** Mean objective score ($S_{obj}$) of the molecules by episode.

- The mean selection score of the molecules in the pool increases by episode as a saturation curve **(Figure 2)**.
- In round 1 and round 2, different chemical spaces are explored, but the increasing trend of the score is similar.

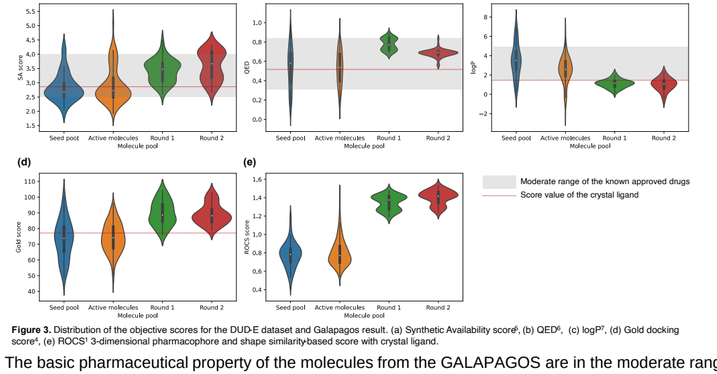


**Figure 3.** Distribution of the objective scores for the DUD-E dataset and Galapagos result. (a) Synthetic Availability score[5], (b) QED[6], (c) logP[7], (d) Gold docking score[4], (e) ROCS[1] 3-dimensional pharmacophore and shape similarity-based score with crystal ligand.

- The basic pharmaceutical property of the molecules from the GALAPAGOS are in the moderate range of the approved oral drugs or got higher scores than seed pool molecules **(Figure 3a, 3b and 3c)**.
- The activity scores, such as gold docking score and ROCS score, of the GALAPAGOS pool overwhelmed the seed pool and active molecules, which is derived from the ChEMBL dataset or DUD-E dataset respectively **(Figure 3d and 3e).**

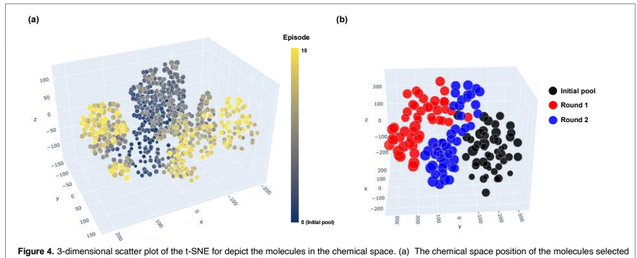


**Figure 4.** 3-dimensional scatter plot of the t-SNE for depict the molecules in the chemical space. (a) The chemical space position of the molecules selected by episode during the round 1, (b) The position of the final selected pool in the chemical space by each round.

- The GALAPAGOS algorithm explored the chemical space extensively by episode diverging from the seed pool **(Figure 4a)**.
- In each new round, it was observed that a chemical space exploration was made in another area each other **(Figure 4b)**.

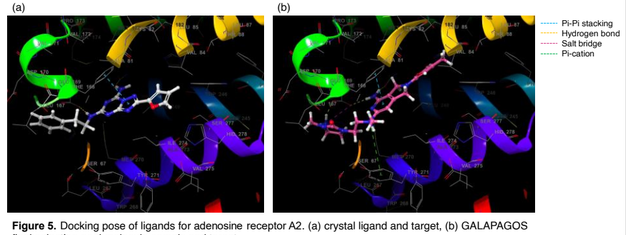


**Figure 5.** Docking pose of ligands for adenosine receptor A2. (a) crystal ligand and target, (b) GALAPAGOS final selection pool molecule sample and target.

- Among the final selected pool from the GALAPAGOS, we sampled randomly one ligand and confirmed docking pose state.
- All interactions of the crystal ligand were also present for the sampled molecule.
- Additional favorable bonds are observed, which increases the affinity of the ligand to target protein **(Figure 5)**.

## Conclusion

- We demonstrated a practical algorithm for utilization in the hit and lead optimization process in the drug development industry.
- We showed a feasibility of GALAPAGOS algorithm in multiple objective optimizing process considering the both of drug-likeness and drug activity.
- By developing a workflow for avoiding being trapped in a local minimum and conducting extensive exploration in the chemical space, it will be facilitated to generate the qualified and scalable drug candidates.

References:1ROCS version 3.5.0. OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com. 2. Baell, Jonathan B., and Georgina A. Holloway. "New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays." Journal of medicinal chemistry 53.7 (2010): 2719-2740. 3. Bruns, Robert F., and Ian A. Watson. "Rules for identifying potentially reactive or promiscuous compounds." Journal of medicinal chemistry 55.22 (2012): 9763-9772. 4. Development and Validation of a Genetic Algorithm for Flexible Docking G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, J. Mol. Biol., 267, 727-748, 1997. 5. Ertl, Peter, and Ansgar Schuffenhauer. "Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions." Journal of cheminformatics 1.1 (2009): 1-11. 6. Bickerton, G. Richard, et al. "Quantifying the chemical beauty of drugs." Nature chemistry 4.2 (2012): 90-98. 7. Wildman, Scott A., and Gordon M. Crippen. "Prediction of physicochemical parameters by atomic contributions." Journal of chemical information and computer sciences 39.5 (1999): 868-873. Conflict of interest: None/ Contact: hokyun.jeon@standigm.com