Standigm
AI for new drug R&D

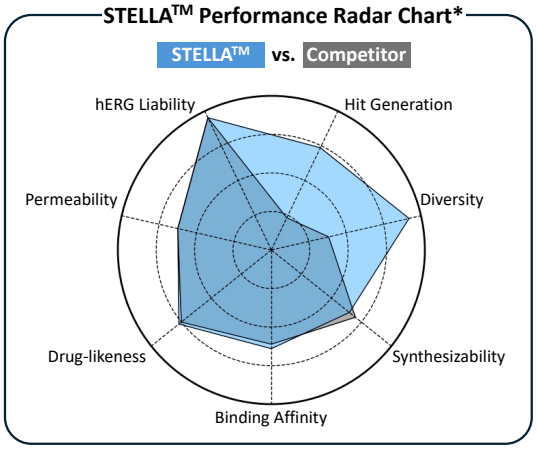
STELLA™

Revolutionizing Drug Discovery with AI-Powered De Novo Molecular Design

Executive Summary

STELLA™ is a next-generation **AI platform for de novo molecular design** that delivers significant business and scientific benefits to drug discovery teams. By optimizing multiple drug-like properties simultaneously and intelligently exploring a broader chemical space, STELLA™ helps pharmaceutical organizations achieve a higher return on investment (ROI) through faster time-to-market, improved hit rates, and the discovery of truly novel compounds. The platform's AI-driven approach consistently yields more high-quality leads in less time, reducing R&D costs and accelerating the path to clinical candidates. Key features and benefits are:

- **Multi-Parameter Optimization**:
  - Optimize against over 20 pharmacological properties in parallel using a customizable scoring function.
  - Integrates advanced predictive models (QIP-ADMET™, SPICA™, GT-AutoML™) to ensure each designed molecule meets complex drug-likeness criteria (potency, ADMET, selectivity, etc.).



- **Hit Expansion**:
  - Achieve ~3.2x more hit compounds and ~2.4x more unique scaffolds than conventional methods, leading to roughly 3x higher hit rates in benchmarks.
- **Proven Novelty**:
  - Consistently generate structurally novel compounds that strengthen IP positioning.
  - Designs exhibit greater scaffold diversity, yielding more patentable candidates and a broader foundation for pipeline expansion.

Introduction

Small-molecule drug discovery faces a fundamental challenge: the immense, near-infinite chemical space. The number of theoretically synthesizable organic compounds is estimated to be on the order of 10^{60}, making it infeasible for traditional experimental screening methods to comprehensively explore this vast space. Furthermore, identifying drug candidates requires satisfying numerous pharmacological properties—many of which are inherently in conflict with each other.

To address these challenges, we present STELLA™, a cutting-edge AI platform designed to rapidly generate molecules with optimized pharmacological profiles. Evaluation results demonstrate STELLA™'s superior performance in both efficient exploration of chemical space and multi-parameter optimization. From concept to market-ready design, STELLA™ facilitates hit-to-lead optimization, ensuring swift progression through the drug discovery pipeline and transforming a traditionally complex, costly, and time-consuming process.

Overview of STELLA™

STELLA™ integrates advanced algorithms and user-centric design to streamline de novo drug discovery. The platform's capabilities span from generative chemistry to predictive modeling and workflow integration.

Advanced Molecular Generation

Figure 1 illustrates an overall workflow of STELLA™. STELLA™ adopts a powerful genetic algorithm to enhance fragment-level chemical space exploration for molecular generation. Given a seed molecule, STELLA™ begins with the *Initialization*, where an initial pool is generated by molecular mutation using FRAGRANCE, our proprietary computational tool for modifying a molecule by substituting its fragments with alternatives from a fragment database. A user-defined pool of molecules can optionally be added to the initial pool. Clustering is conducted to produce a specified number of clusters. The centroid molecules of each cluster are selected as the initial pool, and remaining molecules are assigned to an auxiliary pool. Variants of the molecules in the initial pool are then generated by the FRAGRANCE mutation, maximum common substructure (MCS)-based crossover, and trimming (*Molecule Generation*). Each generated molecule is scored using an objective function (*Scoring*). All generated molecules, along with the initial pool are clustered with a distance cutoff, and molecules with the highest objective score are selected from each cluster (*Clustering-based Selection*). The *Molecule Generation*, *Scoring*, and *Clustering-based Selection* are repeated until any termination condition is met. This molecular

generation workflow enables STELLA™ to efficiently explore a broad chemical space in a reasonable computational cost.

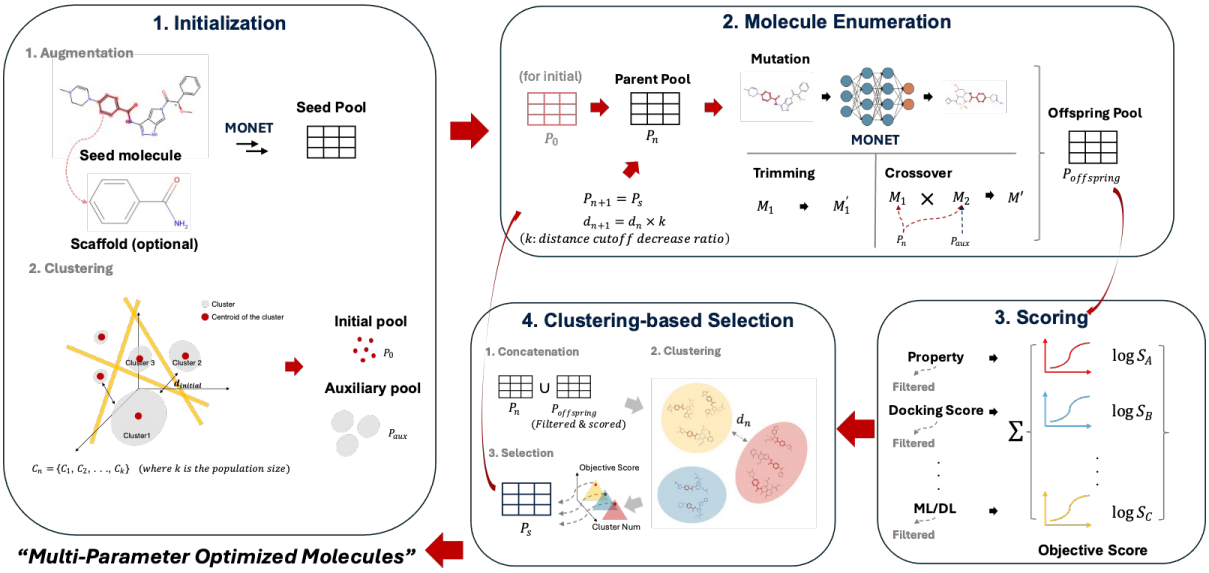


Figure 1. The overall workflow of STELLA™.

Powerful Multi-Parameter Optimization

In STELLA™, we utilize the principles of conformational space annealing (CSA), a highly effective global optimization algorithm, not only to enhance search efficiency but also to maintain structural diversity during chemical space exploration. CSA is a metaheuristic algorithm integrating the strengths of genetic algorithms and simulated annealing to balance exploration and exploitation during the optimization process. During the iterations of *Molecule Generation*, *Scoring*, and *Clustering-based Selection* in STELLA™, the optimization focus gradually shifts from global exploration to local exploitation (Figure 2), thereby enhancing the likelihood of successfully identifying the global optimum.

Conventional CSA employs an individual molecule-based approach, where a generated molecule replaces either the closest molecule with worse objective score (if the distance > cutoff) or the molecule with the worst objective score (if the distance > cutoff). On the other hand, STELLA™ adopts clustering-based CSA, where newly generated pool together with the parent pool are grouped into clusters, and the top-scoring candidates are selected from each cluster. This approach further improves global optimization by efficiently redirecting exploration to other areas when the algorithm struggles to identify promising candidates in a specific local region.

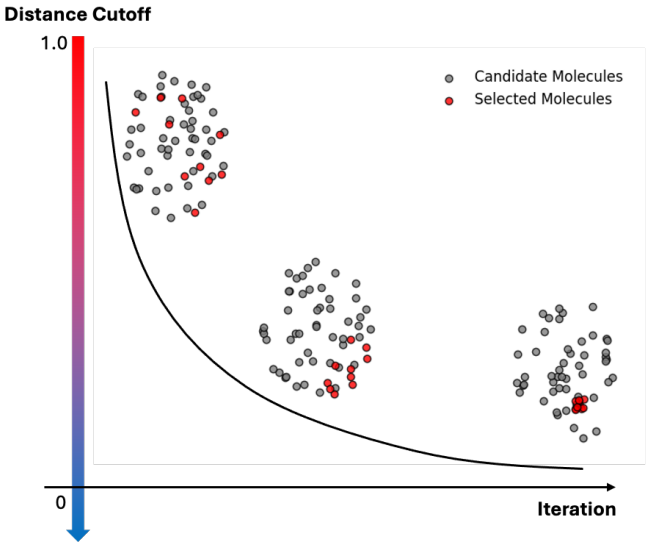


Figure 2. Visualization of molecule selection over iterations in STELLA™.

AI-Based Accurate Pharmacological Property Prediction

Molecules generated by STELLA™ are evaluated using an objective function that consists of a set of pharmacological properties of interest. In addition to conventional molecular descriptors (e.g., RDKit descriptors), STELLA™ incorporates scores from QIP-ADMET™, our proprietary model for predicting ADMET properties. QIP-ADMET™ is a graph transformer-based deep learning model pre-trained on quantum data and fine-tuned through transfer learning. It provides various predicted pharmacological properties, including Caco-2 permeability, aqueous solubility, plasma protein binding ratio, predictions for CYP inhibition and substrate probabilities, hepatocyte clearance, and hERG inhibition probability. STELLA™ also incorporate activity prediction models from GT-AutoML™, our proprietary AI-powered tool for generating QSAR models. These features in STELLA™ enhances the drug discovery process by providing accurate prediction of the pharmacological profiles of potential drug candidates, thus accelerating their progression towards clinical success.

Efficient Research Project Management

STELLA™ provides a user-friendly graphic interface for project management, facilitating cross-functional team collaborations. Users can create projects, set design scenarios, configure molecular generation processes, and monitor results, all through an intuitive web-based interface. The platform is highly customizable and easily scalable to fit any organizational need. This ensures that STELLA™ can be seamlessly integrated into existing

workflows, enhancing its utility across various stages of drug discovery. An example of project management is shown in Figure 3, and Figure 4 illustrates an example of the customizable configuration for molecular generation.

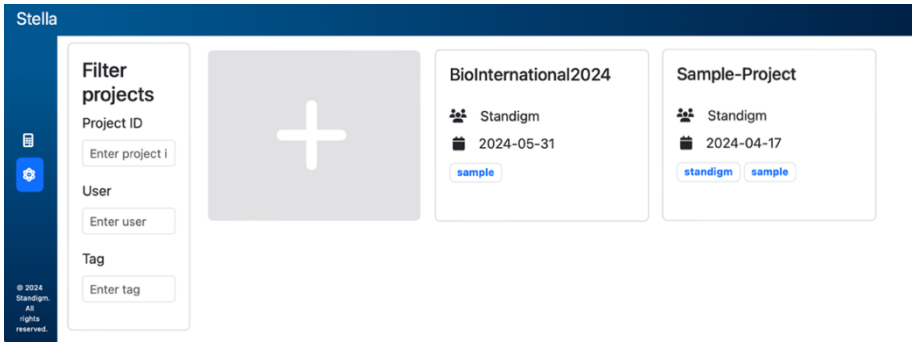


Figure 3. STELLA™'s project management interface.

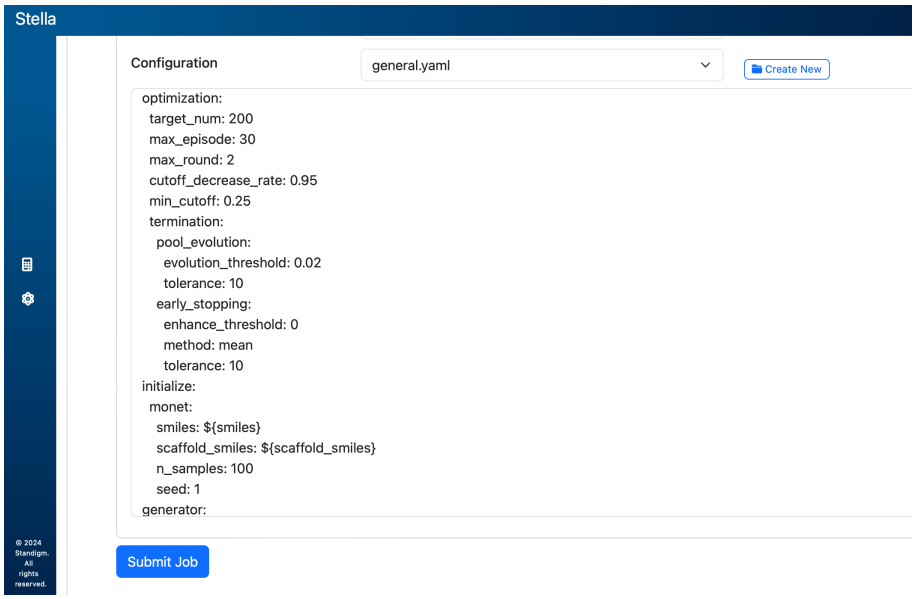


Figure 4. Customizable generation configuration of STELLA™.

Performance of STELLA™

We compare the performance of STELLA™ with REINVENT 4, a widely used tool for de novo molecular generation. REINVENT 4 is a framework for designing small molecules with optimized multiple properties using generative deep learning models, implemented through reinforcement learning and a curriculum learning-based optimization algorithm.

We first apply STELLA™ and REINVENT 4 to a hypothetical virtual screening scenario aimed at identifying novel phosphoinositide-dependent kinase-1 inhibitors. The criteria for hit identification were defined as a GOLD PLP Fitness score of ≥ 70 and a drug-likeness score (QED) of ≥ 0.7. Both metrics were weighted equally in the objective score. In REINVENT 4, we employed 10 epochs of transfer learning followed by 50 epochs of reinforcement learning with a batch size of 128. The batch size represents the number of molecules generated, scored, and incorporated into each epoch. To ensure a comparable computational workload in STELLA™, 128 molecules were generated per iteration with a total of 50 iterations performed (one iteration in STELLA™ corresponds to one epoch in REINVENT 4). Over 50 training iterations/epochs, REINVENT 4 generated 116 hit compounds (average 1.81% hit rate per epoch) with mean scores of 73.37 for GOLD PLP Fitness and 0.75 for QED (Figure 5). In contrast, STELLA™ produced 368 hit compounds (average 5.75% hit rate per iteration) with higher mean scores of 76.80 for GOLD PLP Fitness and 0.77 for QED. STELLA™ maintains a higher cumulative number of hits throughout the optimization process (Figure 5). In addition, STELLA™-generated hits occupy a more favorable Pareto frontier, indicating better optimization of target properties (Figure 5). STELLA™ shows substantially greater scaffold diversity, generating 276 unique generic Murcko scaffolds compared to 115 in REINVENT 4 (Figure 5). Overall, the results indicate that STELLA™ not only explores a broader region of chemical space but also more effectively optimizes the target properties.

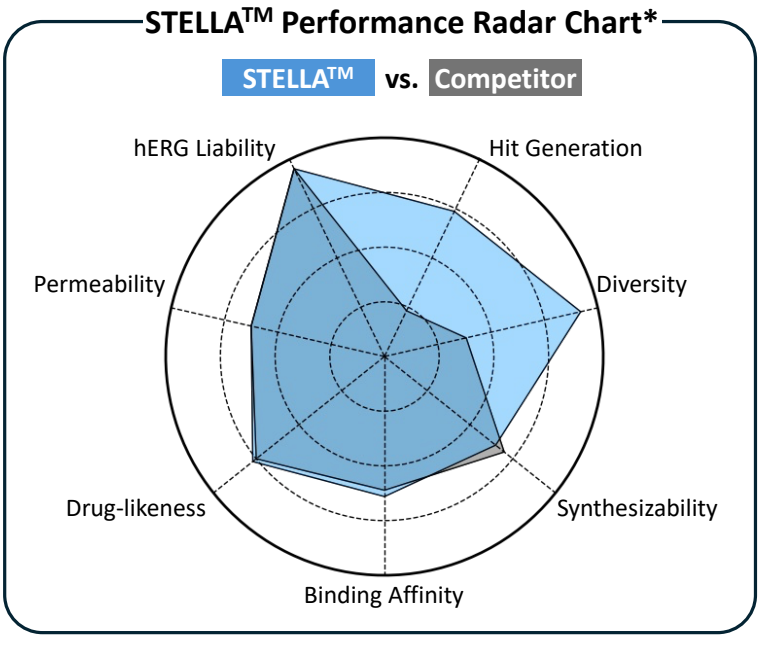


Figure 5. A performance comparison between STELLA™ and REINVENT 4 on identical target and settings.

To further assess performance in multi-parameter optimization, we evaluated the molecular optimization capabilities of STELLA™ and REINVENT 4 for two target proteins: tyrosine-protein kinase Abl1 and cellular tumor antigen p53. The goal of this benchmark was to optimize a single objective score that encompassed 16 properties relevant to drug candidate evaluation, including: (1) GOLD docking score (PDB entries 4TWP and 4AGQ for Abl1 and p53, respectively). (2) 3D similarity to the reference ligand (ligands in 4TWP and 4AGQ) calculated by OpenEye's ROCS. (3) RDKit descriptors—QED, AlogP, and the synthetic accessibility score (SAscore). (4) 11 properties from QIP-ADMET™. Both tools were configured to generate 2,500 molecules per iteration/epoch across 50 iterations for STELLA™ and 50 epochs for REINVENT 4.

We compared the ability of STELLA™ and REINVENT 4 in exploring the chemical space (Figure 6). The molecules selected at each iteration/epoch were transformed into ECFP4 fingerprints and projected onto a 2D space using UMAP for dimensionality reduction. To illustrate the trajectory over exploration, later iterations/epochs are represented by darker points. Both STELLA™ and REINVENT 4 initially select molecules from similar regions of the chemical space. However, as iterations/epochs progress, STELLA™ demonstrates a stronger tendency to expand its search space, whereas REINVENT 4 maintains a more focused search within its initial generation space. The results once again demonstrate STELLA™'s superior performance over REINVENT 4, both in broader exploration of chemical space and in more effective optimization of target properties.

Figure 6. Visualization of top molecules generated over iterations/epochs. Each point represents a single molecule, with color intensity increasing according to iteration/epoch progression. Molecules generated by STELLA™ are shown in a blue gradient, while REINVENT 4 molecules are depicted in a red gradient.

Conclusion

STELLA™ stands out as a state-of-the art computational platform in the field of drug discovery, harnessing the power of AI to optimize the hit-to-lead process. Its advanced molecular design capabilities, rapid compound generation, multi-objective optimization, and integrated ADMET prediction and QSAR modeling make it an invaluable tool for researchers aiming to accelerate the journey from concept to clinic. With its project-based management, customizable design platform, and robust supporting technologies, STELLA™ is well-positioned to transform the future of pharmaceutical research and development.

GT-AutoML™, a part of STELLA™'s suite of tools, enables the creation of tailored Structure-Activity Relationship (SAR) models. These models, both ligand-based and structure-based, are simplified through automated machine learning, making it easier to derive meaningful insights from complex data.

For more information and to explore how STELLA™ can enhance your drug discovery efforts, visit Standigm's STELLA™ page and Standigm's website.

Supporting Technologies

GT-AutoML™

GT-AutoML™, a part of STELLA™'s suite of tools, enables the creation of tailored Structure-Activity Relationship (SAR) models. These models, both ligand-based and structure-based, are simplified through automated machine learning, making it easier to derive meaningful insights from complex data.

QIP-ADMET™

QIP-ADMET™ is Standigm's AI-driven solution designed for rapid and reliable ADMET prediction. The QIP-ADMET™ model leverages extensive datasets that include experimental data, in silico quantum mechanical data, and molecular descriptors to ensure reliable performance and provide comprehensive ADMET insights.

SPICA™

SPICA™ is Standigm's AI-based novelty prediction model, which swiftly evaluates the novelty of new molecular structures. This tool ensures that the innovative compounds generated by STELLA™ can be effectively patented, providing a strategic advantage in the competitive pharmaceutical landscape.

Reference

[1] H. Jeon, H.S. Lee, and I. Joung, GALAPAGOS: Fragment-based Evolutionary Algorithm for Simultaneous Optimization of Drug-likeness and Affinity, ACS Fall 2022.
[2] J. Kim, W. Chang, H. Ji, and I. Joung, Quantum-Informed Molecular Representation Learning Enhancing ADMET Property Prediction, J. Chem. Inf. Model. 2024, 64: 5028-40.
[3] H. Jeon, J.K. Lee, W. Shin, H. Ji, I. Joung, and H.S. Lee, STELLA: A Drug Design Framework Based on Extensive Fragment-Level Chemical Space Exploration and Balanced Multi-Parameter Optimization (*in revision*)

Contact Us

For more information about Standigm and STELLA™ and how it can benefit you, please reach out to us:
- Email: business@standigm.com
- Website: www.standigm.com
- Latest STELLA™ White Paper: https://www.standigm.com/ai-saas/stella

Legal Information

The information contained in this white paper is for informational purposes only and does not constitute a commitment or guarantee of any kind. Standigm assumes no responsibility for errors or omissions.

About Standigm

Standigm is a Korean AI drug discovery company with over 10 years of experience, backed by SK, Pavilion Capital, and Kakao Ventures, with $70M+ in funding and partnerships with global pharma. Standigm's proprietary AI platforms—Standigm ASK™ for target identification and Standigm BEST™ for compound design—power continuous generation of commercially valuable pipelines.