

Платформа для дизайна синтеза молекул с заданными характеристиками

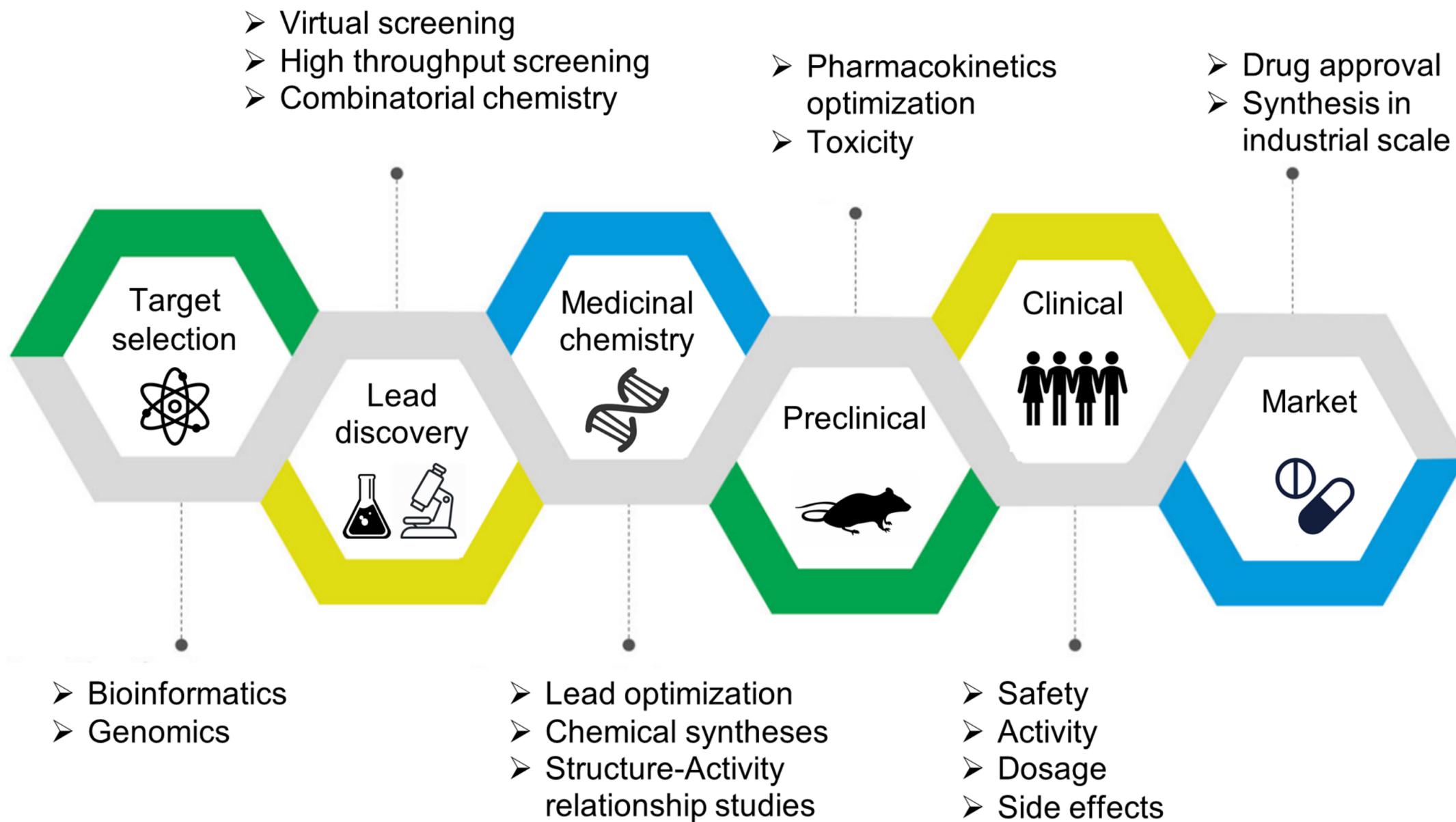
Автор: аспирант 3 г.о. Фатыхова Аделия
adelik21979@gmail.com

Д. Хорват², Д.В. Занков¹, Р.И. Нугманов¹, Т.И. Маджидов¹, А. Варнек²

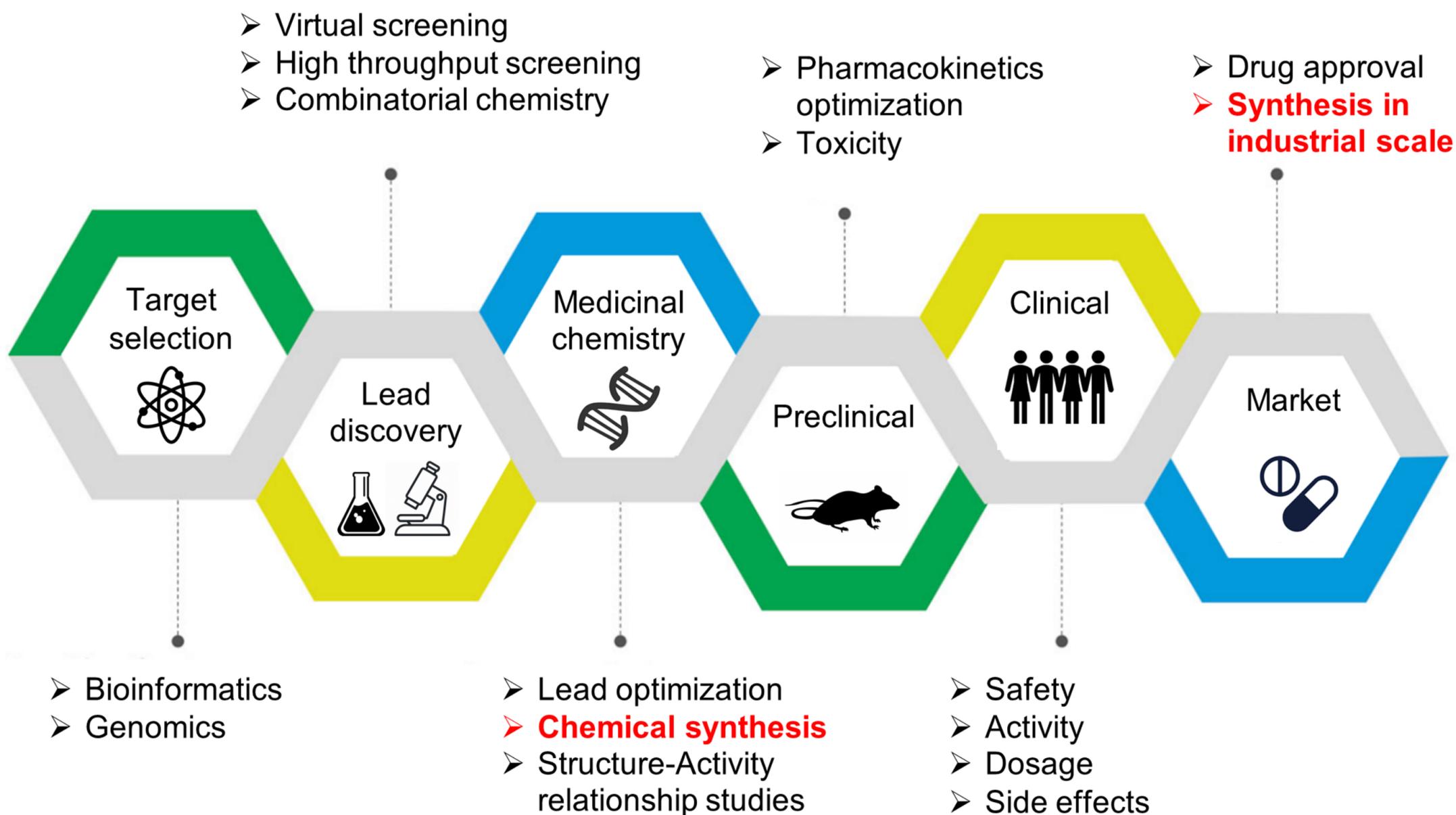
¹Казанский (Приволжский) федеральный университет, Химический
институт им. А.М. Бутлерова, Казань, Россия

²Университет Страсбурга, Страсбург, Франция

Проблема в области разработки лекарственных соединений: как синтезировать?



Проблема в области разработки лекарственных соединений: как синтезировать?



МАШИННОЕ ОБУЧЕНИЕ В РАЗРАБОТКЕ ЛЕКАРСТВ

- Де ново дизайн молекул на основе методов глубокого обучения (Deep Learning)
- Генеративно-сопоставительные сети (Generative adversarial networks - GAN)
- Обучение с подкреплением (Reinforcement Learning)

Kell, Douglas & Samanta, Soumitra & Swainston, Neil. (2020). Deep learning and generative methods in cheminformatics and chemical biology: Navigating small molecule space intelligently. *Biochemical Journal*. 477. 4559-4580. 10.1042/BCJ20200781.

Wenhao Gao and Connor W. Coley. (2020) The Synthesizability of Molecules Proposed by Generative Models

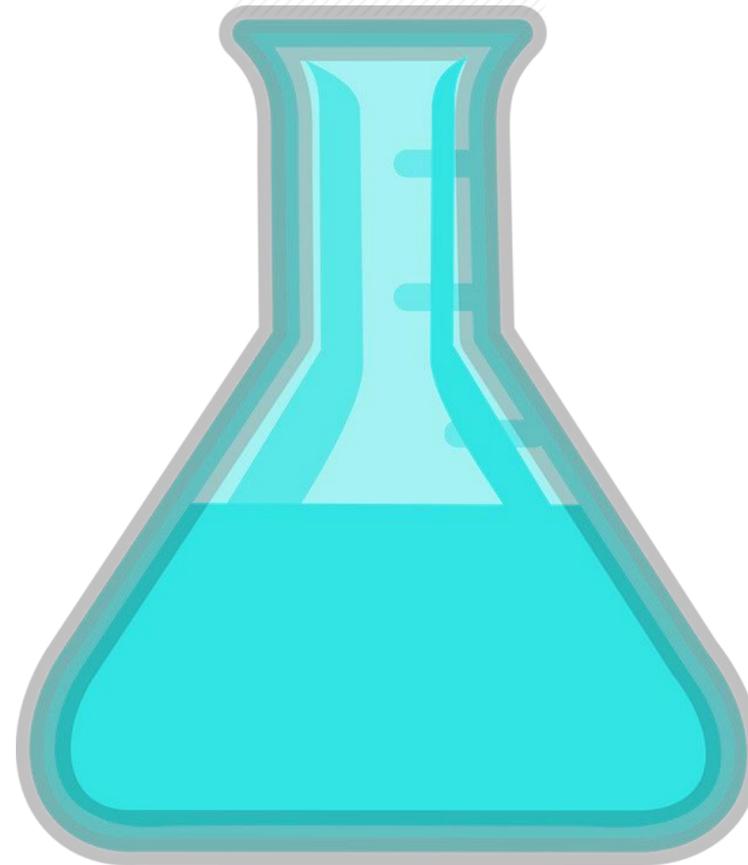
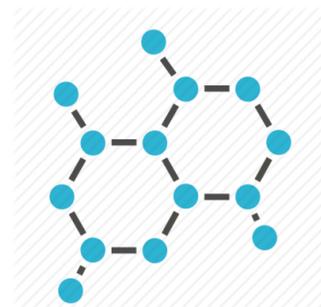
Journal of Chemical Information and Modeling 60 (12), 5714-5723. DOI: 10.1021/acs.jcim.0c00174

ЦЕЛЬ РАБОТЫ

Разработка виртуальной платформы для планирования синтеза химических соединений, которая позволяет предсказать путь синтеза не только **целевой молекулы**, но и ее **аналогов**, а также проводить поиск **молекул с заданными характеристиками**.

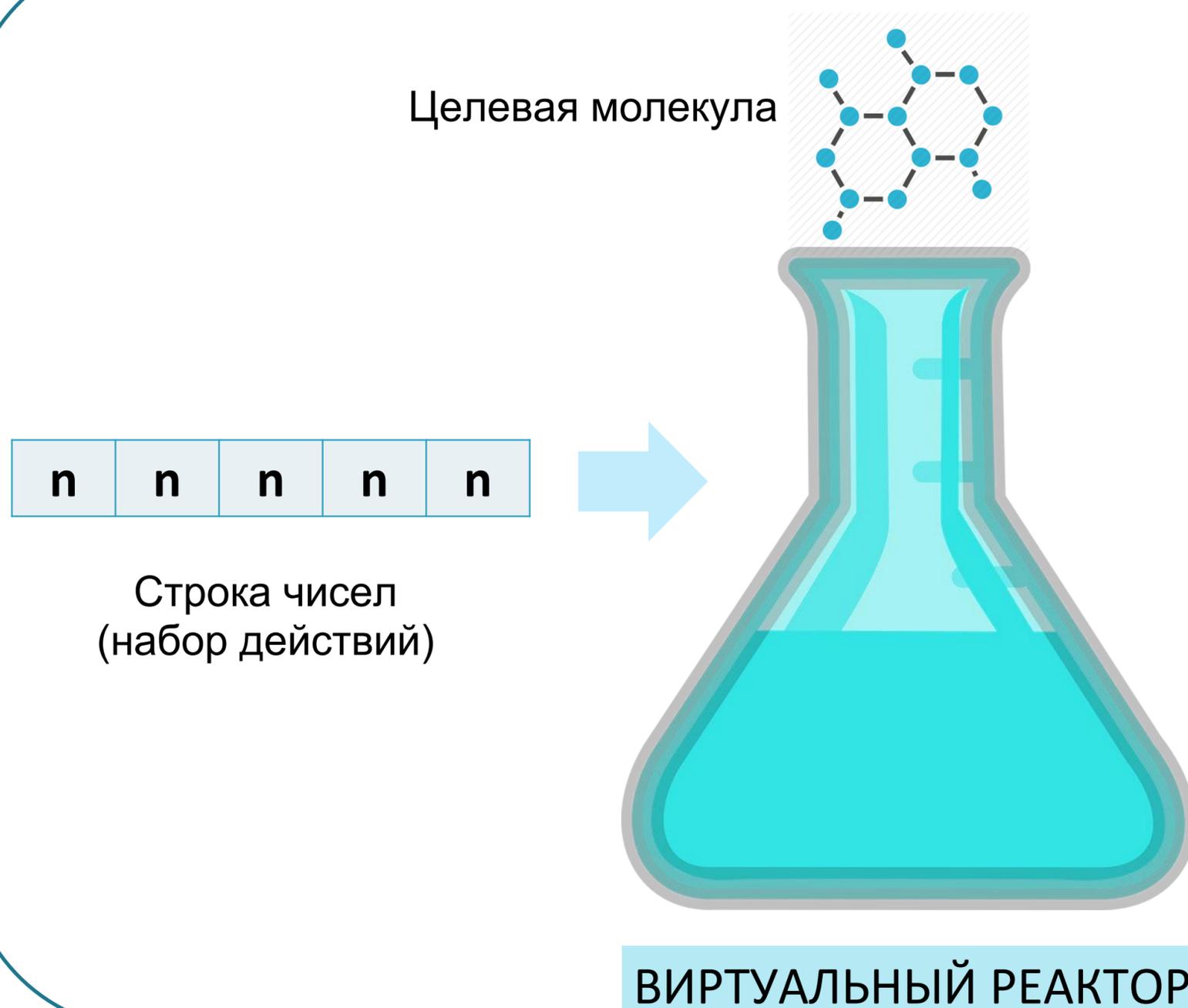
ВИРТУАЛЬНАЯ СРЕДА

Целевая молекула

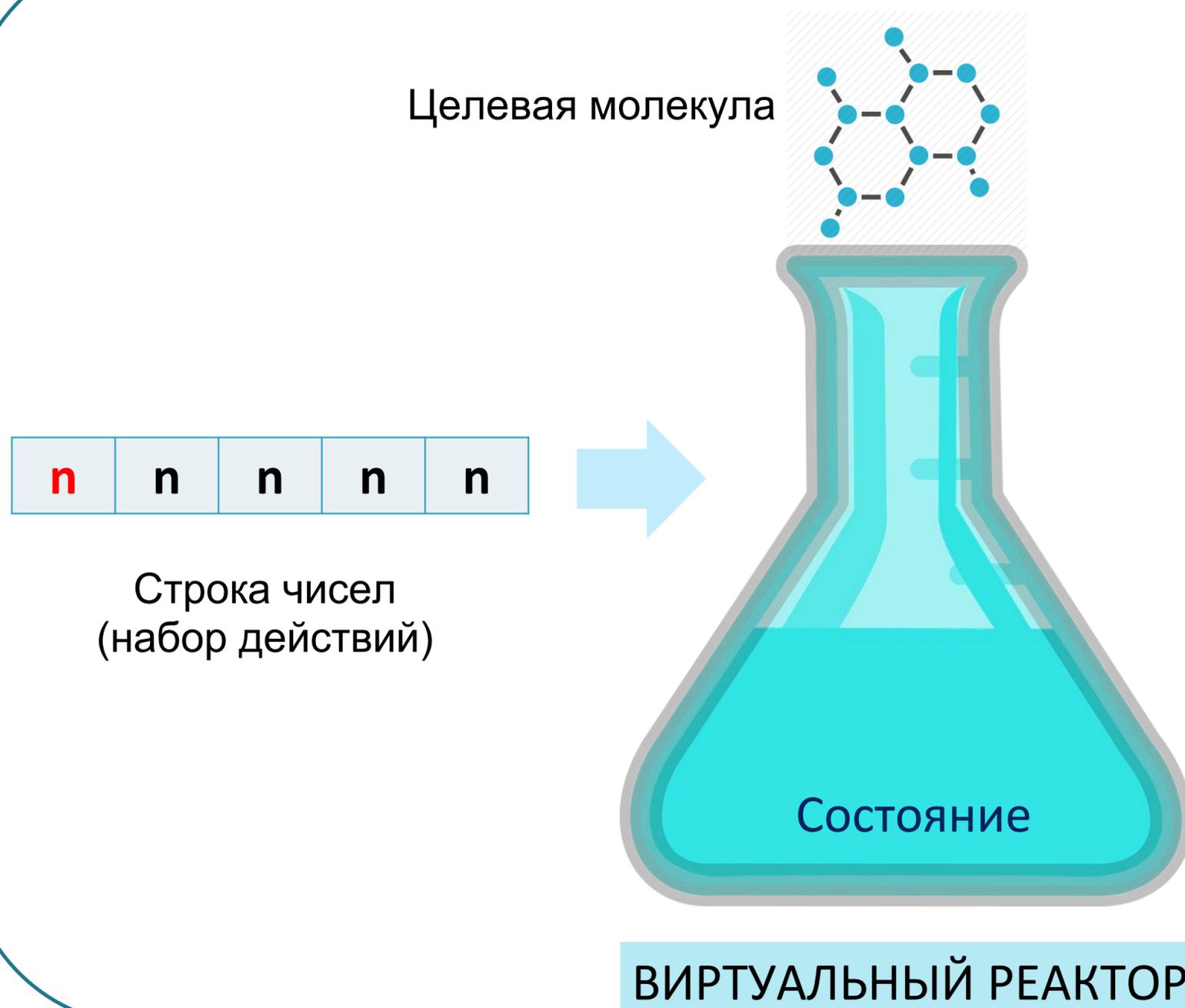


ВИРТУАЛЬНЫЙ РЕАКТОР

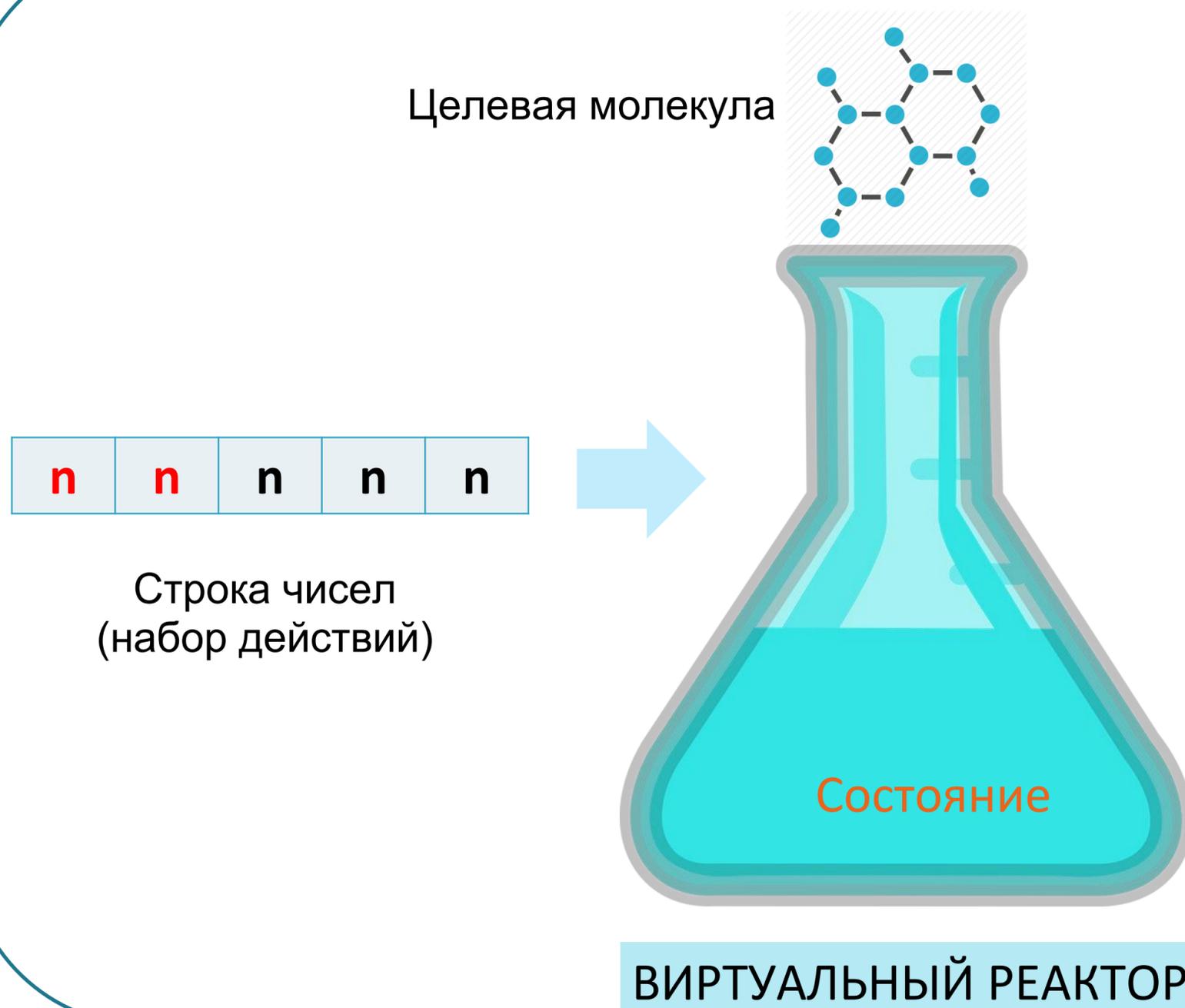
ВИРТУАЛЬНАЯ СРЕДА



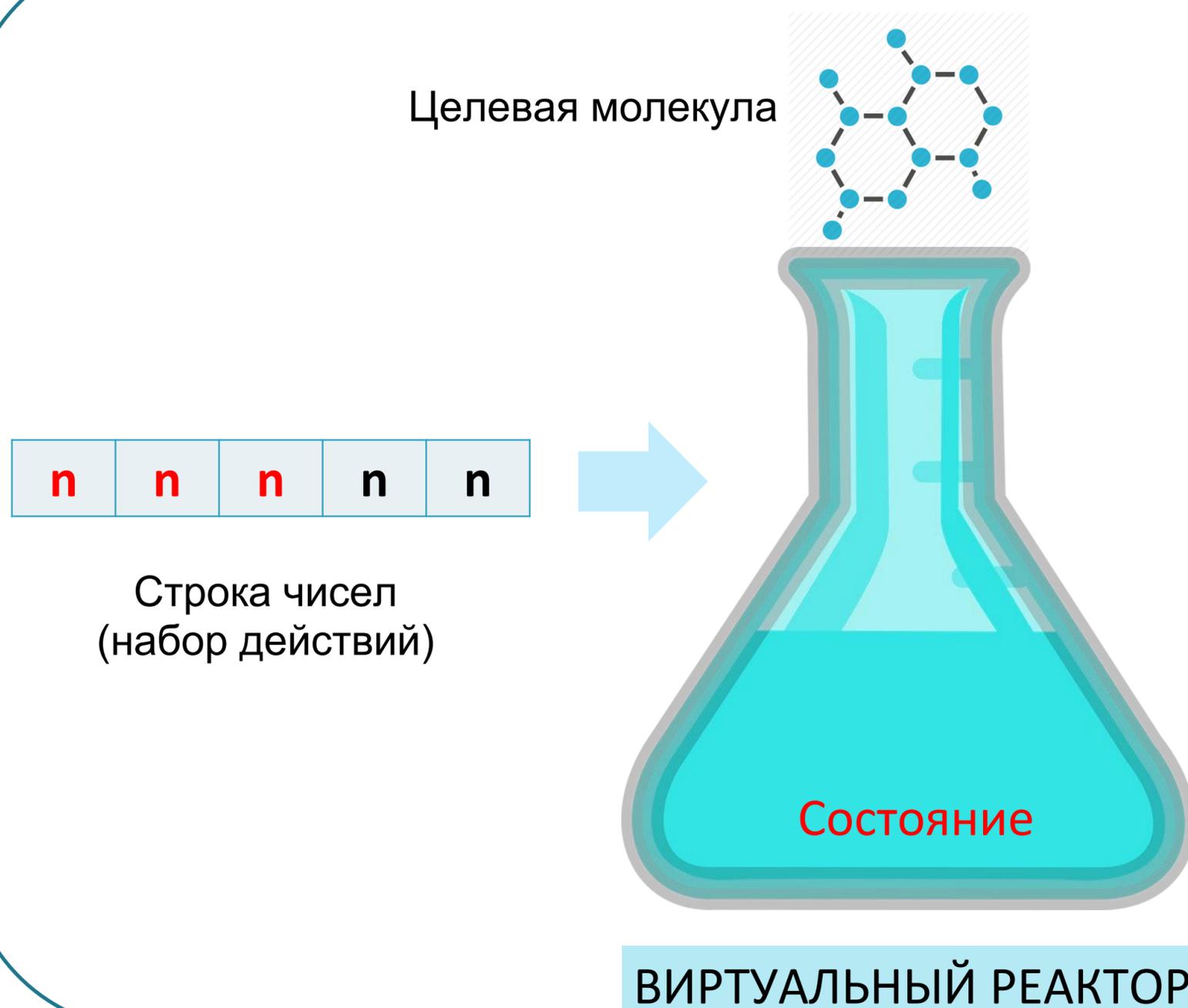
ВИРТУАЛЬНАЯ СРЕДА



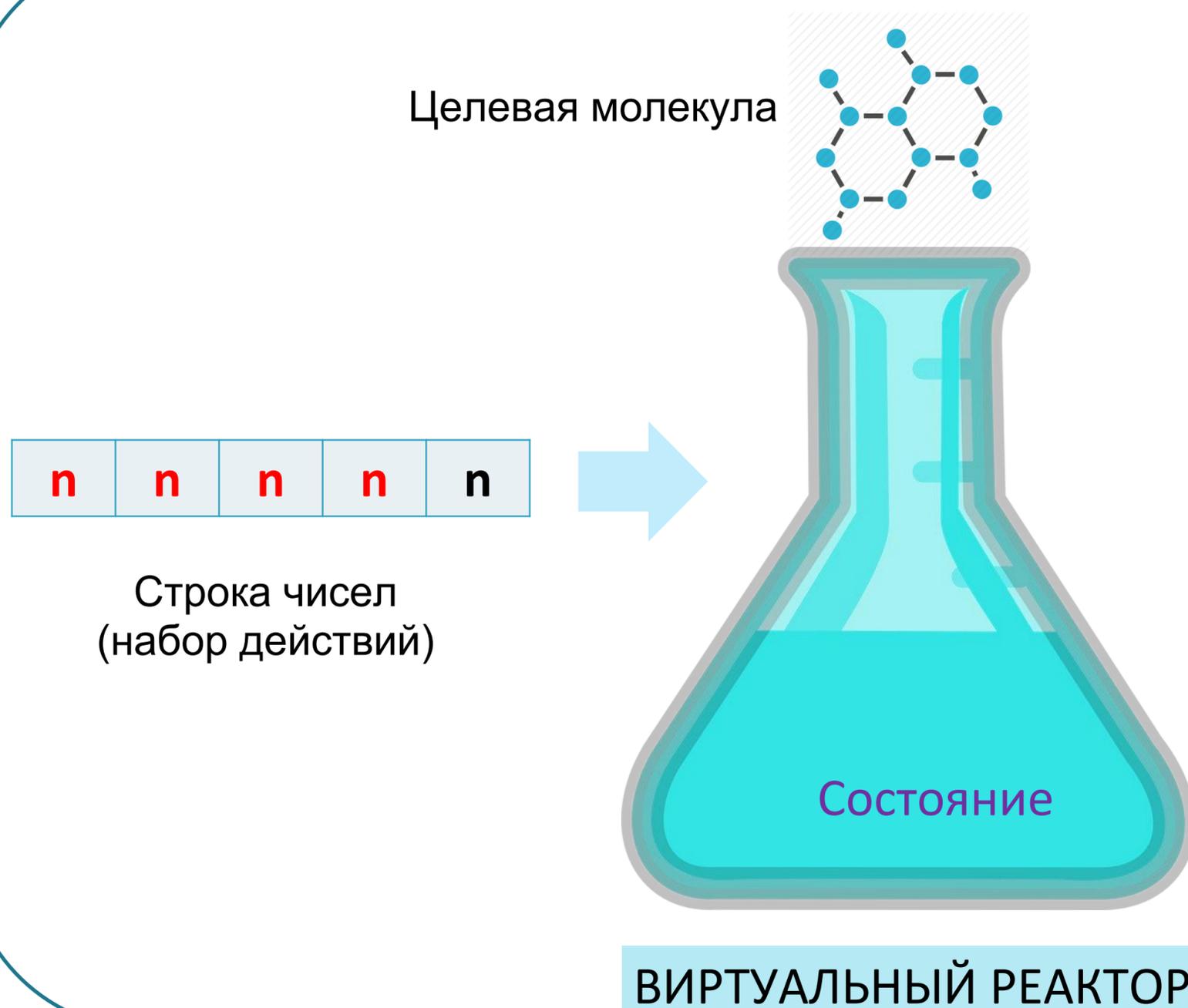
ВИРТУАЛЬНАЯ СРЕДА



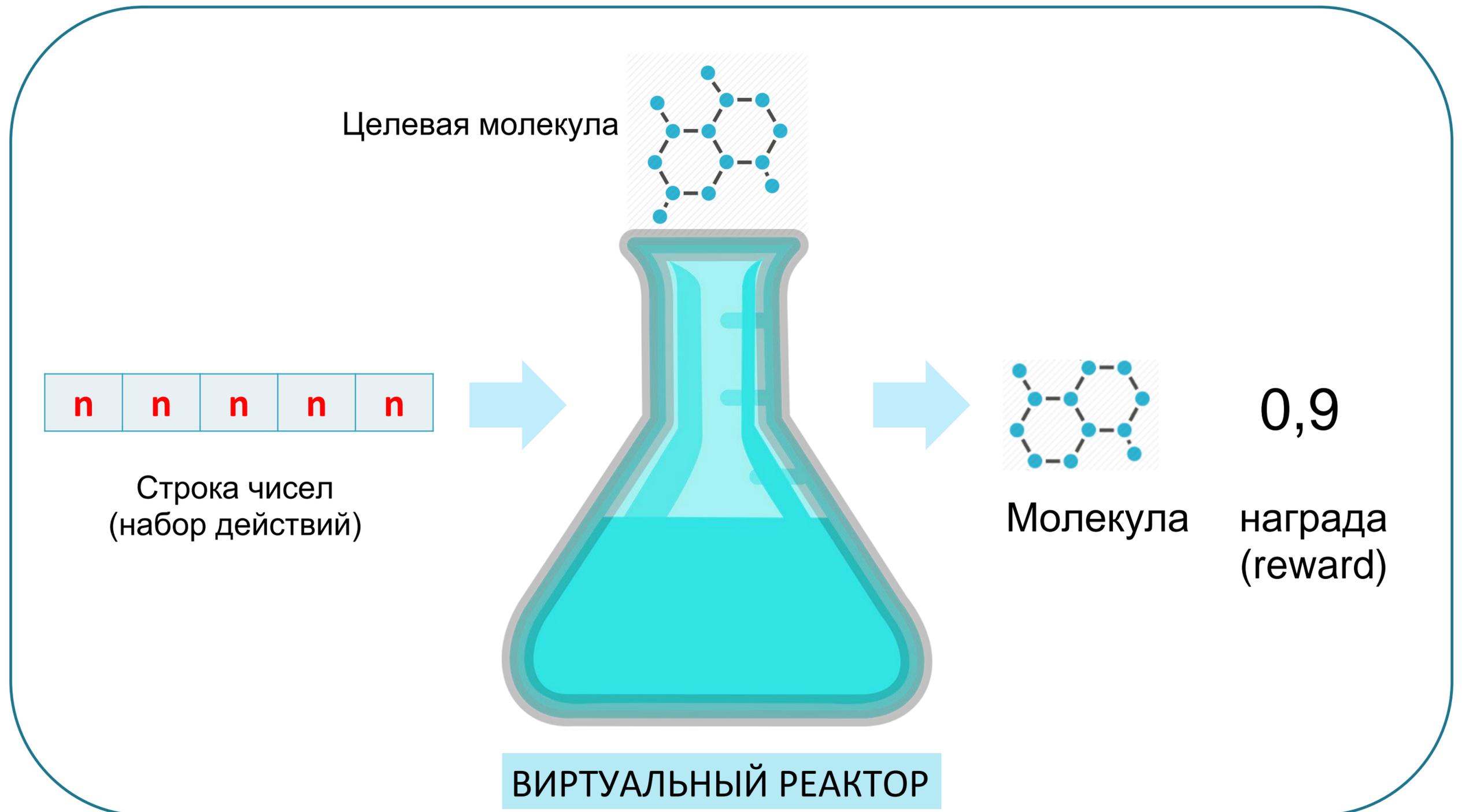
ВИРТУАЛЬНАЯ СРЕДА



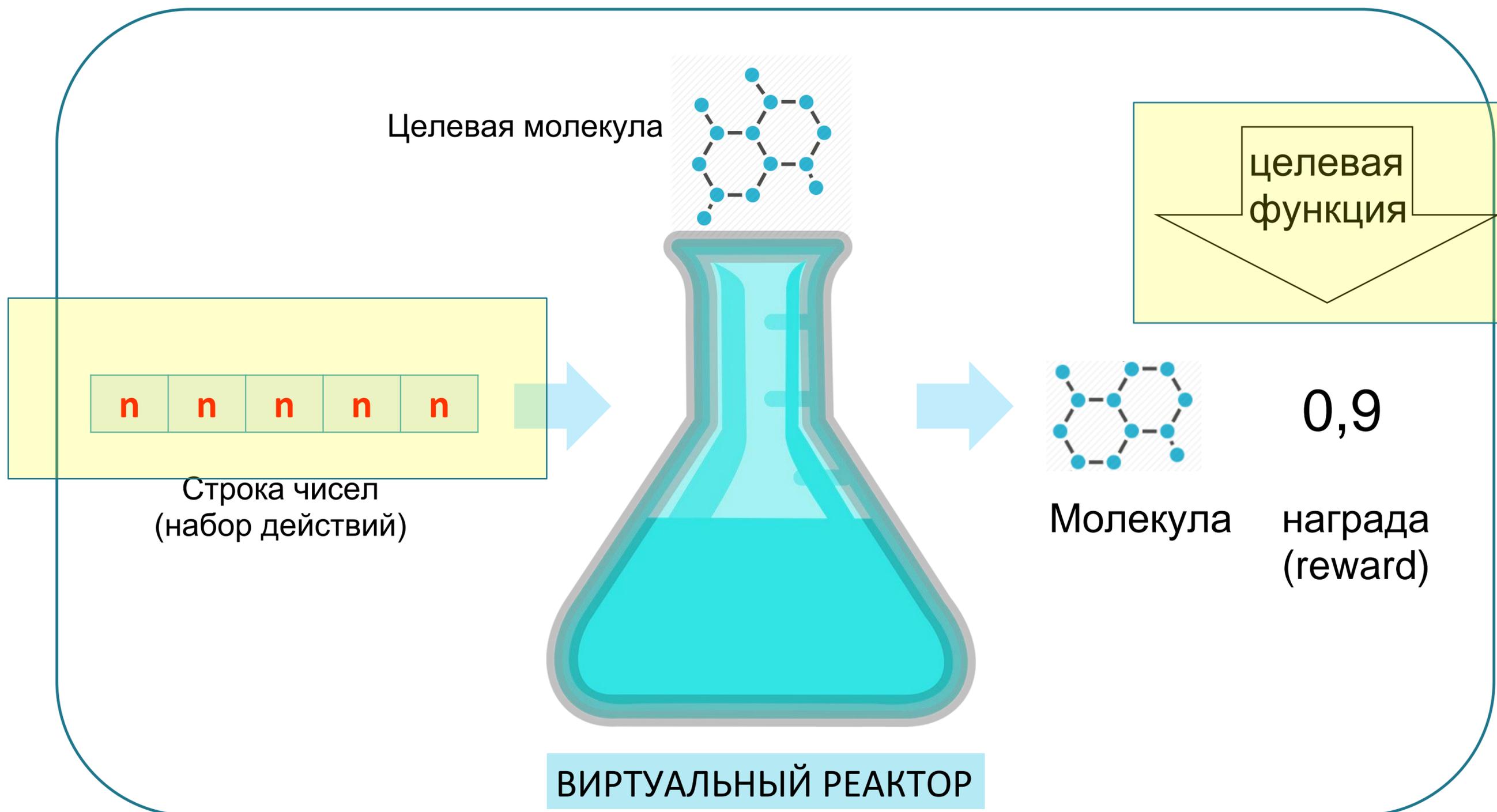
ВИРТУАЛЬНАЯ СРЕДА



ВИРТУАЛЬНАЯ СРЕДА



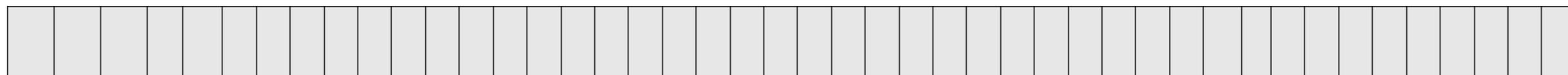
ВИРТУАЛЬНАЯ СРЕДА



ПРОСТРАНСТВО ДЕЙСТВИЙ



ПРОСТРАНСТВО ДЕЙСТВИЙ



Строительные блоки



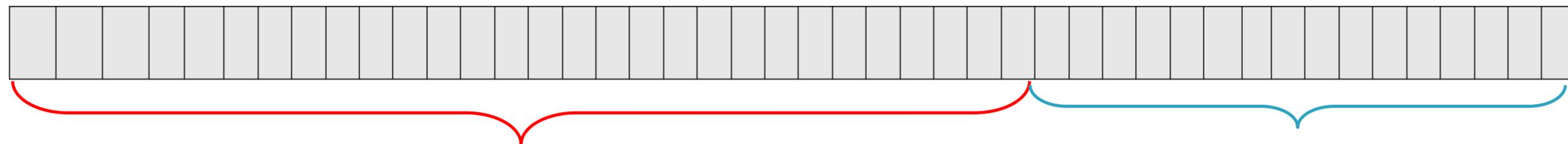
- Добавить блок в качестве реагента к текущему состоянию



- Выбрать молекулу как начальное состояние

Псевдо-реагенты
(однокомпонентные реакции)

ПРОСТРАНСТВО ДЕЙСТВИЙ



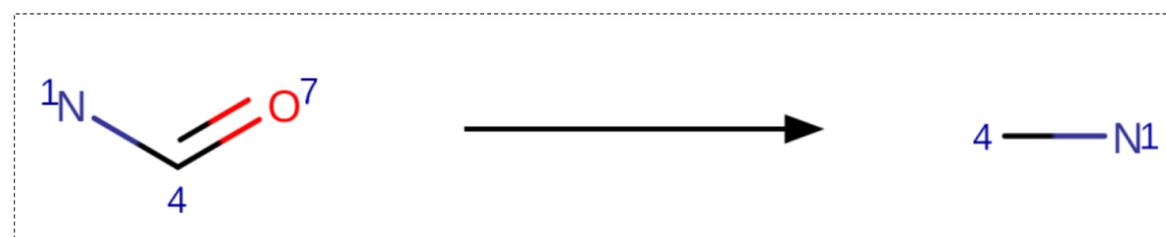
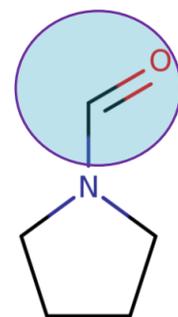
Строительные блоки

Псевдо-реактанты
(однокомпонентные реакции)

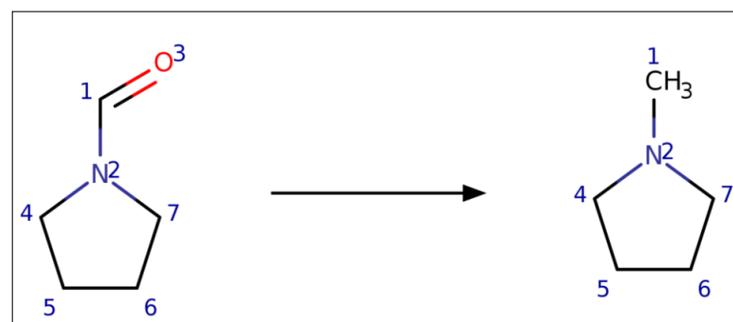


Если правило применимо, реактор генерирует возможные продукты реакций

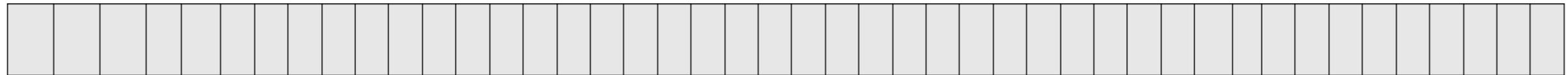
Состояние



Однокомпонентная реакция в качестве восстанавливающего агента



ПРОСТРАНСТВО ДЕЙСТВИЙ

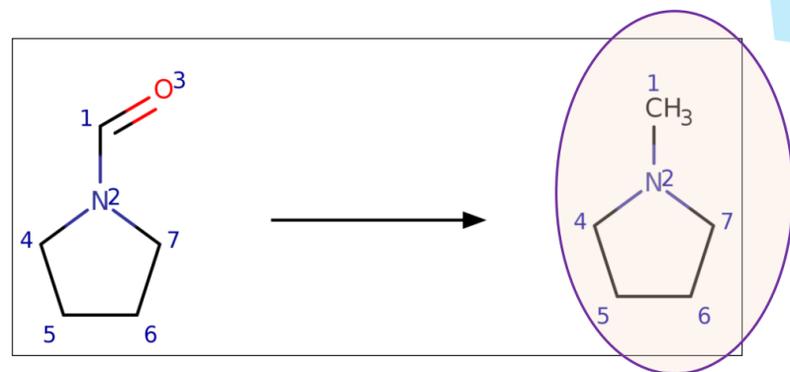


Строительные блоки

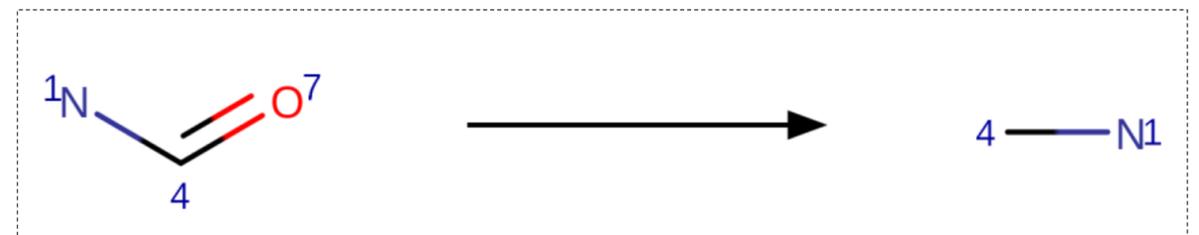
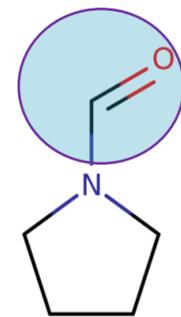
Псевдо-реактанты
(однокомпонентные реакции)



Если правило применимо, реактор генерирует возможные продукты реакций



Состояние



Однокомпонентная реакция в качестве восстанавливающего агента

СТРОИТЕЛЬНЫЕ БЛОКИ



17479



34210

ZINC15

495908

CGRTOOLS
STANDARTIZATIO
N

- imlicify hydrogens, aromatize
- clean isotopes and stereo
- remove duplicates

БД 396к уникальных строительных блоков

ЦЕЛЕВАЯ ФУНКЦИЯ

Данная платформа поддерживает интеграцию пользовательских оценочных функций, в зависимости от того, какое свойство нужно максимизировать:

- Структурное сходство (поиск целевой молекулы и ее аналогов - как альтернатива ретросинтетическому анализу)
- Биологическая активность (de novo дизайн)
- Фармакофорное сходство
- Различные модели QSAR

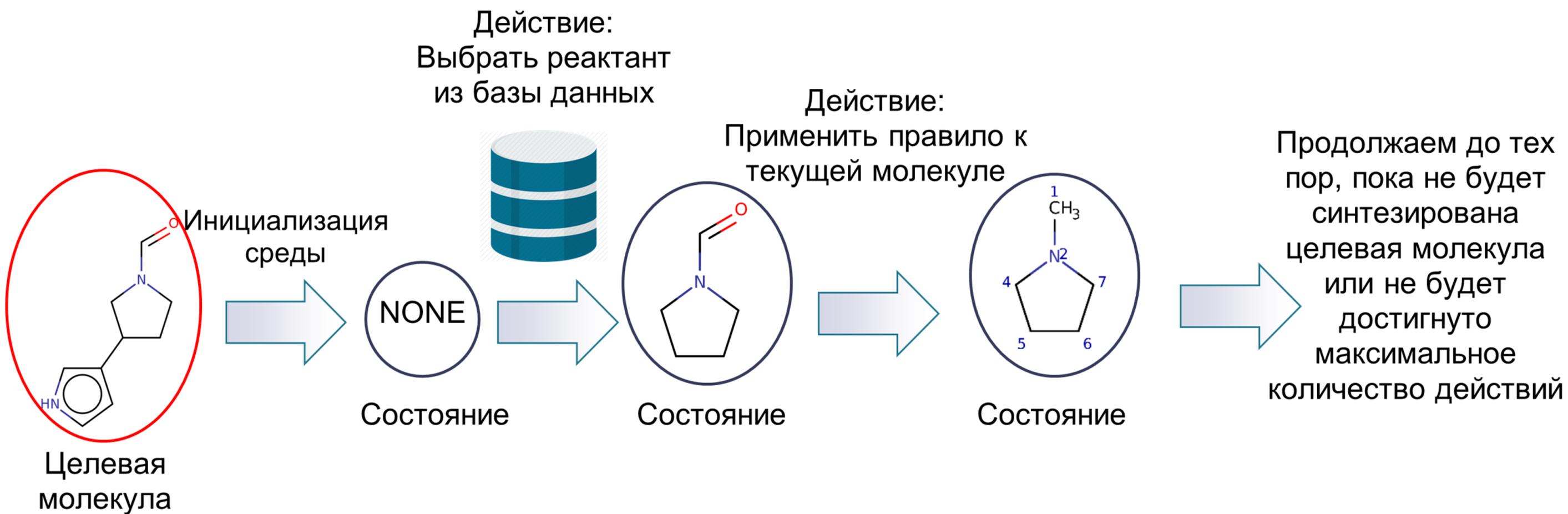
ЦЕЛЕВАЯ ФУНКЦИЯ

Данная платформа поддерживает интеграцию пользовательских оценочных функций, в зависимости от того, какое свойство нужно максимизировать:

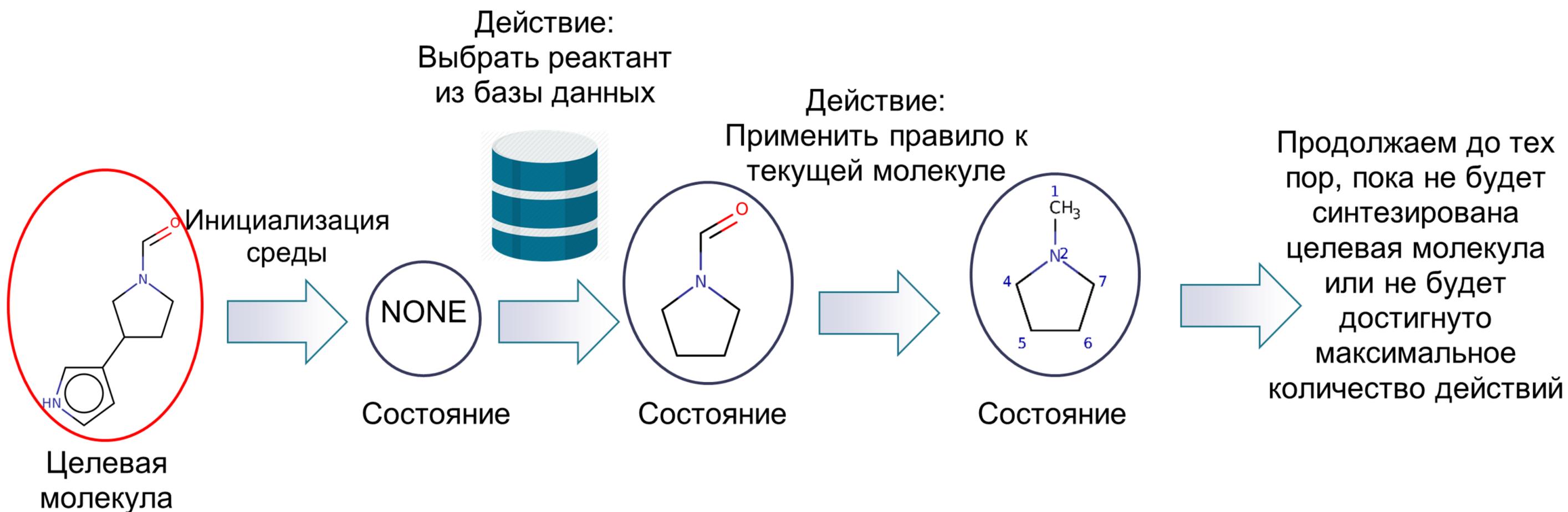
- **Структурное сходство** (поиск целевой молекулы и ее аналогов - как альтернатива ретросинтетическому анализу)
- Биологическая активность (de novo дизайн)
- Фармакофорное сходство
- **Различные модели QSAR (logP)**

В качестве оптимизируемой функции использовались коэффициент липофильности logP и структурное сходство с целевой молекулой на основе коэффициента Танимото.

АЛГОРИТМ ПОИСКА ПУТИ СИНТЕЗА

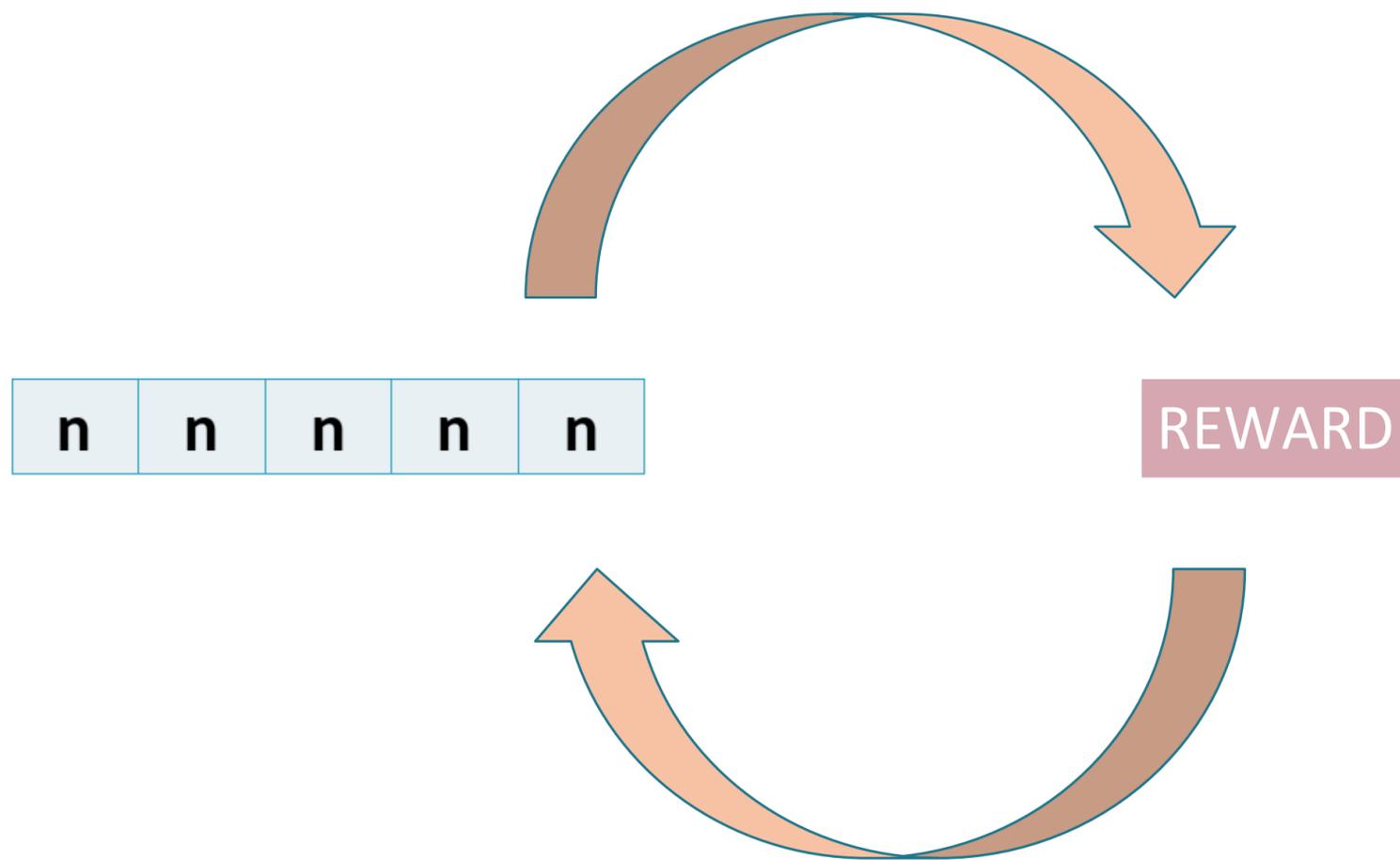


АЛГОРИТМ ПОИСКА ПУТИ СИНТЕЗА

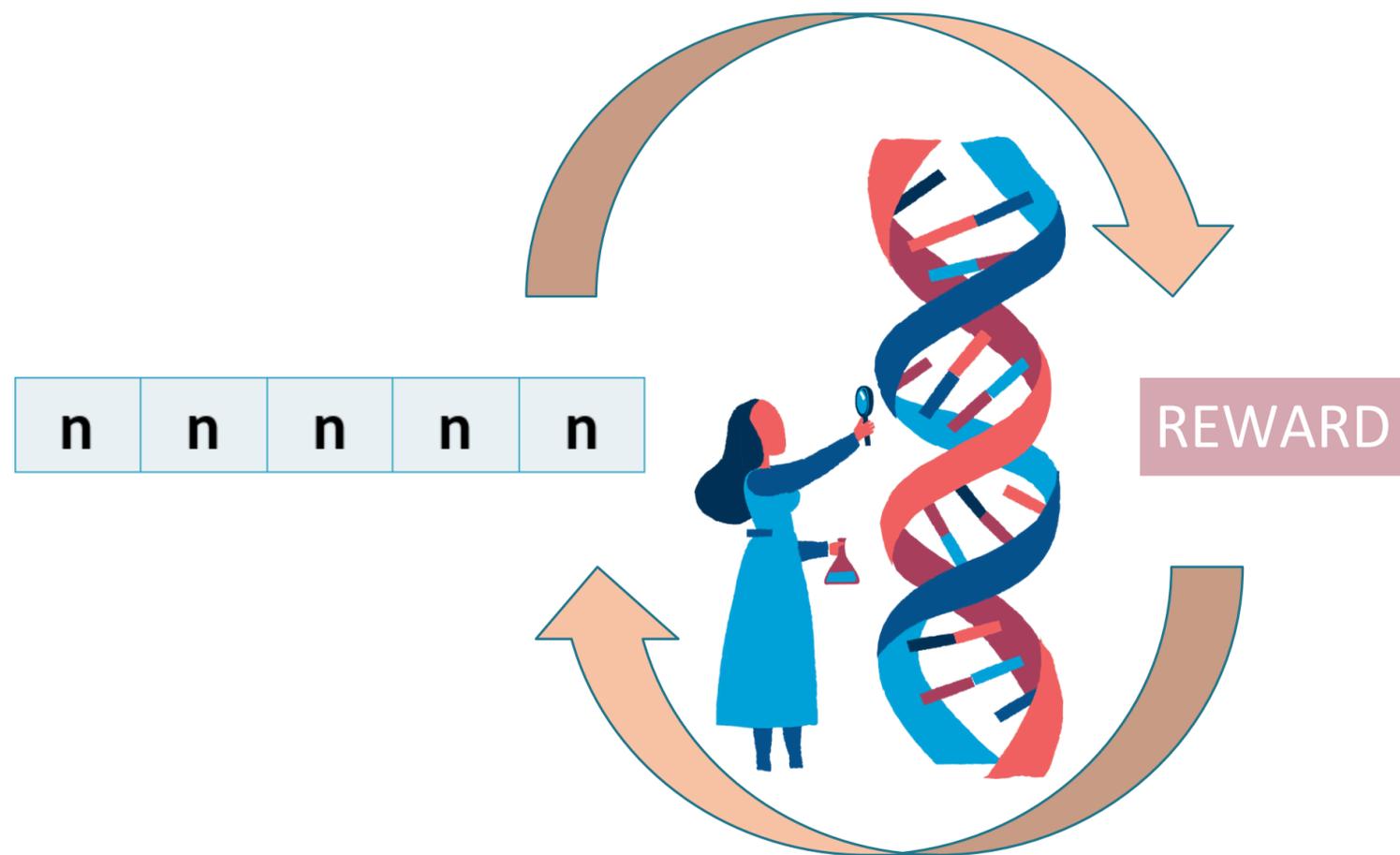


Задача: скомбинировать действия таким образом, чтобы максимизировать награду среды.

ВЫБОР МЕТОДА ОПТИМИЗАЦИИ



ВЫБОР МЕТОДА ОПТИМИЗАЦИИ



Генетический алгоритм оптимизации

ОПТИМИЗАЦИЯ ГЕНЕТИЧЕСКИМ АЛГОРИТМОМ

Хромосома - набор действий для виртуального реактора:



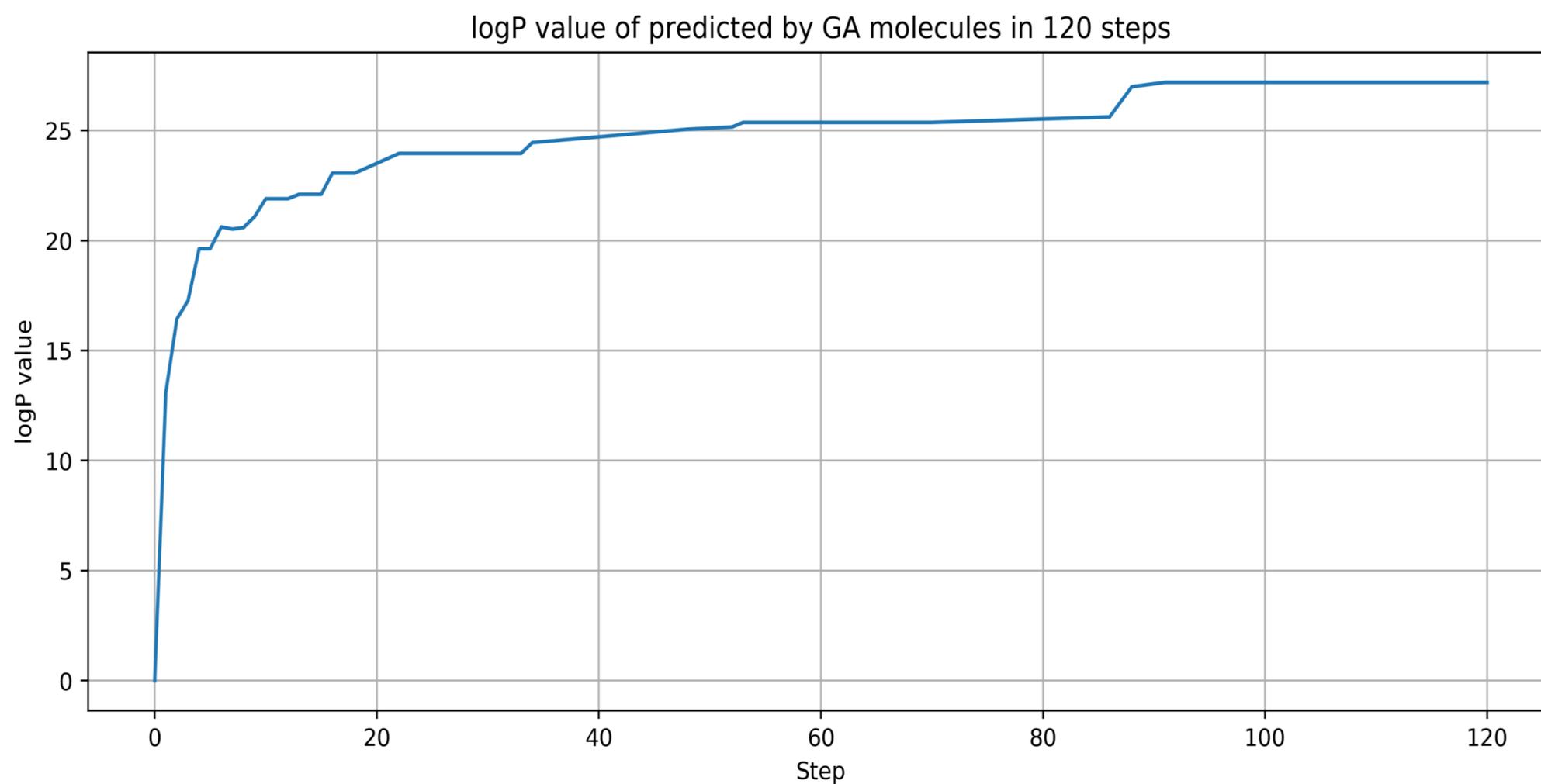
Для расчета коэффициента Танимото использовались структурные отпечатки длиной 4096 бит, отпечатки генерировались с использованием библиотеки CGRtools¹.

Fitness function: REWARD = $-\log(1 - \text{Tanimoto} + 10^{-10})$

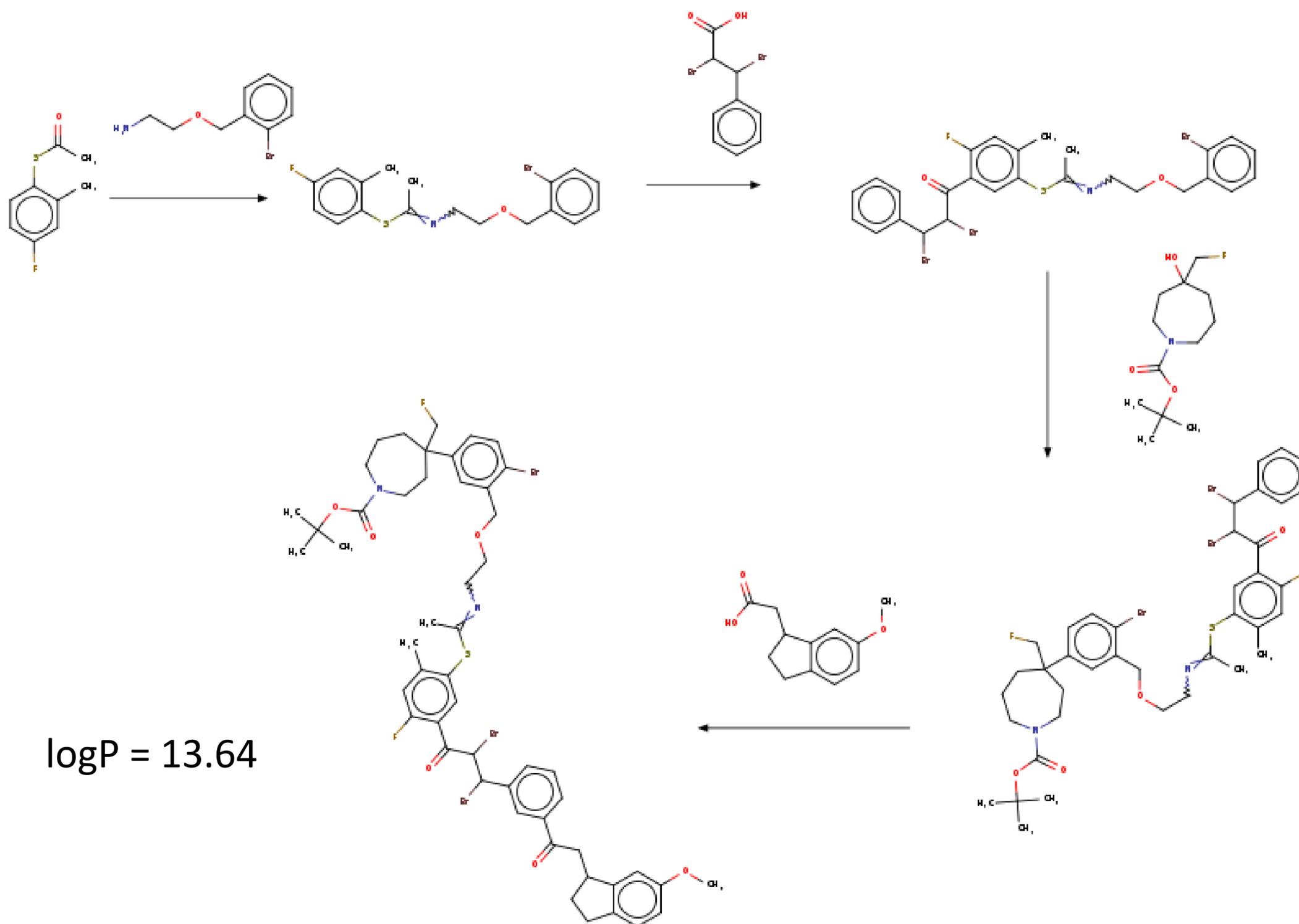
Расчет $\log P$ осуществлялся с использованием дескрипторов библиотеки RDKit.

¹ Nugmanov, R. et al. (2019) CGRtools: Python Library for Molecule, Reaction, and Condensed Graph of Reaction Processing. J. Chem. Inf. Model. 2019 59 (6), 2516-2521. doi: 10.1021/acs.jcim.9b00102.

РЕЗУЛЬТАТЫ: ПОИСК МОЛЕКУЛ С ВЫСОКИМ LOGP И ПУТЬ ЕЕ СИНТЕЗА

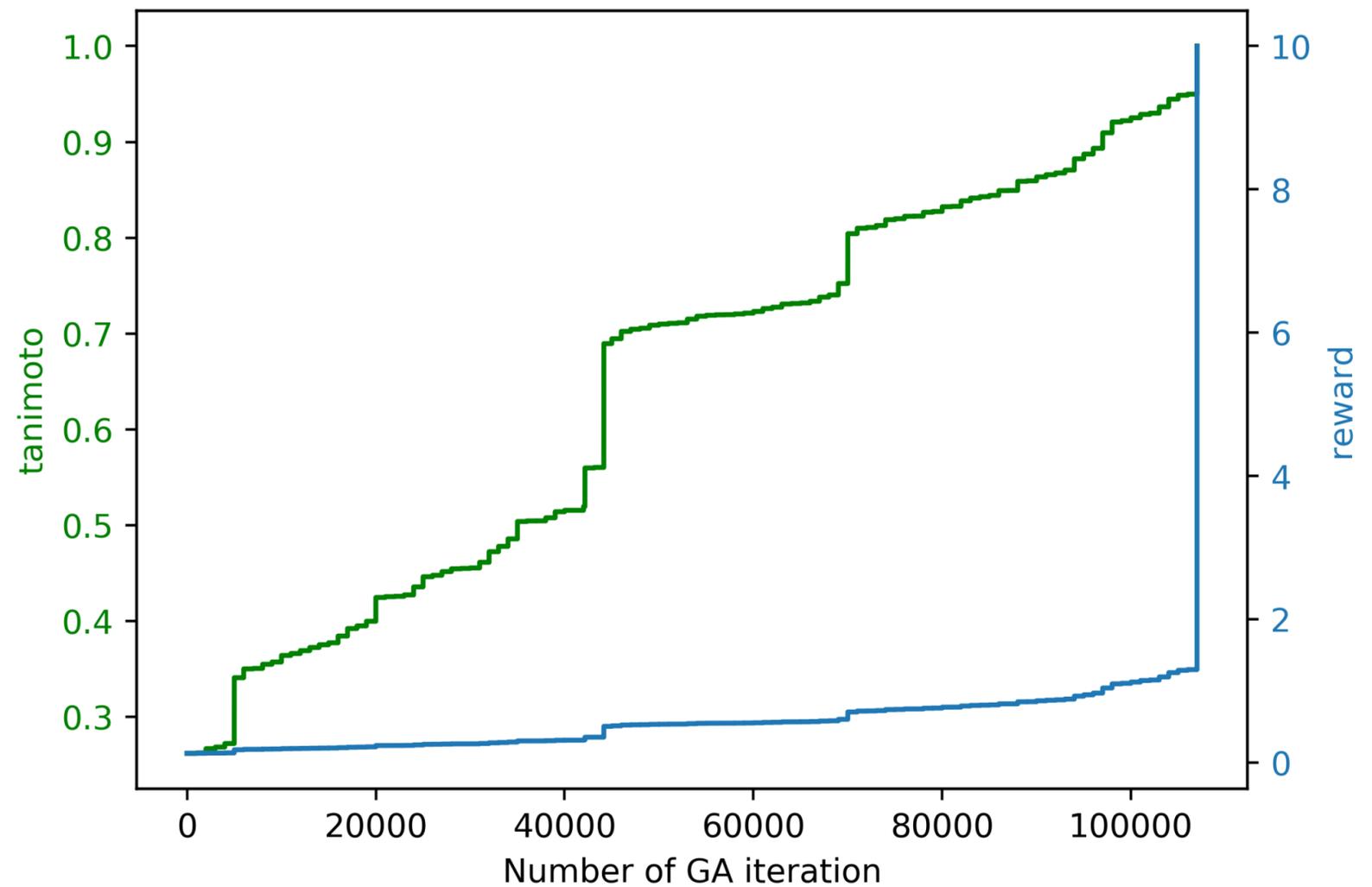
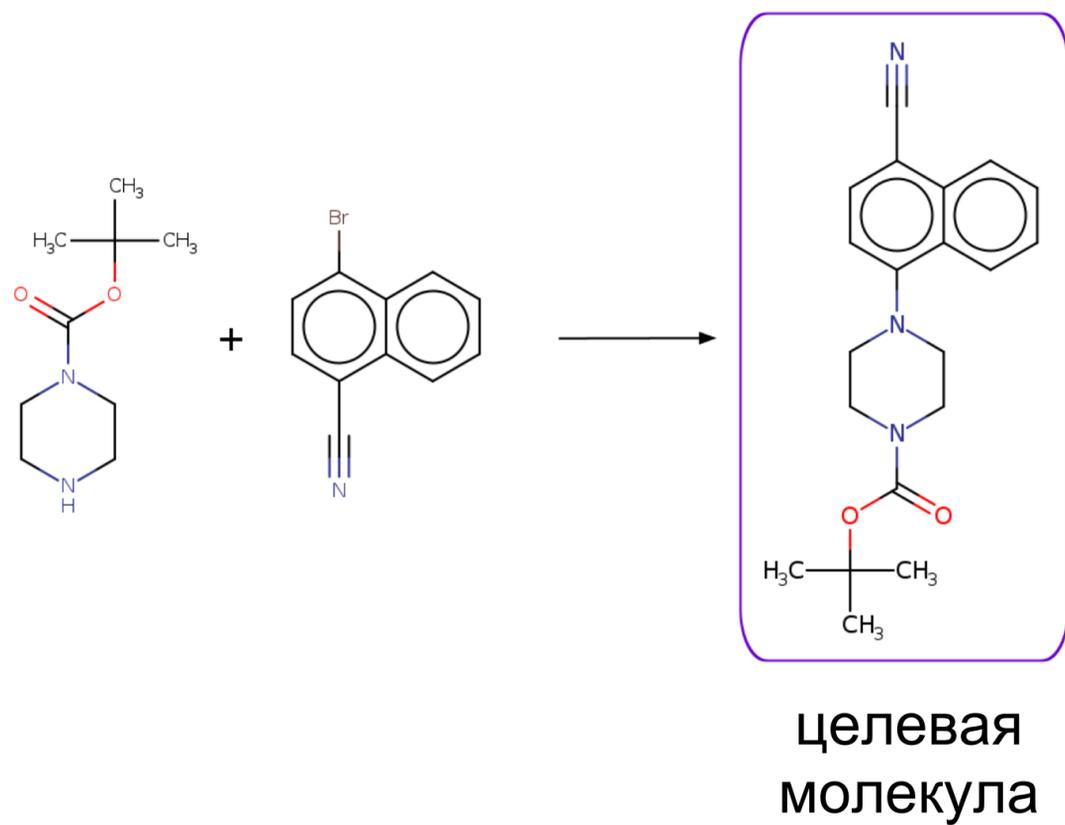


РЕЗУЛЬТАТЫ: ПОИСК МОЛЕКУЛ С ВЫСОКИМ LOGP И ПУТЬ ЕЕ СИНТЕЗА

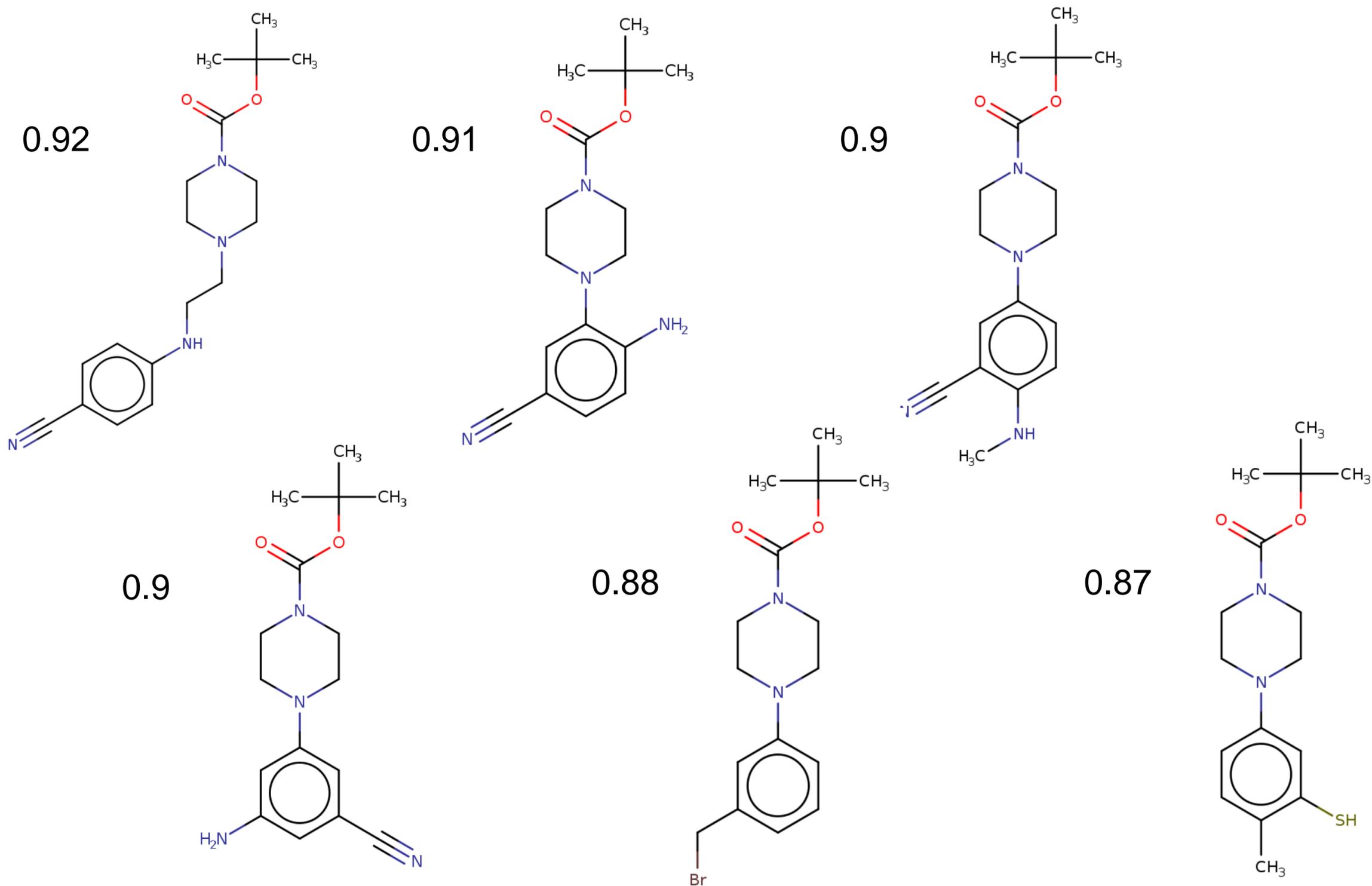


logP = 13.64

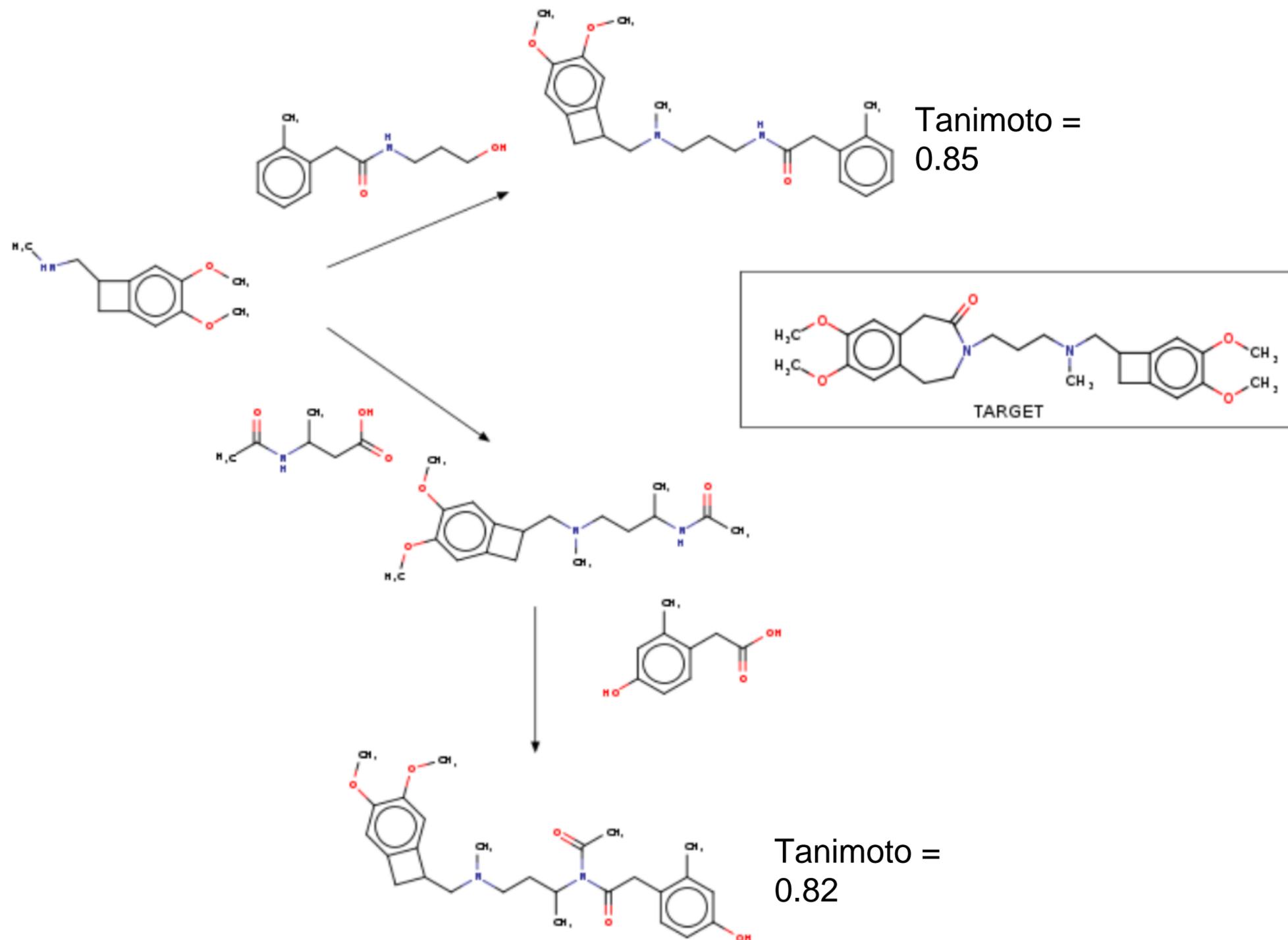
РЕЗУЛЬТАТЫ: ОПТИМИЗАЦИЯ СТРУКТУРНОГО СХОДСТВА



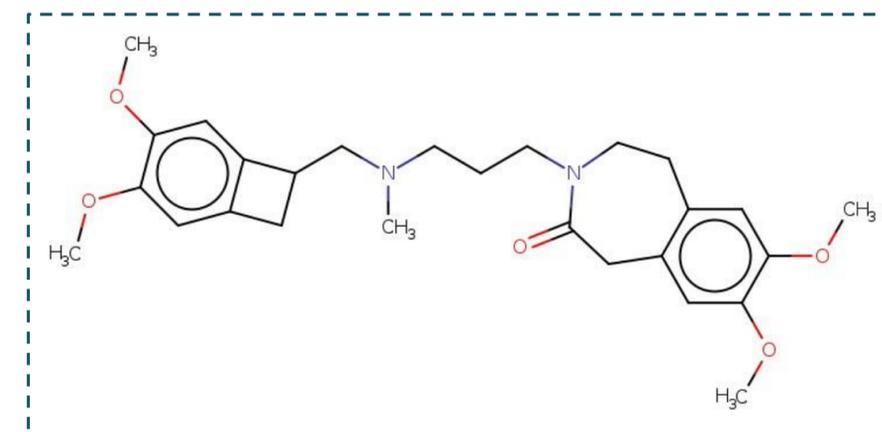
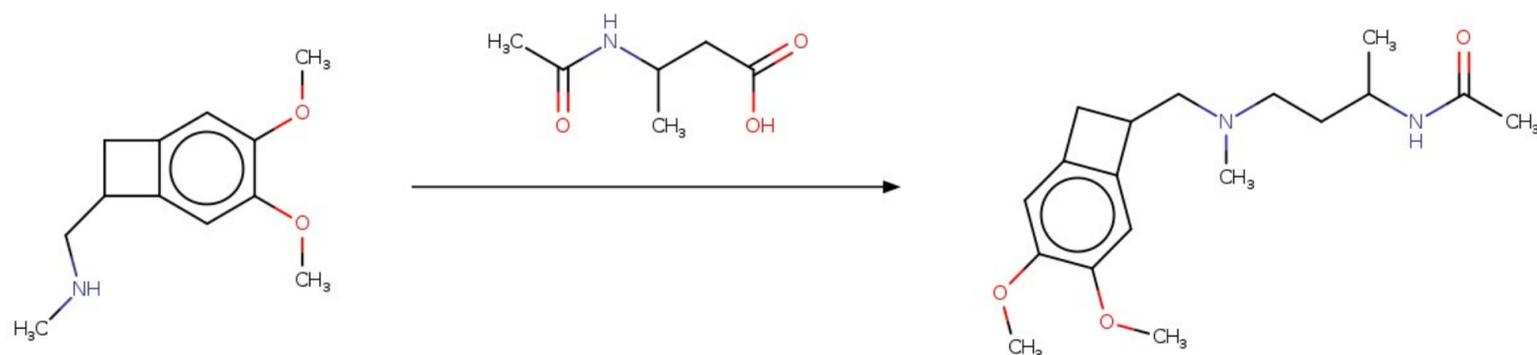
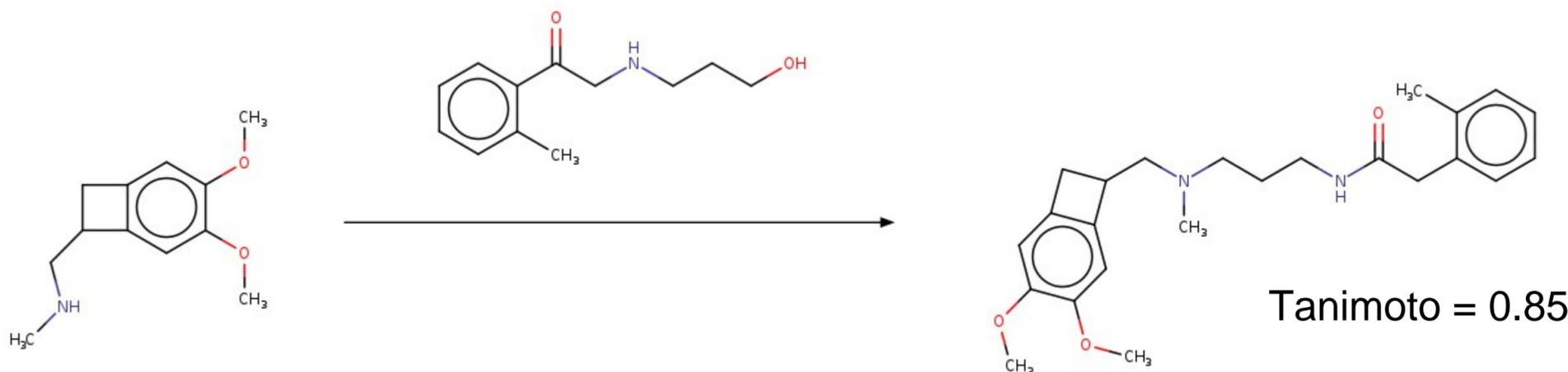
РЕЗУЛЬТАТЫ: ОПТИМИЗАЦИЯ СТРУКТУРНОГО СХОДСТВА



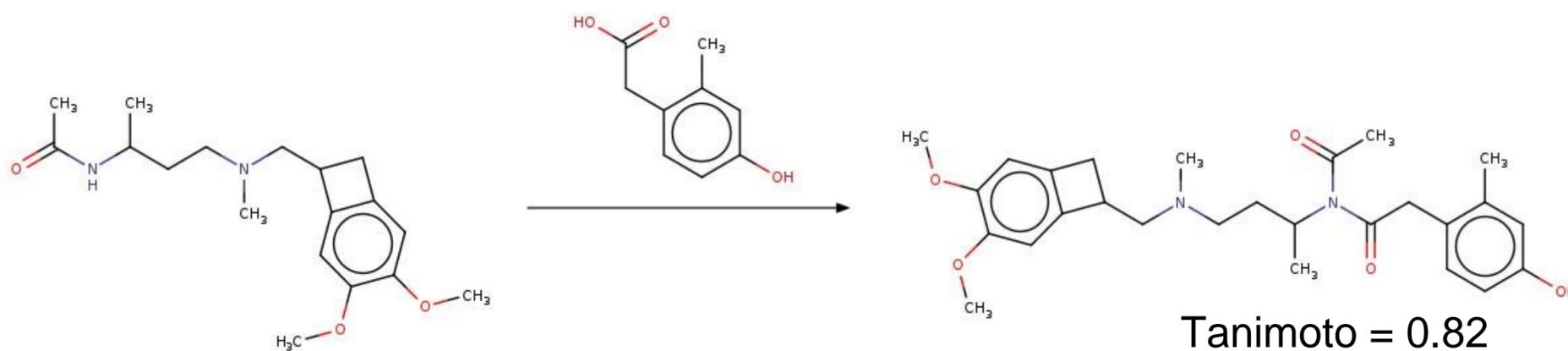
РЕЗУЛЬТАТЫ: ОПТИМИЗАЦИЯ СТРУКТУРНОГО СХОДСТВА



РЕЗУЛЬТАТЫ: ОПТИМИЗАЦИЯ СТРУКТУРНОГО СХОДСТВА



ivabradine



ЗАКЛЮЧЕНИЕ

- Разработана виртуальная платформа для планирования синтеза химических соединений, которая позволяет предсказать путь синтеза не только целевой молекулы, но и ее аналогов, а также проводить поиск молекул с заданными характеристиками;
- Была реализована поддержка оптимизации заданной целевой функции: использование характеристик, связанных с биологической активностью, позволяет использовать предложенную платформу как инструмент *de novo* дизайна молекул. При использовании в качестве оценочной функции индекса фармакофорного сходства, можно предсказывать пути синтеза молекул с желаемым фармакофором. Использование в качестве оценочной функции предсказанных QSAR моделями параметров активности молекул позволяет искать молекулы с желаемыми свойствами.
- Таким образом, разработанная платформа представляет собой универсальный инструмент, позволяющий планировать синтез молекулы, проводить поиск синтетически доступных аналогов молекулы и осуществлять *de novo* дизайн молекул с желаемыми свойствами.

Разработанный инструмент реализован на языке Python3.
Генетический алгоритм имплементирован на скриптовом языке C shell.

БЛАГОДАРНОСТИ



к.х.н. Маджидов Т.И.¹



Занков Д.В.^{1,2}



Dr. Dragos Horvath²



Prof. Alexandre Varnek²

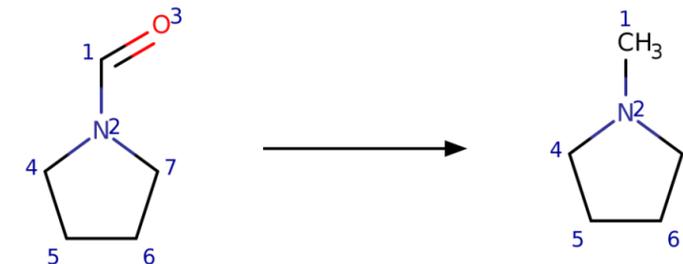
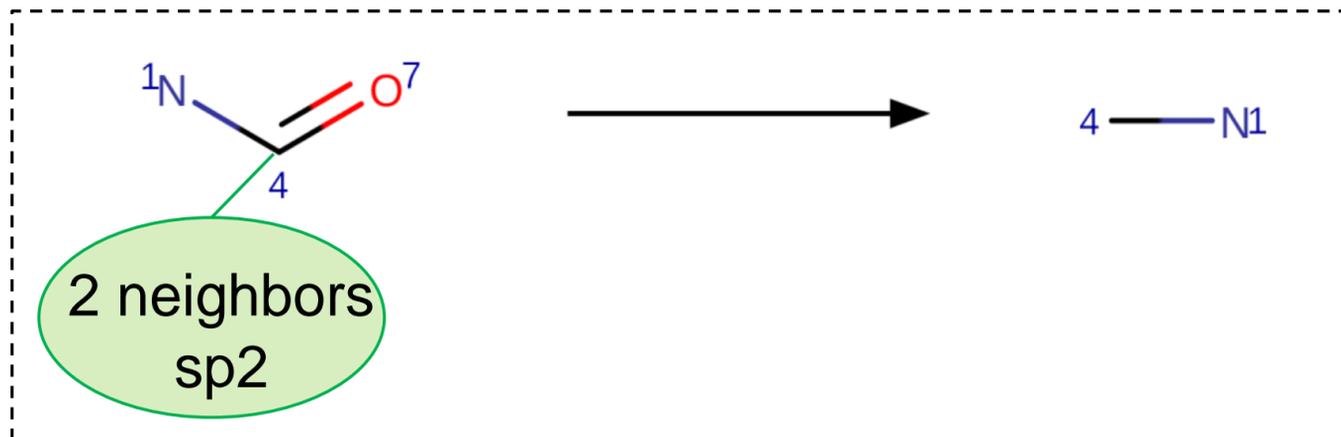
¹Казанский (Приволжский) федеральный университет, Химический институт им. А.М. Бутлерова, Казань, Россия

²Университет Страсбурга, Страсбург, Франция

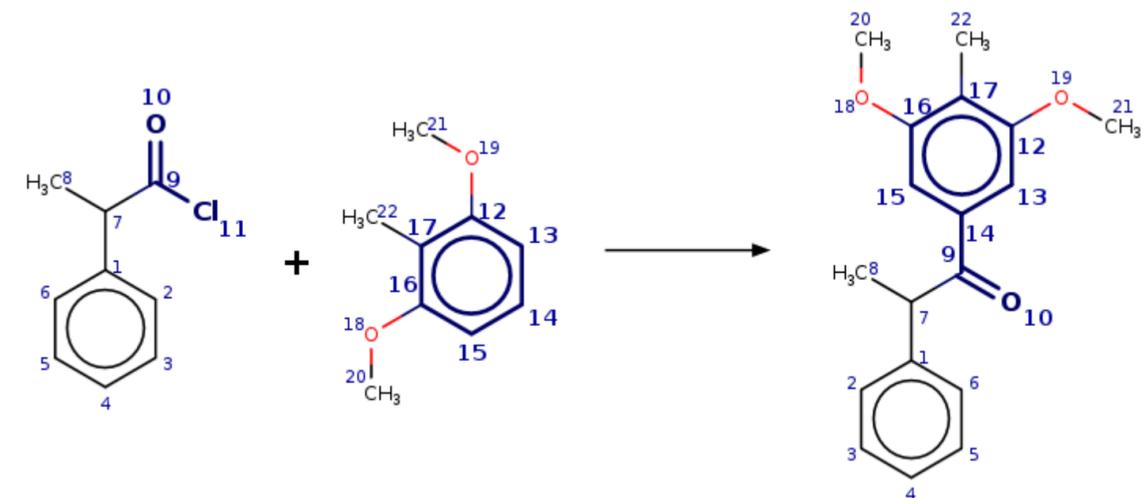
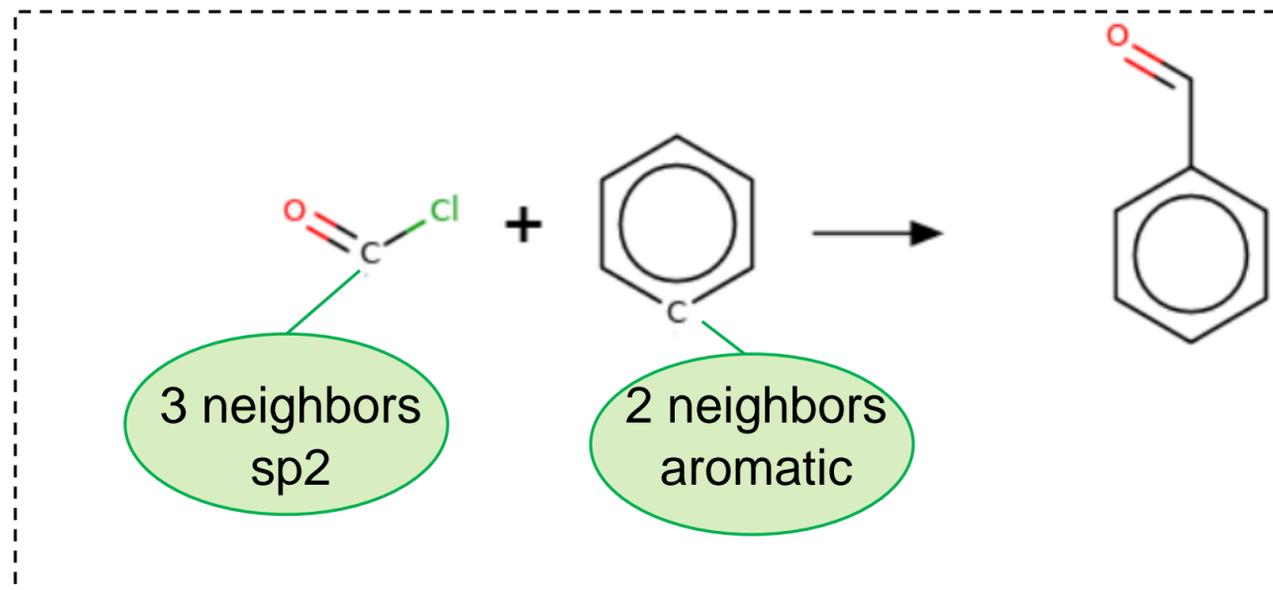
Работа финансировалась в рамках проекта, поддержанного Российским научным фондом, номер 19-73-10137 (Код ГРНТИ 31.01.77)

Правила химических трансформаций

Однокомпонентное правило



Двухкомпонентное правило



В работе используется:

> 550 однокомпонентных правил

> 1700 двухкомпонентных правил