

# АВТОМАТИЧЕСКАЯ РЕКОНСТРУКЦИЯ МЕТАБОЛИЧЕСКИХ ПУТЕЙ БАКТЕРИАЛЬНЫХ СИСТЕМ И ИХ МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Ермак Т.В., Акбердин И.Р., Хлебодарова Т.М., Лихошвай В.А.<sup>1</sup>

Институт цитологии и генетики СО РАН, Новосибирск, Россия, timofei.ermak@live.com

<sup>1</sup>Новосибирский Государственный Университет, Новосибирск, Россия

Мотивация и цель: *Escherichia coli* является одним из наиболее исследованных организмов относительно структурно-функциональной организации генома и механизмов фундаментальных клеточных процессов. Однако до сих пор для данного организма не разработана динамическая модель полного метаболизма с учетом молекулярно-генетического уровня регуляции, так как для этого требуются точные знания о механизмах ферментативных реакций и процессов генетической регуляции, данные о которых, несмотря на множественные исследования, не обладают необходимой полнотой. Такая модель могла бы не только способствовать в решении фундаментальных проблем в области исследования бактериальных метаболических систем, но имела бы широкое применение в более практических сферах для создания штаммов-суперпродуцентов в фармацевтической, агропромышленности, при производстве биотоплива и т.д. В настоящее время задача разработки модели полного метаболизма *E.coli* является одной из ключевых сверхпроблем системной биологии, в рамках которой необходимо разработать ряд новых и усовершенствовать существующие теоретические и экспериментально-теоретические подходы. В частности, ощущается острый недостаток в наличии мощных, высокопроизводительных методов интеграции разнородных данных о метаболитах, ферментативных реакциях, представленных в публичных базах данных, таких как KEGG, Brenda, EcoCyc, BioModels, MGSMODELSDB, KiNET, Sabio-RK и т.д.

Результаты: В рамках данной работы разработаны методы и алгоритмы для интеграции гетерогенных данных, описывающих функционирование метаболических путей *E. coli*, в единую компьютерную систему. Также в систему интегрированы модули, позволяющие производить автоматическую сборку или генерацию динамической математической модели выбранного участка метаболической сети, что позволяет существенно сократить временные и трудовые затраты на создание модели. Тестирование предложенной системы проводили на метаболических путях биосинтеза ароматических аминокислот и нуклеотидов.

Методы и Алгоритмы: В качестве хранилища данных используется графовая СУБД Neo4j. Формат хранения данных в виде графа является более естественным, поскольку структура данных, описывающая взаимодействие фермента, субстрата и продукта биохимической реакции, представляет ничто иное, как граф. Кроме того, что немаловажно, использование графовой СУБД дает прирост производительности при использовании алгоритмов анализа графа, таких как поиск кратчайшего пути между вершинами.