

**РАЗРАБОТКА PYTHON-ПАКЕТА АЛГОРИТМОВ РЕШЕНИЯ ЗАДАЧИ
FLUX VARIABILITY ANALYSIS С ТЕРМОДИНАМИЧЕСКИМИ ОГРАНИЧЕНИЯМИ
Сафонов И.А. (ИТМО)**

Научный руководитель – кандидат технических наук, доцент Сергушичев А.А. (ИТМО)

Введение. Применение вычислительного анализа метаболических моделей является важным способом решения многих практических вопросов в биологии. Задача Flux Variability Analysis (FVA) [1] состоит в определении минимальных и максимальных значений потоков для каждой реакции. В задачу могут быть добавлены термодинамические ограничения [2], запрещающие наличие бесполезных циклов. Такие циклы являются биохимически неосуществимыми, но позволяют получать сколь угодно большие потоки. Задача FVA с ациклическими ограничениями является NP-полной [5]. Однако существуют различные алгоритмы её решения, основанные на решении задач целочисленного линейного программирования (MILP) с помощью солверов. Самые эффективные на данный момент алгоритмы реализованы только в пакете COBRA Toolbox для языка MATLAB. Мы разработали пакет для языка Python, совместимый с библиотекой COBRApy [3], в котором реализованы основные алгоритмы из MATLAB-пакета. Разработанный пакет дает возможность быстро внедрять новые алгоритмы, легко комбинировать их с некоторыми шагами из уже существующих методов, а также сравнивать эффективность и корректность на различных моделях. В будущем планируется интеграция пакета в COBRApy.

Основная часть. При загрузке модели в пакете COBRApy строится задача линейного программирования (LP), в которую добавляют ограничения, заданные в метаболической модели. Для учёта ациклических ограничений в эту LP-задачу можно добавить дополнительные ограничения (в том числе с целочисленными переменными). Алгоритмы решения FVA предлагают различные способы формировать такие ограничения. Полученная задача целочисленного линейного программирования (MILP) затем оптимизируется с различными целевыми функциями.

В разработанном пакете доступны следующие основные алгоритмы добавления ациклических ограничений:

1. **II-COBRA** [4]. Этот алгоритм уже реализован в библиотеке COBRApy и служит базовым способом задать ациклические ограничения. В нем вводятся вспомогательные ненулевые переменные G для каждой реакции, знак которых противоположен знакам самих реакций. При умножении вектора G на матрицу базиса ядра внутренней стехиометрической матрицы должен получаться нулевой вектор. Для нахождения базиса ядра используется SVD.
2. **Fast-tFVA** [5]. В этом алгоритме не требуется искать базис ядра матрицы. Вместо этого вектор G задается как произведение внутренней стехиометрической матрицы и вектора потенциалов метаболитов. Благодаря разреженной структуре итоговая MILP-задача решается быстрее по сравнению с II-COBRA.
3. **Fast-SNP** [6]. В данном алгоритме используется подход II-COBRA с двумя дополнительными оптимизациями:
 - Поиск разреженного базиса ядра матрицы.
 - Определение подмножества циклических реакций, которые потенциально могут участвовать в цикле. На практике таких реакций меньше, чем общее количество внутренних реакций, поэтому ограничивать нужно лишь их, что сокращает размер MILP-задачи.
4. **LLC-NS** [7]. Этот алгоритм добавляет к Fast-SNP предварительный поиск разбиения множества циклических реакций на независимые подмножества, так что любой возможный цикл полностью лежит в одном из них. Это позволяет ограничивать только

нужное подмножество в зависимости от оптимизируемой реакции, тем самым уменьшая размер MILP-задачи.

На данный момент наиболее эффективным (state-of-the-art) алгоритмом считается Fast-SNP с примененной оптимизацией LLC-NS.

Было проверено, что время работы MATLAB-версий алгоритмов приблизительно совпадает со временем работы разработанных Python-версий.

В пакете возможно использование всех MILP солверов, поддерживаемых в библиотеке COBRApy optlang, например, CPLEX и Gurobi.

Важным свойством разработанного пакета стала возможность автоматически комбинировать различные оптимизации из разных алгоритмов с помощью задания соответствующих флагов при запуске. Выяснилось, что при использовании метода Fast-tFVA, который ограничивает только циклические реакции (идея из Fast-SNP), и одновременно ограничивает лишь одну из независимых по циклам групп реакций (идея из LLC-NS), удаётся ускорить вычисления примерно на 20% на большинстве тестируемых моделей.

Выводы. С помощью разработанного пакета основные state-of-the-art алгоритмы решения задачи FVA с ациклическими ограничениями стали доступны при работе с библиотекой COBRApy на языке Python. Кроме того, пакет дает возможность создавать новые алгоритмы, автоматически используя дополнительные оптимизации из уже существующих передовых методов и сравнивая результаты с ними.

Список использованных источников:

1. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat Biotechnol* 28, 245–248 (2010).
2. Beard, D. A., Babson, E., Curtis, E. & Qian, H. Thermodynamic constraints for biochemical networks. *Journal of Theoretical Biology* 228, 327–333 (2004).
3. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology* 7, 74 (2013).
4. Schellenberger, J., Lewis, N. E. & Palsson, B. Ø. Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models. *Biophysical Journal* 100, 544–553 (2011).
5. Müller, A. C. & Bockmayr, A. Fast thermodynamically constrained flux variability analysis. *Bioinformatics* 29, 903–909 (2013).
6. Saa, P. A. & Nielsen, L. K. Fast-SNP: a fast matrix pre-processing algorithm for efficient loopless flux optimization of metabolic models. *Bioinformatics* 32, 3807–3814 (2016).
7. Chan, S. H. J., Wang, L., Dash, S. & Maranas, C. D. Accelerating flux balance calculations in genome-scale metabolic models by localizing the application of loopless constraints. *Bioinformatics* 34, 4248–4255 (2018).