

Im Rahmen des 5-jährigen, kooperativen Forschungsprojektes „omics-integrated pathway analysis in metabolic networks“ zwischen der FH Wiener Neustadt @ Campus Tulln und des Austrian Centres of Industrial Biotechnology gefördert durch das FFG-COMET Programm wird der Ansatz verfolgt, zelluläre metabolische Prozesse systematisch durch die Kombination von elementaren funktionellen Einheiten zu verstehen. Mathematisch lassen sich solche funktionellen Einheiten definieren und es wurde gezeigt, dass diese Einheiten, genannt elementare Flußmoden (EFM), biologisch relevant sind. Insbesondere kann gezeigt werden, dass eine Zelle nur dann eine bestimmte Funktion erfüllen kann, wenn sie auch über mindestens einen entsprechenden EFM verfügt. Umgekehrt bildet die Gesamtheit aller EFMs das mögliche Funktionsspektrum einer Zelle ab und erlaubt daher Einblicke z.B. in funktionelle Zustände von Krankheiten. Man kann aber auch versuchen durch gezielte Elimination von Funktionen bestimmte (unerwünschte) Eigenschaften, wie z.B. die Produktion von unerwünschten Nebenprodukten, zu unterdrücken. Dies ist besonders für das computergestützte Design von Zellfabriken interessant. Die Identifikation dieser elementaren, funktionellen Einheiten ist aber rechentechnisch sehr ressourcenintensiv. Aktuell können daher nur mittelgroße Systeme mit bis zu 100 biochemischen Reaktionen analysiert werden. Genom-weite metabolische Modelle die heutzutage bis zu 7,000 biochemische Reaktionen enthalten können, sind aber derzeit außer Reichweite. Ziel dieses Forschungsprojektes, das noch bis Ende 2021 läuft, ist es daher neue Methoden zu entwickeln, die es uns auch erlauben, größere Modelle zu analysieren und diese Information für das Design von optimalen Zellfabriken zu verwenden. Erste theoretische Erfolge konnten in den letzten Jahren bereits publiziert werden [1-5].

[1] Gerstl et al., Which sets of elementary flux modes form thermodynamically feasible flux distributions? *The FEBS Journal* **283**:1782 (2016)

[2] Zanghellini et al., Toward Genome-Scale Metabolic Pathway Analysis. *Industrial Biotechnology: Microorganisms*, 111 (2017)

[3] Klamt et al., From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. *PLoS Computational Biology* **13**:e1005409 (2018)

[4] Klamt et al., A mathematical framework for yield (versus rate) optimization in constraint-based modeling and applications in metabolic engineering. *Metabolic Engineering* **47**:153 (2018)

[5] Gerstl et al., Flux tope analysis: studying the coordination of reaction directions in metabolic networks. *Bioinformatics*, in press (2018)

Titel: Massiv parallele Analyse minimaler Stoffwechselwege im zellulären Metabolismus

Einleitung: Die (mathematische) Charakterisierung des zellulären Stoffwechsels durch elementare Flußmoden (EFMs) ist ein Schlüsselproblem im Bereich der Systembiologie. EFMs sind minimale Funktionseinheiten eines metabolischen Netzwerks im Fließgleichgewicht. Nur dann, wenn eine Zelle über einen entsprechenden EFM verfügt, kann auch die dazugehörige Funktion ausgeführt werden. Die Gesamtheit der EFMs beschreibt daher vollständig das metabolische Funktionsspektrum einer Zelle [1] und erlaubt damit z.B. wichtige Einblicke in funktionelle Zustände von Krankheiten oder Vorhersagen zum optimalen Aufbau von mikrobiellen Zellfabriken. Die Analyse aller EFMs ist jedoch extrem ressourcenintensiv. Derzeit können nur kleinere und mittlere metabolische Modelle mit bis zu 100 biochemischen Reaktionen vollständig berechnet werden, nicht aber genom-weite Modelle, die bis zu 7,000 biochemische Reaktionen enthalten. In diesem fünfjährigen, kollaborativen Projekt zwischen der FH Wiener Neustadt @ Campus Tulln und dem Austrian Centre of Industrial Biotechnology, einem FFG geförderten K2-Zentrum, versuchen wir neue Methoden zur Analyse genom-weiter Modelle zu entwickeln.

Methoden: Mathematisch ist das Problem der EFM-Bestimmung identisch mit der Aufzählung von Ecken und Kanten in einem Polyeder für die sich zwei Ansätze, die „double description“ Methode (DDM) und die „lexicographic reverse search“ (lrs), als besonders nützlich erwiesen haben. Hier untersuchen und vergleichen wir zwei verschiedene Implementierungen (efmtool [2] und mplrs [3]) dieser Algorithmen in Hinblick auf ihre rechnerische Performance und ihre Eignung für systembiologische Anwendungen.

Ergebnisse: Wir zeigen, dass die Aufzählung von EFMs, entgegen der gängiger Lehrmeinung, auch mit lrs basierten Methoden möglich ist. Insbesondere dann, wenn die Möglichkeiten (mittel)großer Rechenclustern ausgenützt werden, und das metabolische Netzwerk zuvor stark komprimiert wird. Im speziellen können wir zeigen, dass mplrs nahezu ideales Parallelisierungsverhalten aufweist und – im Gegensatz zu DDM basierten Methoden – zu keinem exponentiell wachsenden Speicherverbrauch führt (siehe Abbildung 1). Diese Ergebnisse sind unabhängig vom zugrundeliegenden metabolischen Modell, wie wir durch Vergleich verschiedener Modellen für *E. coli*, *M. genitalium* und *B. cenocti* demonstrieren.

Diskussion/Conclusio: mplrs gemeinsam mit effizienten Komprimierungsmethoden ist eine attraktive Alternative zu aktuellen Standardwerkzeugen wie das efmtool. Aufgrund des niedrigen Speicherbedarfs und des idealen Skalierungsverhalten kommt damit erstmals die Zerlegung von genom-weiten metabolischen Modellen in ihre funktionellen Bestandteile mit aktueller high-performance computing Infrastruktur in Reichweite, welches beispiellose Anwendungsmöglichkeiten eröffnet.

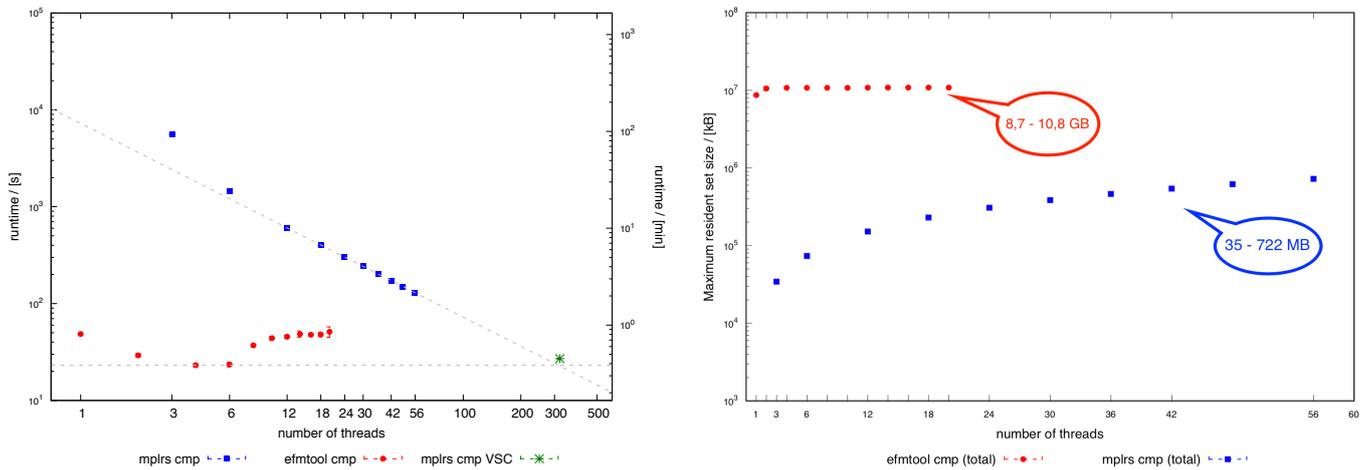


Abbildung 1: Leistungsvergleich zwischen efmtool und mprls in *E. coli*. Bild links, Laufzeit als Funktion der Rechenkern für efmtool (rot) und mprls (blau). Der Schnittpunkt der gestrichelten Linien (320 threads und 23 Sekunden) zeigt den erwarteten Punkt, an dem beide Methoden gleich schnell sind. Der grüne Stern zeigt die tatsächlich erreichte Laufzeit des mprls. Bild rechts, Hauptspeicherverbrauch als Funktion der Funktion der Rechenkern für efmtool (rot) und mprls (blau).

Quellen:

- [1] Zanghellini J, et al. 2013. Elementary flux modes in a nutshell: properties, calculation and applications. *Biotechnol J* **8**:1009
- [2] Terzer M and Stelling J. 2008. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* **24**:2229
- [3] Avis D and Jordan C. 2017. mprls: A scalable parallel vertex/facet enumeration code. *Math Program Comput* **10**:267