

Datenaustausch und Datenintegration zur Modellierung und Analyse metabolischer Netzwerke am Beispiel von Kulturpflanzen

Stephan Weise, Christian Colmsee, Eva Grafahrend-Belau, Björn Junker,
Christian Klukas, Matthias Lange, Uwe Scholz und Falk Schreiber

Leibniz Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK),
Corrensstr. 3, 06466 Gatersleben, Germany
{weise,colmsee,grafahr,junker,klukas,lange,
scholz,schreibe}@ipk-gatersleben.de

Abstract: Die Rekonstruktion und Analyse detaillierter metabolischer Modelle bildet eine wichtige Grundlage für das Verständnis komplexer biologischer Prozesse in Organismen. Um dieses Ziel zu erreichen, wurde eine Pipeline etabliert, die Software-Werkzeuge miteinander verbindet um (1) metabolische Netzwerkdaten zu speichern und metabolische Modelle zu rekonstruieren, (2) stöchiometrische und kinetische Modelle zu simulieren und zu analysieren und (3) die von den Modellen generierten Daten zu visualisieren. Die Anwendbarkeit dieses Ansatzes wird anhand einer Fallstudie zum Metabolismus von Kulturpflanzen gezeigt.

1 Einleitung

Mathematische Modellierung ermöglicht die Analyse der Struktur, der Dynamik und des Verhaltens metabolischer Netzwerke. Mithilfe solcher Modelle kann das Wissen über komplexe biologische Vorgänge überprüft und erweitert werden. Die Generierung neuer Hypothesen wird ebenso ermöglicht wie die Identifikation geeigneter Ziele für das Metabolic Engineering. Für die Verwendung solcher Modelle müssen verschiedene Aspekte berücksichtigt werden: Repräsentation von Modellen, Modellaustausch, Modellanalyse und Visualisierung von Modellen.

Daten zu biologischen Prozessen stammen aus verschiedenen, oftmals widersprüchlichen Quellen und müssen in einem zentralen, gut strukturierten Repository integriert und persistent gespeichert werden. Um hohe Datenqualität sicherzustellen, sollte eine manuelle Kuratierung der integrierten Daten erfolgen. Die aufbereiteten Daten können mit verschiedenen Werkzeugen zur Visualisierung, Simulation und Analyse verwendet werden. Dafür sollten standardisierte Austauschformate wie die Systems Biology Markup Language (SBML) oder die Biological Pathways Exchange Language (BioPax) eingesetzt werden.

In Abhängigkeit von der Zielstellung und den zur Modellrekonstruktion verwendeten Daten können verschiedene Analysetechniken auf die jeweiligen Modelle angewandt werden. Das Spektrum reicht dabei von der reinen stöchiometrischen Analyse bis hin zu kineti-

schen Ansätzen. Jeder dieser Ansätze ist mit Vor- und Nachteilen verbunden. Daher ist eine Analyse-Pipeline erforderlich, die verschiedene Techniken sowohl zur nutzerfreundlichen als auch flexiblen Modellsimulation und -analyse zur Verfügung stellt.

Dieser Artikel beschreibt eine Pipeline zur Unterstützung der Forschung auf dem Gebiet metabolischer Modelle. Sie umfasst die Schritte: (1) Datenmanagement und individuelle, nutzerspezifische Modellrekonstruktion, (2) stöchiometrische und kinetische Modellanalyse sowie (3) Modell- und Fluss-Visualisierung. Jeder dieser Schritte wird anhand von Werkzeugen und Methoden beschrieben. Die Anwendbarkeit der Pipeline wird anhand einer Fallstudie über den Speichermetabolismus der Kulturpflanze Gerste demonstriert.

2 Modellspeicherung und -rekonstruktion

Die Rekonstruktion von Modellen des Stoffwechsels ist ein wichtiger Schritt auf dem Weg zu einem besseren Verständnis biologischer Prozesse. Dazu müssen detaillierte metabolische Informationen bis auf Kompartimentebene erfasst und wohl strukturiert verwaltet werden. Hierfür wurde MetaAll (<http://meta-all.ipk-gatersleben.de>) [W⁺06] als Repository für Stoffwechselprozesse und für Kulturpflanzen das abgeleitete System MetaCrop (<http://metacrop.ipk-gatersleben.de>) [GB⁺08] entwickelt. Dabei handelt es sich um ein manuell kuriiertes Repository für qualitativ hochwertige Daten über sieben Kulturpflanzen mit hoher Bedeutung: *Hordeum vulgare* (Gerste), *Triticum aestivum* (Weizen), *Oryza sativa* (Reis), *Zea mays* (Mais), *Beta vulgaris* (Rübe), *Solanum tuberosum* (Kartoffel) und *Brassica napus* (Raps) und zwei Modellpflanzen: *Arabidopsis thaliana* (Acker-Schmalwand) und *Medicago truncatula* (barrel medic).

Zu jedem biochemischen Umwandlungsprozess (Reaktionen, Translokationen) können in MetaCrop feingranulare Informationen, beispielsweise der Reaktions- oder Translokationstyp, die Reaktionsformel, Synonyme, EC-Nummern, Literaturreferenzen und auch kinetische Daten wie V_{max} -Werte, Affinitäts- oder Inhibitorkonstanten abgespeichert werden. Weiterhin sind all diese Informationen dem Kompartiment zugeordnet, in dem der jeweilige Prozess stattfindet. Netzwerke können als verschiedene parallele Versionen abgespeichert und die einzelnen Elemente können darüber hinaus mit verschiedenen Qualitäts-Tags versehen werden.

Da bisher nur sehr wenige Modelle für Kulturpflanzen existieren, ist der zeitintensive Prozess der manuellen Datenkuration unabdingbar. Daten, die im standardisierten SBML-Format [H⁺03] vorliegen, können in MetaCrop importiert werden; andere Daten müssen manuell erfasst werden. Zur Kuration steht eine Web-Schnittstelle zur Verfügung. Um in MetaCrop integrierte Daten zu visualisieren oder zu analysieren, können Daten im SBML-Format exportiert werden. Anwender des Systems können individuelle Netzwerkmodelle erstellen, indem Elemente (Reaktionen, Translokationen oder auch ganze Pathways) in einem Shopping Cart abgelegt werden. Wizard-gestützt können anschließend Exporteinstellungen vorgenommen werden, wie beispielsweise die Verwendung bestimmter Kinetiken oder die Einschränkung, ausschließlich Daten einer bestimmten Spezies zu verwenden.

3 Modellanalyse und -simulation

Stöchiometrische Modellanalyse oder constraintbasierte Modellierung basiert auf dem Wissen über topologische Strukturen metabolischer Netzwerke. Da hierbei kein Wissen über die Enzymkinetik erforderlich ist, sind constraintbasierte Modellierungsmethoden wie beispielsweise Flux-Balance-Analysis (FBA) zu einem wichtigen Ansatz geworden, um die Möglichkeiten und Eigenschaften metabolischer Netzwerke zu verstehen.

Als Werkzeug für die constraintbasierte Analyse metabolischer Modelle wurde FBA-SimVis (<http://fbasimvis.ipk-gatersleben.de>) entwickelt. Dabei handelt es sich um ein VANTED-Plug-in ([JKS06], vgl. Abschnitt 4) für integrierte constraintbasierte Modellanalyse und -visualisierung. Das Plug-in stellt verschiedene constraintbasierte Analysetechniken (Flux-Balance-Analysis, Knock-out-Analysis, Robustness-Analysis und Flux-Variability-Analysis) und interaktive Visualisierungsroutinen zur Verfügung, um die quantitative Analyse von stöchiometrischen Modellen des Metabolismus zu unterstützen. FBA-SimVis ermöglicht ein umfassendes Verständnis constraintbasierter metabolischer Modelle. Zur Durchführung einer constraintbasierten Analyse eines aus MetaCrop exportierten Modells wird dieses im SBML-Format geladen.

Kinetische Modellanalyse Kinetische Modelle sind sehr detailliert und ermöglichen feingranulare Vorhersagen über metabolisches Verhalten. Jedoch erfordern diese Modelle detaillierte kinetische Kenntnisse über die modellierten Enzyme und Berechnungen erreichen schnell die Grenzen der bezahlbaren Rechenleistung. Weiterhin ist die Etablierung kinetischer Modelle ein langwieriger Prozess, welcher die Erfassung verschiedener Typen von detaillierten Informationen erforderlich macht. Die aus MetaCrop exportierten SBML-Dateien enthalten alle erforderlichen Informationen zur Erstellung eines kinetischen Modells und zur Verwendung mit verschiedenen Simulationswerkzeugen. Beispielsweise sei COPASI [H⁺06a] genannt, ein GUI-basiertes Werkzeug, welches die einfache Bearbeitung und Simulation eines Modells ermöglicht. Ein importiertes Modell kann vom Benutzer bearbeitet und weiterentwickelt werden, sodass auch Daten hinzugefügt werden können, die nicht in MetaCrop enthalten sind. COPASI ermöglicht es dann, das Modell als Zeitverlauf zu simulieren, ein Flussgleichgewicht zu berechnen und dessen Stabilität zu analysieren, auf experimentellen Daten beruhende Parameterschätzungen durchzuführen, metabolische Kontrollanalysen auszuführen usw. Die berechneten Resultate werden in Textdateien gespeichert, welche in Tabellenkalkulationsprogrammen geöffnet und, wie im nachfolgenden Abschnitt beschrieben, an Visualisierungswerkzeuge übermittelt werden können. Weiterhin können Daten in das SBML-Format exportiert und in MetaCrop zum Update verwendet werden.

4 Modellvisualisierung

VANTED [JKS06] ist eine plattformunabhängige Open-Source-Software, die es ermöglicht, sowohl experimentelle Daten mehrerer *-omics*-Gebiete (Transcriptomics, Proteomics, Me-

tabolomics) als auch Resultate aus Simulationsstudien umfassend zu evaluieren. Zur Analyse und Visualisierung metabolischer Flux-Daten wird die Daten-Mapping-Methode verwendet. Diese wurde ursprünglich dafür entwickelt, experimentelle Daten Netzwerkelementen zuzuordnen. Flexible Visualisierungstechniken können zur Darstellung von Daten verwendet werden, die mit den Netzwerkknoten oder -kanten verbunden sind. Die verfügbaren Methoden umfassen Linien-, Balken- und Kreisdiagramme ebenso wie Umwandlungsfunktionen, die Daten auf visuelle Eigenschaften, beispielsweise Knotengröße, Kantenbreite oder Graphenelementfarben, abbilden. Zum flexiblen Austausch von Modellen mit verschiedenen Softwarewerkzeugen und Datenquellen werden verschiedene Datenformate wie GML, GraphML, DOT, KGML und SBML unterstützt.

5 Fallstudie

Um ein systematisches Verständnis des Speichermetabolismus in Getreidekörnern zu entwickeln, wurde ein stöchiometrisches Modell des zentralen Metabolismus im sich entwickelnden Endosperm der Gerste (*Hordeum vulgare*) rekonstruiert. Dazu wurden biochemische, physiologische und proteomische Daten, die aus der Literatur und aus öffentlichen Datenbanken abgeleitet wurden, in MetaCrop integriert. Das daraus resultierende kompartimentierte stöchiometrische Modell beinhaltet 257 Umwandlungsprozesse (193 Reaktionen, 64 Transportprozesse) und 234 Metabolite in den intrazellulären Kompartimenten Cytosol, Mitochondrium und Plastid.

Um Kornertag und metabolische Fluss-Verteilungen unter sich ändernden Wachstumsbedingungen zu studieren, wurde das Modell mit FBA-SimVis untersucht. Die zur FBA erforderlichen Parameter (z. B. maximale Aufnahme- und Absonderungsraten) wurden aus der Literatur extrahiert und das Modell wurde durch den Vergleich der Simulationsergebnisse mit literaturbasierten Resultaten validiert. Im Allgemeinen zeigte sich bei den Simulationsergebnissen eine gute Übereinstimmung mit den wichtigsten qualitativen physiologischen Eigenschaften des Speichermetabolismus. Beispielsweise stimmten die vom Modell vorhergesagte Wachstumsrate und die metabolische Netzwerkstruktur unter vollständigen aeroben Wachstumsbedingungen mit publizierten experimentellen Resultaten überein (siehe Abbildung 1 für ein Beispiel). Dieses Modell kann in zukünftigen *In-silico*-Studien verwendet werden, um Hypothesen zur Verbesserung der Korn- und Ertragsqualität aufzustellen und zu testen.

6 Zusammenfassung

In diesem Artikel wurde eine Pipeline zur Integration und Analyse von Daten über Stoffwechselwege an einem Beispiel aus der Kulturpflanzenforschung vorgestellt. Diese umfasst die Schritte der Modellspeicherung und -rekonstruktion, Modellanalyse sowie Modellvisualisierung durch die Verwendung von MetaCrop, FBA-SimVis, COPASI und VANTED. Alle eingesetzten Werkzeuge sind öffentlich verfügbar und die vorgestellte Metho-

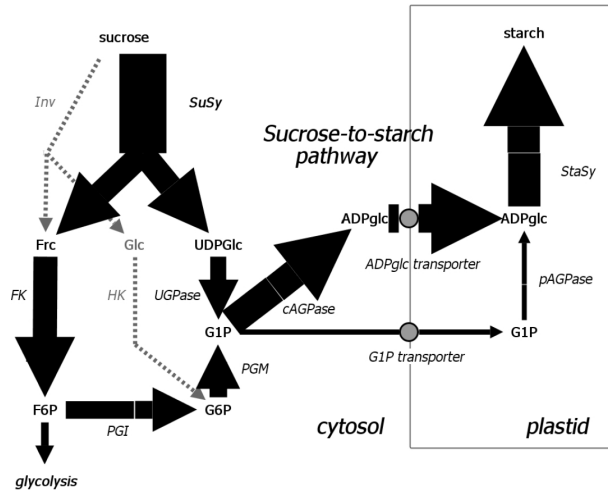


Abbildung 1: Carbon-Flux-Map des Sucrose-to-Starch-Pathways eines Getreidekorns unter vollständig aeroben Wachstumsbedingungen (visualisiert mit VANTED)

dik lässt sich auch auf Stoffwechselmodelle aus der Medizin anwenden.

Literatur

- [GB⁺08] E. Grafahrend-Belau et al. MetaCrop: a detailed database of crop plant metabolism. *NAR*, 36(DB issue):D954–D958, 2008.
- [H⁺03] M. Hucka et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinf.*, 19(4):524–531, 2003.
- [H⁺06a] S. Hoops et al. COPASI—a COMplex PATHway SIMulator. *Bioinf.*, 22(24):3067–3074, 2006.
- [H⁺06b] D. Hull et al. Taverna: a tool for building and running workflows of services. *NAR*, 34(Webserver issue):W729–W732, 2006.
- [JKS06] B. Junker, C. Klukas und F. Schreiber. VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinf.*, 7(1):e109.1–13, 2006.
- [KPE03] K.J. Kauffman, P. Prakash und J.S. Edwards. Advances in flux balance analysis. *Cur. Op. Biotech.*, 14(5):491–496, 2003.
- [M⁺09] T. McPhillips et al. Scientific workflow design for mere mortals. *Future Generation Computer Systems*, 25(5):541–551, 2009.
- [VP94] A. Varma und B.Ø. Palsson. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Nature Biotech.*, 12(10):994–998, 1994.
- [W⁺06] S. Weise et al. Meta-All: a system for managing metabolic pathway information. *BMC Bioinf.*, 7(1):e465.1–9, 2006.