



Doctoral Thesis

Concepts and methods for the automated reconstruction of genome-scale metabolic network models

Author(s):

Ganter, Mathias

Publication Date:

2012

Permanent Link:

<https://doi.org/10.3929/ethz-a-007604552> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH No. 20773

Concepts and Methods for the Automated Reconstruction of Genome-Scale Metabolic Network Models

A dissertation submitted to
ETH Zurich

for the degree of
Doctor of Sciences

presented by
Mathias Ganter

Diplom-Informatiker (Bioinformatik), Eberhard Karls Universität Tübingen

born 5 December 1981
citizen of Germany

accepted on the recommendation of
Prof. Dr. Jörg Stelling, examiner
Prof. Dr. Uwe Sauer, co-examiner
Prof. Dr. Dennis Vitkup, co-examiner

2012

Abstract

Systems biology, enabled by the availability of completely sequenced genomes and new high-throughput technologies, studies complex molecular interactions within biological systems to acquire a systems-level understanding of them. In this context, genome-scale metabolic models represent cellular networks by the interactions between genes, proteins, and biochemical reactions. These models allow predictions of, for example, the maximal growth of a specific organism under various environmental conditions and of the identification of drug targets in human pathogens. However, model reconstruction is challenging and still relies on time-consuming manual curation. Here we present novel concepts and methods for the automated reconciliation and reconstruction of metabolic models. An essential characteristic of these methods is the explicit use of the network structure, primarily defined by metabolic reactions.

We first introduce the concept of structural network consistency and show its applicability for the evaluation and reconciliation of existing models from all three kingdoms of life as well as for the *de novo* generation of draft models for a new species. Our results demonstrate that a network consistency score can assess a model's quality solely based on the network structure and that large fractions of existing models are highly inconsistent. Therefore, an automated method for model consolidation is introduced and proved successful by independent validation using single-gene knockouts. Additionally, we demonstrate the applicability of structural network consistency for model transfer, that is, the generation of a model for a new species based on an existing predictive model.

The second concept yields a constructive approach for analyzing biological network features based on reaction motifs, basic building blocks of these networks. We show how reaction motifs relate to probabilistic models and how to predict metabolic features by systematically exploiting the knowledge encoded in existing models. To this end, we mainly focus on the prediction of reaction directions and gap-filling. Our results reveal that our method for reaction direction prediction proves more accurate than other available computational methods. In addition, our gap-filling method predicts novel candidate reactions for an existing model.

Finally, we present a method for the automatic compartmentalization of a given genome-scale metabolic model. For this, we predict protein subcellular localizations by integrating experiments and predictions with higher accuracy than existing methods. We then predict metabolite transporters based on an extensive feature collection describing transportable compounds. Finally, we integrate these predictions with the network structure and compute efficiently an optimal compartmentalization (using graph cuts). By applying this framework to the model plant *Arabidopsis thaliana*, we show that the high-throughput generation of compartmentalized eukaryotic models is possible.

Zusammenfassung

Systembiologie, durch die Verfügbarkeit von vollständig sequenzierten Genomen und neuen Hochdurchsatz-Technologien ermöglicht, untersucht komplexe molekulare Zusammenhänge innerhalb biologischer Systeme, um sie in ihrer Gesamtheit zu verstehen. In diesem Zusammenhang repräsentieren genomweite metabolische Modelle zelluläre Netzwerke durch die Wechselwirkungen von Genen, Proteinen und biochemischen Reaktionen. Diese Modelle ermöglichen zum Beispiel Vorhersagen über das maximale Wachstum eines bestimmten Organismus unter verschiedenen Umweltbedingungen und die Identifizierung von Angriffspunkten von Medikamenten in menschlichen Krankheitserregern. Allerdings ist die Modellrekonstruktion anspruchsvoll und immer noch auf zeitraubende manuelle Sichtung und Korrektur angewiesen. In dieser Arbeit präsentieren wir neue Konzepte und Methoden für den automatischen Abgleich und die automatische Rekonstruktion von metabolischen Modellen. Ein wesentliches Merkmal dieser Methoden ist die explizite Verwendung der Netzwerkstruktur, die vor allem durch Stoffwechselreaktionen definiert ist.

Wir führen zunächst das Konzept der strukturellen Netzwerkkonsistenz ein und zeigen ihre Anwendbarkeit für die Auswertung und den Abgleich von bestehenden Modellen aus allen drei Reichen des Lebens als auch für die Generierung von neuen Entwurfsmodellen für eine neue Spezies. Unsere Ergebnisse zeigen, dass ein Netzwerkkonsistenzmaß die Qualität eines Modells ausschließlich auf der Grundlage der Netzwerkstruktur bewerten kann und dass große Teile von bestehenden Modellen in sich widersprüchlich sind. Daher wurde eine automatisierte Methode für die Modellkonsolidierung eingeführt, die sich durch die unabhängige Bewertung durch einzelne Gen-Knockouts als erfolgreich herausstellte. Darüber hinaus zeigen wir die Anwendbarkeit der strukturellen Netzwerkkonsistenz für die Modellübertragung, das heißt die Erzeugung eines Modells für eine neue Spezies, basierend auf einem vorhandenen Modell, mit dem man bereits Vorhersagen machen konnte.

Das zweite Konzept liefert einen konstruktiven Ansatz zur Analyse biologischer Netzwerkeigenschaften basierend auf Reaktionsmotiven, die Grundbausteine dieser Netzwerke beschreiben. Wir zeigen, wie Reaktionsmotive zu Wahrscheinlichkeitsmodellen in Beziehung stehen und wie man metabolische Funktionen durch das systematische Ausnutzen des Wissens in bestehenden Modellen vorhersagen kann. Zu diesem Zweck konzentrieren wir uns in erster Linie auf die Vorhersage von Reaktionsrichtungen und dem Schließen von fehlenden Netzwerkverbindungen. Unsere Ergebnisse zeigen, dass unsere Methode zur Vorhersage von Reaktionsrichtungen genauer ist als andere rechnergestützte Methoden. Darüber hinaus sagt unsere Methode zum Schließen von fehlenden Netzwerkverbindungen neuartige Kandidatenreaktionen für ein bestehendes Modell vorher.

Zuletzt präsentieren wir eine Methode zur automatischen Umwandlung eines gegebenen genomweiten metabolischen Modells in ein Modell mit membranbegrenzten Organellen. Hi-

erfür sagen wir die subzelluläre Lokalisationen von Proteinen mit höherer Genauigkeit als bestehende Verfahren vorher, indem wir die Ergebnissen von Experimenten und Vorhersagen verflechten. Als nächstes sagen wir Transporter für Metabolite vorher. Hierzu verwenden wir eine umfangreiche Sammlung von Charakteristiken, die einen Metaboliten beschreiben, der über Membranen transportiert wird. Schließlich integrieren wir diese Vorhersagen mit der Netzwerkstruktur und berechnen effizient eine optimale Aufteilung in Organelle (durch Anwendung von "Graph Cuts"). Durch die Anwendung dieses Verfahrens auf die Modellpflanze *Arabidopsis thaliana* zeigen wir, dass die Hochdurchsatz-Erzeugung von eukaryotischen Modellen unter Berücksichtigung von Organellen möglich ist.