

Siegfried Schloißnig  
Dr. sc. hum.

## **Data Management, Storage and Analytics for High-Throughput Screenings**

Geboren am 10.07.1979 in Klagenfurt, Österreich.

Diplom der Fachrichtung Informatik am 27.10.2005 an der Universität Karlsruhe.

Promotionsfach: DKFZ (Deutsches Krebsforschungszentrum)

Doktorvater: Prof. Dr. Thomas Wetter

Biologie ist komplex. Eine Unmenge an Teilen verbunden durch komplizierte Mechanismen stellen eine große Hürde bei der Analyse biologischer Systeme dar. Im Laufe der letzten Jahre führte dies zu der Entwicklung von neuartigen analytischen und experimentellen Ansätzen, welche dieser Komplexität Herr werden sollen. Dieser Trend ist zur Zeit am besten sichtbar in einer Welle an sogenannten Hochdurchsatzverfahren, die parallel in verschiedenen Bereichen der Biologie und Chemie entwickelt wurden.

Das Problem bei der Anwendung von Hochdurchsatzverfahren ist die enorme Menge an Daten, die dabei in relativ kurzer Zeit produziert werden. Manchmal kann dieses Problem teilweise ignoriert werden, falls die anfallenden Daten nicht langlebig sein sollen und vielleicht nur als Eingabe für weitere Experimente dienen sollen. Aber selbst hier sind oft analytische Ansätze notwendig, um die Daten auszuwerten.

Informatik und deren Rolle in der Glykobiologie, ein Forschungsgebiet das sich mit den Eigenschaften und Funktionen von Kohlenhydraten beschäftigt, ist der Fokus des ersten Teils dieser Arbeit. In diesem Rahmen wurde eine neue Datenbank zur Speicherung von Lektin-Kohlenhydrat Interaktionsdaten, die von verschiedenen Laboratorien mittels diverser Ansätze gemessen wurden, durch den Autor entwickelt. Die erstmalige Verfügbarkeit einer Sammlung von gepflegten Daten dieser Art, ermöglichte den Autor weiterhin die Entwicklung eines algorithmischen Ansatzes, basierend auf der Fragmentierung von Kohlenhydraten, zur Entdeckung von potentiellen Kohlenhydrat-Epitopen, die eine Rolle in der Bindung mit Lektinen spielen. Probleme bei der Integration und Aufbereitung der durch die Laboratorien bereit gestellten Forschungsdaten zeigten schnell auf, dass Softwaresysteme zur Verwaltung von Forschungsdaten (und des Arbeitsablaufes im Labor) eine Notwendigkeit in Hochdurchsatz-Umgebungen sind, falls Daten von hoher Qualität produziert werden sollen.

Ein Labor-Informationen- und Management-System (LIMS) das sich der selben Probleme, die der erste Teil dieser Arbeit bei der Integration von Daten aufzeigte, annimmt, wurde durch den Autor entwickelt und im Rahmen dieser Arbeit vorgestellt. Diese Systeme werden von immer größerer Bedeutung mit dem zunehmenden Einsatz von Hochdurchsatzverfahren. Ein LIMS ist einerseits ein unterstützendes System bei der Verwaltung und Organisation der täglichen Abläufe im Labor, aber auch eine wichtige Sammelstelle für alle anfallenden experimentellen Daten und stellt dadurch einen zentralen Punkt dar, an dem Standards für die Datensammlung durchgesetzt werden können. Das vorgestellte System wurde durch den Autor für das Max Planck Institut für molekulare Zellbiologie und Genetik entwickelt, verwaltet

sämtliche experimentellen Abläufe bei der Erstellung von transgenen BAC Zelllinien und deren Analyse durch diverse analytische Verfahren.

Experimentelle Daten müssen nach deren Sammlung in der höchsten Qualität und mit dem geringsten Pflegeaufwand verfügbar gemacht werden. Im erste Teil dieser Arbeit werden die gängigen Problem, die bei der Integration von großen Datenmengen anfallen, behandelt und es wurde gezeigt, dass diese Probleme meist auf fehlende Sorgfalt bei der Datensammlung durch die Labore zurückzuführen sind. Dies führt unweigerlich dazu, dass sobald Hochdurchsatzverfahren eingesetzt werden, ein unterstützendes Softwaresystem notwendig ist, um die Arbeit im Labor zu organisieren, Unterstützung bei der Sammlung und Auswertung der Daten zu bieten und schlussendlich die Daten der Forschungsgemeinschaft zur Verfügung zu stellen.