

Veranstaltung zum Tagungsmotto: Bioinformatik und Biotechnologie

Volker Heun und Ralf Zimmer (Organisatoren)
Institut für Informatik der Ludwig-Maximilians-Universität München
{ Volker.Heun|Ralf.Zimmer }@bio.ifi.lmu.de

1 Vorwort

Life Sciences meet Computer Science: In den letzten Jahren hat der Einsatz von Informatik-Methoden die Lebenswissenschaften revolutioniert und dort gänzlich neue Denkweisen und Verfahren eingeführt. In dieser Veranstaltung stellen Experten der Bioinformatik und Biotechnologie aktuelle Problemstellungen und neue Herausforderungen in den Lebenswissenschaften sowie informatische Ansätze zu deren Lösung vor. Der Schwerpunkt liegt dabei auf bioinformatischen und systembiologischen Methoden zur Analyse von Daten aus modernen (Hochdurchsatz-)verfahren, um komplexe biologische Systeme qualitativ und quantitativ zu modellieren und damit besser zu verstehen. Am Nachmittag veranschaulichen Exkursionen in die Labor- und Technikwelt die Komplexität der Problemstellungen der modernen Lebenswissenschaften.

2 Strukturelle Bioinformatik: Die Architektur des Lebendigen *Karl-Peter Hopfner (Ludwig-Maximilians-Universität München)*

Das Verständnis der Funktionsweise von zellulären Prozessen und deren molekularer Pathologie in verschiedenen Krankheiten bedarf der genauen Analyse der zugrunde liegenden Biomoleküle. Z.B. werden pharmakologische Wirkstoffe oft anhand der Kenntnis der dreidimensionalen Struktur von Enzymen entwickelt. Die Bestimmung der dreidimensionalen Struktur von Biomolekülen mittels Röntgenkristallographie, nuklearmagnetischer Resonanz oder Elektronenmikroskopie ist sehr aufwendig und jede dieser Methoden hat ihre Stärken und Schwächen. Die Kombination dieser und weiterer Methoden in Hybridansätzen unter Einbindung von bioinformatischen Analysen und der daraus gewonnenen Informationen ist ein zukunftssträchtiger Ansatz, die Wirkweise komplexer zellulärer Mechanismen auf atomarer Ebene aufzuklären und zu verstehen.

Karl-Peter Hopfner promovierte 1997 an der Technischen Universität München. Seit 2007 ist er ordentlicher Professor für Biochemie am Genzentrum und an der Fakultät für Chemie und Pharmazie der Ludwig-Maximilians-Universität München. Sein Forschungsschwerpunkt liegt in der strukturellen Biologie der DNA-Reparatur.

3 Systembiologie halophiler Archaeen

Dieter Oesterhelt (Max-Planck-Institut Martinsried)

Systembiologie ist ein neues Forschungsgebiet der Lebenswissenschaften, das Biologen Mathematiker, Physiker, Informatiker und Ingenieure mit der Vision anzieht, ein möglichst vollständiges und quantitatives Modell biologischer Systeme, z.B. eukaryotischer Zellen, zu erstellen, um testbare Vorhersagen zu machen.

Wir verwenden halophile Archaeobakterien, insbesondere Halobakterium Salinarum für systembiologische Modelle und Experimente. Ein Modell der Signalübertragung simuliert die Experimente mit dem Flagellarmotor von *H. salinarum*. Ein zweites Modell mit ca. 800 Reaktionen des zentralen Energiestoffwechsels erlaubt es, Wachstumsbedingungen und die Nutzung von Kohlenstoffquellen quantitativ zu modellieren.

Dieter Oesterhelt promovierte 1967 und habilitierte 1973 an der Ludwig-Maximilians-Universität München. Seit 1979 ist er Direktor und wissenschaftliches Mitglied am Max-Planck-Institut für Biochemie in Martinsried und ordentlicher Professor für Biochemie an der Ludwig-Maximilians-Universität München. Forschungsschwerpunkte: Retinalproteine als Lichtwandler und Lichtsensoren; Struktur- und Funktionsdynamik von Membranproteinen; Signaltransduktionsketten in Haloarchaea; Kristallisation von Membranproteinen; Systembiologie halophiler Archaeen.

4 Molekulare Mechanismen komplexer Erkrankungen

Hans-Werner Mewes (Technische Universität München, Helmholtz-Zentrum München)

Traditionell hat die Medizin Pathophenotypen als Krankheiten klassifiziert und behandelt. Seit die Genomforschung den Zugriff auf die Primärinformationen der molekularen Komponenten ermöglicht hat, rückt die Interaktion zwischen dem Genom, seiner Ausprägung im Phänotyp und der Umwelt als *Conditio Humana* in den Mittelpunkt der Forschung. Experimentell spielen molekulare Hochdurchsatztechnologien und die Interpretation der aus ihnen gewonnenen Daten eine zentrale Rolle für die Bioinformatik. Große und komplexe, vieldimensionale Datenmengen müssen im Kontext aktuellen biologischen Wissens interpretiert werden. Eine rationale Interpretation des Übergangs vom gesunden in den pathogenen Status eines Systems gelingt nur, wenn es möglich ist, funktionelle Einzelinformationen in Form von Netzwerken zu strukturieren. Nur so kann die Wechselwirkung zwischen dem menschlichen Organismus und den vielfältigen, seine Gesundheit beeinflussenden Faktoren in kausalen Zusammenhang gebracht werden. Der Vortrag stellt grundlegende Konzepte einer medizinischen Systembiologie vor und beschreibt verschiedene experimentelle und theoretische Ansätze zur systemischen Erforschung komplexer Erkrankungen.

Hans-Werner Mewes promovierte 1987 an der Philipps-Universität Marburg. Seit 2001 ist er ordentlicher Professor für genomorientierte Bioinformatik an der Technischen Uni-

versität München und Direktor des Instituts für Bioinformatik und Systembiologie am Helmholtz-Zentrum in München. Forschungsschwerpunkte: Systematische Funktions- und Strukturanalyse von Proteinen; Methodenentwicklung zur Interpretation molekularbiologische Daten und Informationen; Charakterisierung von Modellgenomen wie Hefe, *Arabidopsis thaliana* und Maus.

5 Algorithmische Massenspektrometrie: Informatikmethoden in der Proteomik und Metabolomik *Oliver Kohlbacher (Universität Tübingen)*

Die erfolgreiche Sequenzierung vieler Genome hat die Grundlage für die tiefergehende Analysen komplexer biologischer Systeme gelegt. Mit Hilfe moderner Methoden, insbesondere auch der Massenspektrometrie, gelingt es nun auch, Proteome (die Gesamtheit aller in einer Probe vorhandenen Proteine) oder Metabolome (die Stoffwechselprodukte) zu analysieren. Das Volumen und die Komplexität der dabei anfallenden Datensätze machen eine Auswertung ohne Informatikmethoden unmöglich. Anhand einiger Beispiele wird gezeigt, wie Methoden aus vielen Bereichen der Informatik die Analyse dieser Daten ermöglichen und zu einem systematischen Verständnis der Biologie beitragen.

Oliver Kohlbacher promovierte 2001 an der Universität des Saarlandes mit einer Arbeit über Protein Docking. Seit 2003 ist er ordentlicher Professor für die Simulation biologischer Systeme an der Universität Tübingen und Mitglied des Zentrums für Bioinformatik Tübingen (ZBIT). Sein Forschungsschwerpunkte liegen in den Bereichen Proteinstruktur, Wirkstoffentwurf und Proteomik.

6 Taxonomische und funktionelle Charakterisierung mikrobieller Metagenome mittels Hochdurchsatz-Sequenzierung *Jens Stoye (Universität Bielefeld)*

Unter Metagenomik versteht man das neue Forschungsgebiet, in dem große heterogene DNA-Datensätze auf ihre mikrobielle Zusammensetzung hin untersucht werden. Neue Sequenzieretechniken wie die von Roche vertriebene 454-Sequenzierung oder die Illumina/Solexa-Technologie versprechen hier viele neue Möglichkeiten der molekularbiologischen Analyse, da mit ihnen deutlich größere DNA-Mengen in kürzerer Zeit und zu günstigeren Kosten ermittelt werden können als mit der etablierten Sanger-Technik. Auf der anderen Seite sind die erhaltenen DNA-Abschnitte (Reads) deutlich kürzer, 35-50 Basenpaare (Bp) mit Solexa/Illumina und 100-300 Bp mit 454-Sequenzierung im Vergleich zu durchschnittlich über 500 Bp bei der Sanger-Sequenzierung, was die Auswertung und Einordnung der Daten deutlich schwieriger macht.

Im Anschluss an die Sequenzierung ist die tatsächliche Analyse eines metagenomischen Datensatzes im Wesentlichen eine bioinformatische Aufgabe, bei der verschiedene Ziel-

stellungen im Vordergrund stehen können. Bei der in Bielefeld entwickelten CARMA-Pipeline haben wir uns speziell die Aufgabe gestellt, die Spezies-Komposition und die wesentlichen aktiven Genfunktionen in einem metagenomischen Datensatz zu charakterisieren. Die Zusammensetzung der Spezies wird beispielsweise ermittelt, indem zunächst die am meisten informationstragenden Reads eines Datensatzes identifiziert und dann in taxonomische Gruppen gemäß ihrer wahrscheinlichsten Herkunft eingeordnet werden. Auf diese Weise lässt sich ein Profil erstellen, das Auskunft über die Verteilung der phylogenetischen Zusammensetzung der untersuchten mikrobiellen Gemeinschaft gibt.

Die CARMA-Pipeline wurde bereits in verschiedenen mit der 454-Technologie sequenzierten mikrobiellen Gemeinschaften erfolgreich eingesetzt, u.a. bei der Charakterisierung von Antibiotika-Resistenzplasmiden in einem Datensatz aus einer Krankenhaus-Kläranlage und bei der Untersuchung der bakteriellen Gemeinschaft in einer Biogas-Anlage, um ein besseres Verständnis der Zersetzungs-Abläufe in einer solchen Anlage in Abhängigkeit von ihrer Beschickung mit verschiedenen Arten von Biomasse zu erlangen.

Anhand der DNA-Proben aus der Biogas-Anlage haben wir nun erstmalig die CARMA-Pipeline auch auf DNA-Reads angewandt, die mit der Illumina/Solexa-Technologie sequenziert wurden. Die vorläufigen Ergebnisse an 5,4 Millionen 50 Bp-Reads sind sehr vielversprechend, da sie ein mit den 454-Daten sehr konsistentes Bild bezüglich der phylogenetischen Zusammensetzung gezeigt haben. Allerdings stellt die weit höhere Zahl an Reads bei gleichzeitig verkürzter Länge eine große Herausforderung an die bioinformatische Analyse dar. Der wesentliche Beitrag dieses Vortrags wird es sein zu beschreiben, wie wir mit informatischen Methoden in der Lage sind, diese Herausforderung zu beherrschen.

Jens Stoye promovierte 1997 an der Universität Bielefeld mit einer Arbeit über Divide-and-Conquer-Algorithmen für Alignments mehrerer Sequenzen. Seit 2002 ist er ordentlicher Professor für Genom-informatik an der Universität Bielefeld und Mitglied des Center for Biotechnology (CeBiTec). Sein Forschungsschwerpunkt liegt im Entwurf und in der Analyse von Algorithmen in der Bioinformatik und in der Genomanalyse.

7 Informatik und neue Sequenzierungstechnologien SeqAn: Eine generische C++ Bibliothek zu Sequenzanalyse Knut Reinert (Freie Universität Berlin)

Die technologische Entwicklung der letzten Jahre erlaubt es nun, recht kostengünstig Milliarden von Sequenzstücken zu entschlüsseln. In diesem Vortrag stellen wir kurz einige der Probleme vor, welche dadurch bei der Sequenzanalyse von genomischen Sequenzen entstehen und stellen eine generische C++ Bibliothek – SeqAn – vor, welche viele effizient implementierte algorithmische Komponenten dazu enthält.

Knut Reinert promovierte 1999 an der Universität des Saarlandes mit einer Arbeit über einen polyhedralen Ansatz zum Alignment von Sequenzen. Seit 2002 ist er ordentlicher Professor für Bioinformatik an der Freien Universität Berlin. Sein Forschungsschwerpunkt liegt in der Entwicklung mathematischer Modelle und Algorithmen für die Analyse großer Genome und Massenspektrometriedaten.