

INFORMATIKMETHODEN ZUR ANALYSE UND INTERPRETATION GROSSER GENOMISCHER DATENMENGEN

Ziel des Schwerpunktprogramms *Informatikmethoden zur Analyse und Interpretation großer genomischer Datenmengen* ist die Entwicklung von Informatikmethoden für die Exploration großer genomischer Datensätze, damit Beziehungen zwischen Struktur und Funktion aufgeklärt und so zelluläre Komponenten in metabolische oder regulatorische Netzwerke eingeordnet werden können. Auf der methodischen Seite spielt die Modellierung komplexer biologischer Interaktionen, die Entwicklung effizienter Algorithmen für den geforderten Datenzugriff und -durchsatz sowie die Visualisierung eine Rolle.

Dieser Zielsetzung ordnen sich folgende Problembereiche unter:

- Analyse molekularbiologischer Sequenzen, insbesondere Annotation, Sequenzalignment und Analyse von Sequenzvariabilität, Strukturierung des Genoms sowie Polymorphismen
- Systematische Genomvergleiche (z.B. Analyse der genomischen Topologie verwandter pathogener/nichtpathogener Organismen)
- Molekulare Strukturbestimmung mit Informatikmethoden (Proteine, RNA, Komplexe)
- Bestimmung molekularer Funktionen auf der Basis von Sequenz- und/oder Strukturvergleichen
- Molekularbiologische Datenbanken (Organisation, Zugriff, Datenvalidierung, Suche nach Mustern, Klassifikation)
- Rechnermodellierung regulatorischer und metabolischer Netzwerke
- Visualisierung molekularbiologischer Daten

Projekte sollen auf der methodischen Seite durch Inhalte aus der Informatik und/oder der Mathematik gekennzeichnet sein und gleichzeitig eine direkt nachvollziehbare Relevanz für wichtige molekularbiologische Fragestellungen haben. Deshalb wird z.B. die Validierung von Softwarewerkzeugen an realen biologischen Daten vorausgesetzt und eine Implementierung in allgemein nutzbarer Form gefordert