

Erstes deutsches Community-Grid-Projekt

## SIMAP – Tempo für die Proteinforschung

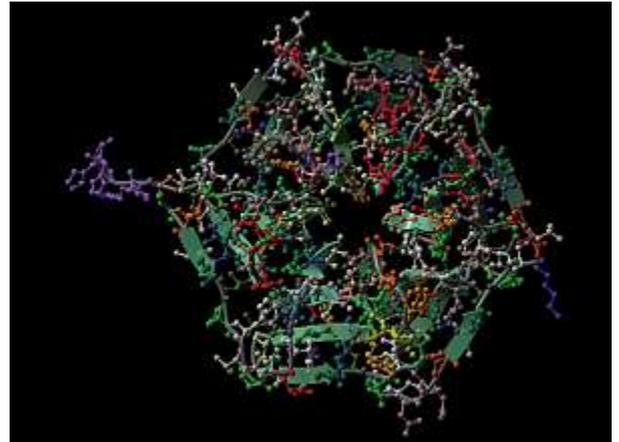
Auf der ganzen Welt arbeiten Forscher daran, die Aminosäuresequenzen von Proteinen zu entschlüsseln und miteinander zu vergleichen, um mehr über die Funktion der Eiweiße zu erfahren. Angesichts der enormen Datenmengen – schon heute kennt man die Sequenz von rund vier Millionen Proteinen – ist dazu eine riesige Rechnerkapazität vonnöten. Dieses Problem lässt sich mit der Datenbank SIMAP lösen, gemeinsam entwickelt von Wissenschaftlern des Lehrstuhls für Genomorientierte Bioinformatik am TUM-Wissenschaftszentrum Weihenstephan (Prof. Hans-Werner Mewes) und des Instituts für Bioinformatik des GSF-Forschungszentrums für Umwelt und Gesundheit in Neuherberg. SIMAP@HOME nutzt als erstes deutsches Projekt das System des »Community Grid Computing«.

Proteine als die wichtigsten Bausteine des Lebens sind Produkte einer Jahrmillionen langen Optimierung. Ihre Eigenschaften – entscheidend für ihre Rolle im zellulären Geschehen – lassen sich nicht einfach aus der Folge ihrer Bausteine ablesen, der Aminosäuren. Das wichtigste Werkzeug für die Entschlüsselung der Funktion ist der Vergleich der Aminosäuresequenzen untereinander. Ähnlichkeiten geben Hinweise auf die Verwandtschaftsverhältnisse zwischen Proteinen. Verwandte Eiweiße haben oft gleiche oder ähnliche Eigenschaften und Funktionen im Organismus, da Proteine sich im Lauf der Evolution nur langsam verändern. Derzeit kennt man viel mehr Proteinsequenzen, als man eingehend in Labors untersuchen kann, weshalb man die experimentellen Erkenntnisse zu einem Protein auch auf dessen Verwandte überträgt.

Die Biologen-Frage »Welche Proteine sind mit *meinem* Protein verwandt?« wird täglich hunderttausendfach von Computern überall auf der Welt bearbeitet. Daraus entstand die Idee, alle bekannten Proteine mitein-

ander zu vergleichen (»alle gegen alle«) und damit die ultimative Lösung des Problems anzubieten. Das Ergebnis, die in München entwickelte Datenbank SIMAP, kann nicht nur Anfragen rasch und effizient beantworten, der entstandene Datensatz ist auch eine ideale Basis, um viele Fragen von Molekularbiologen zu bearbeiten. Mit SIMAP steht eine vollständige Matrix aller bekannten Proteinverwandtschaften zur Verfügung, die ständig aktualisiert wird.

Um die Sequenzähnlichkeiten der bisher bekannten rund vier Millionen Proteine, also 16 000 Milliarden Vergleiche, vollständig zu berechnen, würde ein einzelner Computer etwa 80 Jahre brauchen. Darum griff das Team um Ordinarius Hans-Werner Mewes – Dr. Thomas Rattei und Roland Arnold – auf die Hilfe von Freiwilligen zurück, die die ungenutzte Rechnerkapazität ihrer Computer zur Verfügung stellten. Diese Technologie nennt sich Community Grid Computing, prominentestes Beispiel ist das Projekt SETI@HOME, das seit vielen Jahren Radiowellen



Proteinstruktur der Sensor-Domain einer Protein-Kinase aus *Mycobacterium tuberculosis*.

aus dem Weltraum nach Signalen außerirdischer Intelligenz durchsucht. SIMAP@HOME ist das erste deutsche Community-Grid-Projekt und hat seit Dezember 2005 rund 10 000 Freiwillige mit über 20 000 Computern aus mehr als 50 Ländern gewinnen können. Die momentane Rechnerkapazität liegt bereits im Leistungsbereich eines Supercomputers mit weit mehr als dreieinhalb Teraflops – eine Leistung, die die Kapazität vieler Rechenzentren bei weitem übertrifft.

Die schnell wachsende Zahl an Genomsequenzierprojekten lässt die Kenntnisse über Gen- und somit Proteinsequenzen rapide ansteigen. Durch die Hilfe der weltweiten Community wird das SIMAP-Projekt auch in Zukunft mit dieser Entwicklung Schritt halten und wertvolle Daten für zahlreiche Analysen in der Genom- und Gesundheitsforschung liefern können.

Die Informationen stehen jedem Wissenschaftler zur Verfügung unter: <http://mips.gsf.de/simap>

Wer mit seinem Rechner an SIMAP teilnehmen möchte, erfährt Näheres unter: <http://boinc.bio.wzw.tum.de/boincsimap>

Thomas Rattei

Dr. Thomas Rattei  
Lehrstuhl für  
Genomorientierte  
Bioinformatik  
Tel.: 08161/71-2132  
t.raitei@wzw.tum.de