

■ G. MATTHIAS ULLMANN

Mit Computern die Bausteine des Lebens erschließen

DER WANDEL DER BIOWISSENSCHAFTEN IM DIGITALEN ZEITALTER

■ Screen-Spot eines Linux-Computers mit dem Linux-Maskottchen Tux, dem Pinguin, in der linken oberen Ecke. Weiterhin sind gezeigt eine Moleküldarstellung mit dem interaktiven Programm VMD, ein Protein-Sequenzalignment, das Ergebnis einer Suche in der Cambridge Structural Database sowie verschiedene Fenster (sogenannte Konsolen) zur Kommandoingabe (Bild: G. Matthias Ullmann).

Die biologischen Wissenschaften haben sich in den letzten zwanzig Jahren grundlegend geändert. Hauptgründe dafür sind der Einsatz von Automatisierungstechnologien und die Vernetzung der Forschenden und der wissenschaftlichen Daten im Internet. Zudem sind die Kosten für die Generierung enormer Datenmengen mittlerweile stark gesunken. Ein Beispiel ist das 1990 gegründete und 2003 offiziell beendete *Human Genome Project*. Dieses von einem internationalen Forschungsverbund getragene Vorhaben zielte darauf ab, das Genom des Menschen vollständig zu analysieren und kostete rund 3 Milliarden Dollar. Heute hingegen kann man das Genom eines einzelnen Menschen schon für weniger als 10.000 Dollar bestimmen.

Die Sequenzierungstechniken, die bei diesen Forschungsarbeiten zum Einsatz kommen, haben auch ein breites Anwendungsfeld in der Ökologie. So werden bei der Untersuchung mikrobieller Lebensgemeinschaften nicht die Genome einzelner Bakterien bestimmt, sondern die Genome aller Organismen der Lebensgemeinschaft, die sogenannten Metagenome. Die strukturellen Untersuchungen von Biomolekülen haben sich ebenfalls stark gewandelt, wie das Beispiel der Kristallisationsroboter zeigt. Diese kommen unter anderem bei der Züchtung von Proteinkristallen zum Einsatz, die oftmals unentbehrlich sind, um die dreidimensionalen Strukturen von Biomolekülen bestimmen zu können.

Für die rasanten Fortschritte auf diesen Gebieten ist neben den technologischen Innovationen auf der apparativen Seite auch die Entwicklung von Software entscheidend. Als die Strukturbiologie in den späten 1950er Jahren in den Kinderschuhen steckte, baute man Modelle von Molekülen noch aus Draht und Kugeln in monatelanger Kleinarbeit. Heute dagegen können Proteinstrukturen in manchen Fällen schon innerhalb weniger Stunden auf einem Laptop berechnet werden, sobald die erforderlichen experimentellen Daten vorliegen.

INTERNET – DAS MEDIUM FÜR DEN WISSENSCHAFTLICHEN DATENAUSTAUSCH

Seit den Ursprüngen des Internets ist der Austausch wissenschaftlicher Daten eine seiner wichtigsten Aufgaben. Viele Daten sind heute über Datenbanken, die von großen Forschungseinrichtungen wie dem European Bioinformatics Institute (EBI) oder

dem National Center for Biotechnology Information (USA) gepflegt werden, weltweit online verfügbar. Dabei stellen die meisten dieser Institute nicht nur die Rohdaten zur Verfügung, sondern bieten über Webserver zahlreiche Suchmöglichkeiten und Analysewerkzeuge an. Wer die Daten für die eigene Forschung nutzen will, ist deshalb nicht darauf angewiesen, auf dem eigenen Computer spezielle Software zu installieren und große Datenbanken vorzuhalten.

Allerdings bedürfen die Daten der Analyse und der Interpretation. Darauf wies bereits der britische Biologe Sydney Brenner im Jahr 2002 hin, als er den Nobelpreis für Medizin und Physiologie erhielt: "Die große Herausforderung in der biologischen Forschung heute ist es, Daten in Wissen zu verwandeln. Ich habe Menschen getroffen, die denken, Daten sind Wissen, aber diese Menschen streben dann danach, Wissen in Verstehen zu wandeln."¹

MATHEMATISCHE MODELLE FORDERN UND FÖRDERN INTERDISZIPLINARITÄT

Um einen schwierigen Sachverhalt zu verstehen, ist es in den Naturwissenschaften notwendig, ein konzeptionelles Modell zu entwickeln, das eine vereinfachte bildliche Vorstellung des Sachverhalts ermöglicht. Sobald es jedoch darum geht, tiefere und oft weniger offensichtliche Zusammenhänge aufzuzeigen, sind computergestützte mathematische Modelle unverzichtbar. Zwei Chemie-Nobelpreise unterstreichen deren zentrale Bedeutung für die Forschung: 1998 wurden John Pople und Walter Kohn für die Entwicklung quantenchemischer Methoden gewürdigt; 2013 ging die Auszeichnung an Martin Karplus, Michael Levitt und Arieh Warshel, die sich um die Entwicklung von Multiskalen-Modellen für komplexe chemische Systeme verdient gemacht hatten.

„SOBALD ES DARUM GEHT, TIEFERE UND OFT WENIGER OFFENSICHTLICHE ZUSAMMENHÄNGE AUFZUZEIGEN, SIND COMPUTERGESTÜTZTE MATHEMATISCHE MODELLE UNVERZICHTBAR.“

Solche computergestützten Methoden auf biochemische und biologische Fragestellungen anwenden zu können, erfordert ein breites fächerübergreifendes Fundament in den Naturwissenschaften. Wer

AUTOR



Prof. Dr. G. Matthias Ullmann ist Professor für Bioinformatik/Strukturbiologie an der Universität Bayreuth.

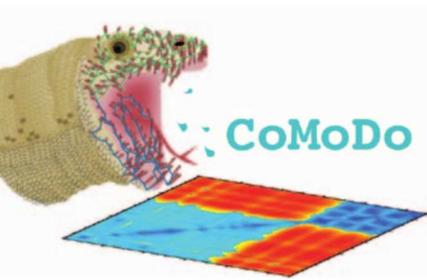
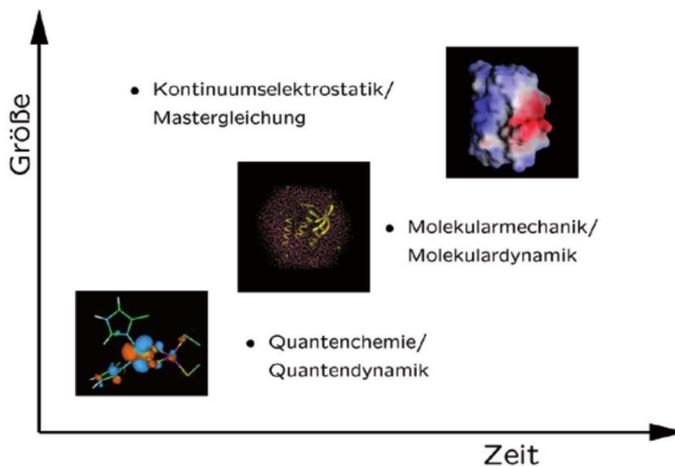


Abb. 1: Abbildung aus dem Inhaltsverzeichnis zum Artikel über das Programm CoMoDo.² Der Comodo-Waran hat in seinem Maul ein Protein aus zwei Domänen. Die Pfeile deuten die korrelierte Bewegung des Proteins an. Rechts ein Korrelationsdiagramm (Grafik: Mit freundlicher Genehmigung der American Chemical Society).

Abb. 2: Die Modellierung von biomolekularen Systemen kann auf verschiedenen Zeit- und Längenskalen erfolgen. Je kleiner die Zeit- und Längenskala, desto genauer ist die Modellierung, aber desto größer ist auch der Rechenaufwand. Oft ist es auch notwendig, verschiedene Methoden zu kombinieren, um die biologisch wichtigen Prozesse simulieren zu können (Grafik: G. Matthias Ullmann).

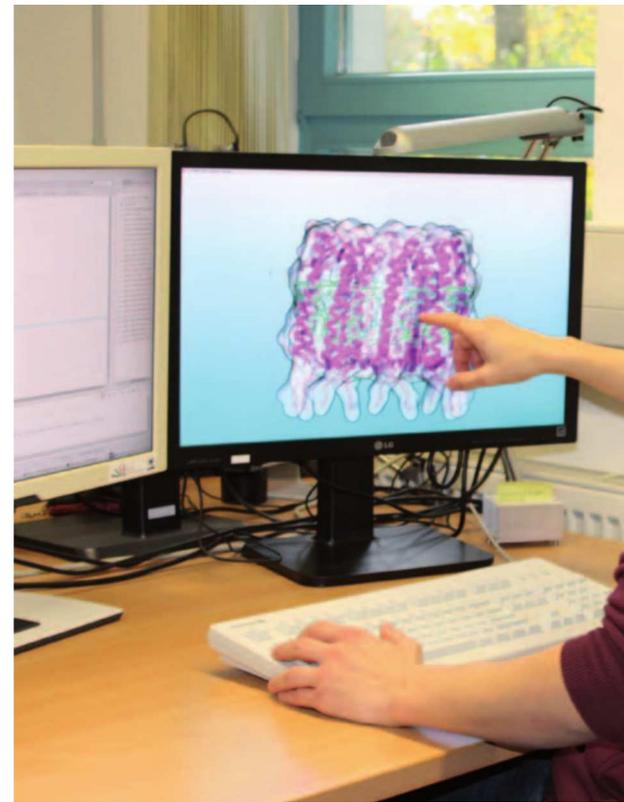


die Biologie nicht versteht, kann nicht die richtigen Fragen stellen. Wer Physik und Chemie nicht beherrscht, dem fehlt das Hintergrundwissen, um die Fragen zu beantworten. Wer Angst hat vor Computern und Mathematik, hat nicht das Handwerkszeug, um die Probleme anzugehen. Interdisziplinarität ist daher in den modernen molekularen Biowissenschaften von überragender Bedeutung.

COMPUTERSIMULATIONEN ZUR DYNAMIK VON PROTEINEN

An der Universität Bayreuth hat sich bereits vor 12 Jahren eine Forschungsgruppe für Bioinformatik und Strukturbiologie etabliert. Sie untersucht mit theoretischen Methoden die chemischen Mechanismen von Enzymen. Die Vielfalt der Fragestellungen spiegelt sich dabei in der Unterschiedlichkeit der jeweiligen Verfahren, die für deren Bearbeitung eingesetzt werden. Ist man beispielsweise am genauen Ablauf einer chemischen Reaktion interessiert, führt kein Weg an quantenchemischen Methoden vorbei. Will man mehr über die Bewegungsabläufe eines einzelnen Proteins erfahren, ist Molekulardynamik die Methode der Wahl. Ist man an sehr großen Enzymsystemen interessiert, können Kontinuumelektrostatik-Rechnungen in Kombination mit einem Master-Gleichungsansatz die Lösung sein. Der Erfolg einer Simulation hängt immer davon ab, dass die jeweils passenden Methoden sinnvoll kombiniert werden und die dabei relevanten Größen- und Zeitskalen korrekt behandelt werden (Abb. 2).

Eine Kombination von Methoden, die auf verschiedene Größen- und Zeitskalen Bezug nehmen, ist häufig eine Grundlage für Computersimulationen.



So wurde in Bayreuth beispielsweise ein neuer Ansatz erarbeitet, um die Dynamik von Proteinen zu simulieren. Die meisten großen Proteine sind aus mehreren Domänen aufgebaut. Es handelt sich dabei um kompakte molekulare Einheiten, welche die Proteine beweglich machen – und zwar so, dass die Proteine ihre jeweiligen Funktionen erfüllen. Die Organisation dieser Domänen muss man deshalb verstehen, um die Dynamik der Proteine erkennen und vorhersagen zu können. Der neue Ansatz hilft, funktionelle Proteinbewegungen in biologischen Prozessen besser zu verstehen, beispielsweise bei der Übertragung von Signalen oder der Ligandenbindung. Das in Bayreuth entwickelte Programm namens „CoMoDo“ – die Abkürzung steht für „Covariance of Motion Domains“ – ist als Open Source Software allgemein zugänglich.²

LINUX – DER VERSTECKTE MOTOR DES WISSENSCHAFTLICHEN RECHNENS

Als der Einsatz von Computern in der Wissenschaft noch am Anfang stand, beschränkte er sich meist auf wenige Institute mit Großrechnern. Diese Computer arbeiteten nicht mit Windows oder DOS, sondern dem Betriebssystem Unix. Es ist auf wissenschaftliche Bedürfnisse sehr gut zugeschnitten, da es erlaubt, Daten sehr einfach zu bearbeiten, dazustellen und zu transferieren. Im Bereich der in den 1990er Jahren entwickelten Heimcomputer dominierte hingegen das Betriebssystem Windows, das eine der Ursachen für deren einge-



erfolgreich arbeiten wollen, mit dem Computer sicher umgehen und grundlegende Kenntnisse in der Computerprogrammierung besitzen müssen. Steve Jobs, der Mitbegründer von Apple Inc., hat einst in einem Interview gesagt, dass jeder lernen sollte, wie man einen Computer programmiert, da dies einem beibringt zu denken.³ So wie heute niemand sein Abitur ohne Grundkenntnisse der Integral- und Differentialrechnung besteht, wird in wenigen Jahren ein naturwissenschaftlicher Universitätsabschluss ohne Kenntnisse in der Computerprogrammierung und der automatisierten Datenanalyse nicht möglich sein.

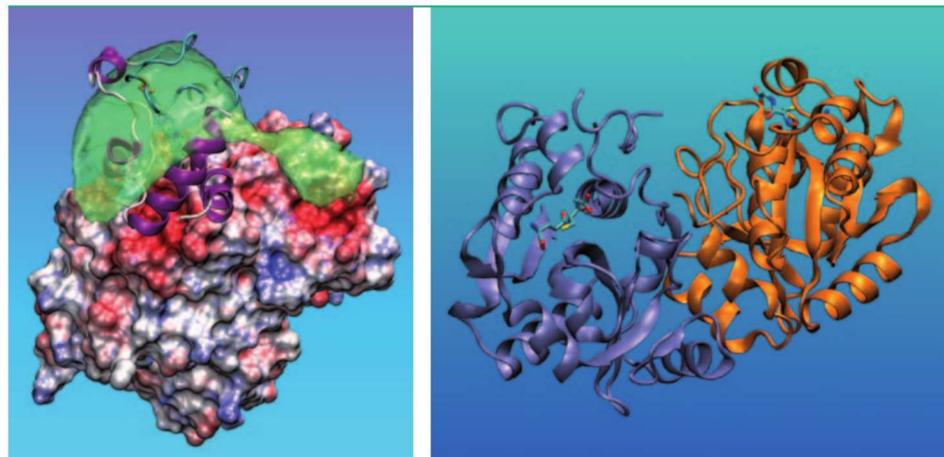
Abb. 3: Lars Müller, Doktorand und Mitglied der Computational Biochemistry Group an der Universität Bayreuth, zeigt auf das Modell eines Lichtsammelkomplexes, der in der bakteriellen Photosynthese eine Rolle spielt (Foto: Christian Wißler).

schränkte Funktionalität war. In der computergestützten Forschung wurden deshalb zu dieser Zeit immer noch teure Workstations, insbesondere der Firmen Sun, Silicon Graphics oder IBM, eingesetzt. Denn diese Computer liefen unter dem Betriebssystem Unix, dessen Funktionalität für das wissenschaftliche Arbeiten mit Computern im größeren Maßstab unverzichtbar ist.

Ab Mitte der 1990er Jahre setzte sich jedoch mehr und mehr das Betriebssystem Linux durch, das es ermöglichte, die volle Funktionalität eines Unix-Systems auf normalen Heimcomputern zu nutzen. Die Entwicklung wurde durch den Umstand begünstigt, dass Linux kostenlos zur Verfügung gestellt wird. Zumeist ist auch der Quellcode der Programme zugänglich, so dass es möglich ist, die theoretischen Grundlagen der Programme nachzuvollziehen und Teile der Programme in eigener Software wiederzuverwenden. Deshalb ist die Bedeutung von Linux für die Entwicklungen der Bioinformatik und anderer computergestützter Forschungsrichtungen kaum zu überschätzen.

HERAUSFORDERUNGEN DER ZUKUNFT

In den letzten Jahren wurden in den Biowissenschaften viele experimentelle Methoden entwickelt, die sehr große Datenmengen in kürzester Zeit produzieren können. Um diese großen Datenmengen sinnvoll analysieren zu können, werden in Zukunft alle, die in den Biowissenschaften



Die zentrale Rolle der Verarbeitung digitaler Daten und die Bedeutung des Betriebssystems Linux im wissenschaftlichen Rechnen sollte sich daher auch stärker in der universitären Lehre widerspiegeln. Das erfordert jedoch ein Umdenken auf Seiten der Studierenden, die gerade in den Biowissenschaften nicht besonders computer- und theorieaffin sind, aber auch auf Seiten der Lehrenden. Die Computerausbildung muss neben mathematisch-physikalischen Themen einen breiteren Raum in den biowissenschaftlichen Studiengängen einnehmen und eine größere Tiefe erreichen. Nur so wird es den Universitäten auch weiterhin möglich sein, mit dem Wandel in den Biowissenschaften Schritt zu halten und ihn kreativ mitzugestalten.

Abb. 4 (links): Elektrostatisches Potenzial an der Oberfläche des Proteins Cytochrom-c-Peroxidase. Dieses Protein wechselwirkt mit Cytochrom c (als Kette dargestellt). Die grüne Wolke deutet die Bewegung von Cytochrom c auf der Oberfläche an. Die Wechselwirkung der beiden Proteine ist sehr dynamisch. Cytochrom-c-Peroxidase wandelt Wasserstoffperoxid in ungiftiges Wasser um (Bild: G. Matthias Ullmann).

Abb. 5 (rechts): Das Protein Phytochelatin-Synthase setzt sich aus zwei gleichartigen Bausteinen zusammen und ist daher ein Dimer. Es stellt aus Glutathion (links und rechts jeweils als Stabmodell dargestellt) Phytochelatin her, das bei der Entgiftung von Cadmium-Ionen in Pflanzen eine Rolle spielt (Bild: G. Matthias Ullmann).

¹ "The great challenge in biological research today is how to turn data into knowledge. I have met people who think data is knowledge but these people are then striving for a means of turning knowledge into understanding."
Sydney Brenner: *Ontology Recapitulates Philology*, in: *The Scientist* (2002) 16(6), S. 12.
² Silke A. Wieninger und G. Matthias Ullmann: CoMoDo: Identifying Dynamic Protein Domains Based on Covariances of Motion, in: *Journal of Chemical Theory and Computation* (2015), 11 (6), S. 2841–2854.
³ www.youtube.com/watch?v=mCDkxUbalCw, aufgerufen am 19.11.2015.