



Informatik

Höchstleistungsrechnen für die Gesundheit

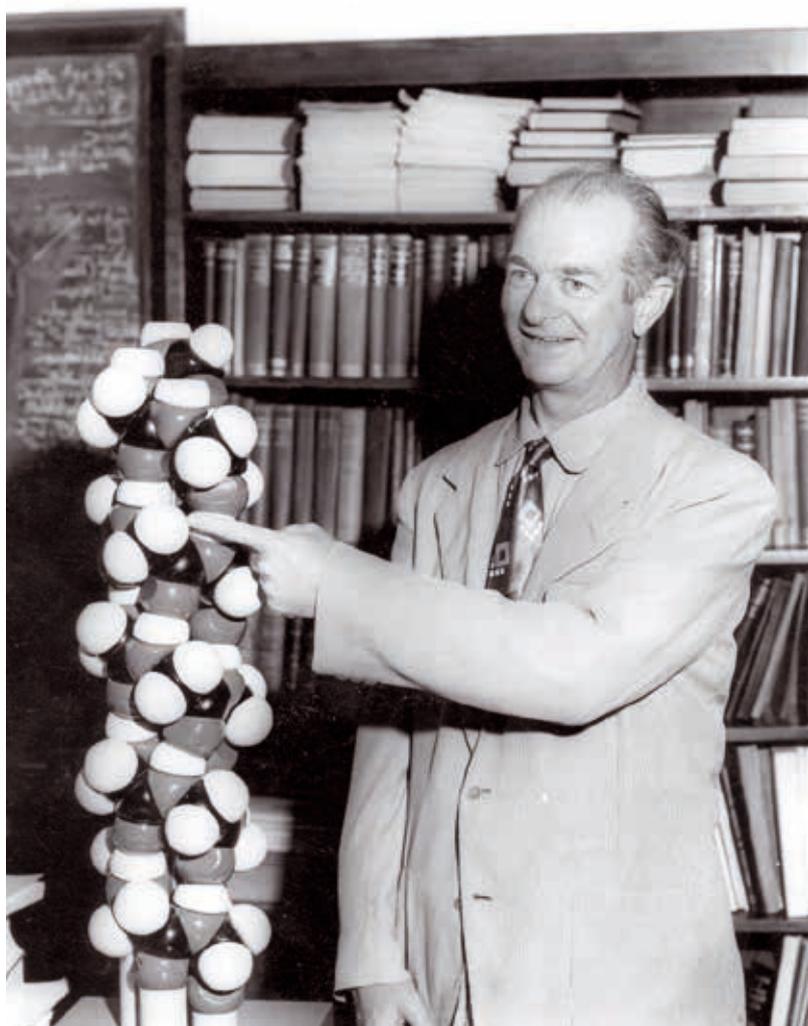
Die ersten Modelle der Proteine, der Bausteine allen Lebens, entstanden noch aus Holz. Heute werden immer komplexere Strukturen und Fragestellungen aus ganz verschiedenen Bereichen der Biologie, Chemie und Medizin auf Höchstleistungsrechnern wie SuperMUC simuliert. Die Forschungserkenntnisse dienen dem besseren Verständnis des menschlichen Körpers ebenso wie neuen Therapieformen und Medikamenten, etwa in der individualisierten Medizin.



Von Ludger Palm

LANGE BEVOR ES die ersten elektronischen Rechner gab, untersuchte der Chemiker Linus Pauling in seiner Doktorarbeit am California Institute of Technology in Pasadena 1925 anorganische Kristalle mit der neuen Röntgenstrukturanalyse. Als einer der ersten Guggenheim-Stipendiaten erhielt er ein Jahr später bei Arnold Sommerfeld am Institut für Theoretische Physik in München Gelegenheit, die neue Quantentheorie der Atome zu erlernen. Bei Erwin Schrödinger in Zürich traf er dessen Assistenten Walter Heitler und Fritz London, die als Erste eine Theorie der chemischen Bindung formulierten.

Linus Pauling mit einem Holzmodell der von ihm entdeckten alpha-Helix eines Proteins (1954).



Vom Holzmodell ...

Zurück am „Caltech“, kombinierte Pauling die Quantentheorie mit neuen experimentellen Techniken. So gelang es ihm, immer kompliziertere Molekülstrukturen aufzuklären. An seinem 50. Geburtstag konnte er 1951 als grundlegende Strukturen der Proteine die alpha-Helix und das beta-Faltblatt vorschlagen. Er bastelte sie aus Holzmodellen, deren Herstellung er wissenschaftlich genau beschrieb, und verglich die Abstände und Winkel zwischen den Atomen im Modell mit den in Experimenten gemessenen Abständen. 1954 wurde er für seine Forschungen mit dem Nobelpreis für Chemie ausgezeichnet.

1962 erhielten Max Perutz und John C. Kendrew den Nobelpreis für Chemie, und zwar für ihre Strukturaufklärung der Proteine Myoglobin und Hämoglobin. Auch sie untersuchten die Beugung von Röntgenstrahlen an Kristallen. In den folgenden Jahrzehnten wurden immer

mehr Strukturen von Proteinen aufgeklärt, stets jedoch ging es dabei um Kristalle. In der Natur hingegen schwimmen die Proteine in wässrigen Lösungen, so auch in uns allen – in den Zellen, im Blut, überall. Erst die Kernresonanzspektroskopie (NMR-Spektroskopie) ermöglichte es in den letzten 25 Jahren, Proteine in wässrigen Lösungen detailliert zu untersuchen. Horst Kessler, Mitglied der Bayerischen Akademie der Wissenschaften, und seine Arbeitsgruppe an der TU München lieferten wichtige Beiträge zu unserem Verständnis der Dynamik von Biomolekülen in Lösungen.

Inzwischen waren auch die Rechner leistungsfähig genug, um die Dynamik der Proteine zu berechnen. Die Atome in den Proteinen sind mehr oder weniger beweglich, abhängig von

ihrer näheren Umgebung. Diese Bewegung lässt sich mit den Gleichungen der klassischen Mechanik beschreiben, wenn keine chemischen Reaktionen stattfinden. Dabei falten sich die Proteine in ihre Sekundärstrukturen alpha-Helix und beta-Faltblatt sowie die globulären Tertiärstrukturen. Diese Strukturen lassen sich mit Hilfe des NMR-Spektrometers bestimmen. Anschließend kann man die simulierten Strukturen mit den gemessenen vergleichen. Und es wurde nun auch möglich, immer größere Biomoleküle zu berechnen. Damit stiegen die Anforderungen an die Rechner, immer häufiger kamen Supercomputer zum Einsatz.

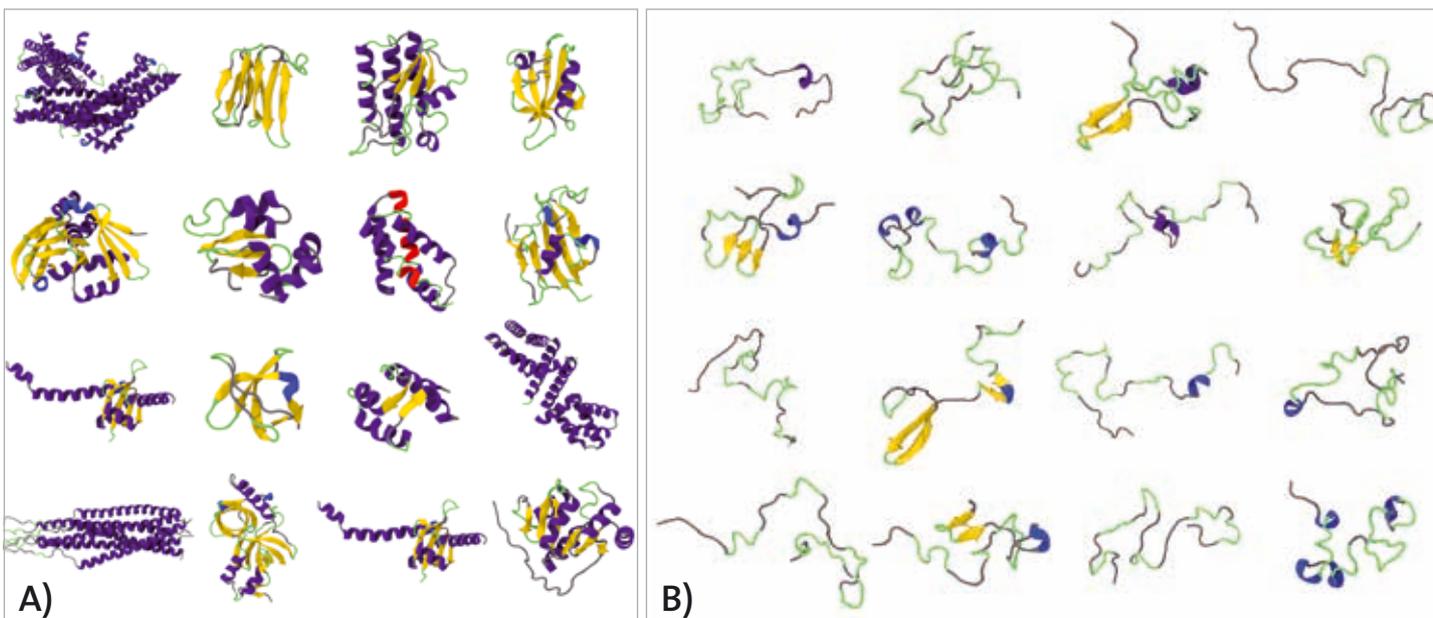
... zum SuperMUC

Allen diesen Proteinen ist gemeinsam, dass sie wohldefinierte Strukturen haben, die sich aus alpha-Helices, beta-Faltblättern usw. zusammensetzen. An noch wesentlich kompliziertere Proteine wagt sich dank des Höchstleistungsrechners SuperMUC am Leibniz-Rechenzentrum das Team um Helmut Grubmüller vom Max-Planck-Institut für biophysikalische Chemie in Göttingen. Die Forscher untersuchen ungeordnete FG-Nukleoporine. Bei diesen ist selbst die experimentelle Zuordnung einzelner Strukturen kaum noch möglich. Sie enthalten sehr große Bereiche geringer Ordnung und sind sehr flexibel. Das macht auch ihre Simulation sehr aufwändig. Aber es ist wichtig, sie zu

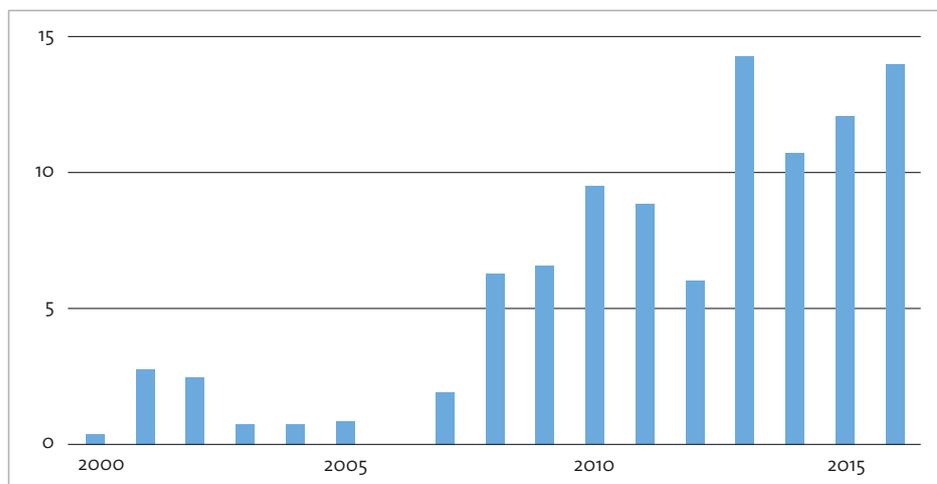
verstehen, denn die FG-Nukleoporine sind die „Türsteher“ in den Membranen unserer Zellkerne: Sie entscheiden, wer in den Zellkern hineindarf und wer hinaus. Es ist daher lebenswichtig für uns, dass die FG-Nukleoporine und andere ungeordnete Proteine ihre Aufgaben erfüllen. Ein Drittel der Proteine in Eukaryoten, also auch in uns Menschen, und drei Viertel der Proteine, die mit der Entstehung von Krebskrankungen in Verbindung gebracht werden, gehören zu diesen „Intrinsically Disordered Proteins“ (IDPs). Erst ein Supercomputer wie SuperMUC bietet genug Rechenleistung, um die Dynamik der FG-Nukleoporine so genau zu simulieren, dass sie mit den wenigen verfügbaren experimentellen Daten verglichen werden können. Wie schon bei der Röntgenstrukturanalyse der Proteinkristalle und der NMR-Spektroskopie der gelösten Proteine geht es Schritt für Schritt vom Experiment zur Simulation und zurück. So erhalten wir ein detaillierteres Verständnis komplexer Strukturen des Lebens, die über Gesundheit und Krankheit entscheiden.

Rechnen für die Gesundheit

Nicht nur komplexe Proteine lassen sich mit Höchstleistungsrechnern untersuchen. Der Anteil von Themen aus den Lebenswissenschaften, etwa Medizin, Biologie oder Biophysik, die auf Höchstleistungsrechnern wie dem SuperMUC gerechnet werden, ist in den ver-

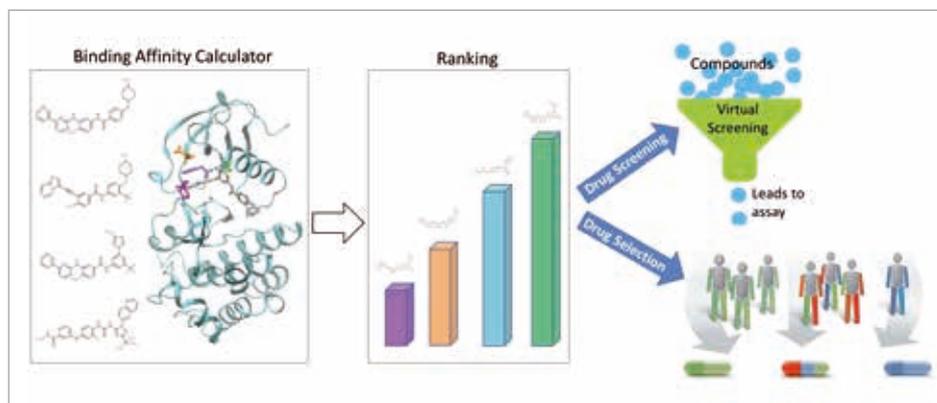


A) 16 Strukturen globulärer Proteine in ihren gefalteten Strukturen, entnommen der Protein Data Bank (violett: alpha-Helix, gelb: beta-Faltblatt).
B) Im Gegensatz dazu findet man für ein einzelnes ungeordnetes Protein ein ganzes Ensemble verschiedener Strukturen.



In den letzten zehn Jahren nahm der Anteil der Rechenzeit von Projekten aus den Lebenswissenschaften auf den Höchstleistungsrechnern am LRZ stark zu.

gangenen Jahren stark gestiegen. Nachdem derartige Projekte bis zum Jahr 2007 weniger als drei Prozent der Rechenleistung beanspruchten, stieg ihr Anteil deutlich auf heute zehn bis 15 Prozent an. Im Folgenden werden exemplarisch einige aktuelle Forschungsvorhaben vorgestellt, die auf SuperMUC gerechnet wurden.



Der Binding Affinity Calculator (BAC) ordnet eine Reihe von Medikamenten nach ihrer Bindung an ein großes Protein. Er hilft so bei der Auswahl geeigneter Substanzen für die Behandlung von Personen mit unterschiedlichem Genom.

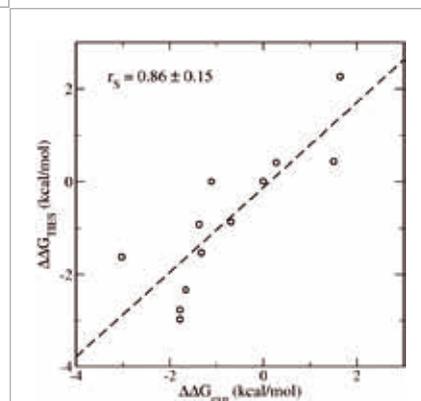
BAC: für jede Person die passende Medizin

Die genaue Kenntnis der Molekülstrukturen in ihrer natürlichen Umgebung und ihrer Energien ist auch die Basis für den Beitrag Peter Coveney's (University College London) zur personalisierten Medizin, wie man den gezielten Einsatz von Medikamenten für bestimmte Personen nennt. Manche Medikamente sind bei einigen Brustkrebspatientinnen wirksam, bei anderen aber nicht. Coveney's Team entwickelte den Binding Affinity Calculator (BAC), mit dem sich die Affinität berechnen lässt, mit der verschiedene kleinere Moleküle an das Protein Abelson-Kinase binden. Verschiedene Gen-Mutationen der Patientinnen führen zu

verschiedenen Strukturen der Abelson-Kinase, und verschiedene Medikamenten-Moleküle binden unterschiedlich stark daran. Coveney wählte vier Medikamente und fünf Varianten des Proteins aus und nahm viele Proben ihrer Wechselwirkungen, um die Bindungsaffinität zu berechnen. Diese Simulationen liefen über ein ganzes Wochenende auf allen Prozessoren des SuperMUC. Der Binding Affinity Calculator erlaubt es, eine Rangordnung der Wirksamkeit von Medikamenten aufzustellen, um die Patientinnen gezielt und wirksam zu behandeln.

Coveney's Simulationen auf dem SuperMUC ermöglichten es auch, für andere Proteine die Ergebnisse des Binding Affinity Calculators mit experimentellen Bindungsaffinitäten zu vergleichen. Wie er in zwei zum Jahreswechsel erschienenen Veröffentlichungen beschreibt, gelingt die Berechnung mit einer Genauigkeit, die die Experimente vorhersagen kann. Man ist damit dem Ziel einen großen Schritt näher gekommen, aus der Struktur des Proteins und dem ihr zu Grunde liegenden Genom vorhersagen zu können, welches Medikament bei welcher Patientin wirksam sein sollte.

Weil das Genom einer Patientin heute schnell analysiert werden kann, wäre es möglich, das für die

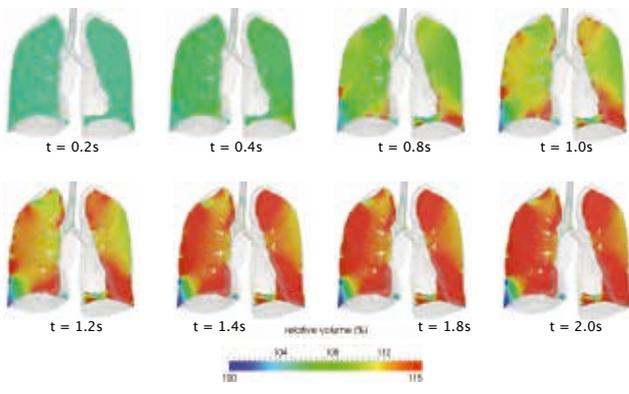


Korrelation zwischen der berechneten und gemessenen Bindungsaffinität.

einzelne Person passende Medikament schnell auszuwählen, wenn sich auch die Wirksamkeit schnell abschätzen ließe. Wofür heute noch der SuperMUC nötig ist, das werden in naher Zukunft wesentlich kleinere Rechner in der Klinik erledigen können.

BACI hilft, wenn die Luft weg bleibt

Doch auch Wissenschaftsbereiche, die man üblicherweise zu den Ingenieurwissenschaften zählt, greifen Fragen aus der Medizin auf und tragen dazu bei, Leiden zu lindern. So beschäftigt sich die Arbeitsgruppe von Wolfgang Wall an der Technischen Universität München seit vielen Jahren mit einer virtuellen Lunge. Die Forscher simulieren den Fluss des Atems bei natürlicher und künstlicher Atmung, angefangen im Großen bis hinab in die kleinsten Lungenbläschen. „Multiphysics“ nennt man solche Szenarien, bei denen die Physik über viele verschiedene Größenordnungen hinweg berechnet wird. Dafür sind spezielle Methoden nötig, um den Schritt zu immer feineren Auflösungen machen zu können.



Zunahme des Volumens des Lungengewebes beim normalen Einatmen.

So wie sich hinter Peter Coveney's „BAC“ viele verschiedene Arbeits- und Rechenschritte verbergen, so sind in Wolfgang Walls Finite-Elemente-basierter Software-Plattform „BACI“ die verschiedenen Größenordnungen zusammengefasst. Seit vielen Jahren schon nutzen Wall, Michael Gee und ihre Mitarbeiterinnen und Mitarbeiter die Supercomputer des LRZ und tragen so dazu bei, dass bei künstlicher Beatmung Verletzungen des Lungengewebes vermieden werden können. Wer denkt schon beim Lehrstuhl für Numerische Mechanik der Fakultät für Maschinenwesen an der Techni-

Schneller, höher, stärker

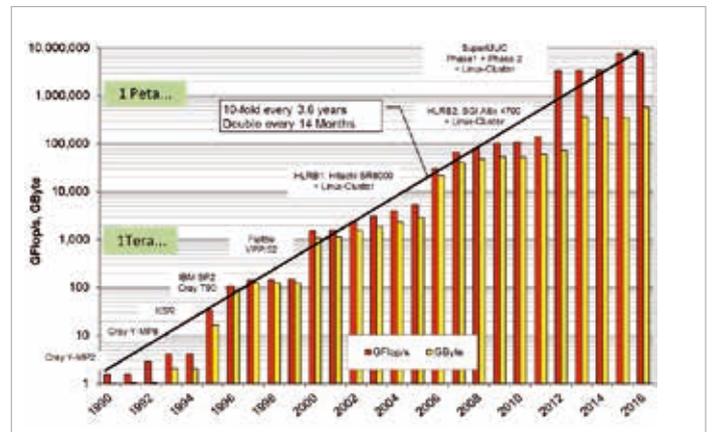
Computer verändern unser Leben seit den letzten 50 Jahren dramatisch. In jedem Smartphone steckt heute ähnlich viel Rechenleistung wie im ersten Höchstleistungsrechner Cray Y-MP2, den das Leibniz-Rechenzentrum (LRZ) der Bayerischen Akademie der Wissenschaften 1990 in Betrieb nahm. Seitdem stieg die Rechenleistung exponentiell an. Auch die Supercomputer am LRZ hielten mit dieser Entwicklung Schritt: von ca. 1 Gigaflop (1.000.000.000 Fließkommaoperationen pro Sekunde), die die Cray Y-MP2 am LRZ lieferte, auf die heute von SuperMUC insgesamt gelieferten 6,8 Petaflops (6.800.000.000.000.000 Fließkommaoperationen pro Sekunde), also auf das 6-Millionenfache.

Jedes Jahr bietet das LRZ am **Tag der Offenen Tür** die Möglichkeit, seine Rechner und das Zentrum für Virtuelle Realität und Visualisierung zu besichtigen, so auch am **21. Oktober 2017**.

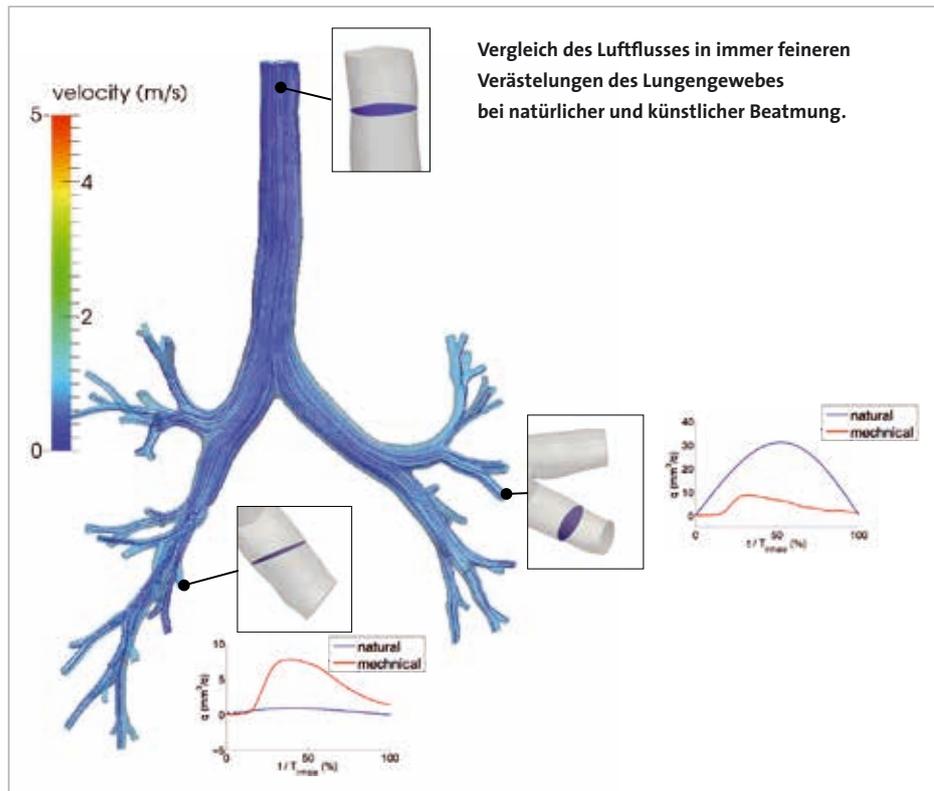
www.lrz.de/wir/tag-der-offenen-tuer



Die 2015 in Betrieb genommene Phase 2 des Höchstleistungsrechners SuperMUC am LRZ.



Die Rechenleistung der Höchstleistungsrechner am LRZ steigt seit 1990 exponentiell.



schen Universität München an die Intensivstation? Aber genau diese Forschungen können für jeden von uns lebensrettend sein.

Schneller, höher, stärker: auf dem Weg zu SuperMUC-NG

2018 läuft SuperMUC schon sechs Jahre! Dann ist es höchste Zeit für einen neuen Höchstleistungsrechner, denn für die Rechenleistung, die er bietet, sind dann die Betriebskosten zu hoch. So ist es eine gute Nachricht, dass Anfang 2017 die politischen Weichen gestellt wurden und nun alles Nötige in die Wege geleitet ist, damit bis dahin SuperMUC-NG, „SuperMUC Next Generation“, in Betrieb genommen werden kann.

Dann wird für Simulationen wieder ein Vielfaches an Prozessoren mehr zur Verfügung stehen – eine neue Herausforderung für die Programmierer. Denn es ist eine Wissenschaft für sich, viele Tausende Prozessoren dazu zu bewegen, gleichzeitig sinnvoll an demselben Problem zu rechnen (s. Kasten, S. 55). Dieses Wissen ist eine Voraussetzung dafür, dass immer leistungsfähigere Hardware und Software dabei helfen, nicht nur auf einem isolierten Supercomputer, sondern auch auf der Intensivstation und in der ärztlichen Praxis die für die jeweilige Person beste Medizin zu empfehlen. ■

Literatur

L. Pauling, R. B. Corey, Atomic Coordinates and Structure Factors for Two Helical Configurations of Polypeptide Chains, in: Proceedings of the National Academy of Sciences, vol. 37 (1951), 235.

L. Pauling, R. B. Corey, The Structure of Synthetic Polypeptides, in: Proceedings of the National Academy of Sciences, vol. 37 (1951), 241.

S. Wagner, A. Bode, H. Brühle, M. Brehm (Hrsg.), High Performance Computing in Science and Engineering, Garching/Munich 2016, ISBN 978-3-9816675-1-6, www.lrz.de/services/compute/supermuc/magazinesbooks oder QR-Code einscannen:



A. P. Bhati, S. Wan, D. W. Wright und P. V. Coveney, Rapid, Accurate, Precise, and Reliable Relative Free Energy Prediction Using Ensemble Based Thermodynamic Integration, in: Journal of Chemical Theory and Computation, vol. 13 (2017), 210; DOI: 10.1021/acs.jctc.6b00979.

A. P. Bhati, S. Wan, D. W. Wright und P. V. Coveney, Rapid and Reliable Binding Affinity Prediction of Bromodomain Inhibitors: A Computational Study, in: Journal of Chemical Theory and Computation, vol. 13 (2017); DOI: 10.1021/acs.jctc.6b00794.

L. Yoshihara, W. A. Wall, M. W. Gee, Die virtuelle Lunge, in: Akademie Aktuell 2/2012, 58.

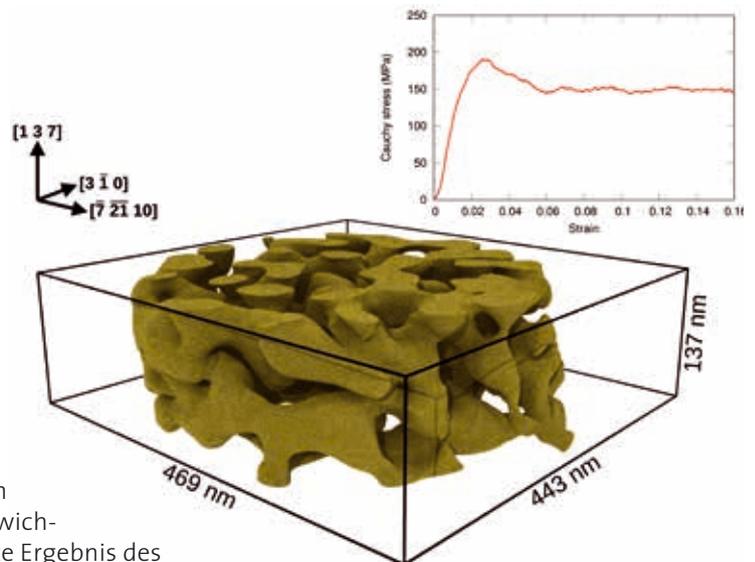
Einen Flohzirkus bändigen

Seit 2012 bietet SuperMUC 147.456 Prozessoren und weitere 86.016 seit 2015. Nicht nur die Lebenswissenschaften, sondern Vorhaben aus allen Disziplinen stehen vor der Frage: Wie bringt man so viele Prozessoren dazu, gleichzeitig effizient am selben Problem zu rechnen?

DAS BAYERISCHE Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst startete zeitgleich mit der ersten Beschaffung das Kompetenznetzwerk für wissenschaftliches Höchstleistungsrechnen in Bayern (KONWIHR) mit Mitteln aus der High-Tech-Offensive Bayern, um „die Nutzung von Hoch- und Höchstleistungsrechnern, insbesondere die des HLRB (Höchstleistungsrechner in Bayern), fachlich zu unterstützen und deren Einsatzpotential durch Forschungs- und Entwicklungsvorhaben auszuweiten.“

Unabhängig davon bietet das LRZ mit der „Partnership Initiative in Computational Science“ π^{CS} seit vielen Jahren umfassende Unterstützung an. Im Mai 2016 rief das „BioLab@LRZ“ zum „Summer of Simulation“, um junge Wissenschaftlerinnen und Wissenschaftler zu ermutigen, ihre Forschungsfragen aus Molekulardynamik und Quantenchemie auf dem SuperMUC zu bearbeiten. Acht Gruppen aus Bochum, Bonn, Erlangen-Nürnberg und München nutzten die vorlesungsfreie Zeit, um mit intensiver Unterstützung des LRZ ihre Programme auf einen richtig großen Rechner zu übertragen. Dabei ist es die zentrale Aufgabe, die Probleme und Programme so zu formulieren, dass die Tausenden von Prozessoren nicht unnötig viel aufeinander warten müssen, sondern so viele wie möglich gleichzeitig sinnvoll beschäftigt sind.

Bei einigen Projekten ging es um die Simulation der Dynamik großer Biomoleküle. Andere untersuchten chemische Reaktionen mit den heute häufig angewandten Dichtefunktional-Methoden z. B. an Metalloxyd-Oberflächen, die für technische Prozesse grundlegend sind. Eines der untersuchten Systeme dürfte eines der größten jemals simulierten sein: 450 Millionen Gold-Atome wurden, ausgehend von experimentellen nanoporösen Strukturen, untersucht. Die im SuperMUC simulierte Verformung durch Druck reproduzierte äußerst eindrucksvoll die experimentell gemessene Verformung.



Doch das wichtigste Ergebnis des „Summer of Simulation“ sind die Kenntnisse und Erfahrungen, die die Nachwuchswissenschaftler erwarben. Eine neue Organisation der Programmstruktur, geschickte Ausnutzung der Rechnerarchitektur und der Kommunikationsmuster zwischen den verschiedenen Prozessoren, die verschiedene Aufgaben übernehmen, sowie die Anwendung neuer Parallelisierungsverfahren konnten etliche der Programme auf nur ein Fünftel der ursprünglichen Laufzeit beschleunigen: „A little bit of brain time can save lots of computer time.“

Nanoporöses Gold, hier aus 450.000.000 Atomen, wird komprimiert. Der gemessene Effekt kann in der Simulation reproduziert werden.

Literatur

G. Mathias, Summer of Simulation 2016, in: Quartl Nr. 80, 4/2016, 21–27; https://www5.in.tum.de/quartl/2016/quartl80_n.pdf

DER AUTOR

Der Diplom-Chemiker Dr. Ludger Palm begann seine Arbeit für das Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften 2001 in der Gruppe Benutzerbetreuung für die Hochleistungsrechner. Seit 2006 ist er für die Öffentlichkeitsarbeit des LRZ verantwortlich.