

FORMVOLLENDET

Proteine steuern das gesamte Leben – wenn sie richtig gefaltet sind. Falsch gefaltete Eiweiße stehen hinter schweren Krankheiten wie Alzheimer und Parkinson. Mit Supercomputern versuchen Forscher, die Raumstruktur der Schlüssel­moleküle immer genauer vorherzusagen.



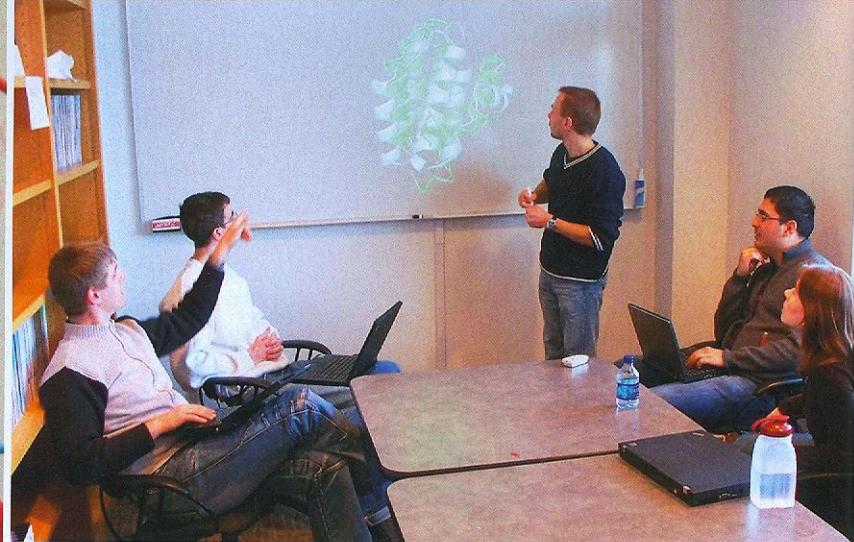
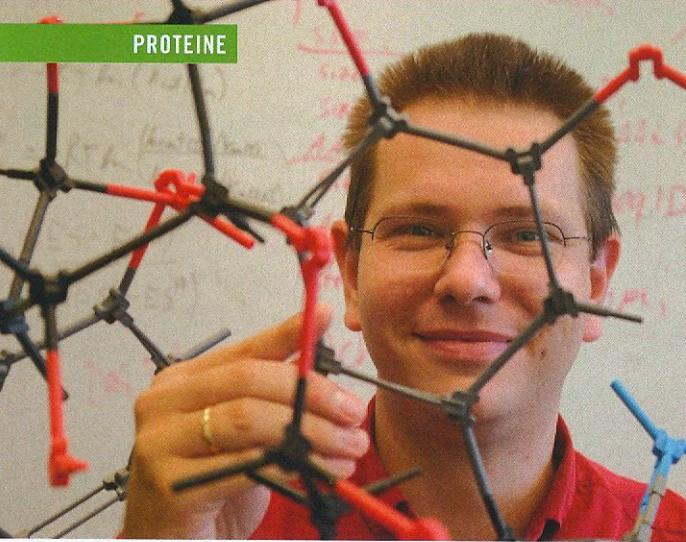


Ketten aus Aminosäuren haben die Tendenz, sich in Spiralen oder „beta-Faltblätter“ – hier durch Pfeile symbolisiert – zu formieren. Doch das ist erst die Sekundärstruktur. Wie sich Spiralen und Faltblätter im dreidimensionalen Raum umeinander winden, ist noch ein großes Rätsel.

KOMPAKT

- Die Entschlüsselung der dreidimensionalen Struktur von Proteinen ist so schwierig, dass Wissenschaftler erst von 0,7 Prozent der bekannten Eiweiße die Form kennen.
- Biologen und Informatiker setzen darum Grips und Computer ein, um mehr Eiweißstrukturen aufzuklären.
- In einem internationalen Vorhersage-Wettbewerb messen sie ihre Kräfte.

N. Speicher/Laguna Design



Professor Jens Meiler (links) und sein ehrgeiziges Doktoranden-Team in Nashville (rechts): Im Jahr 2008 gingen sie mit einem neuen, selbst geschriebenen Vorhersage-Programm an den Start zu einem internationalen Wettbewerb.

von Frederik Jötten

ES IST DER 5. MAI 2008, 14.30 Uhr in Nashville, Tennessee, als das Rennen beginnt. Draußen vor dem Fenster brennt die Südstaatensonne auf den Campus der Vanderbilt University. Hier drinnen in einem vollklimatisierten Raum voller Computer gehen die Wettkämpfer in Jeans und T-Shirt an den Start. Nils Wötzel, Julia Köhler, Nathan Alexander und Mert Karakas, alleamt Doktoranden Mitte 20, blicken auf einen der Bildschirme und warten. Endlich erscheint eine Abfolge von Buchstaben: MFSLRDAKC... Ein unaussprechliches Wort, 103 Buchstaben lang. Es bezeichnet ein Protein, also ein Eiweiß, das aus einer Kette von 103 Aminosäuren besteht. Mit diesem Signal beginnt der Wettbewerb, an dessen Ende die jungen Wissenschaftler aus der Gruppe des deutschen Strukturbiologen Jens Meiler vorhersagen wollen, wie dieses und mehr als 100 weitere Proteine aussehen. Mit ihnen wetteifern 236 Teams weltweit.

Die Struktur von Proteinen aufzuklären, ist eine der zentralen Fragen der modernen Biologie. Proteine steuern alle Lebensprozesse – bei Pflanzen, Tieren und Menschen. Ob es um die Kontraktion unserer Muskeln, die Verdauung von Nahrung oder unser Sehen, Hören und Schmecken geht, Proteine vermitteln fast alle Vorgänge in Lebewesen – auch bei Krankheitserregern. Die Reihenfolge der Aminosäuren von Proteinen zu bestimmen, ist heute kein Pro-

blem mehr. Aber nur bei 0,7 Prozent der Makromoleküle ist die dreidimensionale Struktur bekannt – und allein die entscheidet darüber, ob sie in den Lebewesen ihre Funktion erfüllen können.

Falsch gefaltete Proteine verursachen Krankheiten, das weiß man spätestens, seit die Ursachen für BSE und die Creutzfeld-Jakob-Krankheit bekannt sind. Aus harmlosen Eiweißen werden durch eine Veränderung der dreidimensionalen Struktur tödliche Prionen, die ihrerseits wieder harmlose Proteine in die gefährliche Form umwandeln. Prionen-Erkrankungen sind selten, aber sie zeigen, was Strukturveränderungen von Proteinen auslösen können. Weitaus häufiger sind Krankheiten, bei denen Proteine verklumpen, statt ihre richtige Form anzunehmen: die Alzheimer'sche und die Parkinson'sche Krankheit, an denen Millionen Menschen im Alter leiden, gehen auf solche Fehler in der Proteinfaltung zurück. Neueste Studien legen nahe, dass „kranke“ Proteine innerhalb des Körpers – ähnlich wie Prionen – herkömmliche Formen bestimmter Proteine umfalten.

Und dann gibt es noch die Erkrankungen, die von einzelnen Mutationen ausgelöst werden – Gen-Veränderungen, die ein funktionsuntüchtiges Protein zur Folge haben. Bekannt ist das Protein p53: „Wächter des Genoms“ wird es genannt, denn es hilft

dabei, Tumore zu verhindern. Treten Fehler bei der Verdoppelung der Erbsubstanz auf, hält es den Prozess an und sorgt dafür, dass das Reparatursystem in Gang kommt. Sind die Schäden in der DNA schwerwiegend, aktiviert es die Selbsttötung der Zelle. Wenn p53 aber durch eine Mutation falsch gefaltet ist, kann es das Genom nicht mehr überwachen. Zellen können sich dann unbeschränkt teilen – Krebs entsteht. Bei 50 Prozent aller menschlichen Tumore ist p53 mutiert.

ROSETTAS GRENZEN

Weltweit versucht man heute Proteinstrukturen aufzuklären, um die Ursachen von Krankheiten zu verstehen und neue Medikamente zu entwickeln. Allerdings binden 50 Prozent aller Arzneimittel an Proteine, die in der Zellmembran verankert sind – und bei diesen lässt sich die dreidimensionale Form noch schwerer bestimmen als bei löslichen Proteinen: Nur von 0,04 Prozent ist die Struktur bekannt.

Jens Meilers Arbeitsgruppe in Nashville ist entschlossen, das zu ändern. Der internationale Wettbewerb, bei dem die vier Doktoranden an den Start gehen, heißt „CASP“. Die englische Abkürzung steht für „Kritische Überprüfung von Techniken zur Vorhersage von Proteinstrukturen“. Er läuft so ab: Forscher, die eine Proteinstruktur experimentell bestimmt haben, halten die Veröffentlichung ihrer Daten zurück. Die Vor-

Target 409, ein Protein aus der Bakterien-Zellwand: Geht es nach dem Computerprogramm des Meiler-Teams, könnte es die Formen 1 bis 5 annehmen. Bild 6 zeigt in grau die reale Struktur, in Farbe die besten Vorhersage-Details des Teams aus Nashville.

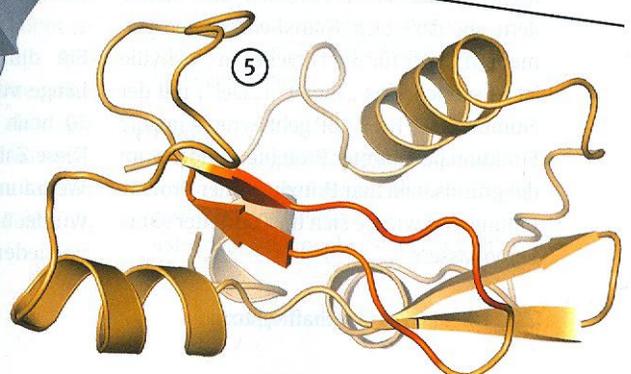
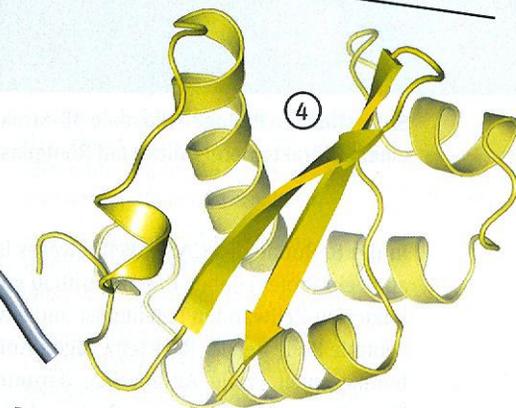
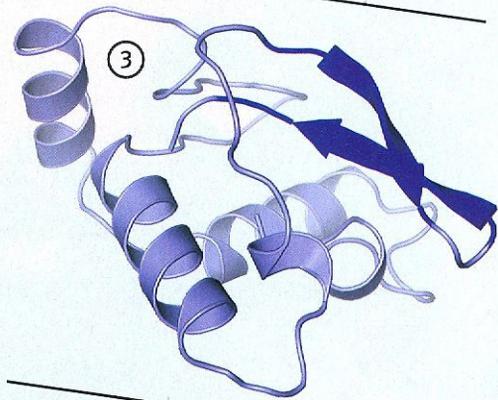
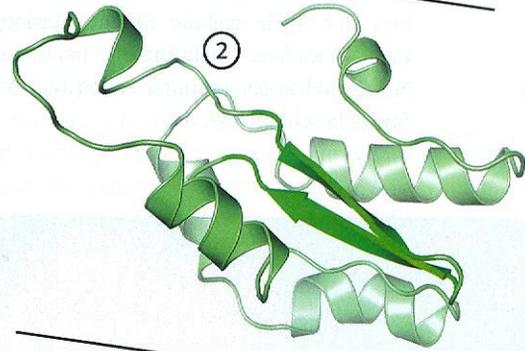
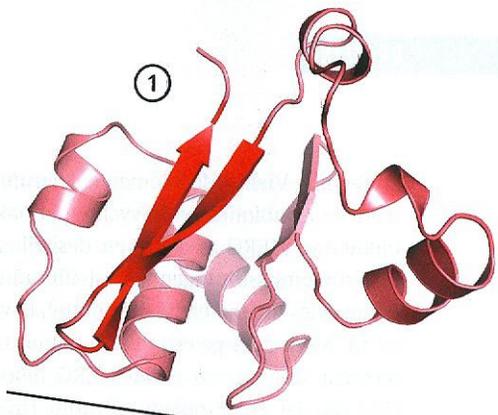
hersage-Teams bekommen die Buchstabenfolge, die sogenannte Aminosäuresequenz, erstellen mit Computer-Programmen eine hypothetische Struktur – und Wochen später werden Modelle und Realität miteinander verglichen. Auf einer Konferenz sieben Monate nach Beginn des Wettbewerbs wird verkündet, welche Teams am besten abgeschnitten haben.

Jens Meiler hat schon als Postdoktorand an CASP teilgenommen. Er arbeitete vier Jahre beim Erfinder der computergestützten Proteinstruktur-Vorhersage, David Baker, in Seattle. Dort war er beteiligt an der Entwicklung des bekanntesten und bislang erfolgreichsten Vorhersage-Programms: „Rosetta“. Mit Rosetta errang Meiler zwei erste Plätze bei CASP. Seine Arbeitsgruppe nimmt nun zum ersten Mal mit einem selbstentwickelten Programm teil. „Bei Membranproteinen und bei Proteinen, die keinem bisher bekannten Molekül ähneln, hat

Rosetta seine Grenzen“, sagt Meiler. „Die wollen wir mit unserem Programm überwinden.“

KANAL IM HERZEN

Ein Beispiel für ein Membranprotein ist HERG, das in Herzmuskelzellen vorkommt. Es bildet einen Kanal zwischen dem Inneren und dem Äußeren der Zelle. Durch ihn strömen nach der Kontraktion Kalium-Ionen aus der Zelle, eine Voraussetzung dafür, dass das Herz bald wieder Blut pumpen kann. Funktioniert der Kanal nicht richtig, kann es zu Herzrhythmusstörungen



kommen. Viele Medikamente, darunter manche Antibiotika und Psychopharmaka, binden an HERG und führen deshalb zu Herzrhythmusstörungen. Deshalb muss heute jedes potenzielle Arzneimittel, bevor es am Menschen getestet wird, darauf untersucht werden, ob es an HERG bindet. Und das ist ein Problem: Obwohl HERG eines der meistuntersuchten Proteine ist, ließ sich seine genaue Struktur bislang nicht aufklären. Algorithmen, neuronale Netze und Supercomputer sollen nun die Suche beschleunigen.

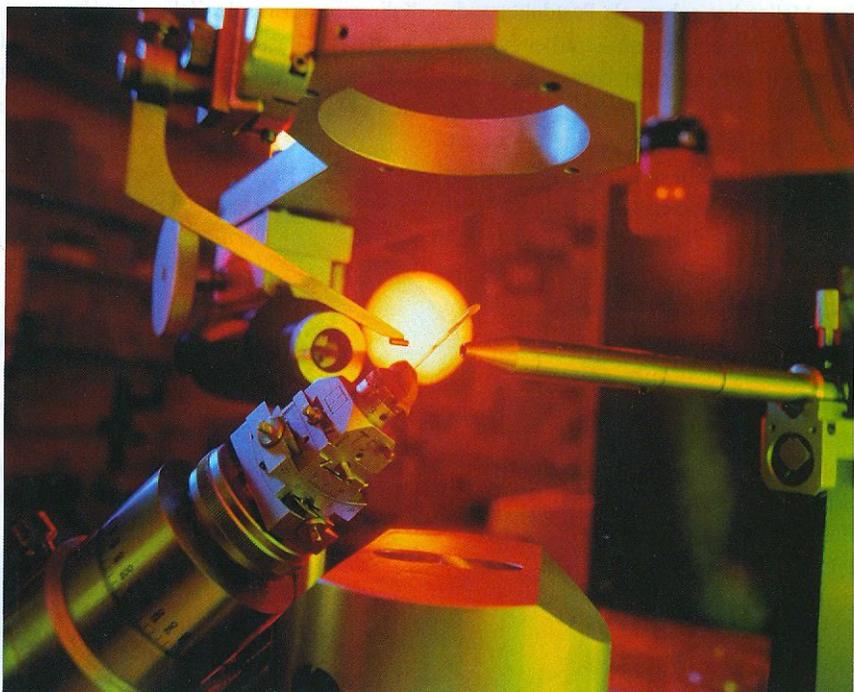
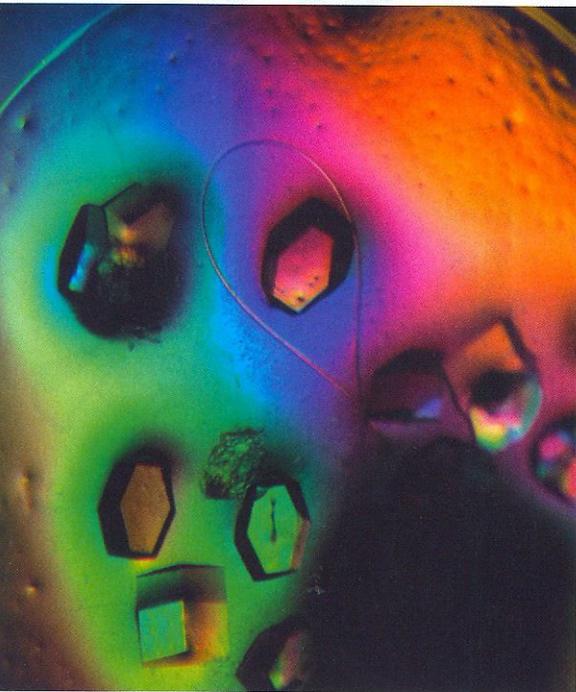
Nils Wötzel startet das Programm, mit dem die Struktur vorhergesagt werden soll. Der Supercomputer, der einen Raum in einem Nachbargebäude füllt und etwa die gleiche Leistung hat wie 2000 herkömmliche PCs, beginnt zu rechnen. In wenigen Tagen werden die ersten Modelle vorliegen.

Proteine gehören zu den komplexesten Molekülen überhaupt. Sie enthalten bis zu 100 000 Atome und bilden die verschiedensten Strukturen – vom reißfesten Baumaterial Kollagen bis zum variablen Anti-

hat die Natur nur wenige dieser Kombinationen verwirklicht – und zwar solche, die in stabilen dreidimensionalen Formen in Lebewesen vorkommen: zum Beispiel in einer Grundform und davon abweichenden Varianten, die sich durch die Wechselwirkung mit anderen Molekülen ausbilden.

WUNSCHTRAUM SEIT JAHRZEHNEN

Proteine falten sich schon während der Synthese. Sie gehorchen dabei inneren Kräften, die sie in den energetisch günstig-



So bestimmen Biologen die reale 3D-Struktur eines Proteins: Sie lassen es Kristalle bilden (links), bombardieren ein solches Kristall in einem Diffraktometer (Mitte) mit Röntgenstrahlen und bestimmen aus dem Beugungsmuster (rechts) die Struktur.

In der 8. Runde des CASP-Wettbewerbs ist auch ein Protein dabei, das am Aufbau von Bakterien-Zellwänden beteiligt ist und eine zentrale Rolle bei der Resistenz gegen Antibiotika spielt. Viele Antibiotika, darunter Penicillin, binden an dieses Protein, hemmen dadurch seine Funktion und verhindern so, dass sich Krankheitserreger vermehren. Aber für die Forscher in Nashville ist es schlicht das „Target“ („Ziel“) mit der Nummer 409. Bei CASP geht es nicht um die Funktion bestimmter Proteine, sondern um die grundsätzlichen Prinzipien der Proteinfaltung und wie sie sich im Computer simulieren lassen.

körper. Auf einer Tafel, die neben dem Computer an der Wand hängt, sind mit Filzstift die Strukturformeln der 20 Aminosäuren geschrieben, aus denen Organismen Eiweiße aufbauen. Alle haben einen identischen Teil – und einen sogenannten Rest, in dem sie sich unterscheiden.

Ein durchschnittliches Protein hat eine Länge von 300 Aminosäuren, es gibt also 20 hoch 300 Kombinations-Möglichkeiten. Diese Zahl ist so groß, dass es im gesamten Weltraum nicht genügend Atome geben würde, um auch nur ein einziges Exemplar von jedem Molekül herzustellen. Trotzdem

ten Zustand zwingen. Allen diesen Kräften ist gemeinsam: Sie sind bereits in der Aminosäuresequenz angelegt. Seit Jahrzehnten haben Wissenschaftler deshalb den Wunschtraum, allein anhand der Abfolge der Aminosäuren die Struktur eines Proteins vorherzusagen. Doch trotz vieler Jahre intensiver Forschung sind die Computermodelle bislang nur Annäherungen. Einzelne Atome innerhalb eines Proteins haben einen Durchmesser von etwa einem Ångström (einem 10 Milliardstel Meter). Die besten Schätzungen liegen im Durchschnitt bis 5 Ångström daneben, mittelmäßige bis 10 und schlechte 20 und mehr.

Zeitschrift Science die Struktur der Diacylglycerol-Kinase verkünden konnten, eines wichtigen Stoffwechsel-Enzyms, das in der Zellmembran sitzt. Die Schwierigkeiten bei der Aufklärung von Proteinstrukturen zeigen, welchen Aufwand man sich ersparen könnte, wenn eine verlässliche Vorhersage am Computer möglich wäre.

ROTE SPIRALEN UND GRÜNE PFEILE

Nashville, drei Wochen nach dem Beginn des Wettbewerbs: Besprechung der Wissenschaftler vor einem Bildschirm. Aus der Buchstaben-Folge sind dreidimensionale Strukturen geworden. Zehn verschiedene Vorhersagen von Target 409 hat die Gruppe mit verschiedenen Methoden erstellt. Jetzt diskutieren die Forscher, welche eingereicht werden sollen. Auf dem Bildschirm erscheint zunächst ein Bündel aus roten Spiralen und grünen Pfeilen. Sie stellen die sogenannten Sekundärstrukturen des Proteins dar. Bevor die Aminosäurekette sich in ihre räumliche Struktur faltet, windet sie sich entweder wie eine Wendeltreppe zu einer Alpha-Helix (in der Computersimulation: Spirale) oder bildet eine Fläche, das beta-Faltblatt (Pfeil). Auf dem Bildschirm sieht es sehr unübersichtlich aus. „Das wird hart, aber wir versuchen es“, sagt Nils Wötzel. Das Molekül ist mit 103 Aminosäuren zwar nicht sehr groß, besteht aber aus vielen Faltblättern, die viel schwerer zu simulieren sind als Helices.

Der Doktorand dreht das Protein am Bildschirm und ruft dann eine zweite Vorhersage auf. Die sieht etwas anders aus als die erste. Die zweite Prognose weist an einer Stelle ein Faltblatt auf, wo die erste eine Helix zeigt. Drei nachdenkliche Gesichter spiegeln sich auf dem Bildschirm. „Wir müssen entscheiden, welche Modelle wir einreichen“, drängt Nathan Alexander. Fünf Modelle pro Target dürfen am Wettbewerb teilnehmen. „Das hier sieht einem Protein am ähnlichsten“, sagt Julia Köhler und deutet auf eines der bunten Knäuel. „Bei diesem Modell ragt die Helix nicht so heraus wie bei dem anderen Modell.“ Die Gruppe beschließt, Köhlers Favoriten einzureichen – und das Modell, das sich von ihm am stärksten unterscheidet. „Wir wollen ja bei dem Wettbewerb etwas lernen, und am besten lernt man aus Fehlern“, meint die Doktorandin.

Sieben Monate später, im Dezember 2008, werden die Ergebnisse präsentiert. Die Computer-Simulationen von Target 409 aller beteiligter Gruppen unterscheiden sich stark von der echten Struktur. Es ist noch ein weiter Weg, bis man die Natur der Proteinfaltung per Computer voraussehen kann, das wird hier einmal mehr klar. Die besten Simulationen liegen 3,1 Ångström daneben, die der Meiler-Gruppe 13,5. Das bedeutet für dieses Protein Platz 75 unter 129 Gruppen, die dieses Target bearbeitet haben. Es gibt auch Moleküle, bei denen die Forscher aus Nashville besser abschneiden: Bei Target 396 liegen sie nur 4 Ångström daneben. Insgesamt rangieren die Wissenschaftler aus Nashville im Mittelfeld. Enttäuscht? „Nein. Wir haben mit unserem neuen Programm das erste Mal beim Wettbewerb teilgenommen. Ausgereift ist es noch nicht. Beim nächsten Mal landen wir weiter vorn“, sagt Wötzel. Die Vorbereitungen laufen längst. Am 3. Mai 2010 beginnt CASP 9, der neue Wettbewerb. ■



FREDERIK JÖTTEN ist Biologe und freier Journalist. Während des Studiums mühte er sich vergebens, ein Protein des Malaria-Erregers zu isolieren.

MEHR ZUM THEMA

LESEN

Dietmar Zimmer

JENSEITS DER GENE

Proteine – Schlüssel zum Verständnis des Lebens

Klett Cotta, Stuttgart 2005, € 19,50 (nur noch antiquarisch erhältlich)

John W. Moore und Tina Blasche

BIOCHEMIE FÜR DUMMIES

Wiley-VCH, Weinheim 2009, € 19,95

Vivienne Baillie Gerritsen

TALES FROM A SMALL WORLD

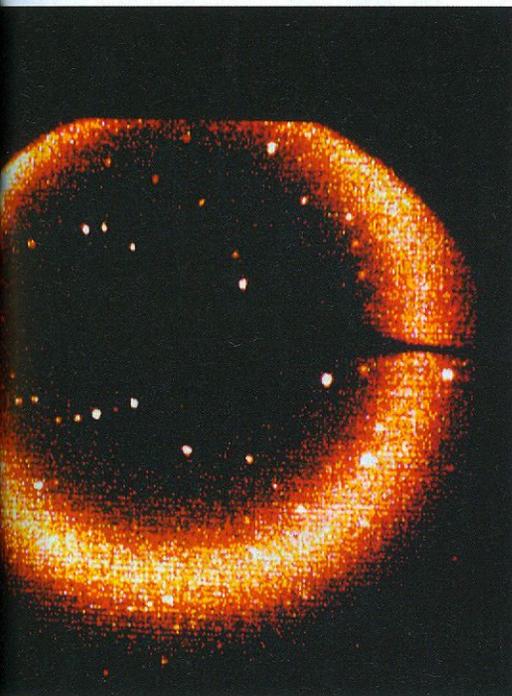
Swiss Institute of Bioinformatics 2009

Zu beziehen über www.proteinspotlight.org

Auf dieser Website wird jeden Monat ein Protein populärwissenschaftlich vorgestellt. In dem Buch sind Texte aus der Serie gesammelt.

INTERNET

Wie Laien sich am virtuellen Proteinfalten beteiligen können, erfahren Sie auf www.wissenschaft.de unter dem Menüpunkt „Aktuelles Heft“.



V. Steger/SPL/Agentur Focus; J. King-Holmes/OCMS/SPL/Agentur Focus (2)

Es gibt heute vor allem zwei Verfahren, Proteinstrukturen experimentell aufzuklären. Die Kernresonanzspektroskopie und die weitaus am häufigsten genutzte Röntgenstrukturanalyse. Bei Letzterer wird zunächst ein Proteinkristall hergestellt. Durch dieses wird dann Röntgenlicht geschickt – und anhand der Beugungsmuster lässt sich die Struktur berechnen. Aber viele Proteine, besonders solche, die in Membranen sitzen, haben bis heute keine Kristalle gebildet. Bei ihnen kann die Kernresonanzspektroskopie angewendet werden. Doch das ist ein sehr langwieriges Verfahren, wie ein aktuelles Beispiel belegt: 13 Jahre brauchte eine weltweite Kooperation von Forschern, bis diese im Juni 2009 in der

PROTEINS

>>*Life & Environment*

PERFECTLY STRUCTURED

Proteins control the entire life – if they are folded properly. Incorrectly folded proteins are responsible for serious illnesses such as Alzheimer's and Parkinson's disease. Researchers are using supercomputers for the ever more accurate prediction of the spatial structure of key molecules.

Amino acid chains have a tendency to form into spirals or “beta-pleated sheets” – symbolized with arrows here. However, this is only the secondary structure. How spirals and pleated sheets are winding around each other remains a mystery.

IN A NUTSHELL

- Decoding the three-dimensional structure of proteins is so difficult that researchers have only identified the shape of 0.7 percent of known proteins.
- Therefore, biologists and computer scientists are using savvy and computers to shed light on more protein structures.
- They are pitting their strength against each other in an international prediction competition.

PROTEINS

Professor Jens Meiler (left) and his ambitious team of PhD students in Nashville (right): in 2008, they entered an international competition with a new self-developed prediction program.

IT IS MAY 5, 2008, 02:30 pm in Nashville, Tennessee when the race begins. Outside the window, the sun of the southern state is scorching the Vanderbilt University campus. On the inside, in a comfortably air-conditioned room full of computers, the competitors wearing jeans and T-shirts are ready to get started. Nils Wötzel, Julia Köhler, Nathan Alexander and Merr Karakas, all PhD students in their mid-twenties are staring at one of the monitors, waiting. Finally, a sequence of letters appears: MFSLRDAKC... An unpronounceable word composed of 103 letters. It identifies a protein consisting of a chain with 103 amino acids. This is the starting signal of the competition at the end of which the young scientists belonging to the group of German structural biologist Jens Meiler are hoping to predict what this and more than 100 other proteins look like. They are joined by 236 other competing teams from around the world.

One of the main challenges of modern biology is to define the structure of proteins. Proteins control all processes of life – in plants, animals and humans. Whether it relates to the contraction of our muscles, the digestion of food or our sense of seeing, hearing and tasting, proteins mediate almost all processes in living creatures – including pathogens. These days, defining the order of the amino acids of proteins is no longer a

problem. But the three-dimensional structure is only known for 0.7 percent of macromolecules – and this structure is the sole decisive factor for whether they can fulfill their function in living creatures.

Incorrectly folded proteins cause illnesses; this is generally known at least since the causes for BSE and the Creutzfeldt-Jakob disease have been identified. The modification of their three-dimensional structure turns harmless proteins into deadly prions which in turn convert harmless proteins into the dangerous form. Prion diseases are rare, but they illustrate the potential consequences of structural alterations of proteins. Illnesses in which proteins clump together rather than adopting their proper shape are much more common: the Alzheimer's and Parkinson's diseases affecting millions of people as they get older are the result of this defective protein folding. Recent studies suggest that "sick" proteins are refolding traditional shapes of certain proteins within the body – similar to prions.

And then there are the disorders triggered by isolated mutations – genetic changes resulting in a non-functional protein. Protein p53 is known as the "master watchman of the genome" for its ability to prevent tumors. If errors in the duplication of the genotype occur, it

by Frederik Jötten
stops the process and makes sure that the repair system is initiated. If the damages involving the DNA are severe, it activates the self-destruction of the cell. However, if p53 is incorrectly folded due to a mutation, it is no longer able to monitor the genome. Cells can then divide in an uncontrolled way – cancer develops. p53 is mutated in 50 percent of all tumors involving humans.

ROSETTA'S LIMITATIONS

Attempts to determine protein structures are ongoing around the world, to understand the causes of illnesses and to develop new medications. However, 50 percent of all drugs bind to proteins integrated in the cell membrane – and for these, it is even more difficult to determine the three-dimensional shape compared to soluble proteins: the structure is only known for 0.04 percent.

Jens Meiler's working group in Nashville is committed to changing this. The international competition the four PhD students are participating in is called "CASP". The English abbreviation stands for "Critical Assessment of Techniques for Protein Structure Prediction". The procedure is as follows: Researchers who defined a protein's structure experimentally are holding off with the publication of their data. The prediction teams are provided with the letter sequence, the so-called

Target 409, a protein from the bacterial cell wall: According to the computer program of Meiler's team, it could adopt the shapes 1 to 5. Picture 6 illustrates the actual structure in grey along with the details of the best predictions from the Nashville team in color.

amino acid sequence, use the computer programs to create a hypothetical structure – and the models and reality are compared weeks later. At a conference held seven months after the start of the competition, it will be announced which teams were most successful.

Jens Meiler already participated in CASP as a post doc. He worked for the inventor of the computer-aided protein structure prediction tool, David Baker in Seattle for four years. While there, he was involved in the development of the best known and thus far most successful prediction program: "Rosetta". With Rosetta, Meiler won first place in CASP twice. His working group is now participating with a self-developed program for the first time. "For membrane proteins and proteins not resembling any previously known molecules,

Rosetta's capabilities are limited", says Meiler. "Our program is designed to overcome these limitations."

CHANNEL IN THE HEART

HERG is an example of a membrane protein; it occurs in cells of the heart muscle and forms a channel between the inside and the outside of the cell. Potassium ions are flowing out of the cell through this channel after the contraction, a requirement for the heart to be able to pump blood soon again. If the channel does not work properly, it may result in

PROTEINS

arrhythmia. Numerous drugs, including some antibiotics and antipsychotic drugs bind to HERG, thus causing arrhythmia. Therefore, every potential drug must now be investigated first to find out whether it binds to HERG before it can be tested in humans. And this is a problem: although HERG is one of the most studied proteins, the determination of its exact structure has been impossible to this day. Algorithms, neuronal networks and supercomputers are currently being used to speed up the search.

Nils Wötzel starts the program designed to predict the structure. The supercomputer which fills an entire room in the neighboring building and has a capacity comparable to 2000 traditional PCs starts the calculation. Preliminary models will be available in a few days.

Proteins are part of the most complex molecules overall. They contain up to 100,000 atoms and create the most diverse structures – ranging from the tear-resistant construction material collagen to the variable antibody.

only realized a few of these combinations – namely ones that occur in living creatures in stable three-dimensional shapes: for example in a basic shape and in deviating variations created by the interaction with other molecules.

CENTURY-OLD DREAM

Proteins fold as early as during synthesis. In doing so, they are driven by inner energies forcing them into the most favorable energetic status. These forces have

This is how biologists determine the real 3D structure of a protein: they allow crystals to form (left), bombard such a crystal with X-rays in a diffractometer (center) and determine the structure from the diffraction pattern (right).

The 8th round of the CASP competition also includes a protein which is involved in the structure of bacterial cell walls and plays a key role in the resistance against antibiotics. Many antibiotics, including penicillin, bind to this protein, thus inhibiting its function and preventing the pathogens from propagating. However, for the researchers in Nashville, this is simply “target” number 409. CASP is not about the function of certain proteins, but rather about the basic principles of protein folding and how they can be computer-simulated.

The structural formulas of the 20 amino acids from which organisms establish proteins are written with marker on a board hanging on the wall next to the computer. All of them have an identical part – and a so-called residue which makes them different.

The length of an average protein is 300 amino acids, in other words, there are 20 to the power of 300 possible combinations. This figure is so large that the number of atoms in the entire universe would not suffice to create only a single sample of each molecule. Nevertheless, nature

one thing in common: they are already present in the amino acid sequence. Therefore, researchers have been dreaming for centuries to predict the structure of a protein solely based on the sequence of the amino acids. But in spite of years of intensive research, computer models are only approximations so far. Individual atoms within a protein have an estimated diameter of one angstrom (1.0×10^{-10} meter). On average, the best estimates are off by up to 5 angstroms, moderate estimates by 10 and poor ones by 20 and more.

These days, mainly two methods are used for the experimental determination of protein structures. They are the nuclear magnetic resonance spectroscopy and the by far most commonly used structural X-ray analysis. For the latter, a protein crystal is first produced. X-ray light is sent through the crystal and the structure can be calculated based on the diffraction pattern. Yet many proteins, particularly ones sitting in membranes have not formed any crystals to this day. Nuclear magnetic resonance spectroscopy can be used for those. However, this is a very tedious procedure as illustrated by a recent example: a worldwide cooperation of researchers needed 13 years, until they were finally able to publish

the structure of Diacylglycerol kinase, one of the important metabolic enzymes sitting in the cell membrane, in the *Journal Science* in June 2009. The difficulties involved in the determination of protein structures illustrate the scope of efforts that could be saved if a reliable prediction by means of the computer were possible.

RED SPIRALS AND GREEN ARROWS

Nashville, three weeks after the start of the competition: discussion between the researchers in front of a monitor. The letter sequence turned into three-dimensional structures. The group compiled ten different predictions of target 409 using a variety of methods. The researchers are now discussing which ones to submit. A bundle of red spirals and green arrows first appears on the monitor. They represent the so-called secondary structures of the protein. Before the amino acid chain folds into its spatial structure, it either winds into an alpha helix similar to a spiral staircase (in the computer simulation: a spiral) or forms a surface, the beta-pleated sheet (arrow). It looks very confusing on the monitor. "This will be difficult, but we will try", says Nils Wötzel. With 103 amino acids, the molecule is not very big indeed, but it consists of many pleated sheets which are much more difficult to simulate than helices.

The doctoral candidate turns the protein on the monitor and opens a second prediction. It looks slightly different than the first one. The second prediction shows a pleated sheet in the location where the first one shows a helix. Three thoughtful faces are reflected on the monitor. "We have to decide which models we want to submit", urges Nathan Alexander. Five models per day are eligible to participate in the competition. "This one most resembles a protein", says Julia Köhler, pointing to one of the colorful tufts. "In this model, the helix is not protruding as much as in the other model." The group decides to submit Köhler's favorite – along with the model that deviates most from it. "The competition is about learning something, and it is best to learn from mistakes", the doctoral candidate explains.

Seven months later, in December 2008, the results are presented. The computer simulations of target 409 from all the participating groups greatly deviate from the real structure. Once again it becomes blatantly obvious that it is still a long way until the nature of protein folding can be predicted by computers. The best simulations are off by 3.1 angstroms, the one from the Meiler group is off by 13.5. In terms of this protein, this means rank 75 of 129 groups who worked on this target. There are other molecules, for which the researchers from Nashville scored better: for target 396, they are only off by 4 angstroms. Overall, the researchers from Nashville are ranked in the middle. Disappointed? "No. We participated for the first time in the competition with our new program. It is not yet mature. Next time, we will do better", says Wötzel. CASP 9, the new competition will start on May 3, 2010.

FREDERIK JÖTTEN is a biologist and freelance journalist. While pursuing his studies, he attempted in vain to isolate a protein of the malaria pathogen.

MORE ABOUT THE TOPIC

READING MATERIAL

Dietmar Zimmer
JENSEITS DER GENE
Proteine – Schlüssel zum Verständnis des Lebens
(BEYOND GENES
Proteins – the key to understanding life)
Klett Cotta, Stuttgart, 2005, €19.50 (available from second hand bookshops only)

John W. Moore and Tina Blasche
BIOCHEMIE FÜR DUMMIES
(BIOCHEMISTRY FOR DUMMIES)
Wiley-VCH, Weinheim 2009, €19.95

Vivienne Baillie Gerritsen
TALES FROM A SMALL WORLD
Swiss Institute of Bioinformatics 2009
Available from www.proteinspotlight.org
A protein is introduced in popular science terms every month on this website. The book contains a collection of texts from this series.

INTERNET

Learn how lay persons can get involved in the virtual protein folding by visiting www.wissenschaft.de under menu item "Current Issue".