

Proteinfaltung am Computer

Die vor kurzem vollendete Sequenzierung des menschlichen Genoms stellt einen Meilenstein in der biomedizinischen Forschung dar. Dennoch kann von einer eigentlichen Entschlüsselung des Genoms keine Rede sein. Erst die Aufklärung der dreidimensionalen Struktur und der Funktion der Gene, der Proteine, wird es ermöglichen, die Information, die in den Genen codiert vorliegt, vollständig zu verstehen und dieses Wissen für medizinische Zwecke effizient zu nutzen. Computergestützten Methoden, die Voraussagen über die Struktur von Proteinen und ihre Wechselwirkungen mit Medikamenten ermöglichen, wird dabei eine Schlüsselrolle zukommen.

VON JÖRG GSPONER, PETER KOLB
UND AMEDEO CAFLISCH

Proteine (Eiweisse) sind Makromoleküle, die durch ihr gegenseitiges Wechselspiel die Abläufe in und ausserhalb der Zellen im ganzen Organismus steuern und bestimmen. Pfortnermoleküle an der Zelloberfläche, die nur bestimmten Stoffen Einlass gewähren, Kontrollproteine, welche die Produktion im Zellinnern überwachen, oder etwa Abwehrmoleküle, so genannte Antikörper, welche vor Eindringlingen schützen, haben unterschiedliche Aufgaben zu erfüllen. Die genaue Funktion, die dabei jedem Protein im komplexen «Zellbetrieb» zukommt, ist eng mit seiner spezifischen dreidimensionalen Struktur verknüpft.

Durch die Kenntnis der genauen Struktur und Funktionsweise der Proteine können ihnen Moleküle auf den Leib geschneidert werden, die mit ihnen wechselwirken und sie in ihrer Aktivität bestärken oder auch bremsen können. Da die grosse Mehrheit der Zielmoleküle, mit denen Medikamente interagieren, Proteine sind, ist die Strukturaufklärung daher von enormer medizinischer und wirtschaftlicher Bedeutung.

Proteinstrukturen können heute mittels Röntgenkristallographie und multidimen-

sionaler NMR-Spektroskopie (eine Methode, für deren Entwicklung Professor Kurt Wüthrich von der ETH den diesjährigen Chemienobelpreis erhielt) bestimmt werden. Weltweit ins Leben gerufene «Structural Genomics»-Programme verfolgen das Ziel, möglichst schnell von allen wichtigen Proteinstrukturen ein Muster zu ermitteln. Die oben genannten Methoden bedürfen aber grosser Proteinmengen und müssen noch verbessert werden, um schneller zu Strukturinformationen zu kommen. Deshalb wird heute intensiv an der Weiterentwicklung von Computerprogrammen und Algorithmen gearbeitet, welche die Strukturbildung, das so genannte Falten der Proteine, simulieren oder voraussagen können.

Dies ist kein triviales Unterfangen, da die Gene keine für den Menschen verständliche Bauanleitung für die Architektur der Proteine enthalten. Sie geben nur an, in welcher Reihenfolge die verschiedenen Bausteine (Aminosäuren) in einem Protein angeordnet sind. Diese für jeden Proteintyp spezifische Aminosäuresequenz scheint aber trotzdem all die Information zu enthalten, die es braucht, damit ein Protein schnell seine charakteristische, dreidimensionale Struktur findet. Dies ist umso erstaunlicher, als einfache statistische Überlegungen dieser Tatsache scheinbar widersprechen.

Grob vereinfacht könnte man Proteine mit einem Konstrukt aus Legosteinen vergleichen, in dem die verschiedenen Bausteine unterschiedlicher Länge und Farbe den Aminosäuren entsprechen. Bildet man beispielsweise aus drei Legosteinen eine rot-blau-grüne Kette, so hat man zuerst die Möglichkeit, den blauen Stein quer oder längs, oben oder unten mit dem roten Anfangsstein zu verbinden. Der nächste, grüne, Legostein kann nun auf die gleichen vier Arten mit dem blauen verknüpft werden. Dies ergibt insgesamt vier mal vier mögliche Strukturen für eine rot-blau-grüne Legokette. Hat man aber – an Stelle von nur drei – hundert Legosteine aneinander zu fügen, ergeben sich 4^{99} ($\approx 10^{60}$) Möglichkeiten.

Selbst wenn man äusserst flink wäre und für die Umplatzierung eines einzelnen Bausteins nur ein Milliardstel einer Sekunde bräuchte, wäre ein Leben nicht lang genug, um alle möglichen Anordnungen einmal

Jörg Gsponer ist MD/PhD, Peter Kolb Doktorand in der Forschungsgruppe von Professor Amedeo Caflisch am Biochemischen Institut der Universität Zürich.

herzustellen. Ähnliche Überlegungen lassen sich auch für Proteine anstellen. Es würde daher keinen Sinn ergeben, ein Programm zu schreiben, das alle möglichen dreidimensionalen Anordnungen einer Aminosäurekette evaluiert.

Zielstrebige Konformation

Stattdessen hat man in den letzten 30 Jahren so genannte Moleküldynamik-Program-

me entwickelt, mit denen die Newtonschen Bewegungsgleichungen für Moleküle gelöst werden können. Ausgehend von der initialen Position und Geschwindigkeit eines jeden Atoms in einem Protein können seine potenzielle und kinetische Energie und damit auch seine Bewegung im Verbund mit anderen Atomen berechnet werden. Solche Moleküldynamik-Programme haben es unserer und anderen Forschungsgruppen ermöglicht, die Faltung sowohl einer ganzen Reihe von Strukturmotiven, die häufig in Proteinen vorkommen, als auch kleinerer Proteine zu simulieren.

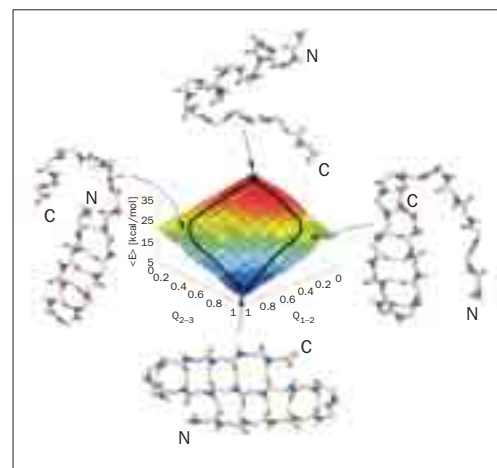


Abb. 1:
Energiefläche für die Faltung eines Proteinmotivs. Die Fläche weist eine gewisse Symmetrie auf, da auch die gefaltete Struktur (unten) symmetrisch ist.

me entwickelt, mit denen die Newtonschen Bewegungsgleichungen für Moleküle gelöst werden können. Ausgehend von der initialen Position und Geschwindigkeit eines jeden Atoms in einem Protein können seine potenzielle und kinetische Energie und damit auch seine Bewegung im Verbund mit anderen Atomen berechnet werden. Solche Moleküldynamik-Programme haben es unserer und anderen Forschungsgruppen ermöglicht, die Faltung sowohl einer ganzen Reihe von Strukturmotiven, die häufig in Proteinen vorkommen, als auch kleinerer Proteine zu simulieren.

Dank dieser Simulationen und anderer theoretischer Studien gelang es herauszufinden, wie es ein Protein – bestehend aus beispielsweise 100 Aminosäuren – in einigen Sekunden oder noch schneller schaffen kann, unter all den möglichen dreidimensionalen Anordnungen (den so genannten Konformationen) diejenige zu finden, welche für dieses Protein die richtige ist. Unterschiedliche Proteinkonformationen, wie auch verschiedene Legokonstrukte, unterscheiden sich in ihrer Stabilität. Wechselwirkungen zwischen den Atomen der Aminosäuren führen dazu, dass einige Strukturformen energetisch günstiger sind als andere. Man nimmt heute deshalb an, dass

die energetisch günstigste aller möglichen Konformationen auch die richtige, weil funktionstüchtige, ist. Proteine bewegen sich bei ihrer Faltung auf so genannten Energieflächen, auf denen jeder einzelne Punkt einer anderen Konformation entspricht. Bei der «Fahrt» auf dieser Energiefläche versuchen die Proteine, möglichst schnell den energetisch günstigsten, das heisst tiefsten, Punkt zu erreichen (siehe Abbildung 1). Die grosse Mehrzahl der möglichen Konformationen nehmen die Proteine bei ihrer Faltung gar nicht ein, da sie zielstrebig «talwärts fahren». Vergleichen lässt sich dies mit einem Wassertropfen, der seinen Weg vom Berg ins Tal finden muss. Theoretisch hätte der Tropfen eine astronomisch grosse Anzahl von möglichen Wegen ins Tal zur Verfügung. Da er aber den Drang hat, stets nach unten zu fließen, wird er einen Weg finden, der ihn schnell in die Talsohle führt und auf dem er nur mit einem Bruchteil der Bergoberfläche in Kontakt kommen wird. Neue Erkenntnisse über das Faltungsverhalten von Proteinen und deren Dynamik ergeben sich vermehrt auch aus der Kombination von experimentellen Daten mit Moleküldynamik-Simulationen. Man baut Informationen, die aus Experimenten gewonnen werden, in Simulationen ein, um gezielt Eigenschaften und Verhaltensmuster von Proteinen zu studieren. Durch die Zusammenarbeit zwischen experimentell und theoretisch arbeitenden Forschern ergeben sich nicht nur neue Erkenntnisse, sondern auch neue Fragestellungen und Ansatzpunkte für weitere Experimente und Simulationen.

Diese fruchtbare Zusammenarbeit wird es auch ermöglichen, die bestehenden Programme weiter zu verbessern. Mit der Hilfe von leistungsstarken Computern wird man dann die Faltung von grossen Proteinen simulieren und somit ihre Struktur voraussagen können. Da für die Faltungen nur die Aminosäuresequenz eines Proteins – also die Information, die von der Gensequenzierung her bekannt ist – benötigt wird, könnten Simulationen in ferner Zukunft eine wichtige Rolle bei der Strukturvoraussage spielen.

Computerdesign für Medikamente

Beim computergestützten Entwerfen von Medikamenten versucht man nicht nur die Faltung von Proteinen – das heisst deren Wechselwirkung mit sich selbst – zu ergründen, sondern auch ihre Interaktionen

mit kleinen Molekülen, wie eben potenziellen Arzneimitteln. Drei Wege lassen sich nun denken, mit denen man dies untersuchen und eventuell sogar neue Arzneien entwickeln kann (siehe Abbildung 2).

Der erste lässt sich beschreiten, wenn schon mehrere Substanzen bekannt sind, die mit diesem Protein interagieren. Durch deren Untersuchung kann man herausarbeiten, welche ihrer Teile wahrscheinlich die stärksten Wechselwirkungen haben, und kann eine Art Landkarte der Interaktionen erstellen. Da diese bei jedem Molekül ein wenig anders aussieht, vermischt man alle Karten zu einer gemeinsamen und entwirft anschliessend ein Molekül, das dieser gerecht wird. Es wird also aus mehreren suboptimalen Molekülen eine Art Supermolekül geschaffen. Der Nachteil dieses Verfahrens: Es müssen schon relativ viele Informationen vorliegen; diese sind aber bei der Mehrzahl der menschlichen Krankheiten noch nicht vorhanden.

Auf dem zweiten Weg ist man schon etwas unabhängiger von tatsächlich ermittelten Daten. Hier versucht man eine ganze Bibliothek von Molekülen, deren Struktur als Dateien vorliegen, nacheinander in die 3D-Struktur des Zielproteins zu setzen (Docking) und die Bindungsenergie zu berechnen. Auf diese Weise sollen Moleküle herausgefiltert werden, die eine hohe Affinität zu dem Protein aufweisen. Dies ist der heute gebräuchlichste Weg, da er von herstellbaren Atomverbindungen ausgeht. Auch unsere Arbeitsgruppe versucht mit Docking von grossen Bibliotheken pharmakologisch aktive Substanzen zu entdecken.

Hat man keine geeignete Bibliothek zur Verfügung, so bleibt immer noch das «De novo»-Design. Dabei baut man aus kleinen Teilen ein grösseres Molekül, wobei bei jedem Verknüpfungsschritt der neu anzufügende Teil so ausgewählt wird, dass er in Bezug auf die Wechselwirkung mit dem Protein optimal ist. Als Resultat erhält man im Idealfall ein Molekül, das sich richtiggehend an das Protein anschmiegt. Auch diese Strategie wird in unserer Arbeitsgruppe verfolgt, und es wurde ein Programm entwickelt, mit dem man potenzielle Hemmstoffe entwerfen kann.

Ressourcenschonende Methoden

Die Vorteile all dieser Methoden liegen auf der Hand: da jeder Schritt *in silico* passiert, kann ein einzelner Wissenschaftler Tausen-

de von Molekülen untersuchen. Erst wenn ein Molekül vielversprechend erscheint, wird es auch tatsächlich zur Synthese vorgeschlagen. Dadurch benötigt man wesentlich geringere Mengen an Chemikalien, denn die ganze Bibliothek muss nicht stofflich vorhanden sein. Dies führt sowohl zu einer Kostenreduktion als auch zu einer signifikanten Erhöhung der Effizienz. Ob aus einem Molekül, das auf einem dieser

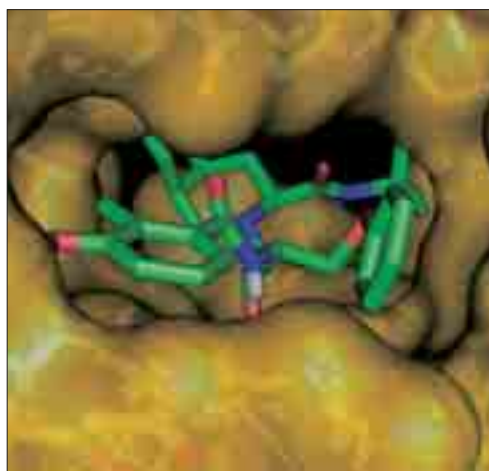


Abb. 2: Grafische Darstellung der Position des Arzneimittels Viracept®, eines Hemmstoffs für ein essentielles Enzym des AIDS verursachenden Virus (HIV). Die goldene Oberfläche stellt die Umrisse des Proteins dar.

Wege gefunden wurde, dann aber tatsächlich ein Medikament wird, hängt dennoch von weiteren In-vitro- und In-vivo-Tests (Zellkulturen, Tierversuche) und klinischen Studien ab.

Wie alle genannten Beispiele zeigen, haben Computer sich in der Biochemie und der biomedizinischen Forschung zu einem wichtigen Werkzeug entwickelt. Es braucht zwar auch heute noch das Experiment, um die errechneten Daten abzusichern. Jedoch können die Computermodelle das Verständnis für Zusammenhänge erhöhen. Von grosser Bedeutung ist ferner die Möglichkeit, grosse Datenmengen zu verarbeiten und Sinn darin zu suchen.

11-51 26219M

