

Physik der Proteine

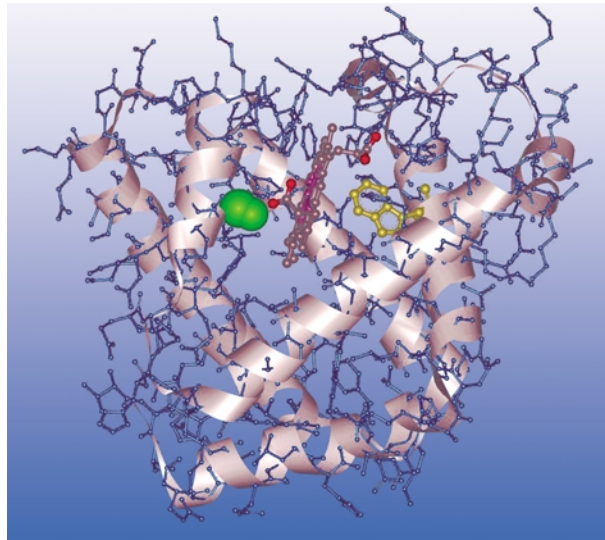
Für das physikalische Verständnis von Proteinfaltung und Proteinfunktion hat sich das Konzept der Energielandschaft als äußerst nützlich erwiesen

Gerd Ulrich Nienhaus

Proteine sind biologische Nanomaschinen, die für die komplexen Abläufe in der lebenden Zelle von essenzieller Bedeutung sind. Sie werden in der Zelle als lineare Polymere synthetisiert, die sich spontan in eine kompakte, räumliche Struktur falten. Im gefalteten Zustand bleiben jedoch flüssigkeitsähnliche Bewegungen mit großer Amplitude möglich, die für die Funktion der Proteine notwendig sind. Diese bislang nur unzureichend verstandenen Strukturfluktuationen lassen sich als Übergänge zwischen vielen verschiedenen Mikrozuständen beschreiben, die lokalen Minima in einer hochdimensionalen Energielandschaft entsprechen.

Proteine sind an allen Prozessen in der lebenden Zelle beteiligt. Regulatorische Proteine steuern die Genaktivität, Enzyme katalysieren biochemische Reaktionen, Antikörper verteidigen uns gegen eindringende Krankheitserreger, und Ionenpumpen kontrollieren die Transmembranpotentiale, auf denen unsere Denkprozesse physiologisch basieren. Das generelle Design der Proteine hat sich seit Milliarden von Jahren bewährt, und ihre detaillierte Struktur unterliegt in der Evolution einem kontinuierlichen Entwicklungsprozess. Ihre Baupläne sind in den Genen in Form einer linearen Polymerkette der Desoxyribonukleinsäure (DNA) in der Abfolge von vier verschiedenen Basen (Adenin, Cytosin, Guanin und Thymin) gespeichert. Seit der Entschlüsselung des menschlichen Genoms im Jahre 2003 weiß man, dass der Zellkern des Menschen etwa 30 000 Gene enthält. Bei der Proteinsynthese in der lebenden Zelle wird die Sequenz der vier Nukleinsäurebasen in eine Sequenz von zwanzig verschiedenen Aminosäuren umgesetzt, wobei jede Aminosäure durch jeweils drei aufeinander folgende Basen kodiert wird. Die so entstehende lineare Kette aus Aminosäuren selbst bildet jedoch noch kein funktionierendes Protein – erst nach der zumeist spontanen Faltung in seine spezifische, dreidimensionale (3D) Architektur liegt das Protein in seiner biologisch aktiven Form vor.

Die genetische Information zu kennen ist nur ein erster Schritt für ein detailliertes Verständnis der Lebensprozesse. Diese Prozesse basieren auf der spezifischen Wechselwirkung von Biomolekülen, insbesondere Proteinen, aus denen sich komplexe Signalkaskaden und Netzwerke aufbauen und stetig verändern. Daher gilt es in einem zweiten Schritt, diese elementaren



Das Protein Myoglobin dient zur Speicherung und zum Transport von Sauerstoff in Muskeln. Um die Funktionsweise von Proteinen genau verstehen zu können, muss man nicht nur die Struktur, sondern auch die Struktur-dynamik kennen.

Wechselwirkungen zu verstehen. Umfassende, quantitative Daten sind bislang nur für wenige Proteine verfügbar, aber notwendig, um die beiden fundamentalen Fragen nach (i) dem Mechanismus der Proteinfaltung sowie (ii) dem Zusammenhang zwischen Struktur und Funktion zu beantworten. Beide ergeben sich aus der enormen Komplexität dieser Vielteilchensysteme [1].

In welche räumliche Struktur faltet sich eine vorgegebene Sequenz von Aminosäuren? – Diese Frage zu beantworten ist eine große intellektuelle Herausforderung und zudem von großem praktischen Nutzen: Die automatisierte Analyse ganzer Genome produziert riesige Datenmengen von Proteinsequenzen. Die 3D-Molekülstrukturen müssen hingegen in vergleichsweise langwierigen Messungen, insbesondere durch Röntgenbeugung an Einkristallen und Kernresonanzspektroskopie an Lösungen, bestimmt werden, die mit dem Tempo der Gen-Sequenzierung nicht Schritt halten können. Eng verbunden mit dem Proteinfaltungsproblem ist das inverse Faltungsproblem. Wie muss eine Aminosäuresequenz gestaltet werden, damit sie sich präzise in eine gewünschte Architektur faltet? – Diese Frage ist für die Biotechnologie relevant, um Proteine für technische Anwendungen *de novo* synthetisieren zu können. Viele bekannte Krankheiten (u. a. BSE, Creutzfeld-Jakob, Alzheimer) sind das direkte Resultat einer fehlerhaften Faltung von Proteinen, und ein besseres Verständnis der Grundlagen wäre hilfreich für Prävention und Therapie.

Die zweite zentrale Frage der Proteinforschung, die nach der Beziehung zwischen Funktion und Struktur, zeigt sich bei Betrachtung der 3D-Struktur des gefalteten Proteins. Die biologische Funktion ist daraus im Allgemeinen nicht ersichtlich, sondern muss unabhän-

Prof. Dr. Gerd Ulrich Nienhaus, Abteilung Biophysik, Universität Ulm, 89069 Ulm

gig bestimmt werden. Für die Funktion ist die Dynamik ein ganz wesentlicher Aspekt. Proteine fluktuieren im thermischen Gleichgewicht zwischen vielen unterschiedlichen Strukturen und nehmen bei Funktionsprozessen angeregte Zustände ein. Im strikten Sinne gibt es also überhaupt nicht „die“ Struktur eines Proteins; vielmehr repräsentiert ein Proteinmodell lediglich die mittleren Positionen der Atome. Gewöhnlich handelt es sich dabei um eine Grundzustandsstruktur; die Struk-

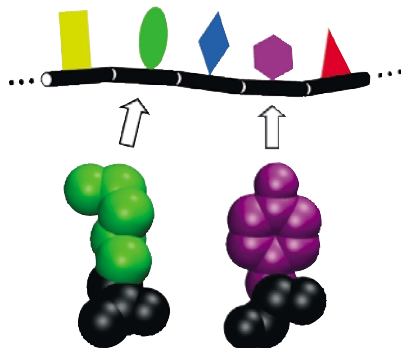


Abb. 1:
Die Primärstruktur eines Proteins besteht aus einer Kette von 20 verschiedenen Aminosäuren, die sich in den Seitenketten unterscheiden. Zwei Aminosäuren sind in Form von Kalottenmodellen dargestellt.

tur von Nichtgleichgewichtszuständen ist bislang nur in ganz wenigen Fällen bekannt. Zum Verständnis der Funktion muss jedoch die zeitliche Entwicklung der Struktur, einschließlich transient eingenommener Zwischenzustände, aufgeklärt werden [2].

Vom Genom zum Proteom

Durch den massiven Einsatz von Robotern und Computern ist die Sequenzierung des gesamten menschlichen Genoms in nur wenigen Jahren gelungen. In den Lebenswissenschaften tragen sich viele Wissenschaftler mit der Hoffnung, dass der vollautomatisierte Ansatz der Genomik insbesondere auf die Erforschung der Proteine übertragbar ist. Im Rahmen dieses Konzepts, der Proteomik, soll die ganze Proteinausstattung einer Zelle, das Proteom, parallel erfasst werden, wobei Trennverfahren wie zweidimensionale Gel-Elektrophorese, chromatographische Methoden, Massenspektroskopie und Biochips eine wichtige Rolle spielen. Proteinstrukturen sollen dann in „Strukturfabriken“ mit Hochdurchsatzverfahren bestimmt und in Datenbanken gespeichert werden. Geeignete Werkzeuge der Bioinformatik („data mining“) sollen es schließlich erlauben, daraus die komplexen Prozesse in der lebenden Zelle aufzuklären. Zum Beispiel hofft man, durch Mustererkennung aus vielen bekannten Proteinstrukturen die Struktur einer unbekannt Sequenz bestimmen zu können.

„Es ist schön, dass der Computer das Problem versteht. Aber ich möchte es auch verstehen.“ Diese Bemerkung soll Eugen Wigner anlässlich einer langwierigen, quantenmechanischen Berechnung gemacht haben. Sie ist repräsentativ für die Denkweise der Physik, die stets nach einfachen, generalisierenden Beschreibungen von Naturphänomenen sucht. Anstatt eine „black box“ zu bauen, in die man eine Sequenz eines Proteins hineinfüttert und aus der am Ende eine gefaltete Struktur herauskommt, bevorzugen es Physiker, Theorien und Modelle zu entwickeln, die die komplexen Wechselwirkungen in Proteinen generell beschreiben. Diese Kenntnisse sollten es ermöglichen, Vorhersagen zur Faltung und Funktion unbekannter Proteine zu machen. Einige der großen Fortschritte, die in den vergangenen Jahren durch ein lebhaftes Wechselspiel von Theorie, Computersimulation und Experiment erzielt wurden, werden in diesem Beitrag skizziert.

Die Struktur der Proteine

Proteine sind Kettenpolymere aus 20 verschiedenen Aminosäuren. Wie Abb. 1 zeigt, haben diese eine identische Rückgratstruktur, unterscheiden sich jedoch in den Seitenketten. Die Sequenz der linearen Peptidkette wird Primärstruktur genannt. Zwei frei drehbare Winkel pro Monomer ermöglichen die Faltung der Kette. Abb. 2 zeigt Beispiele kleiner globulärer Proteine mit charakteristischen periodischen Strukturen, schraubenförmigen α -Helices und flächigen Anordnungen ähnlich einem Wellblech, den β -Faltblättern. Diese so genannten Sekundärstrukturelemente falten sich in die dreidimensionale, extrem kompakte Tertiärstruktur des Proteins. Viele Proteine weisen trotz völlig verschiedener Sequenzen das gleiche Faltungsmuster auf. Derzeit kennt man 560 verschiedene Faltungsmuster, und es wird geschätzt, dass es höchstens einige tausend davon gibt [3]. Trotz der hohen Packungsdichte der Atome, vergleichbar der in einem Molekülkristall, sind Proteine dennoch flexible Gebilde. In einer Vielzahl verschiedener Konformationen können die Atome um 1–2 Å um ihre mittlere Position herum fluktuieren.

Synthetische Polymere hingegen haben nicht die Tendenz, eine kompakte räumliche Struktur einzunehmen. Sie liegen in vielen verschiedenen offenen Strukturen vor, die sich auch miteinander verheddern. Ähnliches Verhalten findet man bei Proteinen unter unphysiologischen Bedingungen. Diese so genannte Denaturierung tritt auf bei zu hoher oder zu tiefer Temperatur, zu hohem oder zu tiefem pH-Wert, bei hohem Druck oder in Gegenwart von Kosolventien (z. B. Harnstoff, Säure).

Thermodynamische Stabilität

Proteine können in wässriger Lösung sehr leicht denaturiert werden; ihre funktionale, native Form ist also nur marginal stabil. Die Stabilität hängt offenbar stark von den Eigenschaften der Umgebung ab, im Falle globulärer Proteine vom wässrigen Medium. In der Tat kann man die Hydrathülle als Teil der Struktur auffassen, sie ist für die physiologischen Eigenschaften essenziell. Seit den 60er-Jahren ist bekannt, dass Proteine nach der Denaturierung wieder in ihre gefaltete Form überführt werden können, zum Beispiel, indem man eine Lösung des Proteins, die Harnstoff enthält, mit Wasser verdünnt. Unter physiologischen Bedingungen ist die native Struktur also thermodynamisch stabil (oder zumindest metastabil auf sehr langen Zeitskalen). Unter den üblichen Bedingungen von konstantem Druck und Temperatur ist dieser Zustand durch ein Minimum der Gibbsschen freien Enthalpie charakterisiert. Die 3D-Struktur ist also bereits durch die Primärstruktur vollständig bestimmt. Die DNA kodiert damit nicht nur die Sequenz der Aminosäuren, sondern enthält auch die Information über die räumliche Struktur, was zuweilen als die „zweite Hälfte des genetischen Codes“ bezeichnet wird. Für kleine, globuläre Proteine lässt sich der Übergang zwischen dem nativen und dem denaturierten Zustand vielfach durch ein thermodynamisches Modell mit zwei Makrozuständen beschreiben; er ähnelt einem Phasenübergang erster Ordnung. Zuweilen treten weitere Makrozustände auf, die man als Faltungsintermediate bezeichnet.

Welcher Natur sind die Wechselwirkungen, die die native Form stabilisieren? Die Gibbssche freie Enthalpie, $G = H - TS$, setzt sich aus Beiträgen der Enthalpie H und der Entropie S zusammen. Letztere sind mit der absoluten Temperatur T gewichtet. Die Beiträge zur

Enthalpie rühren von den Bindungen zwischen den Atomen her. Für die Polymerkettendynamik spielen die kovalenten Bindungen keine Rolle, da sie bei physiologischen Temperaturen stabil sind und nicht aufgebrochen werden. Wichtig sind hingegen die Bindungen, deren Energien nur wenig größer sind als die thermische Energie kT und die damit thermischen Fluktuationen unterliegen. Hierunter fallen die dispersiven (van-der-Waals) Bindungen, die auf Ladungsfluktuationen der Elektronenhüllen zurückgeführt werden, die elektrostatischen Wechselwirkungen zwischen polaren Seitenketten sowie die Wasserstoffbrücken innerhalb der Peptidkette einerseits und zwischen der Peptidkette und dem Lösungsmittel andererseits.

Enthalpie und Energielandschaft

Physikern weniger geläufig, jedoch für die Strukturbildung in biologischen Systemen von herausragender Bedeutung, ist die hydrophobe Wechselwirkung. Diese zeigt sich am einfachsten in dem Phänomen, dass Öl und Wasser nicht mischbar sind. Öl enthält hydrophobe Moleküle, die keine Wasserstoffbrückenbindungen mit Wasser eingehen können. Um hydrophobe Moleküle herum bilden sich Käfige aus Wassermolekülen. Struktur und Dynamik der an den Käfigen beteiligten Wassermoleküle unterscheiden sich deutlich von denen in der normalen Flüssigkeit. Die spontane Phasentrennung von Öl und Wasser zeigt, dass das System Kontaktflächen zwischen Öl und Wasser vermeiden möchte. Die Ausbildung von Käfigstrukturen führt also zu einer höheren freien Enthalpie des Systems. Bei Raumtemperatur ist die hydrophobe Wechselwirkung im Wesentlichen entropischen Ursprungs; der Ordnungszustand der Wassermoleküle im Käfig ist höher als in der normalen Flüssigkeit. Bei der Faltung von globulären Proteinen führt die hydrophobe Wechselwirkung nun dazu, dass Aminosäuren mit apolaren, hydrophoben Seitenketten im Inneren des Proteins einen hydrophoben Kern ausbilden, der für Wasser nicht zugänglich ist, während sich die hydrophilen Seitenketten bevorzugt an der Proteinoberfläche in Kontakt mit dem wässrigen Medium befinden. Ein einfaches geometrisches Modell, das die Polypeptidkette als kontinuierliche Röhre beschreibt, zeigt, dass es nur relativ wenige Topologien für den Verlauf des Rückgrats gibt, die einen kompakten hydrophoben Kern ermöglichen [4]. Da die gefaltete Polymerkette viel weniger Mikrozustände (Konformationen) einnehmen kann als die gestreckte, wirkt sich der Beitrag der entsprechenden Konformationsentropie stark destabilisierend auf die freie Enthalpie aus.

Bei der quantitativen Diskussion der stabilisierenden Beiträge betrachtet man eine effektive Energiefunktion.¹⁾ Diese enthält sowohl die intramolekulare Enthalpie der Kette als auch die freie Enthalpie der Wechselwirkung mit dem Lösungsmittel, also alle Beiträge zur freien Enthalpie mit Ausnahme der Konformationsentropie der Polypeptidkette. Sie ist eine hochdimensionale Funktion der Konformationsfreiheitsgrade der Kette (selbst bei kleinen Proteinen typisch mehrere tausend) und wird als Energielandschaft („energy landscape“) bezeichnet.

Unter physiologischen Bedingungen ist die freie Enthalpie des nativen Zustandes nur minimal niedriger als die des denaturierten Zustandes, was auf die fast vollständige Aufhebung großer enthalpischer und entropischer Beiträge zurückgeführt wird. Für das Protein Myoglobin kann man einen Beitrag der Konformations-

entropie von 2,60 MJ/mol zur freien Enthalpie bei 25 °C abschätzen. Dieser Beitrag wird durch die anderen Beiträge zur freien Enthalpie nur geringfügig überkompensiert, sodass der gesamte Verband von etwa 2500 Atomen mit nur 40 kJ/mol stabilisiert ist. Die Erzeugung der geordneten, nativen Struktur (Entropieabsenkung!) braucht also fast die gesamte übrige freie Enthalpie auf. Der Wert von einigen zehn kJ/mol ist typisch für kleine globuläre Proteine; er entspricht der Bindungsenergie von wenigen Wasserstoffbrücken. Diese marginale Stabilität ist essenziell für die Proteinfunktion; sie ermöglicht die Flexibilität der Struktur. Die fast vollständige Kompensation großer Energiebeiträge macht die quantitative Beschreibung der Energetik allerdings sehr schwierig.

1) Genauer: freie Enthalpie, „potential of mean force“.

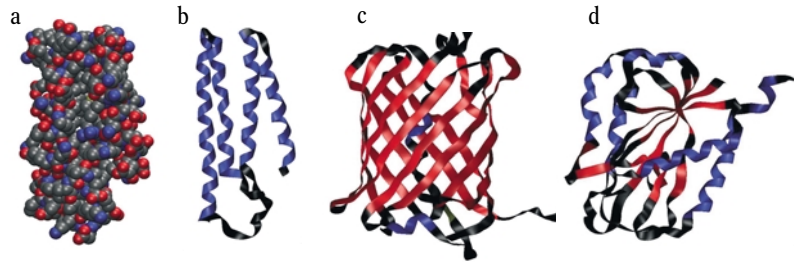


Abb. 2: Modelle von drei kleinen globulären Proteinen. Das raumfüllende Modell (a) zeigt die kompakte Struktur, die Cartoon-Repräsentation des Peptidrückgrates (b) das Faltungsmuster von Cytochrom b_{562} mit vier α -Helices (blau), die durch kurze

Schleifen (schwarz) verbunden sind. Das rot fluoreszierende Protein eqFP 611 (c) besteht fast ausschließlich aus β -Faltblattstruktur, Flavodoxin (d) weist α -helikale und Faltblattbereiche auf.

Die Dynamik der Proteinfaltung

Im Prinzip könnte man die gefaltete Struktur bestimmen, indem man für jede mögliche Konformation der Polypeptidkette alle Wechselwirkungen innerhalb des Proteins sowie zwischen Protein und Lösungsmittel berechnet. Jedoch steigt die Zahl der Konformationen exponentiell mit der Anzahl der Monomere an. Nimmt man fünf mögliche Konformationen für jede Aminosäure an, kommt man bei einem kleinen Protein mit 100 Aminosäuren bereits auf $5^{100} \approx 10^{70}$ mögliche Konformationen. Der Prozess der Faltung lässt sich als Trajektorie eines Punktes im Konformationsraum beschreiben. Wenn man nun die Zeit für eine elementare Konformationsänderung mit ca. 10 ps abschätzt, würde es 10^{59} s dauern, bis die Peptidkette in einer statistischen Suche („random walk“) die native Konformation gefunden hat. Wie kann es nach dieser Abschätzung beim gegenwärtigen Alter des Weltalls von 5×10^{17} s überhaupt gefaltete Proteine geben? Dies ist das bekannte Levinthal-Paradox [5]; es zeigt deutlich, dass die Faltung der Polypeptidkette nicht das Ergebnis einer zufälligen Suche im Konformationsraum sein kann.

Wie aber schafft es ein kleines globuläres Protein, die native Konformation in weniger als einer Sekunde zu finden? Levinthal schlug vor, dass es spezifische Faltungspfade im Konformationsraum gibt, sodass die zufällige Suche vermieden würde. Inzwischen ist jedoch klar geworden, dass die Proteinfaltung sehr wohl eine stochastische Suche zwischen vielen Konformationen darstellt – allerdings ist diese nicht zufällig, sondern durch die Energie stark gerichtet. In den theoretischen Modellen spielen stabilisierende Wechselwirkungen der gefalteten Polymerkette, die so genannten nativen

Kontakte, eine besondere Rolle. Diese entstehen, wenn Fluktuationen der Peptidkette beim Faltungsprozess auch in der Sequenz weit auseinander liegende Teile der Peptidkette zufällig in Kontakt bringen. Ein nativer Kontakt stabilisiert die Kette und schränkt den Konformationsraum ein, sodass die zufällige Suche nach dem nächsten nativen Kontakt in dem verbleibenden Unter- raum schneller geht. Durch die sukzessive Ausbildung nativer Kontakte wird die Peptidkette effizient in die

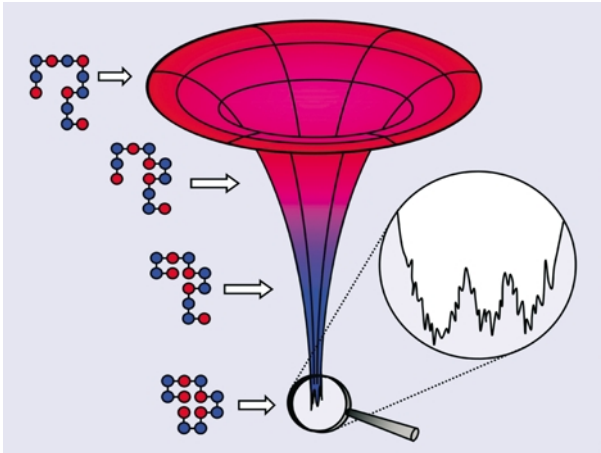


Abb. 3: Globale, trichterförmige Energielandschaft eines Proteins. Die Vertikale repräsentiert die effektive Energie der Kette, die schematisch gegen zwei der vielen Konformationsfreiheitsgrade aufgetragen ist. Auf der linken Seite sind typische Modellsequenzen auf einem zweidimensionalen Gitter dargestellt. Die Vergrößerung zeigt die raue Energielandschaft des nativen Proteins.

native Konformation geführt. Der Kollaps der Polymerkette in die kompakte Form lässt sich mit einer Energielandschaft veranschaulichen, die global die Form eines Trichters („folding funnel“) hat [6, 7] (Abb. 3). Der weite Rand des Trichters beschreibt die vielen möglichen offenen Formen und damit die hohe Konformationsentropie der entfalteten Peptidkette. Das Protein faltet sich entlang einer geneigten Trajektorie in Richtung auf das Minimum, das die gefaltete Spezies darstellt.

Frustrierte Sequenzen

Wie muss eine Aminosäuresequenz beschaffen sein, damit sich die globale Form eines Faltungstrichters ergibt? Angesichts der relativ kleinen Zahl bislang beobachteter Faltungsmuster und der riesigen Zahl unterschiedlicher Proteinsequenzen sollten Details der Sequenz nicht relevant sein (für die Funktion kann allerdings der Austausch einer einzigen Aminosäure fatal sein!). Wie verhalten sich Zufallssequenzen im Vergleich zu Proteinsequenzen? Diese Fragen veran-

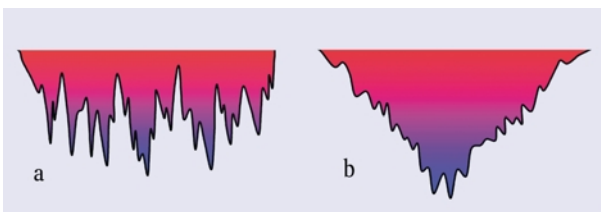


Abb. 4: Die Energielandschaften einer zufälligen und damit stark frustrierten Peptidsequenz (a) ist rau mit vielen lokalen Minima, die sehr unterschiedlichen Strukturen entsprechen. Bei natürlichen Proteinen hat die Energielandschaft hingegen die Form eines Trichters (b).

lassten Bryngelson und Wolynes, das Problem der Proteinfaltung mit Ansätzen aus der statistischen Physik ungeordneter Systeme zu betrachten [7, 8].

Ausgangspunkt war das „Random Energy Model“ (REM), das aus der Theorie der Spingläser bekannt ist. Spingläser sind ungeordnete, verdünnte Anordnungen magnetischer Atome in diamagnetischen Matrizen. Abhängig von der lokalen Geometrie koppeln benachbarte Atome entweder ferro- oder antiferromagnetisch. Beim Abkühlen einer Spinglasprobe richten sich die Spins so aus, dass insgesamt eine Anordnung mit minimaler Energie entsteht. Dabei gibt es Konflikte, wie aus der Betrachtung dreier antiferromagnetisch gekoppelter Spins ersichtlich ist: zwei können sich entgegengesetzt und damit energetisch günstig einstellen, während der dritte nur zu jeweils einem der beiden anderen energetisch günstig stehen kann. Da es nicht möglich ist, alle lokalen Wechselwirkungen simultan zu optimieren, spricht man von „Frustration“. Für das gesamte Spinsystem führt Frustration zu einer Quasi-Entartung des Grundzustandes; es gibt viele Konfigurationen mit ähnlicher Energie. Der durch Abkühlung eingefrorene Mikrozustand hängt von der Vorgeschichte ab; jeder Abkühlzyklus führt zu einer anderen Spinkonfiguration. Außerdem existiert ein Glasübergang: Da die Anzahl thermisch zugänglicher Konfigurationen mit der Temperatur dramatisch absinkt, friert das Spinsystem unterhalb der Glas temperatur in einer zufälligen Konfiguration ein.

Die Tendenz der Spins im Spinglas, sich in Bezug auf die Nachbarspins optimal ausrichten zu wollen, entspricht dem Bestreben der Aminosäuren im Protein, sich im kompakten, gefalteten Zustand energetisch optimal zu ihren Nachbarn anzuordnen. Zufällige Aminosäuresequenzen haben einen hohen Grad an Frustration. Sie haben eine raue Energielandschaft mit vielen lokalen Minima mit höchst unterschiedlichen Strukturen (Abb. 4a) anstelle einer wohldefinierten, nativen Konformation. Die Sequenzen in der Natur vorkommender Proteine sehen zwar zufällig aus, ihre Energielandschaft hat jedoch global die Form eines Trichters. Dabei sollte die Rauigkeit möglichst gering sein, damit die vielen Faltungstrajektorien nicht durch Energiebarrieren behindert werden (Abb. 4b). Man spricht vom „Prinzip der minimalen Frustration“. Die dennoch vorhandene Rauigkeit erklärt die Zwischenzustände oder Faltungsintermediate, die bei der Untersuchung von Proteinfaltungsreaktionen vielfach gefunden werden. Im Innern des Trichters gibt es eine große Anzahl von Mikrozuständen, die der nativen Form zugeordnet sind (Abb. 3). Diese sind durch eine Energielücke von den denaturierten Mikrozuständen separiert.

Modellrechnungen zur Proteinfaltung

Selbst bei Verwendung von Supercomputern sind derzeit Faltungssimulationen bei vollständiger mikroskopischer Beschreibung aller Wechselwirkungen der Peptidkette sowie des umgebenden Lösungsmittels noch kaum möglich.²⁾ Zur Vereinfachung der Rechnungen bedient man sich oft kurzer Modellsequenzen, die aus zwei verschiedenen Monomeren (hydrophob/hydrophil) bestehen. Oft werden nur Konformationen auf 2D- oder 3D-Gittern betrachtet. Abb. 3 zeigt einige Beispiele. Obwohl diese Modelle extreme Karikaturen realer Proteine sind, erweisen sie sich als äußerst nützlich für das Verständnis der physikalischen Prinzipien. So kann man die verschiedenen Konformationen

2) Pionierarbeit wird derzeit mit dezentralen Rechnern im Internet („folding@home“) geleistet. Ab 2006 soll der von IBM entwickelte und 100 Mio. US\$ teure Supercomputer Blue Gene mit massiv paralleler Architektur und einer Rechenleistung von 1 PetaFlop/s speziell für Simulationen der Proteindynamik eingesetzt werden.

vollständig aufzählen und ihre Energie und damit die Wahrscheinlichkeit ihrer Realisierung über den Boltzmann-Faktor berechnen. Durch entsprechendes Design der Sequenzen kann man analysieren, wie sich minimal bzw. stark frustrierte Sequenzen verhalten. Bei minimal frustrierten Sequenzen zeigt sich, dass angeregte Zustände dem Grundzustand ähneln und durch wenige Rekonfigurationen aus diesem hervorgehen. Bei Zufallssequenzen hingegen unterscheiden sich die tiefliegenden angeregten Konformationen strukturell vollständig vom Grundzustand und bilden „kinetische Fallen“: Um den Grundzustand aus einem dieser Zustände zu erreichen, muss das Protein erst wieder fast vollständig auseinandergefaltet werden.

Bei der Analyse der Dynamik des Faltungsprozesses an Modellsequenzen findet man nicht eine wahllose, langwierige Suche à la Levinthal, sondern viele mögliche Trajektorien, die von der riesigen Zahl ungefalteter Zustände zu den wenigen nativen Zuständen führen. Die Faltungsdynamik in dieser rauen Energielandschaft entspricht einer gerichteten Diffusion eines Teilchens unter dem Einfluss des Potentialgradienten des Faltungstrichters. Im oberen Teil ist der Trichter weit und es gibt eine Vielzahl von alternativen Wegen. Selbst wenn ein bestimmter Weg durch eine hohe Energiebarriere blockiert ist, gibt es mit hoher Wahrscheinlichkeit alternative Routen, auf denen die Barriere umgangen werden kann, sodass der anfängliche Kollaps schnell erfolgt. Erst in einer späteren Phase wird der Prozess verlangsamt, da der Trichter immer enger wird und nur noch wenige Pfade zur nativen Struktur führen. Die Polymerketten müssen die wenigen Konformationen entlang dieser Pfade finden, um in der Faltung voranzuschreiten. Solche entropischen Engpässe können lange charakteristische Zeiten der Faltung hervorrufen. Zusätzliche Rauigkeit der Energiefläche kann den Prozess gerade in dieser Phase behindern, sodass es zur Ausbildung von transienten Zwischenzuständen kommt.

Proteinfaltungsexperimente

In klassischen Faltungsexperimenten wird der kleine Unterschied in der freien Enthalpie zwischen nativer und denaturierter Form, der unter physiologischen Bedingungen existiert, auf chemischem (Zusatz von denaturierenden Stoffen) oder physikalischem (Temperatur, Druck) Wege variiert und damit das Gleichgewicht zwischen beiden makroskopischen Zuständen gezielt beeinflusst. Eine Reihe von physikalischen Messmethoden (u. a. optische Absorptions- und Fluoreszenzspektroskopie, Infrarotabsorption, Zirkulardichroismus, NMR, Kleinwinkelstreuung) erlaubt es, den Anteil an denaturiertem Protein im Ensemble zu bestimmen. Um die Zeitabhängigkeit (Kinetik) der Faltung an einem Ensemble von Proteinen zu untersuchen, müssen alle Proteine der Probe gleichzeitig dazu gebracht werden, sich zu falten. Diese Synchronisation wird durch eine externe Störung erzwungen und lässt sich beispielsweise durch schnelle Verdünnung einer Lösung, die das Protein in Anwesenheit eines denaturierenden Stoffes enthält, mit Wasser erzielen. Mikrostrukturierte Flüssigkeitsmischer und Experimente mit kurzen Laserpulsen ermöglichen sogar zeitaufgelöste Messungen im Nano- bis Millisekundenbereich [9].

Die Vielzahl der möglichen mikroskopischen Faltungstrajektorien impliziert, dass sich Details der Energielandschaft nicht durch Messungen an ganzen Ensembles, sondern nur mit Einzelmolekülmethoden

gewinnen lassen. Eindrucksvolle Experimente zur Faltung und Entfaltung wurden mit Hilfe von Rasterkraftmikroskopen durchgeführt [10]. Optische Methoden wie die konfokale Fluoreszenzmikroskopie ermöglichen es, die Struktur und Faltungsdynamik der Peptidkette direkt zu beobachten (Abb. 5).

Energielandschaften gefalteter Proteine

Der native Zustand bildet ein tiefes Minimum in der Energielandschaft, das durch eine Energielücke von der Vielzahl der nicht-nativen Zustände abgetrennt ist. Innerhalb dieses Zustandes kann das gefaltete Protein eine große Anzahl von Mikrozuständen einnehmen, die auch als Konformationssubzustände bezeichnet werden. Sie haben leicht unterschiedliche Strukturen, können aber dennoch große Unterschiede in ihren Funktionseigenschaften aufweisen. Sie werden durch lokale Minima in einer rauen Energiefläche des gefalteten Proteins repräsentiert (Abb. 3). Konformationsübergänge finden in der Regel durch thermische Aktivierung statt, sodass man aus der Temperaturabhängigkeit dynamischer Prozesse die Höhe der Energiebarrieren zwischen den Subzuständen bestimmen kann. Tunnelprozesse sind prinzipiell ebenfalls möglich. Eine Vielzahl von Experimenten hat gezeigt, dass es ein breites Spektrum von Energiebarrieren mit sehr unterschiedlichen Höhen gibt. Selbst nahe am absoluten Temperaturnullpunkt gibt es Konformationsübergänge, wie sich aus dem linearen Anstieg der spezifischen Wärmekapazität, mit

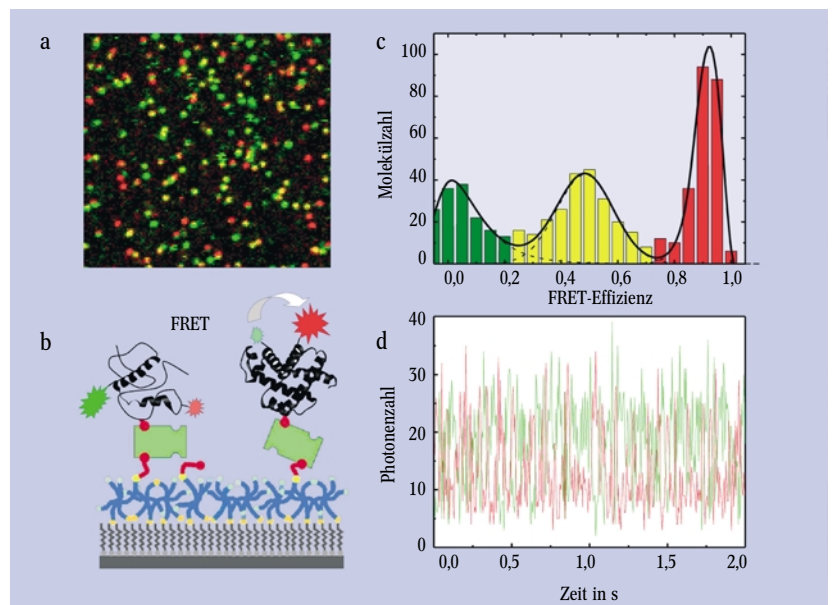


Abb. 5: Faltungsexperimente an einzelnen Proteinen lassen sich mit Hilfe konfokaler Fluoreszenzmikroskopie durchführen. ▶ a) Im Rasterbild stellt jeder Punkt ein einzelnes Enzym (RNase H) dar, das gezielt mit je einem grünen und einem roten Fluoreszenzfarbstoff gentechnisch markiert und zur stationären Beobachtung an einer Glasoberfläche fixiert wurde. ▶ b) Beim gefalteten Proteinmolekül (rechts) sind beide Farbstoffe nah beieinander, sodass der grüne Donor-Farbstoff nach Absorption eines Photons seine Anregungsenergie durch Fluoreszenzresonanz-Energietransfer (FRET) effizient auf den roten Akzeptor-Farbstoff übertragen kann, der anschließend ein rotes Photon

emittiert. In der entfalteten Form (links) sind die Farbstoffe weiter auseinander. Strukturfluktuationen beeinflussen den Abstand der Farbstoffe und damit die Wahrscheinlichkeit für FRET.

▶ c) Aus der simultanen Messung der Emission von Donor und Akzeptor eines jeden einzelnen Proteins erhält man Histogramme der Molekülzahlen als Funktion der FRET-Effizienz E , in denen die Populationen der gefalteten ($E \approx 1$, rot), der entfalteten Proteine ($E \approx 0,5$, gelb) sowie der Proteine mit fehlendem Akzeptor-Farbstoff ($E \approx 0$, grün) hervorgehen. ▶ d) Die Strukturfluktuationen lassen sich aus der Fluoreszenzemission einzelner Biomoleküle zeitaufgelöst verfolgen.

optischer Lochbrennspektroskopie bzw. Photonenechoexperimenten und neuerdings auch mit Einzelmolekülspektroskopie [11] zeigen ließ. Großräumige Fluktuationen der Struktur treten jedoch erst oberhalb einer Temperatur von etwa 180 K auf. Dieses Verhalten ist besonders deutlich aus dem signifikanten Anstieg der mittleren quadratischen Fluktuationen der Atome, $\langle x^2 \rangle$, zu erkennen, die mit der Mössbauer- und Neutronenspektroskopie beobachtet werden [2, 12]. Unterhalb der so genannten dynamischen Übergangstemperatur T_d steigt $\langle x^2 \rangle$ proportional zur Temperatur an; das Protein verhält sich wie ein harmonischer Festkörper. Oberhalb von T_d wachsen die $\langle x^2 \rangle$ -Werte weitaus stärker mit der Temperatur an. Daraus kann man schließen, dass oberhalb von T_d thermisch aktivierte, flüssigkeitsähnliche Bewegungen innerhalb der nativen Konformation stattfinden. Dieses Verhalten ähnelt der Dynamik viskoser Flüssigkeiten oberhalb des Glasübergangs. Das breite Spektrum von Energiebarrieren zwischen den Konformationszuständen führt dazu, dass Proteine unter physiologischen Bedingungen dynamische Prozesse mit charakteristischen Zeiten über viele Größenordnungen in der Zeit aufweisen. Am schnellsten sind die harmonischen Schwingungen innerhalb der Minima, die für

3) Übrigens wird die rote Farbe eines Steaks durch die starke Absorption der Hämgruppe des Myoglobins im blauen Spektralbereich des Lichts hervorgerufen.

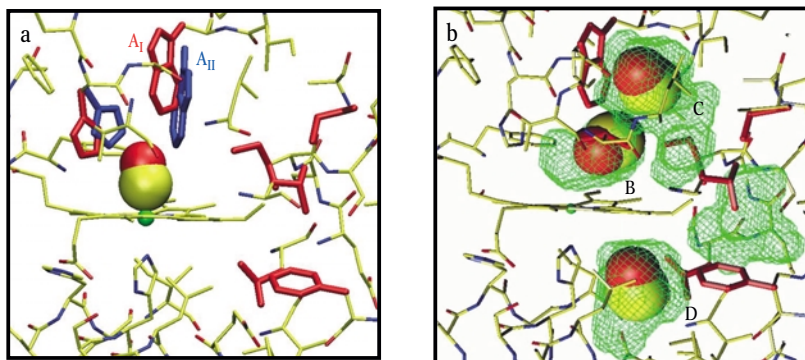


Abb. 6:
 ► a) Das aktive Zentrum von CO-ligandiertem Myoglobin (MbCO) L29W mit dem am Hämeisen gebundenen CO-Molekül: In den Konformationen A_I und A_{II} haben zwei rot und blau markierte Aminosäuren deutlich unterschiedliche Positionen.
 ► b) Nach Photodissoziation wandert das CO-Molekül durch Kavitäten im Innern des Proteins (in grün) zu verschiedenen Andockstellen (B, C, D) und kann auch aus dem Protein entweichen.

Wasserstoffatome eine Periode von ca. 10 fs aufweisen; kollektive Schwingungen ganzer Gruppen finden auf der Pikosekundska statt. Kooperative Bewegungen kleinerer Bereiche werden auf der Nano- bis Mikrosekundska beobachtet. Die für die Proteinfunktion wichtigen Umlagerungen ganzer Domänen finden sich auf der Zeitskala von Mikrosekunden bis Sekunden (und darüber hinaus) [2].

Die Vielzahl von Konformationen und die breite Verteilung der Aktivierungsbarrieren werfen Fragen nach einer Struktur oder generellen Topologie des Konformationsraums auf. Frauenfelder postulierte auf der Basis experimenteller Daten an Myoglobin eine hierarchische Anordnung der Subzustände auf mehreren Ebenen [2, 13]. Auf der obersten Ebene findet man eine kleine Anzahl diskreter, strukturell wohlunterscheidbarer Konformationszustände. Innerhalb dieser „taxonomischen“ Subzustände gibt es weitere, nur statistisch erfassbare Subzustände, die durch kleinere Energiebarrieren separiert sind. Diese ineinander geschachtelte Anordnung mit sukzessive kleiner werdenden Barrieren setzt sich über mehrere Ebenen fort. Hinweise auf ei-

nen hierarchischen Aufbau des Konformationsraums wurden inzwischen für viele Proteine experimentell [11, 14, 15] sowie durch Analysen der Trajektorien von (klassischen) Molekulardynamiksimulationen gefunden.

Strukturdynamik im gefalteten Protein

Großräumige Strukturänderungen sind auf molekularer Ebene nur für wenige Proteine im Detail untersucht. Myoglobin, ein kleines globuläres Protein im Muskelgewebe, hat sich seit vielen Jahren bei der Untersuchung der Energielandschaft als Modellsystem bewährt [2, 13]. Seine Funktion beruht auf der Fähigkeit, ein kleines Gasmolekül (O_2 , CO , NO) am Eisenatom einer Hämgruppe binden zu können.³⁾ Lichtabsorption führt zur Photodissoziation des Liganden vom Hämeisen, worauf sich der Ligand fortbewegt. Anschließend kann derselbe oder auch ein anderer in der Lösung vorhandener Ligand wieder am Hämeisen andocken. Im Labor lässt sich dieser Prozess durch Laserpulse beliebig oft starten, sodass Myoglobin ein ausgezeichnetes Modellsystem für die Untersuchung der Proteindynamik bei der Ligandenbindung ist. Die Strukturänderungen können mit Hilfe verschiedenster spektroskopischer Methoden charakterisiert werden. Erst in den letzten Jahren ist es gelungen, diese Beobachtungen direkt mit strukturellen Änderungen zu korrelieren.

Abb. 6a zeigt Teile der Struktur des aktiven Zentrums von CO-ligandiertem Myoglobin (MbCO), Mutante L29W [16]. Ein CO-Molekül ist am Eisen in der Mitte der planaren Hämgruppe gebunden. Das Protein hat zwei mögliche Konformationen, A_I und A_{II} , mit signifikant unterschiedlicher Struktur eines Teils des Proteins. Dies ist an den alternativen Positionen zweier Aminosäureseitenketten (blau und rot) zu sehen.

In kryokristallographischen Untersuchungen von Proteinkristallen nach Photolyse des Liganden wurde gezeigt, dass das CO-Molekül durch Hohlräume im Protein wandern und an verschiedenen Stellen (B, C, D in Abb. 6b) andocken kann. Außerdem tritt bei Photodissoziation oberhalb der dynamischen Übergangstemperatur eine globale Strukturänderung von A_I nach A_{II} auf. Die transienten Andockstellen im Innern des Proteins sind wichtig für die Funktion. Es gibt keinen permanenten Kanal, auf dem der Ligand das Protein verlassen kann. Pfade bilden sich nur transient durch großräumige Strukturfluktuationen, sodass das CO zum Verlassen des Proteins bis zu $1 \mu s$ braucht. Aufgrund der Andockstellen ist die direkte Rückbindung des Liganden während dieser Zeit stark unterdrückt.

Kürzlich wurde dieser Prozess der Ligandenwanderung bei Raumtemperatur mit 150 ps Zeitauflösung verfolgt [17]. In einer Röntgenstrukturanalyse an einem Protein-Einkristall wurde der Röntgenpuls des ESRF-Synchrotrons in Grenoble mit dem Laser synchronisiert, der das CO photodissoziiert. Durch zeitliche Versetzung der beiden Pulse gelang es so, die Struktur zu bestimmten Zeiten nach der Laseranregung zu messen; es entstand erstmals ein „Film“ einer Proteinstruktur.

Optische und Infrarotspektroskopie geben zwar nur indirekt Informationen über Strukturänderungen, dafür sind sie jedoch weniger aufwändig, vielseitig einsetzbar, und können mit höchster Zeitauflösung und einem großen Dynamikbereich durchgeführt werden. So eignet sich zum Beispiel der CO-Ligand in Hämproteinen hervorragend als Sonde, da seine starke Infrarotbande stark von der Umgebung abhängt, insbesondere vom lokalen elektrischen Feld (Schwingungs-Stark-Effekt).

Die Konformationen A_I und A_{II} in Abb. 6a lassen sich leicht durch separate Infrarotbanden des CO-Liganden (1945 und 1955 cm^{-1}) unterscheiden. Zeitauflösende Messungen an diesen Banden nach Photodissoziation ermöglichen es, die Ligandenrückbindung am Häm-eisen, Details der Ligandenbewegungen im Protein und thermisch aktivierte Übergänge zwischen den Konformationen A_I und A_{II} über weite Bereiche der Zeit und Temperatur zu untersuchen [18].

Dabei treten interessante Eigenschaften zutage, die für eine raue Energielandschaft typisch sind. Relaxationsprozesse werden in der Physik im einfachsten Fall durch den exponentiellen Zerfall eines Ordnungsparameters (z. B. Polarisation, Magnetisierung), der so genannten Debye-Relaxation, beschrieben. In viskosen Flüssigkeiten, Polymeren und Biomolekülen hingegen sind Relaxationen oftmals nicht exponentiell, sondern zeigen ein breites Spektrum charakteristischer Zeiten. Dies ist eine direkte Konsequenz der Heterogenität der Energielandschaft. Oft ergibt die gestreckte Exponentialfunktion oder Kohlrausch-Funktion, $\Phi(t) = \exp[-(\kappa(T)t)^\beta]$, eine gute Beschreibung der Relaxation. So weist zum Beispiel der Übergang zwischen A_I und A_{II} in Abb. 6a ein $\beta = 0,6$ zwischen 200 und 300 K auf. Der charakteristische Ratenkoeffizient $\kappa(T)$ ist stark temperaturabhängig. Üblicherweise werden thermisch aktivierte Ratenprozesse mit dem Arrhenius-Gesetz modelliert, $\kappa(T) = A \exp[-E/kT]$. Die Beschreibung von kollektiven Proteinbewegungen mit dieser Relation ergibt jedoch in der Regel unphysikalische Werte für den Vorfaktor A und die Aktivierungsenergie E . Die Ferry-Bässler-Relation, $\kappa(T) = A \exp[-(E/kT)^2]$, wird vielfach in der Dynamik viskoser Flüssigkeiten und synthetischer Polymere verwendet; sie wurde für das „Random Energy Model“ theoretisch hergeleitet. Der Energieparameter E darin ist ein Maß für die statistische Rauigkeit der Energielandschaft. Für verschiedenste globale Strukturänderungen wurde eine Rauigkeit $E \approx 10$ kJ/mol gefunden [14, 19], welche auch typisch für viskose Lösungen von Wasserstoffbrückenbildnern wie Glycerin ist.

Ausblick

Proteine sind Makromoleküle von enormer Komplexität. Sie stehen auf der untersten Ebene der biologischen Selbstorganisation; dennoch ist die Physik der Proteinfaltung und -funktion noch nicht hinreichend verstanden. Biologische Systeme unterliegen der Evolution und damit dem Zufall. Muss man nun jedes Protein für sich im Detail betrachten – oder wird man in der Vielfalt der Proteine vereinfachende, vereinheitlichende Prinzipien finden können, die die Struktur, Dynamik und Funktion dieser Systeme generell erklären? Diese Erwartung wird durch die Überlegung gestützt, dass auch die Evolution physikalischen Gesetzen unterliegt, die sich den Kreationen der Evolution, von den Biomolekülen bis hin zu ganzen Organismen, aufprägen. Ein schönes Beispiel ist die kleine Zahl möglicher Faltungstopologien von Proteinketten, die mit „platonischen Körpern“ verglichen werden [3].

Weltweit werden insbesondere aufgrund der medizinischen Relevanz riesige Summen in die Untersuchung biomolekularer Prozesse investiert. Die Menge an experimentellen Daten über Proteine und andere Biomoleküle wird zukünftig rasant anwachsen. Um dieser Datenflut Herr zu werden, ist es unbedingt notwendig, die beschreibende, klassifizierende Betrachtung der Biologie mit quantitativen, von mathematischen Modellen und

Theorien geprägten Analysen der Physik zu ergänzen. Hier eröffnet sich für Physiker ein breites Arbeitsfeld mit hervorragenden wissenschaftlichen Perspektiven. Um dieser Entwicklung Rechnung zu tragen, werden zurzeit insbesondere in den USA, aber auch in Deutschland, verstärkt Professuren für biologische Physik eingerichtet. In nicht allzu ferner Zukunft sollte es möglich sein, Proteinstrukturen präzise vorherzusagen, Proteine durch genetische Modifikation gezielt und quantitativ in ihrer Funktion zu beeinflussen und Proteine *de novo* zu synthetisieren. Die Erkenntnisse dieser Forschung werden sich nicht auf die Biologie beschränken. Vielmehr sind wichtige Impulse auch für andere Gebiete der Physik zu erwarten, insbesondere bei der Herstellung und Untersuchung von Strukturen auf der Nanoskala.

Literatur

- [1] A. V. Finkelstein und O. B. Ptitsyn, Protein Physics, Academic Press, San Diego (2002)
- [2] G. U. Nienhaus und R. D. Young, in: Encyclopedia of Applied Physics, Vol. 15, hrsg. von G. L. Trigg, VCH Publishers, New York (1996), S. 163
- [3] M. Denton und C. Marshall, Nature **410**, 417 (2001)
- [4] J. R. Banavar und A. Maritan, Rev. Mod. Phys. **75**, 23 (2003)
- [5] C. Levinthal, in: Mössbauer Spectroscopy in Biological Systems, hrsg. von P. Debrunner, J. Tsibris und E. Münck, University of Illinois Press, Urbana, IL (1969), S. 22
- [6] K. A. Dill und H. S. Chan, Nature Struct. Biol. **4**, 10 (1997)
- [7] P. G. Wolynes und W. A. Eaton, Physics World, September 1999, S. 39
- [8] J. D. Bryngelson et al., Proteins **21**, 167 (1995)
- [9] M. Grübele, Ann. Rev. Phys. Chem. **50**, 485 (1999)
- [10] M. Rief und H. Grubmüller, Phys. Blätter, Februar 2001, S. 55
- [11] C. Hofmann et al., Proc. Natl. Acad. Sci. USA **100**, 15534 (2003)
- [12] F. G. Parak und G. U. Nienhaus, ChemPhysChem **3**, 249 (2002)
- [13] H. Frauenfelder, Proc. Natl. Acad. Sci. USA **99**, Suppl. 1, 2479 (2002)
- [14] G. U. Nienhaus, in: Biological Physics, hrsg. von H. Frauenfelder, G. Hummer und R. Garcia, APS Conference Proceedings 487, American Institute of Physics, Melville (1999), S. 99
- [15] A. J. Wand, Nature Struct. Biol. **8**, 926 (2001)
- [16] A. Ostermann et al., Nature **404**, 208 (2000)
- [17] F. Schotte et al., Science **300**, 1944 (2003)
- [18] J. M. Kriegl et al., Proc. Natl. Acad. Sci. USA **100**, 7069 (2003)
- [19] V. P. Denisov et al., Nature Struct. Biol. **3**, 505 (1996)

Der Autor

Gerd Ulrich Nienhaus hat an der Uni Münster Physik studiert und dort 1988 promoviert. Als Postdoc ging er anschließend an die University of Illinois at Urbana-Champaign (UIUC), wo er 1991 Assistant und später Associate Professor für Physik und Biophysik wurde. 1996 folgte er einem Ruf auf eine C4-Professur an die Uni Ulm und pendelt seither zwischen Ulm und der UIUC, mit der er als Adjunct Professor weiterhin verbunden ist. G. Ulrich Nienhaus ist Vorstandsmitglied der Deutschen Gesellschaft für Biophysik und Sekretär der IUPAP-Kommission für Biologische Physik.

