

- Huisgen u. H. Leitermann, Tetrahedron Lett. 12, 9 (1960); c) N. A. LeBel u. J. J. Whang, J. Am. Chem. Soc. 81, 6334 (1959).
- [49] N. A. LeBel, M. E. Post u. J. J. Whang, J. Am. Chem. Soc. 86, 3759 (1964); N. A. LeBel, Trans. N. Y. Acad. Sci. 27, 858 (1965).
- [50] N. A. LeBel u. E. G. Manucci, J. Org. Chem. 36, 2440 (1971).
- [51] a) W. Oppolzer u. K. Keller, Tetrahedron Lett. 1970, 1117; b) W. Oppolzer u. H. P. Weber, ibid. 1970, 1121; c) W. Oppolzer u. K. Keller, ibid. 1970, 4313; d) J. E. Baldwin, D. H. R. Barton, N. J. A. Gutteridge u. R. J. Martin, J. Chem. Soc. C 1971, 2184.
- [52] E. Gössinger, R. Imhof u. H. Wehrli, Helv. Chim. Acta 58, 96 (1975).
- [53] W. C. Lumma, Jr., J. Am. Chem. Soc. 91, 2820 (1969).
- [54] a) J. B. Bapat, D. S. C. Black, R. F. C. Brown u. C. Ichlòv, Aust. J. Chem. 25, 2445 (1972); b) J. J. Tufariello u. E. J. Trybulski, J. Chem. Soc. Chem. Commun. 1973, 720.
- [55] W. Oppolzer, M. Petrzilka, B. Bakker u. S. Siles, unveröffentlichte Versuche.
- [56] a) W. Oppolzer u. M. Petrzilka, J. Am. Chem. Soc. 98, 6722 (1976); b) unveröffentlichte Versuche; c) W. L. Scott u. D. A. Evans, J. Am. Chem. Soc. 94, 4779 (1972).
- [57] R. Huisgen, R. Fleischmann u. A. Eckell, Tetrahedron Lett. 1960, Nr. 12, S. 1; R. Huisgen u. A. Eckell, ibid. 1960, Nr. 12, S. 5.
- [58] W. Oppolzer u. H. P. Weber, Tetrahedron Lett. 1972, 1711.
- [59] W. Oppolzer, Tetrahedron Lett. 1970, 2199.
- [60] W. Oppolzer, Tetrahedron Lett. 1970, 3091.
- [61] R. Grashey, R. Huisgen, K. K. Sun u. R. M. Moriarty, J. Org. Chem. 30, 74 (1965).
- [62] W. Oppolzer, Tetrahedron Lett. 1972, 1707.
- [63] J. Bjørgo, D. R. Boyd u. D. C. Neill, J. Chem. Soc. Chem. Commun. 1974, 478.

## Regeln für Strukturen globulärer Proteine

Von Georg E. Schulz<sup>[\*]</sup>

Kann der Aufbau der so überaus komplexen dreidimensionalen Strukturen von nativen Polypeptidketten im Detail verstanden werden? Durch die Fülle der bisher gefundenen physiko-chemischen und phylogenetischen Gemeinsamkeiten zwischen Proteinen ist diese Frage zulässig geworden. Der augenblickliche Stand dieser Diskussion wird hier dargelegt.

### 1. Einleitung

Vor 16 Jahren wurde erstmals die Struktur eines globulären Proteins<sup>[1]</sup> aufgeklärt. Inzwischen sind etwa 40 Proteine bis zu atomarer Auflösung analysiert, bei weiteren zehn kennt man im wesentlichen die Gestalt der Polypeptidkette. Zweifellos ist damit die Zeit der Pioniere und des ersten Erstaunens vorüber. Fast selbstverständlich gehören heutzutage Proteinstrukturmodelle zum Handwerkszeug des Biochemikers, der seine Experimente mit ihrer Hilfe wesentlich fundierter planen und interpretieren kann als früher.

Die Fülle der vorhandenen Strukturdaten begünstigt aber nicht nur biochemisches Arbeiten, sie fordert uns auch heraus, die beobachteten Strukturen physikalisch zu verstehen und damit die Biologie in einem weit grundlegenderen Maße als bisher auf ihre physikalische Basis zurückzuführen. Nach genauem Verständnis der Architektur ist außerdem mit einer wesentlichen Ausweitung des Bereiches der analysierbaren Proteine zu rechnen, weil dann auch Proteine zugänglich werden, die nicht kristallisieren. Aber nicht nur rezente biologische Systeme, auch historische Zusammenhänge, insbesondere sehr frühe Stadien der Evolution, werden uns auf diese Weise erschlossen. Also lohnt es sich, hier einige Mühe zu investieren.

### 2. Energiebilanz

Globuläre Proteine bestehen aus linearen Polypeptidketten (Abb. 1), die durch gleichartiges Zusammenfügen von Aminosäuren synthetisiert werden<sup>[4]</sup>. Nach oder während der Synthese falten sich diese Ketten spontan zur exakten Raumstruktur. Die Spontaneität der Faltung wird heute als allgemeingültig

angesehen, weil sie an mehreren Proteinen in Renaturierungs-experimenten demonstriert werden konnte<sup>[5–7]</sup>. Implizit ist also die gesamte Information über die Raumstruktur bereits in der chemischen Struktur der Kette, d.h. in der Sequenz der Aminosäuren vorhanden. Die Faltung ist lediglich der Übergang in einen energetisch günstigeren Zustand der Kette.

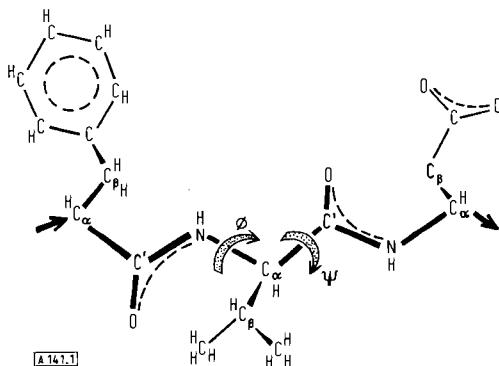


Abb. 1. Teil einer Peptidkette. Wegen Mesomerie sind die Peptidbindungen planar [2]. Die Diederwinkel  $\phi$  und  $\psi$  an den  $C_\alpha$ -Atomen bestimmen deshalb die Kettengestalt. Sterische Behinderungen lassen nur etwa 15 % aller möglichen  $\phi, \psi$ -Einstellungen zu [3]. Man beachte die relativ geringe Bewegungsfreiheit der verwendeten Seitenketten!

Dabei spielen im wesentlichen drei Wechselwirkungstypen eine Rolle: van-der-Waals-Kräfte zwischen unpolaren Gruppen, Dipolkräfte zwischen polaren Gruppen und hier speziell Wasserstoffbrücken, sowie „hydrophobe Kräfte“ (ein Synonym für Wasserentropie, welches die Tendenz der apolaren Seitenketten zur Bildung einer separaten hydrophoben Phase – also einer Art Ölträpfchen – beschreibt)<sup>[8]</sup>. In der Bilanz für die freie Enthalpie

$$\Delta G = \Delta H_{\text{Bindung}} - T \cdot \Delta S_{\text{Wasser}} - T \cdot \Delta S_{\text{Kette}} \quad (1)$$

[\*] Dr. G. E. Schulz  
Max-Planck-Institut für medizinische Forschung  
Jahnstraße 29, D-6900 Heidelberg

sind Bindungsenthalpien und Wasserentropie gemeinsame Gegenspieler der Kettentropie. Nur wenn sie genügend freie Enthalpie zusammenbringen, gelingt es, die Kettentropie zu übertreffen und die Kette stillzulegen, d. h. eine definierte Raumstruktur zu bilden. Schätzt man beide Parteien überschlagsmäßig ab, so findet man für Ketten aus etwa 100 Gliedern Werte von jeweils mehreren hundert kcal/mol. Das resultierende  $\Delta G$  liegt aber in der Größenordnung 10 kcal/mol<sup>[9]</sup>. Wir haben es also mit einem delikat ausbalancierten System zu tun, wodurch ein detailliertes Verständnis erschwert wird.

Die strukturell bekannten Proteine enthalten hydrophobe Kerne und bestätigen damit, wie wichtig der Wasserentropieterm in Gl. (1) ist. Die Packungsdichte dieser Kerne ist so hoch wie bei den dichtesten durch van-der-Waals-Kräfte stabilisierten Kristallen<sup>[10, 11]</sup>, d.h. die Anzahl der van-der-Waals-Kontakte und damit der zugehörige Bindungsenthalpieterm ist maximiert. Das spiegelt sich auch in der relativ geringen Kompressibilität der Proteine<sup>[12, 13]</sup> wider ( $\kappa_{\text{O}}/\kappa_{\text{Protein}} \approx 20$ ). Allerdings ist das Innere der Proteine nicht völlig apolar, sondern enthält etwa die Hälfte aller polaren Gruppen – vor allem die C=O- und N—H-Gruppen der Peptidbindungen<sup>[14]</sup>. Diese dem Wasser entzogenen Dipole bilden zu etwa 90 % Wasserstoffbrücken mit relativ hohen Energiebeiträgen<sup>[15]</sup>. Demnach werden Proteine zwar durch den Öltröpfcheneffekt stabilisiert, aber sie haben wenig mit einem lose und undefiniert gepackten Öltröpfchen gemeinsam – eher ähneln sie einem dicht gepackten Kristall.

Anhand von Gl. (1) wird plausibel, warum  $\alpha$ - und nicht  $\beta$ -Aminosäuren ausgesucht wurden, obwohl letztere in Versuchen zur Simulation der präbiotischen Synthese in vergleichbarer Menge produziert werden<sup>[16]</sup>. In einer Peptidkette aus  $\beta$ -Aminosäuren würde man fast freie Drehbarkeit um die  $C_\alpha$ — $C_\beta$ -Bindung und damit eine sehr viel größere Kettentropie erhalten. Zur Raumstrukturstabilisierung wären also erheblich höhere Bindungsenthalpien und/oder Wasserentropien notwendig. Hingegen ist die Kette aus  $\alpha$ -Aminosäuren wesentlich weniger beweglich (Abb. 1), kann also bequemer stillgelegt werden. Aus analogen Gründen dürften auch die verwendeten Seitenketten relativ unbeweglich sein (Abb. 1).

### 3. Faltungsweg

Bei der Faltung einer Kette von 100  $\alpha$ -Aminosäureresten sind etwa 300 Parameter einzustellen (Abb. 1). Selbst wenn jeder Parameter nur zwei Werte annehmen kann, sind  $2^{300}$  Konformationen möglich. Für ungezieltes Durchprobieren bis zur Auffindung einer energetisch stabilen Konformation würde das Erdalter nicht ausreichen, auch wenn nur  $10^{-13}$  s für eine Probe angesetzt werden. Also müssen die Parameter gezielt durchlaufen werden: Es muß einen Faltungsweg geben.

Die Existenz eines solchen Weges konnte für den Pankreatischen Trypsin-Inhibitor auch experimentell bewiesen werden. Dazu wurde vom statistischen Knäuel bis zum nativen Zustand verfolgt, in welcher Reihenfolge sich die möglichen S—S-Brücken bilden<sup>[17]</sup>. Unter den  $(\frac{1}{2} \cdot 100) = 15$  Möglichkeiten erscheint hier mit höchster Ausbeute zunächst eine der nativen Brücken (Abb. 2). Die Cysteine der begleitenden, nicht-nativen Brücken sind in der nativen Raumstruktur (Abb. 2) dicht benachbart. Demnach nähern sich im ersten Stadium der Faltung die Kettenbereiche um die Reste 5, 30 und 53 einander. Auch im weiteren Verlauf werden eine Reihe von falschen Verbin-

dungen geknüpft (Abb. 2). Das Konformationsgemisch ist also relativ breit, die Faltung ist nicht absolut zielsstrebig.

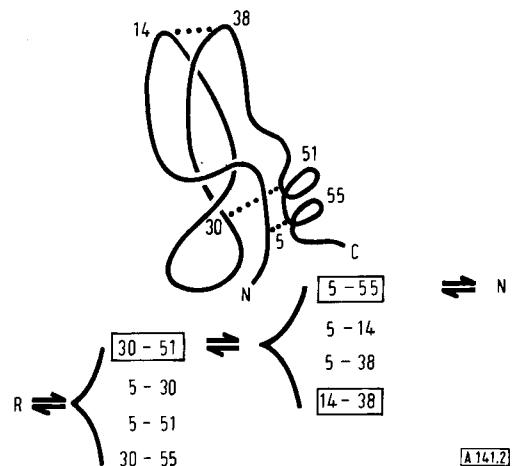


Abb. 2. Reihenfolge der S—S-Brückenbildung beim Renaturierungsprozeß des Pankreatischen Trypsin-Inhibitors nach Creighton [17]. Die Cysteins Seitenketten werden durch Disulfidaustausch-Reagentien verknüpft. Allgemein wird angenommen, daß die Brückenbildung nicht für die Kettenfaltung notwendig ist, also nur die Nachbarschaft der Cysteine anzeigen. Die nativen Brücken sind eingekrahmt. Die verfrühte Bildung der nativen Brücke 14–38 verhindert die weitere Faltung. Damit ist eindeutig gezeigt, daß ein bestimmter Faltungsweg eingehten werden muß. – Die Kettengestalt ist nach Huber et al. [18] gezeichnet. R = entfaltet, N = nativ.

Am Lysozym wurden Renaturierungsversuche bei 84°C unternommen<sup>[19]</sup>. Obwohl man damit weit oberhalb der Denaturierungs temperatur des Lysozyms von etwa 50°C arbeitet, zeigt sich eine vorübergehende Renaturierung. Demnach ist selbst bei dieser hohen Temperatur die Faltung noch schneller als die thermische Entfaltung. Eine schnelle Faltung ist aber nur möglich, wenn nicht lange probiert werden muß, weil ein ausgeprägter Faltungsweg existiert. Da solch ein Weg die Größe des zu berücksichtigenden Konformationsraumes erheblich verringert, können wir hoffen, daß er einmal mit theoretischen Ansätzen gefunden und durchlaufen – die Kettenfaltung also verstanden werden kann.

Da die native Raumstruktur auf einem bestimmten Weg erreicht wird, muß sie nicht dem absoluten Energieminimum im gesamten Konformationsraum entsprechen. Gerade bei komplizierten Strukturen sind sehr unebene Energiehyperflächen mit vielen lokalen Minima zu erwarten<sup>[7]</sup>. Mit relativ großer Wahrscheinlichkeit dürfte ein abgegrenzter Weg also zu einem lokalen Minimum führen. Jedoch kann hier prinzipiell nicht zwischen den Alternativen unterschieden werden, denn dazu wäre es notwendig, den ganzen Konformationsraum zu durchsuchen. Das kann weder die Kette selbst noch irgend eine Simulationsrechnung. Überdies sind alle kinetisch unerreichbaren Minima irrelevant, selbst wenn das absolute Minimum dazugehört. Die Diskussion über diese Frage<sup>[6, 19]</sup> gleicht also dem Streit um des Kaisers Bart.

### 4. Hierarchie

In den strukturell bekannten Proteinen sind mehrere Organisationsstufen voneinander unterscheidbar (Tabelle 1). Die Sequenz der Aminosäuren bildet die Basis. Wie oben erwähnt enthält sie implizit die gesamte Strukturinformation. Als nächst höhere Stufe sind die in Abbildung 3 dargestellten Sekundärstrukturen anzusprechen. Hierbei handelt es sich um

regelmäßige Kettenanordnungen, die eine optimale Bildung von Wasserstoffbrücken ermöglichen. Sekundärstrukturen sind außerordentlich häufig, in manchen Proteinen umfassen sie 75 % der Kette. Das dürfte mit der energetischen Notwendigkeit zusammenhängen, interne polare Gruppen durch Wasserstoffbrücken abzusättigen.

Tabelle 1. Organisationsstufen in Proteinen. Die Ordnung ist vertikal und partiell hierarchisch, d. h. im wesentlichen werden die Strukturelemente einer Stufe, z. B.  $\alpha$ -Helices (eine Art der Sekundärstruktur), nur von den jeweils darin enthaltenen Elementen der direkt darunterliegenden Stufe (z. B. Aminosäurereste) beeinflußt.

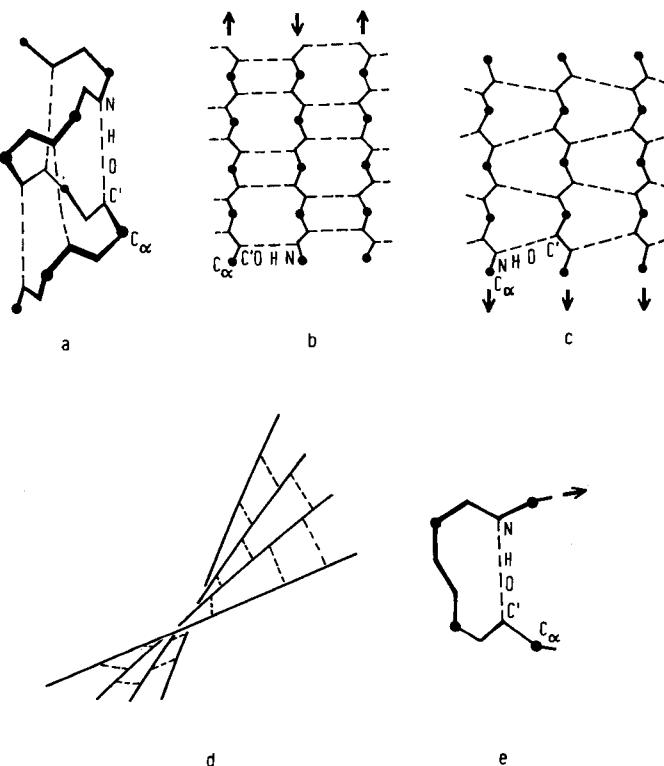
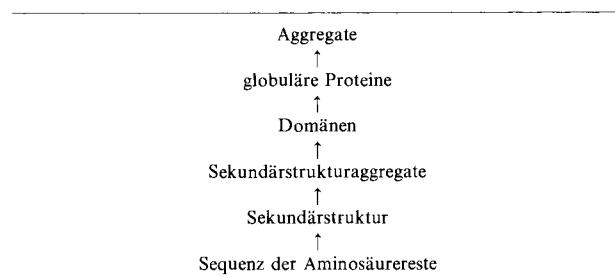


Abb. 3. Sekundärstrukturen in Polypeptidketten. a)  $\alpha$ -Helix, b) antiparalleles  $\beta$ -Faltblatt, c) paralleles  $\beta$ -Faltblatt, d) Verdrehung aller Faltblätter, e) Haarnadelbiegung. Bei allen Strukturen werden Wasserstoffbrücken zwischen den  $\text{C}=\text{O}$ - und den  $\text{N}-\text{H}$ -Gruppen der Peptidbindungen gebildet.

In einigen Fällen bilden Sekundärstrukturen regelmäßige Aggregate („Supersekundärstrukturen“). Ein unumstrittenes Beispiel ist die Doppel- $\alpha$ -Helix<sup>[20]</sup>, bei der zwei  $\alpha$ -Helices über Seitenketten miteinander verzahnt sind und sich mit einer Periode von etwa 180 Å umeinanderwinden (Abb. 4). Eine weitere Supersekundärstruktur dürfte die in einigen Dehydrogenasen<sup>[27]</sup> und anderen Proteinen vorkommende Faltblatt-Helix-Kombination sein (Abb. 4). Neben diesen klar definierten Supersekundärstrukturen gibt es eine Reihe von Strukturpräferenzen, die weniger bestimmt den gleichen Tatbestand ausdrücken. Sie werden weiter unten beschrieben.

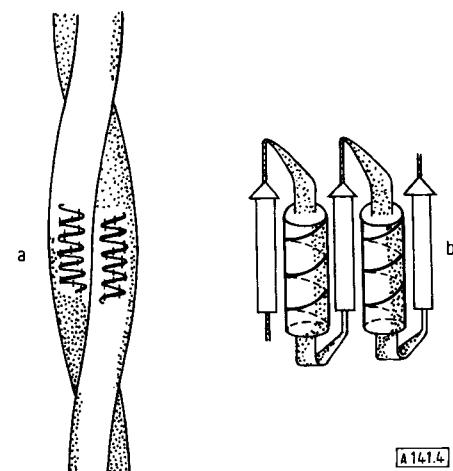


Abb. 4. Sekundärstrukturaggregate oder „Supersekundärstrukturen“. a) Doppel- $\alpha$ -Helix nach Crick [20], gefunden in Paramyosin [21], Myosin [22],  $\alpha$ -Keratin [23] und Tropomyosin [24]. Ähnliche Strukturen liegen wahrscheinlich in Hämerythrin [25] und Bakteriorhodopsin [26] vor. b) Faltblatt-Helix-Struktur, bestehend aus drei Faltblattsträngen und zwei  $\alpha$ -Helices, gefunden in Lactat- [27], Malat- [28], Alkohol- [29] und Glyceraldehyd-3-phosphat-Dehydrogenase [30] sowie Adenylyl-Kinase [31], Flavodoxin [32, 33], Subtilisin [34], Phosphoglycerat-Kinase [35] und Phosphorylase [36].

Der nächst höhere Komplexitätsgrad wird in der Domäne erreicht. Vor allem die größeren Proteine mit mehr als 150 Aminosäureresten können strukturell meist in zwei oder mehr räumlich getrennte Bereiche unterteilt werden. Am deutlichsten tritt das bei den Immunoglobulinen zutage, deren leichte und schwere Ketten zwei bzw. vier separate Domänen bilden (Abb. 5). Die Domänen scheinen diejenigen Bereiche der Kette zu enthalten, die sich unabhängig voneinander zur Raumstruktur falten.

Die aus einer oder mehreren Domänen gebildeten globulären Proteine (Tabelle 1) müssen nicht die Endstufe des Selbstaufbaus darstellen. In vielen Fällen sind ihre Oberflächen so konstruiert, daß sie zu größeren Gebilden aggregieren, z. B. Enzymkomplexen, Ribosomen, Virushüllen und Muskelfasern. Dabei sind die Aggregate üblicherweise symmetrisch, wenn sich nur ein oder zwei Komponententypen zusammenfinden (z. B. Untereinheiten zur Glutathion-Reduktase<sup>[38]</sup> oder Lactat-Dehydrogenase<sup>[39]</sup>, regulative und katalytische Untereinheiten zur Aspartat-Transcarbamylase<sup>[40]</sup>, Penton- und Hexon-Untereinheiten zur Adenovirushülle<sup>[41]</sup>, Myosin zum Myosinfilament<sup>[22]</sup>). Die Symmetrie ergibt sich, weil die Oberflächenkontakte optimiert werden und überall gleich sind.

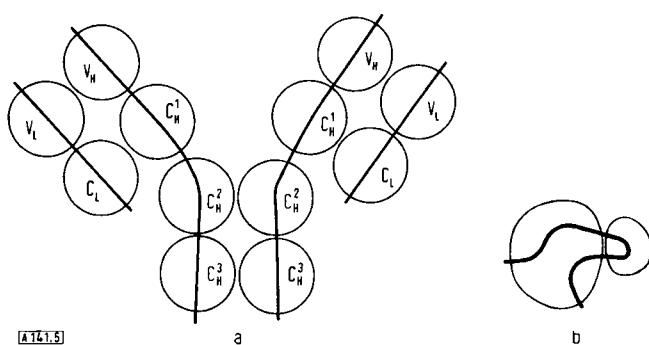


Abb. 5. Domänenstruktur globulärer Proteine. a) Die leichten (Index L) und schweren Polypeptidketten (Index H) der Immunoglobuline falten sich zu zwei bzw. vier Domänen [37], d. h. räumlich voneinander abgegrenzten Strukturbereichen. b) Die Adenylyl-Kinase [31] bildet eine große und eine kleine Domäne, die nicht wie üblich durch einen, sondern durch zwei Kettenstränge verbunden sind. Dadurch zeigt die große Domäne eine geringere Nachbarschaftskorrelation als die kleine.

Zum Verständnis der Kettenfaltung wäre es sehr hilfreich, wenn diese vertikale Ordnung streng hierarchisch wäre, d. h. wenn die Strukturelemente einer Stufe (z. B.  $\alpha$ -Helix) nur von den jeweils darin enthaltenen Elementen der direkt darunterliegenden Stufe (z. B. Aminosäurereste) beeinflußt wären und nicht von anderen Teilen des Proteins oder anderen Strukturstufen. Dann könnte die Faltung für jeden Schritt getrennt analysiert werden, und das schwierige Gesamtproblem wäre in mehrere, wesentlich leichter lösbare Probleme separiert. Am ehesten gilt das für die Aggregation, denn in den meisten Fällen wird die Struktur des freien Monomers dabei nur unwe sentlich verändert. Man kann also erst die Struktur des Monomers bestimmen und dann anhand der Oberflächen die der Aggregate. Jedoch gibt es auch Gegenbeispiele wie etwa ribosomale Proteine, deren Kette fast völlig gestreckt ist<sup>[42]</sup>. Sie können allein keine definierte Struktur annehmen, sondern erst nach Assoziation mit anderen Teilen des Ribosoms. Wahrscheinlich ist auch die Bildung der  $\alpha$ -Helices relativ unabhängig von der später eingenommenen Raumstruktur. Gleichtes kann man für das Zusammenfinden der Domänen zum globulären Protein annehmen. Demnach ist die vertikale Ordnung (Tabelle 1) zwar nicht streng hierarchisch, aber doch so weitgehend hierarchisch, daß sie eine akzeptable Arbeitshypothese ist.

## 5. Separation der Sekundärstruktur

Nachdem die von *Pauling* und *Corey* postulierten Sekundärstrukturen<sup>[43]</sup> in mehreren globulären Proteinen gefunden waren, lag ein Hierarchietest zwischen Sekundärstrukturen und Aminosäuresequenz (Tabelle 1) nahe. Bei strenger Hierarchie dürfte z. B. eine  $\alpha$ -Helix nur von den in ihr enthaltenen Aminosäuren und von keinem anderen Teil der Kette determiniert sein. Anhand der experimentellen Daten sollte die Korrelation zwischen Helixbildung und Aminosäuresequenz auffindbar sein. Mit bekannter Korrelation könnten dann die Amino-

säuresequenzen strukturell unbekannter Proteine untersucht und helicale Bereiche vorhergesagt werden.

Dieser Weg wird seit einiger Zeit von mehreren Gruppen verfolgt. Vorhergesagt werden  $\alpha$ -Helices,  $\beta$ -Faltblattstränge und Haarnadelbiegungen (Tabelle 2). Die rein statistischen, leicht nachvollziehbaren Methoden stammen aus der Zeit des ersten Optimismus. Sie wurden inzwischen von wesentlich verfeinerten Systemen überwachsen. Weiterhin wurde versucht, die Korrelation durch Hinzuziehung anderer physikalischer Daten zu verbessern. Insgesamt sind die leicht nachvollziehbaren Methoden wesentlich populärer als diejenigen, die praktisch nur von den Autoren selbst angewendet werden können. Durch ihre vereinfachende Formulierung haben sie jedoch an Informationsgehalt verloren.

Tabelle 2. Klassifizierung der Methoden zur Vorhersage von Strukturedaten.

	Rein statistische Auswertung der Strukturdaten	Hinzunahme weiterer Daten: Energiebilanz, Strukturen synthetischer Polymere, sterische Betrachtungen
ohne Rechenprogramm und Datei nachvollziehbar	Literaturzitate [44–47]	Literaturzitate [48–50]
nur mit Rechenprogramm und Datei nachvollziehbar	Literaturzitate [51–54]	Literaturzitat [55]

Für lange Zeit fanden Vorhersagen wenig Beachtung, weil sie entweder bereits bekannte Sekundärstrukturen *a posteriori* bestimmten oder bei strukturell unbekannten Proteinen ins Leere schossen. Diese Situation änderte sich mit dem Testfall Adenylylat-Kinase, wo eine gerade aufgeklärte, aber noch allgemein unbekannte Struktur bearbeitet wurde<sup>[56, 57]</sup>. Die in Abbildung 6 dargestellten Ergebnisse zeigen deutlich, daß der auf strenger Hierarchie basierende Methodenansatz erfolgreich ist (Tabelle 3). Ein zweiter Test am T4-Phagen-Lyszym<sup>[58]</sup> mit weniger ausgeprägter Sekundärstruktur ergab zwar

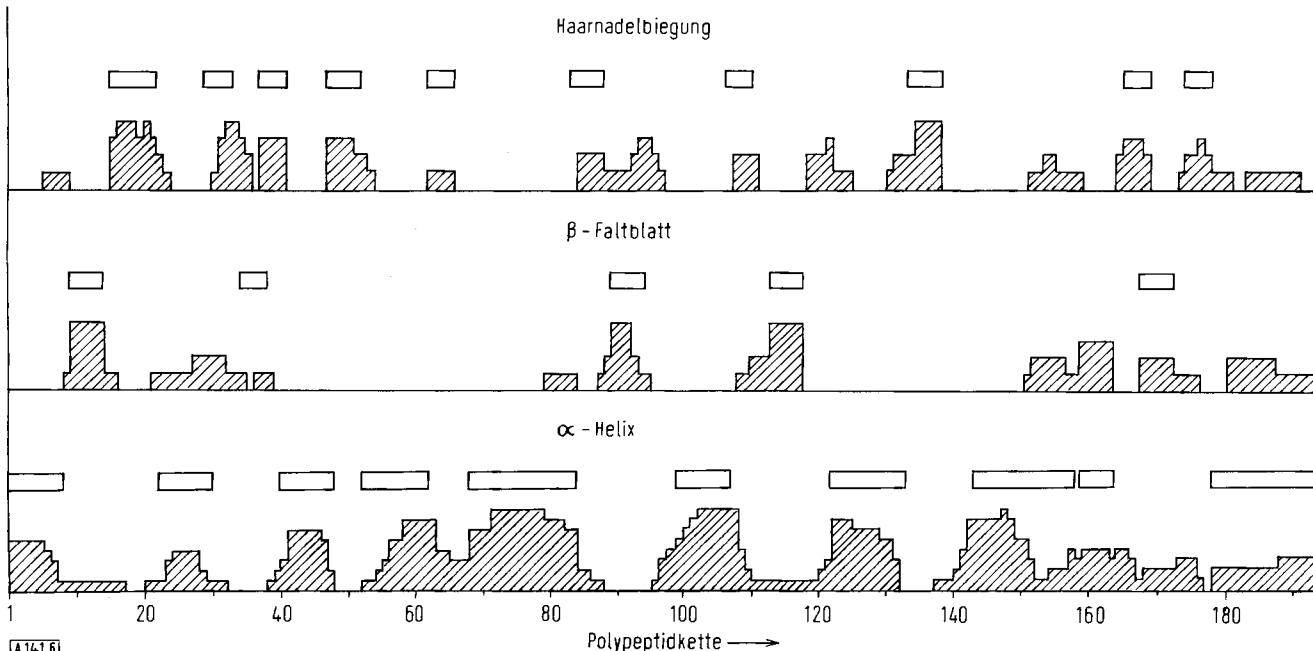


Abb. 6. Zusammenfassung aller Sekundärstrukturvorhersagen (schraffierte) für die Adenylylat-Kinase durch Addition [56]. Dabei wurde jeder Methode das gleiche Gewicht zugeschlagen. Die experimentell beobachtete Sekundärstruktur ist zum Vergleich eingetragen (Rechtecke).

nur geringe Korrelationen (Tabelle 3) und modifizierte damit den Gesamteindruck, verwischte ihn jedoch nicht.

Tabelle 3. Korrelationskoeffizienten zwischen vorhergesagter und experimentell beobachteter Sekundärstruktur [56, 58]. In der Adenyat-Kinase blieben durch den hohen Helixgehalt (54 %) bei relativ guter Helixvorhersage nur kleinere Bereiche für die anderen Sekundärstrukturen übrig. Deshalb sind deren Korrelationskoeffizienten so hoch.

	$\alpha$ -Helix	$\beta$ -Faltblatt	Haarnadelbiegung
Adenyat-Kinase	+0.44	+0.51	+0.49
T4-Phagen-Lysozym	+0.28	+0.05	+0.13

Insgesamt werden die  $\alpha$ -Helices wesentlich exakter gefunden als  $\beta$ -Faltblätter und Haarnadelbiegungen (Tabelle 3). Da die Datenbasen für alle drei Sekundärstrukturen vergleichbare Größe haben, ist demnach die Hierarchieforderung (Tabelle 1) bei den  $\alpha$ -Helices am besten erfüllt. Die Helixbildung wird von entfernten Kettenbereichen am wenigsten beeinflußt; Helices können sich am ehesten selbst falten. Damit sind sie prominente Kandidaten für Faltungsstartbereiche oder Nucleationszentren, und zwar umso mehr, je besser sie vorhergesagt werden können.

Im N-terminalen Bereich der Kette wurden in beiden Tests die  $\alpha$ -Helices besser ermittelt als im C-terminalen; die Faltung sollte also im N-terminalen Bereich beginnen. Da die Kette an Ribosomen vom N- zum C-terminalen Ende synthetisiert wird, dürfte sie sich bereits während der Synthese falten. In diesem Falle simulieren Renaturierungsexperimente an fertigen Ketten nicht ganz die natürlichen Verhältnisse. Man sollte sich also nicht wundern, wenn einige Ketten sich der Faltung *in vitro* widersetzen. In der Adenyat-Kinase wurden drei Helices (Reste 41–84), die eine kleine Domäne bilden, außerordentlich klar erkannt (Abb. 6). Möglicherweise ist diese Domäne ein Kern, an den sich der Rest der Kette bei der Faltung anlehnt.

$\beta$ -Faltblätter werden aus Kettensträngen aufgebaut, die entlang der Kette weit voneinander entfernt sein können. Sie können nur zusammenfinden, wenn das dazwischenliegende Kettenstück bestimmte Konformationen annimmt. Bei solcher Abhängigkeit von anderen Teilen der Kette kann man kaum strenge Hierarchie erwarten. Entsprechend ungenau sind auch die Vorhersagen (Tabelle 3).

Die Haarnadelbiegungen befinden sich meist an der Proteinoberfläche<sup>[59]</sup>. Sie sind energetisch nicht ausgezeichnet, denn sie enthalten nur eine einzige Wasserstoffbrücke, die überdies oft in Konkurrenz mit dem umgebenden Wasser steht, und sind wohl nur als passive Knickstellen anzusehen. Wahrscheinlich wird die Position des Knicks nicht nur von der lokalen Aminosäuresequenz, sondern auch von der Faltung der benachbarten Kettenbereiche beeinflußt, wodurch die Hierarchieforderung ungenügend erfüllt ist. Bei der Adenyat-Kinase (Abb. 6) wurden diese Stellen mit großer, beim T4-Phagen-Lysozym jedoch nur mit geringer Genauigkeit gefunden (Tabelle 3).

Der unterschiedliche Erfolg bei diesen beiden Proteinen spiegelt die Unterschiede zwischen deren Strukturen wider. Während Adenyat-Kinase ein zentrales Faltblatt mit umgebenden Helices enthält, also recht gut geordnet ist, hat T4-Phagen-Lysozym weniger und kaum geordnete Sekundärstruktur. Hier dürften nicht nur die resultierenden Strukturen, sondern auch die Faltungswege qualitativ verschieden sein. Es erscheint

Tabelle 4. Strukturklassen globulärer Proteine.

Strukturklasse	Beispiele
Proteine mit wenig und mit kaum geordneter Sekundärstruktur	Ribonuclease [60], Lysozym [61], Ferredoxin [62]
Proteine mit überwiegend antiparallelem Faltblatt	Immunoglobuline [63–66], Superoxid-Dismutase [67], Concanavalin A [68], Präalbumin [69], Chymotrypsin [70], Bakteriochlorophyll-Protein [71]
Proteine mit zentralem, meist parallelem Faltblatt und umgebenden Helices	Malat-Dehydrogenase [28], Lactat-Dehydrogenase [39], Adenyat-Kinase [31], Phosphoglycerat-Kinase [35], Triose-phosphat-Isomerase [72], Carboxypeptidase [74], viele Enzyme der Glykolyse [73]
Proteine mit überwiegender $\alpha$ -Helix	Myoglobin [1], Hämoglobin [75], Hämerythrin [25], Bakteriorhodopsin [26]

also sinnvoll, die Proteine aufgrund ihrer Sekundärstruktur zu klassifizieren (Tabelle 4), denn wahrscheinlich muß das Faltungsproblem bei den einzelnen Klassen unterschiedlich angefaßt werden.

## 6. Strukturpräferenzen

Sekundärstrukturen kommen in globulären Proteinen sehr häufig vor, weil sie strukturell ausgezeichnet sind. Unter den Helices dominiert dabei die  $\alpha$ -Helix mit geradlinigen Wasserstoffbrücken zwischen den Kettengliedern  $n$  und  $n+4$ . In geringem Maße beobachtet man auch  $3_{10}$ -Helices mit  $n$  und  $n+3$ , aber nie  $\pi$ -Helices mit  $n$  und  $n+5$ . Dieser Befund läßt sich dadurch erklären, daß bei der  $3_{10}$ -Helix die Wasserstoffbrücken gewinkelt, also energetisch ungünstiger sind. Die  $\pi$ -Helix hat zwar geradlinige Wasserstoffbrücken, aber sie läßt um ihre Achse einen großen Hohlraum frei, nutzt also längst nicht alle möglichen van-der-Waals-Kontakte aus.

Bei den Faltblättern hat sich bisher keine Bevorzugung des parallelen, antiparallelen oder gemischten Typs (Abb. 3) herauskristallisiert. Es fällt jedoch auf, daß kein Faltblatt flach ist; alle sind rechtshändig verdrillt (Abb. 3). Durch Blattverdrillung kann hier die Verdrillung der einzelnen Ketten um ihre Längssachse (von etwa 20° pro Glied) für ca. fünf Aminosäurereste ohne erhebliche Störung der Wasserstoffbrücken kompensiert werden. Die Kettenverdrillung wiederum entspricht einer energetisch besonders günstigen Kettenkonformation, d. h. besonders geringer sterischer Behinderung an den  $C_\alpha$ -Atomen (Abb. 1). Energetisch ist das verdrillte Blatt um ca. 0.5 kcal/mol

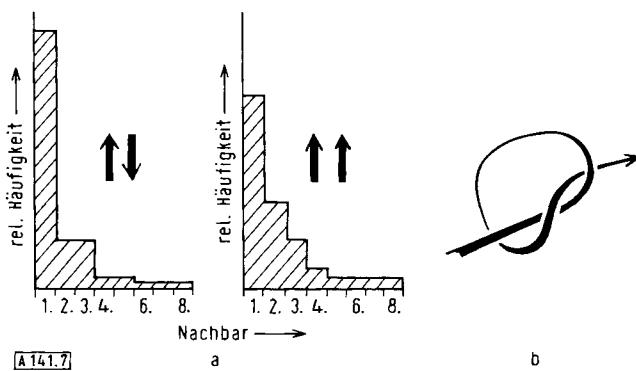


Abb. 7. Strukturpräferenzen in globulären Proteinen. a) Beobachtete Nachbarschaftskorrelation in antiparallelen und parallelen  $\beta$ -Faltblättern, d. h. relative Häufigkeit, mit der die Kette vom Referenzstrang eines Faltblatts als nächstes den 1., 2., ... Nachbarstrang durchläuft. b) Knoten in der Polypeptidkette; bisher nicht beobachtet.

pro Kettenglied günstiger als das flache Blatt<sup>[76]</sup>. Das demonstriert erneut, wie wichtig kleine Energiebeiträge bei der Proteinfaltung sind.

Unter den Sekundärstrukturaggregaten (Tabelle 1) ist nur die Doppel- $\alpha$ -Helix umstritten. Daneben existieren eine Reihe von weniger bestimmten Faltungspräferenzen der Kette. So zeigen die einzelnen Stränge in  $\beta$ -Faltblättern eine sehr klare Nachbarschaftskorrelation<sup>[67]</sup>, die bei den antiparallelen Strängen noch stärker ausgeprägt ist als bei den parallelen (Abb. 7). In der Raumstruktur finden sich jeweils Bereiche zusammen, die entlang der Kette nicht weit voneinander getrennt sind. Grob gesprochen verhält sich die Polypeptidkette wie eine senkrecht gehaltene Schnur, die man auf eine Unterlage fallen läßt. Sie verheddet sich nicht und kann problemlos wieder aufgenommen werden. Dieses Bild wird insofern bestätigt, als bisher keine der bekannten Strukturen einen „Knoten“ enthält. Dabei wird der Begriff „Knoten“ nicht im mathematischen Sinne (mathematische Knoten gibt es nur bei geschlossenen Fäden), sondern im herkömmlichen Sinne verwendet (Abb. 7).

In parallelen Faltblättern muß die Kette zwischen den Strängen jeweils von der Oberseite zur Unterseite des Blattes zurückkehren. Hierbei kann die Verbindung entweder über die Vorder- oder die Rückseite geführt werden (Abb. 8) und bildet dann mit den beiden Faltblattsträngen eine Rechts- bzw. Linksschraube. In den bekannten Strukturen überwiegt die Rechtsschraube etwa im Verhältnis 50:1<sup>[77]</sup>. In Abbildung 8 ist demonstriert, daß die Rechtsschraube wegen der rechtsdrehigen Verdrillung der Faltblätter (Abb. 3) eine bessere Nachbarschaftskorrelation ermöglicht als die Linksschraube. Umgekehrt zeigt die beobachtete Dominanz, wie wichtig die Nachbarschaftskorrelation ist.

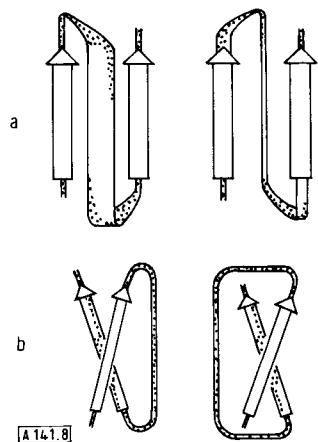


Abb. 8. Chiralität der  $\beta$ -Strang- $\alpha$ -Helix- $\beta$ -Strang-Gruppe in parallelen Faltblättern. Die  $\alpha$ -Helix ist nur als glattes Band dargestellt. Die Faltblattstränge brauchen keine direkten Nachbarn zu sein. Durch die Faltblattverdrillung zeigt die rechtshändige Gruppe eine wesentlich bessere Nachbarschaftskorrelation als die linkshändige. a) Von vorn, b) von der Seite gesehen.

Weitere Strukturpräferenzen zeigen sich bei der Bildung von Domänen. Immunoglobulinmoleküle sind hier das Paradebeispiel (Abb. 5). Gewöhnlich sind Domänen nur durch einen einzigen Strang der Peptidkette miteinander verbunden, sie sind also nicht nur räumlich separiert, sondern auch entlang der Kette<sup>[37]</sup>. Bei Einzelverbindung ergibt sich eine bessere Nachbarschaftskorrelation in den Domänen als bei Mehrfachverbindungen (Abb. 5). Da die Nachbarschaftskorrela-

tion innerhalb der Domänen wesentlich höher ist als zwischen den Domänen, ist anzunehmen, daß sich die Domänen getrennt falten. Dafür spricht auch die Struktur des Chymotrypsins<sup>[70]</sup>. Hier wurden im Inneren des Proteins zwischen den beiden Faltblattzylinern, die als Domänen anzusprechen sind, 13 Wassermoleküle gefunden (Abb. 9). Also falten sich die Zylinder wahrscheinlich getrennt und setzen bei der Aggregation die Wassermoleküle fest. Charakteristischerweise befindet

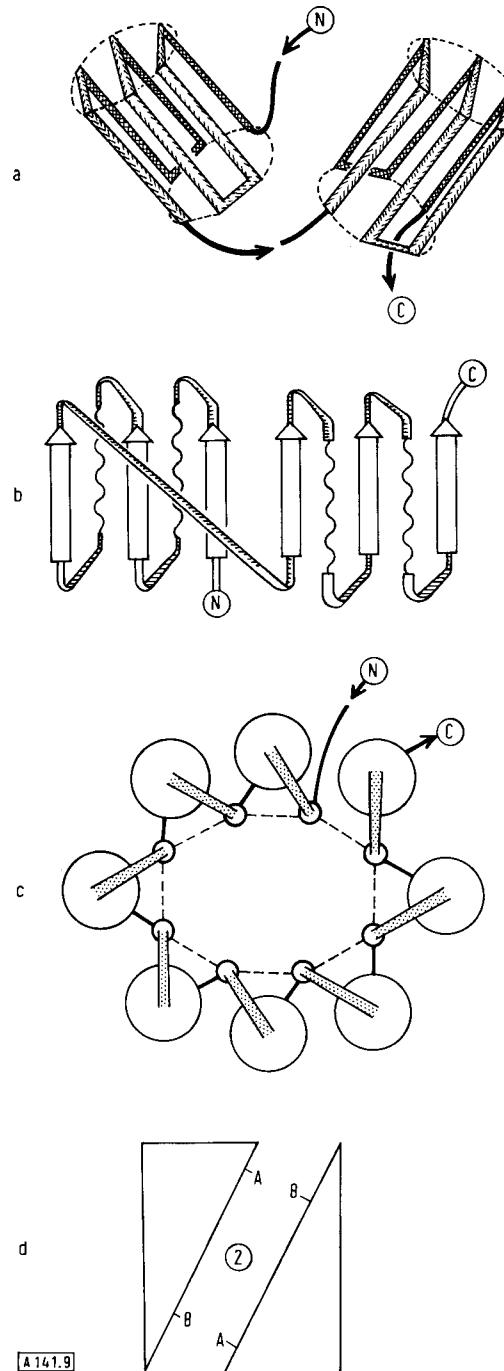


Abb. 9. Symmetrie in der Kettengestalt (schematische Darstellung). a) Chymotrypsin [70]. Die Faltblattzyliner sind je eine Domäne; sie sind antiparallel und haben interne zweizählige Symmetrie. b) Die doppelte nucleotidbindende Domäne der Lactat-Dehydrogenase und anderer Proteine [39, 79, 93]. Hier existiert eine vertikale zweizählige Achse. c) Die außerordentlich symmetrische Struktur der Triose-phosphat-Isomerase [72], bestehend aus einem parallelen Faltblattzyliner und acht umgebenden  $\alpha$ -Helices. d) Kontaktoptimierung und Symmetrie: Wenn der Kontakt A-B besonders günstig ist, dann wird er doppelt ausgeführt und ergibt zweizählige Symmetrie. In Monomeren wurden bisher nur zweizählige Achsen beobachtet, in Aggregaten dagegen die Punktgruppen 2, 3, 17 (Protein des Tabakmosaikvirus), 222, 32, 432 (Apoferretin) und 532 (Kugelviren).

sich in diesen wie auch in vielen anderen Fällen das aktive Zentrum zwischen den Domänen.

Die übliche Domänengrenze liegt unter etwa 150 Aminosäureresten. Nach der Energiebilanz [Gl. (1)] sollte eigentlich die Bildung von größeren Domänen günstig sein, denn in größeren Globuli liegen relativ gesehen weniger Seitenketten an der Oberfläche, die Energie der hydrophoben Bindung ( $\Delta S_{\text{Wasse}}^*$ ) und die Energie der internen Wasserstoffbrücken kann also erhöht werden. Trotzdem wird bei langen Ketten immer die Faltung mehrerer kleiner Domänen mit hoher Nachbarschaftskorrelation beobachtet. Das scheint einfacher zu sein als die Faltung einer einzigen großen und damit komplexeren Domäne.

Zwischen kleineren Domänen, wie im Falle der Adenylyl-Kinase, gibt es oft zwei Verbindungen, was unterschiedliche Nachbarschaftskorrelationen in den Domänen mit sich bringt (Abb. 5). Hier faltet sich wahrscheinlich zunächst die Domäne mit der höheren Korrelation und dient dann als Stützgerüst für den Aufbau des restlichen Moleküls. Diese Annahme wird bei der Adenylyl-Kinase durch die besonders hohe Vorhersagegenauigkeit in der Domäne mit der höheren Nachbarschaftskorrelation gestützt (Abb. 6).

Wie oben beschrieben bilden sich bei der Aggregation von monomeren Proteinen symmetrische Strukturen, falls nur wenige Bausteinsorten verwendet werden. Diese Symmetrie zeigt sich auch bei der Aggregation von Domänen, besonders wenn sie strukturell ähnlich sind wie diejenigen der leichten und schweren Ketten der Immunoglobuline<sup>[63–66]</sup>, der Rhodanese<sup>[78]</sup>, des Ferredoxins<sup>[62]</sup>, der Dehydrogenasen<sup>[79]</sup> und der Triose-3-phosphat-Isomerase<sup>[72]</sup> (Abb. 9). Aber nicht nur Domänen, sondern auch Sekundärstrukturen scheinen symmetrisch zu aggregieren, z. B. im Parvalbumin<sup>[80, 81]</sup> und in den Faltblattzylinern des Chymotrypsins (Abb. 9). Ursächlich entsteht Symmetrie, weil Kontakte optimiert werden (Abb. 9). Auch in sehr frühen Stadien der Kettenfaltung scheint also diese Optimierung wichtig zu sein.

## 7. Strukturregeln durch Phylogenie

Eine große Hilfe beim Verständnis der Proteinstrukturen ist deren historischer Hintergrund. Schließlich haben sich alle diese Strukturen im Laufe der biologischen Evolution kontinuierlich auseinander und miteinander entwickelt. Hier hat uns die Natur eine ganze Reihe von bereits ausgeführten Experimenten beschert. Wir müssen nur die Ergebnisse interpretieren, d. h. aus gemeinsamen Strukturmerkmalen allgemeine Aussagen extrahieren.

Der elementare Schritt in der Evolution der Proteine ist die Änderung einer Seitenkette. Daneben gibt es Insertionen und Deletionen von Aminosäureresten, bei denen auch das Kettenrückgrat betroffen ist und die auch längere Kettenabschnitte umfassen können. Analysiert man Änderungen im Proteininneren, so zeigt sich, daß konservativ ausgetauscht wird (apolare gegen apolare Seitenketten, etc.), die Energiebilanz also nur wenig gestört wird. Außerdem werden entstehende Lücken (z. B. durch Auslassen einer Methylgruppe bei Isoleucin→Valin) durch einen anderen Austausch kompensiert, d. h. die hohe Packungsdichte wird beibehalten. Sie ist offenbar sehr wesentlich für die Energiebilanz.

Anhand der Globinfamilie wurde der Einfluß des Aminosäuren austauschs auf den Faltungsvorgang untersucht<sup>[82]</sup>. Hier

ergab sich, daß alle  $\alpha$ -Helices an ihrem N-terminalen Ende entweder ein Prolin oder eine kurze polare Seitenkette tragen, die Wasserstoffbrücken zum Kettenrückgrat bilden kann (Asn, Asp, His, Ser, Thr). Wahrscheinlich werden diese Wasserstoffbrücken und/oder das Prolin, welches den in Abbildung 1 definierten Diederwinkel  $\phi$  auf den  $\alpha$ -Helixwert festlegt, zur Initiierung der Helix benötigt, weil sie frühzeitig die Anzahl der Konformationen an dieser Stelle vermindern, also einen Miniaturfaltungskern bilden.

Im Laufe der Zeit akkumulieren sich die Veränderungen, und wenn etwa 80 % der Aminosäuren ausgetauscht sind, können Gemeinsamkeiten anhand der Sequenz nur noch in Spezialfällen erkannt werden. Hier hat sich zunächst beim Vergleich des Myoglobins<sup>[1]</sup> mit den Hämoglobinketten<sup>[83]</sup> und später beim Cytochrom c<sup>[84]</sup> und bei den Serin-Proteasen<sup>[85]</sup> herausgestellt, daß sich zwar die Aminosäuren, nicht aber die Kettengestalt ändert. Der Verlauf der Kette sowie die Lage von prosthetischen Gruppen, beispielsweise des Häms, wurden außerordentlich gut konserviert. Daran ist ablesbar, daß nicht nur die Energiebilanz delikat ausbalanciert ist, sondern auch die Faltungsdynamik, denn zur Erfüllung der Energieerfordernisse sollte die Kettengestalt weniger wichtig sein als die verwendeten Aminosäuren. Wohl aber dürfte die Kettengestalt sehr eng mit dem Faltungsweg und demnach mit der Faltungsdynamik zusammenhängen.

Neben den kontinuierlichen kleineren evolutionären Veränderungen gibt es auch große diskontinuierliche, die durch Genfusionen und Genduplikationen hervorgerufen wurden. Ein rezentes Beispiel für Fusion sind die Immunoglobuline, wo wahrscheinlich die Gene der variablen mit denen der konstanten Kettenbereiche verknüpft werden<sup>[86]</sup>. Genseparation und -fusion müßte auch in Dehydrogenasen stattgefunden haben, denn die Lactat-Dehydrogenase hat ihre NAD-bindenden Domänen als N-terminale Hälften, während eine außerordentlich ähnliche Domäne in der C-terminalen Hälfte der Alkohol-Dehydrogenase<sup>[29]</sup> verwendet wird. Die meisten Genduplikationen führen zu separaten Ketten und damit zu sich getrennt entwickelnden Proteinen<sup>[87]</sup>: Globine, Serin-Proteasen, Mittelhirn-Hormone, Lysozym,  $\alpha$ -Lactalbumin, etc. Einige Duplikationen verdoppeln jedoch die Kettenlänge, und zwar in Ferredoxin<sup>[62]</sup>, Rhodanese<sup>[78]</sup>, Chymotrypsin<sup>[88]</sup>, Parvalbumin<sup>[80, 81]</sup>, Troponin C<sup>[89]</sup>, innerhalb der NAD-bindenden Domänen der Dehydrogenasen<sup>[79]</sup>, in Tropomyosin<sup>[90]</sup> etc. Wie oben beschrieben zeigen solche Ketten innere Symmetrie. Wahrscheinlich wird der Faltungsmechanismus hier nur insofern modifiziert, als sich nun zwei Kerne gleichartig, aber separat falten und dann aggregieren. Die Fülle der bisher gefundenen Beispiele zeigt<sup>[87]</sup>, daß diese diskontinuierlichen Änderungen sehr wesentlich an der Herstellung neuer Proteinstrukturen beteiligt sind.

## 8. Phylogenie anhand von Strukturen

Im vorhergehenden Kapitel wurde dargelegt, daß die historische Entwicklung der Proteine erlaubt, Regeln für deren Strukturen zu extrahieren. Umgekehrt vermittelt aber die Strukturanalyse auch Erkenntnisse über phylogenetische Zusammenhänge. Sie wird vor allem dann wichtig, wenn Proteine sich soweit auseinander entwickelt haben, daß Ähnlichkeiten nicht mehr zwischen den Aminosäuresequenzen, sondern nur noch zwischen den Kettengestalten existieren. Da offenbar

die Kettenketten ganz besonders gut konserviert werden, deuten solche Ähnlichkeiten auf Verbindungslien in sehr frühen Stadien der Evolution hin. Insbesondere interessiert hier die Evolution der Stoffwechselwege, die kaum anders zugänglich sind.

Strukturähnlichkeit muß nicht unbedingt phylogenetisch bedingt sein. Sie kann auch der Ausdruck einer Strukturregel sein. So glaubt beispielsweise niemand, daß  $\alpha$ -Keratin<sup>[23]</sup> mit Tropomyosin<sup>[24]</sup> verwandt ist, nur weil beide die Doppel- $\alpha$ -Helixstruktur (Abb. 4) bilden. Dieses Sekundärstrukturaggregat ist eben energetisch besonders günstig und wird deshalb von mehreren Proteinen unabhängig voneinander verwendet. Die Koinzidenz der Strukturen ist nicht signifikant. Anders verhält es sich beim T4-Phagen-Lysozym<sup>[91]</sup>, Hühnereiweiß-Lysozym<sup>[61]</sup> und  $\alpha$ -Lactalbumin<sup>[92]</sup>. Hier sind die Strukturen kompliziert und trotzdem einander ähnlich. Ihre Kettenketten scheint aber in keiner Weise ausgezeichnet zu sein, denn es gibt auch viele andersartige, komplizierte Proteinstrukturen. Die Koinzidenz der Strukturen ist demnach ein signifikantes Ereignis, welches nur durch phylogenetische Beziehung erklärt werden kann. An diesen Beispielen ist zu sehen, daß bei Strukturähnlichkeit die Höhe der Signifikanz entscheidet, und daß diese nur mit Kenntnis der Strukturregeln ermittelt werden kann.

Diese Signifikanz wird außerordentlich hoch, wenn noch Gemeinsamkeiten in den Aminosäuresequenzen vorhanden sind. Beispielsweise sind 45 Aminosäuren im Hühnereiweiß-Lysozym und im  $\alpha$ -Lactalbumin gleich<sup>[92]</sup>. Die Wahrscheinlichkeit dafür, daß das zufällig ist, beträgt 1:20<sup>45</sup>; die Signifikanz ist 20<sup>45</sup>:1. Zwar muß man hier noch Abstriche machen, weil 84 weitere Aminosäuren verschieden sind, weil die Sequenzen zur besten Anpassung abschnittsweise gegeneinander verschoben wurden und weil die Aminosäuren nicht statistisch verteilt sind. Trotzdem verbleibt eine so hohe Signifikanz, daß eindeutig auf einen phylogenetischen Zusammenhang geschlossen werden kann.

Allein auf Raumstrukturmerkmalen basierende Signifikanzen wurden für eine Gruppe von Proteinen mit zentralen Faltblättern berechnet<sup>[93]</sup>, die keine Ähnlichkeit der Aminosäuresequenz mehr zeigen (Abb. 9). Hier wurde das Faltblatt als Referenzstruktur definiert und die Kettenfaltung darauf bezogen. So betrachtet kann ein n-strängiges Faltblatt  $M = 2^{n-2} \cdot n!$  verschiedene Kettenketten annehmen. Werden in zwei Proteinen gleiche Topologien gefunden, dann ist die Signifikanz dieser Übereinstimmung<sup>[93]</sup> also  $M : 1$  (Tabelle 5). Diese Zahlen werden allerdings durch die beobachtete Dominanz der Rechtsschraube bei Verbindungen zwischen Faltblattsträngen (Abb. 8) verkleinert<sup>[77]</sup>. Eine weitere Verminde rung ergibt sich durch die Nachbarschaftskorrelation in Faltblättern (Abb. 7), denn im vorliegenden Falle sind die meisten Stränge benachbart. Auch die eventuell durch Genduplikation verursachte Symmetrie im Faltblatt verringert die Signifikanz. Andererseits wird sie erhöht, wenn nicht nur die Topologie der Kette, sondern auch die Lage des aktiven Zentrums in die Zählung einbezogen wird (Tabelle 5). Außerdem muß beachtet werden, daß hier nur innerhalb einer Strukturklasse (Tabelle 4) verglichen wurde. Erweitert man den Vergleich auf alle Proteine, dann erhöht sich die Signifikanz. Einige dieser Beiträge sind schwer abzuschätzen. Insgesamt dürften die resultierenden Signifikanzwerte aber nur innerhalb der Gruppe der Dehydrogenasen so groß sein, daß auf phylogenetische Verwandtschaft geschlossen werden kann (Tabelle 5);

die anderen Strukturähnlichkeiten können auch rein physikalischen Ursprungs sein. Auf analoge Weise wurden die  $\beta$ -Faltblätter der Superoxid-Dismutase und der Immunoglobulin-Domänen topologisch verglichen<sup>[67]</sup>. Hier ergab sich unter Berücksichtigung der Nachbarschaftskorrelation eine Signifikanz von etwa 3000:1, was deutlich phylogenetische Verwandtschaft anzeigt.

Tabelle 5. Signifikanzen von Strukturähnlichkeiten beim topologischen Vergleich von Proteinen mit parallelen Faltblättern.

	Signifikanz unter Berück- sichtigung der Lage des aktiven Zentrums [93]	Signifikanz nach Reduktion auf- grund der bevor- zugten Händig- keit der $\beta$ - $\alpha$ - $\beta$ - Gruppen [77]
Adenylyl-Kinase/Subtilisin	2400:1	340:1
Adenylyl-Kinase/Flavodoxin	240:1	30:1
Adenylyl-Kinase/beliebige		
Dehydrogenase	160:1	20:1
Subtilisin/Flavodoxin	4800:1	350:1
Subtilisin/beliebige		
Dehydrogenase	2400:1	180:1
Flavodoxin/beliebige		
Dehydrogenase	4800:1	350:1
beliebige Dehydrogenase/andere		
Dehydrogenase	115200:1	4420:1

Für die obigen Rechnungen müssen Topologien definierbar sein, was die Existenz einer Substruktur wie beispielsweise eines Faltblatts voraussetzt. Die Vergleiche sind dann auf die Proteinklasse mit dieser Substruktur (Tabelle 4) beschränkt. Hingegen bietet der direkte geometrische Vergleich von Kettenketten einen wesentlichen allgemeineren Ansatz. Hier wird die Ähnlichkeit durch die mittlere Distanz der  $C_{\alpha}$ -Atome  $\langle \Delta C_{\alpha} \rangle$  oder nahe verwandte Indices ausgedrückt. Dieser Vergleich hat allerdings den Nachteil, daß Strukturregeln schwer zu berücksichtigen sind und daß keine Aussage über die Signifikanz gemacht wird, welche schließlich über phylogenetische Verwandtschaft entscheidet.

Im Prinzip kann eine Beziehung zwischen  $\langle \Delta C_{\alpha} \rangle$ -Wert, Strukturregeln und Signifikanz hergestellt werden, indem auf einem Rechner mit Hilfe eines Zufallszahlengenerators Kettenketten simuliert und miteinander verglichen werden. Dabei müßten alle bekannten Strukturregeln, Nachbarschaftskorrelationen etc. berücksichtigt werden. Die Kettenlänge sollte etwa der einer Domäne (Abb. 5) entsprechen. Beim Vergleich müßte zunächst diejenige relative Orientierung gefunden werden, die dem minimalen  $\langle \Delta C_{\alpha} \rangle$ -Wert entspricht. Man würde dann etwa die in Abbildung 10 skizzierte Häufigkeitsverteilung erhalten. Wenn wir uns nun für die Signifikanz der Ähnlichkeit zweier Proteinstrukturen mit beispielsweise  $\langle \Delta C_{\alpha} \rangle = 3 \text{ \AA}$  interessieren, so fragen wir im Grunde genommen danach, wieviele Strukturen  $N$  es gibt, deren mittlere  $\langle \Delta C_{\alpha} \rangle$ -Distanzen voneinander mehr als  $3 \text{ \AA}$  betragen, denn dieses  $N$  entspricht der Signifikanz. Um  $N$  zu ermitteln, müssen nun längst nicht alle Strukturen simuliert werden. Wenn man statistisch eine Anzahl  $G$  von Strukturen simuliert und miteinander vergleicht, dann erhält man  $Z$  Zufallskoinzidenzen mit  $\langle \Delta C_{\alpha} \rangle \leq 3 \text{ \AA}$  (Abb. 10), die wie bei der Zählung von Strahlungsquanten<sup>[94]</sup> dem Poissonschen Gesetz gehorchen, d. h. das Zufallskoinzidenzverhältnis  $Z/G$  ist gleich der mittleren Besetzung  $G/N$ . Damit ist die Gesamtzahl der mit  $\langle \Delta C_{\alpha} \rangle \geq 3 \text{ \AA}$  verschiedenen Strukturen, das ist die interessierende Signifikanz, aus  $N = G^2/Z$  bestimmbar. Aus einer Häufigkeitsverteilung der  $\langle \Delta C_{\alpha} \rangle$  (Abb.

10) kann demnach eine Beziehung zwischen  $\langle \Delta C_\alpha \rangle$  und der Signifikanz für beliebige  $\langle \Delta C_\alpha \rangle$  abgelesen und beim Vergleich von beobachteten Proteinstrukturen angewendet werden.

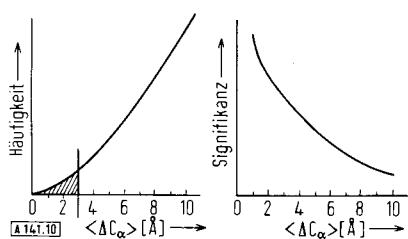


Abb. 10. Schematische Häufigkeitsverteilung von mittleren  $C_\alpha$ -Distanzen zwischen Paaren von simulierten Kettenketten. Die Signifikanz, der Indikator für phylogenetische Verwandtschaft, ergibt sich nach der Formel  $N = G^2/Z$ , wobei  $Z$  das Integral über die Häufigkeitsverteilung von Null zum betrachteten  $\langle \Delta C_\alpha \rangle$  ist (schraffiert).

## 9. Ausblick: Ermittlung von Proteinstrukturen ohne Kristalle

Üblicherweise müssen die Proteine für eine Strukturanalyse kristallisiert und anschließend röntgenographisch untersucht werden, doch ist schon heute eines der Fernziele – die Bestimmung der Raumstruktur allein aus der Sequenz – in einigen Spezialfällen erreicht. Beispielsweise können aus der bekannten Raumstruktur eines Proteins die Strukturen all seiner phylogenetisch Verwandten ermittelt werden, sobald deren Sequenz bekannt ist. Hierzu wird die Sequenz in die bekannte Kettenketten „eingebaut“. Verbleibende Differenzen werden durch anschließende Energieminimierung<sup>[95]</sup> gelindert. Da die Kettenketten während der Evolution außerordentlich gut konserviert worden ist, läßt sich diese Methode auch noch bei sehr entfernten Verwandten wie etwa Thrombin und Trypsin anwenden<sup>[88]</sup>. Im Falle der Muskelproteine Parvalbumin und Troponin C postuliert man nicht nur gleiche Kettenketten, sondern darüber hinaus auch interne Molekülsymmetrie. Hier deuten Sequenzähnlichkeiten darauf hin, daß Troponin C viermal die im Parvalbumin mit einer zweizähligen Achse symmetrisch aggregierte Substruktur (eine  $\text{Ca}^{2+}$ -Bindungsstelle und zwei benachbarte  $\alpha$ -Helices) enthält<sup>[89]</sup>. Für Troponin C liegt also die Annahme nahe, daß die beiden weiteren Substrukturen ebenfalls symmetrisch angefügt sind, also insgesamt die Symmetrie 222 ( $D_2$ ) ergeben. Damit kann ein Strukturmodell anhand der Sequenz gebaut werden<sup>[96]</sup>. Diese Beispiele demonstrieren erneut, wie hilfreich die immer deutlicher werdende phylogenetische Verflechtung der Proteine für deren Strukturanalyse ist.

Von der unmittelbaren Bestimmung einer Raumstruktur aus der Sequenz ist man heute zwar noch weit entfernt, jedoch existieren hierzu erste Ansätze: Beim Myoglobin und beim Pankreatischen Trypsin-Inhibitor wurde versucht, die bereits bekannten Raumstrukturen mit einfachen Regeln abzuleiten. In beiden Fällen wurde für die  $\alpha$ -Helices strenge Hierarchie vorausgesetzt, also angenommen, daß sie durch Vorhersagemethoden bestimmbar sind und daß sie sich zuerst falten. Die Faltungssimulation setzt dann bereits fertige  $\alpha$ -Helices voraus. Da Myoglobin praktisch nur Helices enthält, ist in diesem Fall die Faltung lediglich eine Aggregation dieser Helices; der Konformationsraum ist also extrem verkleinert. Hier wurde ein einfaches Berechnungsschema für die Packungsener-

gie der Helices postuliert und der Faltungsweg als direkte Falllinie zum Energieminimum ermittelt. Im Energieminimum ergab sich dann die native Struktur<sup>[97]</sup>. Diese Methode ist jedoch auf rein helicale Proteine beschränkt. Beim Pankreatischen Trypsin-Inhibitor hingegen gibt es nur eine kurze  $\alpha$ -Helix. Hier wurden die Geometrie der Peptidbindungen und die Seitenketten sowie die Bestimmung der Bindungsenergie vereinfacht. Mit einem iterativen Rechenverfahren wurde dann die Faltung zum Energieminimum simuliert. Dabei ergab sich im besten Falle eine mittlere  $C_\alpha$ -Distanz zur nativen Struktur von nur 6.3 Å<sup>[98]</sup>.

Diese Simulationsrechnungen sowie all die vorher beschriebenen Analysen von Proteinstrukturen geben uns die Hoffnung, daß in ferner Zukunft die Bestimmung der Raumstruktur aus der Sequenz möglich wird. Auf jeden Fall zeigen sie, wie gründlich der erste Schock über die Komplexität der Proteinstrukturen bereits überwunden ist. Allenthalben wird Ordnung in die Datenfülle gebracht. Man darf erwarten, daß diese Ordnung umso klarer und die Strukturregeln umso detaillierter werden, desto mehr Daten vorhanden sind, d. h. desto mehr röntgenographische Strukturanalysen gelingen.

Eingegangen am 29. Juni 1976 [A 141]

- [1] J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips u. V. C. Shore, *Nature* 185, 422 (1960).
- [2] R. E. Marsh u. J. Donohue, *Adv. Protein Chem.* 22, 235 (1967).
- [3] G. N. Ramachandran u. V. Sasisekharan, *Adv. Protein Chem.* 23, 283 (1968).
- [4] A. L. Lehninger: *Biochemie*. Verlag Chemie, Weinheim 1975.
- [5] C. B. Anfinsen, *Science* 181, 223 (1973).
- [6] D. B. Wetlauffer u. S. Ristow, *Annu. Rev. Biochem.* 42, 135 (1973).
- [7] C. B. Anfinsen u. H. A. Scheraga, *Adv. Protein Chem.* 29, 205 (1975).
- [8] W. Kauzmann, *Adv. Protein Chem.* 14, 1 (1959).
- [9] B. Furie, A. N. Schechter, D. H. Sachs u. C. B. Anfinsen, *J. Mol. Biol.* 92, 497 (1975).
- [10] F. M. Richards, *J. Mol. Biol.* 82, 1 (1974).
- [11] A. I. Kitaigorodsky: *Molecular Crystals and Molecules*. Academic Press, New York 1973.
- [12] J. F. Brandts, R. J. Oliveira u. C. Westort, *Biochemistry* 9, 1038 (1970).
- [13] d'Ans-Lax: *Taschenbuch für Chemiker und Physiker*. Bd. 1. Springer-Verlag, Berlin 1967.
- [14] B. Lee u. F. M. Richards, *J. Mol. Biol.* 55, 379 (1971).
- [15] C. Chothia, *Nature* 248, 338 (1974).
- [16] S. L. Miller, *J. Am. Chem. Soc.* 77, 2351 (1955).
- [17] T. E. Creighton, *J. Mol. Biol.* 95, 167 (1975).
- [18] R. Huber, D. Kukla, A. Rühlmann, O. Epp u. H. Formanek, *Naturwissenschaften* 57, 389 (1970).
- [19] D. Wetlauffer, E. Kwolek, W. L. Anderson u. E. R. Johnson, *Biochem. Biophys. Res. Commun.* 56, 380 (1974).
- [20] F. H. C. Crick, *Acta Crystallogr.* 6, 689 (1953).
- [21] C. Cohen u. K. C. Holmes, *J. Mol. Biol.* 6, 423 (1963).
- [22] H. E. Huxley, *Science* 164, 1356 (1969).
- [23] R. D. B. Fraser u. T. P. MacRae, *Nature* 233, 138 (1971).
- [24] C. Cohen, *Sci. Am.* 234, 36 (Nov. 1975).
- [25] W. A. Hendrickson u. K. B. Ward, *Biochem. Biophys. Res. Commun.* 66, 1349 (1975).
- [26] R. Henderson u. P. N. T. Unwin, *Nature* 257, 28 (1975).
- [27] S. T. Rao u. M. G. Rossmann, *J. Mol. Biol.* 76, 241 (1973).
- [28] E. Hill, D. Tsernouglou, L. Webb u. L. J. Banaszak, *J. Mol. Biol.* 72, 577 (1972).
- [29] I. Ohlsson, B. Nordström u. C. I. Brändén, *J. Mol. Biol.* 89, 339 (1974).
- [30] M. Buehner, G. C. Ford, D. Moras, K. W. Olsen u. M. G. Rossmann, *Proc. Natl. Acad. Sci. USA* 70, 3052 (1973).
- [31] G. E. Schulz, M. Elzinga, F. Marx u. R. H. Schirmer, *Nature* 250, 120 (1974).
- [32] K. D. Waterpau, L. C. Sieker, L. H. Jensen, J. Legall u. M. Dubourdieu, *Proc. Natl. Acad. Sci. USA* 69, 3185 (1972).
- [33] R. D. Andersen, P. A. Apgar, R. M. Burnett, G. D. Darling, M. E. Lequesne, S. G. Mayhew u. M. L. Ludwig, *Proc. Natl. Acad. Sci. USA* 69, 3189 (1972).
- [34] C. Schubert-Wright, R. A. Alden u. J. Kraut, *Nature* 221, 235 (1969).
- [35] C. C. F. Blake u. P. R. Evans, *J. Mol. Biol.* 84, 585 (1974).
- [36] R. J. Fletterick, persönliche Mitteilung (1976).
- [37] P. A. Peterson, B. A. Cunningham, I. Berggård u. G. M. Edelman, *Proc. Natl. Acad. Sci. USA* 69, 1697 (1972).

- [38] G. E. Schulz, H. Zappe, D. J. Worthington u. M. A. Rosemeyer, FEBS Lett. 54, 86 (1975).
- [39] M. J. Adams, G. C. Ford, R. Koekoek, P. J. Lentz, A. McPherson, M. G. Rossmann, I. E. Smiley, R. W. Schevitz u. A. J. Wonacott, Nature 227, 1098 (1970).
- [40] S. G. Warren, B. F. P. Edwards, D. R. Evans, D. C. Wiley u. W. N. Lipscomb, Proc. Natl. Acad. Sci. USA 70, 117 (1973).
- [41] M. Grüter u. R. M. Franklin, J. Mol. Biol. 89, 163 (1974).
- [42] G. W. Tischendorf, H. Zeichardt u. G. Stöffler, Proc. Natl. Acad. Sci. USA 72, 4820 (1975).
- [43] L. Pauling u. R. B. Corey, Proc. Natl. Acad. Sci. USA 37, 729 (1951).
- [44] A. V. Guzzo, Biophys. J. 5, 809 (1965).
- [45] J. W. Prothero, Biophys. J. 6, 367 (1966).
- [46] R. Leberman, J. Mol. Biol. 5, 23 (1971).
- [47] P. N. Lewis, F. A. Momany u. H. A. Scheraga, Proc. Natl. Acad. Sci. USA 68, 2293 (1971).
- [48] M. Schiffer u. A. B. Edmundson, Biophys. J. 8, 29 (1968).
- [49] P. Y. Chou u. G. D. Fasman, Biochemistry 13, 222 (1974).
- [50] V. I. Lim, J. Mol. Biol. 88, 873 (1974).
- [51] B. Robson u. R. H. Pain, Biochem. J. 141, 883 (1974).
- [52] K. Nagano, J. Mol. Biol. 84, 337 (1974).
- [53] E. A. Kabat u. T. T. Wu, Proc. Natl. Acad. Sci. USA 70, 1473 (1973).
- [54] A. V. Finkelstein u. O. B. Ptitsyn, J. Mol. Biol. 62, 613 (1971).
- [55] P. N. Lewis, N. Gö, M. Gö, D. Kotelchuck u. H. A. Scheraga, Proc. Natl. Acad. Sci. USA 65, 810 (1970).
- [56] G. E. Schulz, C. D. Barry, J. Friedman, P. Y. Chou, G. D. Fasman, A. V. Finkelstein, V. I. Lim, O. B. Ptitsyn, E. A. Kabat, T. T. Wu, M. Levitt, B. Robson u. K. Nagano, Nature 250, 140 (1974).
- [57] E. A. Kabat u. T. T. Wu, Proc. Natl. Acad. Sci. USA 71, 4217 (1974).
- [58] B. W. Matthews, Biochim. Biophys. Acta 405, 442 (1975).
- [59] I. D. Kuntz, J. Am. Chem. Soc. 94, 4009 (1972).
- [60] H. W. Wyckoff, D. Tsernouglou, A. W. Hanson, J. R. Knox, B. Lee u. F. M. Richards, J. Biol. Chem. 245, 305 (1970).
- [61] C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips u. V. R. Sarma, Nature 206, 757 (1965).
- [62] E. T. Adman, L. C. Sieker u. L. H. Jensen, J. Biol. Chem. 248, 3987 (1973).
- [63] P. M. Colman, J. Deisenhofer, R. Huber u. W. Palm, J. Mol. Biol. 100, 257 (1976).
- [64] R. J. Poljak, L. M. Amzel, B. L. Chen, R. P. Phizackerley u. F. Saul, Proc. Natl. Acad. Sci. USA 71, 3440 (1974).
- [65] D. M. Segal, E. A. Padlan, G. H. Cohen, S. Rudikoff, M. Potter u. D. R. Davies, Proc. Natl. Acad. Sci. USA 71, 4298 (1974).
- [66] A. B. Edmundson, K. R. Ely, E. E. Abola, M. Schiffer u. N. Panagiotopoulos, Biochemistry 14, 3953 (1975).
- [67] J. S. Richardson, D. C. Richardson, K. A. Thomas, E. W. Silverton u. D. R. Davies, J. Mol. Biol. 102, 221 (1976).
- [68] K. D. Hardman u. C. F. Ainsworth, Biochemistry 11, 4910 (1972).
- [69] C. C. F. Blake, M. F. Geisow, I. D. A. Swan, C. Rerat u. B. Rerat, J. Mol. Biol. 88, 1 (1974).
- [70] J. J. Birktoft u. D. M. Blow, J. Mol. Biol. 68, 187 (1972).
- [71] R. E. Fenna u. B. W. Matthews, Nature 258, 573 (1975).
- [72] D. W. Banner, A. C. Bloomer, G. A. Petsko, D. C. Phillips, C. I. Pogson, I. A. Wilson, P. H. Corran, A. J. Furth, J. D. Milman, R. E. Offord, J. D. Priddle u. S. G. Waley, Nature 255, 609 (1975).
- [73] C. C. F. Blake, Essays Biochem. 11, 37 (1975).
- [74] F. A. Quirocho u. W. N. Lipscomb, Adv. Protein Chem. 25, 1 (1971).
- [75] M. F. Perutz, H. Muirhead, J. M. Cox u. L. C. G. Goaman, Nature 219, 131 (1968).
- [76] C. Chothia, J. Mol. Biol. 75, 295 (1973).
- [77] M. J. E. Sternberg u. J. M. Thornton, J. Mol. Biol. 105, 367 (1976).
- [78] J. Bergsma, W. G. J. Hol, J. N. Jansonius, K. H. Kalk, J. H. Ploegman u. J. D. G. Smit, J. Mol. Biol. 98, 637 (1975).
- [79] M. G. Rossmann, D. Moras u. K. W. Olsen, Nature 250, 194 (1974).
- [80] A. D. McLachlan, Nature New Biol. 240, 83 (1972).
- [81] R. H. Kretsinger, Nature New Biol. 240, 85 (1972).
- [82] O. B. Ptitsyn, J. Mol. Biol. 88, 287 (1974).
- [83] M. F. Perutz, J. Mol. Biol. 13, 646 (1965).
- [84] T. Takano, O. B. Kallai, R. Swanson u. R. E. Dickerson, J. Biol. Chem. 248, 5234 (1973).
- [85] R. M. Stroud, Sci. Am. 233, 74 (Juli 1974).
- [86] G. M. Edelman, Science 180, 830 (1973).
- [87] M. O. Dayhoff: Atlas of Protein Structure. Natl. Biomed. Res. Foundation, Washington 1972.
- [88] B. S. Hartley, Philos. Trans. R. Soc. London Ser. B: 257, 77 (1970).
- [89] J. H. Collins, J. D. Potter, M. J. Horn, G. Wilshire u. N. Jackman, FEBS Lett. 36, 268 (1973).
- [90] A. D. McLachlan, M. Stewart u. L. B. Smillie, J. Mol. Biol. 98, 281 (1975).
- [91] B. W. Matthews u. S. J. Remington, Proc. Natl. Acad. Sci. USA 71, 4178 (1974).
- [92] W. J. Browne, A. C. T. North, D. C. Phillips, K. Brew, T. C. Vanaman u. R. L. Hill, J. Mol. Biol. 42, 65 (1969).
- [93] G. E. Schulz u. R. H. Schirmer, Nature 250, 142 (1974).
- [94] G. Friedlander u. J. W. Kennedy: Introduction to Radiochemistry. Wiley, New York 1949.
- [95] M. Levitt, J. Mol. Biol. 82, 393 (1974).
- [96] R. H. Kretsinger u. C. D. Barry, Biochim. Biophys. Acta 405, 40 (1975).
- [97] O. B. Ptitsyn u. A. A. Rashin, Biophys. Chem. 3, 1 (1975).
- [98] M. Levitt u. A. Warshel, Nature 253, 694 (1975).

## Experimentelle Elektronendichten und chemische Bindung

Von Philip Coppens<sup>[\*]</sup>

In diesem Fortschrittsbericht werden einige neue Ergebnisse der Analyse von Ladungsdichten durch Röntgen- und Neutronenbeugung diskutiert. Zu den Problemen, die auf solche Weise in mehreren Laboratorien untersucht werden, gehören die Natur von Einfach-, Doppel- und Dreifachbindungen, die Hybridisierung einsamer Elektronenpaare, die Bindungseigenschaften von Metallen, Metallegierungen und Organometallverbindungen sowie die Ableitung physikalischer Eigenschaften aus Röntgenbeugungsdichten. Beim jetzigen Entwicklungsstand der Methodik sind bereits Untersuchungen an Serien von verwandten Verbindungen möglich, was zu einer breiteren Anwendung führen sollte.

Offensichtlich ist ein Molekül viel komplizierter als ein Atom. Es stellt sich daher die Frage: Was geschieht mit einem Atom, wenn es Teil eines Moleküls wird? [C. A. Coulson, Mem. Soc. R. Sci. Liege, Collect. 8, 6 (2), 143 (1971)].

[\*] Prof. Dr. P. Coppens  
Chemistry Department  
State University of New York at Buffalo  
Buffalo, New York 14214 (USA)

### 1. Einführung

Einen Teil der Antwort auf diese trügerisch einfache Frage kann man durch die kristallographische Auswertung der Röntgen- und Neutronenbeugung bekommen. Da die Röntgenstreuung primär ein Produkt der Wechselwirkung zwischen Röntgenstrahlen und Elektronen ist, beinhaltet sie Information über die Elektronenverteilung in dem streuenden Stoff. Schon