

# Protein-Differenzierung: Entwicklung neuartiger Proteine im Laufe der Evolution

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

Die fast unübersehbare Fülle der Proteine läßt sich auf etwa 200 Grundstrukturen, d. h. Geometrien der Polypeptidkette, zurückführen; das ist die Basis der verfolgbaren Protein-Differenzierung. Diese Differenzierung beginnt mit einer Gen-Duplikation. Das eine Gen-Produkt muß die ursprüngliche Funktion weiterhin erfüllen und wird deshalb konserviert. Das andere Gen-Produkt wird konserviert, sobald es eine neuartige Funktion übernommen hat. Der Vergleich von Aminosäuresequenzen äquivalenter Proteine zweier Spezies ergibt den Entwicklungsabstand zwischen diesen Spezies. Daraus kann der Zeitpunkt abgeleitet werden, zu dem die Verzweigung stattgefunden hat („molekulare Paläontologie“). Diese Vergleiche werden auf Molekülgeometrien nicht äquivalenter Proteine ausgedehnt, wodurch die Entwicklungswege der Protein-Differenzierung aufgeklärt werden können.

## 1. Einleitung

Auf der Erde werden zur Zeit etwa  $10^{11}$  verschiedene Proteine produziert. Sieht man von allen Unterschieden ab, welche sich zwischen Proteinen mit gleicher Funktion aufgrund der separaten Evolution der jeweiligen Arten gebildet haben, dann sinkt diese Zahl auf etwa  $10^5$ . Das sind immer noch unüberschbar viele. Vernachläßigt man darüber hinaus die Funktionen der Proteine und konzentriert sich auf ihre Grundstrukturen, das sind die Geometrien der Polypeptidketten, so endet man bei nur etwa  $10^2$  verschiedenen Proteinen. Das korrespondiert ungefähr mit der Anzahl von Bildern, die man beim Besuch einer Ausstellung betrachtet. Durch die hier vorgeschlagenen Vereinfachungen wird die Proteinwelt mithin so überschaubar wie eine Gemäldegalerie.

Die beiden Reduktionen von  $10^{11}$  zu  $10^5$  und von  $10^5$  zu  $10^2$  entsprechen historischen biologischen Prozessen, die umgekehrt verlaufen sind: von geringerer zu größerer Vielfalt. Der erste Reduktionsschritt entspricht der wohlbekannten Evolution der Arten, während der zweite Schritt die Evolution der Proteine – die Protein-Differenzierung – widerspiegelt. Die Bezeichnung Protein-Differenzierung wurde gewählt<sup>[1]</sup>, weil dieser Prozeß dem Differenzierungsprozeß von Gewebezellen ähnelt: Zwar im Zeitmaß der Evolution, aber ansonsten wie Zellen, werden einzelne Proteine dupliziert. Die Duplikate entwickeln sich dann separat, um neuartige Aufgaben zu erfüllen.

Will man Proteine voneinander unterscheiden, so müssen sie zunächst charakterisiert werden. Hierzu reichen globale Merkmale wie Molekulargewicht, Sedimentationsgeschwindigkeit, Aminosäurenzusammensetzung usw. nicht aus. Für die folgende Diskussion wird eine exakte strukturelle Beschreibung benötigt: die Kenntnis der genauen Reihenfolge aller Aminosäurereste in der Polypeptidkette – die „Sequenz des Proteins“ – oder die Kenntnis des Kettenverlaufs im Raum – die „Geometrie des Proteins“. Am besten hat man natürlich beide Informationen; damit kennt man die räumliche Anordnung aller Atome – die „Struktur des Proteins“<sup>[2]</sup>. Heute sind etwa  $10^3$  Sequenzen und  $10^2$  Geometrien und Strukturen bekannt<sup>[3,4]</sup>. Auf dieser Datenbasis konnten nicht

nur die Wege, auf denen sich Proteine artspezifisch verändern, sondern darüber hinaus auch die Entwicklungswege der Protein-Differenzierung enthüllt werden.

## 2. Artspezifische Veränderungen

Sehr bald nach den ersten gelungenen Analysen wurde klar, daß die Sequenzen homologer Proteine in unterschiedlichen Arten nicht in allen Positionen übereinstimmen. So weist das antibakterielle Protein Lysozym in Mensch und Pavian Unterschiede in 14 von 131 Positionen auf. Im Lauf der separaten Evolution von Mensch und Pavian haben also ungefähr je sieben Aminosäureaustausche stattgefunden. Je weiter der gemeinsame Vorfahre zurückliegt, desto mehr Austausche sind zu erwarten; die Lysozyme von Mensch und Ente unterscheiden sich beispielsweise in 53 Aminosäureresten.

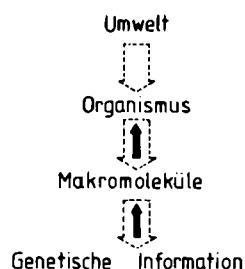


Abb. 1. Einbettung der Makromoleküle im biologischen System. Die genetische Information in den Desoxyribonucleinsäure-Molekülen (DNA) definiert exakt die Struktur der Makromoleküle; das sind zumeist Proteine sowie einige Ribonucleinsäure-Moleküle (RNA). Die Makromoleküle ihrerseits definieren exakt den entsprechenden Organismus. Im Zeitmaß der Evolution kehrt sich die Determination allerdings durch Mutation und Selektion um. Letztlich bestimmen die ökologischen Bedingungen die in einem Organismus vorhandene genetische Information.

Um diese artspezifischen Veränderungen zu verstehen, schaut man sich am besten die Einbettung der Makromoleküle im biologischen System an (Abb. 1). Die in der DNA als Nucleotidsequenz niedergelegte genetische Information bestimmt die Aminosäuresequenz und über den spontanen Faltungsprozeß der Polypeptidkette auch die Struktur der Proteine. Entsprechendes gilt für RNA-Moleküle, wie etwa ribosomale RNA. Ein Organismus wird schließlich durch die Gesamtheit seiner Makromoleküle definiert. Jede Verände-

[\*] Prof. Dr. G. E. Schulz

Max-Planck-Institut für Medizinische Forschung  
Jahnstraße 29, D-6900 Heidelberg 1

rung beginnt mit einer Mutation der genetischen Information, die sich dann in den Makromolekülen und damit im Organismus ausdrückt.

Die Häufigkeit spontaner Mutationen im Menschen ist aus Reihenuntersuchungen am Hämoglobin zu etwa  $10^{-3}$  pro Protein und Individuum bestimmt worden<sup>[5]</sup>. Würden alle diese Mutationen akzeptiert und tradiert, dann sollten die etwa  $10^5$  Generationen, die den Menschen vom Gorilla trennen, zu völlig verschiedenen Makromolekülen geführt haben. Dennoch sind die  $\alpha$ - und  $\beta$ -Ketten der Hämoglobine von Gorilla und Mensch nur in je einer Position verschieden. So geringe Akzeptierung bei so zahlreich angebotenen spontanen Mutationen zeigt, daß auf den Niveaus der Makromoleküle, der Organismen und der Umwelt aus praktisch beliebigen Veränderungen in der DNA selektioniert werden kann. Im Zeitmaß der Evolution geht der Informationsfluß demnach in umgekehrter Richtung (in Abb. 1 von oben nach unten).

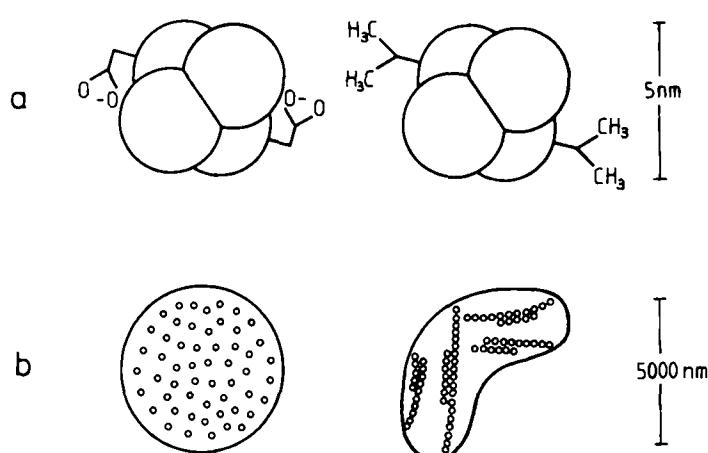


Abb. 2. Sichelzellen des Hämoglobins. a) Normales Hämoglobin und Sichelzellhämoglobin unterscheiden sich durch den Austausch von Glutaminsäure gegen Valin in Position 6 der  $\beta$ -Kette. b) Im Erythrocyten liegt normales Hämoglobin trotz der sehr hohen Konzentration von  $300 \text{ mg/cm}^3$  monodispers vor, während Sichelzellhämoglobin zu Fasern aggregiert.

Ein im Detail geklärtes Beispiel für die Akzeptierung einer Mutation ist das Sichelzellan der  $\beta$ -Kette des Hämoglobins, das in Malariagebieten mit Häufigkeiten bis zu 40% auftritt<sup>[6]</sup>. Wie in Abbildung 2 dargestellt, unterscheiden sich Sichelzellan und normales Gen durch den Austausch einer Glutaminsäure gegen ein Valin an der Proteinoberfläche. Dadurch wird eine lineare Aggregation der Hämoglobinmoleküle energetisch begünstigt. Die entstehenden Fasern strecken die Erythrozyten zur Sichelform, und der Bruch dieser Sicheln in den Kapillaren führt schließlich zu anämischen Zuständen. Trotz der ungünstigen Folgen wurde die Mutation akzeptiert, weil die Faserbildung das Eindringen von Plasmodien in Erythrozyten erschwert und damit den Trägern des Sichelzellan einen gewissen Schutz gegen Malaria verleiht. Die Evolution folgt in diesem Falle einer Route zwischen Scylla und Charybdis - einerseits Tod durch Malaria und andererseits Tod durch Anämie.

Hier hat also die Umwelt in Gestalt einer Bedrohung durch Parasiten die Annahme einer Mutation erzwungen, die den Organismus selbst schädigt (Abb. 1). Auf dem Niveau der Makromoleküle ist der Austausch neutral, er beeinträchtigt keineswegs das Einzelmolekül in der Erfüllung sei-

ner Aufgabe, dem O<sub>2</sub>-Transport. Das zeigt, wie sehr die heute vorhandenen Proteinveränderungen von Details in der Umwelt vergangener Zeiten abhängig sind. Deshalb ist unsere Chance außerordentlich gering, diese Änderungen jemals genau zu verstehen. Möglich ist aber der Versuch, durch statistische Analysen generelle Aspekte und Regeln herauszuarbeiten.

Werden beispielsweise alle Aminosäureaustausche als gleichwertig angesehen, so ergibt sich ein sehr einfaches Maß für den Unterschied zwischen zwei Sequenzen: die Anzahl der Austausche oder in normierter Form - die Anzahl der Austausche pro Position. Nimmt man darüber hinaus an, daß die Austauschhäufigkeit während der Evolution konstant ist, dann ist die Anzahl der Austausche ein Maß für den Entwicklungsabstand und somit auch ein Maß für den Zeitpunkt, an dem die Verzweigung bei einem gemeinsamen Vorfahren stattfand. Damit wird eine „molekulare Paläontologie“ möglich. Kennt man für ein Protein mit gegebener Funktion - wie etwa Lysozym - die in N Spezies auftretenden Sequenzen, so ergeben sich N (N - 1)/2 Entwicklungsabstände zwischen diesen Spezies. Dieses Datenmaterial genügt zur Aufstellung eines Stammbaumes<sup>[3]</sup>. Die molekulare Paläontologie reicht in wesentlich frühere Zeiten zurück als ihr konventioneller Gegenpart. Anhand der Sequenzen Cytochrome-c-ähnlicher Proteine lassen sich z. B. Bakterien phylogenetisch miteinander verknüpfen, bei denen die Verzweigungen vor mehr als  $3 \cdot 10^9$  Jahren stattfanden<sup>[7]</sup>. Da  $10^5$  Proteine gegebener Funktion existieren, können letztlich  $10^5$  Stammbäume aufgestellt und zusammengefaßt werden. Der molekularen Paläontologie kommt also hohe Genauigkeit zu.

In vergleichenden Sequenzanalysen zeigte sich, daß die Veränderungsgeschwindigkeit stark von der Einbettung des jeweiligen Proteins in seine Umgebung abhängt. Beispielsweise bilden die tief im Chromatin verborgenen Histone eine Unzahl von Kontakt mit Nachbarmolekülen. Als Bestandteil des Chromatins sind Histone weiterhin wesentlich für so zentrale Aufgaben wie Mitosis und Meiosis. Entsprechend verändern sie sich nur mit 0.1 Austauschen pro Position und  $10^{10}$  Jahren. Demgegenüber tauschen die „leichtlebigen“ Fibrinopeptide, deren Aufgabe im wesentlichen die Bedeckung der Fibrinvernetzungsfläche zur Verhinderung spontaner Blutgerinnung ist, ihre Aminosäuren tausendmal schneller aus. Fibrinopeptide von Spezies, die sich vor nur zehn Millionen Jahren getrennt haben, zeigen heute schon etwa 20% Sequenzunterschiede.

Diese Regel gilt keineswegs nur für Strukturproteine wie Histone und Fibrin, sondern auch für Enzyme. Hier sind weniger die Kontakte zu anderen Makromolekülen als vielmehr die enzymatisch katalysierten Reaktionen wichtig. Um diese aufrechtzuerhalten, sollten sich die Aminosäurereste im katalytischen Zentrum nur zögernd ändern. Genau dieser Effekt wird in allen bekannten Enzymstrukturen beobachtet: Das katalytische Zentrum wird streng konserviert. In einigen Fällen wurde umgekehrt an der Konservierung erst das katalytische Zentrum erkannt. In beiden Regeln - Erhaltung von Kontakten und Aufrechterhaltung von Reaktionen spiegelt sich der selektionsierende Einfluß des Organismus (Abb. 1) auf die Veränderung der Makromoleküle wider.

Stabilitätsvermindernde Aminosäureaustausche zerstören die Proteine. Solche Austausche werden bereits auf dem Niveau der Makromoleküle selbst diskriminiert (Abb. 1). Auch

das schlägt sich in Regeln nieder: Vergleicht man Mutationshäufigkeiten in Kettenpositionen an der Proteinoberfläche mit denen im Inneren, so zeigt sich deutlich, daß die Oberfläche wesentlich schneller verändert wird. Eine Erklärung dafür liegt auf der Hand. Proteine sind so dicht gepackt wie Kristalle kleiner Moleküle, d. h. die Kontakte zwischen den Kettenteilen sind optimiert<sup>[4]</sup>. Austausche im Proteininneren stören diese Packung, verringern damit die Stabilität und werden deshalb kaum akzeptiert.

Bisher haben wir lediglich Austausche von Aminosäureresten diskutiert, d. h. Veränderungen der Seitenketten unter Beibehaltung des Hauptstranges der Polypeptidkette. Die Evolution macht jedoch nicht vor der Hauptkette halt. Nur kommen Hauptkettenänderungen, das sind Einfügungen und Auslassungen von Aminosäureresten, wesentlich seltener vor. Sie tauchen erst auf, wenn schon etwa die Hälfte der Seitenketten verändert worden ist. Bei 30% und 15% Sequenzidentität steigt der Anteil von Einfügungen und Auslassungen auf 10% bzw. 30%. Damit sind Hauptkettenänderungen mehr als zehnmal seltener als Seitenkettenänderungen.

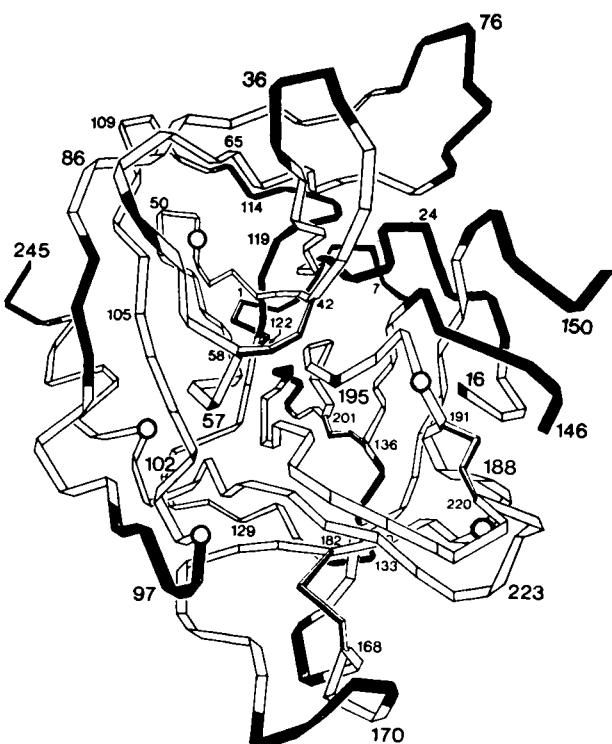


Abb. 3. Überlagerung der Polypeptidketten von Chymotrypsin und Protease-B von *Streptomyces griseus*. Dargestellt ist die Kettenfaltung und Numerierung von Chymotrypsin [8]. Schwarze Partien fehlen in Protease-B. Kreise markieren Einfügungen in Protease-B. Die Sequenzen sind in lediglich 12% ihrer Positionen gleich und konnten nur anhand der Geometrien überlagert werden [9]. Die Topologien der Kettenteile 25 bis 115 und 135 bis 225 stimmen überein.

Die beiden in Abbildung 3 dargestellten Proteine weisen 38% Hauptkettenveränderungen auf. Bei so zahlreichen Einfügungen und Auslassungen können die Sequenzen nicht mehr miteinander in Beziehung gesetzt werden. Nur die Aufklärung der Strukturen ermöglicht die Lokalisierung der Hauptkettenunterschiede und damit die zum Abzählen der Aminosäureaustausche notwendige Kenntnis der relativen Anordnung der Sequenzen. Ein Blick auf Abbildung 3 zeigt uns darüber hinaus, daß die Hauptkettenveränderungen nur

an der Oberfläche stattfinden. Im Proteininneren würden sie die Stabilität noch stärker untergraben als Aminosäureaustausche, hätten also noch geringere Chancen, akzeptiert zu werden.

Da Hauptkettenveränderungen gewöhnlich Aufweitungen und Verkürzungen von Kettenschleifen an der Oberfläche sind, lassen sie den generellen Verlauf der Polypeptidkette – auch „Kettentopologie“ genannt – unverändert. Die Kettentopologie ist die am besten konservierte Eigenschaft eines Proteins. Selbst nachdem alle Sequenzidentitäten verschwunden sind, ist sie noch vorhanden. Bisher ist kein Beispiel bekannt, in welchem artspezifische Veränderungen zur Modifikation einer Kettentopologie geführt haben.

### 3. Funktionelle Veränderungen

Bei den Proteinstrukturanalysen wurden nicht nur Ähnlichkeiten zwischen Proteinen gleicher Funktion in verschiedenen Spezies, sondern auch Ähnlichkeiten zwischen Proteinen unterschiedlicher Funktion im gleichen Organismus entdeckt. Als Beispiel vergleichen wir in Abbildung 4 die Sequenz des Verdauungsenzyms Elastase mit der Sequenz des Plasmaenzymes Thrombin. Die Polypeptidketten stimmen in 30% der Aminosäurereste überein. Damit erhebt sich natürlich die Frage, ob diese beiden im gleichen Organismus parallel produzierten Enzyme eventuell einen gemeinsamen Vorfahren haben, d. h. ob es eine eigenständige Evolution der Proteine gibt, die über die Proteinveränderungen während der Evolution der Arten hinausgeht.

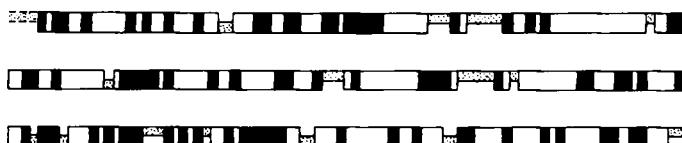


Abb. 4. Vergleich der Aminosäuresequenzen von Elastase [10] und Thrombin [11]. In den schwarzen bzw. weißen Bereichen sind die Aminosäurereste gleich bzw. verschieden. Oben bzw. unten angeschlagene gepunktete Bereiche sind Einfügungen in Thrombin bzw. Elastase. Die skizzierte Kette besteht aus 266 Resten, von denen 79 gleich und 187 verschieden sind. Unter den 187 verschiedenen Resten sind 33 Einfügungen oder Auslassungen. Daraus ergeben sich  $79/266 = 30\%$  Sequenzidentität und  $33/266 = 12\%$  Hauptkettenveränderungen.

Zur Beantwortung dieser Frage schätzen wir die Wahrscheinlichkeit ab, mit der eine solche Sequenzübereinstimmung ein Zufall sein kann. Unter der Annahme, daß alle Aminosäureaustausche gleich häufig auftreten und daß die Einfügungen und Auslassungen sicher lokalisiert sind (Abb. 4), ergibt sich eine Zufalls wahrscheinlichkeit von

$$\left(\frac{1}{20}\right)^{79} \cdot \left(\frac{19}{20}\right)^{187} \cdot \frac{266!}{79! \cdot 187!} \approx 10^{-38}$$

(Zahlen siehe Legende zu Abb. 4; 20 steht für die 20 Standard-Aminosäuren.)

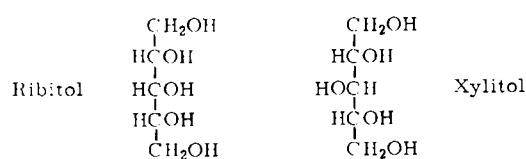
Ohne die vereinfachenden Annahmen erhöht sich dieser Wert zwar um einige Größenordnungen; jedoch bleibt die Wahrscheinlichkeit immer noch klein genug, um einen Zufall praktisch auszuschließen. Folglich gibt es einen tieferliegenden Grund für die entdeckte Übereinstimmung: Entweder ist die zu stabilen Proteinen führende Anzahl von Aminosäuresequenzen eng beschränkt und die obige Rechnung

völlig falsch, oder die beiden Proteine sind durch eine historische Entwicklung miteinander verknüpft.

Da zahlreiche völlig verschiedene Sequenzen bekannt sind, wird die Rechnung bestätigt. Somit kann auf eine historische Verbindung geschlossen werden. Beide Proteine haben einen gemeinsamen Vorfahren, oder umgekehrt ausgedrückt: Nach der Verzweigung haben sich die beiden Proteine separat weiterentwickelt und unterschiedliche Funktionen im Organismus übernommen. Auf dem Niveau der Zellen wird solche Entwicklung „Differenzierung“ genannt. Diesen Begriff auf das Niveau der Makromoleküle übertragend<sup>(1)</sup> sprechen wir von „Protein-Differenzierung“.

Für einen so grundlegenden Prozeß wünscht man sich natürlich Beweise, die etwas unmittelbarer sind als approximativ berechnete Wahrscheinlichkeiten. In der Tat kann die Veränderung von Proteinfunktionen direkt verfolgt werden. Den Medizinern ist geläufig, daß Parasiten allmählich gegen ein ursprünglich wirksames Medikament resistent werden und daß diese Resistenz vererbbar ist. Höchstwahrscheinlich haben hier eines oder mehrere Proteine ihre Funktionen durch Mutationen so abgewandelt, daß der Parasit nun das Medikament abwehren kann.

Noch deutlicher fielen Evolutionsversuche an *Klebsiella aerogenes* aus<sup>[12]</sup>. Dieses Bakterium akzeptiert Ribitol als Energie- und Kohlenstoffspender, wogegen das epimere Xylitol kaum verwendet werden kann. Durch ausschließliches



Angebot von Xylitol wurde nun versucht, die Spezifität des Enzyms Ribitol-Dehydrogenase zu verändern, das dem Bakterium sowohl Ribitol als auch Xylitol zur Weiterverarbeitung vorberichtet, letzteres allerdings hundertmal langsamer. Unter diesem Druck traten nach einiger Zeit Gen-Multiplikationen der Ribitol-Dehydrogenase auf, wodurch das Enzym in größeren Mengen produziert wurde – eine kostspielige Methode, um eine geringere Enzymaktivität auszugleichen. Später erschienen jedoch auch Mutationen im Gen und damit Aminosäureaustausche im Enzym, die dessen Aktivität gegenüber Xylitol erhöhten. Diese Experimente zeigen erstens, daß Gen-Multiplikationen stattfinden können, und zweitens, daß eine Änderung der enzymatischen Spezifität und damit der Funktion des Enzyms möglich ist.

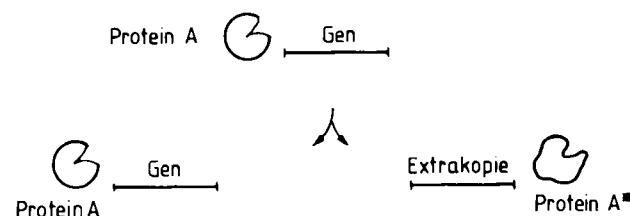


Abb. 5. Schema einer Gen-Duplikation. Nach der Verdoppelung des Gens muß eines der Produkte die ursprüngliche Funktion weiterhin erfüllen und wird deshalb konserviert. Das andere Produkt ist relativ frei in seinen Veränderungen. Es wird jedoch fixiert, sobald es eine neuartige Funktion übernommen hat. In einigen Fällen dient die Gen-Duplikation nur zur Produktionserhöhung; dann werden alle Exemplare konserviert. Bei den Tandem-Gen-Duplikationen (siehe Tabelle 1) bilden beide Produkte eine gemeinsame Kette.

Die wesentlichen Schritte der Protein-Differenzierung konnten also direkt beobachtet werden. Wie Abbildung 5 zeigt, muß für diesen Prozeß zunächst ein Gen dupliziert oder multipliziert werden. In der weiteren Entwicklung bleibt zumindest ein Gen-Exemplar für die ursprüngliche Aufgabe reserviert. Die zusätzlichen Gen-Exemplare befinden sich zunächst auf einer „Spielwiese“. Weil ihr Gen-Produkt nicht benötigt wird, erfahren sie kaum Druck von „oben“ (Abb. 1). Sie können sich solange frei verändern, bis ihr Produkt fähig ist, neuartige Aufgaben zu erfüllen. Mit Übernahme der neuen Funktion durch das Gen-Produkt ist die Freizügigkeit für dieses Gen allerdings beendet.

Auf diese Weise entstehen Isoenzyme; das sind Enzyme, die in einem Organismus parallel vorkommen und deren funktionelle und strukturelle Unterschiede noch relativ gering sind. Sie befinden sich in frühen Stadien der Protein-Differenzierung. Bekannt sind die Skelett- und Herzmuskel-varianten der Lactat-Dehydrogenase, bei denen noch drei Viertel aller Aminosäuren übereinstimmen<sup>[13, 14]</sup>. Dagegen sind die Unterschiede zwischen den drei Isoenzymen der Adenylyl-Kinase wesentlich größer<sup>[15]</sup>. Diese Isoenzyme haben sich jeweils für das Cytosol, den Außen- und den Innenraum der Mitochondrien spezialisiert. Ein klarer Fall von Protein-Differenzierung liegt auch beim Paar Lysozym und Lactalbumin vor. Während der Entstehung der Säugetiere wurde das Lysozym-Gen dupliziert. Das eine Gen-Produkt bewahrte seine antibakteriellen Fähigkeiten, während sich das zweite zu einem wesentlichen Bestandteil der Milch entwickelte und neuartige Funktionen erfüllte. An der Verwandtschaft beider Enzyme besteht kein Zweifel, denn sie sind in 38% aller Aminosäuren gleich. Beide Enzyme bearbeiten übrigens Saccharide; dieser Aspekt ihrer Funktion blieb also erhalten.

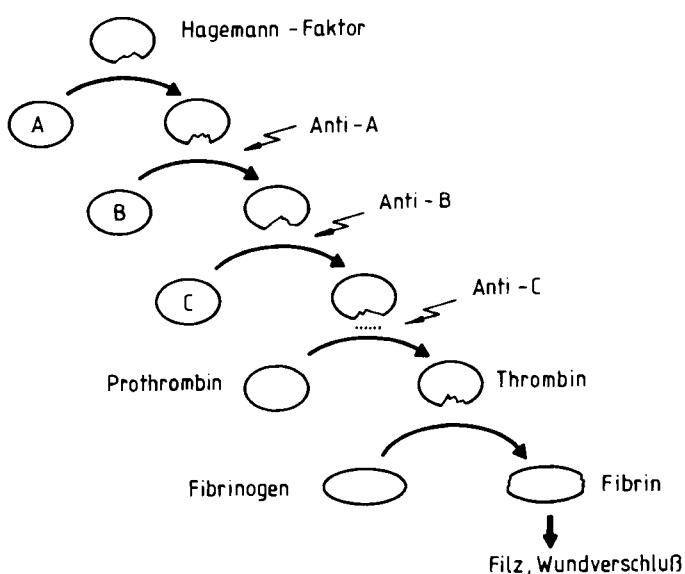


Abb. 6. Vereinfachtes Schema der Enzymkaskade in der Blutgerinnung [16]. Der Hagemann-Faktor aktiviert Enzym-A, dieses aktiviert Enzym-B usw., bis schließlich der Fibrinfilz die Wunde verschließt. Auf jeder Stufe wird das entsprechende Enzym spezifisch gehemmt. Da mehrere, sehr selektive Erkennungsprozesse beteiligt sind, ist die Regulation außerordentlich exakt.

Ein charakteristisches Beispiel für die Protein-Differenzierung ist die in Abbildung 6 skizzierte Blutgerinnungskette.

Mit der „Erfindung“ des Blutkreislaufs wurde auch ein Reparatursystem benötigt; andernfalls liefe alle Flüssigkeit bei der kleinsten Verletzung aus. Zur Reparatur bot sich Verstopfung durch Gerinnung an. Zwar ist Gerinnung leicht zu bewerkstelligen, doch dürfen nur äußerliche Verletzungen verstopft werden und nicht der Kreislauf selbst. Dazu ist genaue Kontrolle notwendig. Für dieses Problem wurde eine ganze Mannschaft von Enzymen entwickelt, die streng hierarchisch geordnet ist. Diese Kaskade wird auf jedem Niveau durch Effektoren reguliert. Alle Enzyme sind außerordentlich spezifisch für ihr jeweiliges Substrat.

Sequenzanalysen zeigten nun, daß die Enzyme dieser Kaskade miteinander verwandt sind. Da Thrombin dazugehört, besteht auch eine Verwandtschaft mit Elastase und weiterhin auch mit den Verdauungsenzymen Trypsin und Chymotrypsin sowie mit bakteriellen Proteasen (Abb. 3 und 4). Wir entdecken hier also hochspezialisierte Abkömmlinge relativ primitiver Verdauungsenzyme. Während die neu entwickelten Proteine jeweils nur ein einziges Substrat in definierter Weise bearbeiten, spalten die „ursprünglicheren“ Verdauungsenzyme alle Proteine, deren sie habhaft werden können, in kleine Stücke. Wahrscheinlich sind die Enzyme der Kaskade nicht gemeinsam, sondern von unten beginnend (Abb. 6) – aus einem einzigen Vorläufer – sukzessiv entstanden. Zunächst gab es nur eine schlecht regulierte Gerinnung. Im Laufe der Evolution wurde die Kontrolle immer präziser; die Protein-Differenzierung ermöglichte immer gezieltere Eingriffe. Eine ähnliche Kaskade ist das Komplementsystem der Immunabwehr<sup>[17]</sup>, das ebenfalls auf Trypsin-ähnlichen Proteinen basiert.

Die Entwicklung höherer Tierformen spiegelt sich auch in den durch Protein-Differenzierung entstandenen Globinen wider. Mit der „Erfindung“ des Kreislaufs trennten sich Hämoglobin und Myoglobin; Hämoglobin wurde der weitreichende O<sub>2</sub>-Träger im Plasma, während Myoglobin als lokaler O<sub>2</sub>-Träger und -Speicher in der Zelle blieb. Mit der Entwicklung der Placenta wurde schließlich die O<sub>2</sub>-Übertragung von der Mutter zum Embryo notwendig. Dafür entstanden die  $\gamma$ -,  $\epsilon$ - und  $\zeta$ -Ketten des Hämoglobins, die sich von den  $\alpha$ - und  $\beta$ -Ketten unterscheiden. Die  $\gamma$ -,  $\epsilon$ - und  $\zeta$ -Ketten werden in frühen Stadien der ontologischen Entwicklung gebildet. Nach der Geburt werden sie durch die adulten  $\alpha$ - und  $\beta$ -Formen ersetzt.

Eine weitere Verwandtschaft findet sich zwischen Immunglobulinen und HL-A-Proteinen<sup>[18]</sup>, welche die Gewebe individuell markieren und damit Transplantationen von einem Individuum zum anderen erheblich erschweren. Die Muskelproteine Parvalbumin, Troponin-C und die leichten Ketten des Myosins sowie das Calcium-abhängige Regulatorprotein<sup>[19]</sup> stammen ebenfalls von einem gemeinsamen Vorfahren ab. Bekannt ist auch die historische Verknüpfung mehrerer Protease-Inhibitoren und eines Nervenwachstumsfaktors<sup>[3]</sup> sowie die Verwandtschaft der Hormone Insulin, Relaxin und zweier Gewebewachstumsfaktoren<sup>[20]</sup>.

Bis auf wenige Ausnahmen sind in all diesen Fällen funktionelle Ähnlichkeiten der differenzierten Proteine erhalten geblieben. Allerdings basieren alle bisherigen Beispiele auf Sequenzvergleichen, die ja nur die Anfangsstadien getrennter Entwicklungen enthüllen. Anhand von Strukturvergleichen wie in Abbildung 3 oder reinen Geometrievergleichen sollten sich noch wesentlich entferntere Verwandte erkennen lassen.

#### 4. Nachweis entfernter Verwandtschaften

Durch Röntgen-Strukturanalysen wurden nun Proteine entdeckt, die keinerlei Sequenzidentität und keine funktionelle Ähnlichkeit, jedoch eine ähnliche Geometrie der Hauptkette haben. Ein solches Paar bilden Azurin<sup>[21]</sup> und Superoxid-Dismutase<sup>[22]</sup>. In beiden Proteinen hat sich die Kette zu einem  $\beta$ -Faltblatt-Fuß gefaltet, wie es in Abbildung 7 skizziert ist. Die Ketten topologien der Faltblätter sind

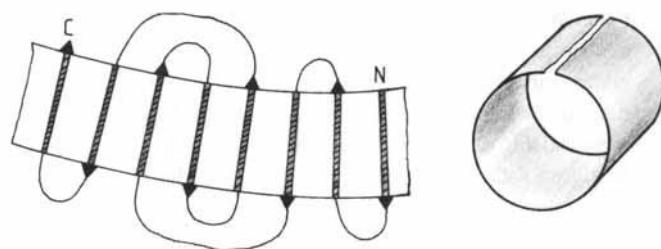


Abb. 7. Übereinstimmende Topologie der Polypeptidketten von Azurin [21] und Superoxid-Dismutase [22]. Die Stränge des antiparallelen Faltblatts sind durch Pfeile symbolisiert. Wie skizziert liegt dieses Faltblatt zusammengerollt in Form eines Fasses vor.

gleich. Würde jedes dieser Proteine nur aus einer einzigen  $\alpha$ -Helix oder einem einfachen  $\beta$ -Faltblatt bestehen, so wäre eine Erklärung schnell zur Hand: Solche Strukturen werden bevorzugt angenommen, weil sie energetisch günstig sind. Die vorhandene Komplexität der Kettenfaltung oder „Kettenfaltung“ spricht jedoch gegen einen physikalischen Effekt und ebenso klar gegen eine zufällige Ähnlichkeit. Folglich muß eine Verwandtschaft zwischen diesen Proteinen vermutet werden, die so entfernt ist, daß zwar die streng konservierte Kettenfaltung, aber keinerlei Aminosäurerest erhalten blieb.

Tabelle 1. Strukturwiederholungen innerhalb einer Kette.

Protein	Anzahl der Wiederholungen
Ferredoxin [27]	2
Parvalbumin [28]	3
Weizenkeim- $\gamma$ -Agglutinin [29]	4
Saure Proteasen [30, 31]	4
Hexokinase [32]	2
Dehydrogenasen [33]	2
Serin-Proteasen [8]	2
Immunglobuline [34]	4
Arabinose-bindendes Protein [35]	2
Glutathion-Reduktase [36]	2
Rhodanese [37]	2

Bei einigen weiteren Proteinen treten einander ähnelnde Kettenfaltungen zwei- oder mehrmals nacheinander in der Polypeptidkette auf, ohne daß irgendeine Sequenzidentität festgestellt werden kann. Die bekannten Beispiele sind in Tabelle 1 aufgeführt. Am deutlichsten zeigt sich diese Ähnlichkeit beim Enzym Rhodanese, bei dem eine komplizierte Kettenfaltung aus 140 Aminosäureresten recht präzise wiederholt wird. Hier beobachten wir offenbar das Resultat einer Gen-Duplikation mit darauffolgendem Zusammenschluß der beiden entstehenden Gen-Exemplare, auch „Tan-

dem-Gen-Duplikation“ genannt. Da solche Duplikationen in Proteinen wie Serumalbumin<sup>[23,24]</sup>, Haptoglobin<sup>[25]</sup> und Protein-A von *Staphylococcus aureus*<sup>[26]</sup> durch Sequenzvergleiche sicher nachgewiesen wurden, gelten sie als etabliert und werden hier nicht weiter diskutiert. Neben den Ähnlichkeiten innerhalb einer Kette werden auch ähnliche Kettenfaltungen von Teilbereichen verschiedener Proteine beobachtet (siehe Abschnitt 5).

Die Beispiele ähnlicher Kettenfaltungen geben zwar deutliche Hinweise auf Verwandtschaften, aber diese Hinweise sind intuitiv und nicht quantitativ wie etwa die oben berechneten Wahrscheinlichkeiten aus Sequenzidentitäten. Um solche entfernten Verwandtschaften nachzuweisen, muß man erstens ein Maß für die Ähnlichkeit von Kettenfaltungen finden und zweitens dieses Maß in eine Wahrscheinlichkeit umwandeln. Als Ähnlichkeitsmaß bietet sich die minimale mittlere Distanz zweier Ketten an. Das ist die Distanz, welche verbleibt, nachdem die beiden Kettenfaltungen mit einem Computer optimal ineinander gedreht und aufeinander geschoben worden sind. Abbildung 8a skizziert das Prinzip.

Für sich allein sagt dieses Ähnlichkeitsmaß wenig aus. Um es sinnvoll anzuwenden, muß bekannt sein, wie wahrschein-

lich diese Ähnlichkeit ist. Wahrscheinlichkeiten wiederum liest man aus der Verteilung dieses Maßes beim Vergleich beliebiger Strukturen ab. Eine solche Verteilung muß also berechnet werden. Da kurze Ketten im Mittel kleinere Distanzen voneinander haben als lange, müssen sogar mehrere Verteilungen bei jeweils konstanter Kettenlänge berechnet werden.

Im Prinzip sollte man diese Verteilungen durch Vergleiche zwischen Geometrien bekannter Proteine ermitteln. Jedoch sind zu wenig Proteingeometrien aufgeklärt, und die Proteine enthalten überdies Ketten unterschiedlicher Länge. Damit genügt die Anzahl der möglichen Vergleiche nicht zur Aufstellung einer verwendbaren Verteilungskurve. Noch schwerwiegender ist allerdings die Störung durch unerkannte Verwandtschaften, die eine solche natürliche Verteilung gerade im interessierenden Bereich kleiner Distanzen wertlos macht: Anhand von Verteilungen, die aus verwandten Proteinen abgeleitet wurden, können keine Verwandtschaften nachgewiesen werden.

Es bleibt nur der Ausweg, Kettenfaltungen unter möglichst natürlichen Bedingungen statistisch zu erzeugen (Abb. 8b) und die simulierten gefalteten Ketten miteinander zu vergleichen. Die resultierenden Gebilde sind globulär, ihre Dichte entspricht derjenigen von Proteinen; die Peptidkettenglieder bilden nur stereochemisch mögliche Winkel miteinander und halten die natürliche Winkelhäufigkeit ein. Es kommen  $\alpha$ -Helices mit natürlicher Längenverteilung und Häufigkeit vor. Tausende solcher Kettengeometrien wurden erzeugt, und Millionen von Paaren wurden verglichen. Die resultierenden Verteilungen sind in Abbildung 8c dargestellt. Mit diesen Verteilungen kann nun die Distanz zwischen zwei Kettenfaltungen (Abb. 8a) in eine Wahrscheinlichkeit umgewandelt werden. Dazu integriert man die Verteilung links vom beobachteten Distanzwert und dividiert durch das Integral über die gesamte Verteilung<sup>[38]</sup>.

Wendet man diese Methode auf die beiden ähnlichen Kettenfaltungen im Enzym Glutathion-Reduktase an (Tabelle 1), so erhält man bei einer mittleren Kettenlänge von 120 eine mittlere Distanz von etwa 6 Å und damit eine Zufalls-wahrscheinlichkeit von  $2 \cdot 10^{-7}$ . Ein Vergleich der beiden ähnlichen Kettenfaltungen in Serin-Proteasen (Tabelle 1) ergibt eine Distanz von etwa 5 Å bei einer mittleren Kettenlänge von 90 und damit eine Zufalls-wahrscheinlichkeit von  $5 \cdot 10^{-7}$ . Diese Werte sind so niedrig, daß man getrost auf Verwandtschaften schließen kann. Die Eichkurven von Abbildung 8c ermöglichen also eine Quantifizierung aller Geometrievergleiche zwischen globulären Proteinen und damit eine Quantifizierung von Verwandtschaften, die so entfernt sind, daß alle Spuren in der Sequenz bereits verwischt sind.

## 5. Anzahl ursprünglicher Proteinstrukturen

Sobald Verwandtschaften quantifizierbar sind, kann man die kühne Frage aufwerfen, ob letztlich alle Proteine miteinander verwandt sind, d. h. ob es eine allen gemeinsame Urstruktur gibt. In dieser Schärfe ist die Frage anhand von Abbildung 9 schnell beantwortet. Die hier dargestellten Protein-geometrien haben absolut nichts miteinander gemeinsam; Präalbumin besteht aus einem durchgehend antiparallelen  $\beta$ -Faltblatt<sup>[39]</sup>, während Adenylyl-Kinase<sup>[40]</sup> zahlreiche  $\alpha$ -Helices und ein paralleles  $\beta$ -Faltblatt enthält. Sollten diese Proteine jemals einen gemeinsamen Vorfahren gehabt ha-

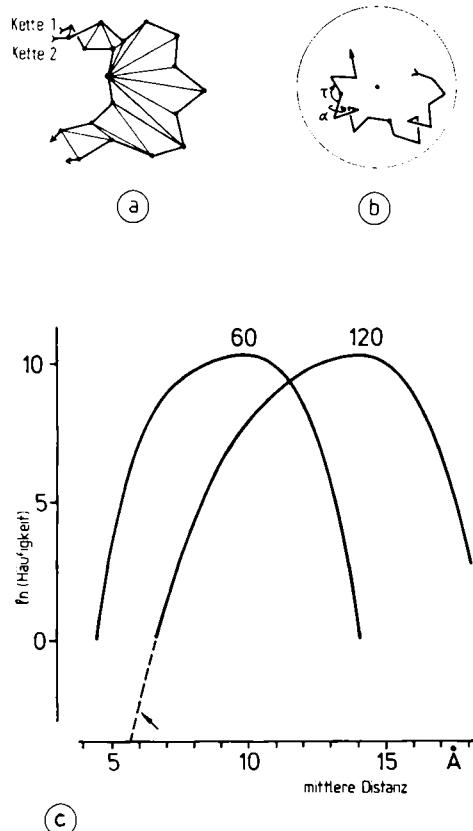


Abb. 8. Verteilung der mittleren Distanzen zwischen je zwei Polypeptidketten [38]. a) Nach optimaler relativer Verschiebung und Drehung verbleibt eine restliche Abweichung der Kettengeometrien. Die mittlere Distanz zwischen den Polypeptidketten wird durch Mittelung der Abstände zwischen den entsprechenden C<sub>α</sub>-Atomen bestimmt. b) Die Erzeugung statistischer Kettengeometrien geschieht unter Einhaltung möglichst natürlicher Verhältnisse. Die virtuellen Bindungswinkel  $\tau$  und Diederwinkel  $\alpha$  zwischen C<sub>α</sub>-Atomen werden mit ihren beobachteten Häufigkeiten verwendet. Weiterhin wird der beobachtete Anteil von  $\alpha$ -Helices und Globularität sowie die bekannte Peptidkonzentration eingestellt. c) Logarithmische Darstellung der Verteilungen für Ketten mit 60 und 120 Gliedern. Zur Ableseung der Wahrscheinlichkeit für die beiden Domänen der Glutathion-Reduktase (Pfeil) mußte die entsprechende Verteilung noch extrapoliert werden.

ben, so besteht keinerlei Hoffnung, diesen Vorfahren aus den heute vorhandenen Proteinstrukturen zu ermitteln. Durch Ähnlichkeitsanalysen können wir die rezenten Proteine bestenfalls auf eine größere Gruppe von Urstrukturen zurückführen.

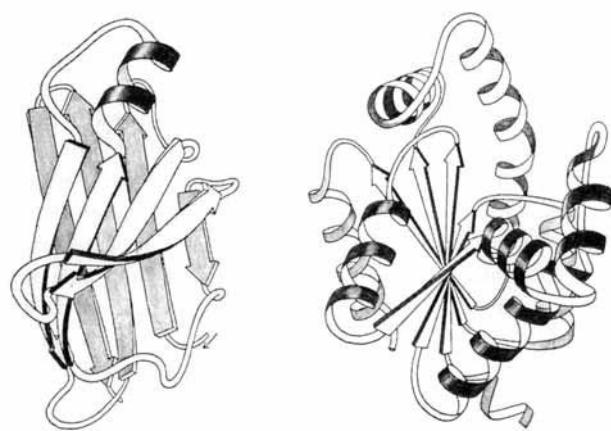


Abb. 9. Topologien der Polypeptidketten von Präalbumin (links) und Adenylat-Kinase (rechts) (mit freundlicher Genehmigung von Dr. Jane S. Richardson, Little River Institute, Rt. 1, Box 92B, Bahama, N.C. 27503 (USA)).

Wie groß ist diese Gruppe? Wir haben in Abschnitt 1 erwähnt, daß es etwa  $10^5$  Proteine verschiedener Funktion gibt. Während der bisher etwa  $G = 150$  Sequenzanalysen solcher Proteine wurden etwa  $Z = 50$  unerwartete Verwandtschaften entdeckt. Nach der hierbei anzuwendenden Poisson-Statistik ist die Koinzidenzhäufigkeit  $Z/G$  gleich der mittleren Belegung  $G/M$ , wobei  $M$  die gesuchte Anzahl von Gruppen nichtverwandter Proteine ist<sup>[41]</sup>. Demnach kann die Anzahl von Urstrukturen  $M$ , in welche die  $10^5$  Proteine anhand ihrer Sequenzen gruppiert werden können, auf etwa 500 geschätzt werden<sup>[31]</sup>. Eine analoge Rechnung können wir für die etwa 70 bekannten Kettenfaltungen von Proteinen verschiedener Funktion und die zwischen diesen entdeckten Verwandtschaften aufstellen. Unter der Annahme, daß eine Verwandtschaft besteht, sobald die Zufallswahrscheinlichkeit (Abb. 8) unter 1% sinkt, wird die Koinzidenzhäufigkeit (zwischen Domänen, siehe Abschnitt 6) auf etwa 0.3 geschätzt. Damit ergeben sich etwa 200 ursprüngliche Kettenfaltungen von Proteinen.

Die Kettenfaltung ist die am stärksten konservierte Eigenschaft eines Proteins und daher das beste bekannte Merkmal zur Unterscheidung ursprünglicher Proteine. Somit erscheint es sinnvoll, die Kettenfaltung oder Kettengeometrie als Basis zur Einordnung aller Proteine zu verwenden.

## 6. Domänen als Bausteine

In Abschnitt 5 wurden die Proteine zu Gruppen geordnet, die wahrscheinlich jeweils einen gemeinsamen Vorfahren besitzen. Damit stellt sich implizit die Frage nach der Größe dieser Vorfahren; waren sie kleiner oder größer als die heutigen Proteine, oder waren sie ebenso groß? In den Fällen, in denen die Evolution extrem weit zurückverfolgt werden kann wie etwa bei den Cytochromen<sup>[7]</sup>, findet sich keinerlei Anzeichen dafür, daß frühe Proteine andere Größen haben als späte. Proteine wachsen und schrumpfen nicht generell, sondern es kommt lediglich zu ungerichteten Schwankungen – den oben erwähnten Einfügungen und Auslassungen in der

Polypeptidkette. Wahrscheinlich sind die Cytochrome direkt in ihrer heutigen Größe entstanden.

Für diese Hypothese sprechen besonders die allenthalben beobachteten Domänen. Das sind Kettenbereiche, die zahlreiche Kontakte intern, aber nur sehr wenige Kontakte mit anderen Kettenabschnitten bilden<sup>[4]</sup>. Solche Domänen kommen in allen Proteinen mit einem Molekulargewicht über 20000 vor. Sehr deutliche Beispiele sind die Domänen der Trypsin-ähnlichen Proteine<sup>[8]</sup> und der Glutathion-Reduktase<sup>[36]</sup>, deren Geometrien in Abschnitt 4 jeweils miteinander verglichen wurden. Höchstwahrscheinlich falten sich die Domänen nach der Synthese der Polypeptidkette am Ribosom unabhängig voneinander zur endgültigen Geometrie.

Die Größe der Domänen liegt zwischen 50 und 150 Aminosäureresten. Die untere Grenze dürfte sich aus der Forderung ergeben, daß Proteine – und so auch Domänen – stabil und fest sein müssen, wenn sie spezifische Funktionen ausüben sollen. Peptidketten mit weniger als 50 Gliedern haben auch als globuläre Gebilde noch eine relativ große Oberfläche und damit intensiven Kontakt mit Wasser, welches die Ketten aufgrund ihrer polaren Molekülgruppen (Peptidbindungen) zu spreiten versucht. Ein Grund für die obere Grenze ist nicht bekannt. Es wird vermutet, daß die spontane Faltung einer größeren Domäne schwierig wird.

Proteine bestehen also aus (einem oder) mehreren strukturell relativ autarken Teilen definierter Größe, die man als Bausteine ansehen kann. Größenveränderungen der Proteine während der Evolution sind in erster Linie auf Änderungen der Bausteinanzahl und nur in geringem Maße auf Änderungen der Bausteingrößen zurückzuführen. Die Gruppeneinteilungen in Abschnitt 5 gelten für die Domänen und nicht für Proteine aus mehreren Domänen.

Eine Reihe von Strukturanalysen hat ergeben, daß gleiche (d. h. gemäß Abschnitt 4 geometrisch ähnliche) Domänen in unterschiedlichen Proteinen auftreten. Das spricht dafür, daß größere Proteine wie mit einem Baukastensystem zusammengesetzt worden sind (Abb. 10). Der Domänenbaustein A der Abbildung 10 ist aus Abbildung 7 bekannt. Er kommt einzeln in der Superoxid-Dismutase und im Azurin vor, doppelt in den leichten und vierfach in den schweren Ketten der Immunoglobuline. Alle Beispiele in Tabelle 1 gehören zu dieser Art von Proteinen, denn in allen Fällen handelt es sich um die Wiederholung eines Strukturbereichs, den man als Domäne bezeichnen kann.

Domäne B eröffnet einen weiteren Aspekt. Diese Domäne taucht nicht nur einzeln als Triosephosphat-Isomerase auf, sondern auch eingerahmt von zwei anderen Domänen in der Pyruvat-Kinase und schließlich als Teil der Glycolat-Oxidase. Hier wird der gleiche Baustein also nicht wiederholt; vielmehr wird er mit unterschiedlichen anderen Bausteinen kombiniert. Entsprechendes gilt für die Domänen F und N. Daß es sich wirklich um eine Kombination von Domänen handelt und nicht um eine separate Evolution, bei welcher die Domänen B, F und N lediglich besonders gut konserviert wurden, schließt man erstens aus den total verschiedenen Strukturen der restlichen Domänen und zweitens aus den völlig unterschiedlichen Positionen der wiederholten Domänen in den einzelnen Ketten: Domäne F kommt beispielsweise am Anfang, in der Mitte und am Ende der jeweiligen Polypeptidkette vor.

An den Domänen F und N ist noch eine weitere Eigenschaft des Baukastensystems abzulesen. Domäne F bindet

das Substrat NAD in allen aufgeführten Proteinen an gleicher Stelle der Kettenfaltung. Domäne N bindet die prosthetische Gruppe FAD in der ersten Position der Glutathion-Reduktase und in der *p*-Hydroxybenzoat-Hydroxylase und außerdem das Substrat NADPH in der zweiten Position der Glutathion-Reduktase ebenfalls an gleicher Stelle der Kettenfaltung. Diese Domänen können somit als Träger enzymatischer Teifunktionen angesehen werden.

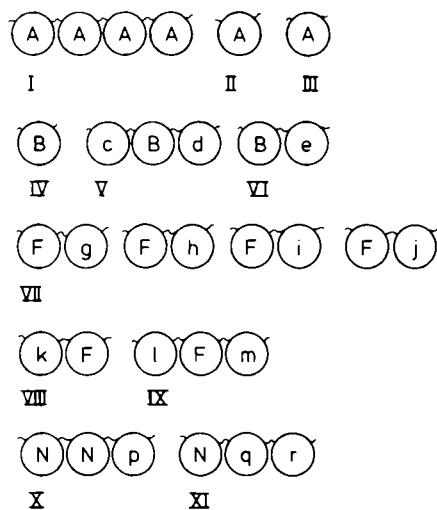


Abb. 10. Domänenaufbau einiger Proteine. Mit gleicher Topologie wiederkehrende Domänen sind mit Großbuchstaben gekennzeichnet. Domäne F bindet das Dinucleotid NAD, während Domäne N in zwei Fällen das Dinucleotid FAD und in einem Fall NADPH bindet. I: Schwere Ketten der Immunoglobuline [34]. II: Azurin [21]. III: Superoxid-Dismutase [22]. IV: Triosephosphat-Isomerase [42]. V: Pyruvat-Kinase [43]. VI: Glycolat-Oxidase [44]. VII: Lactat-, Malat-, Glyceraldehyd-3-phosphat-Dehydrogenase [45] und Phosphoglycerat-Kinase [46, 47]. VIII: Alkohol-Dehydrogenase [45]. IX: Phosphorylase [48]. X: Glutathion-Reduktase [36]. XI: *p*-Hydroxybenzoat-Hydroxylase [49].

Besonders deutlich wird dieser Aspekt in der Glutathion-Reduktase. Dieses Enzym verwendet FAD als Brücke bei der Übertragung von Reduktionsäquivalenten von NADPH auf oxidiertes Glutathion. Hier nehmen drei Partner an der katalytischen Reaktion teil, und jeder Partner bindet an seine Domäne: FAD an die erste, NADPH an die zweite und Glutathion an die dritte<sup>[36]</sup>. Man kann daher die Hypothese aufstellen, daß die zur Bewältigung komplizierter chemischer Reaktionen im Laufe der Evolution benötigten Teifunktionen durch Kombination der entsprechenden Domänen zusammengestellt werden. Die Bausteinenauswahl wird durch die erwünschte Reaktion dirigiert. Diese Hypothese setzt voraus, daß die zu den Domänen gehörenden Gene im langsamsten Zeitmaß der Evolution beliebig auf dem Genom transferiert und dann fusioniert werden können. Da solche Prozesse von der Differenzierung der Antikörper-produzierenden Zellen während der Entwicklung eines Individuums bekannt sind<sup>[50]</sup>, wirft diese Annahme keine prinzipiellen Probleme auf.

## 7. Zusammenfassung

Während der Evolution verändern sich Proteine so kontinuierlich wie die Arten. Die beobachteten Veränderungen

enthüllen wesentliche Aspekte der Proteinstrukturen und ermöglichen vor allem die Unterscheidung zwischen wichtigen und unwichtigen Merkmalen. Wenn neuartige Aufgaben gestellt werden, dann werden sie durch neue Proteine erfüllt, die durch Gen-Duplikation vorhandener Proteine und anschließende Veränderungen entstehen. Die ursprünglichen Proteine behalten dabei ihre alten Funktionen. Dieser Prozeß wird Protein-Differenzierung genannt. Nicht nur aus Sequenzvergleichen, sondern auch aus geometrischen Vergleichen können quantitative Angaben über Entwicklungsdistanzen abgeleitet werden. Da die Geometrien außerordentlich streng konserviert werden, ermöglichen solche Vergleiche das Aufspüren auch sehr entfernter Verwandter. Es ist recht unwahrscheinlich, daß sich alle Proteine aus einem einzigen Vorfahren entwickelt haben. Vielmehr kann man abschätzen, daß es etwa  $10^2$  Urstrukturen gegeben hat. Diese Urstrukturen beziehen sich auf Domänen. Proteine wachsen oder schrumpfen nicht kontinuierlich während der Evolution. Die beobachteten Größenänderungen ergeben sich durch Kombination von Domänen. Große Proteine scheinen nach einem Baukastenprinzip zusammengestellt zu werden, wobei Bausteine mit den benötigten Teifunktionen ausgewählt werden. Im Lichte der sich allmählich enthüllenden Ordnung erscheint ein zukünftiges Gesamtverständnis aller Proteinstrukturen nicht mehr utopisch.

Eingegangen am 3. Oktober 1980 [A 350]

- [1] R. E. Dickerson in M. Kimura: Molecular Evolution and Polymorphism. National Institute of Genetics, Mishima, Japan 1977.
- [2] T. L. Blundell, L. N. Johnson: Protein Crystallography. Academic Press, London 1976.
- [3] M. O. Dayhoff: Atlas of Protein Sequence and Structure. Vol. 5. Suppl. 3 und frühere. National Biomedical Research Foundation, Washington, D.C. 1978.
- [4] G. E. Schulz, R. H. Schirmer: Principles of Protein Structure. Springer, New York 1979.
- [5] K. Sick, D. Beale, D. Irvine, H. Lehmann, P. T. Goodall, S. MacDougall, Biochim. Biophys. Acta 140, 231 (1967).
- [6] A. C. Allison, Sci. Am. 195 (2), 87 (1956).
- [7] R. E. Dickerson, R. Timkovich, R. J. Almasy, J. Mol. Biol. 100, 473 (1976).
- [8] J. J. Birktoft, D. M. Blow, J. Mol. Biol. 68, 187 (1972).
- [9] L. T. J. Delbaere, W. L. B. Illictheon, M. N. G. James, W. E. Thiessen, Nature 257, 758 (1975).
- [10] D. M. Shotton, B. S. Hartley, Biochem. J. 131, 643 (1973).
- [11] S. Magnusson, Enzymes 3, 277 (1971).
- [12] B. S. Hartley, Proc. R. Soc. London B 205, 443 (1979).
- [13] S. S. Taylor, J. Biol. Chem. 252, 1799 (1977).
- [14] H. H. Kiltz, W. Keil, M. Griesbach, K. Petry, H. Meyer, Hoppe-Seylers Z. Physiol. Chem. 358, 123 (1977).
- [15] A. G. Tomasselli, R. H. Schirmer, L. H. Noda, Eur. J. Biochem. 93, 257 (1979).
- [16] E. W. Davie, K. Fujikawa, Annu. Rev. Biochem. 44, 799 (1975).
- [17] J. Müller-Eberhard, Annu. Rev. Biochem. 44, 697 (1975).
- [18] P. A. Peterson, L. Rask, K. Sege, L. Klareskog, H. Anundi, L. Ostberg, Proc. Natl. Acad. Sci. USA 72, 1612 (1975).
- [19] R. H. Kretsinger, CRC Crit. Rev. Biochem. 8, 119 (1980).
- [20] E. Rinderknecht, R. E. Hlumel, FEBS Lett. 89, 283 (1978).
- [21] E. T. Adman, R. E. Stenkamp, L. C. Sieker, L. H. Jensen, J. Mol. Biol. 123, 35 (1978).
- [22] J. S. Richardson, K. A. Thomas, B. H. Rubin, D. C. Richardson, Proc. Natl. Acad. Sci. USA 72, 1349 (1975).
- [23] J. R. Brown, Fed. Proc. Fed. Am. Soc. Exp. Biol. 35, 2141 (1976).
- [24] A. D. McLachlan, J. E. Walker, J. Mol. Biol. 112, 543 (1977).
- [25] J. A. Black, G. H. Dixon, Nature 218, 736 (1968).
- [26] J. Sjödahl, Eur. J. Biochem. 78, 471 (1977).
- [27] E. T. Adman, L. C. Sieker, L. H. Jensen, J. Biol. Chem. 248, 3987 (1973).
- [28] R. H. Kretsinger, Nature New Biol. 240, 85 (1972).
- [29] C. S. Wright, J. Mol. Biol. 141, 267 (1980).
- [30] J. Tang, M. N. G. James, I. N. Hsu, J. A. Jenkins, T. L. Blundell, Nature 271, 618 (1978).
- [31] N. S. Andreeva, A. E. Gustchina, Biochem. Biophys. Res. Commun. 87, 32 (1979).

- [32] A. D. McLachlan, Eur. J. Biochem. 100, 181 (1979).  
 [33] M. G. Rossmann, D. Moras, K. W. Olsen, Nature 250, 194 (1974).  
 [34] R. Huber, Trends Biochem. Sci. 1, 174 (1976).  
 [35] F. A. Quirocho, G. L. Gilliland, G. N. Phillips, Jr., J. Biol. Chem. 252, 5142 (1977).  
 [36] G. E. Schulz, R. H. Schirmer, W. Sachsenheimer, E. F. Pai, Nature 273, 120 (1978).  
 [37] J. H. Ploegman, G. Drenth, K. H. Kalk, W. G. J. Hol, R. L. Heinrikson, P. Keim, L. Weng, J. Russell, Nature 273, 124 (1978).  
 [38] G. E. Schulz, J. Mol. Biol. 138, 335 (1980).  
 [39] C. C. F. Blake, M. J. Geisow, I. D. Swan, C. Rerat, B. Rerat, J. Mol. Biol. 88, 1 (1974).  
 [40] G. E. Schulz, M. Elzinga, F. Marx, R. H. Schirmer, Nature 250, 120 (1974).  
 [41] G. E. Schulz, J. Mol. Evol. 9, 339 (1977).  
 [42] D. W. Banner, A. C. Bloomer, G. A. Petsko, D. C. Phillips, C. J. Pogson, I. A. Wilson, P. H. Corran, A. J. Furth, J. D. Milman, R. E. Offord, J. D. Priddle, S. G. Waley, Nature 255, 609 (1975).  
 [43] M. Levine, H. Muirhead, D. K. Stammers, D. I. Stuart, Nature 271, 626 (1978).  
 [44] Y. Lindquist, C.-I. Brändén, J. Mol. Biol. 143, 201 (1980).  
 [45] M. G. Rossmann, A. Liljas, C.-I. Brändén, I. J. Banaszak, Enzymes 11, 61 (1975).  
 [46] C. C. F. Blake, P. R. Evans, J. Mol. Biol. 84, 585 (1974).  
 [47] T. N. Bryant, H. C. Watson, P. L. Wendell, Nature 247, 14 (1974).  
 [48] S. Sprang, R. J. Fletterick, J. Mol. Biol. 131, 523 (1979).  
 [49] R. K. Wierenga, R. J. DeJong, K. H. Kalk, W. G. J. Hol, J. Drenth, J. Mol. Biol. 131, 55 (1979).  
 [50] N. Hilschmann, H. U. Barnikol, H. Kratzin, P. Altevogt, M. Engelhard, S. Barnikol-Watanabe, Naturwissenschaften 65, 616 (1978).

## Alkylsulfamidsäurechloride als Schlüsselbausteine für neue Pflanzenschutzwirkstoffe

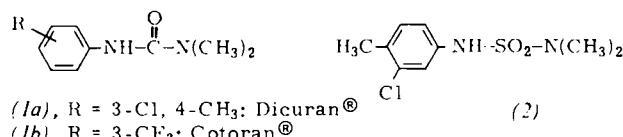
Von Gerhard Hamprecht, Karl-Heinz König und Gerd Stubenrauch<sup>[\*]</sup>

Professor Matthias Seefelder zum 60. Geburtstag gewidmet

Die difunktionellen Alkylsulfamidsäurechloride sind variable Bausteine für die Synthese von Heterocyclen, polaren Sulfamidsäureestern und Sulfonamiden. In den letzten Jahren wurden präparativ allgemein anwendbare Methoden entwickelt, nach denen neue, verschiedenartig substituierte Alkylsulfamidsäurechloride durch Umsetzung von Aminhydrochloriden, Isocyanaten, Aziridinen oder tertiären Alkoholen mit Schwefelsäure-Derivaten synthetisiert werden können. In Folgereaktionen lassen sich diese Zwischenprodukte einerseits zu früher unzugänglichen *N*-Alkoxyalkyl-*N*-alkylsulfamidsäurechloriden oder neuen Heterocyclen vom Typ der 1*H*-2,1,3-Benzothiadiazin-4-on-2,2-dioxide, 2*H*-1,2,6-Thiadiazin-3-on-1,1-dioxide oder 2*H*-1,2,4,6-Thiatriazin-5-on-1,1-dioxide umwandeln; diese sind Beispiele interessanter Struktur-Wirkungs-Modelle und führen zum Teil zu sehr selektiven, umweltfreundlichen Herbiziden. Andererseits sind durch *N*-Aclylierung der Alkylsulfamidsäurechloride weitere drei- bis fünfatomige difunktionelle Synthesebausteine zugänglich, die sich für neuartige Heterocyclensynthesen eignen. Weitere Anwendungsmöglichkeiten sind die Herstellung biologisch aktiver Sulfamidsäureester sowie Cycloadditionen in situ erzeugter *N*-Sulfonylamine (Sulfene).

### 1. Einleitung

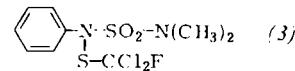
Im modernen chemischen Pflanzenschutz nehmen Kohlensäure-Derivate eine bedeutende Stellung ein; Produkte, die sich beispielsweise von Isocyanaten ableiten, wie die *N*-Alkyl-*N'*-arylharnstoffe (1), sind selektivherbizide Verbindungen<sup>[2a]</sup>, die in Großkulturen wie Getreide, Zuckerrohr und Baumwolle für die Sicherung des wachsenden Nahrungsmittel- und Rohfaserbedarfs heute weltweit angewendet werden.



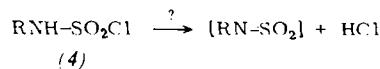
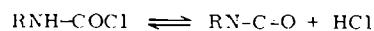
Sulfonsäure-Derivate fanden hingegen bisher wenig Verwendung. Zum einen waren Monoalkylsulfamidsäurechloride (Monoalkylsulfamoylchloride) als Ausgangsmaterialien

nicht bekannt, zum anderen ergaben erste Struktur-Wirkungsuntersuchungen, daß beispielsweise das Sulfamid (2) – durch Ersatz der Dimethylcarbamoylgruppe in (1) durch die Dimethylsulfamoylgruppe erhalten – herbizid nur wenig wirksam ist.

Lediglich auf dem Fungizidgebiet konnte das aus *N,N*-Dimethylsulfamidsäurechlorid (18) (siehe Abschnitt 2.7) hergestellte Euparen® (3) in der Praxis Bedeutung bei der Bekämpfung von Botrytis im Erdbeerbau<sup>[2b]</sup> gewinnen.



Monoalkylsulfamidsäurechloride (4) hielt man noch in den sechziger Jahren für eine nicht existenzfähige Verbindungsklasse<sup>[3]</sup>. In Anlehnung an das Verhalten von Carb-



[\*] Dr. G. Hamprecht, Dr. K.-H. König, Dr. G. Stubenrauch  
 BASF AG, Hauptlabor  
 D-6700 Ludwigshafen