

# ANGEWANDTE CHEMIE

93. Jahrgang 1981  
Heft 12  
Seite 1037-1124

## DNA-Sequenzierung und Gen-Struktur (Nobel-Vortrag)<sup>[\*\*]</sup>

Von Walter Gilbert<sup>[\*]</sup>

### DNA-Sequenzierung

Analysieren wir die Struktur von DNA-Molekülen, so untersuchen wir ein fundamentales Prinzip, das allen Prozessen in lebenden Zellen zugrundeliegt. DNA ist der Informationsspeicher, der letztlich die Struktur eines jeden Genproduktes und damit jeden Teil des Organismus bestimmt. Die Basenanordnung entlang der DNA enthält den kompletten Satz an Instruktionen, die das genetische Erbgut ausmachen. Wir sind nicht imstande, diese Instruktionen zu interpretieren; wie ein kleines Kind können wir Buchstaben identifizieren, ohne mehr als ein paar Wörter auf jeder Seite zu verstehen.

Ich stieß zufällig auf die chemische DNA-Sequenzanalyse (im Fachjargon DNA-Sequenzierung). Seit Mitte der sechziger Jahre konzentrierte sich meine Arbeit auf die Gen-Kontrolle in Bakterien; ich untersuchte ein spezifisches Gen-Produkt, einen Repressor, der von dem Regulatoren für das lac-Operon (ein Gen-Cluster, dessen Produkte den Zucker Lactose metabolisieren) codiert wird. *Benno Müller-Hill* und ich isolierten und charakterisierten dieses Molekül während der späten sechziger Jahre; wir wiesen nach, daß dieses Protein an die bakterielle DNA direkt am Anfang des ersten Gens des aus drei Genen bestehenden Clusters, der von diesem Repressor kontrolliert wird, bindet<sup>[1,2]</sup>. In den Jahren danach zeigte meine Arbeitsgruppe, daß dieses Protein die RNA-Polymerase daran hindert, die lac-Operon-Gene in RNA zu kopieren. Der lac-Repressor bindet an eine spezifische Stelle der DNA – den Operator. Diese Tatsache nutzte ich zur Isolierung der DNA dieser Region; mit DNase wurde

die restliche DNA abhydrolysiert, und nur ein kleines Fragment, an das der Repressor gebunden war, blieb vor dem Angriff des Enzyms geschützt. So konnte ein 25 Basenpaare enthaltendes DNA-Fragment aus den 3 Millionen Basenpaaren des Bakterienchromosoms isoliert werden. In den frühen siebziger Jahren bestimmten *Allan Maxam* und ich die Sequenz dieses kleinen Fragmentes<sup>[3]</sup>, indem wir diese DNA in kurze RNA-Fragmente transkribierten und auf diese RNA-Kopien die Sequenzierungsmethoden anwandten, die *Sanger* et al. in den späten sechziger Jahren entwickelt hatten. Das war ein mühsamer Prozeß, der einige Jahre in Anspruch nahm. Als eine Studentin, *Nancy Maizels*, die Sequenz der ersten 63 Basen der messenger-RNA der lac-Operon-Gene bestimmt hatte, entdeckten wir, daß der lac-Repressor sich an die DNA direkt vor dem Startpunkt der messenger-RNA heftet<sup>[4]</sup>, an eine Region also, die an der Stelle liegt, an die sich auch die RNA-Polymerase bindet, um die RNA-Synthese zu initiieren. Wir charakterisierten den lac-Operator weiter, indem wir eine Reihe von Mutationen (Operator-konstitutive Mutationen) sequenzierten, in denen die Bindung des Repressors an die DNA nicht möglich ist. Wir wollten DNA-Sequenzierungen in dieser Region vornehmen, um die Bindungsstelle der Polymerase zu definieren und andere Elemente des Kontrollabschnitts für die lac-Gene kennenzulernen; diese Sequenz wurde jedoch in einem anderen Labor von *Dickson, Abelson, Barnes* und *Reznikoff* bestimmt<sup>[5]</sup>. Mitte der siebziger Jahre kannte ich also alle Sequenzen, die mich interessierten, und meine Studenten (*David Pribnow* und *John Majors*) und ich versuchten, die Frage nach der Interaktion der RNA-Polymerase und anderer Kontrollfaktoren mit der DNA zu beantworten.

An diesem Punkt wurde durch eine neue Anregung mit einer anderen Versuchsreihe begonnen. *Andrei Mirzabekov* besuchte mich Anfang 1975. Sein Besuch hatte zwei Gründe: Er wollte über seine Experimente berichten, in denen er mit Dimethylsulfat Guanin- und Adeninreste der DNA methyliert hatte; daneben wollte er mir nahelegen, ein

[\*] Prof. Dr. W. Gilbert  
Harvard University, The Biological Laboratories  
Cambridge, Massachusetts 02138 (USA)

[\*\*] Copyright © The Nobel Foundation 1981. – Wir danken der Nobel-Stiftung, Stockholm, für die Genehmigung zum Druck dieser Übersetzung.

ähnliches Experiment mit der DNA-Region, die den lac-Repressor bindet, durchzuführen. Dimethylsulfat methyliert die Guaninreste nur in der Position N7, die Adeninreste nur in der Position N3; diese Positionen sind in der großen bzw. kleinen Rille der DNA-Doppelhelix exponiert (Abb. 1). *Mirzabekov* hatte versucht, diese Eigenschaften zu nutzen, um die Lage von Histonen und bestimmten Antibiotica am DNA-Molekül zu bestimmen, indem er beobachtete, wo die Einführung radioaktiver Methylgruppen in Guanin- und Adeninreste innerhalb der gesamten DNA blockiert war. Er drängte mich, diese Rillenspezifität zu nutzen, um etwas über die Wechselwirkung zwischen lac-Repressor und lac-Operator zu erfahren. Die verfügbare Menge an lac-Operator war jedoch extrem klein; außerdem schien es unmöglich, Aussagen über das Protein auf der DNA machen zu können, indem man die Basen der Sequenz bestimmte, die durch das Protein gegen den Angriff des Dimethylsulfates geschützt waren.

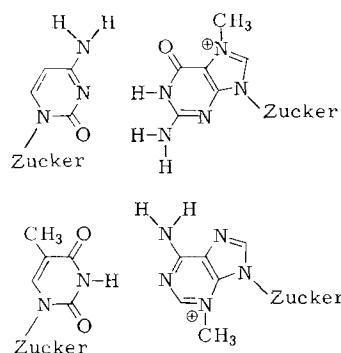


Abb. 1. Die methylierten Basenpaare Cytosin-Guanin und Thymin-Adenin. Die Abbildung zeigt oben das Basenpaar Cytosin-Guanin, das an der Position N7 des Guanins methyliert ist. Unten ist das Basenpaar Thymin-Adenin zu sehen, das an der Position N3 des Adenins methyliert ist. Die obere Region jedes Basenpaares ist in der großen Rille der DNA exponiert. Die untere Region jedes Basenpaares liegt zwischen den Zucker-Phosphat-Gerüsten in der kleinen Rille der DNA-Doppelhelix.

Erst nach dem zweiten Besuch von *Mirzabekov* kristallisierte sich eine Idee heraus. Er, ich, *Allan Maxam* und *Jay Gralla* aßen gemeinsam zu Mittag. Während unserer Unterhaltung fiel mir das Experiment ein, das unserer Sequenzierungsmethode letztlich zugrunde liegt. Wir wußten, wir können ein definiertes DNA-Fragment – 55 Basenpaare lang – erhalten, das etwa in der Mitte die Region trägt, an die der lac-Repressor bindet. Dieses Fragment läßt sich durch aufeinanderfolgendes Spalten der DNA mit zwei verschiedenen Restriktionsenzymen herstellen; jedes Enzym definiert ein Ende des Fragmentes (Abb. 2). Zweitens wußte ich, daß jede methylierte Base der DNA durch Erhitzen abgespalten werden kann. Danach würde die DNA-Kette nur noch durch einen Zuckerrest zusammengehalten; durch alkalische Hydrolyse ließe sich dann prinzipiell der DNA-Strang an dieser Stelle brechen. Ich vermutete, daß wir, wenn wir ein Ende des DNA-Fragmentes mit radioaktivem Phosphat markierten, den Ort der Methylierung durch „Messen“ der Distanz zwischen dem markierten Ende und dem Spaltort bestimmen könnten. Die markierten DNA-Stücke sollten sich folgendermaßen herstellen lassen: Ein DNA-Fragment, das durch ein Restriktionsenzym aus der DNA „herausgeschnitten“ wird, wäre an beiden Enden zu markieren; nach Spaltung mit ei-

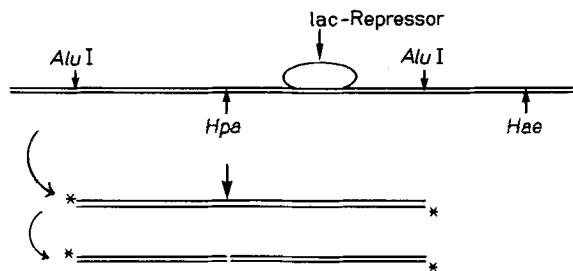


Abb. 2. Verfahren zur Gewinnung doppelsträngiger DNA-Fragmente, die nur an einem Ende eines Stranges markiert sind. Die Abbildung zeigt die Restriktionsstellen für die Enzyme *Alu*I (an AG/CT), *Hpa* (an C/CGG; das Enzym hinterläßt ungleiche Enden) und *Hae*; die Stellen für *Alu*I und *Hpa* befinden sich in der Nähe des lac-Operators. Der lac-Repressor ist an die DNA gebunden. Wir können durch folgende Schritte ein DNA-Fragment isolieren, das die Bindungsstelle für den Repressor enthält und nur an einem Ende eines Stranges markiert ist; zuerst wird die DNA dieser Region durch das Enzym *Alu*I gespalten, dann werden durch die Polynukleotid-Kinase die 5'-Enden beider DNA-Strände mit radioaktivem  $^{32}\text{P}$  markiert und schließlich wird mit dem Enzym *Hpa* gespalten.

nem zweiten Restriktionsenzym blieben zwei (aufgrund ihrer unterschiedlichen Länge durch Elektrophorese in Polyacrylamid-Gelen) voneinander trennbare, an nur einem Ende markierte Doppelstrang-Fragmente übrig. Mit diesem Verfahren könnte durch die Polynukleotid-Kinase das 5'-Ende von einem der DNA-Stränge des Fragmentes, das den Operator trägt, markiert werden, während das 3'-Ende unmarkiert bliebe (Abb. 2). Wenn wir dann die DNA mit Dimethylsulfat so modifizierten, so daß nur einige der Adenin- und Guaninreste methyliert würden, anschließend erhitzten und danach die DNA alkalisch an den depurinierten Stellen spalteten, erhielten wir neben anderen Fragmenten ein markiertes Bruchstück, das sich vom Punkt der Markierung bis zur ersten Bruchstelle erstreckte. Abbildung 3 illustriert diese Idee. Alle Fragmente des an-

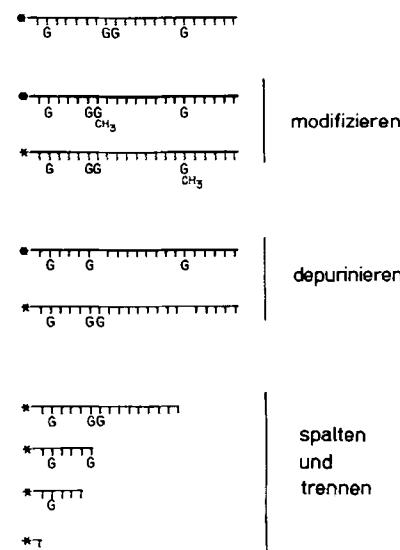


Abb. 3. Skizzenhafte Darstellung des Verfahrens zur Herstellung von DNA-Fragmenten durch Spalten der DNA an Guaninresten. Stellen wir uns einen DNA-Strang vor, der an einem Ende markiert ist, und in dem wir vereinzelte Guaninreste durch Methylierung mit Dimethylsulfat modifizieren. Durch Erhitzen der DNA werden diese Guaninreste aus dem DNA-Strang entfernt; an der Stelle bleibt ein Zuckerrest zurück. Die DNA kann dann an dieser Position gespalten werden, und die Fragmente werden schließlich nach Größe getrennt; die Länge des Fragmentes, das radioaktiv markiert ist und somit nachgewiesen werden kann, gibt Auskunft über die Position des modifizierten Guanins.

deren Stranges wären unmarkiert, ebenso alle Fragmente, die aus dem Teil jenseits der ersten Bruchstelle entstünden. Trennten wir diese Fragmente nach ihrer Länge, wie es prinzipiell durch Elektrophorese in Polyacrylamid-Gelen möglich ist, dann könnten wir die markierten Fragmente der bekannten Sequenz zuordnen und auf diese Weise jedes Guanin und Adenin des Operators identifizieren, das methyliert worden war. Wenn wir diese Modifikation in Anwesenheit des – an das DNA-Fragment gebundenen – lac-Repressors durchführten, und wenn der Repressor nahe an einem N7 eines Guanins läge, dann würden wir die DNA an dieser Base nicht modifizieren, und das entsprechende Fragment tauchte im analytischen Muster nicht auf.

Ich begab mich an die Durchführung dieses Experiments. *Allan Maxam* stellte die markierten DNA-Fragmente her, und ich begann das Modifizieren und Spalten von DNA zu erlernen. Dazu mußten das Abspalten der Basen aus der DNA und die zum Strangbruch führenden Schritte getrennt untersucht werden. Abbildung 4 zeigt die Ergebnisse des gesamten Experiments: Ein Autoradiogramm des Elektrophorese-Musters enthält eine Reihe von Bändern, die von Fragmenten hervorgerufen werden, deren Größe von oben nach unten abnimmt; jede Bande ist die Folge einer DNA-Spaltung an einem Adenin- oder Guaninrest. Wird das DNA-Fragment mit Dimethylsulfat in Anwesenheit des Repressors umgesetzt, so entsteht ein ähnliches Muster, dem jedoch einige Bänder fehlen (Reihe 1 im Vergleich zu Reihe 2 in Abb. 4). Das Experiment war insofern ein deutlicher Erfolg, als die Anwesenheit des Repressors den Angriff von Dimethylsulfat auf einige Guanin- und Adeninreste des Operators verhinderte<sup>[6]</sup>. Ich hoffte, die Auf trennung nach der Größe sei genau genug, um die Zuordnung einer jeden Bande im Muster zu einer spezifischen Base zu erlauben. Das erwies sich als zutreffend, da die Abstände im Verteilungsmuster und das Vorhandensein von hellen und dunklen Banden – die dunklen Bänder entsprachen den Guaninresten, die hellen den Adeninresten – hinreichend charakteristisch waren, um beide zu korrelieren. Die Guaninreste reagieren etwa fünfmal so schnell mit Dimethylsulfat wie die Adeninreste, während methylierte Adeninreste durch Erhitzen schneller als methylierte Guaninreste aus der DNA entfernt werden; der Unterschied der Intensitäten als Funktion der Dauer der Hitzeeinwirkung konnte genutzt werden, um jede Base eindeutig zuzuordnen. Das Gel-Muster ist so deutlich, daß selbst Fragmente, die sich in der Länge nur durch eine Base unterscheiden, gut getrennte Bänder ergeben. Zu diesem Zeitpunkt war klar, daß mit dieser Technik die Adenin- und Guaninreste in DNA-Fragmenten, die aus etwa 40 Nucleotiden bestehen, bestimmt werden können. Durch Identifizierung der Purine in den beiden komplementären Strängen verfügt man prinzipiell über eine vollständige Sequenzierungsmethode (vgl. Abb. 4).

Wir hatten nun eine Methode, mit der wir Adenin- und Guaninreste bestimmen und voneinander unterscheiden konnten; würden wir Reaktionen finden, mit denen sich Cytosin- und Thyminreste charakterisieren lassen? *Allan Maxam* und ich beschäftigten uns mit diesem Problem. (Zunächst jedoch untersuchten wir eine zweite Bindungsstelle für den lac-Repressor, die einige Hundert Basen entfernt im ersten Gen des Operons liegt. Diese Bindungs-

stelle hat keine physiologische Funktion. Wir konnten sie in einem Restriktionsfragment lokalisieren, mit dem wir das „Methylierungs-Schutz-Experiment“ wiederholten und die Basen identifizierten, die durch den lac-Repressor geschützt waren. Ich benutzte das Methylierungs-Muster, um die Positionen der Adenin- und Guaninreste der unbekannten DNA-Sequenz vorherzusagen. *Allan Maxam* se-



Abb. 4. „Methylierungs-Schutz-Experiment“ mit dem lac-Repressor. Die Reihen zeigen das Spaltungsmuster für jeden Strang des 53-55 Basen enthaltenden Fragmentes, das den lac-Operator trägt. Die zweite Reihe von links repräsentiert die DNA (sie wurde am 5'-Ende des 53 Basen langen Stranges markiert), die mit Dimethylsulfat umgesetzt, erhitzt und dann durch ein alkalisches Reagens gespalten wurde. Die dunklen Bänder entsprechen Brüchen an Guaninresten; die hellen Bänder entsprechen Brüchen an Adeninresten. Die zweite Reihe ist von unten zu lesen:

-G-GAA-G-A-G-A-AA---A-A-AA-A-A-...

Die erste Reihe zeigt das gleiche doppelsträngige DNA-Stück, das in Gegenwart des lac-Repressors mit Dimethylsulfat umgesetzt wurde. Der Repressor verhindert die Reaktion von Dimethylsulfat mit dem Guaninrest, dessen Bande sich im Muster der 2. Reihe am Ende des unteren Drittels befindet; er blockiert auch die Reaktion mit den beiden Adeninresten, deren Bande im oberen Drittel des Musters erscheint; diese beiden Bänder repräsentieren Fragmente, die sich in der Länge nur durch eine Base unterscheiden und 30 bzw. 31 Basen enthalten. Die rechte Seite des Autoradiogramms wurde durch das gleiche Experiment erhalten, nur befand sich die Markierung am Ende des anderen DNA-Stranges. Dem Muster rechts außen entspricht die Sequenz:

-G-G-GGAA-G-GAG-GGA-AA-AA-....

quenzierte dann mit der „wandering-spot“-Methode von *Sanger et al.* die DNA-Region, um unsere Vorhersage zu verifizieren.) *Allan Maxam* tat dann den nächsten Schritt. Wir wußten, daß Hydrazin die Cytosin- und Thyminreste der DNA angreift und sie entweder zerstört oder unter Hydrazonbildung eliminiert, so daß eine Reaktion mit Benzaldehyd und danach mit einem alkalischen Reagens (oder eine Umsetzung mit einem Amin) die DNA an den beschädigten Stellen spalten würde. So erhielten wir zwar bald ein ähnliches Muster, die DNA wurde aber gleichermaßen sowohl an Cytosin als auch an Thymin gespalten. *Allan Maxam* fand dann, daß die Thyminreste nicht mehr reagierten, wenn man dem 15 M Hydrazin 1 mol/L Kochsalz

zusetzte. Beide Reaktionen zusammen ermöglichten die Unterscheidung und Positionsbestimmung der Thymin- und Cytosinreste in der DNA-Sequenz und vervollständigten damit unsere Methode. Um die Unterscheidung zwischen den Purinen zu verbessern und um zusätzliche Informationen zu erhalten, die die Sequenzierung gegen Fehler absichern würden, nutzten wir die Tatsache, daß die methylierten Adenosinreste in saurem Milieu schneller depurinieren, so daß bevorzugt Adenosin freigesetzt wird. Wir hatten also vier Reaktionen zur Verfügung: eine für die As, eine bevorzugt für die Gs, eine für die Cs und Ts sowie eine für die Cs; die Ts wurden durch Vergleich bestimmt. Dieser Arbeitsabschnitt war nach einigen Monaten beendet. Als wir die Sequenzierung auf Fragmente mit bis zu 100 Basen ausdehnten und den Auflösungsbereich des Gels entsprechend vergrößerten, stellten wir fest, daß die Spaltung an den Pyrimidinen nicht zufriedenstellend war; die Folge dieser unvollständigen Spaltung waren längere Fragmente mit einer Reihe interner Schädigungen, die ein verschwommenes Muster erzeugten. Nach vielen Monaten der Suche konnte eine Lösung gefunden werden: Das primäre Amin Anilin verdrängt die Produkte der Umsetzung mit Hydrazin und entfernt durch  $\beta$ -Eliminierung den Phosphatrest aus der 3'-Position des Zuckers; der andere Phosphatrest wird nicht freigesetzt. Das Laufverhalten eines DNA-Fragments mit blockiertem 3'-Phosphat, das einen Zucker-Anilin-Rest enthält, unterscheidet sich von dem der Ketten aus den anderen Reaktionen, die mit freiem Phosphat enden. Das sekundäre Amin Piperidin ist noch effektiver und bewirkt sowohl  $\beta$ -Eliminierungen als auch die Eliminierung aller Abbauprodukte der Hydrazin-Reaktion aus dem Zucker. Dieses Reagens vervollständigte die Techniken zur DNA-Sequenzierung<sup>[7]</sup>. Obwohl die Ausarbeitung von Details weitere neun Monate in Anspruch nahm, teilten wir das Verfahren anderen Gruppen mit, die die Methoden benutzen wollten. Abbildung 5 zeigt ein Sequenzierungsmuster aus der Zeit um 1978, das während der in<sup>[8]</sup> beschriebenen Arbeit erhalten wurde. Abbildung 6 zeigt an zwei Beispielen den Reaktionsablauf<sup>[9]</sup>.

Der Sinn dieser chemischen Methode ist es, den Angriff in zwei Schritte zu unterteilen. Im ersten Schritt benutzen wir ein spezifisches Reagens für die jeweilige Base, aber wir führen dessen Reaktion mit dem DNA-Fragment so durch, daß nur etwa eine Base aus den Hunderten der möglichen Angriffsstellen in jedem DNA-Fragment modifiziert wird. So kann die Reaktion im Bereich größter Spezifität eingesetzt werden: Nur die sehr frühen Stadien einer chemischen Reaktion laufen ab. Der zweite Schritt, die Spaltung des DNA-Stranges, muß vollständig sein. Da der Angriffsort schon durch die Modifikation im ersten Schritt von den anderen Basen der DNA-Kette unterschieden ist, können wir die Umsetzung unter extremen Bedingungen durchführen und einen vollständigen Umsatz erzielen. Die Folge ist ein sauberer Bruch, durch den ein Fragment ohne verdeckte Schädigung entsteht. Die Banden verschwimmen daher nicht, und so können Fragmente klar korreliert werden, deren Mobilität sich nur wenig unterscheidet. (Schon ein Faktor 10 für die Spezifität genügt, um die Sequenz eindeutig abzulesen.)

Die Technik<sup>[9]</sup> wurde in der Zwischenzeit weiterentwickelt; die Guanin-Reaktion wurde modifiziert und die Dimethylsulfat-Adenosin-Reaktion durch eine direkte Depu-

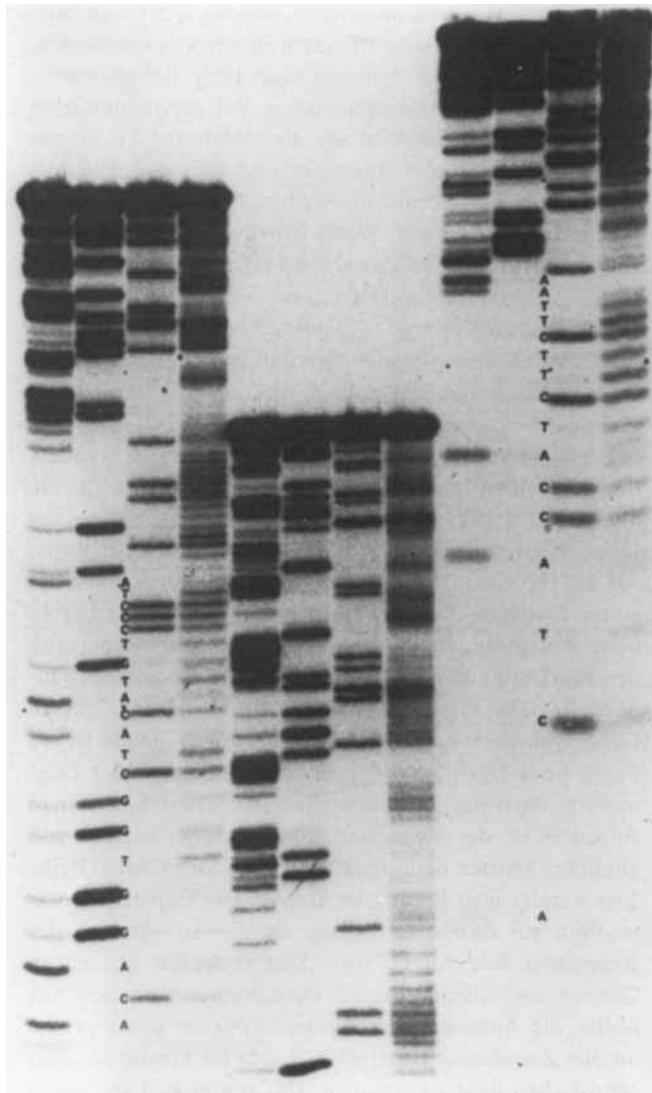


Abb. 5. Sequenzierungsmuster aus der Zeit um 1978. Die Produkte aus vier verschiedenen chemischen Reaktionen von DNA-Fragmenten, bestehend aus etwa 150 Basen, werden elektrophoretisch in einem Polyacrylamid-Gel getrennt; drei Proben ergeben drei Muster-Sets, die verschieden weit durch das Gel gewandert sind. Die jeweils vier Reihen entsprechen den vier Reaktionen, durch die die DNA gespalten wird: 1) hauptsächlich an den Adeninresten, 2) nur an den Guaninresten, 3) an den Cytosin-, aber nicht an den Thyminresten, 4) sowohl an den Cytosin- als auch an den Thyminresten. Die kürzesten Fragmente rufen die Banden unten rechts hervor, und die Sequenz wird im Gel von unten nach oben abgelesen: Als erste ist die Bande in der linken Reihe erkennbar, die einem A entspricht; eine Bande, die in den beiden rechten Reihen auftritt, röhrt von einem C her; eine Bande, die nur in der Reihe außen rechts erscheint, entspricht einem T; eine Bande in der linken Reihe entspricht einem A usw. Nachdem man so weit wie möglich abgelesen hat, wird die Sequenzierung mit den Banden links im Gel und dann mit dem Muster in der Mitte des Gels fortgesetzt. Aus der Photographie kann direkt die gesamte Sequenz des Fragments abgelesen werden. Das Fragment stammt aus dem Teil des Genoms, das die variable Region der leichten  $\lambda$ -Kette des Mäuse-Immunoglobulins codiert [8].

rinierung, die sowohl Adenin- als auch Guaninreste freisetzt, ersetzt. Diese Veränderungen und die Einführung eines sehr dünnen Gels durch die Arbeitsgruppe von Sanger<sup>[10]</sup> ermöglichen gegenwärtig die Sequenzierung von 200-400 Basen vom Markierungspunkt an. Die chemische Aufarbeitung, die Gelelektrophorese und die Autoradiographie sind die weniger zeitaufwendigen Teile des gesamten Prozesses. Die meiste Zeit bei der Sequenzierung von DNA erfordert die Herstellung der DNA-Fragmente und die Ausarbeitung einer Strategie. Die Möglichkeit, einige

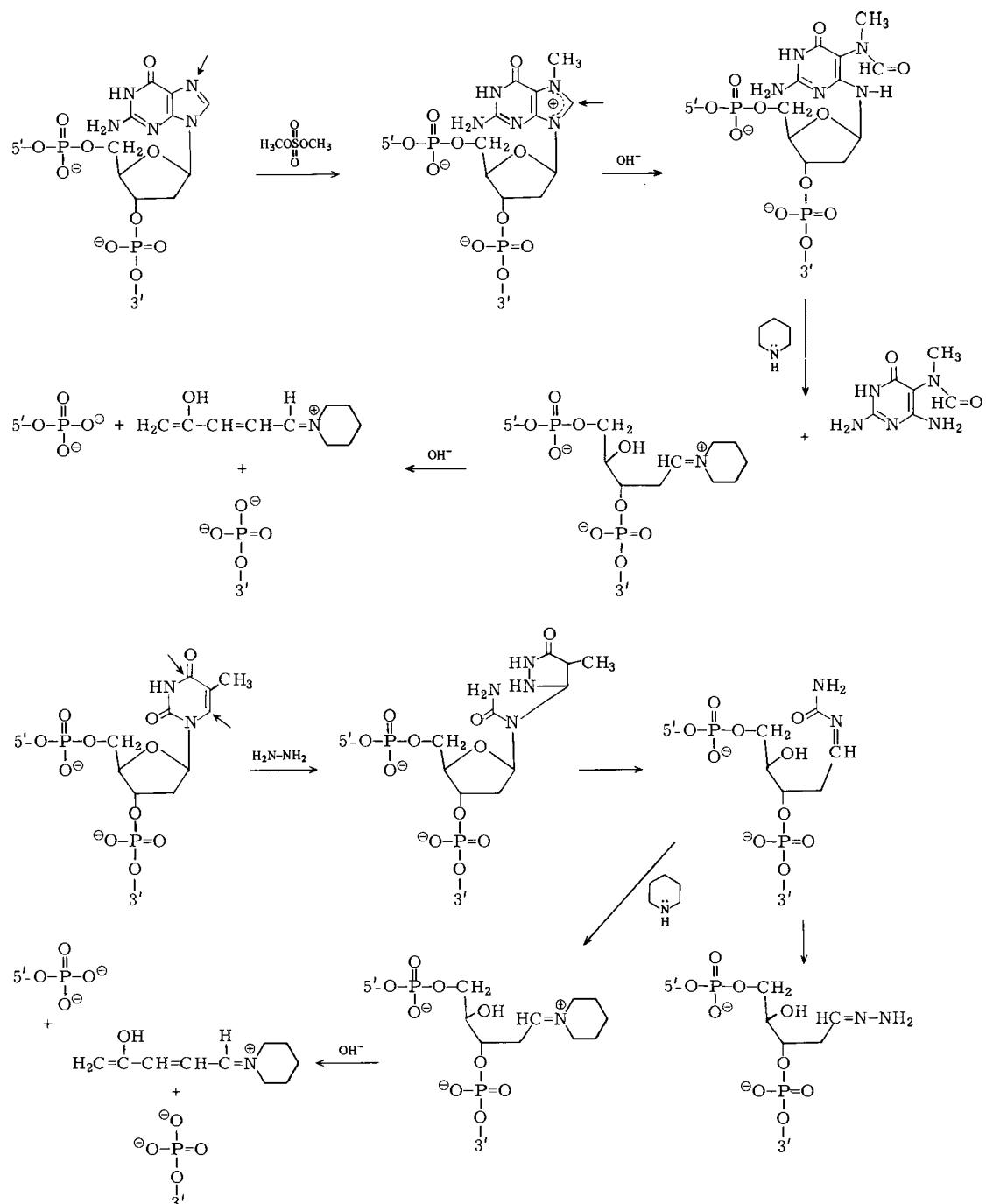


Abb. 6. Beispiele für die DNA-splittenden chemischen Reaktionen. – Oben: Spaltung an Guaninresten. Diese werden erst mit Dimethylsulfat methyliert. Der Imidazolring wird alkalisch geöffnet; durch gleichzeitige Umsetzung mit Piperidin wird die Base verdrängt, und es werden zwei  $\beta$ -Eliminierungen ausgelöst, durch die beide Phosphatreste vom Zucker abgespalten werden, was zum Bruch der DNA führt. Es bleiben ein 3'-Phosphat und ein 5'-Phosphat zurück. – Unten: Angriff des Hydrazins auf einen Thyminrest, wodurch die DNA an den Pyrimidinbasen gespalten wird.

hundert Basen der DNA auf einen Blick abzulesen, ist nur teilweise der Grund für die Schnelligkeit der Sequenzierung. Wichtiger ist, daß ein geradliniges Vorgehen ermöglicht wird: Anstatt zufällige Sequenzen zu bestimmen, kann man an einem Ende der Restriktionskarte beginnen und weiter entlang eines Genes sequenzieren, – oder gar während des Sequenzierens erst die Restriktionskarte aufstellen.

Die erste Bestimmung einer langen Sequenz wurde von einem Studenten, *Phillip Farabaugh*, durchgeführt, der mit den neuen Techniken das Gen für den lac-Repressor sequenzierte<sup>[11]</sup>. Die Aminosäure-Sequenz dieses Gen-Products wurde in den frühen siebziger Jahren von *Beyreut*

*ther et al.*<sup>[12]</sup> erarbeitet. So war es ihm möglich, schnell (in ein paar Monaten) die DNA zu sequenzieren. Die DNA-Sequenz zeigte, daß die Proteinsequenzierung fehlerhaft war; an einer Stelle fehlten zwei, an einer anderen elf Aminosäuren. Da das Protein aus 360 Aminosäuren zusammengesetzt ist, mußte er ein Gen, das aus 1080 Basen besteht, untersuchen. DNA-Sequenzierung ist schneller und exakter als Protein-Sequenzierung, weil die DNA ein linearer Informationsspeicher ist. Da sich alle Restriktionsfragmente chemisch gleich verhalten und sich nur durch ihre Länge unterscheiden, können sie nicht verlorengehen, es sei denn, sie sind sehr klein. Durch ein Sequenzieren auch der überlappenden Teile der DNA-Fragmente kann

eine eindeutige Reihenfolge aufgestellt werden. Proteine hingegen bestehen aus einer Kette von Aminosäuren, die sich chemisch mitunter sehr unterscheiden und von der Natur kombiniert werden, um eine Vielfalt von Strukturen zu schaffen. Wenn ein Protein zerlegt wird, können die Fragmente unterschiedliche Eigenschaften haben; einige sind schwerlöslich, andere sogar vollkommen unauffindbar. Es ist nicht einfach, die Übersicht über den Gesamtgehalt an Aminosäuren und die Reihenfolge der Fragmente zu behalten; bei der DNA kann dagegen die Länge der Restriktionsfragmente ohne Schwierigkeit „gemessen“ werden.

Jeffrey Miller et al. untersuchten sehr genau die Häufigkeit von Mutationen im Gen für den lac-Repressor. An drei Stellen im Gen („hotspots“) treten Mutationen etwa zehnmal so häufig auf wie an anderen Stellen. Durch DNA-Sequenzierung konnte gezeigt werden, daß an jeder der Stellen die modifizierte Base 5-Methylcytosin in der Sequenz vorkommt<sup>[13]</sup>. (Durch chemische Sequenzierung kann die Anwesenheit von 5-Methylcytosin direkt nachgewiesen werden, da diese Base aufgrund der Methylgruppe mit Hydrazin nicht reagiert. Ihr Platz in der Sequenz bleibt frei, und im anderen Strang findet man ein Guanin.) Die hohe Mutationsrate ist durch Umwandlung zu einem Thyminrest bedingt. 5-Methylcytosin taucht selten in der DNA auf, was darauf hindeutet, daß es ein Mutagen ist. Was könnte die Erklärung sein? Desaminierung von Cytosin zu Uracil ist ein natürlicher Prozeß, der, tritt er in der DNA auf, zu einer Mutation führen kann. Dies wird meistens dadurch verhindert, daß die DNA von einem Enzym abgetastet und auf Desoxyuridin geprüft wird<sup>[14]</sup>. Wenn es diese Base in der DNA findet, ob sie an ihren jeweiligen Ort paßt oder nicht, spaltet es die glycosylische Bindung und entfernt das Uracil. Diese Pyrimidin-freie Stelle wird von einem anderen Enzymsystem als Defekt erkannt und repariert. 5-Methylcytosin wird jedoch zu Thymin desaminiert – einem natürlichen Bestandteil der DNA. Nach dem Reparatur- und Resyntheseschritt hat sich eine Umwandlung ergeben. All das erklärt, warum Thymin in der DNA vorkommt: Die zusätzliche Methylgruppe dient dazu, die Folgen der natürlichen Desaminierung, d. h. die Mutationsrate, zu unterdrücken.

Um herauszufinden, wie leicht und genau die DNA-Sequenzierung war, bat ich einen Studenten, *Gregor Sutcliffe*, das Gen für Ampicillin-Resistenz in *E. coli*, das  $\beta$ -Lactamase-Gen, zu sequenzieren. Dieses Gen befindet sich in vielen Plasmiden, einschließlich dem kleinen künstlichen Plasmid pBR322 in *E. coli*. Er kannte nur das ungefähre Molekulargewicht des Proteins und wußte, daß das Gen durch einen bestimmten Restriktionsschnitt im Plasmid inaktiviert würde. Er hatte keine Erfahrung mit der Sequenzierung von DNA als er mit der Analyse dieses Gens begann. Nach sieben Monaten hatte er etwa 1000 Basen der doppelsträngigen DNA identifiziert, indem er erst einen Strang sequenzierte und danach zur Bestätigung des Resultats den anderen. Ein außerordentlich langes Leseraster bestimmte die Sequenz des Proteinproduktes dieses Gens – ein Protein bestehend aus 286 Aminosäuren<sup>[15]</sup>. Wir hielten die DNA-Sequenz für eindeutig. Glücklicherweise standen uns aus dem Labor von *Amblter* Informationen über Teilsequenzen des Proteins zur Verfügung, die dort in jahrelanger Arbeit mit dem Ziel der Sequenzierung der  $\beta$ -

Lactamase erhalten worden waren<sup>[16]</sup>. Diese Information war zwar nicht ausreichend, um die Proteinsequenz direkt zu bestimmen, aber doch hinreichend, um die Voraussage durch die DNA-Sequenzierung zu bestätigen. *Sutcliffe* begeisterte sich sehr für die Sache und sequenzierte den Rest des Plasmides pBR322 innerhalb der folgenden sechs Monate, um seine Doktorarbeit zu beenden. Um sicher zu gehen, sequenzierte er beide Stränge des 4362 Basenpaare umfassenden Plasmids<sup>[17]</sup>. Die chemische Sequenzierung ist in der Regel eindeutig; nur gelegentlich auftretende charakteristische Eigenschaften der DNA-Fragmente selbst verursachen eine anomale Mobilität bei der Gelektrophorese. Je länger die analysierten Stränge werden, desto eher kann sich eine Haarnadelschleife („hairpin-loop“) am Ende des Fragmentes bilden, wenn die Sequenzen hinreichend komplementär sind. Gelangt man bei der Fragmentierung des Moleküls in diesen Teil, dann verringert sich die Wanderungsgeschwindigkeit der Fragmente nicht mehr mit zunehmender Länge, sondern einige Moleküle bewegen sich auch abweichend. Dieses Phänomen wird Kompression genannt, da die Banden auf dem Autoradiogramm näher zusammenrücken oder gar überlappen und so eine oder mehrere Basen unentdeckt bleiben. Dieser seltene Fall tritt im Durchschnitt nach jeder tausendsten Base ein. Um das zu vermeiden, wird der komplementäre Strang in der umgekehrten Richtung sequenziert (oder derselbe Strang in der chemisch umgekehrten Richtung), da dann die Haarnadel in einer anderen Region der Sequenz entsteht und die Kompression daher auch an einer anderen Stelle eintritt. Werden beide Stränge der DNA-Helix sequenziert, dann ist die Sequenz eindeutig.

## Die Struktur von Genen

Die ersten Gene, die sequenziert wurden – sie stammten aus Bakterien –, zeigten die erwartete Struktur: Ihre DNA besteht aus einer kontinuierlichen Reihe von Codons, die zwischen einem Initiationssignal und einem der Terminationssignale liegen. Vor dem Startpunkt der DNA-Transkription befindet sich eine Region mit Sequenzhomologie, die mit der RNA-Polymerase eine Wechselwirkung eingeht; diese Pribnow-Box genannte Sequenz befindet sich eine Helix-Windung vor der Anfangsbase für die messenger-RNA. Eine weitere Wechselwirkung mit der RNA-Polymerase besteht an einer anderen Homologieregion, die 35 Basen vor dem Start liegt. Das bakterielle Gen war also definiert durch seine Bindungsstelle für die RNA-Polymerase und durch weitere Bindungsstellen für Repressoren und Aktivator-Proteine, die in der Nähe der Polymerase-bindenden Region lokalisiert sind. Alternativ könnte die Transkription auch durch Kontrolle der Terminationsfunktionen geregelt werden: Neue Proteine oder eine elegante Translationskontrolle<sup>[18]</sup> könnten entscheiden, ob die Polymerase über ein Stop-Signal hinweg in ein neues Gen weiterliest.

Als durch Rekombinationstechniken die ersten Gene aus Wirbeltieren in Bakterien eingeführt und sequenziert wurden, entdeckte man eine völlig andere Struktur. Die Globin-<sup>[19,20]</sup>, Immunoglobulin-<sup>[21]</sup> und Albumin-<sup>[22]</sup>codierenden Sequenzen bilden keine kontinuierliche Codonreihe, sondern sie werden von großen Strecken nicht codie-

render DNA unterbrochen. Die Entdeckung des RNA-Spleißens in Adenovirus durch Sharp et al.<sup>[23]</sup> und durch Broker und Roberts et al.<sup>[24]</sup> beseitigte noch vorhandene Zweifel an derartigen Strukturen. Sie zeigten, daß nach der Transkription von DNA in RNA diese RNA gespleißt wird: Einige Stücke werden herausgeschnitten, und die verbleibenden Abschnitte werden durch einen noch unbekannten enzymatischen Prozeß wieder verbunden. Die Exons<sup>[25]</sup> – DNA-Regionen, die exprimiert werden – sind voneinander durch Introns – DNA-Regionen, die innerhalb der genetischen Elemente liegen, deren Transkripte aber „herausgespleißt“ werden – getrennt. Abbildung 7

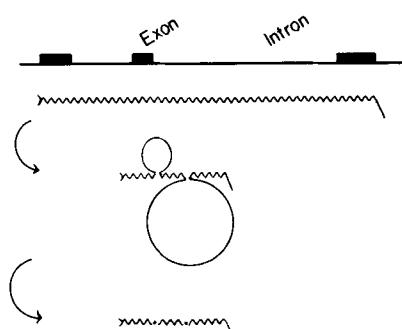


Abb. 7. Eine Transkriptions-Einheit mit alternierenden Exons und Introns. Das ganze Gen, eine Transkriptions-Einheit, wird kopiert; die entstehende RNA endet mit einer poly(A)-Region. Die Teile, die den Introns entsprechen, werden herausgespleißt und hinterlassen eine reife messenger-RNA, die aus drei Exons besteht, d. h. aus den Regionen, die exprimiert werden.

zeigt diesen Vorgang: Das ursprüngliche Transkript eines Gens (heute wird es Transkriptions-Einheit genannt) wird mehrere Male gespleißt, bevor es als reife mRNA im Cytoplasma in Aktion tritt. Abbildung 8 zeigt einige Beispiele. Vertebraten-Gene können viele – 8, 15 oder sogar 50 – Exons<sup>[29,30]</sup> enthalten; es sind meistens kurze codierende DNA-Stücke, die durch mehrere Hundert oder gar mehrere Tausend Basenpaare – die Introns – voneinander getrennt werden. Mit den schnellen Sequenzierungsverfahren können wir zwar die DNA-Sequenz jeder dieser komplexen Gen-Strukturen bestimmen; aber können wir sie auch verstehen?

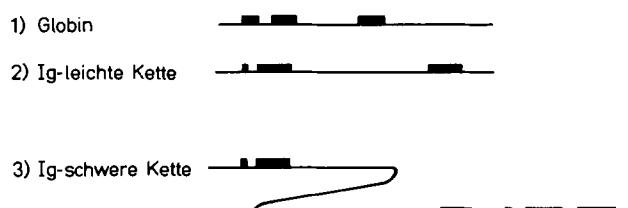


Abb. 8. Beispiele für die Intron-Exon-Struktur einiger Gene. 1) Das Gen für Globin wird durch zwei Introns in drei Exons aufgeteilt [20]. 2) Das funktionelle Gen für die leichte  $\lambda$ -Kette des Immunoglobulins in einer Myelomzelle besteht aus einem kurzen Exon, das der hydrophoben „leader“-Sequenz entspricht, einem Exon für die V-Region und weiter – nach einem Intron aus einigen tausend Basen – einem Exon, das die 112 Aminosäuren der konstanten Region codiert [26]. 3) Ein typisches Gen für die schwere  $\gamma$ -Kette des Immunoglobulins [27, 28]. Das reife Gen besteht aus dem Exon für eine hydrophobe „leader“-Sequenz, einem Exon, das der variablen Region entspricht und – nach einem langen Intron – aus einer Reihe von Exons; von denen codiert das erste die erste Domäne der konstanten Region, das zweite einen 15 Aminosäuren enthaltenden Brückenzug („hinge-Region“), das dritte und vierte die zweite bzw. dritte Domäne der konstanten Region.

Verallgemeinernd kann gesagt werden, daß Gene von Prokaryoten über kontinuierliche codierende Sequenzen verfügen, während die Gene höherer Eukaryoten durch eine komplexe Exon/Intron-Struktur charakterisiert sind. Die einfachsten Eukaryoten, z. B. Hefe, haben wenige Introns, höhere Eukaryoten besitzen Gene, die öfter unterbrochen sind. (Hefe-Mitochondrien verfügen über Introns, sind sie eine Ausnahme?) Erreichen die Intron-Exon-Strukturen immer größere Komplexität, je mehr wir uns den Vertebraten nähern, oder sind die niederen Invertebraten und die Prokaryoten durch den Verlust an präexistierenden Intron-Exon-Strukturen gekennzeichnet? Nach der einen Ansicht ist das Spleißen eine Adaptation, die für höher entwickelte Organismen unumgänglich ist. Nach der anderen Auffassung geht das Spleißen verloren, wenn sich der Organismus vereinfacht und seine DNA möglichst schnell replizieren will, um in kurzer Zeit mehr Generationen zu produzieren; er wäre so unter einem Selektionsdruck gezwungen, seinen Gehalt an DNA zu begrenzen<sup>[31]</sup>.

Welche Rolle kann diese allgemein verbreitete Intron-Exon-Struktur in den Genen höherer Organismen spielen? Obwohl die meisten der untersuchten Gene diese Struktur aufweisen, gibt es zwei bemerkenswerte Ausnahmen: die Gene für die Histone und die für die Interferone. Das deutet darauf hin, daß die Introns nicht essentiell sind, und daß das Spleißen nicht unbedingt erforderlich ist, um ein Gen in Säugetierzellen zu exprimieren. Obwohl Versuchsergebnisse darauf hinweisen, daß einige messenger-RNAs mindestens einmal gespleißt werden müssen, um exprimiert zu werden, gibt es keinen Beweis dafür, daß mehrmaliges Spleißen immer notwendig ist. In der Ratte gibt es zwei Gene für Insulin, die sich durch ihre Anzahl an Introns unterscheiden; beide werden exprimiert – ein Zeichen dafür, daß das Intron, welches die codierende Region des einen Gens teilt, keine essentielle Rolle „in cis“ bei der Expression des Gens spielt<sup>[32]</sup>. Trotz der allgemeinen Vermutung, daß das Spleißen eine regulatorische Funktion ausübt, wurde bis heute kein gewebeabhängiges Spleiß-Muster identifiziert, aus dem auf die Existenz eines Gen- (oder Gewebe-)spezifischen Spleiß-Enzyms hätte geschlossen werden können.

Die Introns sind viel länger als die Exons. Ihre DNA-Sequenz verändert sich rapide durch Mutationen, kleine Additionen und Deletionen (die Veränderungen gehen sehr schnell vor sich, mit der gleichen Rate wie die „silent changes“ in den Codons). Es scheint, daß nicht ihre Sequenz relevant ist, sondern einfach ihre Länge. Ihre Funktion liegt darin, die Exons auf dem Chromosom voneinander zu trennen.

Als Folge dieser Trennung von Exons durch Introns kommt es zu einer größeren Zahl von legitimen oder illegitimen Rekombinationen pro Zeiteinheit zwischen den Exons<sup>[25]</sup>. Im Zeitmaß der Evolution wird dadurch die Häufigkeit erhöht, mit der die Exons – sie bestimmen letztlich die Proteinstruktur – rearrangiert und neukombiniert werden. Man betrachte z. B. den Prozeß, durch den eine Domäne verdoppelt wird, um die Zwei-Domänen-Struktur der leichten Kette des Immunoglobulins zu bilden (wird sie nochmals verdoppelt, dann entsteht die Vier-Domänen-Struktur der schweren Kette; die Kombination beider Prozesse kann zur Dreifach-Struktur des Ovomucoids führen).

ren<sup>[29]</sup>). Nach klassischer Vorstellung geschieht das durch ein ungleiches „crossing over“, durch das die beiden Kopien des ursprünglichen Gens verbunden werden und danach ein Gen mit doppelter Länge bilden. Dieser Prozeß setzt voraus, daß der extrem seltene Fall einer präzisen illegitimen Rekombination eintritt (eine Rekombination, die zur Fusion von zwei DNA-Sequenzen an einer Stelle führt, an der die Sequenzen nicht übereinstimmen), deren Folge die Synthese des neuen und möglicherweise nützlicheren Gen-Produktes mit doppelter Länge ist; Abbildung 9 veranschaulicht diesen Prozeß. Wird der gleiche Vorgang unter der Annahme eines allgemeinen Spleiß-Mechanismus

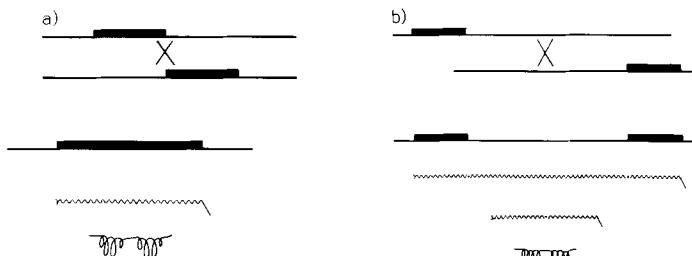


Abb. 9. Ein Gen-Produkt mit doppelter Länge entsteht durch ungleiches crossing over. a) Der klassische Vorgang, durch den ein Gen - es codiert eine einzige Polypeptidkette - seine Länge durch ein crossing over verdoppeln kann. Die beiden oberen Konturen stellen zwei codierende Regionen dar, die durch eine illegitime Rekombination zufällig nebeneinander zu liegen kommen; dadurch wird das Carboxy-Ende (3'-Ende) der einen Kopie mit dem Amino-Ende (5'-Ende) der anderen Kopie verbunden. Dieser seltene Fall einer illegitimen Rekombination (es gibt keine Sequenzen, die übereinstimmen) hätte, trüte er phasengleich ein, ein Gen mit doppelter Länge zur Folge; dieses könnte eine RNA doppelter Länge codieren, welche wiederum in ein Protein doppelter Länge mit sich wiederholender Domäne übersetzt würde. b) Der gleiche Vorgang mit der Möglichkeit des Spleißens. Hier kann das ungleiche crossing over überall hinter dem 3'-Ende der einen Gen-Kopie und überall vor dem 5'-Ende der anderen Gen-Kopie eintreten; die Folge wäre ein Gen, das zwei Exons - getrennt durch ein langes Intron - enthielte. Ich vermute, daß das lange Transkript dieser Region mit niedriger Frequenz gespleißt wird und eine reife messenger-RNA hinterläßt, die das verdoppelte Protein codiert.

betrachtet, so muß auch hier die Bildung eines doppelten Gens durch eine illegitime Rekombination eingeleitet werden; aber dieses Mal kann das überall innerhalb einer 1000 bis 10000 Basen langen Region eintreten, wenn diese an das 3'-Ende einer Kopie und an das 5'-Ende der anderen Kopie anschließt und das Intron bildet, das die Gene für die beiden Domänen trennt. Selbst ein ungenügendes Spleißen des Transkriptes dieser Region führt zum Gen-Produkt mit doppelter Länge. Dieser Prozeß verläuft 10<sup>6</sup>- bis 10<sup>8</sup> mal schneller als der klassische, da die Stellen, an denen die Rekombination stattfinden kann, vielfältig kombinierbar sind. Wird das lange Transkript gespleißt, kann - selbst bei niedriger Frequenz - ein Teil des Gen-Produktes synthetisiert werden. Die Evolution findet so einen schnelleren Weg zur Bildung des endgültigen Gens: Ein schneller Schritt führt zu einer Struktur, die einen kleinen Teil des brauchbaren Gen-Produktes liefert. Durch kleine Mutationen werden bessere Spleiß-Signale gebildet und damit mehr Gen-Produkt. Wenn die Spleiß-Signale schon existieren, dann können durch Rekombination innerhalb der Introns polymere Strukturen aus einfachen Einheiten direkt gebildet werden. Danach wäre vorauszusagen, daß polymere Strukturen, die aus einfachen Einheiten zusammengesetzt sind, durch Gene codiert werden, in denen die Intron-Exon-Struktur der einfachen Einheiten wiederholt vorkommt, und diese Strukturen wiederum selbst durch Introns getrennt sind. Das ist in der Tat der Fall.

Die Häufigkeit der legitimen Rekombination zwischen den Exons eines Gens wird durch das Vorhandensein von Introns erhöht. Man stelle sich zwei Mutationen vor, die zu einer verbesserten Funktion führen, in verschiedenen Gen-Teilen entstehen und sich durch Selektion in einer Population ausbreiten. Durch homologe Rekombination innerhalb eines Gens können dann - nach klassischer Ansicht - beide Mutationen in einer Polypeptidkette austauschen - vorausgesetzt, sie befinden sich in einem diploiden Organismus. Abbildung 10 zeigt, daß dieser Prozeß auf das 10- bis 100fache beschleunigt werden kann, wenn die Exons weiter voneinander entfernt liegen. Der Effekt ist am stärksten, wenn die Exons sich getrennt entwickeln können - wenn sie Strukturen repräsentieren, die erfolgreiche Veränderungen unabhängig voneinander akkumulieren.

Das Exon-Muster läßt sich verändern, indem der Initiations- oder der Terminationsort für die RNA-Transkription verlegt wird; auf diese Weise werden zusätzliche Exons angefügt oder Exons aus einer DNA-Region mit Exons aus einer anderen Region zusammengezogen. Das wurde im Adenovirus beobachtet und findet sich bei einigen bemerkenswerten Immunoglobulinen, bei denen Exons an den dem Carboxy-Ende der schweren Kette entsprechenden Genort angefügt oder entfernt werden können, um das Protein zu modifizieren. Untersuchungen im Labor von *Hood* haben gezeigt, daß durch diesen Prozeß zwischen zwei Formen der schweren Kette eines IgM gewechselt werden kann<sup>[33]</sup>. Eine membrangebundene Form wird durch ein längeres Transkript synthetisiert, dem zwei zusätzliche Exons „angespleißt“ und ein Teil des letzten Exons des kürzeren Transkriptes „abgespleißt“ werden.

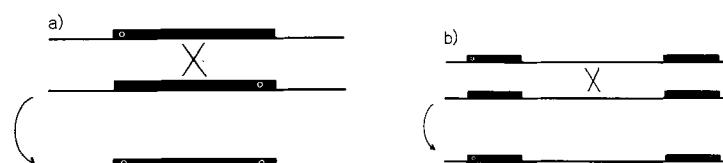


Abb. 10. Introns beschleunigen legitime Rekombinationen. a) Der klassische Vorgang, bei dem zwei Mutationen - eine in der einen Gen-Kopie nahe dem linken Ende, die andere in der anderen Gen-Kopie nahe dem rechten Ende - durch Rekombination innerhalb des homologen DNA-Stückes, das beide Mutationen trennt, zusammenkommen. Durch diese Rekombination kann ein Gen entstehen, das beide Mutationen trägt. b) Der gleiche Vorgang in einem Gen, in dem die Mutationen in verschiedenen Exons - sie sind durch ein Intron getrennt - auftreten. Hier kann die Rekombination überall eintreten, sowohl innerhalb des Exons, als auch innerhalb des Introns, und ein neues Gen, das beide Mutationen trägt, entsteht. Da die Rekombinationshäufigkeit direkt proportional der Länge der DNA-Strecke zwischen den Mutationen ist, ist dieser Vorgang schneller.

Durch das kürzere Transkript wird eine Form des Proteins synthetisiert, die abgesondert wird. In ähnlicher Weise ist der Wechsel von der V-Region des IgM zu einer konstanten Region des IgD wahrscheinlich die Folge eines anderen, noch längeren Transkriptes, bei dem die Exons der V-Region an die Exons der neuen konstanten Region der  $\delta$ -Klasse angefügt werden. Diese Gen-Kombinationen wurden sicherlich durch Rekombinationen in der DNA-Region geschaffen, die letztlich zum Intron des längeren Transkriptes wurde.

Die verblüffendste Voraussage bezüglich evolutionärer Prozesse ist danach, daß die durch Exons definierten ein-

zellen Elemente funktionelle Bedeutung haben, daß sie gruppiert und zu neuen Kombinationen zusammengefügt wurden, um die uns bekannten Proteine zu bilden. Aus früher gefundenen Lösungen des Struktur-Funktion-Problems wurden bestimmte Gen-Produkte ausgewählt. Deutliche Beispiele dafür sind noch immer rar. Die hydrophobe „leader“-Sequenz, die den Transfer von Proteinen durch Membranen erlaubt, und die nach der Sekretion abgespalten wird, ist oft in besonderen Exons codiert – am auffälligsten bei den Immunoglobulinen (siehe Abb. 8), aber auch beim Ovomucoid<sup>[29]</sup>. In der Ratte sind in den beiden Genen für Insulin – Produkte einer „jungen“ Duplikation<sup>[32]</sup> – die beiden Ketten des Insulins in einem Gen in verschiedenen Exons und im anderen Gen in nur einem Exon codiert. Das Ur-Gen (die in anderen Spezies übliche

Struktur<sup>[34]</sup>) verfügt über das zusätzliche Intron; das läßt vermuten, daß das Gen aus ursprünglich voneinander getrennten Stücken zusammengesetzt wurde. Das Gen für Lysozym ist auf vier Exons aufgeteilt; das zweite trägt die Information für die entscheidenden Aminosäuren des aktiven Zentrums und für die meisten Stellen, die mit dem Substrat Kontakt haben<sup>[35]</sup>. Im Gen für Globin codiert das zentrale Exon fast alle Stellen, die mit dem Häm in Kontakt treten. Abbildung 11 zeigt das Molekül schematisch. Kürzlich konnte durch einen Versuch<sup>[36]</sup> gezeigt werden, daß die Polypeptidkette, die dem zentralen Exon entspricht, ein Häm-bindendes „Mini-Globin“ ist; die seitlichen Exons codieren Polypeptid-Material, das das Protein stabilisiert.

In dem Moment, in dem es möglich schien, mit den schnellen Sequenzierungsmethoden und dem molekularen Clonieren die Struktur eines jeden Gens aufzuklären und damit das genetische Material zu verstehen, wurde klar, daß die Natur komplexer ist, als wir annahmen. Wir konnten das Gen-Produkt nicht direkt vom Chromosom einzig durch DNA-Sequenzierung ablesen. Wir mußten auf die Sequenz des Proteins oder auf die der reifen messenger-RNA zurückgreifen, um die Intron-Exon-Struktur des Gens zu erkennen. Trotzdem besteht die Hoffnung, daß wir, wenn wir die DNA-Sequenz des Chromosoms kennen, nicht nur etwas über die Primärstruktur der Gen-Produkte erfahren, sondern auch etwas über die funktionelle Struktur der Proteine; diese sind die Gen-Produkte von im Verlauf der Evolution zusammengesetzten Exons, die durch Introns verbunden sind.

Mein Interesse an der Biologie kreiste immer um zwei Probleme: Wie manifestiert sich die genetische Information, und wodurch wird sie kontrolliert? Wir wissen viel über die Art und Weise, in der ein Gen in ein Protein übersetzt wird. Zahlreiche Aspekte der Gen-Kontrolle in Prokaryoten sind bekannt, aber die entscheidenden Kontrollmechanismen in Eukaryoten liegen noch immer im Dunkeln. Das Ziel der Forschung ist die Untersuchung des Unbekannten. Das Streben nach neuen Erkenntnissen verlangt Antworten auf neue Fragen.

*Viel verdanke ich meinen Studenten und Mitarbeitern, besonders aber Jim Watson, der mein Interesse an der Molekularbiologie weckte, Benno Müller-Hill, mit dem ich am lac-Repressor arbeite und Allan Maxam, mit dem ich die Methode zur Sequenzierung von DNA entwickelte.*

Eingegangen am 14. April 1981 [A 383]  
Übersetzt von Christiane Kozska, Hamburg

- [1] W. Gilbert, B. Müller-Hill: Isolation of the Lac Repressor, Proc. Natl. Acad. Sci. USA 56, 1891 (1966).
- [2] W. Gilbert, B. Müller-Hill: The Lac Operator is DNA, Proc. Natl. Acad. Sci. USA 58, 2415 (1967).
- [3] W. Gilbert, A. M. Maxam: The Nucleotide Sequence of the lac Operator, Proc. Natl. Acad. Sci. USA 70, 3581 (1973).
- [4] W. Gilbert, N. Maizels, A. M. Maxam: Sequences of Controlling Regions of the *E. coli* Lactose Operator, Genetics 79, 227 (1975).
- [5] R. Dickson, J. Abelson, W. Barnes, W. Reznikoff: Genetic Regulation: the Lac Control Region, Science 187, 27 (1975).
- [6] W. Gilbert, A. M. Maxam, A. Mirzabekov: Contacts Between the lac Repressor and DNA Revealed by Methylation in *N. O. Kjelgaard, O. Maaløe*: Control of Ribosome Synthesis, Alfred Benzon Symp. IX, Munksgaard, Kopenhagen 1976, S. 139.
- [7] A. M. Maxam, W. Gilbert: A New Method for Sequencing DNA, Proc. Natl. Acad. Sci. USA 74, 560 (1977).

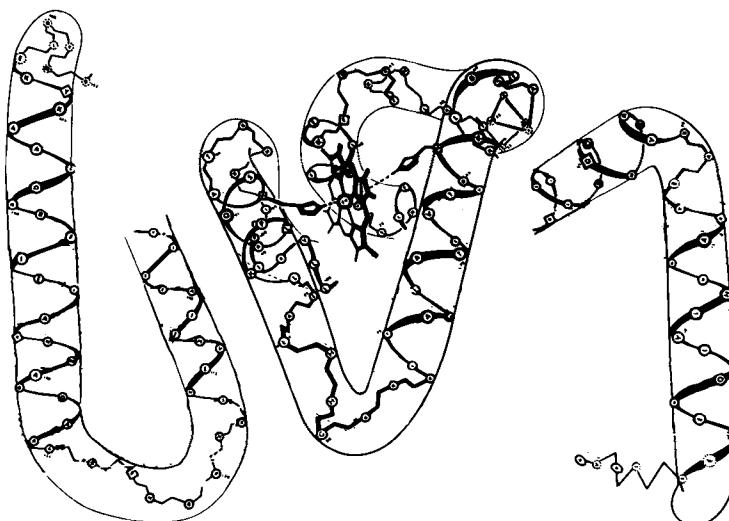
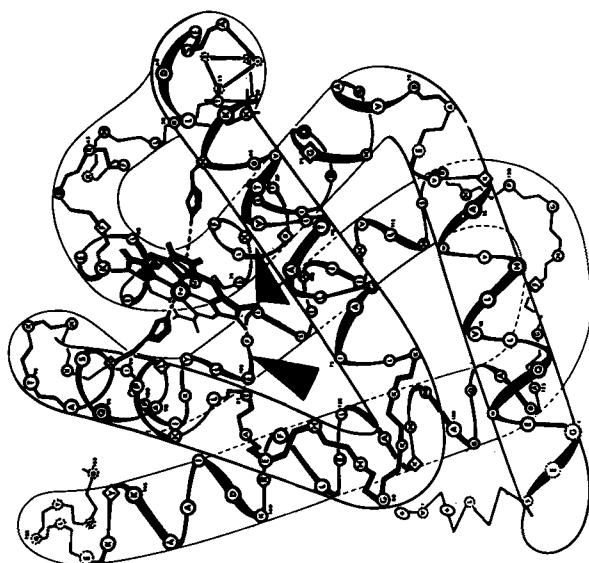


Abb. 11. Eine schematische Unterteilung von Globin in die Produkte der einzelnen Exons. Oben zeigen die schwarzen Pfeile die Punkte, an denen die Struktur einer Globin-Kette von Introns unterbrochen wird. (Die Strukturen der Globin- und Myoglobinketten sind sich sehr ähnlich; die abgebildete schematische Struktur ist die des Myoglobins.) Die Introns unterbrechen das Protein in den  $\alpha$ -helikalen Regionen. Unten: Die Introns teilen das Protein in drei Stücke. Das Produkt des zentralen Exons umgibt das Häm; ich vermute, daß die Produkte der anderen beiden Exons den zentralen Proteinteil einhüllen und stabilisieren.

- [8] S. Tonegawa, A. M. Maxam, R. Tizard, O. Bernard, W. Gilbert: Sequence of a Mouse Germ-line Gene for a Variable Region of an Immunoglobulin Light Chain, Proc. Natl. Acad. Sci. USA 75, 1485 (1978).
- [9] A. M. Maxam, W. Gilbert: Sequencing End-Labeled DNA with Base-Specific Chemical Cleavages, Meth. Enzym. 65, 499 (1980).
- [10] F. Sanger, A. R. Coulson: The Use of Thin Acrylamide Gels for DNA Sequencing, FEBS Lett. 87, 107 (1978).
- [11] Ph. J. Farabaugh: Sequence of the lacI Gene, Nature 274, 765 (1978).
- [12] K. Beyreuther, K. Adler, E. Fanning, C. Murray, A. Klemm, N. Geisler: Amino-Acid Sequence of lac Repressor from *Escherichia coli*, Eur. J. Biochem. 59, 491 (1975).
- [13] Ch. Coulondre, J. H. Miller, Ph. J. Farabaugh, W. Gilbert: Molecular Basis of Base Substitution Hotspots in *Escherichia coli*, Nature 274, 775 (1978).
- [14] T. Lindahl, S. Ljungquist, W. Siegert, B. Nyberg, B. Sperers: DNA N-Glycosidases, J. Biol. Chem. 252, 3286 (1977).
- [15] J. G. Sutcliffe: Nucleotide Sequence of the Ampicillin Resistance Gene of *Escherichia coli* Plasmid pBR322, Proc. Natl. Acad. Sci. USA 75, 3737 (1978).
- [16] R. P. Ambler, G. K. Scott: The Partial Amino Acid Sequence of the Penicillinase Coded by the *Escherichia coli* Plasmid R6K, Proc. Natl. Acad. Sci. USA 75, 3732 (1978).
- [17] J. G. Sutcliffe: Complete Nucleotide Sequence of the *Escherichia coli* Plasmid pBR322, Cold Spring Harbor Symp. 43, 77 (1978).
- [18] Übersicht: C. Yanofsky: Attenuation in the Control of Expression of Bacterial Operons, Nature 289, 751 (1981).
- [19] S. M. Tilghman, D. C. Tiemeier, J. G. Seidman, B. M. Peterlin, M. Sullivan, J. V. Maizel, P. Leder: Intervening Sequence of DNA Identified in the Structural Portion of a Mouse Beta-Globin Gene, Proc. Natl. Sci. USA 75, 725 (1978).
- [20] D. A. Konkel, S. M. Tilghman, P. Leder: The Sequence of the Chromosomal Mouse Beta-Globin Major Gene, Cell 15, 1125 (1978).
- [21] C. Brack, S. Tonegawa: Variable and Constant Parts of the Immunoglobulin Light Chain of a Mouse Myeloma Cell are 1250 Nontranslated Bases Apart, Proc. Natl. Acad. Sci. USA 74, 5652 (1977).
- [22] R. Breathnach, J. L. Mandel, P. Chambon: Ovalbumin Gene is Split in Chicken DNA, Nature 270, 314 (1977).
- [23] S. M. Berget, C. Moore, Ph. A. Sharp: Spliced Segments at the 5' Terminal of Adenovirus 2 Late mRNA, Proc. Natl. Acad. Sci. USA 74, 3171 (1977).
- [24] L. T. Chow, R. E. Gelinas, T. R. Broker, R. J. Roberts: An Amazing Sequence Arrangement at the 5' Ends of Adenovirus 2 Messenger RNA, Cell 12, 1 (1977).
- [25] W. Gilbert: Why Genes in Pieces?, Nature 271, 501 (1978).
- [26] O. Bernard, N. Hozumi, S. Tonegawa: Sequences of Mouse Immunoglobulin Light Chain Genes Before and After Somatic Changes, Cell 15, 1133 (1978).
- [27] H. Sakano, J. H. Rogers, K. Hüppi, C. Brack, A. Traunecker, R. Maki, R. Wall, S. Tonegawa: Domains and the Hinge Region of an Immunoglobulin Heavy Chain Are Encoded in Separate DNA Segments, Nature 277, 627 (1979).
- [28] T. Honjo, M. Obata, Y. Yamawaki-Kataoka, T. Kataoka, T. Kawakami, N. Takahashi, Y. Mano: Cloning and Complete Nucleotide Sequence of Mouse Gamma 1 Chain Gene, Cell 18, 559 (1979).
- [29] J. P. Stein, J. F. Catterall, P. Kristo, A. R. Means, B. W. O'Malley: Ovomucoid Intervening Sequences Specify Functional Domains and Generate Protein Polymorphisms, Cell 21, 681 (1980).
- [30] Y. Yamada, V. E. Avvedimento, M. Mudryj, H. Ohkubo, G. Vogeli, M. Irani, I. Pastan, B. de Crombrugghe: The Collagen Gene: Evidence for its Evolutionary Assembly by Amplification of a DNA segment Containing an Exon of 54 bp., Cell 22, 887 (1980).
- [31] W. F. Doolittle: Genes in Pieces: Were They Ever Together?, Nature 272, 581 (1978).
- [32] P. Lomedico, N. Rosenthal, A. Efstratiadis, W. Gilbert, R. Kolodner, R. Tizard: The Structure and Evolution of the Two Non-Allelic Rat Preproinsulin Genes, Cell 18, 545 (1979).
- [33] P. Early, F. Rogers, M. Davis, K. Calami, M. Bond, R. Wall, L. Hood: Two mRNA's Can be Produced from a Single Immunoglobulin Gene by Alternative RNA Processing Pathways, Cell 20, 313 (1980).
- [34] F. Perler, A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, J. Dodgson: The Evolution of Genes: The Chicken Preproinsulin Gene, Cell 20, 555 (1980).
- [35] A. Jung, A. E. Sippel, M. Grez, G. Schütz: Exons Encode Functional and Structural Units of Chicken Lysozyme, Proc. Natl. Acad. Sci. USA 77, 5759 (1980).
- [36] C. S. Craik, S. R. Buchman, S. Beychok: Characterization of Globin Domains: Heme Binding to the Central Exon Product, Proc. Natl. Acad. Sci. USA 77, 1384 (1980).

## Pyramide Carbokationen<sup>\*\*</sup>

Von Helmut Schwarz<sup>†</sup>

Professor Ferdinand Bohlmann zum 60. Geburtstag gewidmet

Carbokationen mit pyramidenförmiger Struktur können als die Bindeglieder zwischen der Organischen und der Anorganischen Chemie angesehen werden. Aus Vorstellungen vom elektronischen Aufbau dieser Spezies werden Konsequenzen für Struktur, Ladungsverteilung und Reaktivität (Afsangreaktionen mit Nucleophilen) abgeleitet. Phänomene der Gasphasenchemie von Kationen, z. B. das Kohlenstoff-„Scrambling“ bei Carbenium-Ionen, lassen sich durch pyramidal strukturierte Zwischenstufen oder Übergangszustände erklären. Die Möglichkeit von Übergangszuständen mit  $H_2$  als „side-on“- oder „end-on“-Liganden und ihre Bedeutung für den Mechanismus unimolekularer Prozesse werden diskutiert.

### 1. Einleitung

Sucht man nach Bindegliedern zwischen der Molekülwelt der Organischen und der der Metallorganischen Chemie, so fällt das Augenmerk auf Carbokationen<sup>[1]</sup>: Als Koh-

lenwasserstoffe gehören sie zweifellos zur Organischen Chemie; Bindungseigenschaften und strukturelle Aspekte – darunter nicht zuletzt die Koordinationszahl des Kohlenstoffs – lassen jedoch oft eine beträchtliche Verwandtschaft zu typischen Organometall-Verbindungen erkennen. Bei organischen Molekülen ist die Koordinationszahl normalerweise nicht größer als vier, bei Carbaboranen (oder allgemein Cluster-Verbindungen<sup>[2]</sup>) sind fünf oder sechs Liganden nicht ungewöhnlich, und Carbenium-Ionen sind per definitionem mindestens fünfzehn koordiniert. Der Gedanke, ein C-Atom könne auch fünfzehn koordiniert sein,

[†] Prof. Dr. H. Schwarz

Institut für Organische Chemie der Technischen Universität  
Straße des 17. Juni 135, D-1000 Berlin 12

[\*\*] Nach einem Vortrag bei der Verleihung des Otto-Klung-Preises für Chemie an der Freien Universität Berlin im Januar 1981.