

Bestimmung von Nucleotidsequenzen der DNA (Nobel-Vortrag)^[**]

Von Frederick Sanger^[*]

1. Einleitung^[***]

Obwohl DNA-Sequenzen in der belebten Natur eine wichtige Rolle spielen, gibt es erst seit kurzer Zeit allgemeine Methoden zu ihrer Bestimmung. Das beruht hauptsächlich auf der enormen Größe der DNA-Moleküle; die kleinsten sind die einfacher Bakteriophagen wie Φ X174, dessen DNA etwa 5000 Basenpaare enthält. Es war schwierig, Methoden für solche komplizierten Systeme zu entwickeln. Jedoch gibt es relativ kleine RNA-Moleküle wie die transfer-RNAs mit nur etwa 75 Nucleotiden; an ihnen wurden die ersten Bestimmungen von Nucleinsäuresequenzen durchgeführt^[1].

Nach meiner Arbeit an Aminosäuresequenzen in Proteinen^[2] richtete ich mein Interesse auf RNA und entwickelte in Zusammenarbeit mit G. G. Brownlee und B. G. Barrell eine relativ schnelle Methode, mit der in kleinem Maßstab ³²P-markierte Oligonucleotide fraktioniert werden konnten^[3]. Spätere Bestimmungen von RNA-Sequenzen basierten darauf. Die Methode in diesen Untersuchungen wie in jenen über Proteine beruhte auf dem Prinzip der partiellen Zerlegung. Aus großen Molekülen wurden – meistens durch Enzyme – kleinere Bruchstücke gewonnen, diese voneinander getrennt und ihre Sequenz bestimmt. Die erhaltene Sequenz der Bruchstücke ließ auf die vollständige Sequenz schließen. Da dieses Verfahren viele aufeinanderfolgende Zerlegungs- und Fraktionierungsschritte benötigte, war es notwendigerweise langsam und umständlich und nicht leicht auf größere DNA-Moleküle anwendbar. In er-

sten Untersuchungen konnten wir so Sequenzen von 50 Nucleotiden bestimmen^[4,5]; um aber genetisches Material zu sequenzieren, schien eine andere Methode wünschenswert. Wir begannen, unser Augenmerk auf Kopier-Verfahren (copying procedures) zu richten.

2. Kopier-Verfahren

Diese Verfahren wurden zur Untersuchung von RNA von C. Weissmann et al.^[6] während ihrer Arbeit an der RNA-Sequenz des Bakteriophagen Q β entwickelt. Q β enthält eine Replikase, die – am 3'-Ende beginnend – eine komplementäre Kopie einer einzelsträngigen RNA synthetisieren kann. Diese Gruppe arbeitete elegante Verfahren aus; so wurde unter anderem Puls-Markierung mit radioaktiv markierten Nucleotiden verwendet, aus denen dann Sequenzen abgeleitet werden konnten.

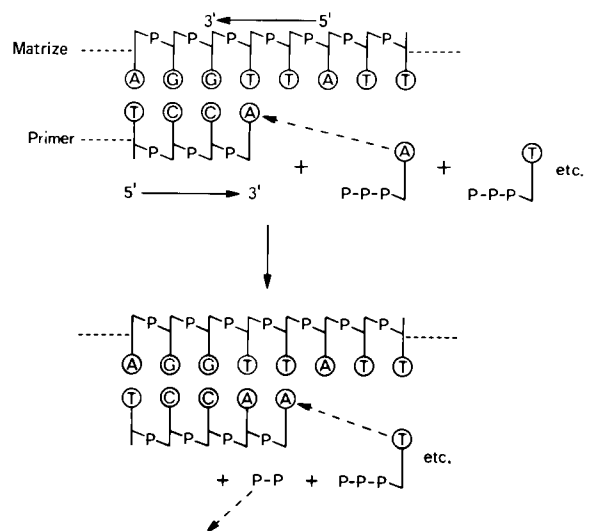


Abb. 1. Spezifitätsbedingungen der DNA-Polymerase.

[*] Prof. Dr. F. Sanger
Medical Research Council Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH (England)

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

[***] Abkürzungen: Die Abkürzungen C, A und G werden – dem jeweiligen Zusammenhang entsprechend – sowohl für Ribonucleotide als auch für Desoxyribonucleotide verwendet.

Wir benutzten für DNA-Sequenzen das Enzym DNA-Polymerase, das Einzelstrang-DNA kopiert (Abb. 1). Dieses Enzym benötigt einen Primer, d. h. ein einzelsträngiges Oligonucleotid, das zu einer Region des zu sequenzierenden DNA-Teils (der Matrize) komplementär ist und folglich mit ihm hybridisieren kann. An das 3'-Ende des Primers werden nacheinander Mononucleotidreste der entsprechenden Desoxynucleosidtriphosphate geknüpft, so daß eine komplementäre Kopie der Matrizen-DNA entsteht. Durch Verwendung von Triphosphaten mit ^{32}P in der α -Position kann die neusynthetisierte DNA markiert werden. In den ersten Experimenten dienten synthetische Oligonucleotide als Primer. Nach der Entdeckung von Restriktionsenzymen erschien es jedoch zweckmäßig, die von ihnen gelieferten, leicht zugänglichen Fragmente zu benutzen.

Anfänglich wurde das Kopier-Verfahren angewendet, um kurze spezifische Stücke markierter DNA zu synthetisieren, die dann partiell zerlegt werden konnten. Eine Schwierigkeit bei der Sequenzierung von DNA war, geeignete Methoden zu deren Zerlegung in kleine Fragmente zu finden. Es war kein Enzym bekannt, das nur ein Nucleotid erkennt. *Berg, Fanher* und *Chamberlin*^[7] hatten jedoch schon früher gezeigt, daß die DNA-Polymerase unter bestimmten Umständen auch Ribonucleotide anstelle der normalen Desoxyribonucleotide in DNA-Ketten einbauen kann. Wurden beim Kopieren z. B. *ribo*-CTP und die anderen drei Desoxynucleosidtriphosphate benutzt, konnten Ketten synthetisiert werden, deren Cytidin-Reste in der *ribo*-Form vorlagen. Unter definierten Bedingungen konnten dann mit Alkali die Bindungen gebrochen werden, die von Ribonucleotiden ausgingen; die Bindungen der Desoxynucleotide blieben hingegen unversehrt. Auf diese Weise war ein spezifisches Spalten an Cytidin-Resten möglich. Mit dieser Methode konnten wir unsere Sequenzuntersuchungen ausweiten^[8]. Trotzdem waren weiterhin vollständige Fraktionierungs- und Analysemethoden nötig.

3. „Plus-und-Minus“-Methode

Für diese Experimente benötigten wir DNA-Kopien mit hoher spezifischer Radioaktivität; um sie herzustellen, mußten wir die hoch markierten Substrate in niedrigen Konzentrationen verwenden. Wurde mit α - ^{32}P -dATP markiert, so wurde es in viel niedrigerer Konzentration zugegeben als die anderen drei Triphosphate. Analysierten wir die neu synthetisierten DNA-Ketten, fanden wir, daß sie oft an einer Position endeten, nach der ein A hätte eingebaut werden müssen. Folglich erhielten wir eine Mischung aus Produkten, die alle das gleiche 5'-Ende hatten (das 5'-Ende des Primers) und am 3'-Ende an der Position eines A-Restes endeten. Könnten diese Produkte nun aufgrund ihrer unterschiedlichen Kettenlänge fraktioniert werden, so sollte das dabei erhaltene Verteilungsmuster proportional zur Verteilung von A entlang der DNA-Kette sein. Die Kenntnis dieser Verteilung und die der anderen drei Mononucleotide ermöglichen eine Sequenzbestimmung. Die ersten, mit *J. E. Donelson* durchgeführten Experimente zeigten, daß dieser Weg die Grundlage für eine Methode zur schnellen Sequenzanalyse bildete. Zugleich erwies sich

die Ionophorese auf Acrylamid-Gel als gut geeignet zur Fraktionierung der DNA-Fragmente nach ihrer Molekülgröße.

Die oben beschriebene Methode brachte nur begrenzten Erfolg; wir konnten jedoch zwei auf demselben Prinzip beruhende, modifizierte Techniken erarbeiten, die eine schnellere und einfachere DNA-Sequenzierung ermöglichen^[9]. Mit der als „Plus-und-Minus“-Methode bezeichneten Technik wurde fast die vollständige, aus 5386 Nucleotiden bestehende Sequenz der DNA des Bakteriophagen ΦX174 bestimmt^[10].

4. „Didesoxy“- oder „Kettenabbruch“-Methode

Später entwickelten wir eine ähnliche Methode, die schneller und exakter als die „Plus-und-Minus“-Technik ist, und bei der spezifische, kettenbeendende Analoga der normalen Desoxynucleosidtriphosphate verwendet werden^[11]; sie ermöglichte die vollständige Sequenzierung von ΦX174 ^[12] und von G4^[13], einem verwandten Bakteriophagen. Kürzlich wurde sie auf mitochondriale Säugetier-DNA angewendet.

Die am häufigsten benutzten Analoga sind die Didesoxynucleosidtriphosphate (Abb. 2). Sie gleichen den nor-

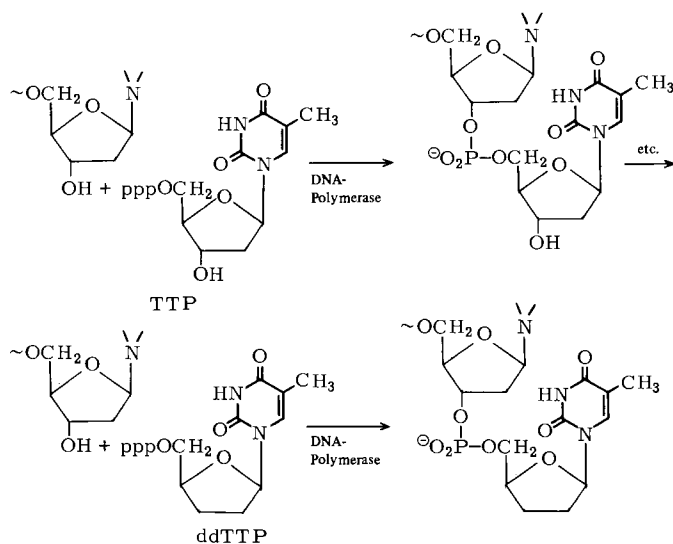


Abb. 2. Kettenbeendigung mit Didesoxythymidintriphosphat (ddTTP). Die obere Reihe zeigt die von der DNA-Polymerase katalysierte Reaktion des normalen Desoxynucleosidtriphosphates (TTP) mit dem 3'-terminalen Nucleotid des Primers; die untere Reihe zeigt die entsprechende Reaktion mit ddTTP.

malen Desoxynucleosidtriphosphaten, doch fehlt ihnen die 3'-Hydroxy-Gruppe. Daher können sie zwar von der DNA-Polymerase in eine wachsende DNA-Kette inkorporiert werden, doch terminieren sie diese mit dem Einbau, da die Kette wegen der fehlenden 3'-Hydroxy-Gruppe nicht um ein weiteres Nucleotid verlängert werden kann.

Das Prinzip dieser Methode ist in Abbildung 3 dargestellt. Zuerst bildet man aus dem Primer – gewöhnlich ein Restriktionsfragment – und der Matrize einen Komplex (primer-template-complex). Das Reaktionsgemisch wird dann in vier Proben geteilt. Eine davon (die „T-Probe“) wird in Gegenwart von ddTTP (Didesoxythymidintriphosphat) und wenig TTP zusammen mit den anderen drei, in

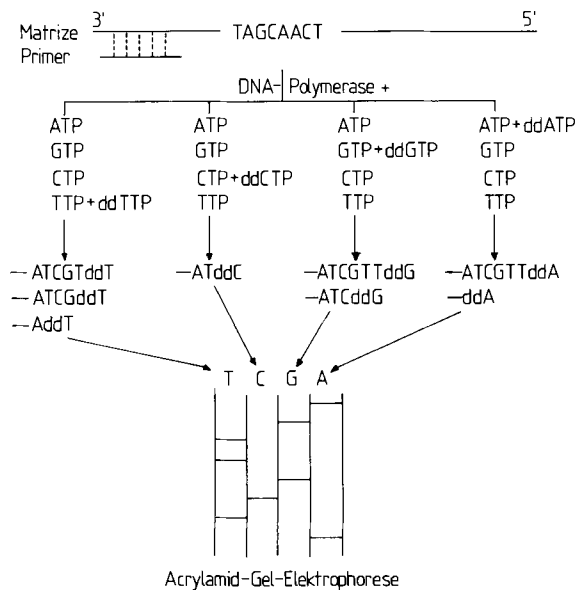


Abb. 3. Prinzip der Dideoxy- oder Kettenabbruch-Methode („chain terminating method“).

normalen Konzentrationen vorliegenden Desoxynucleosidtriphosphaten (eines davon ist ^{32}P -markiert) mit DNA-Polymerase inkubiert. Die DNA-Kette wird vom 3'-Ende des Primers aus aufgebaut; dabei wird die T-Position in den meisten Fällen vom normalen Substrat T eingenommen werden, und die Kette wächst weiter. Wird jedoch ddT eingebaut, bewirkt dies den Kettenabbruch. Nach Beendigung der Inkubation liegt eine Mischung von Ketten mit dem gleichen 5'-Ende (dem 5'-Ende des Primers) und mit ddT am 3'-Ende vor. Ähnliche Inkubationen mit jedem der drei anderen Dideoxy-Derivate liefern Mischungen von Ketten, die jeweils mit ddG, ddA oder ddC enden. Alle vier Mischungen werden nebeneinander gelelektrophoretisch auf Acrylamid unter denaturierenden Bedingungen fraktioniert. Dabei werden die Ketten aufgrund ihrer Länge getrennt; die kleinen laufen schnell, die großen langsam. Da alle Ketten in der „T-Mischung“ mit ddT enden, definiert die relative Position der Ts in der Kette die relative Länge der Ketten und damit auch deren Lage im Gel nach der Fraktionierung. Die richtige Sequenz kann dann einfach aus einem Autoradiogramm des Gels gelesen werden (Abb. 4). Diese Methode ist vergleichsweise schnell und exakt, vom 3'-Ende des Primers an können Sequenzen bis zu 300 Nucleotiden bestimmt werden. Die Bestimmung wesentlich längerer Sequenzen gelang ebenfalls, wenn auch nicht mit hoher Genauigkeit.

Problematisch an dieser Methode ist, daß die DNA-Matrize als Einzelstrang vorliegen muß. In den Bakteriophagen ΦX174 und G4 ist diese DNA-Form natürlich, meistens besteht DNA jedoch aus einem Doppelstrang; es ist gewöhnlich schwierig, beide Stränge zu trennen. Nach einer Methode von A. J. H. Smith^[14] kann doppelsträngige, lineare DNA mit Exonuclease III (eine doppelstrangspezifische Exonuclease) behandelt werden; dabei wird jede Kette vom 3'-Ende her abgebaut (Abb. 5), und es entsteht eine Struktur, die über große Strecken einzelsträngig ist und – mit einem passenden kleinen Primer – als Matrize für die DNA-Polymerase dienen kann. Diese Methode läßt sich besonders gut bei Fragmenten anwenden, die in Plas-

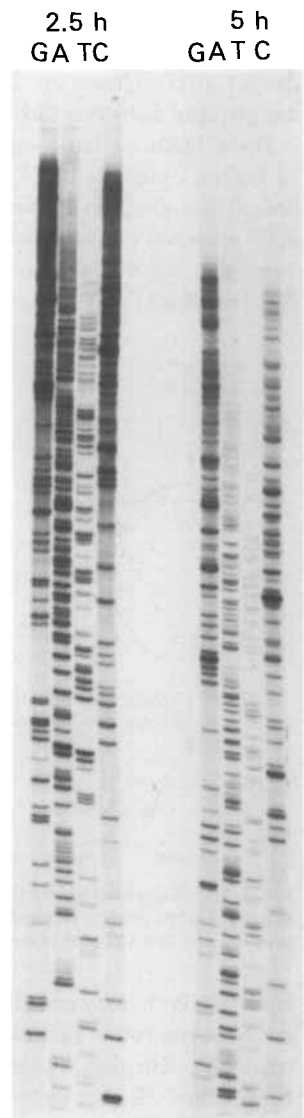


Abb. 4. Autoradiogramm eines Gels, auf dem DNA-Fragmente elektrophoretisch getrennt wurden. Der Startpunkt liegt oben, und die DNA-Ketten werden nach ihrer Größe getrennt. Die Polymer-Mischung wurde geteilt; man ließ einen Teil (links) 2,5 h, den anderen Teil (rechts) 5 h laufen.

mid-Vektoren cloniert werden; sie wurde insbesondere bei der Untersuchung menschlicher mitochondrialer DNA genutzt.

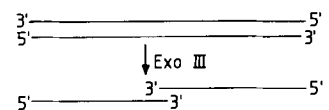


Abb. 5. Zerlegung von Doppelstrang-DNA mit Exonuclease III.

5. Clonierung in Einzelstrang-Bakteriophagen

Ein anderes, weit verbreitetes Verfahren zur Herstellung passender DNA-Matrizen ist die Clonierung von Fragmenten in Einzelstrang-Bakteriophagen^[15-17]. Dieser Weg ist in Abbildung 6 gezeigt. Verschiedene Vektoren wurden schon beschrieben. Wir benutzten einen von Gronenborn und Messing^[16] entwickelten Abkömmling des Bakteriophagen M 13, der das β -Galactosidase-Gen mit einem Angriffsort für das Restriktionsenzym *EcoRI* enthält. Ein passendes, farbbildendes Substrat (X-gal) macht die Anwe-

senheit von β -Galactosidase in einem Plaque schnell sichtbar. Eine Insertion in den Angriffsort von *EcoRI* zerstört das β -Galactosidase-Gen, der Plaque bleibt in Gegenwart des gleichen Substrats farblos.

Diese Methode hat – außer schnell Einzelstrang-DNA zu liefern – noch andere Vorteile. Sie ermöglicht den Gebrauch des gleichen Primers für alle Clone. Heidecker et al.^[18] erhielten ein 96 Nucleotide langes Restriktionsfragment, das sich von der Position im M-13-Vektor herleitet, die dem *EcoRI*-Angriffsort benachbart liegt (Abb. 6). Es

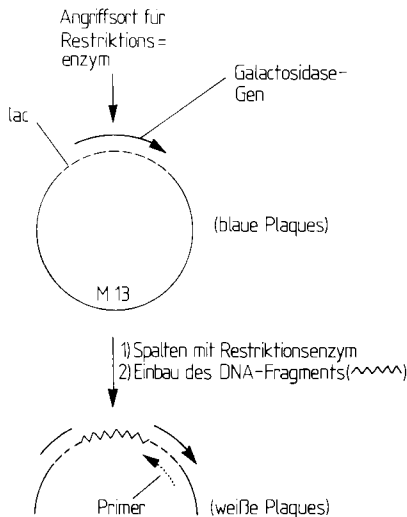


Abb. 6. Das Diagramm zeigt die Clonierung von DNA-Fragmenten im Einzelstrang-Bakteriophagen M13mp2 [16] und die Sequenzierung der Einfügung mit dem flankierenden Primer.

kann als Primer dienen und ermöglicht die Bestimmung der Sequenz von etwa 200 Nucleotiden in einer eingefügten DNA. Kürzlich wurden kleinere, synthetische Primer hergestellt^[19,20], mit denen längere DNA-Sequenzen bestimmt werden konnten. Wir stellten Zufallsclone aus ganzen, durch Restriktionsenzymen erzeugten Spaltansätzen her und bestimmten die Sequenz mit dem flankierenden Primer. Mit Computern ließen sich die Daten speichern und mit entsprechenden Programmen ordnen^[21].

Die Clonierungs-Technik zeichnet sich auch durch schnelle und effektive Fraktionierung von DNA-Fragmenten aus. Ein wichtiger Teil aller Sequenzierungsmethoden war immer der Fraktionierungsschritt; das gilt sowohl für Proteine als auch für Nucleinsäuren. Große Fortschritte beruhten oft auf der Entwicklung neuer Fraktionierungstechniken. Da mit den neuen Methoden die eigentliche Sequenzierung relativ schnell verläuft, beansprucht die Fraktionierung der Restriktionsfragmente durch Gelelektrophorese auf Acrylamid einen großen Teil der Arbeit. Sie wird auch zunehmend schwieriger, da immer größere DNA-Moleküle untersucht werden, die mehreren aufeinanderfolgenden Fraktionierungen unterworfen werden müssen, bevor reine Fragmente erhalten werden. Bei unserer neuen Methode werden diese Fraktionierungen durch Clonierungs-Vorgänge ersetzt. Die Mischung wird auf Agar ausplattiert und bebrütet. Jeder Clon ist die Nachkommenschaft eines einzigen Moleküls und ist – ungeachtet der Komplexität des ursprünglichen Gemisches – daher rein. Besonders große DNA-Moleküle können untersucht werden. Es existiert theoretisch kein Größenlimit für die zu sequenzierende DNA.

Wir haben diese Methode auf mitochondriale DNA-Fragmente^[22,23] und auf die DNA des Bakteriophagen λ angewendet. Zu Beginn der Untersuchungen können sehr schnell Ergebnisse erzielt werden (unter idealen Bedingungen lassen sich Aussagen über 500–1000 Nucleotide pro Tag machen). Später informiert ein großer Teil der Ergebnisse nur über solche DNA-Regionen, die bereits sequenziert wurden; man scheint langsamer voranzukommen. Jedoch liefern die meisten der untersuchten Clone auch dann noch nützliche Daten, da alte Sequenzen bestätigt oder korrigiert werden können. So konnten wir 90% der DNA des Bakteriophagen λ identifizieren; viele der von uns jetzt untersuchten Clone steuern weiter neue Ergebnisse bei. Die meisten DNA-Untersuchungen befassen sich mit der Identifizierung von Protein-Genen und erfordern daher eine korrekte Sequenzierung. Da Fehler bei der Analyse solcher großer Sequenzen leicht möglich sind, müssen Ergebnisse immer wieder bestätigt werden. Wir betrachten es als notwendig, die Sequenz jeder Region immer an beiden DNA-Strängen zu bestimmen.

Theoretisch ist es möglich, eine Sequenz mit Zufallsclonen vollständig zu bestimmen; eine spezifische Methode zur Bestimmung der verbleibenden endständigen Nucleotide erscheint sinnvoller. Zwar sind mehrere Methoden möglich^[22,24], sie sind jedoch verglichen mit dem Weg über Zufallsclone (random cloning approach) sehr langwierig.

6. DNA des Bakteriophagen $\Phi X174$

Die DNA des Bakteriophagen $\Phi X174$ wurde als erste mit dem Kopier-Verfahren vollständig sequenziert^[10,11]; sie ist einzelsträngig und circular, besteht aus 5386 Nucleotiden und codiert zehn Gene. Das überraschendste Ergebnis war, „überlappende Gene“ zu finden. Frühere Untersuchungen ließen vermuten, daß Gene entlang der DNA linear aneinander gereiht vorliegen, jedes mit seiner eigenen, definierten DNA-Region. Die Sequenzuntersuchungen wiesen jedoch darauf hin, daß bestimmte Nucleotidfolgen in der DNA von $\Phi X174$ die Information für zwei Gene enthalten. Ermöglicht wird das durch die Beschaffenheit des genetischen Codes: Da eine Folge von drei Nucleotiden (ein Codon) eine Aminosäure codiert, kann theoretisch jede DNA-Region – abhängig vom Translationsbeginn – drei verschiedene Aminosäuresequenzen codieren (Abb. 7). Das Leseraster, die Reihenfolge, in der ab-

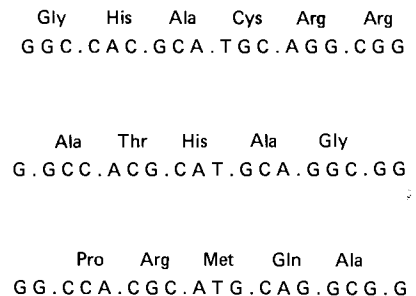


Abb. 7. Eine DNA-Sequenz kann drei verschiedene Aminosäuresequenzen codieren. Die Punkte trennen die Triplets, die eine Aminosäure codieren.

gelesen werden soll, wird durch die Position des Initiationscodons ATG bestimmt; danach werden immer drei Nucleotide gleichzeitig abgelesen. In $\Phi X174$ liegt ein initi-

ierendes Codon ATG inmitten des Gens für das D-Protein, aber in einem anderen Leseraster („reading frame“). Das bedingt eine völlig andere Aminosäuresequenz, nämlich die des E-Proteins. Abbildung 8 zeigt die Genkarte von

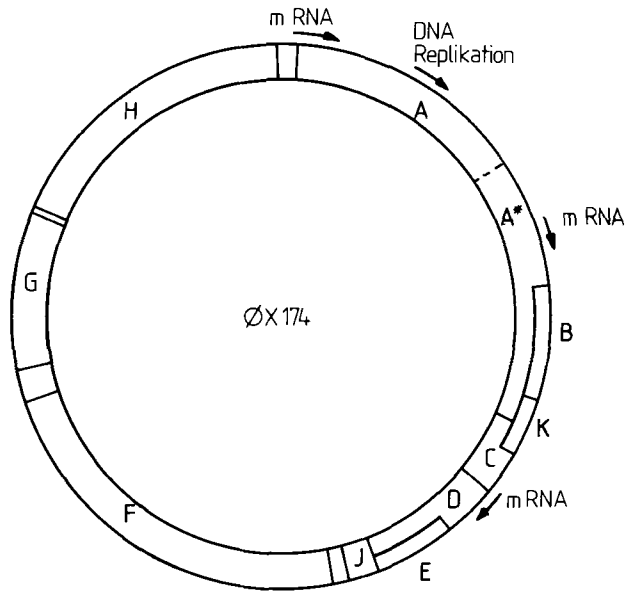


Abb. 8. Genkarte des Bakteriophagen ΦX174.

ΦX174. Das D-Gen enthält das vollständige E-Gen, während das B-Gen in das A-Gen eingebettet ist.

Untersuchungen am verwandten Bakteriophagen G4^[25] enthüllten die Existenz eines früher übersehenen Gens, das K genannt wurde. Es überlappt mit den Genen A und C. Eine Sequenz von vier Nucleotiden codiert sogar einen Teil aller drei Proteine A, C und K, indem alle drei möglichen Leseraster genutzt werden.

Es ist ungewiß, ob überlappende Gene ein allgemeines Phänomen sind oder ob sie nur bei Viren auftreten, deren Überleben von ihrer Replikationsfrequenz und damit von ihrer DNA-Größe abhängt: überlappende Gene ermöglichen die Speicherung von mehr genetischer Information bei vorgegebener DNA-Größe.

Weitere Details über die DNA-Sequenz des Bakteriophagen ΦX174 wurden andernorts publiziert^[10, 12].

7. Mitochondriale DNA der Säugetiere

Mitochondrien enthalten eine kleine Doppelstrang-DNA (mtDNA), die zwei ribosomale RNAs (rRNAs) codiert, 22–23 transfer-RNAs (tRNAs) und etwa 10–13 hydrophobe Proteine, die sich als Bestandteile der inneren Mitochondrienmembran erwiesen. Andere Proteine der Mitochondrien sind im Zellkern codiert und werden in spezifischer Weise in die Mitochondrien transportiert. Wir benutzten die oben beschriebenen Methoden zur vollständigen Bestimmung der Nucleotidsequenz menschlicher mtDNA^[23] und zur fast vollständigen Bestimmung von Rinder-mtDNA. Die Sequenz wies eine Reihe unerwarteter Charakteristika auf, die auf einen im Vergleich zu anderen biologischen Systemen unterschiedlichen Transkriptions- und Translationsapparat in Mitochondrien hindeuten.

7.1. Der genetische Code in Mitochondrien

Bisher wurde der genetische Code für universal gehalten. Die untersuchten Systeme in *Escherichia coli*, Hefe und Säugetieren unterschieden sich nicht voneinander. Unsere ersten Sequenzuntersuchungen befaßten sich mit menschlicher mtDNA. Von keinem der Proteine, die durch menschliche mtDNA codiert werden, war die Aminosäuresequenz bekannt. *Steffens* und *Buse*^[26] hatten die Sequenz der Untereinheit II der Cytochrom-Oxidase (COII) in Rinder-Mitochondrien bestimmt, und *Barrell*, *Bankier* sowie *Drouin*^[27] entdeckten, daß die von ihnen untersuchte Region der menschlichen mtDNA eine Sequenz enthielt, die ein zu dieser Aminosäuresequenz homologes Protein codiert. Man glaubte, das Gen für die menschliche COII gefunden zu haben. Überraschenderweise enthielt diese DNA-Sequenz TGA-Triplets im Leseraster für das homologe Protein. Das Triplet TGA ist aber das Terminationscodon des normalen genetischen Codes; es gibt das Signal für die Beendigung der Polypeptidkette. Es fiel auf, daß an den Stellen, an denen in der menschlichen mtDNA-Sequenz TGA vorkam, in der Rinder-Proteinsequenz die Aminosäure Tryptophan eingebaut war. Die einzig mögliche Erklärung schien zu sein, daß in Mitochondrien TGA kein Terminationscodon ist, sondern Tryptophan codiert. Auf gleiche Weise wurde gefunden, daß ATA nicht – wie normal – Isoleucin codiert, sondern Methionin. Da in diesen Untersuchungen menschliche DNA und Rinderprotein verglichen wurden, hätten Spezies-Unterschiede die Ursache solcher Differenzen sein können. Zur endgültigen Bestimmung des mitochondrialen Codes mußte die DNA-Sequenz des Gens mit der Aminosäuresequenz des von ihm codierten Proteins verglichen werden. *Young* und *Anderson*^[28] isolierten Rinder-mtDNA, bestimmten die Sequenz des COII-Gens und bestätigten die oben genannten Unterschiede.

Abbildung 9 zeigt den genetischen Code in Mitochondrien von Mensch und Rind und die Häufigkeit der ver-

		Zweiter Buchstabe				
		U	C	A	G	
U	UUU Phe 77	UCU 32	UAU Tyr 46	UGU Cys 5	U	
	UUC 140	UCC 99	UAC Tyr 89	UGC Cys 17		C
	UUA 73	UCA Ser 85	UAA Ter -	UGA Trp 93		A
	UUG Leu 17	UCC 7	UAG Ter -	UGG Trp 10		C
C	CUU 65	CCU 41	CAU His 18	CCU 7	U	
	CUC Leu 167	CCC 119	CAC His 79	CCG Arg 25		C
	CUA 276	CCA Pro 52	CAA Gln 81	CGA Arg 29		A
	CUG 45	CCG 7	CAG Gln 9	CGG 2		G
A	AUU Ile 125	ACU 51	AAU Asn 35	AGU Ser 14	U	
	AUC 196	ACC 155	AAC Asn 130	AGC Ser 39		C
	AUA 166	ACA Thr 135	AAA Lys 85	AGA Ter -		A
	AUG Met 40	ACC 10	AAG Lys 10	AGG Ter -		G
G	GUU 30	GCU 43	GAU Asp 15	GGU 24	U	
	GUC Val 49	GCC Ala 124	GAC Asp 55	GGC Gly 88		C
	GUA 71	GCA Ala 80	GAA Glu 64	GGG Gly 67		A
	GUG 18	GCC 8	GAG Glu 24	GGG 34		C

Abb. 9. Der genetische Code menschlicher Mitochondrien. Die codierenden Eigenschaften der tRNAs (eingerahmte Codons) werden ebenso gezeigt wie die Gesamtzahl der Codons, die im gesamten Genom (Abb. 10) verwendet werden. (Eine Methionin-tRNA wurde bisher entdeckt, aber es herrscht Unsicherheit über deren Anzahl und codierende Eigenschaften; diese Codons sind nicht eingerahmt.)

schiedenen Codons in menschlichen Mitochondrien. Alle Codons mit Ausnahme der Terminationscodons UAA und UAG sowie der normalerweise Arginin codierenden AGA und AGG werden genutzt; wahrscheinlich sind letztere in Mitochondrien auch Terminationscodons. Dieser Befund wird durch die Tatsache gestützt, daß keine tRNA zu finden war, die diese Codons erkennt, und daß einige der bisher nicht identifizierten Leseraster in der DNA mit diesen Codons enden.

Zur gleichen Zeit, als wir Säugetier-mtDNA untersuchten, arbeiteten *Tzagoloff et al.*^[29,30] mit mtDNA aus Hefe. Auch sie fanden Änderungen des genetischen Codes, aber nicht die gleichen wie die in Säugetier-Mitochondrien. Tabelle 1 faßt die Unterschiede zusammen.

Tabelle 1. Der genetische Code in Mitochondrien und anderen („normalen“) biologischen Systemen.

Codon	codierte Aminosäure		
	normal	Säugetier-Mitochondrien	Hefe-Mitochondrien
TGA	Term [a]	Trp	Trp
ATA	Ile	Met	Ile
CTN [b]	Leu	Leu	Thr
AGA, AGG	Arg	Term?	Arg
CGN [b]	Arg	Arg	Term?

[a] Term = Terminationscodon. [b] N = irgendeine der vier Basen.

7.2. Transfer-RNAs

Transfer-RNAs verfügen über die charakteristische „Kleeblattstruktur“. Durch Computer-unterstützte DNA-Sequenzierung^[31] konnten die Gene, die mt-tRNAs codieren, identifiziert werden.

Normale cytoplasmatische tRNAs haben außer der Kleeblattstruktur eine Reihe sogenannter „unveränderlicher“ Merkmale, die wichtig für ihre biologische Funktion zu sein scheinen. Die meisten der mt-tRNAs in Säugetieren sind insofern anomal, als ihnen einige dieser Eigenschaften fehlen. Am merkwürdigsten ist eine der Serin-tRNAs, die trotz Fehlen einer kompletten Schleife der Kleeblattstruktur^[32,33] noch als tRNA arbeitet.

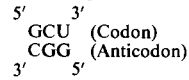
Im Cytoplasma werden die Aminosäuren durch mindestens 32 tRNAs codiert; dies ist eine Folge des „wobble“-Effekts. Codon-Anticodon-Wechselwirkungen werden nur in der ersten und zweiten Position des Codons von den Gesetzen der Basenpaarung bestimmt, in der dritten Position aber kann sich auch G mit U paaren. So kann eine tRNA zwei Codons erkennen. Es existieren viele Fälle im genetischen Code, in denen alle vier Codons, die mit denselben beiden Nucleotiden beginnen, dieselbe Aminosäure codieren. Sie sind bekannt als „family boxes“. Tabelle 2 enthält die „family box“ für Alanin; es zeigt sich, daß im normalen „wobble“-System zwei tRNAs für die vier Alanin-Codons benötigt werden.

In der mtDNA von Säugetieren konnten nur 22 tRNA-Gene gefunden werden, und in allen „family boxes“ existiert nur ein tRNA-Gen, das ein T an der Stelle hat, die der dritten Position des Codons entspricht^[34]. Es schien unwahrscheinlich, daß weitere der vermuteten tRNAs unentdeckt geblieben wären; so war die einfachste Erklärung

Tabelle 2. Normale codierende Eigenschaften der Alanin-tRNAs.

Codon	Anticodon („wobble“)	Anticodon (Mitochondrien)
GCU	GGC	UGC
GCC		
GCA		
GCG		

Es ist zu beachten, daß die erste Position des Codons mit der dritten Position des Anticodons paart und vice versa; z. B.:



die, daß in den Mitochondrien eine tRNA alle vier Codons einer „family box“ erkennen kann, und daß ein U in der ersten Position des Anticodons mit U, C, A und G in der dritten Position des Codons paaren kann. In den Fällen, in denen zwei Codons eine Aminosäure codieren und zwei eine andere, muß es zwei tRNAs geben, so daß der normale „wobble“-Effekt hier wieder gilt. In mitochondrialen Genen wurden solche erwarteten tRNAs gefunden. Abbildung 9 zeigt die codierenden Eigenschaften der mt-tRNAs. Ähnliche Schlüsse wurden von *Heckmann et al.*^[35] und *Bonitz et al.*^[36] gezogen, die mit *Neurospora*- bzw. Hefemitochondrien arbeiteten.

7.3. Verteilung der Protein-Gene

Es war bekannt, daß mitochondriale DNA drei Untereinheiten der Cytochrom-Oxidase codiert, wahrscheinlich auch drei Untereinheiten des ATPase-Komplexes, das Cytochrom b und eine Anzahl nicht identifizierter Proteine. Um die Protein-codierenden Gene zu finden, wurde die DNA nach Leserastern abgesucht, d. h. nach langen DNA-Stücken ohne Terminationscodons, die lange Polypeptidketten codieren können. Diese Leseraster sollten mit einem Initiationscodon – in normalen Systemen fast immer ATG – beginnen und mit einem Terminationscodon enden. Abbildung 10 zeigt die Verteilung der Leseraster auf der DNA, die die Protein-codierenden Gene zu sein scheinen. Das Gen für die Untereinheit II der Cytochrom-Oxidase (CoII) wurde über dessen bekannte Aminosäuresequenz (Abschnitt 7.1) identifiziert; das Gen für die Untereinheit I wurde gefunden, nachdem die Aminosäuresequenz an Rinderprotein von *J. E. Walker* bestimmt worden war (persönliche Mitteilung); die Gene für COIII, Cytochrom b und wahrscheinlich ATPase 6 wurden durch Vergleich mit den entsprechenden DNA-Sequenzen in Hefe-Mitochondrien identifiziert. *Tzagoloff et al.* konnten diese Gene in Hefe mit genetischen Methoden finden. Noch ist es nicht möglich, den anderen Leserastern Proteine zuzuordnen.

Eine ungewöhnliche Eigenschaft der mtDNA ist ihre außergewöhnlich kompakte Struktur. Die Leseraster, die Proteine und rRNA codieren, scheinen jeweils von tRNA-Genen flankiert zu sein, wobei keine oder nur wenige Nucleotide dazwischenliegen. Das läßt ein einfaches Transkriptionsmodell vermuten, in dem eine lange RNA an der DNA kopiert wird und dann die tRNAs durch „processing“-Enzyme herausgeschnitten werden. Das gleiche „processing“ liefert rRNAs und messenger-RNAs (mRNAs), von denen die meisten monocistronisch sein sollten. Dieses Modell wird gestützt durch die Ergebnisse der Arbeiten von *Attardi et al.*^[37,38], die die RNA-Sequenzen an den 3'-

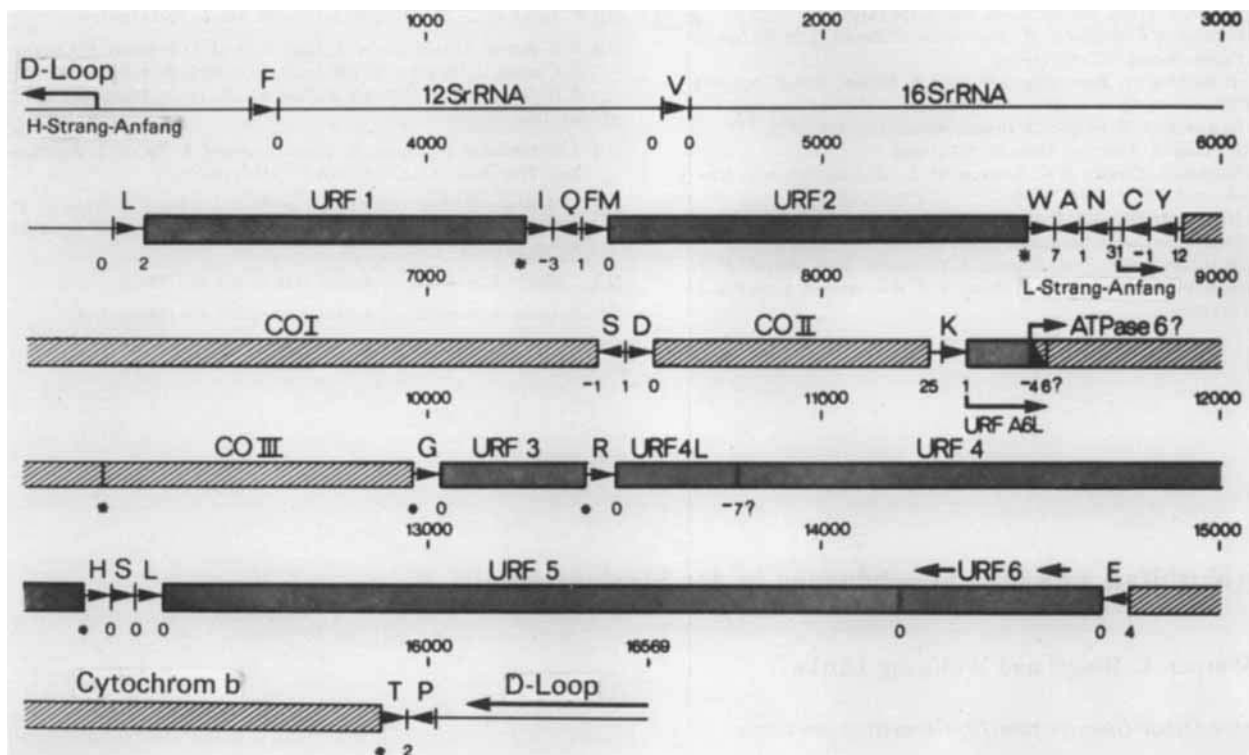


Abb. 10. Genkarte der menschlichen mtDNA, wie sie aus der DNA-Sequenz erhalten wird. Schräg gestrichelte Regionen stellen die erwarteten Leseraster für die Proteine dar. URF = nichtidentifizierte Leseraster (unassigned reading frames). tRNA-Gene sind gekennzeichnet durch die Ein-Buchstaben-Symbole für die entsprechenden Aminosäuren und sind entweder auf dem H-Strang (◄) oder auf dem L-Strang (►) codiert. Die Zahlen über den Genen zeigen die Nucleotid-Skala, die unter den Genen geben die Zahl nichtcodierender Nucleotide zwischen den Genen an. Eine negative Zahl weist auf eine Überlappung hin, ein * bedeutet, daß Terminationscodons durch Polyadenylierung der mRNA geschaffen werden.

und 5'-Enden der mRNAs identifizierten und damit ihren DNA-Ort lokalisierten. Aus der gefundenen Verteilung folgt, daß das Initiationscodon nahe oder direkt am 5'-Ende der mRNA liegt. Das deutet auf einen sich von anderen Systemen unterscheidenden Initiationsmechanismus hin. In Bakterien existiert gewöhnlich eine ribosomale Bindungsstelle vor dem Initiationscodon AUG, in eukaryotischen Systemen scheint die „Kappen“(cap)-Struktur am 5'-Ende der mRNA eine ähnliche Funktion zu haben, und das erste AUG, das auf die Kappen-Struktur folgt, fungiert als Initiator. Mitochondrien haben wahrscheinlich ein einfacheres und primitiveres System, mit dem der Translationsapparat direkt das 5'-Ende der mRNA erkennt. Eine weitere ungewöhnliche Eigenschaft der Mitochondrien ist, daß außer AUG auch AUA und vielleicht auch AUU als Initiationscodon fungieren.

Nach diesem Modell sollten einige der mRNAs kein Terminationscodon enthalten, nachdem die tRNAs herausgeschnitten wurden. An ihrem 3'-Ende tragen sie jedoch U oder UA. Da die mRNAs an ihrem 3'-Ende gewöhnlich polyadenyliert sind, entsteht ein Codon UAA, das das Leseraster für diese Proteine abschließt.

Die kompakte Struktur des mitochondrialen Säugetiergenoms steht in scharfem Kontrast zu dem der Hefe, das etwa fünfmal so groß ist, aber die gleiche Anzahl an Proteinen und RNAs codiert. Die Gene werden dort durch lange AT-reiche DNA-Stücke getrennt, die ohne offensichtliche biologische Funktion sind; einige Gene scheinen auch Insertionselemente (Introns) zu enthalten, die bei Säugetier-mtDNA unbekannt sind.

Eingegangen am 19. Januar 1981 [A 381]
Übersetzt von Christiane Koszka, Hamburg

- [1] R. W. Holley, Les Prix Nobel, Almquist and Wiksell International, Stockholm 1968, S. 183.
- [2] F. Sanger, Les Prix Nobel, Almquist and Wiksell International, Stockholm 1958, S. 134.
- [3] F. Sanger, G. G. Brownlee, B. G. Barrell, J. Mol. Biol. 13, 373 (1965).
- [4] H. D. Robertson, B. G. Barrell, H. L. Weith, J. E. Donelson, Nature, New Biol. 241, 38 (1973).
- [5] E. B. Ziff, J. W. Sedat, F. Galibert, Nature, New Biol. 241, 34 (1973).
- [6] M. A. Billeter, J. E. Dahlberg, H. M. Goodman, J. Hindley, C. Weissmann, Nature 224, 1083 (1969).
- [7] P. Berg, H. Fancher, M. Chamberlin, Symp. Inf. Macromol., Academic Press, New York 1963, S. 467.
- [8] F. Sanger, J. E. Donelson, A. R. Coulson, H. Kössel, D. Fischer, Proc. Natl. Acad. Sci. USA 70, 1209 (1973).
- [9] F. Sanger, A. R. Coulson, J. Mol. Biol. 94, 441 (1975).
- [10] F. Sanger, G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe, M. Smith, Nature 265, 687 (1977).
- [11] F. Sanger, S. Nicklen, A. R. Coulson, Proc. Natl. Acad. Sci. USA 74, 5463 (1977).
- [12] F. Sanger, A. R. Coulson, T. Friedmann, G. M. Air, B. G. Barrell, N. L. Brown, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe, M. Smith, J. Mol. Biol. 125, 225 (1978).
- [13] G. N. Godson, B. G. Barrell, R. Staden, J. C. Fiddes, Nature 276, 236 (1978).
- [14] A. J. H. Smith, Nucleic Acids Res. 6, 831 (1979).
- [15] W. M. Barneš, Gene 5, 127 (1979).
- [16] B. Gronenborn, J. Messing, Nature 272, 375 (1978).
- [17] R. Herrmann, K. Neugebauer, H. Schaller, H. Zenitgraf in D. T. Denhardt, D. Dressler, D. S. Ray: The Single-stranded DNA Phages, Cold Spring Harbor Laboratory, New York 1978, S. 473.
- [18] G. Heidecker, J. Messing, B. Gronenborn, Gene 10, 69 (1980).
- [19] S. Anderson, M. J. Gait, L. Mayol, I. G. Young, Nucleic Acids Res. 8, 1731 (1980).
- [20] M. J. Gait, unveröffentlichte Ergebnisse (1980).
- [21] R. Staden, Nucleic Acids Res. 8, 3673 (1980).
- [22] F. Sanger, A. R. Coulson, B. G. Barrell, A. J. H. Smith, B. A. Roe, J. Mol. Biol. 143, 161 (1980).
- [23] B. G. Barrell, S. Anderson, A. T. Bankier, M. H. L. de Bruijn, E. Chen, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden, I. G. Young, 31. Mosbach Kolloq.: Biological Chemistry of Organelle Formation, Springer, Berlin 1980.

- [24] G. Winter, S. Fields, *Nucleic Acids Res.* 8, 1965 (1980).
 [25] D. C. Shaw, J. E. Walker, F. D. Northrop, B. G. Barrell, G. N. Godson, J. C. Fiddes, *Nature* 272, 510 (1978).
 [26] G. J. Steffens, G. Buse, Hoppe-Seyler's Z. Physiol. Chem. 360, 613 (1979).
 [27] B. G. Barrell, A. T. Bankier, J. Drouin, *Nature* 282, 189 (1979).
 [28] I. G. Young, S. Anderson, *Gene* 12, 257 (1980).
 [29] G. Macino, G. Coruzzi, F. G. Nobrega, M. Li, A. Tzagoloff, *Proc. Natl. Acad. Sci. USA* 76, 3784 (1979).
 [30] G. Macino, A. Tzagoloff, *Proc. Natl. Acad. Sci. USA* 76, 131 (1979).
 [31] R. Staden, *Nucleic Acids Res.* 8, 817 (1980).
 [32] M. H. L. de Bruijn, P. H. Schreier, I. C. Eperon, B. G. Barrell, E. Y. Chen, P. W. Armstrong, J. F. H. Wong, B. A. Roe, *Nucleic Acids Res.* 8, 5213 (1980).
 [33] P. Arcari, G. G. Brownlee, *Nucleic Acids Res.* 8, 5207 (1980).
 [34] B. G. Barrell, S. Anderson, A. T. Bankier, M. H. L. de Bruijn, E. Chen, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden, I. G. Young, *Proc. Natl. Acad. Sci. USA* 77, 3164 (1980).
 [35] J. E. Heckman, J. Sarnoff, B. Alzner-De Weerd, S. Yin, U. L. RajBhandary, *Proc. Natl. Acad. Sci. USA* 77, 3159 (1980).
 [36] S. G. Bonitz, R. Berlini, G. Coruzzi, M. Li, G. Macino, F. G. Nobrega, M. P. Nobrega, B. E. Thalensfeld, A. Tzagoloff, *Proc. Natl. Acad. Sci. USA* 77, 3167 (1980).
 [37] J. Montoya, D. Ojala, G. Attardi, *Nature* 290, 465 (1980).
 [38] D. Ojala, J. Montoya, G. Attardi, *Nature* 290, 470 (1980).

Hybridorbitale und ihre Anwendungen in der Strukturchemie

Von Werner A. Bingel und Wolfgang Lüttke^[*]

Professor Oskar Glemser zum 70. Geburtstag gewidmet

Es wird eine einheitliche, systematische und vollständige Übersicht über die s-p-Hybridorbitale gegeben, mit denen ein Zentralatom Z über σ -Bindungen mit zwei bis vier Liganden L verknüpft ist. Dabei werden sämtliche möglichen Symmetriefälle behandelt: bei Viererkoordination von T_d -Symmetrie bei ZL_4 -Molekülen bis zum asymmetrischen Fall bei vier verschiedenen Liganden; bei Dreierkoordination von D_{3h} -Symmetrie bei ZL_3 - bis zu C_s -Symmetrie bei $ZLL'L''$ -Molekülen. Für einige Fälle werden die Formeln erstmalig angegeben. Dabei wird gezeigt, daß die allgemeinsten sp^3 - und sp^2 -Hybride noch drei bzw. zwei freie Parameter enthalten. Das Auftreten von Hybridorbitalen beim Übergang von der „delokalisierten“ zu einer „lokalisierten“ Beschreibung der Bindungsverhältnisse wird anhand von Beispielen erläutert. Schließlich werden Anwendungsmöglichkeiten bei der Beschreibung der Bindungsverhältnisse von Molekülen und der Interpretation der Ergebnisse spektroskopischer Methoden zur Bestimmung der Molekülstruktur aufgezeigt sowie durch einige Beispiele illustriert.

1. Hybridorbitale früher und heute

Die s-p-Hybridorbitale, mit denen ein Zentralatom Z zwei, drei oder vier gleiche Ligandenatome L bindet, dürfen jedem Chemiker bekannt sein. Abbildungen ihrer räumlichen Anordnung und auch Formeln, mit denen sie durch das s- und die drei p-Atomorbitale (AOs) des Zentralatoms Z ausgedrückt werden, sind in vielen einführenden Lehrbüchern der Allgemeinen, der Anorganischen und der Organischen Chemie zu finden^[1]. Es sind dies die speziellen digonalen (sp), trigonalen (sp^2) und tetraedrischen (sp^3) Hybrid-AOs, die schon 1928 von Pauling zur Erklärung der gerichteten Valenz in linearen ZL_2 -, ebenen ZL_3 - und tetraedrischen ZL_4 -Molekülen eingeführt wurden^[2].

Wir nennen sie *speziell*, weil bei ihnen die Valenz- oder Interorbitalwinkel^[3] alle gleich sind (180° , 120° bzw.

109.5°). Da damit auch der prozentuale s- und p-Charakter festgelegt ist, enthalten diese speziellen Hybrid-AOs keinerlei freie Parameter mehr. Sie sind außerdem dadurch ausgezeichnet, daß bei ihnen aus Symmetriegründen Bindungs- und Valenzwinkel^[3] übereinstimmen müssen.

Bindet das Zentralatom Z dagegen *verschiedenartige* Liganden, L, L', L'' usw., so werden die Bindungswinkel je nach der Natur der Liganden mehr oder weniger stark von diesen speziellen Werten abweichen. Man benötigt *allgemeinere* Hybridorbitale, bei denen die Parameter (Valenzwinkel sowie s- und p-Charakter) noch – innerhalb gewisser Grenzen – frei variabel sind, um sie dem jeweiligen Molekül anpassen zu können.

Eine Reihe solcher Hybridorbitale für ebene ZL_2L' -Moleküle wie $F_2C=O$ oder pyramidale ZL_3L' -Moleküle wie CH_3Cl wurde in den Jahren zwischen 1930 und 1950 beschrieben^[4-6,88], doch fehlt bis heute eine *einheitliche* und *vollständige* Darstellung. Sie soll in diesem Beitrag gegeben werden.

Im zweiten Abschnitt werden zunächst die Formeln für beliebige Hybridorbitale, die aus s- und p-Orbitalen erhalten werden, zusammengestellt, einschließlich der Neben-

[*] Prof. Dr. W. A. Bingel
 Lehrstuhl für Theoretische Chemie der Universität
 Tammannstraße 6, D-3400 Göttingen
 Prof. Dr. W. Lüttke
 Organisch-chemisches Institut der Universität
 Tammannstraße 2, D-3400 Göttingen