

Über den genetischen Code

Nobel-Vortrag am 11. Dezember 1962 [*]

VON DR. F. H. CRICK

MEDICAL RESEARCH COUNCIL LABORATORY OF MOLECULAR BIOLOGY,
UNIVERSITY POSTGRADUATE MEDICAL SCHOOL, CAMBRIDGE, ENGLAND

Einleitung

Es darf heute als sicher gelten, daß die Aminosäure-Sequenz eines jeden Proteins durch eine Basen-Sequenz innerhalb eines Nucleinsäure-Moleküls bestimmt wird. Gewöhnlich findet man in den Proteinen 20 verschiedene Aminosäuren, während es in den Nucleinsäuren nur vier verschiedene Basen gibt. Der genetische Code beschreibt, wie die Aufeinanderfolge von zwanzig oder mehr verschiedenen Aminosäuren durch eine Sequenz von vier verschiedenen Basen festgelegt wird.

Die biologische Bedeutung dieser Frage braucht kaum betont zu werden. Wahrscheinlich ist bei jedem Organismus die genetische Information zum größten Teil – wenn nicht sogar vollständig – in der Nucleinsäure enthalten. Gewöhnlich ist Desoxyribonucleinsäure (DNS) das genetische Material, doch kann es bei einigen kleinen Viren auch Ribonucleinsäure (RNS) sein. Die genetische Information bestimmt offenbar vor allem die Aminosäure-Sequenz der Proteine eines Organismus (ob sie noch andere Funktionen hat, wissen wir nicht). Diesen Zusammenhang hat *Beadle* in die Worte „ein Gen – ein Enzym“ gefaßt. In der heutigen, mehr verfeinerten, aber umständlicheren Terminologie würde man sagen „ein Cistron – eine Polypeptidkette“.

Es ist eine der bedeutenden Verallgemeinerungen in der Biochemie, daß die zwanzig Aminosäuren und die vier Basen mit wenigen Ausnahmen in der ganzen belebten Natur die gleichen sind. Erstaunlicherweise wird das in Lehrbüchern kaum erwähnt. Soweit ich es beurteilen kann, haben *Watson* und ich im Sommer 1953 als Antwort auf einen Brief *Gamows* zum erstenmal diese Liste von zwanzig Aminosäuren zusammengestellt.

Ich werde mich im folgenden nicht mit technischen Einzelheiten der Untersuchungen über den genetischen

Code befassen. Dies ist in einer kürzlich geschriebenen Zusammenfassung [1] geschehen, die demnächst erscheinen wird. Ich werde auch nicht auf die biochemischen Details der messenger-RNS oder der Proteinsynthese eingehen. Ich möchte vielmehr an Hand einiger allgemeiner Fragen zeigen, wieviel wir heute über den genetischen Code wissen.

Größe eines Codons

Wir wollen annehmen, daß der genetische Code einfach ist, und fragen, wieviele Basen einer Aminosäure entsprechen. Zwei Basen dürften kaum genügen, denn aus vier verschiedenen Dingen kann man nur $4 \times 4 = 16$ verschiedene Paare bilden. Wir würden aber mindestens zwanzig Paare brauchen und wahrscheinlich eines oder zwei mehr als „Zwischenraum“ oder für andere Zwecke. Mit Basen-Triplets hätten wir 64 Möglichkeiten. Der leichteren Verständigung halber wollen wir einen Satz Basen, der einer Aminosäure entspricht, als „Codon“ bezeichnen.

Damit sind wir bei unserer ersten Frage: Überlappen sich die Codons, d. h. gibt es Basen, die zwei oder mehr Codons zugleich angehören? Es ist ziemlich sicher, daß solche Überlappungen nicht auftreten. Kämen sie vor, so müßte es möglich sein, durch Änderung einer Base (bei einer Mutation) zwei oder mehr benachbarte Aminosäuren zu verändern. Tatsächlich hat aber die Änderung einer Base nur die Änderung einer Aminosäure zur Folge, und dies gilt sowohl bei spontanen Mutationen (... B. der, die beim Menschen zur Bildung ano-

[*] Wir danken dem Autor und dem Nobelkomitee, Stockholm, für die Genehmigung zur Veröffentlichung dieses Vortrags.

[1] F. H. C. Crick in *Davidson u. W. E. Cohn: Progress in Nucleic Acid Research*. Academic Press, New York, im Druck; dort auch weitere Literaturzitate.

malen Hämoglobins führt) als auch bei chemisch induzierten Mutationen, die man etwa beim Tabakmosaikvirus mit salpetriger Säure und anderen Reagentien hervorrufen kann [2]. Mit größter Wahrscheinlichkeit überlappen sich die Codons also nicht.

Wie aber ist die Basen-Sequenz einer Nucleinsäure in Codons unterteilt? Es gibt in einer Nucleinsäure-Kette nichts, was vollkommen regelmäßig wäre und uns zeigen könnte, wie die Basen in Codons zu gruppieren sind. Wären beispielsweise alle Codons Triplets, so gäbe es eine korrekte Lesart der genetischen Information und daneben aber zwei falsche Lesarten, zu denen man kommt, wenn man die Teilung in Dreiergruppen nicht an der richtigen Stelle beginnt. Wir [3] haben kürzlich experimentelle Hinweise dafür erhalten, daß jeder Abschnitt der genetischen Information von einem bestimmten Punkt aus gelesen werden muß, wahrscheinlich von einem Ende der Nucleinsäure-Kette aus. Dies entspricht der Beobachtung – besonders klar in den Arbeiten von Dintzis [4] –, daß Aminosäuren in linearer Reihenfolge zur Polypeptidkette zusammengefügt werden und daß dieser Aufbau am Aminoende der Kette beginnt.

Unsere nächste allgemeine Frage lautet: wie groß ist ein Codon, d. h. wieviele Basen enthält es? Die eben erwähnten Untersuchungen [3] sprechen dafür, daß alle Codons (oder doch fast alle) aus drei Basen bestehen (allerdings schließen unsere Daten ein kleines Vielfaches von drei, d. h. sechs oder neun, nicht vollkommen aus). Wir kamen zu diesem Schluß, als wir beim Bakteriophagen T4 die Mutationen im A- und B-Cistron der r_{II} -Region untersuchten. Bei diesen Mutationen werden der genetischen Information offenbar Basen hinzugefügt oder entzogen. Sie lassen sich mit Acridinderivaten hervorrufen und können mit Mutagenen, die lediglich Basen ineinander umwandeln, nicht rückgängig gemacht werden. Außerdem führen solche Mutationen zur Inaktivität des gesamten Gens.

Alle so entstehenden Mutanten lassen sich in zwei Gruppen ordnen, die wir mit (+) und (–) bezeichnen wollen. Der Einfachheit halber stellen wir uns vor, daß die Mutanten der (+)-Gruppe an irgendeiner Stelle der genetischen Information eine zusätzliche Base haben, während den Mutanten der (–)-Gruppe eine Base fehlt. Das experimentum crucis ist nun, das Genmaterial von drei Mutanten der gleichen Gruppe, d. h. (+) mit (+) und (+) oder (–) mit (–) und (–), durch genetische Rekombination zu einem neuen Gen zusammenzufügen. Während eine oder zwei (+)-Mutationen ein Gen vollkommen unwirksam machen, hat ein Gen mit drei (geeignet gewählten) (+)-Mutationen einige Aktivität. Die detaillierte Prüfung dieser Resultate ergab, daß sie genau dem entsprechen, was wir zu erwarten haben, wenn die genetische Information von einem Ende aus in Triplets gelesen werden muß.

[2] H. G. Wittmann, Z. Vererbungslehre 93, 491 (1962); A. Tsugita, J. molec. Biol. 5, 284, 293 (1962).

[3] F. H. C. Crick, L. Barnett, S. Brenner u. R. J. Watts-Tobin, Nature (London) 192, 1227 (1961).

[4] M. A. Naughton u. Howard M. Dintzis, Proc. nat. Acad. Sci. USA 48, 1822, (1962).

Wir werden manchmal gefragt, was geschieht, wenn sich vier (+)-Mutationen in einem Gen befinden. Um dies zu beantworten, haben meine Mitarbeiter kürzlich Gene mit bis zu sechs (+)-Mutationen erzeugt. Eine solche Kombination ist aktiv, wie unsere Theorie es erwarten läßt, wogegen Gene mit vier oder fünf (+)-Mutationen unwirksam sind. Wir haben uns auch intensiv bemüht, die Entstehung von Kombinationen zu erklären, in denen das Gen zwar wirksam ist, aber nur mit einem sehr geringen Wirkungsgrad arbeitet. Unsere Ergebnisse stimmen mit der Hypothese überein, daß der Mechanismus, der in der Zelle die genetische Information abliest, in einigen sehr seltenen Fällen von einem Basen-Triplett, das keiner Aminosäure entspricht und das also keinen „Sinn“ ergibt, statt aller drei Basen nur zwei „registriert“. Diese Beobachtung versetzt uns in die Lage festzustellen, in welcher Richtung die genetische Information abgelesen wird: in der Schreibweise, in der die r_{II} -Region gewöhnlich aufgezeichnet wird, ist es die Richtung von links nach rechts. Wir werden über die technischen Einzelheiten dieser Arbeiten in Kürze berichten. Eine endgültige Bestätigung unserer Theorien können nur detaillierte Untersuchungen über Veränderungen in der Aminosäure-Sequenz liefern, die durch Mutationen der hier skizzierten Art hervorgerufen werden.

Unsere Ergebnisse gestatten noch einen weiteren allgemeinen Schluß: es scheint, daß die Zahl derjenigen Triplets, die keiner Aminosäure entsprechen („Unsinn“-Triplets), gering ist, da man solche Dreiergruppen nur gelegentlich antrifft. Allerdings ist diese Folgerung weniger sicher als die anderen Aussagen über die Natur des genetischen Code.

Man hat noch keinen direkten Nachweis dafür, daß die genetische Information mit ihrem Produkt colinear ist, d. h. daß ein Ende des Gens dem Aminoende der Polypeptidkette, das andere dem Carboxylende entspricht und daß dazwischen die Codons im Gen die gleiche Reihenfolge haben wie die Aminosäuren im Protein. Eine solche Colinearität ist jedoch sehr wahrscheinlich, umso mehr, als in vielen Fällen gezeigt werden konnte, daß Mutationen, welche die gleiche Aminosäure betreffen, auf der Genkarte äußerst nahe beieinander liegen. Der experimentelle Beweis für die Colinearität von Gen und Polypeptidkette ist aber sicher in Kürze zu erwarten.

Universalität des Code

Wir können hier gleich eine weitere Frage stellen: ist der genetische Code universell, d. h. ist er in allen Organismen der gleiche? Erste Befunde sprechen dafür. Beispielsweise kann man mit einem zellfreien System, das teils von Kaninchen-Reticulocyten, teils von *Escherichia coli* stammt [5], ein Protein synthetisieren, das Kaninchen-Hämoglobin sehr ähnlich ist. Dies wäre kaum möglich, wenn sich der genetische Code in beiden Or-

[5] G. von Ehrenstein u. F. Lipmann, Proc. nat. Acad. Sci. USA 47, 941 (1961).

ganismen stark unterschiede. Man kann die Universalität des Code jedoch noch durch direktere Versuche prüfen:

In Zellen, in denen DNS das genetische Material ist, regelt die DNS die Proteinsynthese sicher nicht unmittelbar. Es ist vielmehr anzunehmen, daß die Basen-Sequenz der DNS – wahrscheinlich nur die Sequenz einer ihrer Ketten – in der RNS kopiert wird und daß diese spezielle RNS als Überträger der genetischen Information wirkt, indem sie die Verknüpfung von Aminosäuren zu Polypeptidketten lenkt. Entscheidend für die Untersuchung dieser Fragen war die Entdeckung von *Nirenberg* und *Matthaei* [6], daß auch synthetische RNS die Peptidsynthese lenken kann. Beispielsweise bewirkt Polyuridylsäure, d. h. eine RNS, in der jede Base Uracil ist, die Synthese von Polyphenylalanin, wenn man sie einem zellfreien System zusetzt, das Polypeptidketten synthetisieren kann. Ein Codon für Phenylalanin scheint also die Sequenz UUU zu sein (U bedeutet Uracil, so wie wir im folgenden A für Adenin, G für Guanin und C für Cytosin verwenden werden). Diese Entdeckung hat den Weg für einen raschen, wenn auch etwas verworrenen Angriff auf das Problem des genetischen Code geöffnet.

Es würde den Rahmen dieser Übersicht sprengen, wollte ich hier im Detail über diese Arbeiten berichten. Ich habe dies in der bereits erwähnten Zusammenfassung [1] getan, die allerdings infolge der raschen Entwicklung auf diesem Gebiet zum Teil schon wieder überholt ist. Einige allgemeine Folgerungen lassen sich trotz des schnellen Fortschritts aber mit einiger Sicherheit ziehen.

Untersuchungen mit zellfreien Systemen

Die bisher von *Nirenberg* [6] und *Ochoa* [7] und ihren Mitarbeitern hauptsächlich angewendete Technik besteht darin, enzymatisch Polynucleotide zu synthetisieren, in denen zwei oder drei der vier möglichen Basen statistisch verteilt sind, und diese einem zellfreien, peptid-synthetisierenden System zuzusetzen. Beispielsweise werden unter dem Einfluß eines Polynucleotids, das etwa gleiche Mengen Uracil und Cytosin in (vermutlich) statistischer Verteilung enthält und das wir Poly-(U,C) nennen wollen, die Aminosäuren Phenylalanin, Serin, Leucin, Prolin und möglicherweise Threonin verstärkt zur Peptidsynthese verwendet. Mit Hilfe von Polynucleotiden unterschiedlicher Zusammensetzung und mit der Annahme eines Triplet-Code erhält man also begrenzte Informationen über die Zusammensetzung einzelner Triplets.

Diese Untersuchungen ergaben, daß – mit geringfügigen Ausnahmen – jedes Polynucleotid die Verwendung

einer charakteristischen Gruppe von Aminosäuren zur Proteinsynthese fördert. Die vier Basen haben durchaus unterschiedliche Wirkungen. Ein Vergleich der nach dieser Methode ermittelten Triplets mit den durch Mutationen hervorgerufenen Änderungen in der Aminosäure-Sequenz ergibt eine brauchbare Übereinstimmung. Der unter dem Einfluß der Polynucleotide stärkere Einbau von Aminosäuren verlangt die Anwesenheit aller Komponenten, die auch für die Proteinsynthese benötigt werden, und wird durch die gleichen Inhibitoren gehemmt. Das zellfreie System ist also sicher kein vollständiges Kunstprodukt, und seine Syntheseleistung steht wahrscheinlich in enger Beziehung zur echten Proteinsynthese.

Man hat begonnen, mit Polynucleotiden und zellfreien Systemen von verschiedenen Species zu prüfen, ob der Code in allen Organismen der gleiche ist. Es sollte sich auf diese Weise verhältnismäßig leicht feststellen lassen, ob der Code universell ist oder von Organismus zu Organismus variiert. Erste Resultate zeigten zwischen *E. coli* und Säugetieren keinen deutlichen Unterschied [8,9].

Zunächst nahm man an, daß jedes Basen-Triplett Uracil enthalten müsse, doch war dies weder theoretisch plausibel noch ließ es sich experimentell beweisen. Den ersten direkten Hinweis dafür, daß diese Annahme nicht zutrifft, erbrachten meine Mitarbeiter *Bretscher* und *Grunberg-Manago* [10]. Sie zeigten, daß Poly-(C,A) den Einbau mehrerer Aminosäuren stimuliert. Andere Laboratorien [8,11] haben kürzlich ähnliche Ergebnisse für andere uracil-freie Polynucleotide mitgeteilt. Es ist daher heute wahrscheinlich, daß viele der 64 Triplets, sicher die meisten von ihnen, einer Aminosäure entsprechen und daß im allgemeinen mehrere Triplets Chiffren für die gleiche Aminosäure sind. Beispielsweise spricht ein sehr elegantes Experiment [12] dafür, daß Leucin sowohl in dem Triplett (UUC) als auch in der Gruppe (UUG) chiffriert ist (die Klammern deuten an, daß man die Basen-Sequenz innerhalb der Triplets noch nicht kennt). Mehrere indirekte Hinweise, die hier nicht im einzelnen beschrieben werden können, bestätigen diese allgemeine Hypothese. Unglücklicherweise wird damit aber die eindeutige Zuordnung der Triplets sehr viel schwieriger als sie es wäre, wenn jeder Aminosäure nur eine Dreiergruppe entspräche. Hinzu kommt, daß man mit Polynucleotiden, in denen die Basen statistisch verteilt sind, nicht die Basen-Sequenz innerhalb der Triplets bestimmen kann. Es ist zwar gelungen, Polynucleotide zu synthetisieren, deren Basen-Sequenz an einem Ende der Kette genau bekannt ist, aber die damit erzielten Ergebnisse geben vorerst nur Hinweise und ge-

[6] J. H. Matthaei u. M. W. Nirenberg, Proc. nat. Acad. Sci. USA 47, 1580 (1961); M. W. Nirenberg u. J. H. Matthaei, ibid. 47, 1588 (1961); M. W. Nirenberg, J. H. Matthaei u. O. W. Jones, ibid. 48, 104 (1962); J. H. Matthaei, O. W. Jones, R. G. Martin u. M. W. Nirenberg, ibid. 48, 666 (1962).

[7] P. Lengyel, J. F. Speyer u. S. Ochoa, Proc. nat. Acad. Sci. USA 47, 1936 (1961); J. F. Speyer, P. Lengyel, C. Basilio u. S. Ochoa, ibid. 48, 63 (1962); P. Lengyel, J. F. Speyer, C. Basilio u. S. Ochoa, ibid. 48, 282 (1962); J. F. Speyer, P. Lengyel, C. Basilio u. S. Ochoa, ibid. 48, 441 (1961); C. Basilio, A. J. Wahba, P. Lengyel, J. F. Speyer u. S. Ochoa, ibid. 48, 613 (1962).

[8] R. S. Gardner, A. J. Wahba, C. Basilio, R. S. Miller, P. Lengyel u. J. F. Speyer, Proc. nat. Acad. Sci. USA 48, 2087 (1962).

[9] H. R. V. Arnstein, R. A. Cox u. J. A. Hunt, Nature (London) 194, 1042 (1962); E. S. Maxwell, Proc. nat. Acad. Sci. USA 48, 1639 (1962); I. B. Weinstein u. A. N. Schechter, ibid. 48, 1686 (1962).

[10] M. S. Bretcher u. M. Grunberg-Manago, Nature (London) 195, 283 (1962).

[11] O. W. Jones und M. W. Nirenberg, Proc. nat. Acad. Sci. USA 48, 2115 (1962).

[12] B. Weisblum, S. Benzer u. R. W. Holley, Proc. nat. Acad. Sci. USA 48, 1449 (1962).

statten noch keine Schlüsse [13]. Immerhin scheint aus diesen und anderen, noch unveröffentlichten Resultaten hervorzugehen, daß das Aminoende einer Polypeptidkette dem in der üblichen Schreibweise rechts stehenden Ende einer Polynucleotidkette entspricht, d. h. dem Ende, an dem der Zucker mit freier 2'- und freier 3'-OH-Gruppe steht.

Es darf als sicher gelten, daß eine einzelne RNS-Kette als messenger-RNS dienen kann, denn Poly-U ist eine solche einzelne Kette und hat keine Sekundärstruktur. Gibt man Poly-A und Poly-U zusammen, so bilden sich Doppel- oder Tripelhelices, und die Kombination ist unwirksam. Man hat außerdem erste Hinweise [11] dafür, daß eine Sekundärstruktur innerhalb einer Polynucleotidkette deren Fähigkeit, die Proteinsynthese zu stimulieren, herabsetzt.

Daß sich der genetische Code aus Triplets zusammensetzt, ist bisher nur durch indirekte genetische Versuche gezeigt worden, wie wir sie beschrieben haben. Ein direkter Beweis mit biochemischen Methoden steht noch aus.

Man hat versucht, von den Änderungen, die durch Mutationen hervorgerufen werden, auf die relative Basen-Sequenz innerhalb einiger Triplets zu schließen. Ich glaube, daß solche Versuche erst dann einen Sinn haben, wenn man mehr und Zuverlässigeres über die Zusammensetzung der Triplets weiß.

Untersuchungsergebnisse mehrerer Arbeitskreise [10-12] lassen den Schluß zu, daß Poly-U sowohl den Einbau von Phenylalanin als auch – in geringerem Maß – den von Leucin stimuliert. Die Bedeutung dieser Beobachtung ist unklar. Aber man muß danach mit der unglücklichen Möglichkeit rechnen, daß es mehrdeutige Triplets gibt, d. h. solche, die als Chiffre für mehr als eine Aminosäure stehen können. Es ist jedoch zu erwarten, daß derartige Triplets in der Minderzahl sind.

Offenbar lassen sich also die meisten der 64 möglichen Triplets in zwanzig Gruppen ordnen, die den zwanzig Aminosäuren entsprechen. Sowohl aus Untersuchungen mit dem zellfreien System als auch aus Mutationsversuchen geht hervor, daß die Triplets sich auf diese zwanzig Gruppen nicht statistisch verteilen, sondern daß Triplets, die als Chiffre für die gleiche Aminosäure stehen, untereinander sehr ähnlich sein können. Damit aber sind wir bei der zur Zeit wichtigsten Frage: Läßt sich die Verteilung der Triplets aus theoretischen Postulaten ableiten? Man kann sich vorstellen – und leider ist es nicht schwer, das zu tun –, daß die Zuordnung zwischen Triplets und Aminosäuren im Verlauf der Evolution schon sehr früh durch statistische Mutationen entstanden ist, so daß der genetische Code, wie wir ihn heute haben, nur das Ergebnis vieler zufälliger Ereignisse ist. Diese Frage hat eine mehr als nur abstrakte Bedeutung. Wenn nämlich dem Code irgendeine logische Beziehung zugrunde liegt, so ist es gerechtfertigt, jeden Hinweis, sei er gut oder schlecht, zu verwerten, um diese Beziehung zu finden. Dies gilt nicht, wenn die Codons miteinander nicht in einem einfachen logischen Zusammenhang stehen. In einem solchen Fall hätte es wenig Sinn, zu versuchen, ein Codon vorherzusagen. In

jedem Fall aber ist es wichtig, genügend experimentelles Material zu sammeln, um jedes Codon einzeln zu beweisen. Allerdings ist noch nicht klar, wie ein Versuchsergebnis beschaffen sein muß, um als sicherer Beweis für ein Codon gelten zu können. Klar ist lediglich, daß das bisher zusammengetragene Material praktisch in keinem Fall zum Beweis genügt.

Nicht zutreffende Codes

Trotz der Unsicherheit vieler experimenteller Daten kann man einige Codes, die früher vorgeschlagen wurden, verwerfen:

Kommafrie [*] Triplett-Codes sind sowohl auf Grund genetischer Experimente als auch nach den Untersuchungen am zellfreien System unwahrscheinlich.

Codes mit nur zwei oder drei Elementen, beispielsweise ein Code, in dem A mit C und U mit G äquivalent ist, sind mit den Ergebnissen von Versuchen am zellfreien System, wie schon erwähnt, nicht zu vereinbaren.

Der Code der Triplett-Kombinationen (combination triplet code) ist mit den experimentellen Ergebnissen nur in Einklang zu bringen, wenn man sehr spezielle Annahmen macht. In diesem Code stehen alle Sequenzen einer Kombination von drei Basen als Chiffren für die gleiche Aminosäure.

Komplementäre Codes. Hier gibt es mehrere Arten. Das Prinzip ist das folgende: zu jedem Triplett gibt es in der anderen Kette einer Doppelhelix ein komplementäres Triplett. Dieses zweite Triplett kann nun entweder in der gleichen Richtung gelesen werden wie das erste oder in der entgegengesetzten Richtung. Lautet das erste Triplett UCC, so hieße das andere AGG oder (entgegengesetzt gelesen) GGA. Es ist vorgeschlagen worden, daß das erste und das komplementäre Triplett für die gleiche Aminosäure stehen oder daß das komplementäre Triplett keiner Aminosäure entspricht, d. h. Unsinn bedeutet. *Ochoa* und Mitarbeiter [8] haben nun kürzlich zeigen können, daß Poly-A die Verwendung von Lysin zur Proteinsynthese stimuliert. Vermutlich ist AAA also eine Chiffre für Lysin. Da aber UUU dem Phenylalanin entspricht, können die komplementären Codes nicht richtig sein. Gegen diese Codes spricht weiterhin, daß Poly-(U,G) den Einbau einer ganz anderen Aminosäure stimuliert als Poly-(A,C), und ebenso unterscheidet sich Poly-(U,C) von Poly-(A,G) [8,11].

Die Chance, daß sich einer dieser Codes als richtig erweist, ist also sehr gering. Meines Erachtens sind sie auch aus theoretischen Gründen unwahrscheinlich.

Zusammenfassung

Aus unseren heutigen Kenntnissen ergibt sich für den genetischen Code das folgende Bild:

1. Die meisten oder sogar alle Codons bestehen aus drei (einander benachbarten) Basen.

[13] A. J. Wahba, C. Basilio, J. F. Speyer, P. Lengyel, R. S. Miller u. S. Ochoa, Proc. nat. Acad. Sci. USA 48, 1683 (1962).

[*] Als Kommata bezeichnet man Stellen in der DNS, die bestimmen, wie die Triplets ausgewählt werden.

2. Benachbarte Codons überlappen sich nicht.
3. Die im Sinne der genetischen Information richtigen Dreiergruppen ergeben sich, wenn man beim Ablesen der Information an einem bestimmten Punkt beginnt.
4. Die Codon-Sequenz im Gen ist mit der Aminosäure-Sequenz in der Polypeptidkette colinear. Die Aminosäure-Sequenz der Polypeptidkette wird vom Amino-ende her aufgebaut.
5. Im allgemeinen gibt es für eine Aminosäure mehr als ein Triplet.
6. Es ist möglich, daß einige Triplets mehr als einer Aminosäure entsprechen, d. h. mehrdeutig sind.
7. Triplets, die als Chiffren für die gleiche Aminosäure stehen, sind wahrscheinlich einander ähnlich.
8. Es ist nicht bekannt, ob die Zuordnung von Triplets und Aminosäuren irgendeiner allgemeinen Regel folgt oder lediglich das Ergebnis zufälliger Ereignisse ist.

9. Die Zahl der Triplets, die nicht für eine Aminosäure stehen, ist wahrscheinlich klein.

10. Einige früher vorgeschlagene Codes (z. B. komma-freie Codes, Codes mit zwei oder drei Elementen, der Code der Triplet-Kombinationen und einige komplementäre Codes) sind wahrscheinlich nicht richtig.

11. Die genetischen Codes verschiedener Organismen sind einander wahrscheinlich ähnlich. Möglicherweise haben alle Organismen den gleichen Code, aber das ist noch nicht bekannt.

Trotz der Vielschichtigkeit des Problems der Proteinsynthese und trotz der beträchtlichen Schwierigkeiten, die der Synthese von Polynucleotiden mit bekannter Nucleotid-Sequenz noch entgegenstehen, darf man hoffen, daß sich alle hier aufgeworfenen Fragen in naher Zukunft klären lassen, und daß der genetische Code innerhalb weniger Jahre vollständig bewiesen sein wird.

Eingegangen am 2. Januar 1963 [A 291]
Übersetzt von Dr. H. Grünwald, Heidelberg

Die molekulare Konfiguration der Nucleinsäuren

Nobel-Vortrag am 11. Dezember 1962 [*]

VON PROF. DR. M. H. F. WILKINS

MEDICAL RESEARCH COUNCIL, BIOPHYSICS RESEARCH UNIT,
BIOPHYSICS DEPARTMENT, KING'S COLLEGE, UNIVERSITY OF LONDON (ENGLAND)

Nucleinsäuren sind im Prinzip einfach gebaut. Sie stehen an der Wurzel so bedeutender biologischer Prozesse wie Wachstum und Vererbung. Die Einfachheit ihrer Struktur und der Beziehung zwischen Struktur und Funktion entspricht einer tieferliegenden Einfachheit der biologischen Vorgänge und hat eine erste weitgehende Interpretation solcher Vorgänge an Hand makromolekularer Strukturen ermöglicht. All dies gelang nur durch eine vorher nicht dagewesene Kombination biologischer, chemischer und physikalischer Untersuchungen, von der Genetik bis zur Stereochemie der Wasserstoffbrückenbindung. Ich werde darauf nicht weiter eingehen, sondern mich auf mein Arbeitsgebiet beschränken und zeigen, welchen Beitrag Röntgenstrukturanalysen geleistet haben. Ich möchte auch einiges über die Hintergründe meiner Arbeiten sagen, denn ich bin sicher nicht der einzige, der solche Berichte oft interessanter findet als eine allgemeine Zusammenfassung.

1. Anfänge

Ich promovierte 1938 als Physiker in Cambridge und hatte unter dem Einfluß von *J. D. Bernal* einige Ausbildung als Röntgenkristallograph erhalten. In Birmingham arbeitete ich dann bei *J. T. Randall* über Fra-

gen der Lumineszenz und der Elektronenbewegung in Kristallen. Meine Kollegen in Cambridge hatten sich vor allem mit Elementarteilchen beschäftigt. Mir erschienen die Organisation des festen Zustandes und die darauf beruhenden besonderen Eigenschaften bemerkenswerter. Vielleicht war dies eine Vorstufe meines Interesses an biologischen Makromolekülen und am Zusammenhang zwischen ihrer Struktur und der Fähigkeit, den Ablauf der Lebensvorgänge so weitgehend zu bestimmen.

Während des Krieges beteiligte ich mich an der Herstellung der Atombombe und suchte, als er vorüber war, wie so viele andere nach einem neuen Arbeitsgebiet. Nicht zuletzt infolge der Bombe war meine Freude an der Physik geringer geworden. Ich las damals *Schrödingers* Buch „What is Life“ und wurde von der Vorstellung, daß eine sehr komplexe molekulare Struktur die Lebensvorgänge kontrolliere, in den Bann gezogen. Die Untersuchung solcher Fragen erschien mir lohnender als die Festkörperphysik. Damals glaubten viele bedeutende Physiker, u. a. *Massey*, *Oliphant* und *Randall*, daß die Physik einen wesentlichen Beitrag zur Biologie leisten könne (ich erfuhr später, daß auch *Nils Bohr* dieser Meinung war). Ihr Rat ermutigte mich zur Beschäftigung mit der Biologie.

Randall lud mich ein, an biophysikalischen Untersuchungen teilzunehmen, die er am physikalischen Institut der St. Andrews-Universität in Schottland begonnen hatte. Angeregt durch *Müllers* Arbeiten über die

[*] Wir danken dem Autor und dem Nobel-Komitee, Stockholm, für die Genehmigung zum Druck dieses Vortrags.