

Unverkäufliche Leseprobe aus:

Arndt von Haeseler, Dorit Liebers
Molekulare Evolution

Alle Rechte vorbehalten. Die Verwendung von Text und Bildern, auch auszugsweise, ist ohne schriftliche Zustimmung des Verlags urheberrechtswidrig und strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

© S. Fischer Verlag GmbH, Frankfurt am Main

MOLEKULARE EVOLUTION

GRUNDRISS

1. Einige Grundlagen der Vererbung	3
DNA – ein Bote aus der Vergangenheit	3
Die DNA und ihr Code	5
Gene und Genom	7
Die Protein-Biosynthese	10
2. Veränderungen einer DNA-Sequenz im Laufe der Zeit	14
Mutationen in einer DNA-Sequenz	14
Evolution einer DNA-Sequenz	17
Lücken und Ergänzungen in DNA-Sequenzen	22
3. Eine kleine Baumschule	25
Allgemeine Terminologie	25
Phylogenetische Klassifikation	28
Die Zahl der Bäume	31
4. Molekulare Phylogenie	32
Maximum-Parsimonie	36
Distanzbasierte Methoden	41
Maximum-Likelihood	45
Experimentelle und theoretische Phylogenien	48
Der Bootstrap	55
5. Gen-Bäume in der Phylogenie	56
Gen-Bäume in Spezies-Bäumen	56
Widersprüche zwischen Gen-Bäumen und Spezies-Bäumen ...	58
Auswirkungen von Gen-Duplikationen auf Gen-Bäume	63
Gen-Duplikationen als Motor der physiologischen Feinabstimmung	64

6. Gen-Bäume in Populationen	66
Rekonstruktion der Populationsgeschichte	
anhand von DNA-Sequenzen	68
Die Genealogie einer Stichprobe	70
Wann lebte der jüngste Vorfahre	
der menschlichen Population?	72
Demographie	76
Wo kommen wir her?	85
7. Die Zukunft	87

VERTIEFUNGEN

Molekulare Techniken	91
Genetische Drift	100
Die neutrale Theorie der molekularen Evolution	101
Die molekulare Uhr	103
Der Coalescent-Prozess	105
Die genetische Variabilität einer Population	109
Das Jukes-Cantor-Modell der Sequenzevolution	110
Wer sind die nächsten Verwandten der Wale? Ein nicht-	
sequenzbasierter Ansatz zur Aufklärung der Phylogenie	112
»Fossile DNA« – eine Zeitreise in die Vergangenheit	115

ANHANG

Glossar	122
Literaturhinweise	127

1 EINIGE GRUNDLAGEN DER VERERBUNG

DNA – ein Bote aus der Vergangenheit

Die Menschheit ist nicht nur daran interessiert, ihre Zukunft zu deuten, sondern auch ihre Herkunft zu rekonstruieren. Zentrale Fragen sind: Wo kommen wir her? Wie ordnen wir uns in die belebte Welt ein? Und wie können wir alle Lebewesen in einem einheitlichen Schema gruppieren, das uns hilft die Vielfalt des Lebens auf der Erde und ihre Genese zu verstehen?

Carl von Linné (1707–1778) »sortierte« die Natur und gab den meisten Tieren und Pflanzen mittels der binären Nomenklatur eine systematische Zuordnung. Die Dynamik in der Entstehung der Vielfalt blieb jedoch verborgen. Erst ein Jahrhundert später, auf einem Treffen der Londoner Linnean Society am 1. Juli 1858, haben Charles Darwin (1809 – 1882) und Alfred Russel Wallace (1823 – 1913) ihre Ideen zur Entstehung der Arten vorgetragen. Im Jahr darauf publizierte Darwin sein bahnbrechendes Buch *On the Origin of Species*. Darwins Theorie basierte entscheidend auf der Weitergabe von vererbaren Merkmalen. Die zugrunde liegenden Mechanismen waren zu dieser Zeit jedoch noch unklar. Erst 1944 gelang dem amerikanischen Team um Oswald Theodore Avery (1877 – 1955) der eindeutige Nachweis, dass die Desoxyribonukleinsäure (*desoxyribonucleic acid*, abgekürzt DNA) die erblichen Eigenschaften von den Eltern auf die Nachkommen überträgt.

Vererbung beruht also auf einer stofflichen Weitergabe in Form einer Umsetzung von Molekülen. Die DNA besteht aus vier Grundbausteinen, nämlich den Basen Adenin (A), Cytosin (C), Guanin (G)

und Thymin (T). Die Molekular-Genetik beschreibt, wie aus der sequentiellen Abfolge dieser vier Grundbausteine der gesamte Bauplan für einen vollständigen Organismus entsteht und wie die Weitergabe der genetischen Information erfolgt. Da die DNA in fast allen Lebewesen als Träger der Erbinformation fungiert, sind die gefundenen Prinzipien für die Umsetzung der genetischen Information in den Phänotypen nahezu allgemeingültig.

Rasch wurde klar, dass die DNA nicht nur für die Weitergabe der Eigenschaften von Eltern auf ihre Kinder verantwortlich ist. Sie ist auch ein »Dokument der Evolutionsgeschichte«, so Emile Zuckerkandl und Linus Pauling. Die DNA, die in heute lebenden Organismen zu finden ist, gab in grauer Vorzeit ein Vorfahr an den Nächsten weiter. Im Laufe dieser Weitergabe wurde die DNA modifiziert. Nicht mehr benötigte Segmentabschnitte gingen verloren, neue Sequenzabschnitte wurden erworben und wieder andere Abschnitte erfuhren kleine Veränderungen, da der Prozess der Informationsweitergabe nicht fehlerfrei ist. Welche Modifizierungen auch immer eine DNA erfahren hat, die heutigen Organismen zeigen Spuren dieser Änderungen in ihrem Genom. Die Forschung zur molekularen Evolution versucht diesen Prozess zu rekonstruieren und die Mechanismen herauszuarbeiten, die zu der heutigen Vielfalt der Organismen geführt haben.

Besonders in den letzten Jahrzehnten wurden völlig neue Forschungstechniken entwickelt. Beispiele sind die Klonierung von DNA-Segmenten, die Sequenzierung der DNA und die Polymerase-Kettenreaktion (PCR) (**Molekulare Techniken**). Mit der rasanten Entwicklung dieser molekularbiologischen Techniken gelang es, ein detailliertes und immer umfassenderes Bild von den der Evolution zugrunde liegenden Mechanismen zu entwickeln.

Insbesondere bei der Frage nach einem biologischen System der Organismen erweist sich die Analyse von DNA- und Aminosäuresequenzen als eine wertvolle Methode, um sowohl die Verwandt-

schaftsverhältnisse zwischen nah verwandten Arten zu studieren, als auch einen Baum zu rekonstruieren, der die Evolutionsgeschichte aller Organismen der Erde darstellt.

Die DNA und ihr Code

Die genetische Zusammensetzung eines Organismus wird im Wesentlichen durch die Nukleinsäuren bestimmt. Sie enthalten den Bauplan, der die verschiedenen Bauphasen im Organismus steuert und der als Kopie an die nächste Generation weitergegeben wird. Es gibt in den Zellen zwei Arten von Nukleinsäuren, die Desoxyribonukleinsäure (DNA) und die Ribonukleinsäure (RNA). Nukleinsäuren sind Makromoleküle. Ihre Grundbausteine sind Nukleotide, die kettenförmig miteinander verbunden sind. Jedes Nukleotid ist aus drei Molekülen aufgebaut: einer stickstoffhaltigen, heterozyklischen Base (N), einem Zucker (Z) und einer Phosphorsäure (P) (Abbildung 1a).

Als Zuckerbaustein dient bei der DNA die Desoxyribose, bei der RNA ist es die Ribose. Als Basenanteile treten die Pyrimidinderivate Thymin (T) und Cytosin (C) (einfache Ringstruktur) und die Purinderivate Adenin (A) und Guanin (G) (doppelte Ringstruktur) auf (Abbildung 1b). In der RNA kommt statt Thymin die Base Uracil (U) vor, die chemisch nah verwandt ist mit Thymin.

Ein vollständiges DNA-Molekül besteht aus zwei gegenläufigen Polynukleotid-Strängen (Abbildung 1c). Diese sind über Wasserstoffbrücken-Bindungen zu einem Doppelstrang so verknüpft, dass sich immer Thymin beziehungsweise Cytosin des einen Strangs mit Adenin beziehungsweise Guanin vom anderen Strang paaren (Watson-Crick-Basenpaarung). Dabei werden zwischen Adenin und Thymin zwei Wasserstoffbrücken-Bindungen ($A=T$) ausgebildet, zwischen Guanin und Cytosin sind es drei ($G=C$). Zusätzlich sind diese zwei Polynukleotid-Stränge noch spiralförmig umeinander gewunden, und es entsteht die charakteristische Gestalt der DNA-Doppelhelix.

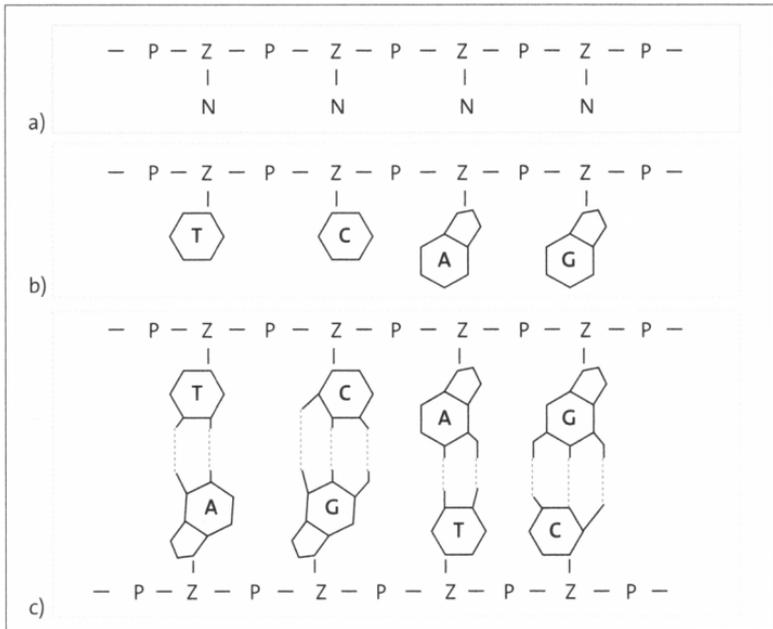


Abb.1: Schematischer Aufbau der DNA

In einer Nucleinsäure sind mehrere tausend Nucleotide zu einem langen, unverzweigten Strang angeordnet. Die genetischen Informationen werden in der unterschiedlichen Abfolge der vier Basen entlang der jeweiligen Sequenz kodiert.

Die genetische Information einer DNA-Sequenz wird in der Protein-Biosynthese an die Eiweißmoleküle (Proteine) weitergegeben, die wiederum die spezifischen Merkmale eines Organismus prägen. Die Bausteine der Proteine sind zwanzig verschiedene Aminosäuren; am Aufbau der Nucleinsäuren hingegen sind nur die vier Basen beteiligt. Zur Kodierung der zwanzig Aminosäuren sind daher spezifische »Codewörter« notwendig. Sie geben an, aus welchen Nucleotiden die Aminosäuren bestehen. Eine einfache Überlegung zeigt, dass mindestens drei Nucleotide (zum Beispiel ACG) notwendig sind, um

alle Aminosäuren zu kodieren. Aus vier Nukleotiden können nur 16 Kombinationen von Zweiergruppen (zum Beispiel GC), wohl aber 64 Kombinationen von Dreiergruppen (wie CGA) gebildet werden. Jeweils drei Nukleotide lassen sich also zu einem Wort aus drei Buchstaben oder in der Sprache der Molekularbiologie zu einem Codon oder Tripletts zusammenfassen. So stehen genügend Wörter zur Verfügung, um jede Aminosäure durch die Abfolge von drei Nukleotiden zu verschlüsseln. Die Übersetzung der Tripletts der DNA in Aminosäuren der mRNA zeigt Abbildung 2.

Die meisten Aminosäuren werden durch mehrere Tripletts kodiert. So dienen als Codon für die Aminosäure Serin (Ser) die sechs Wörter AGC, AGU, UCU, UCC, UCA oder UCG. Man spricht daher von der Redundanz des genetischen Codes. Eindeutig kann nämlich nur von der DNA- beziehungsweise RNA-Sequenz auf die Aminosäure geschlossen werden, nicht jedoch umgekehrt von der Aminosäure auf die Nukleotid-Sequenz. Lediglich für Tryptophan (Trp) und Methionin (Met) gibt es allein ein einziges Schlüsselwort, nämlich UGG respektive AUG. Das AUG-Triplett hat weiterhin die Funktion eines so genannten Startcodons, das den Beginn der kodierenden Sequenz anzeigt. Jedes neu synthetisierte Protein beginnt also mit Methionin. Zu den so genannten Stoppcodons UGA, UAA, und UAG gehören keine Aminosäuren. Diese Tripletts beenden die Protein-Biosynthese.

Gene und Genom

Der DNA-Strang enthält viele tausend Nukleotide. Aber nur einige Abschnitte der DNA tragen die Informationen für den Bauplan eines Organismus in sich. Diese Abschnitte heißen Gene. Sie enthalten die Informationen zur Herstellung von spezifischen Proteinen und sind daher im Wesentlichen für die Gestalt eines Lebewesens verantwortlich. Für die Herstellung der Proteine müssen die Gene ihre Kodierung weitergeben, sie bilden die kodierende DNA.

Darüber hinaus gibt es in dem DNA-Strang zwischen den kodierenden Bereichen weitere, häufig sehr lange Abschnitte, die keine Informationen zur Herstellung von Proteinen tragen. Damit differenziert sich der DNA-Strang in verschiedene Domänen: Gene, die ihre Kodierung weitergeben, bilden die funktionellen Bereiche der DNA. Die anderen Abschnitte, die keine Kodierung tragen und daher auch keine Kodierung weitergeben, bilden die nicht-kodierende DNA. Letztere machen bei den Lebewesen, die einen echten Zellkern haben (Eukaryoten), den Großteil des Genoms aus (siehe Abbildung 3).

Seit langem bekannt ist die Unterteilung des Genoms in Chromosomen. Im Kern einer menschlichen Körperzelle befinden sich 22 autosomale Chromosomen (griechisch *auto* = selbst, *soma* = Körper). Sie steuern hauptsächlich die körpereigenen Prozesse. Die Chromosomen unterscheiden sich in Form und Größe und liegen je zweimal vor. Ein Pärchen gleichartiger Chromosomen heißt homolog (übereinstimmend). Hinzu kommen die Geschlechtschromosomen X und Y. Bei männlichen Individuen gibt es ein X- und ein Y-Chromosom, bei weiblichen Individuen zwei X-Chromosomen. Zellen, in denen die Chromosomen doppelt vorliegen, heißen diploid (zweifach). Eine diploide Zelle des Menschen enthält daher immer 46 Chromosomen, zweimal 22 homologe Autosomen und zwei Geschlechtschromosomen, entweder XY bei männlichen oder XX bei weiblichen Organismen. In den menschlichen Keimzellen (Spermien und Ei) ist die Anzahl der Chromosomen halbiert, sie enthalten nur je einen autosomalen Chromosomensatz ($n=22$) und von den Geschlechtschromosomen entweder das Y- oder das X-Chromosom. In den reifen Geschlechtszellen befinden sich daher 23 Chromosomen. Im Gegensatz zu den diploiden Körperzellen sind die Keimzellen haploid (einfach).

Die molekulare Differenzierung des menschlichen Genoms ist erst in den letzten Jahren aufgeklärt worden. Nach der vollständigen Bestimmung der Abfolge und Anzahl der Nukleotide wurde mit Erstaunen festgestellt, dass ca. 97% der drei Milliarden Basen nicht-

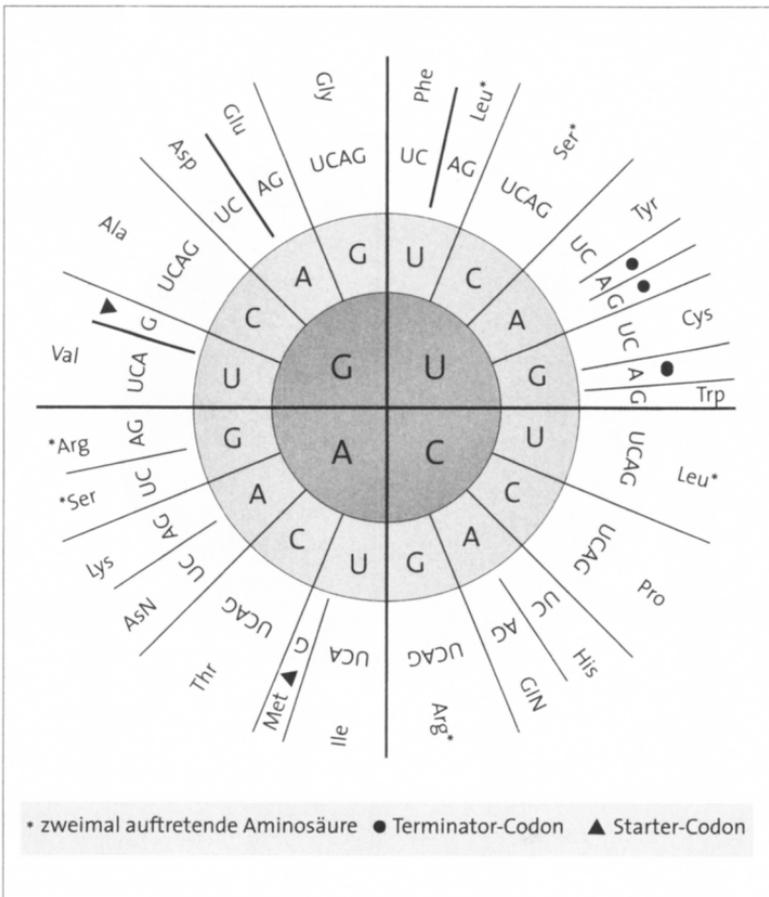


Abb. 2: Der genetische Code für die mRNA. Die Codons sind von innen nach außen zu lesen und geben die Basenabfolge der mRNA-Codons wieder. Außerhalb des Kreises stehen die Aminosäuren, die vom Triplet kodiert werden.

kodierende DNA sind. Sich vielfach wiederholende Abschnitte (repetitive DNA) mit bisher unbekannter Funktion machen ca. 40% des gesamten Genoms aus. In der Forschung werden sie je nach Länge als SINE, LINE oder Satelliten-DNA bezeichnet. Lediglich 3% des menschlichen Genoms sind kodierende DNA. Neben den schät-

zungsweise 30 000–40 000 Genen sind auch regulatorische Sequenzen und spezielle RNA-Moleküle bei der Synthese von Proteinen wichtig.

Das eukaryotische Genom enthält neben den im Zellkern lokalisierten Chromosomen (auch Kern-DNA genannt) zusätzliche extrachromosomale DNA, die in den Mitochondrien der Zellen vorkommt. Sie heißt mitochondriale DNA (mtDNA). Beim Menschen handelt es sich hierbei um ein kleines, ca. 16 000 Basenpaare langes, ringförmiges DNA-Molekül. Die mtDNA wird in der Regel maternal vererbt. Nur die Mütter geben das Mitochondrien-Genom an ihre Kinder weiter, die väterliche mtDNA wird nicht vererbt. Im Unterschied zur Kern-DNA ist die Abfolge der kodierenden Abschnitte nahezu lückenlos. Der einzige nicht-kodierende Abschnitt ist die Kontrollregion (Abbildung 3). Sie steuert die Replikation (originalgetreue Nachbildung) des ringförmigen Genoms. Eine weitere Eigenheit der mitochondrialen DNA besteht darin, dass es nach dem derzeitigen Wissensstand so gut wie keine Rekombination gibt, das heißt es findet kein Austausch zwischen verschiedenen DNA-Abschnitten statt. Diese Tatsache macht die mtDNA besonders geeignet für evolutionsbiologische Analysen.

Pflanzen besitzen noch ein weiteres, extra-chromosomales Genom, das in den Plastiden der Zellen vorkommt und daher Plastiden-Genom heißt. Zu den Plastiden zählen unter anderem die grünen Chloroplasten, die maßgeblich für die Photosynthese verantwortlich sind sowie die rötlich bis gelben Chromoplasten der reifen Früchte und Blüten. Das Plastiden-Genom ist ebenfalls ringförmig geschlossen und hat eine Länge von 85 000–190 000 Basenpaaren.

Die Protein-Biosynthese

Ein Gen trägt die Information zur Bildung eines spezifischen Eiweißmoleküls (Protein). Diese sind vorwiegend aus Aminosäuren aufge-

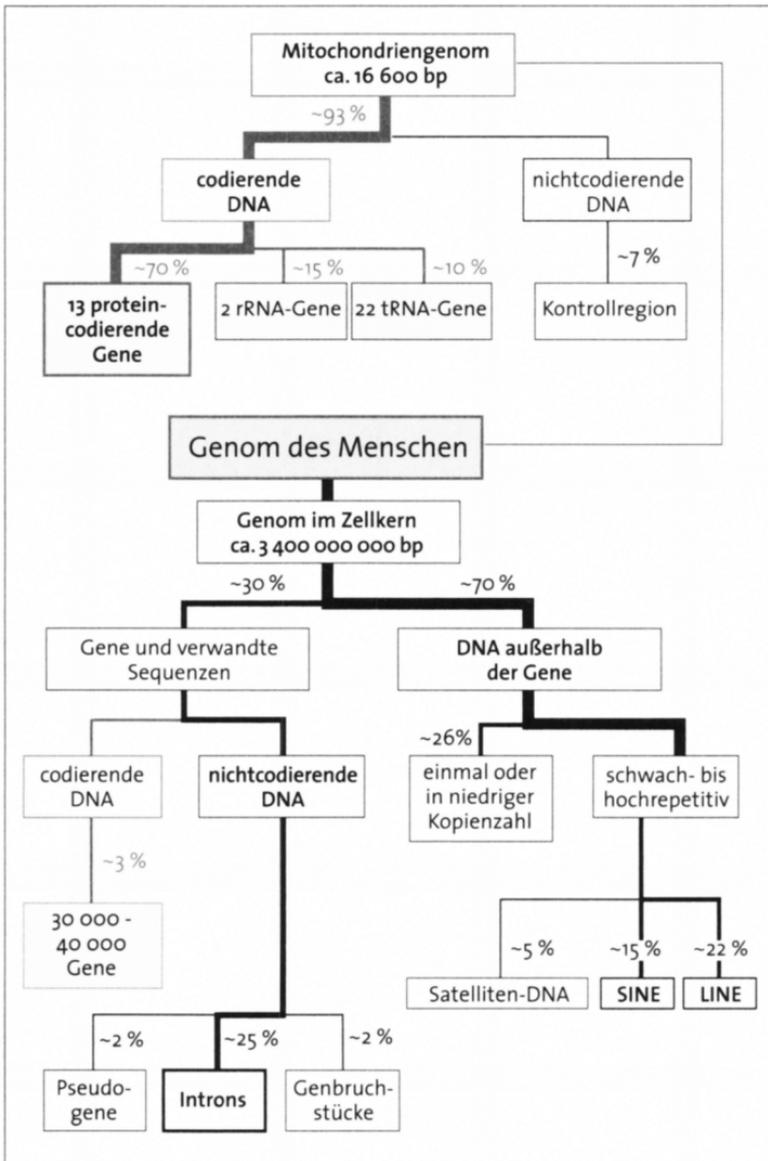


Abb.3: Anteil kodierender und nicht-kodierender DNA im menschlichen Genom

baut. Ganz ähnlich wie bei der DNA bestimmt die Abfolge der Aminosäuren in der Proteinsequenz die charakteristischen Eigenschaften dieses Proteins. Da die Gene in der Abfolge ihrer Nukleinsäuren die Informationen zum Aufbau der Proteine enthalten, muss es einen Mechanismus geben, der die Abfolge dieser Nukleinsäuren eines Gens in eine Abfolge von Aminosäuren übersetzt. Dieser Vorgang ist die Protein-Biosynthese, sie verläuft in zwei Schritten.

Bei den Eukaryoten befindet sich die DNA im Zellkern. Die Eiweißsynthese erfolgt aber außerhalb des Zellkerns an den Ribosomen im Cytoplasma. Daher muss in einem ersten Schritt die Information aus dem Kern durch die Kernhülle zu den Ribosomen im Cytoplasma transportiert werden. Diese Übertragung übernimmt ein einsträngiges RNA-Molekül. Da dieses die »Botschaft« nach außen überträgt, heißt das Molekül *messenger*-RNA (mRNA oder Boten-RNA). Die mRNA wird im Zellkern an der Kern-DNA gebildet. Die Basenfolge (die genetische Information) der Kern-DNA wird dabei auf das neu gebildete mRNA-Molekül übertragen (kopiert). Dieser erste Schritt der Protein-Biosynthese wird Transkription (Abbildung 4) genannt. Anstelle von Thymin in der DNA wird in die mRNA jedoch die Base Uracil eingebaut.

Die mRNA gelangt durch die Kernporen in das Cytoplasma. Jetzt beginnt der zweite Schritt in der Übertragung der genetischen Information der Kern-DNA auf die Proteinbildung. Im Cytoplasma heften sich zwei Teile eines Ribosoms an die mRNA an und bilden ein funktionsfähiges Ribosom. Zugleich binden weitere RNA-Moleküle je eine bestimmte, in den Zellen frei existierende Aminosäure an sich. Diese RNA-Moleküle nennt man *transfer*-RNA oder tRNA. Sie transportieren die Aminosäuren zum Ribosom, wo sie unter Mitwirkung der mRNA zu einem Polypeptid verknüpft werden. Die Reihenfolge, in der die Aminosäuren zu einem bestimmten Protein zusammengesetzt werden, wird durch die Abfolge der Codons in der mRNA bestimmt. Dieser zweite Schritt, die Übersetzung der in der Basen-