Teil I
Allgemeine Grundlagen und Präanalytik

COPYRIGHTED MATERIAL

1
Grundlagen der molekularen Diagnostik

Frank Thiemann

Viren, Bakterien oder Parasiten können aus verschiedensten Probenmaterialien spezifisch und sensitiv mit molekularbiologischen Methoden nachgewiesen werden. Auch Mutationen im menschlichen Genom werden, je nach Fragestellung, mit unterschiedlichen Techniken detektiert.

Das entsprechende Erbgut, die DNA oder RNA, muss dazu zunächst isoliert und anschließend spezifisch vermehrt (amplifiziert) werden.

Für den Nachweis und die Vervielfältigung der Nukleinsäuren existieren mannigfaltige Verfahren und Methoden. Diese verschiedenen Techniken bedienen sich dabei der spezifischen Eigenschaften dieser Moleküle, deren chemische Grundstruktur immer die gleiche ist.

1.1
Die DNA

Träger der Erbinformation ist in allen Lebewesen die *D*esoxyribo*n*ukleins*ä*ure (DNS; das A in DNA steht für engl. acid = Säure). Die DNA setzt sich aus vier unterschiedlichen *Nukleotiden* zusammen. Ein Nukleotid besteht aus einem *Zucker*, einem *Phosphat* und aus einer der vier stickstoffhaltigen Basen *Adenin, Thymin, Cytosin oder Guanin*. Adenin und Guanin bezeichnet man aufgrund ihrer chemischen Struktur als *Purine*, Thymin und Cytosin als *Pyrimidine*.

Ein *Nukleosid* ist eine Verbindung aus Base und Zucker ohne Phosphat.

Der Zucker ist eine *Ribose*, deren Grundgerüst aus fünf Kohlenstoffatomen (C) aufgebaut ist (Pentose). Diese Kohlenstoffatome sind von 1' bis 5' gekennzeichnet. Am 1'-C-Atom ist jeweils eine Base mit dem Zucker verknüpft. Am 2'-C-Atom der Ribose ist keine OH-Gruppe vorhanden, sondern ein Wasserstoffatom. Aus diesem Grund bezeichnet man diese Ribose auch als *Desoxyribose*. Die Nukleotide sind über eine *Phosphodiesterbindung*, die zwischen der OH-(Hydroxyl-) Gruppe des 3'-C-Atoms und dem Phosphat des 5'-C-Atoms des benachbarten Nukleotids gebildet wird, kovalent miteinander verbunden. Es entsteht somit eine Polynukleotidkette mit einem *Zucker-Phosphat-Rückgrat*, deren beiden Enden sich chemisch voneinander unterscheiden. An dem einen Ende ist die *3'-OH-*

Molekulare Diagnostik: Grundlagen der Molekularbiologie, Genetik, Analytik, 2. Auflage.
Herausgegeben von Dr. F. Thiemann, Prof. Dr. P.M. Cullen und Dr. H.-G. Klein
© 2015 WILEY-VCH Verlag GmbH & Co. KGaA. Published 2015 by Wiley-VCH Verlag GmbH & Co. KGaA

Gruppe, an dem anderen Ende die *freie Phosphatgruppe* des 5'-C-Atoms lokalisiert. Jeder DNA-Strang hat somit eine chemische Polarität. Konventionell bezeichnet man die beiden Enden als *5'- und 3'-Ende* (Abb. 1.1, Abb. 1.2).

Die dreidimensionale Struktur der DNA wurde 1953 von Watson und Crick mithilfe von Röntgenstrukturanalysen, die Rosalind E. Franklin und Maurice Wilkins am Londoner King's College durchführten, aufgeklärt. Sie postulierten anhand der erzielten Ergebnisse die doppelhelikale (schraubenförmig gewundene) Struktur der DNA. In der *DNA-Doppelhelix* winden sich zwei DNA-Stränge unterschiedlicher Polarität um eine senkrecht im Raum stehende Achse. Das

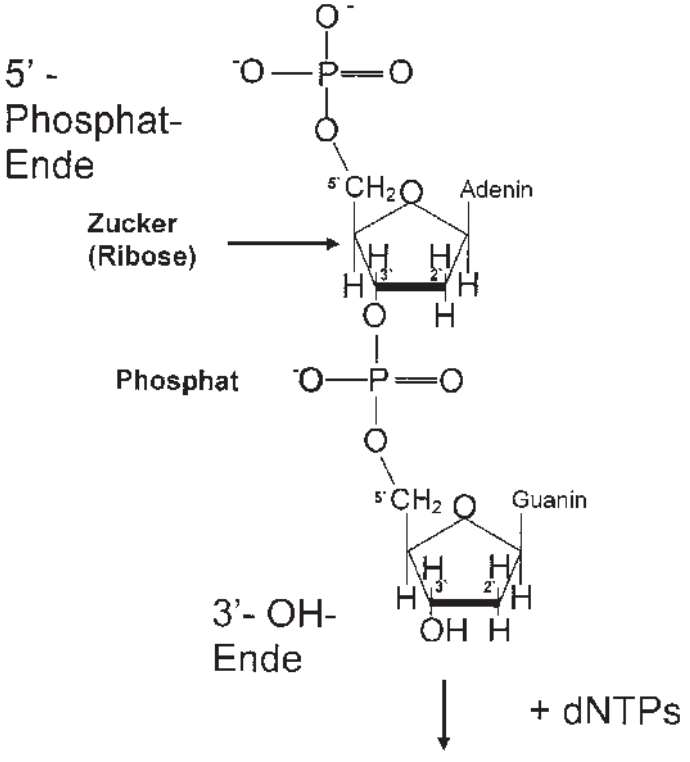


Abb. 1.1 Ausschnitt eines DNA-Einzelstranges.

Die DNA besteht aus einem Zucker (Ribose), der jeweils am 5'- und 3'-Kohlenstoff (C) über ein Phosphat (P) miteinander verknüpft ist (Zucker-Phosphat-Rückgrat). Mit dem ersten Kohlenstoffatom des Zuckers sind die einzelnen Basen Guanin, Cytosin, Thymin oder Adenin verbunden. Am 5'-Ende einer DNA steht immer ein Phosphat, am 3'-Ende ist die OH-Gruppe lokalisiert. An dieser 3'-OH-Gruppe wird von der DNA-Polymerase ein weiteres Nukleotid eingebaut.

Am 2'-Kohlenstoff der Ribose sind bei der DNA zwei Wasserstoffatome (H) gebunden. Bei der RNA ist hier eine weitere OH-Gruppe vorhanden. Daher die Bezeichnungen RNA = *R*ibon*u*kleins*ä*ure und DNA = *D*esoxyribon*u*kleins*ä*ure (*A* = *acid* = *engl.: Säure*) (Abbildung aus: *MTA-Dialog*, Heft 4, 2000, mit freundlicher Genehmigung der Hoppenstedt Bonnier Zeitschriften GmbH, Darmstadt).

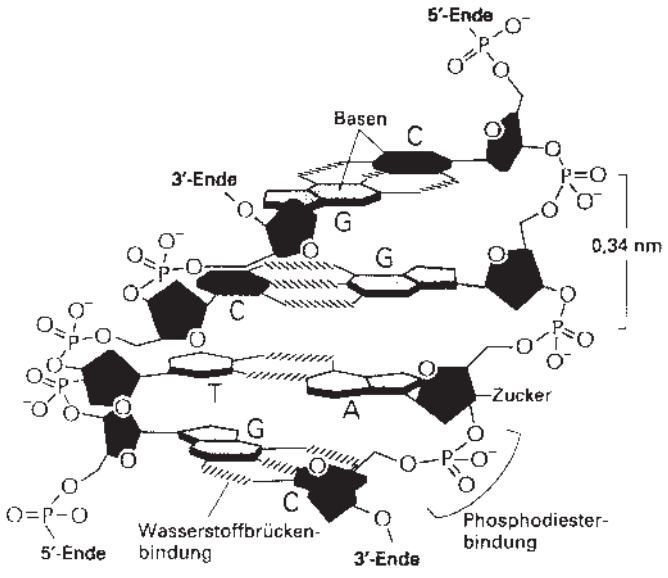


Abb. 1.2 Ausschnitt eines DNA-Doppelstranges.

In einem DNA-Doppelstrang sind Guanin und Cytosin über drei, Adenin und Thymin über zwei Wasserstoffbrückenbindungen komplementär miteinander verbunden. Daher sind DNA-Abschnitte mit einem hohen GC-Gehalt stabiler als Abschnitte, die reich an Adenin und Thymin sind (Abbildung verändert aus: Alberts, B. et al., 2012, *Lehrbuch der molekularen Zellbiologie*, 4. Auflage, Wiley-VCH Weinheim).

Zucker-Phosphat-Rückgrat zeigt dabei nach außen. Die Phosphatreste sind bei neutralem pH-Wert negativ geladen und geben damit dem gesamten Molekül in wässriger Lösung eine negative Ladung. Die Basen liegen im Innern der Doppelhelix.

Die Stabilität der Doppelhelix wird durch die Stapelwechselwirkung der übereinander liegenden Basen und zu einem geringeren Teil auch durch die *Wasserstoffbrückenbindungen* zwischen den Basen der gegenüberliegenden DNA-Stränge gewährleistet.

Diese nicht kovalenten Bindungen verbinden die einzelnen DNA-Stränge. Die chemische Struktur der Basen erlaubt dabei nur eine Paarung zwischen Adenin und Thymin (zwei Wasserstoffbrückenbindungen) und zwischen Cytosin und Guanin (drei Wasserstoffbrückenbindungen). Die Basenabfolge des einen Stranges determiniert somit die Basenabfolge des anderen Stranges. Man bezeichnet zwei so miteinander verbundene DNA-Stränge als *komplementäre* DNA-Stränge und die entsprechenden Basen als komplementäre Basen oder Basenpaar.

Eine Umwindung in der DNA-Doppelhelix umfasst 10,4 Basenpaare. Dreidimensional ergibt sich daraus ein lang gestrecktes gedrehtes Makromolekül mit zwei „Furchen“ (Abb. 1.3). Die „kleine Furche“ hat eine Breite von 1,2 nm, die

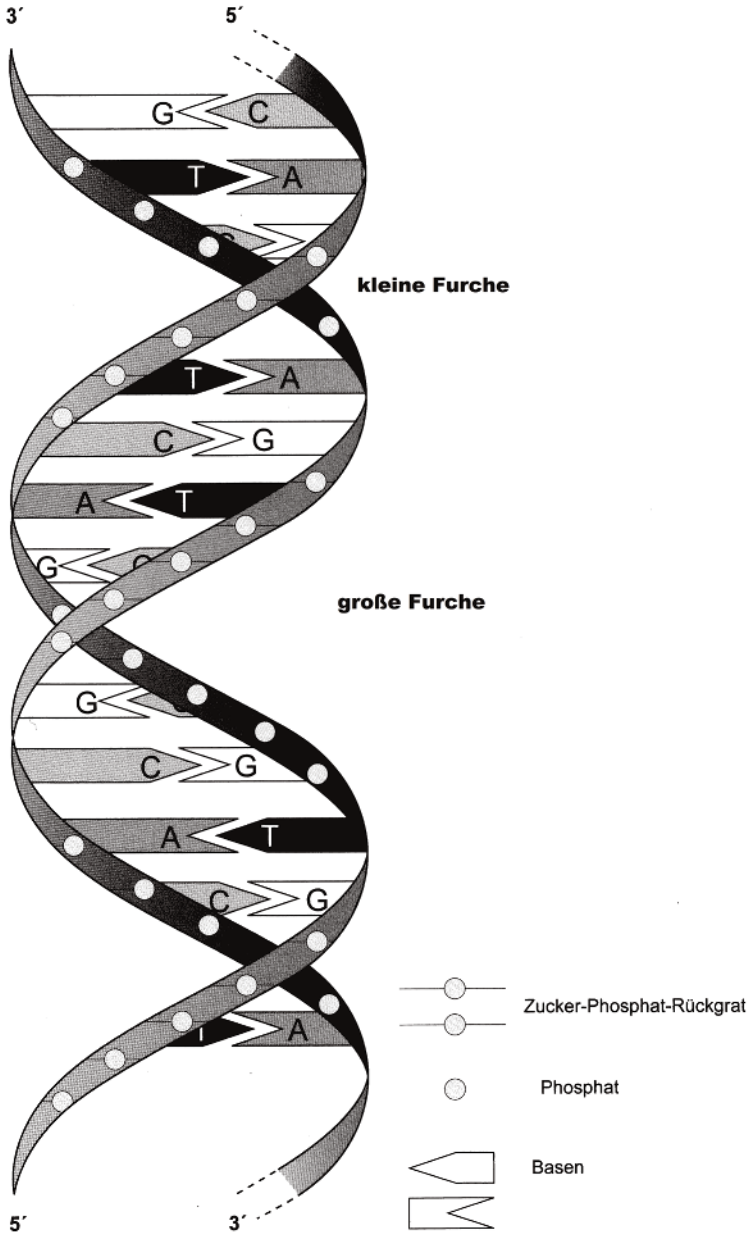


Abb. 1.3 Die DNA-Doppelhelix (Watson-Crick-Modell).

Die beiden gegenläufigen DNA-Stränge winden sich in einer rechtshändigen Spirale umeinander, wobei die Basen ins Innere der Helix weisen und komplementär über Wasserstoffbrückenbindungen miteinander verbunden sind. Eine Umwindung in der Doppelhelix umfasst 10,4 Basenpaare (Abbildung aus: *MTA-Praxis*, Sonderheft 1, 2002, mit freundlicher Genehmigung der Hoppenstedt Bonnier Zeitschriften GmbH, Darmstadt).

„große Furche“ eine Breite von 2,2 nm. Diese Strukturen bieten Proteinen oder Farbstoffen Bindungsmöglichkeiten.

In der Abfolge der Basen auf der DNA ist die genetische Information verschlüsselt. Da die DNA-Stränge in einer Doppelhelix komplementär sind, kann jeder Strang als Vorlage (*Matrize*) zur Herstellung eines neuen DNA-Stranges dienen. Werden die Basen der beiden Matrizenstränge komplementär ergänzt, erhält man zwei identische dopppelsträngige DNA-Moleküle. Die genetische Information auf der DNA wird so verdoppelt oder *repliziert*. Da bei der *Replikation* jeweils ein DNA-Strang der DNA-Doppelhelix erhalten bleibt und als Vorlage dient, spricht man auch von der *semikonservativen Replikation* der DNA. In der Zelle (*in vivo*) läuft dieser Vorgang während jeder Zellteilung ab, damit gewährleistet wird, dass jede Tochterzelle die gleiche genetische Information erhält (s. Abschnitt 1.3).

In Kapitel 4 wird gezeigt, wie man sich diese Eigenschaft der DNA zunutze macht, um diese *in vitro* zu vervielfältigen (*amplifizieren*) und nachzuweisen.

1.2 Die RNA

Die in der DNA gespeicherte genetische Information muss in jeder Zelle abgerufen und in die Bildung von Proteinen umgesetzt werden. Diese Umsetzung und Informationsübertragung ist eine der wesentlichen Funktionen der RNA (s. Abschnitt 1.7). Es existieren aber nicht nur RNA-Moleküle, die der Informationsübertragung dienen, sondern auch solche, die weitere spezielle Aufgaben in der Zelle übernehmen.

Von ihrer chemischen Zusammensetzung unterscheidet sich die RNA nur in zwei wesentlichen Punkten von der DNA (Abb. 1.4).

An der 2’-Position der Ribose ist eine Hydroxygruppe (OH-Gruppe) anstatt eines Wasserstoffatoms lokalisiert. Aus diesem Grund ist RNA weniger stabil als DNA, weil sie an dieser Stelle hydrolysiert und in ihre Nukleotide zerlegt werden kann.

Außerdem enthält RNA *Uracil* anstelle von Thymin als Base. Uracil kann durch *Desaminierung* (Entfernung des Stickstoffs) und Hydrolyse relativ einfach aus Cytosin entstehen. Würde dies an der DNA in der Zelle geschehen, entspräche es einem Basenaustausch und somit einer Punktmutation an dieser Stelle, da Uracil komplementär zu Adenin und nicht wie Cytosin komplementär zu Guanin ist.

Thymin enthält im Gegensatz zu Uracil eine zusätzliche Methylgruppe. Eine chemische Umwandlung von Cytosin zu Thymin kann daher nicht so einfach stattfinden. Der Einbau dieser zusätzlichen Methylgruppe ist für die Zelle zwar energetisch aufwendiger, schützt sie aber vor Mutationen, denn Uracil wird in der DNA durch spezifische Reparaturenzyme entfernt und entsprechend ausgetauscht.

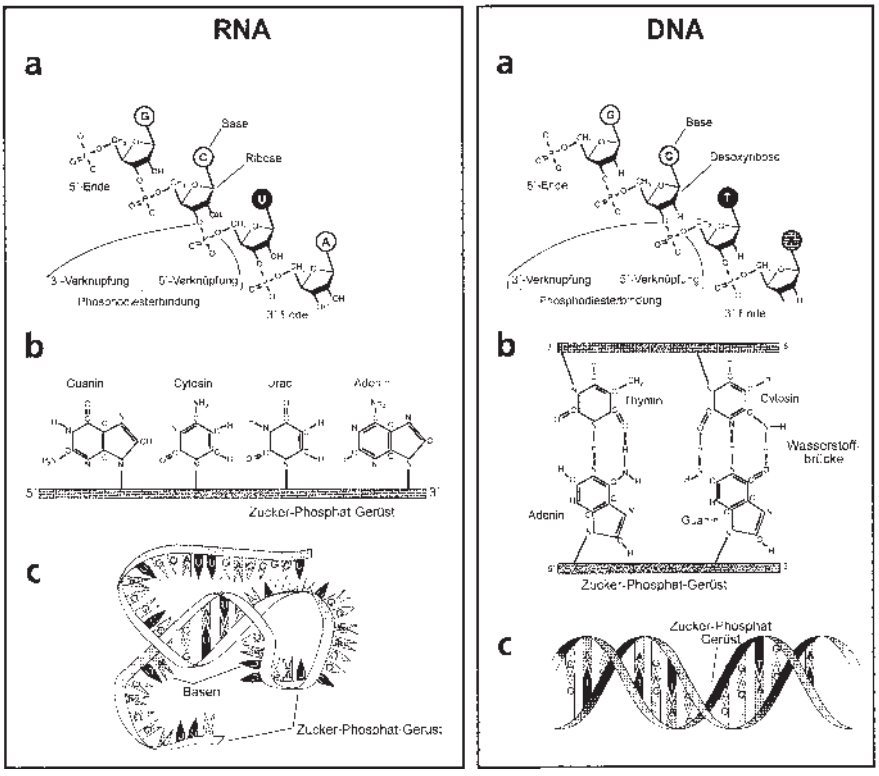


Abb. 1.4 Vergleich von RNA und DNA.

(a) *Das Zucker-Phosphat-Rückgrat*:

Die DNA enthält im Zucker-Phosphat-Gerüst eine Desoxyribose. Die RNA enthält eine Ribose.

*D*esoxyribonukleinsäure und *R*ibonukleinsäure (Säure = engl. acid)

In beiden Molekülen sind die Zucker über eine Phosphodiesterbindung miteinander verknüpft.

(b) *Die Basen*:

Beide Nukleinsäuren beinhalten 4 Basen, von denen sich 2 jeweils komplementär miteinander paaren. In der DNA paart sich Adenin mit Thymin und Guanin mit Cytosin. Die RNA enthält statt Thymin die Base Uracil. Die Basenpaarungen erfolgen in beiden Molekülen über spezifische, nicht kovalente Wasserstoffbrückenbindungen.

(c) *Die Struktur*:

Die DNA liegt in der Zelle als doppelsträngiges Molekül vor.

Die RNA ist ein einzelsträngiges Molekül. Kurze Abschnitte können sich aber zufällig komplementär miteinander paaren und Sekundärstrukturen ausbilden (Abbildung aus: *MTA-Praxis*, Sonderheft 1, 2002, mit freundlicher Genehmigung der Hoppenstedt Bonnier Zeitschriften GmbH, Darmstadt).

Die Nukleotide der RNA sind, wie in der DNA, über eine Phosphodiesterbindung miteinander verknüpft.

RNA ist im Gegensatz zur DNA ein *einzelsträngiges* Molekül. Bedingt durch interne komplementäre Abschnitte innerhalb eines RNA-Einzelstranges, die sich über Wasserstoffbrücken miteinander verbinden, kann die RNA verschiedene Formen und Strukturen annehmen. Diese Loop- (Schleifen) oder Hairpin (Haarnadel)-Strukturen weisen einzel- sowie doppelsträngige Abschnitte auf. Die wohl

bekannteste zweidimensionale Struktur ist die Kleeblatt-Struktur der Transfer-RNA (tRNA).

Aufgrund ihrer Funktionalität werden RNA-Moleküle in zwei Klassen eingeteilt:

RNAs der Klasse I werden nicht translatiert, d. h. sie enthalten keine direkte Information für die Proteinbiosynthese, können aber beteiligt sein (s. Abschnitt 1.7).

Zu diesen nicht codierenden oder untranslatierten RNAs der Klasse I (*utRNAs*) werden u. a. folgende RNA-Moleküle gezählt:

- **tRNA (Transfer-RNA):** bindet und transportiert eine für sie spezifische Aminosäure zu den Ribosomen, dem Ort der Proteinbiosynthese.
- **rRNA (ribosomale RNA):** struktureller Bestandteil der Ribosomen. Die rRNA dient als Katalysator bei der Knüpfung der Peptidbindung.
- **snRNA (small nuclear RNA):** snRNA-Moleküle haben eine Länge von ca. 100 bis 300 Basen. Die snRNA ist im Zellkern im *Spleißosom* mit spezifischen Proteinen komplexiert und katalytisch aktiv. Konservierte Abschnitte an der Übergangsstelle Intron/Exon der hnRNA (s. u.) binden zur Entfernung der Introns an die snRNA. Das Enzym Telomerase, verantwortlich für die Stabilität der Chromosomenenden (Telomere), enthält ebenfalls snRNA.

Neben der snRNA existieren noch weitere kleine RNA-Moleküle z. B. im Nukleolus, die

- **snoRNA (small nucleolar RNA):** Diese Moleküle sind hauptsächlich an der Modifikation der rRNA beteiligt. Auch sie liegen komplexiert mit spezifischen Proteinen vor, den snoRNPs (*small nucleolar ribonucleoprotein particles*).
- **siRNA (small interfering RNA):** Die siRNA ist eine ca. 19 bis 23 bp doppelsträngige kleine RNA mit jeweils zwei endständig überstehenden Nukleotiden. Sie entsteht durch eine Spaltung eines großen endogenen oder exogenen doppelsträngigen RNA-Moleküls. siRNA spielt eine wesentliche Rolle bei der RNA-Interferenz (RNAi), ein Mechanismus in Eukaryonten zur Ausschaltung von Genen. Experimentell lassen sich mithilfe der siRNA gezielt Gene stilllegen.
- **miRNA (micro RNA):** miRNA ist im Gegensatz zur siRNA auf der genomischen DNA codiert, also immer endogenen Ursprungs. miRNAs regulieren die Genexpression hochspezifisch durch Bindung an die mRNA (posttranskriptional).

RNAs der Klasse II werden transkribiert und translatiert, d. h sie enthalten eine spezifische genetische Information, die in ein Peptid oder Protein übersetzt wird. Hauptsächlicher Vertreter dieser Klasse ist die **mRNA** (Boten-RNA, engl.: *m*essenger). Diese entsteht aus der **hnRNA** (heterogener Kern, engl.: heterogeneous nuclear)-RNA. Die hnRNA wird daher synonym auch als prä-mRNA bezeichnet (engl.: precursor-mRNA).

Nach der Transkription werden die auf der hnRNA lokalisierten nicht codierenden Sequenzen (Introns) durch das Spleißen entfernt (Abb. 1.10). Nach weiterer Prozessierung entsteht die reife messenger-RNA (mRNA).

1.3 DNA-Replikation

Während der Zellteilung wird die genetische Information der Elternzelle auf die zwei neu entstehenden Tochterzellen verteilt. Damit jede Tochterzelle die identische genetische Ausstattung erhält, muss die DNA vorher verdoppelt werden. Dieser Vorgang wird als *DNA-Replikation* bezeichnet (Abb. 1.5).

Die DNA-Replikation ist ein enzymatisch gesteuerter Prozess. Die DNA-Doppelhelix wird zunächst durch das Enzym *Helikase* aufgetrennt. Die beiden nun vorliegenden DNA-Einzelstränge der *Replikationsgabel* werden durch Einzelstrang bindende Proteine (*s*ingle *s*tranded *b*inding proteins = SSB) stabilisiert.

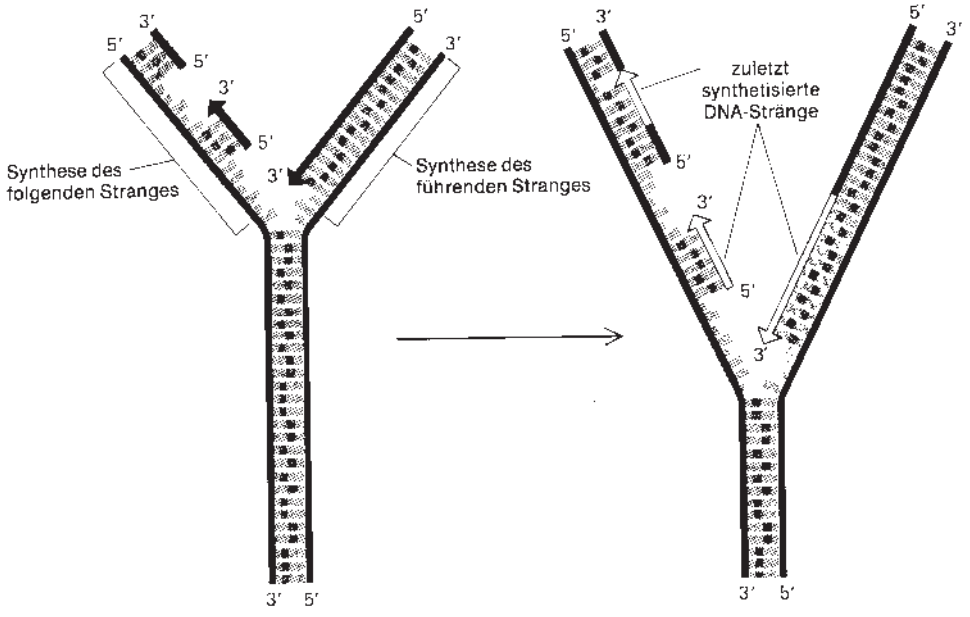


Abb. 1.5 DNA-Replikation.
Die beiden DNA-Stränge der DNA-Helix haben gegensätzliche Polaritäten. Die DNA-Polymerase kann aber einen DNA-Strang nur in 5′→3′-Richtung synthetisieren. Daher verläuft die DNA-Replikation des Leit (Leading)-Stranges kontinuierlich und die des Folge (Lagging)-Stranges diskontinuierlich. Die Replikationsgabel ist asymmetrisch. Während der Replikation bleibt ein parentaler DNA-Strang erhalten (semikonservative Replikation) (Abbildung verändert aus: Alberts, B. et al., 2011, *Molekularbiologie der Zelle*, 5. Auflage, Wiley-VCH Weinheim).

Jeder Einzelstrang dient als Vorlage (*Matrize, Template*) für den neu zu synthetisierenden Strang.

Das Enzym *DNA-Polymerase* benötigt für diese komplementäre Ergänzung der Matrize ein Startermolekül (*Primer*), das eine freie 3'-OH-Gruppe für die chemische Anknüpfung des komplementären Nukleotids zur Verfügung stellt. Der Primer ist ein kurzes RNA-Molekül, das von einer *Primase* synthetisiert wird.

Die DNA-Polymerase erkennt das Startermolekül und beginnt mit der komplementären Strangsynthese in 5'→3'-Richtung, indem sie die vorliegenden Desoxyribonukleotidtriphosphate unter Abspaltung zweier Phosphatgruppen (*PPi = Pyrophosphat*) über eine *Phosphodiesterbindung* miteinander verknüpft. Durch die Abspaltung des Pyrophosphats wird die Energie für diese Verbindung zur Verfügung gestellt.

Eine kontinuierliche Synthese in 5'→3'-Richtung der DNA ist aber nur an einem Strang möglich (*Leitstrangsynthese, Leading-Strand*). Die Ergänzung des anderen Stranges (*Folgestrangsynthese, Lagging-Strand*) erfolgt diskontinuierlich. Die Primase synthetisiert hier im Abstand von 100–200 Nukleotiden die RNA-Primer, die von der DNA-Polymerase verlängert werden. Die so entstandenen DNA-Fragmente nennt man *Okazaki-Fragmente*. Die Synthese eines Okazaki-Fragmentes endet, wenn die DNA-Polymerase auf einen vor ihr liegenden RNA-Primer stößt.

Die RNA-Primer werden anschließend von einem DNA-Reparatursystem entfernt und durch DNA ersetzt. Eine *DNA-Ligase* knüpft abschließend die Phosphodiesterbindung zwischen den benachbarten Okazaki-Fragmenten.

Da bei der enzymatischen Verdopplung immer ein „alter" DNA-Strang erhalten bleibt, bezeichnet man diesen Prozess als *semikonservative DNA-Replikation*.

Die DNA-Replikation läuft in Pro-und Eukaryonten ähnlich ab. Sie startet immer an einer definierten Stelle im Genom, dem *Replikationsursprung* (ORI = origin of replication). Prokaryonten haben pro Genom meist nur einen ORI, während Eukaryonten aufgrund ihres größeren Genoms die Replikation an mehreren ORIs gleichzeitig initiieren müssen.

Die an der DNA-Verdopplung beteiligten Enzyme liegen in der Zelle in einem Multi-Enzym-Komplex vor, um diesen Prozess effizient ablaufen zu lassen. Eukaryonten replizieren ihre DNA mit einer Geschwindigkeit von etwa 50 Nukleotiden/Sekunde, Prokaryonten mit 500–1000 Nukleotiden/Sekunde.

1.4 Das Gen

Proteine sind Polypeptide, die jeweils eine bestimmte Anzahl von spezifischen Aminosäuren enthalten. Sie können entweder aus einem Polypeptid und/oder verschiedenen Polypeptiden bestehen. Jedes Polypeptid wird von einem definierten Abschnitt auf der DNA codiert. Diesen Abschnitt bezeichnet man als *Gen*, die Gesamtheit der Gene eines Organismus als *Genom*.

Die Genome bei Pro- und Eukaryonten sind unterschiedlich organisiert (s. Abschnitt 1.5, 1.6). Aufgrund dieser Unterschiede ergeben sich auch andere Abläufe bei der *Proteinbiosynthese*.

1.5 Genomorganisation bei Prokaryonten

Zu den *Prokaryonten* zählt man die *Eubacteria* und die *Archaea*. Diese einzelligen Mikroorganismen besitzen keinen abgegrenzten, membranumschlossenen Zellkern.

Das Genom besteht aus einem großen, *ringförmigen doppelsträngigen* DNA-Molekül, das *superhelikal* (überspiralisiert) aufgewunden ist. Es wird auch als *Kernäquivalent* oder *Nukleoid* bezeichnet, da es sich in elektronenmikroskopischen Abbildungen deutlich von dem es umgebenden Zytoplasma abhebt.

Bakterien besitzen außerdem kleinere ringförmige doppelsträngige DNA-Moleküle, die *Plasmide*. Plasmide enthalten immer einen Replikationsursprung und verschiedene Gene, die z. B. für Virulenz-, Konjugationsfaktoren oder Antibiotikaresistenzen codieren. Sie replizieren unabhängig vom Bakteriengenom und verschaffen dem Bakterium manchmal einen Selektionsvorteil. Plasmide, die sich in das Wirtsgenom integrieren können, bezeichnet man als *Episomen*.

Plasmide sind in der Molekularbiologie und Gentechnik häufig verwendete Werkzeuge zur Vermehrung oder Expression von Genen. Diese sogenannten *Vektoren* sind entweder kommerziell erhältlich (z. B. pUC 18/19) oder werden vom Anwender im Labor modifiziert und so auf die spezifischen experimentellen Bedürfnisse angepasst. Die erforderlichen Gene oder Genabschnitte werden in das Plasmid kloniert. Dazu wird mithilfe von *Restriktionsendonukleasen* in der enthaltenen *Multiple Cloning Site* (MCS) eine definierte Schnittstelle generiert. In diese Schnittstelle passt das zu klonierende Gen. Ligasen verbinden anschließend die jeweiligen Enden kovalent miteinander. Durch eine Transformation werden die Vektoren in dafür vorbereitete Bakterienzellen verbracht, wo sie dann replizieren.

Die Genomgröße bei Prokaryonten variiert zwischen 1×10^6 und 5×10^6 Basenpaaren (bp). 85 % des prokaryotischen Genoms, das entspricht circa 4000–5000 Genen, codiert für Proteine oder RNA, der Rest sind regulatorische DNA-Sequenzen und nur sehr wenig nicht codierende Sequenzen.

1.6 Genomorganisation bei Eukaryonten

Tiere, Pflanzen, Pilze und Protisten (Algen, Protozoen, Schleimpilze) werden zu den *Eukaryonten* gezählt.

Eukaryonten besitzen im Gegensatz zu den Prokaryonten einen membranumschlossenen *Zellkern* (*Nukleus*) und weitere durch Membranen voneinan-

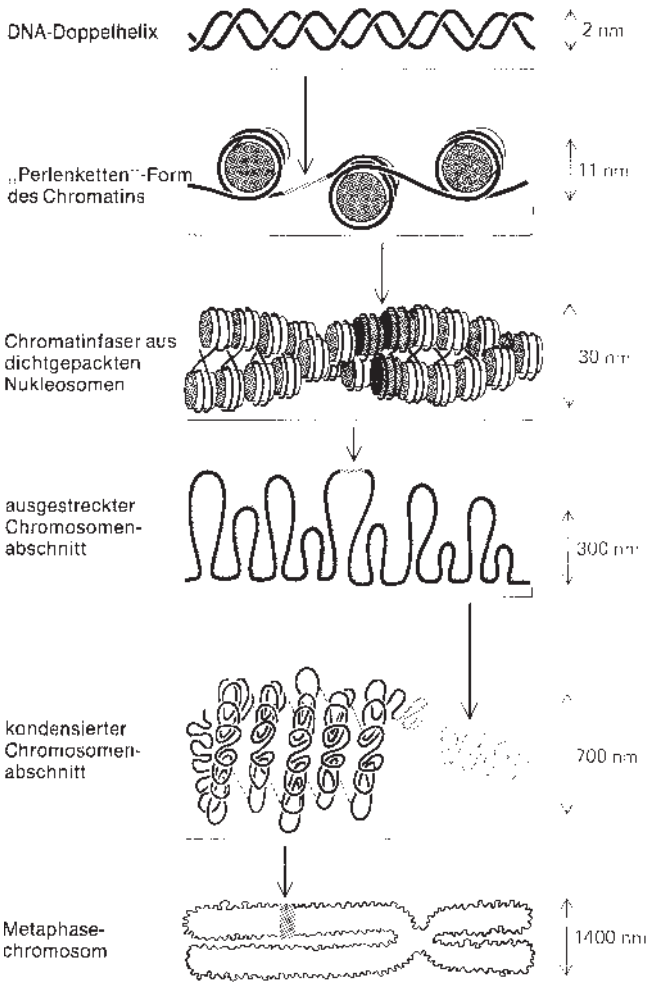


Abb. 1.6 DNA-Verpackung.

Übersicht über die Verpackung der DNA (von oben nach unten):

Die DNA ist in regelmäßigen Abständen um basische Proteinkomplexe, die Histone, gewunden. Diese sogenannten Nukleosomen werden durch Nicht-Histon-Proteine weiter zur 30 nm-Faser gepackt. Durch Faltung der 30 nm-Faser entstehen Schleifendomänen, die weiter kondensieren und das Chromosom in der Metaphase sichtbar werden lassen (Abbildung verändert aus: Alberts, B. et al., 2012, *Lehrbuch der molekularen Zellbiologie*, 4. Auflage, Wiley-VCH Weinheim).

der abgegrenzte *Zellkompartimente* oder Organellen wie Mitochondrien, Golgi-Vesikel, Endoplasmatisches Retikulum etc.

Im Nukleus ist die DNA einer Zelle lokalisiert. Fast alle zellulären Prozesse werden von hier aus koordiniert. Neben dem Zellkern enthalten auch Mitochondrien und Chloroplasten jeweils eigene DNA, die für einen Teil ihrer Stoffwechselenzyme codiert. Mitochondriale DNA (*mtDNA*) wird maternal vererbt. Diese

Tatsache macht man sich bei der Untersuchung evolutionärer Verwandtschaftsverhältnisse zunutze. Stammbäume werden anhand von Sequenzunterschieden der mtDNA verwandter Populationen aufgestellt.

Das menschliche Genom, das hier als Beispiel eines typischen eukaryotischen Genoms dienen soll, besteht aus ca. $3{,}2 \times 10^9$ Basenpaaren (bp), die für ca. 30.000 Gene codieren. Die DNA einer menschlichen Zelle wäre als Faden abgewickelt etwa 1,80 m lang. Sie liegt im Zellkern komplexiert mit basischen Proteinen, den *Histonen*, vor, um die sie in definierten Abständen gewickelt ist. Durch andere, sogenannte *Nicht-Histon-Proteine*, wird die DNA weiter verpackt. Während der Zellteilung wird die DNA so kompakt verdichtet, das sie im Lichtmikroskop in Form der *Chromosomen* sichtbar wird (chromos = Farbe, soma = Körper) (Abb. 1.6).

Der Mensch hat 46 Chromosomen, davon zwei geschlechtsbestimmende *Gonosomen* (XX oder XY) und 44 *Autosomen* (diploider Chromosomensatz).

70 % der DNA ($2{,}1 \times 10^9$ bp) entfallen auf Bereiche, die keine Gene enthalten. Nur etwa 30 % ($0{,}9 \times 10^9$ bp) der DNA lässt sich bestimmten Genen zuordnen, wobei hiervon nur 3 % (0.09×10^9 bp) direkt codierenden Sequenzen entsprechen. Dies muss bedeuten, dass die Gene Sequenzen enthalten, *auf denen keine Information codiert ist*. Tatsächlich sind solche Abschnitte identifiziert worden. Sie werden als *Introns* (engl.: ***int**ervening **r**egi**ons***) bezeichnet. Die codierenden Abschnitte eines Gens nennt man *Exons*.

Introns werden nach der Transkription aus der *prä-mRNA* herausgeschnitten (s. Abschnitt 1.7.1.3).

1.7 Die Proteinbiosynthese

Wird ein Protein in der Zelle benötigt, muss die in dem entsprechenden Gen enthaltene Information abgelesen und weiter prozessiert werden, damit das Protein synthetisiert werden kann. Die Nukleotidsequenz eines Gens wird zuerst abgeschrieben (*transkribiert*) und mittels eines Boten (*messenger*) zu dem Ort der Proteinsynthese, den *Ribosomen*, transportiert. Anschließend wird sie in die entsprechende Aminosäuresequenz übersetzt (*translatiert*) und das Genprodukt synthetisiert.

1.7.1 Die Transkription

Die Transkription ist ein Prozess, der hauptsächlich von den *RNA-Polymerasen* katalysiert wird (Abb. 1.7). Prokaryonten besitzen nur eine, Eukaryonten hingegen drei verschiedene RNA-Polymerasen.

Die RNA-Polymerase I transkribiert die meisten Gene, die für die ribosomale RNA (rRNA) codieren (s. Abschnitt 1.7.2.3), die RNA-Polymerase III transkribiert hauptsächlich die Gene für die tRNAs (s. Abschnitt 1.7.2.2). Für die Tran-

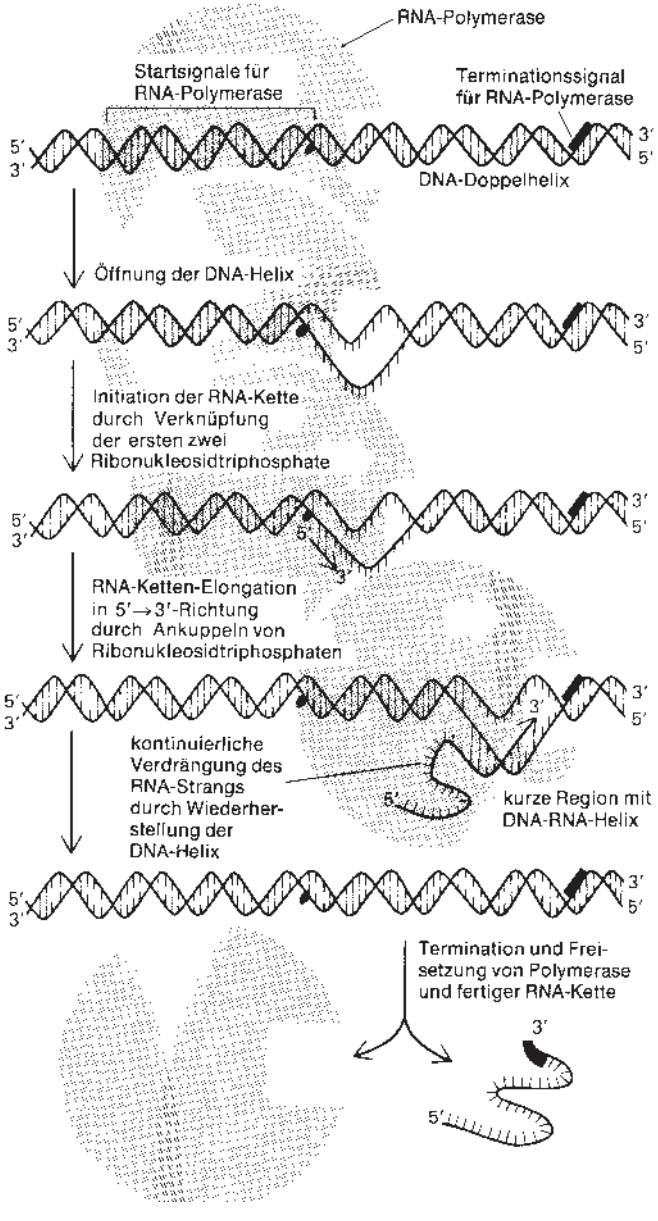


Abb. 1.7 Transkription.

Die RNA-Polymerase bindet an den Promotor eines Gens und beginnt mit der Transkription. Die DNA-Helix wird aufgewunden und die mRNA synthetisiert. Die Ribonukleotidtriphosphate werden komplementär zum abgelesenen DNA-Strang eingebaut (codogener Strang). Die DNA-Helix schließt sich hinter der RNA-Polymerase und das Transkript wird verdrängt. Erreicht das Enzym den Terminator, endet die Synthese. Polymerase und mRNA werden freigesetzt (Abbildung verändert aus: Alberts, B. et al., 2011, *Molekularbiologie der Zelle*, 5. Auflage, Wiley-VCH Weinheim).

skription der proteincodierenden Gene in Eukaryonten ist die *RNA-Polymerase II* zuständig.

Generell bestehen diese Enzyme aus mehreren Untereinheiten, die als *Transkriptionsfaktoren* bezeichnet werden.

Bakterielle RNA-Polymerasen benötigen den *Sigma (σ)-Faktor* zur Transkription.

Die RNA-Polymerase II der Eukaryonten benötigt dagegen mehrere, als *allgemeine Transkriptionsfaktoren (TFIIA, TFIIB etc.*) bezeichnete Untereinheiten.

Diese erkennen und binden an spezifische DNA-Sequenzen im Genom, die *Promotoren*. Ein Promotor liegt oberhalb (*upstream*) am 5'-Ende eines Gens. Die DNA-Sequenz verschiedener bakterieller Promotoren ist sehr heterogen. Diese Heterogenität ist dafür verantwortlich, wie häufig ein Gen transkribiert wird, denn RNA-Polymerasen haben zu bestimmten Sequenzen eine größere Affinität. Gene mit „starken" Promotoren werden demnach häufiger abgelesen. Sie codieren für Genprodukte, die in der Zelle häufiger oder in größerer Anzahl benötigt werden.

Ein Promotor enthält zwei kurze DNA-Abschnitte, die im Laufe der Evolution konserviert wurden (*Konsensussequenzen*). Eine Konsensussequenz (TTGACA) liegt ca. 35 Basenpaare, die andere (TATATT; sog. (TATA-Box) ca. 10–25 Basenpaare upstream vom Startpunkt des jeweiligen Gens entfernt.

Promotoren bei Eukaryonten enthalten noch weitere Konsensussequenzen (BRE-, INR- oder DPE-Element), die für die jeweilige „Stärke" eines Promotors verantwortlich sind. Weitere regulatorische DNA-Abschnitte können sehr weit vom Startpunkt entfernt liegen.

Hat die RNA-Polymerase mithilfe der Transkriptionsfaktoren an den Promotor gebunden, wird die DNA-Doppelhelix geöffnet und entwunden. Dieser Prozess benötigt keine Energie. Die Lage und Richtung des Promotors entscheidet darüber, welcher der nun vorliegenden DNA-Einzelstränge als Vorlage (Matrize) für die RNA-Polymerase dient. Das Enzym synthetisiert aus den hinzukommenden Nukleotidtriphosphaten ATP, GTP, UTP und CTP ein RNA-Polymer, die *messenger-RNA (mRNA*). Die Nukleotide werden komplementär zum Matrizenstrang in 5'→3'-Richtung eingebaut und enzymatisch durch die Phosphodiesterbindung kovalent miteinander verbunden. Die Energie für diese Reaktion wird durch die Abspaltung des Pyrophosphats (PPi) von den Nukleotidtriphosphaten bereitgestellt. Der DNA-Doppelstrang schließt sich hinter der RNA-Polymerase wieder.

Das Enzym synthetisiert so lange, bis es auf eine weitere spezielle DNA-Sequenz, den *Terminator* trifft. Diese DNA-Sequenz veranlasst die Polymerase sich von der DNA zu lösen. Die synthetisierte mRNA und der Matrizenstrang werden freigesetzt.

Die mRNA der Prokaryonten wird direkt anschließend translatiert.

Bei Eukaryonten wird sie weiter prozessiert, bevor sie vom Zellkern in das Zytoplasma zu den Ribosomen transportiert wird. Man bezeichnet diese mRNA daher auch als *prä-mRNA* (s. Abschnitt 1.1, Abb. 1.8a, b).

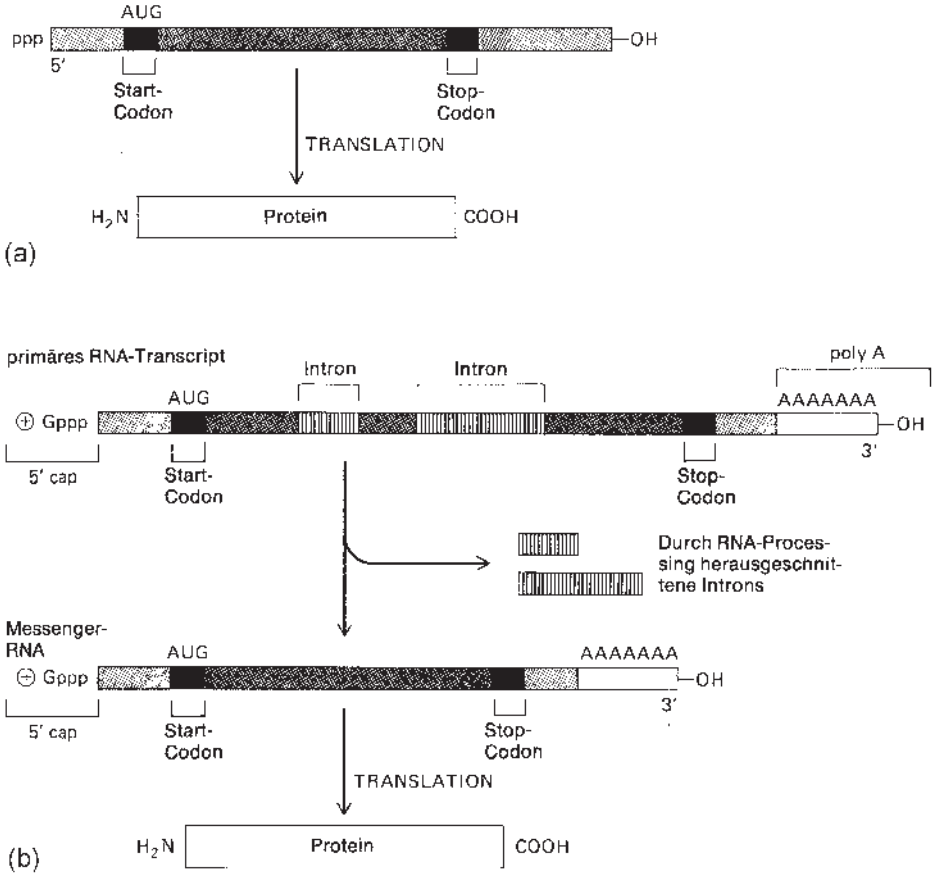


Abb. 1.8 (a) Die mRNA der Prokaryonten. Die mRNA der Prokaryonten enthält keine Introns und wird nicht weiter prozessiert, sondern direkt translatiert.

(b) Die mRNA der Eukaryonten. Die prä-mRNA der Eukaryonten enthält Introns, die aus dem Spleißosom herausgeschnitten werden. Erst dann wird das Transkript translatiert. Außerdem erhält sie eine 5'-Kappe (*engl. „cap"*) und einen poly A-Schwanz (Abbildung verändert aus: Alberts, B. et al., 2012, *Lehrbuch der molekularen Zellbiologie*, 4. Auflage, Wiley-VCH Weinheim).

1.7.1.1 Die 5'-Cap-Struktur (Capping)

Der erste Schritt hin zu einer „reifen" eukaryotischen mRNA ist das sogenannte *„Capping"*. Dabei wird ein modifiziertes Guaninnukleotid an das 5'-Ende der prä-mRNA gebunden (Abb. 1.9). Dieser enzymatisch gesteuerte Prozess beginnt, wenn die RNA-Polymerase ca. 25 Nukleotide der prä-mRNA synthetisiert hat, und lässt sich in drei Schritte unterteilen:

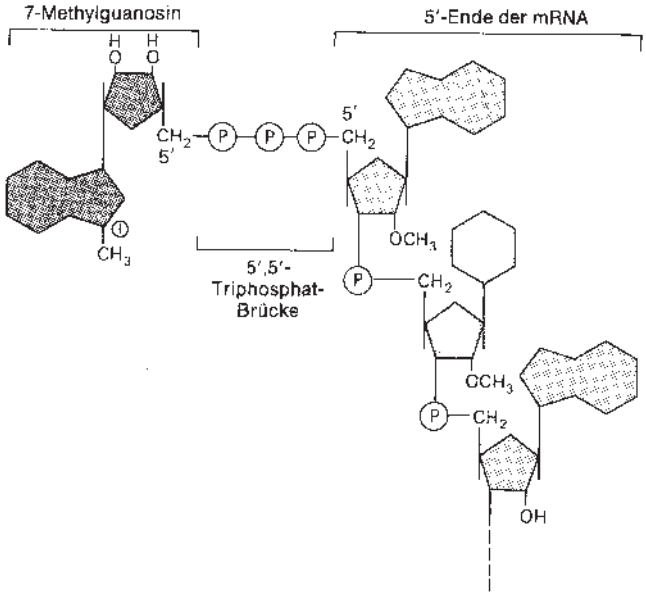


Abb. 1.9 Chemische Struktur der 5'-Kappe einer eukaryotischen mRNA (Abbildung verändert aus: Alberts, B. et al., 2012, *Lehrbuch der molekularen Zellbiologie*, 4. Auflage, Wiley-VCH Weinheim).

1) Das 5'-Phosphat der prä-mRNA wird von einer *Phosphatase* entfernt.
2) Das Enzym *Guanyltransferase* überträgt ein Guanosinnukleotid.
3) Eine *Methyltransferase* methyliert das Guanosin (*7-Methyl-Guanosin*).

Die 5'-Kappe markiert die mRNA und kann so in der Zelle von anderen RNAs unterschieden werden, denn die von den anderen RNA-Polymerasen synthetisierten RNAs werden nicht auf diese Art modifiziert.

Dieser chemisch modifizierte Teil der mRNA wird von dem *„Cap-Binding-Komplex"* gebunden. Dieser Komplex spielt eine wichtige Rolle beim Export der RNA aus dem Zellkern in das Zytoplasma, zum Ort der *Translation*. Die 5'-Kappe dient dann als Bindungsstelle der Ribosomen (s. Abschnitt 1.7.2.4).

1.7.1.2 Der Poly(A)-Schwanz

Das 3'-Ende einer eukaryotischen prä-mRNA wird ebenfalls modifiziert. Am Ende eines transkribierten DNA-Abschnitts befinden sich spezifische Sequenzen. Diese Signalsequenzen werden von mit der RNA-Polymerase II assoziierten Proteinen auf der RNA erkannt. Die mRNA wird am 3'-Ende spezifisch geschnitten und das Enzym *Poly-A-Polymerase* synthetisiert ca. 200–250 Adeninnukleotide an, wobei die Länge von den *Poly-A-Bindungsproteinen* festgelegt wird.

Poly-A-Polymerasen benötigen keine Matrize. Der Poly-A-Schwanz der mRNA ist also nicht im Genom codiert.

Abb. 1.10 Der Mechanismus des RNA-Spleißens.

Das Spleißen von RNA wird durch die snRNPs (small nuclear ribonuclearproteins) katalysiert, die mit anderen, nicht dargestellten Proteinen, das Spleißosom bilden. Das Spleißosom erkennt die Spleißsignale auf der prä-mRNA und führt die beiden Enden des Introns zusammen. Durch mehrfache Umlagerungen der in den RNPs enthaltenen RNA wird das aktive Zentrum innerhalb des Spleißosoms geschaffen und die jeweils passenden Abschnitte der prä-mRNA zusammengeführt. An der Reaktion sind insgesamt über 50 Proteine beteiligt, von denen einige ATP hydrolisieren, um die Umlagerungen zwischen den RNAs anzutreiben. Die ausgeschnittene Intronsequenz wird im Zellkern abgebaut. Die snRNPs werden wieder verwendet (Abbildung verändert aus: Alberts, B. et al., 2012, *Lehrbuch der molekularen Zellbiologie*, 4. Auflage, Wiley-VCH Weinheim).

Der Poly(A)-Schwanz wird von Poly(A)-Bindeproteinen (PABP) besetzt und ist notwendig für die Initiation der Translation. Er schützt die RNA außerdem vor RNA- abbauenden Enzymen (RNasen).

1.7.1.3 Spleißen von RNA

Die weitere Prozessierung der während der Transkription synthetisierten prä-mRNA wird als *Spleißen* bezeichnet. Die in der prä-mRNA enthaltenen *Introns* werden herausgeschnitten und die *Exons* direkt miteinander verbunden (Abb. 1.10).

An diesem Prozess sind hauptsächlich kurze, mit Proteinen assoziierte RNA-Moleküle beteiligt, die sogenannten *„small nuclear ribonuclearproteins" (snRNP*). Sie bilden mit weiteren Proteinen das *Spleißosom*.

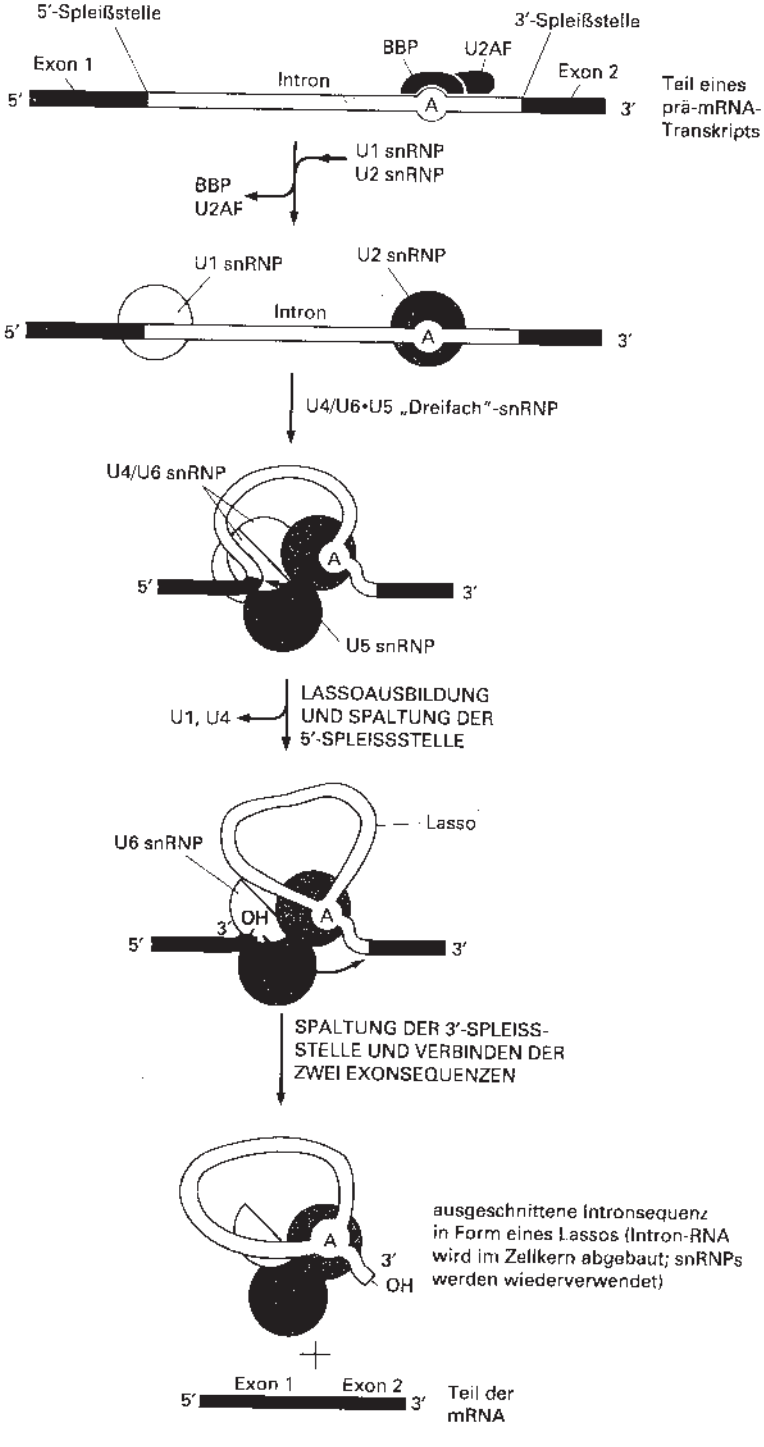
Das Spleißosom erkennt die Sequenzen auf der prä-mRNA, die als *Spleißsignale* dienen. Die beiden Enden des Introns werden durch das Spleißosom in räumliche Nähe zueinander gebracht. Nach zwei *Phosphoryltransfer-Reaktionen* und mannigfaltigen Umlagerungen der RNPs werden die Exons exakt miteinander verknüpft. Die ausgeschnittene Intronsequenz liegt am Ende dieses Prozesses in Form eines *„Lassos"* oder *„Lariats"* vor. Sie wird im Zellkern abgebaut.

Die so prozessierte RNA wird jetzt als *mRNA* bezeichnet und kann im Zytoplasma der eukaryotischen Zelle translatiert werden.

Ein eukaryotisches Gen kann *alternativ gespleißt* werden. Die Anzahl der herausgespleißten Introns kann dabei variieren. Aus einem *Primärtranskript* können so unterschiedliche mRNAs für verschiedene Proteine entstehen.

1.7.2 Die Translation

Die auf der DNA codierte genetische Information ist von der RNA-Polymerase abgeschrieben (*transkribiert*) worden und muss jetzt übersetzt (*translatiert*) werden, um ein funktionelles Protein zu erhalten.


5′-Spleißstelle
3′-Spleißstelle
BBP
U2AF
Exon 1
Intron
Exon 2
5′
A
3′
Teil eines prä-mRNA-Transkripts
U1 snRNP
U2 snRNP
BBP
U2AF
U1 snRNP
U2 snRNP
Intron
U4/U6•U5 „Dreifach"-snRNP
U4/U6 snRNP
U5 snRNP
U1, U4
LASSOAUSBILDUNG UND SPALTUNG DER 5′-SPLEISSSTELLE
Lasso
U6 snRNP
OH
SPALTUNG DER 3′-SPLEISS-STELLE UND VERBINDEN DER ZWEI EXONSEQUENZEN
ausgeschnittene Intronsequenz in Form eines Lassos (Intron-RNA wird im Zellkern abgebaut; snRNPs werden wiederverwendet)
Exon 1
Exon 2
Teil der mRNA

Der genetische Code

Position 1 (5'-Ende)	Position 2 U	Position 2 C	Position 2 A	Position 2 G	Position 3 (3'-Ende)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Aminosäuren und ihre Symbole			Codons					
A	Ala	Alanin	GCA	GCC	GCG	GCU		
C	Cys	Cystein	UGC	UGU				
D	Asp	Asparaginsäure	GAC	GAU				
E	Glu	Glutaminsäure	GAA	GAG				
F	Phe	Phenylalanin	UUC	UUU				
G	Gly	Glycin	GGA	GGC	GGG	GGU		
H	His	Histidin	CAC	CAU				
I	Ile	Isoleucin	AUA	AUC	AUU			
K	Lys	Lysin	AAA	AAG				
L	Leu	Leucin	UUA	UUG	CUA	CUC	CUG	CUU
M	Met	Methionin	AUG					
N	Asn	Asparagin	AAC	AAU				
P	Pro	Prolin	CCA	CCC	CCG	CCU		
Q	Gln	Glutamin	CAA	CAG				
R	Arg	Arginin	AGA	AGG	CGA	CGC	CGG	CGU
S	Ser	Serin	AGC	AGU	UCA	UCC	UCG	UCU
T	Thr	Threonin	ACA	ACC	ACG	ACU		
V	Val	Valin	GUA	GUC	GUG	GUU		
W	Trp	Tryptophan	UGG					
Y	Tyr	Tyrosin	UAC	UAU				

Abb. 1.11 Der genetische Code und die Aminosäuren.

Ein Codon besteht aus drei Nukleotiden (Triplett). Jedes Codon codiert für eine spezifische Aminosäure. Einige Aminosäuren werden allerdings von mehreren Codons determiniert. Dabei ist immer die dritte Base eines Tripletts variabel. Dargestellt ist hier die Sequenz der mRNA.

Die Aminosäuren werden konventionell durch einen oder drei Buchstaben abgekürzt wiedergegeben (Abbildung verändert aus: Alberts, B. et al., 2011, *Molekularbiologie der Zelle*, 5. Auflage, Wiley-VCH Weinheim).

1.7.2.1 Der genetische Code

Die DNA ist der Speicher der genetischen Information. Diese Information ist in der Abfolge der Basen, der Basen- oder Nukleotidsequenz, verschlüsselt oder *codiert*. Wie wird dieser Code in Proteine übersetzt?

Proteine bestehen aus einer bestimmten Abfolge von Aminosäuren. Der Code für die Abfolge der Aminosäuren ist durch die Nukleotidsequenz der DNA festgelegt. Eine spezifische Sequenz von drei Nukleotiden (*Triplet-Codon*) entspricht jeweils einer Aminosäure. Jedes Codon ist durch die drei Basen genau definiert und kann auf eine bestimmte Aminosäure zurückgeführt werden. Allerdings kann eine Aminosäure durch mehr als ein Codon codiert sein. Die größte Variabilität weisen die Triplets dabei an ihrer dritten Position auf. Diese Variabilität bezeichnet man als *Redundanz*. Etwas verschiedene Nukleotidsequenzen können demnach für dieselben Aminosäuresequenzen codieren (Abb. 1.11).

Spezifische Triplets werden als *Start- oder Stop-Codons* bezeichnet. Dies sind spezifische Codons, an denen die Proteinbiosynthese startet bzw. stoppt. Sie definieren den *offenen Leserahmen (ORF)* einer DNA-Sequenz (Abb. 1.12).

Der genetische Code ist universal, d. h. alle Organismen benutzen, bis auf einige Ausnahmen, die gleichen Codons für dieselben Aminosäuren.

1.7.2.2 Transfer-RNA

Transfer-RNAs oder tRNAs sind RNA-Moleküle, die die Aminosäuren zum Ort der Proteinsynthese transportieren. Sie haben eine durchschnittliche Länge von ca. 80 Nukleotiden und zeichnen sich durch ihre besondere Struktur aus. Diese kommt durch interne komplementäre Basenpaarungen zustande. Dadurch entsteht ein Molekül mit doppelhelikalen und einzelsträngigen Bereichen. Zweidimensional erscheint die tRNA als *kleeblattförmige Struktur*, die durch weitere Auffaltungen eine *dreidimensionale L-förmige Struktur* annimmt (Abb. 1.13).

Transfer-RNAs enthalten Introns, die wie bei einer prä-mRNA herausgespleißt werden müssen. Der Mechanismus des tRNA-Spleißens kommt bei Pro- und

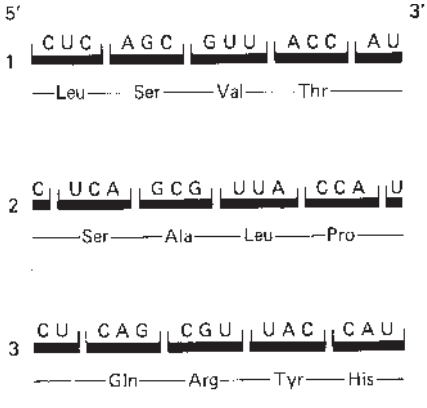


Abb. 1.12 Leseraster (*O*pen *R*eading *F*rame).

Das Leseraster einer DNA ist durch die Tripletts, die jeweils für eine spezifische Aminosäure codieren, festgelegt. Die Verschiebung des Leserasters um nur ein Nukleotid ergibt in diesem Beispiel eine andere Abfolge von Aminosäuren (Abbildung verändert aus: Alberts, B. et al., 2012, *Lehrbuch der molekularen Zellbiologie*, 4. Auflage, Wiley-VCH Weinheim).

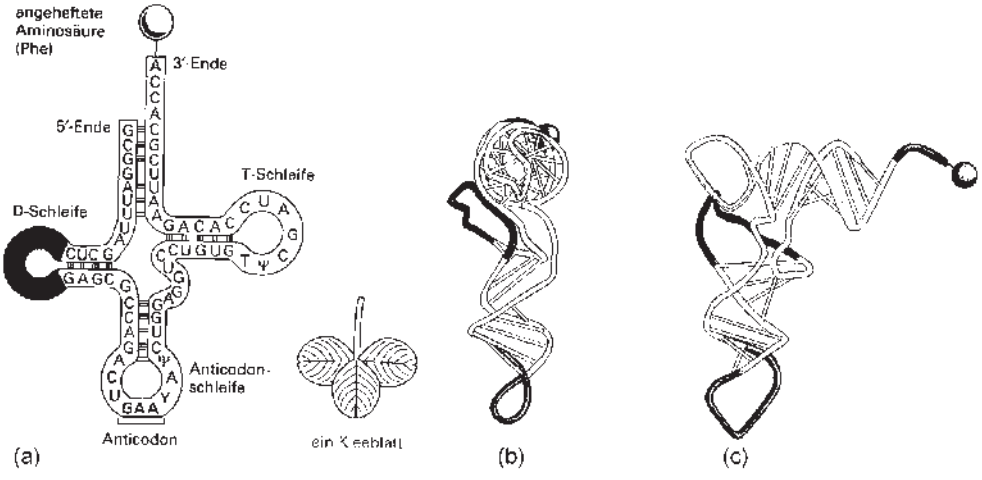


Abb. 1.13 Die tRNA.

In der Abbildung ist die spezifische tRNA für Phenylalanin dargestellt. (a) Die Kleeblattstruktur: In der zweidimensionalen Kleeblattstruktur werden die komplementären Basenpaarungen innerhalb eines tRNA-Moleküls sichtbar, welche die doppelhelikalen Strukturen des Moleküls ausmachen. Das Anticodon besteht aus drei Nukleotiden. Dieser Bereich bindet über komplementäre Basenpaarungen während der Translation mit dem Codon der mRNA. An das 3'-Akzeptor-Ende bindet die Aminoacyl-Transferase die spezifische Aminosäure. Die nicht gepaarten Bereiche bilden Schleifen aus (Anticodon-, D- und T-Schleife). tRNAs enthalten ungewöhnliche Basen, die nach der tRNA-Synthese durch chemische Modifikationen entstehen (Ψ = Pseudouridin, D = Dihydrouridin). (b) und (c) In der dreidimensionalen Ansicht wird die L-förmige Struktur der tRNA sichtbar (Abbildung verändert aus: Alberts, B. et al., 2012, *Lehrbuch der molekularen Zellbiologie*, 4. Auflage, Wiley-VCH Weinheim).

Eukaryonten vor. Der Spleißvorgang unterscheidet sich aber von dem des mRNA-Spleißens.

Transfer-RNAs enthalten außerdem *chemisch modifizierte* RNA-Nukleotide, die entweder zur Konformationsstabilisierung oder zur besseren Erkennung des mRNA-Codons dienen.

Die wichtigsten funktionellen Bereiche einer t-RNA sind das *3'-Akzeptorende*, an der sie mit der für sie spezifischen Aminosäure beladen wird, und das *Anticodon*, das komplementär zum Codon der mRNA ist.

Die Beladung der tRNA mit der spezifischen Aminosäure erfolgt durch Enzyme, die *Aminoacyl-tRNA-Synthetasen*. Eine beladene tRNA wird allgemein als *Aminoacyl-tRNA* bezeichnet. Die jeweilige Aminosäure beziehungsweise das Anticodon legt die spezifische Bezeichnung fest. Eine tRNA, die die Aminosäure *Lysin* transportiert, wird als *Lysyl-tRNA* bezeichnet.

1.7.2.3 Die Ribosomen

Die eigentliche Proteinsynthese findet im Zytoplasma der Zelle an den Ribosomen statt. Ribosomen enthalten mehr als *50 verschiedene Proteine*, die mit den *ribosomalen RNAs (rRNA*) assoziiert sind.

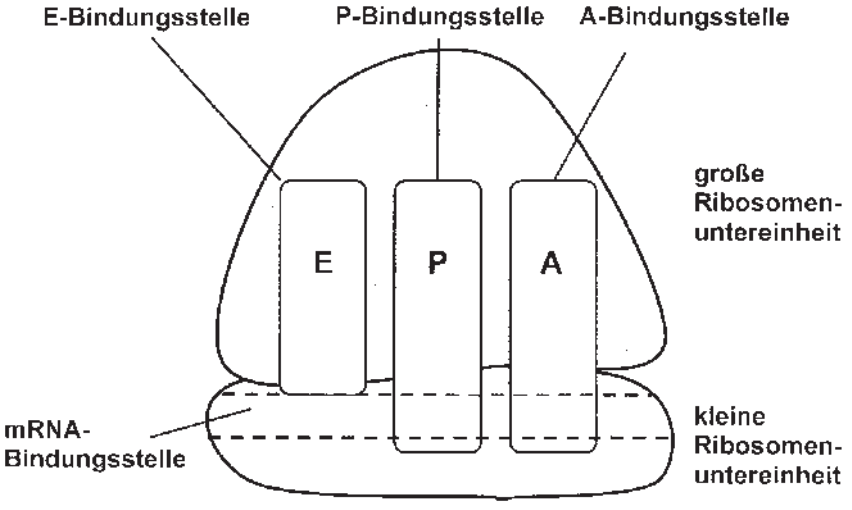


Abb. 1.14 Das Ribosom mit seinen RNA-Bindungsstellen.

Jedes Ribosom hat drei Bindungsstellen für die tRNA: A-Bindungsstelle (*A*minoacyl-tRNA), P-Bindungsstelle (*P*eptidyl-tRNA) und Ausgang (*E*xit). Die mRNA bindet an die kleine Untereinheit.

Ribosomen der Pro- und Eukaryonten sind sehr ähnlich und bestehen aus einer *großen* und einer *kleinen Untereinheit* (Abb. 1.14). Die einzelnen Untereinheiten als auch das gesamte Ribosom werden oft mit der *S-Einheit* versehen (z. B. 50S und 30S-Untereinheit beim Prokaryontenribosom). Die S-Einheit gibt die Sedimentationsgeschwindigkeit der Ribosomen während einer Ultrazentrifugation an.

Die Untereinheiten lagern sich erst unmittelbar vor Beginn der Proteinsynthese in der Nähe des 5'-Bereichs der mRNA zusammen. Ein funktionelles Ribosom hat eine *Bindungsstelle für die mRNA* und drei Bindungsstellen (*A-, P-und E-Bindungsstelle*) für die Aminoacyl-tRNAs.

1.7.2.4 Der Beginn der Proteinsynthese

Für den Translationsstart wird sowohl in Pro- als auch in Eukaryonten eine *Initiator-tRNA* benötigt. Die spezielle tRNA ist immer mit der Aminosäure *Methionin* beladen. In Prokaryonten ist diese Aminosäure allerdings modifiziert (*N-Formylmethionin*). Das entsprechende *Startcodon* auf der mRNA ist *AUG*.

Die mRNA der Prokaryonten enthält eine Konsensussequenz, die als *Bindungsstelle* für die kleine *16S-Einheit des Ribosoms* fungiert (*Shine-Dalgarno-Sequenz*). Ist die mRNA an diese Untereinheit gebunden, kommt die große Untereinheit dazu.

Die Initiator-tRNA der Eukaryonten bindet zuerst mit *eukaryotischen Initiationsfaktoren (eIFs)* an die kleine ribosomale Untereinheit. Die mRNA wird mittels der 5'-Cap-Struktur (s. Abschnitt 1.7.1.1) erkannt und ebenfalls gebunden. Die

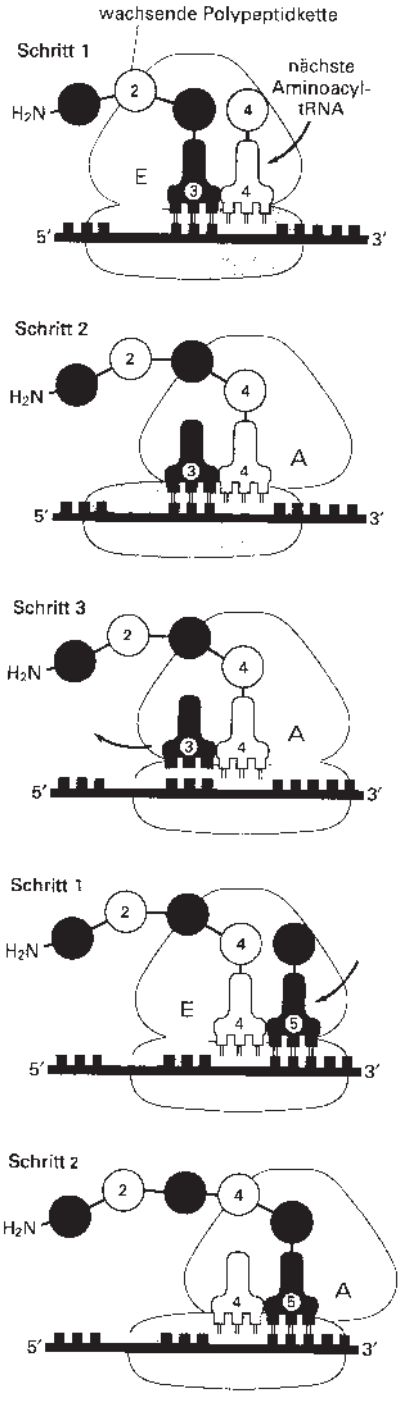


Abb. 1.15 Die Translation.

Die Proteinsynthese läuft zyklisch in drei sich immer wiederholenden Schritten ab:

Schritt 1: Die Aminoacyl-tRNA bindet an die freie A-Bindungsstelle des Ribosoms. Durch die komplementäre Basenpaarung zwischen dem Codon der mRNA und der Aminoacyl-tRNA wird die Aminosäure bestimmt, die an das Ende der Polypeptidkette angehängt wird.

Schritt 2: Die Peptidyltransferase-Aktivität der großen Ribosomenuntereinheit knüpft die Peptidbindung zwischen dem Carboxyende der in der P-Bindungstelle lokalisierten Aminosäure und dem Aminoende der Aminosäure in der A-Bindungsstelle. Das Carboxyende der Polypeptidkette wird dabei von der in der P-Bindungsstelle liegenden tRNA abgekoppelt. Das Ribosom verändert dabei seine Konformation, wobei die beiden tRNAs in die E- und P-Bindungsstelle gezwungen werden.

Schritt 3: Die mRNA wird durch weitere Konformationsänderungen des Ribosoms genau drei Nukleotide weiter bewegt. Die nicht mehr beladene tRNA wird aus der E-Bindungsstelle entlassen und die ursprüngliche Konformation des Ribosoms wird wieder hergestellt. Die Polypeptidkette wächst so von ihrem Amino- zu ihrem Carboxyende. Der Zyklus endet, wenn ein Stop-Codon erreicht wird (Abbildung verändert aus: Alberts, B. et al., 2012, *Lehrbuch der molekularen Zellbiologie*, 4. Auflage, Wiley-VCH Weinheim).

kleine Untereinheit wandert nun an der mRNA bis zum ersten AUG-Codon, wo sich die große Untereinheit anlagert.

Die Translation der mRNA kann beginnen.

1.7.2.5 Das Protein entsteht

Die mit Methionin beladene Initiator-tRNA befindet sich in der *P-Bindungsstelle* des Ribosoms. Eine mit ihrer spezifischen Aminosäure beladene Aminoacyl-tRNA bindet in der *A-Bindungsstelle*. Welche tRNA binden kann, wird durch die komplementäre Basenpaarung zwischen der mRNA und dem Anticodon der tRNA festgelegt. Zwischen der Aminosäure der zuvor gebundenen tRNA in der P-Stelle und der Aminosäure der zuletzt gebundenen tRNA in der A-Stelle wird durch die *Pepidyltransferase-Aktivität* der großen Ribosomenuntereinheit die *Peptidbindung* geknüpft. Durch Konformationsänderung bewegt sich mRNA genau drei Nukleotide, also ein Codon, weiter. Die verbrauchte tRNA wird aus der *E-Stelle* entlassen und kann neu beladen werden. Die mit der wachsenden Polypeptidkette beladene tRNA befindet sich nun in der P-Bindungsstelle. Die A-Bindungsstelle ist wieder frei für die nächste Aminoacyl-tRNA (Abb. 1.15).

1.7.2.6 Das Ende der Translation

Drei spezifische Codons markieren das Ende der Translation. Diese *Stop-Codons UAA, UGA oder UAG* bewirken die Bindung von Freisetzungsfaktoren in der A-Bindungsstelle des Ribosoms. Die Peptidyltransferase transferiert daraufhin ein Wassermolekül anstatt eine Aminosäure auf das *carboxyterminale Ende* der Polypeptidkette, was wiederum zur Folge hat, dass die Bindung zwischen tRNA und Polypeptidkette sofort gelöst und das fertig synthetisierte Protein in das Zytoplasma entlassen wird.

Dort wird es weiter prozessiert und in seine funktionelle Form gebracht. Die *Sekundär- und Tertiärstruktur* ist dabei von der Aminosäurezusammensetzung abhängig. Hydrophobe Bereiche werden sich in das Innere des Moleküls falten, hydrophile nach außen. Hilfsproteine (*Chaperone*) helfen dabei, die Proteine in die richtige Struktur zu falten.

Signalsequenzen zeigen an, ob die Proteine in bestimmte Organellen transportiert oder sezerniert werden sollen.

Die Abb. 1.16 zeigt den gesamten Prozess von der DNA zum Protein für die Prokaryonten und Eukaryonten.

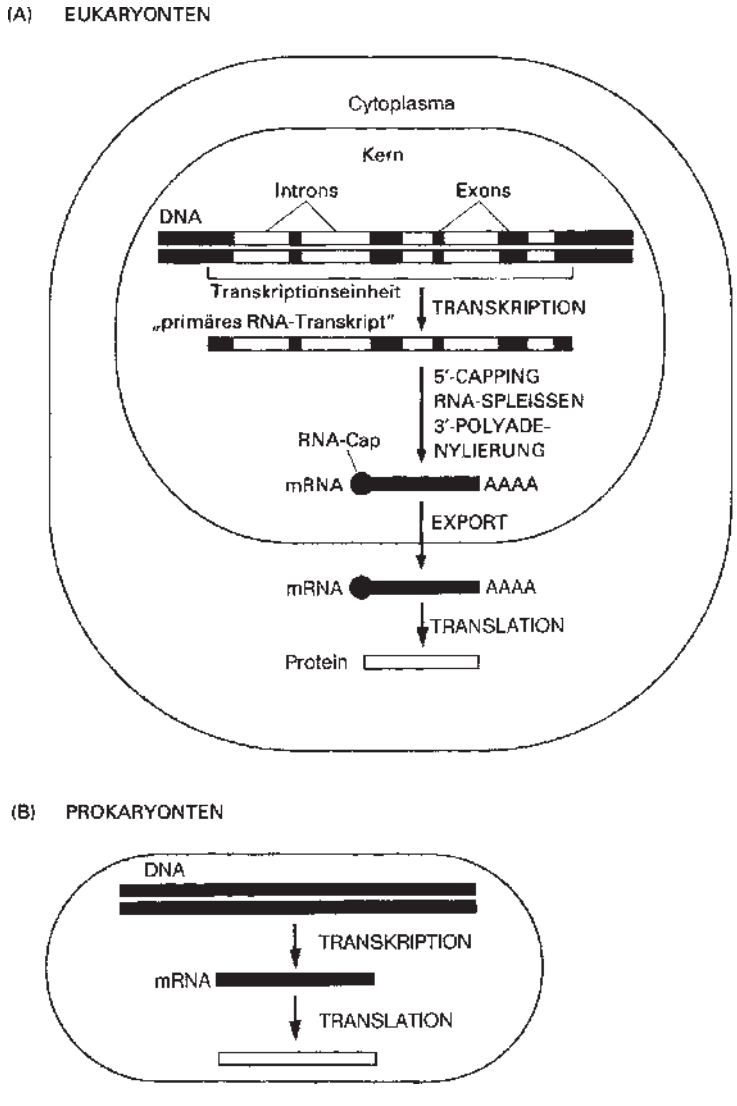


Abb. 1.16 Von der DNA zum Protein.

(a) Eukaryonten: Die in dem Zellkern lokalisierte DNA enthält codierende (Exons) und nicht codierende (Introns) Abschnitte. Alle Abschnitte eines Gens werden transkribiert und das primäre Transkript, die prä-mRNA, synthetisiert. Die prä-mRNA wird weiter prozessiert: Die Introns werden durch das Spleißosom herausgeschnitten. Die mRNA erhält an ihrem 5'-Ende eine chemische Kappe, an ihrem 3'-Ende einen poly A-Schwanz. Sie wird in das Zytoplasma der Zelle transportiert und an den Ribosomen translatiert.

(a) Prokaryonten besitzen keinen Zellkern. Die Transkription und Translation der mRNA findet daher in einem gemeinsamen Zellkompartiment statt. Noch während der Transkription kann die mRNA translatiert werden (Abbildung verändert aus: Alberts, B. et al., 2012, *Lehrbuch der molekularen Zellbiologie*, 4. Auflage, Wiley-VCH Weinheim).

1.8 Grundbegriffe in der molekularen Diagnostik

Alle hier aufgeführten Begriffe und Definitionen sind allgemeingültig und werden nicht nur in der molekularen Diagnostik verwendet, sondern finden ihre Anwendung auch in der allgemeinen Labordiagnostik.

Die Begriffserläuterungen der Abschnitte 1.8.1–1.8.4 beziehen sich auf epidemiologische Betrachtungen und beschreiben Begriffe, die in Bezug auf eine Population, eine definierte Gruppe einer Population oder unter Einbeziehung einer bestimmten Krankheit verwendet werden. So können mithilfe der erhobenen statistischen Daten Krankheitsursachen, zeitliche Verläufe oder lokale Häufungen von Krankheiten erkannt und im besten Fall zukünftig rechtzeitig präventive Maßnahmen zur Eindämmung von Epidemien oder langfristigen Erkrankungen (z. B. Krebs) eingeleitet werden.

Die in Abschnitt 1.8.5–1.8.8 aufgeführten Begriffe beziehen sich auf die in der Diagnostik zur Anwendung kommenden Tests und Testsysteme.

Die in der molekularen Diagnostik verwendeten Tests und Testsysteme zum Nachweis von Krankheitserregern sind mannigfaltig und müssen vielfältigen Ansprüchen genügen. Zur Beurteilung der Leistungsfähigkeit und zur besseren Vergleichbarkeit der verschiedenen Testsysteme werden spezifische Leistungsmerkmale definiert. Diese Merkmale werden im Rahmen von Evaluierungsstudien und den entsprechenden Zulassungsverfahren bestimmt und überprüft (s. a. Kapitel 8).

Zur Beschreibung dieser Leistungsmerkmale werden Begriffe und Definitionen herangezogen, die sich durch statistische Berechnungen und Auswertungen von Testreihen und durch länger angelegte Studien und Beobachtungen ergeben.

1.8.1 Inzidenz

Die Inzidenz beschreibt die Wahrscheinlichkeit für eine beliebige Person innerhalb eines definierten Beobachtungszeitraumes zu erkranken. Sie definiert das *Erkrankungsrisiko* innerhalb einer Population. Die Inzidenz wird als relative Häufigkeit prozentual (0–100 %) oder als Rate (z. B. Anzahl von 1000 Fällen) angegeben.

1.8.2 Prävalenz

Die *Prävalenz* beschreibt die *Erkrankungsrate*, also den Anteil erkrankter Personen einer definierten Population zu einem festen Zeitpunkt (*Punktprävalenz*) oder einem festen Beobachtungszeitraum (*Periodenprävalenz*). Die Prävalenz wird ebenfalls als relative Häufigkeit prozentual (0–100 %) oder als Rate (Anzahl von 1000 Fällen) angegeben.

1.8.3 Mortalität

Die *Mortalität* gibt die Anzahl der Todesfälle in einem bestimmten, festgelegten Zeitraum an und wird auch als *Todesrate* bezeichnet (*rohe Mortalitätsrate*).

Wird die Rate auf eine bestimmte Krankheit bezogen, wird sie als *krankheitsspezifische* Mortalitätsrate definiert. Bei der *fallspezifischen* Mortalitätsrate werden bestimmte Ursachen in Bezug genommen, während bei der *altersspezifischen* Mortalitätsrate definierte Altersgruppen mit einbezogen werden.

Auch hier wird die Rate als relative Häufigkeit prozentual (0–100 %) oder bezogen auf 1000 Individuen einer Population angegeben.

1.8.4 Letalität

Die *Letalität* oder *Tödlichkeitsrate* steht in direktem Zusammenhang mit der Mortalität und Inzidenz. Sie stellt die Anzahl der Todesfälle, die durch eine bestimmte Krankheit bedingt sind, in direkten Bezug zueinander.

Letalität = Mortalität/Inzidenz

Sie gibt also die Wahrscheinlichkeit wieder, innerhalb eines definierten Zeitraums an einer Krankheit, z. B. Influenza, zu versterben.

1.8.5 Goldstandard

Der Goldstandard ist kein fest definierter Begriff, sondern bezeichnet eine Methode, die bezüglich der u. a. Leistungsmerkmalen allen anderen überlegen ist und/oder z. B. aufgrund der Exklusivität der Methode von der Mehrheit der Anwender durchgeführt wird. Neue Testverfahren werden mit dem Goldstandard verglichen. Beweist sich ein neuer Test auf Dauer dem bisherigen Goldstandard als überlegen, wird er zum neuen Goldstandard deklariert.

1.8.6 Richtigkeit und Präzision

Die *Richtigkeit* und *Präzision* eines Verfahrens sind grundlegend für jedes Messverfahren. Die Bestimmung erfolgt über wiederholte Messungen und den anschließenden Vergleich der Mittelwerte der Messergebnisse mit dem wahren zuvor bekannten Wert. Der wahre Wert kann mithilfe von Referenzmaterialien und/oder Referenzmethoden bestimmt oder in Erfahrung gebracht werden. Erfasst wird so auch die systematische Messwertabweichung.

Durch Vergleich der Werte und unter Berücksichtigung vorher definierter Grenzen der Messwertabweichung werden Richtigkeit und Präzision bestimmt (Abb. 1.17).

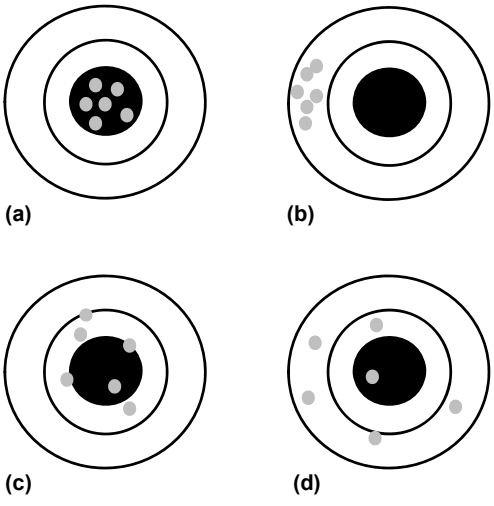


Abb. 1.17 Richtigkeit und Präzision.
(a): Richtig und Präzise: Die Messwerte weichen untereinander nicht signifikant voneinander ab und streuen innerhalb der definierten Grenze um den wahren Wert.
(b): Präzise aber Unrichtig: Die Messwerte weichen untereinander nicht signifikant voneinander ab, streuen aber oberhalb der definierten Grenze vom wahren Wert.
(c): Richtig aber Unpräzise: Die Messwerte weichen untereinander signifikant voneinander ab, streuen aber innerhalb der definierten Grenze um den wahren Wert.
(d): Unrichtig und Unpräzise: Die Messwerte weichen untereinander signifikant voneinander ab und streuen oberhalb der definierten Grenze um den wahren Wert.

1.8.7 Sensitivität und Spezifität

Diagnostische Tests sollen Testresultate liefern, die eine Unterscheidung zwischen erkrankten und gesunden Personen hinsichtlich einer bestimmten Erkrankung erlauben. In der Realität gibt es allerdings Einflussgrößen, die eine solche Unterscheidung schwierig machen können. So können bei Gesunden z. B. aufgrund von anderen Begleiterkrankungen, Medikamenteneinnahme oder Umwelteinflüssen pathologische Werte gemessen werden, währenddessen bei Erkrankten normale Messergebnisse generiert werden. Sensitivität und Spezifität sind statistisch ermittelte Parameter, die die Fähigkeit eines diagnostischen Tests beschreiben, zwischen gesunden und erkrankten Personen innerhalb einer Population zu unterscheiden, und als Gütekriterium dienen.

Man unterscheidet zwischen der *klinischen* und *analytischen* Sensitivität bzw. zwischen der *klinischen* und der *analytischen* Spezifität eines diagnostischen Tests.

1.8.7.1 Klinische Sensitivität

Die *klinische* oder auch *diagnostische Sensitivität* eines Tests gibt die Wahrscheinlichkeit an, mit der bei einer erkrankten Person aus der Gruppe aller Erkrankten das Testergebnis „richtig positiv" ausfällt. Sie ergibt sich aus der Formel:

$$\text{Sensitivität}\,[\%] = \frac{\text{Anzahl richtig positiver Befunde}}{(\text{Richtig positive} + \text{falsch negative Befunde})} \times 100$$

Bei dieser Vorgehensweise muss anhand einer Referenzmethode (Goldstandard; s. Abschnitt 1.8.5) die Summe der erkrankten Personen bestimmt werden.

1.8.7.2 Analytische Sensitivität

Die *analytische Sensitivität* ist eine direkte Testspezifikation. Synonym werden auch die Begriffe „*Nachweisgrenze*" oder „*Limit of Detection" (LoD*) verwendet.

Die analytische Sensitivität bezeichnet den unteren Wert eines Messverfahrens, zu dem der Analyt gerade noch zuverlässig nachgewiesen werden kann. Hierzu wird zuerst eine definierte Anzahl von Messungen mit einem Leerwert (*engl.:* blank) durchgeführt. Anhand dieser Messergebnisse wird die Schwankung des Messsignals definiert, wenn kein Analyt vorhanden ist (Standardabweichung des Leer- oder Blindwertes; *engl.:* Limit of Blank (LoB). Danach erfolgen die Messreihen mit niedrigen, definierten Konzentrationen der Proben.

Der Messwert an der Nachweisgrenze ist mit einer erhöhten Ungenauigkeit behaftet. Diese Ungenauigkeit wird statistisch ermittelt und darf einen bestimmten Vertrauensbereich (*Konfidenzintervall*) nicht überschreiten. Daher wird die analytische Sensitivität häufig mit einer 95 %igen Wahrscheinlichkeit angegeben. Messwerte, die eine größere Ungenauigkeit aufweisen als der vorgegebene Bereich, liegen unterhalb der Nachweisgrenze und müssen als „*nicht nachweisbar*" gekennzeichnet werden.

Quantitative Messverfahren geben die Konzentration eines Analyten in einer Probe an. In der molekularen Diagnostik werden diese Verfahren zumeist zur Therapieüberwachung (z. B. bei HIV-, Hepatitis B oder Hepatitis C-Infektionen) eingesetzt. Gemessen wird hier der sogenannte „*virus-load*", also die Konzentration der Viren im peripheren Blut. Diese wird anhand interner oder externer Standards und Standardkurven von den Testsystemen errechnet. Zu beachten ist hierbei, dass eine genaue Quantifizierung nur innerhalb eines bestimmten und definierten Bereichs möglich ist. Unterhalb dieses Bereichs ist keine valide Quantifizierung möglich. Diese Grenze wird als „*Limit of Quantification" (LoQ*) bezeichnet. LoD und LoQ können identisch sein, sich aber auch bei einigen Testsystemen voneinander unterscheiden. Unterhalb des LoQ ist dann nur eine qualitative Angabe („nachweisbar" oder „nicht nachweisbar") möglich.

Nachweisgrenzen und Quantifizierungsbereiche müssen mit dem verwendeten Testverfahren auf dem Befundbericht angegeben werden.

1.8.7.3 Klinische Spezifität

Die *klinische* oder auch *diagnostische Spezifität* eines Tests gibt die Wahrscheinlichkeit an, mit der bei einer gesunden Person aus der Gruppe aller Gesunden das Testergebnis „richtig negativ" ausfällt. Sie ergibt sich aus der Formel:

$$\text{Spezifität}\,[\%] = \frac{\text{Anzahl richtig negative Befunde}}{(\text{Anzahl der negativen} + \text{Anzahl der falsch positiven Befunde})} \times 100$$

Auch hier muss anhand genau definierter Kriterien und/oder mit einer Referenzmethode die Anzahl der tatsächlich gesunden Personen bestimmt werden, um die falsch positiven Resultate zu selektieren.

1.8.7.4 Analytische Spezifität

Die *analytische Spezifität* bezeichnet die Fähigkeit eines Tests, den nachzuweisenden Analyten, in der molekularen Diagnostik zumeist Krankheitserreger, eindeutig zu bestimmen. Kreuzreaktivitäten mit anderen Erregern sollen ausgeschlossen werden, um falsch positive Befunde zu vermeiden.

Ermittelt wird die analytische Spezifität anhand von Testreihen, in denen Erreger getestet werden, die entweder nah verwandt sind, die im selben Probenmaterial nachgewiesen werden können oder die ähnliche Krankheitsbilder und Symptome hervorrufen.

1.8.8 Prädiktiver Wert

Die klinische Sensitivität und Spezifität sind Gütekriterien für einen Test, die meistens vom Anwender zurate gezogen werden, um einen Test zu beurteilen oder zu vergleichen. Im klinischen Alltag interessieren eher Aussagen bezüglich der Wahrscheinlichkeit, dass eine Erkrankung bei positivem Ergebnis tatsächlich vorliegt. Diese Wahrscheinlichkeit wird durch den *positiven prädiktiven Vorhersagewert* ausgedrückt. Der *negative prädiktive Wert* drückt folgerichtig die Wahrscheinlichkeit aus, dass der untersuchte Patient bei negativem Testergebnis tatsächlich gesund ist.

$$\textit{Positiver Vorhersagewert} = \frac{\text{richtig positive Befunde}}{\text{richtig positive} + \text{falsch positive Befunde}}$$

$$\textit{Negativer Vorhersagewert} = \frac{\text{richtig negative Befunde}}{\text{richtig negative} + \text{falsch negative Befunde}}$$

Die Vorhersagewerte sind abhängig von der Prävalenz eines Patientenkollektivs (s. Abschnitt 1.8.2). Werden bei einer niedrigen Prävalenz mehr gesunde als kranke Patienten getestet, ist die Wahrscheinlichkeit falsch positiver Resultate höher als die der falsch negativen.

Die in diesem Kapitel aufgeführten Definitionen und Erklärungen beziehen sich auf den Nachweis von Krankheitserregern. In der Humangenetik werden Tests und Diagnostika natürlicherweise auch miteinander verglichen und unter-

liegen ebenfalls strengen Regularien. So können die hier verwendeten Begriffe gleiche oder ähnliche Bedeutung haben. Allerdings unterscheiden sich die Kriterien hinsichtlich der Beurteilung eines Tests oder Testsystems oftmals von den hier aufgeführten und werden daher in Kapitel 7 beschrieben.