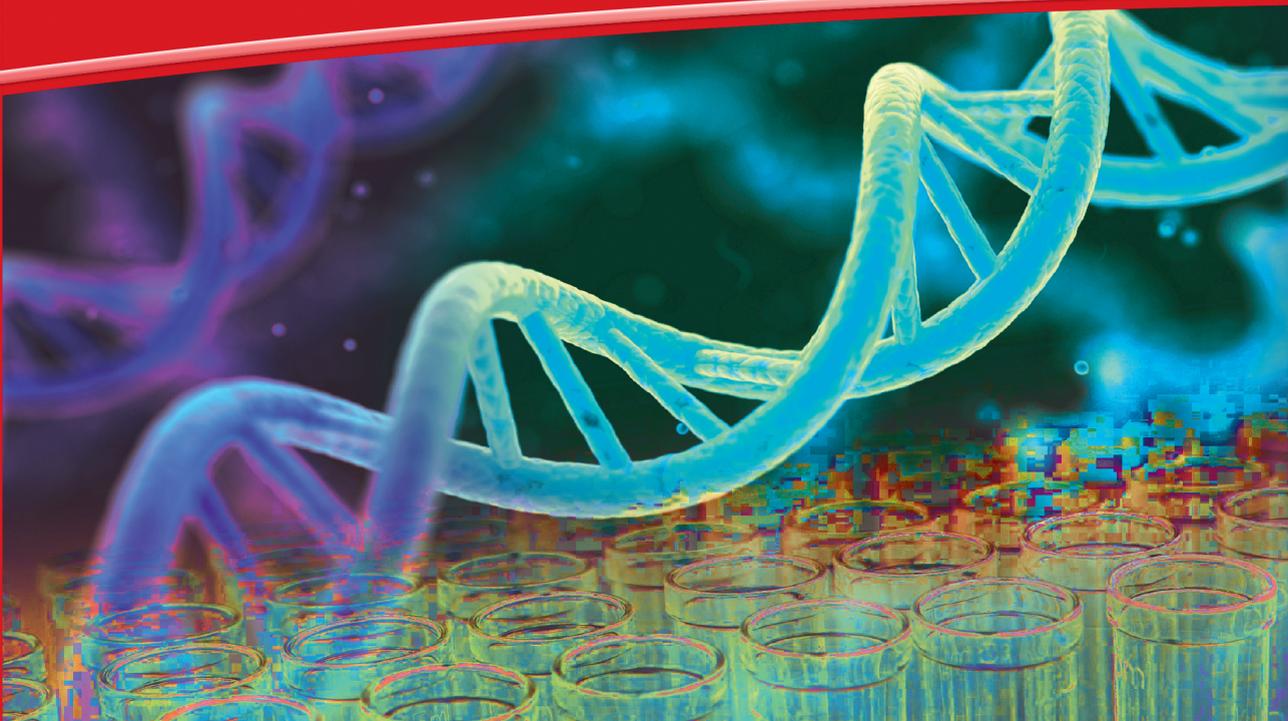


utb.

Thomas Reinard

Molekularbiologische Methoden 2.0

3. Auflage



utb 8449



Eine Arbeitsgemeinschaft der Verlage

Brill | Schöningh – Fink · Paderborn

Brill | Vandenhoeck & Ruprecht · Göttingen – Böhlau Verlag · Wien · Köln

Verlag Barbara Budrich · Opladen · Toronto

facultas · Wien

Haupt Verlag · Bern

Verlag Julius Klinkhardt · Bad Heilbrunn

Mohr Siebeck · Tübingen

Narr Francke Attempto Verlag – expert Verlag · Tübingen

Ernst Reinhardt Verlag · München

transcript Verlag · Bielefeld

Verlag Eugen Ulmer · Stuttgart

UVK Verlag · München

Waxmann · Münster · New York

wbv Publikation · Bielefeld

Wochenschau Verlag · Frankfurt am Main

Thomas Reinard

Molekularbiologische Methoden 2.0

3., aktualisierte Auflage

145 Abbildungen

12 Maps

50 Tabellen

Verlag Eugen Ulmer Stuttgart

Thomas Reinard wurde 1963 in Leverkusen geboren und studierte nach dem Abitur 1982 an der Friedrich-Wilhelms-Universität Bonn Biologie. Das Thema seiner Dissertation war die Charakterisierung von Auxin-bindenden Proteinen und wurde am Institut für Genetik und einem längerem Aufenthalt an der University of North Carolina at Chapel Hill durchgeführt. Anschließend wechselte er an das Institut für Pflanzengenetik, Abt. II, der Leibniz Universität Hannover (<https://www.genetik.uni-hannover.de/de/pflanzenbiotechnologie/ag-reinard/>), wo er als Akademischer Oberrat arbeitet. Die von ihm geleitete Arbeitsgruppe beschäftigt sich mit der Herstellung pharmazeutisch relevanter Proteine in Pflanzen, vor allem aus Wasserlinsen und Mikroalgen. Er bietet in den Studiengängen Life Sciences, Pflanzenbiotechnologie und Biologie eigenverantwortlich Lehrveranstaltungen in den Fächern Molekularbiologie und Bioinformatik an und ist für Bachelor- und Masterstudiengänge prüfungsberechtigt.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

© 2010/2021 Eugen Ulmer KG
Wollgrasweg 41, 70599 Stuttgart (Hohenheim)
E-Mail: info@ulmer.de
Internet: www.ulmer.de
Lektorat: Sabine Mann
Herstellung: Jürgen Sprengel
Umschlagbild: ©enot-poloskun (eingefärbt)
Umschlaggestaltung: Atelier Reichert, Stuttgart
Satz: Bernd Burkart; www.form-und-produktion.de
Druck und Bindung: Pustet, Regensburg
Printed in Germany

UTB-Band-Nr. 8449

ISBN 978-3-8252-8795-5 (Print)

ISBN 978-3-8385-8795-0 (E-Book)

Inhalt

Vorwort zur 1. Auflage	12
Vorwort zur 2. Auflage	13
Vorwort zur 3. Auflage	13

1 Das Leben und seine Bestandteile

1.1	Der Aufbau von DNA und RNA	17	1.4	Proteine	26
1.2	Der genetische Code	20	1.5	Die Zelle	30
1.3	Die Gene	22			
1.3.1	Die Genstruktur in Prokaryoten ..	22			
1.3.2	Genstruktur in Eukaryoten	24			

2 Grundlagen der Arbeit im Labor

2.1	Wasser	34	2.9	Probenlagerung	46
2.2	Messung des pH-Werts	35	2.10	Steriles Arbeiten	46
2.3	Puffer	36	2.11	Geräte für den Aufschluss von Geweben	49
2.4	Waagen	36	2.12	Pflege und Aufzucht von <i>Escherichia coli</i>	51
2.5	Mikropipetten	38	2.12.1	Nährmedien	52
2.6	Gefäße im Labor	40	2.12.2	Antibiotika	53
2.7	Zentrifugen	41	2.12.3	Lagerung von Bakterien	55
2.8	Mischen und Konzentrieren ..	44			
2.8.1	Konzentrieren	45			
2.8.2	Pufferwechsel	45			

3 Aufreinigung von Nukleinsäuren

3.1 Gegenseitler erfolgreicher DNA- und RNA-Isolationen . . .	59	3.4.2 Übersicht über den Einsatz von Kits	70
3.1.1 Nukleasen	59	3.5 Ausgewählte DNA-Aufreinigungsverfahren	71
3.1.2 Scherkräfte	60	3.5.1 Die Plasmidpräparation aus <i>E. coli</i>	71
3.1.3 Chemische Verunreinigungen	60	3.5.2 Die Isolation von DNA aus Pflanzen mit CTAB	72
3.1.4 EDTA und zweiwertige Ionen: das Yin und Yang der Molekularbiologie	61	3.5.3 Isolation von DNA aus Blut oder Zellkulturen	74
3.2 Extraktion von Nukleinsäuren	62	3.5.4 Isolation hochmolekularer DNA	75
3.3 Die weitere Aufreinigung der DNA	64	3.6 Die Isolation von RNA	76
3.3.1 Phenolextraktion	64	3.7 Tipps zum Erzielen hoher Ausbeuten bei der Isolation von Nukleinsäuren	78
3.3.2 Ethanolfällung	65	3.8 Quantifizierung von Nukleinsäuren	79
3.3.3 Isopropanolfällung	67	3.8.1 DNA-Bestimmung im Photometer	79
3.3.4 PEG-Fällung	68	3.8.2 Konzentrationsbestimmung mittels optischer Dichte	81
3.3.5 Entfernen von Polysacchariden mit CTAB	68		
3.3.6 Tropfendialyse	68		
3.4 Silica Matrices	69		
3.4.1 Von Nukleinsäuren und Silica Matrices	69		

4 Polymerase-Kettenreaktion

4.1 Prinzip der Polymerase-Kettenreaktion	85	4.5.2 Gradienten-PCR und <i>Touch-Down</i> PCR	96
4.2 Die Komponenten der PCR	86	4.5.3 <i>Nested</i> PCR	97
4.2.1 Die Ausgangs-DNA	86	4.5.4 Multiplex PCR	97
4.2.2 Thermostabile DNA-Polymerasen	86	4.5.5 Einführen von Restriktionsschnittstellen	98
4.2.3 Puffer, Magnesium und dNTPs	89	4.5.6 Reverse Transkriptions-PCR (RT-PCR)	99
4.2.4 Primer	91	4.5.7 Kolonie-PCR	102
4.3 Geräte für die PCR	93	4.5.8 Quantitative PCR	103
4.4 Die Standard-PCR	94	4.6 Anwendungen der PCR	105
4.5 Ausgewählte PCR-Methoden	94	4.7 Optimierung der PCR-Reaktion	105
4.5.1 <i>Two-Step</i> PCR	94		

5 Klonieren für Einsteiger

5.1	Restriktionsenzyme	109	5.6	Der Weg zum klonierten Gen.	123
5.1.1	Restriktionsenzyme des Typs II.	110	5.7	Der ungerichtete Einbau einer amplifizierten DNA in einen Vektor	123
5.1.2	Restriktionsenzyme im Labor	113	5.7.1	Die Klonierung über eine Restriktionsstelle	124
5.1.3	Methylasen und Typ IIM-Enzyme	115	5.7.2	TA-Klonierung und TOPO-Cloning	125
5.1.4	Typ IIS-Enzyme	116	5.7.3	Klonierung in ein Letal-Plasmid	126
5.2	Ligation	116	5.7.4	Zyklische <i>Blunt End</i> Klonierung.	126
5.3	(De)Phosphorylierung von DNA	118	5.8	Der gerichtete Einbau von DNA in einen Vektor	127
5.3.1	Dephosphorylierung	118	5.9	Wenn es mal nicht klappt	129
5.3.2	Phosphorylierung	119			
5.4	Enzyme für spezielle Aufgaben	121			
5.5	Die Transformation von <i>E. coli</i>	121			

6 Vektoren

6.1	Plasmide	132	6.8	Künstliche Chromosome: YACs, BACs und PACs	148
6.2	Klonierungsplasmide	135	6.9	Hefen als Klonierungs- und Expressionssystem	149
6.2.1	Blau-Weiß-Screening	135	6.9.1	Hefe als Modellorganismus	149
6.2.2	Letalvektoren	138	6.9.2	Vektoren für Hefe	151
6.3	Expressionsplasmide	139	6.9.3	Transformation von Hefe	152
6.3.1	Promotoren für Expressionsvektoren	140	6.10	Pflanzen als Bioreaktoren	152
6.3.2	Weitere regulative Sequenzen	141	6.10.1	Die Transformation von Pflanzen	153
6.3.3	Expression in das Periplasma	141	6.10.2	Transiente Transformationsverfahren	155
6.3.4	Einschlusskörperchen [<i>Inclusion Bodies</i>]	141	6.11	Die Transformation von Säugetieren und tierischen Zellkulturen	156
6.3.5	<i>Tag</i> -Sequenzen	142	6.11.1	Säugetiere als Modellorganismen	156
6.4	Reportergene	142	6.11.2	Säuger-Zellkulturen	157
6.5	Phagen	146	6.11.3	Vektoren für tierische Zellen	158
6.6	Phagemide	147	6.11.4	Transfektion tierischer Zellkulturen	159
6.7	Shuttle-Vektoren	147			

7 Elektrophorese und Hybridisierung von Nukleinsäuren

7.1	Agarose-Gelelektrophorese von DNA	162	7.5	Fehlersuche bei Agarose-Gelelektrophorese	170
7.2.	Die Detektion der DNA im Gel	167	7.6	Hybridisierung von Nukleinsäuren	171
7.3	Präparative Agarosegele	169	7.6.1	Southern und Northern Blot	173
7.4	Auftrennen von RNA im Agarosegel	170	7.6.2	Herstellung der Sonde und Hybridisierung	174
			7.7	Microarrays	174

8 Fortgeschrittene Klonierung

8.1	Klonsammlungen und synthetische Gene	179	8.6	Biobricks	189
8.1.1	Klonsammlungen	180	8.7	Nahtlose Klonierungsverfahren	189
8.1.2	Synthetische Gene und Genfragmente	180	8.7.1	Golden Gate Klonierung	190
8.2	Rekombinase-basierte Klonierung	181	8.7.2	<i>Cut-Ligation</i> Verfahren	191
8.3	Mutageneseverfahren	183	8.7.3	Multiple Insertionen mittels Golden Gate Klonierung	192
8.4	PCR-basierte Klonierungsverfahren: RF-Cloning und oePCR	186	8.7.4	Golden Gate basierte Tool Kits am Beispiel des MoClo Kits	192
8.5	Synthetische Biologie	188	8.8	Gibson Assembly	195
			8.9	Vergleich der Verfahren	198

9 Proteinaufreinigung

9.1 Homogenisation	203	9.5 Quantifizierung von Proteinen	217
9.1.1 Proteasen	204	9.5.1 Proteinbestimmung nach Bradford	217
9.1.2 Phenoloxidasen	206	9.5.2 BCA-Test	219
9.2 Extraktion von Proteinen	207	9.5.3 Welchen Test benutzen?	220
9.2.1 Homogenisationspuffer	207	9.6 Aufreinigungsverfahren für getaggte Proteine aus <i>E. coli</i>	221
9.2.2 Abtrennen von Zell- und Gewebetrümmern	208	9.6.1 Bakterienanzucht und Homogenisation	222
9.2.3 Fraktionierung des Rohextrakts	208	9.6.2 Weitere Aufreinigung des heterologen Proteins	224
9.3 Dialyse und Konzentrierung	211		
9.4 Weitere Aufreinigungsschritte	212		
9.4.1 Chromatographie	213		
9.4.2 Chromatographische Trennprinzipien	215		
9.4.3 <i>Magnetic Beads</i>	216		

10 Gelelektrophorese von Proteinen

10.1 Die Polyacrylamid-Gelelektrophorese (PAGE)	229	10.4 Weitere Elektrophoreseverfahren für Proteine	239
10.2 Die denaturierende SDS-PAGE	229	10.4.1 Native PAGE	239
10.2.1 Herstellung eines Proteinextrakts für die SDS-Gelelektrophorese	229	10.4.2 Isoelektrische Fokussierung	239
10.2.2 Herstellen des Gels	232	10.4.3 2-D-Gelelektrophorese	239
10.2.3 Der Gellauf	235	10.5 Western Blotting	241
10.3 Das Färben der Gele	237	10.6 Massenspektrometrische Verfahren	246

11 Immunbiochemische Methoden

11.1	Antikörper	251	11.3	ELISA	261
11.2	Prinzipien immun- biochemischer Verfahren	254	11.3.1	Sandwich-ELISA	263
11.2.1	Immunpräzipitation	254	11.3.2	Kompetitiver ELISA	265
11.2.2	Trägerbasierte Immunfärbung ...	256	11.4	Immunfärbung nach Western Blot	266
11.2.3	Immobilisierung des Antigens ...	257	11.5	Immunaffinitäts- chromatographie	269
11.2.4	Blocken überschüssiger Bindestellen	257	11.6	Weitere Antikörper- basierte Methoden	270
11.2.5	Detektion der Bindung	257	11.6.1	Rekombinante Antikörper	270
11.2.6	Antikörper-Kaskaden	259	11.6.2	<i>Phage Display</i> und scFv	271
11.2.7	Spezifische und unspezifische Bindungen	260	11.6.3	Immunhistochemische Verfahren .	273
			11.6.4	Markierung von Zellen	274

12 Fortgeschrittene Verfahren

12.1	DNA-Sequenzierung	278	12.4	Analyse der Genfunktion	286
12.1.1	Sequenzierung nach Sanger	278	12.4.1	Reverse Genetik und <i>Gene Targeting</i>	286
12.1.2	Herstellen von Proben für die DNA-Sequenzierung	280	12.4.2	RNAi	287
12.2	Next Generation Sequenzierung	281	12.4.3	Durchführung von siRNA-Experimenten	288
12.3	Von den „Omics“ zur Systembiologie	285	12.5	Genom Editierung	289

13 Bioinformatik

13.1	Datenbanken	296	13.4	Sequenzsuche mit einer Sequenz (BLAST)	300
13.2	Dateiformate	298	13.5	Bioinformatik zur Planung der Laborarbeit	305
13.3	Sequenzsuche anhand von Stichwörtern (Entrez)	300			

14 Anhang

A1	Sicherheit im Labor.....	308	Verzeichnis der „Gut zu wissen“-Boxen	313
A2	Die 10 goldenen Laborregeln ..	309	Abbildungsverzeichnis.....	314
A3	Das Laborbuch und die finale Arbeit.....	311	Verzeichnis der Maps	316
			Verzeichnis der Protokolle	316
			Tabellenverzeichnis.....	317
			Abkürzungsverzeichnis	318
			Sachverzeichnis.....	321
			Quellenverzeichnis.....	328

Vorwort zur 1. Auflage

Dieses Buch basiert auf meiner gleichnamigen Vorlesung, die sich an fortgeschrittene Studierende der Lebenswissenschaften richtet. Es soll die Lücke zwischen den allgemeinen Lehrbüchern der ersten Semester und reinen Methodenbüchern schließen.

Darüber hinaus ist es ein Reiseführer durch das zunächst unentdeckte Land molekularbiologisch geprägter Lebenswissenschaften mit all ihren Facetten. Daher wird Ihnen dieses Buch auch später im Labor ein hoffentlich wertvoller Berater sein. Denn obwohl die Arbeit im Labor heute durch die Verwendung von Kits bestimmt ist, entheben diese den experimentierfreudigen Studierenden nicht von der Notwendigkeit, die zugrundeliegenden Prinzipien zu kennen. Nur so kann der Fehler gefunden werden, wenn ein Verfahren nicht wie vorgesehen funktioniert. Noch viel wichtiger ist es jedoch, dass das Fehlen dieser Kenntnisse dazu führen kann, Ergebnisse falsch zu interpretieren oder dass wichtige Kontrollen vergessen werden. Dieses Buch kann nur beispielhaft die große Zahl der molekularbiologischen Methoden beleuchten, daher wurde der Fokus auf die Aktualität und Verbreitung der Verfahren gesetzt. Oft lassen sich bei unterschiedlichen Methoden starke Parallelen ausmachen, weshalb die Concept-Maps der zunächst verwirrenden Vielfalt der Verfahren eine Struktur geben und Zusammenhänge zwischen ihnen aufdecken. Der Text wird durch verschiedene Textboxen aufgelockert:

GUT ZU WISSEN-Boxen

liefern Hintergrundwissen.



TIPP-Boxen geben wertvolle Tipps für die praktische Arbeit.



PROTOKOLL-Boxen vervollständigen besonders wichtige Verfahren mit zusätzlichen und detaillierten Informationen.

Um das Buch preislich erschwinglich zu halten, mussten leider einige interessante Methoden ausgelagert werden und stehen daher nur online zur Verfügung. Es lohnt sich sehr, in den umfangreichen Online-Bereich mehr als nur einen Blick zu werfen.

Natürlich entsteht ein Buch nicht im stillen Kämmerlein. Ohne die Hilfe von meinen Kollegen und Freunden wäre das vorliegende Werk sicher unfertig dahergekommen, daher möchte ich André Frenzel, Wiebke Rathje und Maren Wichmann ganz herzlich für die viele konstruktiven Diskussionen und Vorschläge danken. Martina Famulla hat die Ergebnisse, die in vielen Fotos zu sehen sind, im Labor produziert. Meinen Mitarbeitern und Kollegen, die ihre Fotos zur Verfügung gestellt haben, möchte ich ebenfalls herzlich danken.

Ohne meine Familie, Elvira, Henrik und Clara, die mich die ganze Zeit unterstützt hat, wäre dieses Projekt niemals verwirklicht worden. Danke! So, und dies soll genug der Vorrede sein. Steigen wir nun ein in die Welt der molekularbiologischen Methoden, aber nicht ohne einen letzten wichtigen Hinweis zu geben:

Sicher wird die ein oder andere hier beschriebene Methode in manchen Laboren abweichend durchgeführt. Wenn das Protokoll funktioniert, gibt es keinerlei Grund daran etwas zu ändern. Bevor wir uns nun auf den spannenden Weg durch die Methoden der Molekularbiologie begeben, bleiben mir nur noch zwei Worte mitzuteilen:

KEINE PANIK!

Vorwort zur 2. Auflage

Es ist unglaublich, in welchem Tempo sich die Molekularbiologie seit Erscheinen der ersten Auflage weiterentwickelt hat. Technologien, die 2010 noch nicht einmal erfunden waren, sind heute bereits in vielen Laboren anzutreffen:

Verfahren wie die Modulare Klonierung, PCR-basierte Klonierungen oder das Gibson Assembly stellen heute Standardverfahren der Molekularbiologie dar. Die atemberaubenden Fortschritte bei der Genomsequenzierung und der chemischen DNA-Synthese haben die Denkansätze und Forschungsstrategien in den letzten Jahren grundlegend verändert. Anstatt ein einzelnes Gen mit seinen Funktionen zu beschreiben, stehen heute oft sogenannte massiv parallele Verfahren, die viele Millionen Daten gleichzeitig generieren, im Vordergrund. Sie sollen es ermöglichen, ein detailliertes Bild eines komplexen biologischen Systems mit all seinen internen Interaktionen zu liefern. Obwohl die Auswirkungen der „Omics“-Technologien, die ein ganzes Genom oder Proteom in wenigen Tagen sequenzieren und analysieren kön-

nen, noch nicht vollständig abzuschätzen ist, steht schon die nächste Revolution in den Startlöchern: Erst vor wenigen Jahren wurden die Techniken zur Genom Editierung wie TALEN oder CRISPR/Cas entwickelt. Schon heute sind sie weit über die Biologie hinaus bekannt, und eines ist klar: diese Verfahren haben schon jetzt einen enormen Einfluss auf die Lebenswissenschaften und darüber hinaus.

All das ist Grund genug, die zweite Auflage der „Molekularbiologischen Methoden“ komplett zu überarbeiten. Die schnelle methodische Entwicklung der Molekularbiologie schlägt sich auch im Umfang des Buches nieder, der um fast ein Drittel gestiegen ist.

Die Überarbeitung eines Lehrbuchs erfordert kompetente Hilfe, daher möchte ich besonders Dr. Maren Wichmann und Dr. Jana Streubel für ihre sehr hilfreichen Korrekturen und konstruktiven Vorschläge danken.

Vorwort zur 3. Auflage

Dass die „Molekularbiologischen Methoden 2.0“ so schnell nach ihrer Veröffentlichung bereits eine 3. Auflage erfordern, zeigt, dass die Bedeutung von Lehrbüchern trotz der vielen digitalen Lernangebote noch immer hoch ist. Ich habe in dieser Auflage neben einigen wenigen Korrekturen vor

allem kleinere Aktualisierungen sowie zusätzliche Tipps und Tricks aufgenommen, welche die Fehlersuche im Labor erleichtern können. Manche beruhen auf Hinweisen von Leser*innen, was mich besonders freut.

1 Das Leben und seine Bestandteile

- 1.1 Der Aufbau von DNA und RNA
- 1.2 Der genetische Code
- 1.3 Die Gene
- 1.4 Proteine
- 1.5 Die Zelle

Auf den ersten Blick erscheint das Leben auf der Erde extrem vielfältig. Schon in der kleinsten biologischen Einheit, der Zelle, finden wir komplexe strukturelle und funktionale Zusammenhänge. Auf den zweiten Blick lassen sich allgemeine Prinzipien entdecken, die es erleichtern, Gemeinsamkeiten zwischen den vielfältigen biologischen Vorgängen zu entdecken:

1. Alle biologischen Systeme sind hierarchisch strukturiert.
2. Es gibt in biologischen Systemen eine sehr enge Verknüpfung zwischen der Struktur und der Funktion.

Gerade einmal vier verschiedene Arten an Makromolekülen sind für die enorme Vielfalt verantwortlich. In diesem Buch befassen wir uns mit den Nucleinsäuren und den Proteinen. Polysaccharide und Lipide stellen ebenfalls wichtige Makromoleküle dar, werden hier jedoch nicht behandelt.

Alle Makromoleküle sind hierarchisch aufgebaut, denn sie bestehen aus Monomeren, also kleineren Molekülen. Die Monomere werden über eine Kondensation unter Abspaltung von Wasser zu immer größeren Einheiten zusammengefügt: Nucleotide werden zu Nucleinsäuren kondensiert, Aminosäuren zu Peptiden und Proteinen. Solche Makromoleküle können durch Hydrolyse wieder in kleinere Einheiten, bis hin zu den Monomeren zerlegt werden.

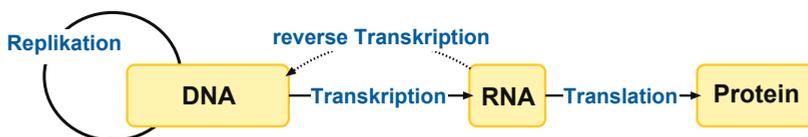
Heute ist allgemein akzeptiert, dass Ribonucleinsäuren (RNA) die ältesten Biomoleküle darstellen. Einerseits können sie in ihrer Sequenz genetische Informationen speichern, wie dies in RNA-Viren der Fall ist, andererseits sind sie auf vielfältige Weise enzymatisch aktiv, wie in Ribosomen oder

Spliceosomen. Später in der Evolution haben sich andere Biomoleküle auf jeweils eine der beiden Funktionen spezialisiert. Die Desoxyribonucleinsäure (DNA) ist chemisch viel stabiler und fungiert als Speicher der genetischen Information, Proteine dienen der Strukturbildung (Strukturproteine), als Speicherproteine sowie als Katalysatoren (Enzyme) in der Zelle.

Prinzipiell verläuft der genetische Informationsfluss in eine Richtung: von der DNA über die mRNA als Zwischenstufe hin zum Protein. Dieser universelle Informationsfluss wird „Zentrales Dogma der Molekularbiologie“ genannt (→ Abb. 1.1). Die genetische Information wird durch vier verschiedene, stickstoffhaltige Basen in der DNA gespeichert: den Purinen Adenin (A) und Guanin (G) sowie den Pyrimidinen Cytosin (C) und Thymin (T). Das Umschreiben dieser genetischen Information von DNA auf mRNA heißt Transkription. Auch die dafür zuständige mRNA nutzt für die Informationsvermittlung vier verschiedene Basen, wobei statt Thymin hier Uracil (U) verwendet wird. Im zweiten Schritt wird die Information der mRNA in eine Proteinsequenz umgeschrieben, dieser Prozess wird Translation genannt. Bereits Anfang der 70er Jahre wurde eine Ausnahme dieses universellen Dogmas entdeckt, denn Retroviren sind in der Lage, ihre als RNA gespeicherte Information in DNA umzuschreiben. Daher wurde das Zentrale Dogma um diese Möglichkeit erweitert. Eine weitere zentrale Voraussetzung für das Leben soll hier nicht unerwähnt bleiben: die genetische Information muss regelmäßig dupliziert werden (→ Abb. 1.1). Diesen Prozess heißt Replikation und er ist eng verknüpft mit der Teilung der Zelle. Durch die vorherige Duplikation der

Abb. 1.1

Zentrales Dogma der Molekularbiologie. Durch die Replikation der DNA wird die genetische Information vervielfältigt und auf die Tochterzellen weitergegeben. Der genetische Informationsfluss läuft von der DNA über die RNA zum Protein. Durch die Entdeckung der Reversen Transkription in Retroviren wurde das Konzept entsprechend erweitert.



genetischen Information wird gewährleistet, dass beide Tochterzellen nach einer Zellteilung die gleiche genetische Information erhalten.

1.1 Der Aufbau von DNA und RNA

In allen Lebewesen besteht das genetische Material aus Nucleinsäuren. Diese stellen Polymere aus einer sehr kleinen Zahl unterschiedlicher Monomere dar, den Nucleotiden. Die Nucleotide wiederum bestehen aus drei verschiedenen Komponenten: einem Zucker und einem Phosphat, die das Rückgrat der Nucleinsäure bilden, sowie einer stickstoffhaltigen Base, die den eigentlichen Informationsträger darstellt (→ Abb. 1.2).

Die einzelnen Nucleotide sind über 5'→3'-Phosphodiesterbindungen miteinander verknüpft, wodurch das Molekül eine bestimmte Ausrichtung aufweist: vom freien 5'-Phosphatende hin zum

freien 3'-OH-Ende (→ Abb. 1.3). Alle DNA-spaltenden oder synthetisierenden Enzyme wirken an dieser Phosphodiesterbindung, d.h. DNA oder RNA-Moleküle werden zunächst an der Phosphodiesterbindung kondensiert (= ligiert, also zusammengefügt) oder hydrolysiert (gespalten).

Während Zucker und Phosphat für die strukturelle Integrität der DNA sorgen, ist die dritte Komponente, die stickstoffhaltigen Basen, für die Codierung der genetischen Information verantwortlich. In der DNA sind dies die beiden Purine Adenin (A) und Guanin (G) sowie die Pyrimidine Cytosin (C) und Thymin (T). Bei RNA wird Thymin durch Uracil (U), ebenfalls ein Pyrimidin, ersetzt.

DNA liegt in den meisten Organismen doppelsträngig vor, wobei die beiden Stränge durch Wasserstoffbrücken zwischen den beiden Strängen zusammengehalten werden. Es bilden immer A und T eine Paarung, die aus zwei Wasserstoffbrücken besteht, G und C bilden Paarungen mit drei Wasserstoffbrücken. Es erfolgt also immer eine Paarung zwischen einem Purin und einem Pyrimidin. Andere Paarungen sind in der DNA

Abb. 1.2

Die DNA ist aus drei verschiedenen Komponenten aufgebaut: einer von vier verschiedenen Basen (Adenin, Guanin, Cytosin oder Thymin), die an einen Zucker, (eine Desoxyribose) gebunden sind. Diese Kombination aus Base und Zucker wird als Desoxynucleosid bezeichnet. Abhängig von der Base ergibt sich so Desoxyadenosin, -cytidin, -guanosin oder -thymidin. Durch die Bindung eines Monophosphats an den Zucker entsteht ein Desoxynucleotid, welches allgemein als dNMP abgekürzt wird. Für die verschiedenen Basen ergibt sich daher Desoxyadenosinmonophosphat, (dAMP) etc. Freie, nicht eingebaute Nucleotide liegen als energiereiche Triphosphate vor, welche allgemein als dNTP und je nach Base z. B. Desoxyguanosintriphosphat (dGTP) genannt werden. Beim Einbau werden diese dNTPs dephosphoryliert, sodass Monophosphate (dNMP) entstehen. Im Gegensatz zur DNA trägt RNA einen anderen Zuckeranteil, eine Ribose. Daher fällt bei RNA immer die Vorsilbe „Desoxy“ weg (z. B. Cytosinmonophosphat, CMP).

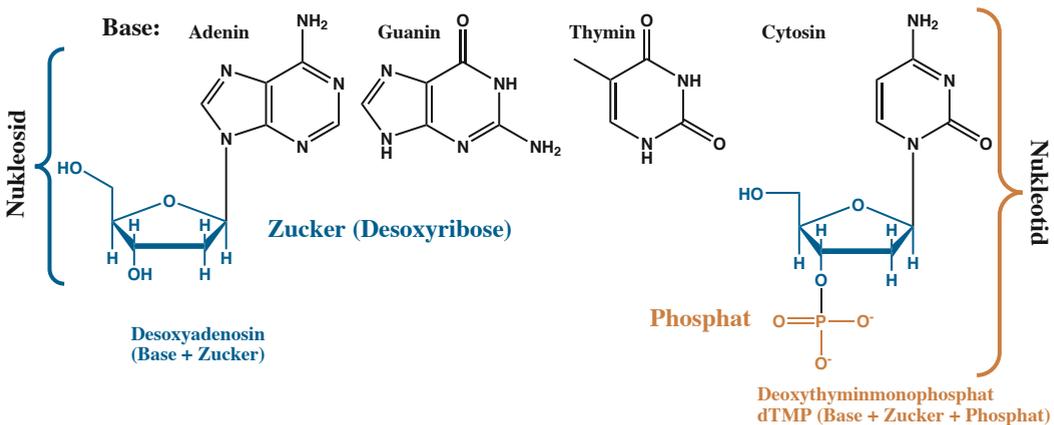
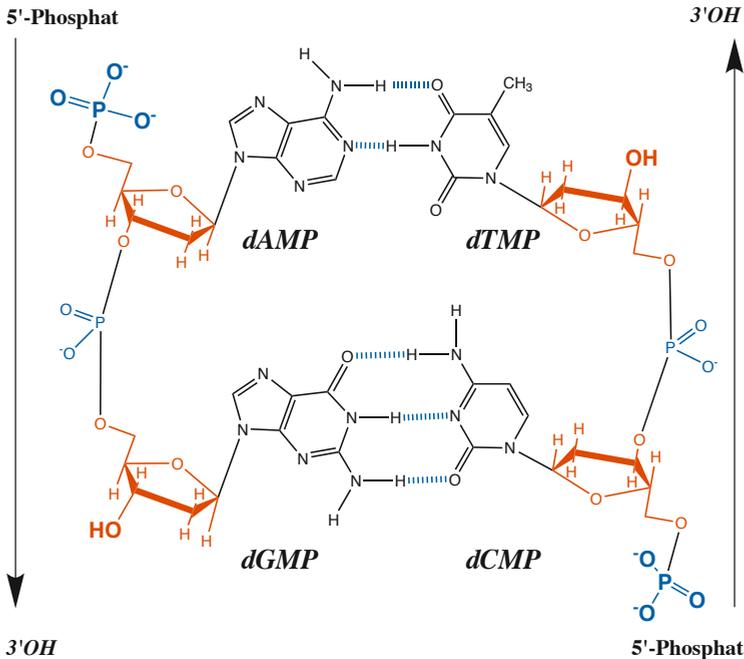


Abb. 1.3

Vier verschiedene Monomere bauen die DNA auf. Jedes dieser Monomere besteht aus einer Desoxyribose (rot), einem Phosphat (Blau) und einer der vier stickstoffhaltigen Basen: dAMP, dCMP, dGMP und dTMP. Innerhalb eines Stranges sind die Monomere über ein Zucker-Phosphat-Rückgrat miteinander verbunden. Das Phosphat ist immer an das C5-Atom des Zuckers gebunden. Bei der Polymerisation wird eine kovalente Bindung zum C3-Atom des nächsten Nukleotids gebildet. Dadurch erhält der DNA Strang eine Richtung vom 5'-Phosphatende auf der einen Seite zum 3'-OH-Ende auf der anderen Seite. DNA liegt als antiparalleler Doppelstrang vor, wobei immer Paarungen zwischen den Basen Adenin und Thymin (Paarung oben) oder zwischen Guanin und Cytosin (Paarung unten) erfolgen. Zwischen dAMP und dTMP werden zwei Wasserstoffbrücken und zwischen dGMP und dCMP drei Wasserstoffbrücken gebildet.



nicht möglich. Durch diese Paarungen bildet DNA normalerweise eine rechtdrehende Doppelhelix, die sogenannte B-Form.

Da immer die gleichen Basen miteinander Wasserstoffbrücken ausbilden, ist die Information auf den beiden Strängen gegenläufig (antiparallel). Die beiden DNA-Stränge sind zueinander komplementär, wie an dieser kurzen DNA-Sequenz verdeutlicht werden kann:



Da T immer mit A paart bzw. C mit G, ergibt sich die Basenfolge des einen Stranges zwangsläufig aus der des anderen Stranges. Daher bleibt die Information auch nach der Trennung der beiden Stränge erhalten (→ Abb. 1.4).

! TIPP

Doppelsträngige DNA wird mit dsDNA [*ds* = *double stranded*], einzelsträngige als ssDNA [*ss* = *single stranded*] abgekürzt. Diese Abkürzungen werden analog auch für RNA (dsRNA und ssRNA) verwendet.

GUT ZU WISSEN

► Unübliche DNA-Strukturen

DNA kann auch andere Strukturen ausbilden, wie die rechtdrehende A-Form oder die linksdrehende Z-Form. Diese sind jedoch selten.

Abb. 1.4

Doppelsträngige DNA besteht immer aus den Paaren A und T bzw. C und G und bildet eine rechtsdrehende Helix aus. In dieser Struktur bilden sich alternierend eine zugängliche, große Furche [Major Groove] sowie eine kleine Furche [Minor Groove] aus.

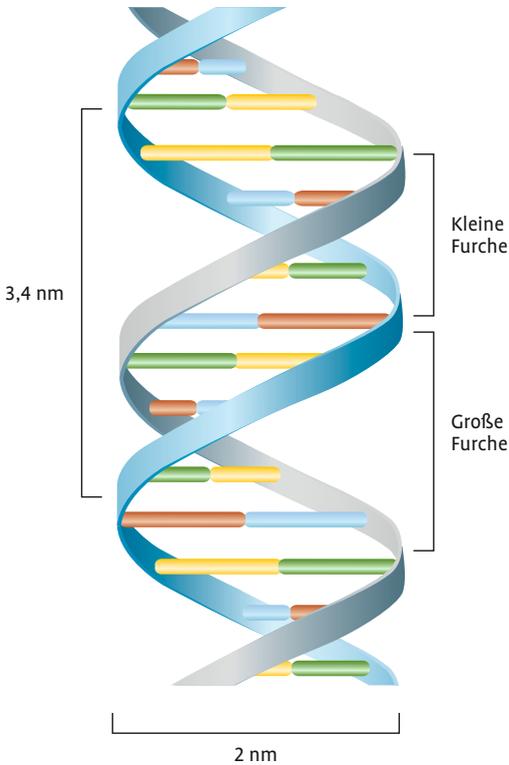
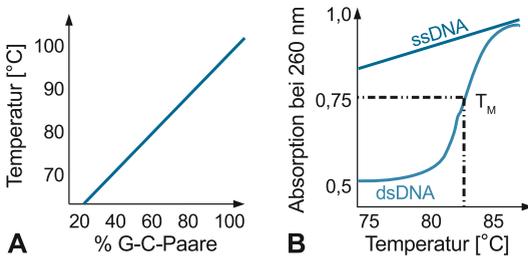


Abb. 1.5

Abhängig vom Gehalt an GC-Paarungen steigt die Schmelztemperatur einer DNA an (A). Die Denaturierung von dsDNA kann photometrisch bei einer Wellenlänge $\lambda = 260$ nm verfolgt werden (B). Die Temperatur, bei der die Hälfte der DNA Moleküle einzelsträngig vorliegen, heißt Schmelztemperatur T_M . In der Abbildung liegt der T_M bei 83 °C.



Die dsDNA ist ein sehr stabiles Molekül, was ein-drucksvoll das Neandertaler-Genomprojekt belegt (www.eva.mpg.de/neandertal), dessen zu sequenzierende drei Milliarden Basenpaare aus DNA stammt, die ungefähr 35 000–70 000 Jahre alt ist.

Bei hohen Temperaturen werden die Wasserstoffbrücken zwischen den Basen aufgebrochen, was zu einer Trennung der DNA Stränge führt. Dieser Vorgang wird als Denaturierung der DNA bezeichnet. DNA denaturiert bei Temperaturen zwischen 70–90 °C. Die Temperatur, bei der die Hälfte der DNA einzelsträngig vorliegt, wird Schmelztemperatur (T_M) genannt. Da die A-T-Paarung zwei Wasserstoffbrücken ausbildet und C-G drei, ist die Schmelztemperatur nicht nur von der Länge der DNA Moleküle abhängig, sondern auch vom A-T- bzw. G-C-Gehalt. Die DNA-Denaturierung kann photometrisch (\rightarrow Kapitel 3.8) nachverfolgt (\rightarrow Abb. 1.5) werden, denn ssDNA absorbiert das Licht stärker als dsDNA. Dieses Verhalten wird Hypochromer Effekt genannt und in \rightarrow Kapitel 3.8.1 eingehender behandelt.

Die Denaturierung der DNA ist reversibel, was bedeutet, dass sich bei Abkühlung wieder die beiden DNA-Stränge zusammenlagern. Dieser Vorgang der Renaturierung ist stark abhängig von der Länge der beiden DNA-Moleküle. Bei kurzen DNA Molekülen von einigen Tausend Basen läuft dieser Vorgang innerhalb von Minuten recht schnell und effizient ab, für DNA Stränge von über 10^4 – 10^5 Basen werden Stunden bis Tage benötigt.

Der Aufbau der RNA ist dem der DNA sehr ähnlich. Durch die Ribose im Rückgrat ist die Struktur der RNA flexibler als die der DNA. Weiterhin findet sich in der RNA statt Thymin das Pyrimidin Uracil (U) (\rightarrow Abb. 1.6).

Eigentlich ist RNA einzelsträngig, sie tendiert aber dazu, ebenfalls doppelsträngige Auffaltungen durchzuführen. Diese doppelsträngigen Bereiche werden nur über relativ kurze Sequenzabschnitte ausgebildet, weil der antiparallele Partnerstrang fehlt. Diese dsRNA-Regionen werden gebildet, wenn zwei Abschnitte auf dem gleichen Molekül oder auf zwei verschiedenen RNA-Molekülen zueinander komplementär sind. Sie haben eine enorme biologische Bedeutung, denn im Gegensatz zur DNA kann RNA durch ihre flexible

Struktur auch enzymatische Funktionen übernehmen¹.

Es gibt drei wichtige Arten an RNAs:

- Die ribosomale RNA (rRNA) in den Ribosomen macht ungefähr 85 % der gesamten RNA einer Zelle aus und stellt vor allem die enzymatische Komponente der Ribosomen dar.
- Die Transfer-RNA (tRNA) kommt auf ungefähr 10 % der Gesamt-RNA und ist ebenfalls an der Translation beteiligt. Sie vermittelt zwischen den Codons der mRNA und den zugehörigen Aminosäuren. Die tRNA besitzt einige seltene Basen und ungewöhnliche Basenpaarungen. Sie bildet sehr charakteristische Strukturen aus.
- Die Messenger-RNA (mRNA) überträgt die Sequenzinformation von der DNA auf den Translationsapparat und macht weniger als 5 % der Gesamt-RNA aus.

1.2 Der genetische Code

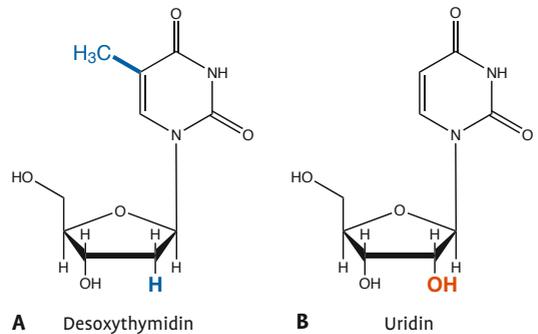
Da den vier Basen der DNA bzw. mRNA zwanzig verschiedene Aminosäuren der Proteine gegenüberstehen, bilden immer drei aufeinanderfolgende Basen eine Informationseinheit für eine bestimmte Aminosäure. Die DNA-Abfolge ATC wird bei der Transkription in AUC umgeschrieben und anschließend in die Aminosäure Isoleucin translatiert.

Durch die Verwendung von Triplets können maximal 64 verschiedene Kombinationen entstehen ($4 \times 4 \times 4$ Varianten = 64), genug um die zwanzig Aminosäuren zu codieren. Daher können einige Aminosäuren durch verschiedene Triplets, die sogenannten Codons, repräsentiert werden (→ Abb. 1.7). Oft unterscheiden sich Codons, die für die gleiche Aminosäure codieren, in der dritten Base. Beispielsweise codieren GCA, GCC, GCG und GCT alle für die Aminosäure Alanin (→ Abb. 1.7). Man spricht in diesem Zusammenhang vom degenerierten Code, bei dem die dritte Base eines

1 Katalytisch aktive RNAs werden als Ribozyme bezeichnet und sind in der Natur weit verbreitet, z. B. in Ribosomen und Spliceosomen.

Abb. 1.6

Die DNA enthält 2-Desoxyribose (A), während RNA aus Ribose (B) besteht. Die Base Thymin (A) ist typisch für DNA und wird in RNA durch die Base Uracil (B) ersetzt.



! TIPP

Da immer drei Basen eine Aminosäure repräsentieren, kann aus der Länge einer DNA-Sequenz auf die ungefähre Größe des zugehörigen Proteins geschlossen werden: Eine DNA-Sequenz von 900 Basenpaaren (bp) codiert für ein Protein aus 300 Aminosäuren. Als Faustregel kann für eine Aminosäure ein durchschnittliches Molekulargewicht von 110 Da angenommen werden. Die 900 bp DNA codieren also für ein Protein aus 300 Aminosäuren mit ungefähr 33 kDa. Dies ist nur eine grobe Abschätzung, denn einzelne Aminosäuren weichen in ihrem Molekulargewicht stark ab, außerdem sind Proteinmodifikationen wie Glykosylierungen nicht berücksichtigt.

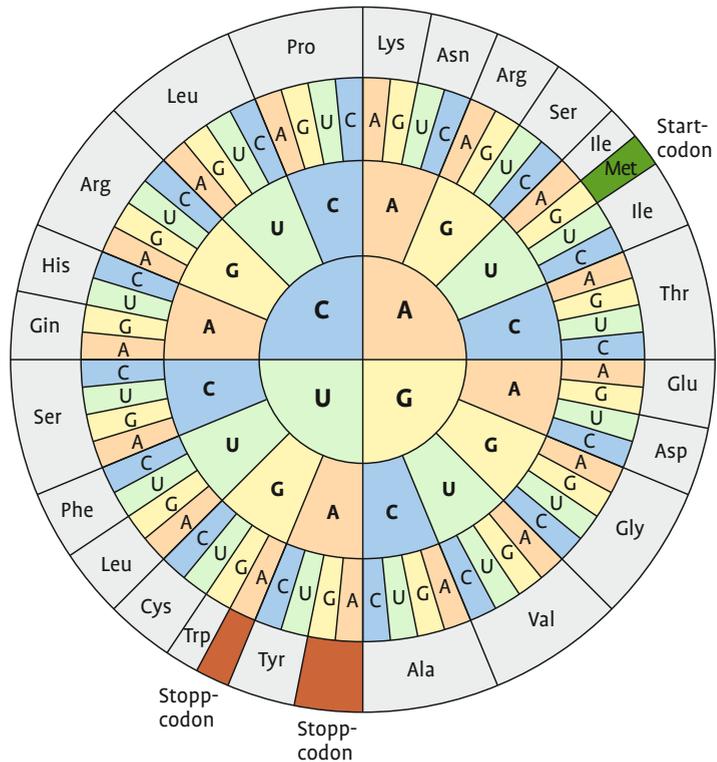
GUT ZU WISSEN

► Universelle biologische Prinzipien

Die Grundbausteine, welche DNA, RNA oder Proteine aufbauen, sind in allen Lebewesen gleich. Auch der genetische Code ist – bis auf wenige Ausnahmen – universell. Dies gilt auch für zelluläre Prozesse wie Transkription, Translation oder dem primären Stoffwechsel.

Abb. 1.7

Codon Sonne. Da RNA statt Thymin die Base Uracil enthält, verwenden Codon-Tabellen U statt T. Die Aminosäuren sind in ihrem 3-Buchstaben-code angegeben. Es ist zu beachten, dass manche Aminosäuren nur durch einen Code repräsentiert werden, z. B. Tryptophan (Trp: UGC) oder Methionin (Met: AUG). Letzteres dient auch als Startcodon für die Translation. Codons ohne zugehörige Aminosäure dienen als Stoppcodon. Die Codon-Sonne wird von innen nach außen gelesen.



Codons variieren kann („wobbeln“ [to wobble = wackeln, schwanken]). Leider nutzten die verschiedenen Spezies die unterschiedlichen Codons einer Aminosäure verschieden häufig. Die Codon-Nutzung [Codon Usage] variiert i. d. R. umso deutlicher, je weiter zwei Arten evolutionär voneinander entfernt sind. Soll ein Gen aus einem Säuger oder einer Pflanze in *E. coli* produziert werden, muss berücksichtigt werden, dass *E. coli* einige Codons, beispielsweise für die Aminosäure Arginin oder Prolin, nur selten nutzt, da es nur wenig t-RNAs für diese Codons hat. Demzufolge können Gene aus Eukaryoten, die diese Codon verwenden, nicht ohne Weiteres für die heterologe Produktion des Proteins in *E. coli* verwendet werden. Drei Codons codieren nicht für eine Aminosäure. Diese Codons dienen als Stoppcodons, denn bei ihnen wird die Translation beendet.

Von diesen allgemeinen Prinzipien gibt es nur wenige Ausnahmen. Beispielsweise verläuft der

Informationsfluss bei Retroviren von der RNA zur DNA. Einige Mitochondrien und wenige niedere Eukaryoten² nutzen abweichende Codons für bestimmte Aminosäuren. Einige Proteine enthalten besondere Aminosäuren.

Diese universellen Prinzipien ermöglichen erst die molekularbiologischen Arbeitstechniken, wie wir sie heute kennen. Ohne diese könnte kein Protein in einem artfremden Organismus hergestellt werden. Dies ist auch ein deutlicher Hinweis darauf, dass das Leben auf dieser Erde nur ein einziges Mal entstanden ist.

2 Sowohl „Eukaryont“ als auch „Eukaryot“ sind mögliche Schreibweisen und beide seit den 70er-Jahren gebräuchlich. Während früher die erste Variante häufiger anzutreffen war, setzt sich inzwischen die Schreibweise ohne „n“ durch, denn im Englischen heißt es immer *eucaryot* bzw. *procaryot*.

1.3 Die Gene

Die genetische Information eines Organismus ist in bestimmte DNA-Bereiche organisiert, den Genen. Dies sind die Abschnitte einer Nukleinsäure, welche die genetische Information für ein funktionales Polypeptid oder ein RNA-Molekül (mRNA, rRNA, tRNA) tragen³.

Die DNA-Sequenz wird sowohl von den RNA-Polymerasen als auch von den DNA-Polymerasen von 5' nach 3' des DNA-codierenden (sense-) Strangs bzw. von 3' nach 5' des Matrizen- oder (antisense-) Strangs gelesen. Die RNA-Polymerasen sind für die Erstellung der RNAs verantwortlich, die DNA-Polymerasen für die Replikation der DNA. Wie wir bereits erfahren haben, verläuft die Umsetzung der genetischen Information, die in der DNA-Sequenz gespeichert ist, hin zu der Aminosäuresequenz der Proteine über ein zweiteiliges Verfahren (Transkription und Translation, → Abb. 1.1). Es ergibt sich somit folgender Zusammenhang zwischen DNA-Sequenz und Proteinsequenz:

- Das 5' Ende der Nukleinsäure entspricht dem Aminoende des Proteins.
- Das 3' Ende der Nukleinsäure entspricht dem Carboxyende des Proteins.

Da die DNA aus zwei gegenläufigen, komplementären Strängen besteht (→ Abb. 1.4), die RNA-Polymerase aber nur einen dieser Stränge als Vorlage für die Synthese der mRNA verwendet, ist die Sequenz des einen DNA-Stranges komplementär zur davon abgeleiteten mRNA. Der andere DNA Strang weist genau die gleiche Sequenz wie die mRNA auf. Leider ist die Nomenklatur dieser beiden Stränge nicht einheitlich, da die obigen Begriffe von einigen Autoren auch anders definiert werden. Dieses Buch folgt den Empfehlungen der International Union of Biochemistry. In der Proteinsequenz symbolisiert N das Amino- und C das Carboxyende des Proteins.

Neben der eigentlichen proteindefinierenden Sequenz gibt es eine Reihe weiterer DNA Abschnit-

3 Es gibt auch die Situation, dass eine bestimmte Sequenz von zwei Genen genutzt wird. Gene können auch nicht-codierte Sequenzen enthalten. Diese Szenarien sind hier nicht wichtig.

! TIPP

Leserichtung und Strangbezeichnungen:

DNA codierender Strang:

5'-ATG CCA GTA GGC CAC TTG TCA-3'
(= sense Strang; = + Strang)

DNA Template Strang:

3'-TAC GGT CAT CCG GTG AAC AGT-5'
(= codogener Strang; = antisense Strang; = - Strang)

mRNA:

5'-AUG CCA GUA GGC CAC UUG UCA-3'

Protein:

N - Met - Pro - Val - Gly - His - Leu - Ser - C

te, die vor allem der Regulation der Genexpression dienen. Die Organisation der Gene und der regulierenden DNA-Bereiche ist in Prokaryoten und Eukaryoten signifikant verschieden.

1.3.1 Die Genstruktur in Prokaryoten

In Prokaryoten sind die Gene in Gruppen unter der Kontrolle eines Promotors⁴ [*Promoter*] zusammengefasst. Eine solche Gengruppe heißt Operon und die Produkte der gemeinsam regulierten Gene sind meistens auch im gleichen Stoffwechselweg involviert. Bekannte Beispiele sind das bakterielle Lac-Operon oder das Tryptophan-Operon.

Ist das Operon aktiviert bzw. nicht reprimiert⁵, bindet die RNA-Polymerase am Promotor. Alle Gene eines Operons werden in eine mRNA umgeschrieben (→ Abb. 1.8). Die Transkription endet durch eine spezielle Sequenz, dem Terminator, welche zu einer Schlaufenbildung in der mRNA führt. Dadurch fällt die Polymerase ab⁶.

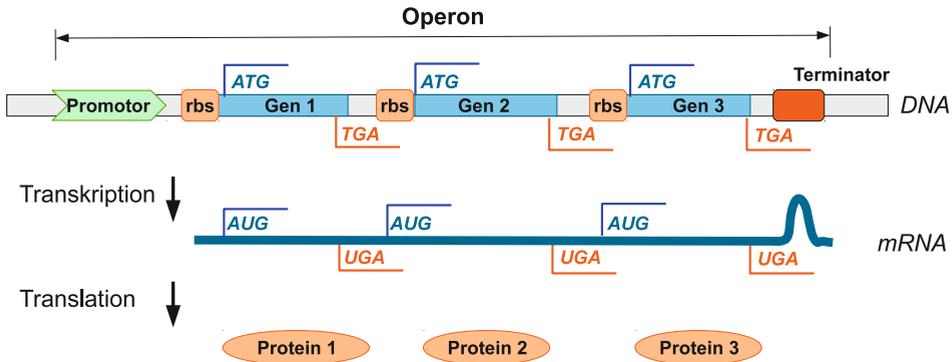
4 Ein Promotor ist eine in 5'-Richtung vom Gen lokalisierte DNA-Sequenz, welche für die Stelle des Beginns der Transkription verantwortlich ist. Sie kann die Menge an gebildeter mRNA beeinflussen.

5 Die Expression eines Operons wird durch die Bindung eines Aktivators an den Operator aktiviert (bzw. durch Bindung eines Repressors abgeschaltet). Diese sind in → Abb. 1.8 nicht weiter berücksichtigt. Das Gen des Repressors bzw. Aktivators liegt außerhalb des regulierten Operons.

6 Es gibt alternative Terminationsverfahren in Prokaryoten.

Abb. 1.8

Aufbau eines prokaryotischen Operons. Die Gene eines Operons werden gemeinsam durch einen Promotor reguliert und bilden eine polycistronische mRNA. Die Gene stehen oft auch in einem funktionalen Zusammenhang: Sie sind beispielsweise im gleichen Stoffwechselweg involviert. Am Ende des Operons entsteht in der transkribierten mRNA eine Schlaufe, die zur Abtrennung der RNA-Polymerase führt. Jedem Gen ist eine ribosomale Bindestelle (rbs) vorgeschaltet, die durch die Bindung an ein Ribosom erst die Translation ermöglicht. Das dabei entstehende Peptid beginnt am ersten AUG nach der rbs. Sobald ein Stopcodon (hier: UGA) in der Sequenz auftritt, endet die Translation und somit das Peptid.



Die nachfolgende Translation beginnt mit der Bindung der mRNA an die kleine Untereinheit des Ribosoms. Diese Bindung wird durch die sogenannte Shine-Dalgarno-Sequenz oder ribosomale Bindestelle (rbs), die komplementär zu einer Sequenz in den Ribosomen ist, am 5'-Ende der mRNA vermittelt. Die Proteinsynthese beginnt am ersten AUG, welches dem Transkriptionsstart folgt und endet am ersten Stopcodon. Es folgen weitere Gruppen, sogenannte Cistrons, aus rbs, Startcodon, zu translatierender Sequenz und Stopcodon. Da Bakterien mRNAs bilden, die mehrere solcher Cistrons besitzen, sind bakterielle mRNAs polycistronisch.

Sobald die mRNA an die kleine ribosomale Untereinheit gebunden ist, bindet auch die große Untereinheit. Ribosomen sind sehr komplex aufgebaute Strukturen aus drei verschiedenen rRNAs (welche nach ihrer Größe 5S, 16S und 23S genannt werden⁷) und 52 ribosomalen Proteinen. Die benötigten Aminosäuren werden durch tRNAs, an deren Ende sie gebunden sind, geliefert. Die tRNAs besitzen eine Kleeblattstruktur, an deren mittlerem Blatt das Anticodon aus drei Basen liegt.

Dieses Anticodon definiert, welche Aminosäure an der t-RNA gebunden ist.

Die Ribosomen dirigieren die mRNA und die tRNA so, dass es zu einer Paarung zwischen einem Codon der mRNA und dem zugehörigen Anticodon der tRNA kommt. Durch die Vermittlung der tRNA wird so eine bestimmte Aminosäure, die mit dem Codon auf der mRNA korrespondiert, in Position gebracht. Beispielsweise paart das Codon AGC der mRNA (→ Abb. 1.7) mit der tRNA, welche das Anticodon UCG besitzt und demzufolge mit der Aminosäure Serin beladen ist. Die Zuordnung von Codon bzw. Anticodon zu bestimmten Aminosäuren ist in (fast) allen Lebewesen gleich (→ Kapitel 1.2).

Bei der Proteinsynthese befinden sich immer zwei tRNAs am Ribosom. Nachdem das Codon der mRNA mit dem passenden Anticodon der tRNA gepaart wurde, werden die beiden Aminosäuren miteinander verknüpft. Das Ribosom rutscht auf der mRNA ein Codon weiter. Es folgt die Paarung zwischen Codon und Anticodon und anschließend die Verknüpfung der Aminosäure an die Peptidkette. Dieser Prozess wiederholt sich bis zum Stopcodon. Auf diese Weise wird die mRNA in 5'-3' Richtung in das zugehörige Protein umgeschrieben, welches dabei vom Aminoende (N-Ter-

⁷ Das „S“ steht für Svedberg-Einheit, welche durch Dichtenzentrifugation ermittelt wird (→ Kapitel 2.7).

Tab. 1.1
Gengrößen und Anzahl der Introns einiger menschlicher Gene.

Gen	Gengröße [kbp]	Größe der mRNA [kbp]	Zahl der Introns
Insulin	1,7	0,4	2
Kollagen	38	5	50
Albumin	25	2,1	14
Dystrophin	2000	17	50

minus) zum Carboxyende (C-Terminus) synthetisiert wird.

1.3.2 Genstruktur in Eukaryoten

Während bei Prokaryoten die Genstruktur und Genregulation noch relativ übersichtlich ist, sind bei Eukaryoten eine Vielzahl von Transkriptionsfaktoren und regulierenden Gensequenzen auf der DNA an der Steuerung der Expression beteiligt. Dazu kommen drei verschiedene RNA-Polymerasen, von denen wir hier nur die RNA-Polymerase II behandeln.

Es gibt DNA-Abschnitte, deren Lage zum Gen in bestimmten Grenzen variabel ist, die sogenannten Enhancer, welche oft die Gewebespezifität der Expression vermitteln. Andere Elemente liegen nahe am Gen und werden als Promotor bezeichnet. Sie dirigieren die RNA-Polymerase II zusammen mit einer großen Zahl weiterer genereller und spezifischer Transkriptionsfaktoren an die richtige Position. Spezifische Transkriptionsfaktoren sind nur bei der Transkription bestimmter Gene beteiligt, die generellen bei (fast) jeder Transkription. Solche generellen Faktoren binden beispielsweise an die CAAT-Box (deren Sequenz ähnlich dieser ist: 5'-GGCCAAATC-3'), andere an die TATA-Box (mit einer Sequenz ähnlich 5'-TATAAAA-3'). Die TATA-Box definiert den Transkriptionsstartpunkt, der die Koordinate +1 zugewiesen bekommt⁸. Erst wenn alle benötigten Transkriptionsfaktoren im Bereich des Promotors gebunden sind, kann die RNA-Polymerase II mit

der Synthese der hnRNA⁹ beginnen. Die hnRNA stellt ein Abbild der DNA-Sequenz dar, doch wird diese, noch während die RNA-Synthese läuft, bis zur mRNA prozessiert, die dann den Zellkern verlässt.

Am 5' Ende der hnRNA wird ein methyliertes Guanin angehängt, das sogenannte ^{m7}G-Cap. Dieses stabilisiert einerseits die RNA, dient aber auch als Bindestelle für den eIF4 Komplex, der später noch für die Initiation der Translation wichtig wird.

Eine weitere Prozessierung stellt das sog. Spleißen [*Splicing*] dar, bei dem bestimmte Bereiche der gerade entstandenen hnRNA wieder herausgeschnitten werden. Die herausgeschnittenen Bereiche heißen Introns. Diejenigen Bereiche des Gens, die noch in der reifen RNA vorhanden sind, heißen Exons. In niederen Eukaryoten sind Introns eher selten, in höheren Eukaryoten stellen sie die Regel dar und können einen großen Anteil am Gen einnehmen (→ Tab. 1.1).

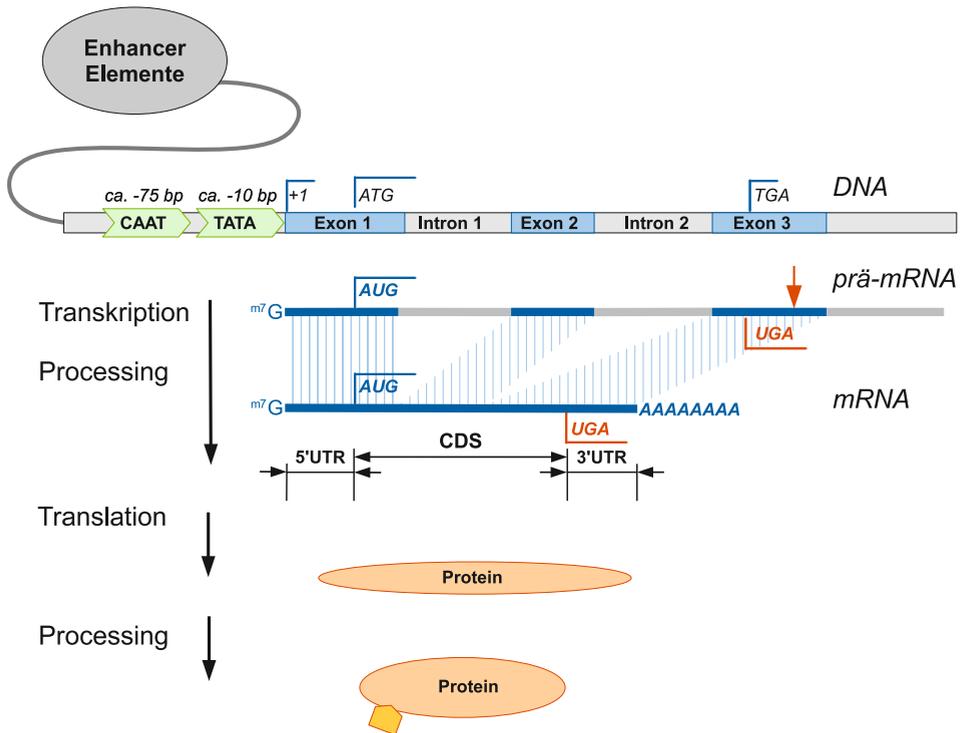
Das Spleißen, welches nur bei Eukaryoten vorkommt, hat eine wichtige Konsequenz für den Molekularbiologen. Um ein eukaryotisches Protein in Bakterien heterolog zu exprimieren, muss die Sequenzinformation der mRNA intronfrei in das Bakterium eingebracht werden, da die Bakterien die Introns der eukaryotischen, genomischen DNA nicht entfernen können. Um intronfreie DNA zu erhalten, wird eukaryotische mRNA verwendet, die durch das Enzym Reverse Transkriptase (→ Kapitel 4.5.5) in DNA umgeschrieben wird (→ Abb. 1.1). Das entstandene

⁸ Es gibt auch Gene ohne TATA- oder CAAT-Boxen.

⁹ hnRNA steht für heterogene nukleäre RNA und ist die Form der RNA, welche durch die RNA-Polymerase II in Eukaryoten synthetisiert, aber schnell zu mRNA prozessiert wird.

Abb. 1.9

Genstruktur und Expression bei Eukaryoten (RNA-Polymerase II Gene). Neben einer viel komplizierteren Regulation durch Enhancer und verschiedener Promotorelemente sowie einer Vielzahl an Transkriptionsfaktoren verläuft die Bildung der mRNA über verschiedene Zwischenschritte. Die durch die RNA-Polymerase II gebildete hnRNA wird sofort umfassend prozessiert. Neben dem m^7 G-Cap am 5'-Ende sorgt das Polyadenylierungssignal (roter Pfeil) dafür, dass ca. 200 A's, der sogenannte PolyA⁺-Strang, am 3'-Ende angehängt wird. Ein wesentlicher Unterschied zu Prokaryoten liegt in der Intron-Exon-Struktur der Gene. Die Intronbereiche werden aus der hnRNA ausgeschnitten. Erst die vollständig prozessierte mRNA verlässt den Zellkern. Nur der mittlere Bereich der mRNA, die CDS, wird in ein Protein translatiert. Dieser Bereich liegt zwischen dem Startcodon AUG und dem Stopcodon (hier: UGA). An beiden Enden gibt es untranslatierte Bereiche, die UTR's [UnTranslated Regions]. Auch das Protein kann weiter modifiziert werden, am bekanntesten ist die Glykosylierung von Proteinen (gelbes Fünfeck).



Abbild der mRNA wird cDNA [*copy-DNA*] genannt.

Eine direkte Terminationssequenz, wie bei Prokaryoten, gibt es in Eukaryoten nicht. Erreicht die RNA-Polymerase das Polyadenylierungssignal, wird sie wahrscheinlich von einem Komplex, der für die Polyadenylierung von der DNA zuständig ist, abgedrängt. Bei der Polyadenylierung werden ca. 20–30 Basen nach dem Polyadenylierungssignal (meist 5'-AAUAAA-3') eine Sequenz von 100–200 A's (Adenin) angehängt (→ Abb. 1.9).

Alle Schritte der RNA-Prozessierung finden im Zellkern statt. Erst die reife mRNA verlässt diesen und wird in das Zytoplasma transportiert, wo die Translation stattfindet.

Im Zytoplasma erfolgt die Translation ähnlich wie oben für die Prokaryoten beschrieben, allerdings um ein Vielfaches komplexer. Beispielsweise erfolgt die Bindung der mRNA an die kleine ribosomale Untereinheit nicht direkt, sondern wird durch den Proteinkomplex eIF4F am m^7 G-Cap des 5' Endes vermittelt. Das eukaryotische Ribosom besteht aus vier rRNAs (5S, 5,8S, 18S und 28S)

sowie 88 ribosomalen Proteinen. Auch der polyA-Strang besitzt eine wichtige Funktion bei der Translation. Dort bindet PABP 1 (polyA-bindendes Protein 1). Durch eine Interaktion mit eIF4F entsteht so ein ringförmiges Gebilde aus mRNA, eIF4F und PABP1. Erst dieses kann sehr effizient translatiert werden.

Nicht die gesamte mRNA wird in Protein translatiert. Die Translation beginnt beim ersten Startcodon auf der mRNA, meist ein AUG (welches für die Aminosäure Methionin codiert) und sie endet beim Stopcodon (→ Abb. 1.9). Es soll an dieser Stelle betont werden, dass sich Start- und Stopcodon exklusiv auf die Translation beziehen und nichts mit der Transkription zu tun haben!

Der Bereich, welcher in ein Protein translatiert wird, heißt CDS [*Coding DNA Sequence*]. Die Bereiche außerhalb heißen 5'-UTR und 3'-UTR [UTR=*UnTranslated Region*]. Sie haben wichtige Funktionen für die Regulation und die Stabilität der mRNA.

Der Vollständigkeit halber sei erwähnt, dass der Aufbau und die Regulation der rRNA und tRNA-Gene von dem hier beschriebenen System abweichen. Auch die dort involvierten RNA-Polymerasen sind andere.

GUT ZU WISSEN

► Alternatives Spleißen

Eine mRNA kann an unterschiedlichen Stellen gespleißt werden. Der Vorgang des Alternativen Spleißens ist in höheren Eukaryoten ein gängiger Prozess (22–59% der menschlichen Gene sollen alternativ gespleißt werden). Aus einer DNA-Sequenz können so verschiedene mRNAs und demzufolge verschiedene Proteine entstehen. Welche Spleißvariante der mRNA gebildet wird, ist oft gewebeabhängig.

1.4 Proteine

Proteine sind die biologischen Makromoleküle, welche in der Natur am häufigsten vorkommen. Im Gegensatz zu den Nukleinsäuren sind sie in Sequenz, Struktur und Funktion äußerst vielfältig. Sie werden aus 20 verschiedenen Monomeren, den Aminosäuren, aufgebaut. Aminosäuren sind in ihrer Grundstruktur sehr ähnlich: ein zentrales C-Atom (das C α -Atom) wird von einer Aminogruppe (NH $_2$) auf der einen Seite und einer organischen Säure, der Carboxylgruppe (COOH), auf der anderen Seite eingefasst. Die verschiedenen Aminosäuren unterscheiden sich durch die Seitenkette (→ Abb. 1.10). Wegen der vier unterschiedlichen Gruppen am C α -Atom sind Aminosäuren asymmetrisch (=chiral), denn ihre Seitenketten können im Raum verschieden angeordnet sein¹⁰, weshalb die D-Form von der L-Form unterschieden wird. Allerdings kommen in Lebewesen fast ausschließlich Aminosäuren der L-Form vor.

Aminosäuren können in wässrigen Lösungen, je nach pH-Wert, als Kation oder Anion vorliegen. Sie sind Ampholyte, die bei niedrigem pH (Protonenüberschuss) eine H $_3^+$ N-Gruppe tragen und positiv geladen sind. Bei Protonenmangel tragen sie eine negativ geladene Carboxylgruppe COO $^-$. An ihrem isoelektrischen Punkt ist eine Aminosäure nach außen hin ungeladen, sie wird zum Zwitterion. Dabei ist zu beachten, dass die Seitengruppen ebenfalls Amino- und Carboxygruppen tragen und die Ladung einer Aminosäure maßgeblich beeinflussen (→ Abb. 1.11).

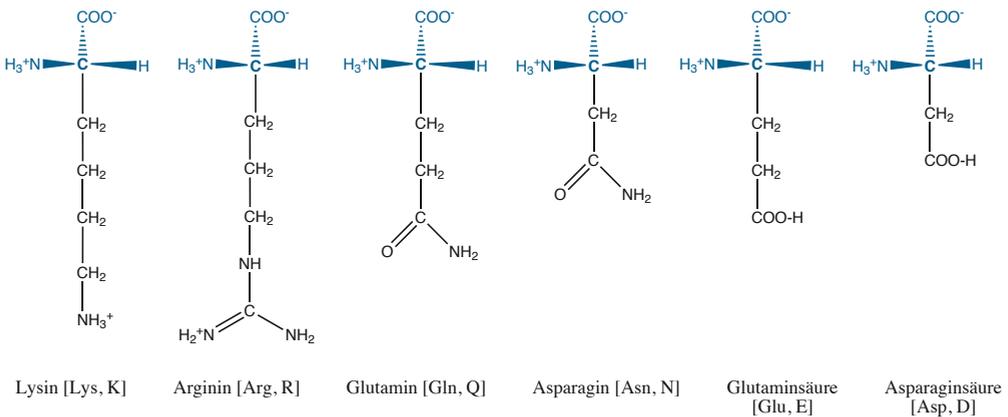
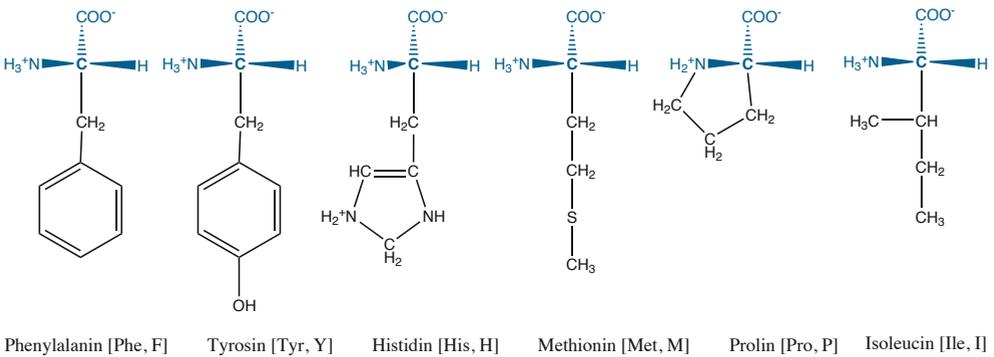
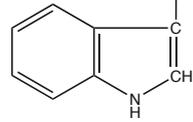
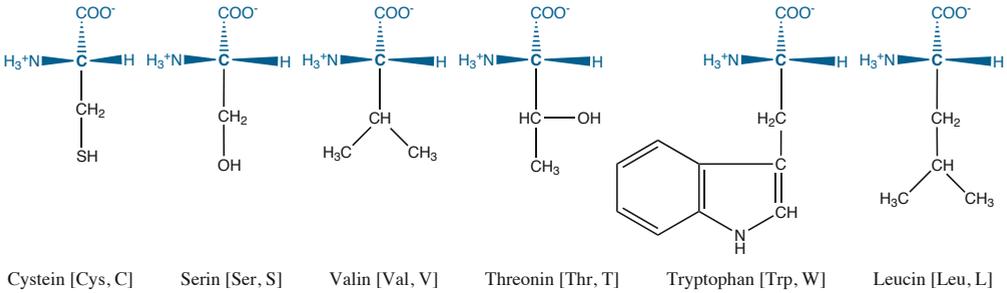
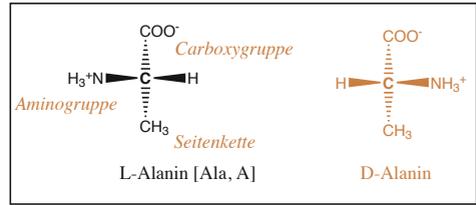
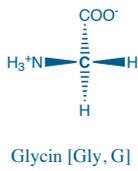
Aufgrund der unterschiedlichen Seitenketten können die Aminosäuren in verschiedene Gruppen mit ähnlichen chemischen Eigenschaften zusammengefasst werden. Zu den basischen Aminosäuren gehören Lysin, Arginin und Histidin. Aminosäuren mit sauren Seitenketten sind Asparaginsäure und Glutaminsäure.

Durch die Translation werden die einzelnen Aminosäuren zu Peptiden verbunden. Dabei entsteht zwischen der Carboxygruppe der einen und der Aminogruppe der anderen Aminosäure unter

10 Die einzige Ausnahme stellt das achirale, symmetrische Glycin dar.

Abb. 1.10

Alle proteinogenen Aminosäuren außer Glycin besitzen ein asymmetrisch substituiertes C α -Atom als chirales Zentrum, hier am Beispiel von L-Alanin und D-Alanin dargestellt.



Abspaltung von H_2O eine Peptidbindung, die recht starr ist und das Rückgrat des Proteins darstellt. Das Peptidrückgrat [Backbone] kann über Wasserstoffbrückenbildung zwischen Amino- und Carboxylgruppe eine α -Helix, ein β -Faltblatt [β -sheet] oder eine Schleife [Loop] ausbilden, das sind die bekannten Sekundärstrukturen. Die Ausbildung

von Sekundärstrukturen ist immer lokal begrenzt, was bedeutet, dass ein Protein verschiedene Sekundärstrukturelemente enthält (\rightarrow Abb. 1.12). Die nächsthöhere strukturelle Ebene ist die Tertiärstruktur (\rightarrow Abb. 1.13), die sich aus Interaktionen zwischen den Seitenketten des Peptids ergibt. Sie stellt in der Regel die funktionale Form

Abb. 1.11

Die Ladung einer Aminosäure hängt vom pH-Wert der umgebenden Lösung ab. An ihrem isoelektrischen Punkt (pI) liegt die Aminosäure ungeladen vor – man spricht vom Zwitterion. Ist der pH < pI ist die Aminosäure positiv geladen, bei pH > pI ist sie negativ geladen.

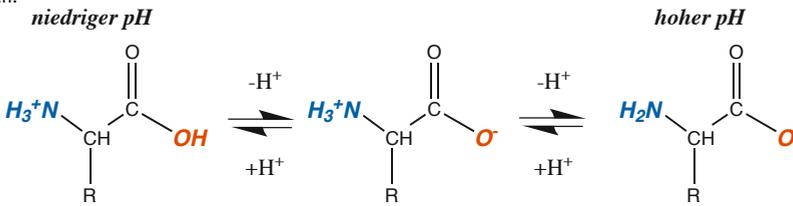


Abb. 1.12

Die Sekundärstruktur eines Proteins ergibt sich aus der Interaktion von Carboxy- und Aminogruppe im Rückgrat. Die abgebildete Interaktion in (A) führt zur Ausbildung einer α -Helix, die in (B) zur Ausbildung eines β -Faltblattes.

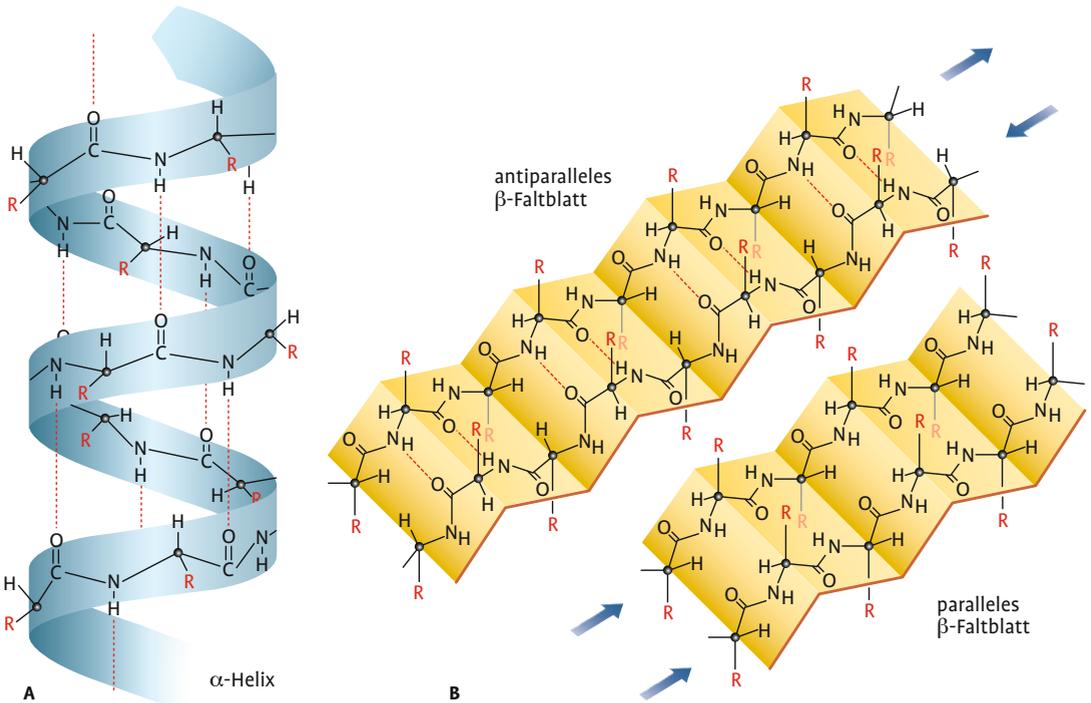


Abb. 1.13

Ein Protein (hier: die Hämophore HasAp von *Pseudomonas aeruginosa*) enthält meistens unterschiedliche Sekundärstrukturen wie α -Helix (rot), β -Faltblatt (gelb) und Schleifen (grün). Die Tertiärstruktur eines Proteins beruht auf Interaktionen zwischen den Seitenketten der Aminosäuren und kann auch weitere Nicht-Proteingruppen (hier: Hämgruppe (rosa)) enthalten. Mehrere Peptide, die in einer Tertiärstruktur vorliegen, können sich zu einem multimeren Protein zusammenschließen. Hier sind die beiden Untereinheiten, die die Quartärstruktur bilden, deutlich zu erkennen.

Quelle der abgebildeten Struktur:
Alontaga *et al.*: Structural characterization of the hemophore HasAp from *Pseudomonas aeruginosa*: NMR spectroscopy reveals protein-protein interactions between Holo-HasAp and hemoglobin. *Biochem.* 2009, 48(1):96–109.



eines Proteins dar. Die Interaktionen zwischen den Seitenketten können entweder auf Wasserstoffbrückenbildung oder ionischen Wechselwirkungen beruhen oder sogar kovalenter Natur sein, wie bei der Ausbildung von Disulfidbrücken zwischen zwei Cysteinen. Auch van der Waals-Kräfte beeinflussen die Tertiärstruktur. Da die zugrunde liegenden Bindungen unterschiedlich stark sind, ist die Stabilität der Tertiärstruktur eines Proteins sehr variabel. Der Verlust der Tertiärstruktur¹¹ führt zum Verlust der biologischen Funktion eines Proteins, es wird denaturiert und in seine Primärstruktur überführt. Die Denaturierung zerstört zwar die Struktur eines Proteins, führt jedoch nicht zu einer Spaltung der Peptidkette. Die Denaturierung kann durch Hitze oder bestimmte Chemikalien erfolgen. Je nach Protein und denaturierendem Agens kann die Denaturierung irreversibel oder reversibel sein. Ein gekochtes Ei wird nicht mehr flüssig, hier liegt eine irreversible Denaturierung vor. Die RNase A

kann mit 8 M Harnstoff und 2-Mercaptoethanol denaturiert werden. Werden die Chemikalien durch Dialyse (\rightarrow Kapitel 2.10) entfernt, faltet sich die RNase A selbstständig wieder in ihre biologisch aktive Tertiärstruktur. Hier liegt daher eine reversible Denaturierung vor.

Die höchste strukturelle Organisationsform von Proteinen stellt die Quartärstruktur dar, die durch die Interaktion verschiedener Peptidketten in einem multimeren Protein entsteht. Die zugrundeliegenden Bindungskräfte sind die gleichen, die zur Ausbildung der Tertiärstruktur führen.

Proteine mit einer mehr oder weniger kugelförmigen Tertiär- oder Quartärstruktur werden als globuläre Proteine bezeichnet. Die meisten Enzyme gehören in diese Gruppe. Oft sind die Aminosäuren aufgrund ihrer Hydrophobizität unterschiedlich im Protein verteilt: Innen befinden sich bevorzugt hydrophobe (wasserfeindliche) Aminosäuren, während sich polare Aminosäuren eher auf der Oberfläche befinden. Daher sind globuläre Proteine in der Regel gut wasserlöslich.

Proteine können auch faserige Strukturen ausbilden, meist finden sich solche bei Strukturproteinen. Eine weitere Gruppe stellen die Membran-

11 Mit dem Verlust der Tertiärstruktur geht i. d. R. auch der Verlust der auf Wasserstoffbrücken beruhenden Sekundärstruktur einher.