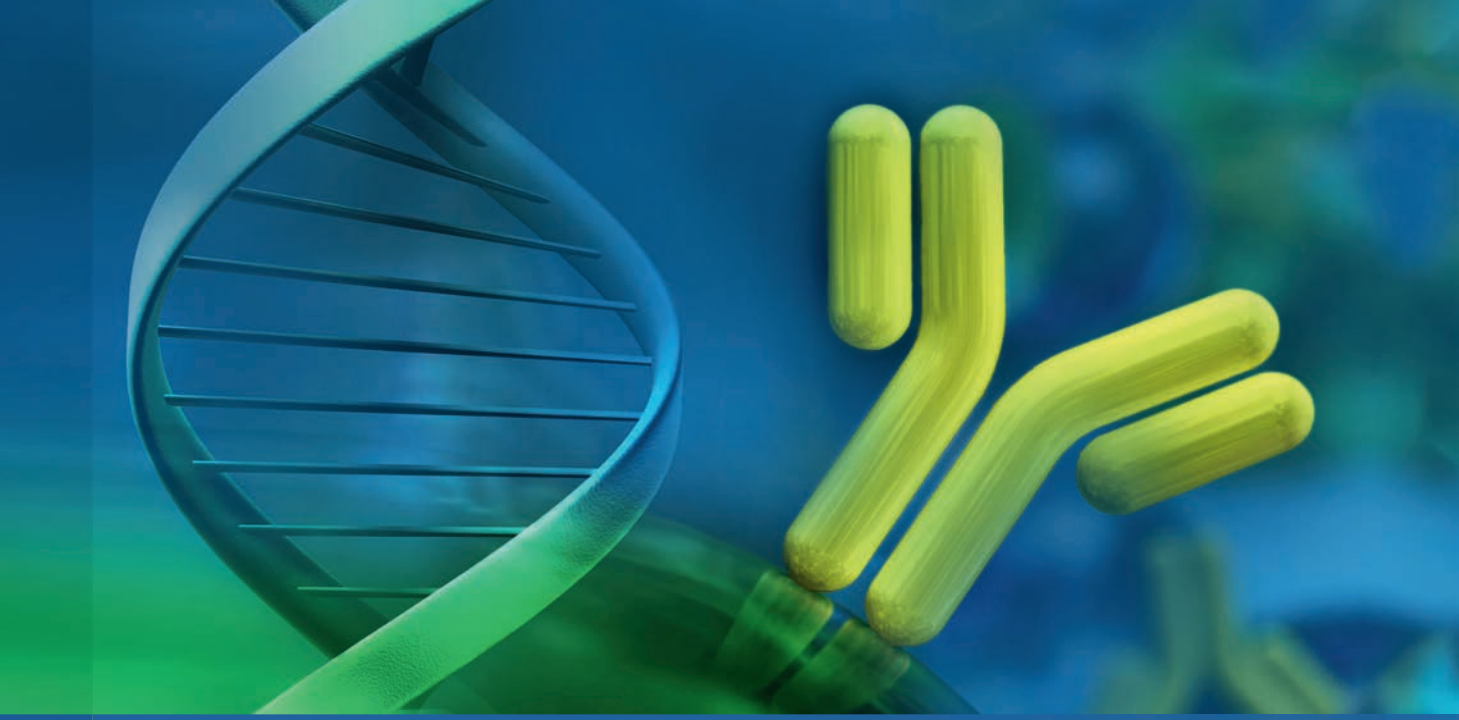


Dingermann / Winckler / Zündorf

Gentechnik Biotechnik

Grundlagen und Wirkstoffe

3. AUFLAGE

WVG
Wissenschaftliche Verlagsgesellschaft Stuttgart

Phosphatgruppe

chemisch modifiziertes RNA-Molekül

Protein, nicht glycosyliert

Protein, glycosyliert

Protein, chemisch modifiziert

Protein, glycosyliert, chemisch modifiziert

Protein mit Aminosäure-änderungen, nicht glycosyliert

Protein mit Aminosäure-änderungen, glycosyliert

Protein mit Aminosäure-änderungen, chemisch modifiziert

chemisch-synthetisches Peptid

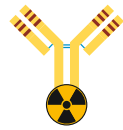
Fusionsprotein

Fusionsprotein eines Glycoproteins

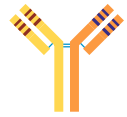
Protein, Zuckerkonjugat

Virus

Zelltransplantat



Maus-Antikörper, radioaktiv markiert



Maus/Ratte-Hybrid-Antikörper

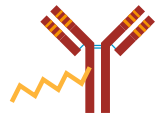
muriner, single-chain *Diabody*

chimärer Antikörper

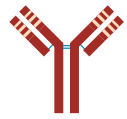
chimärer Antikörper, chemisch modifiziert

humanisierter Antikörper

bispezifischer humanisierter Antikörper



humanisierter Antikörper, chemisch modifiziert



humaner Antikörper

humanisiertes Fab-Fragment

humanisiertes Fab-Fragment, chemisch modifiziert

monomeres Antikörper-Fusionsprotein

dimeres Antikörper-Fusionsprotein

Dingermann / Winckler / Zündorf

Gentechnik Biotechnik

Dingermann / Winckler / Zündorf

Gentechnik Biotechnik

Grundlagen und Wirkstoffe

Theodor Dingermann, Frankfurt/M.
Thomas Winckler, Jena
Ilse Zündorf, Frankfurt/M.

Unter Mitarbeit von
Hanns-Christian Mahler, Basel

3., völlig neu bearbeitete Auflage

Mit 514 Abbildungen und 56 Tabellen

WVG
Wissenschaftliche
Verlagsgesellschaft
Stuttgart

Zuschriften an
lektorat@dav-medien.de

Anschriften der Autoren
Prof. Dr. Theodor Dingermann
Institut für Pharmazeutische Biologie
Johann Wolfgang Goethe-Universität
Max-von-Laue-Str. 9
60438 Frankfurt/Main

Dr. Ilse Zündorf
Institut für Pharmazeutische Biologie
Johann Wolfgang Goethe-Universität
Max-von-Laue-Str. 9
60438 Frankfurt/Main

Prof. Dr. Thomas Winckler
Friedrich-Schiller-Universität Jena
Institut für Pharmazie
Semmelweisstr. 10
07743 Jena

PD Dr. Hanns-Christian Mahler
Lonza Group Ltd
Drug Product Services
Hochbergerstr. 60A
CH-4057 Basel

Alle Angaben in diesem Werk wurden sorgfältig geprüft. Dennoch können die Autoren und der Verlag keine Gewähr für deren Richtigkeit übernehmen.

Ein Markenzeichen kann markenrechtlich geschützt sein, auch wenn ein Hinweis auf etwa bestehende Schutzrechte fehlt.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter https://portal.dnb.de abrufbar.

Jede Verwertung des Werkes außerhalb der Grenzen des Urheberrechtsgesetzes ist unzulässig und strafbar. Das gilt insbesondere für Übersetzungen, Nachdrucke, Mikroverfilmungen oder vergleichbare Verfahren sowie für die Speicherung in Datenverarbeitungsanlagen.

3. Auflage 2019
ISBN 978-3-8047-3708-2 (Print)
ISBN 978-3-8047-3961-1 (E-Book, PDF)

© 2019 Wissenschaftliche Verlagsgesellschaft mbH
Birkenwaldstraße 44, 70191 Stuttgart
www.wissenschaftliche-verlagsgesellschaft.de
Printed in Germany

Satz: abavo GmbH, Buchloe
Druck und Bindung: Firmengruppe Appl, Wemding
Umschlagabbildung: ustas/stock.adobe.com und JohanSwanepoel/stock.adobe.com
Umschlaggestaltung: deblik, Berlin
Grafiken: Dr. Ilse Zündorf
Indexer: Dr. Angelika Fallert-Müller, Dr. Beatrix Föllner, Walter Greulich (verantwortlich)

Vorwort

Nach 1999 und 2011 halten Sie nun die 3. Auflage des Lehrbuchs „Gentechnik – Biotechnik: Grundlagen und Wirkstoffe“ in den Händen. Wir sind auch in dieser 3. Auflage unserem Ursprungskonzept aus dem Jahr 1999 treu geblieben. „Gentechnik – Biotechnik: Grundlagen und Wirkstoffe“ bleibt auch in der 3. Auflage ein Lehrbuch und Nachschlagewerk. Und dennoch halten Sie ein völlig neues Buch und nicht etwa eine leicht ergänzte Version der 2. Auflage in Ihren Händen. Was hat uns umgetrieben?

Zum einen wollten wir noch deutlicher machen, dass der Fokus des Buches auf dem pharmazeutischen Wirkstoff liegt. Das stand nie außer Frage im ursprünglichen, zweiten Teil des Buches, der sich quasi exklusiv den zugelassenen rekombinanten Arzneimitteln widmet. Längst nicht so klar war das bisher im ersten Teil des Buches.

Der „erste Teil“ grenzt sich jetzt nicht mehr so deutlich von den Wirkstoffkapiteln ab, wie noch in der letzten Auflage des Buches und umfasst die Kapitel 1 bis 10. Dieser Teil, der sich generell mit den Methoden und Konzepten der Gentechnologie und der Biotechnologie befasst, konnte in den beiden ersten Auflagen noch so konzipiert werden, dass ein recht kompletter Überblick über das einschlägige Methodenspektrum gegeben wurde. Offensichtlich ist aber hier die Entwicklung nicht stehen geblieben: ganz im Gegenteil. Die methodische Weiterentwicklung verlief rasant, und was vor acht Jahren noch methodisch *state of the art* war, ist heute teilweise obsolet. Da war es nicht einfach zu entscheiden, was bleiben sollte und was verzichtbar war. An allem konnten wir nicht festhalten, da wichtige neue Methoden natürlich aufgenommen werden mussten und der Umfang des Buches im Rahmen bleiben sollte.

Unsere Lösung bestand darin, den Fokus der Methoden, die in diesem Buch beschrieben werden, noch mehr dahingehend zu schärfen, was zur Herstellung von Biologicals erforderlich ist. Der Leitgedanke bei der Konzeption des ersten Teils von „Gentechnik – Biotechnik: Grundlagen und Wirkstoffe“ war: „Vom Gen zum Wirkstoff“. Allerdings haben wir uns auch die Freiheit genommen, durchaus Methoden zu erklären, die sicherlich heute so nicht mehr zur Anwendung kommen. Beispielsweise beschreiben wir nach wie vor die konventionelle DNA-Sequenzierung nach Sanger (Kapitel 2), die noch angewendet wird, obwohl heute viel häufiger nach neueren Methoden im Hochdurchsatz maschinell sequenziert wird. Auch haben wir uns entschlossen, die klassische Identifizierung von krankheitsrelevanten Genen z.B. durch positionelles Klonieren oder Screenen von Genbanken zu beschreiben (Kapitel 9), obwohl diese Methoden durchaus bereits als historisch angesehen werden können. Allerdings wurden beispielsweise die Methoden zum molekularbiologischen Arbeiten mit dem Bakteriophagen Lambda gestrichen.

Klar ist jedoch, dass die Inhalte der ersten 10 Kapitel dieser dritten Auflage im Vergleich zu den ersten beiden Auflagen deutlich gestrafft wurden, nicht zuletzt auch, um die große Zahl neuer Wirkstoffe unterbringen zu können.

Deutlich gestrafft haben wir die Texte im Kapitel über die Fermentation und Produktreinigung (Kapitel 6). Hier haben wir versucht, den Fokus klar auf die Wirkstoffherstellung zu legen. Details zu dem sicherlich nicht unwichtigen Thema „Bioprozesstechnik“ müssen somit in entsprechend einschlägigen Lehrbüchern nachgelesen werden.

Stattdessen findet man wesentlich detailliertere Inhalte zu den Themen „Antikörper“ (Kapitel 5) und „Biosimilars“ (Kapitel 8) sowie zu den Themen „Glycosylierung“ bzw. „Glycoengeneering“ (z.B. bei Humanisierung von Hefen und Fucosylierung von Antikörpern in Kapitel 3, bezüglich Mannosephosphat-Rezeptoren in Kapitel 15). Neu aufgenommen wurde auch das Thema „Genomeditierung“.

Die meisten Anpassungen findet man im ursprünglichen 2. Teil des Buches, der die Kapitel 11 bis 21 einschließt.

Zum einen haben wir die Struktur der einzelnen Produktmonographien vereinheitlicht. Zum anderen wurden 21 Wirkstoffe entfernt, die nicht mehr verfügbar sind. Im Gegenzug wurden 119 Arzneimittel hinzugefügt, die während der vergangenen sieben Jahre neu zugelassen wurden.

Aktuell enthält die „Gentechnik – Biotechnik: Grundlagen und Wirkstoffe“ 232 Arzneimittel. Wirkstoffe, die für mehrere Indikationen zugelassen sind, werden nur einmal beschrieben. Ansonsten wird auf die jeweilige Monographie verwiesen. Dies betrifft 25 Einträge.

Sechs Arzneimittel wurden eingeschlossen, die nicht gentechnisch hergestellt werden. Vier Wirkstoffe sind aus unterschiedlichen Gründen als Infokästen aufgeführt.

Zwei Punkte, die auch schon für die 2. Auflage galten, wollen wir hier noch einmal hervorheben, um Missverständnisse zu vermeiden:

Mehr als jede andere Wissenschaftsdisziplin hat die Gentechnik ihre Wurzeln im anglo-amerikanischen Sprachraum. Das bedeutet nicht, dass nicht auch Wissenschaftler aus nicht-englischsprachigen Ländern bahnbrechende Beiträge geleistet hätten. Allerdings war die Sprache, mit deren Hilfe man sich verständigte, ausnahmslos Englisch. Daher bleibt es nicht aus, dass auch ein in deutscher Sprache abgefasstes Lehrbuch der Gen-

technik und Biotechnik eine Vielzahl von Anglizismen enthält. Wir haben uns bemüht, Anglizismen dann zu vermeiden, wenn es möglich ist, sie aber andererseits auch dort zu gebrauchen, wo der Sinn verstellt würde oder wo der Fachmann verständnislos den Kopf schütteln würde, wenn man eine deutsche Übersetzung gewählt hätte. In wissenschaftlichen Diskussionen in deutscher Sprache sind Ausdrücke wie *upstream*, *screenen*, *Leader*, *Enhancer* oder *inclusion bodies* absolut gebräuchlich. Derartige Ausdrücke wurden daher auch in diesem Buch verwendet. Sie wurden allerdings durch Kursivdruck als Anglizismen gekennzeichnet.

Bestimmte IUPAC-Regeln sind in der praktischen Wissenschaft auch noch nicht angekommen. So sollte man das Symbol „Da" für das Molekulargewicht nicht mehr nutzen, sondern stattdessen „U" verwenden. Da aber auch in Dokumenten, die für Wirkstoffe zulassungsrelevant sind immer noch an der alten Bezeichnung festgehalten wird, haben wir uns entschlossen, dies ebenfalls zu tun.

Ausdrücklich möchten wir betonen, dass dieses Buch nicht den Anspruch erhebt, ein pharmakologisches Lehrbuch zu sein. Vielmehr stellen wir das Molekül, die Konzeption seiner Herstellung, das Verständnis der relevanten Technologien, die damit verbundenen Chancen aber auch Gefahren und die Ansprüche an Qualität, Wirksamkeit und Verträglichkeit in den Mittelpunkt. All diese Punkte sollten gerade den Arzneimittelfachmann ansprechen, wie wir meinen. Leider werden wichtige Informationen in dieser Richtung von den Firmen immer stärker zurückgehalten.

Und schließlich soll nicht unerwähnt bleiben, dass wir uns ernste Gedanken darüber gemacht haben, Literaturzitate mit aufzunehmen. Wir haben uns – wie bei den beiden letzten Auflagen – gegen diese Option entschieden. Es ist heute sehr leicht möglich, über eine schnelle Recherche im Internet unendlich viel vertiefende Literatur zu einem Thema zu finden, auf das man in diesem Buch stößt und für das man sich noch tiefer informieren will. Wir fordern Sie auf, genau dies zu tun und bitten um Nachsicht, dass wir davon abgesehen haben, Ihnen eine sicherlich subjektive Auswahl an vertiefender Literatur zu präsentieren.

Wir hoffen dennoch, eine interessante Lektüre zusammengestellt zu haben, nicht nur für diejenigen, die im Rahmen ihrer Ausbildung vieles von dem, was hier beschrieben ist, lernen müssen, sondern auch für diejenigen, die selbst die Initiative ergreifen wollen, das genauer verstehen zu lernen, was es zum Zeitpunkt ihrer Ausbildung noch gar nicht gab.

Frankfurt/M. und Jena im Herbst 2018

Theodor Dingermann
Thomas Winckler
Ilse Zündorf

Inhalt

Vorwort ... V

Abkürzungsverzeichnis ... XV

1 Die bio(techno)logische Revolution ... 1

1.1 Von der klassischen Genetik zur Gentechnik ... 1

1.2 Gentechnik und Biotechnik ... 6

2 DNA-Rekombinationstechnologie ... 9

2.1 Die DNA ... 9
2.1.1 Aufbau der DNA ... 9
2.1.2 Denaturierung von DNA ... 11
2.1.3 Größe und Topologie der DNA ... 12
2.1.4 DNA-Analytik ... 14

2.2 Chemische DNA-Synthese ... 18

2.3 Enzymatische DNA-Synthese ... 20
2.3.1 DNA-Polymerasen ... 20
2.3.2 Polymerase-Kettenreaktion (PCR) ... 20
2.3.3 DNA-Sequenzierung ... 28

2.4 Prinzipien der Gentechnik ... 37
2.4.1 Restriktionsendonukleasen ... 38
2.4.2 Transformationsmethoden für Bakterien ... 41
2.4.3 Vektoren für die Transformation von Bakterien ... 41
2.4.4 Rekombinieren von DNA-Fragmenten ... 46

3 Grundsätze der Proteinexpression ... 51

3.1 Umsetzung genetischer Information in Proteine ... 51

3.2 Genexpression in Prokaryonten ... 53
3.2.1 Transkription ... 53
3.2.2 Translation ... 55

3.3 Genexpression in Eukaryonten ... 62

3.4 Posttranslationale Modifikationen an Proteinen ... 65
3.4.1 Proteolytische Prozessierung von Proteinen ... 65
3.4.2 Glycosylierung von Proteinen ... 68
3.4.3 Glycoengineering ... 73

4 Wirt-Vektor-Systeme zur Protein-Herstellung ... 81

4.1 Rekombinante Expressionseinheiten ... 81
4.1.1 Informationen über Gene und Proteine: Wissenschaftliche Datenbanken ... 81
4.1.2 Isolierung genetischer Information ... 82
4.1.3 Auswahl geeigneter Produktionszellen ... 84

4.2 Proteinherstellung in Bakterien ... 89
4.2.1 Steuerung der Transkription von Transgenen ... 89
4.2.2 Optimierung der Translation von Transgenen ... 93
4.2.3 Sekretion und N-Glycosylierung von Proteinen durch Bakterien ... 94
4.2.4 Fusionssysteme zur effizienten Produktreinigung ... 94
4.2.5 Plasmidstabilisierung in Produktionsstämmen ... 96

4.3 Proteinherstellung in Hefen ... 98
4.3.1 Vektorsysteme für die Proteinherstellung in Hefen ... 99
4.3.2 Transformation von Hefezellen ... 101
4.3.3 Heterologe Expression von Proteinen in Hefen ... 102

4.4 Proteinherstellung in Insekten-Zellkulturen ... 105
4.4.1 Baculovirus als Gentransfer-System ... 106
4.4.2 Transfektion von Insekten-Zellkulturzellen ... 108

4.5 Transgene Pflanzen ... 108
4.5.1 Transformationsmethoden ... 109
4.5.2 Selektionsmarker zur Etablierung transgener Pflanzen ... 113
4.5.3 Regeneration transgener Pflanzen ... 114
4.5.4 Verhinderung der Ausbreitung transgener Pflanzen ... 114
4.5.5 Manipulation von Sekundärstoffmustern in Medizinalpflanzen ... 117
4.5.6 Produktion von Naturstoffen in Pflanzen-Zellkulturen ... 117
4.5.7 Proteinexpression in transgenen Pflanzen-Zellkulturen ... 118
4.5.8 Proteinexpression in transgenen Pflanzen ... 119

4.6 Proteinherstellung in Säugetier-Zellkulturen ... 121
4.6.1 Transiente versus stabile Expression ... 121
4.6.2 Methoden zum Einführen von DNA in Säugetierzellen ... 122
4.6.3 Selektionsmarker zur Identifizierung transfizierter Säugetierzellen ... 128
4.6.4 Regulation heterologer Genexpression in Säugetierzellen ... 128
4.6.5 Produktions-Zelllinien ... 131

4.7 Genetisch veränderte Säugetiere ... 134
4.7.1 Transgene Tiere ... 135
4.7.2 Knockout-Mäuse ... 136
4.7.3 Wirkstoffproduktion in transgenen Säugetieren ... 141

5 Monoklonale Antikörper ... 144

5.1 Struktur von Antikörpern ... 144

5.2 Hybridoma-Technologie ... 147
5.2.1 Immunisierung von Mäusen ... 147
5.2.2 Fusion von Milzzellen mit Myelomzellen ... 148
5.2.3 Selektion der fusionierten Hybridomzellen ... 148
5.2.4 Etablierung der antikörperproduzierenden Hybridomzellen ... 151

5.3 Humanisierte Antikörper ... 153

5.4 Vollständig humane Antikörper ... 154
5.4.1 Phage-Display-Technologie ... 155
5.4.2 HuCal®-System ... 157
5.4.3 Transgene Mäuse zur Herstellung humaner Antikörper ... 160

5.5 Antikörper-Fragmente ... 162

5.6 Bispezifische Antikörper ... 163

5.7 Namensgebung bei Antikörper-Wirkstoffen ... 165

5.8 Fusionsproteine mit Fc-Teilen von Antikörpern ... 167

6 Biotechnische Produktion von Proteinwirkstoffen ... 170

6.1 Bioreaktoren ... 170

6.2 Fermentation ... 173
6.2.1 Zellkulturmedien ... 173
6.2.2 Wachstumskinetik und Produktbildung ... 174
6.2.3 Fermentationsverfahren ... 177

6.3 Produkt-Aufarbeitung ... 178
6.3.1 Zellernte ... 179
6.3.2 Zellaufschluss ... 180
6.3.3 Produkt-Reinigung ... 181

7 Vom biotechnologischen Wirkstoff zum Arzneimittel ... 192

7.1 Anforderungen der Arzneibücher an sterile Arzneimittel mit biotechnologischen Wirkstoffen ... 192

7.2 Instabilitätsreaktionen und Formulierung biotechnologischer Arzneistoffe ... 193
7.2.1 Instabilitätsreaktionen ... 193
7.2.2 Formulierung und Darreichungsform ... 195
7.2.3 Produktstabilität ... 196

7.3 Besonderheiten biotechnologischer Wirkstoffe bei Herstellung, Transport, Lagerung und Anwendung ... 198
7.3.1 Herstellung steriler Fertigprodukte von Biopharmazeutika ... 198
7.3.2 Transport ... 202
7.3.3 Lagerung ... 203
7.3.4 Anwendung ... 203

7.4 Analytische Untersuchungen am Fertigprodukt ... 205
7.4.1 Analytische Methoden ... 205
7.4.2 Laufzeit, Lagerbedingungen, Lagerhinweise ... 207

7.5 Die Anforderungen des Europäischen Arzneibuchs an gentechnisch hergestellte Produkte ... 207
7.5.1 Definition ... 207
7.5.2 Herstellung ... 207
7.5.3 Wirt-Vektor-System ... 208
7.5.4 Zellbanksysteme: Validierung und Kontrolle ... 209
7.5.5 Validierung des Herstellungsprozesses ... 210
7.5.6 Identitätsprüfung ... 211
7.5.7 Reinheitsprüfung ... 211
7.5.8 Gehalt und Wirksamkeit ... 212
7.5.9 Biologische Wirksamkeit ... 212
7.5.10 Struktur ... 212
7.5.11 Weitere Methoden ... 213

7.6 Die Europäische Zulassungsbehörde (EMA) ... 213
7.6.1 Aufgabe ... 213
7.6.2 Europäischer Öffentlicher Beurteilungsbericht ... 213
7.6.3 Der Orphan-Drug-Status ... 214

8 Biosimilars ... 216

8.1 Einführung ... 216

8.2 Definition ... 216

8.3 Zulassung eines Biosimilars ... 217

8.4 Die EMA als zentrales Kompetenzzentrum für Biopharmazeutika ... 218

8.5 Der Nachweis von Qualität, Wirksamkeit und Sicherheit ... 219

8.6 Eindeutige Identifizierung des Arzneimittels ... 220

8.7 Extrapolation der Indikationen ... 220

8.8 Derzeit in der EU zugelassene Biosimilars ... 221

8.9 Austausch von Biosimilars (Switching) ... 222

9 Gentransfer-Arzneimittel und Stammzelltherapie ... 224

9.1 Gentherapie ... 224
9.1.1 Einführung ... 224
9.1.2 Nichtvirale Gentransfer-Arzneimittel ... 228
9.1.3 Virale Gentransfer-Arzneimittel ... 229

9.2 Genreparatur ... 241
9.2.1 Zelluläre Reparatur von DNA-Doppelstrangbrüchen ... 241
9.2.2 Zinkfinger-Nukleasen und TALENs ... 243
9.2.3 CRISPR-Cas9-System ... 245
9.2.4 Anwendungen der Genom-Editierung mit CRISPR-Cas9 ... 246

9.3 Stammzelltherapie ... 250
9.3.1 Adulte humane Stammzellen ... 251
9.3.2 Humane embryonale Stammzellen ... 252
9.3.3 Somatischer Kerntransfer zum Klonen adulter Zellen ... 253
9.3.4 Induzierte pluripotente Stammzellen ... 256

10 Genomik, Genfunktionsanalyse und stratifizierte Pharmakotherapie ... 258

10.1 Genomik ... 258
10.1.1 Durchführung von Genomprojekten ... 259
10.1.2 Erkenntnisse aus Genomprojekten ... 261
10.1.3 Das Genom des Menschen ... 263
10.1.4 Das individuelle menschliche Genom ... 265

10.2 Gendiagnostische Methoden ... 267
10.2.1 Zytogenetische Diagnostik genomweiter Aberrationen ... 268
10.2.2 Molekulargenetische Diagnostik von Genmutationen ... 271

10.3 Identifizierung krankheitsrelevanter Gene ... 278
10.3.1 Vererbungsregeln ... 278
10.3.2 Positionsunabhängige Klonierung von Kandidatengenen ... 280
10.3.3 Positionelle Klonierung von Kandidatengenen ... 286
10.3.4 Genomweite Assoziationsstudien ... 294

10.4 Methoden zur Genfunktionsanalyse ... 295
10.4.1 Genetische Komplementation ... 295
10.4.2 Genexpressionsanalyse ... 296
10.4.3 Proteomik ... 304
10.4.4 Reportergene ... 306
10.4.5 Modellorganismen ... 312
10.4.6 Mausmodelle ... 313

10.5 Stratifizierte Pharmakotherapie ... 314
10.5.1 Arzneistoffmetabolisierende Enzyme ... 315
10.5.2 Phänotypisierung ... 316
10.5.3 Genotypisierung ... 318
10.5.4 Beispiele für pharmakogenetische Interaktionen ... 319

11 Anämie und Blutgerinnungsstörungen ... 325

11.1 Therapie der Anämie ... 325
11.1.1 Erythropoetin ... 325
11.1.2 Struktur des menschlichen Erythropoetins ... 327
11.1.3 Erythropoetin-Rezeptor ... 328
11.1.4 Erythropoetin-Arzneimittel ... 329

11.2 Therapie und Prophylaxe von Blutgerinnungsstörungen ... 334
11.2.1 Blutgerinnungskaskade ... 334
11.2.2 Blutgerinnungsfaktor VIII ... 338

11.2.3 Arzneimittel zur Substitution von FVIII 341
11.2.4 Blutgerinnungsfaktor IX 350
11.2.5 Arzneimittel zur Substitution von FIX 351
11.2.6 Blutgerinnungsfaktor VII 355
11.2.7 FVII-Arzneimittel 356
11.2.8 Blutgerinnungsfaktor XIII 356
11.2.9 Arzneimittel zur Substitution von FXIII 358

11.3 Therapie und Prophylaxe von Thrombosen 358
11.3.1 Regulation der Thrombozyten-aggregation 360
11.3.2 GPIIb/IIIa-Rezeptor 361
11.3.3 Thrombozytenaggregationshemmer 362
11.3.4 Hemmung der plasmatischen Blut-gerinnung 363
11.3.5 Antithrombin 363
11.3.6 Hirudin 364
11.3.7 Rekombinante Antikoagulanzien 365
11.3.8 Regulation der Fibrinolyse 367
11.3.9 Gewebe-Plasminogenaktivator 369
11.3.10 Rekombinante Fibrinolytika 371
11.3.11 Antidote oraler Antikoagulanzien 374

12 Allergische Entzündungs-reaktionen 377

12.1 Therapie des allergischen Asthmas 377
12.1.1 Asthma bronchiale 377
12.1.2 Allergisches Asthma 378
12.1.3 Senkung des IgE-Spiegels 380
12.1.4 Eosinophiles Asthma 381
12.1.5 Senkung des IL-5-Spiegels 382

12.2 Therapie der atopischen Dermatitis 384
12.2.1 Atopische Dermatitis 384
12.2.2 Neutralisierung von IL-4 und IL-13 384

13 Augenkrankheiten 386

13.1 Therapie der Altersabhängigen Makuladegeneration 386
13.1.1 Altersabhängige feuchte Makula-degeneration 386
13.1.2 Neutralisierung von VEGF 386

13.2 Therapie der vitreomakulären Traktion 392
13.2.1 Vitreomakuläre Traktion 392
13.2.2 Therapieoptionen 392

13.3 Therapie der Limbusstammzellen-insuffizienz 393
13.3.1 Limbales Stammzellreservoir 393
13.3.2 Therapie mit Limbusstammzellen 394

13.4 Therapie der neurotrophen Keratitis 395
13.4.1 Nervenwachstumsfaktor 395
13.4.2 Therapeutische Verwendung von NGF 396

14 Autoimmunerkrankungen 398

14.1 Therapie von rheumatischen Erkrankungen und Kollagenosen 398
14.1.1 Rheumatoide Arthritis 398
14.1.2 Juvenile idiopathische Arthritis 400
14.1.3 Ankylosierende Spondylitis 400
14.1.4 Psoriasis-Arthritis 401
14.1.5 Systemischer Lupus erythematodes 401
14.1.6 Tumornekrosefaktor alpha 401
14.1.7 Neutralisierung von TNF-α 403
14.1.8 Interleukin-1β 414
14.1.9 Neutralisierung von IL-1β 414
14.1.10 Interleukin-6 416
14.1.11 Neutralisierung von IL-6 416
14.1.10 Neutralisierung von IL-17A 419
14.1.13 Reduktion der T-Zell-Aktivierung 419
14.1.14 Depletion von B-Zellen 422
14.1.15 B-Zell-aktivierender Faktor (BLys) 424
14.1.16 Neutralisierung von BLys 424

14.2 Therapie der Plaque-Psoriasis 425
14.2.1 Neutralisierung von TNF-α 425
14.2.2 Interleukine IL-12 und IL-23 426
14.2.3 Neutralisierung von IL-12 und IL-23 427
14.2.4 Interleukin-17 429
14.2.5 Neutralisierung von IL-17A 430

14.3 Therapie entzündlicher Darm-erkrankungen 433
14.3.1 Neutralisierung von TNF-α 433
14.3.2 Neutralisierung von IL-12 und IL-23 434
14.3.3 Blockade von $\alpha_4\beta_7$-Integrin 434

14.4 Therapie der Multiplen Sklerose 436
14.4.1 Klinik 436
14.4.2 Immunopathogenese 439
14.4.3 Therapeutische Konzepte 440
14.4.4 Interferon beta 440
14.4.5 Glatirameracetat 445
14.4.6 Elimination aktivierter T-Zellen 447
14.4.7 Elimination von B- und T-Zellen 448
14.4.8 Blockade von α_4-Integrinen 450
14.4.9 Depletion von B-Zellen 453

14.5 Therapie der Immunthrombozytopenie 453
14.5.1 Pathophysiologie 454
14.5.2 Thrombopoetin 454

15 Erbkrankheiten ... 457

15.1 Therapie der Mukoviszidose ... 457
15.1.1 Krankheitsbild ... 457
15.1.2 Chloridkanal CFTR ... 458
15.1.3 Desoxyribonuklease I ... 459

15.2 Therapie lysosomaler Speicherkrankheiten ... 460
15.2.1 Das lysosomale Kompartiment ... 461
15.2.2 Mannose-6-Phosphat-Rezeptoren ... 461
15.2.3 β-Glucocerebrosidase ... 464
15.2.4 α-Galactosidase A ... 467
15.2.5 α-L-Iduronidase ... 469
15.2.6 Iduronat-2-Sulfatase ... 471
15.2.7 N-Acetylgalactosamin-4-Sulfatase ... 472
15.2.8 N-Acetylgalactosamin-6-Sulfatase ... 474
15.2.9 Saure α-Glucosidase ... 474
15.2.10 Saure Lipase A ... 475
15.2.11 Tripeptidyl-Peptidase 1 ... 476

15.3 Therapie Cryopyrin-assoziierter periodischer Syndrome (CAPS) ... 478
15.3.1 Familiäres kälteinduziertes autoinflammatorisches Syndrom (FCAS) ... 478
15.3.2 Muckle-Wells-Syndrom (MWS) ... 479
15.3.3 Neonatal-onset multisystem inflammatory disease (NOMID) ... 479
15.3.4 Bedeutung von IL-1 bei CAPS ... 479

15.4 Therapie der paroxysmalen nächtlichen Hämoglobinurie ... 481
15.4.1 Krankheitsbild ... 481
15.4.2 Komplementkaskade ... 482
15.4.3 Blockade des Komplementsystems ... 484

15.5 Therapie des hereditären Angioödems ... 485
15.5.1 C1-Inhibitor ... 485
15.5.2 Substitution des C1-Inhibitors ... 486

15.6 Therapie der Hypophosphatasie ... 487
15.6.1 Alkalische Phosphatase ... 487
15.6.2 Substitution der Alkalischen Phosphatase ... 487

15.7 Therapie der X-chromosomalen Hypophosphatämie ... 488

15.8 Therapie der schweren kombinierten Immundefizienz ... 489
15.8.1 Genetische Ursachen ... 489
15.8.2 Therapieoptionen ... 490

15.9 Therapie der spinalen Muskelatrophie ... 490
15.9.1 Molekularbiologische Grundlagen der SMA ... 491
15.9.2 Therapeutische Korrektur der SMN-Expression ... 493

15.10 Therapie der chronischen Granulomatose ... 494
15.10.1 Genetische Ursachen ... 494
15.10.2 Interferon gamma ... 494

15.11 Therapie der malignen Osteopetrose ... 496
15.11.1 Genetische Ursachen ... 496

16 Fertilisationsstörungen ... 497

16.1 Regulation des ovariellen Zyklus ... 497
16.1.1 Gonadotropin-Releasing-Hormon ... 499
16.1.2 Follikelstimulierendes Hormon ... 499
16.1.3 Luteinisierendes Hormon ... 500
16.1.4 Choriongonadotropin ... 501

16.2 Assistierte Reproduktion ... 502
16.2.1 Rekombinante Wirkstoffe für die hormonelle Hyperstimulation ... 504

17 Infektionskrankheiten ... 510

17.1 Prophylaxe und Therapie von Virus-Hepatitiden ... 511
17.1.1 Hepatitis-A-Virus ... 511
17.1.2 Hepatitis-B-Virus ... 512
17.1.3 Hepatitis-C-Virus ... 514
17.1.4 Impfstoffe gegen Hepatitis-B-Viren ... 515
17.1.5 Interferone ... 523
17.1.6 Interferon-Rezeptoren ... 523
17.1.7 Wirkungen von Interferonen ... 525
17.1.8 Alpha-Interferon ... 527
17.1.9 Interferon-alfa-2-Arzneimittel ... 527

17.2 Impfung gegen Gebärmutterhalskrebs ... 532
17.2.1 Humane Papillomviren ... 532
17.2.2 Impfstoffe gegen humane Papillomviren ... 533

17.3 Impfung gegen Influenza ... 537
17.3.1 Influenza-Viren ... 537
17.3.2 Impfstoffe gegen Influenza-Viren ... 538

17.4 Impfung gegen Respiratorisches Synzytial-Virus ... 540
17.4.1 Respiratorisches Synzytial-Virus ... 540
17.4.2 Antiinfektiva gegen RSV ... 542

17.5 Impfung gegen Pneumokokken ... 542
17.5.1 Streptococcus pneumoniae ... 542
17.5.2 Impfstoffe gegen Pneumokokken ... 543

17.6 Impfung gegen Meningokokken ... 545
17.6.1 Neisseria meningitidis ... 545
17.6.2 Impfstoffe gegen Meningokokken ... 546

17.7 Impfung gegen Cholera ... 547
17.7.1 Vibrio cholerae ... 547
17.7.2 Cholera-Toxin ... 548
17.7.3 Impfstoffe gegen Cholera-Bakterien ... 548

17.8 Impfung gegen Clostridium-difficile-Infektionen ... 549
17.8.1 Clostridium difficile ... 549
17.8.2 Clostridium-difficile-Antitoxin ... 550

18 Organtransplantation ... 552

18.1 Ursache von Gewebe-Unverträglichkeit ... 552

18.2 T-Zellen und Abstoßungsreaktion ... 554

18.3 Reduktion der Aktivität von T-Zellen ... 556
18.3.1 Depletion aktivierter T-Zellen ... 556
18.3.2 Reduktion der T-Zell-Aktivierung ... 559

19 Stoffwechselstörungen ... 561

19.1 Therapie des Diabetes mellitus ... 561
19.1.1 Insulin ... 562
19.1.2 Insulinrezeptor ... 563
19.1.3 Regulation der Insulinfreisetzung ... 564
19.1.4 Konsequenzen eines Insulinmangels ... 566
19.1.5 Tierisches Insulin als Grundlage von Insulin-Arzneimitteln ... 566
19.1.6 Biochemische „Humanisierung" von Schweineinsulin ... 566
19.1.7 Gentechnische Herstellung von Humaninsulin ... 567
19.1.8 Expression von Humaninsulin in Bakterien ... 568
19.1.9 Expression von Humaninsulin in Hefen ... 571
19.1.10 Gentechnische Herstellung von Humaninsulin-Analoga ... 573
19.1.11 Schnell wirkende Insulin-Analoga ... 574
19.1.12 Lang wirksame Humaninsulin-Analoga ... 577
19.1.13 Glucagon ... 580
19.1.14 Inkretine ... 582

19.2 Therapie des Kurzdarmsyndroms ... 588
19.2.1 Glucagon-like peptide-2 ... 589

19.3 Therapie von Lipidstoffwechselstörungen ... 590
19.3.1 Hypercholesterinämie ... 590
19.3.2 Proproteinkonvertase Subtilisin/Kexin Typ 9 ... 591
19.3.3 Lipoproteinlipase ... 593

19.4 Therapie der Osteoporose ... 595
19.4.1 Regulation der Mineralisierung des Knochens ... 595
19.4.2 Osteoporose ... 597
19.4.3 Parathormon ... 597
19.4.4 Calcitonin ... 600
19.4.5 RANK-Ligand ... 601

19.5 Knochenwachstumsfaktoren ... 602
19.5.1 Bone morphogenetic proteins (BMPs) ... 602
19.5.2 Rezeptoren und Aktivierung ... 603

19.6 Hypoparathyreoidismus ... 605
19.6.1 Parathormon ... 605

20 Tumorerkrankungen ... 607

20.1 Tumorentstehung ... 607
20.1.1 Onkogene ... 609
20.1.2 Tumorsuppressor-Gene ... 613
20.1.3 MicroRNA-Gene ... 614

20.2 Zytokine als Tumortherapeutika ... 615
20.2.1 Interleukin-2 ... 616
20.2.2 Tumornekrosefaktor-α ... 617
20.2.3 Interferon alfa-2a ... 618

20.3 Anti-Zytokin-Strategien ... 621
20.3.1 Castleman-Krankheit ... 621
20.3.2 Neutralisierung von IL-6 ... 621

20.4 Inhibition der Signaltransduktion ... 622
20.4.1 Zielstruktur HER1 ... 622
20.4.2 Zielstruktur HER2 ... 627
20.4.3 Zielstruktur RANK-Ligand ... 632
20.4.4 Zielstruktur PDGF-Rezeptor ... 632

20.5 Zellen markierende Wirkstoffe ... 633
20.5.1 Zielstruktur Tumormarker SLAMF7 ... 633
20.5.2 Zielstruktur Gangliosid GD2 ... 634
20.5.3 Zielstruktur CD20 ... 635
20.5.4 Zielstruktur CD22 ... 643
20.5.5 Zielstruktur CD30 ... 644
20.5.6 Zielstruktur CD38 ... 645
20.5.7 Zielstruktur CD52 ... 645

20.6 Inhibitoren der Nährstoffzufuhr ... 646
20.6.1 Prinzip der Angiogenese ... 646
20.6.2 Zielstruktur VEGF/VEGF-Rezeptor ... 646
20.6.3 VEGF-Blocker ... 648
20.6.4 Asparaginase ... 651

20.7 Immunonkologika ... 653
20.7.1 Bispezifische Antikörper ... 653
20.7.2 Immun-Checkpoint-Inhibitoren ... 654
20.7.3 Autologe CAR-T-Zellen ... 659
20.7.4 Allogene Donor-T-Zellen ... 661

20.8 Virale Onkolyse ... **662**

20.9 Adjuvante Therapeutika ... **663**
20.9.1 Erythropoetin ... 663
20.9.2 Granulozyten-Wachstumsfaktor ... 664
20.9.3 Urat-Oxidase ... 669

20.10 Diagnostikum nach einer Thyreoidektomie ... **671**
20.10.1 Schilddrüsenkarzinom ... 671
20.10.2 Thyreotropin (TSH) ... 672

21 Therapie von Wachstumsstörungen ... **674**

21.1 Ursachen des Minderwuchses ... **674**
21.1.1 Ullrich-Turner-Syndrom und SHOX-Mangel ... 674
21.1.2 Prader-Willi-Syndrom ... 676
21.1.3 Chronische Niereninsuffizienz bei Kindern ... 678
21.1.4 Vorgeburtliche Wachstumsverzögerung (SGA) ... 678
21.1.5 Primärer IGF-1-Mangel ... 679
21.1.6 Regulation der Somatropin-Freisetzung ... 679
21.1.7 Somatropin ... 680
21.1.8 Insulin-ähnlicher Wachstumsfaktor IGF-1 ... 688

21.2 Therapie der Akromegalie ... **689**
21.2.1 Krankheitsbild ... 689
21.2.2 Somatropin-Antagonisten ... 691

Übersicht der genannten Wirkstoffe ... **693**

Sachregister ... **699**

Die Autoren ... **725**

Abkürzungsverzeichnis

A

µM	mikromolar
A_{260}/A_{280}	Absorption bei 260 nm bzw. 280 nm
AAV	Adenoassoziiertes Virus
AcMNPV	*Autographa californica Multiple Nucleopolyhedrovirus*
ACR	*American College of Rheumatology*
ADA	Adenosin-Deaminase
ADCC	*antibody-dependent cell-mediated cytotoxicity*; antikörperabhängige zellvermittelte Zytotoxizität
ADCP	*antibody-dependent cell-mediated phagocytosis*; antikörperabhängige zellvermittelte Phagozytose
ADA	*anti-drug antibodies*
ALL	akute lymphatische Leukämie
AMD	altersabhängige feuchte Makuladegeneration
AMP	Adenosinmonophosphat
amp^R	Ampicillin-Resistenzgen (β-Lactamase)
AS04	Adjuvans aus hydratisierten Alumiumhydroxid und MPL
ASCT	autologe Stammzelltransplantation

B

BAC	*bacterial artificial chromosome*, bakterielles artifizielles Chromosom
BAFF	*B cell activating factor of the TNF family*
BDD-rFVIII	*B-domain deleted recombinant FVIII*
BLyS	*B lymphocyte stimulator* (auch BAFF genannt)
BFU-E	*erythroid-burst forming units*
BHK	*baby hamster kidney cells*
BiTE	*Bi-specific T cell Engager*
BMP	*bone morphogenetic protein*; Knochenwachstumsfaktor
bp	*base pairs*, Basenpaare in einem DNA-Molekül
Bq	Becquerel
BSA	bovines Serumalbumin
BSC	*best supportive care*; bestmögliche Begleitbehandlung

C

C1INH	C1-Esterase-Inhibitor
CAPS	Cryopyrin-assoziiertes periodisches Syndrom
CAR	*chimeric antigen receptor* (chimärer Antigenrezeptor)
CD	*cluster of differentiation*
CDC	*complement-dependent cytotoxicity*; komplementabhängige Zytotoxizität
CDK	Cyclin-abhängige Proteinkinase
cDNA	*complementary DNA*, auch *copy DNA*, DNA-Kopie einer mRNA
CDR	*complementarity determining region*, hypervariable Region eines Antikörpers
CF	Cystische Fibrose oder Mucoviszidose
CFTR	*cystic fibrosis transmembrane conductance regulator*
CFU-E	*erythroid colony forming unit*
CGD	chronische Granulomatose
CHOP	Chemotherapie mit Cyclophosphamid, Doxorubicin, Vincristin, Prednisolon
CHO	*chinese hamster ovary*
CIN	*cervical intraepithelial neoplasia*, intraepitheliale Neoplasie
CIS	*clinically isolated syndrome*; klinisch isoliertes Syndrom
CLL	chronische lymphatische Leukämie
CML	chronische myeloische Leukämie
CMP	Cytidinmonophosphat
CMV	Cytomegalievirus
CTLA-4	zytotoxisches T-Lymphozyten-Antigen-4
CXCL	*C-X-C motif chemokine receptor ligand*
CXCR	*C-X-C motif chemokine receptor*
CYP	Cytochrom-P450-Enzym

D

Da	Dalton
dAMP	Desoxyadenosin-5'-monophosphat
DAP I	Dipeptidylaminopeptidase I
dCMP	Desoxycytidin-5'-monophosphat
ddNTPs	2',3'-Didesoxynukleosidtriphosphate, Mischung der artifiziellen Nukleotide ddATP, ddCTP, ddGTP und ddTTP
dGMP	Desoxyguanosin-5'-monophosphat
DHFR	Dihydrofolat-Reduktase
DMARD	*disease-modifying antirheumatic drug*; Basistherapeutikum bei rheumatoider Arthritis
DNA	Desoxyribonukleinsäure
DNase	Desoxyribonuklease
dNTPs	2'-Desoxynukleosidtriphosphate, Mischung der DNA-Bausteine dATP, dCTP, dGTP und dTTP
DOI	*Digital Object Identifier*
dsRNA	doppelsträngige RNA
DU	Polio-D-Antigen-Einheit
dTMP	Desoxythymidin-5'-monophosphat
dUMP	Desoxyuridin-5'-monophosphat

E

E. coli	*Escherichia coli*

ECOG	*Eastern Cooperative Oncology Group*
EDSS	*expanded disability status scale*
EGF(R)	*epidermal growth factor (receptor)*
eIF	eukaryontischer Initiationsfaktor der Translation
ELISA	*enzyme-linked immunosorbent assay*
EMA	*European Medicines Agency*, zentrale europäische Arzneimittelzulassungsstelle
ENV	*envelope* (Gen in Retroviren)
EPAR	*European Public Assessment Report*, Europäischer Öffentlicher Beurteilungsbericht
EpCAM	*epithelial cell adhesion molecule*
EPO	Erythropoetin
ER	Endoplasmatisches Retikulum
ERK	*extracellular signal-regulated kinase*
ES	*embryonic stem cells*; embryonale Stammzellen
ESE	*exonic splicing enhancer*
ESchG	Embryonenschutzgesetz
ESS	*Exonic splicing silencer*

F

Fab	*fragment antigen-binding* (von Antikörpern)
Fc	*fragment crystallizable* (Effektordomäne von Antikörpern)
FcγR	Rezeptor für Fc-Teil von IgG-Antikörpern
FCS	*fetal calf serum*; fetales Kälberserum
FDA	*Food and Drug Administration*; US-amerikanische Arzneimittelzulassungsstelle
FGE	*formylglycine-generating enzyme*
FISH	*fluorescent in situ hybridization*
FOLFIRI	Chemotherapie mit Folinsäure + 5-Fluorouracil + Irinotecan
FOLFOX	Chemotherapie mit Folinsäure + 5-Fluorouracil + Oxaliplatin
FRET	Fluoreszenz-Resonanz-Energietransfer
FSH	follikelstimulierendes Hormon
Fuc	Fucose
FVC	forcierte Vitalkapazität

G

GA	*gene activation* (Genaktivierung in humanen Zellen)
GAG	*group-specific antigen* (Retroviren)
Gal	Galactose
GalNAc	*N*-Acetylgalactosamin
GALNS	*N*-Acetylgalactosamin-6-Sulfatase
GBq	Giga-Becquerel
gc	*genome copies*; Genomkopien (Dosierung viraler Gentherapeutika)
GC	Gemcitabin
G-CSF	Granulozyten-Kolonie-stimulierender Faktor
GF	*growth factor* (Wachstumsfaktor)
Glc	Glucose
GlcNAc	*N*-Acetylglucosamin
GLP	*glucagon-like-peptide*
GM-CSF	Granulozyten-Monozyten-Kolonie-stimulierender Faktor
GMP	*Good Manufacturing Practice*
GHRH	*growth hormone releasing hormone*
GnT	*N*-Acetylglucosamin-Transferase
gp	Glycoprotein
GPI-Anker	Glycosylphosphatidylinositol-Anker
GRAS	*generally recognized as safe*
GST	Glutathion-*S*-Transferase
GvHD	*graft-versus-host disease*
GVO	gentechnisch veränderter Organismus

H

HAE	hereditäres Angioödem
HAMA	humaner Anti-Maus-Antikörper
HAT	Hypoxanthin, Aminopterin, Thymidin
HBV	Hepatitis-B-Virus
hCG	humanes Choriongonadotropin
HEK	*human embryonic kidney cells*
HER	*human epidermal growth factor receptor*
HERV	endogenes humanes Retrovirus
hGH	*human growth hormone*; humanes Wachstumshormon
HGPRT	Hypoxanthin-Guanin-Phosphoribosyltransferase
HHV	humanes Herpesvirus
Hib	*Haemophilus influenzae* Typ b
His_6	Hexahistidin-Peptid
HIT	Heparin-induzierte Thrombozytopenie
HIV	humanes Immundefizienz-Virus
HLA	humanes Leukozyten-Antigen
hMG	humanes Menopausen-Gonadotropin
HPV	humanpathogenes Papillomavirus
HR	*hazard ratio*
HSV	Herpes-simplex-Virus

I

ICAM-1	*intercellular adhesion molecule-1*
ICH	*International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*
IE	Internationale Einheit
IFN	Interferon
IgA	Immunglobulin A
IgD	Immunglobulin D
IgE	Immunglobulin E
IGF	insulinähnlicher Wachstumsfaktor
IGFBP	*insulin-like growth factor-binding protein*
IGFR	*insulin-like growth factor receptor*
IgG	Immunglobulin G
IgM	Immunglobulin M
IHRES	interne ribosomale Eintrittsstelle

IL	Interleukin
IL-1Ra	Interleukin-1-Rezeptorantagonist
INN	*International Nonproprietary Name*
iPS	*induced pluripotent stem cells*
IPTG	Isopropyl-β-D-thiogalactosid
IRS	Insulinrezeptor-Substrat
ISE	*Intronic splicing enhancer*
ISS	*Intronic splicing silencer*
ITP	idiopathische thrombozytopenische Purpura

J

JAK	Januskinase
JNK	Jun-N-terminale Kinase

K

kbp	*kilo base pairs*, 1000 Basenpaare (bp)
KGF(R)	*keratinocyte growth factor (receptor)*
kg KG	Kilogramm Körpergewicht
KI	Konfidenz-Intervall
KRAS	*Kirsten rat sarcoma*

L

LAK	Lymphokin-aktivierte Killerzellen
LH	luteinisierendes Hormon
LTR	*long terminal repeat* (in Retroviren)

M

MAE-hGH	rekombinantes Wachstumshormon mit den zusätzlichen N-terminalen Aminosäuren Methionin, Alanin und Glutaminsäure
Man	Mannose
MAP-Kinase	Mitogen-aktivierte Proteinkinase
Mbp	Megabasenpaare, 10^6 bp
MBq	Mega-Becquerel
mCi	Milli-Curie
MCS	*multiple cloning site*
MHC	*major histocompatibility complex*
M-hGH	rekombinantes Wachstumshormon mit einem zusätzlichen N-terminalem Methionin
MLV	Maus-Leukämie-Virus
mM	millimolar
MMAE	Monomethylauristatin E
MMP	Matrixmetalloproteinase
MoMLV	Moloney-Maus-Leukämievirus
MPL	3-*O*-Desacyl-4'-monophosphoryl-Lipid A
MPR	Mannose-6-phosphat-Rezeptor
MPS	Mukopolysaccharidose
MRT	Magnetresonanztomographie
MS	Multiple Sklerose
MTX	Methotrexat
mU	Milli-Units

N

NAT	Nukleinsäure-Amplifikationstechniken
NBCD	*non-biological complex drug*; nicht-biologisches komplexes Arzneimittel
NeuAc	*N*-Acetylneuraminsäure (Sialinsäure)
NeuGC	*N*-Glycolylneuraminsäure
neo^R	Resistenz gegen Neomycin bzw. G418 (Aminoglycosid-Phosphotransferase)
NFκB	*nuclear factor κ B*
NGF(R)	*nerve growth factor (receptor)*
NHL	Non-Hodgkin-Lymphom
NIH	*National Institutes of Health*
NK	Natürliche Killerzellen
nM	nanomolar
NSCLC	*non-small cell lung cancer*; nicht-kleinzelliges Lungenkarzinom
NYHA I-IV	*New York Heart Association* I–IV

O

OPG	Osteoprotegerin
ORF	offener Leserahmen
ori	*origin of replication*, Replikationsursprung

P

p53	Tumorsuppressorprotein 53 kDa
PACE	*paired basic amino acid cleaving enzyme*
PBMC	*periperal blood mononuclear cells*
PCR	Polymerase-Kettenreaktion
PD-1(L)	*programmed death-1 (ligand)*
PDGF(R)	*platelet-derived growth factor (receptor)*; Plättchen-Wachstumsfaktor (-Rezeptor)
pdFIX	*plasma-derived factor IX*
PEG	Polyethylenglykol
PFU	plaquebildende Einheiten
PI3K	Phosphatidylinositol-3-Kinase
PlGF	*placental growth factor*; plazentarer Wachstumsfaktor
POL	*polymerase* (Gen in Retroviren)
PPMS	primär progrediente MS
pRB	Retinoblastom-Protein
PTH	Parathormon
PVX	Potatovirus X
PWS	Prader-Willi-Syndrom

R

RANK(L)	*receptor activator of nuclear factor kappa B (ligand)*
RNAi	*RNA interference*
RNase	Ribonuklease
RRE	*rev response element*
RMS	*relapsing multiple sclerosis*
RRMS	*relapsing remitting multiple sclerosis*; schubförmig remittierende MS
rsCT	rekombinant hergestelltes Lachs-Calcitonin

RSV	Respiratorisches Synzytial-Virus
RT	Reverse Transkriptase
RT-PCR	Reverse Transkription mit anschließender Amplifikation einer DNA durch PCR

S

scFv	*single-chain variable fragment* (eines Antikörpers)
SCID	*severe combined immunodeficiency*; schwere kombinierte Immundefizienz
SCNT	*somatic cell nuclear transfer*, somatischer Kerntransfer
SD	*standard deviation*, Standardabweichung
SDF-1	*stroma cell-derived factor 1*, auch CXCL-12
SDS	*standard deviation score*
Sf	Zelllinien aus *Spodoptera frugiperda*
SGA	*small for gestational age*; vorgeburtliche Wachstumsverzögerung
SHOX	*short stature homeobox*
siRNA	*small interfering RNA*
SLAMF7	*signaling lymphocyte activation molecule family member 7*
SLE	systemischer Lupus erythematodes
SNP	(sprich: „snip"): *single nucleotide polymorphism*
SPF	*specific pathogen-free*; spezifiziert pathogenfrei (Status von Tierherden für die Wirkstoffherstellung)
SPMS	sekundär progrediente MS
SRIF	*somatotropin release-inhibiting factor*; Somatostatin
SRP	*signal recognition particle*
STAT	*signal transducers and activators of transcription*
STIKO	Ständige Impfkommission des Robert Koch-Instituts
StZG	Stammzellgesetz

T

TGF	*transforming growth factor*
TIL	Tumor-infiltrierende Lymphozyten
TK	Thymidinkinase
T_m	halbmaximale Denaturierungstemperatur von DNA
TMV	Tabakmosaikvirus
Tn	Zelllinien aus *Trichoplusia ni*
TNF	Tumornekrosefaktor
TNSALP	*tissue non-specific alkaline phosphatase*; gewebeunspezifische Alkalische Phosphatase
tPA	Gewebe-Plasminogenaktivator
TPMT	Thiopurin-*S*-Methyltransferase
TPO	Thrombopoetin
TRH	*thyreotropin-releasing hormone*
TSH	schilddrüsenstimulierendes Hormon
TSLP	*thymic stromal lymphopoietin*

U

UDP	Uridindiphosphat
U	unit, Einheit

V

VCAM	*vascular cell adhesion molecule*
VEGF(R)	*vascular endothelial growth factor (receptor)*; vaskulärer endothelialer Wachstumsfaktor (-Rezeptor)
VLA-4	*very late antigen-4*, ein Protein aus der Integrin-Familie
vWF	von-Willebrand-Faktor

W

WCB	*working cell bank*, Arbeitszellbank
WHO	*World Health Organisation*, Weltgesundheitsorganisation

X

X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranosid
Xyl	Xylose

Y

YAC	*yeast artificial chromosome,* Hefe-artifizielles Chromosom
YEp	*Yeast Episomal plasmid*
YIp	*Yeast Integrating plasmid*

Z

ZNS	Zentralnervensystem

1 Die bio(techno)logische Revolution

1.1 Von der klassischen Genetik zur Gentechnik 1
1.2 Gentechnik und Biotechnik 6

Wir befinden uns inmitten einer „biologischen Revolution", oder besser gesagt, einer „molekulargenetischen Revolution". Seit gerade einmal etwas mehr als 60 Jahren wissen wir um die prinzipielle Struktur der Erbinformation, der DNA. Bereits kurze Zeit später haben wir verstanden, dass der „genetische Code" in allen auf der Erde lebenden Organismen gleich ist und in allen Organismen auf ähnliche Weise in Proteine übersetzt wird. In den frühen 1970er Jahren wurden die ersten DNA-Moleküle verschiedener Organismen, über Artgrenzen hinweg, zu neuen funktionalen Einheiten zusammengefügt. Diese neu entwickelte „DNA-Rekombinationstechnologie", die auch unter der vielleicht geläufigeren Bezeichnung „Gentechnik" bekannt ist, ist heute die Grundlage für die Erforschung komplexer physiologischer und pathophysiologischer Zusammenhänge auf molekularer Ebene, der Entwicklung neuer biomedizinischer Theorien und Therapieansätze, der Suche nach neuen Zielen für Wirkstoffe und der großtechnischen Produktion von Proteinwirkstoffen. „Biopharmazeutika", die in gentechnisch veränderten Organismen mit biotechnologischen Verfahren hergestellte Proteine enthalten, nehmen ein stetig wachsendes Segment im Arzneimittelschatz ein, sodass sich Arzneimittelfachleute neben ihrem Fachwissen über die Gewinnung und Anwendung chemisch-synthetischer oder natürlicher niedermolekularer Wirkstoffe zunehmend auch mit den Konzepten zur Produktion von und dem Umgang mit derartigen Makromolekülen befassen müssen. Immerhin haben wir es heute mit mehr als 200 zugelassenen Biopharmazeutika zu tun, die ca. 23 % des gesamten deutschen Pharmamarktes ausmachen – Tendenz steigend.

Die Methoden und Konzepte zur gentechnischen Herstellung von Proteinwirkstoffen nehmen in diesem Buch einen breiten Raum ein. Allerdings hat die Gentechnik eine weitaus größere Bedeutung in der biomedizinischen und pharmazeutischen Anwendung als „nur" ihre Nutzung zur Herstellung von Proteinwirkstoffen. In diesem einführenden Kapitel wollen wir daher die „biologische Revolution" zusammenfassen, die vor gut 150 Jahren mit Darwins Evolutionstheorie begann, deren Fundament die Etablierung zunächst der „klassischen" und danach der „molekularen" Genetik ist, die zunächst in der Erfindung der Gentechnik gipfelte und die möglicherweise erst mit dem gezielten „Editieren" des menschlichen Genoms und dem Klonen des Menschen endet. Nicht alles heute oder morgen Machbare der neuen molekularmedizinischen Entwicklungen (z. B. reproduktionsmedizinische Maßnahmen) hat etwas mit Gentechnik zu tun – allerdings wäre die Erforschung neuer molekularmedizinischer Therapiekonzepte ohne die Gentechnik mit Sicherheit nicht möglich gewesen. Heute wird es kein Medikament geben – übrigens auch keines mit einem chemisch-synthetischen Wirkstoff – in dessen Entwicklung von der biologischen Grundlagenforschung bis zur klinischen Entwicklung nicht zu irgendeinem Zeitpunkt gentechnische Methoden zum Einsatz kommen. Nachfolgend wollen wir kurz die Entwicklung der Genetik und der Gentechnik rekapitulieren, bevor wir den wichtigen Unterschied zwischen „Gentechnik" und „Biotechnik" thematisieren, der für das Verständnis der Entwicklung und Herstellung von Proteinwirkstoffen von Bedeutung ist.

1.1 Von der klassischen Genetik zur Gentechnik

Molekularbiologie ist von seinen Ursprüngen her ein Zweig der Biologie, der darauf abzielt, durch das Studium (makro)molekularer Strukturen Erkenntnisse über zelluläre Vorgänge zu erlangen. Die Geschichte der Molekularbiologie, der modernen biomedizinischen Forschung und der Entwicklung von Biopharmazeutika ist eng verbunden mit der Entdeckungsgeschichte der DNA als dem Träger der Erbinformation. Alles begann im Jahr 1865, als Gregor Mendel (1822–1884) seine Entdeckung veröffentlichte, dass Merkmale

von Organismen als vererbbare Einheiten in klaren mathematischen Zusammenhängen über Generationen weitergegeben werden, auch wenn sich die Anwesenheit dieser Merkmale nicht immer am Erscheinungsbild der Organismen ablesen lässt. Mendels „Versuche über Pflanzenhybriden", die im Jahr 1866 veröffentlicht wurden, zeigten, dass Merkmale in den Nachkommen entweder hervortreten und *dominant* sind oder zurücktreten und *rezessiv* sind. Mendel war seiner Zeit weit voraus, und seine Entdeckungen fanden zunächst kaum Beachtung. Sie sind jedoch heute ebenso fundamental wie damals, gelten sie doch beispielsweise auch für die Vererbung von Krankheiten des Menschen und bilden die Grundlage für die Identifizierung von „Krankheitsgenen" durch Stammbaum-Analysen in betroffenen Familien.

Ende des 19. Jahrhunderts wurde allmählich klar, dass die Zellkerne die Träger der Erbinformation sind. Im Zellkern fand Walter Flemming (1843–1901) eine anfärbbare Substanz, die er *Chromatin* nannte. Weitere Forschungen ergaben, dass sich das Chromatin während der Zellteilung verdichtet. Heinrich Wilhelm Waldeyer (1836–1921) sprach erstmals von *Chromosomen* und beschrieb 1888, dass sich Chromosomen während der Zellteilung aufspalten und die einzelnen Chromosomen-Fäden auf die entstehenden Tochterzellen verteilt werden. In diesem Zeitraum wurde auch klar, dass jedes Chromosom in einer Körperzelle zweimal vorhanden ist – außer in den Keimzellen, die jeweils nur eine Kopie enthalten.

Obwohl bereits andere Wissenschaftler um die Wende zum 20. Jahrhundert herum die Arbeiten von Mendel wiederentdeckt hatten, war es vor allem William Bateson (1861–1926), der im Jahr 1900 Mendels Publikation gelesen und die „Mendel'schen Regeln" in Wissenschaftskreisen bekannt gemacht hatte. Bateson hat auch den Begriff *Genetik* eingeführt (1906). Es wurde klar, dass die Chromosomen die Träger der Mendel'schen Merkmale sein müssen, und es war unter anderen Theodor Boveri (1862–1915), der aus dem Wissen der damaligen Zeit ableitete, dass die vererbbaren Merkmale im Mendel'schen Sinne auf den Chromosomen lokalisiert sein müssten. Walter S. Sutton (1877–1916) erkannte 1902, dass während der Bildung der Keimzellen und der Reduktion des doppelten Chromosomensatzes (heute nennen wir diesen Vorgang Meiose) die Chromosomenpaare in zufälliger Verteilung in die Keimzellen gelangen und dass dies die Grundlage für die von Mendel beschriebenen statistischen Verteilungsmuster der beobachteten Merkmale sein könnte. Diese Hypothese ist als *Chromosomentheorie der Vererbung* bzw. *Boveri-Sutton-Theorie* in die Geschichtsbücher eingegangen.

Der Botaniker Wilhelm Johannsen (1857–1927) benutzte in seinem Buch „Elemente der exakten Erblichkeitslehre" im Jahre 1909 erstmals den Begriff *Gen*. Das war damals etwas Abstraktes, ein Symbol, das dafür stand, dass die Merkmale von Organismen in den Keimzellen festgelegt sind und als unabhängige Einheiten an nachfolgende Generationen vererbt werden. Von Johannsen stammt auch der abgeleitete Begriff *Genotyp*, also die Gesamtheit der Gene eines Organismus. Am *Phänotyp*, dem Erscheinungsbild eines Organismus, ist der Genotyp nicht unbedingt zu erkennen, was Mendel bereits als dominante und rezessive Merkmale beschrieben hatte.

Zur endgültigen Akzeptanz der Mendel'schen Regeln trug insbesondere Thomas Hunt Morgan (1866–1945) bei. Morgan arbeitete mit der Fruchtfliege *Drosophila melanogaster*, einem überaus wertvollen Modellsystem der biologischen Forschung. Morgan identifizierte viele *Drosophila*-Mutanten mit bestimmten Merkmalen und zeigte, dass diese teilweise gekoppelt vererbt werden. Da Morgan genauso viele Kopplungsgruppen fand wie die Fruchtfliege Chromosomen hat, lag der Schluss nahe, dass Gene auf den Chromosomen liegen und daher in der Regel gemeinsam vererbt werden, sofern sie sich auf demselben Chromosom befinden. Morgan bemerkte aber auch Ausnahmen von dieser Regel, also Kopplungsbrüche, die offenbar dadurch zustande kamen, dass homologe Chromosomen während der Meiose Stücke austauschen. Schließlich war es Alfred H. Sturtevant (1891–1970), ein Mitarbeiter von Morgan, der 1913 erkannte, dass die Wahrscheinlichkeit, dass gekoppelt vererbte Merkmale während der Keimzellbildung voneinander getrennt werden, mit der Entfernung der Gene zueinander auf den jeweiligen Chromosomen zusammenhängt. Morgan fasste die Ergebnisse seiner Forschungen im Jahr 1928 in dem berühmten Buch *The Theory of the Gene* zusammen. Noch heute werden Entfernungen zwischen Genen auf Chromosomen als „Centimorgan" angegeben.

Auf dem Weg zur Beantwortung der Frage, was ein Gen eigentlich ist, hatte zunächst die Zytogenetikerin Barbara McClintock (1902–1992) im Jahr 1931 erstmals am Beispiel der Maispflanze gezeigt, dass die Entkopplung genetischer Marker mit einem Austausch von genetischem Material zwischen homologen Chromosomen einhergeht. Etwa um die gleiche Zeit (1927) berichtete Herman J. Muller (1890–1967), ein Schüler von Morgan, über seine Experimente zur mutationsauslösenden Wirkung von Röntgenstrahlung. Die Erbinformation konnte also von außen beeinflusst (verändert) werden. Max Delbrück (1906–1981) und Salvador E. Luria (1912–1991) zeigen 1943 in bahnbrechenden Experimenten mit Bakteriophagen, dass Mutationen zufällig auftreten.

Oswald T. Avery (1877–1955) hat 1944 gefunden, dass Desoxyribonukleinsäure (DNA) Träger geneti-

scher Information ist. Avery prägte den Begriff *Transformation* und meinte damit die Übertragung neuer Eigenschaften auf einen Organismus durch DNA. Alfred Kühn (1885–1968) und später George W. Beadle (1903–1989) und Edward Tatum (1909–1975) konnten um 1950 durch biochemische Experimente zeigen, dass Gene die Information zur Bildung von Enzymen (Proteinen) tragen: Die Hypothese „Ein Gen = ein Enzym" war geboren. Im Jahr 1952 zeigte Alfred Hershey (1908–1997), dass bei der Infektion von Bakterien mit Bakteriophagen nur die virale DNA, nicht aber seine Proteine, in die infizierte Zelle übertragen werden. Somit war klar, dass in der DNA die genetische Information zur Bildung neuer Phagenpartikel enthalten sein musste.

James D. Watson (geb. 1928) und Francis Crick (1916–2004) trafen sich erstmals im Jahr 1951 und waren fortan von der Idee getrieben, die Struktur der DNA als dem offensichtlichen Träger der genetischen Information zu entschlüsseln. Die experimentellen Daten dazu kamen eigentlich von der DNA-Strukturforscherin Rosalind E. Franklin (1920–1958) und wurden Watson und Crick von Maurice H. F. Wilkens (1916–2004) ohne Wissen Franklins zur Verfügung gestellt. Die berühmte, von Franklin 1952 erstellte Röntgenbeugungsaufnahme der DNA mit der Laborbezeichnung *Photograph 51* war letztlich die Grundlage für die Lösung der DNA-Struktur durch Watson und Crick, die in einer unscheinbaren einseitigen Publikation in *Nature* im Jahr 1953 erschien: Die DNA besteht aus zwei DNA-Strängen, die sich helixartig umeinander winden und antiparallel verlaufen; die Phosphatreste befinden sich außen und die DNA-Basen zeigen nach innen.

Für die endgültige Lösung der DNA-Struktur war noch eine weitere Information unerlässlich, und zwar die Lage der DNA-Basen beider Einzelstränge zueinander. Erwin Chargaff (1905–2002) hatte mit eleganten biochemischen Methoden etwa um das Jahr 1950 gezeigt, dass die Verhältnisse der DNA-Basen Thymin (T) und Adenin (A) sowie Cytosin (C) und Guanin (G) in einer DNA immer gleich sind. Daraus schlossen Watson und Crick treffend, dass sich T und A bzw. C und G in der Doppelhelix gegenüberliegen und über spezifische Wasserstoffbrückenbindungen verbunden sind. Sie schlussfolgerten auch, dass die spezifische Basenpaarung einen Replikationsmechanismus der DNA beinhaltet, denn durch die spezifische Basenpaarung sind die DNA-Stränge „komplementär". Wenn die Basen-Sequenz des einen DNA-Strangs bekannt ist, kann man aufgrund der *Komplementaritätsregeln* auf die Basenfolge des gegenüber liegenden Strangs schließen. Heute sind die Chargaff'schen Regeln bzw. die Komplementaritätsregeln Grundlage der DNA-Rekombinationstechnologie und damit der biologischen Revolution, die nach der Entdeckung der DNA-Doppelhelix erst so richtig in Gang kam.

Kurz nach der Entdeckung der DNA-Struktur führten Matthew Meselson (geb. 1930) und Franklin W. Stahl (geb. 1929) im Jahr 1958 ein heute als klassisch anerkanntes Experiment durch, mit dem sie zeigten, dass die DNA durch semikonservative Replikation vermehrt wird. Damit ist gemeint, dass die beiden DNA-Stränge geöffnet werden und ein Strang jeweils als Matrize zur Neubildung des Gegenstrangs verwendet wird. Etwa um die gleiche Zeit (1956) isolierte Arthur Kornberg (1918–2007) erstmals eine DNA-abhängige DNA-Polymerase, die diese Art der enzymatischen DNA-Synthese durchführen kann. DNA-Polymerasen gehören heute zu den unverzichtbaren Werkzeugen in der Gentechnologie.

Da nun klar war, dass die genetische Information in der DNA gespeichert ist und bei der Zellteilung von Zelle zu Zelle weitergegeben werden kann, stand die Frage im Raum, wie die in der Basenfolge der DNA „codierte" Information in die Aminosäureabfolge der Proteine übersetzt wird. Wieder war es Francis Crick, der den entscheidenden Hinweis gab. Basierend auf Experimenten mit Bakteriophagen schlug Crick in einer 1961 erschienenen *Nature*-Publikation vor, dass die genetische Information für Aminosäuren durch jeweils drei DNA-Basen codiert wird und dass diese Information „degeneriert" sei, also dass mehrere Basen-Tripletts für eine bestimmte Aminosäure stehen könnten. Welches Triplett für welche Aminosäure codiert, fanden zuerst Marshall Nirenberg (1927–2010) und Heinrich Matthaei (geb. 1929) heraus, die im Jahr 1961 ein kurzes Stück RNA herstellten, das ausschließlich Uracil-Basen enthielt. Als sie mit dieser RNA ein In-vitro-Translations-Experiment durchführten, erhielten sie ein kurzes Protein, das ausschließlich aus Phenylalanin bestand: Das Triplett UUU codiert für die Aminosäure Phenylalanin. Der gesamte genetische Code wurde in den folgenden Jahren entziffert, und es wurde festgeschrieben, was heute eine weitere Grundlage der Anwendung der DNA-Rekombinationstechnologie bildet: Der genetische Code ist auf der Erde universell gültig.

Das zunehmende Interesse an der Erforschung von Gen-Funktionen und ihrer Beteiligung am Auftreten von Krankheiten wurde zunächst gedämpft durch einen Mangel an Werkzeugen, mit denen DNA außerhalb von Zellen bearbeitet und Gene gezielt isoliert werden konnten. Man brauchte Methoden, um die langen chromosomalen DNA-Abschnitte gezielt in kürzere Fragmente zu zerlegen, solche Fragmente zu trennen und die gesuchten Gene zu charakterisieren. Das war die Geburtsstunde der Gentechnik. Zunächst wurden von Werner Arber (geb. 1929), Daniel Nathans (1928–1999) und Hamilton O. Smith (geb. 1931) die sogenannten

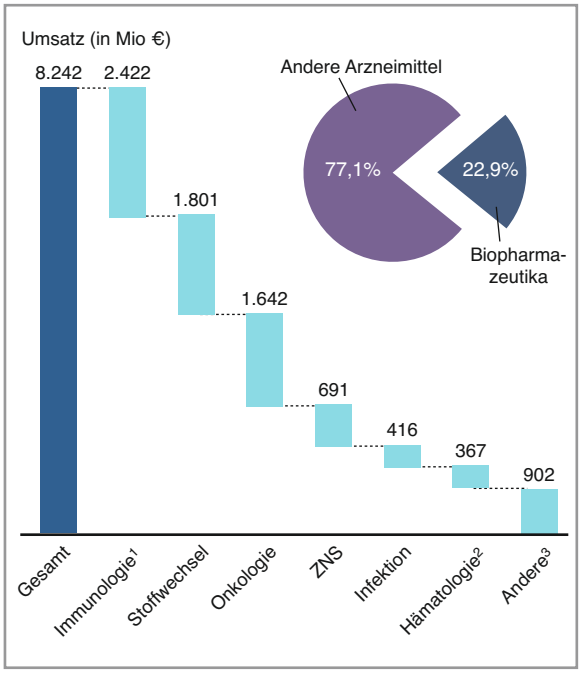


Abb. 1.1 Umsatz mit Biopharmazeutika in Deutschland nach Indikationsgebieten. Kreisdiagramm: Biopharmazeutika haben 2015 einen Netto-Umsatz im Apotheken und Krankenhaus-Markt unter Berücksichtigung des Herstellerabschlags von 22,9 % des Gesamt-Arzneimittelmarktes (35,9 Mrd. Euro) eingenommen. Balkendiagramm: Umsätze aufgeschlüsselt nach Indikationsgruppen.
[1] Enthalten u. a. TNF-α-Hemmer, ohne Multiple Sklerose (siehe ZNS),
[2] nicht enthalten: hämatologische Onkologie,
[3] enthalten u. a. Osteoporose, Erkrankungen der Augen und der Atemwege. Quelle: Biotech-Report 2016 der *Boston Consulting Group*, abrufbar unter www.vfa-bio.de.

Restriktionsendonukleasen als fundamentale Werkzeuge der Gentechnologie entdeckt und zum gezielten Fragmentieren von DNA verwendet. Paul Berg (geb. 1926) erzeugte 1972 erstmals rekombinante DNA, indem er Stücke des SV40-Virus mithilfe DNA-modifizierender Enzyme in bakterielle Plasmide einbaute. Kurz darauf (1973) zeigten Herbert W. Boyer (geb. 1936) und Stanley Cohen (geb. 1935), dass man in vitro rekombinierte Plasmide in Bakterienzellen einbringen und dort vermehren kann. Etwas später (1974) zeigten sie, dass man auch DNA einer fremden Spezies in solche Plasmide einbauen und von Bakterien replizieren und auch transkribieren lassen kann.

Bereits kurz nach ihrer Entwicklung wurde das kommerzielle Potenzial der DNA-Rekombinationstechnologie erkannt. Im Jahr 1976 gründete Herbert Boyer die Firma Genentech (***Genetic engineering technology***), die bereits ab 1978 mit Humaninsulin den weltweit ersten rekombinanten Proteinwirkstoff herstellte (Zulassung 1982). Ebenfalls im Jahr 1978 wurde von Walter Gilbert (geb. 1932), einem Pionier der DNA-Sequenzierung, die Firma Biogen gegründet, die mit Erfolg Proteine wie Hepatitis-B-Antigen, Alpha-Interferon und Beta-Interferon produzierte. Die 1980 gegründete Firma Amgen produzierte unter anderem Erythropoetin. Am 14. Oktober 1980 ging die damals vier Jahre alte kalifornische Firma Genentech an die New Yorker Börse – also noch zwei Jahre vor der Zulassung des ersten von der Firma produzierten Produkts. Damit hatte man sozusagen die kommerzielle Nagelprobe der Gentechnik bestanden. Unter Investoren initiierten diese Erfolge unglaubliche Phantasien: Binnen 20 Minuten stieg der Preis der Genentech-Aktie von 35 auf 89 Dollar. Bei Börsenschluss lag der Wert der Aktie bei 71,25 Dollar. Die Möglichkeiten der Biotechnik, einer sehr alten, aber keineswegs populären Wissenschaft, waren durch die Gentechnik deutlich erweitert worden, und das erregte damals unglaubliche Aufmerksamkeit.

Ein bedeutendes Ziel aller biotechnologischen Forschung ist die Entwicklung kommerzieller Produkte. Aus diesem Grund wird im Gegensatz zu vielen anderen wissenschaftlichen Disziplinen auch die gentechnische bzw. biotechnische Forschung zu einem großen Teil durch eine ökonomische Motivation getragen. Die Erwartungen eines signifikanten finanziellen Gewinns war auch der Grund für das große Interesse und die Begeisterung während der Anfangszeit der industriellen Gentechnik. Bereits im Jahre 1985, also kaum zehn Jahre nach der Entwicklung der DNA-Rekombinationstechnologie, existierten über 400 Biotech-Firmen in den USA. Im Jahr 2015 waren es nach dem *Biotechnology Industry Report 2016* von Ernst & Young insgesamt 2336 Firmen, davon 436 börsennotiert. Die weltweite Biotechnologie-Industrie konnte im Jahr 2015 mit ca. 132 Milliarden Dollar Betriebseinkommen aufwarten und beschäftigte über 200 000 Mitarbeiter.

Auch in Deutschland hat der Umsatz mit biotechnologisch erzeugten Proteinwirkstoffen stark an Bedeutung gewonnen. Im Biotech-Report 2016 der *Boston Consulting Group* wird erwähnt, dass der Anteil an Biopharmazeutika am Gesamtpharmamarkt (Apotheken und Kliniken) bei 22,9 % lag, das entsprach einem Netto-Umsatz von ca. 8,2 Milliarden Euro (Abb. 1.1) und einer Steigerung von 2014–2015 von knapp 10 %. Den größten Anteil am Umsatz haben Therapeutika für chronische Entzündungserkrankungen wie rheumatoide Arthritis oder Psoriasis sowie Stoffwechselerkrankungen (vor allem Diabetes mellitus; Abb. 1.1). Im Bereich der medizinischen Biotechnologie gab es 2015 in Deutschland 391 Unternehmen mit zusammen etwa 40 252 Beschäftigten. Im September 2016 waren 205 Biopharmazeutika (161 Wirkstoffe) auf dem deutschen Markt zugelassen. Biopharmazeutika stellten in den Jahren 2010–2016 einen Anteil von 22–30 % der jährlichen Arzneimittel-Neuzulassungen in Deutschland

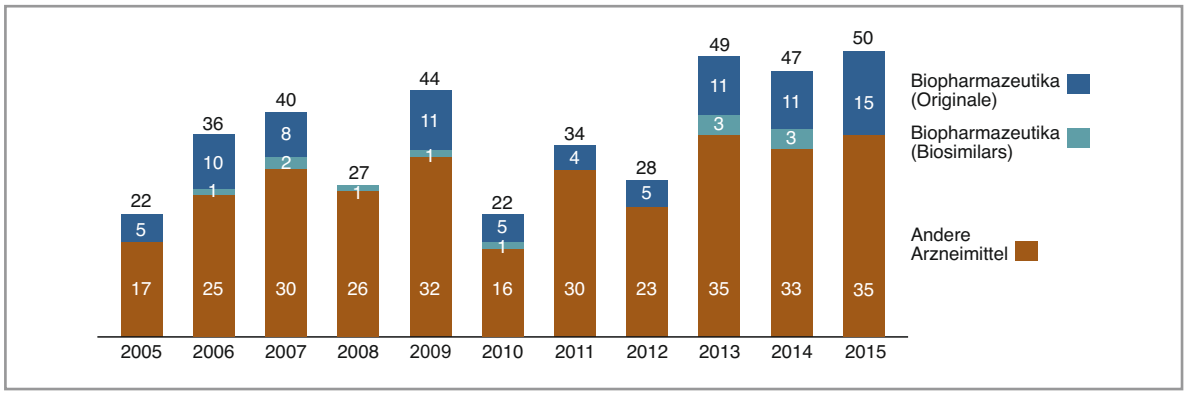


Abb. 1.2 Anteil von Biopharmazeutika an jährlichen Neuzulassungen in Deutschland. Aufgegliedert in „Andere Arzneimittel" (keine Proteinwirkstoffe) und Biopharmazeutika (Originalia und *Biosimilars*). Quelle: Biotech-Report 2016 der *Boston Consulting Group*, abrufbar unter www.vfa-bio.de.

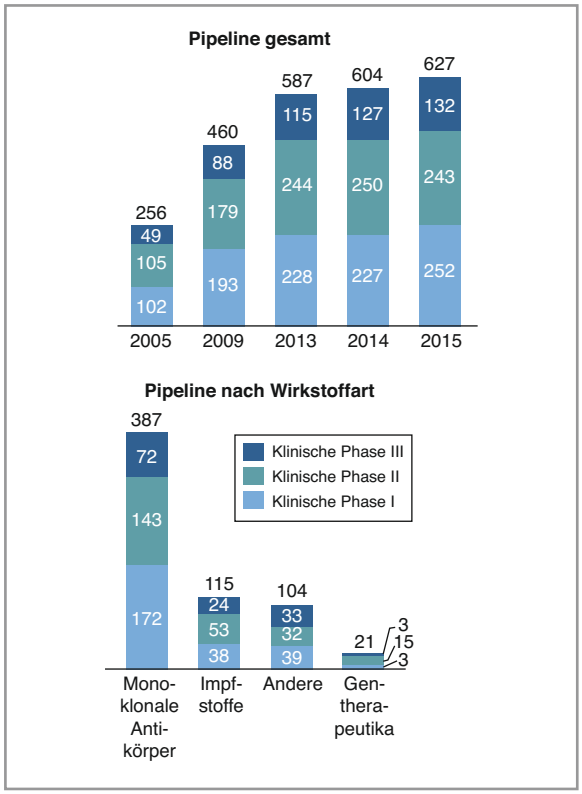


Abb. 1.3 Biopharmazeutika-*Pipeline*. Oben: Anzahl von Wirkstoffen in der klinischen Entwicklung zwischen 2005 und 2015. Unten: Aufgliederung der Arzneistoffe in Entwicklung im Jahr 2015 nach Wirkstoffarten. Quelle: Biotech-Report 2016 der *Boston Consulting Group*, abrufbar unter www.vfa-bio.de.

(Abb. 1.2). Und die *Pipeline* der pharmazeutischen Industrie ist gut gefüllt: 2016 befanden sich 627 Biopharmazeutika in klinischen Prüfungen (Phasen I-III), davon 132 in Phase III (Abb. 1.3). Mehr als die Hälfte der Neuentwicklungen fallen in den Bereich der monoklonalen Antikörper (Abb. 1.3), *dem* Wachstumsmarkt in den vergangenen Jahren und auch in absehbarer Zukunft.

Zusammenfassung

- Die bio(techno)logische Revolution begann vor etwa 150 Jahren mit Charles Darwins Theorie über die Entstehung der Arten. Die Vererbungslehre (Genetik) gründet auf klassischen Arbeiten von Gregor Mendel aus dem Jahr 1865.
- In den folgenden etwa 100 Jahren wurden die Grundlagen für neue naturwissenschaftliche Disziplinen geschaffen, die wir heute als Molekularbiologie und Molekulare Genetik kennen.
- Die Erfindung der DNA-Rekombinationstechnologie (Gentechnik) durch Herbert Boyer und Stanley Cohen Anfang der 1970er Jahre hat nicht nur den Weg für die systematische Erforschung von Genfunktionen und ganzen Genomen geebnet.
- Gentechnische Methoden sind heute unerlässlich für die biomedizinische Forschung über pathophysiologische Mechanismen und die Identifizierung klinisch relevanter Zielstrukturen für Wirkstoffe. Sie bilden ebenso die Grundlage für die Herstellung von Proteinwirkstoffen (Biopharmazeutika).
- Eine ganze Industrie hat sich seit Ende der 1970er Jahre etabliert, die heute weltweit ca. 130 Milliarden Dollar erwirtschaftet und ca. 200 000 Angestellte beschäftigt und die mit biotechnologischen Verfahren potente Arzneimittel gegen Krankheiten liefert, die bis vor Kurzem noch als wenig oder gar nicht therapierbar galten.

1.2 Gentechnik und Biotechnik

Häufig werden die Begriffe *Molekularbiologie* und *Gentechnologie* (bzw. *Gentechnik*) synonym verwendet. Das ist jedoch nicht korrekt. Der Begriff *Molekulare Biologie* wurde zuerst von dem amerikanischen Wissenschaftler Warren Weaver benutzt – und zwar bereits im Jahr 1938. Damals wollte die Rockefeller-Stiftung grundlegende biologische Fragestellungen im Bereich der biologischen Forschung und insbesondere die Frage nach der Natur des Gens finanziell unterstützen und dazu die zwei zu der Zeit vorherrschenden Denkweisen und experimentellen Ansätze zusammenbringen: die klassische Genetik und die Strukturbiologie. Daraus hat sich eine ganz neue Disziplin der biologischen Forschung geformt, und zwar die *Molekulare Genetik*, die sich mit den molekularen Grundlagen der Vererbung von Genen befasst, wie etwa der Struktur und Funktion von Nukleinsäuren oder den Replikations- und Genexpressionsmechanismen. Der Begriff *Molekulare Biologie* hat sich erst im Lauf der Zeit entwickelt und überlappt heute mit der Molekularen Genetik; Molekularbiologie ist heute eine Disziplin innerhalb der Biologie, die sich mit der Struktur und dem Stoffwechsel von Nukleinsäuren und deren Interaktionen mit zellulären Komponenten auf molekularer Ebene befasst. Die Überlappungen mit anderen biologischen Disziplinen wie der Zellbiologie, Genetik oder Biochemie sind dabei fließend. Wenn sich heute fast alle Forscher aus egal welchen biologischen Disziplinen übergreifend als *Molekularbiologen* bezeichnen, dann meinen sie damit vermutlich, dass sie Vorgänge innerhalb einer Zelle oder auch Interaktionen zwischen Zellen auf molekularer Ebene untersuchen. Dafür ist die Anwendung gentechnischer Methoden zwar unerlässlich, allerdings ist Molekularbiologie nicht mit Gentechnik gleichzusetzen. Das erkennt man auch daran, dass Molekularbiologie eine Wissenschaft ist, während Gentechnik eine Methodensammlung darstellt, die es uns erlaubt, die DNA eines Organismus zu isolieren, gezielt zu analysieren und gegebenenfalls auch zu rekombinieren.

Auch die Begriffe *Gentechnik* und *Biotechnik* bedeuten nicht das Gleiche. Der Begriff Biotechnik wurde im Jahr 1917 durch den ungarischen Ingenieur Karl Ereky (1878–1952) geprägt. Er beschrieb damit einen von ihm entwickelten Prozess, mit dem er die Schweinemast verbesserte, indem er aus billigen Zuckerrüben wertvolles Schweinefutter herstellte. Unter Biotechnik verstand Ereky solche Prozesse, in denen ein Rohmaterial mithilfe lebender Organismen zu einem Produkt veredelt werden konnte. Fast 45 Jahre lang blieb der Begriff Biotechnik recht vage definiert. Das änderte sich im Jahr 1961, als der schwedische Mikrobiologe Carl-Göran Hedén (1920–2009) für ein wissenschaftliches Journal, das sich der angewandten Mikrobiologie und industriellen Fermentation widmete, den Titel *Biotechnology and Bioengineering* vorschlug. Seit dieser Zeit verstehen wir die Biotechnologie (Biotechnik) als industriellen Prozess, bei dem biologische Organismen, biologische Systeme und biochemische Reaktionen zur Herstellung von Produkten eine entscheidende Rolle spielen. Biotechnik gründet sich also unter anderem auf den Erfahrungen aus Mikrobiologie, Biochemie und Verfahrenstechnik.

Ein typischer biotechnischer Prozess mit Zellen als Produzenten eines Produkts besteht in Anlehnung an Karl Erekys ursprüngliche Beschreibung aus drei Phasen:

- Im *upstream processing* genannten Abschnitt der biotechnischen Produktion wird die verwendete Produktionskultur zunächst für das Wachstum und die Produktion vorbereitet.
- Anschließend folgt die Fermentation, in der die Produktionskultur in großem Maßstab vermehrt wird und in der ein Produkt gebildet oder durch Biotransformation biochemisch verändert wird.
- Es schließt sich das *downstream processing* an, also die Reinigung des Produkts entweder aus den Produktionszellen oder dem Kulturmedium.

In der biotechnischen Forschung wird versucht, jeden einzelnen dieser Schritte zu optimieren. In den 1960er und 1970er Jahren konzentrierte sich die Forschung auf das *upstream processing*, auf Bioreaktor-Konzepte und das *downstream processing*. Diese Studien führten zu enormen Fortschritten, vor allem im Anlagenbau und in der Überwachung und Steuerung von Fermentationsprozessen. Als weitaus schwieriger erwies sich die Suche nach neuen Organismen für die Optimierung von Biotransformationsprozessen. Zwar bieten natürlich vorkommende Stämme ein unglaubliches Spektrum biosynthetischer Möglichkeiten mit bemerkenswerter Spezifität und Selektivität. Allerdings sind diese biochemischen Reaktionen für den biotechnischen Einsatz keineswegs optimiert.

Dies wurde teilweise mit Methoden der klassischen Genetik korrigiert. Durch die Einführung von Mutationen in die Genome von Organismen gelang es, die Produktivität bestimmter Prozesse beachtlich zu steigern. Allerdings wird in solchen klassischen Selektionsverfahren die DNA nicht aus dem ausgewählten Organismus isoliert, um sie gezielt zu verändern (dies würde gentechnische Methoden voraussetzen). Mutationen werden vielmehr ungerichtet durch physikalische oder chemische Mutagene erzeugt und die gewünschte Eigenschaft danach durch Selektion von Mutanten isoliert. Das sind langwierige Prozesse, die noch dazu dar-

auf limitiert sind, dass man lediglich die vorhandenen Syntheseleistungen der ausgewählten Organismen optimieren, aber keine „neuen Syntheseleistungen“ hinzufügen kann.

Die Produkte biotechnischer Verfahren sind aus dem täglichen Leben nicht mehr wegzudenken. Sie reichen von Bier, Wein und anderen Lebens- und Genussmitteln über Metabolite wie Vitamine und Aminosäuren zu Produkten des Sekundarstoffwechsels wie Antibiotika oder Alkaloiden. Durch Biotransformation können häufig leicht zugängliche Ausgangsstoffe in hochwertige Therapeutika umgewandelt oder komplexe sekundäre Pflanzenstoffe wie Taxane hergestellt werden, die aus natürlichen Quellen nicht in ausreichenden Mengen zur Verfügung stünden. Als biologische Quellen für biotechnische Verfahren wird heute eine breite Palette von Organismen in Form sogenannter Hochleistungsstämme etabliert – Bakterien, Pilze, Algen sowie pflanzliche und tierische Zellen sind in solchen Prozessen weit verbreitet.

Was verstehen wir unter Gentechnik? Gentechnik ist eine Sammlung von Methoden, mit denen man DNA aus einem Organismus isolieren, in vitro rekombinieren und dann in einen anderen Organismus einschleusen kann, beispielsweise um dort ein bestimmtes Produkt herzustellen (Abb. 1.4). Insofern können die Bereiche der Biotechnik und Gentechnik durchaus überlappen. Das Ergebnis einer mit gentechnischen Methoden optimierten biotechnischen Produktion kann dem entsprechen, was man auch durch den Einsatz von klassischen Selektionsverfahren erreichen kann. Beispielsweise kann ein Kontrollmechanismus sowohl durch ungerichtete Mutagenese als auch durch gezielte Modifikation an einer isolierten DNA außer Kraft gesetzt werden, um so eine Produktionssteigerung eines intrazellulären Metaboliten zu bewirken. Darüber hinaus können durch gentechnische Methoden aber auch die natürlicherweise vorhandenen Syntheseleistungen von Organismen erweitert werden. So kann mit gentechnischen Methoden erreicht werden, dass etablierte Produktionsstämme durch das Einführen fremder genetischer Information gezielt in einer Weise verändert werden, dass sie praktisch jede beliebige fremde Syntheseleistung durchführen können (Abb. 1.4). Woher die genetischen Korrelate für diese Syntheseleistungen stammen, ist letztlich nicht mehr von Bedeutung, da mit gentechnischen Methoden der Austausch genetischer Informationen nicht nur über Artgrenzen hinweg möglich ist, sondern selbst zwischen sehr unterschiedlichen Organismen beispielsweise aus dem Tier- und Pflanzenreich oder zwischen Pro- und Eukaryonten.

Die Vereinigung von Gentechnik und Biotechnik zu einer *Molekularen Biotechnologie* ist besonders markant

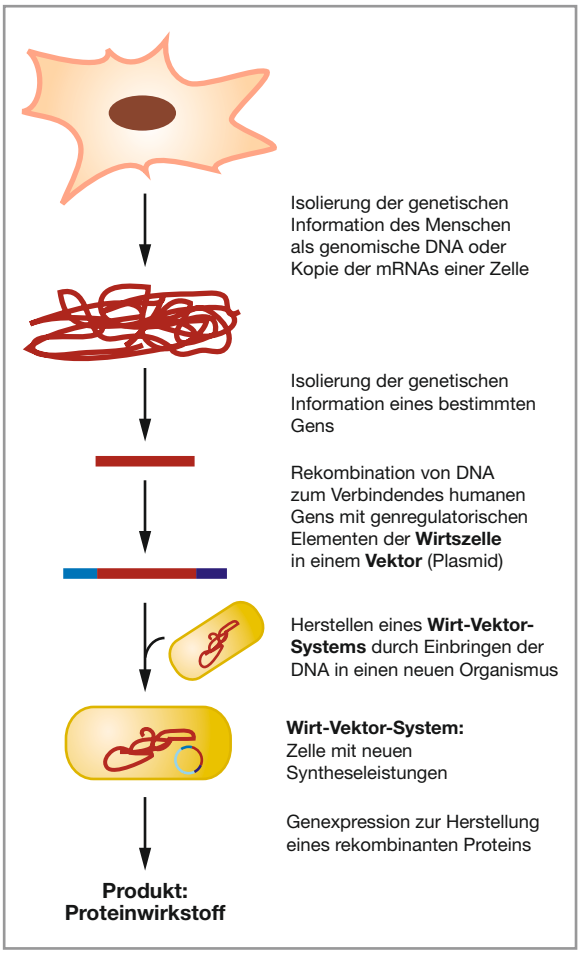


Abb. 1.4 Kombination von Gentechnik und Biotechnik zur Herstellung von Biopharmazeutika. Zunächst muss die genetische Information für das als Wirkstoff entwickelte Protein isoliert werden, in der Regel aus dem menschlichen Genom. Die genetische Information dafür steht entweder in Form der genomischen DNA im Zellkern oder als *Pool* aller mRNAs im Zytoplasma einer Zelle zur Verfügung. Die genetische Information für das gesuchte Gen wird aus diesem *Pool* isoliert und in einem Vektor (Plasmid) mit Kontrollelementen für die Expression in einer geeigneten Wirtszelle rekombiniert. Diese rekombinierte DNA wird dann in eine geeignete Produktionszelle eingeschleust. Es entsteht ein Wirt-Vektor-System, das in einem geeigneten biotechnischen Verfahren zur Produktion des „rekombinanten Proteins“ genutzt werden kann.

in der Produktion von Proteinwirkstoffen (Biopharmazeutika; Abb. 1.4). Die Nutzung dieser Wirkstoffe, die sich in der Herstellung so fundamental von chemisch-synthetischen „kleinen Molekülen“ unterscheiden, wäre ohne Anwendung gentechnischer Methoden nicht denkbar, denn fast immer geht es um die Produktion menschlicher Proteine in nicht-menschlichen Produktionszellen, seien es Bakterien, eukaryontische Mikroorganismen oder kultivierte tierische oder pflanzliche

Zellen, und damit um die Hinzufügung einer artfremden Syntheseleistung in den Produzenten. In den folgenden Kapiteln werden wir die Konzepte der Entwicklung und Herstellung von Biopharmazeutika mit gentechnischen Methoden ausführlich darstellen – vom Gen zum Wirkstoff.

Zusammenfassung

- Die Begriffe *Molekularbiologie*, *Biotechnik* und *Gentechnik* sollten nicht verwechselt oder synonym verwendet werden.
- Der Begriff Molekularbiologie beschreibt in der engen Auslegung eine Wissenschaft, die Nukleinsäuren von Zellen und die Mechanismen der DNA-Replikation und Genexpression untersucht. Eine Erweiterung des Begriffs schließt im Prinzip alle Konzepte und Methoden mit ein, die uns heute die Analyse zellulärer Vorgänge – die ja letztlich immer genetisch gesteuert sind – auf molekularer Ebene ermöglichen.
- Biotechnik (Biotechnologie) befasst sich mit der Etablierung industrieller Prozesse zur Produktion von Stoffen mit biologischen Organismen oder biochemischen Reaktionen zur Herstellung von Produkten. Biotechnik kommt im Prinzip ohne Gentechnik aus.
- Allerdings kann die Kombination von Gentechnik und Biotechnik die Selektion von Hochleistungs-Produktionsstämmen erleichtern und die Produktion von Stoffen ermöglichen, die eigentlich nicht der natürlichen Syntheseleistung eines Organismus entsprechen. Letzteres ist insbesondere im Bereich der Produktion therapeutischer Proteine von Belang.
- Die Produktion von Proteinwirkstoffen in gentechnisch veränderten Organismen ist nur durch eine Symbiose von Biotechnik und Gentechnik möglich, die heute mit Begriffen wie *Pharmazeutische Biotechnologie* oder *Molekulare Biotechnologie* umschrieben wird.

2 DNA-Rekombinationstechnologie

2.1 Die DNA 9
2.2 Chemische DNA-Synthese 18
2.3 Enzymatische DNA-Synthese 20
2.4 Prinzipien der Gentechnik 37

2.1 Die DNA

In allen lebenden Organismen bildet die Desoxyribonukleinsäure (DNS) den molekularen Träger der Erbinformation. DNA ist die heute gebräuchlichere Abkürzung der englischen Bezeichnung *desoxyribonucleic acid*. Der Begriff DNA beschreibt nicht etwa ein bestimmtes Molekül, sondern lediglich eine Molekülart. Es ist klar, dass die genetische Information in jeder Spezies verschieden sein muss. Doch ist es nicht die Molekülart DNA an sich, die in jeder Spezies variiert, sondern die in der DNA codierte Information.

DNA-Moleküle sind das Zielobjekt der Gentechnik. Dabei verstehen wir Gentechnik als eine Technologie, die es uns erlaubt, DNA aus einem Organismus zu isolieren, die in der DNA codierte Information zu verstehen, DNA-Moleküle gegebenenfalls miteinander zu rekombinieren und diese Information in einem anderen Organismus zu realisieren. Insbesondere streben wir zwei Ziele an, wenn wir das Instrumentarium der Gentechnik einsetzen:

- Die DNA wird molekular charakterisiert. Das bedeutet, dass wir die Abfolge der DNA-Basen ermitteln, also den Informationsinhalt einer DNA verstehen wollen und untersuchen, wann und wo im komplexen Leben eines Organismus diese Information abgerufen wird.
- Teilbereiche von DNA-Molekülen werden neu kombiniert. Ziel ist es entweder, den Informationsgehalt der DNA zu variieren oder ein bestimmtes Gen mit neuen Kontrollelementen auszustatten, um zum Beispiel die in der DNA gespeicherte Information in einem anderen Organismus abrufbar zu machen.

2.1.1 Aufbau der DNA

Bei der DNA handelt es sich um sehr große, strukturell sehr einheitlich gebaute Moleküle. Die monomeren Bausteine der DNA sind die Desoxynukleotide

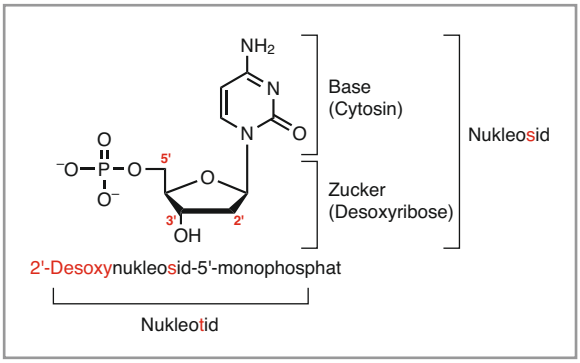


Abb. 2.1 DNA-Bausteine. Als Beispiel für einen DNA-Baustein ist das Desoxynukleotid dCMP dargestellt. Die Base ist Cytosin, das Nukleosid ist die Base mit dem Zuckeranteil (Desoxyribose), also 2'-Desoxycytidin. Das Nukleotid umfasst die Base, den Zuckeranteil und die Phosphatgruppe, also 2'-Desoxycytidin-5'-monophosphat.

(Abb. 2.1). Jeder dieser Bausteine besteht aus drei Komponenten:

- einer Purin- oder Pyrimidinbase,
- einer Desoxyribose, die *N*-glycosidisch mit der Purin- bzw. Pyrimidinbase verknüpft ist,
- einem anorganischen Phosphatrest, der esterartig mit der 5'-OH-Gruppe der Desoxyribose verknüpft ist.

Durch die *N*-glycosidische Verknüpfung einer DNA-Base mit einer Desoxyribose entsteht ein Nukleosid. Wird an die 5'-OH-Gruppe der Desoxyribose eines Nukleosids mindestens eine Phosphatgruppe gebunden, spricht man von einem Nukleotid. In der DNA kommen jeweils zwei Purinbasen (Adenin und Guanin) und zwei Pyrimidinbasen (Thymin und Cytosin) vor (Abb. 2.2). Die Bezeichnungen der in einem DNA-Molekül vorkommenden Nukleotide lauten:

- Desoxycytidin-5'-monophosphat (dCMP),
- Desoxyguanosin-5'-monophosphat (dGMP),

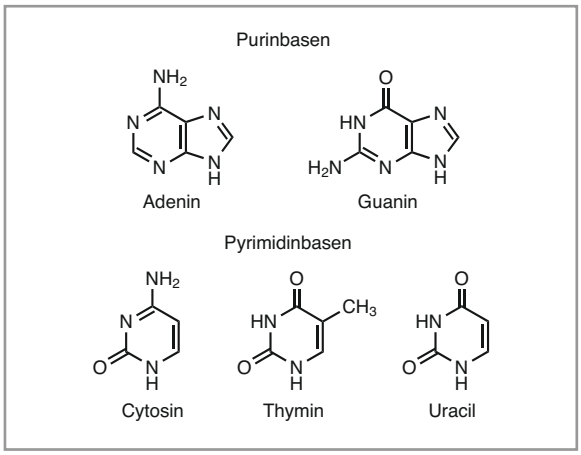


Abb. 2.2 Nukleinsäurebasen. Dargestellt sind die fünf Nukleinsäurebasen Adenin, Guanin, Cytosin, Thymin und Uracil. Thymin und Uracil sind funktionell äquivalent. Allerdings kommt Thymin nur in DNA und Uracil nur in RNA vor. Dies liegt daran, dass die DNA-Polymerase nicht dUTP, sondern dTTP als Substrat erkennt, wohingegen die RNA-Polymerase UTP und nicht TTP als Baustein verwendet.

- Desoxythymidin-5'-monophosphat (dTMP),
- Desoxyadenosin-5'-monophosphat (dAMP).

Im normalen Sprachgebrauch werden die DNA-Nukleotide mit den Buchstaben C, G, T und A abgekürzt.

Ribonukleinsäure (RNA) unterscheidet sich von DNA durch das Vorhandensein von Ribose anstelle der Desoxyribose. Außerdem ist in RNA Thymin durch die dritte Pyrimidinbase, Uracil, ersetzt. Bis auf die Tatsache, dass die DNA-Polymerasen 2'-Desoxythymidintriphosphat und die RNA-Polymerasen Uridintriphosphat als Substrat erkennen, sind Thymin und Uracil biologisch äquivalent. Dies gilt insbesondere für die spezifische Basenpaarung mit Adenin (siehe unten).

Innerhalb der DNA sind die einzelnen Nukleotide durch Phosphodiesterbindungen zwischen dem 5'-C-Atom der Desoxyribose des **einen** Bausteins und dem 3'-C-Atom der Desoxyribose des **anderen** Bausteins verknüpft (Abb. 2.3). Das bedeutet, dass jede DNA-Kette eine eindeutige Richtung hat. Sie hat ein 5'-Ende und ein 3'-Ende, wobei das 5'-Ende der DNA als „DNA-Anfang" und das 3'-Ende als „DNA-Ende" definiert wird. Dies ist auch biologisch sinnvoll, da alle Nukleinsäuren – DNA ebenso wie RNA – vom 5'-Ende zum 3'-Ende synthetisiert werden. Daraus leitet sich eine allgemein gültige Konvention ab: Ein einzelner DNA-Strang oder auch eine RNA-Sequenz wird von links nach rechts immer vom 5'-Ende zum 3'-Ende geschrieben. Bei doppelsträngigen DNA-Fragmenten gilt diese Regel immer für den „oberen" Strang. Wird von dieser Konvention abgewichen, ist es wichtig, das 5'-Ende und das 3'-Ende entsprechend zu kennzeichnen.

Zwei einander komplementäre DNA-Moleküle bilden eine DNA-Doppelhelix, wobei die Zucker-Phosphatketten in entgegengesetzter, antiparalleler Richtung verlaufen. Mit **komplementär** ist gemeint, dass jedem dCMP ein dGMP und jedem dTMP ein dAMP gegenübersteht. Für die gegenseitige Bindung bilden die komplementären Basen schwache Wechselwirkungen in Form von Wasserstoffbrücken aus. Ein G/C-Basenpaar bildet drei und ein A/T-Basenpaar zwei Wasserstoffbrückenbindungen (Abb. 2.4).

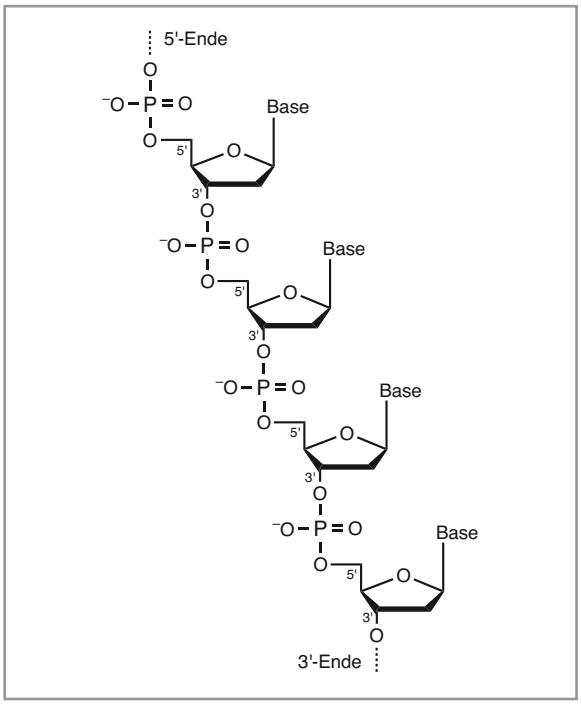


Abb. 2.3 Verknüpfung von Nukleotiden. Die Verknüpfung der Nukleotide erfolgt in der RNA wie in der DNA als Phosphorsäurediester über die 5'- und 3'-OH-Gruppen der Ribosen bzw. Desoxyribosen. Damit erhalten Nukleinsäuren eine eindeutige Polarität, da sie durch diesen Verknüpfungsmechanismus ein 5'-Ende und ein 3'-Ende besitzen.

Wegen der Komplementarität von A/T und G/C ist innerhalb eines DNA-Moleküls die Anzahl der G-Nukleotide identisch mit der Anzahl der C-Nukleotide. Ebenso ist die Anzahl der A-Nukleotide gleich der Anzahl der T-Nukleotide. Andererseits ist aber in der Regel die Anzahl der A+T-Basenpaare nicht gleich der Anzahl der G+C-Basenpaare. Vielmehr unterscheidet sich der sogenannte G/C-Gehalt, ausgedrückt in Prozent, in Organismen verschiedener Arten. Sehr G/C-arme Genome besitzt beispielsweise der Erreger der Malaria tropica *Plasmodium falciparum* (19 %). Hingegen enthält die DNA aus bestimmten *Micrococcus*-Arten bis zu 75 % G/C-Basenpaare. Genome aus Säugetierzellen besitzen einen G/C-Gehalt von ca. 45 % und das Genom von *Escherichia coli* enthält nahezu gleich viele G/C- wie A/T-Basenpaare. Generell können wir festhalten, dass der G/C-Gehalt des genetischen Materi-

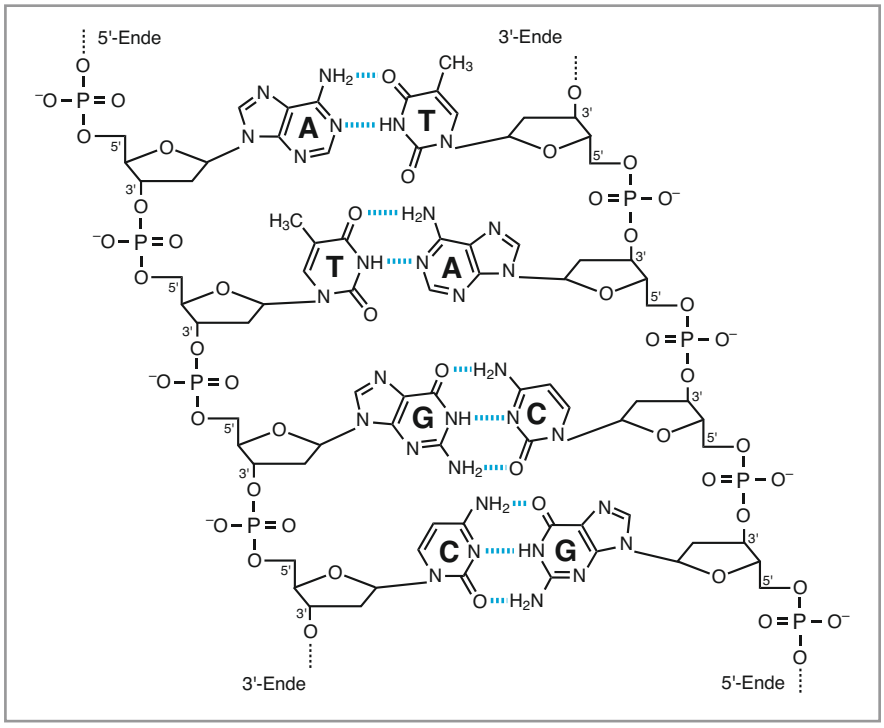


Abb. 2.4 Komplementarität von Nukleinsäuren. Nukleotide sind paarweise zueinander komplementär. Die komplementären Nukleotidpaare A/T (in DNA) bzw. A/U (in RNA) werden durch zwei Wasserstoffbrücken zwischen den Basen stabilisiert. Dagegen bildet das komplementäre Nukleotidpaar G/C drei Wasserstoffbrücken zwischen den Basen aus. Komplementäre Nukleinsäurestränge sind immer antiparallel zueinander angeordnet, d. h. am Ende eines Nukleinsäuredoppelstrangs befindet sich das 5'-Ende des einen und das 3'-Ende des komplementären anderen Nukleinsäurestrangs.

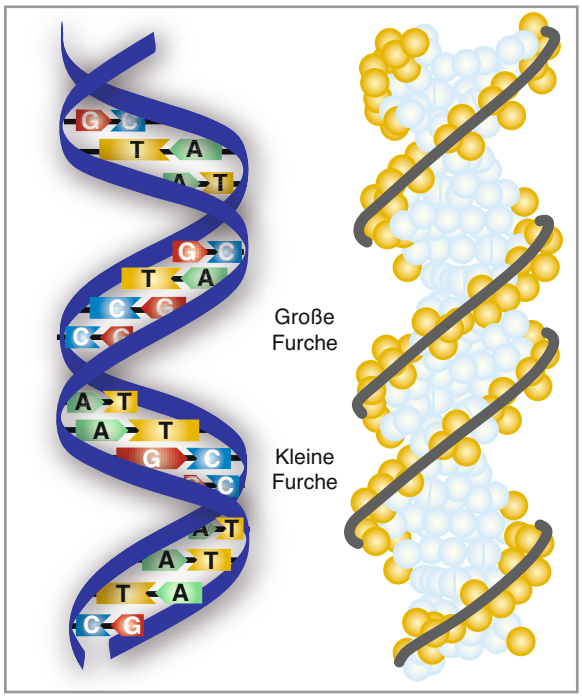


Abb. 2.5 DNA-Doppelhelix. Schematische Darstellung (links) und raumfüllende Darstellung aller Atome (rechts) einer DNA-Doppelhelix. Es wechseln sich immer eine große Furche und eine kleine Furche ab. Eine helikale Windung enthält ca. 10,5 Basenpaare und jedes Basenpaar ist um ca. 36° zueinander verdreht. Die Basenpaare sind im 90°-Winkel zur Helixachse angeordnet. Die Doppelhelix wird durch die parallele Anordnung der flachen Basenpaare signifikant über sogenannte *Stacking*-Kräfte zwischen den Basen stabilisiert.

als eines Organismus nichts über dessen Komplexität aussagt. Denn letztlich ist es die Abfolge der DNA-Basen in den einzelnen, für Informationen codierenden Einheiten (den Genen), die die Individualität eines Organismus bestimmen.

Durch Verdrillung der beiden komplementären DNA-Stränge wird die typische DNA-Doppelhelix gebildet (Abb. 2.5). Sie erinnert an eine sich nach rechts windende Wendeltreppe. Die Basenpaare bilden in diesem Bild die Stufen der Treppe. Das Rückgrat aus Phosphat- und Zuckermolekülen bildet gewissermaßen das Geländer. Jede helikale Windung der DNA-Doppelhelix besteht aus ca. 10,5 Basenpaaren, sodass jedes Basenpaar gegenüber seinem benachbarten Basenpaar um einen Winkel von ca. 36° verdreht ist. Die Basenpaare bilden einen Winkel von 90° zur Helixlängsachse. Durch diese typische Verdrillung entstehen eine kleine und eine große Furche (Abb. 2.5). Die beiden Doppelstränge werden ausschließlich über die angesprochenen Wasserstoffbrückenbindungen zwischen G/C- und A/T-Basenpaaren zusammengehalten. Wegen der Vielzahl dieser Wasserstoffbrückenbindungen entlang einer DNA-Doppelhelix ist der Energiegehalt so groß, dass eine DNA-Doppelhelix häufig als ein einziges Molekül angesehen wird, obwohl es sich formal um zwei, nicht kovalent miteinander verbundene Moleküle handelt.

2.1.2 Denaturierung von DNA

Die in der DNA-Doppelhelix angeordneten Einzelstränge lassen sich wie beim Öffnen eines Reißverschlusses voneinander trennen. Dieser Prozess wird bei

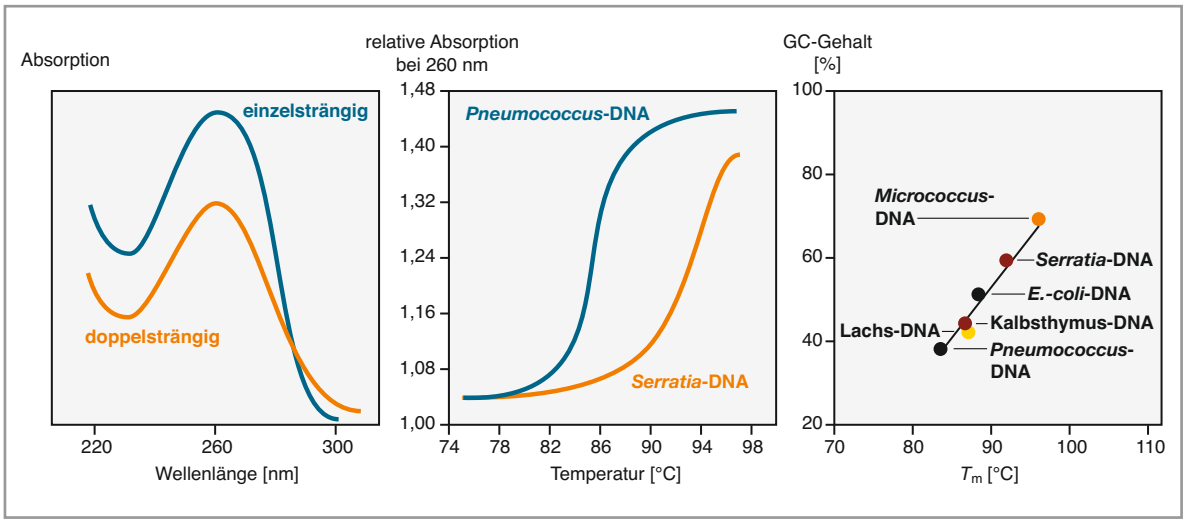


Abb. 2.6 Bestimmung des G/C-Gehalts eines Genoms. Links: Absorptionsspektren einzelsträngiger und doppelsträngiger DNA. Man erkennt den sogenannten Hyperchromizität-Effekt, also die Zunahme der Absorption bei 260 nm beim Übergang von doppelsträngiger zu einzelsträngiger DNA. Mitte: Den Hyperchromizität-Effekt kann man ausnutzen, um den G/C-Gehalt eines Genoms experimentell zu bestimmen. Dazu setzt man eine genomische DNA ansteigenden Temperaturen aus und misst die Absorption bei 260 nm. Ändert sich die Absorption bei steigender Temperatur nicht mehr, ist die DNA maximal denaturiert. Aus der Kurve kann man die halbmaximale Denaturierungstemperatur T_m ableiten. Rechts: Der G/C-Gehalt eines Genoms steht in linearer Beziehung zu T_m. Trägt man den Messwert T_m gegen die bereits zuvor bestimmten G/C-Gehalte anderer Organismen auf, kann man den G/C-Gehalt eines untersuchten Organismus auf der Geraden ablesen.

der Replikation und Transkription durch spezielle Enzyme, die DNA-Helikasen, vermittelt. Doch auch durch Erhitzen und durch Zugabe bestimmter Chemikalien ist eine Aufspaltung von DNA-Doppelsträngen experimentell möglich. Diesen Vorgang nennt man in der Gentechnik „Denaturieren" von DNA.

Die Stabilität der DNA hängt vom prozentualen Anteil an G/C-Nukleotidpaaren ab, da diese drei Wasserstoffbrücken ausbilden, während A/T-Paare nur über zwei Wasserstoffbrücken miteinander wechselwirken (Abb. 2.4). Die Temperatur, bei der 50 % einer doppelsträngigen DNA denaturiert vorliegen, ist die Schmelztemperatur T_m. Je größer der G/C-Gehalt einer DNA ist, desto höher liegt ihr T_m-Wert.

Die Gesamtheit der im Zellkern eukaryontischer Zellen vorhandenen DNA ist das Genom. Man kann den G/C-Gehalt ganzer Genome experimentell bestimmen (Abb. 2.6). Hierfür wird das als Hyperchromizität bezeichnete Phänomen ausgenutzt, dass einzelsträngige DNA eine höhere Absorption bei 260 nm zeigt als doppelsträngig vorliegende DNA. Setzt man die genomische DNA eines Organismus steigenden Temperaturen aus, wird die DNA zunehmend denaturiert und man misst eine ansteigende Absorption. Aus der sich ergebenden Kurve lässt sich dann der T_m-Wert ablesen. Vergleicht man diesen gemessenen T_m-Wert mit dem anderer Organismen, deren G/C-Gehalte vorher schon einmal bestimmt wurden, kann man den G/C-Gehalt des untersuchten Genoms an einer Standardgeraden ablesen (Abb. 2.6). Heute wird der G/C-Gehalt vieler Organismen im Rahmen von Genomprojekten – also der Aufklärung der DNA-Sequenzen von Gesamt-Genomen – quasi nebenbei bestimmt.

Ebenso wie DNA-Doppelstränge denaturiert werden können, lassen sich komplementäre Einzelstränge auch wieder zu einer Doppelhelix zusammenfügen. Dieser Vorgang wird „Renaturierung" genannt. Die Kinetik, mit der komplementäre DNA-Stränge renaturieren, erlaubt Rückschlüsse auf den G/C-Gehalt und auf die Komplexität der DNA-Moleküle. Erstaunlicherweise erfolgt die Renaturierung auch in den komplexesten Gemischen verschiedener DNA-Spezies hochspezifisch und effizient. Dabei bildet sich immer dann – und nur dann – ein DNA-Doppelstrang aus, wenn zwei komplementäre Partner in einem DNA-Gemisch vorhanden sind. Dieses Prinzip wird uns in vielen Bereichen der Gentechnik und der Gendiagnostik wieder begegnen, wie zum Beispiel bei der Initiation enzymatischer DNA-Synthese durch sogenannte Primer oder dem spezifischen Nachweis von DNA-Sequenzen in der Gendiagnostik.

2.1.3 Größe und Topologie der DNA

Die Größe der DNA-Moleküle stellt beim gentechnischen Arbeiten mit isolierter DNA eine besondere Herausforderung dar (Abb. 2.7). Ein Genom wie das des Darmbakteriums *Escherichia coli* besteht aus einem einzigen Molekül in der Größenordnung von ca. 4,6 Milli-

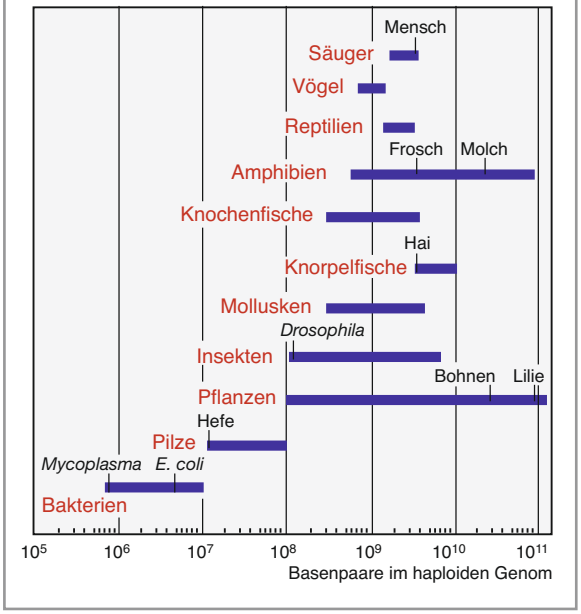


Abb. 2.7 Vergleich der Genomgrößen verschiedener Organismen.

onen Basenpaaren. Dieses Molekül ergibt gespreizt eine Kantenlänge von ca. 1,5 mm. Die abgekürzte Schreibweise für die Länge einer DNA in Basenpaaren ist „bp" (*base pairs*). Das Genom von *E. coli* ist demnach ca. 4 600 000 bp oder 4600 kbp oder 4,6 Mbp groß. Das menschliche Genom ist ungefähr 600-mal größer als das von *E. coli*. Es besteht aus 23 einzelnen Molekülen (Chromosomen), die zusammen etwa 3 270 000 000 bp (3270 Mbp) enthalten und gespreizt ca. 99 cm lang sind. Da die Zellen diploider Organismen jeweils zwei Chromosomensätze enthalten, befindet sich in jeder einzelnen menschlichen Zelle die genetische Information auf DNA-Molekülen, die aneinandergereiht ca. 2 m lang sind.

DNA-Moleküle gibt es in verschiedenen Formen. Sie können fadenförmig linear sein, wie beispielsweise die DNA einiger Viren, die Bakterien befallen (z. B. die Bakteriophagen T7 und T4). Die DNA der Bakteriophagen M13 und f1, die als Vektorsysteme eine Rolle in der Gentechnik spielen, ist ein ringförmiger DNA-Einzelstrang, der nach der Infektion von den Enzymen der Wirtszelle zu einem DNA-Doppelstrang komplettiert wird. Dagegen liegt die DNA des Bakteriophagen Lambda in einem linearen, doppelsträngigen DNA-Molekül vor, das erst nach der Infektion ringförmig geschlossen wird. Bakterielle Genome sind in der Regel ringförmig geschlossene, doppelsträngige DNA-Moleküle, während sich die genomische DNA im Zellkern der Eukaryonten auf mehrere lineare Moleküle, die Chromosomen, verteilt. Allerdings kommen in Eukaryonten zusätzlich auch noch ringförmige DNA-Moleküle in Form der DNA der Mitochondrien und der Chloroplasten vor.

Ringförmige DNA-Moleküle wie beispielsweise die Genome von Bakterien, Mitochondrien oder Chloroplasten haben eine wichtige topologische Besonderheit: Die beiden DNA-Stränge der Doppelhelix können nicht vollständig voneinander getrennt werden. Auch lineare DNA-Moleküle sind ab einer gewissen Länge nicht mehr vollständig in zwei Einzelstränge trennbar. Ringförmig geschlossene DNA liegt zudem meist verdrillt vor (Abb. 2.8). Diese Verdrillungen überlagern die eigentliche DNA-Doppelhelix und sind als Superhelices sichtbar. Die Zahl der Verdrillungen (*Supercoils*) ist dynamisch und kann durch zelluläre Enzyme beeinflusst werden. DNA-Moleküle, die bei identischer DNA-Sequenz in verschiedenen Konformationen vorliegen, nennt man DNA-Topoisomere. Dementsprechend nennt man Enzyme, die die Konformation einer DNA verändern können, DNA-Topoisomerasen.

Wir finden *Supercoils* nicht nur in zirkulären DNA-Molekülen, sondern auch innerhalb von Teilbereichen sehr großer, linearer Moleküle, beispielsweise der Chromosomen von Säugetierzellen: Die Verdrillungen entstehen im Prinzip durch eine Drehung der DNA-Doppelhelix um ihre Längsachse. Dies führt zu Drehungsspannungen, die sich in der Bildung einer Superhelix äußern. Anders ausgedrückt entstehen *Supercoils*, wenn man das eine Ende der DNA um ihre Längsachse dreht, während das andere Ende festgehalten wird. Eine Konsequenz der *Supercoils* ist, dass ein DNA-Doppelstrang nicht über längere Strecken in Einzelstränge getrennt werden kann, da die Drehungsspannung innerhalb der Superhelix dem Prozess entgegenwirkt.

In der Zelle kann die Überdrillung einer DNA nicht dadurch verringert werden, dass die DNA-Doppelhelix um ihre Längsachse rotiert. Dazu ist das DNA-Molekül in der Regel schlicht zu lang. Damit es zur Einführung oder Entfernung einer Überdrillung und zur Ausbildung superhelikaler Formen kommen kann, muss an bestimmten Stellen der DNA mindestens einer der beiden DNA-Stränge hydrolytisch geöffnet werden. Dieser geöffnete Strang kann sich nun lokal frei um den intakten Strang drehen und damit den Verdrillungsgrad ändern. Dieser Vorgang wird durch DNA-Topoisomerasen katalysiert, die somit eine große Rolle bei der Transkription, Rekombination und Replikation spielen und für jede Zelle überlebenswichtig sind. Aus diesem Grund sind sie auch wichtige therapeutische Ziele für Antibiotika, wie beispielsweise die Gyrase-Inhibitoren. Inaktiviert man die bakterielle Topoisomerase durch einen Gyrase-Hemmer, werden die DNA-Replikation und andere wichtige Stoffwechselprozesse pathogener Prokaryonten rasch blockiert.

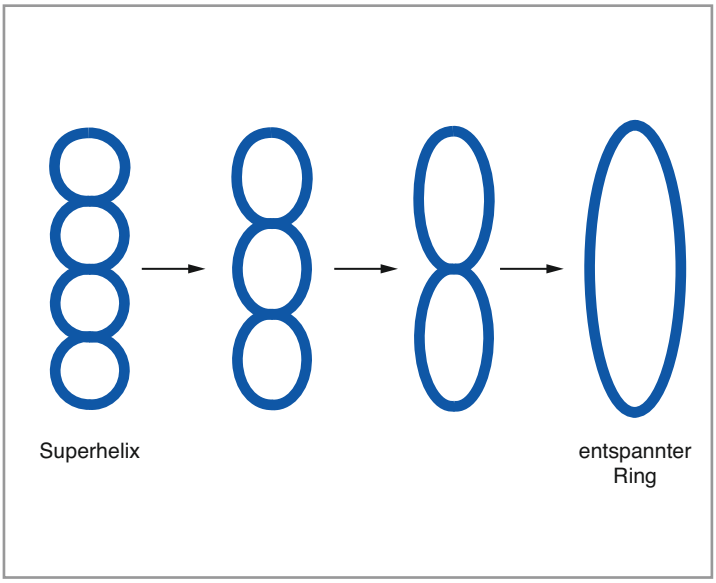


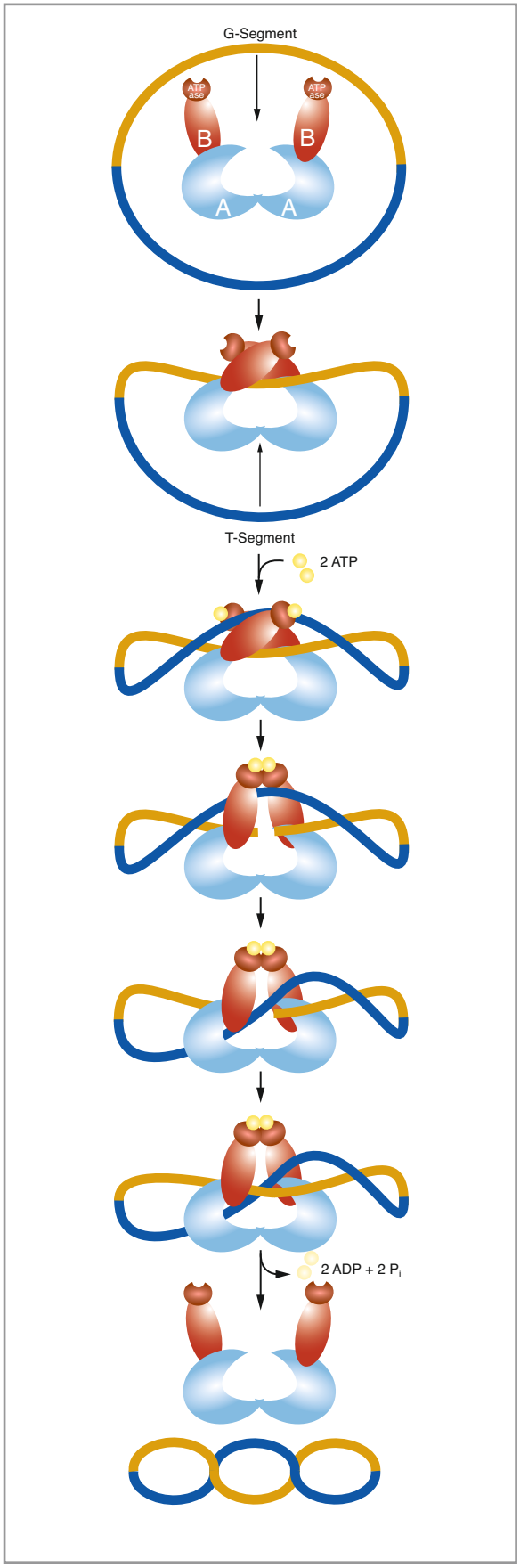
Abb. 2.8 DNA-Topoisomere. Ringförmig geschlossene, doppelsträngige DNA-Moleküle liegen häufig verdrillt als Superhelix vor. Die Verdrillungen verhindern, dass die beiden Stränge einer DNA-Helix über eine längere Distanz hinweg zu Einzelsträngen geöffnet werden können, wie es beispielsweise für die Replikation oder Transkription notwendig ist. Um Verdrillungen aus der DNA zu entfernen, öffnen Typ-1-Topoisomerasen einen der beiden DNA-Stränge, drehen diesen Strang einmal um die helikale Achse und verknüpfen die freien DNA-Enden wieder kovalent. Dadurch werden sukzessive Windungen aus der DNA entfernt. Typ-2-Topoisomerasen spalten beide DNA-Stränge gleichzeitig und fädeln den intakten DNA-Doppelstrang durch den entstandenen DNA-Doppelstrangbruch, sodass ebenfalls Verdrillungen entfernt werden.

Es gibt im Wesentlichen zwei Typen von DNA-Topoisomerasen:

- Typ-1-DNA-Topoisomerasen binden an die DNA und öffnen *einen* der beiden DNA-Stränge der Doppelhelix. Es lassen sich zwei Subtypen von Typ-1-DNA-Topoisomerasen unterscheiden:
  - Die bakteriellen Typ-1A-Topoisomerasen spalten einen DNA-Strang, indem sie eine Phosphodiesterbindung öffnen und eine Phosphatbrücke zwischen dem 5'-Ende der DNA und einem Tyrosinrest im aktiven Zentrum des Enzyms bilden. Durch diese kovalente Bindung an das Enzym wird die Energie der Phosphodiesterbindung konserviert und der geöffnete DNA-Strang wird am Enzym fixiert. Das Enzym veranlasst eine vollständige Drehung der DNA an der Stelle der Strangöffnung um seine eigene zentrale Achse und entfernt somit eine superhelikale Windung. Durch Transesterifizierung wird dann unter Freisetzung des Enzyms die 5'-Phosphatgruppe wieder auf die freie 3'-OH-Gruppe des gespaltenen DNA-Strangs übertragen und dadurch der zuvor geöffnete DNA-Strang wieder geschlossen. Dieser Vorgang benötigt keine chemische Energie in Form von ATP, da bei der Übertragung des DNA-Strangs auf den Tyrosinrest des Enzyms keine Energie verloren geht.
  - Typ-1B-Topoisomerasen kommen in Eukaryonten vor. Sie spalten ebenfalls einen der beiden Stränge einer DNA, allerdings übertragen sie dabei eine 3'-Phosphatgruppe auf einen Tyrosinrest im aktiven Zentrum des Enzyms. Nach der Rotation um die Längsachse der DNA wird die 3'-Phosphatgruppe wieder auf die verbliebene 5'-OH-Gruppe des gespaltenen DNA-Strangs zurück übertragen und der DNA-Strang dadurch geschlossen.
- Eine Typ-2-DNA-Topoisomerase bindet an die DNA und öffnet *beide* DNA-Stränge gleichzeitig. Die bakterielle Typ-2-DNA-Topoisomerase, die auch als Gyrase bekannt ist, besteht aus einem Tetramer aus jeweils zwei GyrA- und GyrB-Untereinheiten. Gyrase wirkt über folgende Teilschritte (Abb. 2.9): Zunächst bindet das Enzym an die DNA, wobei sich eine Schleife der DNA von ca. 140 Basenpaaren um das Enzym schlingt. Dann spaltet das Enzym den DNA-Doppelstrang und die GyrA-Untereinheit des Enzyms wird kovalent an das 5'-Ende des gespaltenen DNA-Strangs gebunden. Als nächster Schritt wird ein benachbart gelegenes DNA-Segment durch die Öffnung gezogen. Schließlich wird das kovalent gebundene Protein wieder abgelöst und die gespaltenen Doppelstränge werden wieder verknüpft. Bei dieser Reaktionsfolge wird unter Verbrauch eines Moleküls Adenosintriphosphat (ATP) eine superhelikale Windung in die DNA eingeführt, wie in Abb. 2.9 dargestellt. In Abwesenheit von ATP kann das Enzym auch die umgekehrte Reaktion katalysieren und somit zur Relaxation superhelikaler DNA führen.

2.1.4 DNA-Analytik

Beim gentechnischen Arbeiten wird ständig mit DNA umgegangen. Dabei werden verschiedene Standardmethoden der DNA-Analytik angewendet, die wir hier kurz besprechen wollen. Als geladene Makromoleküle können DNA oder RNA durch Gelelektrophorese getrennt werden. In der Regel werden dazu Matrices aus gelierter Agarose oder aus einem Polyacrylamid-



Netzwerk verwendet. Der Logarithmus der relativen Mobilität des Makromoleküls ist linear abhängig von der Gelkonzentration und kann durch einen sogenannten Ferguson-Plot dargestellt werden (o Abb. 2.10). Dabei trennt man gleiche Proben in Gelen unterschiedlicher Porengröße auf. Trägt man dann die Logarithmen der relativen Laufstrecken gegen die eingesetzten Gelkonzentrationen auf, erhält man eine Gerade. Die Steigung der Geraden im Ferguson-Plot ist damit ein Maß für die Molekülgröße.

Polyelektrolyte wie Nukleinsäuren tragen an jeder Monomereinheit eine Ladungseinheit. Daraus folgt, dass das Verhältnis der Molekülmasse der Nukleinsäure zur Gesamtladung stets konstant ist. Für die relative Mobilität in Gelmatrices ist somit der Siebeffekt – und damit die Moleküllänge – bestimmend (o Abb. 2.10). Über einen relativ weiten Größenbereich gibt es eine lineare, experimentell bestimmbare Abhängigkeit zwischen dem Logarithmus ($\log_{10}$) der Länge (in Basenpaaren) von DNA-Molekülen und ihrer relativen Wanderungsdistanz (in cm).

In der Praxis verwendet man Agarosegele von 0,5 % bis ca. 2 %. In 0,5%igen Gelen lassen sich DNA-Fragmente im Bereich von 1000–15 000 bp gut auftrennen. Gele mit einer höheren Agarosekonzentration sind geeignet, um DNA-Fragmente zwischen 100 und 2000 bp zu trennen. Trennt man ringförmig geschlossene Plasmide, so wirkt sich die Konformation der DNA-Moleküle stärker auf die relative Mobilität aus als bei der Trennung linearer Nukleinsäurefragmente. Ringförmige, superhelikale DNA wandert nämlich schneller als ringförmig offene (relaxierte) oder lineare DNA. Das kommt daher, dass sich die dichter gepackte superhelikale DNA schneller durch die Poren des Agarosegels bewegen kann (o Abb. 2.11).

Nachweisen lassen sich die DNA-Fragmente in Gelmatrices beispielsweise durch Anfärbung mit Ethidiumbromid (o Abb. 2.11). Einzelne Ethidiumbromid-Moleküle interkalieren zwischen die DNA-Basen, wobei die aromatischen Ringe des Farbstoffs mit den heteroaromatischen Ringen der Nukleinsäurebasen wechselwirken. Der Farbstoff zeigt nach Anregung mit UV-Licht

o **Abb. 2.9** Reaktionsmechanismus der Gyrase. Gyrase ist eine Typ-2-Topoisomerase, die nacheinander zwei doppelsträngige DNA-Bereiche (G- und T-Segment) bindet. Im Verlauf weiterer Konformationsänderungen des Enzyms, für die zwei Moleküle ATP benötigt werden, wird das G-Segment durch die GyrA-Untereinheit des Enzyms gespalten. Das T-Segment wird dann durch die Lücke im G-Segment gefädelt, sodass – wie in diesem Beispiel gezeigt – eine Verdrillung eingeführt wird. Das Enzym kann auf gleiche Weise auch Verdrillungen aus der DNA entfernen.

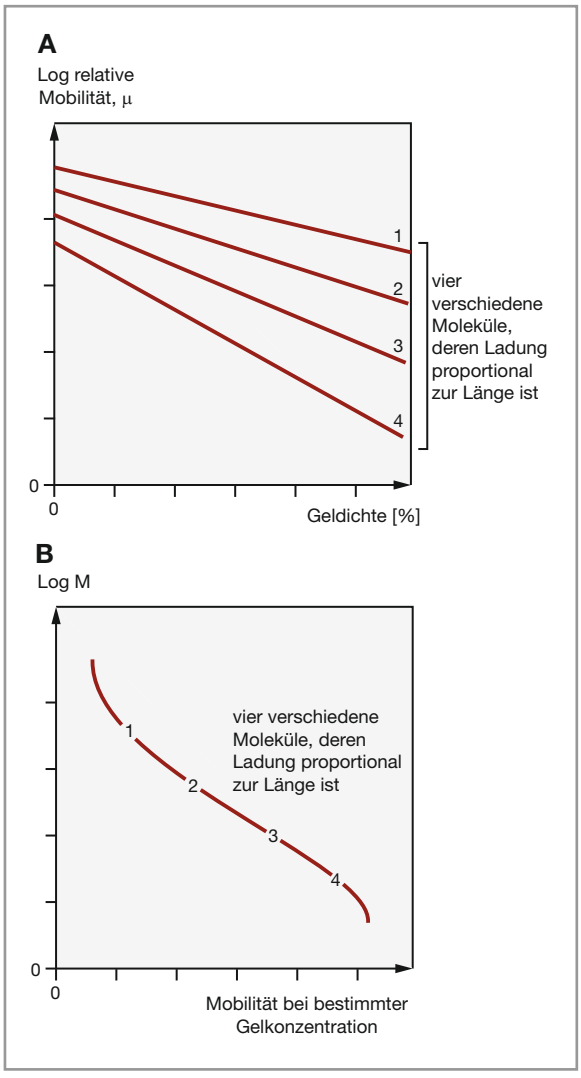


Abb. 2.10 Ferguson-Plot für verschiedene Polyelektrolyte. **A** Polyelektrolyte wie Nukleinsäuren tragen an jeder Monomereinheit eine Ladungseinheit. Demnach tragen verschieden große Nukleinsäuren eine Ladung, die zu der jeweiligen Moleküllänge proportional ist. Für die relative Mobilität in Gelmatrices ist somit der Siebeffekt – und damit die Moleküllänge – bestimmend. **B** Bei einer konstanten Gelkonzentration ist in einem bestimmten Bereich der Logarithmus der Molekülmasse (Log M) proportional zur Mobilität der Moleküle in der Gelmatrix.

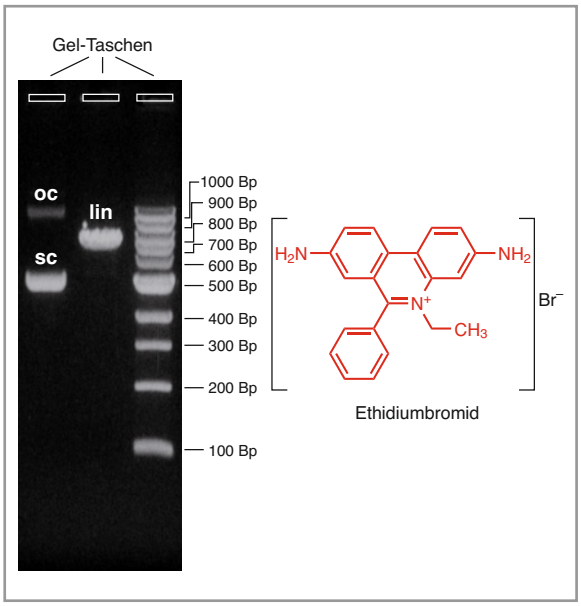


Abb. 2.11 Gelelektrophoretische Trennung von DNA-Topoisomeren. Die DNA-Fragmente werden in „Taschen" am oberen Gelrand aufgetragen. Dann wird an das Gel eine Gleichspannung angelegt, wobei die Anode unten und die Kathode oben angeschlossen wird. Als Polyanionen wandern die DNA-Fragmente nach unten auf die Anode zu, wobei kleinere Fragmente schneller wandern als größere. Zum Nachweis der DNA enthält das Agarosegel gleichmäßig verteilt Ethidiumbromid. Während der Elektrophorese „sammelt" die DNA Ethidiumbromid-Moleküle auf, da diese bei Kontakt mit DNA zwischen die Basenpaare interkalieren. Durch die Interkalation wird die Fluoreszenz des Ethidiumbromids stark erhöht, sodass im UV-Licht nur dort eine intensiv orange Fluoreszenz sichtbar ist, wo sich im Gel DNA befindet. In der rechten Spur ist ein DNA-Größenstandard aufgetragen worden. In der linken Spur lief eine Präparation von Plasmid-DNA. Die ringförmigen Plasmide liegen entweder als entspannte Ringe (*open circle*, oc) oder als superspiralisierte Form (*supercoiled*, sc) vor. Da die *Supercoiled*-DNA wesentlich kompakter ist als entspannte DNA-Ringe, läuft sie im Agarosegel schneller. In der mittleren Spur haben wir die Plasmid-DNA vorher mit einer Restriktionsendonuklease geschnitten, sodass ein linearer DNA-Strang (lin) resultiert. Die Mobilität dieser DNA liegt zwischen der *open circle* und *supercoiled*-Konformation.

(254–366 nm) eine Fluoreszenz bei 590 nm. Da die Fluoreszenz des Farbstoffs nach der Bindung an DNA stark zunimmt, ist die Orange-Fluoreszenz nur an den Stellen eines Agarosegels deutlich sichtbar, an denen sich DNA befindet, selbst wenn der Farbstoff zuvor gleichmäßig in das Agarosegel eingegossen wurde.

Neben der Größenbestimmung ist ein weiterer wichtiger Parameter im Umgang mit DNA die DNA-Konzentration in einer DNA-Lösung. Nukleinsäuren absorbieren im UV-Bereich mit einem Maximum bei 260 nm (angegeben als A_{260}). Demzufolge erfolgt die Bestimmung der DNA-Konzentration in der Regel durch UV-Spektroskopie. Aufgrund des Phänomens der Hyperchromizität (Einzelstrang-DNA absorbiert stärker als Doppelstrang-DNA) ergeben sich aus der UV-spektroskopischen Messung verschiedener Nukleinsäure-Typen unterschiedliche Konzentrationen. Danach entspricht 1 A_{260}

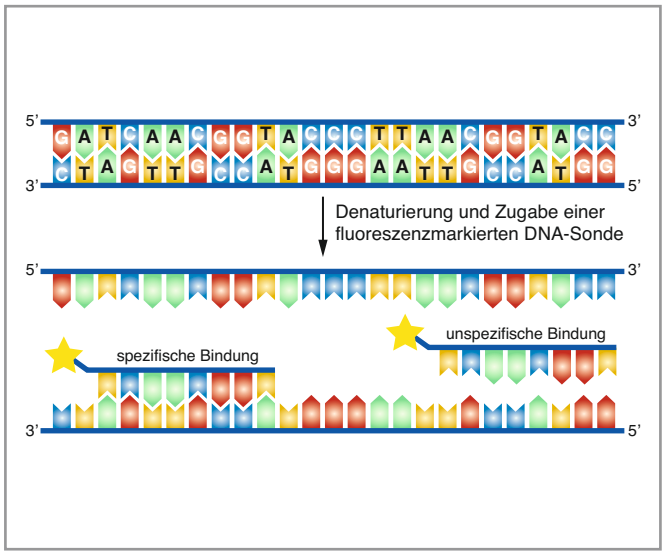


Abb. 2.12 Prinzip der Hybridisierung. Die vier DNA-Basen sind durch die unterschiedlichen Formen und Farben der Symbole dargestellt. Durch Erhitzen wird der DNA-Doppelstrang in die beiden Einzelstränge denaturiert. Wir geben ein kurzes Oligonukleotid hinzu, das für einen kleinen Bereich eines der beiden DNA-Doppelstränge exakt komplementär ist und am 5'-Ende mit einem Fluoreszenzfarbstoff gekoppelt ist. Die markierte Sonde kann über Basenpaarungen an den denaturierten DNA-Strang binden. Durch Waschen mit geeigneten Puffern und bei steigenden Temperaturen entfernt man die unspezifisch gebundenen Sonden (die Fehlpaarungen aufweisen) früher als die spezifisch gebundenen Proben (weil jede Base paaren kann). Der große Überschuss der zugegebenen Sonde sorgt in der Konkurrenz mit der Reassoziation der beiden DNA-Stränge für eine zuverlässige Bindung der Sonde.

- 50 µg/ml Doppelstrang-DNA,
- 40 µg/ml Einzelstrang-DNA,
- oder 33 µg/ml (Einzelstrang)-RNA.

Bezüglich der Reinheit einer gegebenen DNA sind Verunreinigungen mit anderen DNAs oder mit Proteinen von Interesse. Verunreinigungen einer DNA mit anderen DNAs können in Grenzen durch Agarosegel-Elektrophorese nachgewiesen werden. Die Nachweisbarkeitsgrenze für eine einzelne Bande in Ethidiumbromid-Gelen liegt bei ca. 10 ng Doppelstrang-DNA. Verunreinigungen mit Proteinen können über den Quotienten A_{260}/A_{280} abgeschätzt werden. Proteine absorbieren über ihre aromatischen Aminosäuren bei einem Maximum von 280 nm. Eine „saubere" DNA sollte einen A_{260}/A_{280}-Quotienten nahe 1,8 aufweisen.

Die Nachweisbarkeitsgrenze der UV-spektroskopischen Vermessung von DNA liegt bei ca. 250 ng/ml. In vielen Fällen hat man es mit wesentlich weniger DNA zu tun, sodass empfindlichere Methoden zur Bestimmung einer DNA-Konzentration erforderlich sind. Eine solche Methode basiert auf der DNA-Hybridisierung. Wir haben bereits besprochen, dass man DNA-Doppelstränge durch die Zufuhr von Energie in Einzelstränge überführen kann. Dieser Vorgang wird Denaturierung genannt. Senken wir die Temperatur wieder, werden die Einzelstränge renaturieren. Wenn wir nun während der Renaturierung ein kurzes DNA-Molekül im Überschuss zusetzen, das zu einem bestimmten Teil der denaturierten DNA komplementär ist, kann das kurze Oligonukleotid mit einem der längeren DNA-Einzelstränge bei der Bildung von Doppelsträngen konkurrieren. Bindet auf diese Weise gezielt ein kurzes DNA-Molekül an einen längeren DNA-Einzelstrang, nennen wir diesen Vorgang Hybridisieren. Die Hybridisierung von Nukleinsäuren ist eine fundamentale Methode der Gentechnologie. Wir verwenden sie beispielsweise zum qualitativen Nachweis von DNA, für die Konzentrationsbestimmung von DNA und um die enzymatische In-vitro-DNA-Synthese mit sogenannten Primern zu starten.

Wie können wir die Hybridisierung einsetzen, um die Konzentration einer DNA-Lösung zu bestimmen? Stellen wir uns vor, dass wir von der DNA-Lösung, deren Konzentration bestimmt werden soll, eine Serienverdünnung herstellen und jeweils ein Aliquot davon auf eine Nitrocellulosemembran tropfen (*dot blot*). Außerdem tropfen wir auf das Filter eine Konzentrationsreihe einer DNA, deren Konzentration uns bekannt ist (Standard-Reihe). Stellen wir uns weiter vor, dass wir ein kurzes DNA-Oligonukleotid zur Verfügung haben, das an einem Ende mit einem Fluoreszenzfarbstoff gekoppelt ist (Abb. 2.12). Wir inkubieren die Nitrocellulosemembran mit einer Lösung des markierten Oligonukleotids. Mehrere Waschschritte in Puffern mit definiertem Salzgehalt und bei definierten Temperaturen ermöglichen die Unterscheidung zwischen spezifisch und unspezifisch gebundener DNA-Sonde: Eine Sonde, die nicht „passt" (in Abb. 2.12 rechts), wird bei geringeren Temperaturen „abgewaschen" als die Sonde, die über die gesamte Länge spezifische Basenpaarungen eingehen kann. Nach den Waschschritten wird die Fluoreszenz des spezifisch an der Proben-DNA gebundenen Oligonukleotids gemessen und mit der Standardreihe in Beziehung gesetzt, um die Konzentration der Ausgangslösung zu errechnen.

Zusammenfassung

- Das Zielmolekül aller gentechnischen Methoden ist die DNA. Der Begriff DNA beschreibt allgemein eine Molekülart, während die Einzigartigkeit eines DNA-Moleküls durch die Sequenz der vier DNA-Basen bestimmt wird.
- DNA besteht aus drei Grundbausteinen: einer Purin- oder Pyrimidinbase, einer *N*-glycosidisch daran gebundenen Desoxyribose und einem Phosphatrest, der esterartig mit dem Zucker verbunden ist. Der Zucker und das Phosphat bilden das Rückgrat der DNA.
- Die DNA-Basen können mit einem anderen DNA-Strang komplementärer Basen-Sequenz einen DNA-Doppelstrang ausbilden, wobei die beiden DNA-Stränge durch die Bildung von Wasserstoffbrückenbindungen zwischen komplementären Basen stabil miteinander verbunden werden. Die Einzelstränge in einer DNA-Doppelhelix verlaufen antiparallel und haben definierte Enden, die als 5'-Ende bzw. als 3'-Ende bezeichnet werden.
- Wegen der Komplementarität enthalten DNA-Moleküle immer ebenso viele A- wie T-Nukleotide und ebenso viele G- wie C-Nukleotide. Allerdings kann die Anzahl der A/T- und der G/C-Paare je nach Organismus deutlich variieren.
- Da die beiden komplementären DNA-Stränge nur durch schwache Wechselwirkungen zusammengehalten werden, lassen sie sich durch Energiezufuhr trennen („denaturieren").
- Die Reorganisation zweier komplementärer Einzelstränge zu einem DNA-Doppelstrang wird als „renaturieren" bezeichnet. Denaturierung und Renaturierung bilden die Basis für die wichtige gentechnische Technik des Hybridisierens.

2.2 Chemische DNA-Synthese

Der Umgang mit DNA in der Gentechnik erfordert oft die In-vitro-Synthese von DNA. Dabei imitieren wir die natürliche DNA-Replikation und machen uns Enzyme zunutze, die in der Lage sind, einzelsträngige DNA-Abschnitte zu kopieren und zu einem neuen DNA-Doppelstrang aufzubauen. Ein wichtiger Aspekt der „enzymatischen DNA-Synthese" ist die Tatsache, dass die verwendeten DNA-Polymerasen zwar eine einzelsträngige DNA als Matrize für den Kopiervorgang benötigen (▸ Kap. 2.3), allerdings einen kurzen Bereich doppelsträngiger DNA als Starter der DNA-Synthese benötigen. Soll also im Labor ein beliebiger DNA-Abschnitt enzymatisch kopiert werden, muss die Startstelle künstlich festgelegt werden. Dafür dienen sogenannte Primer, die an den Anfang des zu kopierenden DNA-Abschnitts hybridisieren. Die Primer sind kurze, einzelsträngige DNA-Moleküle (Oligonukleotide), die sich durch chemische DNA-Synthese an Syntheseautomaten herstellen lassen. Viele molekularbiologisch orientierte Firmen haben sich heute auf die kommerzielle Primer-Synthese spezialisiert, sodass man beliebige Primer-Sequenzen im heimischen Labor am Computer entwerfen und eine entsprechende Firma via Internet mit der Synthese beauftragen kann.

Ausgangspunkt für eine chemische Synthese eines DNA-Oligonukleotids ist ein ausreichend geschütztes Nukleosid, das mit seiner 3'-OH-Gruppe über einen *Spacer* und eine Dicarbonsäure kovalent an eine feste Matrix gebunden ist. Die Matrix besteht in der Regel aus Silica-Gel oder aus besonderen Glas-Kügelchen (CPG, *controlled pore glass material*). Die Synthese erfolgt in Zyklen, wobei jeder Zyklus aus einer Reihe von Syntheseschritten besteht, die erforderlich sind, um das matrixgebundene Nukleotid bzw. Oligonukleotid um einen Nukleotidbaustein zu verlängern (○ Abb. 2.13). Anders als die biologische DNA-Synthese wird bei der chemischen DNA-Synthese die Nukleotidkette vom 3'- zum 5'-Ende hin verlängert.

Der Vorteil der matrixgebundenen Oligonukleotidsynthese besteht darin, dass nach jeder chemischen Reaktion die überschüssigen Reagenzien durch intensives Waschen der Säule von dem immobilisierten Produkt getrennt werden können. Folgende Einzelschritte werden im Lauf eines Synthesezyklus bei der heute überwiegend verwendeten Desoxynukleotid-Phosphoramidit-Synthese durchlaufen (○ Abb. 2.13):

- Im ersten Schritt wird die Dimethoxytrityl-Schutzgruppe (DMT) am 5'-Ende der mit der Matrix verbundenen Oligonukleotidkette durch Säurebehandlung entfernt.
- Das freigewordene 5'-Ende des Oligonukleotids wird an das 3'-Phosphoramidit-Derivat eines Desoxyribonukleosids gekoppelt, das die nächste Position in der Kette einnehmen soll. Das Kopplungsreagenz in dieser Reaktion ist Tetrazol.
- Alle 5'-Enden, die nicht reagiert haben, werden durch Acetylierung (*capping*) inaktiviert, um sie von den folgenden Kopplungsreaktionen auszuschließen. Dadurch wird die Verlängerung „falscher" Oligonukleotide unterbunden.

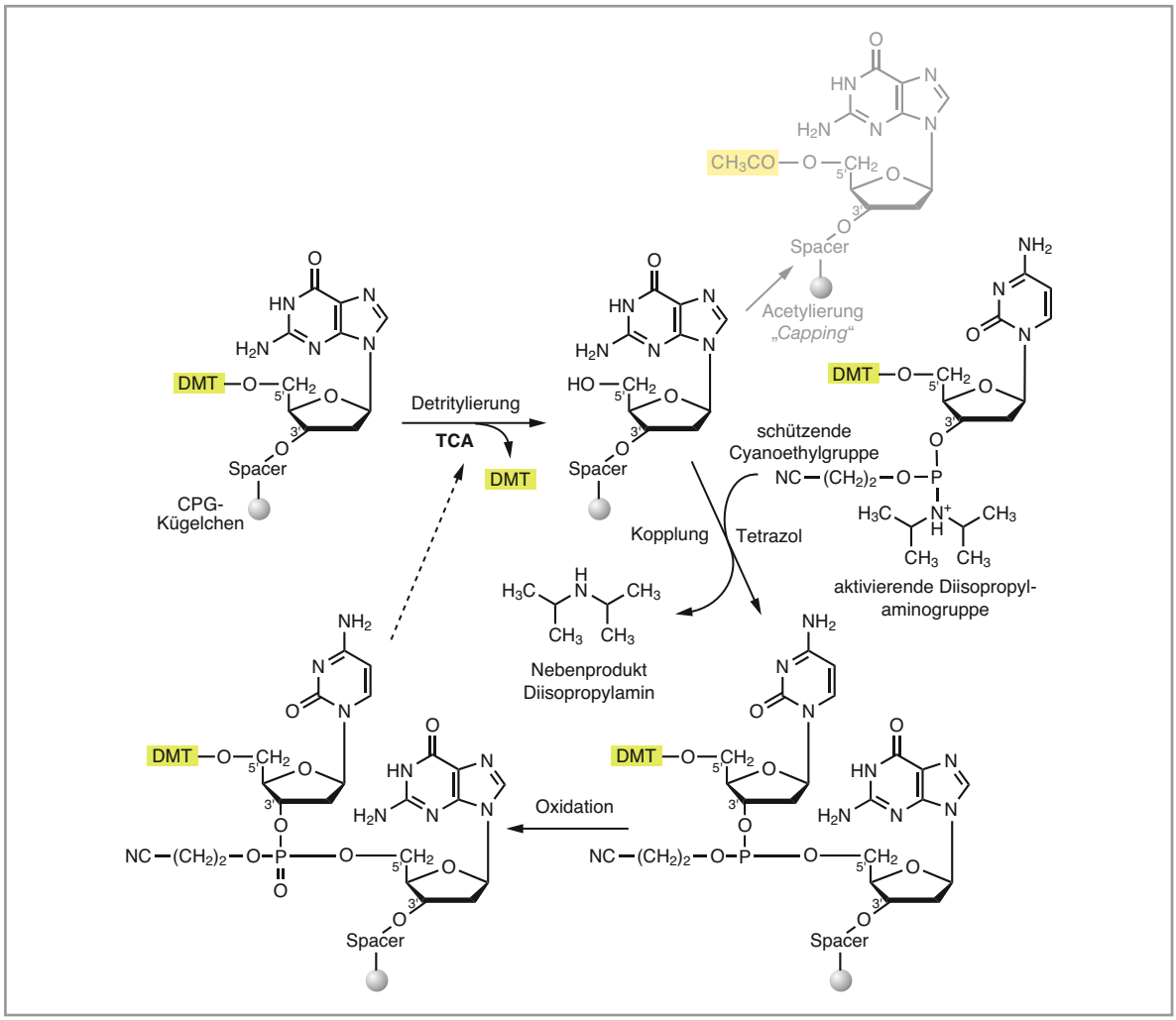


Abb. 2.13 Chemische DNA-Synthese nach dem Phosphoramidit-Verfahren. Es handelt sich bei dieser Methode um eine matrixgebundene Oligonukleotidsynthese. Der Zyklus beginnt mit der Detritylierung. Dabei wird die Dimethoxytrityl-Schutzgruppe (DMT) am 5'-Ende der Oligonukleotidkette, die mit der Matrix (CPG: *controlled pore glass material*) verbunden ist, in Gegenwart von Trichloressigsäure (TCA) abgespalten. Im folgenden Schritt wird das 3'-Phosphoramidit-Derivat, das die nächste Position in der Kette einnehmen soll, an das entschützte 5'-Ende des matrixgebundenen Oligonukleotids gekoppelt. Das Kopplungsreagenz in dieser Reaktion ist Tetrazol. Alle 5'-Enden, die nicht reagiert haben, werden durch Acetylierung (*capping*) inaktiviert, um sie von den folgenden Kopplungsreaktionen auszuschließen (grauer Nebenweg). Der bei der Kopplungsreaktion entstandene Phosphit-Triester wird schließlich zum Phosphat-Triester oxidiert.

- Der bei der Kopplungsreaktion entstandene Phosphit-Triester wird zum Phosphat-Triester oxidiert. Dabei entsteht die um ein Nukleotid verlängerte Kette.
- Dieser Reaktionszyklus kann heute in modernen Syntheseautomaten routinemäßig bis zu 100 Mal wiederholt werden.
- Zum Abschluss einer Oligonukleotidsynthese wird das Oligonukleotid von der Festphase abgelöst und die diversen Schutzgruppen werden entfernt. Danach sollte das Oligonukleotid durch eine HPLC-Trennung gereinigt werden, um zu kurze Oligonukleotide abzutrennen und damit die nachfolgenden enzymatischen Reaktionen nicht durch verunreinigende Chemikalien in ihrer Effizienz herabgesetzt werden.

Trotz der vielen Einzelschritte sind die Ausbeuten erstaunlich gut, da die Effizienzen der Einzelreaktionen für chemische Reaktionen so extrem hoch sind. Trotzdem können chemisch hergestellte Oligonukleotide nicht beliebig lang sein, da die kumulativen Ausbeuten der einzelnen Synthesezyklen für das Endprodukt maßgebend sind. Dies wird an folgenden Beispielrechnungen deutlich:

- Wenn ein 20-mer-Oligonukleotid hergestellt wird, haben selbst bei 99 % Kopplungseffizienz der Einzelzyklen nur 82 % ($0{,}99^{20} \bullet 100$) die korrekte Länge von

20 Nukleotiden. Wäre die Kopplungseffizienz jedoch in jedem Schritt nur 90 %, hätten nur noch 12 % der Produkte die korrekte Länge.
- Die Ausbeute für ein 60-mer-Oligonukleotid liegt selbst bei 99 % Kopplungseffizienz nur bei 55 % ($0{,}99^{60} \bullet 100$), bei einem 100-mer sind es nur noch 36 %.

Da die Kopplungseffizienz selbst in gut eingestellten Syntheseautomaten nicht in jedem Schritt nahe 100 % liegen wird, ist die Gesamtausbeute bei Oligonukleotiden über 100 Basen Länge sehr begrenzt. Es sind also in jedem Fall aufwendige Reinigungsverfahren notwendig, um korrekt lange Oligonukleotide von weniger langen Produkten zu trennen. In der Praxis ist die chemische DNA-Synthese heute auf ca. 200 Basen lange Oligonukleotide beschränkt.

2.3 Enzymatische DNA-Synthese

2.3.1 DNA-Polymerasen

Die DNA-synthetisierenden Enzyme nennt man DNA-Polymerasen. Eine ihrer Funktionen ist die semikonservative Replikation der DNA vor der Zellteilung. Für die Details wollen wir an dieser Stelle auf Lehrbücher der molekularen Genetik verweisen. Eine wichtige Eigenschaft der DNA-Polymerasen ist, dass sie DNA-Doppelstränge nie *de novo* herstellen. Vielmehr können sie lediglich einen einzelsträngigen DNA-Matrizenstrang entsprechend den Komplementaritätsregeln kopieren und dadurch zum DNA-Doppelstrang komplettieren. Daher nennt man diese Enzyme korrekt DNA-abhängige DNA-Polymerasen. Die Enzyme benötigen ein kurzes Stück doppelsträngiger DNA mit einem freien 3'-Hydroxylende als Startpunkt für die DNA-Synthese. Dieses kurze Stück eines doppelsträngigen DNA-Moleküls wird dadurch generiert, dass ein einzelsträngiges Oligonukleotid (Primer) an die 5'-Seite des zu synthetisierenden DNA-Abschnitts hybridisiert wird (o Abb. 2.14). Die DNA-Synthese durch eine DNA-Polymerase erfolgt dann durch Verlängerung des Primers an dessen 3'-Ende, d. h. die enzymatische DNA-Synthese verläuft immer in 5'→3'-Richtung bezogen auf den neu synthetisierten DNA-Strang.

Die DNA-Polymerase I aus *E. coli* besitzt neben einer 5'→3'-DNA-Polymeraseaktivität auch zwei Exonuklease-Aktivitäten. Diese bauen die DNA entweder in 5'→3'-Richtung oder in 3'→5'-Richtung von den Enden her ab. Die 3'→5'-Exonuklease-Aktivität wird auch *Proofreading*-Aktivität genannt, da durch diese enzymatische Aktivität vom Enzym quasi überprüft werden kann, ob der erfolgte Polymerisationsschritt korrekt war. Das sogenannte Klenow-Fragment der DNA-Polymerase I aus *E. coli* ist eine modifizierte Variante der DNA-Polymerase I. Es handelt sich um das größere von zwei enzymatisch aktiven Peptidfragmenten, das bei der partiellen Proteolyse der DNA-Polymerase I durch eine bestimmte Protease entsteht. Das Klenow-Fragment besitzt nur noch die 5'→3'-DNA-Polymeraseaktivität und die 3'→5'-Exonukleaseaktivität. Die 5'→3'-Exonukleaseaktivität ist hingegen verlorengegangen.

Welche Bestandteile müssen sich in einem Reaktionsansatz befinden, damit eine In-vitro-DNA-Synthese effizient abläuft? Obwohl jede DNA-Polymerase ihre individuellen Anforderungen an ein Puffersystem stellt, gibt es doch einige allgemeine Regeln:

- Zum einen wird natürlich die DNA-Polymerase selbst gebraucht. Heute stammen kommerziell erhältliche DNA-Polymerasen übrigens aus gentechnischer Produktion.
- Dann muss eine einzelsträngige DNA vorhanden sein, die als Matrize (im Englischen *template*) zum Kopieren durch die DNA-Polymerase dienen kann.
- Ein kurzes Oligonukleotid ist als Primer der DNA-Polymerisation nötig (o Abb. 2.14). Diese Primer stammen heute aus der kommerziell von spezialisierten Firmen durchgeführten chemischen DNA-Synthese.
- Außerdem müssen alle vier DNA-Bausteine im Reaktionsansatz vorliegen. Eine Mischung aller vier DNA-Nukleotide wird im molekularbiologischen Sprachgebrauch als Desoxynukleosidtriphosphate oder kurz als dNTPs bezeichnet.
- Alle enzymatischen Reaktionen benötigen ein geeignetes Puffersystem, das in diesem Fall auch Mg^{2+}-Ionen enthalten muss, da die eigentlichen Substrate der DNA-Polymerase die Mg^{2+}-Komplexe der dNTPs sind.

2.3.2 Polymerase-Kettenreaktion (PCR)

Eine der wichtigsten Methoden in der modernen Molekularbiologie ist die Polymerase-Kettenreaktion (PCR). Daher wollen wir diese Methode ausführlich besprechen. Die PCR wurde Mitte der 1980er Jahre von dem Amerikaner Kary Mullis erfunden, der dafür 1993 zusammen mit Michael Smith mit dem Nobelpreis für Chemie geehrt wurde. Die Abkürzung PCR leitet sich von der englischen Bezeichnung *polymerase chain reaction* ab. Das wirklich Bemerkenswerte an der PCR besteht darin, dass Mullis ein Prinzip aufzeigte, mit dem DNA im Reagenzglas nicht einfach nur synthetisiert, sondern zugleich auch vervielfältigt (amplifiziert) werden kann.

Wie wir bereits gehört haben, waren DNA-Polymerasen lange bekannt und die enzymatische In-vitro-Synthese von DNA gehörte damals bereits zum Standard-Repertoire der Molekularbiologen. Das Neue war die exponentielle Amplifikation der DNA auf der Basis

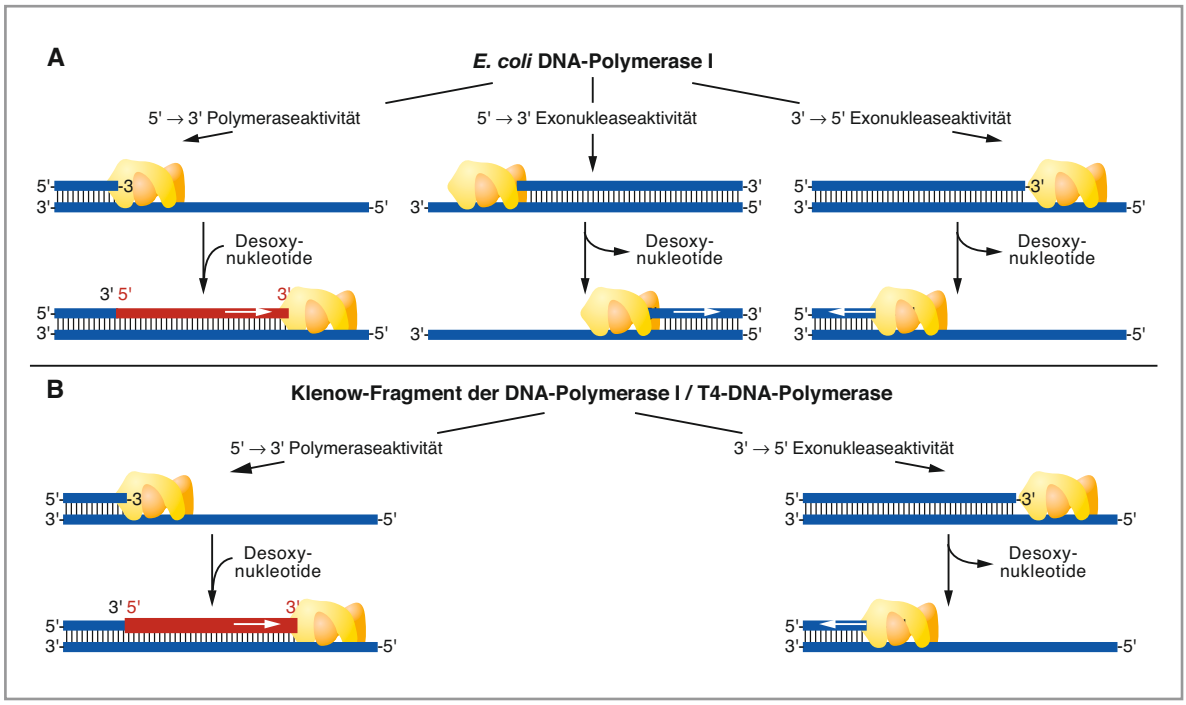


Abb. 2.14 DNA-Polymerasen aus *E. coli*. **A** Das Holoenzym DNA-Polymerase I aus *E. coli* besitzt drei unterschiedliche Aktivitäten. Die Polymerase-Aktivität synthetisiert matrizenabhängig DNA vom 5'- zum 3'-Ende. Dementsprechend wird die Matrize vom 3'-Ende zum 5'-Ende hin abgelesen (oben links). Die 5'→3'-Exonukleaseaktivität hydrolysiert DNA vom 5'-Ende in Richtung des 3'-Endes, Nukleotid für Nukleotid (oben Mitte). Demgegenüber greift die 3'→5'-Exonukleaseaktivität die DNA an den 3'-Enden an und hydrolysiert die entsprechenden DNA-Stränge in Richtung des 5'-Endes (oben rechts). Die 5'-Enden selbst bleiben intakt. **B** Durch partielle Proteolyse wird eine Enzymvariante generiert, die als das Klenow-Fragment der DNA-Polymerase I bezeichnet wird. Diesem Fragment fehlt, ähnlich wie der DNA-Polymerase aus dem Bakteriophagen T4, die 5'→3'-Exonukleaseaktivität des Holoenzyms. Hingegen besitzen die Klenow-Polymerase und die T4-DNA-Polymerase sowohl die 5'→3'-Polymeraseaktivität als auch die 3'→5'-Exonukleaseaktivität (unten links und rechts).

eines verblüffend einfachen Prinzips. Ausgehend von ganz wenigen Einzelmolekülen ließ sich DNA nun im Reagenzglas in einen Größenordnungsbereich vervielfältigen, der ausreichte, um die DNA nachzuweisen, zu charakterisieren oder für DNA-Rekombinationen einzusetzen.

Die PCR ist heute aus den gentechnischen Forschungslaboratorien nicht mehr wegzudenken – sie hat die molekularbiologische Forschung revolutioniert. Sie wird nicht nur eingesetzt, um DNA-Fragmente für DNA-Rekombinationen zu gewinnen, sondern ist darüber hinaus im Lauf der Zeit eine der wichtigsten molekularbiologischen Nachweismethoden geworden, die in praktisch jedem Zweig der Biowissenschaften eingesetzt werden kann. Unter anderem hat diese Technik auch die humangenetische Diagnostik und die forensische Analytik bereichert.

Die Basis der DNA-Amplifikation durch die PCR ist die enzymatische DNA-Synthese. Es werden daher für den Reaktionsansatz alle Komponenten gebraucht, die bei jeder DNA-Synthese nötig sind:

- Wie bei jeder durch DNA-Polymerasen katalysierten Reaktion ist auch bei der PCR eine Matrize nötig, also ein Abschnitt einer DNA, der kopiert werden soll. Diese als Matrize oder *template* bezeichnete DNA muss in diesem Fall nicht einzelsträngig vorliegen, da während der PCR zyklische Temperaturerhöhungen eine Denaturierung der Matrizen-DNA in Einzelstränge bewirken.
- Die DNA-Polymerase braucht für die DNA-Synthese einen Startpunkt. Nachdem die in die Reaktion eingesetzte doppelsträngige DNA denaturiert wurde, muss ein Primer an die einzelsträngige Matrize hybridisieren, um einen kurzen doppelsträngigen Bereich als Startpunkt für die DNA-Polymerisation zu bilden.
- Die vier Desoxynukleosidtriphosphate dATP, dCTP, dGTP und dTTP müssen als Substrate für die DNA-Polymerase vorhanden sein.
- Geeignete Puffer- und Salzbedingungen ermöglichen die Arbeit des Enzyms. So dienen Mg^{2+}-Ionen als Kofaktoren für die DNA-Polymerase, deren

eigentliche Nukleotid-Substrate die Mg^{2+}/dNTP-Komplexe sind.
- Eine geeignete DNA-abhängige DNA-Polymerase katalysiert dann die DNA-Synthese.

Die Genialität der von Mullis gemachten Erfindung bestand darin, dass bei der In-vitro-DNA-Synthese gleichzeitig zwei Primer eingesetzt werden und die DNA-Polymerisation mehrmals hintereinander durchgeführt wird. Eigentlich griff Mullis eine Idee des indisch/amerikanischen Biochemikers und Nobelpreisträgers Har Gobind Khorana auf, der das Prinzip einer wiederholten DNA-Polymerisation bereits 1971 publiziert hatte. Allerdings variierte Mullis diese Idee dadurch, dass er zwei Startpunkte für die DNA-Synthese definierte. Diese beiden Startpunkte liegen in einem gewissen Abstand voneinander entfernt und begrenzen dadurch den DNA-Bereich, der amplifiziert werden soll. Außerdem werden die Primer so gewählt, dass die beiden von ihnen ausgehenden DNA-Synthesereaktionen aufeinander zulaufen.

Prinzip der PCR

Eine schematische Übersicht über das Prinzip der PCR-Technik ist in Abb. 2.15 gezeigt. Wir können nur DNA-Bereiche zwischen bekannten Nukleotidsequenzen amplifizieren. Für diese bekannten Nukleotidsequenzen werden Primer synthetisiert, die sich optimal an die bekannten Sequenzen anlagern können. Die Sequenz des amplifizierten DNA-Bereichs zwischen den beiden Primern muss hingegen nicht bekannt sein.

Wir benötigen immer zwei PCR-Primer. Der eine Primer bindet an den oberen Strang der Ziel-DNA, wohingegen der andere Primer in einigem Abstand vom ersten Primer an den unteren Strang bindet. Dadurch haben wir zwei Startpunkte für DNA-Polymerase-Reaktionen definiert, wobei die Richtungen dieser beiden Polymerisations-Reaktionen aufeinander zu und schließlich überlappend verlaufen.

Die Primer und die zu amplifizierende DNA mischen wir in einer Pufferlösung zusammen mit den vier Desoxynukleosidtriphosphaten, Mg^{2+} und einer thermostabilen DNA-Polymerase. Solche thermostabilen DNA-Polymerasen hat man ursprünglich aus verschiedenen thermophilen Bakterien isoliert, beispielsweise die „Taq-Polymerase“ aus dem thermophilen Bakterium *Thermus aquaticus*. Heute werden diese Polymerasen gentechnisch erzeugt und sind kommerziell erhältlich. Sobald dem Reaktionsansatz alle erforderlichen Komponenten zugesetzt wurden, stellt man das Gefäß in einen sogenannten *Thermocycler*. Mithilfe eines solchen Geräts wird ein komplexes Inkubationsprogramm kontrolliert und wiederholt durchlaufen. Typisch für dieses Inkubationsprogramm sind periodisch wechselnde Inkubationstemperaturen (Abb. 2.15).

Worin unterscheiden sich die unterschiedlichen Inkubationsperioden, und was geschieht dort?

- Zunächst wird die Temperatur für ca. 5 min auf **95 °C** eingestellt. Unter den vorgegebenen Pufferbedingungen werden während dieser Zeit alle doppelsträngigen DNA-Bereiche zu Einzelsträngen denaturiert.
- Danach wird die Temperatur auf einen Bereich von **50–60 °C** gesenkt, wodurch die Primer mit den komplementären Bereichen auf der Ziel-DNA doppelsträngige Hybride ausbilden können. Die beiden Primer liegen im großen Überschuss zur DNA-Matrize vor. Deshalb bilden sich bevorzugt Doppelstrangbereiche zwischen den Primern und dem jeweiligen DNA-Einzelstrang der Ziel-DNA. Die Renaturierung der beiden denaturierten Einzelstränge der Ziel-DNA ist hingegen unter diesen Bedingungen nicht favorisiert. Die Hybridisierung der Primer (*annealing*) ist in der Regel bereits nach 15–30 Sekunden abgeschlossen.
- Im nächsten Schritt wird die Inkubationstemperatur auf ca. **72 °C** angehoben. Bei dieser Temperatur arbeiten die thermostabilen DNA-Polymerasen optimal. Je nach Länge des zu amplifizierenden DNA-Bereichs wird die Temperatur zwischen 30 Sekunden und mehreren Minuten bei 72 °C gehalten.
- Nun wird durch Anheben der Temperatur für 15–30 Sekunden auf > 90 °C wieder die gesamte im Reaktionsansatz vorhandene doppelsträngige DNA denaturiert.
- Den Wechsel zwischen den drei verschiedenen Temperaturen (50–60 °C → 72 °C → 95 °C bzw. 95 °C → 50–60 °C → 72 °C) wird als Zyklus bezeichnet. Je nach gewünschtem Grad der Amplifikation der Ausgangs-DNA werden in der Praxis meist zwischen 25 und 35 solcher Zyklen nacheinander durchlaufen.

Die Produkte des ersten Amplifikationszyklus sind nur an einem Ende durch den eingesetzten Primer klar begrenzt. Sie besitzen jedoch keine genau definierte Länge; sie werden deshalb als „lange Produkte“ bezeichnet (Abb. 2.15). Ab der zweiten Amplifikationsrunde fungieren neben der Originalmatrize auch die „langen Produkte“ als Matrizen. Deshalb werden zusätzlich zu zwei weiteren „langen Produkten“ auch DNA-Fragmente definierter Länge gebildet (Abb. 2.15). Diese neu entstehenden Transkripte sind an der einen Seite durch den eingesetzten Primer und an der anderen Seite durch das Ende der Matrize begrenzt. Somit liegen am Ende der PCR-basierten DNA-Amplifikation drei verschiedene Molekülarten vor (Abb. 2.16):

- die Stränge der Originalmatrize,
- einseitig durch jeweils einen Primer begrenzte „lange Produkte“ und

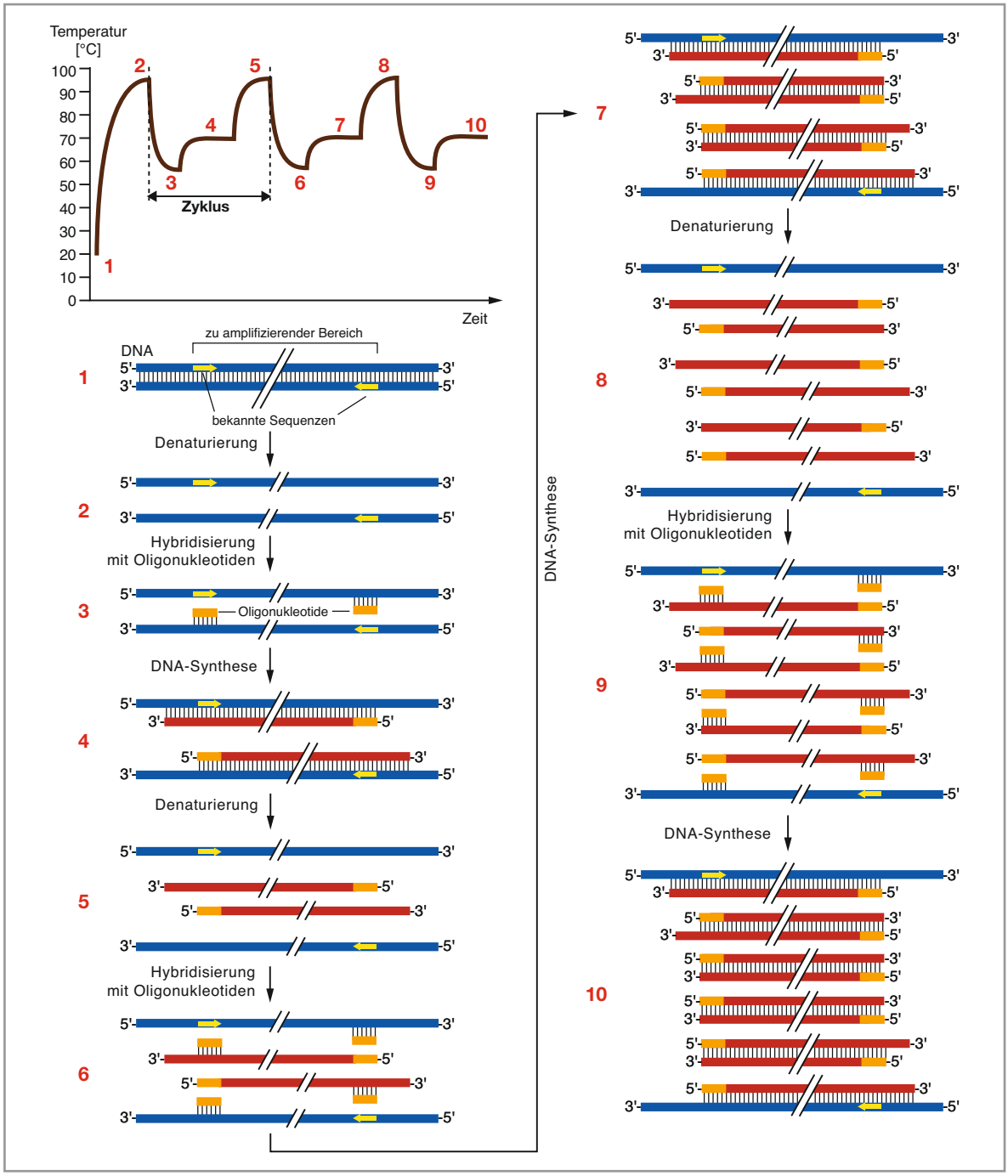


Abb. 2.15 Prinzip der Polymerase-Kettenreaktion (PCR). In der ersten Abbildung ist der Temperaturverlauf der PCR dargestellt. Wir bezeichnen die Menge an Matrizen-Molekülen mit N_0 und die Zahl der Zyklen mit z. (1) Die dunklen Balken zeigen einen Ausschnitt aus einer DNA, die als Matrize dient. Der Bereich, der amplifiziert werden soll, ist mit gelben Pfeilen markiert. Dieser Bereich wird von zwei Primern (orangefarbene Balken) eingerahmt, die jeweils entweder am oberen oder am unteren Strang der Matrizen-DNA binden. (2) Denaturieren der DNA durch Erhitzen auf > 95 °C. (3) Abkühlen auf 50–60 °C, um das Anlagern der Primer (*annealing*) zu ermöglichen. (4) Einstellen der Temperatur auf 72 °C, um die DNA-Synthese durch die Taq-Polymerase zu starten. Einen Umlauf von Denaturieren/*annealing*/DNA-Synthese (95 °C → 60 °C → 72 °C) bezeichnet man als PCR-Zyklus. Beachte, dass die DNA-Synthese weiter läuft als bis zur gegenüberliegenden Primer-Bindestelle. Solche „langen Produkte" vermehren sich linear ($N_0 \cdot z$). (5) Erhitzen auf 95 °C. (6) Erneutes *annealing* der Primer. (7) Erneute DNA-Synthese. Beachte, dass die ersten DNA-Moleküle entstehen, die genau so lang sind wie das gewünschte Endprodukt. Diese Moleküle amplifizieren sich jetzt exponentiell ($N_0 \cdot 2^z$). (8) Denaturieren. (9) *Annealing*. (10) DNA-Synthese. Das Endprodukt ist genau so lang wie der Abstand zwischen den PCR-Primern (inklusive der Primer selbst).

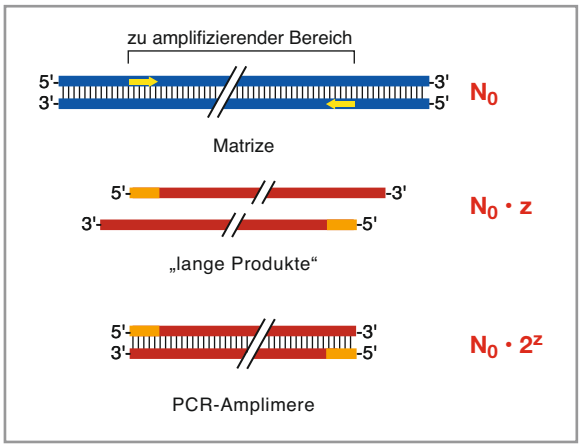


Abb. 2.16 Gegenüberstellung der in einer PCR gebildeten Produkte. Die blauen Balken zeigen einen Ausschnitt aus einer DNA, die als Matrize dient. Der zu amplifizierende Bereich ist mit gelben Pfeilen begrenzt. Die Primer sind orange dargestellt. Die Zahl der eingesetzten Matrizen-Moleküle ist (N_0) und bleibt während der PCR konstant. Beachte, dass die eingesetzten Primer Teil des PCR-Amplimers sind. Die sogenannten „langen Produkte“ bilden sich durch Abschreiben der Matrizen-Moleküle in jedem Zyklus (z). Da sich die Matrizen-Menge nicht ändert, vermehren sich die langen Produkte linear ($N_0 \cdot z$). Die langen Produkte wie auch die eigentlichen PCR-Amplimere vermehren sich in jedem Zyklus und dienen wieder selbst als Matrize. Daher vermehren sich die Amplifikate exponentiell ($N_0 \cdot 2^z$).

- beidseitig durch die Primer begrenzte Transkripte, die eigentlichen PCR-Amplimere.

Alle diese Moleküle dienen auch in den jeweils folgenden Amplifikationszyklen als Matrizen für die DNA-Polymerisation.

Effizienz und Limitation der PCR

Nachdem wir im vorhergehenden Abschnitt das Prinzip der PCR diskutiert haben, ist natürlich die Frage interessant, wie stark eine DNA mit dieser Methode amplifiziert werden kann. Ausgehend von N_0 Molekülen DNA, die für eine PCR-Amplifikation einsetzt wurden, sieht die theoretische Bilanz wie folgt aus: Die Zahl der Ausgangsmatrizen ist konstant geblieben, während die Zahl der „langen Produkte“ (Abb. 2.16) auf $N_0 \cdot z$ gestiegen ist (z = Anzahl der Zyklen). Die Zahl der PCR-Amplimere definierter Länge nach z Zyklen (N_z) errechnet sich als $N_z = N_0 \cdot 2^z$. Ein Beispiel soll den Amplifikationsgrad verdeutlichen: Nehmen wir an, dass wir 1000 Moleküle einer DNA als Ausgangsmaterial (Matrize, N_0) haben. Nehmen wir weiterhin an, dass wir ein Stück aus dieser DNA mit 30 PCR-Zyklen amplifizieren. Wir bekommen nach der oben genannten Formel $1000 \cdot 2^{30}$, also ca. 10^{12} PCR-Amplimere. Das entspricht einem Amplifikationsgrad von 10^9. Interessant für die Praxis ist, dass wir 1000 Moleküle auch mit modernsten Methoden der DNA-Analytik kaum werden nachweisen können, während nach der Amplifikation durch die PCR aus diesen 1000 Molekülen ca. $2 \cdot 10^{-12}$ mol oder $1{,}3 \cdot 10^{-6}$ g (1,3 µg) DNA geworden sind (bei einer angenommenen Amplimer-Länge von 1000 bp). Diese Menge DNA kann über konventionelle Agarosegel-Elektrophorese leicht nachgewiesen oder in ihrer DNA-Sequenz bestimmt werden. Ebenso bekommen wir mit dieser Methode genug DNA „in die Hand“, um sie mit gentechnologischen Methoden bearbeiten zu können.

Verschiedenste Parameter können die Effizienz der PCR beeinflussen. Einer der limitierenden Faktoren ist die Länge der DNA, die durch die Primer eingeschlossen wird. Generell nimmt die Effizienz der PCR mit steigender Länge des zu amplifizierenden DNA-Bereichs kontinuierlich ab. Dies kann teilweise durch die Inkubationsdauer während des Polymerisationsschrittes ausgeglichen werden. Optimal lassen sich Bereiche in der PCR vervielfältigen, die kleiner als 2000 bp sind. Allerdings können bei entsprechender Auswahl der DNA-Polymerasen und der Inkubationsbedingungen auch bis zu 40 000 bp amplifiziert werden.

Was ist bei der PCR weiterhin zu beachten? Es sollten keine Primer-Paare verwendet werden, die partiell oder komplett zueinander komplementär sind. Solche Primer-Paare können mit hoher Effizienz Primer-Dimere bilden, statt sich an die Matrizen-DNA anzulagern. Dadurch wird die eigentliche Polymerase-Kettenreaktion sehr ineffizient. Nicht jede Matrizen-DNA lässt sich optimal über PCR vervielfältigen. Beispielsweise kann die DNA-Amplifikation erheblich gestört werden, wenn die beiden Einzelstränge der Matrizen-DNA nach dem Denaturieren jeweils intramolekulare Sekundärstrukturen ausbilden können. Dies kann vor allem dann passieren, wenn die DNA-Einzelstränge hohe G/C-Anteile aufweisen. Diese Komplikation können wir unter Umständen dadurch umgehen, dass wir dem Reaktionsansatz einen geringen Anteil 7-Aza-dGTP beimischen. Dieses Nukleotid-Analogon verhindert teilweise die Ausbildung von Sekundärstrukturbereichen, ohne die Matrizenfunktion zu stark zu stören. Auch durch Zusätze von Glycerin oder Dimethylsulfoxid können wir die Ausbildung von Einzelstrang-Sekundärstrukturen möglicherweise unterdrücken. Dabei ist allerdings zu beachten, dass unter Umständen auch die Enzymaktivität abnimmt. Dem können wir durch den Zusatz beispielsweise von Rinderserumalbumin (BSA) entgegenwirken. BSA schützt die DNA-Polymerase in ihrer Aktivität und bindet Inhibitoren.

In der Regel werden die theoretischen Ausbeuten an PCR-Produkt nicht erreicht, weil nach ca. 20 Zyklen die Effizienz der Reaktion kontinuierlich abnimmt. Das kann folgende Gründe haben:

- Erstens werden Substrate und Primer langsam aber stetig verbraucht.
- Zweitens lässt das Enzym durch die starken Temperaturschwankungen in seiner Aktivität langsam nach.
- Drittens lagern sich vermehrt komplementäre Stränge zu Doppelsträngen aneinander und verhindern damit, dass sich neue Initiationsstellen für die DNA-Polymerisation bilden.

Ein Aspekt ist bei der PCR noch wichtig: Die Taq-Polymerase aus *Thermus aquaticus* besitzt keine 3'→5'-Exonuklease-Aktivität. Diese Aktivität haben wir bereits bei der DNA-Polymerase I als *Proofreading*-Aktivität kennengelernt. Da die Taq-Polymerase diese Aktivität nicht hat, sind DNA-Synthesen, die durch Taq-Polymerase katalysiert werden, oft ungewöhnlich fehlerhaft. Dies kann – muss aber nicht – Probleme bereiten. Signifikant wird dieses Problem, wenn Fehler bereits während der ersten PCR-Zyklen auftreten und wenn sehr wenig Matrizenmaterial eingesetzt wird. Dann dienen die fehlerhaften Kopien als Hauptmatrizen für weitere PCR-Zyklen. Derartige Schwierigkeiten lassen sich allerdings heute teilweise umgehen, da beispielsweise mit der Pwo-Polymerase aus *Pyrococcus woesei* oder der Pfu-Polymerase aus *Pyrococcus furiosus* thermostabile DNA-Polymerasen verfügbar sind, die auch *Proofreading*-Aktivität besitzen.

Eine der großen Stärken der Polymerase-Kettenreaktion – nämlich ihre enorme Sensitivität – kann auch zu einem großen Problem werden. Das muss besonders dann bedacht werden, wenn die PCR im Rahmen diagnostischer Verfahren eingesetzt wird. Unter Sensitivität versteht man die Möglichkeit der PCR, aus sehr geringen Matrizen-Mengen analysierbare Mengen an Amplimeren zu generieren. Man kann sich leicht vorstellen, dass hierbei Kontaminationen der Analysenprobe mit fremdem biologischem Material unter allen Umständen vermieden werden müssen. Ansonsten besteht die große Gefahr der Isolierung einer falschen Amplimer-DNA und daraus folgend eines falsch positiven Ergebnisses in der genetischen Diagnostik. Dies ist gerade für solche Krankheiten eine inakzeptable Komplikation, bei denen heute bevorzugt die PCR-Analytik als Diagnostikmethode eingesetzt wird, wie beispielsweise bei Krebserkrankungen oder viralen Infektionen. Ähnliches gilt natürlich auch für die Erstellung sogenannter genetischer Fingerabdrücke, die im Rahmen forensischer Untersuchungen zur Identifizierung bestimmter Personen eingesetzt werden. Die hohe Sensitivität der Methode bedingt, dass eine Kontamination mit DNA-Spuren, die zum Beispiel vom forensischen Personal eingetragen werden, eine Analyse sofort unbrauchbar machen würde.

Reverse-Transkription-PCR

Mittlerweile gibt es so viele Anwendungsvarianten der PCR-Technik, dass eine vollständige Aufzählung und Erklärung dieser Varianten den Rahmen dieses Buches sprengen würde. Deshalb sollen nur einige Beispiele erwähnt werden, die bei der Isolierung von Genen zur Gewinnung rekombinanter Proteinwirkstoffe oder in der molekularen Diagnostik besondere Bedeutung haben.

Die Reverse Transkriptase der Retroviren (meist als RT abgekürzt) ist insofern eine ungewöhnliche DNA-Polymerase, als sie vor allem RNA als Matrize erkennt und diese in DNA kopiert: Dieses Enzym ist also eine RNA-abhängige DNA-Polymerase. Im Labor ist die Reverse Transkriptase das maßgebliche Enzym zur Herstellung von mRNA-Kopien, den cDNAs. Der Begriff „cDNA" steht für *complementary* DNA und deutet an, dass das erste Produkt der Reversen Transkription ein DNA-Einzelstrang ist, der zur kopierten mRNA komplementär ist. Im Sprachgebrauch der Molekularbiologen wird der Begriff „cDNA" oft auch als *copy* DNA übersetzt, was als Synonym für „kopierte mRNA" gesehen werden kann.

Obwohl das Prinzip der PCR darauf beruht, DNA in vitro zu amplifizieren, schließt dies nicht aus, dass wir auch RNA vervielfältigen können. Das geschieht allerdings indirekt, da die RNA dazu in einem ersten Schritt zunächst in DNA umgeschrieben werden muss (reverse Transkription). Danach können wir die cDNA als Matrize in einer normalen PCR einsetzen, um sie zu amplifizieren. Dieses Verfahren einer gekoppelten reversen Transkription mit der PCR nennt man Reverse-Transkription-PCR, kurz RT-PCR. Diese Methode hat in der molekularbiologischen Forschung große Bedeutung:

- Mit relativ geringem Aufwand lässt sich für ein beliebiges Gen die Expressionsrate, also die Gleichgewichtsmenge von mRNA in einer Zelle, messen.
- Nur die mRNA enthält die genetische Information für ein Protein ohne Unterbrechung des Triplett-Leserahmens durch Introns. Das ist von enormer praktischer Bedeutung, wenn man therapeutische Proteine aus dem menschlichen Genom beispielsweise in Bakterien exprimieren möchte, da Bakterien keine Introns kennen und eine genetische Information, die Introns enthält, nicht in ein Protein umsetzen können.

Konkret gehen wir folgendermaßen vor, um mRNA in cDNA zu kopieren (o Abb. 2.17): Zunächst lagern wir an das 3'-Ende einer isolierten mRNA einen DNA-Primer an. Eine Möglichkeit, die Reverse Transkription zu starten, ist die Nutzung eines spezifischen Primers, der nur zu einer bestimmten Gensequenz passt. Der Vorteil dieser Methode ist, dass aus der Auswahl von bis zu 50 000 verschiedenen mRNAs einer Zelle theoretisch

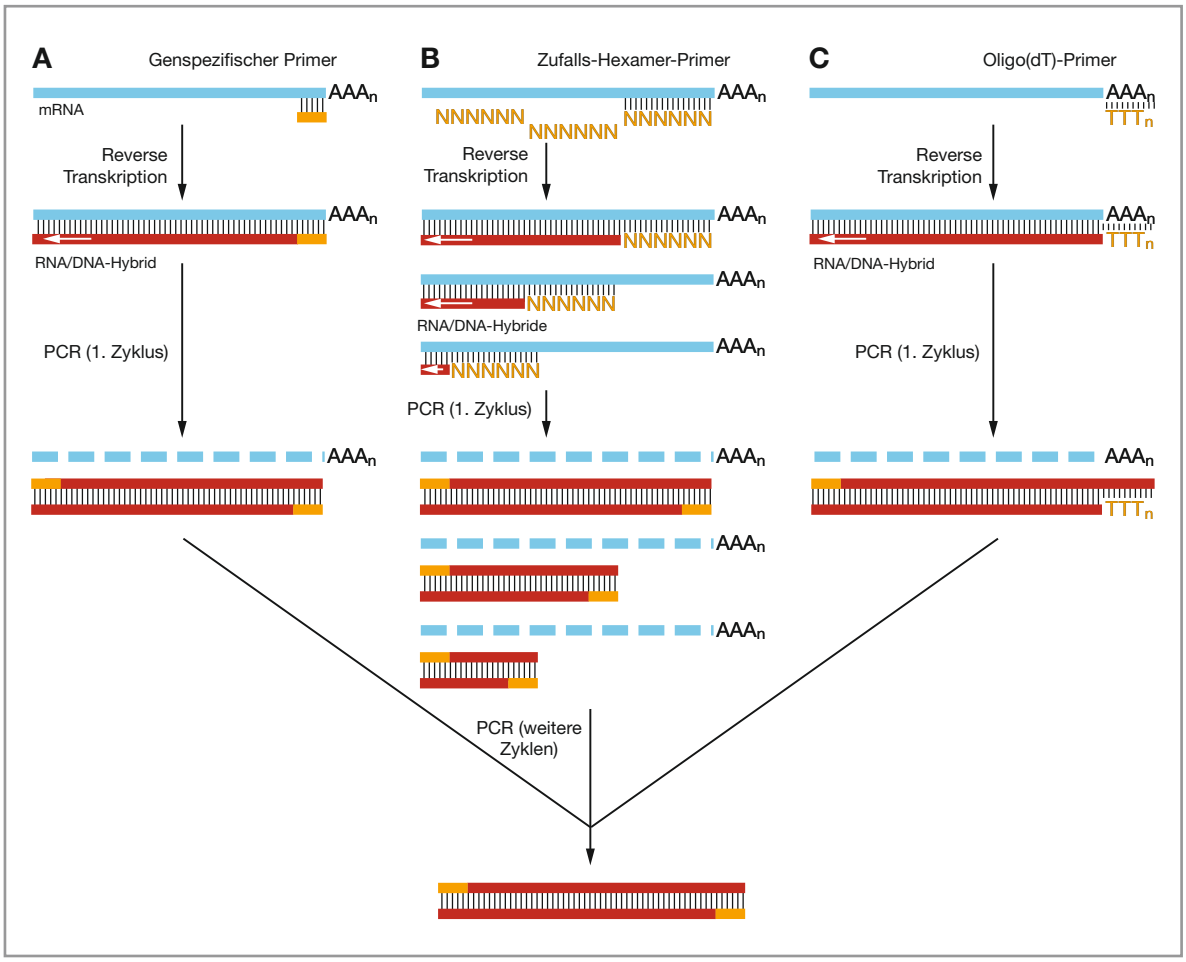


Abb. 2.17 Gekoppelte Reverse-Transkription-PCR (RT-PCR). In einer RT-PCR wird eine RNA (hellblaue Linien) durch das Enzym Reverse Transkriptase (RT) in eine DNA kopiert. Es entsteht eine einzelsträngige *complementary* DNA (cDNA; rote Linien), die zunächst einmal als RNA/DNA-Hybrid vorliegt. In den hier gezeigten Beispielen geht die RT-PCR von polyadenylierter RNA, also mRNA, aus. Primer sind in orange dargestellt. **A** Starten der RT-PCR mit einem genspezifischen Primer. In diesem Beispiel wird aus einem *Pool* isolierter mRNAs entsprechend der Spezifität des Primers nur eine ganz bestimmte RNA amplifiziert. Nach der Erststrang-cDNA-Synthese wird die RNA in dem RNA/DNA-Hybrid durch die RNAse-H-Aktivität der RT abgebaut. Mit einem zweiten Primer, der am 3'-Ende der Erststrang-cDNA bindet, wird durch PCR ein DNA-Doppelstrang erzeugt. **B** Starten der RT-PCR mit Zufalls-Hexameren. Diese Primer bestehen aus einer Mischung von 6-mer-Oligonukleotiden, die in jeder Position jedes der vier möglichen DNA-Basen enthalten können. Zufalls-Hexamere sind demnach theoretisch zu jeder denkbaren Gen-Sequenz komplementär. Allerdings kann dadurch die Reverse Transkription an irgendeiner Position des zu analysierenden Gens gestartet werden, ohne dass man dies beeinflussen kann. Man bekommt also eine Schar von Erststrang-cDNAs, die sich in ihren Längen unterscheiden. Allerdings lässt sich aus dieser Schar die cDNA mit der korrekten Länge amplifizieren, indem die entsprechenden genspezifischen Primer eingesetzt werden. **C** Starten der RT-PCR mit einem Oligo(dT)-Primer. Hier nutzt man aus, dass alle mRNAs polyadenyliert sind. Der mRNA-Pool wird mit einem Oligo(dT)-Primer hybridisiert und so die Reverse Transkription aller mRNAs gestartet. Danach folgt eine PCR mit genspezifischen Primern.

nur eine einzige cDNA-Spezies hergestellt wird, nämlich die des untersuchten Gens. Eine andere Möglichkeit ist, die Reverse Transkription *aller* in der RNA-Präparation enthaltenen mRNAs durch die Nutzung unspezifisch bindender Primer zu starten. Es gibt zwei verschiedene Arten solcher unspezifischer Primer:

- Zum einen kann man ausnutzen, dass alle eukaryontischen mRNAs an ihren 3'-Enden polyadenyliert werden. In diesem Fall wird ein Oligo(dT)-Primer eingesetzt, der an die Oligo(A)-Enden der mRNAs hybridisiert. Damit wird die Reverse Transkription *jeder* zellulären mRNA von deren 3'-Ende aus gestartet und man erhält eine komplexe Mischung von cDNAs.
- Alternativ können sogenannte Zufalls-Hexamere eingesetzt werden. Das sind 6-mer-Oligonukleotide, die in jeder Position jedes der vier möglichen DNA-

Basen enthalten können, also theoretisch zu jeder denkbaren Gen-Sequenz komplementär sind. Die Reverse Transkription wird dann an irgendeiner Position einer mRNA gestartet und man erhält ebenfalls eine komplexe cDNA-Mischung.

Nach der Hybridisierung von Oligo(dT)- oder Hexamer-Primern an die mRNA inkubieren wir den Ansatz mit einer Reversen Trankriptase. Am Ende dieser Reaktion liegen alle mRNAs einer Zelle als RNA/DNA-Hybride vor. Soll aus dieser Mischung die cDNA eines speziellen Gens isoliert werden, führen wir anschließend eine PCR mit zwei genspezifischen Primern durch. Dabei wird im ersten Zyklus der PCR durch Erhitzen auf 95 °C der RNA-Anteil der vorhandenen RNA/DNA-Hybride zersetzt. Die einzelsträngige cDNA dient nun als Matrize für die PCR.

Quantitative PCR

Wir haben bereits in einem der vorausgehenden Abschnitte einige Grenzen der PCR diskutiert. Im Bereich der Anwendung der PCR in der Diagnostik sind folgende Limitationen der konventionellen PCR von praktischer Bedeutung:

- PCRs sind relativ zeitaufwendig, da einerseits die Reaktion selbst einige Zeit beansprucht, andererseits jedoch die Analyse der Amplimere durch eine nachgeschaltete Agarosegel-Elektrophorese erfolgen muss.
- Die konventionelle PCR ist keine quantitative Methode. Sie ist normalerweise so angelegt, dass die Reaktion möglichst vollständig abläuft. Daraus folgt, dass bei starker Amplifikation einer Matrizen-DNA die Ressourcen in der Reaktion schneller limitierend werden als wenn nur eine geringe Amplifikation stattfindet. Mit großer Wahrscheinlichkeit wird die Reaktion am Ende der PCR in der Sättigung gewesen sein. Daraus folgt, dass eine quantitative Abschätzung der zu Beginn der PCR vorliegenden Matrizen-Menge durch Vergleich der am Ende einer PCR vorliegenden Amplimer-Mengen in verschiedenen Proben kaum möglich ist, denn man wird nicht beobachten können, in welchem Zyklus der PCR die Reaktion in einer Probe bereits in die Sättigung gegangen war, während sie in parallelen Proben noch exponentiell ablief.

Hieraus ergibt sich die Überlegung, ob es nicht möglich wäre, die Akkumulation der Produkte in einer PCR in Echtzeit zu verfolgen, sodass die Menge gebildeter Amplimere während der exponentiellen Phase der PCR festgestellt werden kann. Diese Abwandlung der konventionellen PCR wird als *Realtime*-PCR oder auch Quantitative PCR bezeichnet (o Abb. 2.18). Die Grundlage für die *Realtime*-PCR ist, dass dem PCR-Ansatz ein

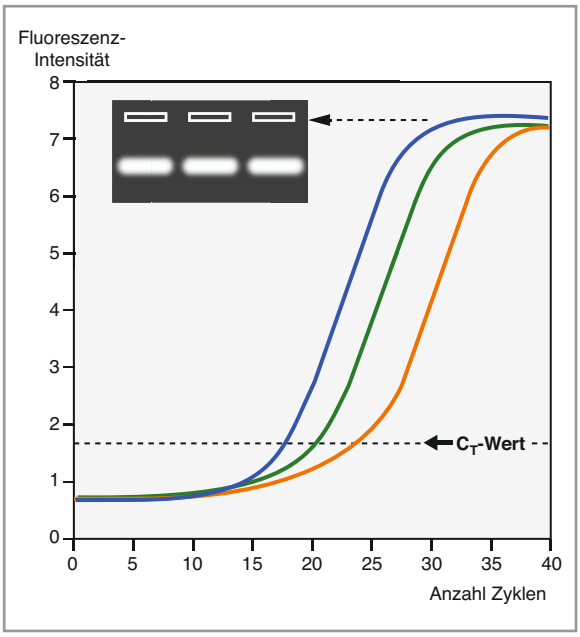


Abb. 2.18 Quantitative PCR. Diese Methode erlaubt es, die Bildung von PCR-Amplimeren in Echtzeit zu verfolgen. Dazu wird dem PCR-Ansatz ein Fluoreszenzfarbstoff zugegeben, der Doppelstrang-DNA spezifisch anfärbt (CYBR-Green®). In der Abbildung ist der theoretische Reaktionsverlauf paralleler PCRs gezeigt, in der in der Reihenfolge braune → grüne → blaue Kurve steigende Menge Matrizen-DNA eingesetzt wurden. Am Ende der Reaktionen wird in allen Reaktionen eine ähnliche Fluoreszenz erreicht, d. h. die Reaktionen waren in der Sättigung. Wenn man die gebildete DNA-Menge am Ende der Reaktionen auf einem Agarosegel analysieren würde, würde man also in etwa gleich starke Banden (Insert) erhalten und würde fälschlicherweise schlussfolgern, dass alle PCRs gleiche Matrizen-Mengen enthalten haben und daher die Reaktionen vergleichbar abgelaufen sind. Tatsächlich sind die Reaktionen aber unterschiedlich schnell verlaufen, was auf unterschiedliche Mengen an vorhandenem Ausgangsmaterial (Matrize) zurückzuführen ist. Den Unterschied an eingesetzter Matrize kann man bestimmen, indem man die Anzahl von PCR-Zyklen vergleicht, die bis zum Erreichen eines zuvor festgelegten Schwellenwerts (C_T) benötigt werden, wobei der Unterschied in Zyklen (z) im Wesentlichen einer Differenz der Matrizen-Menge von 2^z entspricht.

für doppelsträngige DNA spezifischer, interkalierender Farbstoff zugesetzt wird, dessen Fluoreszenz nach Bindung an die gebildeten Amplimere zunimmt. Das gemessene Fluoreszenzsignal in jedem Zyklus ist dann proportional zur gebildeten Menge an Amplimer und kann nach jedem PCR-Zyklus abgefragt werden. Der Farbstoff der Wahl ist hier CYBR-Green®. Wie Ethidiumbromid interkaliert CYBR-Green® in doppelsträngige DNA und bildet einen sehr stabilen Komplex, wobei nach Anregung mit Licht einer Wellenlänge von 497 nm ein Fluoreszenzlicht von 521 nm emittiert wird.

Abb. 2.19 Aufbau eines TaqMan®-Primers. Ein TaqMan®-Primer trägt an beiden Enden chromophore Gruppen. An das 5'-Ende des Primers ist ein Fluoreszenz-Donor gekoppelt, der bei entsprechender Anregung die eingestrahlte Energie durch Fluoreszenz-Resonanz-Energietransfer (FRET) an einen an das 3'-Ende des Primers gekoppelten Chromophor weitergibt. Wenn letzterer nicht selbst fluoresziert, sondern die Fluoreszenz des Donors praktisch auslöscht, spricht man von einem Quencher. Eine Fluoreszenz beobachten wir immer dann, wenn der Donor vom Quencher getrennt wird (siehe 5'-Nukleasetest).

Die Bestimmung von DNA-Mengen mit CYBR-Green® ist wesentlich empfindlicher als die DNA-Färbung mit Ethidiumbromid.

Führen wir eine PCR in Gegenwart von CYBR-Green® durch, erhalten wir einen in Abb. 2.18 dargestellten Verlauf der Reaktion in Abhängigkeit von der eingesetzten Matrizen-Menge. Je mehr Matrize in der PCR eingesetzt wird, desto früher erreicht die Fluoreszenz-Intensität einen Schwellenwert, den wir früh in der exponentiellen Phase der Reaktion festlegen (Abb. 2.18). Wir können nun die Matrizen-Menge in unbekannten Proben abschätzen, wenn wir bestimmen, bei welchem Zyklus der Schwellenwert erreicht wird. Der Unterschied an Zyklen bis zum Erreichen des Schwellenwerts steht über die bereits diskutierte Formel $N_z = N_0 \times 2^z$ (N_z: Zahl der PCR-Amplimere nach z Zyklen, N_0: Ausgangsmenge Matrize, z: Zahl der PCR-Zyklen) in Beziehung zur in der Probe befindlichen Matrizen-Menge, die sich als $N_0 = N_z$ geteilt durch 2^z bzw. $N_0 = 2^{-z} \times N_z$ berechnet. In der Praxis können wir N_0 am besten dadurch bestimmen, dass wir für eine Vergleichs-DNA eine Standardkurve wie in Abb. 2.18 herstellen und die Zykluszahl am Schwellenwert in den unbekannten Proben mit dieser Standardkurve vergleichen.

Die Verwendung von CYBR-Green® hat jedoch auch einen Nachteil: Wir detektieren *jede* in der PCR synthetisierte doppelsträngige DNA, also auch nicht gewünschte Amplimere, die durch unspezifisches Hybridisieren der Primer an falsche Matrizen entstehen können. Um dies zu vermeiden, wurde eine Modifikation der *Realtime*-PCR eingeführt, die als TaqMan®-PCR oder 5'-Nukleasetest bekannt geworden ist. TaqMan®-Primer (Abb. 2.19) tragen fluoreszierende Farbstoffe, sodass die Messung des Reaktionsverlaufs nicht auf dem Zusatz von CYBR-Green® basiert. Ein TaqMan®-Primer wird zusätzlich zu zwei konventionellen Primern eingesetzt, die für die Bildung des Amplimers sorgen. Der TaqMan®-Primer bindet spezifisch an eine DNA-Sequenz zwischen diesen beiden konventionellen Primern.

Das Besondere an TaqMan®-Primern ist die Nutzung des Prinzips des Fluoreszenz-Resonanz-Energietransfers (FRET). Die Fluoreszenz eines Chromophors am 5'-Ende des TaqMan®-Primers (Donor) wird von einem am 3'-Ende des Primers verankerten Quencher unterdrückt. Die Energie-Übertragung vom Donor zum Quencher erfolgt durch FRET. Durch Lichteinstrahlung in einen Chromophor (hier der Donor) wird ein Elektron in einen elektronisch angeregten Zustand versetzt. Die Energie des angeregten Elektrons kann auf ein Akzeptor-Chromophor übertragen werden, das weniger als 10 nm entfernt ist. Der Energietransfer ist strahlungslos, die Geschwindigkeit und Intensität von FRET hängen unter anderem von der Größe der Überlappung des Emissionsspektrums des Donors und des Absorptionsspektrums des Akzeptors ab. In unserem Fall nutzen wir FRET, um die Fluoreszenz des Donors durch Absorption im Akzeptor, der selber nicht fluoresziert, zu unterdrücken. Wenn wir also während der PCR die Zunahme eines Fluoreszenzsignals als Maß für die Menge entstehender Amplimere messen wollen, muss der FRET innerhalb des TaqMan®-Primers irgendwie aufgehoben werden.

Was geschieht nun während der TaqMan®-PCR? Nach der Denaturierung der Matrizen-DNA binden die zwei konventionellen Primer und der TaqMan®-Primer an die Matrizen-DNA (Abb. 2.20). Zu diesem Zeitpunkt ist FRET zwischen Donor- und Quencher-Chromophor möglich, wir bekommen kein Fluoreszenz-Signal. Trifft nun die DNA-Polymerase während der folgenden DNA-Synthese auf den TaqMan®-Primer, wird dieser durch die 5'→3'-Exonuklease-Aktivität der DNA-Polymerase abgebaut. Die dadurch bedingte Trennung der Chromophore hebt FRET auf und der Donor-Chromophor fluoresziert messbar (daher der Name 5'-Nukleasetest). In der TaqMan®-PCR ist also das Fluoreszenz-Signal proportional zur aktuell gebildeten Menge eines Amplimers. Wir messen aufgrund der Spezifität des TaqMan®-Primers nur die Zunahme des gewünschten Amplimers und nicht die Menge unspezifischer Nebenprodukte.

2.3.3 DNA-Sequenzierung

Konventionelle DNA-Sequenzierung nach Sanger

Die von dem britischen Biochemiker Frederick Sanger Mitte der 1970er Jahre eingeführte Sequenzierung von DNA erfolgt mit einer Modifikation der normalen enzymatischen DNA-Synthese. Sangers Konzept besteht darin, DNA in Gegenwart geringer Konzentrationen sogenannter Ketten-Terminatoren zu synthetisieren. Ketten-Terminatoren sind 2',3'-Didesoxynukleosidtriphosphate (ddNTPs), die sich von den „norma-

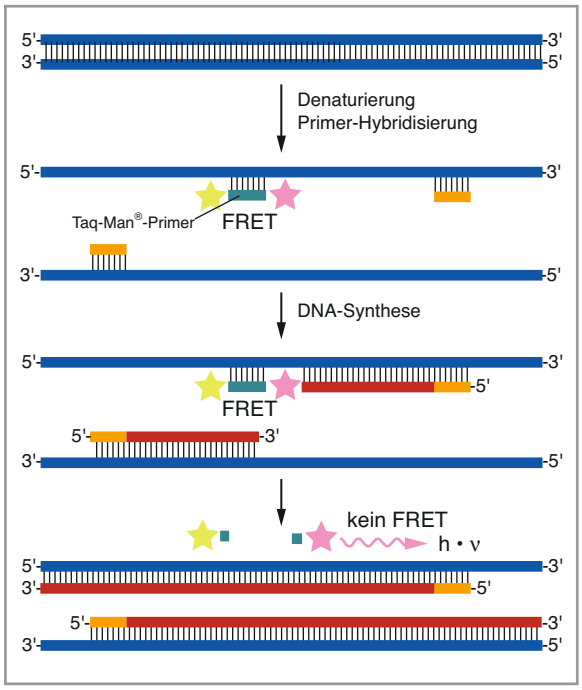


Abb. 2.20 Prinzip der TaqMan®-PCR (5'-Nukleasetest). Der 5'-Nukleasetest ist eine PCR mit drei Primern. Zwei Primer (orange) sind konventionelle PCR-Primer, der dritte ist ein TaqMan®-Primer. Der TaqMan®-Primer bindet zwischen den beiden normalen PCR-Primern. Die Matrizen-DNA (blaue Linien) wird denaturiert und alle drei Primer hybridisieren an die DNA. Wenn die DNA-Polymerase den TaqMan®-Primer erreicht, baut sie diesen über ihre 5'→3'-Exonuklease-Aktivität ab. Dadurch werden Fluoreszenz-Donor und -Quencher voneinander getrennt und es entsteht ein Fluoreszenz-Signal.

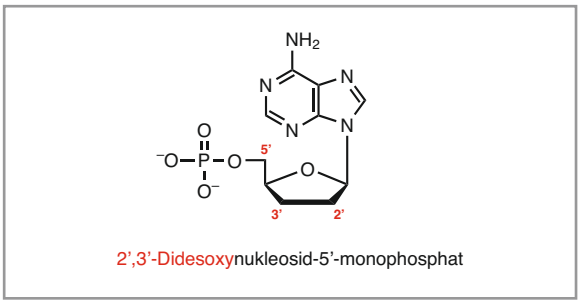


Abb. 2.21 Didesoxynukleotide für die DNA-Sequenzierung. Didesoxynukleotide unterscheiden sich von einem gewöhnlichen Desoxynukleotid durch das Fehlen der 3'-OH-Gruppe. Wird ein solcher DNA-Baustein während der DNA-Synthese in die wachsende Polynukleotidkette eingebaut, ist anschließend keine Kettenverlängerung mehr möglich.

len" 2'-Desoxynukleosidtriphosphaten (dNTPs) darin unterscheiden, dass ihnen neben der 2'-OH-Gruppe auch die 3'-OH-Gruppe der Ribose fehlt (Abb. 2.21). Wird ein ddNTP von der DNA-Polymerase als Substrat verwendet, kann die DNA-Kette an dieser Position wegen des Fehlens einer 3'-OH-Gruppe nicht weiter verlängert werden.

Für die Sanger-Sequenzierung vorbereitete DNAs liegen normalerweise in Plasmiden kloniert vor. Ein Reaktionsansatz zur DNA-Sequenzierung enthält alle Komponenten, die auch für die normale DNA-Synthese benötigt werden. Der zu sequenzierende Bereich einer DNA kann als doppelsträngige Plasmid-DNA eingesetzt werden, da diese DNA im Sinne eines *cycle sequencing* in Einzelstränge überführt wird. *Cycle sequencing* bedeutet, dass DNA-Sequenzierungen ähnlich wie eine PCR durchgeführt werden, jedoch enthält die Sequenzierreaktion nur einen Primer. Im ersten Schritt des jeweiligen *cycle-sequencing*-Zyklus wird also die doppelsträngige DNA denaturiert, im zweiten Schritt der Primer an die DNA hybridisiert und im dritten Schritt wird die DNA-Synthese gestartet. Wie in einer normalen PCR wird dieser Vorgang 25–30-mal wiederholt.

Die Sequenzierreaktion enthält dNTPs und ddNTPs in einem definierten Verhältnis. Als Folge davon kann die DNA-Polymerase statistisch gesehen an jeder Position der Matrizen-DNA mit einer gewissen Wahrscheinlichkeit ein ddNTP statt eines dNTPs einbauen, sodass eine weitere Verlängerung der DNA nicht mehr möglich ist. Als Resultat der Reaktion entsteht eine Schar von DNA-Fragmenten, die alle denselben Anfang haben (den Primer) und mit einem ddNTP enden. Da der Kettenabbruch an jeder Position der synthetisierten DNA erfolgen kann, unterscheiden sich die erhaltenen DNA-Fragmente in der Länge um jeweils eine Base (Abb. 2.22).

Wie lässt sich jetzt herausfinden, mit welchem Didesoxynukleotid der Kettenabbruch erfolgt ist? Der Trick besteht darin, dass jedes der vier in der Reaktion eingesetzten ddNTPs mit einem anderen Fluoreszenzfarbstoff markiert wird. Die Trennung einer Schar von fluoreszenzmarkierten DNA-Molekülen erfolgt durch Elektrophorese entweder in hochauflösenden Polyacrylamidgelen oder durch Kapillarelektrophorese. Die Trennschärfe ist dabei so groß, dass einzelne DNA-Fragmente getrennt werden können, die sich in der Länge um eine einzige Base unterscheiden. Am Ende der Trennstrecke regt ein Laser die vier Fluoreszenzfarbstoffe an (Abb. 2.23). Die jeweilige Emissionswellenlänge korrespondiert mit der Didesoxybase, die das entsprechende DNA-Fragment am Ende durch Kettenabbruch markiert hat. Ein Computer erstellt aus den Daten ein „Elektropherogramm" (Abb. 2.23) sowie ein Textdokument über die Reihenfolge der Basen, also der DNA-Sequenz.

DNA-Sequenzierung nach Sanger ist die Standard-Methode, wenn es beispielsweise darum geht, mit gentechnologischen Methoden bearbeitete DNA-Fragmente als eine Art „Identitätsprüfung" zu charakterisie-

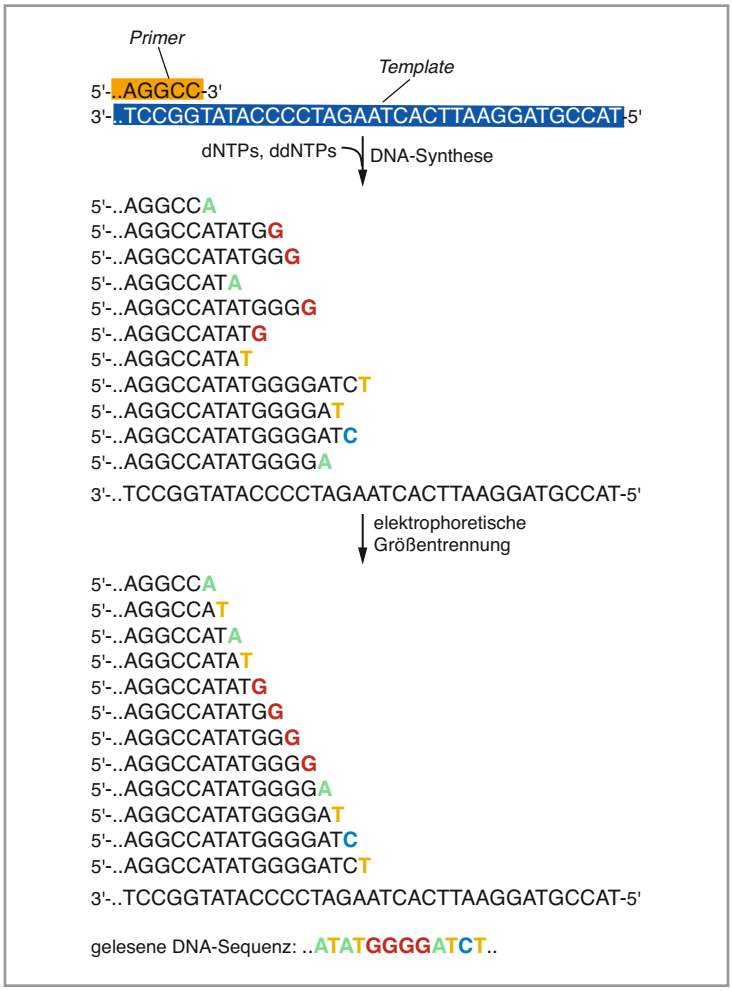


Abb. 2.22 Prinzip der DNA-Sequenzierung nach Sanger. Eine Sequenzierreaktion enthält neben einem Primer und einer Matrizen-DNA eine definierte Mischung aus normalen Desoxynukleotiden (dNTPs) und 2',3'-Didesoxynukleotiden (ddNTPs). Jedes der vier ddNTPs ist mit einem anderen Fluoreszenzfarbstoff gekoppelt. Die DNA-Synthese wird mit einem aus der PCR bekannten zyklischen Protokoll aus Denaturierung, Hybridisierung des Primers und DNA-Synthese durchgeführt. Während der vom Primer ausgehenden DNA-Synthese wird die DNA-Polymerase zu irgendeinem Zeitpunkt statt eines normalen dNTP ein ddNTP in die wachsende Nukleinsäurekette einbauen. An dieser Stelle bricht die DNA-Synthese ab. Da der Zeitpunkt des Kettenabbruchs Zufall ist, bekommt man eine Schar von DNA-Fragmenten, die statistisch alle Längen der Matrize abdecken und sich jeweils um genau eine Base unterscheiden. Diese Schar von DNA-Fragmenten kann man mit geeigneten Methoden auftrennen. An jedem Ende eines solchen Fragments befindet sich ein fluoreszenzmarkiertes ddNTP. Man kann also durch Auslesen der Fluoreszenzsignale vom Primer ausgehend eine DNA-Sequenz ableiten.

ren. Mit einer einzigen Sequenzierreaktion erreicht man mit der Sanger-Methode „Leseweiten" von bis zu 1000 Nukleotiden.

Die Sanger-Sequenzierung spielt aber auch eine große Rolle für die Identifizierung neuer DNA-Sequenzen in der Genfunktionsanalyse. Die weitgehende Automatisierung dieser Technologie hat auch die Entschlüsselung ganzer Genome von Organismen ermöglicht, auch wenn zwischenzeitlich wesentlich effizientere „Sequenziermethoden der nächsten Generation" für die Entschlüsselung kompletter Genome entwickelt wurden.

Hochdurchsatz-DNA-Sequenzierung

Seit der Publikation der Terminator-Technologie zur DNA-Sequenzierung durch Frederick Sanger wurde die Methode schrittweise immer weiterentwickelt (Abb. 2.24). Diese Entwicklung wurde ganz maßgeblich angetrieben durch die Mitte der 1990er Jahre getroffene Vereinbarung, in einem großen internationalen Konsortium die komplette Sequenz des menschlichen Genoms zu erarbeiten. Seit der Erfindung der Sanger-Sequenzierung war man es gewohnt, fluoreszenzmarkierte DNA-Fragmente in hochauflösenden Polyacrylamidgelen aufzutrennen. Diese Methode ist jedoch aufwendig und schlecht automatisierbar. Nachdem die Auftrennung fluoreszenzmarkierter DNA-Fragmente auf Kapillarelektrophorese umgestellt wurde, konnte durch die parallele Sequenzierung mit mehreren Automaten und mit Maschinen, die beispielsweise 96 Kapillaren gleichzeitig enthalten, ungleich größere Datenmengen generiert werden. Ein Vergleich verdeutlicht, welche Fortschritte in den letzten ca. 30 Jahren auf diesem Gebiet gemacht wurden: Das etwa 3 Milliarden Basenpaare umfassende menschliche Genom wurde praktisch komplett nach der Sanger-Methode sequenziert. In dem mit öffentlichen Geldern geförderten Humanen Genomprojekt wurden ungefähr 23 Milliarden Basen sequenziert. Das hat ca. 15 Jahre gedauert und Kosten in der Größenordnung von Milliarden Dollar verursacht. Heute würde ein Sequenzierautomat, der nach dem Sanger-Prinzip arbeitet, ca. 1 Million Basenpaare Sequenzinformation am Tag produzieren – ein einzelner solcher Automat

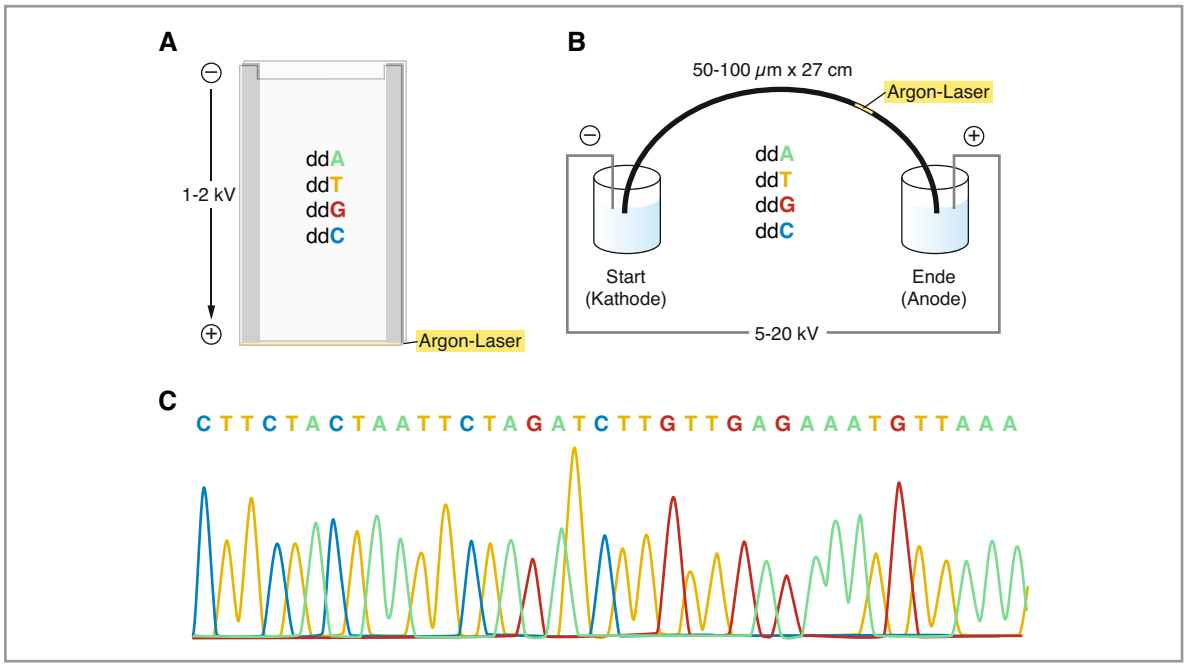


Abb. 2.23 Methoden zur Trennung von DNA-Fragmenten bei der DNA-Sequenzierung. **A** DNA-Fragmente können in hoch auflösenden Polyacrylamidgelen getrennt werden. Die Trennstrecke beträgt ca. 50 cm. Die Gele werden mit einem harnstoffhaltigen Puffer hergestellt, um die einzelsträngigen DNA-Fragmente an der Ausbildung intramolekularer Basenpaarungen zu hindern. Am Ende der Trennstrecke werden die Fluoreszenzfarbstoffe der ddNTPs angeregt und ausgelesen. Da die kleinen DNA-Fragmente schneller durch das Gel laufen als die größeren, kann man eine Reihenfolge der Fragmentlängen (= der DNA-Sequenz) vom Primer ausgehend ablesen. **B** Alternativ zu Polyacrylamidgelen wird in modernen Sequenzierautomaten die Kapillarelektrophorese zur Trennung der fluoreszenzmarkierten Sequenzierprodukte eingesetzt. **C** Die Reihenfolge der aus einem Polyacrylamidgel oder einer Kapillare eluierten DNA-Fragmente werden vom Computer ausgewertet und als Elektropherogramm dargestellt. Die unterschiedliche Farbgebung der einzelnen ddNTPs erleichtert die Interpretation der *Peaks*.

hätte also immer noch 23 000 Tage oder 63 Jahre gebraucht, um die Rohinformation für das menschliche Genom zu erstellen. Bedenken muss man natürlich, dass viele Sequenzierautomaten weltweit mit dem Projekt beschäftigt gewesen wären, sodass beispielsweise 63 Automaten diese große Herausforderung im Prinzip in einem Jahr hätten erledigen können. Wäre das humane Genom jedoch mit einem der modernen Hochdurchsatz-Sequenzierautomaten erstellt worden, wären die 23 Milliarden Basenpaare in der Theorie auf einem *einzigen* Gerät innerhalb *weniger* Tage analysiert worden.

An diesem Beispiel wird klar, dass man die neuen Sequenziermethoden zu Recht als „Hochdurchsatz-Sequenzierung" bezeichnen kann. Da diese Methoden nach der Sanger-Sequenzierung eingeführt wurden, wird manchmal auch von „Nächstgenerationen-Sequenzierung" (*Next Generation Sequencing* oder NGS) gesprochen. Diese Methoden eröffnen heute völlig neue Ansätze in der Genomforschung und dadurch auch in der Arzneistoffforschung. Obwohl es mehrere NGS-Methoden gibt, wollen wir an dieser Stelle nur die derzeit wichtigsten kurz vorstellen: die Pyrosequenzierung, die Illumina-Sequenzierung und die PacBio-Sequenzierung.

Die Methode der Pyrosequenzierung wurde Anfang der 1990er Jahre von dem iranischen Molekularbiologen M. Ronaghi erfunden und später von den Firmen 454 Life Sciences und Roche kommerzialisiert. Daher ist diese Form der Sequenzierung auch als „454-Sequenzierung" bekannt geworden. Wie wir bereits über die Funktion von DNA-Polymerasen gelernt haben, erfolgt die Verlängerung eines bestehenden DNA-Moleküls durch den Einbau eines Nukleosidmonophosphats aus einem Nukleosidtriphosphat (dNTP) unter Abspaltung von anorganischem Pyrophosphat. Bei der Pyrosequenzierung wird eine DNA-Synthese durchgeführt und dabei die Bildung des Pyrophosphats in Echtzeit verfolgt. Im Unterschied zur Sanger-Sequenzierung können durch die Pyrosequenzierung DNA-Moleküle sequenziert werden, ohne dass die DNA zuvor in Vektoren kloniert und dadurch amplifiziert werden muss. Vielmehr greift die Technologie der Pyrosequenzierung allein auf die In-vitro-Amplifikation von DNA durch PCR zurück. Eine entscheidende Innovation war dabei die Amplifikation von einzelnen DNA-Molekülen an

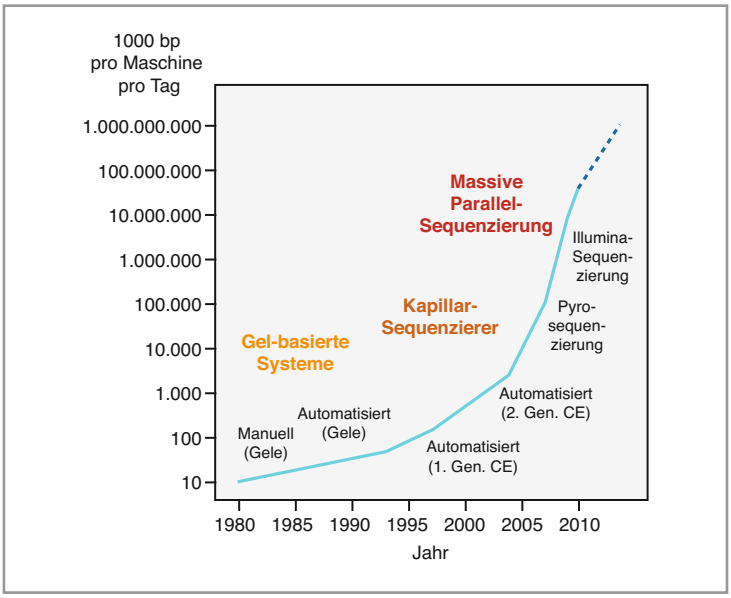


Abb. 2.24 Entwicklung der durch verschiedene Sequenziermethoden generierten Datenmengen. Von den Zeiten Anfang der 1980er Jahre, als Polyacrylamid-Gele für die Trennung der DNA-Fragmente aus Sequenzierreaktionen noch individuell von Hand gefertigt wurden, bis zur Entwicklung von automatisierten Hochdurchsatz-Sequenzierungen, wurde der *Output* an Sequenzinformationen durch sehr viele Wissenschaftler um mehr als das Millionenfache gesteigert. CE: Kapillarelektrophorese. Quelle: Stratton et al.: Nature 458 (2009), 719–724.

funktionalisierten Kügelchen von ca. 20 µM Durchmesser durch die sogenannte Emulsions-PCR. Die einzelnen an den Kügelchen amplifizierten DNA-Moleküle werden in Picotiterplatten verteilt, die 1 Million Löcher (*wells*) pro Platte besitzen. Ebenso viele DNA-Abschnitte können dann gleichzeitig sequenziert werden. Die Länge der in einer Reaktion von einem DNA-Molekül bestimmbaren DNA-Sequenz beträgt meist ca. 250 bp, kann aber in neueren Verfahren auf bis zu 600 bp ausgedehnt werden. Eine „454-Sequenziermaschine" kann in *einem* Lauf, der etwa 10 Stunden dauert, bis zu 600 Millionen Basenpaare Sequenzinformation liefern. Das wäre eine etwa 600-fach größere Datenmenge im Vergleich zur „Tagesleistung" eines fortgeschrittenen Sequenzierautomaten, der mit der Sanger-Technologie arbeitet.

Ebenso wie die bereits diskutierte Sanger-Methode basiert auch die Pyrosequenzierung auf der Neusynthese von DNA mit gleichzeitiger Bestimmung der DNA-Sequenz, es ist also eine *Sequencing-by-synthesis*-Methode. Zunächst wird ein Primer an einen zu sequenzierenden DNA-Einzelstrang hybridisiert, damit die DNA-Polymerase einen Startpunkt für die DNA-Synthese hat. Zur DNA-Polymerase und der zu sequenzierenden DNA wird in vier aufeinander folgenden Reaktionen jeweils nur eines der vier dNTPs zugesetzt (z. B. in der Reihenfolge A-G-C-T wie in Abb. 2.25 gezeigt). Gemeinsam mit dem jeweiligen dNTP gibt man das Enzym Sulfurylase und dessen Substrat Adenosin-5'-phosphosulfat (APS) zu. In der ersten Reaktion, die nur dATP für die DNA-Synthese enthält, kann die Polymerase das Nukleotid nur einbauen, wenn die erste Base der Matrize ein komplementäres Thymin-Nukleotid ist. Wird ein dATP eingebaut, wird das frei werdende Pyrophosphat sofort von der Sulfurylase mit APS zu ATP umgesetzt. Das ATP wiederum ist Substrat für ein weiteres in der Reaktion vorhandenes Enzym, die Luciferase. Luciferase setzt das Substrat Luciferin mithilfe des entstandenen ATPs zu Oxoluciferin um, wobei Lichtquanten emittiert werden. Diese können durch hochempfindliche Kameras detektiert werden. Nun wird das nächste dNTP zugegeben, in unserem Beispiel dGTP (Abb. 2.25).

In dem in Abb. 2.25 gezeigten Beispiel kommen zwei C-Nukleotide nacheinander vor. Das bedeutet, dass die DNA-Polymerase im zweiten Durchgang zwei dGTPs einbauen kann, sodass zwei Moleküle Pyrophosphat freigesetzt werden. Da die Lichtemission in der Reaktion proportional zur entstandenen Menge an Pyrophosphat ist, bekommt man zweimal die Peakfläche eines einzelnen Pyrosphosphats, d. h. man hat die Sequenz AGG bestimmt. Nun folgt im dritten Durchgang die Zugabe von dCTP. Man bekommt Lichtemission, die ebenfalls der Peakfläche von zwei eingebauten Nukleotiden entspricht, die Sequenz lautet also: AGGCC. Nun gibt man dTTP zu und bekommt ein Signal, das so stark ist, wie bei dATP. Das bedeutet, dass die nachfolgende Base in der Matrize ein Adenin-Nukleotid ist. Jetzt beginnt der Zyklus A-G-C-T wieder von vorn. Durch Wiederholungen der Reaktionsfolge A-G-C-T kann die Matrizen-DNA sequenziert werden und es entsteht ein charakteristisches Peakmuster, das als „Pyrogramm" bezeichnet wird (Abb. 2.25).

Ende der 1990er Jahre experimentierten S. Balasubramanian und D. Klenerman mit einem neuen Sequenzierverfahren, das auf dem Einbau von fluoreszenzmarkierten Terminator-Nukleotiden basiert. Das kommt uns bekannt vor, denn so funktioniert dem Prinzip

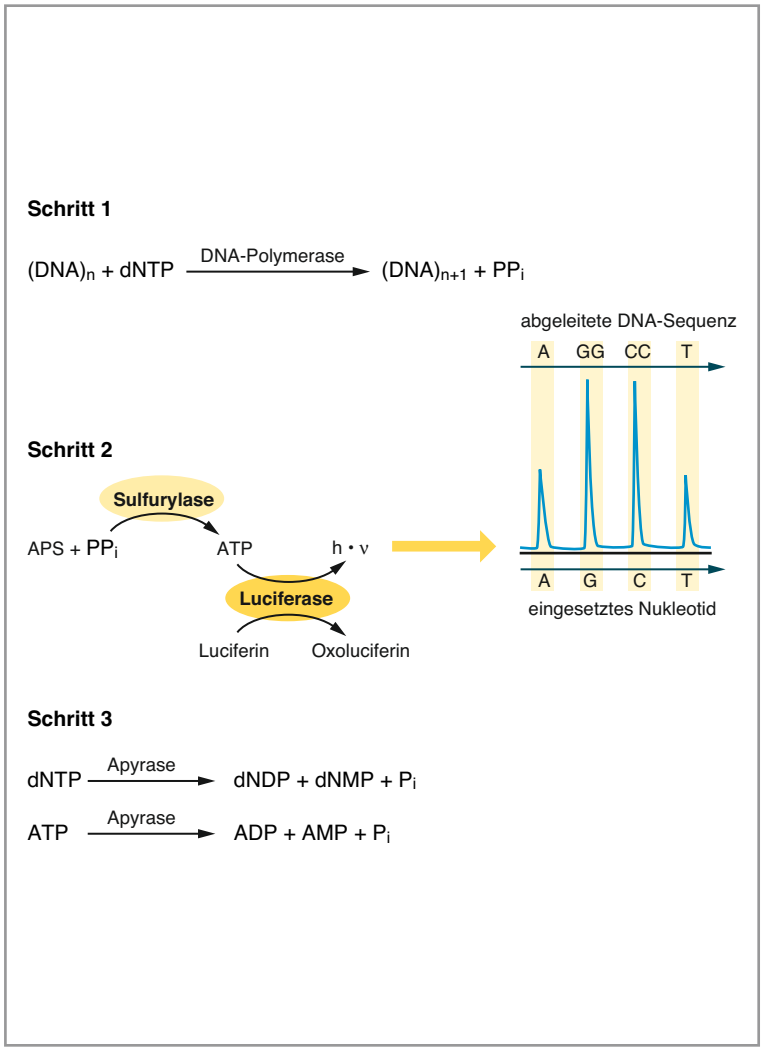


Abb. 2.25 Prinzip der Pyrosequenzierung. **Schritt 1**: DNA-Synthese. Eine DNA-Polymerase verlängert einen Primer, der an eine einzelsträngige DNA hybridisiert wurde. Dabei wird das Substrat Nukleosidtriphosphat (dNTP) als Nukleosidmonophosphat in die DNA eingebaut und es entsteht Pyrophosphat (PP_i). In jedem Reaktionszyklus werden die vier Nukleotide dATP, dGTP, dCTP und dTTP nacheinander zur Reaktion gegeben. Die Polymerase kann nur eines davon einbauen, sodass bekannt ist, beim Einbau welchen Nukleotids das PP_i entsteht. **Schritt 2**: Detektion. Das entstandene PP_i wird mit dem Enzym Sulfurylase und Adenosin-5'-phosphosulfat umgesetzt, als Produkt entsteht ATP. Dieses wird von dem Enzym Luciferase genutzt, um Luciferin zu Oxoluciferin umzusetzen. Es wird Licht emittiert, das von einer hochempfindlichen Kamera detektiert wird. Das Signal wird dem jeweils in die Reaktion eingesetzten dNTP zugeordnet. Die Lichtemission ist proportional zur Menge an freigesetztem PP_i. Ein zweifaches Signal (hier bei Zugabe von dGTP und dCTP) bedeutet, dass die Polymerase pro Reaktion zwei G- bzw. C-Nukleotide nacheinander eingebaut hat. **Schritt 3**: Recycling. Die Nukleotide des ersten Reaktionszyklus werden durch das Enzym Apyrase degradiert, sodass ein neuer Zyklus ohne zwischenzeitliches Waschen erfolgen kann.

nach auch die Sanger-Sequenzierung. Das neue Sequenzierverfahren wurde Anfang der 2000er Jahre von der Firma Solexa zur Markreife weiterentwickelt, sodass 2006 der erste Sequenzierautomat zur Verfügung stand. Das Sequenzierverfahren war demzufolge zunächst als „Solexa-Methode" bekannt, bevor die Firma 2007 von Illumina übernommen wurde. Seither wird meist von „Illumina-Sequenzierung" gesprochen, wenn diese Sequenziertechnologie angewendet wird. Ein solcher Sequenzierautomat produziert heute in einem einzigen Lauf, der etwa 6 Tage dauert, bis zu 8 Milliarden Einzelsequenzen von beispielsweise 125 bp Länge, d. h. bis zu 1000 Milliarden Basenpaare Sequenzinformation pro Woche.

Bei der Illumina-Methode wird eine PCR-Vervielfältigung der zu sequenzierenden DNA auf einem soliden Träger durchgeführt, was als *Bridge*-PCR bezeichnet wird (Abb. 2.26). Zunächst werden an die zu sequenzierenden DNA-Moleküle Adapter ligiert. Die DNAs werden dann denaturiert und mit beiden Enden kovalent an die Oberfläche der Fließzelle gebunden. Nach einer PCR-Runde werden die Amplifikate denaturiert und am Ort ihrer Entstehung über ihre Enden wiederum an die Oberfläche der Fließzelle fixiert. Somit entstehen in der Fließzelle dichte Cluster von doppelsträngigen DNA-Molekülen, die anschließend separat von anderen Clustern parallel sequenziert werden. Die eigentliche DNA-Sequenzierung basiert auf der *Sequencing-by-synthesis*-Methode mit fluoreszenzmarkierten Nukleotiden. Dabei wird mit modifizierten Basen gearbeitet, deren 3'-Hydroxylgruppen durch eine Schutzgruppe blockiert sind, sodass bei der DNA-Synthese jeweils nur ein einziges neues Nukleotid eingebaut werden kann. Die vier Terminator-Nukleotide werden mit unterschiedlichen Fluoreszenzfarbstoffen markiert, wobei der Farbstoff entweder direkt an die 3'-Hydroxylgruppe gehängt wird und gleichzeitig als Schutzgruppe funktioniert, oder der Farbstoff mit der Base des Nukleotids verbunden wird und die 3'-Hydroxylgruppe mit einer separaten Schutzgruppe versehen wird. Man gibt

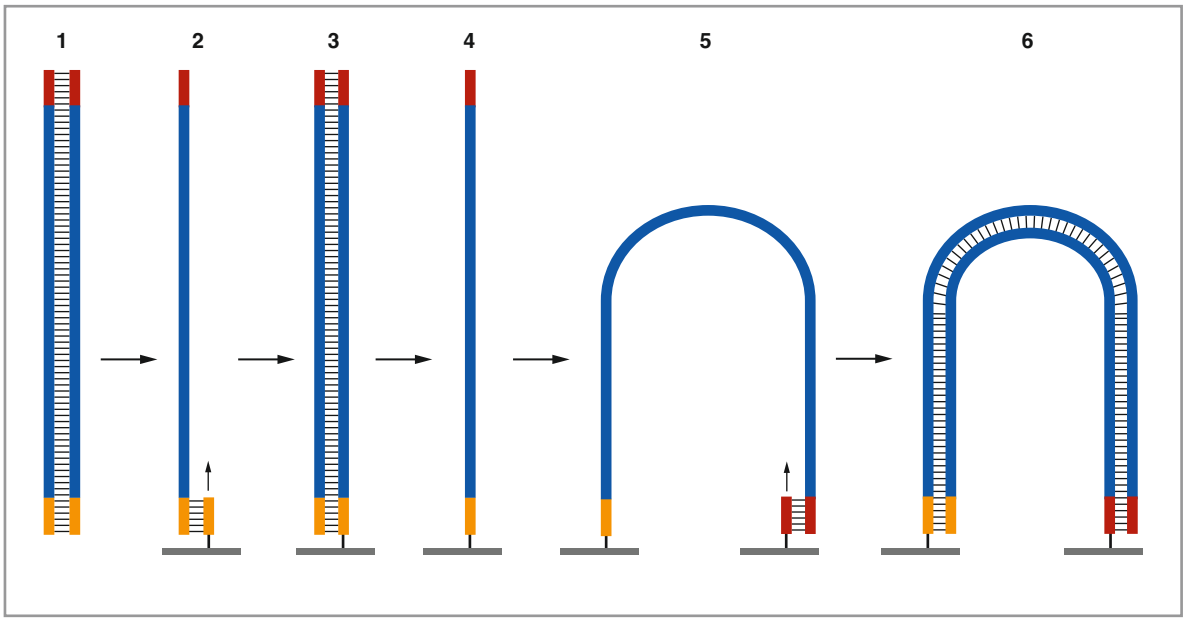


Abb. 2.26 Prinzip der *Bridge*-PCR bei der Illumina-Sequenzierung. (1) An die zu sequenzierende doppelsträngige DNA (blau) werden zunächst zwei unterschiedliche Adapter-Primer ligiert (orange und rot). (2) Die DNA wird dann in eine „Reaktionskammer" (*flow cell*) gegeben, in der die beiden Adapter-Primer als Einzelstränge kovalent an die Oberfläche gebunden sind. Die zu sequenzierende DNA wird denaturiert und über Hybridisierung mit dem fixierten Adapter-Primer (orange) in der *flow cell* gebunden. (3) Der DNA-Strang wird nun von einer DNA-Polymerase kopiert. (4) Danach wird der entstandene DNA-Doppelstrang denaturiert, sodass der entstehende Einzelstrang über den ersten Adapter-Primer (orange) an die Oberfläche gebunden ist und am anderen Ende den zweiten Adapter-Primer trägt (rot). (5) Über diesen kann die DNA jetzt auch mit dem zweiten Ende an die Oberfläche binden (*Bridge*-Bildung), indem die roten Adapterprimer hybridisieren. (6) Nun wird die DNA vom roten Primer ausgehend kopiert. Wird diese DNA wieder denaturiert, hängen die Einzelstränge mit einem Ende jeweils an der Oberfläche und können nach erneuter *Bridge*-Bildung wieder kopiert und die DNA im Sinne einer konventionellen PCR exponentiell vermehrt werden. Es entstehen dichte Cluster von Amplimeren, die dann durch Terminator-Technologie sequenziert werden (siehe Text).

in einer Sequenzreaktion alle vier Terminator-Nukleotide zu und erlaubt den Einbau der jeweils komplementären Base durch eine DNA-Polymerase. Die eingebaute Base wird anhand der Fluoreszenz des Terminator-Nukleotids bestimmt und dann werden die Schutzgruppe am 3'-OH-Ende und der Fluoreszenzfarbstoff abgespalten und ausgewaschen, bevor der nächste Zyklus beginnt.

Aufgrund der im Vergleich mit der Sanger-Sequenzierung großen produzierbaren Datenmengen und der immer wirtschaftlicher werdenden Illumina-Sequenzierung eröffnen sich viele Möglichkeiten der Anwendung dieser Technologie in der Genomforschung. Hier nur einige Anwendungsbeispiele:

- Komplexe Lebensgemeinschaften in interessanten ökologischen Systemen, z. B. in Biofilmen, können allein durch Sequenzierung der darin enthaltenen Genom-DNA charakterisiert werden, d. h. ohne die in den Proben enthaltenen Lebewesen durch Kultivierung anreichern zu müssen. Dieser neue Zweig der Genomik wird auch als „Metagenomik" bezeichnet.
- Genome von Organismen lassen sich komplett aufklären, deren wirtschaftliche Sequenzierung allein aufgrund ihrer Genomgröße bisher nicht denkbar war. Dies gilt insbesondere auch für tierische Genome, deren Analyse unser Wissen über Funktionen menschlicher Gene bereichern wird (Modellorganismen), was wiederum zum Verständnis pathophysiologischer Prozesse und damit zur Arzneimittelentwicklung beitragen kann.
- Für die Suche nach krankheitsrelevanten DNA-Sequenz-Polymorphismen des Menschen können Genome von Individuen ganz oder teilweise resequenziert werden.
- Die Expression aller zu einem bestimmten Zeitpunkt in einer Zelle exprimierten mRNAs kann in einem einzigen Experiment bestimmt werden. Diese als „mRNA-Sequenzierung" oder RNA-seq bekannte Methode hat große Bedeutung, beispielsweise für die Charakterisierung von Tumoren, da man heute weiß, dass jeder Tumor ein individuelles (vom gesunden Gewebe verschiedenes) Genexpressionsmuster aufweist. Das tumorspezifische Genexpressionsmuster könnte dann als Biomarker für die Entscheidung

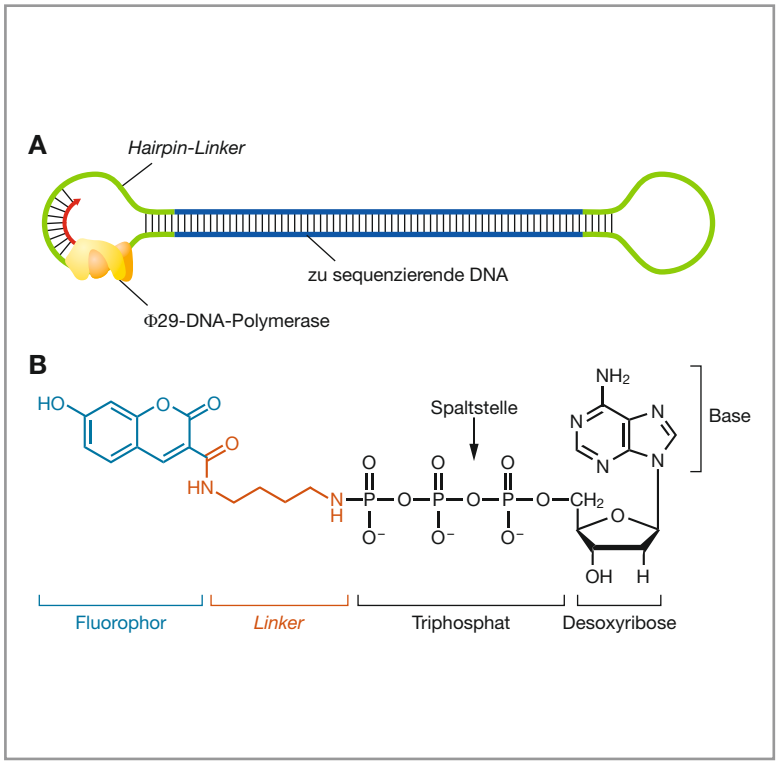


Abb. 2.27 PacBio-Sequenzierung. **A** SMRTbell-Matrizen bestehen aus einem zu sequenzierenden DNA-Doppelstrang (blau), an den *Hairpin-Linker* (grün) ligiert werden. Wenn diese Moleküle in eine ZMW diffundieren, werden sie von der kovalent an die Oberfläche der ZMW fixierte Φ29-Polymerase gebunden. Die Polymerase kann im einzelsträngigen Bereich der *Hairpins* mit der DNA-Polymerase beginnen. Da die SMRTbell-Matrizen über die *Hairpins* geschlossen sind, kann die Polymerase mehrere Umläufe der DNA-Synthese absolvieren, sodass konkatenierte Sequenzen der Matrizen-DNA entstehen. Die Syntheserichtung ist durch den roten Pfeil angedeutet. **B** Beispiel für ein in der PacBio-Sequenzierung genutztes, über das Triphosphat-Ende fluoreszenzmarkiertes Desoxynukleotid. Beim Einbau eines solchen Nukleotids während der DNA-Synthese wird der Pyrophosphatrest mitsamt dem Fluoreszenzfarbstoff abgespalten.

über ein geeignetes Therapieregime herangezogen werden. Übrigens wird bei der RNA-seq nicht die mRNA selbst sequenziert. Vielmehr werden alle mRNAs einer Zelle durch die Reverse Transkriptase in doppelsträngige cDNA umgewandelt und diese wird dann sequenziert.

- Auch zur Validierung der Herstellungsverfahren von Biopharmazeutika wird die Hochdurchsatz-DNA-Sequenzierung wertvolle Beiträge leisten, z. B. durch die vollständige Resequenzierung der Genome von Produktions-Zelllinien vor und nach einem Produktionsprozess.

Hochdurchsatz-Sequenzierung wird ständig weiterentwickelt, meist mit dem Ziel, noch mehr Sequenzinformation in kürzerer Zeit und mit wirtschaftlicheren Methoden zu generieren. Die Illumina-Sequenzierung wird mittlerweile als „2. Generation" der Sequenziermethoden angesehen (die Sanger-Sequenzierung ist die 1. Generation), um sie von weiteren Neuentwicklungen wie beispielsweise der PacBio-Sequenzierung abzugrenzen, die als Drittgenerationen-Sequenziermethode bezeichnet wird.

Bei der PacBio-Sequenzierung steht nicht primär die Menge an produzierter Sequenzinformation im Vordergrund, sondern es wird darauf optimiert, möglichst lange, fortlaufende DNA-Sequenzen zu erzeugen. Das grundsätzliche Problem mit der Illumina-Sequenzierung ist nämlich, dass sie zwar große Menge an Einzelsequenzen liefert, diese aber mit ca. 100–300 bp recht kurz sind. Wenn man beispielsweise in Genomprojekten versucht, unbekannte genomische Sequenzen anhand von überlappenden Einzelsequenzen virtuell zusammenzusetzen (zu „assemblieren"), funktioniert das gut, solange die erzeugten Illumina-Sequenzen einzigartig sind und nicht mehrmals an verschiedenen Stellen des untersuchten Genoms vorkommen. Sind die Illumina-Sequenzen aber so kurz, dass sie nur repetitive und keine einzigartige Sequenzinformation enthalten, kann man sie keiner bestimmten Position im Genom zuordnen. PacBio-Sequenzen schaffen hier einen Ausweg, da sie mehrere 10 000 bp lange, ununterbrochene Sequenzen erzeugen und damit auch repetitive Abschnitte eines Genoms überspannen können.

PacBio-Sequenzierung wurde von der Firma Pacific BioSciences entwickelt, wodurch sich der Name dieser Drittgenerationen-Sequenziermethode erklärt. Bei dieser Technik werden DNA-Einzelmoleküle sequenziert, und die Sequenzen werden in Echtzeit (*real time*) aufgezeichnet, weshalb die Methode auch als *single-molecule real-time sequencing* oder SMRT-Sequenzierung bezeichnet wird. Zur Vorbereitung der Sequenzierung werden an einzelne DNA-Moleküle *Hairpin*-Adapter ligiert (Abb. 2.27). Die dadurch erhaltenen SMRTbell-Matrizen werden in eine sogenannte SMRT-Zelle gegeben, die aus bis zu 150 000 einzelnen Vertiefungen besteht und als *zero-mode waveguides* (ZMW) bezeichnet werden. Am Boden jedes ZMW ist ein *einziges*

Molekül einer DNA-Polymerase gebunden. Bevorzugt wird hier die DNA-Polymerase des Bakteriophagen Φ29, die aus einer einzelnen Proteinkette besteht und eine hohe Prozessivität besitzt, also sehr lange DNA-Polymere herstellten kann. Der Sequenzierreaktion werden die vier erforderlichen Desoxynukleosidtriphosphate (dNTPs) zugesetzt, wobei jedes Nukleotid über das Triphosphat-Ende mit einem anderen Fluoreszenzfarbstoff verbunden ist (Beispiel in ○ Abb. 2.27). Die Φ29-Polymerase bindet im einzelsträngigen *Hairpin*-Bereich der SMRTbell-Matrizen und beginnt die DNA-Synthese. Die ZMW sind so gestaltet, dass das Fluoreszenzlicht der dNTPs nur in dem extrem kleinen Volumen um die DNA-Polymerase herum erfasst wird. Während die DNA-Polymerase die SMRTbell-Matrizen kopiert, wird das jeweils erforderliche dNTP kurzzeitig an das Enzym gebunden und damit das Fluoreszenzsignal des jeweiligen dNTPs gemessen. Sobald das Enzym das gebundene Nukleotid in den bereits polymerisierten DNA-Strang einbaut, wird das Pyrophosphat mitsamt dem Fluoreszenzfarbstoff abgespalten und diffundiert aus dem Detektionsvolumen, sodass das Fluoreszenzsignal erlischt. Aus dem zyklischen Binden und Hydrolysieren eines dNTPs entsprechend der Komplementarität des kopierten Matrizen-Strangs lässt sich also die Sequenzinformation ableiten. Da die SMRTbell-Matrizen nicht jedes ZMW besetzen bzw. auch mehrere DNA-Moleküle in dasselbe ZMW diffundieren können, sind in einem Lauf pro SMRT-Zelle meist nur zwischen 35 000 und 75 000 ZMWs mit Einzelmolekülen besetzt und liefern auswertbare Sequenzinformation. Dennoch werden mit der Methode innerhalb weniger Stunden bis zu 1000 Milliarden Basen an Sequenzinformation gewonnen, wobei jede Einzelsequenz mehrere 10 000 bp lang sein kann.

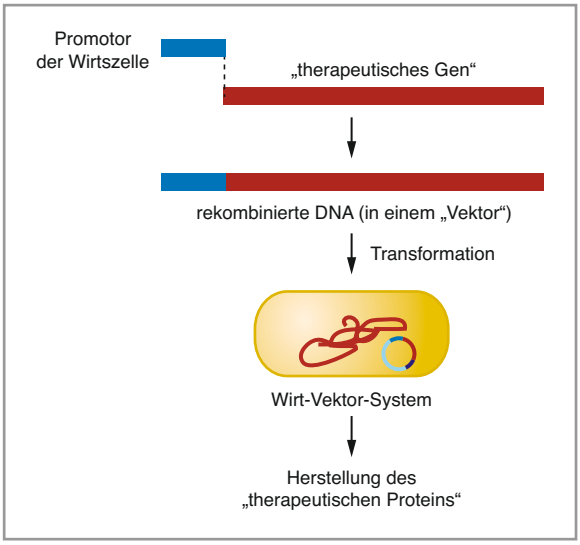
Ein Nachteil der PacBio-Sequenzierung ist derzeit noch die hohe Fehlerrate der Sequenzierung, die bei 10–15 % liegt. Das ist viel zu hoch, um valide Information zu liefern, wenn man bedenkt, dass beispielsweise maximal 0,01 % Fehlerrate in Genomprojekten als akzeptabel angesehen wird. Die hohe Fehlerrate der PacBio-Sequenzierung wird dadurch relativiert, dass die Φ29-Polymerase je nach Länge der kopierten SMRTbell-Matrizen in der Lage ist, mehrere „Runden zu drehen", indem sie am Ende des DNA-Moleküls den *Hairpin* kopiert und auf dem Gegenstrang weiter synthetisiert. Die resultierenden konkatenierten DNA-Sequenzen werden im Computer übereinandergelegt und ein „Konsensus" erstellt, der dann der am häufigsten ermittelten und deshalb fehlerlosen Sequenz entspricht.

Zusammenfassung

- Für die Synthese größerer DNA-Abschnitte in vitro sind DNA-Polymerasen das Mittel der Wahl. Diese Enzyme können einzelsträngige DNA-Abschnitte kopieren, indem sie den fehlenden DNA-Strang entsprechend der Komplementaritätsregeln mit Nukleotiden auffüllen.
- DNA-Polymerasen benötigen dabei einen Startpunkt aus doppelsträngiger DNA. Dieser Startpunkt kann künstlich festgelegt werden, indem gezielt kurze Oligonukleotide (Primer) entworfen werden, die an die zu kopierenden Einzelstrang-DNAs hybridisieren.
- Eine der wichtigsten Methoden der molekularbiologischen Forschung und der Gendiagnostik ist die Polymerase-Kettenreaktion (PCR).
- Die PCR kann eine DNA-Sequenz in einem beliebig komplexen genetischen Hintergrund sehr empfindlich nachweisen und den untersuchten DNA-Abschnitt zu Mengen vervielfältigen, die eine Analyse der DNA-Sequenz und damit eine Feststellung der Identität der DNA ermöglicht.
- Die Analyse der Basensequenz einer DNA beruht auf einer Modifikation der enzymatischen DNA-Synthese, bei der eine definierte Mischung von normalen Desoxynukleotiden und artifiziellen Didesoxynukleotiden während des Kopierens eines DNA-Abschnitts vorliegen. Dabei werden DNA-Fragmente erzeugt, die sich in ihrer Länge um jeweils eine Base unterscheiden.
- Sensitive Nachweismethoden erlauben heute eine automatisierte Hochdurchsatz-Sequenzierung, die die Grundlage für die Durchführung komplexer Genomprojekte bildet. Die Sequenziertechniken werden ständig weiterentwickelt mit dem Ziel, möglichst viel DNA-Information mit geringerem Aufwand und Kosten zu produzieren.

2.4 Prinzipien der Gentechnik

DNA-Rekombinationstechnologie (Gentechnologie) ist für die Herstellung von Proteinwirkstoffen von besonderer Bedeutung. Die Herstellung von Proteinwirkstoffen erfordert die Expression von Genen, die in den meisten Fällen aus dem menschlichen Genom stammen, in einem nicht-menschlichen Zellsystem. Die Genexpression läuft in allen Lebewesen prinzipiell gleich ab, d. h. die in einem Gen codierte Information wird in allen Organismen prinzipiell gleich umgesetzt (▸ Kap. 3). Allerdings braucht jedes Gen ein Steuerelement (Promotor), damit die Transkription gestartet werden kann. Eben diese Promotoren sind aber spezifisch für Organismen und innerhalb mehrzelliger Organismen sogar zelltypspezifisch. Eine Expression eines menschlichen Gens in einem nicht-menschlichen Zellsystem kann also nur gelingen, wenn wir Methoden haben, um die jeweiligen zellspezifischen Promotoren mit der gewünschten Gensequenz zu verbinden – also DNA-Moleküle zu rekombinieren. Nicht zuletzt deshalb ist im Europäischen Arzneibuch in der Monographie „DNA-rekombinationstechnisch hergestellte Produkte“ von „Wirt-Vektor-Systemen“ die Rede, wobei der „Vektor“ jeweils die rekombinante Transkriptionseinheit beschreibt (○ Abb. 2.28).



○ **Abb. 2.28** Wirt-Vektor-System. Eine Einheit zur Herstellung von Biopharmazeutika besteht aus der genetischen Information („therapeutisches Gen“), die als Protein exprimiert und als Wirkstoff etabliert werden soll („therapeutisches Protein“). Zur Herstellung der Expressionseinheit muss das therapeutische Gen mit einem in der Wirtszelle aktiven Promotor durch DNA-Rekombination verbunden werden. Diese funktionelle Expressionseinheit wird dann durch Transformation in eine geeignete Wirtszelle eingebracht, die das therapeutische Protein produziert.

Obwohl die generelle Struktur der DNA seit den klassischen Arbeiten von James Watson und Francis Crick seit Mitte der 1950er Jahre bekannt war und es auch klar wurde, dass die genetische Information in der Sequenz der Basen innerhalb einer DNA festgelegt ist, war es lange undenkbar, die riesigen DNA-Moleküle detailliert zu charakterisieren, die Basenfolgen zu ermitteln oder gar unterschiedliche DNA-Moleküle miteinander zu verbinden. Es fehlte die erforderliche Technologie. Möglich wurde dies erst, als Anfang der 1970er Jahre zwei Voraussetzungen geschaffen waren:

- Enzyme wurden entdeckt und charakterisiert, die DNA als Substrat erkennen und modifizieren.
- Methoden wurden entwickelt, mit denen es möglich war, DNA zu rekombinieren.

Um die Werkzeuge der Gentechnologen kennen zu lernen, wollen wir uns zunächst mit den Enzymen befassen, die DNA als Substrat erkennen. Danach werden wir sehen, dass das Bakterium *Escherichia coli* mit Abstand der wichtigste Organismus innerhalb der molekularbiologischen Forschung mit gentechnischen Methoden ist. Fast jede gentechnische Modifikation wird zunächst in *E. coli* realisiert, auch wenn geplant ist, die modifizierte genetische Information später in einem anderen Organismus zu exprimieren.

Warum müssen wir neu kombinierte DNA in *E. coli* amplifizieren? Rekombinationen von DNA-Fragmenten müssen auf der Ebene *einzelner Moleküle* durchgeführt werden, was ein extrem ineffizienter Vorgang ist. Man steht daher regelmäßig vor dem Problem, dass man untersuchen muss, ob ein gewünschtes Rekombinationsereignis überhaupt stattgefunden hat und man benötigt ein Werkzeug, um diese rekombinierte DNA von negativen Rekombinationsereignissen zu trennen. Es wäre durchaus eine Herausforderung, auf der Ebene einzelner Moleküle zu analysieren, ob das Zusammenfügen eines DNA-Moleküls mit einem anderen DNA-Molekül wirklich stattgefunden hat. Geschweige denn, dass man diese rekombinierte DNA in ausreichender Menge zur weiteren Verwendung in die Hand bekäme. Aus diesem Dilemma gibt es aber einen Ausweg: die Durchführung der DNA-Rekombination in Plasmiden und eine Amplifikation dieser Plasmide in *E.-coli*-Zellen. Wird ein Plasmid in eine Bakterienzelle eingebracht, wird nicht nur das Plasmid in der einzelnen Bakterienzelle amplifiziert, sondern auch die Bakterien selbst. Diese vermehren sich als Kolonie von einer einzelnen Zelle zu mehreren Millionen Zellen. Da es sich innerhalb einer Bakterienkolonie um genetisch identische Zellen, also Klone, handelt, wird der gesamte Vorgang der Amplifikation von DNA-Molekülen in *E. coli* im Rahmen der DNA-Rekombination (des gentechnischen Arbeitens) im Labor auch „Klonieren“ genannt.

An folgender Beispielrechnung wird deutlich, dass eine einfache Kultivierung von Bakterienzellen aus

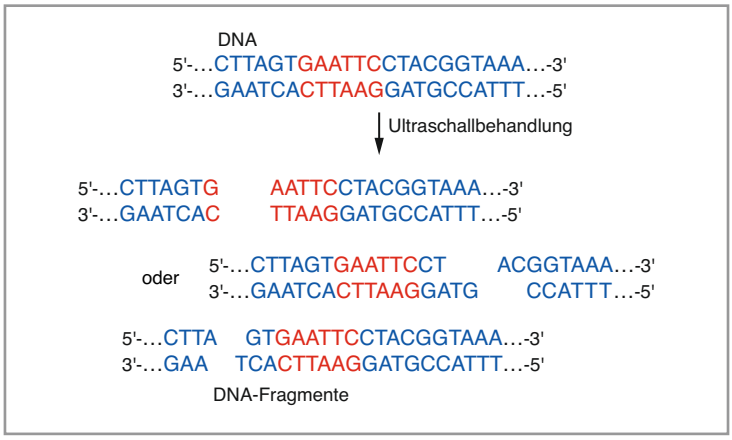


Abb. 2.29 Unspezifisches Fragmentieren von DNA. Längere DNA-Abschnitte können unspezifisch gebrochen werden, z. B. indem man die DNA mit Ultraschall behandelt. Wäre unser Ziel, die rot eingefärbte DNA-Sequenz gezielt zu brechen, wäre dies mit dieser Methode nicht möglich, da es Zufall ist, an welchem Ort ein Doppelstrangbruch jeweils auftritt. Auch die durch die Brüche resultierenden DNA-Enden sind verschieden. Dies erschwert das gezielte Rekombinieren verschiedener DNA-Abschnitte.

einem *Molekül* rekombinierter DNA, die in ein geeignetes Vektorsystem integriert und in eine Bakterienzelle eingeschleust wurde, quasi über Nacht zu einer DNA-*Menge* amplifiziert werden kann, die wir mit den gängigen DNA-modifizierenden Enzymen und DNA-analytischen Verfahren leicht erfassen können:

- Wir gehen davon aus, dass bei jeder Transformation eines Bakteriums lediglich ein *einziges* neu rekombiniertes Plasmid-Molekül in die Zelle gelangt.
- Die Plasmide werden innerhalb *einer* Bakterienzelle stark amplifiziert (100–700 Plasmid-Kopien pro Zelle).
- Wird die transformierte Bakterienzelle in Flüssigkultur vermehrt, lässt sich die Dichte der Bakterienkultur durch die Absorption (A) bei 600 nm messen: 1 ml einer Bakterienkultur der Dichte $A_{600} = 1$ entspricht ca. 8×10^8 Bakterien.
- Enthält jede Bakterienzelle einer solchen Kultur nur ca. 100 Kopien des Plasmids, entspricht das einem Amplifikationsfaktor von 8×10^{10}; das entspricht 1,3 $\times 10^{-13}$ mol Plasmid einer Größe von 3000 bp oder 0,3 µg Plasmid. Diese *Menge* an Plasmid-DNA entspricht einer gut sichtbaren „Bande" im Agarosegel, wenn wir die DNA mit Ethidiumbromid anfärben, d. h. wir können durch herkömmliche DNA-Analytik untersuchen, ob das isolierte Plasmid rekombinant ist oder nicht.

Gentechnisches Arbeiten bedeutet die Isolierung, Modifizierung und Verknüpfung bestimmter DNA-Moleküle zu neuen Funktionseinheiten. Damit die neukombinierte DNA schließlich in *E. coli* stabil erhalten bleibt, sind geeignete Vektorsysteme wie z. B. Plasmide nötig und geeignete Methoden, mit deren Hilfe sich die neu kombinierte DNA effizient in die *E.-coli*-Zellen einschleusen lässt. Bevor wir uns diesen Methoden widmen, wollen wir uns zunächst anschauen, wie wir DNA-Moleküle in einer bestimmten, vorhersagbaren Weise miteinander verbinden können. Der Schlüssel hierfür sind die Restriktionsendonukleasen.

2.4.1 Restriktionsendonukleasen

Wie können wir DNA-Fragmente gezielt neu kombinieren? DNA kann beispielsweise durch mechanisches Scheren durch Behandlung mit Ultraschall zerbrochen werden (Abb. 2.29). An welchen Stellen die DNA bricht, ist jedoch zufällig, sodass wir mit dieser Methode Gene nicht gezielt isolieren und klonieren können. Für die DNA-Rekombination brauchen wir eine Methode, mit der DNA spezifisch, d. h. vorhersagbar, fragmentiert werden kann. Dies erreichen wir durch Hydrolyse doppelsträngiger DNA mit Restriktionsendonukleasen (im Labor-Jargon auch kurz als „Restriktionsenzyme" bezeichnet). Der Einsatz dieser Enzyme hat gegenüber dem unspezifischen Scheren der DNA den Vorteil, dass Restriktionsendonukleasen ausschließlich an spezifischen DNA-Sequenzen schneiden und dabei charakteristische Enden erzeugt werden, die durch die Spezifität der verwendeten Restriktionsenzyme bestimmt werden und daher vorhersagbar sind.

Wir unterscheiden vier große Gruppen von Restriktionsendonukleasen. Allen diesen Enzymen ist gemeinsam, dass sie ausnahmslos doppelsträngige DNA als Substrate verwenden und dass sie bestimmte DNA-Sequenzmotive erkennen und binden, bevor sie die DNA schneiden (Abb. 2.30).

- Typ-I-Restriktionsendonukleasen sind komplexe Enzyme, die sowohl Nuklease- als auch Methyltransferase-Aktivität besitzen. Diese Enzyme benötigen als Cofaktoren ATP, *S*-Adenosylmethionin und Magnesiumionen. Die eigentliche Hydrolyse der DNA geschieht nicht an der Erkennungsstelle des Enzyms auf der DNA, sondern in einer Entfernung von einigen hundert Basenpaaren. Aus diesem Grund werden Typ-I-Restriktionsendonukleasen

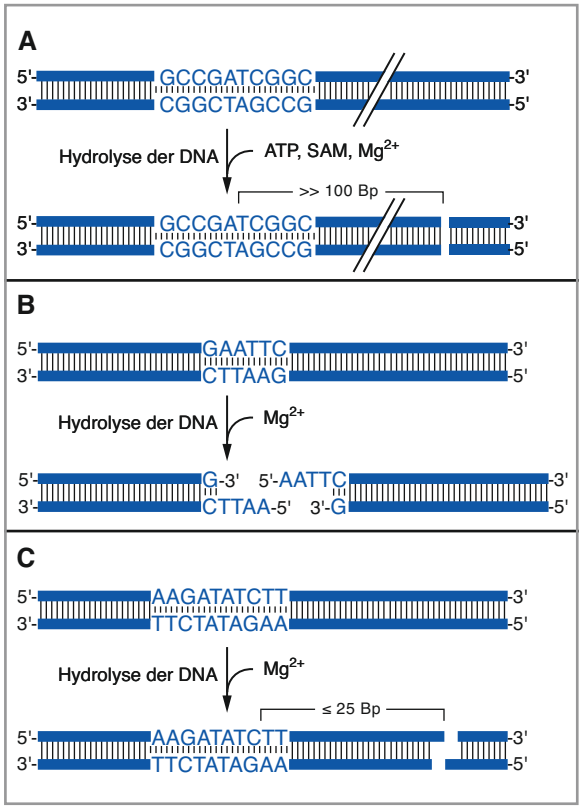


Abb. 2.30 Die drei Hauptgruppen der Restriktionsendonukleasen. **A** Typ-I-Restriktionsendonukleasen benötigen als Cosubstrate ATP, SAM (*S*-Adenosylmethionin) und Mg^{2+}. Die Hydrolyse erfolgt weit entfernt von der Erkennungssequenz. Diese Restriktionsendonukleasen spielen in der Gentechnologie praktisch keine Rolle. **B** Typ-II-Restriktionsendonukleasen hydrolysieren die DNA innerhalb der jeweils spezifischen Erkennungssequenz. Spalten sie diese Sequenz nicht versetzt, entstehen DNA-Fragmente mit „glatten Enden". Spalten die Enzyme die Erkennungssequenz versetzt, entstehen DNA-Fragmente, bei denen entweder das 5'-Ende über das 3'-Ende oder aber das 3'-Ende über das 5'-Ende hinausragt. In der Gentechnologie gehören die Typ-II-Restriktionsendonukleasen zu den wichtigsten Enzymen. **C** Typ-III-Restriktionsendonukleasen hydrolysieren die DNA – ähnlich wie die Typ-I-Restriktionsendonukleasen – außerhalb der Erkennungssequenz. Allerdings liegen die Spaltstellen deutlich näher an der Erkennungssequenz, sodass diese Restriktionsendonukleasen teilweise auch in der Gentechnologie verwendet werden.

praktisch nie für gezielte Klonierungsstrategien verwendet.

- Typ-II-Restriktionsendonukleasen benötigen meist nur Magnesium als Cofaktor und spalten die DNA innerhalb der Erkennungssequenz. Die große Mehrzahl der in der Molekularbiologie verwendeten Enzyme sind Typ-II-Restriktionsendonukleasen. Das hängt damit zusammen, dass beliebige DNA-Moleküle, egal aus welcher Spezies sie stammen, nach Hydrolyse mit demselben Restriktionsenzym die gleichen DNA-Enden aufweisen. Wir können also DNA-Abschnitte, die wir mit derselben Restriktionsendonuklease geschnitten haben, über ihre kompatiblen Enden miteinander verbinden (rekombinieren). Die Typ-II-Restriktionsendonukleasen sind nicht mit Methyltransferasen assoziiert. Allerdings können die Erkennungsstellen für Typ-II-Restriktionsendonukleasen von separaten Methyltransferasen modifiziert werden. Dies kann dazu führen, dass die Restriktionsendonuklease ihre Zielsequenz nicht mehr hydrolysieren kann. Umgekehrt gibt es aber auch Restriktionsendonukleasen, die eine Spaltstelle nur dann erkennen, wenn in dieser Sequenz eine Base methyliert ist. Die unterschiedlichen Eigenschaften und Anforderungen der Typ-II-Restriktionsendonukleasen hat zu ihrer weiteren Aufteilung in elf Untergruppen (A, B, C, E, F, G, H, M, P, S und T) geführt. Die für die Klonierung von DNA relevanten Restriktionsendonuklasen sind überwiegend in der Untergruppe Typ IIP zu finden, in die Enzyme eingeteilt werden, die symmetrische Erkennungssequenzen auch symmetrisch schneiden.
- Die Typ-III-Restriktionsendonukleasen sind komplexe Enzyme, die eine DNA ATP-abhängig in einem spezifischen Abstand – in der Regel bis zu 25 bp – von der Erkennungssequenz entfernt hydrolysieren. Aus diesem Grund sind die überlappenden Sequenzen, die von Typ-III-Restriktionsendonukleasen erzeugt werden, nicht miteinander kompatibel. Diese Restriktionsenzyme werden daher eher selten für Klonierungen verwendet.
- Typ-IV-Restriktionsendonukleasen schneiden nur modifizierte, also methylierte, hydroxymethylierte oder glucosylhydroxyethylierte DNA. Bisher ist nur wenig über diese Enzyme bekannt, ihre Erkennungssequenzen scheinen recht unspezifisch zu sein. Deshalb sind diese Enzyme für den molekularbiologischen Laboralltag irrelevant.

Für das Rekombinieren von DNA haben die Typ-IIP-Restriktionsendonukleasen die größte Bedeutung. An dieser Stelle sollten wir uns fragen, worin denn eigentlich eine Erkennungsstelle für eine Restriktionsendonuklease besteht? In der Regel ist die Erkennungssequenz eine palindromartige Sequenz. Palindrome sind bekanntlich Wörter oder Sätze, die von vorne und von hinten gelesen den gleichen Sinn ergeben. Beispiele sind „Otto" oder „Anna". So ist es auch mit den Erkennungssequenzen der Restriktionsendonukleasen (Abb. 2.31). Allerdings gilt die Spiegelbildlichkeit nicht nur für die Sequenz auf einem einzelnen DNA-Strang, sondern für die Sequenz des DNA-Doppelstrangs. Eine typische Erkennungssequenz wäre z. B. die

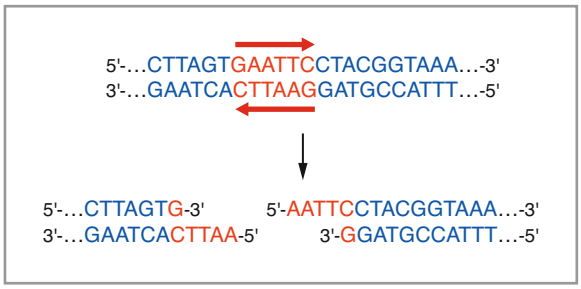


Abb. 2.31 Spezifisches Fragmentieren von DNA mit Typ-II-Restriktionsendonukleasen. Typ-II-Restriktionsendonukleasen erkennen eine spezifische DNA-Sequenz, hier rot markiert. Wir bemerken, dass diese Sequenz ein Palindrom bildet, d. h. sie lautet auf beiden DNA-Strängen – in antiparalleler Richtung gelesen – jeweils gleich. Die Restriktionsendonukleasen schneiden die DNA an der Erkennungsstelle in einer bestimmten Weise, sodass definierte Enden entstehen. Dies erleichtert das Zusammenfügen beliebiger DNA-Ende zu rekombinierten DNAs, sofern wir diese DNAs mit der gleichen Restriktionsendonuklease „geschnitten" haben.

Basenabfolge 5'-GAATTC-3' auf dem „oberen" DNA-Strang. Da die beiden komplementären DNA-Stränge antiparallel verlaufen, lautet die Sequenz des „unteren", komplementären Strangs ebenfalls 5'-GAATTC-3' (Abb. 2.31). Restriktionsendonukleasen funktionieren in der Regel als Homodimere, wobei ein Monomer jeweils die Zielsequenz auf einem DNA-Strang erkennt und schneidet. Wenn beispielsweise die Dimer-Untereinheiten des Enzyms EcoRI gleichzeitig die Sequenz 5'-GAATTC-3' zwischen dem G und dem A schneiden, entsteht wie in Abb. 2.31 gezeigt ein kurzer einzelsträngiger „Überhang" an der Schnittstelle.

Typ-IIP-Restriktionsenzyme hydrolysieren doppelsträngige DNA auf bestimmte Art und Weise, sodass drei unterschiedliche Arten von Enden erzeugt werden:

- Die meisten Enzyme schneiden die Zielsequenz versetzt, aber symmetrisch zur Mitte. Dabei entstehen als Resultat DNA-Fragmente, bei denen ein Ende als einzelsträngige DNA über das andere Ende hinausragt. In Abb. 2.32 ist dies nochmals am Beispiel der Restriktionsendonuklease EcoRI gezeigt, die die DNA-Sequenz 5'-GAATTC-3' schneidet. Durch die versetzte, aber symmetrische Spaltung des DNA-Doppelstrangs entstehen DNA-Enden, bei denen das 5'-Ende übersteht. Die überstehenden Enden von zwei DNA-Fragmenten, die mit demselben Restriktionsenzym geschnitten wurden, können sich durch Hybridisierung der komplementären Basen aneinanderlagern. Man bezeichnet diese Art Endstück als „klebrige" DNA-Enden (*sticky ends*).
- Umgekehrt kann eine Restriktionsendonuklease den DNA-Doppelstrang symmetrisch in der Weise schneiden, dass das 3'-Ende über das 5'-Ende hinausragt. Als Beispiel dieses Enzymtyps ist in Abb. 2.32 KpnI genannt.
- Als dritte Möglichkeit können Enzyme die Erkennungssequenz nicht versetzt hydrolysieren, d. h. die DNA-Stränge sind an den beiden Enden gleich lang. Hierzu zählt das in Abb. 2.32 aufgeführte Enzym PvuII. Solche DNA-Enden werden als „stumpf" (*blunt ends*) bezeichnet.

Restriktionsendonukleasen sind wichtige Werkzeuge in der Gentechnologie, die für viele Aufgaben eingesetzt werden können. Einige dieser Aufgaben werden wir noch kennen lernen:

- Schneiden wir DNA-Abschnitte unterschiedlicher Herkunft mit der gleichen Restriktionsendonuklease, können wir über die kompatiblen Enden z. B. ein menschliches Gen mit den Kontrollregionen eines nicht-humanen Wirtsorganismus verbinden, um ein Wirt-Vektor-System für die Herstellung eines Proteinwirkstoffs für die Therapie einer Krankheit zu etablieren.
- Mithilfe verschiedener Restriktionsendonukleasen können wir das gesamte Genom eines Organismus in einzelne DNA-Fragmente zerlegen, um z. B. krankheitsrelevante Gene zu kartieren und zu isolieren oder Genome in ihrer Ganzheit zu sequenzieren.

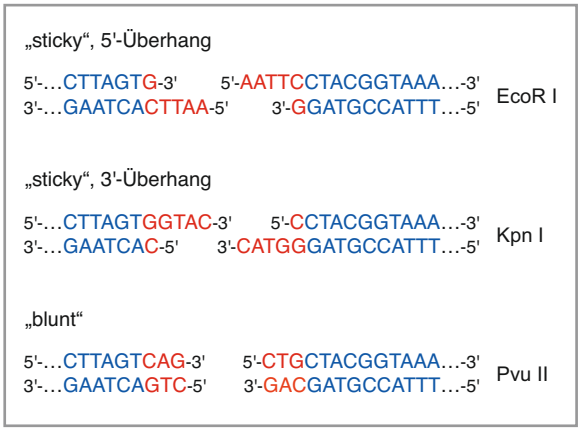


Abb. 2.32 Produkte von Typ-II-Restriktionsendonukleasen. Restriktionsendonukleasen schneiden DNA meistens versetzt, sodass entweder 5'- oder 3'-überhängende einzelsträngige Enden entstehen. Ein Zusammenfügen solcher Enden wird durch Hybridisierung der vier überstehenden Basen erleichtert, weshalb man solche Enden als „klebrig" oder *sticky* bezeichnet. Einige Restriktionsendonukleasen schneiden die DNA gerade durch, sodass „stumpfe" oder *blunt* Enden entstehen. Da das Zusammenfügen stumpfer Enden nicht von der Hybridisierung überstehender Basen erleichtert wird, können wir beliebige stumpfe Enden zusammenfügen, um DNAs miteinander zu rekombinieren.

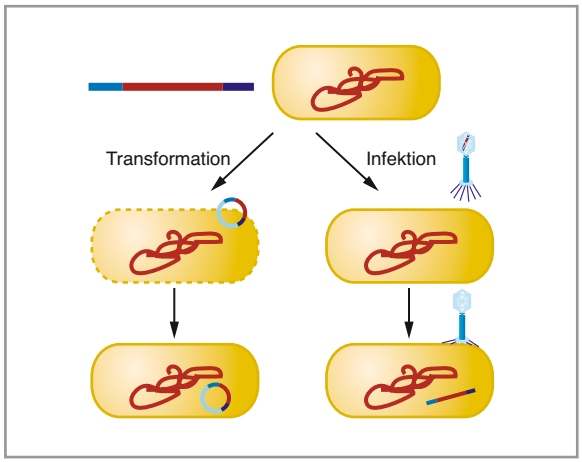


Abb. 2.33 Transformation *versus* Infektion. Wird DNA passiv über eine vorgeschädigte Zellhülle in eine bakterielle Zelle aufgenommen, bezeichnet man das als Transformation. Die aktive Aufnahme von Fremd-DNA mithilfe von Phagen bezeichnet man hingegen als Infektion.

- In der Gendiagnostik können wir das Vorhandensein oder das Fehlen von Erkennungsstellen für Restriktionsendonukleasen dazu nutzen, um Mutationen in Genen zu diagnostizieren. Man spricht von Restriktionsfragmentlängen-Polymorphismen.

2.4.2 Transformationsmethoden für Bakterien

Einer der wichtigsten Schritte bei der Etablierung eines rekombinanten Organismus ist das Einbringen der zuvor in vitro rekombinierten DNA in eine Bakterienzelle. Dieser Prozess wird als Transformation bezeichnet, wenn die DNA auf chemischem oder physikalischem Weg in die Zelle eingebracht wird (Abb. 2.33). Im Gegensatz dazu spricht man von Infektion, wenn die DNA mithilfe eines Bakteriophagen, also auf biologischem Weg, in eine *E.-coli*-Zelle eingeschleust wird (Abb. 2.33).

Transformieren lassen sich sogenannte „kompetente" *E.-coli*-Zellen (Abb. 2.34). Deren Zellwände und Membranen wurden durch Behandlung beispielsweise mit $CaCl_2$ temporär für Fremd-DNA durchlässig gemacht. Für die eigentliche Transformation werden die kompetenten Zellen zusammen mit der zu transformierenden DNA in einer $CaCl_2$-Lösung 30 min bei 0 °C inkubiert. Durch die Ca^{2+}-Ionen in der Zellsuspension wird die DNA als Polyanion mikrokristallin auf der Oberfläche der kompetenten Zellen präzipitiert. Danach folgt eine Inkubation für etwa 4 min bei 37 °C, und schließlich werden die Zellen 10 min bei 25 °C belassen. Durch diese Temperaturwechsel wird die DNA über Mechanismen, die nicht genau bekannt sind, passiv in die Zelle aufgenommen. Zur Regeneration der Bakterienzellen wird der Transformationsansatz mit Medium verdünnt, 60 min bei 37 °C inkubiert und schließlich auf eine Agarplatte transferiert (Abb. 2.34). Die Transformationseffizienzen liegen bei diesem Verfahren bei 10^5–10^7 transformierten Zellen pro µg DNA. Dies hängt von der Güte der kompetenten Zellen und vom Transformationsprotokoll ab.

Deutlich höhere Transformationsraten erhält man durch ein Verfahren, das als Elektroporation bezeichnet wird (Abb. 2.34). Hierzu werden die Zellen, die zuvor salzfrei gewaschen wurden, zusammen mit der zu transformierenden DNA in einem Feld hoher elektrischer Konduktivität inkubiert. Bei der Entladung (z. B. 2500 V, 200 Ω, 25 µF) gelangt die DNA dann in die Zellen. Derartige Systeme erreichen Transformationseffizienzen von bis zu 10^{10} Transformanten pro µg DNA.

2.4.3 Vektoren für die Transformation von Bakterien

Plasmide

Plasmide sind natürlich vorkommende, zirkuläre DNA-Moleküle, die in Bakterienzellen unabhängig vom bakteriellen Chromosom repliziert werden. In den vergangenen Jahren wurde für die gentechnische Anwendung eine Vielzahl von Plasmiden entwickelt, die sich prinzipiell von natürlichen Plasmid-Vorbildern wie den R-Plasmiden ableiten. R-Plasmide tragen eines oder mehrere Resistenz-Gene für Antibiotika („R" steht für „Resistenz"). Außerdem codieren R-Plasmide für die sogenannten *tra*-Gene (tra = transfer), die erforderlich sind, damit Bakterien die R-Plasmide durch einen als Konjugation bezeichneten Prozess untereinander austauschen können. Bakterien können durch Aufnahme eines einzigen R-Plasmids Multiresistenzen gegen Antibiotika erwerben.

Für die gentechnische Nutzung hat man sich von den R-Plasmiden abgeschaut, was man tun muss, damit Plasmide in der Bakterienzelle „episomal", also unabhängig vom Bakterienchromosom, repliziert werden können. Außerdem hat man von R-Plasmiden gelernt, wie man die Expression der auf Plasmiden codierten Antibiotika-Resistenzgene nutzen kann, um das Einbringen artifizieller Plasmide in Bakterienzellen durch Transformation zu kontrollieren. Die natürlichen R-Plasmide selbst sind allerdings aus verschiedenen Gründen für die gentechnische Anwendung ungeeignet:

- R-Plasmide kommen nur in 1–2 Kopien pro Bakterienzelle vor.
- R-Plasmide sind mit ca. 100 000 bp zu groß und zu instabil für eine effiziente Neukombination mit fremden DNA-Fragmenten durch gentechnische Methoden.
- R-Plasmide können durch Konjugation zwischen Bakterien einer Kultur ausgetauscht werden, solange sie die *tra*-Gene tragen; dieser Vorgang ist der

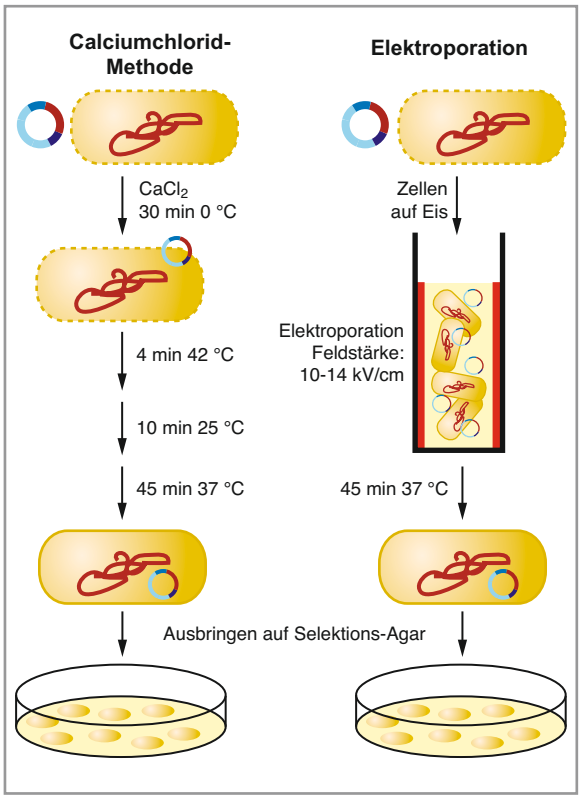


Abb. 2.34 Die bakterielle Transformation. Um Bakterien zu transformieren, müssen ihre Zellwände durch eine Behandlung mit $CaCl_2$ vorgeschädigt sein. Wir bezeichnen derartige Bakterien als „kompetent". Links: Calciumchlorid-Methode. Die Zellen werden in der $CaCl_2$-Lösung zunächst 30 min bei 0 °C inkubiert. Dann erhöhen wir die Inkubationstemperatur schnell auf 37 °C („Hitzeschock") und anschließend auf 25 °C. Zur Regeneration der Zellhülle verdünnen wir danach die $CaCl_2$-Lösung mit Medium und inkubieren bei der optimalen Wachstumstemperatur für *E. coli* von 37 °C. Schließlich werden die Zellen auf eine Schale mit einem geeigneten Selektionsagar ausplattiert. Rechts: Elektroporation. Bei der Elektroporation verwenden wir Zellen, die vor der Transformation salzfrei gewaschen werden. Diese „elektrokompetenten" Zellen werden zusammen mit der DNA in eine Küvette gegeben, die an zwei gegenüberliegenden Seiten mit dünnen Plattenelektroden ausgekleidet sind. Nach Anlegen hoher Feldstärken wird die DNA bei der Entladung in die Bakterienzellen aufgenommen. Zur Regeneration werden die Zellen für 60 min bei 37 °C inkubiert und danach auf Selektionsagar ausplattiert.

gewünschten klonalen Vereinzelung von gentechnisch herbeigeführten Rekombinationsereignissen natürlich abträglich und muss durch das Entfernen der *tra*-Gene unterbunden werden.

Allen in der Gentechnik eingesetzten Plasmiden ist daher gemeinsam, dass sie von R-Plasmiden nur noch die Sequenzen enthalten, die für das gentechnische Arbeiten absolut notwendig sind. Der Vorteil liegt dann darin, dass diese Vektoren wesentlich kleiner sind (3000–10 000 bp) als R-Plasmide, was sie deutlich besser handhabbar macht:

- Alle Vektoren besitzen eine Sequenz, die als Replikationsursprung (*origin of replication*, kurz *ori*) bezeichnet wird (Abb. 2.35). Diese Sequenz wird von der DNA-Polymerase aus *E. coli* als Initiationspunkt für die Replikation erkannt und stellt sicher, dass das Plasmid unabhängig von der genomischen DNA des Bakteriums repliziert werden kann und zusätzlich zur genomischen DNA in der Zelle persistiert.
- Ferner enthalten alle Vektoren einen oder mehrere selektierbare Marker in Form von Resistenzgenen für Antibiotika. Der verbreitetste Marker ist das Gen für die β-Lactamase (Abb. 2.35). Dieses Enzym verleiht Bakterienzellen die Fähigkeit, in Gegenwart des Antibiotikums Ampicillin zu wachsen. Somit sind Zellen, die ein Plasmid mit einem β-Lactamase-Gen aufnehmen, gegen Ampicillin resistent. Auch andere Antibiotikum-Resistenzen – z. B. gegen Tetracyclin, Chloramphenicol, Kanamycin oder Neomycin – können als Marker auf Plasmiden verwendet werden.
- Die dritte wichtige Komponente aller Klonierungsvektoren ist ein speziell ausgewiesener Bereich, der die zu klonierenden fremden DNA-Fragmente aufnehmen soll. Dieser Bereich, die sogenannte multiple Klonierungsstelle (*multiple cloning site*, MCS), enthält auf kleinem Raum eine Vielzahl von Erkennungssequenzen für unterschiedliche Restriktionsendonukleasen, die jeweils nur einmal in dem Plasmid schneiden (Abb. 2.35).

Wir können uns nun vorstellen, dass wir ein beliebiges DNA-Fragment in ein Plasmid einbauen können, indem wir das DNA-Fragment und den Vektor jeweils mit einem Restriktionsenzym schneiden, dessen Erkennungssequenz Teil der MCS des Plasmids ist. Wie können wir in einem DNA-Rekombinationsexperiment unterscheiden, ob ein Plasmid eine uns interessierende DNA aufgenommen hat? Um dies zu untersuchen wird häufig das β-Galactosidase-Gen (*lacZ*) aus *E. coli* als Indikator-Gen auf dem Klonierplasmid verwendet (Abb. 2.36). Die β-Galactosidase ist ein 116-kDa-Protein, das in seiner aktiven Form ein Homotetramer bildet. Jedes Monomer kann funktionell in zwei Teile unterteilt werden: die N-terminale α-Domäne und die C-terminale ω-Domäne. Interessanterweise kann man α und ω gentechnisch voneinander trennen und als zwei getrennte Proteine exprimieren, wobei beide Fragmente für sich jeweils inaktiv sind. Werden aber α und ω in derselben Zelle exprimiert, können sie spontan assoziieren und so eine funktionelle β-Galactosidase rekonstituieren. Deshalb bezeichnet man α auch als das

α-Komplement oder den α-Donor, während ω häufig als α-Akzeptor bezeichnet wird.

Fast alle modernen Klonierungsvektoren enthalten die codierende Region des α-Komplements der β-Galactosidase. Bemerkenswerterweise liegt die MCS eines Klonierungsvektors in der Regel innerhalb des α-Komplements. Dabei wird die MCS so in das Gen für das α-Komplement inseriert, dass sie den Leserahmen für die β-Galactosidase-mRNA nicht zerstört. Deshalb bleibt das α-Komplement trotz der inserierten MCS aktiv und kann mit dem ω-Fragment eine aktive β-Galactosidase bilden. Bauen wir Fremd-DNA in die MCS ein, wird der Leserahmen für das α-Komplement durch die neu eingebrachte DNA unterbrochen und die enzymatische Aktivität der β-Galactosidase geht verloren. Diese Inaktivierung lässt sich durch einen Farbtest nachweisen. Die transformierten Zellen werden hierzu in Kulturschalen kultiviert, deren Agar mit dem Indikator X-Gal (5-Bromo-4-chloro-3-indolyl-β-D-galactopyranosid) imprägniert wurde. Wird die β-galactosidische Bindung dieses Indikators durch die β-Galactosidase hydrolysiert, entsteht nach Luftoxidation der blaue Farbstoff 5,5'-Dibromo-4,4'-dichloroindigo und die Bakterienkolonien färben sich blau. Wurde allerdings durch Insertion einer Fremd-DNA in die MCS das Gen für die β-Galactosidase inaktiviert, bleiben die Kolonien weiß. So lassen sich diese Zellen als Träger eines rekombinanten Plasmids als weiße Kolonien identifizieren (Abb. 2.36).

Klonierungsvektoren liegen in einer Anzahl von mehr als 50 Kopien, in der Regel jedoch zu mehreren Hundert Kopien pro Zelle vor. Solche Plasmide werden als *High-copy*-Vektoren bezeichnet. Hohe Kopienzahlen sind meist vorteilhaft und erwünscht, da man die Plasmid-DNA aus relativ wenigen Bakterienzellen in guter Ausbeute isolieren kann. Manchmal ist es jedoch sinnvoll oder sogar erforderlich, mit speziellen Plasmiden zu arbeiten, die in geringerer Kopienzahl in *E. coli* persistieren. Dies ist beispielsweise dann der Fall, wenn die klonierte DNA besonders groß ist, sodass die vielen Plasmid-Kopien „normaler" Vektoren die Zelle zu stark mit DNA überfrachten würden. Bedeutsam ist der Einsatz von *Low-copy*-Vektoren (Kopienzahlen < 20) aber auch, wenn das Produkt der klonierten DNA für die *E.-coli*-Zelle unverträglich ist. Dies führt dazu, dass die Zellen schlecht oder gar nicht wachsen. Würden hier *High-copy*-Vektoren verwendet, wäre zu viel genetisches Material in der Zelle, das für ein „unverträgliches Protein" codieren würde.

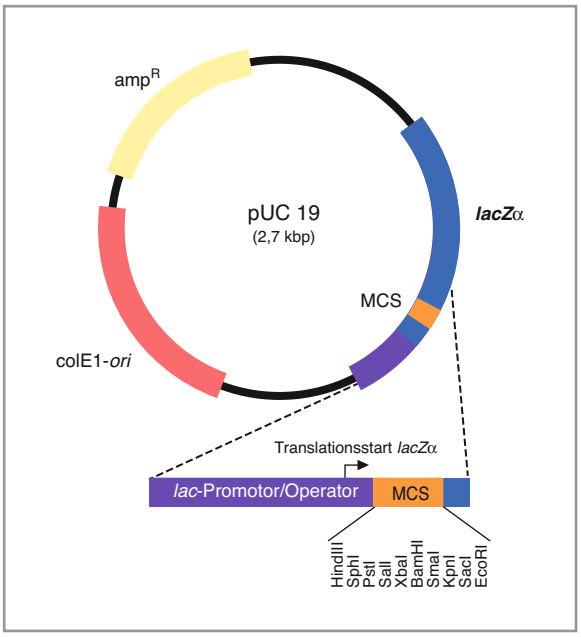


Abb. 2.35 Ein typischer Kloniervektor. Das Plasmid pUC19 enthält den codierenden Bereich für das α-Komplement der β-Galactosidase als Reportergen (*lacZα*, blau). Die Transkription dieses Gens wird durch den regulierbaren *lac*-Promotor/Operator kontrolliert (Abb. 4.9 in ▸Kap. 4.2). Im vorderen Bereich des β-Galactosidase-Gens befindet sich eine multiple Klonierungsstelle (MCS, orange), die sich dadurch auszeichnet, dass sie eine Vielzahl von Erkennungssequenzen für Restriktionsendonukleasen enthält, die das Plasmid jeweils nur einmal schneiden. Wird in die MCS ein DNA-Fragment einligiert, wird das β-Galactosidase-Gen zerstört. Dies ist die Basis für das spezifische Reportersystem („Blau-weiß-Selektion"), das in Abbildung Abb. 2.36 erklärt ist. Das Plasmid enthält weiterhin ein Resistenzgen für ein Antibiotikum (β-Lactamase bzw. amp^R) und einen bakteriellen Replikationsursprung für Plasmide (colE1-*ori*).

M13-Phagen und Phagemid-Vektoren

M13 gehört zu der Gruppe der filamentösen Phagen. Diese enthalten ein einzelsträngiges, ringförmiges DNA-Molekül als Genom. Der Umgang mit M13-Phagen war zu Beginn der Gentechnologie essenziell, da die von Frederick Sanger etablierte Methode der DNA-Sequenzierung zunächst ausschließlich mit Einzelstrang-DNA als Matrize durchgeführt werden konnte. Dies hat sich mit Einführung der *Cycle-Sequencing*-Variante der Sanger-Sequenzierung geändert, sodass der Umgang mit M13 stark zurückging. Wir wollen diese Phagen hier jedoch dennoch erwähnen, weil sie in der Entwicklung von vollständig humanen therapeutischen Antikörpern die Grundlage des sogenannten *Phage Display* bilden (▸Kap. 5).

Die Infektion der Bakterienzelle durch M13 erfolgt über den sogenannten F-Pilus (Abb. 2.37). Über diese Struktur können Bakterien untereinander genetisches Material austauschen. Nur „männliche" Bakterien tragen einen F-Pilus, da nur sie ein sogenanntes Fertilitäts-Plasmid (F-Plasmid) enthalten. Nach dem Eindringen in die Wirtszelle wird die einzelsträngige DNA des Phagen durch bakterielle Enzyme in einen DNA-Doppel-

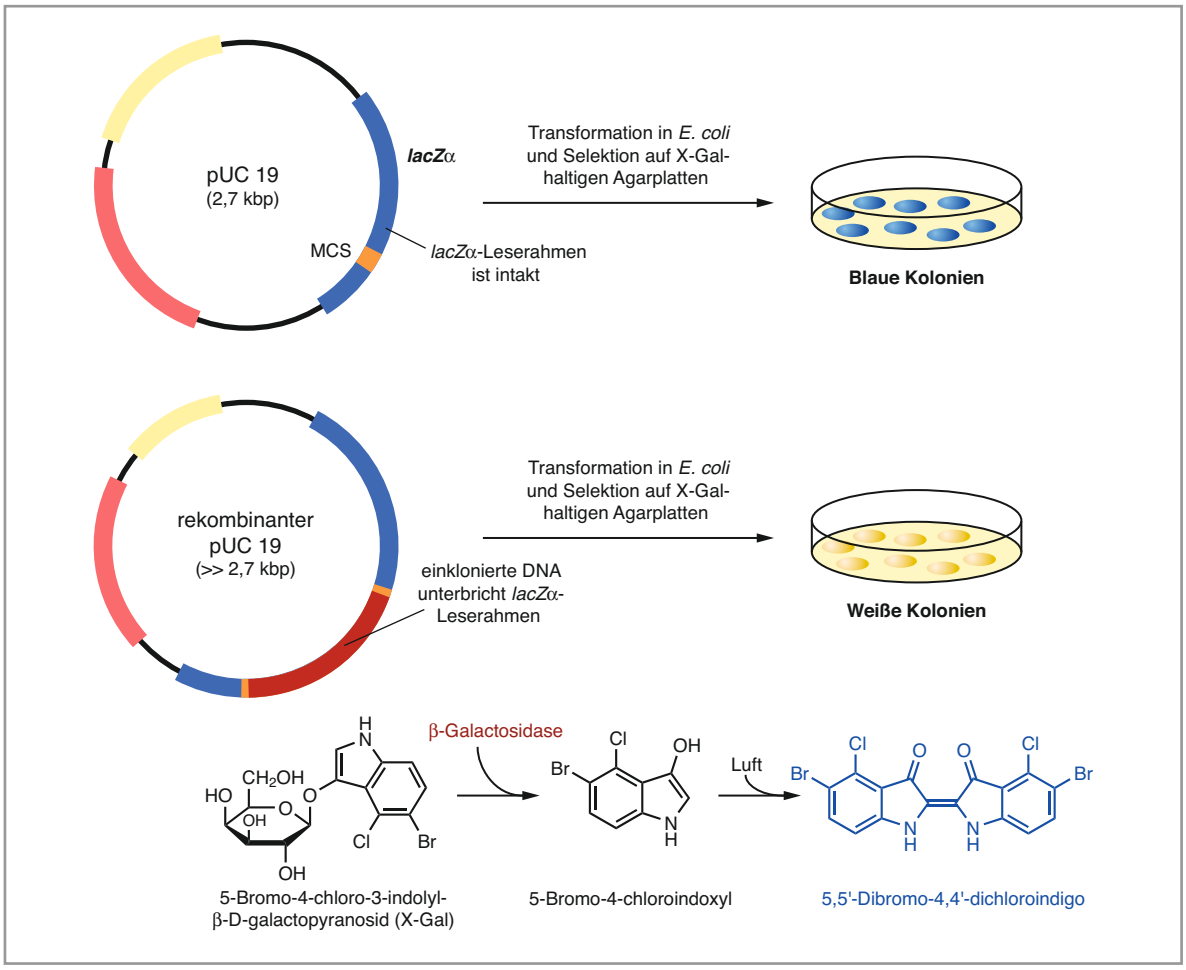


Abb. 2.36 Blau-weiß-Selektion mit dem β-Galactosidase-Reportergen. Zellen, die eine chromosomale Deletion im sogenannten α-Komplement der β-Galactosidase (*lacZα*) aufweisen, können kein aktives Enzym bilden. Das Gen für das fehlende α-Komplement kann durch Plasmide wie pUC19 (Abb. 2.35) in Zellen eingeführt werden (blauer Bereich auf dem pUC19-Plasmid). Zellen mit derartigen Plasmiden können also wieder β-Galactosidase synthetisieren, und wir können die Aktivität des Enzyms durch Zugabe des Substrats 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranosid (X-Gal) nachweisen. X-Gal wird von der β-Galactosidase gespalten und dann zu einem blauen Indigofarbstoff oxidiert. Die Kulturen färben sich also blau. In Plasmiden wie pUC19 ist eine multiple Klonierungsstelle (MCS) in das Gen für das α-Komplement eingebaut. Die MCS unterbricht den Leserahmen für das α-Komplement nicht, sodass mit pUC19 transformierte Bakterienzellen in Gegenwart von X-Gal blaue Kolonien bilden. Die Klonierung eines DNA-Fragments in die MCS zerstört allerdings in der Regel den Leserahmen für das α-Komplement, sodass Bakterienzellen, die ein derartiges rekombinantes Plasmid aufgenommen haben, keine funktionsfähige β-Galactosidase synthetisieren können. Die Kolonien bleiben daher auch in Gegenwart des Indikators X-Gal farblos.

strang umgewandelt. Diese doppelsträngige DNA bezeichnet man als die replikative Form oder auch als RF-DNA (Abb. 2.37). Nur in der replikativen Form kann M13-DNA mit gentechnischen Methoden bearbeitet werden, da ja nur doppelsträngige DNA von Restriktionsenzymen geschnitten werden kann und damit für DNA-Rekombinationen zugänglich ist.

Ähnlich wie andere Vektoren, die für gentechnische Zwecke eingesetzt werden, enthalten M13-Vektoren die genetische Information für das α-Komplement der β-Galactosidase und eine multiple Klonierungsstelle (MCS). So können auch in diesem Vektorsystem rekombinante Klone von nichtrekombinanten Klonen über die Aktivität der β-Galactosidase (Blau-weiß-Selektion; vgl. Abb. 2.36) unterschieden werden. Infizierte Zellen erkennt man daran, dass diese langsamer wachsen als nicht infizierte Zellen. Auf einem konfluenten *E.-coli*-Rasen erscheinen daher punktförmige, lichtere Bereiche infizierter Zellen als sogenannte Plaques, weil diese Bereiche eine geringere Zelldichte aufweisen. Diese Plaques sind trüb, da M13-Phagen die Bakterienzellen nicht lysieren. Sie verlassen vielmehr die infizierten Zellen über den Pilus, also auf dem gleichen Weg, wie sie in die Zelle eingedrungen sind.

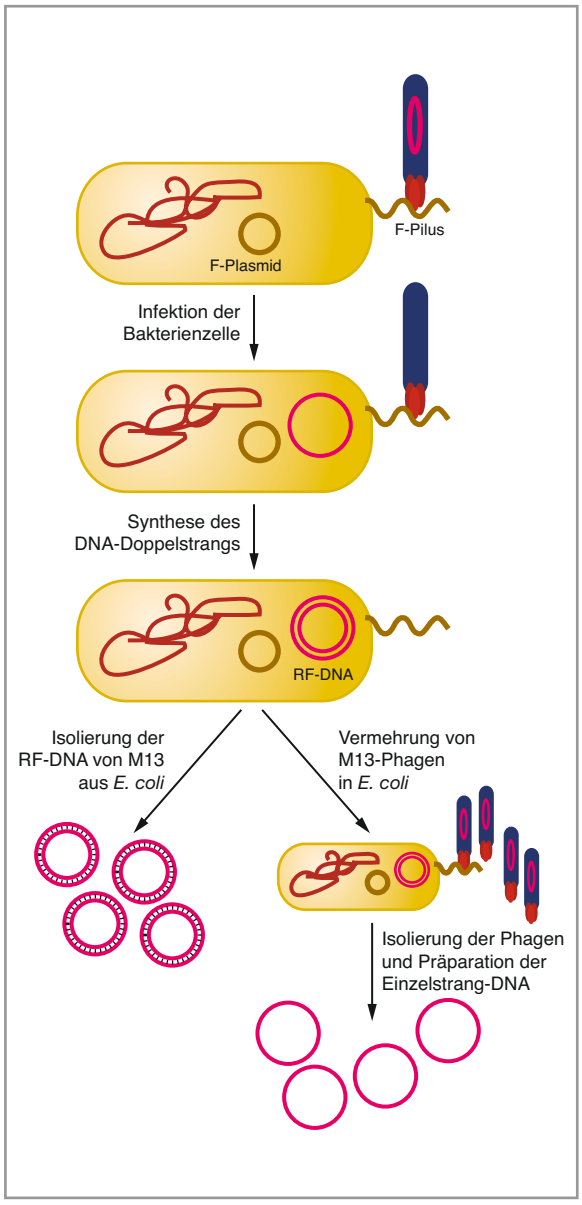


Abb. 2.37 M13-Phagen. M13 gehört zu den filamentösen Phagen. Dies sind *E.-coli*-spezifische Phagen, die jedoch nur solche *E.-coli*-Zellen infizieren, die ein Fertilitätsplasmid (F-Plasmid) besitzen und daher F-Pili ausbilden. Sie binden an diese F-Pili und entlassen über die Pili ihre DNA in das Zellinnere. Die DNA der filamentösen Phagen ist einzelsträngig und wird in der *E.-coli*-Zelle zu einer doppelsträngigen, replikativen Form (RF-DNA) ergänzt. Nur diese Variante ist für die Klonierung brauchbar, da Restriktionsendonukleasen nur doppelsträngige DNA hydrolysieren. Parallel zur Replikation des Phagengenoms werden auch die Strukturproteine der Phagen synthetisiert. Filamentöse Phagen lysieren die infizierten Zellen nicht. Vielmehr verlassen sie die Zellen wieder über den Pilus. Die Phagen – und damit die einzelsträngigen DNA-Formen – können aus dem Kulturüberstand präpariert werden. Die klonierbare, doppelsträngige Phagen-DNA wird dagegen konventionell wie andere Plasmid-DNA aus den infizierten Bakterienzellen isoliert.

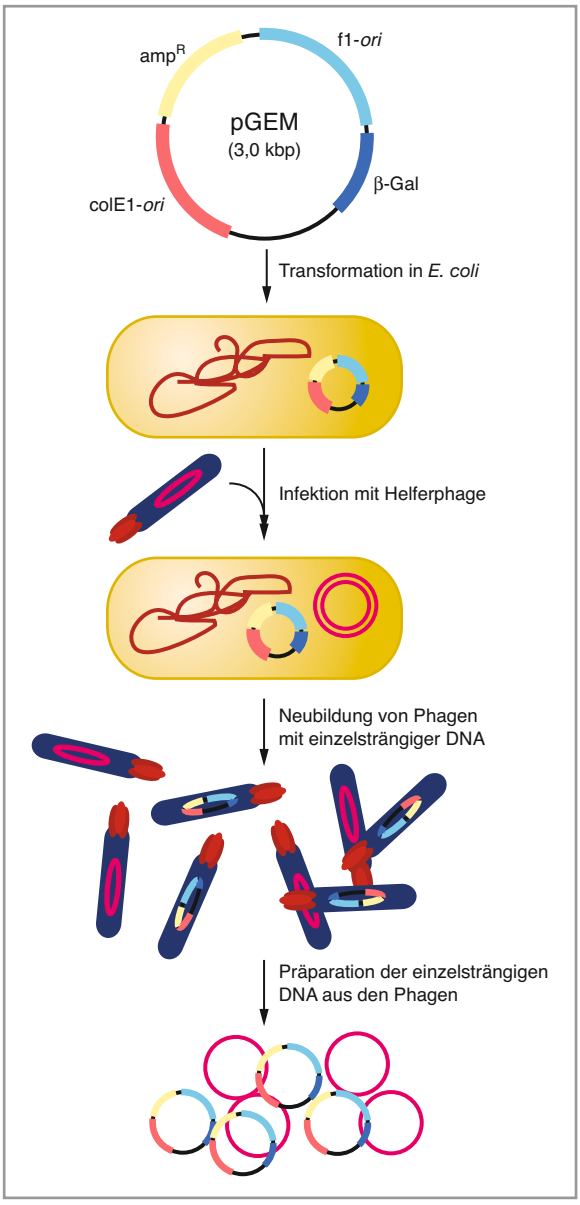


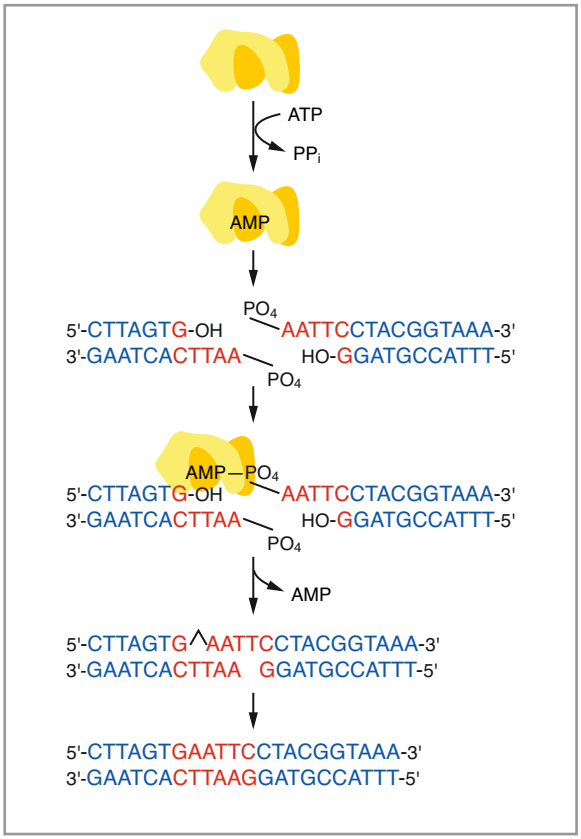
Abb. 2.38 Phagemid-Vektoren. Plasmide, die neben dem bakteriellen Replikationsursprung für Plasmide (ColE1-*ori*) auch einen Replikationsursprung eines filamentösen Bakteriophagen (hellbau) tragen, wie beispielweise die Plasmide der pGEM-Reihe, werden als Phagemide bezeichnet. Von Phagemid-Vektoren kann man auf elegante Art einzelsträngige DNA isolieren. Hierzu muss die Zelle, die ein derartiges Plasmid trägt, mit einem sogenannten Helferphagen infiziert werden. Dieser Phage liefert die Proteine, die zur Herstellung einzelsträngiger DNA und der phagenspezifischen Strukturproteine erforderlich sind. Man erhält dann eine Mischung von Phagen, die das Genom des Helferphagen oder des Phagemids in einzelsträngiger Form tragen. Wenn Helferphagen wie M13K07 benutzt werden, die weitgehend replikationsdefekt sind, isoliert man praktisch reine Phagemid-Einzelstrang-DNA.

Phagemid ist ein Kunstwort aus **Phage** und Plas**mid**. Phagemid-Plasmide (Phagemide) können entweder als normales Plasmid oder als in Phagenpartikel verpackte Einzelstrang-DNA existieren. Solche Plasmide sind der RF-Form filamentöser Phagen nicht unähnlich, tragen jedoch keine phagencodierten Gene und sind deshalb nicht infektiös. Phagemide tragen zwei Replikationsursprünge (*ori*). Der erste ist ein typischer Plasmid-*ori* und sorgt dafür, dass die Bakterienzelle die DNA wie ein normales Plasmid in Doppelstrang-Form repliziert. Dieses Plasmid kann mit den üblichen DNA-modifizierenden Enzymen bearbeitet werden. Zusätzlich trägt ein Phagemid aber noch einen Replikationsursprung für den M13-ähnlichen Phagen f1 (f1-*ori*; ○ Abb. 2.38). Infiziert man eine Bakterienzelle, die einen Phagemid-Vektor enthält, mit einem „f1-Helferphagen", lässt sich die Plasmid-DNA auch als DNA-Einzelstrang isolieren, da ein DNA-Strang des Plasmids durch den Helferphagen verpackt wird.

2.4.4 Rekombinieren von DNA-Fragmenten

Ligation mit der T4-DNA-Ligase

Die weitaus verbreitetste Methode, um freie DNA-Enden miteinander zu verbinden, die zuvor durch Restriktionsendonukleasen geschnitten wurden, ist die Nutzung der T4-DNA-Ligase. Die T4-DNA-Ligase schließt mithilfe der Cofaktoren Mg^{2+} und ATP eine Phosphodiesterbindung zwischen einer 3'-Hydroxylgruppe und einer benachbarten 5'-Phosphatgruppe in einem DNA-Strang; der Vorgang wird „Ligation" genannt. Dabei muss die DNA-Ligase zunächst aktiviert werden, indem ein Molekül AMP nach Hydrolyse von ATP auf einen Lysinrest im aktiven Zentrum des Enzyms übertragen wird. In diesem aktivierten Zustand kann das Enzym eine DNA nach freien 5'-Phosphatenden absuchen. Wird eine solche Stelle identifiziert, wird der AMP-Rest auf das 5'-Phosphatende der DNA in Form einer 5'→5'-Pyrophosphatstruktur übertragen. Ist ein freies 3'-OH-Ende in der DNA verfügbar, erfolgt anschließend eine Transesterifizierung unter Freisetzung von AMP und Verknüpfung der DNA-Enden (○ Abb. 2.39).



○ **Abb. 2.39** Verknüpfung von DNA-Enden durch T4-DNA-Ligase. Die T4-DNA-Ligase schließt mithilfe der Cofaktoren Mg^{2+} und ATP eine Phosphodiesterbindung zwischen einer 3'-Hydroxylgruppe und einer benachbarten 5'-Phosphatgruppe in einem DNA-Strang. Das Enzym wird aktiviert, indem nach Hydrolyse von ATP ein Molekül AMP an einen Lysinrest im aktiven Zentrum der Ligase bindet. Der AMP-Rest wird auf 5'-Phosphatenden von DNA mit Einzelstrangbrüchen übertragen, wobei eine 5'→5'-Pyrophosphatstruktur ausgebildet wird. Danach erfolgt eine Transesterifizierung mit einem benachbarten freien 3'-OH-Ende unter Abspaltung von AMP und Verknüpfung der DNA-Enden.

Ligation mit Rekombinasen

Eine Methode, die besonders geeignet zu sein scheint, große DNA-Fragmente miteinander zu verbinden, ist die sogenannte Gateway®-Klonierung (○ Abb. 2.40). Die Methode beruht auf der DNA-Sequenz-spezifischen Aktivität der Rekombinasen des Bakteriophagen Lambda (λ). Diese Integration wird durch eine vom λ-Phagen codierte Integrase gemeinsam mit dem *E. coli Integration Host Factor* (IHF) katalysiert, die spezifische *attachment*-Sequenzen im λ- und Bakteriengenom erkennen und eine intermolekulare DNA-Rekombination zwischen den beiden Genomen herbeiführen. Die *att*-Sequenz auf dem Bakterien-Chromosom wird als *attB* bezeichnet, die entsprechende Sequenz im λ-Genom heißt *attP*. Beide *att*-Sequenzen haben einen gemeinsamen Kern, der von den Rekombinasen miteinander ausgetauscht wird. Als Produkte der Rekombination werden zwei neue *att*-Regionen generiert, die *attL* und *attR* genannt werden. Die Rekombination zwischen *attB* und *attP* entspricht der Rekombination für die Integration des Phagen in das Bakteriengenom. Die umgekehrte Reaktion, zwischen *attL* und *attR*, entspricht dem Austritt des Prophagen aus dem Bakteriengenom und wird benötigt, um einen lytischen Zyklus zu starten. Das Ausschneiden des Prophagen aus dem *E.-coli*-Genom wird von einer „Excisionase" zusammen mit IHF katalysiert. Beide Reaktionen können in der

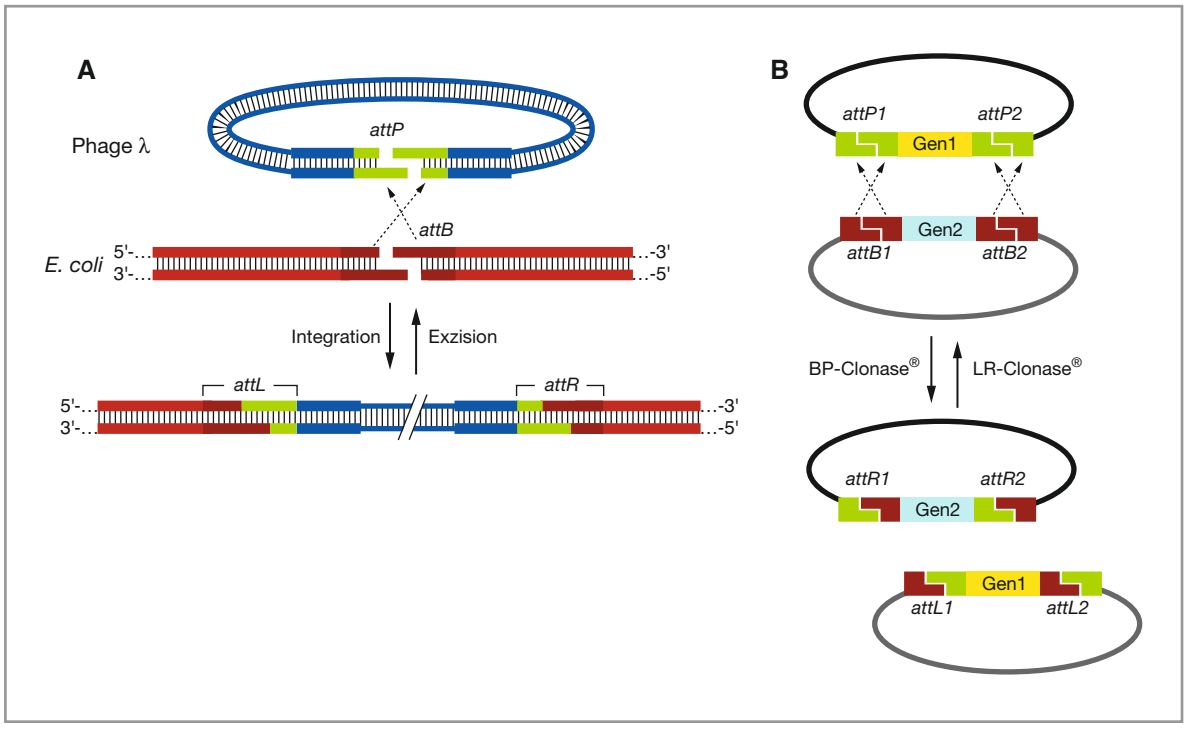


Abb. 2.40 Verknüpfung von DNA-Enden durch Rekombination. Die Gateway®-Klonierung basiert auf der DNA-Sequenz-spezifischen Rekombination durch Rekombinasen des Phagen λ. **A** Die Integration des Phagen λ in ein Bakteriengenom erfolgt über Rekombination zwischen spezifischen Sequenzabschnitten, die als *attachment*-Sequenz *attP* beim Phagen und *attB* im Bakterium bezeichnet werden. Dabei entstehen die neuen *att*-Sequenzen *attL* und *attR*. Für die Exzision der Phagen-DNA wird der umgekehrte Weg eingeschlagen. **B** Bei der Gateway®-Klonierung können größere DNA-Stücke über Rekombination zwischen Plasmiden ausgetauscht oder PCR-Fragmente in einen Vektor integriert werden. Voraussetzung ist, dass die auszutauschenden DNA-Stücke (Gen1 bzw. Gen2) von den *attachment*-Sequenzen *attP* bzw. *attB* flankiert werden. Nach erfolgter Rekombination zwischen *attP1* und *attB1* bzw. *attP2* und *attB2* befinden sich die beiden Gene im jeweils anderen Plasmid und sind von den *att*-Sequenzen *attL1*/*attL2* bzw. *attR1*/*attR2* umgeben. Die erforderliche Enzymmischung für diese Reaktion wird BP-Clonase® genannt. Die umgekehrte Reaktion, also eine Rekombination zwischen *attL* und *attR*, ergibt wieder *attB*- bzw. *attP*-Sequenzen. Auch in dieser Reaktion werden die zwischen den *att*-Sequenzen gelegenen Bereiche ausgetauscht. Das erforderliche Enzymsystem ist LR-Clonase®. Durch mehrmaligen Wechsel zwischen *attB*/*attP*- und *attL*/*attR* können DNA-Fragmente effizient zwischen mehreren Plasmidsystemen transferiert werden.

Gentechnologie genutzt werden, um DNA-Fragmente miteinander zu verbinden (Abb. 2.40).

Klonierung von PCR-Amplimeren durch A/U-Hybridisierung

Wie wir bereits besprochen haben, ist die Polymerase-Kettenreaktion (PCR) die Methode der Wahl zur schnellen Isolierung von DNA. Wollen wir beispielsweise ein bestimmtes menschliches Protein in Bakterien exprimieren, können wir die entsprechende proteincodierende cDNA durch PCR-Amplifikation gewinnen. Wir benötigen im Prinzip nicht einmal Kenntnis von der gesamten DNA-Sequenz des uns interessierenden Gens. Das Einzige, was wir brauchen, sind Informationen über die DNA-Sequenzen am Anfang und am Ende des Gens, um geeignete Bindestellen für die PCR-Primer festzulegen.

Was können wir aber tun, um die erhaltenen Amplimere effizient in ein Vektorsystem zu klonieren? Auf den ersten Blick sind PCR-Fragmente doppelsträngige DNA-Moleküle mit „stumpfen" Enden, also ohne Überhänge, wie sie z. B. die Restriktionsenzyme generieren. Es liegt also nahe, ein PCR-Produkt mithilfe der T4-DNA-Ligase in einen Vektor zu ligieren, den wir zuvor mit einem Restriktionsenzym geschnitten haben, das ebenfalls stumpfe Enden erzeugt. Diese Methode funktioniert allerdings in vielen Fällen nur recht ineffizient, weil PCR-Fragmente oft gar keine stumpfen Enden haben. Das liegt unter anderem daran, dass viele DNA-Polymerasen eine 3'-Adenylierungsaktivität besitzen und dazu neigen, an die 3'-Enden von PCR-Produkten jeweils ein einzelnes Adenosin-Nukleotid anzuhängen. Dieses Adenosin-Nukleotid bildet also quasi einen einzelsträngigen 3'-Überhang.

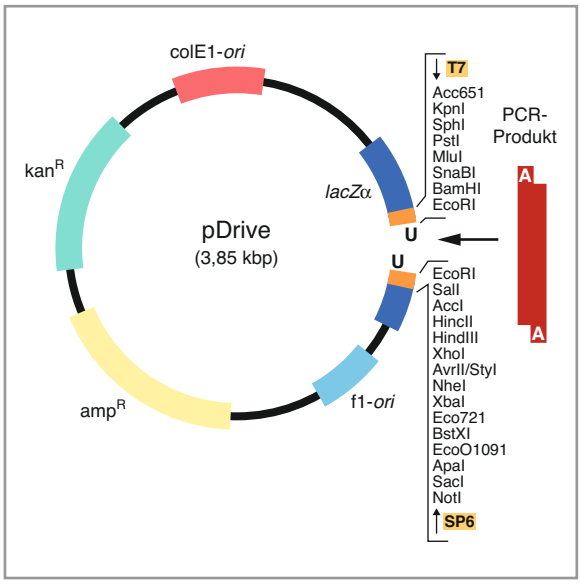


Abb. 2.41 Klonierung von PCR-Fragmenten über A/U-Hybridisierung. Gezeigt ist exemplarisch einer von vielen kommerziell erhältlichen Vektoren, die speziell dafür gemacht sind, effizient PCR-Produkte zu klonieren. Die Methode basiert auf der Beobachtung, dass viele in der PCR eingesetzten DNA-Polymerasen eine 3'-Adenylierungsaktivität besitzen. Die 3'-überhängenden Adeninbasen werden mit Uracil auf Seiten des linearisierten Vektors hybridisiert und dann mit T4-DNA-Ligase ligiert. Der Vektor pDRIVE enthält neben zwei Resistenzgenen für Antibiotika auch einen Replikationsursprung für den Phagen f1 und das Gen für das α-Komplement der β-Galactosidase, sodass eine Blau-weiß-Selektion möglich ist (Abb. 2.36).

Eine elegante Möglichkeit, PCR-Fragmente mit solchen 3'-überstehenden Adenosin-Nukleotiden zu klonieren, ist die sogenannte A/T- bzw. A/U-Hybridisierung. Dazu benutzt man einen linearisierten Vektor, der durch das Enzym Terminale Transferase mit einem 3'-überhängenden Thymidin- oder Uridin-Nukleotid ausgestattet wurde. Mischt man einen solchen Vektor mit einem PCR-Produkt, wird die Bindung der beiden DNA-Moleküle über die A/T- bzw. A/U-Hybridisierung kurzfristig stabilisiert, sodass die T4-DNA-Ligase die Phosphodiesterbindungen schließen und die DNA-Fragmente kovalent verknüpfen kann (Abb. 2.41).

Eine weitere Möglichkeit, PCR-Amplimere effizient in Vektoren zu inserieren, ist die sogenannte TOPO®-Klonierung. Bei dieser Methode werden DNA-Fragmente nicht durch die T4-DNA-Ligase miteinander verbunden, sondern mit einer Typ-1B-Topoisomerase. Die DNA-Topoisomerase 1 des Vacciniavirus bindet spezifisch an die Sequenz 5'-CCCTT-3' eines DNA-Strangs und katalysiert die Spaltung der Phosphodiesterbindung hinter dem letzten Thymidin der Zielsequenz. Die freiwerdende 3'-Phosphatgruppe der DNA wird dabei kovalent mit einem Tyrosinrest im aktiven Zentrum des Enzyms verbunden und dadurch die Energie der Phosphodiesterbindung konserviert. Dies ermöglicht eine Transesterifizierung zum Verbinden zweier DNA-Fragmente, d.h. die Übertragung der 3'-Phosphatgruppe auf das freie 5'-OH-Ende eines PCR-Fragments unter Freisetzung der Topoisomerase (Abb. 2.42).

Ein PCR-Fragment wird am effizientesten mit einem TOPO®-beladenen DNA-Fragment verbunden, wenn beide DNA-Fragmente komplementäre, überhängende Enden aufweisen. Diese kann man ebenfalls ohne Hilfe von Restriktionsenzymen nur durch die Aktivität der TOPO® erzeugen, wenn auch das PCR-Fragment an beiden Enden TOPO®-Bindestellen besitzt. Dazu müssen wir beim Design unserer PCR-Primer bedenken, dass zusätzlich zu dem Bereich des Primers, der komplementär zu der zu amplifizierenden DNA ist, eine um 11 Nukleotide verlängerte Sequenz enthalten sein muss, die sowohl die fünf Basen lange TOPO®-Bindestelle als auch eine sechs Basen lange, später 5'-überhängende Sequenz umfasst (Abb. 2.42).

Um die Effizienz bei der Klonierung von PCR-Produkten weiter zu erhöhen, können die A/T-Hybridisierung und die TOPO®-Ligation miteinander kombiniert werden. Der Vorteil dieser Methode ist, dass die PCR-Primer nicht um die elf Nukleotide längere TOPO®-Bindestelle plus eines 5'-Überhangs verlängert sein müssen. Vielmehr kann die 3'-Adenylierungsaktivität der Taq-Polymerase ausgenutzt werden, um jeweils ein Adenosin-Nukleotid an die 3'-Enden des PCR-Produkts anzufügen. Auf der anderen Seite hat der linearisierte TOPO®-Vektor ein überhängendes Thymidin-Nukleotid, an welches die TOPO® bereits durch den Hersteller des kommerziellen Plasmids kovalent gebunden wurde (Abb. 2.43). Nun wird zum einen die Bindung zwischen dem Vektor und dem PCR-Fragment durch A/T-Hybridisierung stabilisiert und zum anderen sehr effizient die kovalente Verknüpfung der DNA-Enden durch die TOPO®-Ligation erreicht.

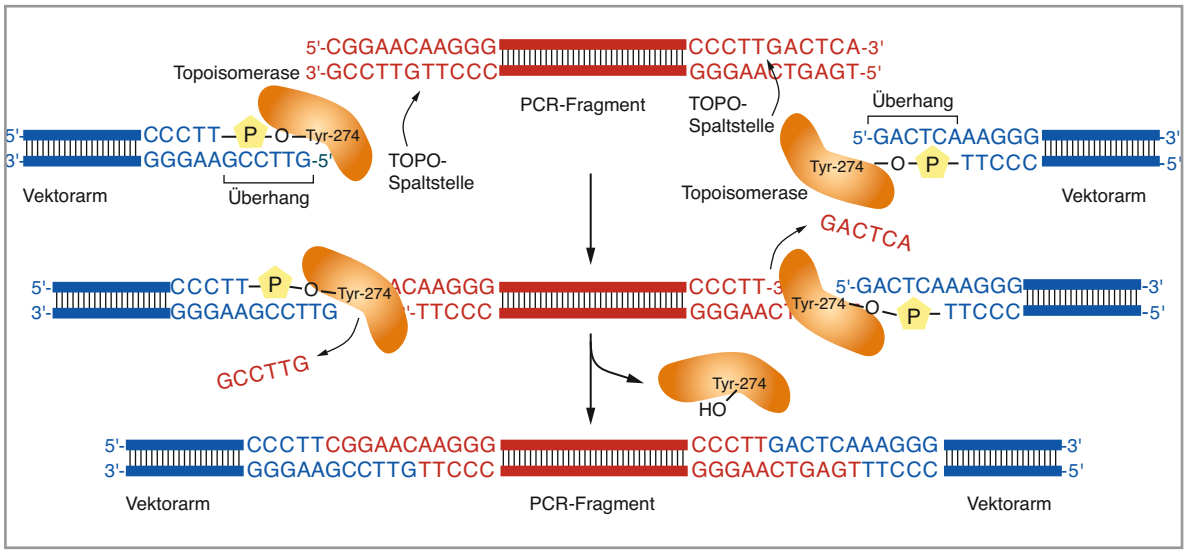


Abb. 2.42 Verknüpfung von DNA-Enden durch eine Typ-1-Topoisomerase. Die sogenannte TOPO®-Klonierung nutzt eine Typ-1B-Topoisomerase. In der Praxis liefert der Hersteller ein lineares Plasmid, dessen 3'-Enden bereits mit der Topoisomerase „beladen" sind. Dabei erfolgt die Bindung der DNA über eine Phosphatbrücke an einen Tyrosinrest (Tyr-274) im aktiven Zentrum des Enzyms. Die PCR-Primer werden so synthetisiert, dass sie jeweils unterschiedliche 6-bp-Sequenzen vor der Erkennungsstelle für die Topoisomerase (5'-CCCTT-3') enthalten. Die Topoisomerase spaltet das PCR-Fragment, sodass 6-bp-Überhänge entstehen, die komplementär zu den Überhängen im TOPO®-Vektor sind. Dies erlaubt eine effiziente und in der Richtung festgelegte Verbindung des PCR-Produkts mit beiden Vektorenden, indem die 3'-Phosphatgruppen des Vektors von der Topoisomerase auf die freien 5'-OH-Gruppen des PCR-Produkts übertragen werden.

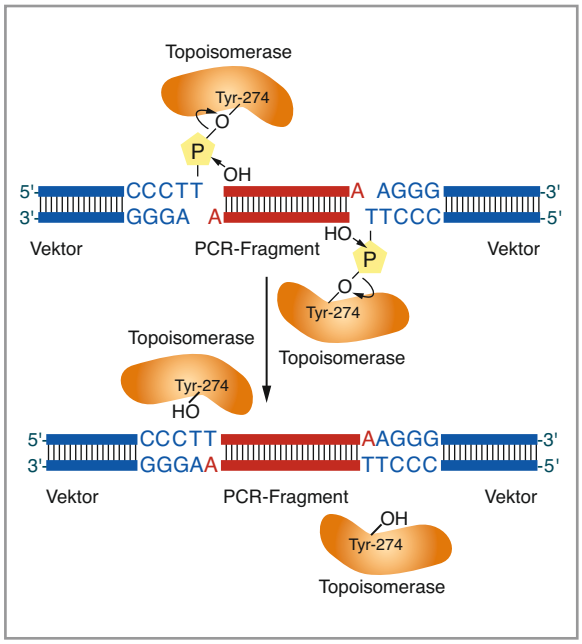


Abb. 2.43 Klonierung von PCR-Produkten durch Kombination von A/T-Hybridisierung und TOPO®-Ligation. Bei dieser Methode wird ein PCR-Produkt mit 3'-überhängenden Adenin-Basen mit einem kommerziell erhältlichen TOPO®-A/U-Vektor inkubiert, der bereits mit der TOPO®-Rekombinase beladen wurde.

Zusammenfassung

- Die DNA-Rekombinationstechnologie erlaubt die Verknüpfung zweier DNA-Fragmente zu neuen funktionellen Einheiten. Obwohl die eigentliche Verknüpfung von DNA-Fragmenten von Enzymen katalysiert wird, ist diese auf der Ebene einzelner Moleküle stattfindende Reaktion sehr ineffizient und benötigt ein Repertoire an Methoden, mit denen unterschieden werden kann, ob ein Plasmid rekombinant ist oder nicht.
- Um DNA-Rekombinationsereignisse analysieren zu können, muss die DNA zunächst von einem Einzelmolekül zu einer DNA-Menge amplifiziert werden, die mit herkömmlichen Methoden der DNA-Analytik erfassbar ist.
- Dabei bilden bakterielle Plasmide die Basis der DNA-Rekombinationstechnologie. Will man überprüfen, ob die Kombination von zwei DNA-Molekülen erfolgreich war, werden die Ligationsreaktionen zwischen Plasmiden und DNA-Fragmenten in Bakterien transformiert und die transformierten Bakterien klonal vereinzelt.
- Daher wird das Rekombinieren von DNA in der Laborsprache auch als „Klonieren" bezeichnet. Die am häufigsten verwendeten Vektoren sind bakterielle Plasmide, obwohl auch von Bakteriophagen abgeleitete Vektorsysteme existieren. Ein Plasmid ist dadurch gekennzeichnet, dass es einen eigenen Replikationsursprung besitzt und unabhängig vom bakteriellen Chromosom repliziert wird.
- In der Gentechnik verwendet man auf den Plasmiden natürlich vorkommende Resistenzgene gegen Antibiotika, um zu gewährleisten, dass die Plasmide in den Bakterienzellen verbleiben und dort repliziert werden.
- Die Techniken, mit denen DNA-Fragmente in vitro rekombiniert werden können, sind prinzipiell sehr ineffizient und werden ständig weiterentwickelt. Neuere Innovationen wie die Gateway®- oder TOPO®-Klonierung nutzen Rekombinasen anstelle von Ligasen für eine effiziente Konstruktion von Vektorsystemen.

3 Grundsätze der Proteinexpression

3.1 Umsetzung genetischer Information in Proteine 51
3.2 Genexpression in Prokaryonten 53
3.3 Genexpression in Eukaryonten 62
3.4 Posttranslationale Modifikationen an Proteinen 65

3.1 Umsetzung genetischer Information in Proteine

Proteine sind Makromoleküle, die aus einer linearen Abfolge von Aminosäuren bestehen. Die Abfolge der Aminosäuren in einem Protein, also die Aminosäuresequenz, wird als die Primärstruktur bezeichnet. Letztlich ist es jedoch die räumliche Struktur eines Proteins, die Sekundär- und Tertiärstruktur, die die Funktionalität eines Proteins ausmacht. Die Natur benutzt 20 verschiedene Aminosäuren zur Herstellung von Proteinen. Die physikochemischen Eigenschaften der Seitenketten der Aminosäuren in einem Protein bestimmen, in welche dreidimensionale Struktur (Konformation) sich ein Protein faltet. Dadurch wiederum wird bestimmt, welche Oberflächen Proteine nutzen können, um z. B. mit anderen Proteinen zu interagieren oder biochemische Reaktionen zu katalysieren.

Die Information, welche Aminosäure in einem Protein an welcher Stelle der Primärsequenz steht, ist in den Genen festgelegt. In der klassischen Genetik wurde der Begriff *Gen* ursprünglich als die kleinste vererbbare Einheit definiert. In der molekularen Genetik ist es eher gebräuchlich, ein Gen als eine Informationseinheit aufzufassen, die für ein Protein codiert. Allerdings ist diese Definition unvollständig, denn manche Gene enthalten Informationen für RNAs, die nicht in Proteine übersetzt werden, sondern als RNAs Funktionen in der Zelle ausüben. Auch neigen wir in Zeiten der molekularbiologischen Erforschung von Zellfunktionen dazu, die Kontrolleinheiten (Promotoren, Terminatoren), die die Umsetzung der genetischen Information in Proteine steuern, zu den Genen hinzuzurechnen. Heute ist der Begriff „Gen" daher besser als eine „Transkriptionseinheit" zu verstehen.

Der Fluss der genetischen Information verläuft in allen Organismen von der DNA (den Genen, die die Information speichern) über RNAs (die die Information zum Ort der Translation transportieren) zu den Proteinen, die die Information ausführen. Da Nukleinsäuren und Proteine biochemisch sehr unterschiedliche Moleküle sind, muss es einen biochemischen Prozess geben, der die „Nukleinsäure-Sprache" in die „Aminosäure-Sprache" übersetzt. Diesen Vorgang nennt man Genexpression, und man unterscheidet die Transkription – also das Abschreiben eines Gens in eine RNA – von der nachfolgenden Translation – also das Übersetzen des Nukleinsäure-Codes in eine Abfolge von Aminosäuren. Nur in wenigen Fällen kann die Richtung des Informationsflusses von RNA zu DNA quasi umgedreht sein. Dieser Vorgang, der ein besonderes Merkmal der Retroviren ist, ist die sogenannte *reverse Transkription*.

Zu den Hauptzielen der Gentechnik im Bereich pharmazeutisch/medizinischer Anwendungen gehört die Expression genetischer Information aus dem Menschen in Mikroorganismen bzw. pflanzlichen oder tierischen Zellen, mit anderen Worten: die Herstellung von Biopharmazeutika. Generell sprechen wir von *heterologer Genexpression*, wenn ein beliebiger Organismus eine für ihn *fremde* genetische Information exprimiert, also z. B. Bakterien ein menschliches Gen.

Um erfolgreich menschliche Proteinwirkstoffe durch heterologe Expression herstellen zu können, müssen wir uns die zum Teil erheblichen mechanistischen Unterschiede in der Genexpression in den jeweiligen für die heterologe Expression infrage kommenden Wirtszellen vor Augen führen. Deshalb wollen wir in den folgenden Abschnitten dieses Kapitels kurz auf die Unterschiede in der Genexpression zwischen Prokaryonten (Bakterien) und Eukaryonten eingehen. Diese Unterschiede sind letztlich der Grund, warum eine heterologe Expression ohne DNA-Rekombinationstechnologie nicht realisierbar ist, denn die speziesspezifischen Kontrollelemente der Genregulation in den Produzenten müssen in der Regel mit den zu exprimierenden Genen verbunden werden.

Eine der wichtigsten Erkenntnisse aus der molekularbiologischen Forschung der letzten 50 Jahre ist diese: Der genetische Code gilt in der belebten Natur auf der Erde universell. Daraus folgt, dass sich die Bedeutung,

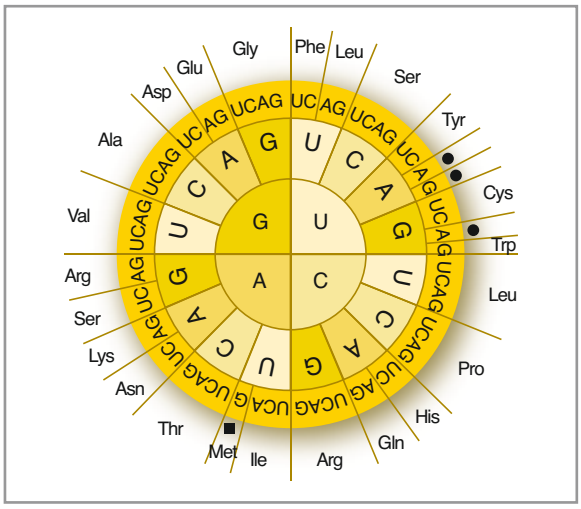


Abb. 3.1 Der universelle genetische Code. Dargestellt sind die 64 Codons und die jeweils codierte Aminosäure im Drei-Buchstaben-Code in Form einer „Code-Sonne". Die Codons sind von innen (5') und außen (3') zu lesen; sie geben die Basensequenz der mRNA-Codons wieder, die für die außerhalb des Kreises stehende Aminosäure codieren. Schwarze Punkte bezeichnen die drei Translations-Stoppcodons. Das Codon AUG (schwarzes Quadrat) steht als Translationsstart sowie innerhalb von Gensequenzen für die Aminosäure Methionin.

also die in der DNA niedergelegte Information für ein Protein, nicht verändert – egal, in welchem Organismus sich diese Information befindet. Nur deshalb können wir überhaupt daran denken, ein menschliches Protein als Wirkstoff z. B. in Bakterien herzustellen. Ein menschliches Gen wird in allen fremden (nicht-menschlichen) Zellen in ein Protein mit identischer Primärsequenz translatiert – allerdings nur dann, wenn wir dafür sorgen, dass der jeweilige Organismus das heterologe Gen als Transkriptionseinheit verstehen und umsetzen kann.

Die in Proteinen verwendeten Aminosäuren werden in den Genen als Basentripletts „codiert" (Abb. 3.1). Im molekulargenetischen Sprachgebrauch heißen diese Tripletts auch Codons. Die Natur verwendet deshalb Basentripletts, weil dadurch mithilfe der vier DNA-Basen alle 20 Aminosäuren codiert werden können. Wäre der genetische Code durch Basendupletts repräsentiert, könnte die Zelle nur 16 (4^2) der 20 benötigten Aminosäuren codieren. Mit Tripletts sind es 64 (4^3) Codierungsmöglichkeiten. Wie wir in Abb. 3.1 erkennen können, bleiben die 44 „überschüssigen" Codierungsmöglichkeiten nicht ungenutzt: Die Natur verwendet diese Codons, um die meisten der 20 benötigten Aminosäuren durch mehrere Tripletts zu definieren. So wird z. B. die Aminosäure Prolin durch vier Codons repräsentiert, Serin sogar durch sechs Codons (Abb. 3.1). Das hat den Vorteil einer Pufferung von Mutationen: Unsere DNA ist ständigen Veränderungen (Mutationen) aufgrund der chemischen Instabilität der Basen und der Exposition der Zellen mit mutagenen Einflüssen (Chemikalien, Strahlung) ausgesetzt. Die Änderung eines Codons in einem Gen durch Mutagenese führt nicht zwangsläufig zu einer Änderung des codierten Proteins, wenn eine Aminosäure durch mehrere, einander ähnliche Tripletts codiert ist.

Im genetischen Code ist auch festgelegt, wo ein translatierter Bereich eines Gens anfängt und wo er endet. Wir werden den Prozess der Translation etwas später besprechen. Wir wollen aber bereits jetzt zur Kenntnis nehmen, dass jede translatierte Region eines Gens mit dem Codon AUG beginnt. Da „AUG" für die Aminosäure Methionin steht, bedeutet das, dass jedes neu translatierte Protein mit der Aminosäure Methionin beginnt. Es gibt außerdem drei Codons, die bestimmen, wann die Translation eines Proteins beendet ist: die sogenannten Stoppcodons UAG, UGA und UAA (Abb. 3.1).

Ein Gen – wenn wir es als Transkriptionseinheit verstehen – besitzt Kontrollelemente, die bestimmen, in welchen Zellen eines Organismus ein Gen wann und in welcher Stärke transkribiert werden soll. Diese Kontrollelemente sind die Promotoren. Sie befinden sich in der Regel direkt vor dem jeweils von ihnen kontrollierten Gen und legen fest, welcher DNA-Strang von der RNA-Polymerase abgelesen werden soll. Am Ende eines Gens befindet sich ein Transkriptionsterminator, der dafür sorgt, dass die Transkription am Ende des proteincodierenden Teils aufhört.

Der Strang der DNA, der als Matrize für den Transkriptionsvorgang dient, ist der *codogene* Strang. Als Produkt der Transkription entsteht eine Boten-RNA (*messenger*-RNA oder kurz mRNA), die komplementär zum codogenen Strang ist. In der Praxis ist es gebräuchlich, als Basensequenz eines Gens die Sequenz der mRNA zu schreiben, also den „oberen" (eigentlich nicht-codogenen) Strang der DNA in 5'→3'-Richtung. Dies ist insofern logisch, weil die Basensequenz der mRNA letztlich bestimmt, welche Aminosäuren während der Translation an welcher Stelle in ein wachsendes Protein eingebaut werden (siehe unten).

Zusammenfassung

- Proteine sind Makromoleküle, die aus einer linearen Abfolge von Aminosäuren bestehen. Die Abfolge der Aminosäuren in einem Protein ist in den Gensequenzen festgelegt.
- Die Natur benutzt 20 Aminosäuren, um daraus Proteine herzustellen. Der genetische Code besteht aus Dreierkombinationen von DNA-Basen, die bestimmen, welche Aminosäure an welcher Stelle in einem Protein eingebaut wird.

- Der genetische Code gilt auf der Erde universell, d. h. die genetische Information für ein bestimmtes Protein ist immer gleich, egal in welchem Organismus sie unabhängig von ihrer eigentlichen Herkunft realisiert wird.
- Daraus folgt, dass für einen Organismus eigentlich fremde genetische Informationen zur Wirkung gebracht werden können, vorausgesetzt, dass die für verschiedene Organismen sehr spezifischen Kontrollelemente für die Transkription mit der fremden genetischen Information kombiniert werden.

3.2 Genexpression in Prokaryonten

3.2.1 Transkription

Das Kopieren der in der DNA gespeicherten genetischen Information in eine Transportform ist notwendige Voraussetzung für die spätere Umsetzung der Information in die Synthese von Proteinen. Der Vorgang der Transkription bezeichnet das „Abschreiben" der DNA-Information in eine RNA-Form. Transkription ist also die Synthese von RNA, das Produkt der Transkription ist komplementär zum abgeschriebenen codogenen DNA-Strang. Die Enzyme, die die Transkription durchführen, sind DNA-abhängige RNA-Polymerasen. In Bakterien ist die für die Transkription verantwortliche RNA-Polymerase ein aus fünf Untereinheiten aufgebauter Multiproteinkomplex. Das Minimalenzym besteht aus vier Untereinheiten: zwei α-, einer β- und einer β'-Untereinheit. Die β-Untereinheiten sind vor allem für die Bindung an die Matrizen-DNA zuständig, während die α-Untereinheiten den Gesamtkomplex stabilisieren. Um jedoch bakterielle Promotoren zu erkennen und effizient zu transkribieren, benötigt das Enzym die fünfte Untereinheit Sigma (σ). Bakterienzellen haben mehrere σ-Untereinheiten, die für jeweils verschiedene Sets von Genen benutzt werden.

Wie erkennt die RNA-Polymerase überhaupt einen Promotor und wie kann sie „entscheiden", welcher der beiden DNA-Stränge transkribiert werden soll? Um diese Frage zu beantworten, müssen wir uns einen typischen bakteriellen Promotor anschauen (Abb. 3.2).

Das erste Nukleotid eines Gens, das in RNA überschrieben wird, befindet sich unterhalb (stromabwärts oder *downstream*) des Promotors. Es wird als Initiationsnukleotid +1 bezeichnet. Das oberhalb (stromaufwärts oder *upstream*) von diesem Initiationsnukleotid gelegene Nukleotid wird als Nukleotid −1 bezeichnet. Wichtig sind zwei DNA-Segmente, die gewöhnlich etwa 35 Basen beziehungsweise etwa 10 Basen vor der eigentlichen Startstelle der Transkription liegen. Diese beiden

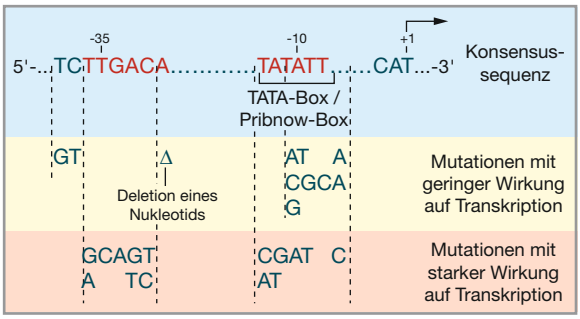


Abb. 3.2 Konsensus-Elemente für Promotoren in *E. coli*. Ein *E.-coli*-Promotor befindet sich immer vor dem transkribierten Sequenzbereich. Das erste transkribierte Nukleotid wird mit +1 bezeichnet. Das Nukleotid vor der +1-Position ist Nukleotid −1. Eine signifikante Übereinstimmung (Konsensus) in allen Promotoren aus *E. coli* finden wir um die Positionen −10 und −35 (rot-markierte Nukleotide, Konsensussequenz). Den Sequenzblock um die Position −10 bezeichnet man als TATA-Box oder Pribnow-Box. Verändert man die Nukleotide innerhalb der Promotorregion, beobachtet man entweder geringe (Mutationen im gelb-unterlegten Bereich) oder aber starke Auswirkungen auf die Transkription (rosa unterlegter Bereich). Wir erkennen daran, dass die Konsensussequenz zwar Variationen zulässt, dass aber die Variationsmöglichkeiten sehr stark eingeschränkt sind.

DNA-Segmente werden −35-Region bzw. als −10-Region genannt. Schreiben wir alle bekannten −35- und −10-Regionen untereinander fällt auf, dass diese Regionen eine deutliche Sequenzähnlichkeit aufweisen. Aus dieser Sequenzähnlichkeit lässt sich eine Konsensussequenz formulieren. Die Konsensussequenz für die −35-Region lautet 5'-TTGACA-3', die Konsensussequenz für die −10-Region lautet 5'-TATATT-3'. Oft wird die −10-Region wegen ihrer A/T-reichen Sequenz auch TATA-Box oder nach ihrem Entdecker Pribnow-Box genannt.

Bindet die RNA-Polymerase an einen Promotor, entwindet sie die DNA in einem Bereich von 12 Basenpaaren um den Transkriptionsstartpunkt herum und beginnt mit der Synthese von RNA. Sobald die Transkription von dieser Initiationsphase in die Elongationsphase übergeht, fällt die σ-Untereinheit von der RNA-Polymerase ab und das Minimalenzym setzt die RNA-Synthese fort.

Veränderungen der Konsensussequenzen innerhalb der −35-Region und der −10-Region beeinträchtigen die Promotor-Stärke. Dabei wirken sich unterschiedliche Basensubstitutionen unterschiedlich stark auf die Funktion des Promotors aus (Abb. 3.2). Letztlich hängt die Stärke eines Promotors, die als Menge an synthetisierter RNA pro Zeiteinheit definiert werden kann, davon ab, wie fest die RNA-Polymerase an den Promotor binden kann und wie schnell sie den Transkriptionsstart wieder freigibt, damit neue RNA-Polymerase-

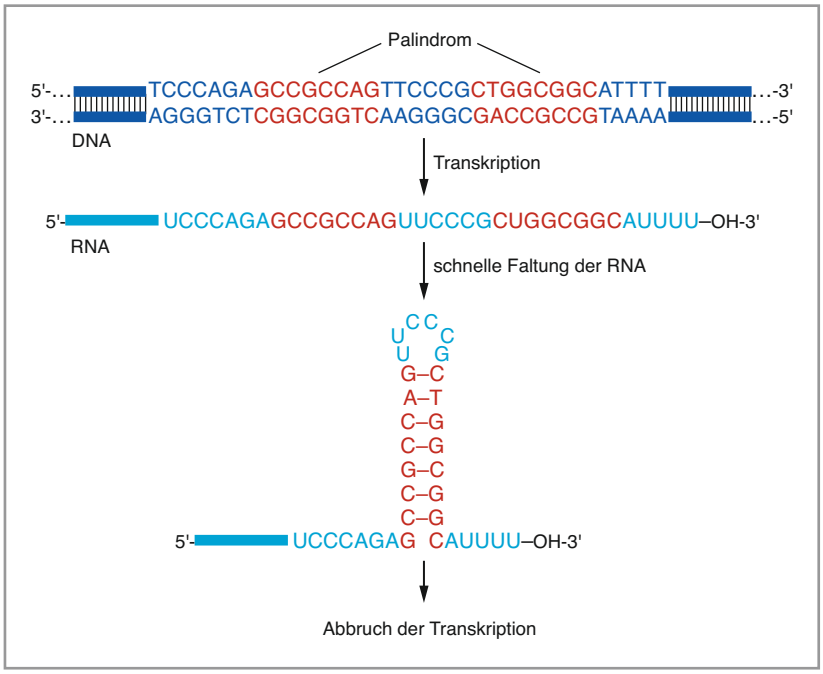


Abb. 3.3 Die Transkriptionstermination wird in *E. coli* über ein RNA-Signal reguliert. Bakterielle Gene besitzen an ihrem Ende eine Sequenz, die nach Transkription – also auf der Ebene der RNA – eine Stamm-/*Loop*-Struktur ausbilden kann, da die zu der *Loop*-Region links und rechts benachbarten Nukleotide palindromartig aufgebaut sind. Unmittelbar auf diese Stamm-/*Loop*-Struktur folgen eine Reihe von Uridin-Nukleotiden. Wird eine solche Struktur von einer bakteriellen RNA-Polymerase synthetisiert, erkennt das Enzym dies als Ende des Gens und fällt von der DNA ab. Wir erkennen, dass das Terminationssignal nicht wie ein Promotor von einer Konsensussequenz sondern von einer Konsensusstruktur gebildet wird.

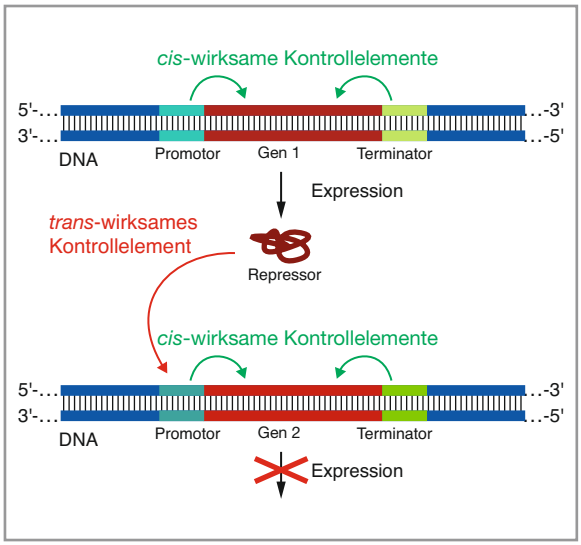


Abb. 3.4 Kontrollelemente für die Transkription. In *cis* wirksame Kontrollelemente befinden sich immer auf dem gleichen Molekül und in relativ enger Nachbarschaft zu dem Element, das sie kontrollieren. Promotoren und Transkriptionsterminatoren sind typische Kontrollelemente, die in *cis* auf ein Gen wirken. Dagegen sind in *trans* wirksame Kontrollelemente eigenständige Moleküle, die durch Bindung an eine Zielsequenz eine kontrollierende Funktion ausüben. Beispielsweise ist ein Repressor-Protein ein typisches in *trans* aktives Kontrollelement. Dieses Repressorprotein wird irgendwo im Genom – bei Eukaryonten eventuell auf einem anderen Chromosom – codiert. Der Repressor bindet dann an eine Zielsequenz in der Nähe des Promotors eines bestimmten Gens und verhindert so die Expression dieses Gens.

Moleküle binden und die nächste Transkriptionsrunde einleiten können. Typischerweise sind immer mehrere RNA-Polymerasekomplexe gleichzeitig an demselben Promoter aktiv.

Die Transkription verläuft immer von Nukleotid +1 bis zu einer Stelle, die von der RNA-Polymerase als Transkriptionsterminator erkannt wird. An dieser Stelle fällt die RNA-Polymerase von der Matrizen-DNA ab und steht dann für eine neue Transkriptions-Runde zur Verfügung. Als Transkriptionsterminator wird in *E. coli* eine Sequenz erkannt, die aus einer Palindromsequenz aus G/C-Basen, gefolgt von mehreren U-Nukleotiden, besteht (Abb. 3.3). Wir erkennen an den U-Nukleotiden, dass dieses Signal eigentlich auf der synthetisierten RNA liegt und nicht etwa auf der DNA. Der doppelsträngige RNA-Bereich bewirkt an der RNA-Polymerase eine Konformationsänderung, die zum Zerfallen des Komplexes aus der RNA-Polymerase, der Matrizen-DNA und der gebildeten RNA führt.

Ein Promotor ist ein typisches Beispiel für ein *cis*-wirksames Kontrollelement (Abb. 3.4). Das bedeutet, dass dieses Element eine ganz bestimmte Position relativ zu dem Bereich einnehmen muss, den es kontrolliert. Promotoren müssen sich nämlich immer auf dem gleichen DNA-Molekül befinden wie das Gen, dessen Transkription sie kontrollieren.

Es gibt aber auch Elemente, die die Expression eines Gens als in *trans* wirksame Kontrollelemente – sozusagen von außen her – beeinflussen (Abb. 3.4). Ein solches Element ist beispielsweise ein Repressor, der auf einem ganz anderen DNA-Molekül codiert sein kann. Ein typischer Repressor entfaltet seine Wirkung als Pro-

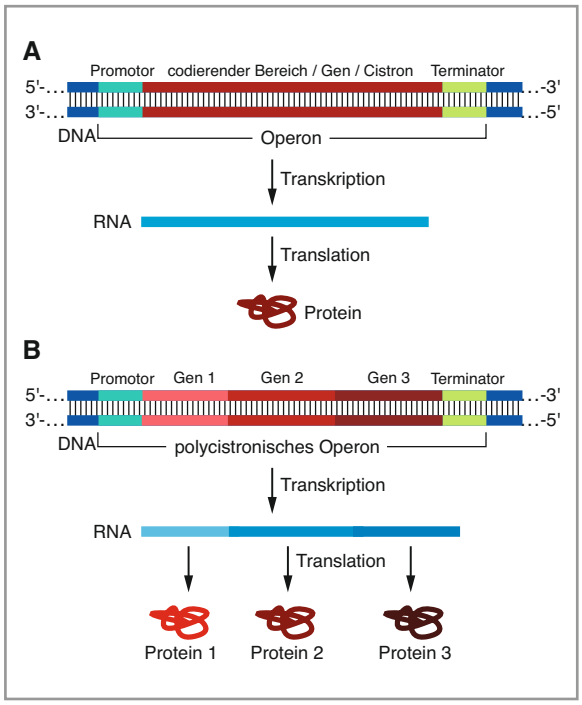


Abb. 3.5 Aufbau eines Operons. Ein Operon ist der transkribierte Bereich, der von einem Promotor und einem Transkriptionsterminator kontrolliert wird. **A** Befindet sich in diesem Bereich ein einzelnes Gen (= Cistron), so handelt es sich um ein monocistronisches Operon. **B** Befinden sich zwischen den Transkriptionskontrollelementen mehrere Gene (verschiedene Rottöne), bezeichnen wir dies als ein polycistronisches Operon. Die durch Transkription von diesem Operon gebildete polycistronische mRNA besitzt für jedes Gen einen eigenen Translationsstart. Daher wird von dieser mRNA nicht etwa ein großes Vorläuferprotein gebildet, das dann in Einzelproteine zerschnitten wird, sondern es werden von der polycistronischen mRNA simultan alle codierten Proteine einzeln gebildet.

tein, das an die Transkriptionseinheit bindet und mit ihr so wechselwirkt, dass das entsprechende Gen nicht mehr exprimiert werden kann.

Den Bereich, der einen Promotor, die codierende Region und einen Transkriptionsterminator beinhaltet, bezeichnet man als Operon (Abb. 3.5). Dieser Bereich stellt gewissermaßen eine genetische Arbeitseinheit dar, die sicherstellt, dass informative DNA in RNA übersetzt werden kann. In Prokaryonten werden häufig mehrere proteincodierende Bereiche von nur einem Promotor und einem Transkriptionsterminator kontrolliert. Ein derartiges Operon wird polycistronisches Operon und die entsprechend große mRNA polycistronische mRNA genannt. Der Hintergrund für diese Bezeichnung ist, dass die funktionelle Informationseinheit des Gens auch mit dem in der klassischen Genetik gebräuchlichen Begriff Cistron bezeichnet werden kann. Zu beachten ist, dass in Bakterien von einer polycistronischen mRNA mehrere unterschiedliche Proteine translatiert werden. In Eukaryonten kommen polycistronische Gene selten vor – in diesen Zellen codiert in aller Regel eine mRNA für ein Protein.

3.2.2 Translation

Bei der Transkription eines Gens entsteht eine mRNA, die dem Proteinsynthese-Apparat während des Vorgangs der Translation als Information für die zu bildenden Proteine zur Verfügung steht. Während der Proteinsynthese werden die Reihenfolge und die Identität der jeweils zu verknüpfenden Aminosäuren durch Adaptermoleküle festgelegt, die als Transfer-RNAs, oder kurz tRNAs, bezeichnet werden. tRNAs haben zwei für diese Aufgabe nötige Eigenschaften (Abb. 3.6): Jede tRNA besitzt eine exponierte Sequenz aus drei Nukleotiden, die komplementär zu einem Codon der mRNA ist. Diese auf tRNAs vorkommende Sequenz wird als Anticodon bezeichnet. Die Basenpaarung eines Anticodons der tRNA mit einem Codon der mRNA am Proteinsynthese-Apparat bestimmt also, welche Aminosäure in eine wachsende Polypeptidkette eingebaut wird.

Durch genaues Studium der Strukturen verschiedener tRNAs wurden gemeinsame Eigenschaften entdeckt. So sind alle tRNAs aus 74–94 Basen aufgebaut. Typischerweise sind tRNAs keine reinen einzelsträngigen Moleküle. Vielmehr werden innerhalb eines tRNA-Moleküls bestimmte Basenpaarungen ausgebildet, die dem tRNA-Molekül in der zweidimensionalen Darstellung die Form eines Kleeblattes geben (Abb. 3.6). tRNAs haben noch andere gemeinsame Eigenschaften:

- Das 5'- und das 3'-Ende bilden über eine Länge von sieben Basen einen doppelsträngigen Bereich aus, der Akzeptor-Arm genannt wird. Die 3'-Enden von tRNAs bestehen charakteristischerweise aus der Basensequenz CCA und ragen über das 5'-Ende der tRNA hinaus. Auf das Adenosin des CCA-Endes der tRNA wird eine Aminosäure übertragen, indem die Carboxylgruppe der Aminosäure mit der freien 3'-Hydroxylgruppe des Adenosins verestert wird.
- Das Anticodon liegt im Zentrum einer sieben Basen langen Schleife auf der dem Akzeptor-Arm entgegen gesetzten Seite der Kleeblattstruktur, die als Anticodon-Arm bezeichnet wird.
- Etwa 10 % der Nukleotide in tRNAs sind posttranskriptionell modifiziert und entsprechen dadurch nicht exakt den vier in RNAs normalerweise vorkommenden Standardnukleotiden. Es sind mehr als 30 solche Modifikationen in tRNAs gefunden worden, aber die Funktionen dieser Modifikationen sind weitgehend unklar.
- Der in der Kleeblattstruktur „links“ gelegene Teil der tRNA wird als D-Arm bezeichnet, weil er häufig die Base Dihydrouracil enthält.