

10 Klonieren und Sequenzieren

Bis etwa zum Jahre 1975 konzentrierte sich die molekulare Genetik im Wesentlichen auf Untersuchungen von Bakterien und Bakteriophagen. Und nur wenige Molekularbiologen wagten sich damals an eine Erforschung der Gene und Genome von Tieren und Pflanzen. Dies blieb ein weitgehend unerforschtes Kapitel der molekularen Genetik.

Die Schwierigkeiten lagen nur teilweise an experimentellen Problemen im Umgang mit Tier- und Pflanzenzellen. Sie bestanden in der geradezu überwältigenden Größe und Komplexität der Genome höherer Eukaryoten.

Zur Erinnerung: Ein Säugetiergenom ist mit 3×10^9 Basenpaaren (bp) gut 1000-mal größer als ein Genom von Bakterien. Die 20 000 – 30 000 Säugetiergene nehmen nur wenige Prozent auf dem langen DNA-Faden ein, dazwischen liegen lange informationsleere Abschnitte mit vielen repetitiven Sequenzen (S. 20, 242 f.). Wie lässt sich ein einzelnes Gen in diesem gewaltigen Überschuss an DNA erkennen und untersuchen?

Die seit 1975 zur Verfügung stehenden Methoden werden unter dem Begriff **Gentechnik** zusammengefasst. Gentechnische Methoden haben die Situation dramatisch verändert. Sie nahmen Einzug in die gesamte biologische Forschung. Heute kommt keine biologische Forschungsrichtung – von der Biophysik bis zur Ökologie – ohne gentechnische Verfahren aus.

Gentechnische Methoden sind im Prinzip einfach, und Leser, die bis zu diesem Punkt des Buches gekommen sind, sollten keine Schwierigkeiten des Verständnisses haben. Aber im Detail sind viele der heute gebräuchlichen gentechnischen Verfahren kompliziert, weil sie aus den einfachen Anfängen immer weiter entwickelt wurden, bis hin zu spezifischen Varianten für spezielle experimentelle Aufgaben. Wir beschreiben hier die Grundlagen und dann in späteren Kapiteln einige Anwendungen.

Wir wollen die Ziele formulieren, auf die es bei den hier beschriebenen Methoden ankommt:

- Zerlegen der langen natürlichen DNA-Fäden in definierte Abschnitte (Restriktion).
- Trennung der einzelnen DNA-Abschnitte voneinander (Herstellen einer Genombibliothek).
- Isolierung und Vermehrung eines gesuchten DNA-Abschnittes mit nachfolgender molekularbiologischer Analyse (Bestimmung der DNA-Sequenz, Untersuchung des kodierten Proteins u. a.).

Genombibliotheken

Zerlegen der DNA

Über den ersten Schritt, Zerlegen der natürlichen DNA-Moleküle in definierte Stücke, haben wir schon gesprochen (S. 25 f.). Die notwendigen Werkzeuge sind **Restriktionsnucleasen** (Tab. 2.6). Die Enzyme erkennen kurze Nucleotid-Sequenzen, an denen sie den DNA-Strang schneiden. Wir nehmen zwei Beispiele aus der Tab. 2.6:

- Das Enzym *EcoRI* schneidet eine DNA, gleich welchen Ursprungs, immer an Stellen mit der Nucleotid-Folge GAATC. Falls solche Sequenzen einigermaßen statistisch verteilt vorkommen, kann eine lange DNA in Stücke von durchschnittlich 4000–5000 Basenpaaren zerlegt werden.
- Das Enzym *AluI* trennt DNA an der Tetranucleotid-Folge AGCT. Deswegen werden natürliche DNA-Moleküle in Stücke mit 250–350 Basenpaaren zerlegt.

Das Enzym *EcoRI* zerlegt also ein Säugetier-Genom oder manche Pflanzen-Genome in $3 \times 10^9 / 5 \times 10^3 = \text{ca. } 600\,000$ verschiedene Stücke. Der angenommene Wert von 5000 Basenpaaren ist ein Mittelwert, und die Länge eines Restriktionsfragmentes kann erheblich davon abweichen, abhängig von der Verteilung der *EcoRI*-Schnittstellen im Genom (Plus 10.1).

Der wichtige Punkt ist hier, dass immer die gleiche Kollektion von DNA-Stücken entsteht, gleichgültig, wann und wo ein gegebenes Genom mit einer gegebenen Restriktionsnuclease behandelt wird.

Aber diese Kollektion ist ein sehr komplexes Gemisch von Molekülen. Es kommt darauf an, die einzelnen Bestandteile des Gemisches sauber voneinander zu trennen. Dies geschieht mithilfe bakteriologischer Verfahren. DNA lässt sich gut in entsprechend vorbehandelte Bakterien übertragen, aber Restriktionsfragmente würden nach Aufnahme in Bakterienzellen schnell abgebaut und verloren gehen. Deswegen müssen die Restriktionsfragmente in Überträger oder – in der Sprache der Gentechnik – in Vektoren eingebaut werden.

Gebräuchliche Vektoren sind Plasmide, Bakteriophagen oder eine Kombination aus beiden. Auf dem Gentechnik-Markt findet man buchstäblich hunderte verschiedener Vektoren, aber hier geht es um eine Einführung in das Gebiet. Deswegen schildern wir die Verhältnisse anhand einiger einfacher Bilder.

Plus 10.1 Häufigkeit von Schnittstellen

Häufigkeit und Verteilung von Schnittstellen für eine bestimmte Restriktionsnuclease sind spezifische Kennzeichen eines Genoms. Die Häufigkeit lässt sich einigermaßen abschätzen. Wir denken uns eine DNA, in der alle vier Nucleotide in gleichem Verhältnis vorkommen. Wenn eine Vierer-Nucleotid-Folge aus allen vier Basen zusammengesetzt ist, wie etwa AGCT, die Erkennungssequenz für das Enzym *AluI*, wird ein Schnitt in Abständen von $(1/4)^4$ erfolgen, d. h. einmal pro 256 Basenpaaren. Entsprechend gilt für die Häufigkeit von Hexanucleotiden $(1/4)^6$, also eine Schnittstelle pro 4096 Basenpaaren. Aber natürliche DNA-Moleküle und viele Schnittstellen bieten nicht diese einfachen Voraussetzungen. Die geschätzten Werte geben nur Anhaltspunkte.

Plasmide als Vektoren

Plasmid-Vektoren sind Derivate von R-Plasmiden (S. 228). Die Abb.10.1 zeigt das Schema eines natürlichen R-Plasmids mit seinen Transfer-Genen, mit einigen Genen, die Resistenz gegen Antibiotika vermitteln (S. 228), mit einem Origin für die Ring-zu-Ring-Replikation (*oriV*) sowie einem zweiten Origin (*oriT*), der die Transfer-Replikation bei der Übertragung des Plasmids von Donor- in Empfänger-Bakterien (S. 97) vermittelt.

Ein Original-R-Plasmid ist für die Gentechnik aus folgenden Gründen ungeeignet:

- Es kommt nur in 1–2 Kopien/Bakterienzelle vor.
- Wegen seiner Größe von 100 kb (Kilobasenpaare) zerbricht es leicht bei biochemischen Prozeduren.
- Vor allem: Ein natürliches R-Plasmid kann von einem Bakterium zum nächsten übertragen werden. Genau dies wäre dem angestrebten Ziel einer effizienten Auftrennung von Restriktionsfragmenten abträglich.

Aus diesen Gründen verwendet man abgeleitete Plasmide (Abb.10.1), die sich innerhalb einer Bakterienzelle über Ring-zu-Ring-Replikation zu höheren Kopienzahlen vermehren und nicht auf andere Bakterien übertragen werden können.

Ein einfaches und zu Beginn des Gentechnik-Zeitalters viel verwendetes, abgeleitetes Plasmid hat die Bezeichnung **pBR322**: Es trägt den *oriV* und die Gene *amp^R* und *tet^R* zur Vermittlung von Antibiotika-Resistenz (Abb. 10.2). Wie üblich in der Nomenklatur von Plasmid-Vektoren bezeichnet p den Plasmid-Charakter, BR sind die Initialen der Konstruk-

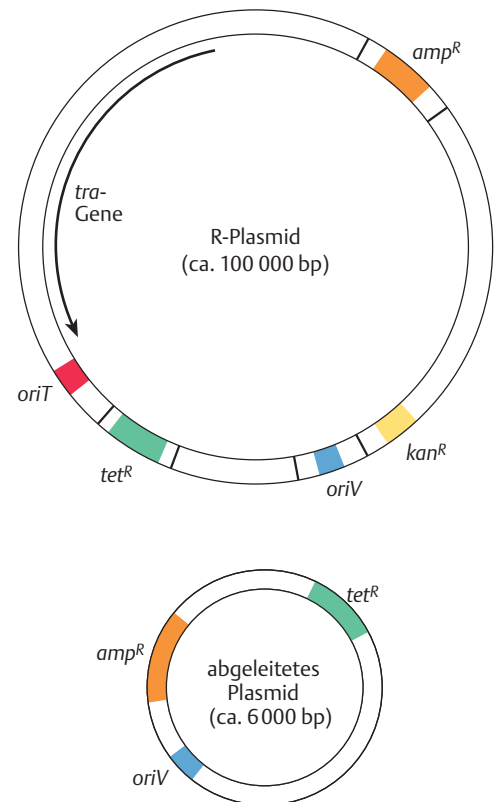


Abb. 10.1 Natürliches R-Plasmid und abgeleitetes Vektorplasmid: Größen- und Strukturvergleich.

amp^R = Ampicillin-Resistenz
kan^R = Kanamycin-Resistenz
tet^R = Tetracyclin-Resistenz
tra = Transfer-Gene

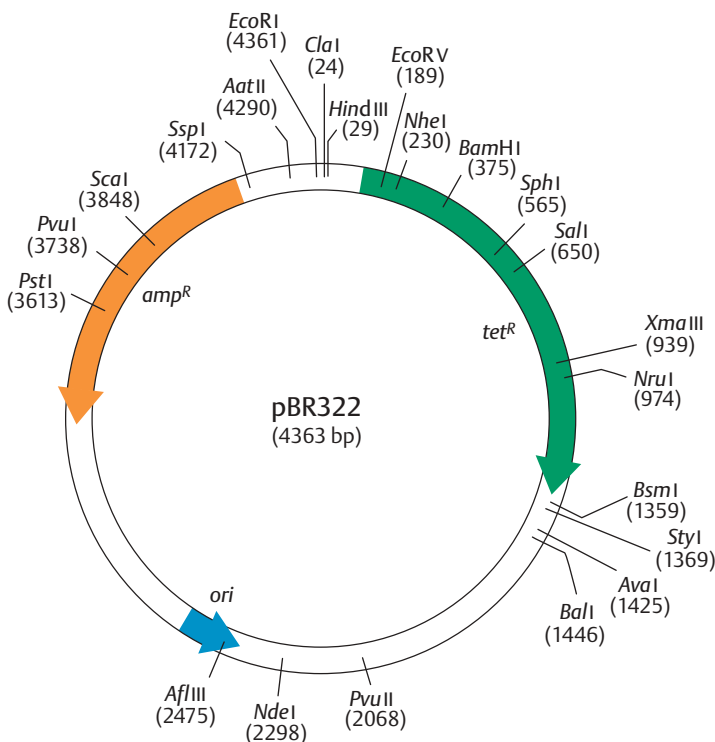


Abb. 10.2 Das Plasmid pBR322. Die Nummerierung der 4363 Basenpaare läuft von der *EcoRI*-Stelle im Uhrzeigersinn. Schnittstellen für einige Restriktionsnucleasen kommen nur je einmal auf dem Molekül vor. Ihre Positionen sind als Nucleotidnummern in Klammern eingetragen [nach 13].

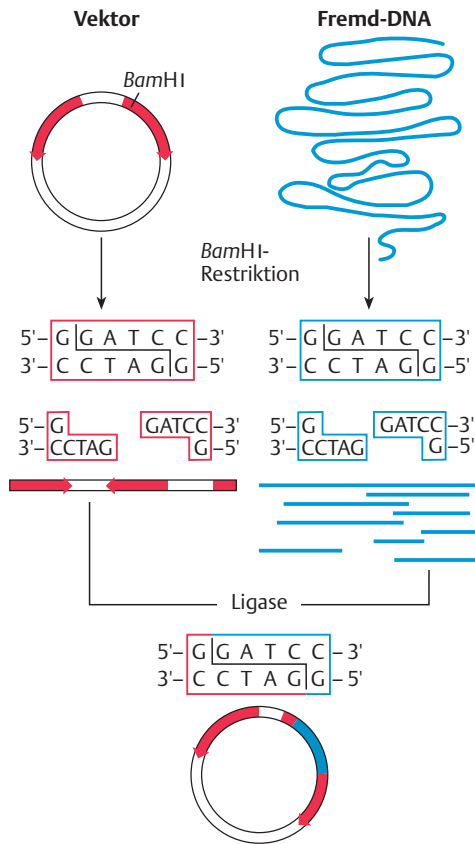


Abb.10.3 Grundlagen des Klonierens. Vektor- und Fremd-DNA werden mit einer Restriktionsnuclease geschnitten, vermischt und durch die Ligase wieder zu Ringen vereinigt.

teure (hier: F. Bolivar und R. L. Rodrigues, 1977), gefolgt von einer Labornummer, die eine Unterscheidung zwischen ähnlichen, aber nicht identischen Plasmiden ermöglicht.

Restriktionsfragmente werden in Vektor-DNA eingebaut. Einige Voraussetzungen erleichtern das Verfahren. Wenn möglich, verwendet man die gleiche Restriktionsnuclease für die Restriktion der fremden DNA und für das Schneiden der Vektor-DNA, und benutzt dazu eine einmal im Vektor vorkommende Schnittstelle. Als Beispiel betrachten wir in der Abb.10.3 die Öffnung („Linearisierung“) von pBR322 an der *Bam*HI-Stelle im *tet*^R-Gen. Zu der linearisierten Vektor-DNA werden die Restriktionsfragmente der Fremd-DNA gegeben. Beide DNA-Arten werden durch eine Ligase (S.185) kovalent verknüpft. Es entstehen DNA-Ringe mit Vektoranteilen und eingebauten Fremd-DNA-Stücken (Abb.10.3).

Experimentatoren müssen darauf achten, dass möglichst eine Wiedervereinigung der Enden geschnittener Vektor-DNA verhindert und der Einbau von Fremd-DNA gefördert wird (s.a. Box Topo-Cloning). Das gelingt, wenn im Reaktionsgemisch Fremd-DNA im Überschuss gegenüber Vektor-DNA vorliegt, oder besser noch, wenn die Vektor-DNA nach ihrer Linearisierung mit dem Enzym Phosphatase zur Entfernung der 5'-Phosphat-Reste behandelt wird. Das Vorhandensein von 5'-Phosphat-Resten ist eine Voraussetzung für die Wiederverknüpfung durch das Enzym Ligase (S.185).

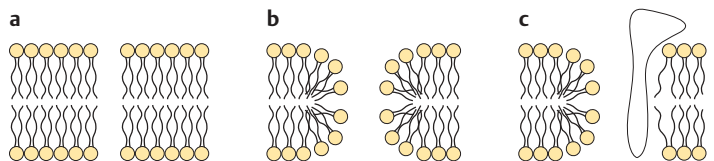
Nach Durchführung der biochemischen Reaktionen erfolgt die Auftrennung der DNA-Moleküle in Bakterien. *E. coli*-Zellen nehmen, nach Vorbehandlung durch Calcium-Salze oder Elektroporation (s. Box), DNA auf. Hier muss die Reaktion so geleitet werden, dass eine Bakterienzelle möglichst nur ein DNA-Molekül erhält. Experimentell lässt sich dies am besten durch einen Überschuss von Bakterien gegenüber DNA erreichen.

Methode: Elektroporation

Ein kurzer elektrischer Puls erzeugt Poren in der Cytoplasmamembran von Bakterien, aber auch von Eukaryotenzellen. Zunächst wird die Lipid-Doppelschicht der Membran einfach unterbrochen (a), dann klappen die Lipide um, so dass die hydrophilen Enden der Lipide die Porenwandung auskleiden (b), wobei membranständige Proteine zur Stabilität der Pore beitragen (c). Durch diese Öffnungen kann DNA in die Zellen eindringen.

Anschließend muss die Cytoplasmamembran wieder hergestellt werden. Andernfalls gingen die Zellen zugrunde. Eine wichtige experimentelle Aufgabe ist hier, für jeden Zelltyp die richtigen Bedingungen zum Wiederverschließen der Poren zu finden. Dabei müssen Temperatur und die Zusammensetzung der Lösung berücksichtigt werden.

Elektroporatoren sind einfache Apparate, die die Prozedur unter kontrollierten Bedingungen ermöglichen. Sie gehören zur Ausrüstung vieler zell- und molekularbiologischer Laboratorien [9].



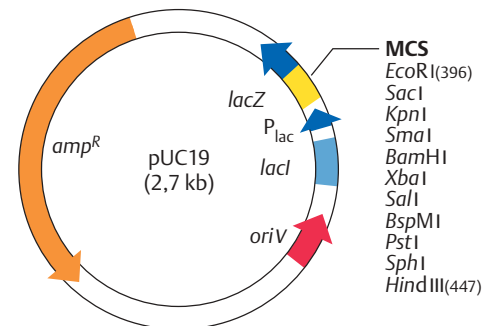
Die Bakterien werden schließlich auf Agar-Platten so verteilt, dass sie sich zu getrennten Kolonien entwickeln können. In unserem Beispiel-Experiment (Abb.10.3) enthalten die Agar-Platten das Antibiotikum Ampicillin, das alle Bakterien ohne Plasmid tötet und nur die Vermehrung von Kolonien mit aufgenommenem pBR322 erlaubt. Wenn man schließlich noch sicher sein will, dass nur Bakterien mit Plasmid plus eingebauter Fremd-DNA weiter untersucht werden, bietet sich noch ein zusätzliches Selektionsverfahren an: Bakterien mit pBR322 bilden Kolonien auf Platten mit Ampicillin **und** Tetracyclin; Bakterien mit pBR322/ Fremd-DNA bilden nur Kolonien auf Ampicillin-Platten, weil das *tet^R*-Gen durch Einbau der Fremd-DNA zerstört ist.

Wir fassen das Ergebnis zusammen: Durch die Verteilung der Bakterien in einzelne unabhängige Kolonien gelingt eine Auftrennung von Restriktionsfragmenten der Fremd-DNA in die einzelnen Komponenten des Gemisches. Jede Kolonie besteht aus zahlreichen Bakterien, die alle das identische Stück Fremd-DNA in ihrem Vektorplasmid tragen. In der Tradition der Biologie bezeichnet man die genetisch gleichen Nachkommen einer Elternzelle als Klone. Entsprechend wird das geschilderte Verfahren oft als **Klonieren** und das Restriktionsfragment in einer Kolonie als **DNA-Klon** bezeichnet. Man sagt auch, dass ein Stück DNA kloniert worden ist. Im Jargon molekularbiologischer Laboratorien wird das eingebaute Stück Fremd-DNA oft auch als **Insert** bezeichnet.

Wenn alle Schritte technisch gut gelungen sind, steht dem Molekularbiologen ein in Einzelstücke aufgeteiltes Genom für weitere Untersuchungen zur Verfügung. Man spricht von einer **Genombibliothek**. Wie ein Buch aus einer Bücherei, lässt sich aus der Genombibliothek jedes interessante Genomstück herausnehmen und genauer untersuchen. Bakterien lassen sich problemlos in großen Mengen züchten. Damit stehen sozusagen beliebige Mengen eines Fremd-DNA-Stückes für die Untersuchung zur Verfügung. Voraussetzung dafür ist, dass aus den Bakterien die Plasmid-DNA gewonnen und das klonierte DNA-Stück durch Restriktion freigesetzt wird. Allerdings ist das Herausnehmen eines „Buches“ aus der Genombibliothek ein Kapitel für sich, das wir uns später ansehen werden.

Das Selektionsverfahren zum Nachweis von eingebauter DNA in pBR322 ist umständlich, deswegen werden heute andere Plasmid-Vektoren bevorzugt. Als Beispiel nennen wir das vielbenutzte Plasmid pUC19, ein Vertreter der Familie von **pUC-Plasmiden** (Abb.10.4). Wie andere Plasmid-Vektoren enthält pUC19 den *oriV* und zusätzlich ein *amp^R*-Gen sowie zur Aufnahme der Fremd-DNA einen besonders eingefügten DNA-Abschnitt mit vielen Restriktionsschnittstellen (MCS, *multiple cloning site*). Die MCS-Sequenz liegt innerhalb eines 5'-Abschnitts des *lacZ*-Gens. Dieser Abschnitt kodiert einen aminoterminalen Teil der β -Galactosidase. Um das Selektionssystem gut ausnutzen zu können, müssen passende Bakterienstämme verwendet werden, die auf ihrem Genom den 3'-Abschnitt des *lacZ*-Gens besitzen und deswegen den carboxyterminalen Teil der β -Galactosidase kodieren.

Bakterien mit pUC19 sind ampicillinresistent und besitzen überdies eine funktionsfähige β -Galactosidase, die sich aus dem plasmidkodierten Anteil und dem genomkodierten Anteil zusammensetzt. Diese Bakterien bilden in Gegenwart des Farbstoffes X-Gal blaue Kolonien (S.114). Der Einbau der Fremd-DNA in MCS zerstört das Leseraster des



MCS-Sequenz

396	410	420		440	450
·	·	·		·	·
			HincII		
			SalI		
EcoRI	KpnI	BamHI	AccI	PstI	HindIII
GAATTCGAGCTCGGTACCCGGGGATCCCTAGAGTCGACCTGCAGGCATGCCAAGCTTGG					
	SacI	XmaI	XbaI	BspMI	SphI
		SmaI			

Abb.10.4 Ein Vektor für Blau-Weiß-Selektion, pUC19. Die Gene sind im Text beschrieben. MCS (*multiple cloning site*) ist ein DNA-Abschnitt im *lacZ*-Gen mit mehreren Schnittstellen für Restriktionsnucleasen. Die MCS-Sequenz ist unten genauer dargestellt [nach 8].

lacZ-Gens. Die betreffenden Bakterienkolonien bleiben in Gegenwart von X-Gal farblos. Oder anders gesagt, Kolonien auf Ampicillin-Agar mit X-Gal sind blau, wenn sie pUC19, und weiß, wenn sie pUC19 plus Fremd-DNA enthalten. Diese **Blau-Weiß-Selektion** wird in vielen gentechnischen Verfahren eingesetzt.

Methode: Topo-Cloning

Wie auf diesen Seiten beschrieben, besteht das Standardverfahren beim Einbau von Fremd-DNA in einen Vektor aus zwei Schritten:

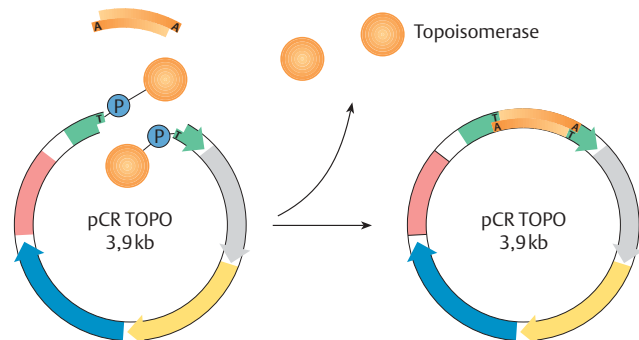
1. Schneiden der Fremd-DNA und der Vektor-DNA mit Hilfe von Restriktionsnucleasen,
2. Verknüpfung von Fremd- und Vektor-DNA durch das Enzym Ligase.

Wenn alles gut vorbereitet ist, benötigt man mehrere Stunden, um das Verfahren ordentlich durchführen zu können. Oft stellt sich die Aufgabe, Insert-DNA ohne kompatible Enden in einen Vektor einzubauen. Dann muss ein Oligonucleotid mit den Schnittstellen für die passende Restriktionsnuclease an das Insert ligiert werden (s. Abb. 10.10), was wiederum Zeit kostet – und experimentelle Geschicklichkeit erfordert.

Eine technische Vereinfachung ist das *Topo-Cloning*, ein Verfahren, das die Biotechnologie-Firma Invitrogen entwickelt und auf den Markt gebracht hat. Dabei nutzt man biochemische Eigenschaften der Typ-I-DNA-Topoisomerase (Topo) aus. Wie beschrieben (S. 190), schneidet das Enzym einen Strang in der DNA-Doppelhelix, indem vorübergehend eine kovalente Bindung zwischen dem Enzym und den Enden der geschnittenen DNA geschlossen wird. Anschließend löst sich das kovalent gebundene Enzym wieder ab, bei gleichzeitiger Versiegelung des Strangbruches.

Beim Topo-Cloning tragen beide Enden des gespaltenen Vektors kovalent gebundene Topoisomerase (was eine Verknüpfung der Enden verhindert, denn dazu sind **freie** 3'-OH-Enden notwendig). Wenn aber der Topo-tragende Vektor mit Insert-DNA in Kontakt kommt, kann das Enzym seinen Reaktionsablauf fortsetzen, indem es die 3'-OH-Enden des Inserts benutzt. Dabei werden Insert und Vektor verknüpft, während das Enzym sich von der DNA ablöst (s. Abb.).

Die Vorteile des Topo-Cloning: Eine Inkubation mit Ligase entfällt und dementsprechend verkürzt sich die Zeit für die experimentelle Arbeit. Vor allem: DNA mit beliebigen Enden kann in den Vektor eingebaut werden, so dass eine vorhergehende Anheftung von Oligonucleotiden nicht erforderlich ist.



Lambda-DNA als Vektor

Klonierung in Plasmiden ist technisch einfach, hat aber den Nachteil, dass nur relativ kurze DNA-Stücke (bis zu einigen tausend Basenpaaren) eingebaut werden können. Längere DNA-Stücke schränken die Plasmid-Vermehrung in Bakterien ein. Funktionell zusammengehörende Genomabschnitte werden deswegen beim Klonieren in Plasmiden auseinander gerissen, was für eine Erforschung von Genen oft sehr störend ist. Viele Molekularbiologen bevorzugen daher Vektoren, die den Einbau von längeren Stücken Fremd-DNA ermöglichen.

Ein einfaches und populäres System ist der Bakteriophage Lambda. Wie wir früher besprochen haben (S.127), ist ein beträchtlicher Teil des Lambda-Genoms für einen lytischen Infektionszyklus entbehrlich: die *b2*-Region, das Integrationsystem (mit den Genen *xis* und *int*) sowie die Gene für viruseigene Rekombinationsenzyme (Abb.5.40 und Abb.10.5). Dieser Teil ist ersetzbar durch Fremd-DNA. Allerdings enthält die natürliche Lambda-DNA keine geeigneten Restriktionsstellen, die eine einfache Entfernung des entbehrlichen Abschnittes ermöglichen. Deswegen wurde eine lange Reihe von Lambda-Derivaten entwickelt, von denen wir nur den Vektor EMBL3 in der Abb.10.5 zeigen. Bei EMBL3 ist eine innere Region des Phagen-genoms durch eine quasi beliebige DNA ersetzt und beiderseits eingerahmt durch günstige Schnittstellen für Restriktionsnucleasen.

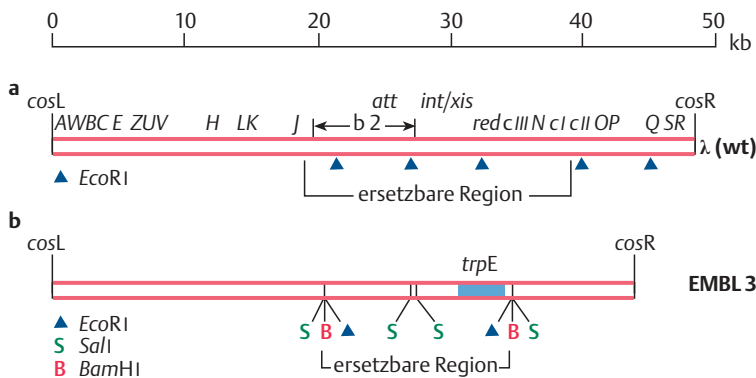


Abb. 10.5 Lambda-DNA als Vektor.

a Die linearisierte Gen-Karte des Wildtyp-Lambda-Genoms mit den überstehenden *cos*-Enden. Die für einen lytischen Infektionsweg entbehrliche Region ist angegeben.

b Ein abgeleitetes Lambda-Genom: Die ersetzbare Region ist beiderseits von Restriktionsstellen eingerahmt.

Das experimentelle Vorgehen (Abb.10.6):

- Die zu untersuchende DNA wird nur unvollständig mit Restriktionsnucleasen behandelt, so dass längere Genomabschnitte erhalten bleiben.
- Die „Arme“ des Lambda-Vektors werden isoliert und mit der zerschnittenen DNA in einem Reaktionsgefäß zusammengebracht und durch Ligase zu langen DNA-Concatemeren (S.136) verknüpft.
- Die concatemere DNA wird *in vitro* in Phagenpartikel eingebaut. Hier erinnern wir daran, dass der Zusammenbau der Phagenstruktur aus den Einzelkomponenten im Reagenzglasversuch möglich ist, sofern die angebotene DNA in concatemerer Form vorliegt und die *cos*-Stellen der DNA-Enden enthält (S.137).

In Lambda-Vektoren können DNA-Fragmente von 10–20 kb eingebaut werden. Ihre Trennung erfolgt durch das Verfahren der Plaque-Bildung auf Bakterienrasen (S.124). Ein Plaque enthält über eine Million identi-

scher Phagen („Phagenklone“), die sich dann ohne weiteres zu noch größeren Mengen vermehren lassen, so dass eine molekularbiologische Untersuchung der eingebauten Fremd-DNA möglich wird.

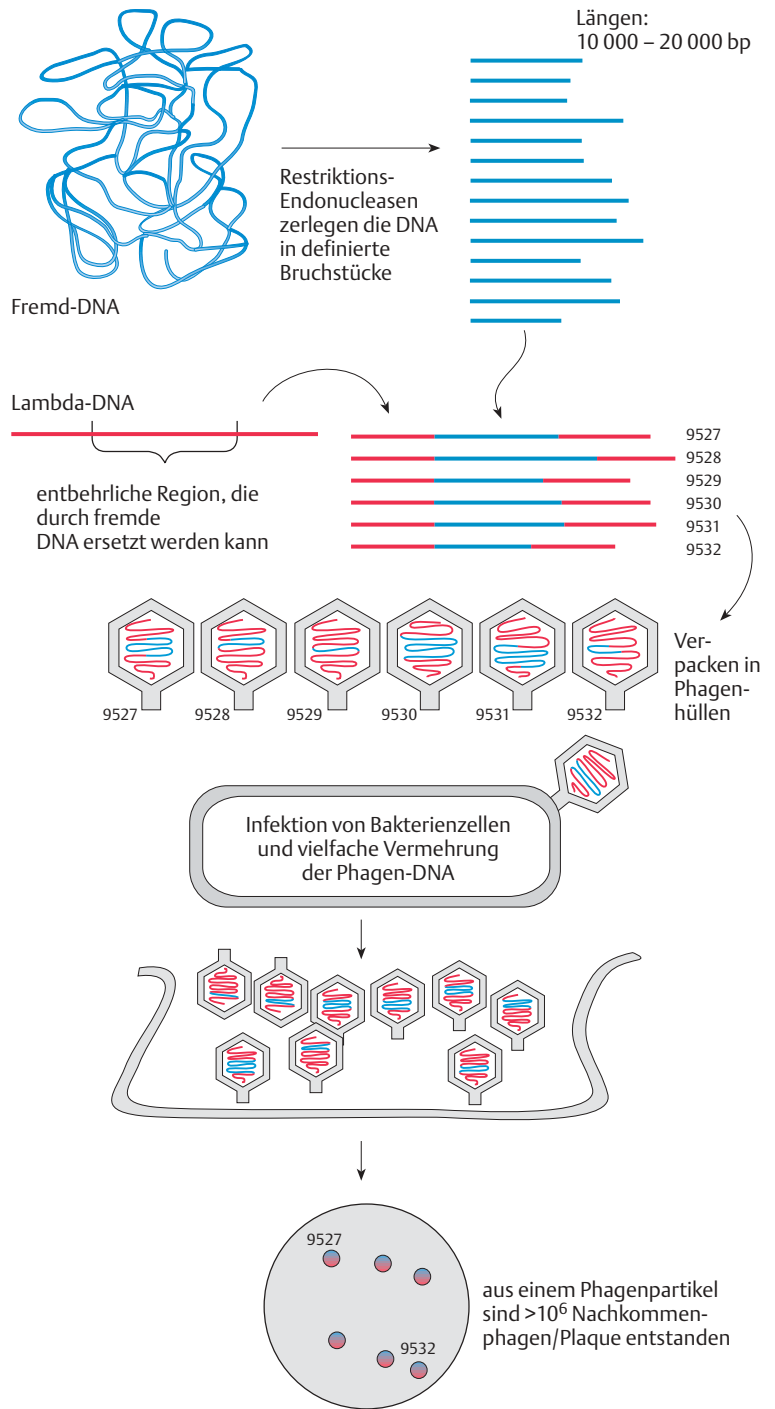


Abb. 10.6 Klonieren in Lambda-Vektoren.

1. Fremd-DNA wird durch vorsichtigen (unvollständigen, „partiellen“) Angriff von Restriktionsnucleasen in Stücke von 10 000 – 20 000 Basenpaare zerlegt.
2. Die Arme der Lambda-Vektor-DNA werden präpariert, mit den Fremd-DNA-Stücken in einem Reaktionsgefäß vereinigt und durch Ligase zu langen End-zu-End-verbundenen Concatemeren verknüpft.
3. Die DNA wird in Phagenpartikel verpackt (s. Abb. 5.50).
4. Phagen werden auf Agar-Platten mit dichten Bakterienrasen verteilt: Wo ursprünglich ein Phagenpartikel einen Infektionsprozess begonnen hat, bildet sich ein Plaque im Bakterienrasen (s. Abb. 5.36). Die Zahlen deuten an, dass wir hier auf das 9527., 9528. usw. Fragment von Millionen von Fragmenten blicken. Mit dem Trick der Plaque-Bildung lässt sich jedes der vielen Fragmente eindeutig von den anderen trennen.

Cosmide als Vektoren

Cosmide vereinigen die Vorteile der Plasmid-Klonierung (einfache technische Handhabung) mit den Vorteilen der Lambda-Vektoren (Einbau langer DNA-Stücke).

Das Verfahren beginnt mit der Verwendung von Plasmiden, die außer einem *amp^R*-Gen die **Cos-Elemente** der Lambda-DNA-Enden enthalten. In geschnittene Cos-Plasmide können Fremd-DNA-Stücke mit Längen von 40 und 50 kb eingebaut werden. Die concatemere Cosmid-DNA wird in Phagenstrukturen verpackt und für eine Infektion verwendet. Die Empfänger-Bakterien erwerben die Fähigkeit zur Koloniebildung auf Agar-Platten mit Ampicillin (Abb.10.7).

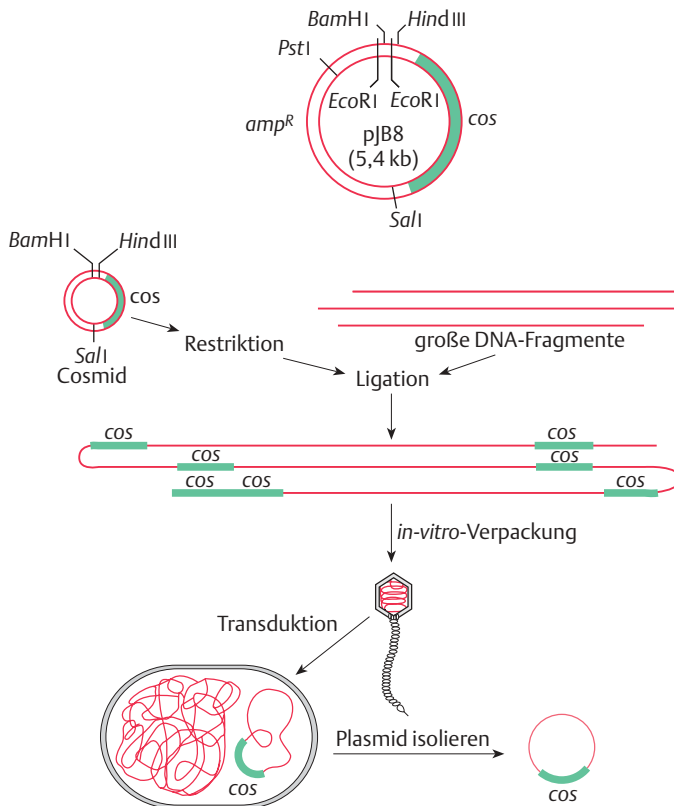


Abb.10.7 Cosmide als Vektoren. Der Ausgang ist ein Plasmid, hier pJB8, das die *cos*-Sequenzen von Lambda enthält. Das Plasmid wird linearisiert und mit Fremd-DNA zu langen Concatemeren verknüpft. Daran schließt sich die Verpackung in Phagenpartikel und die Infektion von Bakterien an. Cosmide verhalten sich wie Plasmide, und cosmidgehaltige Bakterien sind resistent gegen Ampicillin [nach 6].

Künstliche Phagen-, Bakterien- und Hefe-„Chromosomen“ (PAC, BAC und YAC)

Wenn es um die Aufklärung der Struktur großer Genome und die Aufstellung „physikalischer Gen-Karten“ (S.94, 482) geht, ist der Umgang mit Genombibliotheken in Lambda- oder Cosmid-Vektoren mühsam, einfach weil die klonierten DNA-Stücke im Vergleich zum Gesamtgenom zu klein sind (Tab.10.1). Ja, nicht selten kommt man mit Lambda- oder Cosmid-Bibliotheken überhaupt nicht weiter, wenn Genombereiche mit vielen repetitiven Elementen untersucht werden. In diesen

Tab. 10.1 Genombibliotheken

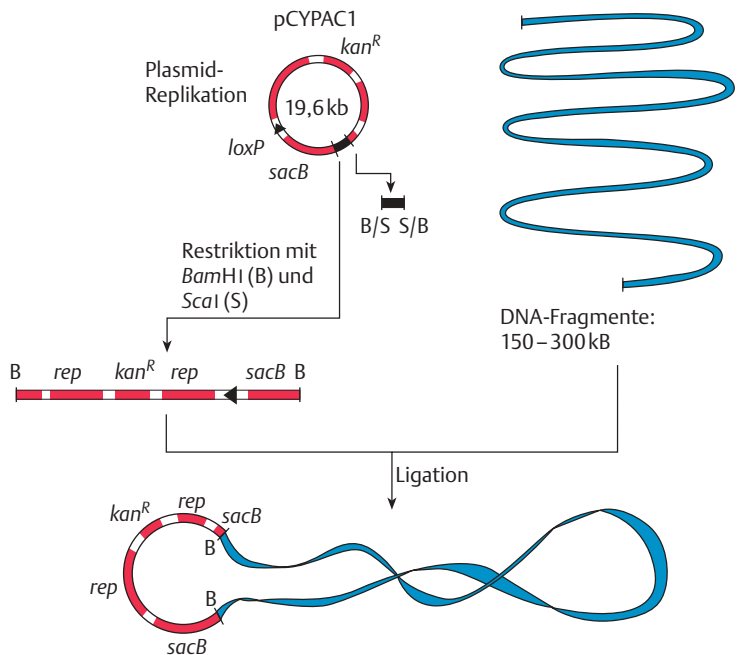
Vektor	Insert	Zahl der Klone mit insgesamt 3×10^9 bp (entsprechend dem Human-Genom)
Phagen	$1 - 2 \times 10^4$	ca. 150 000
Cosmide	$4 - 5 \times 10^4$	ca. 50 000
YAC/BAC	$10^5 - 10^6$	ca. 10 000

Fällen ist man auf Vektoren angewiesen, die das Klonieren von DNA mit mehr als hunderttausend Basenpaaren erlauben.

Ein früher verwendetes System waren künstliche Hefechromosomen (YAC, *yeast artificial chromosomes*), die fremde DNA mit bis zu einer Million Basenpaare enthalten können.

Aber künstliche Hefechromosomen haben einige Nachteile. Vor allem sind sie nicht stabil: es kann zu Rekombinationen kommen, so dass die natürliche Reihenfolge von DNA-Bereichen durcheinander gerät und die Sequenz im Insert nicht mehr der Sequenz im ursprünglichen Genom entspricht. Aus diesen Gründen werden in den großen Genomprojekten YAC-Klone nur noch selten benutzt. Stattdessen kommen so genannte künstliche Bakterienchromosomen (BAC, *bacterial artificial chromosomes*) zur Anwendung. Im wesentlichen handelt es sich dabei um Derivate des bakteriellen F-Faktors (S. 95), der zur Selektion ein Resistenzgen (gegen das Antibiotikum Chloramphenicol) und zum Einbau der Fremd-DNA geeignete Restriktionsschnittstellen enthält. In BAC-Vektoren können DNA-Stücke mit Längen von 300 kb und mehr kloniert werden. BACs können problemlos über Elektroporation in *E. coli*-Zellen übertragen werden. Sie bleiben in niedrigen Kopienzahlen über viele hundert Bakteriengenerationen stabil.

Abb. 10.8 PAC-Vektoren.
Erklärungen im Text [nach 7].



Das ist auch einer der Vorteile eines zweiten populären Klonierungssystems, PAC (*P1 derived artificial chromosome*). Andere Vorteile von PAC: höhere Kopienzahlen und die Möglichkeit einer „positiven“ Selektion für Klone mit Inserts.

Wir betrachten den Vektor pCYPAC1 (Abb.10.8). Er trägt die folgenden Abschnitte aus dem Genom des Bakteriophagen P1 (Plus 10.2):

- einen Abschnitt, der eine Replikation über viele Bakteriengenerationen mit niedrigen Kopienzahlen ermöglicht;
- einen Abschnitt, der unter der Kontrolle des Lac-Repressors steht, also durch IPTG induziert werden kann, und Proteine kodiert, die eine Steigerung der Replikation bewirken (wie beim lytischen Vermehrungszyklus von P1, sodass sich die Zahl der DNA-Kopien stei-

Plus 10.2 Der Bakteriophage P1

Der Bakteriophage P1 ist ein **temperenter Phage**, der – ähnlich wie der Phage Lambda (S.125) – entweder lysogen in *E. coli* vorkommen kann oder einen lytischen Vermehrungsprozess durchläuft. Aber anders als Lambda-DNA, die in das Genom eines Wirtsbakteriums integriert (S.128), bleibt die P1-DNA als geschlossener Ring (aus etwa 10^5 Basenpaaren) erhalten und wird **wie ein Plasmid** von einer Bakteriengeneration zur nächsten weitergegeben. Zur Regulation der Replikation besitzt die P1-DNA einen *Origin* aus mehreren kurzen Wiederholungseinheiten von je 19bp, an welches sich das phageneigene Initiatorprotein RepA bindet und eine Replikationsrunde/Bakteriengeneration einleitet. Beim Übergang vom lysogenen Zustand in den lytischen Vermehrungsprozess erfolgt ein anderer Mechanismus der Replikation, nämlich eine Replikation nach dem Schema des rollenden Rings (S.97), aus dem zuerst DNA-Concatemere hervorgehen, die später als Stücke von Genomlänge in Phagenhüllen verpackt werden.

Auf S.100 wird darauf hingewiesen, dass gelegentlich DNA-Stücke aus dem Genom der Wirtszelle in Phagenhüllen verpackt werden. Dies wird bei der Kartierung von *E. coli*-Genen über Transduktion ausgenutzt.

Der Bakteriophage P1 besitzt ein **eigenes Rekombinationssystem**, das uns später noch einmal interessieren wird (S.519). Eine Funktion des Systems ist die Auflösung von DNA-Dimeren, die bei der Ring-zu-Ring-Replikation der lysogenen DNA entstehen können. Solche Dimere würden die korrekte Weitergabe der P1-Ringe bei der Vermehrung der Bakterien stören. Das Rekombinationssystem besteht aus einem Enzym, der **Cre-Rekombinase** (*cre, causes recombination*), die spezifisch an spezielle Stellen im P1-Genom bindet, **loxP** (*lox, locus of crossing over*) genannt. Die Cre-Rekombinase vermittelt ein Schneiden und kreuzweises Verknüpfen von DNA-Strängen in *loxP*-Bereichen und damit ein Auflösen der Dimer-Struktur.

Der Vorgang hat entfernte Ähnlichkeit mit der Integration und Exzision von Lambda-DNA. Aber er ist einfacher:

- die Cre-Rekombinase kommt ohne Hilfe von Wirtszell-Funktionen aus;
- die *loxP*-Erkennungsstellen sind einfach: sie bestehen aus zwei 13-bp-langen umgekehrten Wiederholungen mit einem mittleren Abschnitt von 8bp. Ein Cre-Molekül bindet an einen 17-bp-Abschnitt (eine Wiederholung plus 4bp aus dem zentralen Abschnitt) und führt dort um sechs Nucleotide versetzte Schnitte in den beiden DNA-Strängen ein.

gern lässt. Weiter trägt der Vektor die *loxP*-Sequenz des Phagen P1, die aber für unsere Besprechung hier keine Bedeutung hat (s. Box auf S. 307).

Der Vektor wird ergänzt durch ein Kanamycin-Resistenz-Gen, das, wie bei anderen Klonierungsvektoren, eine Selektion für Bakterien mit Vektoren ermöglicht und **ein Gen mit der Bezeichnung *sacB***. Das Produkt des *sacB*-Gens benutzt Saccharose aus dem Kulturmedium und bildet daraus u. a. den Zucker Levan, der für die Bakterienzelle hoch toxisch ist. Im Vektor ist das *sacB*-Gen durch eine kurze pUC19-Sequenz stillgelegt.

Die Klonierung beginnt mit der Zubereitung des Vektors durch das Herausschneiden der pUC-DNA und der Bereitstellung von Fremd-DNA-Fragmenten mit einer Länge von 100–300 kb. Das Gemisch wird mit Ligase behandelt und mithilfe der Elektroporation (S. 300) in Bakterien transfiziert. Unter den transfizierten DNA-Molekülen befinden sich auch rückgeschlossene Vektoren ohne eingebaute Fremd-DNA. Solche Vektoren besitzen ein intaktes *sacB*-Gen. Durch Zusatz von Saccharose zum Kulturmedium lassen sich Bakterien mit rückgeschlossenen Vektoren leicht ausschalten.

PAC mit eingebauter Fremd-DNA von bis zu 300 kb bleiben über viele Bakteriengenerationen stabil. Bevor die PAC-DNA präpariert wird, schaltet man durch IPTG den „lytischen“ Replikationsapparat an und sorgt damit für eine intrabakterielle Vermehrung der PAC-DNA.

Eine **Genombibliothek** ist eine Sammlung von DNA-Fragmenten, die möglichst das gesamte Genom eines Organismus umfassen. Die Fragmente sind in Vektoren eingebaut und können deswegen einzeln isoliert und untersucht werden. Wie kann man die Bibliothek sinnvoll nutzen? Ein erstes Ziel ist die Isolierung von Klonen mit interessanten Genen. Oft erfolgt dies über cDNA-Sequenzen.

Bibliotheken von cDNA-Sequenzen

Eine typische Säugetierzelle enthält 10 000–30 000 verschiedene mRNA-Arten, von denen viele selten sind und in zehn oder weniger Kopien/Zelle vorkommen. Eine genaue Untersuchung ihrer Art und Menge ist für Genetiker wichtig, weil sie Aufschluss über den interessantesten Teil des Genoms gibt, nämlich über den Teil, der Gene trägt und aktiv transkribiert wird. Aber jeder Versuch einer direkten biochemischen Analyse solcher mRNA-Arten wäre von vornherein aussichtslos. Hier sind gentechnische Verfahren die einzig mögliche Lösung.

Das Verfahren besteht aus zwei Teilschritten:

1. Die RNA-Sequenzen werden in DNA-Sequenzen überführt, mithilfe des Enzyms reverse Transkriptase (S. 237). So wird von einer RNA eine DNA-Kopie gemacht: **copy-DNA** oder kurz **cdna**.
2. Die cDNAs können dann, wie andere DNA-Stücke, in Plasmid- oder Lambda-Vektoren eingebaut und kloniert werden. Damit wird eine Trennung des sonst unauflösbaren Gemisches von mRNA erreicht. Man erhält die Möglichkeit zur Produktion quasi beliebig großer Mengen von interessanten Sequenzen.

Plus 10.3 Genombibliotheken

Wenn Molekulargenetiker über die Vor- und Nachteile einer gegebenen Genombibliothek reden, werden früher oder später zwei Fragen gestellt:

- Ist die Bibliothek vollständig?
- Ist sie überlappend?

Vollständigkeit ist ein statistisches Problem und kann mit folgender Beziehung geschätzt werden:

$$N = \frac{\ln(1 - p)}{\ln(1 - f)}$$

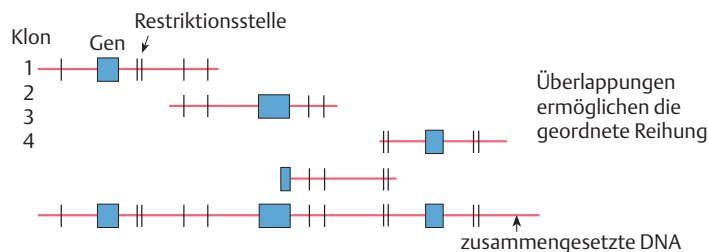
p = Wahrscheinlichkeit, mit der ein gegebenes Stück DNA (ein Gen) in der Bibliothek vorkommt

f = Verhältnis der durchschnittlichen Größe der eingebauten DNA zur Größe des Gesamtgenoms

N = Zahl der Plasmid-, Phagen- oder BAC-Klone

Beispiel: Wenn eine Wahrscheinlichkeit von 0,99 gewünscht ist, muss eine Lambda-Bibliothek (durchschnittliche Größe der eingebauten DNA: 2×10^4 bp) des Humangenoms (3×10^9 bp) etwa 700 000 Klone enthalten.

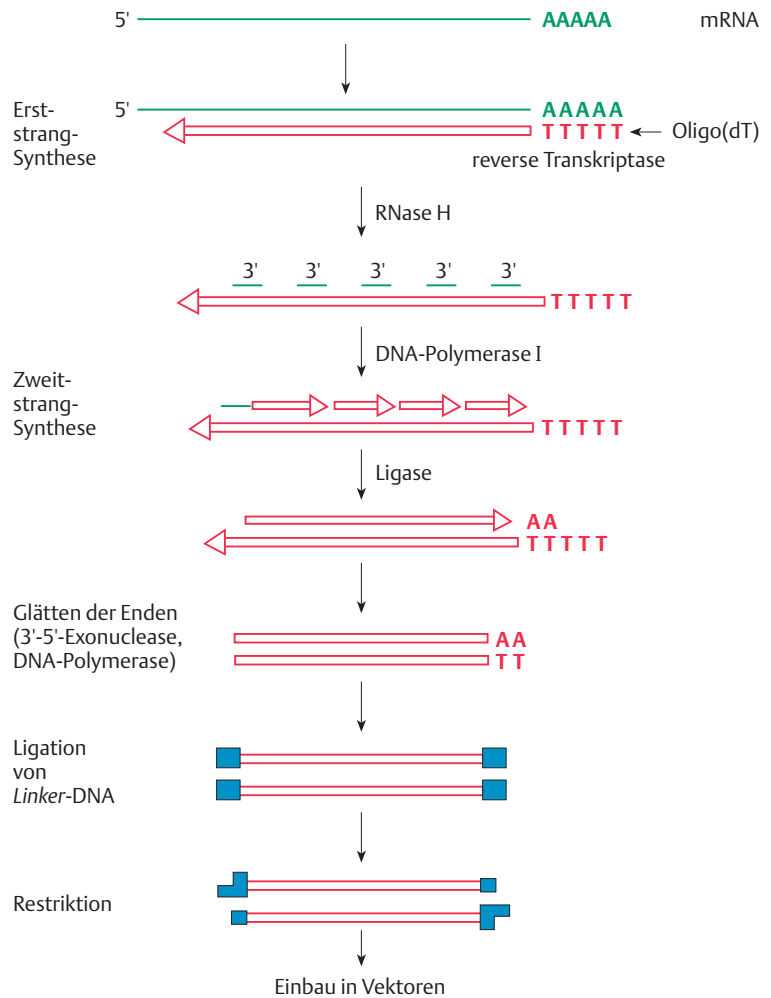
Warum sollte die Genombibliothek überlappend sein? Das ermöglicht oder erleichtert die Analyse zusammenhängender Genomabschnitte. Beachte, dass viele Gene in Eukaryotengenomen weit mehr als einige 10^4 bp umfassen und deswegen nicht auf einem Stück klonierter DNA vorkommen können. Aber wenn ein Ende eines DNA-Stücks sich am Beginn eines anderen wieder findet, kann man durch Aneinanderreihen einen kontinuierlichen Abschnitt des Genoms zusammenfügen, wie in der Skizze verdeutlicht.



Die wichtigsten Schritte bei der Herstellung der cDNA sind in der Abb.10.9 gezeigt:

- Die meisten eukaryotischen mRNAs tragen eine Folge von Adenin-Resten am 3'-Ende [poly(A)-Ende, S.431]. An dieses Ende wird ein komplementäres Oligonucleotid aus Deoxythymidin-Nucleotiden gebunden. Dieses Oligo(dT) dient als Primer für die reverse Transkriptase, die Deoxynucleosid-Triphosphate polymerisiert und die so genannte **Erststrang-Synthese** durchführt.
- Die Synthese des komplementären DNA-Stranges, die **Zweitstrang-Synthese**, kann über mehrere Verfahren erfolgen. Ein heute gebräuchliches Verfahren bedient sich der Spezifität des Enzyms **RNase H**, das den RNA-Anteil in einem RNA-DNA-Hybrid-Molekül angreift. Die Reaktion wird unterbrochen, bevor der gesamte RNA-Strang abgebaut ist. Die Reihe kurzer RNA-Stücke am Matrizenstrang

Abb. 10.9 Herstellung von cDNA. Der wichtigste Teilschritt ist die Erststrang-Synthese durch die reverse Transkriptase. Für alle weiteren Schritte stehen alternative Methoden zur Verfügung. Hier wird als Beispiel die Methode von U. Gubler und B. J. Hofman (1983) gezeigt [nach 5].



bildet eine Struktur, die ein gutes Substrat für die DNA-Polymerase I aus *E. coli* ist (S.177): Die 3'-OH-Enden dienen als Primer für die Synthese, während gleichzeitig die RNA mithilfe der 5'-3'-Exonuclease entfernt wird (in Analogie zum Prozess der *Nick Translation*, s. Abb. 7.6).

- Die hintereinander liegenden DNA-Stücke können dann durch die Ligase verknüpft werden. Im Wesentlichen ist damit die cDNA-Synthese abgeschlossen. Beachte, dass bei diesem Verfahren notwendigerweise das extreme 5'-Ende der mRNA nicht in der cDNA auftaucht. Um das 5'-Ende zu erhalten, müssen besondere Methoden angewendet werden, die wir hier aber übergehen.
- Der letzte Schritt ist die Vorbereitung der cDNA für den Einbau in Vektoren. Zunächst werden die überstehenden Einzelstrang-Bereiche durch geeignete Nucleasen entfernt. Dann geht es um die Anpassung an die Enden im Vektor. Man benutzt dazu kurze synthetische DNA-Stücke (Oligonucleotide) als Verbindung, so genannte **Linker**, die durch Ligase an die Enden der cDNA geheftet werden. Die *Linker*-

DNAs werden so ausgewählt, dass sie geeignete Schnittstellen für Restriktionsnucleasen tragen und kompatibel mit den Enden der Vektor-DNA sind. Alle übrigen Klonierschritte entsprechen der Herstellung von Genombibliotheken.

Aus der großen Zahl der möglichen Vektoren geben wir nur ein Beispiel in der Abb. 10.10. Das gezeigte Plasmid besitzt die üblichen Vektorelemente, nämlich einen Origin und ein *amp^R*-Gen, aber zusätzlich noch ein *lacI*-Gen sowie einen starken Promotor unter der Kontrolle des *lac*-Operators (S. 114) und eine optimale Ribosomen-Bindestelle (Shine-Dalgarno-Sequenz, S. 72). Daran schließt sich ein kurzes offenes Leseraster mit mehreren Histidin-Codons und einer multiplen Klonierungsstelle für den Einbau der cDNA in das richtige Leseraster an.

In Bakterien ist dieser Expressionsvektor zunächst stumm, denn der Lac-Repressor verhindert die Transkription der eingebauten cDNA. Aber durch Zusatz des Induktors IPTG (S. 111) kann die Transkription auf einfache Weise freigegeben werden. Die Bakterien produzieren dann oft große Mengen des cDNA-kodierten Proteins. Das Protein ist allerdings häufiger als gewünscht unlöslich und fällt schon in den Bakterien als Präzipitat aus. Fachleute sprechen von „inclusion bodies“.

Die Histidin-Reste am aminoterminalen Ende des bakteriell exprimierten Eukaryotenproteins dienen zur Reinigung, denn Proteine mit engen Folgen von Histidin binden Nickel-Salze und können durch Säulen mit nickelspezifischen Chelatbildnern von den vielen anderen Proteinen des Bakterienextraktes getrennt werden.

Das gereinigte Protein kann zu Struktur- und Funktionsuntersuchungen benutzt werden oder als Antigen zur Erzeugung von Antikörpern in Kaninchen, Mäusen oder Hühnern dienen. Wie wir gleich sehen werden, sind spezifische Antikörper wichtige Werkzeuge gentechnischer Arbeit.

Zweck des Klonierungsverfahrens ist die Auftrennung eines komplexen Gemisches von Nucleinsäure-Fragmenten in die einzelnen Bestandteile: Man erhält jeweils ein Fragment pro Bakterien- oder Bakteriophagen-Klon.

Benutzung von Bibliotheken

Für manche Forscher ist mit der Herstellung von cDNA-Bibliotheken schon der Zweck erfüllt. Die Sequenzen der cDNAs werden Klon für Klon bestimmt. Das erlaubt Einblicke in die Gen-Aktivität einer Zell- oder einer Gewebsart, denn die Art und Menge von cDNAs in einer Bibliothek geben das Spektrum der mRNAs wieder.

Aber viele Forscher sind an speziellen Genen interessiert. Die betreffenden Gene müssen aus der Bibliothek isoliert werden (Abb. 10.11). Fast immer muss dazu erhebliche Vorarbeit geleistet werden.

Vorarbeiten sind:

1. Herstellung von Antikörpern gegen ein gereinigtes Protein oder
2. Bestimmung von kurzen Aminosäure-Sequenzabschnitten. In günstigen Fällen genügt eine einwandfreie Darstellung des Proteins in der Polyacrylamid-Gel-Elektrophorese, gefolgt von einer Mikro-Sequenzierung einiger kurzer proteolytischer Spaltprodukte (Abb. 10.11). Mithilfe der Codewort-Tabelle der Abb. 4.33 kann die Aminosäure-Sequenz in eine Nucleinsäure-Sequenz umgeschrieben wer-

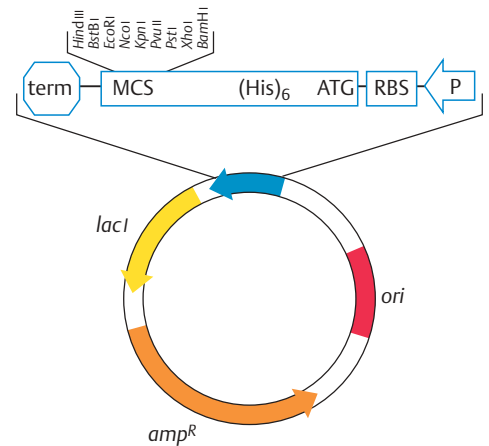


Abb. 10.10 Ein Plasmid-Expressionsvektor. Bei ATG (dem universellen Start-Codon) beginnt das offene Leseraster und setzt sich mit einer Reihe von Histidin-Codons (His) bis zur multiplen Klonierungsstelle (MCS) fort.

P = Promotor

RBS = Ribosomen-Bindestelle

term = Termination

Unser Beispiel ist das Plasmid pRSET (angeboten von der Firma Invitrogen, San Diego, USA). Vereinfachte Darstellung.

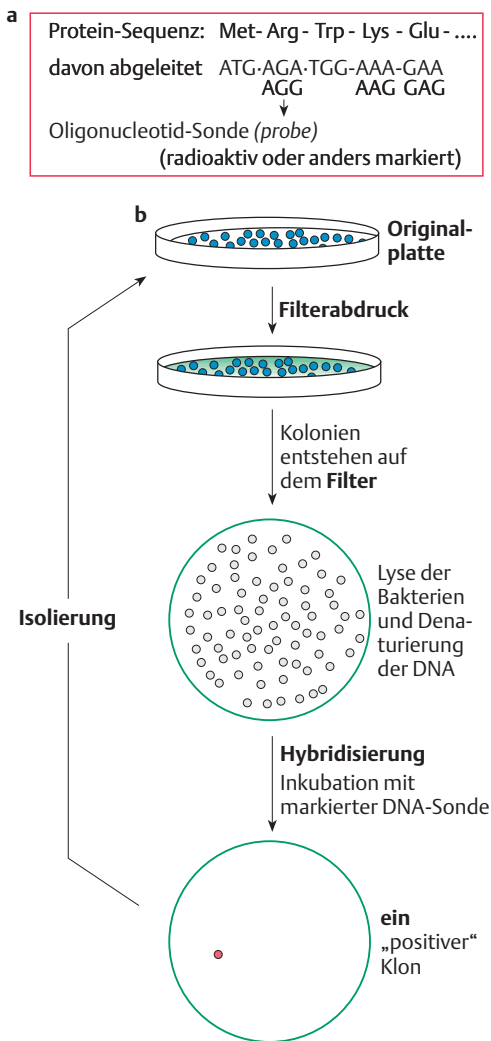


Abb. 10.11 Screening einer cDNA-Bibliothek mit Oligonucleotid-Sonden.

a Aus einer kurzen Peptid-Sequenz kann die Struktur von Oligonucleotid-Sonden abgeleitet werden. Wegen der Redundanz des genetischen Codes müssen mehrere Oligonucleotide verwendet werden.

b Mit einem Nitrocellulose-Filter wird ein Abdruck der Originalplatte genommen. Auf dem Filter entstehen Kolonien. Zur Lyse der Bakterien und zur Denaturierung der DNA wird das Filter mit 0,5 M NaOH behandelt. Die DNA wird bei 80 °C im Vakuum an den Filter fixiert. Dann wird der behandelte Filter mit der markierten Sonde inkubiert. Es erfolgt eine Hybridisierung mit komplementären DNA-Sequenzen. Ein positiver, markierter Klon wird identifiziert. Die positiven Kolonien werden von der Originalplatte isoliert.

Einfacher geht es mit PCR-Methoden (s. S. 318 f.).

den. Das entsprechende Oligonucleotid wird dann mit chemischen Verfahren synthetisiert. Da der genetische Code redundant ist, treten fast immer Zwei- oder Mehrdeutigkeiten auf, so dass möglichst viele Oligonucleotide hergestellt werden müssen.

Spezifische Antikörper und DNA-Oligomere sind **Sonden (probes)**, mit denen man cDNA-Bibliotheken durchsucht (*screening*). Beide Verfahren setzt man prinzipiell ähnlich ein.

In jedem Fall wird zuerst ein Abdruck einer Original-Agarplatte mit Bakterienkolonien oder mit Bakteriophagen-Plaques auf einem Nitrocellulose-Filter hergestellt. Der Filter übernimmt – wie beim Stempeln – das Originalmuster von Kolonien oder Plaques. Am Filter werden die notwendigen Analysen durchgeführt, also entweder eine Reaktion des gebundenen Proteins mit Antikörpern oder eine Hybridisierung der gebundenen DNA mit markierten DNA-Oligonucleotid-Sonden (Abb. 10.11). Reaktionen treten nur im Bereich positiver Kolonien oder positiver Plaques auf. Deswegen entspricht die Position auf dem Filter der Position auf der Originalplatte, wo dann die wenigen positiven unter den vielen hundert oder tausend negativen Klonen identifiziert werden können.

Positive Kolonien oder positive Plaques werden isoliert und mit den Methoden der Mikrobiologie vermehrt:

- Bakterienkolonien werden in Nährmedium aufgelöst und bis zu hohen Bakterien-dichten kultiviert (S. 88), der Plasmid-DNA-Klon wird präpariert.
- Bakteriophagen-Klone werden zur Infektion großer Bakterienmengen verwendet, bis ausreichend Lambda-DNA für weitere Untersuchungen vorhanden ist.

Zu solchen weiteren Untersuchungen gehört oft die Isolierung des Gens. Dazu verwendet man die zuvor isolierte cDNA als Sonde zum Durchsuchen der entsprechenden **Genombibliothek** (Plus 10.3). Das Verfahren entspricht dem Oligonucleotid-Screening der Abb. 10.11. Am Ende liegt dann das isolierte Gen vor.

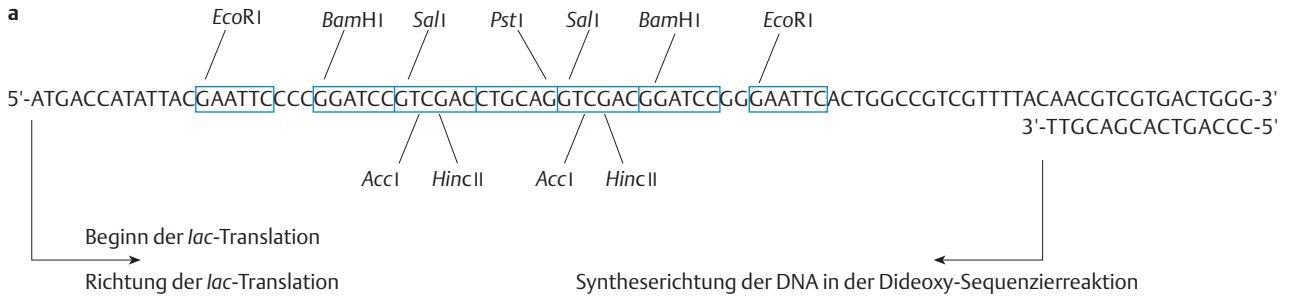
Ergebnisse solcher Bemühungen füllen die nächsten Kapitel dieses Buches und ungezählte Seiten in den Fachzeitschriften der Genetik, Molekularbiologie, Biochemie, Zellbiologie, Tier- und Pflanzenphysiologie.

Aber cDNAs und Gene geben ihre Information nur preis, wenn sich Forscher die Mühe einer Sequenzbestimmung machen. Deswegen muss unser Überblick über die methodischen Grundlagen der molekularen Genetik zumindest eine kurze Beschreibung der gebräuchlichsten Sequenziermethoden enthalten.

Sequenzieren

Die beiden wichtigen Verfahren zur Bestimmung von DNA-Sequenzen sind die chemische Methode von A. M. Maxam und W. Gilbert (1977) und die biochemische Dideoxy- oder Kettenabbruch-Methode von F. Sanger (1977).

Obwohl die Kettenabbruch-Methode umständlich ist, wird sie von den meisten Experimentatoren, wegen ihrer Verlässlichkeit und Genauigkeit, bevorzugt. Mithilfe des Verfahrens wurden fast alle der vielen



DNA-Sequenzen gewonnen, die sich in den internationalen Datenbanken anhäufen. Unter den analysierten Sequenzen sind auch die kompletten Sequenzen ganzer Genome von Bakterien, Pflanzen und Tieren. Deswegen ist es berechtigt, wenn wir hier nur das Sanger-Verfahren betrachten.

Zur Routine-Prozedur gehört das Klonieren der DNA in M13-RF-DNA, mit dem Ziel der Herstellung einzelsträngiger DNA-Ringe.

Molekularbiologen verwenden Derivate des natürlichen M13-Phagen für Klonierungsarbeiten (Plus 10.4). Die intergenische Region trägt Teile des *lacZ*-Gens für eine Blau-Weiß-Selektion (S. 301 f.) und eine multiple Klonierungsstelle (Abb. 10.12). An diesen Stellen wird M13-RF für die Aufnahme von Fremd-DNA linearisiert. RF-/Fremd-DNA-Konstrukte werden in Bakterien eingebracht, und im Zuge eines Infektionsprozesses entstehen Nachkommenphagen mit Einzelstrang- plus Fremd-DNA.

Diese Einzelstrang-Ringe sind Substrate für die *in-vitro*-DNA-Synthese nach dem Schema der Abb. 10.13:

- Ein **DNA-Primer** wird durch Hybridisierung eines Oligonucleotides an einen Bereich vor der multiplen Klonierungsstelle bereitgestellt.
- **DNA-Polymerase** vom Typ des Klenow-Fragmentes der bakteriellen DNA-Polymerase I (S. 177) oder eine andere DNA-Polymerase (ohne 5'-3'-Exonuclease) benutzt das 3'-OH-Ende des Primers für die Synthese eines Komplementärstranges. Ein vielverwendetes Enzym bei den großen Sequenzierprojekten ist die DNA-Polymerase des Bakteriophagen T7 („T7-DNA-Polymerase“).

Der Trick bei der Sequenzierreaktion ist die gezielte, aber statistisch verteilte Unterbrechung der Komplementärstrang-Synthese. Das erreicht man durch Zusatz von **Dideoxynucleotiden** (ddNTP) zu dem Gemisch von normalen dNTPs. Dideoxynucleotide haben keine 3'-OH-Gruppe, sodass die Synthese unterbrochen wird, wo immer ein Dideoxynucleotid in die wachsende DNA-Kette eingebaut wird. Die Anwendung dieses Prinzips erläutert die Abb. 10.13.

In der Praxis lassen sich auf diese Weise einige hundert Nucleotide einer DNA-Sequenz in einem Experiment „lesen“ (Abb. 10.14). Um noch längere DNA-Stücke sequenzieren zu können, werden „Primer“ eingesetzt, die komplementär zu den zuvor bestimmten DNA-Sequenzen sind.

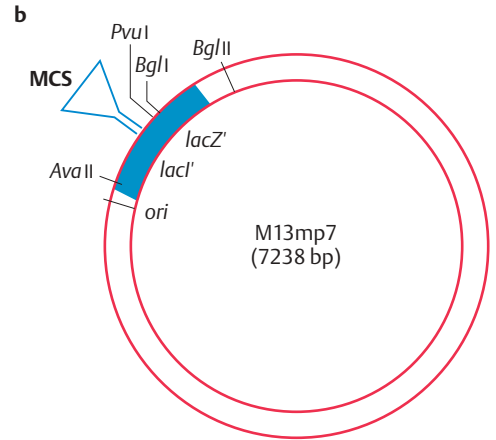


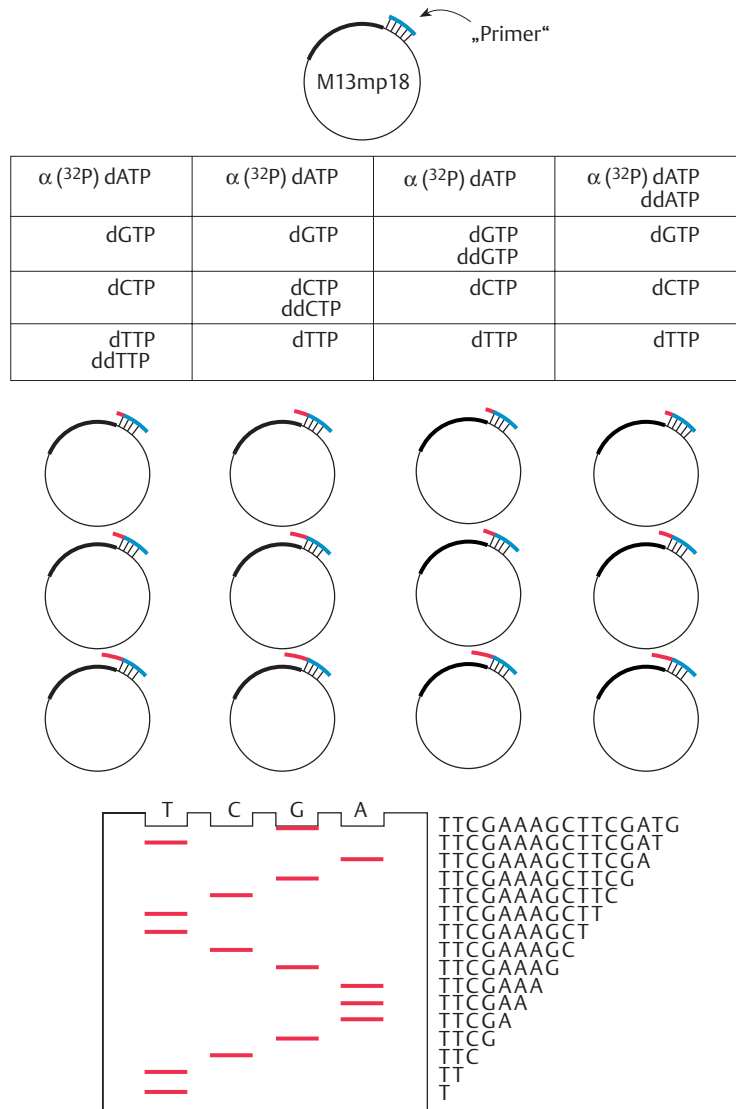
Abb. 10.12 M13 als Klonierungsvektor.

a Die multiple Klonierungsstelle (MCS) von M13 ist hier vergrößert dargestellt, um die Lage des Primers für die Sequenzierreaktion deutlich zu machen [aus 8].

b In die „intergenische Region“ (blau) der natürlichen M13-DNA (s. Plus 10.4) wurden 5'-Bereiche des *lacZ*-Gens und eine MCS eingebaut.

Abb. 10.13 Prinzip der Sequenzreaktion nach der Dideoxy- oder Kettenabbruch-Methode.

- Vier Ansätze werden parallel vorbereitet. Jeder Ansatz enthält ein radioaktiv oder sonst wie markiertes Nucleotid (hier: α -[^{32}P]-dATP) und die drei anderen nichtmarkierten Deoxynucleosid-Triphosphate. Jeder Ansatz enthält zudem ein Dideoxynucleotid, entweder ddTTP, ddCTP, ddGTP oder ddATP. Je nach experimenteller Situation wird ein Verhältnis von 1/50, 1/100 oder 1/200 von ddNTP zu dNTP gewählt.
- Nach Zusatz der DNA-Polymerase beginnt die Komplementärstrang-Synthese. Sie kommt zum Halt, wenn zufällig ein Dideoxynucleotid statt des normalen Deoxynucleotids in das aktive Zentrum der DNA-Polymerase gelangt. Im ersten Ansatz wird das der Fall sein, wenn die Sequenz des Matrizenstranges den Einbau eines Thymin-Nucleotids verlangt. Mit anderen Worten, im ersten Ansatz erhält man eine Kollektion von DNA-Fragmenten, deren Längen die Positionen von Adenin-Resten im Matrizenstrang wiedergeben. Entsprechendes gilt für die Längen der Syntheseprodukte in den anderen Ansätzen.
- Der kritische Teil der Methode ist die genaue Auftrennung der Syntheseprodukte. Dazu werden Matrize und synthetisierte Komplementärstränge durch Denaturierung voneinander gelöst und auf dünnen, besonders zubereiteten Polyacrylamid-Gelen analysiert. Nur die Syntheseprodukte sind markiert, in unserem Beispiel durch radioaktiven Phosphor (^{32}P), heute aber meist durch Fluoreszenzfarbstoffe. Und nur die Syntheseprodukte werden durch geeignete Methoden sichtbar gemacht, in unserem Beispiel durch Autoradiographie auf einen Röntgenfilm. Durch das kleinste, am weitesten gewanderte Fragment wird das erste Nucleotid in der Sequenz angezeigt, durch das nächst größte Fragment das zweite usw. [nach 11].



Sequenziermaschinen

Wie erwähnt, beruhen fast alle Sequenziervorhaben, auch die großen Sequenzierprojekte, auf dem Kettenabbruch-Verfahren, dessen Prinzip die Abb. 10.13 zeigt. F. Sanger und Mitarbeiter konnten mit dieser Methode zum ersten Mal ein ganzes Genom, nämlich das des Bakteriophagen $\Phi\text{X} 174$, mit 5386 Basenpaaren sequenzieren (1977). Damals und in den darauf folgenden Jahren kamen fast ausschließlich radioaktiv markierte Nucleotide zur Anwendung, wie bei unserem Beispiel-Experiment der Abb. 10.14.

Aber später nahmen die Sequenzierarbeiten ein solches Ausmaß an, dass der Umgang mit radioaktiven Verbindungen zu aufwendig und auch zu umständlich wurde. Stattdessen werden Nucleotide eingesetzt, die durch Fluoreszenzfarbstoffe markiert sind.

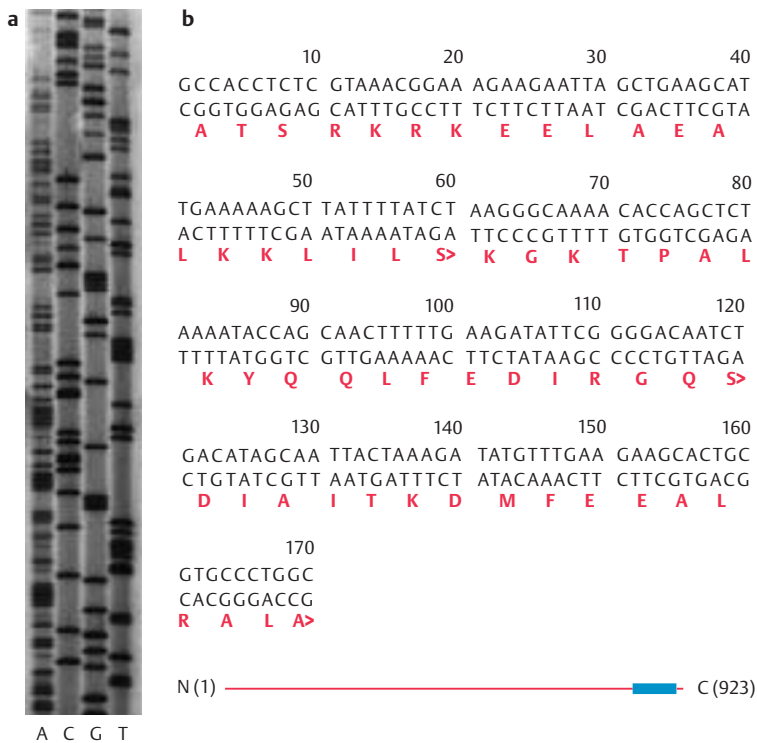


Abb. 10.14 Autoradiographischer Nachweis einer Sequenzreaktion.

a Autoradiographie eines Sequenzgels (A, C, G, T: Ergebnisse mit ddATP, ddCTP, ddGTP bzw. ddTTP in den Reaktionsgemischen).

b Computerunterstützte Auswertung: Ausdruck der Sequenz in Doppelstrang-Form und Ableitung der kodierten Aminosäure-Sequenz (rot). Die abgebildete Sequenz ist Teil der cDNA eines menschlichen Replikationsproteins (Mcm4). Die Strichskizze zeigt, dass die hier abgebildete Sequenz nur die Information für einen kleinen Teil (blaue Box) des Proteins liefert (Christine Musahl, Konstanz 1992).

Ebenso wurde die Enzymologie verbessert: Anstelle des Klenow-Fragments der bakteriellen DNA-Polymerase kamen die T7-DNA-Polymerase oder auch andere DNA-Polymerasen, zur Anwendung; Enzyme, die längere DNA-Abschnitte bei höherer Prozessivität (S. 176) und Genauigkeit synthetisieren können.

Der Einsatz von fluoreszenzmarkierten Nucleotiden ermöglichte vor allem eine entscheidende Weiterentwicklung der Instrumente. Das technische Prinzip dabei ist, dass die Syntheseprodukte (mit eingebauten fluoreszierenden Nucleotiden) am Ende ihrer Wanderung durch ein Polyacrylamid-Gel von geeigneten Laserstrahlen angeregt werden. Ein Computer mit geeigneten Software-Paketen registriert die Fluoreszenzsignale und übersetzt sie gleich in DNA-Sequenzen (Abb. 10.15).

Dieses Prinzip wird in den automatischen Sequenziermaschinen einzelner Herstellerfirmen in unterschiedlicher Weise realisiert. Ein wichtiger Unterschied der Verfahren:

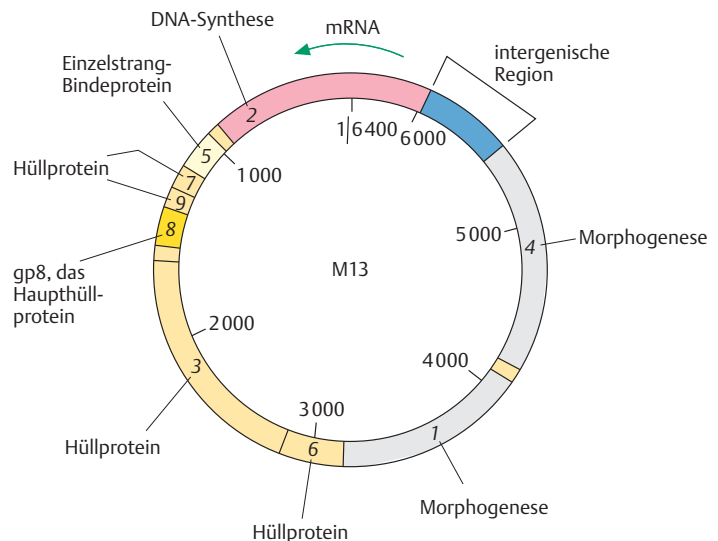
- vier parallele Elektrophorese-Bahnen unter Verwendung je eines Fluoreszenzfarbstoffes;
- oder eine Bahn, wobei die vier eingesetzten Nucleotide jeweils einen anderen Fluoreszenzfarbstoff tragen.

In jedem Fall ist es das Ziel, den Prozess zu automatisieren und so viele Sequenzierreaktionen wie möglich in einem einzigen Elektrophorese-Lauf durchzuführen. Neuere Versionen sind Sequenziermaschinen, die statt der herkömmlichen Flachbettgele viele parallel laufende gelgefüllte Kapillaren besitzen. Mit solchen Geräten lassen sich täglich hunderte, ja tausende Sequenzreaktionen durchführen, wobei jede Reaktion eine DNA-Sequenz von 600–700 Basenpaar-Länge liefert

Plus 10.4 Der filamentöse Phage M13 und seine Verwandten

Der Bakteriophage M13 und seine Verwandten, die Phagen fd, f1 und andere, sind fast 900 nm lang mit einem Durchmesser von nur 6 nm. Die DNA, ein ringförmiger Einzelstrang aus etwa 6400 Nucleotiden, ist von einer Röhre aus 2800 Kopien des Haupthüllproteins gp8 eingeschlossen mit Extra-Hüllproteinen an den Enden.

Phagenpartikel binden an die Spitzen der F-Pili von F⁺- oder Hfr-Bakterien (S. 96) und schleusen ihre DNA durch den Pilus in die Zelle. Dort wird der infizierende Einzelstrang in eine doppelsträngige Form überführt, die als **replikative Form** oder **RF-DNA** bezeichnet wird. Doppelsträngige RF-DNA ist das eigentliche Virusgenom und wird zur Expression der Gene transkribiert. RF-DNA repliziert mehrfach, bevor es später als Matrize zur Herstellung von Nachkommen-Einzelsträngen dient. Einzelstrang-Synthese erfolgt nach dem Modell des rollenden Ringes (S. 97). An den Einzelstrang lagern sich die Hüllproteine. Die fertigen Nachkommenphagen verlassen den infizierten Wirt **ohne Lyse**. Mit anderen Worten, M13-infizierte Bakterien sind lebensfähig und teilen sich bei gleichzeitiger Produktion von Nachkommenphagen [nach 10].



Das Phagen-genom trägt zehn Gene mit Funktionen für die Replikation, die Phagenhülle und den Zusammenbau des Partikels (Morphogenese). In unserem Zusammenhang ist die intergenische Region wichtig, ein Abschnitt von etwa 500 Nucleotiden zwischen dem Gen 2 und dem Gen 4 (s. Abb.). Die intergenische Region kann viel größer als normalerweise sein, ohne dass die Vermehrung beeinflusst wird. Dabei wird das Phagenpartikel länger durch Aufnahme einer größeren Zahl von gp8-Molekülen in die Protein-Hülle. Aufgrund dieser Tatsache sind M13-Phagen für die Gentechnik geeignet.

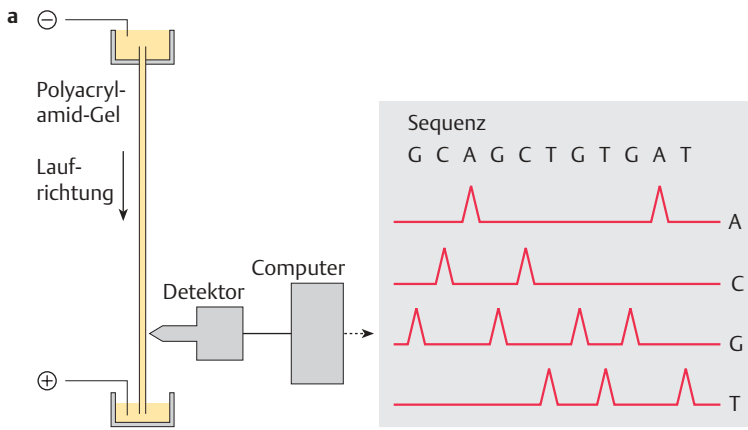
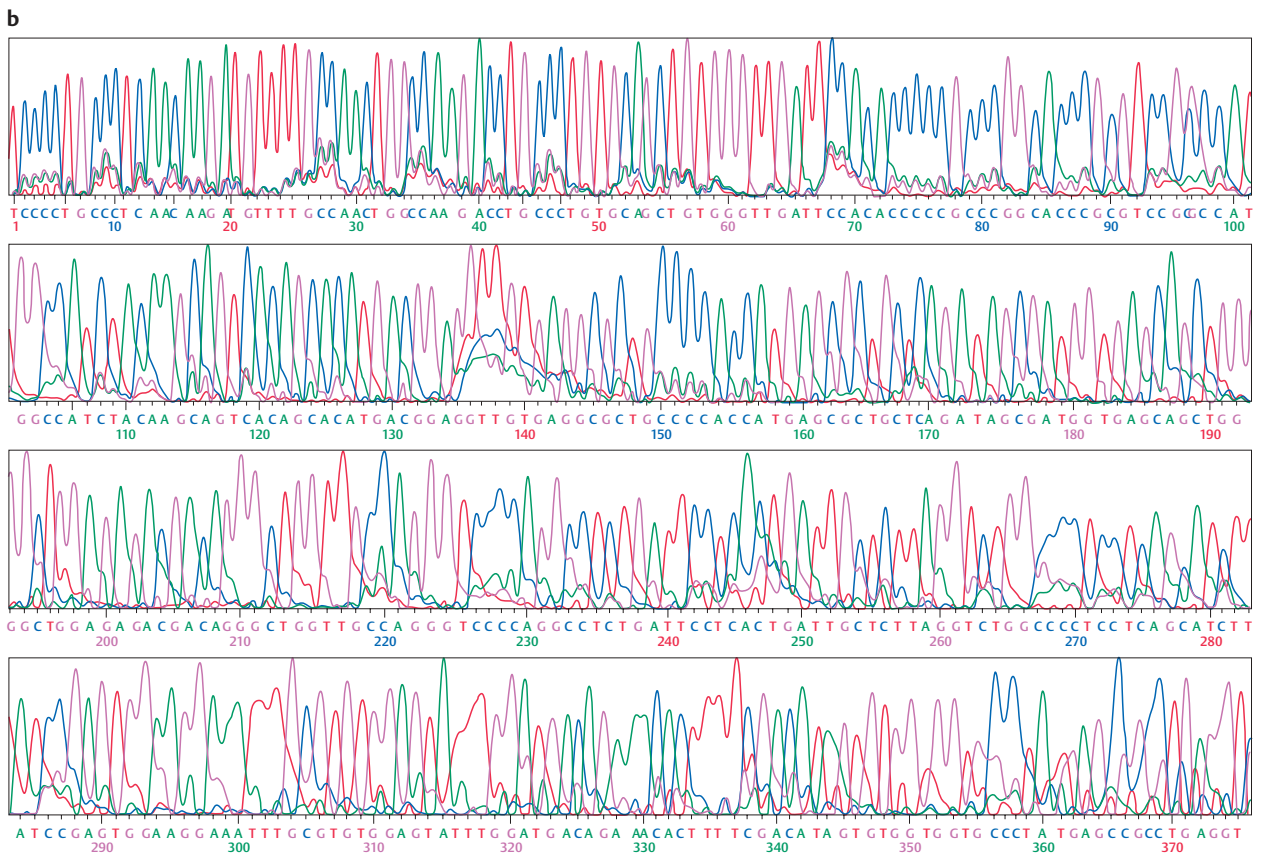


Abb.10.15 Sequenzieren mit fluoreszenzmarkierten Nucleotiden.

a Die DNA-Fragmente wandern im elektrischen Feld an einem Laserstrahl vorbei, der die Fluoreszenzen anregt. Die Signale werden durch einen Detektor aufgenommen und von einem Computer analysiert. **b** Ergebnis: eine Folge von Fluoreszenz-Peaks, die den einzelnen Nucleotiden zugeordnet werden können. Falls Unsicherheiten auftreten, wie etwa im Bereich der Nucleotide 135–140, muss die Sequenzreaktion wiederholt werden. Moderne Sequenziermaschinen liefern verlässliche Sequenzen von 600–700 Basenpaar-Längen.



(Plus 10.5). In den großen Sequenzierzentren sind Hundert und mehr dieser Geräte buchstäblich Tag und Nacht im Einsatz. Nur auf diese Weise lassen sich die enormen Datenmengen gewinnen, die zur Entzifferung der großen Genome von Mensch, Tier und Pflanze notwendig sind (siehe Kap.17).

Plus 10.5 Ein Viertel-Jahrhundert Sequenzierarbeit. Immer mehr Sequenzen

Tagtäglich gelangen Millionen Basenpaare neuer DNA-Sequenzen in die internationalen Datenbanken – Sequenzen der Genome von Viren, Bakterien, Archaea sowie von einzelligen Eukaryoten, von Tieren und Pflanzen. Der computergestützte Zugriff auf diese Datenfülle gehört zu den Routinearbeiten der Molekularbiologen. Wir verweisen darauf an geeigneten Stellen des Buches.

In früheren Jahren wurde die vollständige Sequenzierung eines Virus- oder Bakteriengenoms oder die Sequenzierung eines eukaryotischen Chromosoms als besondere wissenschaftliche Leistung gewürdigt. Die Zusammenstellung gibt einen knappen Rückblick auf ein Vierteljahrhundert Sequenzierarbeit: immer längere Sequenzen.

1977	Bakteriophage Φ X 174	5 386 bp
1978	Simian Virus 40 (SV40)	5 243 bp
	Bakteriophage M13	6 407 bp
1981	Mitochondrien-DNA (Mensch)	16 569 bp
1982	Bakteriophage Lambda	48 502 bp
1986	Chloroplasten-DNA (Tabak)	155 844 bp
1990	Vaccinia-Virus	191 737 bp
1992	Chromosom III (Hefe)	315 316 bp
1995	<i>Haemophilus influenzae</i> (Bakterium)	1 830 000 bp
1996	das gesamte Genom der Hefe <i>Saccharomyces cerevisiae</i>	12 070 000 bp
1997	<i>Escherichia coli</i>	4 290 000 bp
1997	<i>Archaeoglobus fulgidus</i> (Schwefel-metabolisierendes Archaeon)	2 180 000 bp
1998	<i>Mycobacterium tuberculosis</i>	4 410 000 bp
1998	(fast) das gesamte Genom des Nematoden <i>C. elegans</i>	97 000 000 bp
2000	(fast) das gesamte Genom von <i>Drosophila melanogaster</i>	180 000 000 bp
2000	(fast) das gesamte Genom der Pflanze <i>Arabidopsis thaliana</i>	120 000 000 bp
2001	(fast) das gesamte Genom des Menschen	2 700 000 000 bp

Polymerase-Kettenreaktion (PCR)

Unter der Abb. 10.11 haben wir notiert, dass man heute PCR-Methoden zum Screening von Genom- und cDNA-Bibliotheken einsetzt. Was ist PCR?

Die Polymerase-Kettenreaktion, abgekürzt **PCR** (*polymerase chain reaction*), ist eine unentbehrliche Methode für viele Fragestellungen der heutigen molekularen Genetik. Damit gelingt der Nachweis und die Vermehrung kleinster Mengen spezifischer DNA-Sequenzen bei einfacher Handhabung.

Das Prinzip der PCR ist die enzymatische Vermehrung eines DNA-Abschnittes zwischen zwei Oligonucleotid-Primern, die gegenläufig an komplementäre DNA-Stränge gebunden sind. Voraussetzung für den Einsatz von PCR-Methoden ist also die Information über die Nucleotid-Sequenzen beiderseits des DNA-Abschnittes und die Verfügbarkeit von geeigneten Oligonucleotiden. Manche großen molekularbiologischen

Forschungseinrichtungen besitzen Geräte zur organisch-chemischen Synthese von Oligonucleotiden. Die meisten Laboratorien wenden sich an eine der Firmen, die sich auf die Synthese von Oligonucleotiden spezialisiert haben.

Die Oligonucleotid-Primer werden im Überschuss unter Hybridisierungsbedingungen (bei etwa 70 °C) zu einer DNA-Präparation gegeben. DNA-Polymerase heftet Nucleotide an die 3'-OH-Primer-Enden und synthetisiert komplementäre DNA-Sequenzen. Die entstandenen DNA-Syntheseprodukte werden bei 94 °C denaturiert. Dann folgt eine neue Runde mit der Hybridisierung von Oligonucleotid-Primern und DNA-Strangsynthesen. Die Zyklen – Denaturierung, Hybridisierung, DNA-Synthese – werden 25-mal, und wenn nötig häufiger, wiederholt. Es handelt sich um eine Kettenreaktion, bei der winzige Mengen einer gegebenen DNA-Sequenz um das Millionenfache amplifiziert werden können (Abb. 10.16).

Als die PCR-Methode eingeführt wurde (K. B. Mullis, 1986), verwendete man meist das Klenow-Fragment der bakteriellen DNA-Polymerase (S.177). Aber dieses Enzym denaturiert bei den hohen Temperaturen, die für die Denaturierung der Doppelstrang-DNA und für die Hybridisierung von Oligonucleotiden notwendig sind. Deswegen musste man nach jedem Amplifikationsschritt eine neue Enzymprobe zu dem Reaktionsansatz geben. Diese Anfangsschwierigkeit ist inzwischen längst überwunden. Heute verwendet man thermostabile DNA-Polymerasen aus Bakterien, die normalerweise in heißen Quellen vorkommen. Ein Prototyp dieser DNA-Polymerasen ist die **Taq-Polymerase** von *Thermus aquaticus*.

Der Einsatz von thermostabilen DNA-Polymerasen ermöglicht die Automatisierung der PCR-Methode. Einfache PCR-Apparate, die nach programmiertem Muster den Temperaturwechsel durchführen, gehören zur Grundausstattung von molekularbiologischen Laboratorien.

Mit der PCR-Methode kann man spezifische Sequenzen in einzelnen Klonen, aber auch in Klon-Gemischen aus Genom- und cDNA-Bibliotheken nachweisen. Über einige weitere der vielen Anwendungsmöglichkeiten berichten wir in den folgenden Kapiteln.

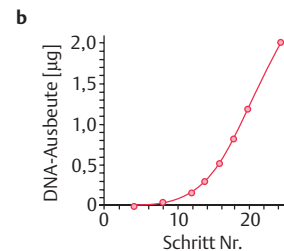
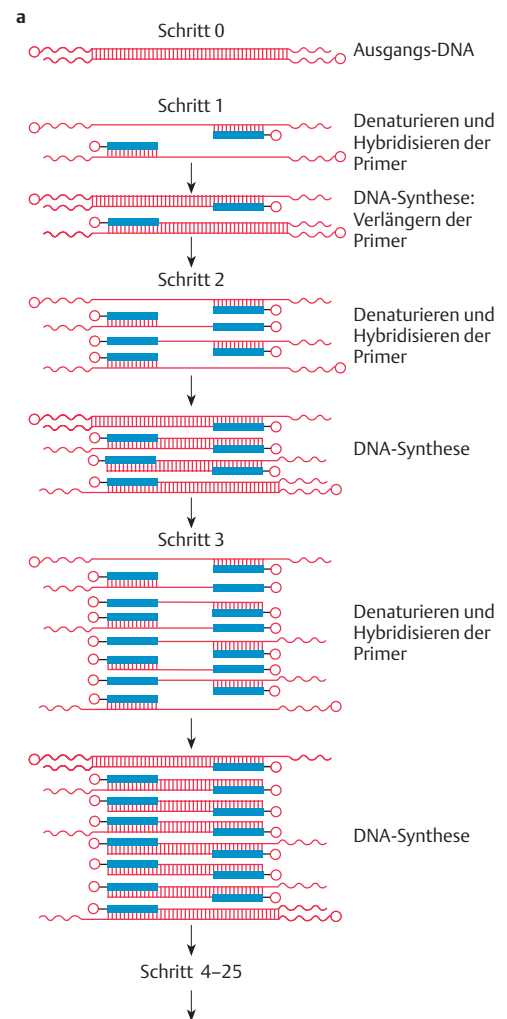


Abb. 10.16 Polymerase-Kettenreaktion (PCR).
a Schritt 1: Sense- und Antisense-Oligonucleotide (Forward- and Reverse-Primers) werden an denaturierte DNA hybridisiert. Schritt 2: Taq-DNA-Polymerase synthetisiert Komplementärstränge. Schritt 3: Denaturierung, erneute Hybridisierung von Oligonucleotiden und Komplementärstrang-Synthese.
b Zunahme der Syntheseprodukte [nach 14].