

13.1 DNA als Träger genetischer Informationen

Mit Ausnahme einiger Viren ist das genetische Material aller Organismen Desoxyribonucleinsäure (DNA), die als Doppelhelix mit zwei antiparallelen Strängen vorliegt. Die Information ist in der Abfolge von vier aromatischen Basen, A, G, C und T gespeichert, die in zwei komplementären, über Wasserstoffbrücken miteinander verbundenen Paaren A=T und G=C vorliegen.

Aus didaktischen Gründen haben wir für dieses Kapitel die Geschichte der Genetik, wie sie in der Box 13.1 skizziert ist, auf den Kopf gestellt. Bevor wir uns den Gesetzen der Vererbung, der Chromosomenverteilung bei der Zellteilung und den Veränderungen im Erbgut (Mutationen) zuwenden, wollen wir uns mit der DNA als Erbsubstanz beschäftigen.

Die Ausarbeitung der DNA-Struktur als Doppelhelix mit zwei antiparallelen und zueinander komplementären Strängen durch Watson und Crick (Abb. 13.1) machte mit einem Schlag deutlich, daß DNA geradezu ideal für den biologischen Zweck der Speicherung großer Mengen von Information auf kleinstem Raum und für die identische Weitergabe ist (Kap. 1.3.2):

- Das **genetische Alphabet** besteht aus nur vier Buchstaben (A, G, C und T), die biochemisch durch die vier **stickstoffhaltigen, heterocyclischen Aromaten** Adenin (A), Guanin (G), Cytosin (C) und Thymin (T) repräsentiert sind. Wegen der positiven Ladung durch die Stickstoffatome in den Ringen werden diese auch als Basen und nach ihrer Grundstruktur mit zwei Ringen als Purinbasen (A und G) bzw. mit einem Ring als Pyrimidinbasen (C und T) bezeichnet (Abb. 13.1b).
- Die regelmäßige Anordnung der Buchstaben beruht auf besonderen Eigenschaften dieser Basen. Jeweils eine Purin- und eine Pyrimidinbase bilden ein **komplementäres Paar**, das über **Wasserstoffbrücken** in Wechselwirkung tritt. Dies gilt sowohl für das G=C-Paar mit drei Wasserstoffbrücken als auch für das A=T-Paar mit zwei Wasserstoffbrücken (Abb. 13.1a). Die Bildung solcher Brücken mit einem Energiegehalt von je etwa 2 kcal · mol⁻¹ beruht auf einer Elektronendefizienz an den Wasserstoffatomen und einem Elektronenüberschuß durch freie Elektronenpaare an den Stickstoff- bzw. Sauerstoffatomen (Abb. 1.37 S. 43).
- Die **Basenpaarung** im Inneren der Doppelhelix, die nach außen von der Desoxyribosephosphatkette abgeschirmt wird, ermöglicht eine sehr **regelmäßige Struktur der Doppelhelix**, da beide Paare fast identische räumliche Strukturen haben. Diese Regelmäßigkeit der DNA-Struktur ist von großer Bedeutung für die frühzeitige Erkennung von Fehlern in der Sequenz, wie sie z.B. bei der Replikation entstehen können. Solche Fehler „verzerren“ die Doppelhelix und werden durch Reparatursysteme beseitigt (Kap. 13.5 und Kap. 13.8.5).
- Die Besonderheiten der Doppelhelix bergen aber auch das Geheimnis der **Informationsweitergabe** bei der Zellteilung nach **DNA-Replikation** ebenso wie bei der **Transkription**. In beiden Fällen wird von den beteiligten Enzymen (DNA-Polymerase bzw. RNA-Polymerase) eine komplementäre Kopie (K) angefertigt, deren Buchstabenfolge eindeutig durch das Gesetz der Komplementarität zum **Matrizenstrang (M)** bestimmt ist. An einem DNA-Strang M mit der Buchstabenfolge 3'-ATGACTG-5' kann nur ein **komplementärer Strang K** mit der Buchstabenfolge 5'-TACTGAC-3' synthetisiert werden. Wir werden bei der Besprechung der Details noch sehen, daß der M-Strang immer vom 3'- zum 5'-Ende

Box 13.1 Von der klassischen zur molekularen Genetik

Die ersten Vererbungsexperimente wurden im 19. Jahrhundert von einem Augustinermönch in Brünn (Brno), Gregor Mendel, durchgeführt. Die daraus abgeleiteten Gesetzmäßigkeiten, die wir heute als Mendel-Regeln bezeichnen, erschienen 1866 in den Verhandlungen des Naturwissenschaftlichen Vereins zu Brünn. Sie blieben jedoch weitgehend unbeachtet, bis sie auf der Grundlage ähnlicher Experimente Anfang des 20. Jahrhunderts durch die Botaniker C. E. Correns (Berlin), E. von Tschermak-Seysenegg (Gent, Wien) und H. de Vries (Amsterdam) wiederentdeckt wurden. Seither hat die Genetik eine stürmische Entwicklung erfahren. Waren die Versuchsobjekte der „klassischen“ Genetik zunächst höhere Pflanzen und dann vor allem Tiere, so begann mit den Experimenten an Pilzen, Bakterien und Bakteriophagen in den USA um 1940 die Ära der molekularen Genetik. Zwei Meilensteine prägen die weitere Entwicklung: 1944 gelingt O. T. Avery und Mitarbeitern der Nachweis, daß die pathogenen Eigenschaften des Bakteriums *Streptococcus pneumoniae* durch Desoxyribonucleinsäure (DNA) auf nichtpathogene Stämme übertragen werden können, und 1953 publizierten J. D. Watson und F. Crick ihr bahnbrechendes Modell der DNA-Struktur als Doppelhelix. In den folgenden Jahren führte die Entdeckung von Enzymen, die DNA an spezifischen Stellen spalten können (Restriktionsendonucleasen, W. Arber 1962), zur Entwicklung der Gentechnik (Kap. 13.11 und Kap. 13.12).

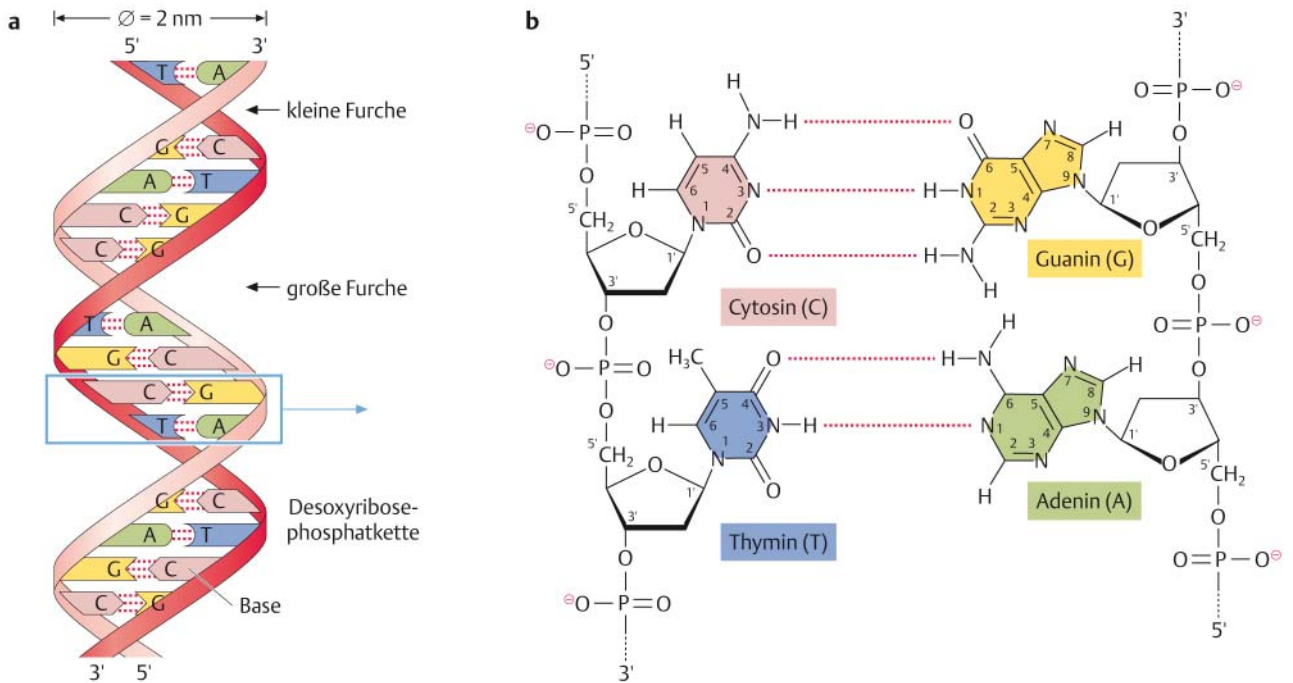


Abb. 13.1 DNA-Struktur.

a Schematische Anordnung der α -Doppelhelix mit zwei antiparallelen Strängen. Die Basenpaare liegen im Inneren und sind nach außen durch die Desoxyribosephosphatkette abgeschirmt. Spezifische Kontakte zu den Basen sind für DNA-bindende Proteine möglich, wenn sie sich in die große oder kleine Furche der DNA einfügen (Plus 15.7 S. 569).

b Detailansicht von zwei benachbarten Basenpaaren (s. Umrandung in **a**) mit drei Wasserstoffbrücken für das G=C-Paar und zwei für das A=T-Paar. Die Phosphatreste tragen eine negative Ladung und verbinden benachbarte Basen über Phosphodiesterbindungen zwischen der 3'-Hydroxylgruppe des letzten Nucleotids und der 5'-Hydroxylgruppe des folgenden. Entsprechend der internationalen Nomenklatur erfolgt die Numerierung der Atome in den aromatischen Ringen von 1–6 bzw. 1–9, während die C-Atome in den Zuckerresten von 1'–5' bezeichnet werden (verändert nach Buchanan et al. 2000).

gelesen wird, während die Neusynthese antiparallel dazu stets vom 5'- zum 3'-Ende erfolgt (Abb. 13.6 S. 388).

- Die Anordnung der komplementären Basenpaare im Inneren der DNA-Doppelhelix führt zu einer beträchtlichen Stabilisierung durch **Stapelung der aromatischen Ringe** (engl.: base stacking), wie etwa bei den Münzen in einer Geldrolle. Der Effekt beruht auf der Wechselwirkung zwischen den freien sog. π -Elektronen der aromatischen Ringe. Wie in Abb. 13.1a zu sehen, sind die beiden Stränge der DNA so angeordnet, daß sie eine **große** und eine **kleine Furche** bilden, in denen DNA-bindende Proteine basenspezifische Kontakte ausbilden können. Die meisten Proteine, die als Regulatoren der Transkription eine Rolle spielen (Plus 15.7), binden in der großen Furche.
- Schließlich birgt die DNA-Doppelhelix in sich zwei scheinbar widersprüchliche Eigenschaften: auf der einen Seite die Fähigkeit zur **unveränderten Weitergabe** der genetischen Information von einer Zelle an die andere bzw. von einer Generation auf die nächste und auf der anderen Seite die Möglichkeit zu **Veränderungen durch Mutationen** (Kap. 13.8). Beides, Stabilität der Information und ihre Veränderlichkeit im richtigen Verhältnis zueinander, sind die Voraussetzungen für Evolution (Kap. 13.10).

13.2 Der genetische Code

Der genetische Code beschreibt die Regeln für die Übersetzung der Nucleinsäureinformation mit 4 Buchstaben (Basen) in die Proteininformation mit 20 Buchstaben (Aminosäuren). Der genetische Code ist ein degenerierter und in allen lebenden Organismen universell gültiger Tripletcode.

Bevor wir uns mit dem genetischen Code im Detail beschäftigen können, müssen wir einen kleinen Vorgriff auf das Kap. 15 (Genexpression) machen. Nucleinsäuren sind zwar die universellen Informationsspeicher, aber Zellstruktur und -funktion und damit das Leben selbst in seiner heutigen Form ist an **Genexpression** und die Wirkung von Proteinen als Genprodukte gebunden. Bei der Umsetzung der genetischen Information haben wir es stark vereinfacht mit folgendem **Informationsverarbeitungsprozeß** zu tun (Kap. 15):



Als erster Schritt wird also im Zuge der **Transkription** eine Abschrift des DNA-Teilabschnitts (Gen) angefertigt. Diese **mRNA-Kopie** (Boten- oder Messenger-RNA) dient dann als Matrize für die Proteinbiosynthesemaschine bei der Translation. Die Informationsspeicherung auf der Nucleinsäureebene kommt mit vier Buchstaben aus (A, G, T, C für die DNA und A, G, U, C für die RNA). Dagegen werden auf der Ebene der **Proteine** wegen der besonderen Anforderungen an die Flexibilität von Struktur und Funktion 20 sehr verschiedene Buchstaben (**Aminosäuren**) benötigt (Tab. 13.1). Bei der Umsetzung der genetischen Information in die Proteininformation muß es also Regeln für den spezifischen Sprachübergang geben (**Translation**). Dies führt zu der Frage, wie viele und welche Nucleotide den Einbau einer bestimmten Aminosäure bei der Proteinbiosynthese bestimmen, d. h. zu der Frage nach dem **genetischen Code**. Eine Zusammenstellung dieser wichtigen Gesetzmäßigkeiten für die Translation findet sich in Tab. 13.1, die zugleich auch die gebräuchlichen Abkürzungen für die 20 Aminosäuren im 3-Buchstaben- bzw. 1-Buchstaben-Code enthält, z. B. Glutamin=Gln=Q.

Der genetische Code ist:

- ein **Tripletcode**, d. h. jeweils drei Nucleotide bilden ein Codon. Für die eindeutige Festlegung der 20 Aminosäuren braucht man mindestens drei Nucleotide als Codewort, da sich bei der Kombination von zwei Nucleotiden nur $4^2 = 16$, bei drei Nucleotiden aber $4^3 = 64$ mögliche Codons ergeben.
- **universell**, d. h. bei allen Pro- und Eukaryoten sowie bei den Viren gelten die gleichen Codon-/Aminosäure-Beziehungen. Lediglich bei Mycoplasmen, Mitochondrien und einigen Ciliaten wurden bisher einzelne Codons gefunden, die von dem universellen Code abweichen.
- **degeneriert**. Die meisten Aminosäuren werden durch mehr als ein Codon determiniert. Insgesamt sind 61 Codons mit dem Einbau einer entsprechenden Aminosäure verbunden („**sense**“-Codons). Nur drei Triplets, die sog. „**nonsense**“- oder „**stop**“-Codons **UAG**, **UGA** und **UAA**, codieren nicht für den Einbau einer Aminosäure. Sie haben aber essentielle Funktionen für die Beendigung der Polypeptidsynthese (Abb. 15.27 S. 540).
- **nicht überlappend und kommafrei**, d. h. er wird von dem 5'-Ende beginnend, kontinuierlich und lückenlos abgelesen; Einfügung oder Fortfall eines einzigen Nucleotids verändert die gesamte Information des nachfolgenden DNA-Abschnitts durch Verschiebung des Leserasters (engl.: frameshift, Abb. 13.19 S. 410).

Tab. 13.1 Zusammenstellung der wichtigsten Gesetzmäßigkeiten für die Translation (s. auch die Tab. in der hinteren Umschlagklappe). Aminosäuren mit ähnlichen Eigenschaften sind durch den Farbcode gekennzeichnet.

Aminosäuren			
Abkürzungen	Name	Eigenschaften der Seitenkette	in der mRNA
A Ala	Alanin	hydrophob, klein	GCA, GCC, GCG, GCU
C Cys	Cystein	hydrophil, SH-Gruppe	UGC, UGU
D Asp	Asparaginsäure	hydrophil, COOH-Gruppe, sauer	GAC, GAU
E Glu	Glutaminsäure	hydrophil, COOH-Gruppe, sauer	GAA, GAG
F Phe	Phenylalanin	hydrophob, aromatisch	UUC, UUU
G Gly	Glycin	kleinste Aminosäure, keine Seitenkette	GGA, GGC, GGG, GGU
H His	Histidin	Imidazolring, basisch	CAC, CAU
I Ile	Isoleucin	hydrophob, groß	AUA, AUC, AUU
K Lys	Lysin	hydrophil, ϵ -NH ₂ -Gruppe, basisch	AAA, AAG
L Leu	Leucin	hydrophob, groß	CUA, CUC, CUG, CUU, UUA, UUG
M Met	Methionin	hydrophob, groß, Initiatoraminosäure	AUG
N Asn	Asparagin	hydrophil, neutral	AAC, AAU
P Pro	Prolin	Iminosäure	CCA, CCC, CCG, CCU
Q Gln	Glutamin	hydrophil, neutral	CAA, CAG
R Arg	Arginin	hydrophil, Guanidino-Gruppe, basisch	AGA, AGG, CGA, CGC, CGG, CGU
S Ser	Serin	hydrophil, OH-Gruppe	AGC, AGU, UCA, UCC, UCG, UCU
T Thr	Threonin	hydrophil, OH-Gruppe	ACA, ACC, ACG, ACU
V Val	Valin	hydrophob, groß	GUA, GUC, GUG, GUU
W Trp	Tryptophan	hydrophob, aromatisch	UGG
Y Tyr	Tyrosin	hydrophob, aromatisch	UAC, UAU
	Stop		UAA, UGA, UAG

13.3 Verpackung von DNA in Chromatin und Chromosomen

Im Kern eukaryotischer Zellen liegt die DNA als Chromatin im Komplex mit den Histonen H1, H2A, H2B, H3 und H4 vor. Die kleinste Verpackungseinheit ist das Nucleosom mit einem zentralen Histonoktamer aus je zwei Molekülen H2A, H2B, H3 und H4, um das 146 bp DNA herumgewickelt sind. Ein abgestimmtes System histonmodifizierender Enzyme hat wesentlichen Anteil an Veränderungen in den Funktionszuständen des Chromatins. Prinzipiell sind aktive Chromatindomänen weniger dicht verpackt als inaktive. Die größte Verpackungsdichte wird transient in den Chromosomen als Transportform für das genetische Material während der Zellteilung erreicht.

Struktur und Funktion der DNA als universellem Speicher genetischer Information machen es notwendig, diese mit Proteinen zu dem sog. **Chromatin** zu verpacken und darüber hinaus spezielle Transportformen (**Chromosomen**) für die geordnete Weitergabe der Gesamtinformation an die Tochterzellen zu entwickeln (Box 13.2). Zwei weitere Gründe für die Ent-

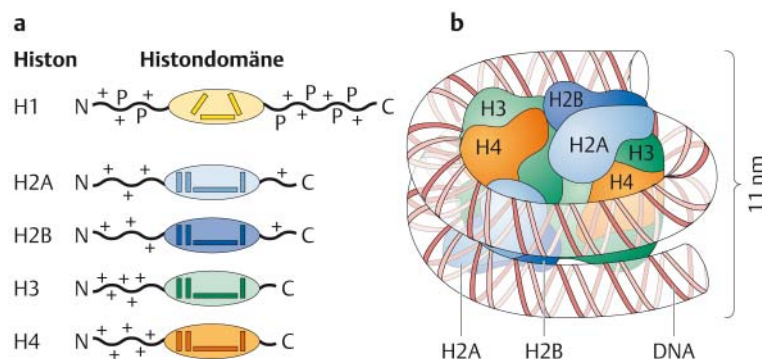
stehung dieses aufwendigen Prozesses im Verlauf der Evolution sind: 1. die Notwendigkeit, eine sehr große Menge genomischer DNA – die in einigen Zellen mehrere cm lang sein würde – in dem vergleichsweise kleinen Volumen eines Zellkerns von 5–25 μm unterzubringen und 2. die Organisation des selektiven Ablesens ausgewählter Teile des Genoms (differenzielle Genexpression) im Verlauf des Lebenszyklus einer Pflanze.

13.3.1 Histone als Verpackungsmaterial

Entsprechend den physikochemischen Eigenschaften der DNA als einem riesigen **Polyanion** mit zahlreichen negativen Ladungen in dem Rückgrat aus Desoxyribosephosphat (Abb. 13.1), sind die Proteine für die Verpackung der DNA stark positiv geladen. Die Rolle übernehmen die **fünf Histone** H1, H2A, H2B, H3 und H4 (Abb. 13.2a). Die **positiven Ladungen** der Histone beruhen auf dem hohen Gehalt an basischen Aminosäureresten, wie Lysin und Arginin, die sich vor allem in den N-terminalen und C-terminalen Domänen finden (Abb. 13.2). Auch bei Bakterien ist die DNA im Nucleoid mit histonartigen Proteinen verbunden (Abb. 15.3 und S. 734). Insbesondere bei den Archaeobakterien finden sich Proteine, die in Struktur und Funktion als unmittelbare Vorläufer der eukaryotischen Histone betrachtet werden. Da der Aufbau der bakteriellen DNA-/Proteinkomplexe sich allerdings von dem eukaryotischer Zellen wesentlich unterscheidet, spricht man nicht von Chromatin.

Grundbaustein des Chromatins in den 10-nm-Nucleofilamenten sind die sog. **Nucleosomen** mit einem **Histonoktamer** als Zentralkörper (jeweils zwei Moleküle der Histone H2A, H2B, H3 und H4) und **146 Basenpaaren** der DNA-Doppelhelix, die in $1\frac{3}{4}$ Windungen um das Histonoktamer herumgewunden ist (Abb. 13.2b). Im Zentrum des Nucleosoms sitzt ein H3/H4-Tetramer, während die beiden H2A/H2B-Dimeren daran angelagert sind. Die Histone treten mit ihren positiv geladenen und in ihrer Struktur flexiblen N-terminalen Domänen in Wechselwirkung mit den negativen Ladungen an den Phosphatgruppen der DNA. Die Nucleosomen sind miteinander durch DNA-Abschnitte von etwa 50 Basenpaaren, den sog. **Linkerregionen** verbunden, sodaß die Nucleofilamente perlchnurartig aufgereiht erscheinen (Abb. 13.3a, b).

Das besonders lysinreiche **Histon H1** kommt nicht im Nucleosom vor, hat aber eine spezielle Funktion bei der weiteren Kondensierung des Chromatins. H1 bindet an die Linkerregionen der DNA und ermöglicht damit die Verkürzung der Abstände zwischen den Nucleosomen und die Ausbildung einer 30 nm dicken **Chromatinfibrille**. Diese, als **Solenoid** bezeichnete Einheit, besteht aus jeweils sechs durch H1 verbundene Nucleo-



Box 13.2 Euchromatin/ Heterochromatin

Die Hauptmasse des Chromatins besteht aus lockerem Euchromatin, das bei Eintritt in die Mitose kondensiert und in der Interphase dekontensiert wird (Kap. 13.7). Es enthält die weitaus überwiegende Menge der im Zellkern liegenden genetischen Information. Wenn man Zellkerne mit Hämatoxylin färbt, dann ist das dichter gepackte Heterochromatin deutlich als sog. Chromozentren von dem lockeren Euchromatin zu unterscheiden. Chromozentren, die hochrepetitive DNA-Sequenzen (Satelliten-DNA) enthalten und genetisch inaktiv sind, bilden sich bei allen Zellen einer Art stets an den gleichen Stellen der Chromosomen, z. B. den Centromeren (Abb. 13.4), unabhängig vom Entwicklungszustand und Gewebetypus.

Abb. 13.2 Histone und Grundstruktur eines Nucleosoms.

a Die vier Histone im Nucleosom (H2A, H2B, H3 und H4) haben eine verwandte Grundstruktur mit einer zentralen Histonodomäne, die eine charakteristische Anordnung von vier Helices (farbige Stäbchen) aufweist, und einer N-terminalen, stark positiv geladenen (+) aber unstrukturierten Domäne. Das Histon H1 ist ebenfalls stark positiv geladen, hat aber eine anders aufgebaute Zentraldomäne, und die ausgedehnten N-terminalen und C-terminalen Enden sind häufig phosphoryliert (P).

b Struktur eines Nucleosoms mit einem Histonoktamer und 146 Basenpaaren der DNA in $1\frac{3}{4}$ Windungen (nach Rindt und Nover 1982).

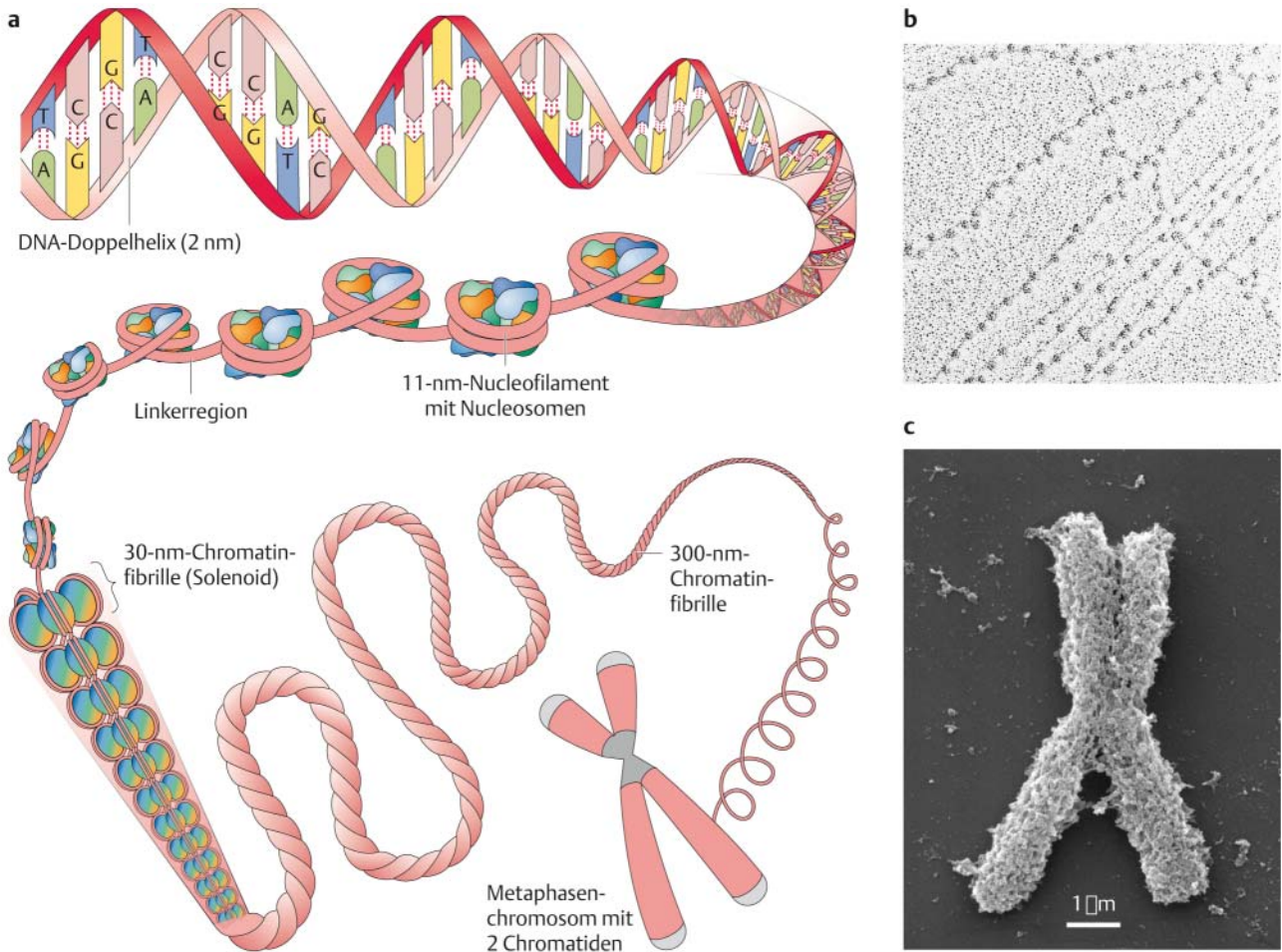


Abb. 13.3 Chromatinverpackung.

a Stufen der Chromatinverpackung von der DNA-Doppelhelix bis zum Chromosom (nach Rindt und Nover). **b** Elektronenoptische Aufnahme der Perlenkette eines 11-nm-Nucleofilamentes (präpariert mit der DNA-Spreitungstechnik nach O. Miller, Originalaufnahme W. Nagel, Chromatin von *Allium cepa* [Küchenzwiebel]; Vergr. 120 000-fach). **c** Metaphasechromosom aus der Wurzelspitze der Gerste (rasterelektronenmikroskopische Originalaufnahme G. Wanner).

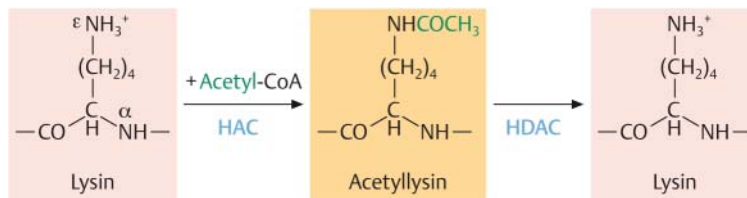
somen pro Windung (Abb. 13.3a). Die weiteren Kondensierungsschritte, die insbesondere in der Prophase von Meiose und Mitose für die Herausbildung der Chromosomen als Transportform eine Rolle spielen (Kap. 13.7), sind in Abb. 13.3a schematisch dargestellt. Da die Gesamtheit der DNA-Doppelhelix eines eukaryotischen Chromosoms eine Länge von vielen Zentimetern haben kann, muß das Chromatin bei der Bildung der nur wenige Mikrometer großen Chromosomen bis 10 000-fach kondensiert werden (Abb. 13.3a, c). Dafür bedarf es allerdings weiterer, als **Condensin** und **Cohesin** bezeichneter Proteinkomplexe (Kap. 13.7.3).

Interessanterweise bleibt die Struktur der Chromosomen selbst auch dann mikroskopisch erkennbar, wenn man alle DNA und Histone entfernt. Sie wird durch eine Art **Chromosomenskelett** aus unlöslichen Proteinen gebildet, an die die Chromatinschleifen (**Chromatindomänen**) durch spezielle Proteine angeheftet werden. Die Chromosomen sind sehr vielgestaltig (Abb. 13.4 und Abb. 13.7 S. 391). Ihre Länge kann zwischen 0,2 und 50 µm und ihre Breite zwischen 0,2 und 2 µm schwanken. In hochkondensierter Form färben sich Chromosomen nach Fixierung intensiv mit Farbstoffen wie Safranin oder Hämatoxylin an. Dieser Eigenschaft verdanken sie ihren Namen (1888, W. Waldeyer). Da die Chromosomen nur in der Transportform als selbständige Elemente in Erscheinung treten, wird

auch ihre Individualität erst zu diesem Zeitpunkt erkennbar. Jedes Chromosom hat eine definierte Größe und eine charakteristische, artspezifische Gestalt, die seine Identifizierung ermöglicht. Die Gesamtheit der Chromosomen wird als **Karyogramm** eines Organismus bezeichnet (Abb. 13.7, S. 391).

13.3.2 Histon-Modifikationen

Wie wir in Kap. 15.7.4 erfahren werden, sind Struktur und Funktion von Proteinen nicht nur durch ihre Aminosäuresequenz und ihre Raumstruktur festgelegt. Proteine können vielmehr durch enzymatische Modifikationen der Aminosäureseitenketten nachhaltig in ihren Eigenschaften verändert werden (Tab. 15.6 S. 543). Ein besonders eindrucksvolles Beispiel finden wir bei den Histonen, die mit ihren flexiblen N-terminalen Domänen entscheidend für die Kontakte zur DNA im Verband der Nucleosomen sind. Das regulatorische Geflecht von Histon-Modifikationen und histonbindenden Proteinen ist wegen seiner Bedeutung für die Chromatinstruktur und -funktion und wegen seiner stabilen Weitergabe an die Tochterzellen bei der Zellteilung auch als **Histoncode** bezeichnet worden (Plus 13.1). Der Modifikationszustand der Histone gibt Auskunft über den Aktivitätszustand des jeweiligen Gens bzw. der Chromatindomäne. Ein charakteristisches Beispiel ist die transiente Aktivierung von Teilen des Chromatins bei der Transkription. Dabei werden die für die Bindung an die negativ geladene DNA wichtigen ϵ -Aminogruppen in den Seitenketten der Lysinreste durch Histon-Acetyltransferasen (HAC) acetyliert. Sie verlieren damit ihre positive Ladung, sodaß die Bindung an die DNA aufgelockert wird. Umgekehrt sind die Lysinreste in inaktivem Chromatin nicht acetyliert. Für die Deacetylierung sind Histon-Deacetylasen (HDAC) zuständig:



13.4 Die drei Genome der Pflanzenzellen

Die Gesamtheit der genetischen Information einer Pflanzenzelle liegt in drei Genomen. Für die Modellpflanze *Arabidopsis thaliana* sind dies das Kerngenom mit 125 Mb in 5 Chromosomen und je 154 kb in den Plastomen bzw. 367 kb in den Chondriomen. *Arabidopsis* hat das kleinste pflanzliche Kerngenom. Bei Weizen und der Kaiserkrone ist das Kerngenom mit 10 Gb bzw. 100 Gb deutlich größer. Die Genomgröße wird wesentlich durch die Menge an nichtcodierender DNA bestimmt (Größe der intergenischen Bereiche, Transposons und andere repetitive Sequenzen, Größe der Introns).

Der bei weitem größte Teil der genetischen Information einer Pflanzenzelle ist im Kern codiert (**Kerngenom**). Daneben gibt es noch zwei sehr viel kleinere Genome in den Chloroplasten und den Mitochondrien. Die Größe eines Genoms wird im Allgemeinen durch die Anzahl der Basenpaare in der DNA angegeben. Zur besseren Darstellung benutzt man die

Plus 13.1 Der Histoncode als epigenetische Information

In den N-terminalen Domänen der Histone können die Serin-, Lysin- und Argininreste durch Acetylierung, Methylierung, Ubiquitinierung bzw. Phosphorylierung in ihren Eigenschaften reversibel verändert werden. Eine beachtliche Zahl von histonmodifizierenden und -demodifizierenden Enzymen sind beteiligt. Dies sind insbesondere **Histon-Acetyltransferasen (HAC)** und **Histon-Deacetylasen (HDAC)**, **Histon-Methyltransferasen (HMT)** und **-Demethylasen (HDM)**, **Ubiquitin-Konjugasen** und **-Isopeptidasen**, **Histon-Kinasen** und **-Phosphatasen**. Histon-Modifikationen bestimmen nicht nur die Kontakte zur DNA, sondern darüber hinaus die Rekrutierung von Proteinen, die Einfluß auf die Aktivität des Chromatins haben (Box 13.2 S. 379). Veränderungen im Modifikationsmuster können einander bedingen, sich verstärken oder wechselseitig ausschließen. Dies sei am Beispiel des Histon H3 verdeutlicht, dem in allen eukaryotischen Organismen eine besondere Rolle in diesem Zusammenhang zukommt. Wir können zwei Grundprozesse unterscheiden:

- **Aktivierung von Chromatin:** Phosphorylierung von Ser10 stimuliert die Acetylierung von Lys14; zusammen mit Acetyl-Lys9 steht dieses Modifikationsmuster für aktives Chromatin mit hoher Affinität für die Assemblierung von Transkriptionskomplexen. Auch die Methylierung von Lys4 durch die Methyltransferase SET9 ist charakteristisch für aktives Chromatin. Methylierung von Lys4 hemmt ihrerseits die Bindung der Lys9-HMT und damit die Desaktivierung des Chromatins durch Rekrutierung von Histon-Deacetylasen.
- **Inaktivierung von Chromatin und Bildung von Heterochromatin:** Methylierung von Lys9 durch die Methyltransferase SUV39 ist der Gegenspieler der unter 1. zusammengefaßten Prozesse. Sie verhindert die Lys9-Acetylierung

und fördert die Bindung des Proteins HP1, das eine fortschreitende Inaktivierung der Chromatindomäne (Heterochromatisierung, Box 13.2) bewirkt.

Zusammenfassend kann man sagen, daß das Muster der Histon-Modifikationen eine Art Code darstellt, der über die Wechselwirkung mit anderen Proteinen den Aktivitätszustand einer Chromatindomäne bestimmt. Dieser Histoncode kann stabil von einer Zelle zur anderen weitergegeben werden und ist häufig mit sequenzspezifischen Modifikationen der DNA durch **Methylierung** von Cytosinresten verbunden. Diese sind ein weiterer Marker für inaktives Chromatin. Wir haben es also mit einem epigenetischen Langzeitgedächtnis der Zellen zu tun, das der Aufrechterhaltung stabiler Genexpressionsmuster dient und damit der Zellidentität innerhalb einer Gruppe spezialisierter Zellen in einem Gewebe oder Organ. Man benutzt den Terminus **Epigenetik** in diesem Zusammenhang, weil es sich um ein stabiles, somatisch vererbbares Muster handelt, das keine Veränderungen im Informationsgehalt selbst einschließt, also nicht durch die Keimzellen an die nächste Generation weitergegeben werden kann (Plus 18.6 S. 734).

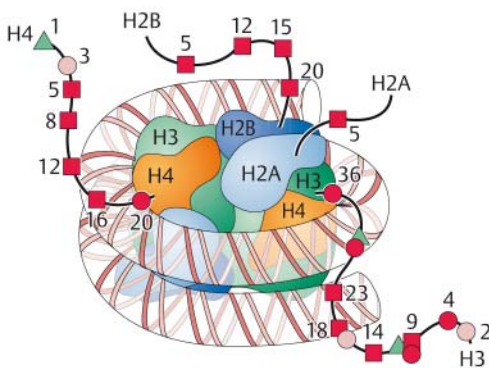
Aktives Chromatin mit Modifikationsmuster:

K4-Me, K9-Ac, S10-P, K14-Ac

- Rekrutierung von Transkriptionsfaktoren und Chromatinaktivierungskomplexen
- Hemmung der Bindung von K4-HMT, HP1 und HDAC

Inaktives Chromatin mit Modifikationsmuster: K9-Me

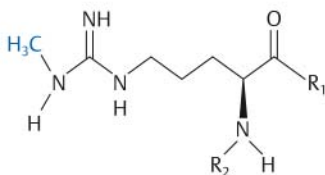
- Rekrutierung des Heterochromatin-spezifischen Proteins HP1
- Hemmung der Acetylierung von K9 und der Phosphorylierung von S10



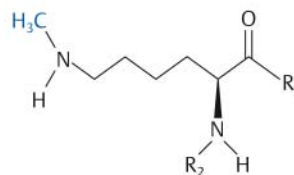
mögliche Modifikationen am N-Terminus von Histon H3



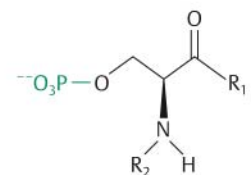
- K-AC: N-Acetyllysin
- R-Me: N-Methylarginin
- K-Me: N-Methyllysin
- ▲ S-P : O-Phosphoserin



R-Me = N-Methylarginin



K-Me = N-Methyllysin

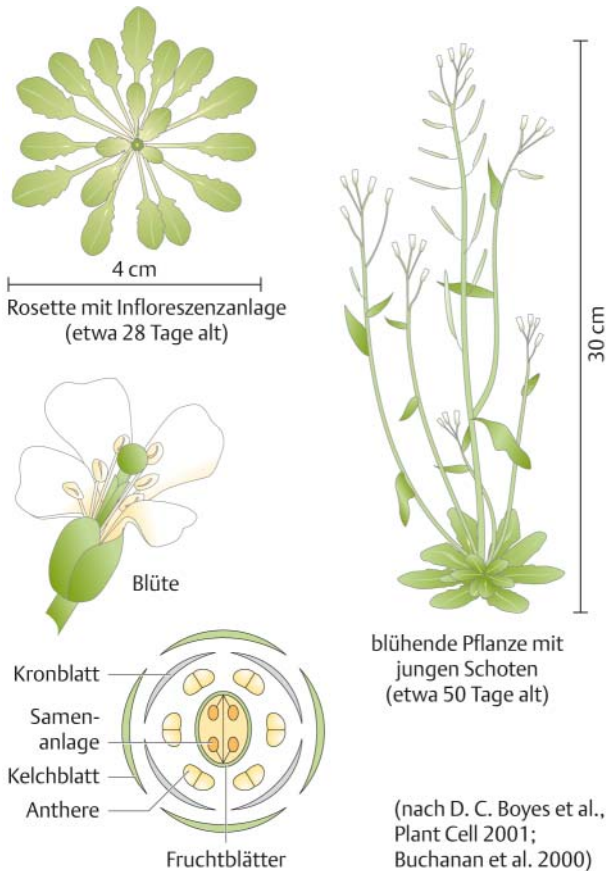


S-P = O-Phosphoserin

R₁, R₂: Peptidkette (Modell nach Turner 2002)

Plus 13.2 Das Genom der Modellpflanze *Arabidopsis thaliana*

Es war ein langer Weg, den das kleine, von den Biologen kaum beachtete Wildkraut *Arabidopsis thaliana* (Ackerschmalwand) von den Anfängen als experimentelles Objekt (F. Laibach 1907) bis in die Spitzenposition einer weltweit operierenden Forscher-Community zurückgelegt hat. Bis zum Jahr 2010 wollen Forscher alle molekularbiologischen Prozesse in dieser Modellpflanze soweit aufklären, daß die Masse der gewonnenen Daten im Computer zur Konstruktion einer virtuellen Pflanze als Muster für andere Pflanzen verwendet werden kann. Wenn dies gelingt, könnte die Computersimulation dieser Pflanze dramatische Auswirkungen auf alle Aspekte der Pflanzenforschung und darüber hinaus auf die Biologie im Allgemeinen haben (Plus 13.3).



Arabidopsis thaliana ist eine Langtagspflanze mit einem Entwicklungszyklus von etwa 60 Tagen (Tab. 18.3 S. 733). Wie bei allen Brassicaceen sind die Blüten 4-zählig und mit 2 langen und 4 kurzen Staubblättern versehen. Je nach Größe enthalten die Schoten bis zu 50 Samen. Alle Eigenschaften zusammen ergeben günstige Voraussetzungen für eine experimentelle Bearbeitung von *Arabidopsis* als Modellpflanze.

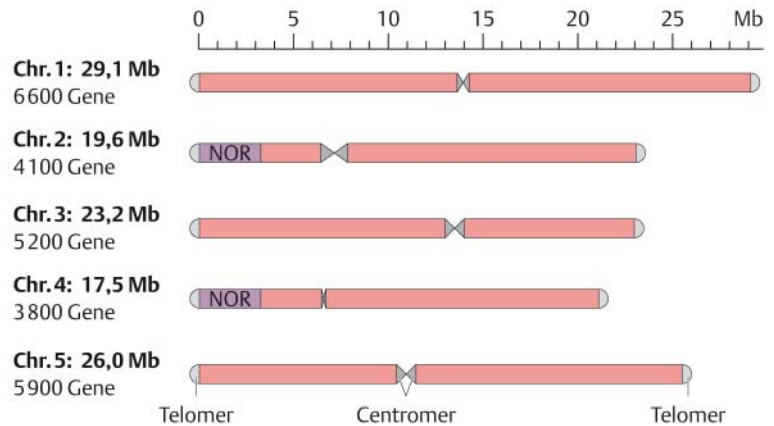
Das Jahr 2000 markiert als eine wichtige Etappe die Publikation der Gesamtsequenz des Genoms von *Arabidopsis* als erstem Pflanzengenom. Die Gesamtinformation des Kerns ist auf **5 Chromosomen** mit **125 Mb DNA** untergebracht (Abb. 13.4). Für die sinnvolle Bearbeitung der Datenfülle, die aus immer mehr Organismen gewonnen werden, benötigt man moderne Methoden der Bioinformatik (Plus 13.3 S. 386). Danach codiert das Genom von *Arabidopsis* für eine Gesamtheit von mehr als 25 000 Proteinen (**Proteom**). Für etwa zwei Drittel der offenen Leseraster (ORF, engl.: open reading frame), d. h. der proteincodierenden Sequenzen, konnte aus dem Vergleich mit Sequenzen anderer Pflanzen sowie der Bäckerhefe *Saccharomyces cerevisiae*, der Fruchtfliege *Drosophila melanogaster* und des Menschen eine Vorstellung gewonnen werden, welche Funktion diese hypothetischen Proteine haben könnten. Für etwa 7000 ORFs gelingt das aber bisher nicht. Insgesamt werden für das *Arabidopsis*-Genom 133 000 Exons mit einer durchschnittlichen Größe von 214 bp und 107 000 Introns mit einer durchschnittlichen Größe von 164 bp vorhergesagt. Das macht etwa 35 Mb an Exon- und 18 Mb an Intronsequenzen. Die durchschnittliche Gengröße wird mit 2,4 kb angegeben. Wenn man etwa 1,5 kb für die flankierenden regulatorischen Bereiche zwischen den Genen annimmt, dann sind insgesamt etwa **100 Mb von proteincodierenden Genen** eingenommen (4 kb · 25 000). Natürlich dienen solche Angaben über Durchschnittsgrößen nur einer ersten Orientierung über die Gendichte im Vergleich zu anderen Pflanzen. Die Details für ein bestimmtes Gen können davon mehr oder weniger stark abweichen.

Die **restlichen 20%** des *Arabidopsis*-Genoms entfallen auf verschiedene Formen sich mehrfach wiederholender Sequenzen (**repetitive Sequenzen**).

- Dazu gehören etwa 9 Mb mit **repetitiven Genen** für Struktur-RNAs mit besonderer Funktion in der Genexpression, z. B. 800 Gene für ribosomale RNA, 590 Gene für tRNAs und etwa 100 Gene für snRNAs und snoRNAs. Details von Struktur und Funktion dieser RNAs werden wir in Kap. 15.4 und Kap. 15.6 kennenlernen.
- Die **Centromerregionen** bestehen primär aus ausgedehnten Regionen mit hochrepetitiven Grundeinheiten von 180 bp (Satelliten-DNA), während die Telomere aus einer Vielzahl sehr kurzer Einheiten mit der Sequenz 5'-CCCTAAA-3' aufgebaut sind.
- Etwa 10% des Genoms werden von etwa 5000 **Transposons und transposonartigen Elementen** (Kap. 13.9.2) eingenommen, die gehäuft in den heterochromatischen Bereichen der Centromere oder in deren Nachbarschaft auftreten. Diese transponierbaren Elemente gehören drei Klassen an: Etwa 200 relativ große Retrotransposons (Klasse I), DNA-Transposons der Klasse II und etwa 1200 kurze transposonartige Elemente der Klasse III (sog. MITEs, engl.: miniature inverted repeat transposable elements).

Das *Arabidopsis*-Genom ist ungewöhnlich dicht mit Informationen gepackt. Man kann aus den Ergebnissen neuerer Untersuchungen mit „Mikroarrays“ ableiten (Box 15.2 S. 492), daß tatsächlich mehr als 90% aller Gene irgendwann im Leben einer *Arabidopsis*-Pflanze abgelesen werden.

Abb. 13.4 Die Chromosomen von *Arabidopsis thaliana*. Kapfenförmige Strukturen an den Enden = Telomere, zentrale Einschnürungen = Centromere, NOR = Nucleolus organisierende Region (engl.: nucleolus organisator region) mit den Batterien für ribosomale RNA(*rRNA*)-Gene. Chromosomen mit etwa gleich großen Armen, d. h. mit dem Centromer in der Mitte, bezeichnet man als **metazentrisch** (Chr. 1), während man solche mit ungleich großen Armen als **acrozentrisch** bezeichnet (Chr. 2–5). (Nach <http://www.arabidopsis.org/servlets/mapper>)



dezimalen Maßeinheiten. 10^3 Basenpaare sind 1 Kb, 10^6 Basenpaare sind 1 Mb und 10^9 Basenpaare sind 1 Gb. Für die bisher am besten untersuchte Pflanze, *Arabidopsis thaliana* (Plus 13.2), gelten folgende Größenverhältnisse für die drei Genome: Das **Kerngenom** umfaßt 125 Mb, die in 5 Chromosomen untergebracht sind (Abb. 13.4). Dagegen enthalten das **Chloroplastengenom** mit etwa 560 Kopien pro Zelle nur 154 Kb und das **Mitochondriengenom** mit etwa 25 Kopien pro Zelle 367 Kb. Das in den Mitochondrien lokalisierte Genom wird auch als **Chondriom**, das in den Plastiden als **Plastom** bezeichnet.

Die Zahl der Gene im Kern, die für Proteine codieren, wird auf 25 000 geschätzt, während in Chloroplasten und Mitochondrien nur 79 bzw. 58 proteincodierende Gene gefunden werden. Alle zusammen bilden sie die Gesamtheit der möglichen Proteine in *Arabidopsis*, das sog. **Proteom**. Dazu kommen natürlich jeweils eine beachtliche Zahl von Genen für die **Struktur-RNAs**, die eine unverzichtbare Rolle bei der Genexpression spielen. Dies sind die Gene für rRNAs, tRNAs, miRNAs, snoRNAs und snRNAs (s. Kap. 15.4 und 15.6 sowie Plus 18.3 S. 722).

Die geringe Zahl der proteincodierenden Gene in den Organellen zeigt an, daß diese trotz des eigenen Genoms in ihrer Struktur und Funktion wesentlich von Proteinen abhängen, die im Kern codiert, im Cytoplasma synthetisiert und in die Organellen importiert werden. Man bezeichnet die Chloroplasten und Mitochondrien daher als **semiautonome Organellen**. Auf der anderen Seite kann eine Pflanzenzelle ohne diese beiden Organellen nicht existieren, da im Verlauf der Evolution eine Reihe wichtiger Biosyntheseprozesse anteilig zwischen Cytoplasma und Organellen organisiert wurden. Das gilt z. B. für die Sulfatreduktion, Nitritreduktion (Abb. 17.22 S. 699) oder die Synthese von aromatischen Aminosäuren, Fettsäuren, Phytohormonen (Abb. 16.15 S. 613, Abb. 16.30 S. 637 und Abb. 16.36 S. 643) und von Tetrapyrrolen (Abb. 15.49 S. 586). Die Vorstellung über die Entstehung dieser interessanten Koexistenz von Kern- und Organellengenom in eukaryotischen Zellen ist in der **Endosymbiontenhypothese** zusammengefaßt (Plus 4.1 S. 130).

Die **Genomgrößen** können sehr unterschiedlich sein, wobei insbesondere im Reich der Pflanzen eine Diskrepanz zwischen dem Gehalt an Genen und der Genomgröße besteht (C-Wert-Paradoxon, Box 13.3). Das haploide Karyogramm von *Arabidopsis* (Abb. 13.4) umfaßt 5 Chromosomen mit einer Größe von 17,5 Mb (Chromosom 4) bis 29,1 Mb (Chromosom 1). Jedes der Chromosomen enthält eine **Centromerregion**, die das Chromosom in einen linken und einen rechten Arm unterschiedlicher

Größe teilt. Diese Region enthält hochrepetitive DNA-Sequenzen von 150–200 bp (Satelliten-DNA). Sie ist arm an codierender Information. Die Modifikation der Nucleosomen durch eine centromerspezifische Variante des Histon H3 (CenH3) ist Ausgangspunkt für die Umwandlung der ganzen Region in Heterochromatin (Box 13.2). Die Centromeren haben aber entscheidenden Anteil an der geordneten Verteilung der Chromosomen auf die Tochterzellen während der Mitose und Meiose, weil sich an ihnen der Proteinkomplex für die Anknüpfung des Spindelapparates (**Kinetochor**) bildet (vgl. Kap. 13.7.3).

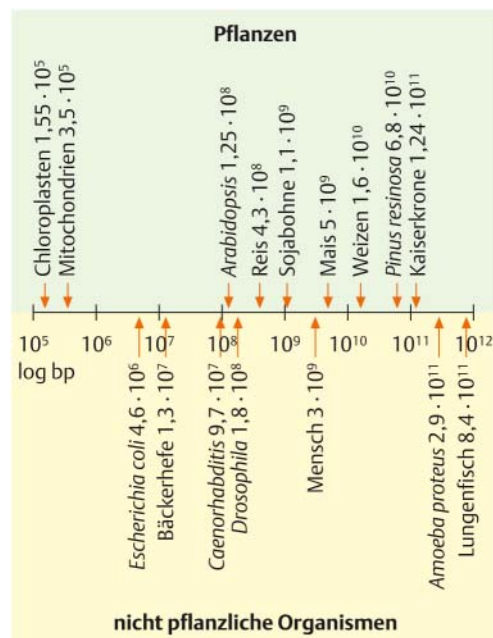
An den Enden der Chromosomen befinden sich die **Telomeren** mit kurzen hochrepetitiven Sequenzen (5'-CCCTAAA-3'). Sie sind für die Stabilität und die vollständige Replikation der Gesamtinformation unerlässlich. Sie werden bei der Replikation durch die Telomerase unter Verwendung einer RNA-Matrize neu hergestellt. Die Telomerase stellt also eine spezielle Form von reverser Transkriptase dar (Abb. 13.28 S. 428). Das gesamte Genom von *Arabidopsis* ist übersät mit Transposons, die etwa 10% der DNA-Sequenz ausmachen und die höchste Dichte in den Centromerbereichen haben. Im linken Arm der Chromosomen 2 und 4 gibt es je eine **nucleolus organisator region** (NOR) von etwa 10 kb mit Batterien von jeweils etwa 400 rRNA-codierenden Genen (Abb. 15.22 S. 532).

Box 13.3 Genomgrößen, C-Wert-Paradoxon

In der Abbildung sind die Größen der haploiden Genome einiger pflanzlicher und nicht pflanzlicher Organismen zusammengetragen. *Arabidopsis* hat das bisher kleinste bekannte pflanzliche Genom. Demgegenüber stehen die Genomriesen wie Weizen mit 10 Gb und die Kaiserkrone (*Fritillaria imperialis*) mit 100 Gb. Die offensichtliche Diskrepanz zwischen Organisationsstufe bzw. dem Gehalt an Genen und der Genomgröße kann bis heute nicht befriedigend geklärt werden (**C-Wert-Paradoxon**). Das gilt auch für die Mitochondriengenome, die mit 24 kb bei der Grünalge *Chlamydomonas eugametos*, 200 kb bei Moosen, 570 kb beim Mais und 2400 kb bei der Melone um den Faktor 100 variieren können. Die Mitochondriengenome der Säugetiere haben dagegen nur 16 kb.

Für das C-Wert-Paradoxon gibt es eine Reihe von Gründen: Der erste hängt mit der Größe und Organisation der Gene selbst zusammen. Wir müssen uns in diesem Zusammenhang erst einmal mit der Definition eines Gens als Funktionseinheit im Genom beschäftigen. Viele Gene der Eukaryoten sind mosaikartig aufgebaut. Sie enthalten codierende Teile (Exons), die durch nichtcodierende Teile (Introns) voneinander getrennt sind. Zum Gen gehören aber auch die regulatorischen Sequenzen am 5'- und am 3'-Ende, die für die Effizienz und die Kontrolle von Transkription und Translation benötigt werden (Abb. 15.2 S. 492). Bei größeren Genomen nimmt die durchschnittliche Größe der Introns und die Menge an repetitiven Sequenzen (Transposons und MITEs, Plus 13.2) drastisch zu. Die Bedeutung dieser artspezifischen, über Jahrmillionen erhaltenen Struktur-DNA ist ungeklärt. Beispielsweise besteht das Maisgenom (2800 Mb) zu 50–80% aus ineinander verschachtelten Retrotransposons, die die intergenischen Bereiche stark ausdehnen. Man nimmt an, daß sich vor etwa

6 Millionen Jahren durch einen Ausbruch von Retrotransposonaktivitäten die DNA-Menge im Genom der Vorläuferformen des Mais verdoppelt hat. Seitdem sind diese Elemente inaktiv, und eine gewisse Reaktivierung kann nur unter Streßbedingungen (osmotischer Streß, Wassermangel, Hitzestreß) beobachtet werden.



Die Genome der Organellen sind wie bei den Bakterien meist **ringförmig**. In Abb. 13.5 ist exemplarisch das **Chloroplastengenom** des Tabaks (*Nicotiana tabacum*) gezeigt. Charakteristisch ist die Duplikation einer Genbatterie, die u. a. die Gene für die ribosomalen RNAs der Organellen enthalten. Chloroplasten wie Mitochondrien enthalten eigene Komponenten für den Genexpressionsapparat mit RNA-Polymerase, Ribosomen, tRNAs usw. Sie machen wesentliche Teile der Gene auf den Orga-

Plus 13.3 Internationale Datenbanken

Die Sequenzierung von Genomen und die damit eng verbundene Funktionsanalyse von Genen bei komplexen Organismen wie Pflanzen und Tieren hat eine bisher nicht gekannte Fülle von wertvollen Daten erzeugt. Dabei ist sehr schnell deutlich geworden, daß weder die Beschaffung noch die Handhabung der Daten durch einzelne Labors oder Forschernetzwerke in einzelnen Ländern zu gewährleisten waren. Es gab eine Reihe unabdingbarer Voraussetzungen für die neue Entwicklung:

- internationale Kooperation im großen Stil,
- entsprechende Hochleistungsrechner,
- neue Methoden der Bioinformatik,
- geeignete internationale Absprachen für den weltweiten und freien Zugang zu allen Daten,
- englisch als einheitliche Wissenschaftssprache,
- das Internet mit seinen einmaligen Möglichkeiten zur weltweiten Kommunikation.

Die im folgenden aufgeführten Datenbanken stehen schwerpunktmäßig für die genannten Stichworte. Sie sind aber im allgemeinen sehr viel komplexer nutzbar und durch zahlreiche Links mit anderen Datenbanken verbunden. Neben den Informationen werden in wichtigen Fällen auch Forschungsmaterialien wie Plasmide oder Samen von Mutantenlinien zur Verfügung gestellt. Dieses gesamte internationale System von Datenbanken und Ressourcen hat Finanzmittel für die weiterführenden Forschungsarbeiten freigesetzt und entscheidend zu dem beispiellosen Fortschritt in den biologischen Wissenschaften und in besonderer Weise den Pflanzenwissenschaften der letzten 15 Jahre beigetragen.

Genomsequenzierung: Alle gewonnenen Daten aus dem weltweit arbeitenden *Arabidopsis*-Konsortium sind in zwei Datenbanken zusammengeführt:

- The Arabidopsis Information Resources (TAIR): <http://www.arabidopsis.org/>
- The Munich Information Center for Protein Sequences (MIPS): <http://mips.gsf.de/>

Bei letzterem finden sich auch Informationen über die vollständigen bzw. fast vollständigen Genomsequenzen für andere Pflanzen (Medicago, Mais, Reis, Lotus, Pappel).

Expressionsdatenbanken: Als wichtige Hilfe für die Genfunktionsanalyse haben sich Datenbanken erwiesen, die aus mRNAs durch Synthese komplementärer DNAs (cDNAs) gewonnen wurden (Abb. 13.28 S. 428). Diese **EST-Datenbanken** (engl.: expressed sequence tags) umfassen für einzelne Pflanzen weit mehr als 100 000 Einträge und lassen Sequenzvergleiche zwischen verschiedenen Pflanzen zu, auch wenn das Genom noch nicht sequenziert wurde. Wichtigstes Hilfsmittel sind Suchverfahren (**Blast searches**), bei denen man

eine Zielsequenz (Nucleinsäure oder Protein) mit allen verfügbaren Teilsequenzen vergleichen kann, um homologe Sequenzen in anderen Organismen zu finden. Zu nennen sind:

- National Center of Biotechnology Information (NCBI): <http://www.ncbi.nlm.nih.gov/>
- Harvard University: <http://compbio.dfci.harvard.edu/tgi/plant.html>

Mikroarray-Expressionsdatenbanken: Die Kenntnisse über das vollständige Genom von *Arabidopsis* haben wie bei anderen Organismen zur Herstellung von etwa 1 cm² großen Chips geführt, auf denen Oligonucleotidproben für fast alle der 25 000 proteincodierenden Gene verankert sind. In einem internationalen Großprojekt wurden Expressionsdaten für alle diese Gene im Verlauf der Entwicklung von *Arabidopsis*, aber auch unter unterschiedlichen Licht- oder Streßbedingungen erarbeitet. Die mehr als 30 Millionen Daten kann man bei folgenden Stellen abrufen und ggf. für einzelne Zielgene analysieren (Genvestigator):

- MPI Tübingen: <http://www.weigelworld.org/resources/microarray/AtGenExpress/>
- ETH Zürich (Genvestigator): <https://www.genevestigator.ethz.ch/>

Knock-out-Linien durch Transposonmutagenese: Bei *Arabidopsis* stehen heute etwa 350 000 T-DNA-Insertionslinien aus verschiedenen Ländern zur Verfügung. Auf einer Homepage am SALK Institute in San Diego (USA) erhält man alle relevanten Informationen für die 25 000 *Arabidopsis*-Gene und kann unmittelbar Knock-out-Linien herausuchen und ggf. online bestellen: <http://signal.salk.edu/> (Plus 13.11 S. 418).

Plastidengenome: Insgesamt wurden inzwischen etwa 50 Plastidengenome aus Algen, Moosen, Farnen und Höheren Pflanzen sequenziert. Die Daten finden sich bei:

- Department of Biology, Penn State University: <http://chloroplast.cbio.psu.edu/cgi-bin/organism.cgi>
- Expert Protein Analysis System: <http://www.expasy.org/sprot/hamap/plastid.html>

Proteinstruktur und -funktion: Die letzte und für die Biologie natürlich wichtigste Ebene der umfassenden Analyse ist die Proteinebene. Hier sind die Daten allerdings bei weitem weniger vollständig als auf der Nucleinsäureebene. Gerade deshalb sind die Anlauf- und Verarbeitungsstellen für Daten zur Struktur und Funktion von Proteinen so wichtig. Zwei Datenbanken sind vor allem zu nennen:

- Expert Protein Analysis System (Swissprot): <http://www.expasy.ch/>
- Protein Data Bank: <http://www.pdb.org/pdb/Welcome.do>