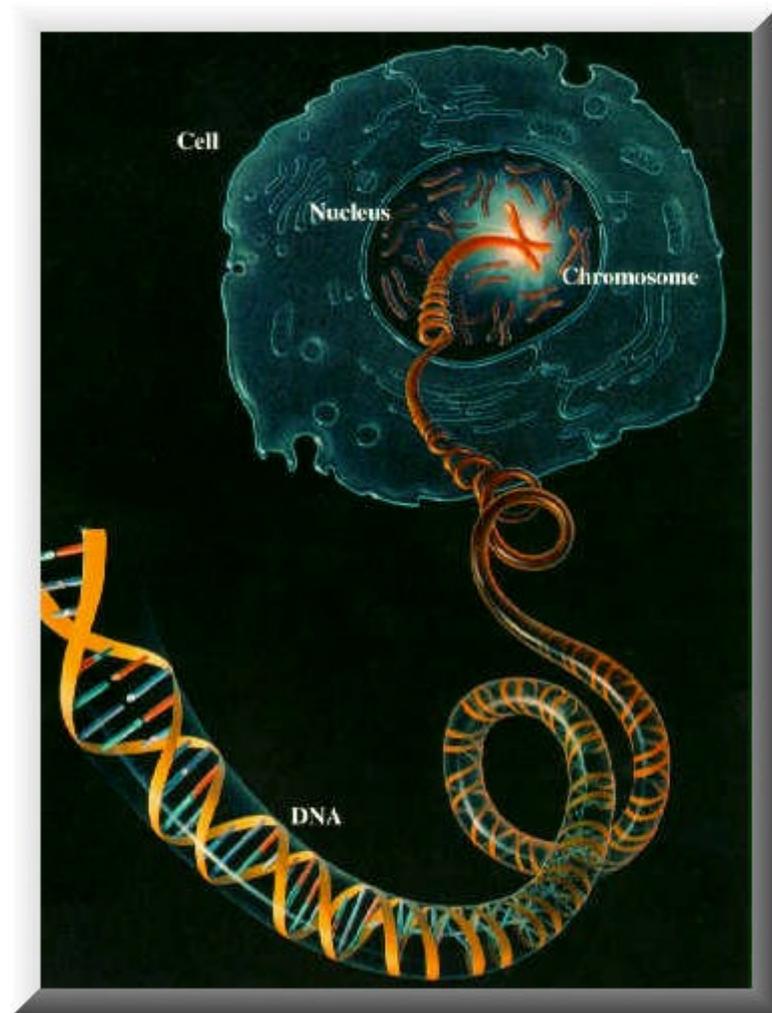


# *Bio-Datenbanken*



## *Einführung in die Bioinformatik*

Bearbeiter: Torsten Glomb    Betreuer: Dr. Dieter Sosna

# *Inhalt*

- Einleitung
- I Proteine
  - I.1 Aminosäuren
  - I.2 Peptidbindung
  - I.3 Primärstruktur: Sequenz der Aminosäuren
  - I.4 Wechselwirkungen zwischen Aminosäure-Seitenketten
  - I.5 Sekundärstrukturen:  $\alpha$ -Helix und  $\beta$ -Blatt
    - I.5.1  $\alpha$ -Helix
    - I.5.2  $\beta$ -Blatt
  - I.6 Tertiärstrukturen: Von der Aminosäure-Sequenz zum gefalteten Protein
    - I.6.1 Protein-Domänen
    - I.6.2 Untereinheiten
    - I.6.3 Faltungen
- II Desoxyribonukleinsäure
  - II.1 Bausteine von Nukleinsäuren
    - II.1.1 Zahlenverhältnis der Nukleotide
  - II.2 Die Doppelhelix
  - II.3 DNA-Helices: Flexible DNA-Strukturen
- III Der genetische Code
- IV Die Transkription
  - IV.1 Initiation (Kettenstart)
    - IV.1.1 Der Promotor
    - IV.1.2 Ereignisse am Promotor
  - IV.2 Elongation (Kettenwachstum)
  - IV.3 Termination (Kettenabbruch)
  - IV.4 mRNA-Reifung
- V Die Translation
  - V.1 Transfer-RNA (tRNA)
  - V.2 Ribosomen
  - V.3 Ablauf der Translation
    - V.3.1 Initiation
    - V.3.2 Elongation
    - V.3.3 Terminierung
- VI Die Genregulation
  - VI.1 Das Operon-Modell
  - VI.2 Substratinduktion
  - VI.3 Endproduktrepression
- VII Genomics
  - VII.1 Zerlegen der DNA
  - VII.2 Herstellen einer Genom-Bibliothek
    - VII.2.1 cDNA-Bibliotheken
  - VII.3 Sequenzieren
    - VII.3.1 Dideoxy- oder Kettenabbruch-Methode
  - VII.4 Monitoring von Genexpressionen mittels DNA-Microarrays
- Zusammenfassung
- Quellen

## *Einleitung*

Vor einigen Monaten wurde die gesamte Abfolge der menschlichen DNA entschlüsselt. Zur Zeit sind die kompletten DNA-Sequenzen von mehr als 60 Organismen in Datenbanken verfügbar. Und diese Zahl steigt fast täglich.

Dieser Zuwachs an genetischer Information wird gewaltige Auswirkungen auf die biomedizinische Forschung und auf die Art und Weise haben, wie Medizin praktiziert wird.

Diese Masse an Informationen in den Datenbanken muss sinnvoll verwaltet werden, um gezielte Suchen nach Gensequenzen durchführen zu können und um neue Erkenntnisse daraus abzuleiten.

Die Grundlagen für das Verständnis der biologischen Strukturen und Prozesse, die in diesem Zusammenhang eine Rolle spielen, sollen an dieser Stelle erläutert werden. Desweiteren wird ein Einblick in die Funktionsweise der Microarrays geliefert, die in der Erforschung der Gene und ihrer Umsetzung eine bedeutende Rolle erlangt haben.

# I. Proteine

Tiere oder Pflanzen einer gegebenen Art haben Proteine mit annähernd gleicher Struktur. Die Baupläne für diese Strukturen werden von Generation zu Generation weitergegeben.

Wir unterscheiden zwei Formen von Proteinen:

- Globuläre Proteine: Mit Abstand betrachtet, haben diese eine annähernd kugelförmige oder ellipsoide Form. Alle Enzyme<sup>1</sup> sind globuläre Proteine.
- Faserförmige Proteine: Sie sind meist Strukturproteine innerhalb und außerhalb der Zelle. Manche faserförmigen Proteine haben einen sehr hohen Gehalt (bis zu 90 %) an  $\alpha$ -Helix-Strukturen<sup>2</sup>.

## I.1 Aminosäuren

Alle Aminosäuren besitzen am zentralen Kohlenstoffatom, dem Alpha-C-Atom (auch  $C_{\alpha}$ -Atom geschrieben), eine Carboxylgruppe, eine Aminogruppe, ein Wasserstoffatom und eine Seitenkette (Abb. 1.1). Die 20 Aminosäuren werden unterschieden durch die Art der Seitenkette, also durch deren Größe, Form und Ladung. In der Abb. 1.2 erhält man einen Überblick über die verschiedenen Aminosäuren mit den üblichen Abkürzungen (im Drei-Buchstaben-Code und im Ein-Buchstaben-Code)

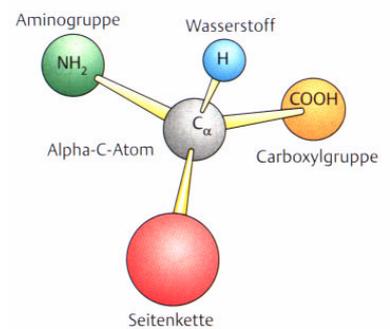


Abb. 1.1 Allgemeine Struktur einer Aminosäure

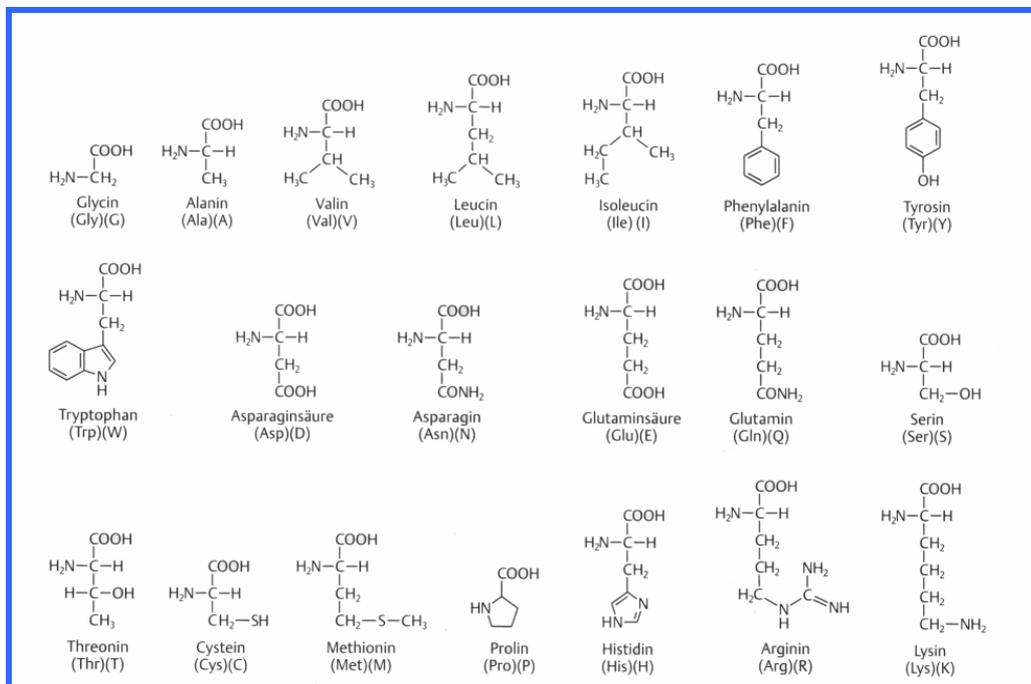


Abb. 1.2 Die zwanzig in Proteinen vorkommenden Aminosäuren

<sup>1</sup> Enzym – biologischer Katalysator, der eine chemische Reaktion in Gang setzt, beeinflusst oder beschleunigt

<sup>2</sup> Der Begriff der  $\alpha$ -Helix wird im Folgenden noch erläutert.

## I.2 Peptidbindung

Im Protein ist die Carboxylgruppe einer gegebenen Aminosäure mit der Aminogruppe der benachbarten Aminosäure durch eine kovalente Bindung verknüpft, die Peptidbindung. Ein Peptidabschnitt hat eine Richtung: Auf der linken Seite liegt eine freie, nicht mit einer benachbarten Aminosäure verknüpfte Aminogruppe, auf der rechten Seite liegt eine freie Carboxylgruppe. Diese Richtung bleibt erhalten im Hinblick auf ein Protein natürlicher Länge mit vielen hundert Bausteinen. Jedes Protein hat also jeweils ein freies Amino- und Carboxyl-Ende, meist kurz mit N-Terminus und C-Terminus benannt.

Normalerweise wird eine Aminosäure-Sequenz nicht als umständliche Formel wiedergegeben, sondern mit Hilfe der Ein- oder Drei-Buchstaben-Abkürzungen. Das Amino-Ende wird dabei links notiert.

## I.3 Primärstruktur: Sequenz der Aminosäuren

Proteine sind Makromoleküle, die aus vielen Einzelbausteinen aufgebaut sind, den Aminosäuren. Proteine sind fast ausschließlich aus 20 Aminosäuren aufgebaut, die in wechselnden Zahlen und in wechselnden Kombinationen zu langen unverzweigten Ketten verknüpft sind.

Als Primärstruktur eines Proteins bezeichnet man die Reihenfolge (Sequenz) der Aminosäuren, aus denen es sich zusammensetzt. Die meisten Proteine bestehen aus 100-800 Aminosäuren, aber auch Proteine mit weit mehr als 800 oder weniger als 100 Aminosäuren sind keine Seltenheit. Aminosäure-Sequenzen mit weniger als 20 Bausteinen bezeichnet man als Peptide.

## I.4 Wechselwirkungen zwischen Aminosäure-Seitenketten

Bei der Betrachtung von einer Folge von Aminosäuren, ohne deren Lage im dreidimensionalen Raum zu beachten, spricht man meist von Polypeptidketten. Zur Beschreibung eines Proteins gehört jedoch auch seine dreidimensionale Struktur, die durch Wechselwirkungen zwischen den einzelnen Aminosäure-Ketten entsteht.

Obwohl schon dreidimensionale Formen zahlreicher Proteine bekannt sind, ist das Wissen über die genauen Regeln, nach denen sich die Faltung einer Polypeptidkette zum nativen Protein vollzieht, noch lückenhaft. Dies ist eine der ungelösten fundamentalen Fragen der Molekularbiologie. Einige allgemeine Prinzipien und Grundregeln können jedoch formuliert werden:

- **Die Peptidbindung selbst ist eine starre Fläche.** Die Faltung erfolgt als Drehung um die Bindungen der Alpha-C-Atome. Aufgrund der chemischen Natur der beteiligten Aminosäuren unterliegen diese Drehungen gewissen Einschränkungen.
- **Form und Art der Seitenketten beeinflussen die Wechselwirkungen zwischen den Aminosäure-Bausteinen.** Diese Wechselwirkungen sind entweder elektrostatischer Art zwischen positiv und negativ geladenen Seitenketten (Abb. 1.3) oder zum größeren Teil Wasserstoffbrückenbindungen zwischen polaren Seitenketten. Aminosäuren mit geladenen oder polaren Seitenketten bestimmen zudem die Wechselwirkung mit dem umgebenden Wasser. Die Folge ist eine Anhäufung von geladenen und polaren Aminosäuren an der Oberfläche des Proteins (hydrophile<sup>1</sup> Aminosäuren), während sich die nichtpolaren Aminosäuren vom Wasser wegwenden

---

<sup>1</sup> hydrophil – Wasser anziehend

(hydrophobe<sup>1</sup> Aminosäuren) und deswegen bevorzugt im Inneren des Proteins gefunden werden.



Abb. 1.3 Die Wasserstoff-Brückenbindung

- **Disulfid-Brücken stellen zusätzliche kovalente Bindungen dar.** Zwischen den SH-Gruppen in den Seitenketten von Cysteinen bilden sich Disulfid-Bindungen aus. Für diese Reaktion ist eine oxidative<sup>2</sup> Umgebung notwendig. Deswegen findet man diese Art der Bindung nicht bei Proteinen innerhalb einer Zelle, aber sehr wohl nach deren Export in die Zellumgebung. Die Form der Immunglobuline (Antikörper) im Serum wird zum Beispiel sehr deutlich durch einige charakteristische Disulfid-Bindungen geprägt, die die dreidimensionale Struktur dieser und anderer extrazellulärer Proteine stabilisieren.
- **Zwei Arten von regelmäßigen Anordnungen bestimmen einen wesentlichen Teil der Proteinstruktur;** die Sekundärstrukturen  $\alpha$ -Helix und  $\beta$ -Blatt.

## I.5 Sekundärstrukturen: $\alpha$ -Helix und $\beta$ -Blatt

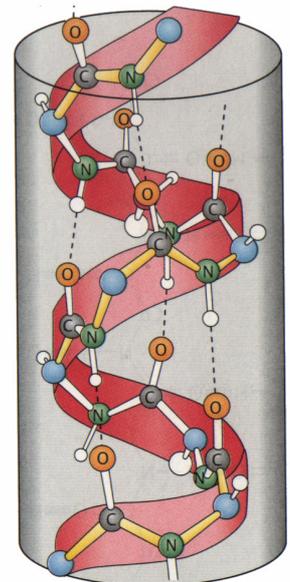
Wasserstoff-Brücken können sich zwischen der CO-Gruppe (Carbonylgruppe) einer Peptidbindung und der NH-Gruppe (Iminogruppe) einer anderen Peptidbindung ausbilden.

### I.5.1 $\alpha$ -Helix

In der  $\alpha$ -Helix besteht eine Wasserstoff-Brücke zwischen der CO-Gruppe einer Aminosäure und der NH-Gruppe der viertnächsten Peptidbindung in der Reihe aufeinanderfolgender Aminosäuren (Abb. 1.4). Dadurch kommt es zu der gut definierten Struktur einer rechtsläufigen Helix:

- In jeder vollständigen Drehung der Helix befinden sich 3,6 Aminosäuren.
- Der Abstand zwischen zwei aufeinanderfolgenden Aminosäuren beträgt 0,15 nm (1,5Å).

Theoretisch können alle Aminosäuren eine  $\alpha$ -Helix bilden, mit der Ausnahme von Prolin. Diese Aminosäure hat statt der üblichen Aminogruppe (NH<sub>2</sub>) eine Iminogruppe (NH), deren Wasserstoff bei der Peptidbindung verloren geht und deswegen in der Aminosäure-Kette für Wasserstoff-Brücken nicht mehr zur Verfügung steht. Deshalb kann man vorhersagen, dass Prolin irgendwo vor oder nach, aber nicht in der Mitte einer  $\alpha$ -Helix vorkommt.



bb. 1.4 Die  $\alpha$ -Helix

<sup>1</sup> hydrophob – Wasser abstoßend

<sup>2</sup> oxidativ – sauerstoffhaltig

## I.5.2 b-Blatt

Das zweite wichtige Strukturelement von Proteinen sind  $\beta$ -Blätter (*b sheets*). Hierbei bilden sich Wasserstoff-Brücken zwischen verschiedenen Abschnitten der Polypeptidkette und nicht zwischen hintereinanderliegenden Aminosäuren, wie bei der  $\alpha$ -Helix.  $\beta$ -Blätter bestehen aus einzelnen  $\beta$ -Strängen (*b strands*), meist 5-10 Aminosäuren lang, in denen jeweils die CO- mit den NH-Gruppen und die NH- mit den CO-Gruppen eines anderen Stranges über Wasserstoff-Brücken verbunden sind. So können mehrere  $\beta$ -Stränge zu einem  $\beta$ -Blatt verbunden sein, wobei die hintereinanderliegenden Alpha-C-Atome abwechselnd unter- oder oberhalb der Ebene zu liegen kommen. Es kommt zur Form des  $\beta$ -Faltblatts (*pleated sheet*). (Abb. 1.5)

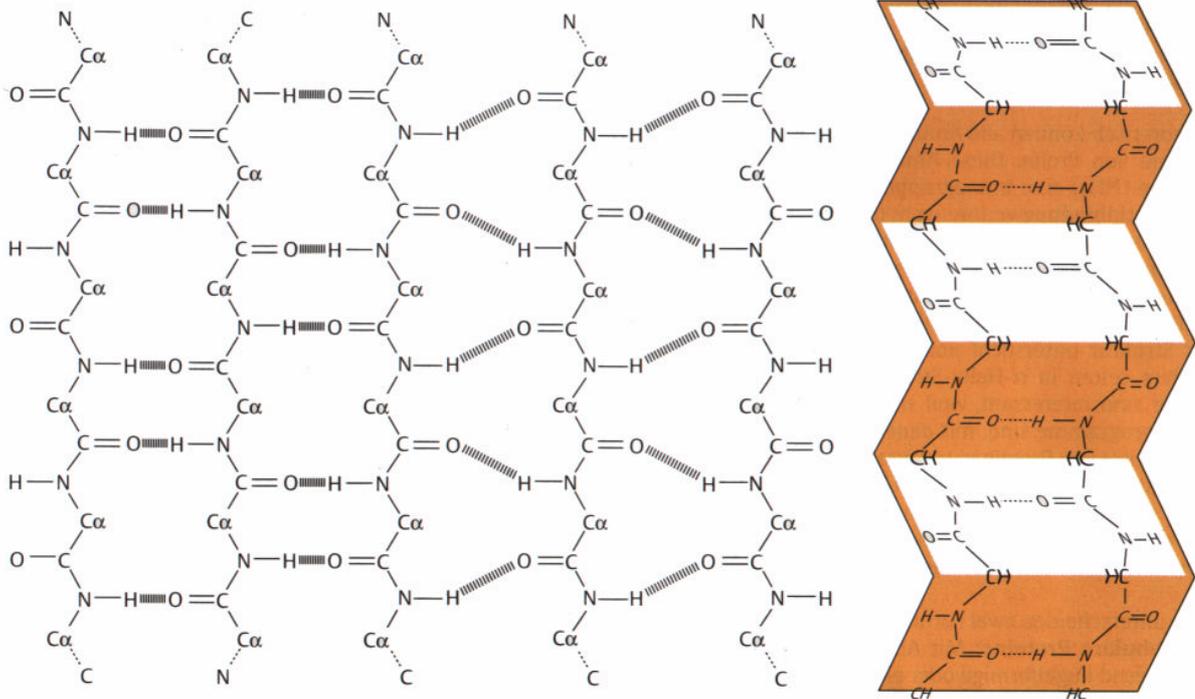


Abb. 1.5 Das  $\beta$ -Blatt

Die  $\beta$ -Stränge eines Faltblattes können in zwei Richtungen verlaufen:

- parallel: die Richtung N-Terminus zu C-Terminus ist die gleiche in den nebeneinanderliegenden Strängen.
- antiparallel: ein Strang geht vom N- zum C-Terminus, und der antiparallele Strang vom C- zum N-Terminus. In globulären Proteinen kommt es nicht selten zur Ausbildung von  $\beta$ -Blättern, in denen einige Stränge parallel und andere antiparallel laufen.

Die Enden einzelner  $\beta$ -Stränge sind auf jeden Fall durch Schleifen (turns) miteinander verbunden. Diese Schleifen (Abb. 1.6) bestehen meist aus 4-8 Aminosäuren, die oft eine definierte Anordnung in der dreidimensionalen Struktur eines Proteins einnehmen. Die Aminosäuren einer Schleife sind häufig geladen oder polar (hydrophil) und liegen an der Oberfläche von Proteinen.

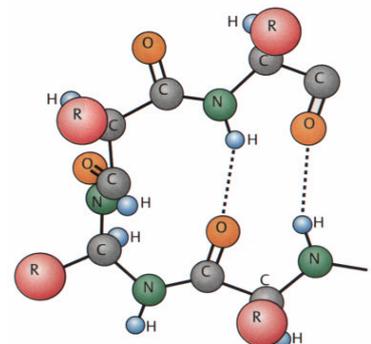
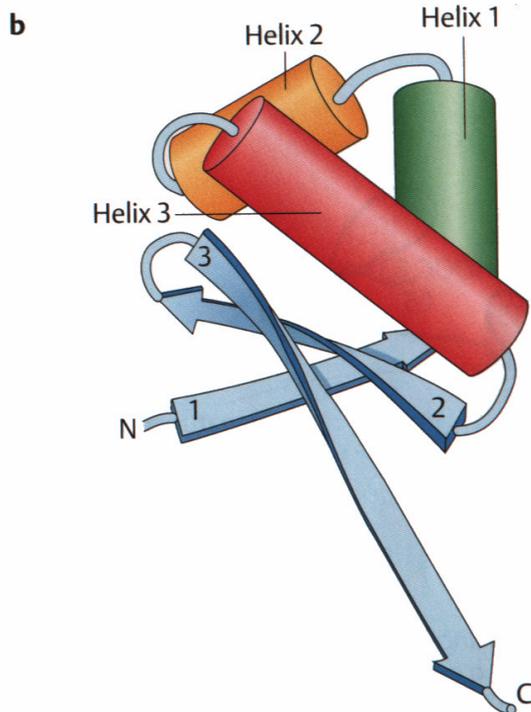
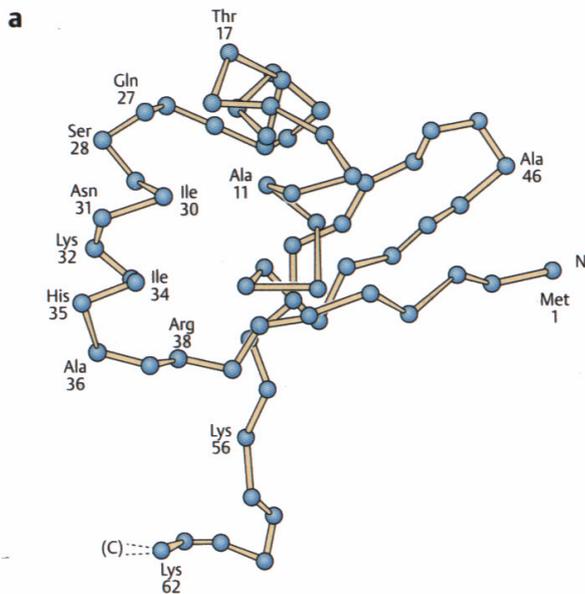


Abb. 1.6 Schleife an Enden von  $\beta$ -Blättern

## I.6 Tertiärstrukturen: Von der Aminosäure-Sequenz zum gefalteten Protein



Bei jeder neuen Aminosäure-Sequenz, die entdeckt wird, stellt sich die Frage nach der Funktion, und diese schließt vor allem die Frage nach der dreidimensionalen Form des Proteins ein. Bevor mit Hilfe physikalischer Methoden die Form aufgeklärt ist, hilft man sich zunächst mit leicht zugänglichen, aber sehr vorläufigen und ungenauen Informationen (Computerberechnungen). Letztendlich können jedoch nur physikalische Verfahren die dreidimensionale Struktur eines Proteins endgültig aufklären.

Als Beispiel sei hier das Cro-Protein des Bakteriophagen Lamda gegeben (Abb. 1.7). Abb. 1.7a zeigt die Anordnung der Alpha-C-Atome im Raum. In der molekularen Genetik hat sich aber eine noch einfachere und übersichtlichere Darstellung von dreidimensionalen Proteinstrukturen eingeführt, nämlich  $\alpha$ -Helix-Abschnitte als Zylinder und  $\beta$ -Stränge als breite Pfeile darzustellen, wobei die Pfeilspitzen in Richtung C-Terminus weisen, wie in Abb. 1.7b zu sehen ist.

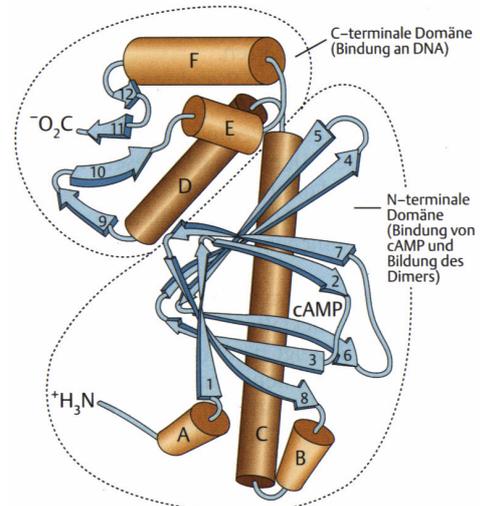


Abb. 1.7 Dreidimensionale Struktur des Cro-Proteins

Abb. 1.8 Protein-Domänen

### I.6.1 Protein-Domänen

Mit einigem Abstand betrachtet, hat das Cro-Protein die kompakte und relativ einheitliche Form eines Ellipsoids. Größere Proteine sind jedoch meist viel weniger einheitlich in ihrer Form. Sie sehen eher aus, als wenn sie aus zwei oder mehr kugelförmig, ellipsoid oder anders geformten Strukturen verschmolzen wären. Die Einzelteile sind oft durch ungeordnetere

Abschnitte der Aminosäure-Sequenz verbunden, und einzelne Struktur-Domänen können unterschieden werden.

Eine Domäne ist die kleinste Protein-Einheit mit einer definierten und unabhängig gefalteten Struktur. Die meisten Domänen bestehen aus 50-150 Aminosäuren. Vielfach führen Domänen eigene Reaktionen aus, deren Zusammenwirken dann die Funktion des Gesamtproteins ausmachen.

Ein Beispiel für die Domänenbildung ist das CAP-Protein (Abb. 1.8). Zu sehen ist die dreidimensionale Anordnung von  $\alpha$ -Helices und  $\beta$ -Blättern. Dementsprechend hat es eine kleine Domäne mit drei  $\alpha$ -Helices und eine zweite größere Domäne.

### **I.6.2 Untereinheiten**

Der Aufbau von Proteinen aus Untereinheiten ist eher die Regel als die Ausnahme.

Gelegentlich bezeichnet man die geordnete Zusammenlagerung von Untereinheiten auch als **Quartärstruktur**. Dabei können die Untereinheiten identisch sein und in zwei- oder mehrfacher Ausführung vorkommen. Sie können aber auch verschieden sein und sich in größerer Quartärstruktur zusammenfinden.

### **I.6.3 Faltungen**

Bei hoher Temperatur, in milden Säuren oder Basen sowie in hochkonzentrierten Harnstofflösungen verliert ein Protein seine charakteristische Faltung, ohne dass die Polypeptidkette zerbricht. Man spricht von Denaturierung eines nativen Proteins. Ein denaturiertes Protein kann - selbstverständlich - seine Funktion als Enzym oder als Strukturkomponente nicht mehr ausführen.

Die Rückfaltung der Polypeptidkette zum nativen, funktionsfähigen Protein nennt man Renaturierung. Dieser Prozess der Renaturierung zeigt, dass im Prinzip die Information zur Ausbildung der dreidimensionalen Struktur in der Polypeptidkette selbst liegt, d.h. von der Art, der Zahl und der Reihenfolge der Aminosäure-Seitenketten bestimmt wird.

Es genügt also, wenn die Information zur Herstellung der Aminosäure-Folgen in den Genen (Bauplänen) vererbt wird. Die dreidimensionale Struktur kann sich daraus sozusagen von selbst bilden (*self assembly*).

Die Zelle verhindert zudem falsche Faltungen mit Hilfe einer Klasse von Hilfsproteinen, die als molekulare Chaperone (*molecular chaperons*) bekannt sind. Diese Proteinkomplexe stabilisieren zunächst ungefaltete, später dann halbwegs gefaltete Übergangszustände.

Die Wirkungsweise von Chaperonen bei der intrazellulären Faltung von Polypeptidketten wird als "*assisted self assembly*" beschrieben. *Assisted self assembly* bedeutet also, dass die Chaperone der Polypeptidkette beim Finden des richtigen und passenden Faltungsweges helfen.

## II. Desoxyribonukleinsäure

Alle Organismen besitzen einen gemeinsamen Stammbaum aus der frühen Evolution der Erde. Die Grundlagen der Molekulargenetik von Bakterien, eukaryotischen Einzellern, Pflanzen und Tieren sind daher gleich, wichtige Unterschiede erkennt man erst im Detail. Besonders deutlich werden die evolutionären Gemeinsamkeiten im Träger der genetischen Information, die in den Zellen aller Lebewesen auf dem selben Makromolekül gespeichert ist. Dieses Makromolekül ist die Desoxyribonukleinsäure, abgekürzt mit DNS oder häufiger mit DNA (**d**esoxyribo**n**ucleic **a**cid).

Zwischen 1940 und 1950 gelang der erste eindeutige Nachweis für die Funktion der DNA als Genträger durch Untersuchungen an Bakterien und deren Viren (Bakteriophagen).

### II.1 Bausteine von Nukleinsäuren

Die Einzelbausteine von Nukleinsäuren werden Nukleotide (Abb. 2.1) genannt und bestehen aus den folgenden drei Komponenten:

- ein Fünf-Kohlenstoff-**Zucker** (Pentose)
- eine **Purin- oder Pyrimidinbase**, die über N-glykosidische Bindung an die Pentose gebunden ist
- ein **Phosphatrest**, der mit dem C5-Atom der Pentose verbunden ist.

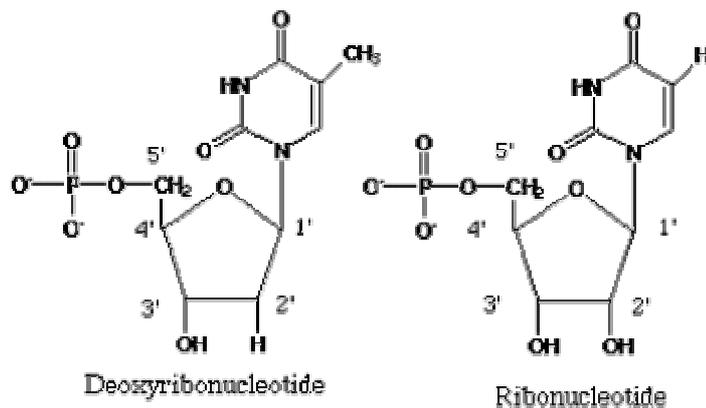
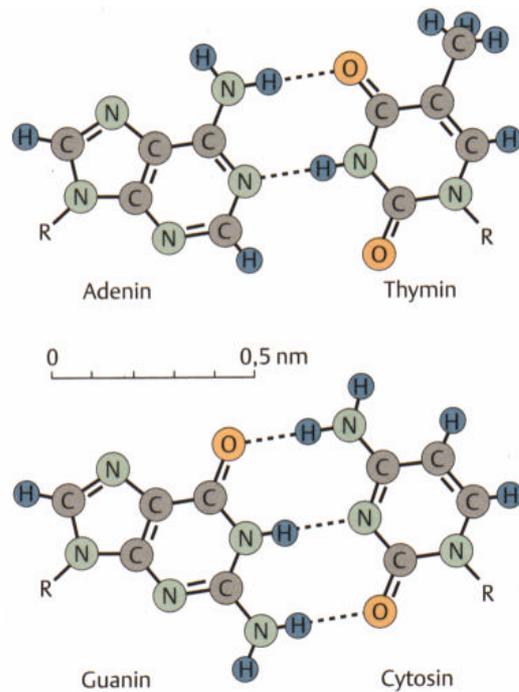


Abb. 2.1 Nukleotide

Zellen enthalten zwei Nukleinsäuren, die DNA und die Ribonukleinsäure (RNA). Die hauptsächlichen Unterschiede liegen bei den Pentosen und Basen, die verwendet werden. Die Pentose der DNA heißt Desoxyribose und die der RNA Ribose. Daher bezeichnet man zur Unterscheidung die Bausteine der DNA als Desoxy(ribo)nukleotide und die Bausteine der RNA als Ribonucleotide. In der DNA kommen jeweils zwei Purinbasen, Adenin (A) und Guanin (G), und zwei Pyrimidinbasen, Thymin (T) und Cytosin (C), vor, wobei sich immer eine Adeninbase mit einer Thyminbase paart und eine Guaninbase mit einer Cytosinbase (Abb. 2.2).



**Abb. 2.2** Basenpaarungen

Bei der RNA steht anstelle des Thymins die Base Uracil. Einen Vergleich der beiden Nucleinsäuren zeigt Tab. 2.1.

	DNA	m-RNA	r-RNA	t-RNA
Struktur	Doppelstrang	Einzelstrang		
Länge	sehr groß	kurz variabel	kurz 16S; 23S	kurz 4S
Zucker	Desoxyribose	Ribose		
Basen	A, T, C, G	A, U, C, G	A, U, C, G	A, U, C, G und viele "seltene"
Vorkommen	Zellkern	Plasma	Ribosom	Plasma
Synthese	Replikation	Transkription	Art Transkription	

**Tab. 2.1** Vergleich zwischen der DNA und den drei RNA-Typen

(A - Adenin, T - Thymin, C - Cytosin, G - Guanin, U - Uracil, S – Sedimentationskonstante<sup>1</sup>)

Auf die Funktionen der unterschiedlichen RNA-Typen gehen wir bei der Transkription und der Translation ein.

In polymeren Nucleinsäuren sind die einzelnen Bausteine durch Phosphatbrücken zwischen dem 5'-C-Atom des einen mit dem 3'-C-Atom des Benachbarten verknüpft. Dadurch entstehen lange unverzweigte Polynukleotidketten, die durch das Zucker-Phosphodiester-Band zusammengehalten werden. Die Polynukleotidkette hat eine definierte Richtung mit einem freiem 5'-Ende und einem freiem 3'-Ende. (Abb. 2.3)

<sup>1</sup> Sedimentationskonstante - 1 S = 10<sup>-13</sup> s

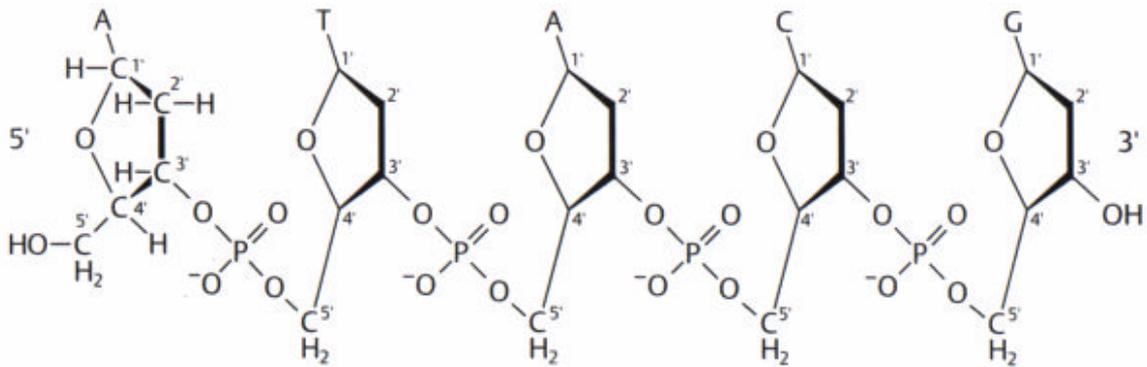


Abb. 2.3 Folgen von Nucleotiden

Natürliche DNA-Moleküle bestehen aus vielen tausenden, oft Millionen aneinandergelinkter Nucleotide.

### II.1.1 Zahlenverhältnis der Nucleotide

Es gilt:  $P(A) = P(T)$  und  $P(G) = P(C)$ . Aufgrund dieser Regel kann man die prozentuale Zusammensetzung jeder DNA voraussagen, wenn der Prozentsatz einer einzigen Base bekannt ist. Als Charakteristikum einer DNA wird daher entweder der AT-Gehalt (in %) angegeben oder das folgende Verhältnis:

$$\frac{A + T}{C + G}$$

## II.2 Die Doppelhelix

Röntgenstruktur-Analysen ergaben, dass die DNA-Fasern eine regelmäßige Periodik zeigen und aus jeweils 2 DNA-Strängen aufgebaut sein müssen. In jeder menschlichen Zelle finden sich etwa 90 Zentimeter solcher gedrehten Stränge. Die Doppelhelix (Abb. 2.4) wird oft als Doppelspirale aus zwei Bändern dargestellt, die ähnlich einer Wendeltreppe durch Stufen miteinander verbunden sind. Die Bänder bilden das Rückgrat der DNA-Kette, das aus den Zucker-Phosphat-Teilen der polymerisierten Nucleotide besteht. Die basischen Ringe der Nucleotide sind nach innen gerichtet, wobei gegenüberliegende Basen durch Wasserstoffbrückenbindungen miteinander verbunden sind. Paarungen erfolgen immer zwischen einer Purin- und einer Pyrimidinbase, also wie schon erwähnt A mit T und G mit C. Zwei gepaarte Purinringe würden zuviel Platz in der Helix in Anspruch nehmen, wohingegen zwei gegenüberliegende Pyrimidinringe den Platz nicht ausfüllen würden. Ein Teil der Stabilität der DNA-Doppelhelix rührt von den spezifischen Basenpaarungen her, ein anderer Teil beruht auf den hydrophoben Bindungen zwischen den aufeinander gestapelten Basen eines Stranges. In dem Kalottenmodell der Abb. 2.4c erscheinen die Zucker-Phosphat-Bänder wie Wülste, die eine kleine und eine große Rinne bilden. Diese Rinnen entstehen dadurch, dass sich die Anheftungsstellen für Desoxyribose in den Basenpaaren nicht direkt gegenüberliegen (s. Abb. 2.2).

Die beiden Basenstränge sind komplementär, d.h., aufgrund der strengen Regeln für die Basenpaarung kann man von der Nucleotid-Sequenz eines DNA-Stranges auf die des gegenüberliegenden Stranges schließen.

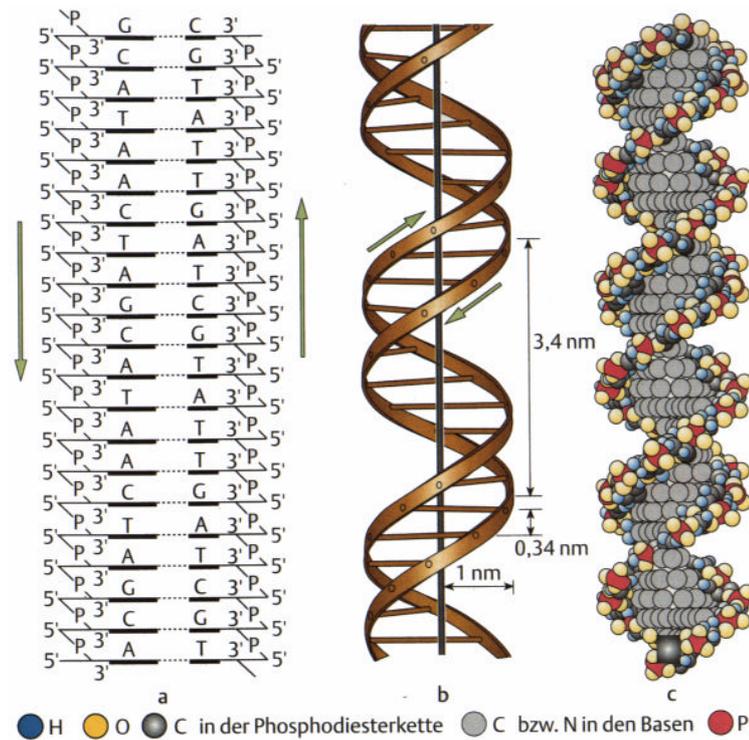


Abb. 2.4 Die DNA-Doppelhelix

### II.3 DNA-Helices: Flexible DNA-Strukturen

Die Struktur von wahrscheinlich 99% der zellulären DNA entspricht der Idealform, wie in Abb. 2.4 gezeigt. Somit machen bei dem größten Teil der natürlich vorkommenden DNA 10,4 bis 10,5 Basenpaare eine komplette Drehung der Helix aus.

Basen können in Abhängigkeit ihrer spezifischen Abfolge unterschiedliche Positionen innerhalb der vorgegebenen Struktur der DNA einnehmen. Diese nehmen sie nach folgenden Kriterien ein:

- Ausbildung optimaler hydrophober Wechselwirkungen zwischen benachbarten Basenpaaren
- Vermeiden eines Zusammentreffens funktioneller Seitengruppen in benachbarten Nucleotiden

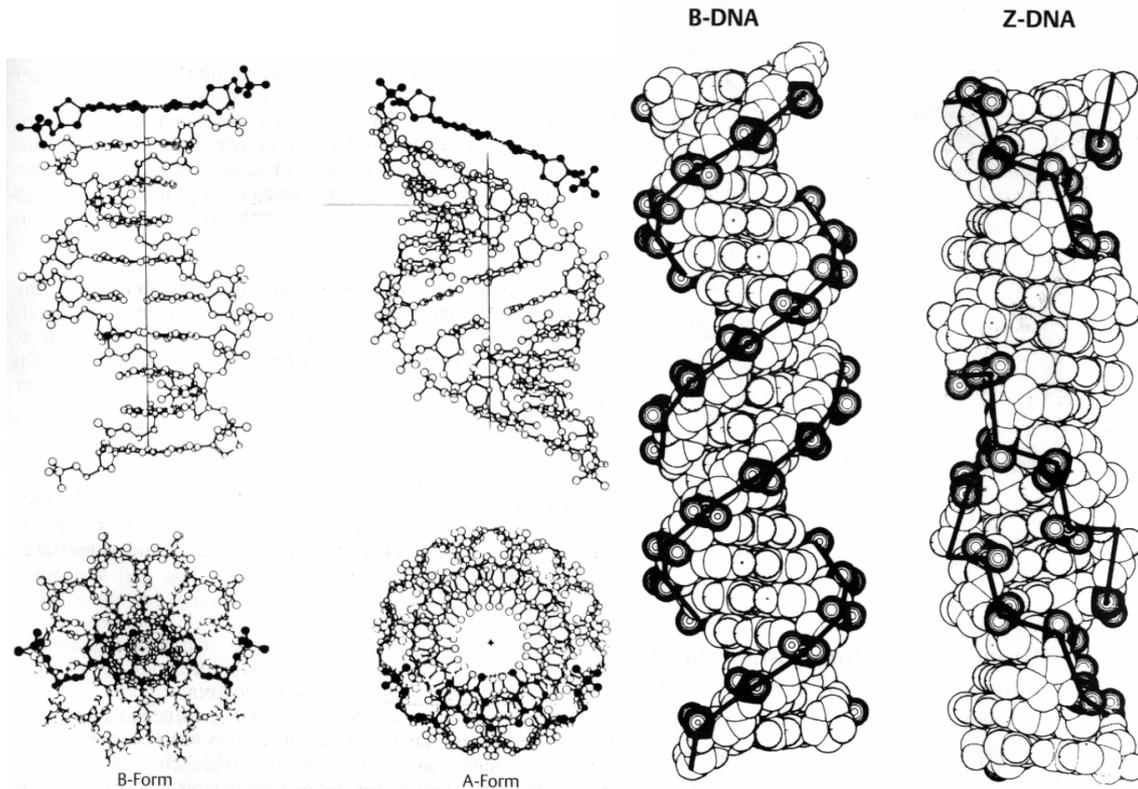
Grundlage für diese Kriterien ist die Flexibilität der chemischen Bindungen im Nucleotid. Die betrifft alle Bindungen im Desoxyribose-Molekül und die Phosphodiesterbindung.

Die DNA besitzt drei verschiedene Formen, und zwar die A-Form, die B-Form (Abb. 2.5) und die Z-Form (Abb. 2.6). Die A- und B-Form sind rechtsläufig (Tab. 2.2), während die Z-Form linksläufig ist. Die B-Form stellt die oben erwähnte Idealform dar, aus der bei drastischer Wasserabnahme die A-Form hervorgeht. Die Bezeichnung für die Z-Form stammt von dem Zick-Zack-Verlauf von dem Zucker-Phosphodiester-Bandes.

In der B-Form gibt es noch einige Variation, die hier aber nur erwähnt werden sollen; das ist die gebogene DNA, die kreuzförmige DNA an Palindromen und die intramolekulare Triplex-Struktur.

	A-Form	B-Form
Basenpaare/Helixwindung	ca. 11	10,4- 10,5
Abstand der Basenpaare	0,26 (+/- 0,04) nm	0,34 (+/- 0,04) nm
Winkel zwischen zwei Basen	33,1° (+/- 5,9)	35,9° (+/- 4,3)
Winkel zwischen Helixachse und Basenpaaren	71-77°	ca. 90°

**Tab. 2.2** Strukturmerkmale von rechtsläufigen DNA-Formen



**Abb. 2.5** DNA-Formen

**Abb. 2.6** Z-DNA

### III. Der genetische Code

Als Gen bezeichnen wir einen DNA-Abschnitt, der die Information zum Aufbau einer Polypeptidkette enthält, die sich dann zum nativen Protein faltet. Die Gesamtzahl der Gene eines Organismus nennen wir Genom.

Die DNA-Moleküle sind die Träger der genetischer Information. Die lineare Folge von Nukleotiden in der DNA bestimmt die lineare Abfolge der Aminosäuren im Protein. Proteine bestehen aus 20 Aminosäuren, die in wechselnder Zahl und Zusammensetzung zu langen Ketten verknüpft sind. Die DNA hingegen besteht aber nur aus 4 Nukleotiden. Somit kann ein Nukleotid nicht die Lage einer Aminosäure im Protein bestimmen, sondern nur eine Gruppe von Nukleotiden. Aber auch eine Gruppe von zwei Nukleotiden ist nur in der Lage  $4^2 = 16$  Aminosäuren zu kodieren, jedoch nicht benötigten 20 Aminosäuren.

Tatsächlich wird eine Aminosäure durch eine Dreierfolge von Nukleotiden kodiert. Diese Dreierfolge wird Triplet oder, als Einheit des genetischen Codes, Codon genannt. Aus drei unterschiedlichen Nukleotiden lassen sich  $4^3 = 64$  Dreierkombinationen bilden. Somit stehen 64 Triplets 20 Aminosäuren gegenüber, so dass mehrere Triplets für ein und dieselbe Aminosäure stehen. Der genetische Code ist also redundant, man sagt dazu, er ist degeneriert. Nur drei Triplets der 64 stehen nicht für Aminosäuren, sondern haben eine andere Funktion bei der Umsetzung genetischer Information.

Eine Folge von drei Nukleotiden auf der DNA steht also für eine Aminosäure im Genprodukt Protein. Daraus folgt die Frage, welche Triplets für welche Aminosäuren kodieren.

Biochemische Untersuchungen ergaben die in Tab. 3.1 angegebenen Kodierungen.

Erste Base 5'-Ende	Zweite Base				Dritte Base 3'-Ende
	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr „Stop“ „Stop“	Cys Cys „Stop“ Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met (Start)	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

**Tab. 3.1** Der genetische Code

Eine weitere bekannte Darstellungsart für die Kodierungen ist die Codesonne (Abb. 3.1), die von innen nach außen zu lesen ist.

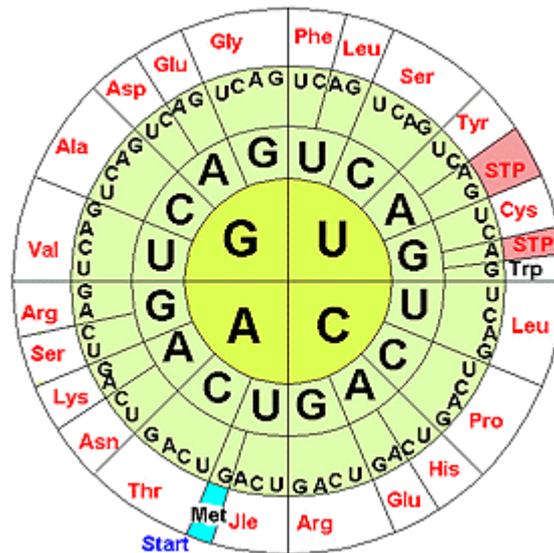


Abb. 3.1 Die Codesonne

Heute besteht kein Zweifel mehr, dass die Code-Wörter der Tab. 3.1 für alle Organismen auf der Erde gelten. Daraus folgt, dass der genetische Code universell ist.

## IV. Die Transkription

Die Transkription ist der erste Schritt der Proteinsynthese. In dieser Phase wird die genetische Information von der DNA auf die Messenger-RNA (mRNA, s. Tab. 2.1) kopiert, um sie dann an anderer Stelle zu verwenden. Die Transkription gliedert sich in drei Abschnitte; die Initiation, die Elongation und die Termination.

### IV.1 Initiation (Kettenstart)

#### IV.1.1 Der Promotor

Die mRNA-Synthese sollte nicht irgendwo auf der DNA beginnen, sondern genau vor einem Gen. Es sollte auch nicht irgendein Strang transkribiert werden, sondern nur der Strang, dessen Transkript die genetische Information trägt, der codogene Strang.

Die RNA-Polymerase, das Enzym zur RNA-Polymerisation, bindet bevorzugt an Stellen auf der DNA, die vor einem Genanfang liegen. Eine solche Erkennungs- und Bindungsstelle nennt man Promotor.

Um die Promotorregion zu finden, bildet die RNA-Polymerase zunächst eine schwache Bindung mit irgendeiner Stelle auf der DNA. Von dort aus gleitet sie an der DNA entlang, d.h. sie löst und bindet sich im Wechsel, bis sie auf eine Promotor-Sequenz trifft, an der sie dann mit großer Stabilität haften bleibt.

#### IV.1.2 Ereignisse am Promotor

Nun selektiert die RNA-Polymerase den codogenen Strang, das ist der Strang, der vom 3'-Ende zum 5'-Ende transkribiert wird, und öffnet die Basenpaare der DNA, um diese zu entwenden (Abb. 4.1).

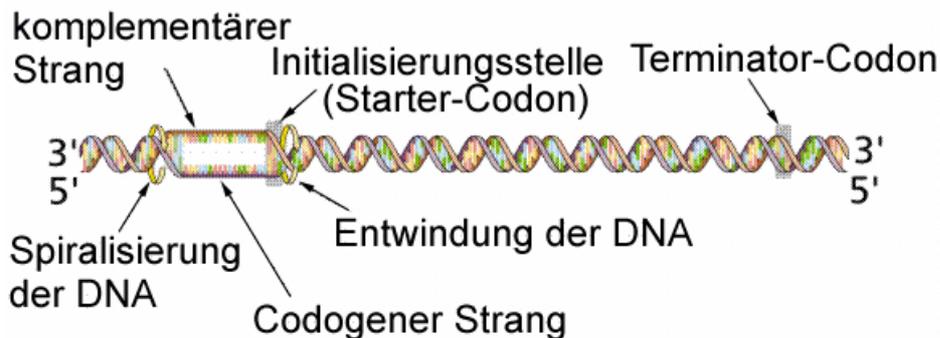


Abb. 4.1 Initiation am Promotor

### IV.2 Elongation (Kettenwachstum)

Nachdem ein Stück der DNA nach dem Promotor entwunden wurde, bilden sich schon die ersten komplementären Ribonukleotide in Richtung 3'-Ende der mRNA-Kette. Aber erst wenn die mRNA-Stücke eine Länge von 12 oder mehr Nukleotiden erreicht haben, geht der Initial-Komplex in den Elongations-Komplex über. Dabei verändert die RNA-Polymerase ihre Struktur, einige Untereinheiten werden abgespalten.

Das Prinzip der Kettenverlängerung (Abb. 4.2) liegt in der ständigen Wiederholung der Reaktionsfolge:

1. Auswahl des passenden Nucleotids,
2. dessen Anheftung an das 3'-Ende und
3. Bewegung des Enzyms relativ zur DNA-Matrize.

Dies läuft mit einer Geschwindigkeit von durchschnittlich 30 - 60 Nucleotiden / Sekunde bei optimalen Bedingungen (37°C) ab.

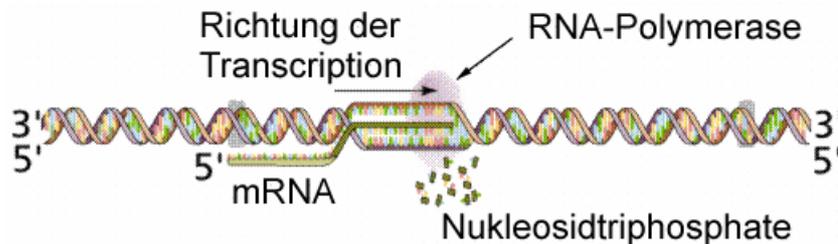


Abb. 4.2 Kettenwachstum

### IV.3 Termination (Kettenabbruch)

Das Ende eines Transkriptionsabschnittes auf der DNA bezeichnet man als Terminator. Dort kommt die RNA-Polymerase zum Stillstand und die synthetisierte mRNA und die RNA-Polymerase werden freigesetzt.

### IV.4 mRNA-Reifung

Eukaryonten<sup>1</sup>-Gene enthalten meist nicht fortlaufend Information, sondern sind durch Nonsens- und andere Sequenzen unterbrochen. Man nennt diese Abschnitte Introns. Kodierende Sequenzen werden als Exons bezeichnet. Je komplexer der Organismus, desto umfangreicher sind die Introns in deren Gene. Nach der Transkription des Gens erfolgt ein Prozess, der die mRNA so bearbeitet, dass die Introns entfernt werden. Diesen Vorgang nennt man mRNA-Reifung, das Herausschneiden der Introns heißt RNA-Splicing. In der Abb. 4.3 ist die mRNA-Reifung am Beispiel der Transkription des Ovalbumin-Gens zu sehen.

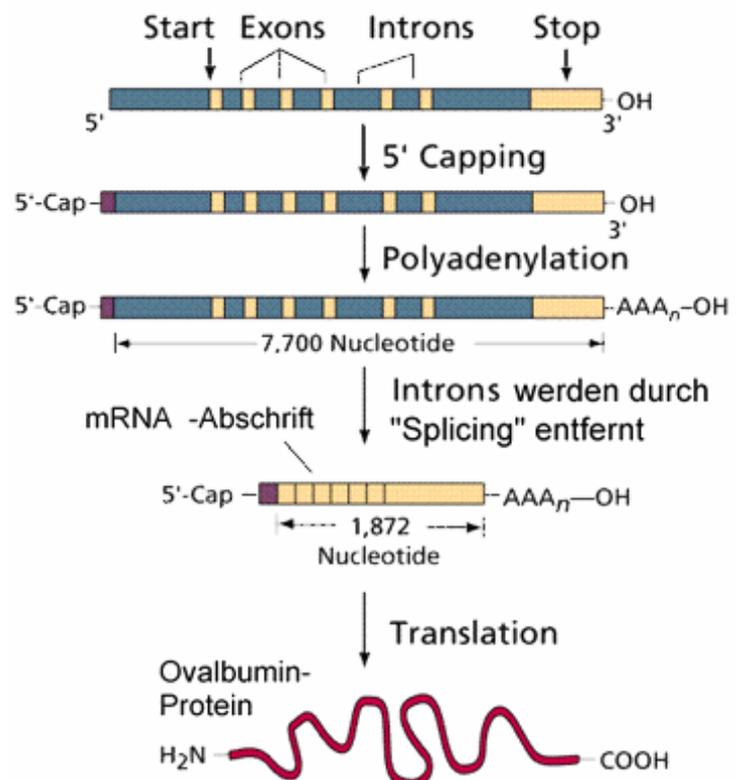


Abb. 4.3 Reifung der mRNA bei Eukaryonten

<sup>1</sup> Eukaryont – Lebewesen, dessen Zelle einen Zellkern mit Doppelmembran und Chromosomen besitzt (alle ein- oder vielzelligen Tiere, Pflanzen und Pilze)

Vor dem Herausschneiden der Introns wird die mRNA noch für den Transport aus dem Zellkern (Nukleus) durch die Kernporen ins Zellplasma (Cytoplasma) etwas aufbereitet. An das 5'-Ende wird eine Art Kappe aus mehreren 7-Methylguanin-Molekülen (einer modifizierten Base) angehängt, die später wichtig für die Bindung an das Ribosom<sup>1</sup> ist. Dieser Vorgang wird als Capping bezeichnet. An das 3'-Ende wird eine Sequenz von bis zu 200 AMP<sup>2</sup>-Nukleotide addiert (Polyadenylation), bekannt als Poly-A. Ihre Funktion ist noch nicht bekannt. Nun binden für den Transport ins Cytoplasma noch einige Proteine an die mRNA und der Ribonukleoprotein-Komplex diffundiert zu den Ribosomen.

Zusammenfassung der mRNA-Reifung:

1. 5' Ende Modifikation der mRNA (Capping)
2. 3' Ende Modifikation der mRNA (Polyadenylation)
3. Splicing

---

<sup>1</sup> Ribosom – siehe V. 2 Ribosomen

<sup>2</sup> AMP - Adenosinmonophosphat

## V. Die Translation

Nachdem die mRNA nach der Transkription zu den Ribosomen diffundiert ist, um dort in Proteine übersetzt zu werden, heften sich diese an die mRNA. Die Aminosäuren werden von speziellen Transfer-Molekülen zu den Ribosomen herangetragen, den tRNAs.

Bei der Translation wird die mRNA-Codonsequenz in eine Aminosäuresequenz übersetzt.

### V.1 Transfer-RNA (tRNA)

In jeder Zelle gibt es über 20 verschiedene tRNA-Typen, für jede Aminosäure eine eigene tRNA.

Die tRNAs fungieren als eine Art Übersetzermoleküle, da sie an einem Ende eine spezifische Aminosäure und am anderen ein Anticodon besitzen (Abb. 5.1), das auf ein bestimmtes Codon der mRNA passt. Die Anticodon-Schleife dient zum Abtasten der mRNA und die beiden anderen Schleifen besitzen eine Funktion bei der Anheftung im Ribosom. Die tRNA in Abb. 5.1 würde sich an das Codon UUC der mRNA binden und die Aminosäure Phenylalanin transportieren.

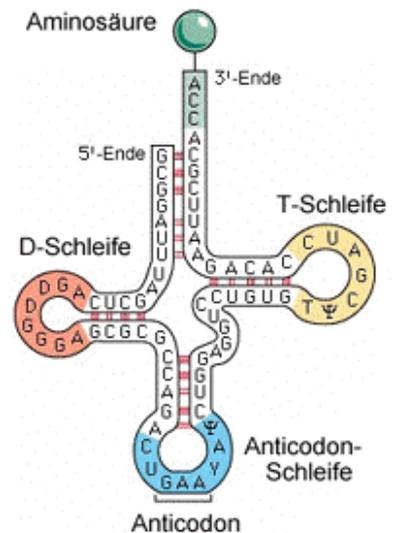


Abb. 5.1 Primärstruktur einer tRNA

### V.2 Ribosomen

Eine Zelle besitzt tausende von Ribosomen. Sie kommen in Eukaryonten frei vor oder gebunden an das endoplasmatische Retikulum<sup>1</sup> (ER) oder die Kernmembran. Auch in Mitochondrien<sup>2</sup> und Chloroplasten<sup>3</sup> findet man sie, allerdings sind sie dort so klein, wie die Ribosomen der Bakterien. Sie bestehen zu  $\frac{2}{3}$  aus ribosomaler RNA (rRNA) und  $\frac{1}{3}$  aus Protein.

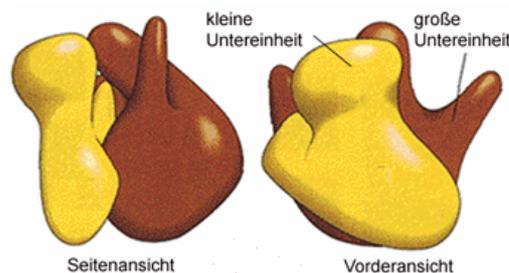


Abb 5.2 Aufbau eines eukaryotischen Ribosoms

<sup>1</sup> endoplasmatisches Retikulum – Röhren- und Zysternensystem im Cytoplasma; dient dem Stofftransport und ist Ort der Proteinsynthese

<sup>2</sup> Mitochondrium – dient der Zelle zur Energiegewinnung

<sup>3</sup> Chloroplast – nur in pflanzlichen Zellen; Ort der Photosynthese

Der typische Aufbau eines eukaryontischen Ribosoms ist in Abb. 5.2 dargestellt. Dabei erkennt man eine große und eine kleine Untereinheit.

An den Ribosomen werden die aminosäurebeladenen tRNAs an die entsprechenden Codons der mRNA angelagert und nebeneinanderliegende Aminosäuren zu einem Polypeptid verknüpft.

### V.3 Ablauf der Translation

Die mRNA enthält am Genanfang ein sogenanntes Starter-Codon und am Ende ein Stop-Codon. Beide sind für den Beginn der Translation am Ribosom und die Beendigung notwendig.

Grundsätzlich läuft die Translation so ab, dass die beiden Untereinheiten der Ribosomen die mRNA umschließen und mit Aminosäuren beladene t-RNAs mit ihrem Anticodon die mRNA Stück für Stück abtasten (Abb. 5.3).

Dieser Vorgang beginnt mit dem Starter-Codon und endet mit dem Terminator-Codon.

Der gesamte Ablauf geschieht in 3 Phasen:

- Initiation
- Elongation und
- Terminierung.

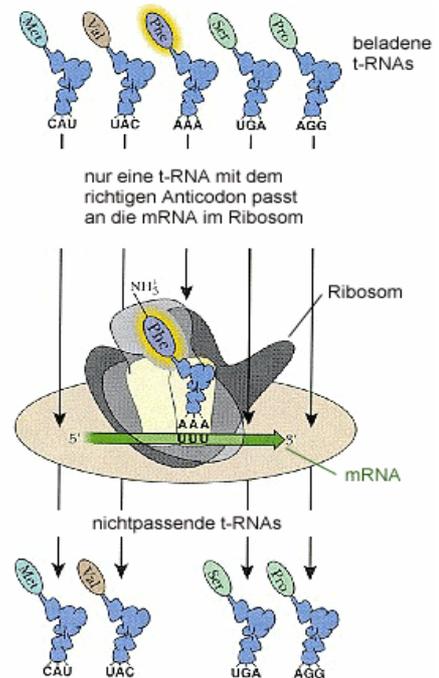


Abb. 5.3 Abtastung der mRNA

#### V.3.1 Initiation

Zu Beginn der Translation bindet sich die kleine Untereinheit des Ribosoms an die mRNA, danach eine Initiator-tRNA, die F-Met-tRNA, an die Stelle, an der sich das Starter-Codon (AUG) befindet. Zwischen dem Anticodon der t-RNA und dem Starter-Codon kommt es mittels Wasserstoffbrückenbindungen zu Basenpaarungen.

Oft binden sich mehrere Ribosomen gleichzeitig hintereinander an eine mRNA, so dass fast parallel Polypeptidketten mit steigender Kettenlänge entstehen. So aufgereihete Ribosomen nennt man Polysomen (Abb 5.4).

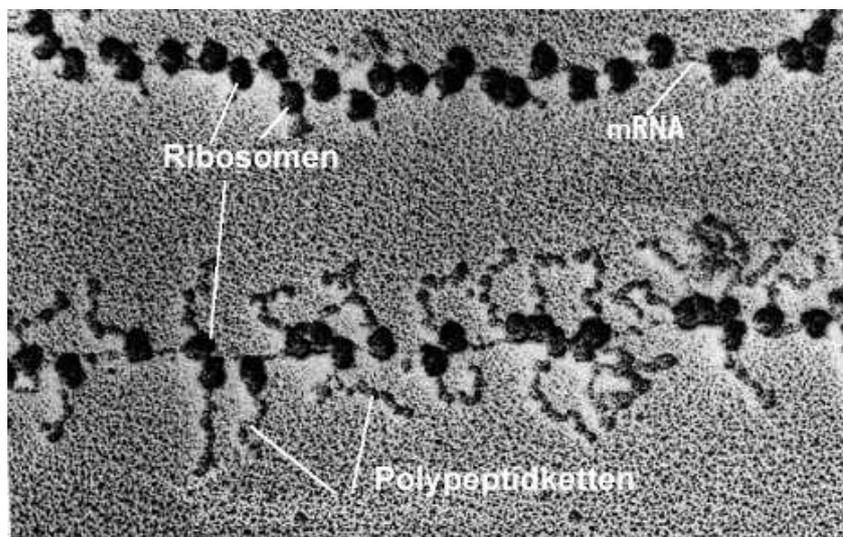


Abb. 5.4 Polysomen an der mRNA unter dem Elektronenmikroskop

Die Initiator-tRNA besitzt als Aminosäure ein am Amminoende formyliertes Methionin, das dafür sorgt, dass die zu knüpfende Peptidbindung nur an der Carboxylgruppe entstehen kann (Abb. 5.5).

Danach vervollständigt die große ribosomale Untereinheit den Initiator-Komplex, indem sie sich an die kleine Einheit anlagert. Die große Untereinheit besitzt 2 Bindungsstellen für t-RNAs:

- den A-Ort (A-Site – Eingang) und
- den P-Ort (P-Site – Ausgang).

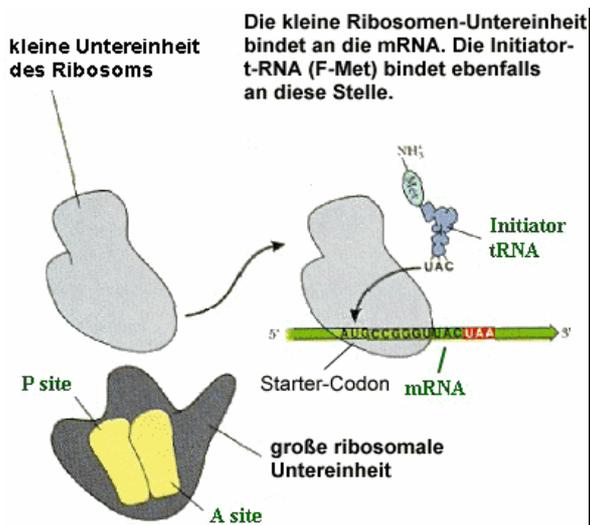


Abb. 5.5 Initiation

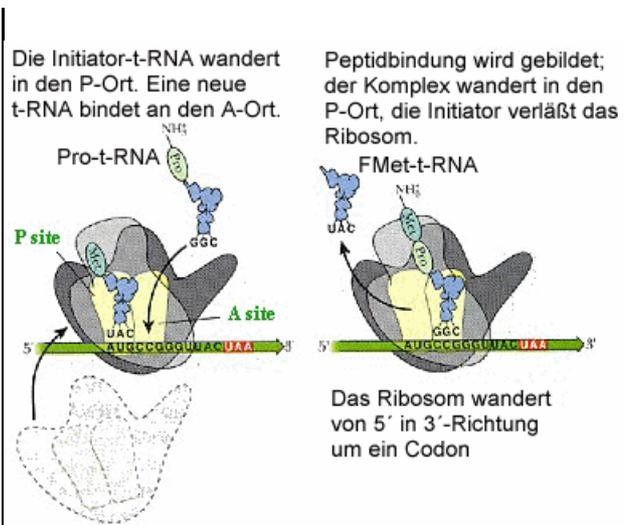


Abb. 5.6 Elongation

### V.3.2 Elongation

Nun wandert der Initiator-tRNA-Komplex ein Codon weiter in den P-Ort (Abb 5.6). Der A-Ort wird frei für die nächste passende tRNA mit dem entsprechenden Anticodon. Nach der Bindung an die mRNA wird zwischen den Aminosäuren der beiden tRNAs eine Peptidbindung geknüpft.

Die unbeladene F-Met-tRNA verläßt das Ribosom und die verbliebene tRNA wandert in den P-Ort. Dieser Vorgang geht in 5'-3'-Richtung weiter. Immer neue passende, beladene t-RNAs binden an den A-Ort, die Peptidbindung wird geknüpft, rutschen ein Codon weiter in den P-Ort usw. bis zum Stop-Codon.

### V.3.3 Terminierung

Am Terminator-Codon (UAA, UAG, oder UGA) bricht die Synthese wegen des Fehlens einer passenden tRNA ab, die beiden Ribosomen-Untereinheiten fallen von der mRNA ab (Abb. 5.7).

Nun liegt das synthetisierte Protein in seiner Primärstruktur vor und kann nun zu höheren Strukturen reifen. Die mRNA wird zerlegt und steht nun wieder für die Transkription zur Verfügung.

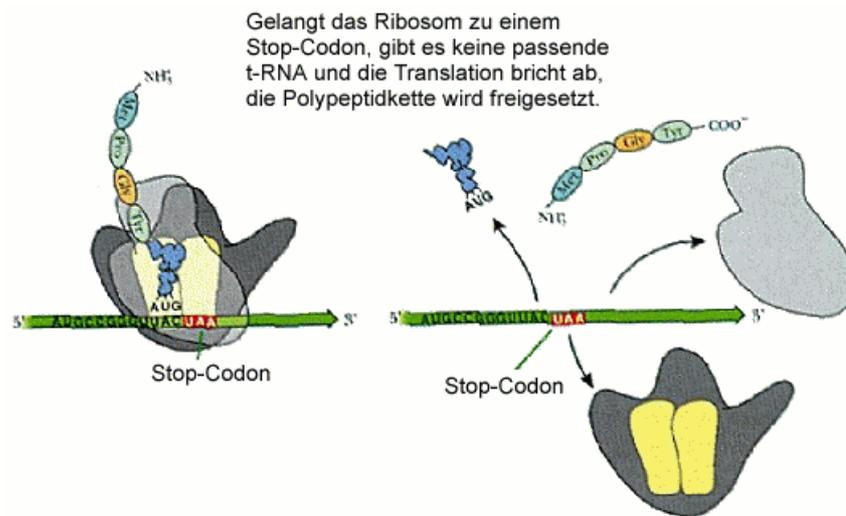


Abb. 5.7 Terminierung

## VI. Die Genregulation

Organismen haben Mechanismen entwickelt, mit denen sie die Gene nach Bedarf regulieren können, da nicht alle Genprodukte ständig gebraucht werden. Gene, die ständig abgelesen werden nennt man konstitutive Gene, sonst regulierte Gene.

Über die Genregulation soll hier nur ein kurzer Überblick gegeben werden.

### VI.1 Das Operon-Modell

Das Operon ist eine Funktionseinheit bestehend aus dem Promotor, dem Operatorgen und den Strukturgenen (Abb. 6.1, Abb. 6.2).

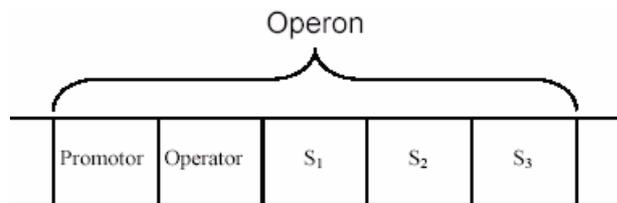


Abb. 6.1 Operon-Modell nach Jacob und Monod

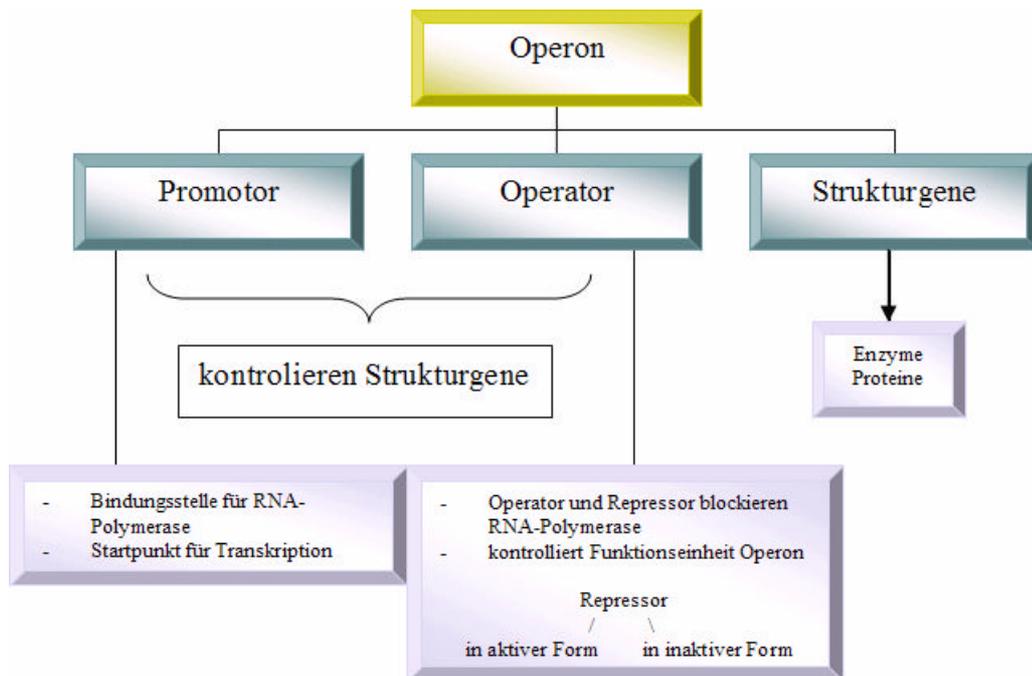
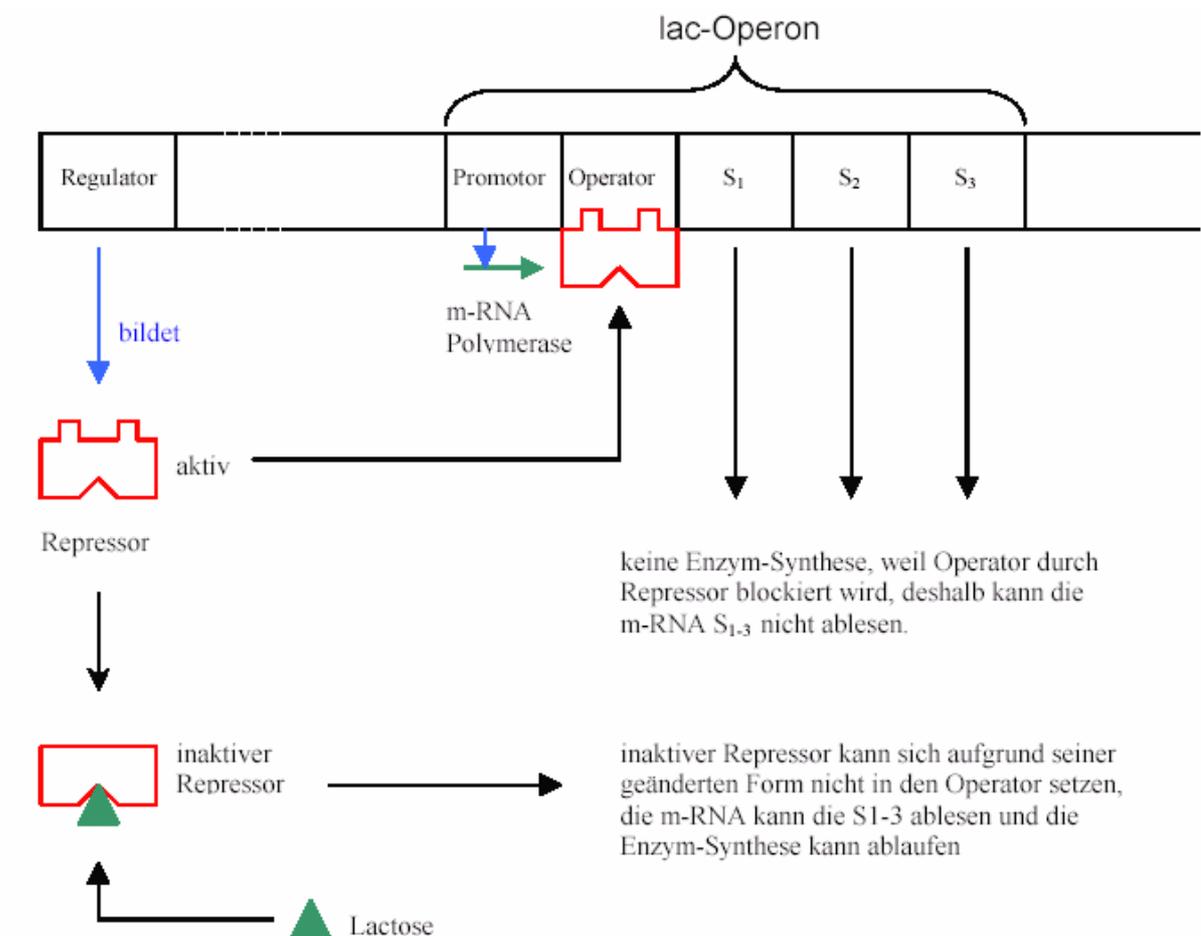


Abb. 6.2 Das Operon

### VI.2 Substratinduktion

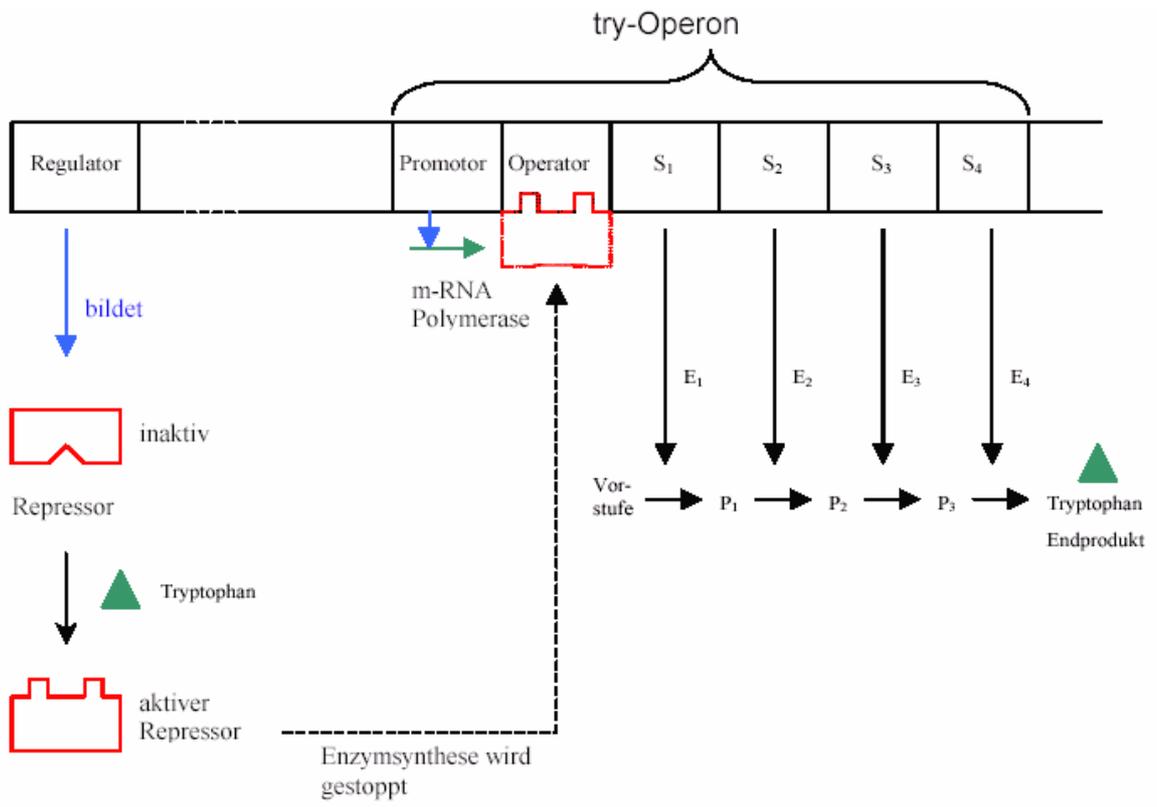
Bei der Substratinduktion wird der am Operator befindliche Repressor, der an anderer Stelle im Genom gebildet wurde, durch das Anlagern eines Substrates inaktiv und die RNA-Polymerase kann nun mit der Transkription der Strukturgenen beginnen. Nachdem die Strukturgenen transkribiert wurden und die entstandene mRNA translatiert, kann das Protein, meist ein Enzym, reifen und das Substrat abbauen. Somit wird der Repressor wieder aktiv und blockiert das Operon am Operator (Abb. 6.3).



**Abb. 6.3** Substratinduktion am Beispiel des lac-Operons  
 Das Substrat Lactose inaktiviert den Repressor, so dass die mRNA-Polymerase die Strukturgene des lac-Operons transkribieren kann und das Enzym Lactase gebildet werden kann, das die Lactose dann abbauen kann.

### VI.3 Endproduktrepression

Bei der Endproduktrepression liegt der Repressor inaktiv vor, so dass die Strukturgene ungehindert transkribiert werden können. Das fertige Genprodukt, also das Protein, ändert dann die Struktur des Repressors und macht diesen somit aktiv. Dadurch wird das Operon am Operatorgen blockiert und die Transkription wird gestoppt. Wird das am Repressorprotein angelagerte Protein durch andere Substanzen verändert (Abbau, Umstrukturierung), geht der Repressor wieder in den inaktiven Zustand und die Synthese beginnt erneut (Abb. 6.4).



**Abb. 6.4** Endproduktrepression am Beispiel des try-Operon  
 Die Strukturgene vom try-Operon werden transkribiert und Tryptophan wird gebildet. Dieses aktiviert jetzt den Repressor, welches wiederum das try-Operon blockiert.

## VII. Genomics<sup>1</sup>

Um die Gene eines Organismus zu untersuchen, muss man folgende Schritte durchführen:

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

### VII.1 Zerlegen der DNA

Die notwendigen Werkzeuge zur Zerlegung natürlicher DNA-Moleküle heißen Restriktionsnukleasen. Diese Enzyme erkennen kurze Nukleotid-Sequenzen, an denen sie einen DNA-Strang schneiden. Bei diesem Schneiden wird immer die gleiche Kollektion von DNA-Stücken produziert, egal, wann und wo ein gegebenes Genom mit einer gegebenen Restriktionsnuklease behandelt wird.

### VII.2 Herstellen einer Genom-Bibliothek

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 mit Hilfe bakteriologischer Verfahren. DNA lässt sich gut in entsprechend vorbehandelte Bakterien übertragen, aber Restriktionsfragmente würden nach Aufnahme in Bakterien schnell abgebaut und verlorengehen. Deswegen müssen die Restriktionsfragmente in Überträger (Vektoren) eingebaut werden.

Gebräuchliche Vektoren sind Plasmiden<sup>2</sup>, Bakteriophagen<sup>3</sup> oder einer Kombination aus beiden.

Die Bakterien mit veränderten Plasmiden oder der injizierten DNA der Bakteriophagen vermehren sich nun auf Nährböden zu Kolonien. Durch die Verteilung der Bakterien in einzelne unabhängige Kolonien gelingt eine Auftrennung des Gemisches von Restriktionsfragmenten der Fremd-DNA in die einzelnen Komponenten. Jede Kolonie besteht aus zahlreichen Bakterien, die alle das identische Stück Fremd-DNA in ihrem Vektor-Plasmid tragen. Dieser Vorgang wird auch als Klonieren bezeichnet und das Restriktionsfragment in einer Kolonie als DNA-Klon.

Jetzt steht dem Molekularbiologen ein in Einzelstücke aufgeteiltes Genom für weitere Untersuchungen zur Verfügung. Man spricht von einer Genom-Bibliothek. Aus dieser Bibliothek kann nun jedes interessante Genom-Stück entnommen und genauer untersucht werden.

#### VII.2.1 cDNA-Bibliotheken

Eine typische Säugetier-Zelle enthält 10.000-30.000 mRNA-Arten, von denen viele selten sind und in zehn oder weniger Kopien/Zelle vorkommen. Eine genaue Untersuchung ihrer Art

---

<sup>1</sup> Genomics – gemeint ist die Untersuchung aller Gene oder einer möglichst großen Anzahl von Genen eines Organismus

<sup>2</sup> Plasmiden – DNA der Bakterien, deren Enden sich zu einem Kreis verbunden haben

<sup>3</sup> Bakteriophagen – Viren der Bakterien, die ihre DNA oder RNA in Bakterien injizieren und sich so vermehren

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 wesentliche dieser Verfahren kann in zwei Aussagen zusammengefasst werden: Erstens, die RNA-Sequenzen werden in DNA-Sequenzen überführt, mit Hilfe des Enzyms reverse Transkriptase. So wird von einer RNA eine DNA-Kopie gemacht: copy-DNA oder kurz: cDNA. Zweitens, können DNAs dann, wie andere DNA-Stücke, in Vektoren eingebaut und kloniert werden. Damit erreicht man eine Trennung des sonst unauflösbaren Gemisches von mRNA und hat die Möglichkeit zur Produktion beliebig großer Mengen von interessanten Sequenzen.

## VII.3 Sequenzieren

Molekularbiologen verwenden hauptsächlich zwei Verfahren zur Bestimmung von DNA-Sequenzen, die chemische Methode von A. M. Maxam und W. Gilbert (1977) und die biochemische Didesoxy- oder Kettenabbruch-Methode von F. Sanger (1977).

### VII.3.1 Didesoxy- oder Kettenabbruch-Methode

Obwohl die Kettenabbruch-Methode umständlich ist, wurde sie von den meisten Experimentatoren, wegen ihrer Verlässlichkeit und Genauigkeit, bevorzugt. Mit Hilfe des Verfahrens wurden mehr als 95% der vielen DNA-Sequenzen gewonnen, die in den internationalen Datenbanken gespeichert sind.

In der Praxis lassen sich auf diese Weise einige hundert Nukleotide einer DNA-Sequenz in einem Experiment bestimmen. Um noch längere DNA-Stücke sequenzieren zu können, muss die Ausgangs-DNA zuerst in handliche Stücke zerlegt werden.

Aufgrund der Komplexität dieses Verfahrens sei an dieser Stelle nur das Endprodukt und dessen Auswertung in Abb. 7.1 gezeigt. Auf der linken Seite ist das Sequenz-Gel zusehen und rechts die computerunterstützte Auswertung. Dabei wurde die Sequenz in Doppelstrang-Form ausgedruckt und die kodierte Aminosäure-Sequenz (rot) abgeleitet.

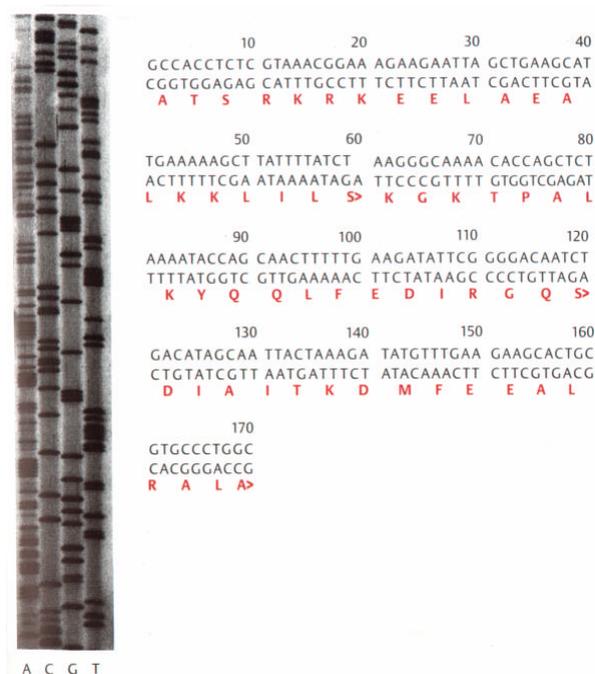


Abb. 7.1 Anwendung der Kettenabbruch-Methode

## VII.4 Monitoring von Genexpressionen mittels DNA-Microarrays

DNA-Chips – auch als Gen-Chips oder DNA-Microarrays bezeichnet – wurden in den frühen 90er-Jahren des 20. Jahrhunderts entwickelt. Die Firma Affymetrix, Inc. in Santa Clara, Kalifornien USA, hatte die Idee, tausende DNA-Sonden auf Glas-Mikrochips aufzubringen, ähnlich wie Transistoren auf Silikon. Dazu wurden die Methoden zur Erzeugung von Computer-Chips neu adaptiert. Affymetrix produziert zur Zeit mehr als 100.000 Chips pro Jahr.

Alle DNA-Chips und -Microarrays sind nicht anderes als eine geordnete Sammlung von DNA-Molekülen von bekannter Sequenz (= Array). So ein Array ist gewöhnlich in Form eines Rechtecks oder Quadrates angeordnet. Es kann aus nur einigen hundert aber auch aus einigen zehntausend Einheiten bestehen (z.B.: 60x40, 100x100 oder 300x500). Jede Einheit ist ein örtlich genau definierter Punkt auf der Glasoberfläche mit einem Durchmesser von weniger als 200µm. Sie enthält Millionen von Kopien eines genau definierten, kurzen DNA-Stückes. Im Computer ist die Information, wo sich welche DNA in einem Array befindet, abrufbar.

Die so auf den Chips angebrachten DNA-Stücke (DNA-Sonden) binden nun ein komplementäres DNA-Fragment aus der Probelösung. Jede DNA-Sonde kann aus einem komplexen Gemisch aus Millionen von verschiedenen DNA-Molekülen genau jenes herausfinden, das genetisch perfekt übereinstimmt (ein Prozess, den man als Hybridisierung<sup>1</sup> bezeichnet). Die DNA-Fragmente bleiben daher auf einem genau definierten Punkt des DNA-Chips fixiert. Die DNA-Fragmente werden vor dem Versuch mit fluoreszierenden Farbstoffen versehen, so dass diese Punkte unter dem Laserlicht aufleuchten und so leicht nachgewiesen werden können. So erhält man dann Bilder von DNA-Arrays, wie in Abb. 7.2 zu sehen ist.

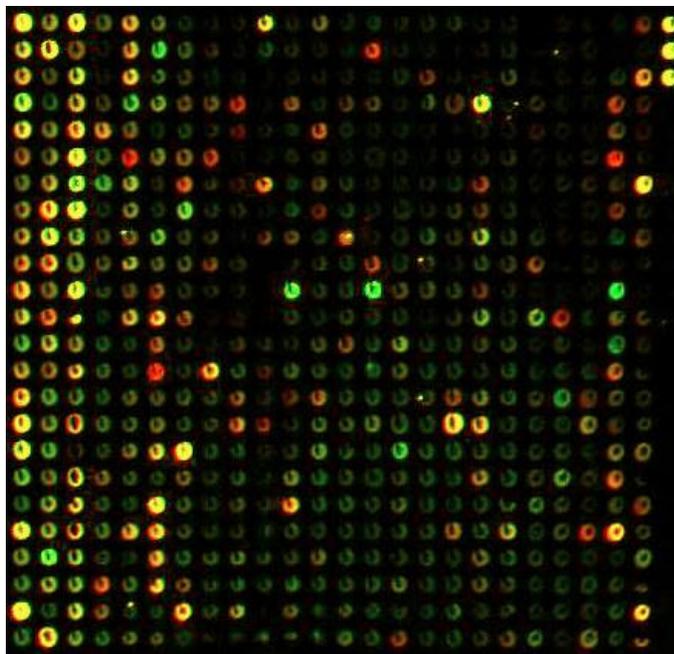


Abb. 7.2 Mit einem Scanner ausgewertetes Array

<sup>1</sup> Hybridisierung - zwei einzelsträngige Nukleinsäuremoleküle bilden Basenpaarungen miteinander aus

Ein typischer Arbeitsablauf bei einem Arraying-Experiment, besteht meist aus den folgenden Schritten:

- Herstellung des Microarrays (manuell, mit einem Arrayer oder bestellen bei einer dafür spezialisierten Firma)
- Präparation der mRNA einer Zellkultur und Umwandlung in cDNA. Die hergestellte cDNA wird mit einem Fluoreszenzfarbstoff markiert
- Hybridisierung der cDNA mit den DNA-Sonden auf dem Microarray
- Auswertung mit einem Laserscanner

Ein Beispiel ist eine Untersuchung, die feststellt, welche Gene in menschlichen Brustkrebs-Zellen und welche in menschlichen Melanom-Zellen angeschaltet sind. Dazu wird zunächst die mRNA aus beiden Zelltypen isoliert und in cDNA umgewandelt und mit Fluoreszenzfarbstoffen markiert, die Brustkrebs-cDNA mit einem rot und die Melanom-cDNA mit einem grün fluoreszierenden Farbstoff. Dann wird die cDNA mit dem Array in Kontakt gebracht, es erfolgt die Hybridisierung. Die DNA-Sonden auf den Microarray-Punkten binden jeweils die cDNA-Moleküle aus dem Gemisch, die genetisch identisch sind. Mit einem Laser wird der Chip abgefahren und ausgewertet. Die Stellen, die cDNAs aus den Brustkrebszellen gebunden haben, leuchten rot auf. Die Stellen mit gebundenen Melanom-cDNAs fluoreszieren grün. Stellen, die cDNAs aus beiden Tumorzellen enthalten, leuchten in der Mischfarbe gelb.

Ein weiteres Beispiel ist in Abb. 7.3 zusehen.

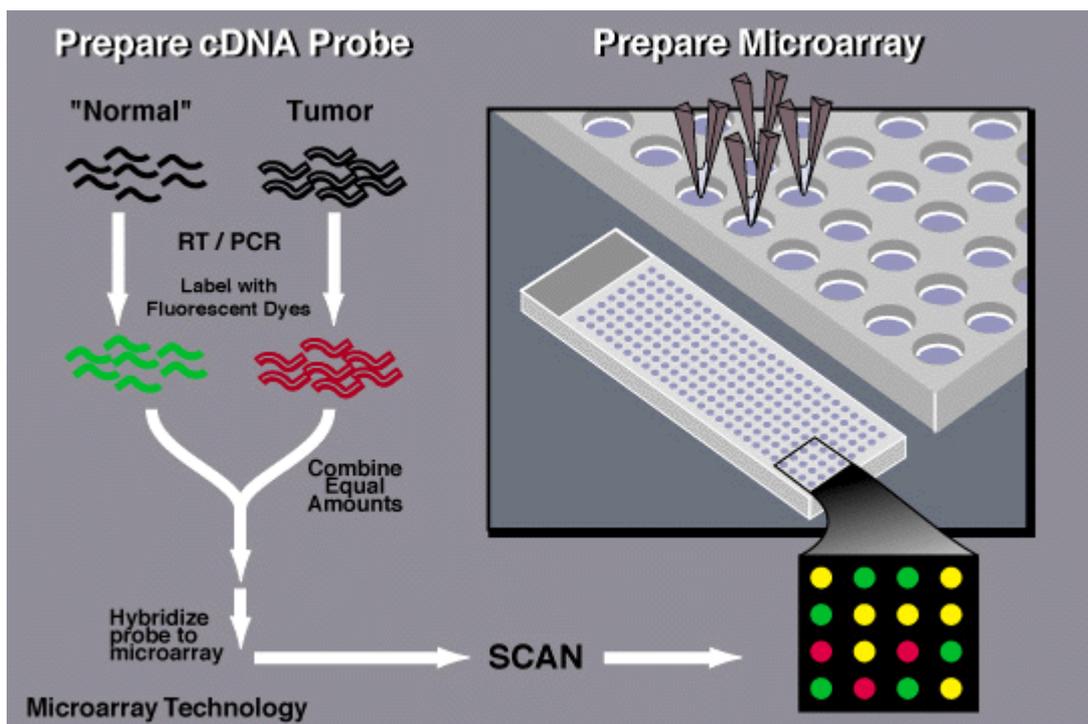


Abb. 7.3 Microarray Technologie

## *Zusammenfassung*

Die Genforschung ist bei weitem kein Gebiet, auf dem schon alles erforscht wurde, sondern wir befinden uns immer noch am Anfang. Wir haben das menschliche Genom „nur“ entschlüsselt, seine Expression aber ist uns noch weitgehend unbekannt.

Die Technologie wird weiterhin Einzug in die Biowissenschaften halten und damit verbunden, wird die Informationsverarbeitung eine stärkere Rolle spielen. Neben vielen anderen Teilen der Informatik werden Datenbanken mehr gebraucht werden denn je, um die Fülle an Informationen zu speichern und zu verwalten, aufzuarbeiten und sich für weiterer Untersuchungen zu nutze zu machen.

# Quellen

## Literaturverzeichnis

Knippers, Rolf: Molekulare Genetik. Georg Thieme Verlag, Stuttgart 1995

Molekulare Genetik

<http://www.drd.de/helmich/bio/gen/reihe2/index.html>

Molekulargenetik

<http://www.biokurs.de/skripten/13/bs13-1.htm>

Molekulare Genetik

<http://www.hausarbeiten.de/archiv/biologie/bio-molgen.pdf>

DNA-Chips und Microarrays

<http://www0.eduhi.at/humangenetik/Projekte/4/Start.htm>

## Bildverzeichnis

Titelbild:

The National Human Genome Research Institute

<http://www.nhgri.nih.gov/>

Abb. 1.1, Abb. 1.2, Abb. 1.3, Abb. 1.4, Abb. 1.5, Abb. 1.6, Abb. 1.7, Abb. 1.8, Abb. 2.2, Abb. 2.3, Abb. 2.4, Abb. 2.5, Abb. 2.6, Abb. 5.1:

Knippers, Rolf: Molekulare Genetik. Georg Thieme Verlag, Stuttgart 1995

Abb. 2.1:

DNA Structure

<http://www.blc.arizona.edu/INTERACTIVE/DNA3/dnastruct.html>

Abb. 3.1:

Molekulare Genetik

<http://www.drd.de/helmich/bio/gen/reihe2/index.html>

Abb. 3.1, Abb. 3.2, Abb. 3.4, Abb. 5.2, Abb. 5.3, Abb. 5.4, Abb. 5.5, Abb. 5.6, Abb. 5.7:  
Molekulargenetik

<http://www.biokurs.de/skripten/13/bs13-1.htm>

Abb. 6.1, Abb. 6.3, Abb. 6.4, Abb. 6.5:

Molekulare Genetik

<http://www.hausarbeiten.de/archiv/biologie/bio-molgen.pdf>

Abb. 7.1:

DNA-Chips und Microarrays

<http://www0.eduhi.at/humangenetik/Projekte/4/Start.htm>

Abb. 7.2

DNA Microarray (Genome Chip)

<http://www.gene-chips.com/>