

# Informationsmakromoleküle

# 1

## In diesem Kapitel ...

... geht es um diese Themen:

- Informationsverarbeitung und Molekularbiologie
- Nukleinsäurestruktur und -funktion
- Proteinstruktur und -funktion

## 1.1 Informationsverarbeitung und Molekularbiologie

### 1.1.1 Das zentrale Dogma

Die Molekularbiologie befasst sich mit den molekularen Wechselwirkungen, die den biologischen Funktionen zugrunde liegen. Sie überschneidet sich beträchtlich mit der Biochemie und der Genetik, und sie scheint sich hauptsächlich mit den strukturellen Grundlagen und der Kontrolle der Informationsverarbeitung in der Zelle zu beschäftigen sowie mit den für deren Untersuchung erforderlichen Technologien. Durch die Pionierarbeiten von Avery, MacLeod und McCarty sowie von Hershey und Chase in den 1940er- und 1950er-Jahren wurde klar bewiesen, dass die genetischen Anweisungen zur Erschaffung einer Zelle im Zellkern sitzen, innerhalb einer linearen Sequenz von **Basen**; diese wiederum sind Bestandteil der Struktur eines langen chemischen Polymers, der **Desoxyribonukleinsäure (DNA)**. Im Jahre 1953 schlugen dann Crick und Watson die berühmte Doppelhelixstruktur der DNA vor, die genau darlegte, wie diese Information gespeichert und an nachfolgende Generationen weitergegeben wird. Um zu erklären, wie Zellen die im **DNA-Genom** verschlüsselten Anweisungen verwenden, postulierte Crick, dass der Fluss der genetischen Information in nur eine Richtung verläuft: von der DNA über eine zwischengeschaltete Nukleinsäure, die **Ribonukleinsäure (RNA)**, zum **Protein** – d. h. „DNA macht RNA macht Protein“. Diese Aussage wurde zum **zentralen Dogma** der Molekularbiologie, ohne dass die einzelnen Schritte groß bewiesen wurden. Wir wissen heute, dass diese Aussage des zentralen Dogmas weitgehend korrekt ist, auch wenn das ursprüngliche Schema inzwischen mehrmals modifiziert worden ist. Abb. 1.1 zeigt ein Diagramm dieses Informationsflusses. Der Hauptweg führt von der DNA über die RNA zum Protein, und man

2 | 1 Informationsmakromoleküle

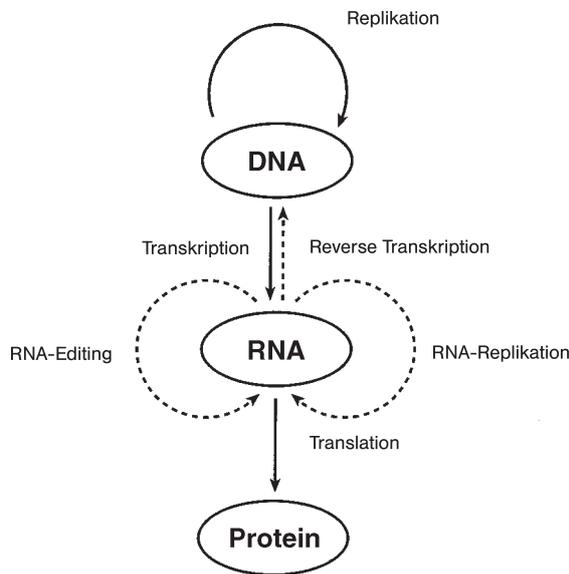


Abb. 1.1 Der Fluss der genetischen Information.

weiß heute, dass dies auch für die DNA in den kleinen unabhängigen Genomen der Mitochondrien und Chloroplasten gilt. In allen Zellen wird die DNA konzeptionell (aber nicht physikalisch) in diskrete codierende Einheiten (**Gene**) eingeteilt, die die Information für die einzelnen Proteine enthalten. Diese DNA wird transkribiert (Kap. 5 und 7), sodass RNA-Moleküle entstehen (**Boten- oder messenger-RNA, mRNA**), die die gleiche Sequenzinformation wie die DNA enthalten; sie können als Arbeitskopien der Gene angesehen werden, die in der Haupt-DNA-Blaupause vorhanden sind. Diese mRNAs werden dann entsprechend dem **genetischen Code** (Abschnitt 8.1) in Aminosäuresequenzen von Proteinen **translatiert** (übersetzt) (Kap. 9). Die Kombination all dieser Prozesse, die erforderlich sind, um die Information der DNA zu entschlüsseln und ein funktionsfähiges Molekül zu erzeugen, heißt **Genexpression**. Wir können auch die **DNA-Replikation** (Kap. 3) in Abb. 1.1 mit einschließen, bei der durch Verdoppelung der Information in der Ausgangs-DNA zwei Tochter-DNA-Moleküle gebildet werden; dies hat den Informationsfluss und die Bewahrung der Information von einer Generation zur nächsten zur Folge.

Man hat jedoch einige Ausnahmen von diesem Grundschemata identifiziert. Viele RNA-Moleküle werden nicht in ein Protein translatiert, sondern funktionieren eigenständig als RNAs (Abschnitt 9.4). Ihre Gene werden als **RNA-Gene** bezeichnet. Eine Reihe von Virusklassen besitzt keine DNA, sondern enthält ein Genom, das aus einem oder mehreren RNA-Molekülen besteht. Bei den **Retroviren**, zu denen auch das **menschliche Immunschwächevirus (HIV, human immunodeficiency virus)** gehört, der Verursacher des erworbenen Immunschwäche-syndroms (AIDS, *acquired immunodeficiency syndrome*), wird das einzelsträngige RNA-Molekül

in eine doppelsträngige DNA-Kopie überführt; diese wird dann in das Genom der Wirtszelle eingebaut. Dieser Vorgang wird als **reverse Transkription** bezeichnet. Man kennt auch eine Reihe von Viren, deren RNA-Genom direkt zu RNA kopiert wird, ohne Zuhilfenahme der DNA als Zwischenstufe (**RNA-Replikation**). Dazu zählen das Influenza- und das Hepatitis-C-Virus. Soweit bekannt, gibt es keine Beispiele dafür, dass ein Protein „rückwärts translatiert“ wird, um eine spezifische RNA- oder DNA-Sequenz zu erzeugen; somit scheint der Translationsschritt des zentralen Dogmas in nur eine Richtung zu gehen. Schließlich gibt es noch eine faszinierende Ausnahme von dem Dogma, dass die RNA- und Proteinsequenzen eindeutig in der DNA verschlüsselt (codiert) sind: der Prozess des **RNA-Editing**. Man kennt Beispiele (hauptsächlich in Eukaryoten), bei denen die Basensequenz einer RNA tatsächlich nach der Transkription der DNA verändert wird, sodass sie, und jedes Proteinprodukt im Falle einer mRNA, nicht mehr exakt der DNA entspricht.

Die Erörterung dieser Systeme im vorliegenden Buch basiert auf dem biologischen **Klassifizierungssystem der drei Domänen**, bei dem der **letzte gemeinsame Vorfahre (LUCA, last universal common ancestor)** allen Lebens sich zunächst in die Bakterien (**Bacteria**) und den gemeinsamen Vorfahren der **Archaea** und **Eukarya** aufspaltete. Die beiden Letztgenannten trennten sich später auf. Zwar sind Bakterien und Archaeen beide **Prokaryoten**, weil ihnen ein echter Zellkern fehlt, hinsichtlich vieler Aspekte der Informationsverarbeitung haben die Archaeen jedoch mehr Gemeinsamkeiten mit den kernhaltigen **Eukaryoten**. Die meisten Beispiele stammen von Bakterien und Eukaryoten.

### 1.1.2 Rekombinante DNA-Technologie

In den späten 1970er-Jahren erlebte die Molekularbiologie große Fortschritte durch die Entwicklung der **rekombinanten DNA-Technologie (Gentechnik)**. Sie erlaubte, dass Gene isoliert, sequenziert, modifiziert und von einem Organismus auf einen anderen übertragen werden; sie war von größter Bedeutung für das zunehmende Verständnis darüber, wie Zellen arbeiten. Zudem werden auf diese Weise erzeugte **transgene** Mikroorganismen heute routinemäßig eingesetzt, um menschliche Therapeutika im Großmaßstab herzustellen. Transgene Tiere und Pflanzen verfügen über ein großes Potenzial, sowohl die Spannbreite nützlicher Produkte zu vergrößern als auch verbessertes Wachstum, Krankheitsresistenz oder Modelle für menschliche Krankheiten usw. zu erreichen (Abschnitt 13.5). Die dauerhafte Korrektur einer Erbkrankheit mithilfe der **Gentherapie** ist jetzt ebenfalls eine realistische Möglichkeit. In den letzten Jahren ist die für die Bestimmung der DNA-Basen verantwortliche Technologie vorangeschritten und die Kosten sind so rasch gesunken, dass es schon bald praktikabel sein wird, das vollständige Genom eines Individuums zu sequenzieren und die Krankheitsanfälligkeit im Rahmen eines routinemäßigen Gesundheitsfürsorgeprogramms zu bestimmen. Inzwischen können sogar neue Gene chemisch synthetisiert und zu vollständigen Genomen zusammengefügt werden. Im Jahre 2010 bildeten J. Craig Venter und seine Kollegen die vollständige chromosomale DNA eines kleinen Mykoplasma-

#### 4 | 1 Informationsmakromoleküle

Bakteriums nach und fügten sie in eine „leere“ Zelle ein, deren eigenes Chromosom entfernt worden war; auf diese Weise schufen sie einen lebenden Organismus (Abschnitt 13.7). Dieses DNA-Molekül besaß auch einige neue Eigenschaften und ebnete damit den Weg für die zukünftige Möglichkeit, echtes **synthetisches Leben** zu erschaffen – „Designer“-Organismen mit künstlichen Genomen, die neue biochemische Funktionen ausführen können, die in der natürlichen Welt nicht vorkommen. Ziel ist dabei die Herstellung neuer Medikamente, Brennstoffe und anderer Produkte. Die Molekularbiologie und die um sie entstandenen Technologien haben eine zentrale Rolle bei der Entwicklung von Medikamenten für Mensch und Tier, in der Landwirtschaft und in der biotechnologischen Industrie gespielt; nun werden sie darauf angesetzt, die Herausforderungen der weltweiten Gesundheit, der Umweltveränderungen und der Lebensmittelsicherheit zu bewältigen, mit denen wir im 21. Jahrhundert konfrontiert sind.



#### Noch einmal in Kürze

##### Das zentrale Dogma

Das zentrale Dogma besagte ursprünglich: „DNA macht RNA macht Protein.“ Dies geschieht mithilfe der Transkription bzw. der Translation. Dies stimmt im Wesentlichen, obwohl es eine Reihe von Beispielen gibt, die Teilen davon widersprechen. Retroviren schreiben RNA zurück in DNA, andere Viren können RNA direkt zu einer RNA-Kopie replizieren, wohingegen manche RNAs nach ihrer Synthese editiert werden können, sodass die sich ergebende Sequenz nicht direkt durch die DNA-Sequenz spezifiziert ist.

##### Rekombinante DNA-Technologie (Gentechnik)

Die Fähigkeit, die Genome von Mikroorganismen, Tieren und Pflanzen zu manipulieren, hat große Fortschritte für das Verständnis der Zellbiologie gebracht. Außerdem haben transgene Organismen, die DNA von anderen Quellen enthalten, viele Anwendungen in der Medizin, der Landwirtschaft und der Industrie gefunden. Die Fähigkeit, neue Genome zu synthetisieren, wird noch größere Fortschritte auf diesen Gebieten mit sich bringen.



#### Tipp

##### Verwandte Themen:

- (Abschnitt 1.2) Nukleinsäurestruktur und -funktion
- (Abschnitt 1.3) Proteinstruktur und -funktion
- (Kap. 5) Transkription in Bakterien
- (Kap. 7) Transkription in Eukaryoten
- (Kap. 8) Der genetische Code und die tRNA
- (Kap. 9) Proteinsynthese

## 1.2 Nukleinsäurestruktur und -funktion

### 1.2.1 Basen

Die Basen der DNA und RNA sind heterozyklische (kohlenstoff- und stickstoffhaltige) aromatische Ringe, mit einer Reihe von Substituenten (Abb. 1.2). Bei Adenin (A) und Guanin (G) handelt es sich um **Purine**, bityklische Strukturen mit zwei fusionierten Ringen; dagegen sind Cytosin (C), Uracil (U) und Thymin (T) **Pyrimidine** mit nur einem Ring. In der DNA ist die Base Uracil, die in der RNA vorkommt, durch Thymin ersetzt. Thymin unterscheidet sich von Uracil nur bezüglich einer Methylgruppe in der 5-Position, d. h. Thymin ist ein 5-Methyluracil.

### 1.2.2 Nukleoside

In den Nukleinsäuren sind die Basen kovalent mit einem Pentosezuckerring an der 1'-Position verknüpft und bilden dabei ein **Nukleosid** (Abb. 1.3). Bei RNA ist der Zucker eine **Ribose**, bei DNA eine **2'-Desoxyribose**, bei der die Hydroxylgruppe an der 2'-Position durch ein Wasserstoffatom ersetzt ist. Die Verknüpfung mit der Base findet an der 1-Position (*N*-1) der Pyrimidine und an der 9-Position (*N*-9) der Purine statt (Abb. 1.2). Die Kennzahl der Atome im Ribosering wird mit 1', 2', usw. bezeichnet, einfach nur um sie von den Atomen der Base zu unterscheiden. Die Bindung zwischen den jeweiligen Basen und Zuckern ist eine **glykosidische** oder **Glykosidbindung**. Handelt es sich bei dem Zucker um Ribose, dann heißen die Nukleoside (technisch **Ribonukleoside**) Adenosin, Guanosin, Cytosin und Uridin. Ist der Zucker aber Desoxyribose (wie in der DNA), dann sind die

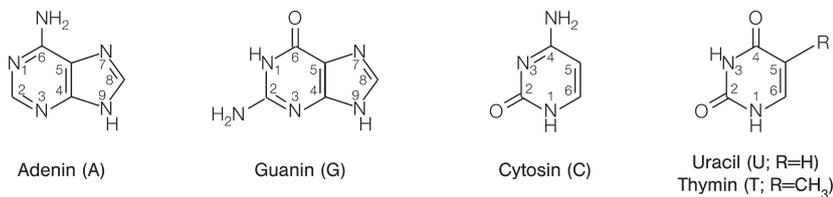


Abb. 1.2 Nukleinsäurebasen.

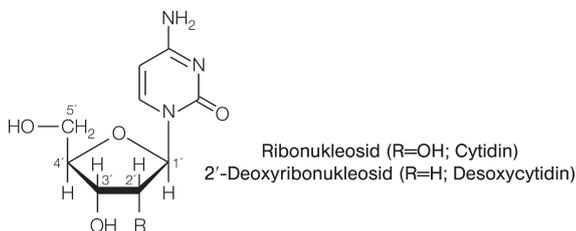


Abb. 1.3 Nukleoside.

## 6 | 1 Informationsmakromoleküle

Nukleoside (**2'-Desoxyribonukleoside**) Desoxyadenosin, Desoxyguanosin usw. Die Bezeichnungen „Thymidin“ und „Desoxythymidin“ können alternativ verwendet werden.

### 1.2.3 Nukleotide

Ein **Nukleotid** ist ein Nukleosid mit einer oder mehreren Phosphatgruppen, die kovalent an der 3'-, 5'- oder (nur in manchen **Ribonukleotiden**) der 2'-Position verknüpft sind. Ist der Zucker Desoxyribose, dann heißen die Verbindungen **2'-Desoxyribonukleotide** oder einfach **Desoxyribonukleotide** (Abb. 1.4). Chemisch gesehen sind die Verbindungen Phosphatester. Im Falle der 5'-Position können bis zu drei Phosphate verknüpft sein und bilden dann z. B. Adenosin-5'-triphosphat oder Desoxyguanosin-5'-triphosphat; die genannten Verbindungen werden üblicherweise mit ATP bzw. dGTP abgekürzt. Auf die gleiche Weise erhalten wir Desoxycytidintriphosphat (dCTP), Uridintriphosphat (UTP) und Desoxythymidintriphosphat (dTTP oder einfach TTP genannt). 5'-Mono- und -Diphosphate werden beispielsweise als AMP bzw. dGDP abgekürzt. Nukleosid-5'-triphosphate (NTPs) oder Desoxynukleosid-5'-triphosphate (dNTPs) sind die Bausteine der polymeren Nucleinsäuren. Im Verlauf der DNA- oder RNA-Synthese werden zwei Phosphate in Form von Pyrophosphat abgespalten und es verbleibt ein Phosphat pro Nukleotid, das in die Nucleinsäurekette eingebaut wird (Abschnitt 3.1 und 5.1). Damit ist die sich wiederholende Einheit einer DNA- oder RNA-Kette ein Nukleotid.

### 1.2.4 Phosphodiesterbindungen

In einem DNA- oder RNA-Molekül sind die Desoxyribonukleotide bzw. die Ribonukleotide durch die kovalente Verknüpfung einer Phosphatgruppe mit der 5'-Hydroxylgruppe einer Ribose und der 3'-Hydroxylgruppe der nächsten Ribose zu einem Polymer verbunden (Abb. 1.5). Eine derartige Verknüpfung nennt man **Phosphodiesterbindung**, da das Phosphat chemisch als ein Diester vorliegt. Damit besitzt eine Nucleinsäure eine Richtung oder **Polarität**. Eine Nucleinsäurekette egal welcher Länge (solange sie nicht ringförmig ist, Abschnitt 2.3) hat ein freies 5'-Ende – das mit einer Phosphatgruppe verknüpft sein kann oder auch nicht –

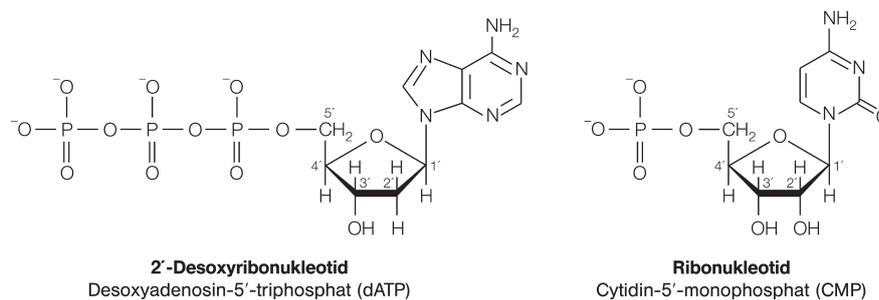


Abb. 1.4 Nukleotide.

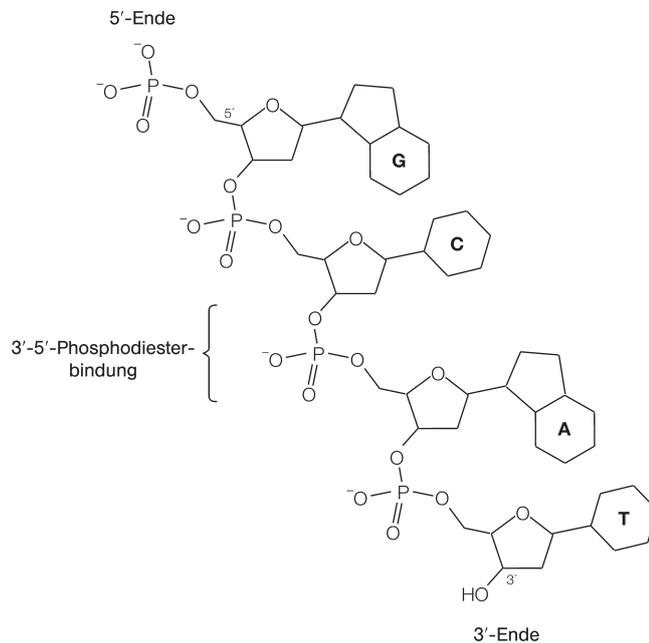


Abb. 1.5 Phosphodiesterbindungen und die kovalente Struktur eines DNA-Strangs.

und ein freies 3'-Ende – das zumeist eine freie Hydroxylgruppe ist. Bei neutralem pH-Wert hat jede Phosphatgruppe eine einzelne negative Ladung. Aus diesem Grund heißen Nukleinsäuren „Säuren“; sie sind die Anionen starker Säuren. Damit sind Nukleinsäuren **stark negativ geladene Polymere**.

### 1.2.5 DNA/RNA-Sequenz

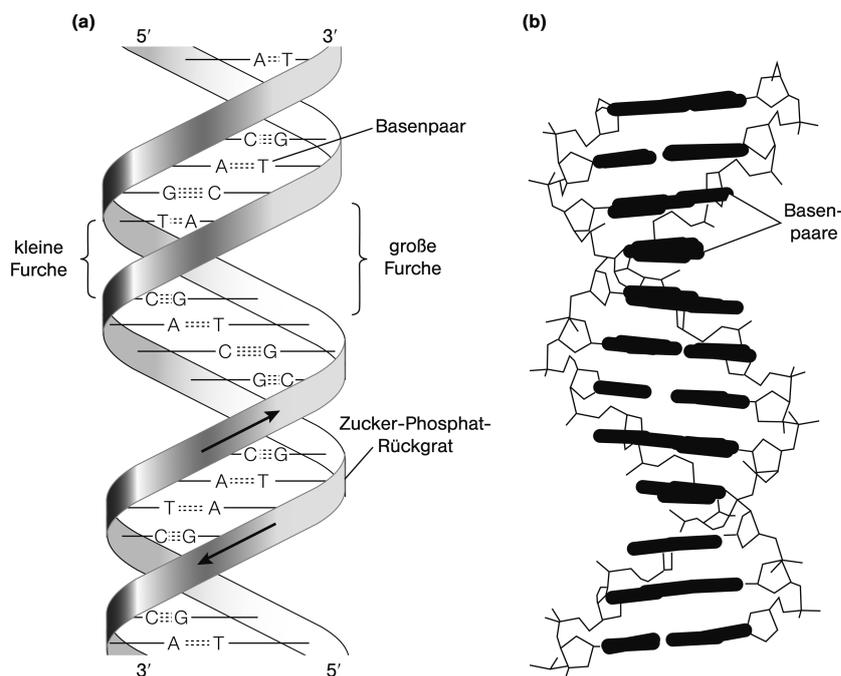
Die sich wiederholenden Monomere der DNA oder RNA werden vereinbarungsgemäß durch ihre Anfangsbuchstaben, A, T, G, C oder U, wiedergegeben. Zudem ist vereinbart, die Sequenzen so zu schreiben, dass sich das 5'-Ende links befindet. Somit kann man eine DNA-Sequenz z. B. als 5'-ATAAGCTC-3' oder einfach ATAAGCTC schreiben. Eine RNA-Sequenz könnte lauten: 5'-AUAGCUUGA-3'. Man beachte, dass die Ausrichtung der Kette bedeutet, dass z. B. ATAAG nicht dasselbe ist wie GAATA.

### 1.2.6 DNA-Doppelhelix

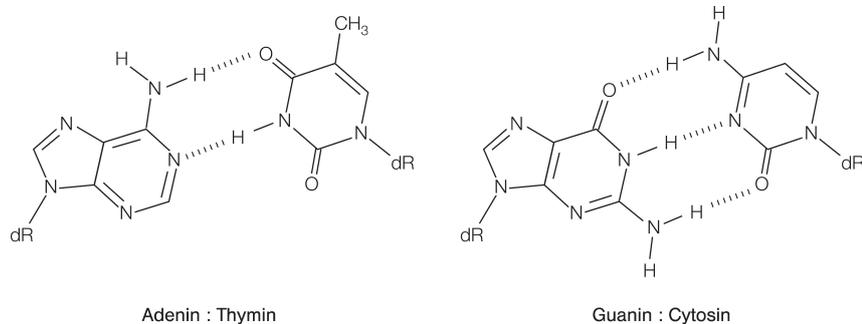
DNA kommt in der Natur zumeist in Form der gut bekannten **Doppelhelix** vor. Die grundlegenden Eigenschaften dieser Struktur wurden im Jahre 1953 von James Watson und Francis Crick abgeleitet. Zwei getrennte DNA-Ketten winden sich umeinander, wobei jede eine Wendel (Spirale) bildet; daraus resultiert eine **rechtsgängige** Doppelhelix (Abb. 1.6a). Das negativ geladene Zucker-Phosphat-Rückgrat der Moleküle befindet sich außen und die planaren Basen jedes Strangs stapeln

8 | 1 Informationsmakromoleküle

sich in der Mitte der Helix übereinander (Abb. 1.6b). Zwischen den Rückgratsträngen verlaufen die **kleine** und die **große Furche**, die ebenfalls einer helikalen Bahn folgen. Die Basen der sich gegenüberliegenden Stränge bilden miteinander Wasserstoffbrückenbindungen aus und verknüpfen damit die beiden Stränge nichtkovalent; auf diese Weise entstehen **Basenpaare (Bp)**. Auf eine Windung der DNA-Doppelhelix kommen etwa 10 Bp. Die beiden Stränge sind hinsichtlich der 5'→3'-Richtung entgegengesetzt ausgerichtet (**antiparallel**) und, was entscheidend ist, sie haben zueinander **komplementäre** Sequenzen. Die letztgenannte Eigenschaft kommt zustande, weil die Basenstrukturen und die Zwänge des DNA-Rückgrats vorschreiben, dass die Basen in Form von Purin-Pyrimidin-Paaren miteinander Wasserstoffbrückenbindungen ausbilden (Abschnitt 1.3); diese Paare besitzen eine sehr ähnliche Geometrie und Ausdehnung (Abb. 1.7). Guanin paart mit Cytosin (drei Wasserstoffbrückenbindungen), und Adenin paart mit Thymin (zwei Wasserstoffbrückenbindungen). So kann jede beliebige Sequenz innerhalb der regulären doppelsträngigen DNA-Struktur untergebracht werden. Die Sequenz des einen Strangs definiert eindeutig die Sequenz des anderen, und Watson und Crick erkannten schnell, dass dieser Umstand einen offensichtlichen Replikationsmechanismus für die DNA impliziert (Abschnitt 3.1). Natürlich beruht auch der Mechanismus der Transkription der DNA-Sequenz in eine RNA (Abschnitt 5.1).



**Abb. 1.6** Die DNA-Doppelhelix. (a) Schematische Darstellung der Struktur; (b) detailliertere Darstellung der Struktur: Die Stapelung der Basenpaare (fett) ist hervorgehoben.



**Abb. 1.7** Die DNA-Basenpaare. Die Wasserstoffbrückenbindungen sind durch gestrichelte Linien dargestellt; dR = Desoxyribose.

### 1.2.7 A-, B- und Z-Helices

In der Tat hat man eine Reihe verschiedener Formen der Nukleinsäuredoppelhelix beobachtet und untersucht; alle haben das Grundmuster von zwei helikal gewundenen antiparallelen Strängen. Die von Watson und Crick identifizierte und oben beschriebene Struktur wird als **B-DNA** bezeichnet (Abb. 1.8a). Man glaubt, dass sie die idealisierte Strukturform ist, die praktisch von jeder DNA *in vivo* angenommen wird. Sie ist charakterisiert durch 10 Bp/Windung, durch die Anwesenheit von Basenpaaren, die auf der Helixachse und nahezu senkrecht zu ihr liegen, und durch die vorhandenen, klar definierten, tiefen großen und kleinen Furchen. Die eigentlichen DNA-Sequenzen haben genau genommen eine helikale Wiederholung, die näher bei 10,5 Bp liegt, und eine Vielzahl anderer struktureller Verformungen, die von der exakten Sequenz abhängen.

Bei geringer Feuchtigkeit kann die DNA dazu gebracht werden, eine alternative Helix, die **A-Form**, zu bilden (Abb. 1.8b). Die A-Form ist rechtsgängig, wie die B-Form, besitzt aber eine breitere, stärker komprimierte Struktur, bei der die Basenpaare zur Helixachse hin gekippt sind und sogar außerhalb der Achse liegen (vom Ende her gesehen hat die A-Helix ein Loch in der Mitte). Die helikale Wiederholung der A-Form liegt bei etwa 11 Bp/Windung. Obwohl es sein kann, dass die A-Form oder eine Form, die ihr nahekommt, von der DNA *in vivo* unter ungewöhnlichen Umständen angenommen wird, hat sie hauptsächlich Bedeutung als Helixform für die RNA (s. unten) und für DNA-RNA-Hybride; es stellte sich heraus, dass die 2'-OH-Gruppe der RNA unmöglich in die ansonsten stabilere B-Form-Struktur passt.

Die DNA kann eine weitere ungewöhnliche Helixstruktur bilden. Die linksgängige **Z-DNA** (Abb. 1.8c) ist in synthetischer doppelsträngiger DNA stabil, die nur aus sich abwechselnden Pyrimidin-Purin-Sequenzen besteht (wie z. B. 5'-CGCGCG-3', und natürlich das Gleiche im anderen Strang). Der Grund ist, dass in dieser Struktur die Pyrimidin- und Purinnukleotide ganz unterschiedliche Konformationen annehmen, im Gegensatz zur A- und B-Form, bei der jedes Nukleotid grundsätzlich immer die gleiche Konformation und die gleiche unmittelbare

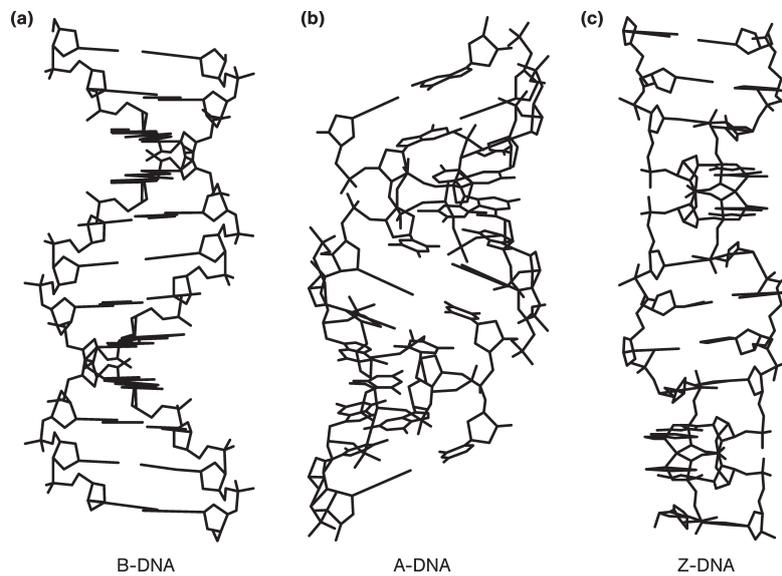


Abb. 1.8 Die alternativen Helixformen der DNA-Doppelhelix.

Umgebung hat. Genauer gesagt nehmen die Purinnukleotide in der Z-Form die *syn*-Konformation an, bei der die Purinbase direkt über dem Desoxyribosering liegt (man stelle sich die Base in Abb. 1.4 180° rotierend um die Glykosidbindung vor; die dort gezeigten Nukleotide sind in der alternativen *anti*-Konformation). Die Pyrimidinnukleotide in der Z-DNA und alle Nukleotide in der A- und B-Form nehmen die *anti*-Konformation an. Die Z-Form hat ein zickzackartiges Aussehen, mit 12 Bp/Umdrehung, obgleich es wahrscheinlich sinnvoll ist anzunehmen, dass sie aus sechs „Basenpaar-Dimeren“ pro Windung besteht. Die Wiederholungs-

Tabelle 1.1 Zusammenfassung der wichtigsten Eigenschaften der A-, B- und Z-Nukleinsäurehelices

	A-Form	B-Form	Z-Form
Helixsinn	rechtsgängig	rechtsgängig	linksgängig
Durchmesser	~2,6 nm	~2,0 nm	~1,8 nm
Basenpaare pro Helixwindung ( <i>n</i> )	11	10	12 (6 Dimere)
Helikale Windung pro Bp (= 360/ <i>n</i> )	33°	36°	60° (pro Dimer)
Helixanstieg pro Bp ( <i>h</i> )	0,26 nm	0,34 nm	0,37 nm
Helixganghöhe ( <i>nh</i> )	2,8 nm	3,4 nm	4,5 nm
Basenneigung zur Helixachse	20°	6°	7°
Große Furche	eng/tief	weit/tief	flach
Kleine Furche	weit/flach	eng/tief	eng/tief
Glykosidbindung	<i>anti</i>	<i>anti</i>	<i>anti</i> (Pyrimidine), <i>syn</i> (Purine)

einheit entlang eines jeden Stranges ist in Wirklichkeit ein Dinukleotid. In normaler DNA bildet sich die Z-Form sogar in Bereichen mit sich wiederholenden CGCGCG-Sequenzen nicht so leicht, da die Grenzbereiche zwischen der linksgängigen Z-Form und der umgebenden B-Form sehr instabil wären. Die Z-Form ist wahrscheinlich keine übliche Eigenschaft der DNA (oder RNA) *in vivo*, obwohl vorgeschlagen wurde, dass sie in einigen speziellen Fällen eine Rolle bei der Auflösung der Verdrehungsbelastung (Abschnitt 2.2) in der DNA während der Transkription spielt. Tabelle 1.1 vergleicht die A-, B- und Z-Helix miteinander.

### 1.2.8 RNA-Sekundärstruktur

RNA kommt normalerweise als einzelsträngiges Molekül vor und nimmt daher keine lange, regelmäßig helikale Struktur an wie die doppelsträngige DNA. Stattdessen bildet RNA relativ globuläre Konformationen, in denen lokale Bereiche mit helikaler Struktur gebildet werden; in diesen ist ein Teil der RNA-Kette komplementär zu einem anderen und bildet über intramolekulare Wasserstoffbrückenbindungen und Basenstapel innerhalb der einzelsträngigen Nukleinsäurekette **Haarnadel-** und **Stamm-Schleife-**Strukturen (Abschnitt 5.1, Abb. 5.3 und Abschnitt 8.2, Abb. 8.2). Diese Konformationsvariabilität spiegelt sich in den vielfältigeren Aufgaben der RNA im Vergleich zur DNA in der Zelle wider (s. unten).

### 1.2.9 Modifizierte Nukleinsäuren

Die chemische Modifikation von Basen oder Nukleotiden in Nukleinsäuren ist weit verbreitet und hat eine Reihe spezieller Aufgaben. In der zellulären DNA beschränken sich die Modifikationen auf die Methylierung der N-6-Position von Adenin und der N-4- und N-5-Position von Cytosin (Abb. 1.2), obgleich in der DNA mancher Phagen komplexere Modifikationen vorkommen. Diese Methylierungen spielen eine Rolle bei der Restriktionsmodifikation (Abschnitt 10.3), der Basenfehlpaarungsreparatur (Abschnitt 4.3) und der eukaryotischen Genomstruktur und -expression. Eine weit vielfältigere Spannbreite von Modifikationen kommt in der RNA nach der Transkription vor, was erneut die verschiedenen Aufgaben der RNA in der Zelle widerspiegelt. Diese werden in Abschnitt 8.2 ausführlicher behandelt.

### 1.2.10 Nukleinsäurefunktion

DNA fungiert ausschließlich als Träger genetischer Information von einer Generation zur nächsten und im Zuge dieser Aufgabe dient sie als Matrize für die Synthese der komplementären RNA-Spezies. Obwohl RNA-Moleküle auch selbst als Genom oder Matrizen dienen können (z. B. bei **RNA-Viren** und **Telomerase-RNA**; Abschnitt 3.3) und als Zwischenschritte im Informationsfluss von der DNA zum Protein fungieren (**mRNA**; Abschnitt 1.1), sind sie im Gegensatz zur DNA wegen ihrer chemischen Instabilität weniger verlässliche Informationsspeicher (Abschnitt 2.1). Sie können jedoch eine große Bandbreite von Tertiärstrukturen annehmen und mit DNA Basenpaarungen eingehen; viele RNAs haben

## 12 | 1 Informationsmakromoleküle

zusätzliche Funktionen, die denen von Proteinen ähnlich sind. Diese zahlreich vorkommenden RNAs heißen **nicht codierende RNAs (ncRNAs)**, da sie nicht in Proteine translatiert werden; ihre Gene werden als **RNA-Gene** bezeichnet. Viele sind Struktur- und Funktionsbestandteile der Prä-mRNA-Verarbeitungs- und Proteinsynthesemaschinerie (z. B. **snRNA**, **tRNA**, **rRNA** und **7SL-RNA**; Kap. 9). Dagegen haben manche, als **Ribozyme** bezeichnete RNAs katalytische Aktivität (z. B. **rRNA** und **RNase P**; Abschnitt 9.2). In jüngster Zeit wurde klar, dass ein überraschend großer Teil des eukaryotischen Genoms für weitere ncRNAs codiert, die für die Kontrolle der Genexpression unverzichtbar sind. Sie werden unterteilt in **lncRNAs** (>200 Nt), die primär an der Transkriptionskontrolle beteiligt sind, und die kleineren (<200 Nt) **miRNAs**, **siRNAs** und **piRNAs**, die hauptsächlich in die Translationskontrolle einbezogen sind (Abschnitt 9.4), obwohl die Größen- und Funktionsunterschiede nicht absolut sind. Da RNA die Fähigkeit zur Speicherung genetischer Information und auch zur Katalyse und Kontrolle chemischer Reaktionen besitzt, basierte vielleicht das Leben zuerst auf RNA und erst später auf dem jetzt vorhandenen System, das auf DNA, RNA und Proteinen fußt.



### Noch einmal in Kürze

#### Basen

In der DNA kommen vier heterozyklische Basen vor: Adenin (A) und Guanin (G) sind Purine, Cytosin (C) und Thymin (T) sind Pyrimidine. In der RNA ist Thymin durch das strukturell sehr ähnliche Pyrimidin Uracil (U) ersetzt.

#### Nukleoside

Ein Nukleosid besteht aus einer Base, die kovalent mit einer Pentose (Zucker) an deren 1'-Position verknüpft ist. In der RNA ist der Zucker Ribose, und die Verbindungen heißen Ribonukleoside oder einfach Nukleoside. In der DNA ist der Zucker dagegen 2'-Desoxyribose, und die Nukleoside bezeichnet man als 2'-Desoxyribonukleoside oder einfach Desoxyribonukleoside. Base + Zucker = Nukleosid.

#### Nukleotide

Nukleotide sind Nukleoside, die an der 3', 5'- oder – in manchen Ribonukleotiden – 2'-Position mit einer oder mehreren Phosphatgruppen kovalent verknüpft sind. Base + Zucker + Phosphat = Nukleotid. Die Nukleosid-5'-triphosphate, NTPs und dNTPs, sind die Bausteine der polymeren RNA bzw. DNA.

#### Phosphodiesterbindungen

In Nukleinsäurepolymeren sind die Ribose- oder Desoxyribosezucker über ein Phosphat zwischen der 5'-Position eines Zuckers und der 3'-Position des nächsten Zuckers verknüpft; dabei bildet sich eine 3',5'-Phosphodiesterbindung aus. Damit besitzen Nukleinsäuren ein Zucker-Phosphat-Rückgrat, das eine Richtung aufweist und bei dem an der 1'-Position eines jeden Zuckers eine Base

sitzt. Die Wiederholungseinheit ist ein Nukleotid. Nukleinsäuren sind stark geladene Polymere mit einer negativen Ladung an jedem Phosphat.

#### **DNA/RNA-Sequenz**

Die Nukleinsäuresequenz ist die Sequenz der Basen A, C, G, T/U in der DNA- oder RNA-Kette. Vereinbarungsgemäß schreibt man die Sequenz vom freien 5'- zum freien 3'-Ende des Moleküls, z. B. 5'-ATAAGCTC-3' (DNA) oder 5'-AUAGCUUGA-3' (RNA).

#### **DNA-Doppelhelix**

DNA kommt zumeist als Doppelhelix vor. Zwei getrennte antiparallele DNA-Ketten winden sich als rechtsgängige Helices (Spiralen) umeinander. Dabei liegen die Zucker-Phosphat-Rückgrate außen und die über Wasserstoffbrückenbindungen gepaarten Basen sind im Innern übereinandergestapelt. Die beiden Ketten ergänzen sich (sie sind komplementär), d. h. eine legt die Sequenz der anderen fest.

#### **A-, B- und Z-Helices**

Die von Watson und Crick entdeckte „Standard“-DNA-Helix, die sogenannte B-Form, wird als die *in vivo* vorherrschende DNA-Struktur angesehen. Nukleinsäuren können aber ebenso die rechtsgängige A-Helix ausbilden, die von RNA-Sequenzen *in vivo* angenommen wird, und die linksgängige Z-Helix, die sich nur bei bestimmten alternierenden Basensequenzen ausbildet und wahrscheinlich *in vivo* keine sehr bedeutsame Konformation darstellt.

#### **RNA-Sekundärstruktur**

Die meisten RNA-Moleküle kommen als Einzelstrang vor, der zu einer komplexen Konformation gefaltet sein kann, mit lokalen Bereichen intramolekularer Basenpaarung und anderen Wechselwirkungen über Wasserstoffbrückenbindungen. Diese Komplexität spiegelt sich in den verschiedenen Rollen der RNA in der Zelle wider.

#### **Modifizierte Nukleinsäuren**

Kovalente Modifikationen der Nukleinsäuren haben spezifische Funktionen in der Zelle. In der DNA beschränken sich diese Modifikationen zumeist auf Methylierungen der Basen Adenin und Cytosin. Die Spannbreite der RNA-Modifikationen ist jedoch weitaus größer.

#### **Nukleinsäurefunktion**

DNA fungiert nur als Träger der exprimierbaren genetischen Information. Die weit vielseitigeren RNAs haben dagegen zahlreiche strukturelle und funktionelle Aufgaben beim Mechanismus und bei der Regulation von Informationsspeicherung, -fluss und -verarbeitung.



### Tipp

#### Verwandte Themen:

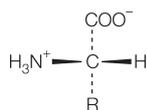
- (Abschnitt 2.1) Chemische und physikalische Eigenschaften von Nucleinsäuren
- (Abschnitt 2.2) Spektroskopische und thermische Eigenschaften von Nucleinsäuren
- (Abschnitt 2.3) DNA-Superspiralisierung

## 1.3 Proteinstruktur und -funktion

### 1.3.1 Aminosäurestruktur

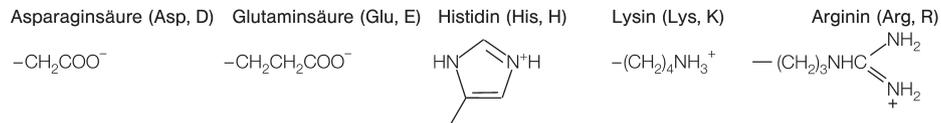
Proteine sind Polymere von L-Aminosäuren. Abgesehen von **Prolin** haben alle der in Proteinen vorkommenden 20 Aminosäuren eine gemeinsame Struktur: Ein Kohlenstoffatom (das  $\alpha$ -Kohlenstoffatom) ist mit einer Carboxylgruppe, einer primären Aminogruppe, einem Proton und einer **Seitenkette** (R) verknüpft. Die Seitenkette ist bei jeder Aminosäure anders (Abb. 1.9). Außer bei **Glycin** ist das  $\alpha$ -Kohlenstoffatom asymmetrisch – es ist mit vier chemisch verschiedenen Gruppen verknüpft. Aminosäuren können somit als optisch aktive Stereoisomerpaare (D- und L-) vorkommen. Nur die L-Isomere sind in Proteinen anzutreffen. Glycin, die einfachste Aminosäure, hat anstelle der Seitenkette ein Wasserstoffatom und ist optisch inaktiv. In wässriger Lösung sind Aminosäuren dipolare Ionen (**Zwitterionen**) und verhalten sich sowohl als Säuren als auch als Basen (sie sind **amphoter**). Die Seitenketten unterscheiden sich hinsichtlich Größe, Gestalt, Ladung und chemischer Reaktivität, und sie sind für die unterschiedlichen Eigenschaften der verschiedenen Proteine verantwortlich (Abb. 1.10). Viele Proteine enthalten auch Nichtstandard-Aminosäuren, wie z. B. 4-Hydroxyprolin und 5-Hydroxylysin in Kollagen. Diese entstehen hauptsächlich durch **posttranslationale Modifikation** der Ausgangsaminosäuren, z. B. Prolin und Lysin, im neu synthetisierten Protein (Abschnitt 9.4). Beim **Selenocystein** (kommt in einer Reihe von Enzymen vor) ist der S des Cysteins durch Se ersetzt. **Pyrrolysin** (ein modifiziertes Lysin, das nur in bestimmten Archaeen-Proteinen vorkommt) und Selenocystein werden beide aufgrund einer feinen Veränderung des genetischen Codes (Abschnitt 8.1) in die wachsenden Proteinketten eingebaut, und manche Wissenschaftler betrachten sie als die 21. und 22. „Standard“-Aminosäure.

Wenn man den pH-Wert von 7 als Bezugspunkt nimmt, dann haben mehrere Aminosäuren ionisierbare Gruppen in ihren Seitenketten, die bei diesem pH-Wert für eine zusätzliche positive oder negative Ladung sorgen. Die „sauren“ Aminosäuren, **Asparaginsäure** und **Glutaminsäure**, besitzen zusätzliche Carboxylgrup-

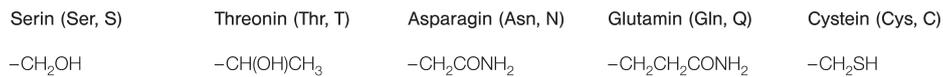


**Abb. 1.9** Allgemeine Struktur einer L-Aminosäure. Die R-Gruppe ist die Seitenkette.

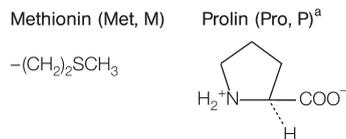
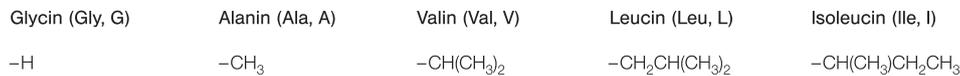
**Geladene Seitenketten**



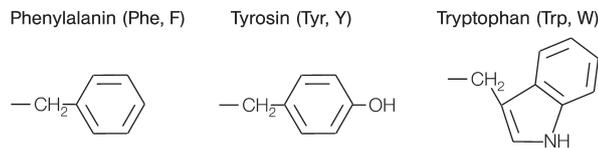
**Polare ungeladene Seitenketten**



**Unpolare aliphatische Seitenketten**



**Aromatische Seitenketten**



**Abb. 1.10** Seitenketten (R) der 20 häufig auftretenden Aminosäuren. In Klammern stehen die Standard-Drei-Buchstaben-Abkürzung und der Ein-Buchstaben-Code. <sup>a</sup>Von Prolin ist die gesamte Struktur gezeigt, da es sich um eine sekundäre Aminosäure handelt.

pen, die in der Regel ionisiert (negativ geladen) sind. Die „basischen“ Aminosäuren haben positiv geladene Gruppen: **Lysin** verfügt über eine zweite Aminogruppe, die mit dem  $\epsilon$ -Kohlenstoffatom verknüpft ist, während **Arginin** eine Guanidinogruppe hat. Die Imidazolgruppe von **Histidin** hat einen nahezu neutralen  $pK_s$ -Wert. Die reversible Protonierung dieser Gruppe unter physiologischen Bedingungen trägt zum Katalysemechanismus vieler Enzyme bei. Saure und basische Aminosäuren können in Proteinen miteinander wichtige Salzbrücken bilden.

Polare, ungeladene Seitenketten enthalten Gruppen, die mit Wasser Wasserstoffbrückenbindungen ausbilden können. Zusammen mit den geladenen Aminosäuren werden sie oft als **hydrophil** („wasserliebend“) beschrieben. **Serin** und **Threonin** haben Hydroxylgruppen, die durch Proteinkinasen reversibel phosphoryliert werden können (s. unten). **Asparagin** und **Glutamin** sind dagegen die Amidderi-

vate der Asparagin- bzw. Glutaminsäure. **Cystein** besitzt eine **Thiol**-(Sulfhydryl)-gruppe, die oft zu **Cystin** oxidiert. In Cystin bilden zwei Cysteine eine strukturell wichtige Disulfidbrücke.

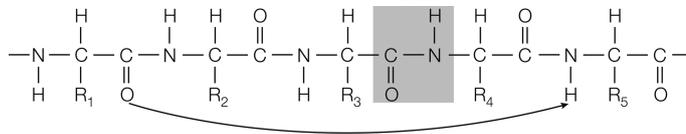
**Phenylalanin**, **Tyrosin** (das ebenfalls phosphoryliert werden kann) und **Tryptophan** haben sperrige **hydrophobe** („wasserabweisende“) Seitenketten, die an **hydrophoben Wechselwirkungen** in der Proteinstruktur teilnehmen (s. unten). Die aromatischen Strukturen von Tyrosin und Tryptophan sind hauptsächlich für die Absorption von ultraviolettem Licht (UV) durch Proteine verantwortlich; deren maximale Absorption liegt bei 280 nm. Die phenolischen Hydroxylgruppen von Tyrosin können ebenfalls Wasserstoffbrückenbindungen ausbilden. Zu den anderen unpolaren, hydrophoben Seitenketten zählen die aliphatischen Alkylgruppen von **Alanin**, **Valin**, **Leucin**, **Isoleucin** und **Methionin** (das ein **Schwefelatom** in einer Thioetherverbindung enthält) sowie der zyklische Ring von Prolin – Prolin ist eine ungewöhnliche Aminosäure, da es sich hier um eine sekundäre Amino- (oder **Imino**)säure handelt.

### 1.3.2 Proteingröße und -formen

Zwei umfangreiche Proteinklassen lassen sich unterscheiden: **Globuläre Proteine** sind kompakt gefaltet und verhalten sich in Lösung mehr oder weniger wie kugelige Partikel; die meisten Enzyme sind globulärer Natur. **Faserproteine** besitzen ein hohes Achsenverhältnis (Länge/Breite) und sind häufig wichtige Strukturproteine, z. B. Seidenfibroin und Keratin in Haaren und Wolle. Die Molekularmasse reicht von wenigen Tausend Dalton (Da), wie z. B. beim Hormon Insulin mit 51 Aminosäuren und einer Molekülmasse von 5734 Da (5,7 Kilodalton, kDa), bis zu nahezu 4 Millionen Da (4 MDa) im Falle des Muskelproteins Titin. Manche Proteine enthalten **Nichtprotein-Moleküle**, entweder in Form kleiner **prothetischer Gruppen**, die als Cofaktoren in enzymatischen Reaktionen mitwirken können, oder als große Anlagerungen (z. B. Lipide in **Lipoproteinen** oder Kohlenhydrate in **Glykoproteinen**).

### 1.3.3 Primärstruktur

Die  $\alpha$ -Carboxylgruppe einer Aminosäure ist kovalent mit der  $\alpha$ -Aminogruppe der nächsten Aminosäure über eine Amidbindung verknüpft; in Proteinen wird eine solche Verknüpfung gewöhnlich als **Peptidbindung** bezeichnet. Sind zwei Aminosäurereste auf diese Weise verknüpft, entsteht ein **Dipeptid**. Viele über Peptidbindungen verknüpfte Aminosäuren bilden ein **Polypeptid** (Abb. 1.11). Die sich wiederholende Abfolge der  $\alpha$ -Kohlenstoffatome und Peptidbindungen liefert das strukturelle **Rückgrat** des Polypeptids, während die verschiedenen Aminosäure-**Seitenketten** dem Protein die Funktionalität verleihen. An einem Ende der Polypeptidkette hat die Aminosäure eine nicht verknüpfte  $\alpha$ -Aminogruppe, während das andere Ende der Kette eine freie  $\alpha$ -Carboxylgruppe besitzt. Somit sind Polypeptide gerichtet, mit einem **N-Terminus** und einem **C-Terminus**. Manchmal ist der N-Terminus **blockiert**, z. B. durch eine Acetylgruppe. Die Abfolge (Sequenz)



**Abb. 1.11** Ausschnitt aus einer Polypeptidkette. Die Peptidbindung ist grau unterlegt. In der  $\alpha$ -Helix ist die CO-Gruppe des Aminosäurerests  $n$  mit der NH-Gruppe des Rests  $n+4$  über eine Wasserstoffbrücke verbunden (angedeutet mit dem Pfeil).

der Aminosäuren vom N- zum C-Terminus ist die **Primärstruktur** des Polypeptids. Die typische Größe einer Polypeptidkette reicht von 100 bis 1500 Aminosäuren, obwohl es auch kürzere oder längere gibt; Titin hat z. B. 34 000 Aminosäuren.

### 1.3.4 Nichtkovalente Wechselwirkungen

Die dreidimensionale Struktur der Proteine und der großen proteinhaltigen Zusammenschlüsse wird durch viele verschiedene Wechselwirkungen aufrechterhalten. Elektrostatische **Ladungs-Ladungs-Wechselwirkungen (Salzbrücken)** wirken zwischen ionisierbaren Gruppen entgegengesetzter Ladung bei physiologischem pH-Wert, d. h. zwischen positiven Lysin- und Arginin-Seitenketten und negativen Glutamin- und Asparagin-Seitenketten oder negativen Phosphaten der DNA bei DNA-bindenden Proteinen wie den Histonen (Abschnitt 2.4). **Ladungs-Dipol- und Dipol-Dipol-Wechselwirkungen** sind schwächer und bilden sich, wenn ein oder beide Teilnehmer wegen der asymmetrischen Ladungsverteilung im Molekül Dipole sind (Abb. 1.12a). Sogar ungeladene Gruppen wie Methylgruppen können sich gegenseitig über vorübergehende Dipole, die durch die Bewegung ihrer Elektronen (**Dispersionskräfte**) entstehen, schwach anziehen.

Nichtkovalente Assoziationen zwischen elektrisch neutralen Molekülen sind allgemein unter der Bezeichnung **van der Waals-Kräfte** bekannt. **Wasserstoffbrückenbindungen** sind sehr wichtig. Sie bilden sich zwischen einem kovalent verknüpften Wasserstoffatom auf einer Donatorgruppe (z. B. –O-H oder –N-H) und einem Paar nicht bindender Elektronen auf einer Akzeptorgruppe (z. B. :O=C– oder :N–) (Abb. 1.12b). Hydrophobe Bindungen und andere Wechselwirkungen, an denen Dipole beteiligt sind, sind gerichtet und tragen so dazu bei, die makromolekulare Gestalt und die Spezifität der molekularen Wechselwirkungen festzulegen. Die Anwesenheit ungeladener und unpolarer Substanzen, z. B. Lipide, in einer wässrigen Umgebung zwingt den umgebenden Wassermolekülen eine hoch geordnete Struktur auf. Dies ist energetisch ungünstig, da hierdurch die Entropie des Systems verringert wird. Somit neigen unpolare Moleküle und funktionelle Gruppen wie die aliphatischen und aromatischen Aminosäure-Seitenketten zum Verklumpen, wodurch die dem Wasser ausgesetzte Gesamtoberfläche geringer wird. Man nennt diese Anziehung **hydrophobe Wechselwirkung**; sie ist die wichtigste stabilisierende Kraft bei Protein-Protein- und Protein-Lipid-Wechselwirkungen sowie in Nukleinsäuren (Abschnitt 1.2).

18 | 1 Informationsmakromoleküle

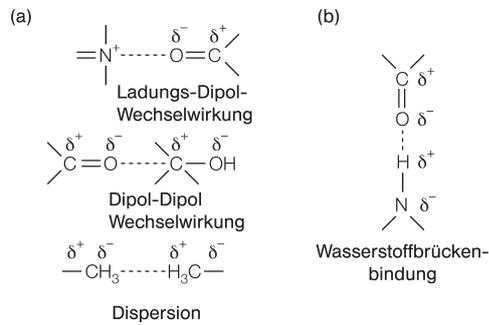


Abb. 1.12 Beispiele für (a) van der Waals-Kräfte und (b) eine Wasserstoffbrückenbindung.

### 1.3.5 Sekundärstruktur

Da die C=O- und N-H-Gruppen in Peptidbindungen stark polar sind, hat die C-N-Bindung einen partiellen Doppelbindungscharakter. Dies macht die Peptidbindung starr und planar, dennoch ist eine freie Rotation zwischen benachbarten Peptidbindungen möglich. Diese Polarität begünstigt auch die Bildung von Wasserstoffbrückenbindungen zwischen Peptidbindungseinheiten mit entsprechendem Abstand und entsprechender Ausrichtung. Somit können sich Polypeptidketten zu einer

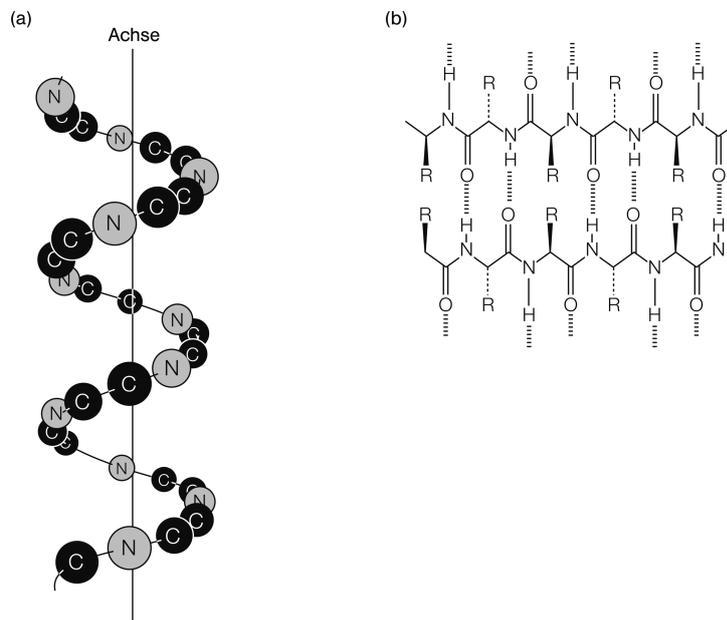


Abb. 1.13 (a)  $\alpha$ -Helix-Sekundärstruktur. Der Übersicht halber sind nur der  $\alpha$ -Kohlenstoff, der Kohlenstoff der Peptidbindung sowie die Stickstoffatome des Polypeptidrückgrats gezeigt. (b) Ausschnitt aus einer  $\beta$ -Faltblatt-Sekundärstruktur.

Reihe regelmäßiger Strukturen falten, die durch diese Wasserstoffbrückenbindungen zusammengehalten werden. Die bestbekannte **Sekundärstruktur** ist die  $\alpha$ -**Helix** (Abb. 1.13a). Das Polypeptidrückgrat bildet eine rechtsgängige Helix mit 3,6 Aminosäureresten pro Windung; dabei bildet jede N–H-Peptidgruppe mit der C=O-Gruppe einer sich im Abstand von drei Aminosäureresten entfernt befindenden Peptidbindung Wasserstoffbrückenbindungen aus (Abb. 1.11). Abschnitte mit einer  $\alpha$ -helikalen Sekundärstruktur kommen oft in globulären Proteinen und in manchen Faserproteinen vor. Die seltenere  **$3_{10}$ -Helix** ist ähnlich, hat aber eine andere Ausdehnung. Das  **$\beta$ -Faltblatt** entsteht durch Wasserstoffbrücken zwischen den N–H- und C=O-Gruppen der Peptidbindung und den komplementären Gruppen eines anderen Abschnitts der Polypeptidkette (Abb. 1.13b). Daran können mehrere Abschnitte der Polypeptidkette – Seite an Seite – beteiligt sein; so entsteht eine Blattstruktur mit Seitenketten (R), die abwechselnd aus dem Blatt nach oben und nach unten ragen. Verlaufen diese Abschnitte in dieselbe Richtung (d. h. N-Terminus  $\rightarrow$  C-Terminus), ist das Faltblatt **parallel**; wechselt die Richtung zwischen N  $\rightarrow$  C und C  $\rightarrow$  N, ist das Blatt **antiparallel**. Es gibt auch **gemischte  $\beta$ -Faltblätter**, die beide Ausrichtungen aufweisen.  $\beta$ -Faltblätter sind kräftig und starr und in Strukturproteinen von Bedeutung, z. B. in Seidenfibroin. Das Bindegewebsprotein **Kollagen** hat eine ungewöhnliche **Dreifach-Helix** als Sekundärstruktur, bei der drei Polypeptidketten ineinander verschlungen sind; dadurch wird es sehr fest.

### 1.3.6 Tertiärstruktur

Die verschiedenen Abschnitte der  $\alpha$ -Helix, des  $\beta$ -Faltblatts, anderer kleiner Sekundärstrukturen und verbindender unstrukturierter Schleifen falten sich dreidimensional zu einer Polypeptid-**Tertiärstruktur** (Abb. 1.14). Die Beschaffenheit der Tertiärstruktur ist in der Primärstruktur angelegt, und die meisten Polypeptide falten sich – die richtigen Bedingungen vorausgesetzt – spontan zur richtigen Tertiärstruktur, da diese im Allgemeinen die Konformation mit der niedrigsten Energie für diese Sequenz ist. *In vivo* wird die korrekte Faltung gewöhnlich durch Proteine, die sogenannten **Chaperone**, gefördert; diese tragen dazu bei, dass sich die neuen Polypeptide nicht fehlfalten, bevor ihre Synthese (und Primärstruktur) abgeschlossen ist. Die Faltung geschieht folgendermaßen: Aminosäuren mit hydrophilen Seitenketten kommen hauptsächlich auf der Oberfläche des Proteins zu liegen, wo sie mit Wasser oder den Ionen des Lösungsmittels wechselwirken; hydrophobe Aminosäuren werden dagegen im Innern verborgen, von dem Wasser ausgeschlossen ist. Dies verleiht der Struktur eine Gesamtstabilität. Verschiedene Arten nichtkovalenter Wechselwirkungen zwischen Seitenketten halten die Tertiärstruktur zusammen: van der Waals-Kräfte, Wasserstoffbrückenbindungen, elektrostatische Salzbrücken zwischen entgegengesetzt geladenen Gruppen und hydrophobe Wechselwirkungen zwischen unpolaren Seitenketten. Zusätzlich können sich zwischen zwei Cysteinresten, die in der Primärstruktur weit auseinander liegen können, in der gefalteten Tertiärstruktur jedoch nahe beieinander sind, kovalente Disulfidbrücken bilden. Werden durch Hitze oder extreme pH-Werte die Sekundär- und Tertiärstruktur zerstört, kommt es zur **Denaturierung** des Proteins

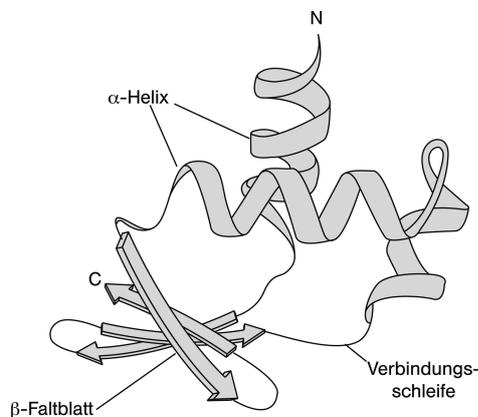


Abb. 1.14 Schematisches Diagramm eines Ausschnitts einer Protein-Tertiärstruktur.

und zur Bildung einer **Random-Coil**-Konformation (eine zufällige Spiralisierung). Jegliche biologische Aktivität geht dabei in der Regel verloren, und die denaturierten Proteine neigen dazu, zu unlöslichen Aggregaten zu verklumpen, da ihre exponierten hydrophoben inneren Strukturen miteinander wechselwirken, um das Wasser auszugrenzen.

Wie wichtig die korrekte Proteinfaltung ist, zeigt sich daran, dass viele neurodegenerative Erkrankungen, wie z. B. die **Alzheimer-Krankheit**, mit der Anhäufung unlöslicher Proteinaggregate in den Nervenzellen, sog. **Amyloidfibrillen**, verbunden sind. Diese Amyloidfibrillen enthalten ausgedehnte β-Faltblattstrukturen. Die tödliche Traberkrankheit (**Scrapie**, beim Schaf), die **bovine spongiforme Enzephalopathie (BSE, „Rinderwahnsinn“)** und die **Creutzfeld-Jakob-Krankheit (CJK, beim Menschen)** werden durch ein infektiöses Agens, ein **Prion**, verursacht, das ebenfalls nur aus einem fehlgefalteten Protein besteht. Gelangt das Prion in eine Nervenzelle, bindet es an ein verwandtes Zellprotein und verursacht dessen Fehlfaltung. Dies setzt eine Kettenreaktion von Fehlfaltungen und die Amyloidbildung in Gang, was zum Verlust der Zellfunktion führt.

### 1.3.7 Quartärstruktur

Viele Proteine sind aus zwei oder mehreren Polypeptidketten (**Untereinheiten**) aufgebaut, die **Oligomere** (mit einigen wenigen Untereinheiten) oder **Multimere** (mit vielen Untereinheiten) bilden. Die Untereinheiten können identisch (**Homomere**) oder verschieden (**Heteromere**) sein. So hat beispielsweise **Hämoglobin** zwei α-Globin- und zwei β-Globinketten ( $\alpha_2\beta_2$ ). Die gleichen Kräfte, die die Tertiärstruktur stabilisieren, halten auch die Untereinheiten zusammen. In manchen Fällen zählen dazu auch Disulfidbrücken zwischen Cysteinen auf getrennten Polypeptiden. Diese Organisationsebene wird als **Quartärstruktur** bezeichnet und sie hat bestimmte Auswirkungen. Zum einen erlaubt sie, dass sehr große Proteinstrukturen gebildet werden können, z. B. die **Mikrotubuli** des Cytoskeletts. Zum

anderen kann sie einem Protein eine größere Funktionsvielfalt verschaffen, indem verschiedene Aktivitäten in einem einzelnen Gebilde kombiniert werden; dieser Fall liegt z. B. beim DNA-Polymerase-III-Holoenzym und dem Replisom vor (Abschnitt 3.2). Oft können die Wechselwirkungen zwischen den Untereinheiten durch die Bindung kleiner Moleküle modifiziert werden; dies kann dann zu **allosterischen** Effekten führen, wie sie bei der Enzymregulation auftreten. Es gibt auch viele Beispiele für vorübergehend auftretende Proteinkomplexe, insbesondere in Zellsignalwegen: Hier bewirkt die posttranslationale Modifikation (etwa eine Phosphorylierung, Abschnitt 9.4) eines Proteins, dass sich dieses Protein kurzzeitig mit einem anderen Protein zusammenschließt (assoziiert), was häufig beim zweiten Protein zu einer Konformationsänderung führt. Dadurch wird dessen Funktion an- oder abgeschaltet.

### 1.3.8 Prothetische Gruppen

Viele Proteine enthalten kovalent oder nichtkovalent verknüpfte kleine Moleküle, sog. **prothetische Gruppen**, die dem Protein eine strukturelle oder chemische Funktionsweise verleihen, die die Aminosäure-Seitenketten nicht verschaffen können. Viele von ihnen sind **Cofaktoren** bei enzymkatalysierten Reaktionen. Beispiele hierfür sind Nicotinamidadeninukleotid (NAD<sup>+</sup>) in vielen Dehydrogenasen, Pyridoxalphosphat in Transaminasen, Häm in Hämoglobin und Cytochromen, Metall-Ionen wie Zn<sup>2+</sup> und Acylgruppen von Fetten, die Proteine durch hydrophobe Wechselwirkungen in Zellmembranen verankern können. Solche Proteine heißen **konjugierte** Proteine, und ohne seine prothetische Gruppe wird das Protein als **Apoprotein** bezeichnet. Andere konjugierte Proteine enthalten assoziierte Makromoleküle in großen Komplexen, z. B. Kohlenhydrate (**Glykoproteine**), Lipide (**Lipoproteine**) oder Nukleinsäuren (**Nukleoproteine**).

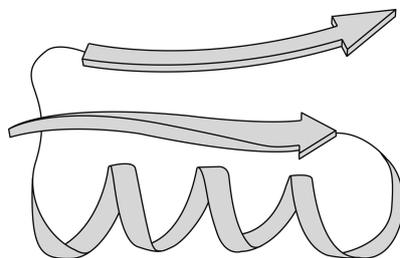
### 1.3.9 Domänen, Motive, Familien und Evolution

Viele einzelne Polypeptide bestehen aus strukturell unabhängigen Untereinheiten oder **Domänen**, die über Abschnitte begrenzter höherer Strukturordnung innerhalb desselben Polypeptids verbunden sind. Diese Verbindungen können als Gelenke wirken, um einzelnen Domänen die Bewegung im Verhältnis zueinander zu ermöglichen. Die Spaltung dieser Verbindungen durch eine begrenzte Proteolyse kann die Domänen häufig abtrennen, die sich dann als unabhängige globuläre Proteine verhalten können. Das aktive Zentrum eines Enzyms bildet sich manchmal in einer Vertiefung zwischen zwei Domänen, die das Substrat umfassen. Domänen können auch eine spezifische Funktion besitzen, wie die Bindung eines häufig verwendeten Moleküls wie z. B. ATP. Wenn eine derartige Funktion in vielen verschiedenen Proteinen erforderlich ist, kommt oft die gleiche Domänenstruktur vor. In Eukaryoten werden Domänen oft durch diskrete Teile von Genen codiert, **Exons** genannt. Aus diesem Grund hat man vorgeschlagen, dass während der Evolution neue Proteine durch Duplikation und Umgruppierung von Exons, die für Domänen codieren, im Genom entstanden sind. Dadurch

wurden in den sich ergebenden neuen Proteinen neue Kombinationen von Bindungsstellen, Katalysezentren und Strukturelementen geschaffen. Auf diese Weise wurde vielleicht die Evolutionsgeschwindigkeit neuer funktionsfähiger Proteine erheblich gesteigert.

**Struktur motive** (auch als **supersekundäre Strukturen** bezeichnet) sind Gruppierungen von Sekundärstrukturelementen, die häufig in globulären Proteinen vorkommen. Sie besitzen oft funktionelle Bedeutung und können die essenziellen Teile von Bindungsstellen oder Katalysezentren darstellen, die während der Evolution der Proteinfamilien – ausgehend von einem gemeinsamen Vorgängermolekül – konserviert wurden. Alternativ können sie auch die beste Lösung einer strukturell-funktionellen Notwendigkeit darstellen, die unabhängig voneinander in nicht verwandten Proteinen erreicht wurde. Ein geläufiges Beispiel ist das  **$\beta\alpha\beta$ -Motiv**, bei dem eine  $\alpha$ -Helix die Verbindung zwischen zwei aufeinanderfolgenden parallelen Strängen eines  $\beta$ -Faltblatts ist (Abb. 1.15). In vielen, ansonsten nicht verwandten Proteinen bilden zwei sich überlappende  $\beta\alpha\beta$ -Motive ( $\beta\alpha\beta\alpha\beta$ ) eine Dinukleotidbindungsstelle (z. B. für  $\text{NAD}^+$ ). **Sequenz motive** bestehen aus linearen Sequenzen konservierter, funktionell wichtiger Aminosäuren, also Primärstruktur anstatt Supersekundärstruktur. Sie können auch Bindungsstellen oder aktive Zentren darstellen.

**Proteinfamilien** entstehen durch aufeinanderfolgende Duplikationen und nachfolgende **divergente Evolution** eines Vorgängergens. Myoglobin, das sauerstoffbindende Protein im Muskel, die  $\alpha$ - und  $\beta$ -Globinketten sowie die kleine  $\delta$ - (delta-) Kette des adulten Hämoglobins und die  $\gamma$ - (gamma-),  $\epsilon$ - (epsilon-) und  $\zeta$ - (zeta-) Globine der embryonalen und fötalen Hämoglobine sind allesamt verwandte Peptide innerhalb der **Globinfamilie** (Abb. 1.16). Man bezeichnet ihre Gene und Proteine als **Homologe**. Familienmitglieder in verschiedenen Spezies, die die gleiche Funktion bewahrt haben und die gleiche biochemische Aufgabe erfüllen (z. B. das Myoglobin der Ratte und der Maus), sind **Orthologe**. Haben Gene bzw. ihre Proteine verschiedene, aber häufig verwandte Funktionen entwickelt (z. B.  $\alpha$ -Globin und  $\beta$ -Globin), sind sie **Paraloge** (Abschnitt 13.6). Der Ähnlichkeitsgrad zwischen den Aminosäuresequenzen orthologer Mitglieder einer Proteinfamilie in verschiedenen Organismen hängt davon ab, vor wie langer Zeit sich die beiden Organismen von ihrem gemeinsamen Vorgängermolekül auseinanderentwickelt haben, und davon, wie wichtig die Konservierung der Sequenz für die Funktion des Proteins ist. Diese Funktion, egal ob strukturell oder katalytisch, ist untrennbar



**Abb. 1.15** Darstellung eines  $\beta\alpha\beta$ -Motivs. Die  $\alpha$ -Helix ist als spiralisiertes Band und die  $\beta$ -Faltblatt-Abschnitte sind als flache Pfeile dargestellt.

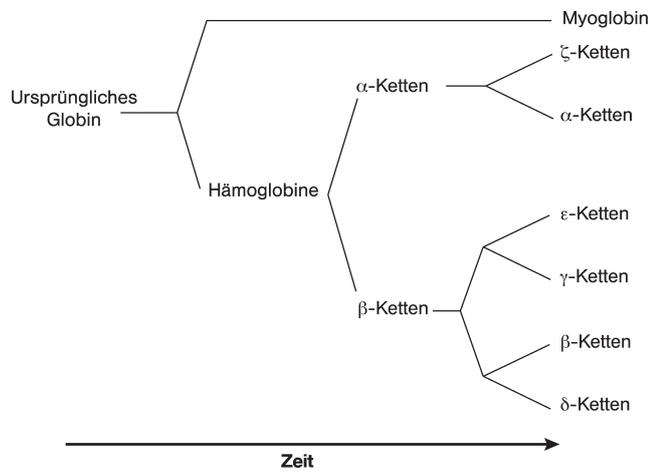


Abb. 1.16 Evolution der Globine aus einem Vorfahren-Globin-Gen.

mit seiner Struktur vorknüpft. Wie oben gezeigt, können ähnliche Strukturen und Funktionen auch durch **konvergente Evolution** erreicht werden, bei der nicht verwandte Gene sich entwickeln und Proteine mit ähnlichen Strukturen oder Katalyseaktivitäten hervorbringen. Ein gutes Beispiel dafür liefern die proteolytischen Enzyme **Subtilisin** (bakteriell) und **Chymotrypsin** (tierisch). Obwohl ihre Aminosäuresequenzen ganz unterschiedlich sind und sie aus unterschiedlichen Strukturmotiven aufgebaut sind, haben sie die gleiche räumliche Ausrichtung der **katalytischen Triade** der Aminosäuren des aktiven Zentrum entwickelt: Serin, Histidin und Asparaginsäure. Und sie benutzen exakt den gleichen Katalysemechanismus, um Peptidbindungen zu hydrolysieren. Solche Proteine heißen **Funktionsanaloge**. Wo sich ähnliche Struktur motive unabhängig voneinander entwickelt haben, sind die resultierenden Proteine **Strukturanaloge**.

### 1.3.10 Proteinfunktion

- **Enzyme:** Abgesehen von ein paar wenigen katalytisch aktiven RNA-Molekülen sind alle Enzyme Proteine. Enzyme können die Geschwindigkeit biochemischer Reaktionen um mehrere Größenordnungen steigern. An der Bindung des **Substrats** sind zahlreiche nichtkovalente Wechselwirkungen mit spezifischen Aminosäure-Seitenketten beteiligt, darunter van der Waals-Kräfte, Wasserstoffbrückenbindungen, Salzbrücken und hydrophobe Kräfte. Die Bindungsspezifität kann mit nur einem einzigen bindenden Substrat extrem hoch sein (Glucose-Oxidase bindet beispielsweise nur Glucose) oder das Enzym ist gruppenspezifisch (z. B. bindet Hexokinase eine Vielzahl von Hexosen). An der Katalyse können auch Seitenketten direkt beteiligt sein, z. B. indem sie als Nucleophile oder als Protonendonatoren oder -akzeptoren agieren.
- **Signalgebung:** Rezeptorproteine in Zellmembranen können **Liganden** (z. B. Hormone) aus dem extrazellulären Medium binden und aufgrund der resultie-

## 24 | 1 Informationsmakromoleküle

renden Konformationsänderung in der Zelle Reaktionen als Antwort auf den Liganden auslösen. Die Bindung des Liganden hat Ähnlichkeit mit der Substratbindung, aber der Ligand bleibt in der Regel unverändert. Manche Hormone sind selbst kleine Proteine, wie z. B. Insulin und das Wachstumshormon. **Proteinkinasen** ändern die Eigenschaften anderer Proteine, indem sie eine Phosphorylgruppe von ATP auf diese Proteine übertragen; sie sind äußerst wichtige Enzyme bei der intrazellulären Signalgebung.

- **Transport und Speicherung: Hämoglobin** transportiert Sauerstoff in den roten Blutkörperchen, während **Transferrin** Eisen zur Leber bringt. Sobald das Eisen in der Leber angekommen ist, wird es, gebunden an das Protein **Ferritin**, gespeichert. Nahrungsfette werden im Blut durch **Lipoproteine** befördert. Viele andere Moleküle und Ionen werden in proteingebundener Form transportiert und gespeichert. Dies kann die Löslichkeit erhöhen und ihre Reaktivität herabsetzen, bis sie benötigt werden.
- **Struktur und Bewegung: Kollagen** ist das wichtigste Protein in der Haut, den Knochen und im Bindegewebe; Haare bestehen dagegen hauptsächlich aus **Keratin**. In der Zelle gibt es auch viele Strukturproteine, z. B. im **Cytoskelett**. Die wichtigsten Muskelproteine, **Aktin** und **Myosin**, bilden Gleitfilamente, die die Grundlage der Muskelkontraktion bilden.
- **Ernährung: Casein** und **Ovalbumin** sind die wichtigsten Proteine der Milch bzw. der Eier. Sie dienen dazu, den Nachkommen die Aminosäuren für Wachstum und Entwicklung zur Verfügung zu stellen. Samenproteine verschaffen dem keimenden Pflanzenembryo ebenfalls Nahrung.
- **Immunität: Antikörper**, die Bakterien, Viren und anderes Fremdmaterial (das **Antigen**) erkennen und binden, sind Proteine.
- **Regulation: Transkriptionsfaktoren** binden an DNA und modulieren deren Funktion. Viele andere Proteine modifizieren die Funktion anderer Moleküle, indem sie an diese binden.



### Noch einmal in Kürze

#### Aminosäurestruktur

Die 20 in Proteinen häufig auftretenden Aminosäuren besitzen ein chirales  $\alpha$ -Kohlenstoffatom, das mit einem Proton, einer Amino- und einer Carboxylgruppe sowie einer spezifische Seitenkette verknüpft ist; die Seitenkette verleiht dem Molekül verschiedene physikalische und chemische Eigenschaften. Diese Seitenketten können basisch (positiv geladen), sauer (negativ geladen) oder hydrophob (sowohl aliphatisch als auch aromatisch) sein oder andere spezielle funktionelle Gruppen besitzen wie z. B. Hydroxyle, Amide oder Thiole. In Lösung verhalten sie sich wie Zwitterionen. Außer zwei bemerkenswerten Ausnahmen werden Nichtstandard-Aminosäuren in Proteinen durch posttranslationale Modifikation gebildet.

### **Proteingröße und -formen**

Globuläre Proteine, einschließlich der meisten Enzyme, verhalten sich in Lösung wie kompakte, grob kugelige Partikel. Faserproteine besitzen ein hohes Achsenverhältnis und haben häufig eine strukturelle Bedeutung, wie z. B. Fibroin und Keratin. Die Größe reicht von wenigen Tausend bis zu mehreren Millionen Dalton. Manche Proteine sind mit Nicht-Proteinen verbunden, z. B. mit Lipiden oder Kohlenhydraten oder auch kleinen Cofaktoren.

### **Primärstruktur**

Aminosäuren sind über Peptidbindungen zwischen  $\alpha$ -Carboxyl- und  $\alpha$ -Amino-gruppen miteinander verknüpft. Die entstehende Polypeptidsequenz besitzt einen N-Terminus und einen C-Terminus. In der Regel bestehen Polypeptide aus 100 bis 1500 Aminosäuren, die auf diese Weise verknüpft sind.

### **Nichtkovalente Wechselwirkungen**

Die dreidimensionale Struktur von Proteinen wird durch eine große Anzahl schwacher Wechselwirkungen aufrechterhalten. Ladungs-Ladungs-, Ladungs-Dipol- und Dipol-Dipol-Wechselwirkungen gehen mit Anziehungen zwischen voll oder teilweise geladenen Atomen einher. Wasserstoffbrückenbindungen sowie hydrophobe Wechselwirkungen, die Wasser ausschließen, sind ebenfalls von Bedeutung.

### **Sekundärstruktur**

Polypeptide können sich zu einer Reihe regelmäßiger Strukturen falten. Die rechtsgängige  $\alpha$ -Helix hat 3,6 Aminosäuren pro Windung und wird über Wasserstoffbrückenbindungen zwischen N-H- und C=O-Gruppen stabilisiert; diese brückenbildenden Gruppen liegen in der Primärstruktur jeweils drei Peptidbindungen auseinander. Parallele und antiparallele  $\beta$ -Faltblätter werden durch Wasserstoffbrückenbindungen zwischen verschiedenen Teilen der Polypeptidkette stabilisiert.

### **Tertiärstruktur**

Die verschiedenen Abschnitte der Sekundärstruktur und verbindende Bereiche falten sich zu einer klar definierten Tertiärstruktur, bei der die hydrophilen Aminosäuren zumeist auf der Oberfläche liegen und die hydrophoben im Innern. Die Struktur wird über nichtkovalente Wechselwirkungen und manchmal über Disulfidbrücken stabilisiert. Bei der Denaturierung gehen Sekundär- und Tertiärstruktur verloren.

### **Quartärstruktur**

Viele Proteine bestehen nicht nur aus einer einzigen Polypeptidkette. Hämoglobin hat z. B. zwei  $\alpha$ - und zwei  $\beta$ -Ketten. Große Komplexe wie die Mikrotubuli sind aus verschiedenen Polypeptidketten aufgebaut, die über eine Quartärstruktur assoziiert sind. Allosterische Effekte hängen gewöhnlich von Wechselwirkungen zwischen den Untereinheiten ab.

### Prothetische Gruppen

Manche Proteine sind mit Nichtprotein-Molekülen (prothetischen Gruppen) assoziiert, die dem Protein zusätzliche chemische Funktionen verleihen. Zu den kleinen prothetischen Gruppen zählen Nicotinamidadeninucleotid (NAD<sup>+</sup>), Häm und Metall-Ionen wie Zn<sup>2+</sup>.

### Domänen, Motive, Familien und Evolution

Domänen bilden halbunabhängige Strukturen und Funktionseinheiten innerhalb einer einzelnen Polypeptidkette. Durch Kombination von Domänen können sich neue Proteine entwickeln. Motive sind Gruppen von Primär- oder Sekundärstrukturelementen, die oft in verwandten Mitgliedern von Proteinfamilien vorkommen. Proteinfamilien entstehen durch Genduplikation und die anschließende divergente Evolution der neuen Gene.

### Proteinfunktion

Proteine besitzen eine große Funktionsvielfalt. Sie können als Enzyme, Antikörper und Strukturkomponenten in und außerhalb der Zelle wirken, als Rezeptoren und Transporter für chemische Liganden, als Regulatoren und als Nahrungsspeicher.



#### Tipp

##### Verwandte Themen:

- (Kap. 9) Proteinsynthese



#### Wissen testen

- 1.1. Die Glykosylierung der sezernierten Proteine findet statt in/im:
  - A) Mitochondrien
  - B) Peroxisomen
  - C) Endoplasmatischen Retikulum
  - D) Zellkern
- 1.2. Welche der folgenden Strukturen ist ein Beispiel für ein Nukleoprotein?
  - A) Keratin
  - B) Chromatin
  - C) Histone
  - D) Proteoglykan
- 1.3. Welche der folgenden Strukturen ist kein Beispiel für ein Polysaccharid?
  - A) Chitin
  - B) Amylopektin
  - C) Glykosaminoglykan
  - D) Glycerin

- 1.4. Transmembranproteine:
  - A) verbinden zwei Lipiddoppelschichten miteinander.
  - B) haben intra- und extrazelluläre Domänen.
  - C) sitzen vollständig innerhalb der Membran.
  - D) lassen sich leicht aus der Membran entfernen.
- 1.5. Welche der folgenden Aminosäuren ist eine Iminosäure?
  - A) Prolin
  - B) Hydroxylysin
  - C) Tryptophan
  - D) Histidin
- 1.6. Mitglieder einer Proteinfamilie in verschiedenen Spezies, die die gleiche biochemische Aufgabe erfüllen, heißen:
  - A) Paraloge.
  - B) Strukturanaloge.
  - C) Heterologe.
  - D) Orthologe.
- 1.7. Bei welcher der folgenden Strukturen handelt es sich nicht um eine Sekundärstruktur?
  - A)  $\alpha$ -Helix
  - B) Dreifachhelix
  - C) Doppelhelix
  - D)  $\beta$ -Faltblatt
- 1.8. Bei der isoelektrischen Fokussierung werden Proteine getrennt mithilfe eines:
  - A) pH-Gradienten.
  - B) Salzgradienten.
  - C) Dichtegradienten.
  - D) Temperaturgradienten.

