INSTITUT FÜR INFORMATIK
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN
LEHRSTUHL FÜR EFFIZIENTE ALGORITHMEN

**Skriptum**

**zur Vorlesung**

# Algorithmische Bioinformatik I/II

*gehalten im Wintersemester 2001/2002*

*und im Sommersemester 2002 von*

*Volker Heun*

*Erstellt unter Mithilfe von:*

*Peter Lücke – Hamed Behrouzi – Michael Engelhardt*

*Sabine Spreer – Hanjo Täubig*

*Jens Ernst – Moritz Maaß*

**14. Mai 2003**

*Version 0.96*

# Vorwort

Dieses Skript entstand parallel zu den Vorlesungen *Algorithmische Bioinformatik I* und *Algorithmische Bioinformatik II*, die im Wintersemester 2001/2002 sowie im Sommersemester 2002 für Studenten der Bioinformatik und Informatik sowie anderer Fachrichtungen an der Technischen Universität München im Rahmen des von der Ludwig-Maximilians-Universität und der Technischen Universität gemeinsam veranstalteten Studiengangs Bioinformatik gehalten wurde. Einige Teile des Skripts basieren auf der bereits im Sommersemester 2000 an der Technischen Universität München gehaltenen Vorlesung *Algorithmen der Bioinformatik* für Studierende der Informatik.

Das Skript selbst umfasst im Wesentlichen die grundlegenden Themen, die man im Bereich Algorithmische Bioinformatik einmal gehört haben sollte. Die vorliegende Version bedarf allerdings noch einer Ergänzung weiterer wichtiger Themen, die leider nicht in den Vorlesungen behandelt werden konnten.

An dieser Stelle möchte ich insbesondere Hamed Behrouzi, Michael Engelhardt und Peter Lücke danken, die an der Erstellung des ersten Teils dieses Skriptes (Kapitel 2 mit 5) maßgeblich beteiligt waren. Bei Sabine Spreer möchte ich mich für die Unterstützung bei Teilen des siebten Kapitels bedanken. Bei meinen Übungsleitern Jens Ernst und Moritz Maaß für deren Unterstützung der Durchführung des Übungsbetriebs, aus der einige Lösungen von Übungsaufgaben in dieses Text eingeflossen sind. Bei Hanjo Täubig möchte ich mich für die Mithilfe zur Fehlerfindung bedanken, insbesondere bei den biologischen Grundlagen.

Falls sich dennoch weitere (Tipp)Fehler unserer Aufmerksamkeit entzogen haben sollten, so bin ich für jeden Hinweis darauf (an heun@in.tum.de) dankbar.

München, im September 2002 Volker Heun

# Inhaltsverzeichnis

# Molekularbiologische Grundlagen

## 1.1 Mendelsche Genetik

In diesem Einführungskapitel wollen wir uns mit den molekularbiologischen Details beschäftigen, die für die informatische und mathematische Modellierung im Folgenden hilfreich sind. Zu Beginn stellen wir noch einmal kurz die Anfänge der systematischen Genetik, die Mendelsche Genetik, dar.

### 1.1.1 Mendelsche Experimente

Eine der ersten systematischen Arbeiten zur Vererbungslehre wurde im 19. Jahrhundert von Gregor Mendel geleistet. Unter anderem untersuchte Mendel die Vererbung einer Eigenschaft von Erbsen, nämlich ob die Erbsen eine glatte oder runzlige Oberfläche besitzen. Wie bei allen Pflanzen besitzt dabei jedes Individuum zwei Eltern (im Gegensatz beispielsweise zu Einzellern, die sich durch Zellteilung fortpflanzen).

Bei einer Untersuchung wurden in der so genannten *Elterngeneration* oder *Parentalgeneration* Erbsen mit *glatter* und Erbsen mit *runzliger* Oberfläche gekreuzt. Somit hatte in der nachfolgenden Generation, der so genannten *ersten Tochtergeneration* oder *ersten Filialgeneration* jede Erbse je ein Elternteil mit glatter und je ein Elternteil mit runzliger Oberfläche.

Überraschenderweise gab es bei den Nachkommen der Erbsen in der ersten Tochtergeneration nur noch glatte Erbsen. Man hätte wohl vermutet, dass sowohl glatte als auch runzlige Erbsen vorkommen oder aber leicht runzlige bzw. unterschiedlich runzlige Erbsen auftauchen würden.

Noch überraschender waren die Ergebnisse bei der nachfolgenden Tochtergeneration, der so genannten *zweiten Tochtergeneration* oder *zweiten Filialgeneration*, bei der nun beide Elternteile aus der ersten Tochtergeneration stammten. Hier kamen sowohl glatte als auch wieder runzlige Erbsen zum Vorschein. Interessanterweise waren jedoch die glatten Erbsen im Übergewicht, und zwar im Verhältnis 3 zu 1. Die Frage, die Mendel damals untersuchte, war, wie sich dieses Phänomen erklären lassen konnte.

### 1.1.2 Modellbildung

Als Modell schlug Gregor Mendel vor, dass die Erbse für ein bestimmtes Merkmal oder eine bestimmte Ausprägung von beiden Elternteilen je eine Erbinformation erhielt. Im Folgenden wollen wir eine kleinste Erbinformation als *Gen* bezeichnen. Zur Formalisierung bezeichen wir das Gen, das die glatte Oberfläche hervorruft mit $G$ und dasjenige für die runzlige Oberfläche mit $g$. Da nun nach unserem Modell jede Erbse von beiden Elternteilen ein Gen erhält, muss jedes Gen für ein Merkmal doppelt vorliegen. Zwei Erbinformationen, also Gene, die für dieselbe Ausprägung verantwortlich sind, werden als *Allel* bezeichnet. Wir nehmen also an, dass unsere glatten Erbsen in der Elterngeneration die Allele $GG$ und die runzligen die Allele $gg$ enthalten.

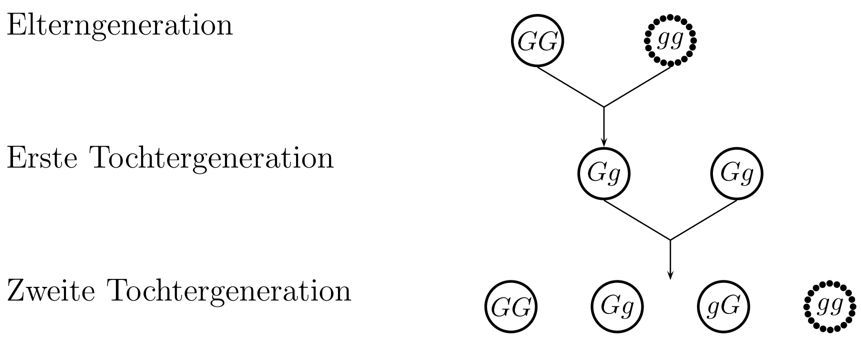



Abbildung 1.1: Skizze: Mendelsche Vererbung

Welche Erbinformation besitzt nun die erste Tochtergeneration? Sie erhält jeweils ein $G$ und ein $g$ von ihren Eltern und trägt als Erbinformation bezüglich der Oberfläche ein $Gg$. Was soll nun $Gg$ eigentlich sein? Wir wissen nur, dass $GG$ glatt und $gg$ runzlig bedeutet. Ein Organismus, der bezüglich einer Ausprägung, dieselbe Erbinformation trägt, wird als *reinerbig* oder *homozygot* bezeichnet.

Wir haben nun mit $Gg$ eine *mischerbige* oder *heterozygote* Erbinformation vorliegen. Wie oben bereits angedeutet, könnte die Ausprägung nun gemischt vorliegen, also ein „wenig runzlig", oder aber einer der beiden Allelen zufällig die Ausprägung bestimmen.

Werden die Merkmale in Mischformen vererbt, wie in „ein wenig runzlig", dann sagt man, dass das Merkmal *intermediär* vererbt wird. Mitglieder der ersten Tochtergeneration tragen dann also eine Mischung von beidem. Beispielsweise können die Nachfahren von Blumen mit roten bzw. weißen Blüten rosa-farbene Blüten besitzen oder aber auch weiße Blüten mit roten Tupfen etc.

Dies ist aber hier, wie die Experimente von Gregor Mendel gezeigt haben, nicht der Fall: Alle Erbsen der ersten Tochtergeneration sind glatt. Das bedeutet, dass beide Gene eines Allels gegeneinander konkurrieren und in Abhängigkeit der Gene sich immer eins der beiden als dominant behauptet und den Wettkampf gewinnt. In unserem Falle, setzt sich also das Gen für die glatte Oberfläche gegenüber dem Gen für die runzlige durch. Das Gen, das sich durchsetzt, wird als *dominant* bezeichnet, und dasjenige, das unterliegt, wird als *rezessiv* bezeichnet.

Da nun sowohl die Erbinformation $GG$ als auch $Gg$ für glatte Erbsen stehen, muss man zwischen den so genannten Phänotypen und den Genotypen unterscheiden. Als *Phänotyp* bezeichnet man die sichtbare Ausprägung, also z.B. glatt. Als *Genotyp* bezeichnet man die Zusammensetzung der Erbinformation, also z.B. $GG$ oder $Gg$ für glatte Erbsen. Insbesondere kann also der Genotyp unterschiedlich, aber der Phänotyp gleich sein, wie bei den glatten Erbsen in der Elterngeneration und in der ersten Tochtergeneration.

Wie kann man jetzt die Erscheinung in der zweiten Tochtergeneration erklären? Betrachten wir nun die Eltern, also die Erbsen der ersten Tochtergeneration, die als Genotyp $Gg$ tragen. Nimmt man nun an, das jedes Elternteil eines seiner Gene eines Allels zufällig (mit gleich hoher Wahrscheinlichkeit) an seine Kinder weitergibt, dann gibt es für die Erbsen der zweiten Tochtergeneration $2 \cdot 2 = 4$ Möglichkeiten, wie sich diese Gene dieses Alles vererben können (siehe Abbildung 1.2).

| | $G$ | $g$ |
|---|---|---|
| $G$ | $GG$ | $Gg$ |
| $g$ | $gG$ | $gg$ |

Abbildung 1.2: Skizze: Vererbung des Genotyps von zwei mischerbigen Eltern

Also sind drei der vier Kombinationen, die im Genotyp möglich sind ($gg$, $gG$ sowie $Gg$), im Phänotyp gleich, nämlich glatt. Nur eine der Kombinationen im Genotyp liefert im Phänotyp eine runzlige Erbse. Dies bestätigt in eindrucksvoller Weise das experimentell ermittelte Ergebnis, dass etwa dreimal so viele glatte wie runzlige Erbsen zu beobachten sind.

An dieser Stelle müssen wir noch anmerken, dass diese Versuche nur möglich sind, wenn man in der Elterngeneration wirklich reinerbige Erbsen zur Verfügung hat und keine mischerbigen. Auf den ersten Blick ist dies nicht einfach, da man ja nur den Phänotyp und nicht den Genotyp einfach ermitteln kann. Durch vielfache Züchtung kann man jedoch die Elternteile identifizieren, die reinerbig sind (nämlich, die runzligen sowie die glatten, deren Kinder und Enkelkinder nicht runzlig sind).

### 1.1.3 Mendelsche Gesetze

Fassen wir hier noch einmal kurz die drei so genannten Mendelschen Gesetze zusammen, auch wenn wir hier nicht alle bis ins Detail erläutert haben:

1) **Uniformitätsregel:** Werden zwei reinerbige Individuen einer Art gekreuzt, die sich in einem einzigen Merkmal unterscheiden, so sind alle Individuen der ersten Tochtergeneration gleich.

2) **Spaltungsregel:** Werden zwei Mischlinge der ersten Tochtergeneration miteinander gekreuzt, so spalten sich die Merkmale in der zweiten Tochtergeneration im Verhältnis 1 zu 3 bei dominant-rezessiven Genen und im Verhältnis 1 zu 2 zu 1 bei intermediären Genen auf.

3) **Unabhängigkeitsregel** Werden zwei mischerbige Individuen, deren Elterngeneration sich in zwei Merkmalen voneinander unterschieden hat, miteinander gekreuzt, so vererben sich die einzelnen Erbanlagen unabhängig voneinander.

Die Unabhängigkeitsregel gilt in der Regel nur, wenn die Gene auf verschiedenen Chromosomen sitzen bzw. innerhalb eines Chromosoms so weit voneinander entfernt sind, dass eine so genannte *Crossing-Over-Mutation* hinreichend wahrscheinlich ist.

### 1.1.4 Wo und wie sind die Erbinformationen gespeichert?

Damit haben wir die Grundlagen der Genetik ein wenig kennen gelernt. Es stellt sich jetzt natürlich noch die Frage, wo und wie die Gene gespeichert werden. Dies werden wir in den folgenden Abschnitten erläutern.

Zum Abschluss noch ein paar Notationen. Wie bereits erwähnt, bezeichnen wir ein *Gen* als den Träger einer kleinsten Erbinformation. Alle Gene eines Organismus zusammen bilden das *Genom*. Wie bereits aus der Schule bekannt sein dürfte, ist das Genom auf dem oder den *Chromosom(en)* gespeichert (je nach Spezies).

## 1.2 Chemische Grundlagen

Bevor wir im Folgenden auf die molekularbiologischen Grundlagen näher eingehen, wiederholen wir noch ein paar elementare Begriffe und Eigenschaften aus der Chemie bzw. speziell aus der organischen und der Biochemie. Die in der Biochemie wichtigsten auftretenden Atome sind Kohlenstoff (C), Sauerstoff (O), Wasserstoff (H), Stickstoff (N), Schwefel (S), Kalzium (Ca), Eisen (Fe), Magnesium (Mg), Kalium (K) und Phosphor (P). Diese Stoffe lassen sich beispielsweise mit folgendem Merkspruch behalten: **COHNS CaFe M***it* **g***roßem* **K***uchen*-**P***aket*. Zunächst einmal wiederholen wir kurz die wichtigsten Grundlagen der chemischen Bindungen.

### 1.2.1 Kovalente Bindungen

Die in der Biochemie wichtigste Bindungsart ist die *kovalente Bindung*. Hierbei steuern zwei Atome je ein Elektron bei, die dann die beiden Atome mittels einer gemeinsamen Bindungswolke zusammenhalten. Im Folgenden wollen wir den Raum, für den die Aufenthaltswahrscheinlichkeit eines Elektrons bzw. eines Elektronenpaares (nach dem Pauli-Prinzip dann mit verschiedenem Spin) am größten ist, als *Orbital* bezeichnen.

Hierbei sind die Kohlenstoffatome von besonderer Bedeutung, die die organische und Biochemie begründen. Die wichtigste Eigenschaft der Kohlenstoffatome ist, dass sie sowohl Einfach-, als auch Doppel- und Dreifachbindungen untereinander ausbilden können. Das Kohlenstoffatom hat in der äußersten Schale 4 Elektronen. Davon befinden sich im Grundzustand zwei in einem so genannten *s-Orbital* und zwei jeweils in einem so genannten *p-Orbital*.

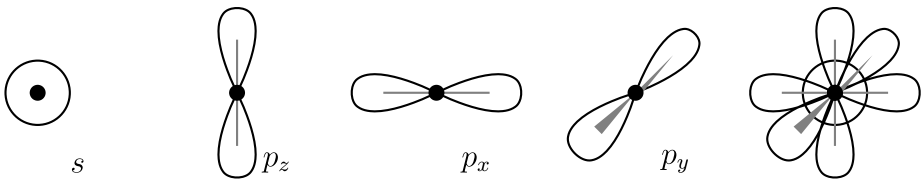



Abbildung 1.3: Skizze: Räumliche Ausdehnung der Orbitale

Das $s$-Orbital ist dabei kugelförmig, während die drei verschiedenen $p$-Orbitale jeweils eine *Doppelhantel* ausbilden, die paarweise orthogonal zueinander sind. In Abbildung 1.3 ist die räumliche Ausdehnung des $s$- und der drei $p$-Orbitale schematisch dargestellt, von denen jedes bis zu zwei Elektronen aufnehmen kann. Ganz rechts sind alle Orbitale gleichzeitig zu sehen, die in der Regel für uns interessant sein werden.

In Einfachbindungen befinden sich beim Kohlenstoffatom die einzelnen Elektronen in so genannten *$sp^3$-hybridisierten Orbitalen*, die auch als *$q$-Orbitale* bezeichnet werden. Hierbei bilden sich aus den 3 Hanteln und der Kugel vier energetisch äquivalente keulenartige Orbitale. Dies ist in der Abbildung 1.4 links dargestellt. Die Endpunkte der vier Keulen bilden dabei ein Tetraeder aus. Bei einer Einfachbindung überlappen sich zwei der Keulen, wie in Abbildung 1.4 rechts dargestellt. Die in der Einfachbindung überlappenden $q$-Orbitale bilden dann ein so genanntes *$\sigma$-Orbital*.

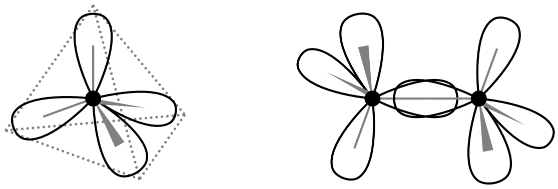


Abbildung 1.4: Skizze: $sp^3$ hybridisierte Orbitale sowie eine Einfachbindung

In Doppelbindungen sind nur zwei $p$-Orbitale und ein $s$-Orbital zu drei Keulen hybridisiert, so genannte *$sp^2$-Orbitale*. Ein $p$-Orbital bleibt dabei bestehen. Dies ist in Abbildung 1.5 links illustriert. Die in Doppelbindungen überlappenden $p$-Orbitale werden dann auch als *$\pi$-Orbital* bezeichnet. Bei einer Doppelbindung überlappen sich zusätzlich zu den zwei keulenförmigen hybridisierten $q$-Orbitalen, die die $\sigma$-Bindung bilden, auch noch die beiden Doppelhanteln der $p$-Orbitale, die dann das $\pi$-Orbital bilden. Dies ist schematisch in der Abbildung 1.5 rechts dargestellt (die Abbildungen sind nicht maßstabsgetreu).

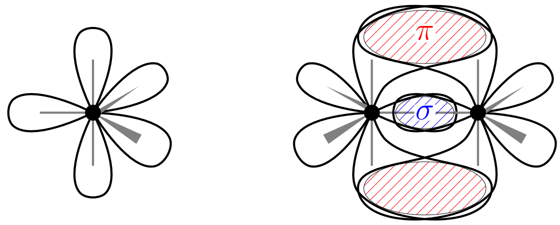



Abbildung 1.5: Skizze: $sp^2$ hybridisierte Orbitale und eine Doppelbindung

Die Bindung der Doppelbindung, die durch Überlappung von $q$-Orbitalen entsteht, wird auch *$\sigma$-Bindung* genannt, die Bindung der Doppelbindung, die durch Überlappung von $p$-Orbitalen entsteht, wird als *$\pi$-Bindung* bezeichnet. Ähnlich verhält es sich bei Dreifachbindungen, wo zwei $p$-Orbitale verbleiben und ein $s$- und nur ein $p$-Orbital zu einem *$sp$-Orbital* hybridisieren. Die konkrete Art der Hybridisierung der Orbitale der Kohlenstoffatome eines bestimmten Moleküls ist deshalb so wichtig, weil dadurch die dreidimensionale Struktur des Moleküls festgelegt wird. Die vier $sp^3$-Orbitale zeigen in die Ecken eines Tetraeders, die drei $sp^2$-Orbitale liegen in einer Ebene, auf der das verbleibende $p$-Orbital senkrecht steht, die zwei $sp$-Orbitale schließen einen Winkel von 180° ein und sind damit gerade gestreckt, auf ihnen stehen die verbleibenden beiden $p$-Orbitale senkrecht.

Bei zwei benachbarten Doppelbindungen (wie im Butadien, $H_2C{=}CH{-}HC{=}CH_2$) verbinden sich in der Regel die beiden benachbarten Orbitale, die die jeweilige $\pi$-Bindung zur Doppelbindung machen, um dann quasi eine $\pi$-Wolke über alle vier Kohlenstoffatome auszubilden, da dies energetisch günstiger ist. Daher spricht man bei den Elektronen in dieser verschmolzenen Wolke auch von *delokalisierten $\pi$-Elektronen*.

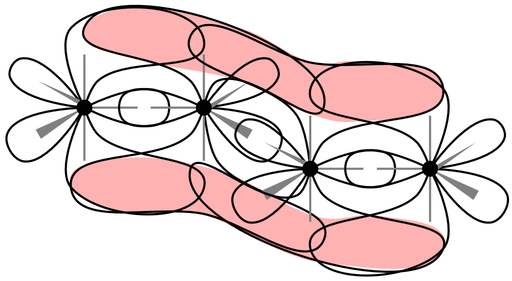


Abbildung 1.6: Skizze: Delokalisierte $\pi$-Bindung im Butadien

Ein Beispiel hierfür ist das *Benzol*-Molekül ($C_6H_6$). Aus energetischen Gründen bilden sich in dem Ring aus sechs Kohlenstoffatomen nicht drei einzelne alternierende Doppelbindungen aus, sondern eine große Wolke aus sechs delokalisierten $\pi$-Elektronen, die die starke Bindung des Benzolrings begründen.

Wie wir später noch sehen werden, kann sich eine Wolke aus delokalisierten $\pi$-Elektronen auch aus den $\pi$-Elektronen einer C=C Doppelbindung und dem nichtbindenden Orbital eines Sauerstoff- oder Stickstoffatoms bilden. Stickstoff bzw. Sauerstoff besitzen in der äußersten Schale mehr als vier Elektronen und daher kann sich ein $p$-Orbital mit zwei Elektronen ausbilden. Dieses hat dann bezüglich der Delokalisation von $\pi$-Elektronen ähnliche Eigenschaften wie eine $\pi$-Bindung.

Die Energie einer kovalenten Bindung variiert zwischen 200kJ/mol und 450kJ/mol (Kilojoule pro Mol), wobei Kohlenstoffatome untereinader relativ starke Bindungen besitzen (etwa 400kJ/mol). Die Angabe dieser absoluten Werte ist für uns eigentlich nicht von Interesse. Wir geben sie hier nur an, um die Stärken der verschiedenen Bindungsarten im Folgenden vergleichen zu können.

### 1.2.2 Ionische Bindungen

Bei *ionischen Bindungen* gibt ein Atom, das so genannte *Donatoratom*, ein Elektron an ein anderes Atom, das so genannte *Akzeptoratom*, ab. Damit sind die Donatoratome positiv und die Akzeptoratome negativ geladen. Durch die elektrostatische Anziehungskraft (und die Abstoßung gleichnamiger Ladung) bildet sich in der Regel ein Kristallgitter aus, das dann abwechselnd aus positiv und negativ geladenen Atomen besteht.

Ein bekanntes Beispiel hierfür ist Kochsalz, d.h. Natriumchlorid (NaCl). Dabei geben die Natriumatome jeweils das äußerste Elektron ab, das dann von den Chloratomen aufgenommen wird. Dadurch sind die Natriumatome positiv und die Chloratome negativ geladen, die sich dann innerhalb eines Kristallgitters anziehen.

Hier wollen wir noch deutlich den Unterschied herausstellen, ob wir diese Bindungen in wässriger Lösung oder ohne Lösungsmittel betrachten. Ohne Wasser als Lösungsmittel sind ionische Bindungen sehr stark. In wässriger Lösung sind sie jedoch sehr schwach, sie werden etwa um den Faktor 80 schwächer. Beispielsweise löst sich das doch recht starke Kristallgitter des Kochsalzes im Wasser nahezu auf. In wässriger Lösung beträgt die Energie einer ionischen Bindung etwa 20 kJ/mol.

### 1.2.3 Wasserstoffbrücken

Eine andere für uns sehr wichtige Anziehungskraft, die keine Bindung im eigentlichen chemischen Sinne ist, sind die *Wasserstoffbrücken*. Diese Anziehungskräfte werden im Wesentlichen durch die unterschiedlichen Elektronegativitäten der einzelnen Atome bedingt.

Die Elektronegativität ist ein Maß dafür, wie stark die Elektronen in der äußersten Schale angezogen werden. Im Periodensystem der Elemente wächst der Elektronegativitätswert innerhalb einer Periode von links nach rechts, weil dabei mit der Anzahl der Protonen auch die Kernladung und damit auch die Anziehungskraft auf jedes einzelne Elektron ansteigt. Innerhalb einer Hauptgruppe sinkt die Elektronegativität mit zunehmender Ordnungszahl, weil die Außenelektronen sich auf immer höheren Energieniveaus befinden und der entgegengesetzt geladene Kern durch die darunterliegenden Elektronen abgeschirmt wird. Deshalb ist z.B. Fluor das Element mit dem größten Elektronegativitätswert.

Eine Liste der für uns wichtigsten Elektronegativitäten in für uns willkürlichen Einheiten ist in Abbildung 1.7 angegeben. Hier bedeutet ein größerer Wert eine größere Affinität zu Elektronen.

| Atom | $C$ | $O$ | $H$ | $N$ | $S$ | $P$ |
|---|---|---|---|---|---|---|
| EN | 2.5 | 3.5 | 2.1 | 3.0 | 2.5 | 2.1 |

Abbildung 1.7: Tabelle: Elektronegativitäten nach Pauling

Bei einer kovalenten Bindung sind die Elektronenwolken in Richtung des Atoms mit der stärkeren Elektronegativität hin verschoben. Dadurch bekommt dieses Atom eine teilweise negative Ladung, während das andere teilweise positiv geladen ist. Ähnlich wie bei der ionischen Bindung, wenn auch bei weitem nicht so stark, wirkt diese Polarisierung der Atome anziehend.

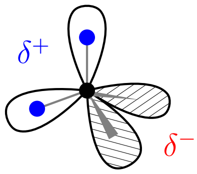



Abbildung 1.8: Skizze: Polarität bei einem Wassermolekül

Insbesondere Wasser ist für die Ausbildung von zahlreichen Wasserstoffbrücken bekannt. In Abbildung 1.8 ist ein Wassermolekül schematisch dargestellt.

Wie beim Kohlenstoffatom sind die drei $p$-Orbitale und das $s$ Orbital zu vier $q$-Orbitalen hybridisiert. Da das Sauerstoffatom in der äußersten Schale sechs anstatt vier Elektronen besitzt, sind bereits zwei der $q$-Orbitale des Sauerstoffatoms mit je zwei Elektronen besetzt und können daher keine kovalente Bindung eingehen. Man bezeichnet diese Orbitale daher auch als *nichtbindend*.

Die beiden anderen werden im Wasser gemäß der Formel $H_2O$ mit jeweils einem Wasserstoffatom protoniert. Da nun die beiden nichtbindenden Orbitale (zumindest aus dieser Richtung auf das Sauerstoffatom) negativ geladen sind und die beiden protonierten bindenden Orbitale positiv geladen sind, wirkt das Wasser als Dipol und die Wassermoleküle hängen sich wie viele kleine Stabmagneten aneinander. Einziger Unterschied ist hier dass die Teilladungen in den Ecken eine Tetraeders sitzen, so dass sich die Wassermoleküle ähnlich wie die Kohlenstoffatome im Kristallgitter des Diamanten anordnen.

Im gefrorenen Zustand ist das Kristallgitter von Wasser (also Eis) nahezu ein Diamantengitter, während im flüssigen Zustand die Wasserstoffbrücken häufig aufbrechen und sich wieder neu bilden. Daher ist es zum einen flüssig, und zum anderen kann es im flüssigen Zustand dichter gepackt werden als im gefrorenen Zustand. Erst dadurch nimmt Wasser den flüssigen Zustand bei Zimmertemperatur an, während sowohl Wasserstoff wie auch Sauerstoff einen sehr niedrigen Siedepunkt besitzen. Eine Wasserstoffbrückenbindung ist mit ca. 21 kJ/mol deutlich schwächer als eine kovalente oder eine ionische Bindung.

### 1.2.4 Van der Waals-Kräfte

Die *Van der Waals-Anziehung* bzw. *Van der Waals-Kräfte* treten insbesondere in großen bzw. langen Molekülen, wie Kettenkohlenwasserstoffen auf. Da der Ort der Elektronen ja nicht festgelegt ist (Heisenbergsche Unschärferelation), können sich durch die Verlagerung der Elektronen kleine Dipolmomente in den einzelnen Bindungen ergeben. Diese beeinflussen sich gegenseitig und durch positive Rückkopplungen können diese sich verstärken. Somit können sich lange Moleküle fester aneinander legen als kürzere. Dies ist mit ein Grund dafür, dass die homologe Reihe der Alkane ($C_nH_{2n+2}$) mit wachsender Kohlenstoffanzahl bei Zimmertemperatur ihren Aggregatzustand von gasförmig über flüssig bis zu fest ändert. Die Energie der Van der Waals-Kraft liegt bei etwa 4kJ/mol.

### 1.2.5 Hydrophobe Kräfte

Nichtpolare Moleküle, wie Fette, können mit Wasser keine Wasserstoffbrücken ausbilden. Dies ist der Grund, warum nichtpolare Stoffe in Wasser unlöslich sind. Solche Stoffe werden auch als *hydrophob* bezeichnet, während polare Stoffe auch als *hydrophil* bezeichnet werden. Aufgrund der Ausbildung von zahlreichen Wasserstoffbrücken innerhalb des Wassers (auch mit hydrophilen Stoffen) tendieren hydrophobe Stoffe dazu, sich möglichst eng zusammenzulagern, um eine möglichst kleine Oberfläche (gleich Trennfläche zum Wasser) auszubilden. Diese Tendenz des Zusammenlagerns hydrophober Stoffe in wässriger Lösung wird als *hydrophobe Kraft* bezeichnet.

### 1.2.6 Funktionelle Gruppen

Wie schon zu Beginn bemerkt spielt in der organischen und Biochemie das Kohlenstoffatom die zentrale Rolle. Dies liegt insbesondere daran, dass es sich mit sich selbst verbinden kann und sich so eine schier unendliche Menge an verschiedenen Molekülen konstruieren lässt. Dabei sind jedoch auch andere Atome beteiligt, sonst erhalten wir bekanntlich Graphit oder den Diamanten.

| Chem. Rest | Gruppe | gewöhnlicher Name |
|---|---|---|
| $-CH_3$ | Methyl | |
| –OH | Hydroxyl | Alkohol |
| $-NH_2$ | Amino | Amine |
| –NH– | Imino | |
| –CHO | Carbonyl | Aldehyde |
| –CO– | Carbonyl | Ketone |
| –COO– | Ester | Ester |
| –COOH | Carboxyl | organische Säure |
| –CN | Cyanid | Nitrile |
| –SH | Sulfhydril | Thiole |

Abbildung 1.9: Tabelle: Einige funktionelle (organische) Gruppen

Um diese anderen vorkommenden Atome bezüglich ihrer dem Molekül verleihenden Eigenschaften ein wenig besser einordnen zu können, beschreiben wir die am meisten vorkommenden *funktionellen Gruppen*. Die häufigsten in der Biochemie auftretenden einfachen funktionellen Gruppen sind in der Tabelle 1.9 zusammengefasst.

### 1.2.7 Stereochemie und Enantiomerie

In diesem Abschnitt wollen wir einen kurzen Einblick in die *Stereochemie*, speziell in die *Enantiomerie* geben. Die Stereochemie beschäftigt sich mit der räumlichen Anordnung der Atome in einem Molekül. Beispielsweise ist entlang einer Einfachbindung die Rotation frei möglich. Bei Doppelbindungen ist diese aufrund der $\pi$-Bindung eingeschränkt und es kann zwei mögliche räumliche Anordnungen desselben Moleküls geben. In Abbildung 1.10 sind zwei Formen für Äthendiol angegeben. Befinden sich beide (der bedeutendsten) funktionellen Gruppen auf derselben Seite der Doppelbindung, so spricht man vom *cis-Isomer* andernfalls von *trans-Isomer*. Bei der cis-trans-Isomerie kann durch Energiezufuhr die Doppelbindung kurzzeitig


cis trans

Abbildung 1.10: Skizze: Cis-Trans-Isomerie bei Äthendiol

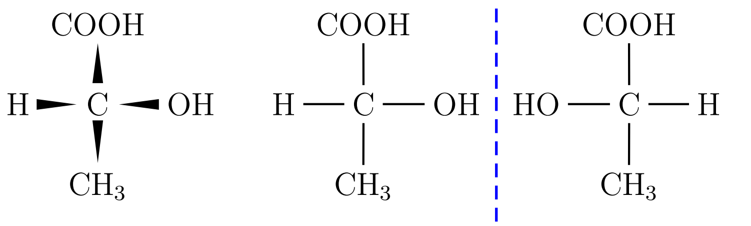
geöffnet werden und um 180° gedreht werden, so dass die beiden Isomere ineinander überführt werden können.

Es hat sich herausgestellt, dass scheinbar identische Stoffe (aufgrund der Summen- und Strukturformel) sich unter bestimmten Bedingungen unterschiedlich verhalten können. Dies sind also solche Isomere, die sich nicht ineinander überführen lassen. Betrachten wir dazu in Abbildung 1.11 ein Kohlenstoffatom (schwarz darge-

Abbildung 1.11: Skizze: Asymmetrisches Kohlenstoffatom

stellt) und vier unterschiedliche funktionelle Gruppen (farbig dargestellt), die jeweils mittels einer Einfachbindung an das Kohlenstoffatom gebunden sind. Das betrachtete Kohlenstoffatom wird hierbei oft als *zentrales Kohlenstoffatom* bezeichnet. Auf den ersten Blick sehen die beiden Moleküle in Abbildung 1.11 gleich aus. Versucht man jedoch, die beiden Moleküle durch Drehungen im dreidimensionalen Raum zur Deckung zu bringen, so wird man feststellen, dass dies gar nicht geht. Die beiden Moleküle sind nämlich Spiegelbilder voneinander.

Daher werden Moleküle als *chiral* (deutsch Händigkeit) bezeichnet, wenn ein Molekül mit seinem Spiegelbild nicht durch Drehung im dreidimensionalen Raum zur Deckung gebracht werden kann. Die beiden möglichen, zueinander spiegelbildlichen Formen nennen wir *Enantiomere*. Beide Formen werden auch als *enantiomorph* zueinander bezeichnet. Für Kohlenstoffatome, die mit vier *unterschiedlichen* Resten verbunden sind, gilt dies immer. Aus diesem Grund nennt man ein solches Kohlenstoffatom auch ein *asymmetrisches Kohlenstoffatom*.



Milchsäure D-(−)-Milchsäure L-(+)-Milchsäure

Abbildung 1.12: Skizze: Milchsäure

Ein einfaches (und bekanntes) Beispiel hierfür ist die Milchsäure. Hierbei sind die funktionellen Gruppen, die an einem zentralen Kohlenstoffatom sitzen, ein Wasserstoffatom, eine Hydroxyl-, eine Methyl und eine Carboxylgruppe. In Abbildung 1.12 sind rechts die beiden Formeln der beiden spiegelbildlichen Formen dargestellt. In einer zweidimensionalen Abbildung muss man jedoch eine Konvention einführen, wie man die Projektion vornimmt. Die längste Kohlenstoffkette wird dabei von oben nach unten dargestellt. Hier also das zentrale Kohlenstoffatom und das Kohlenstoffatom der Methylgruppe. Dabei wird oben von der charakteristischsten funktionellen Gruppe, hier die Carboxylgruppe, bestimmt. Dabei ist zu verstehen, dass die vertikale Kette hinter dem zentralen Kohlenstoffatom unter der Papierebene verschwindet, während die beiden restlichen Seitenketten aus der Papiereben dem Leser entgegen kommen (dies wird als *Fischer-Projektion* bezeichnet). Dies ist in der Abbildung 1.12 ganz links noch einmal illustriert.

Da nun im mittleren Teil der Abbildung 1.12 die bedeutendere funktionelle Gruppe, also die Hydroxylgruppe gegenüber dem Wasserstoffatom, rechts sitzt, wird dieses Enantiomer mit D-Milchsäure (latein. dexter, rechts) bezeichnet. Rechts handelt es sich dann um die L-Milchsäure (latein. laevis, links).

Diese Bezeichnungsweise hat nichts mit den aus der Werbung bekannten links- bzw. rechtsdrehenden Milchsäuren zu tun. Die Namensgebung kommt von der Tatsache, dass eine Lösung von Milchsäure, die nur eines der beiden Enantiomere enthält, polarisiertes Licht dreht. Dies gilt übrigens auch für die meisten Moleküle, die verschiedene Enantiomere besitzen. Je nachdem, ob es polarisiertes Licht nach rechts oder links verdreht, wird es als *rechtsdrehend* oder *linksdrehend* bezeichnet (und im Namen durch (+) bzw. (−) ausgedrückt).

Bei der Milchsäure stimmen zufällig die D- bzw. L-Form mit der rechts- bzw. linksdrehenden Eigenschaft überein. Bei Aminosäuren, die wir noch kennen lernen werden, drehen einige L-Form nach rechts! Hier haben wir einen echten Unterschied gefunden, mit dem sich Enantiomere auf makroskopischer Ebene unterscheiden lassen.

In der Chemie wurde die *DL-Nomenklatur* mittlerweile zugunsten der so genannten *RS-Nomenklatur* aufgegeben. Da jedoch bei Zuckern und Aminosäuren oft noch die DL-Nomenklatur verwendet wird, wurde diese hier angegeben. Für Details bei der DL - sowie der RS-Nomenklatur verweisen wir auf die einschlägige Literatur.

### 1.2.8 Tautomerien

*Tautomerien* sind intramolekulare Umordnungen von Atomen. Dabei werden chemisch zwei verschiedene Moleküle ineinander überführt. Wir wollen dies hier nur exemplarisch am Beispiel der *Keto-Enol-Tautomerie* erklären. Wir betrachten dazu die folgende Abbildung 1.13

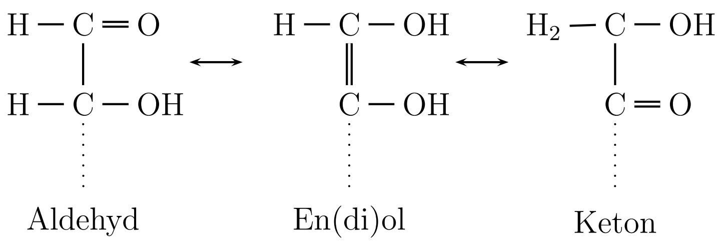



Abbildung 1.13: Skizze: Keto-Enol-Tautomerie

Im Aldehyd sind aufgrund der Doppelbindung in der Carbonylgruppe und der daraus resultierenden starken Elektronegativität die Elektronen zum Sauerstoffatom der Carbonylgruppe verschoben. Dies führt induktiv zu einer Verlagerung der Elektronen in der C-C Bindung zum Kohlenstoffatom der Carbonylgruppe. Auf der anderen Seite sind die anderen Elektronen im Bindungsorbital zur Hydroxylgruppe aufgrund derer starken Elektronegativität zum Sauerstoffatom hin verschoben. Dadurch lässt sich das Wasserstoffatom am zweiten Kohlenstoffatom sehr leicht trennen und kann eines der nichtbindenden Orbitale des Sauerstoffatoms der benachbarten Carbonylgruppe protonieren. Diese wandelt sich somit zu einer Hydroxylgruppe und es entsteht zwischen den beiden Kohlenstoffatomen eine Doppelbindung.

Man beachte hierbei, dass die bindenden Orbitale der $\pi$-Bindung und die nichtbindenden Orbitale der angrenzenden Sauerstoffatome sich jetzt ebenfalls überlappen, um für die darin enthaltenen Elektronen ein größeres Orbital bereitzustellen. Durch eine Delokalisierung dieser Elektronen kann sich zwischen dem zweiten Kohlenstoffatom und dem Sauerstoffatom der Hydroxylgruppe eine Carbonylgruppe ausbilden. Das frei werdende Wasserstoffatom wird dann unter Aufbruch der Doppelbindung am ersten Kohlenstoffatom angelagert.

Aus ähnlichen Gründen kann sich diese intramolekulare Umlagerung auch auf dem Rückweg abspielen, so dass sich hier ein Gleichgewicht einstellt. Das genaue Gleichgewicht kann nicht pauschal angegeben werden, da es natürlich auch von den speziellen Randbedingungen abhängt. Wie schon erwähnt gibt es auch andere Tautomerien, d.h. intramolekulare Umlagerungen bei anderen Stoffklassen.

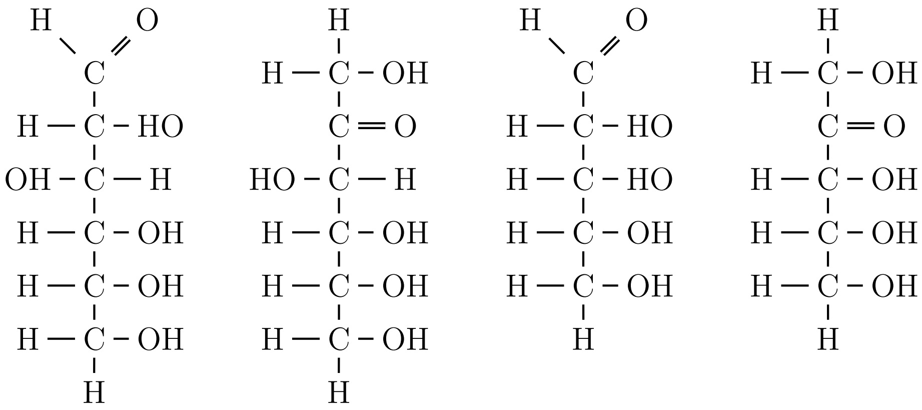
## 1.3 DNS und RNS

In diesem Abschnitt wollen wir uns um die chemische Struktur der *Desoxyribonukleinsäure* oder kurz *DNS* bzw. *Ribonukleinsäure* oder kurz *RNS* (engl. *deoxyribonucleic acid*, *DNA* bzw. *ribonucleic acid*, *RNA*) kümmern. In der DNS wird die eigentliche Erbinformation gespeichert und diese wird durch die RNS zur Verarbeitung weitergegeben. Wie diese Speicherung und Weitergabe im Einzelnen geschieht, werden wir später noch genauer sehen.

### 1.3.1 Zucker

Ein wesentlicher Bestandteil der DNS sind Zucker. Chemisch gesehen sind Zucker Moleküle mit der Summenformel $C_nH_{2n}O_n$ (weshalb sie oft auch als *Kohlenhydrate* bezeichnet werden). In Abbildung 1.14 sind die für uns wichtigsten Zucker in der Strukturformel dargestellt. Für uns sind insbesondere Zucker mit 5 oder 6 Kohlenstoffatomen von Interessen. Zucker mit 5 bzw. 6 Kohlenstoffatomen werden auch *Pentosen* bzw. *Hexosen* genannt.

Jeder Zucker enthält eine Carbonylgruppe, so dass Zucker entweder ein Aldehyd oder ein Keton darstellen. Daher werden Zucker entsprechend auch als *Aldose* oder



| Hexose | Hexose | Pentose | Pentose |
|---|---|---|---|
| Aldose | Ketose | Aldose | Ketose |
| D-Glucose | D-Fructose | D-Ribose | D-Ribulose |

Abbildung 1.14: Skizze: Zucker (Hexosen und Pentosen sowie Aldosen und Ketosen)

*Ketose* bezeichnet. Diese Unterscheidung ist jedoch etwas willkürlich, da aufgrund der Keto-Enol-Tautomerie eine Aldose in eine Ketose überführt werden kann. In der Regel pendelt sich ein Gleichgewicht zwischen beiden Formen ein. An dieser Stelle wollen wir noch anmerken, dass es sich bei allen Kohlenstoffatomen (bis auf das erste und das letzte) um asymmetrische Kohlenstoffatome handelt. Somit bilden Zucker Enantiomere aus.

Warum haben jetzt eigentlich Glucose und Fructose unterschiedliche Namen, obwohl diese aufgrund der Keto-Enol-Tautomerie im Gleichgewicht miteinander stehen? In der Natur treten die Zucker kaum als Aldose oder Ketose auf. Die Aldehyd- bzw. Ketogruppe bildet mit einer der Hydroxylgruppen einen Ring aus. In der Regel sind diese 5er oder 6er Ringe. Als 5er Ringe werden diese Zucker als *Furanosen* (aufgrund ihrer Ähnlichkeit zu *Furan*) bezeichnet, als 6er Ringe als *Pyranosen* (aufgrund ihrer Ähnlichkeit zu *Pyran*).

Bei den Hexosen (die hauptsächlich in gewöhnlichen Zuckern und Stärke vorkommen) wird der Ringschluss über das erste und vierte bzw. fünfte Kohlenstoffatom gebildet. Bei Pentosen (mit denen wir uns im Folgenden näher beschäftigen wollen) über das erste und vierte Kohlenstoffatom. Dabei reagiert die Carbonylgruppe mit der entsprechenden Hydroxylgruppe zu einem so genannten *Halb-Acetal*, wie in der folgenden Abbildung 1.15 dargestellt. Aus der Carbonylgruppe entsteht dabei die so genannte glykosidische OH-Gruppe. Die Ausbildung zum Voll-Acetal geschieht über eine weitere Reaktion (Kondensation) dieser Hydroxylgruppe am zentralen Kohlenstoffatom der ehemaligen Carbonylgruppe.

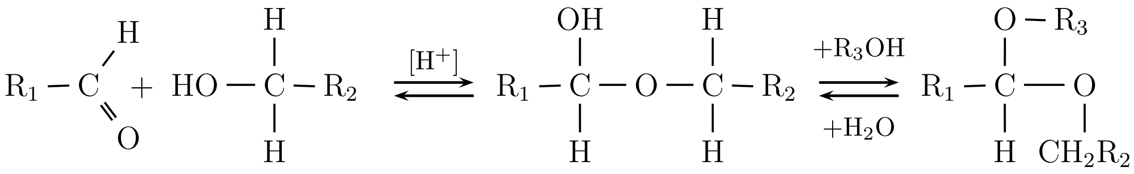



Abbildung 1.15: Skizze: Halb-Acetal- und Voll-Acetal-Bildung

In der Abbildung 1.16 sind zwei Furanosen, nämlich *Ribose* und *Desoxyribose* dargestellt. Der einzige Unterschied ist das quasi fehlende Sauerstoffatom am zweiten Kohlenstoffatom, dort ist eine Hydroxylgruppe durch ein Wasserstoffatom ersetzt. Daher stammt auch der Name *Desoxy*ribose. Wie man aus dem Namen schon vermuten kann tritt die Desoxyribose in der Desoxyribonukleinsäure (DNS) und die Ribose in der Ribonukleinsäure (RNS) auf. Die Kohlenstoffatome werden dabei zur Unterscheidung von 1 bis 5 durchnummeriert. Aus später verständlich werdenden Gründen, verwenden wir eine gestrichene Nummerierung 1' bis 5' (siehe auch Abbildung 1.16).

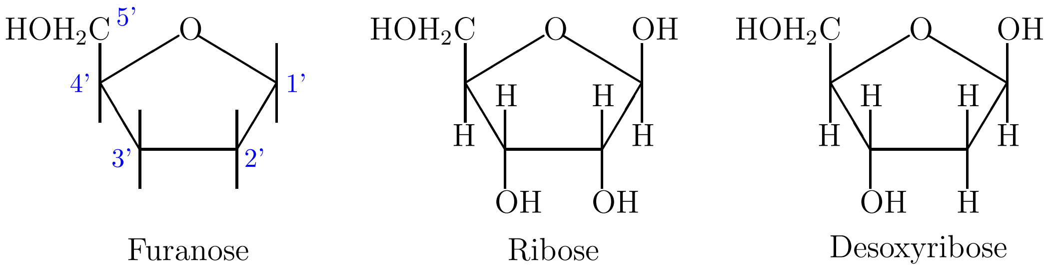



Abbildung 1.16: Skizze: Ribose und Desoxyribose als Furanosen

## 1.3.2 Basen

In diesem Abschnitt wollen wir einen weiteren wesentlichen Bestandteil der DNS bzw. RNS vorstellen, die so genannten *Basen*. Hiervon gibt es fünf verschiedene: Adenin, Guanin, Cytosin, Thymin und Uracil. Betrachten wir zuerst die von Purin abgeleiteten Basen. In Abbildung 1.17 ist links das Purin dargestellt und in der Mitte bzw. rechts *Adenin* bzw. *Guanin*. Die funktionellen Gruppen, die Adenin bzw. Guanin von Purin unterscheiden, sind rot dargestellt. Man beachte auch, dass sich durch die Carbonylgruppe im Guanin auch die Doppelbindungen in den aromatischen Ringen formal ändern. Da es sich hierbei jedoch um alternierende Doppelbindungen und nichtbindende Orbitale handelt, sind die Elektronen sowieso über die aromatischen Ringe delokalisiert.

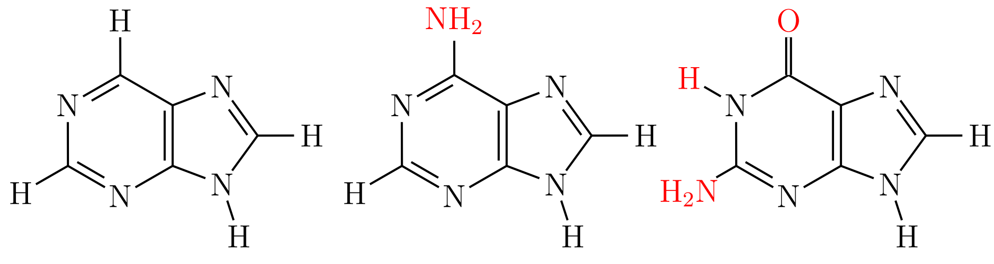



Abbildung 1.17: Skizze: Purine

Eine weitere Gruppe von Basen erhält man aus Pyrimidin, dessen Strukturformel in der Abbildung 1.18 links abgebildet ist. Hiervon werden *Cytosin*, *Thymin* und *Uracil* abgeleitet. Auch hier sind die funktionellen Gruppen, die den wesentlichen Unterschied zu Pyrimidin ausmachen, wieder rot bzw. orange dargestellt. Man beachte, dass sich Thymin und Uracil nur in der orange dargestellten Methylgruppe unterscheiden. Hier ist insbesondere zu beachten, dass Thymin nur in der DNS und Uracil nur in der RNS vorkommt. Auch hier beachte man, dass sich durch die Carbonyl-

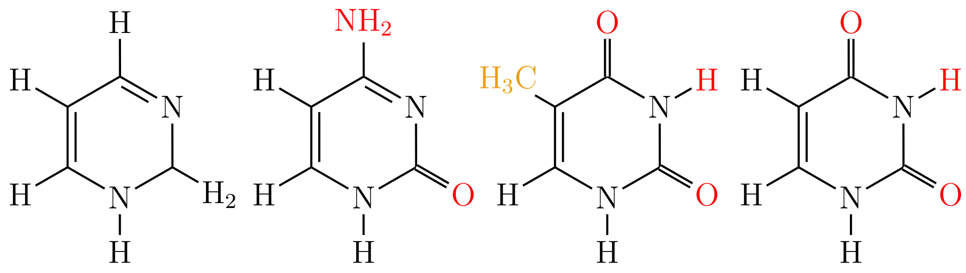



Abbildung 1.18: Skizze: Pyrimidine

gruppe im Thymin bzw. Uracil auch die Doppelbindungen in den aromatischen Ringen formal ändern. Jedoch bleiben auch hier die Elektronen über die aromatischen Ringe delokalisiert.

Ohne an dieser Stelle im Detail darauf einzugehen, merken wir noch an, dass Guanin über die Keto-Enol-Tautomerie mit einem ähnlichen Stoff in Wechselwirkung steht, ebenso Cytosin über eine so genannte Amino-Imino-Tautomerie. Wir kommen später noch einmal kurz darauf zurück.

Wie im Zucker werden auch die Atome in den aromatischen Ringen durchnummeriert. Da wir im Folgenden auf diese Nummerierung nie zurückgreifen werden, geben wir sie an dieser Stelle auch nicht an. Um eine Verwechslung mit der Nummerierung in den Zuckern zu vermeiden, wurde die Nummerierung in den Zuckern gestrichen durchgeführt.

### 1.3.3 Polymerisation

Nun haben wir die wesentlichen Bausteine der DNS bzw. RNS kennen gelernt: die Zucker Desoxyribose bzw. Ribose sowie die Basen Adenin, Guanin, Cytosin und Thymin bzw. Uracil. Zusätzlich spielt noch die Phosphorsäure $H_3PO_4$ eine Rolle. Je ein Zucker, eine Base und eine Phosphorsäure reagieren zu einem so genannten *Nukleotid*, das sich dann seinerseits mit anderen Nukleotiden zu einem Polymer verbinden kann.

Dabei wird das Rückgrat aus der Phosphorsäure und einem Zucker gebildet, d.h. der Desoxyribose bei DNS und der Ribose bei RNS, gebildet. Dabei reagiert die Phosphorsäure (die eine mehrfache Säure ist, da sie als Donator bis zu drei Wasserstoffatome abgeben kann) mit den Hydroxylgruppen der Zucker zu einer Esterbindung. Eine Bindung wird über die Hydroxylgruppe am fünften Kohlenstoffatom, die andere am dritten Kohlenstoffatom der (Desoxy-)Ribose gebildet. Somit ergibt sich für das Zucker-Säure-Rückgrat eine Orientierung.

Die Basen werden am ersten Kohlenstoffatom der (Desoxy-)Ribose über eine glykosidische Bindung (zum bereits erwähnten Voll-Acetal) angebunden. Eine Skizze eines Teilstranges der DNS ist in Abbildung 1.19 dargestellt. Man beachte, dass das Rückgrat für alle DNS-Stränge identisch ist. Die einzige Variabilität besteht in der Anbindung der Basen an die (Desoxy-)Ribose. Eine Kombination aus Zucker und Base (also ohne eine Verbindung mit der Phosphorsäure) wird als *Nukleosid* bezeichnet.

### 1.3.4 Komplementarität der Basen

Zunächst betrachten wir die Basen noch einmal genauer. Wir haben zwei Purin-Basen, Adenin und Guanin, sowie zwei Pyrimidin-Basen, Cytosin und Thymin (bzw. Uracil in der RNS). Je zwei dieser Basen sind *komplementär* zueinander. Zum einen sind Adenin und Thymin komplementär zueinander und zum anderen sind es Guanin und Cytosin. Die Komplementarität erklärt sich daraus, dass diese Paare untereinander Wasserstoffbrücken ausbilden können, wie dies in Abbildung 1.20 illustriert ist.

Dabei stellen wir fest, dass Adenin und Thymin zwei und Cytosin und Guanin drei Wasserstoffbrücken bilden. Aus energetischen Gründen werden diese Basen immer

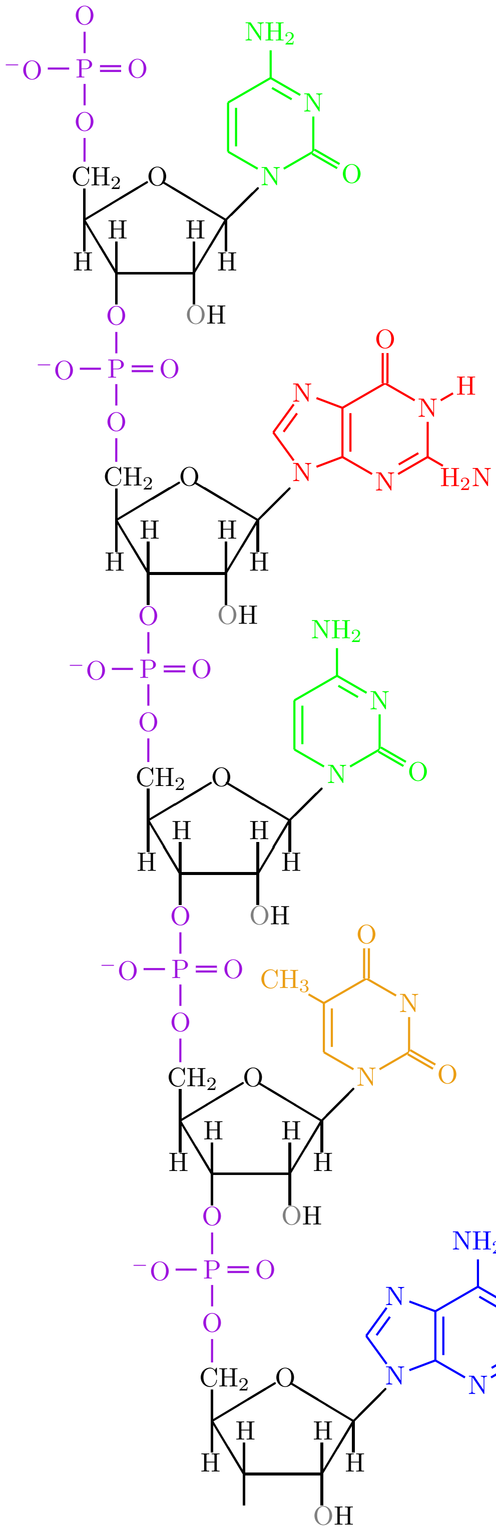


Abbildung 1.19: Skizze: DNS bzw. RNS als Polymerstrang

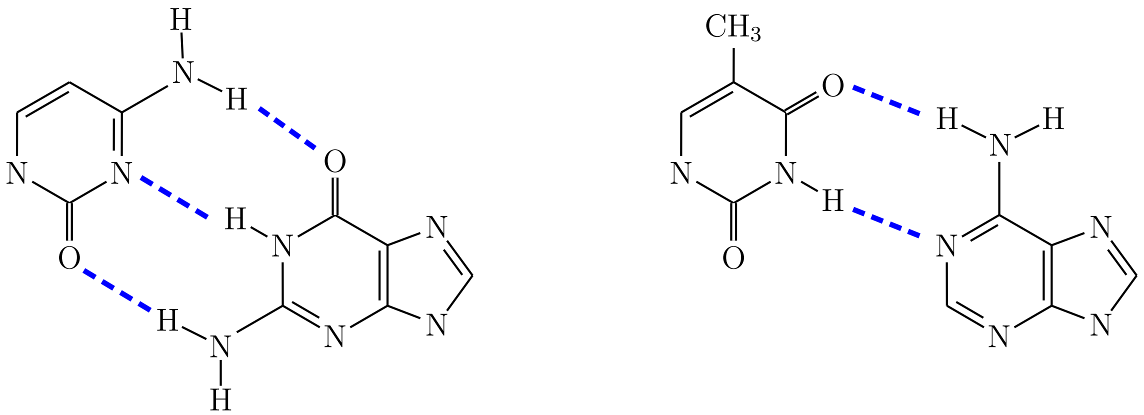


Abbildung 1.20: Skizze: Wasserstoffbrücken der komplementären Basen

versuchen, diese Wasserstoffbrücken auszubilden. Wir merken, an dass sich auch andere Brückenverbindungen ausbilden können, wie zwischen Thymin und Guanin sowie zwischen Adenin und Cytosin. Diese „falschen“ Wasserstoffbrücken sind aufgrund der Keto-Enol-Tautomerie von Guanin und der Amino-Imino-Tautomerie von Cytosin möglich. Diese sind aus energetischen Gründen zwar eher unwahrscheinlich, können aber dennoch zu Mutationen führen.

Als einfache Merkregel kann man sich merken, dass runde Buchstaben (C und G) bzw. eckige (A und T) zueinander komplementär sind. In der RNS ersetzt Uracil die Base Thymin, so dass man die Regel etwas modifizieren muss. Alles was wie C aussieht ist komplementär (C und G) bzw. alles was wie U aussieht (U und A).

### 1.3.5 Doppelhelix

Frühe Untersuchungen haben gezeigt, dass in der DNS einer Zelle die Menge von Adenin und Thymin sowie von Cytosin und Guanin immer gleich groß sind. Daraus kam man auf die Schlussfolgerung, dass diese Basen in der DNS immer in Paaren auftreten. Aus dem vorherigen Abschnitt haben wir mit der Komplementarität aufgrund der Wasserstoffbrücken eine chemische Begründung hierfür gesehen. Daraus wurde die Vermutung abgeleitet, dass die DNS nicht ein Strang ist, sondern aus zwei komplementären Strängen gebildet wird, die einander gegenüber liegen.

Aus sterischen Gründen liegen diese beiden Stränge nicht wie Gleise von Eisenbahnschienen parallel nebeneinader (die Schwellen entsprechen hierbei den Wasserstoffbrücken der Basen), sondern sind gleichförmig miteinander verdrillt. Jedes Rückgrat bildet dabei eine Helix (Schraubenlinie) aus. Eine schematische Darstellung ist in Abbildung 1.21 gegeben.

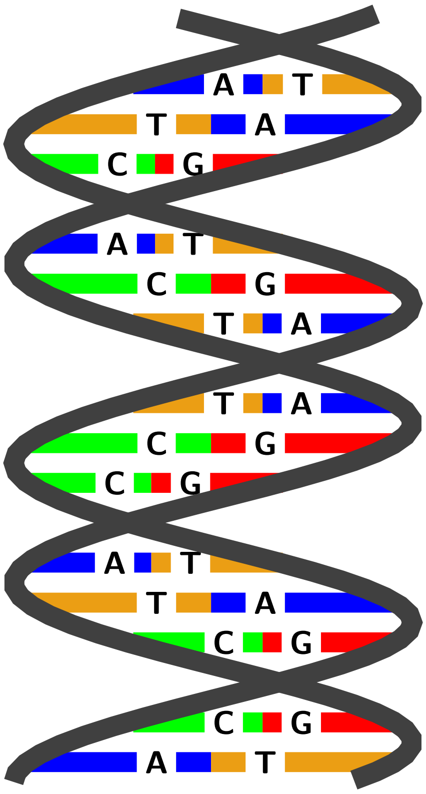



Abbildung 1.21: Skizze: Doppelhelix der DNS

In einer vollen Drehung sind ungefähr 10 Basenpaare involviert, wobei die Ebenen der Purine bzw. Pyrimidine in etwa orthogonal zur Achse der Doppelhelix liegen. Diese Struktur wurde 1953 von Watson und Crick mit Hilfe der Röntgenkristallographie bestätigt. Es sollte auch noch angemerkt werden, dass unter anderen Randbedingungen auch noch andere Formen von Doppelhelices ausgebildet werden können.

Wie wir schon gesehen haben, besitzt das Rückgrat eines DNS-Strangs eine Orientierung (von 5' nach 3'). Genaue sterische Untersuchungen haben gezeigt, dass die beiden Stränge der DNS innerhalb einer Doppelhelix gegenläufig sind. Läuft also der eine Strang quasi von unten nach oben, so läuft der andere von oben nach unten. Ferner haben die beiden Stränge keinen maximalen Abstand voneinander. Betrachtet man die Doppelhelix der DNS aus etwas „größerem" Abstand (wie etwa in der schematischen Zeichnung in Abbildung 1.21), so erkennt man etwas, wie ein kleinere und eine größere Furche auf einer Zylinderoberfläche.

Zum Abschluss noch ein paar Fakten zur menschlichen DNS. Die DNS des Menschen ist nicht eine lange DNS, sondern in 46 unterschiedlich lange Teile zerlegt. Insgesamt sind darin etwa 3 Milliarden Basenpaare gespeichert. Jedes Teil der gesamten DNS ist im Zellkern in einem Chromosom untergebracht. Dazu verdrillt und klumpt sich die DNS noch weiter und wird dabei von Histonen (spezielle Proteine) unterstützt, die zum Aufwickeln dienen. Würde man die gesamte DNS eines Menschen hintereinander ausrollen, so wäre sie etwa 1 Meter(!) lang.

## 1.4 Proteine

In diesem Abschnitt wollen wir uns um die wichtigsten Bausteine des Lebens kümmern, die *Proteine*.

### 1.4.1 Aminosäuren

Zunächst einmal werden wir uns mit den *Aminosäuren* beschäftigen, die den Hauptbestandteil der Proteine darstellen. An einem Kohlenstoffatom (dem so genannten *zentralen Kohlenstoffatom* oder auch *$\alpha$-ständigen Kohlenstoffatom* ist ein Wasserstoffatom, eine Carboxylgruppe (also eine Säure) und eine Aminogruppe gebunden, woraus sich auch der Name ableitet. Die letzte freie Bindung des zentralen Kohlenstoffatoms ist mit einem weiteren Rest gebunden. Hierfür kommen prinzipiell alle

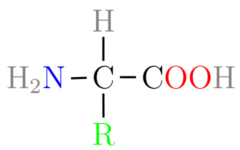



Abbildung 1.22: Aminosäure

möglichen organischen funktionellen Gruppen in Frage. In der Natur der Proteine kommt jedoch nur eine Auswahl von zwanzig verschiedenen Resten in Betracht. Dabei können die Reste so einfach sein wie ein Wasserstoffatom (Glyzin) oder eine Methylgruppe (Alanin), aber auch recht komplex wie zum Beispiel zwei aromatische Ringe (Tryptophan). In Abbildung 1.22 ist die Strukturformel einer generischen Aminosäure dargestellt.

In Abbildung 1.23 sind die Namen der zwanzig in Proteinen auftretenden Aminosäuren und ihre gebräuchlichsten Abkürzungen im so genannten Three-Letter-Code und One-Letter-Code angegeben. Auf die Angabe der genauen chemischen Formeln wollen wir an dieser Stelle verzichten. Hierfür sei auf die einschlägige Literatur verwiesen. In Abbildung 1.24 sind die grundlegendsten Eigenschaften der einzelnen Aminosäuren schematisch zusammengefasst.

Auch hier sind in der Regel (mit Ausnahmen von Glyzin) am zentralen Kohlenstoffatom vier verschiedene Substituenten vorhanden. Somit handelt es sich bei dem

| Aminosäure | 3LC | 1LC |
|---|---|---|
| Alanin | Ala | A |
| Arginin | Arg | R |
| Asparagin | Asn | N |
| Asparaginsäure | Asp | D |
| Cystein | Cys | C |
| Glutamin | Gln | Q |
| Glutaminsäure | Glu | E |
| Glyzin | Gly | G |
| Histidin | His | H |
| Isoleuzin | Ile | I |
| Leuzin | Leu | L |
| Lysin | Lys | K |
| Methionin | Met | M |
| Phenylalanin | Phe | F |
| Prolin | Pro | P |
| Serin | Ser | S |
| Threonin | Thr | T |
| Tryptophan | Trp | W |
| Tyrosin | Tyr | Y |
| Valin | Val | V |
| Selenocystein | Sec | U |
| Aspartamsäure oder Asparagin | Asx | B |
| Glutaminsäure oder Glutamin | Glx | Z |
| Beliebige Aminosäure | Xaa | X |

Abbildung 1.23: Tabelle: Liste der zwanzig Aminosäuren

zentralen Kohlenstoffatom um ein asymmetrisches Kohlenstoffatom und die Aminosäuren können in zwei enantiomorphen Strukturen auftreten. In der Natur tritt jedoch die L-Form auf, die meistens rechtsdrehend ist! Nur diese kann in der Zelle mit den vorhandenen Enzymen verarbeitet werden.

### 1.4.2 Peptidbindungen

Auch Aminosäuren besitzen die Möglichkeit mit sich selbst zu langen Ketten zu polymerisieren. Dies wird möglich durch eine so genannte *säureamidartige Bindung* oder auch *Peptidbindung*. Dabei kondensiert die Aminogruppe einer Aminosäure mit der Carboxylgruppe einer anderen Aminosäure (unter Wasserabspaltung) zu einem

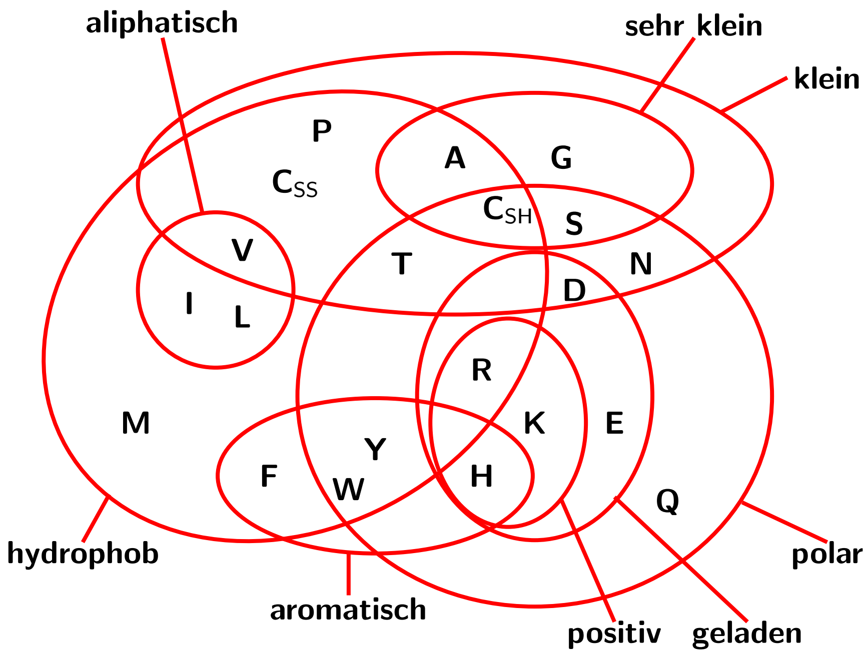



Abbildung 1.24: Skizze: Elementare Eigenschaften von Aminosäuren

neuen Molekül, einem so genannten *Dipeptid*. Die chemische Reaktionsgleichung ist in Abbildung 1.25 illustriert.

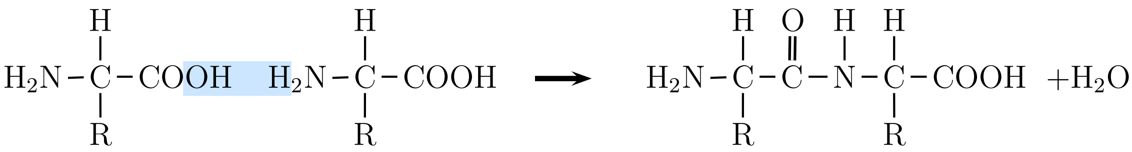


Abbildung 1.25: Skizze: Säureamidartige oder Peptidbindung

Man beachte, dass das Dipeptid an einem Ende weiterhin eine Aminogruppe und am anderen Ende eine Carboxylgruppe besitzt. Dieser Prozess kann also fortgesetzt werden, so dass sich aus Aminosäuren lange unverzweigte Polymere konstruieren lassen. Solche Polymere aus Aminosäuren nennt man *Polypeptide*. Auch hier bemerken wir wieder, dass ein Polypeptid eine Orientierung besitzt. Wir werden Polypeptide, respektive ihre zugehörigen Aminosäuren immer in der Leserichtung von der freien Aminogruppe zur freien Carboxylgruppe hin orientieren. Ein *Protein* selbst besteht dann aus einem oder mehreren miteinander verwundenen Polypeptiden.

Wir wollen uns nun eine solche Peptidbindung etwas genauer anschauen. Betrachten wir hierzu die Abbildung 1.26. Von unten links nach oben rechts durchlaufen wir eine Peptidbindung vom zentralen Kohlenstoffatom der ersten Aminosäure über

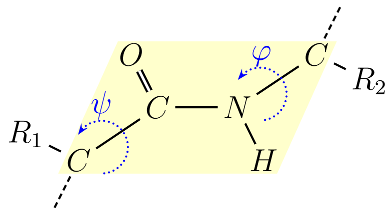



Abbildung 1.26: Skizze: freie Winkel in der Peptidbindung

das Kohlenstoffatom der ehemaligen Carboxylgruppe über das Stickstoffatom der ehemaligen Aminogruppe der zweiten Aminosäure bis hin zum zentralen Kohlenstoffatom der zweiten Aminosäure.

Auf den ersten Blick könnte man meinen, dass Drehungen um alle drei Bindungen C−C, C−N und N−C möglich wären. Eine genaue Betrachtung zeigt jedoch, dass der Winkel um die $C - N$ Bindung nur zwei Werte, nämlich 0° oder 180°, annehmen kann. Dies wird aus der folgenden Abbildung 1.27 deutlicher, die für die Bindungen O=C−N die beteiligten Elektronen-Orbitale darstellt. Man sieht hier, dass nicht nur die C=O-Doppelbindung ein $\pi$-Orbital aufgrund der Doppelbindung ausbildet, sondern dass auch das Stickstoffatom aufgrund seiner fünf freien Außenelektronen zwei davon in einem nichtbindenden $p$-Orbital unterbringt. Aus energetischen Gründen ist es günstiger, wenn sich das $\pi$-Orbital der Doppelbindung und das $p$-Orbital des Stickstoffatoms überlagern.

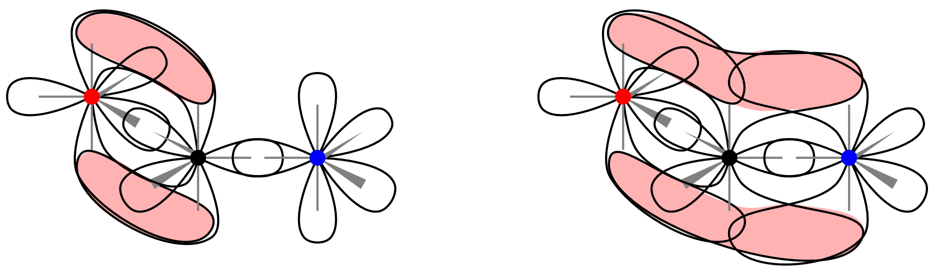


Abbildung 1.27: Skizze: Elektronenwolken in der Peptidbindung

Somit ist die Bindung zwischen dem Kohlenstoff- und dem Stickstoffatom auf 0° oder 180° festgelegt. In der Regel wird die trans-Konformation gegenüber der cis-Konformation bevorzugt, da dann die variablen Reste der Aminosäuren ziemlich weit auseinander liegen. Eine Ausnahme stellt nur Prolin dar, da hier die Seitenkette eine weitere Bindung mit dem Rückgrat eingeht.

Prinzipiell unterliegen die beiden anderen Winkel keinen Einschränkungen. Auch hier haben Untersuchungen gezeigt, dass jedoch nicht alle Winkel eingenommen werden. Ein Plot, der alle Paare von den beiden übrigen Winkeln darstellt, ist der so genannte *Ramachandran-Plot*, der schematisch in der Abbildung 1.28 dargestellt ist. Hier sieht man, dass es gewisse ausgezeichnete Gebiete gibt, die mögliche Winkelkombinationen angeben. Wir kommen auf die Bezeichnungen in diesem Plot später noch einmal zurück.

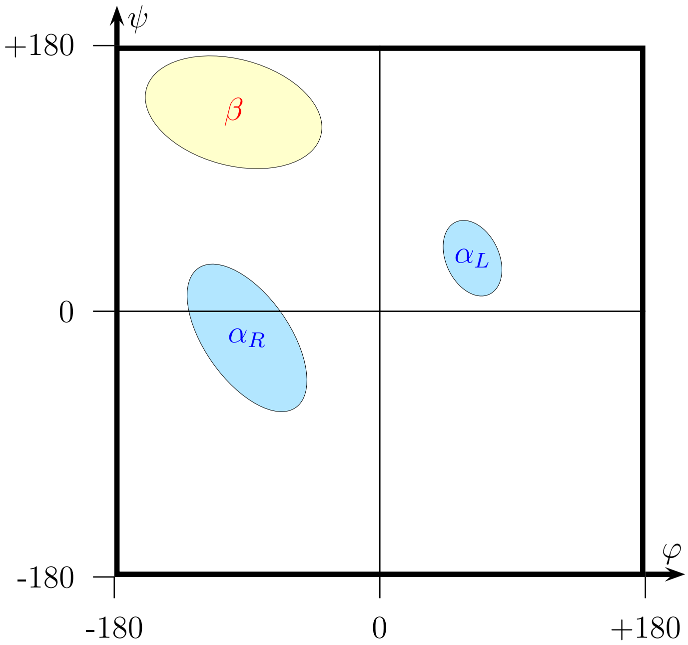



Abbildung 1.28: Skizze: Ramachandran-Plot (schematische Darstellung)

### 1.4.3 Proteinstrukturen

Man betrachtet die *Struktur* der Proteine auf vier verschiedenen Ebenen:

#### 1.4.3.1 Primärstruktur

Die *Primärstruktur* (primary structure) eines Proteins ist die Abfolge der beteiligten Aminosäuren des Polypeptids, also seine *Aminosäuresequenz*. Hierbei hält man die Konvention ein, dass man die Aminosäuren von dem Ende mit der freien Aminogruppe her aufschreibt. Für uns ist dann ein Protein, respektive seine Primärstruktur nichts anderes als eine Zeichenreihe über einem zwanzig-elementigen Alphabet.

#### 1.4.3.2 Sekundärstruktur

Als Sekundärstruktur (secondary structure) bezeichnet man Regelmäßigkeiten in der lokalen Struktur des Proteins, die sich nur über einige wenige Aminosäuren erstrecken. Die prominentesten Vertreter hierfür sind die spiralförmige *$\alpha$-Helix* und der langgestreckte $\beta$-Strang (*$\beta$-strand*). Ursache für die Ausbildung dieser Sekundärstrukturmerkmale ist vor allem die Stabilisierung durch Wasserstoffbrückenbindungen. Zu einem großen Teil wird die Sekundärstruktur eines Proteinabschnitts durch seine eigene Primärstruktur bestimmt, d.h. bestimmte Aminosäuresequenzen bevorzugen (oder vermeiden) $\alpha$-Helices, $\beta$-Strands oder Loops.

**$\alpha$-Helices:** Wie bei der DNS kann ein Protein oder ein kurzes Stück hiervon eine helixartige (spiralförmige) Gestalt ausbilden. Dabei werden die Helices durch Wasserstoffbrücken innerhalb des Polypeptids stabilisiert, die sich zwischen dem Sauerstoffatom der Carbonylgruppe und dem Wasserstoffatom der Aminogruppe der viertnächsten Aminosäure im Peptidstrang ausbilden. Einen solchen Teil eines Peptids nennt man $\alpha$-Helix. Dabei entfallen auf eine volle Drehung etwa 3,6 Aminosäuren. Hierbei hat die Helix in der Regel eine Linksdrehung, weil bei einer Rechtsdrehung die sterische Hinderung deutlich größer ist. Die zugehörigen Winkelpaare der Peptidbindung entsprechen im Ramachandran-Plot in Abbildung 1.28 dem mit $\alpha_L$ markierten Bereich. Einige wenige Helices bilden eine Rechtsdrehung aus. Die zugehörigen Winkelpaare sind im Ramachandran-Plot mit $\alpha_R$ gekennzeichnet.

Beispielsweise sind die Haare aus Proteinen gebildet, die eine Helix bilden. Auch hier können wie bei der DNS mehrere (sogar mehr als zwei) Helices zusammen verdrillt sein. Ebenso seien Proteine erwähnt, die in Muskeln eine wichtige Rolle spielen.

**$\pi$- und $3_{10}$-Helices:** In seltenen Fällen treten Abwandlungen der $\alpha$-Helix auf, bei denen die Wasserstoffbrücken nicht zwischen den Aminosäuren $n$ und $n+4$, sondern zwischen den Aminosäuren $n$ und $n+3$ oder $n+5$ gebildet werden. In diesen Fällen ist die Helix also etwas mehr oder etwas weniger verdreht. Man nennt diese beiden Formen die $3_{10}$- und die $\pi$-Helix. Der Name $3_{10}$-Helix entstammt dabei der Tatsache, dass die Helix 3 Aminosäuren pro Umdrehung enthält und dass zwischen den beiden Enden einer Wasserstoffbrücke zehn Atome (incl. Wasserstoffatom) liegen. In dieser Nomenklatur (nach Linus Pauling und Robert Corey) würde man die $\alpha$-Helix mit $3.6_{13}$ und die $\pi$-Helix mit $4.4_{16}$ bezeichnen.

**$\beta$-Strands:** Eine andere Struktur sind langgezogene Bereiche von Aminosäuren. Meist lagern sich hier mehrere Stränge (oft von verschiedenen Polypeptidketten, aber durchaus auch nur von einer einzigen) nebeneinander an und bilden so genannte $\beta$-Sheets oder $\beta$-Faltblätter aus. Hierbei ist zu beachten, dass sich diese wie Spaghetti nebeneinander lagern. Dies kann entweder parallel oder antiparallel geschehen (Polypeptide haben ja eine Richtung!). Auch hier werden solche Faltblätter durch Wasserstoffbrücken zwischen den Amino- und Carbonylgruppen stabilisiert. Diese Struktur heißt Faltblatt, da es entlang eines Polypeptid-Strangs immer wieder auf und ab geht, ohne insgesamt die Richtung zu ändern. Bilden sich ganze $\beta$-Faltblätter aus, so sehen diese wie ein gefaltetes Blatt aus, wobei die einzelnen Polypeptide quer zu Faltungsrichtung verlaufen. Die zugehörigen Winkelpaare der Peptidbindung entsprechen im Ramachandran-Plot in Abbildung 1.28 dem mit $\beta$ markierten Bereich. Beispielsweise tauchen im Seidenfibroin (Baustoff für Seide) fast nur $\beta$-Faltblätter auf.

**Reverse Turns:** Zum Schluss seien noch kurze Sequenzen (von etwa fünf Aminosäuren) erwähnt, die einfach nur die Richtung des Polypeptids umkehren. Diese sind beispielsweise in antiparallelen $\beta$-Faltblätter zu finden, um die einzelne $\beta$-Strands zu einem $\beta$-Sheet anordnen zu können.

#### 1.4.3.3 Supersekundärstruktur

Oft lagern sich zwei oder drei Sekundärstrukturelemente zu sogenannten *Motifs* zusammen.

**Hairpins** Reverse Turns, die zwischen zwei nebeneinander liegenden antiparallelen $\beta$-Strands liegen und deshalb die Form einer Haarnadel nachbilden

**Coiled coils** bestehen aus zwei verdrillten $\alpha$-Helices und spielen eine wichtige Rolle in Faser-Proteinen

#### 1.4.3.4 Tertiärstruktur

Die *Tertiärstruktur* (tertiary structure) ist die *Konformation*, also die räumliche Gestalt, eines einzelnen Polypeptids. Sie beschreibt, wie die Elemente der Sekundär- und Supersekundärstruktur sich zu so genannten *Domains* zusammensetzen. Hier ist für jedes Atom (oder Aminosäure) die genaue relative Lage zu allen anderen bekannt. Tertiärstrukturen sind insbesondere deshalb wichtig, da für globuläre Proteine die räumliche Struktur für ihre Wirkung wichtig ist (zumindest in fest umschriebenen Reaktionszentren eines Proteins).

#### 1.4.3.5 Quartärstruktur

Besteht ein Protein aus mehreren Polypeptidketten, wie etwa Hämoglobin, dann spricht man von der *Quartärstruktur* (quaternary structure) eines Proteins. Proteine können aus einem einzelnen Polypeptid oder aus mehreren (gleichen oder unterschiedlichen) Polypeptidketten bestehen.

## 1.5 Der genetische Informationsfluss

Bislang haben wir die wichtigsten molekularbiologischen Bausteine kennen gelernt. Die DNS, die die eigentliche Erbinformation speichert, und sich in der Regel zu den bekannten Chromosomen zusammenwickelt. Über die Bedeutung der zur DNS strukturell sehr ähnlichen RNS werden wir später noch kommen. Weiterhin haben wir mit den Proteinen die wesentlichen Bausteine des Lebens kennen gelernt. Jetzt wollen wir aufzeigen, wie die in der DNS gespeicherte genetische Information in den Bau von Proteinen umgesetzt werden kann, d.h. wie die Erbinformation weitergegeben (vererbt) wird.

### 1.5.1 Replikation

Wie wird die genetische Information überhaupt konserviert, oder anders gefragt, wie kann man die Erbinformation kopieren? Dies geschieht durch eine Verdopplung der DNS. An einer bestimmten Stelle wird die DNS entspiralisiert und die beiden Stränge voneinander getrennt, was dem Öffnen eines Reißverschlusses gleicht. Diese Stelle wird auch *Replikationsgabel* genannt. Dann wird mit Hilfe der DNS-Polymerase an beiden Strängen die jeweils komplementäre Base oder genauer das zugehörige Nukleotid mit Hilfe der Wasserstoffbrücken angelagert und die Phosphatsäuren bilden mit den nachfolgenden Zuckern jeweils eine Esterbindung zur Ausbildung des eigentlichen Rückgrates aus. Dies ist schematisch in Abbildung 1.29 dargestellt, wobei die neu konstruierten DNS-Stücke rot dargestellt sind.

Die Polymerase , die neue DNS-Stränge generiert, hat dabei nur ein Problem. Sie kann einen DNS-Strang immer nur in der Richtung vom 5'-Ende zum 3'-Ende synthetisieren. Nach dem Öffnen liegt nun ein Strand in Richtung der Replikationsgabel (in der ja die Doppelhelix entspiralisiert wird und die DNS aufgetrennt wird) in Richtung vom 3'-Ende zum 5'-Ende vor der andere jedoch in Richtung vom 5'-Ende zum 3'-Ende, da die Richtung der DNS-Stränge in der Doppelhelix ja antiparallel ist.

Der Strang in Richtung vom 3'-Ende zum 5'-Ende in Richtung auf die Replikationsgabel zu lässt sich jetzt leicht mit der DNS-Polymerase ergänzen, da dann die

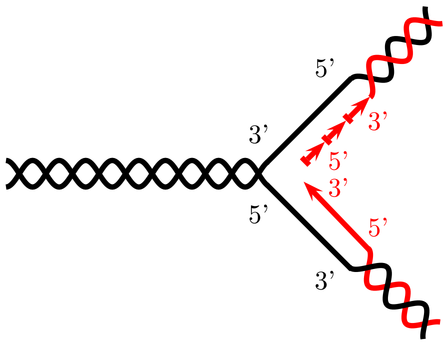



Abbildung 1.29: Skizze: DNS-Replikation

Synthese selbst in Richtung vom 5'-Ende zum 3'-Ende erfolgt und beim weiteren Öffnen der DNS einfach weiter synthetisiert werden kann.

Für den Strang in Richtung vom 5'-Ende zum 3'-Ende hat man herausgefunden, dass auch hier die Synthese in Richtung vom 5'-Ende zum 3'-Ende erfolgt. Die DNS-Polymerase wartet hier solange, bis ein hinreichend langes Stück frei liegt und ergänzt den DNS Strang dann von der Replikationsgabel weg. Das bedeutet, dass die die DNS hier immer in kleinen Stücken synthetisiert wird und nicht im ganzen wie am komplementären Strang. Die dabei generierten kleine DNS-Teilstränge werden nach ihrem Entdecker *Okazaki-Fragmente* genannt.

### 1.5.2 Transkription

Damit die in der DNS gespeicherte Erbinformationen genutzt werden kann, muss diese erst einmal abgeschrieben oder kopiert werden. Der Prozess ist dabei im Wesentlichen derselbe wie bei der Replikation. Hierbei wird allerdings nicht die gesamte DNS abgeschrieben, sondern nur ein Teil. Der hierbei abgeschriebene Teil bildet dann nicht eine Doppelhelix mit der DNS aus, sondern löst sich am nichtaktiven Ende wieder, so dass die beiden aufgetrennten Stränge der DNS wieder eine Doppelhelix bilden können.

Hierbei ist zu beachten, dass der abgeschriebene Teil keine DNS, sondern eine RNS ist. Hier wird also Ribose statt Desoxyribose verwendet und als Base Uracil statt Thymin. Die abgeschriebene RNS wird als *Boten-RNS* bzw. *messenger RNA* bezeichnet, da sie als Überbringer der Erbinformation dient.

In prokaryontischen Zellen ist der Vorgang damit abgeschlossen. In eukaryontischen Zellen ist der Vorgang etwas komplizierter, da die Chromosomen im Zellkern beheimatet sind und damit auch die Boten-RNS. Da die Boten-RNS jedoch außerhalb des Zellkerns weiterverarbeitet wird, muss diese erst noch durch die Membran des Zellkerns wandern.

Des Weiteren hat sich in eukaryontischen Zellen noch eine Besonderheit ausgebildet. Die Erbinformation steht nicht kontinuierlich auf der DNS, sondern beinhaltet dazwischen Teilstücke ohne Erbinformation. Diese müssen vor einer Weiterverarbeitung erst noch entfernt werden.

Es hat sich gezeigt, dass dieses Entfernen, *Spleißen* (engl. *Splicing*) genannt, noch im Zellkern geschieht. Dabei werden die Stücke, die keine Erbinformation tragen und *Introns* genannt werden, aus der Boten-RNS herausgeschnitten. Die anderen Teile, *Exons* genannt, werden dabei in der selben Reihenfolge wie auf der DNS aneinander gereiht. Die nach dem Spleißen entstandene Boten-RNS wird dann als *reife Boten-RNS* oder als *mature messenger RNA* bezeichnet.

Für Experimente wird oft aus der mRNA wieder eine Kopie als DNS dargestellt. Diese wird als *cDNS* bzw. *komplementäre DNS* (engl. *cDNA* bzw. *complementary DNA*) bezeichnet. Diese entspricht dann dem Original aus der DNS, wobei die Introns bereits herausgeschnitten sind. Die originalen Gene aus der DNS mit den Introns wird auch als *genetische DNS* (engl. *genetic DNA*) bezeichnet.

Dazu betrachten wir die schematische Darstellung des genetischen Informationsflusses innerhalb einer Zelle (hier einer eukaryontischen) in Abbildung 1.30.

### 1.5.3 Translation

Während der *Translation* (*Proteinbiosynthese*) wird die in der DNS gespeicherte und in der reifen Boten-RNS zwischengespeicherte komplementäre Erbinformation in Proteine übersetzt. Dies geschieht innerhalb der Ribosomen, die sich wie zwei Semmelhälften auf den Anfang der Boten-RNS setzen und den RNS-Strang in ein Protein übersetzen. Ribosomen selbst sind aus Proteinen und RNS, so genannter *ribosomaler RNS* oder kurz *rRNS* (engl. *ribosomal RNA*, *rRNA*), zusammengesetzt.

Wir erinnern uns, dass die RNS bzw. DNS im Wesentlichen durch die vier Basen Adenin, Guanin, Cytosin und Uracil bzw. Thymin die Information trägt. Nun ist also die in der RNS gespeicherte Information über einem vierelementigen Alphabet codiert, wobei ein Protein ein Polymer ist, das aus zwanzig verschiedenen Aminosäuren gebildet wird. Wir müssen also noch die Codierung eines zwanzig-elementigen durch ein vier-elementiges Alphabet finden.

Offensichtlich lassen sich nicht alle Aminosäuren durch je zwei Basen codieren. Es muss also Aminosäuren geben, die durch mindestens drei Basen codiert werden. Es hat sich herausgestellt, dass der Code immer dieselbe Länge hat und somit jeweils drei Basen, ein so genanntes *Basen-Triplett* oder *Codon*, jeweils eine Aminosäure codiert.

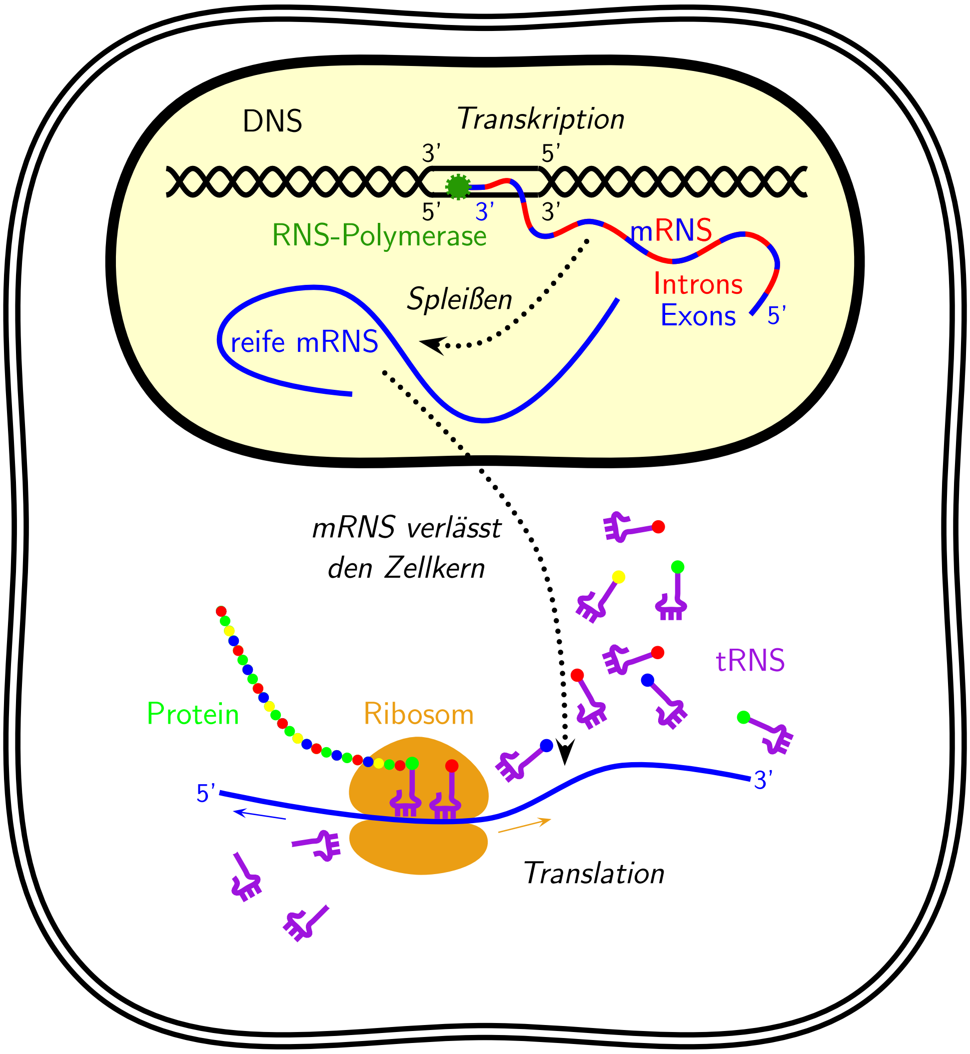



Abbildung 1.30: Skizze: Transkription und Transaltion in einer Zelle

Da es 64 verschiedene Tripletts aber nur zwanzig Aminosäuren gibt, werden einige Aminosäuren durch mehrere Tripletts codiert. In der folgenden Abbildung 1.31 ist der Code für die Umwandlung von Tripletts in Aminosäuren angegeben. In Abbildung 1.31 steht das erste Zeichen des Tripletts links, das zweite oben und das dritte in der rechten Spalte. AGU bzw. AGC codiert also Serin.

Zum genetischen Code ist noch folgendes zu sagen. Es gibt auch spezielle Stopp-Codons, die den Ribosomen mitteilen, dass die Übersetzung der RNS in ein Protein zu Ende ist: Diese sind UAG, UAA und UGA. Ebenfalls gibt es auch ein so genanntes Start-Codon, das jedoch nicht eindeutig ist. AUG Übernimmt sowohl die Rolle der Codierung von Methionin als auch dem Anzeigen an das Ribosom, dass hier mit

| | U | C | A | G | |
|---|---|---|---|---|---|
| U | Phe | Ser | Tyr | Cys | U |
| | Phe | Ser | Tyr | Cys | C |
| | Leu | Ser | STOP | STOP | A |
| | Leu | Ser | STOP | Trp | G |
| C | Leu | Pro | His | Arg | U |
| | Leu | Pro | His | Arg | C |
| | Leu | Pro | Gln | Arg | A |
| | Leu | Pro | Gln | Arg | G |
| A | Ile | Thr | Asn | Ser | U |
| | Ile | Thr | Asn | Ser | C |
| | Ile | Thr | Lys | Arg | A |
| | Met | Thr | Lys | Arg | G |
| G | Val | Ala | Asp | Gly | U |
| | Val | Ala | Asp | Gly | C |
| | Val | Ala | Glu | Gly | A |
| | Val | Ala | Glu | Gly | G |

Abbildung 1.31: Tabelle: Der genetische Code

Übersetzung begonnen werden kann. Ebenso ist dieser Code universell, d.h. fast alle Lebewesen benutzen diesen Code. Bislang sind nur wenige Ausnahmen bekannt, die einen anderen Code verwenden, der aber diesem weitestgehend ähnlich ist.

Auch sollte erwähnt werden, dass die Redundanz des Codes (64 Tripletts für 20 Aminosäuren) zur Fehlerkorrektur ausgenutzt wird. Beispielsweise ist die dritte Base für die Decodierung einiger Aminosäuren völlig irrelevant, wie für Alanin, Glyzin, Valin und andere. Bei anderen Mutationen werden in der Regel Aminosäuren durch weitestgehend ähnliche (in Bezug auf Größe, Hydrophilie, Ladung oder ähnliches) ersetzt. Auch werden häufig auftretende Aminosäuren durch mehrere Tripletts, selten auftretende nur durch eines codiert.

Ebenfalls sollte man hierbei noch darauf hinweisen, dass es für jeden RNS-Strang eigentlich drei verschiedene Leseraster gibt. Dem RNS-Strang $s_1 \cdots s_n$ an sich sieht man nicht, ob das codierte Gen in $(s_1s_2s_3)(s_4s_5s_6)\cdots$, $(s_2s_3s_4)(s_5s_6s_7)\cdots$ oder $(s_3s_4s_5)(s_6s_7s_8)\cdots$ codiert ist. Das Start-Codon kann dabei jedoch Hilfe leisten.

Zum Schluss bleibt nur noch die Übersetzung im Ribosom zu beschreiben. Dabei hilft die so genannte *Transfer-RNS* oder *tRNS*. Die Transfer-RNS besteht im Wesentlichen aus RNS mit drei Basen und einer Bindungsstelle für eine Aminosäure. Dabei entsprechen die drei komplementären Basen der zugehörigen Aminosäure, die an der Bindungsstelle angebunden ist.

Im Ribosom werden dann jeweils die komplementären tRNS zum betrachteten Triplett der mRNS mittels der Wasserstoffbrücken angebunden. Dabei wird anschließend die neue Aminosäure mittels der säureamidartigen Bindung an den bereits synthetisierten Polypeptid-Strang angebunden. Die tRNS, die den bereits synthetisierten Polypeptid-Strang festhielt, wird dann freigegeben, um in der Zelle wieder mit einer neuen, zum zugehörigen Triplett gehörigen Aminosäure aufzuladen.

### 1.5.4 Das zentrale Dogma

Aus der bisherigen Beschreibung lässt sich die folgende Skizze für den genetischen Informationsfluss ableiten. Die genetische Information wird in der DNS gespeichert und mit Hilfe der Replikation vervielfacht. Mit Hilfe der Transkription wird die genetische Information aus der DNS ausgelesen und in die RNS umgeschrieben. Aus der RNS kann dann mit Hilfe der Translation die genetische Information in die eigentlichen Bausteine des Lebens, die Proteine, übersetzt werden. Damit ist der genetische Informationsfluss eindeutig von der DNS über die RNS zu den Proteinen gekennzeichnet. Dies wird als das zentrale Dogma der Molekularbiologie bezeichnet.

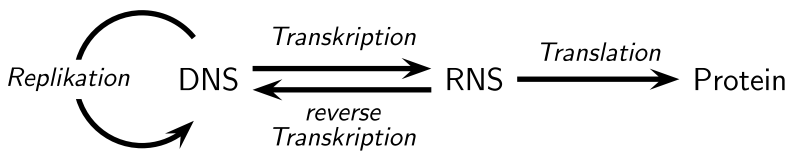



Abbildung 1.32: Skizze: Das zentrale Dogma der Molekularbiologie

In der Biologie gibt es kaum eine Regel ohne Ausnahmen. Trotz des zentralen Dogmas ist auch ein Informationsfluss in die umgekehrte Richtung möglich. Auch aus der RNS kann modifizierte genetische Erbinformation wieder zurück in die DNS eingebaut werden. Das zentrale Dogma ist in Abbildung 1.32 noch einmal schematisch dargestellt.

### 1.5.5 Promotoren

Für den letzten Abschnitt müssen wir uns nur noch überlegen, wie man die eigentliche Erbinformation, die Gene, auf der DNS überhaupt findet. Dazu dienen so genannte *Promotoren*. Dies sind mehrere kurze Sequenzen vor dem Beginn des eigentlichen Gens, die RNS-Polymerase überhaupt dazu veranlassen unter bestimmten Bedingungen die spiralisierte DNS an dieser Stelle aufzuwickeln und die beiden

DNS-Stränge zu trennen, so dass das Gen selbst transkribiert werden kann. In Bakterien ist dies sehr einfach, da es dort im Wesentlichen nur einen Promotor gibt, der in Abbildung 1.33 schematisch dargestellt ist. In höheren Lebewesen und insbesondere in eukaryontischen Zellen sind solche Promotoren weitaus komplexer und es gibt eine ganze Reihe hiervon.

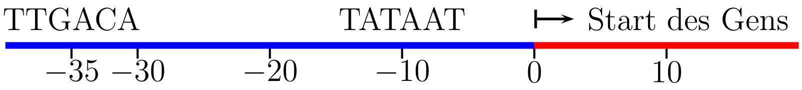



Abbildung 1.33: Skizze: Promotoren in Bakterien

## 1.6 Biotechnologie

In diesem Abschnitt wollen wir einige der wichtigsten biotechnologischen Methoden vorstellen, die für uns im Folgenden für eine informatische Modellbildung wichtig sein werden.

### 1.6.1 Hybridisierung

Mit *Hybridisierung* wird die Aneinanderlagerung zweier DNS-Stränge bezeichnet, die aufgrund der Komplementarität ihrer Basen über Wasserstoffbrücken gebildet wird. Dies haben wir prinzipiell schon bei der Replikation und Transkription kennen gelernt.

Eine Hybridisierung kann dazu ausgenutzt werden, um festzustellen, ob sich eine bestimmte, in der Regel recht kurze Teilsequenz innerhalb eines DNS-Strangs befindet. Dazu wird eine kurze Teilsequenz synthetisiert und an einem Ende markiert, z.B. mit einem fluoreszierenden oder radioaktiven Stoff. Werden die kurzen Teilsequenzen mit den durch Klonierung verfielfachten DNS-Strängen zusammengebracht, so können die kurzen Sequenzen mit dem DNS-Strang hybridisieren. Nach Entfernung der kurzen synthetisierten Stücke kann dann festgestellt werden, ob die langen DNS-Stränge fluoreszent oder radioaktiv sind. Letzteres ist genau dann der Fall, wenn eine Hybridisierung stattgefunden hat.

### 1.6.2 Klonierung

Für biologische Experimente wird oft eine Vielzahl von identischen Kopien eines DNS-Stückes benötigt. Solche Vervielfältigungen lassen sich mit Hilfe niederer Organismen erledigen. Dazu wird das zu vervielfältigende DNS-Stück in die DNS des Organismus eingesetzt und dieser vervielfältigt diese wie seine eigene DNS. Je nachdem, ob Plasmide, Bakterien oder Hefe (engl. yeast) verwendet werden, spricht man vom *plasmid* (*PAC*), *bacterial* (*BAC*) oder *yeast artificial chromosomes* (*YAC*).

Die bei PACs verwendeten Plasmide sind ringförmige DNS-Stränge, die in Bakterien auftreten. Bei jeder Zellteilung wird dabei auch der zu klonierende, neu eingesetzt DNS-Strang vervielfältigt. Hierbei können jedoch nur Stränge bis zu 15.000 Basenpaaren kloniert werden.

Bei BACs werden Phagen (ein Virus) verwendet. Die infizierten Wirtszellen (Bakterien) haben dann das zu klonierende DNS-Stück, das in die Phage eingesetzt wurde, vervielfältigt. Hier sind Vervielfältigungen von bis zu 25.000 Basenpaaren möglich.

Bei YACs wird die gewöhnliche Brauerhefe zur Vervielfältigung ausgenutzt, in die die gewünschten DNS-Teilstücke eingebracht werden. Hierbei sind Vervielfältigungen bis zu 1 Million Basenpaaren möglich.

### 1.6.3 Polymerasekettenreaktion

Eine andere Art der Vervielfältigung ist mit Hilfe der *Polymerasekettenreaktion* (engl. *polymerase chain reaction*) möglich. Diese hatten wir ja schon bei Replikation von DNS-Doppelhelices kennen gelernt. Wir müssen von dem zu vervielfältigenden Bereich nur die Sequenzen der beiden Endstücke von etwa 10 Basenpaaren kennen. Diese kurzen Stücke werden *Primer* genannt

Zuerst werden die DNS-Stränge der Doppelhelix durch Erhitzen aufgespalten. Dann werden sehr viele komplementäre Sequenzen der Primer zugegeben, so dass sich an die Primer der DNS hybridisieren können. Mit Hilfe der Polymerase werden dann ab den Primern in die bekannte Richtung vom 5'-Ende zum 3'-Ende die Einzelstränge der aufgesplitteten DNS zu einem Doppelstrang vervollständigt. Dies ist in Abbildung 1.34 schematisch dargestellt. Dabei ist das zu vervielfältigende DNS-Stück grün, die Primer rot und Rest der DNS grau dargestellt. Die Pfeile geben die Synthetisierungssrichtung der Polymerase an.

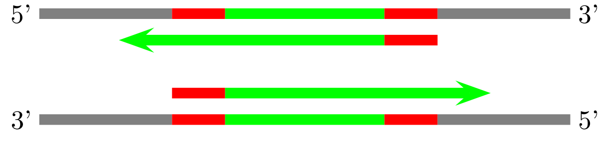



Abbildung 1.34: Skizze: Polymerasekettenreaktion

Einziges Problem ist, dass bei der ersten Anwendung die Polymerase immer bis zum Ende des Strangs läuft. Es wird also mehr dupliziert als gewünscht. Nun kann man dieses Experiment mehrfach (50 Mal) wiederholen. In jedem Schritt werden dabei die vorher synthetisierten Doppelstränge verdoppelt. Nach $n$ Phasen besitzt man also $2^n$ Doppelstränge!

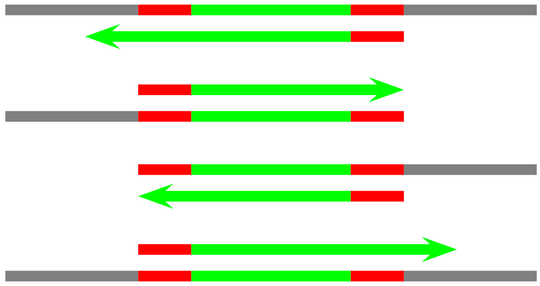


Abbildung 1.35: Skizze: PCR: nach der 2. Verdopplung

Da nach der ersten Verdopplung, bereits jeweils ein unnützes Ende nicht kopiert wurde, überlegt man sich leicht, dass nach der zweiten Verdopplung bereits zwei (von vier) Doppelsträngen nur den gewünschten Bereich verdoppelt haben. Zum Schluss ist jeder zweite DNS-Doppelstrang eine Kopie des gewünschten Bereichs und alle anderen sind nur im gewünschten Bereich doppelsträngig (ansonsten einsträngig, mit zwei Ausnahmen).

Mit dieser Technik lassen sich also sehr schnell und recht einfach eine Vielzahl von Kopien eines gewünschten, durch zwei kurze Primer umschlossenen DNS-Teilstücks herstellen.

### 1.6.4 Restriktionsenzyme

Restriktionsenzyme sind spezielle *Enzyme* (Proteine, die als Katalysator wirken), die eine DNS-Doppelhelix an bestimmten Stellen aufschneiden können. Durch solche Restriktionsenzyme kann also eine lange DNS geordnet in viele kurze Stücke zerlegt werden. Eines der ersten gefundenen Restriktionsenzyme ist EcoRI, das im Bakterium Escherichia Coli auftaucht. Dieses erkennt das Muster GAATTC. In Abbil-

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5’ | G | A | A | T | T | C | 3’ |
| 3’ | C | T | T | A | A | G | 5’ |

Abbildung 1.36: Skizze: Restriktionsenzym mit Muster *GAATTC*

dung 1.36 ist dieses Muster in der Doppelhelix der DNS noch einmal schematisch mit den Bruchstellen dargestellt. Man beachte hierbei, dass die Sequenz zu sich selbst komplementär ist. Es handelt sich also um ein so genanntes *komplementäres Palindrom*.

### 1.6.5 Sequenzierung kurzer DNS-Stücke

In diesem Abschnitt wollen wir kurz die gebräuchlichsten Techniken zur Sequenzierung von DNS darstellen. Mit *Sequenzierung* ist das Herausfinden der Abfolge der vier verschiedenen Basen in einem gegebenen DNS-Strang gemeint.

Die Grundidee ist die Folgende: Zuerst wird für das zu sequenzierende Teilstück mit Hilfe der Klonierung eine ausreichende Anzahl identischer Kopien erzeugt. Dann werden die klonierten Teilstücke in vier Gruppen (quasi für jede Base eine) eingeteilt. In jeder Gruppe werden an jeweils einer Base die Teilstücke aufgebrochen. Zu Beginn hat man eines der Enden mit Hilfe eines fluoreszierenden oder radioaktiven Stoffes markiert. Im Folgenden interessieren nur die Bruchstücke, die das markierte Ende besitzen, wobei die anderen Bruchstücke jedoch nicht entfernt werden. Mit Hilfe der so genannten *Elektrophorese* werden die verschieden langen Bruchstücke getrennt.

Die Elektrophorese nutzt aus, dass unter bestimmten Randbedingungen die DNS-Bruchstücke nicht elektrisch neutral, sondern elektrisch geladen sind. Somit kann man die DNS-Bruchstücke mit Hilfe eines elektrischen Feldes wandern lassen. Dazu werden diese innerhalb eines Gels gehalten, dessen Zähflüssigkeit es erlaubt, dass sie sich überhaupt, aber auch nicht zu schnell bewegen. Da die Bruchstücke alle dieselbe Ladung tragen, aber aufgrund ihrer Länge unterschiedlich schwer sind, haben sie innerhalb des angelegten elektrischen Feldes eine unterschiedliche Wanderungsgeschwindigkeit. Die kurzen wandern naturgemäß sehr schnell, während die langen Bruchstücke sich kaum bewegen.

Führt man dieses Experiment gleichzeitig für alle vier Gruppen (also für jede Base) getrennt aus, so erhält man ein Bild der gewanderten Bruchstücke. Dies ist schematisch in Abbildung 1.37 dargestellt, das man sich als eine Fotografie einer Gruppe radioaktiv markierter Stücke vorstellen kann (auch wenn dies heute mit fluoreszierenden Stoffen durchgeführt wird). Hier ist links die (noch nicht bekannte) Sequenz der einzelnen Bruchstücke angegeben. Mit rot ist speziell das Ergebnis für die Base Adenin hervorgehoben. In den experimentellen Ergebnissen gibt es an sich jedoch keine farblichen Unterscheidungen.

Man kann nun die relativen Positionen einfach feststellen und anhand der Belichtung des Films feststellen in welcher Gruppe sich ein Bruchstück befindet und somit die Base an der entsprechenden Position ablesen. Es ist hierbei zu berücksichtigen, dass die Wanderungsgeschwindigkeit umgekehrt proportional zu dessen Masse (und somit im Wesentlichen zur Länge des betrachteten Bruchstücks ist). Während also der Abstand der Linie der Bruchstücke der Länge eins und der Linie der Bruchstücke

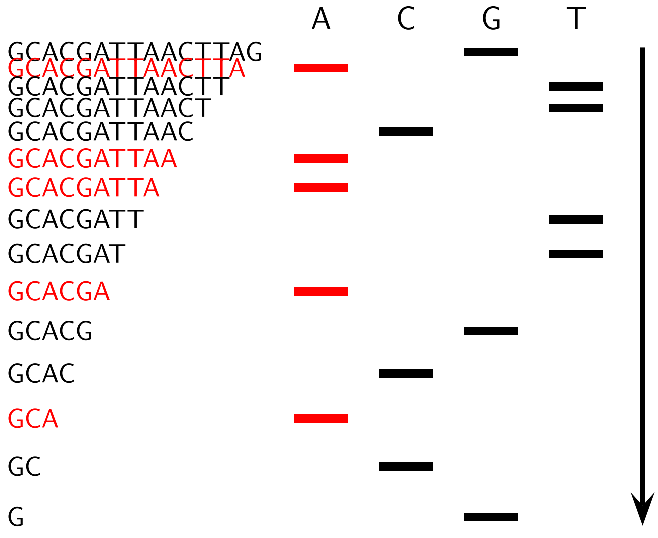



Abbildung 1.37: Skizze: Sequenzierung nach Sanger

der Länge zwei relativ groß sind, ist der Abstand der Linie der Bruchstücke der Länge 100 und der Linie der Bruchstücke der Länge 101 relativ kurz.

#### 1.6.5.1 Sanger-Methode

Das Aufspalten kann über zwei prinzipiell verschiedene Methoden erfolgen. Zum einen kann man bei der Vervielfältigung innerhalb einer Klasse dafür sorgen, dass neben der entsprechenden Base auch eine modifizierte zur Verfügung steht, an der die Polymerase gestoppt wird. Der Nachteil hierbei ist, dass längere Sequenzen immer weniger werden und man daher lange Sequenzen nicht mehr so genau erkennen kann. Diese Methode wird nach ihrem Erfinder auch *Sanger-Methode* genannt.

#### 1.6.5.2 Maxam-Gilbert-Methode

Zum anderen kann man mit Hilfe chemischer Stoffe die Sequenzen entweder nach Adenin, einer Purinbase (Adenin und Guanin), Thymin oder einer Pyrimidin-Base (Thymin und Cytosin) aufbrechen. Man erhält also nicht für jede Base einen charakteristischen Streifen, sondern bei Adenin oder Thymin jeweils zwei, wie in der folgenden Abbildung illustriert. Nichts desto trotz lässt sich daraus die Sequenz ablesen (siehe auch Abbildung 1.38). Diese Methode wird nach ihren Erfinder auch *Maxam-Gilbert-Methode* genannt.

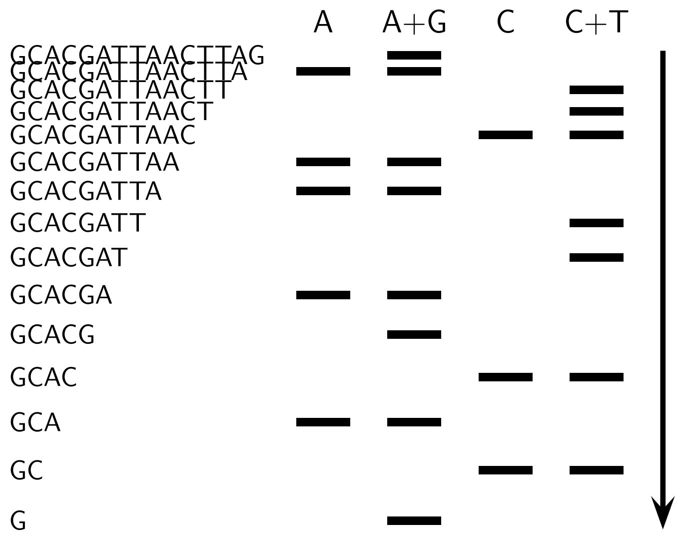



Abbildung 1.38: Skizze: Sequenzierung nach Maxam-Gilbert

Heutzutage wird in der Regel die Sanger-Methode mit fluoreszierenden Stoffen angewendet, die sich dann gleichzeitig in großen Sequenzierautomaten sehr leicht automatisch anwenden lässt. Dennoch lassen sich auch heute nur DNS-Sequenzen der Länge 500 gut sequenzieren. Mit den bekannten Methoden sind auch in Zukunft bei einem entsprechenden technologischen Fortschritt Sequenzierungen von mehr als 1000 Basenpaaren nicht denkbar.

## 1.6.6 Sequenzierung eines Genoms

Im letzten Abschnitt haben wir gesehen, wie sich kurze DNS-Stücke sequenzieren lassen und dass sich diese Methoden nicht auf beliebig lange DNS-Sequenzen ausdehnen lässt. In diesem letzten Abschnitt wollen wir uns überlegen, wie man ein ganzes Genom sequenzieren kann.

### 1.6.6.1 Primer Walking

Beim *Primer Walking* ist die Idee, dass man die ersten 500 Basen des Genoms sequenziert. Kennt man diese, so kann man am Ende einen möglichst eindeutigen Primer der Länge ca. zwanzig ablesen und mit der Polymerasekettenreaktion, die Sequenz ab dieser Stelle vervielfältigen. Der folgende Sequenzierungsschritt beginnt daher am Ende (mit einer kleinen Überlappung) des bereits sequenzierten Anfangsstücks.

Dieses Verfahren lässt sich natürlich beliebig wiederholen. Somit läuft man also mit Primern über die Sequenz und sequenziert die DNS Stück für Stück. In der Anwesenheit von Repeats (Wiederholungen) versagt diese Methode, da man innerhalb eines langen Repeats per Definition keinen eindeutigen Primer mehr finden kann.

#### 1.6.6.2 Nested Sequencing

Auch beim *Nested Sequencing* läuft man Stück für Stück über die Sequenz. Immer wenn man eine Sequenz der Länge 500 sequenziert hat, kann man diese mit Hilfe eines Enzyms (Exonuclease) entfernen und mit der restlichen Sequenz weitermachen.

Beide vorgestellten Verfahren, die die Sequenz Stück für Stück sequenzieren, haben den Nachteil, dass sie inhärent sequentiell und somit bei großen Genomen sehr langsam sind.

#### 1.6.6.3 Sequencing by Hybridization

Beim *Sequenzieren durch Hybridisierung* oder kurz *SBH* (engl. sequencing by hybridization) werden die Sequenzen zuerst vervielfältigt und dann durch Restriktionsenzyme kleingeschnitten. Diese klein geschnittenen Sequenzen werden durch Hybridisierung mit allen Sequenzen der Längen bis etwa 8 verglichen. Somit ist bekannt, welche kurzen Sequenzen in der zu sequenzierenden enthalten sind. Aus diesen kurzen Stücken lässt sich dann mit Informatik-Methoden die Gesamtsequenz wiederherstellen.

Dieses Verfahren ist jedoch sehr aufwendig und hat sich bislang nicht durchgesetzt. Jedoch hat es eine bemerkenswerte Technik, die so genannten *DNA-Microarrays* oder *Gene-Chips* hervorgebracht, die jetzt in ganz anderen Gebieten der Molekularbiologie eingesetzt werden.

Ein DNA-Microarray ist eine kleine Glasplatte (etwa $1\text{cm}^2$), die in etwa 100 mal 100 Zellen aufgeteilt ist. In jeder Zelle wird ein kleines Oligonukleotid der Länge von bis zu 50 Basen aufgebracht. Mit Hilfe der Hybridisierung können nun parallel 10.000 verschiedene Hybridisierungsexperimente gleichzeitig durchgeführt werden. Die zu untersuchenden Sequenzen sind dabei wieder fluoreszent markiert und können nach dem hochparallelen Hybridisierungsexperiment entsprechend ausgewertet werden.

#### 1.6.6.4 Shotgun Sequencing

Eine weitere Möglichkeit, ein ganzes Genome zu sequenzieren, ist das so genannte *Shotgun-Sequencing*. Hierbei werden lange Sequenzen in viele kurze Stücke aufgebrochen. Dabei werden die Sequenzen in mehrere Klassen aufgeteilt, so dass (in der Regel) eine Bruchstelle in einer Klasse mitten in den Fragmenten der anderen Sequenzen liegt.

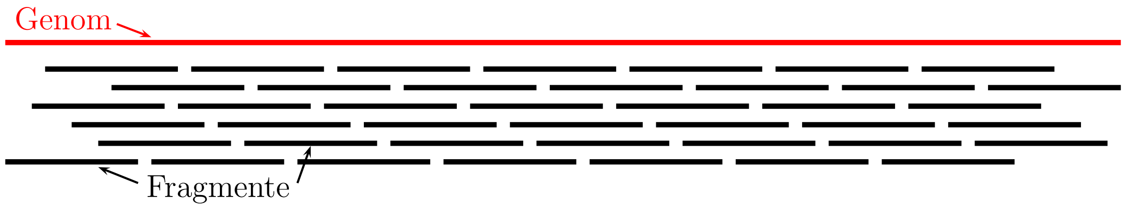



Abbildung 1.39: Skizze: Shotgun-Sequencing

Die kurzen Sequenzen können jetzt wieder direkt automatisch sequenziert werden. Es bleibt nur das Problem, aus der Kenntnis der Sequenzen wieder die lange DNS-Sequenz zu rekonstruieren. Dabei hilft, dass einzelne Positionen (oder sogar kurze DNS-Stücke) von mehreren verschiedenen Fragmenten, die an unterschiedlichen Positionen beginnen, überdeckt werden. In der Regel sind diese Überdeckungen relativ lang. Somit muss man nur noch die Fragmente wie in einem Puzzle-Spiel so anordnen, dass überlappende Bereiche möglichst gleich sind (man muss ja leider immer noch mit Sequenzierfehlern leben).

Zunächst dachte man, dass diese Methode nur für kürzere DNS-Stränge möglich ist. etwa für 100 000 Basenpaare. Celera Genomics zeigte jedoch mit der Sequenzierung des ganzen Genoms der Fruchtfliege (Drosophila melanogaster) und schließlich dem menschlichen Genom, dass diese (bzw. eine geeignet modifizierte) Methode auch für lange DNS-Sequenzen zum Ziel führt.

# Literaturhinweise

## A.1 Lehrbücher zur Vorlesung

Peter Clote, Rolf Backofen: *Introduction to Computational Biology*; John Wiley and Sons, 2000.

Richard Durbin, Sean Eddy, Anders Krogh, Graeme Mitchison: *Biological Sequence Analysis*; Cambridge University Press, 1998.

Dan Gusfield: *Algorithms on Strings, Trees, and Sequences — Computer Science and Computational Biology*; Cambridge University Press, 1997.

David W. Mount: *Bioinformatics — Sequence and Genome Analysis*, Cold Spring Harbor Laboratory Press, 2001.

Pavel A. Pevzner: *Computational Molecular Biology - An Algorithmic Approach*; MIT Press, 2000.

João Carlos Setubal, João Meidanis: *Introduction to Computational Molecular Biology*; PWS Publishing Company, 1997.

Michael S. Waterman: *Introduction to Computational Biology: Maps, Sequences, and Genomes*; Chapman and Hall, 1995.

## A.2 Skripten anderer Universitäten

Bonnie Berger: *Introduction to Computational Molecular Biology*, Massachusetts Institute of Technology, http://theory.lcs.mit.edu/~bab/01-18.417-home.html;

Bonnie Berger, *Topics in Computational Molecular Biology*, Massachusetts Institute of Technology, Spring 2001, http://theory.lcs.mit.edu/~bab/01-18.418-home.html;

Paul Fischer: *Einführung in die Bioinformatik* Universität Dortmund, Lehrstuhl II, WS2001/2002, http://ls2-www.cs.uni-dortmund.de/lehre/winter200102/bioinf/

Richard Karp, Larry Ruzzo: *Algorithms in Molecular Biology*; CSE 590BI, University of Washington, Winter 1998. http://www.cs.washington.edu/education/courses/590bi/98wi/

Larry Ruzzo: *Computational Biology*, CSE 527, University of Washington, Fall 2001; http://www.cs.washington.edu/education/courses/527/01au/

Georg Schnittger: *Algorithmen der Bioinformatik*, Johann Wolfgang Goethe-Universität Frankfurt am Main, Theoretische Informatik, WS 2000/2001, http://www.thi.informatik.uni-frankfurt.de/BIO/skript2.ps.

Ron Shamir: *Algorithms in Molecular Biology* Tel Aviv University, http://www.math.tau.ac.il/~rshamir/algmb.html; http://www.math.tau.ac.il/~rshamir/algmb/01/algmb01.html.

Ron Shamir: *Analysis of Gene Expression Data, DNA Chips and Gene Networks*, Tel Aviv University, 2002; http://www.math.tau.ac.il/~rshamir/ge/02/ge02.html;

Martin Tompa: *Computational Biology*, CSE 527, University of Washington, Winter 2000. http://www.cs.washington.edu/education/courses/527/00wi/

## A.3 Lehrbücher zu angrenzenden Themen

Teresa K. Attwood, David J. Parry-Smith; *Introduction to Bioinformatics*; Prentice Hall, 1999.

Maxime Crochemore, Wojciech Rytter: *Text Algorithms*; Oxford University Press: New York, Oxford, 1994.

Martin C. Golumbic: *Algorithmic Graph Theory and perfect Graphs*; Academic Press, 1980.

Benjamin Lewin: *Genes*; Oxford University Press, 2000.

Milton B. Ormerod: *Struktur und Eigenschaften chemischer Verbindungen*; Verlag Chemie, 1976.

Hooman H. Rashidi, Lukas K. Bühler: *Grundriss der Bioinformatik — Anwendungen in den Biowissenschaften und der Medizin*,

Klaus Simon: *Effiziente Algorithmen für perfekte Graphen*; Teubner, 1992.

Maxine Singer, Paul Berg: *Gene und Genome*; Spektrum Akademischer Verlag, 2000.

Lubert Stryer: *Biochemie*, Spektrum Akademischer Verlag, 4. Auflage, 1996.

## A.4 Originalarbeiten

Kellogg S. Booth, George S. Lueker: Testing for the Consecutive Ones property, Interval Graphs, and Graph Planarity Using PS-Tree Algorithms; *Journal of Computer and System Science*, Vol.13, 335–379, 1976.

Ting Chen, Ming-Yang Kao: On the Informational Asymmetry Between Upper and Lower Bounds for Ultrametric Evolutionary Trees, *Proceedings of the 7th Annual European Symposium on Algorithms, ESA'99*, Lecture Notes in Computer Science 1643, 248–256, Springer-Verlag, 1999.

Richard Cole: Tight Bounds on the Complexity of the Boyer-Moore String Matching Algorithm; *SIAM Journal on Computing*, Vol. 23, No. 5, 1075–1091, 1994.
s.a. *Technical Report*, Department of Computer Science, Courant Institute for Mathematical Sciences, New York University, TR1990-512, June, 1990, http://csdocs.cs.nyu.edu/Dienst/UI/2.0/Describe/ncstrl.nyu_cs%2fTR1990-512

Martin Farach, Sampath Kannan, Tandy Warnow: A Robust Model for Finding Optimal Evolutionary Trees, *Algorithmica*, Vol. 13, 155–179, 1995.

Wen-Lian Hsu: PC-Trees vs. PQ-Trees; *Proceedings of the 7th Annual International Conference on Computing and Combinatorics, COCOON 2001*, Lecture Notes in Computer Science 2108, 207–217, Springer-Verlag, 2001.

Wen-Lian Hsu: A Simple Test for the Consecutive Ones Property; *Journal of Algorithms*, Vol.43, No.1, 1–16, 2002.

Haim Kaplan, Ron Shamir: Bounded Degree Interval Sandwich Problems; *Algorithmica*, Vol. 24, 96–104, 1999.

Edward M. McCreight: A Space-Economical Suffix Tree Construction Algorithm; *Journal of the ACM*, Vol. 23, 262–272,1976.

Moritz Maaß: *Suffix Trees and Their Applications*, Ausarbeitung von der Ferienakademie '99, Kurs 2, Bäume: Algorithmik und Kombinatorik, 1999. http://www14.in.tum.de/konferenzen/Ferienakademie99/

Esko Ukkonen: On-Line Construction of Suffix Tress, *Algorithmica*, Vol. 14, 149–260, 1995.

# Index

## F



## G



## H

## P



## Q



## R

## S



## T