MASTER THESIS | MASTER'S THESIS

Titel | Title

Synonyme Codons

verfasst von | submitted by

Mag.rer.nat. Regine Zawodsky

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 992 499
Universitätslehrgang lt. Studienblatt | Postgraduate programme as it appears on the student record sheet:	Studium Generale - Das nachberufliche Studium an der Universität Wien (MA)
Betreut von | Supervisor:	ao. Univ.-Prof. i.R. Dr. Friedrich Propst

Danksagung

Das Thema dieser Masterarbeit *»Synonyme Codons«* hat sich im Laufe von Hausarbeiten während des Moduls Molekularbiologie im Studium Generale ergeben. Im Zuge meiner Masterarbeit durfte ich dieses spannende Thema weiter untersuchen.

Ich möchte mich ganz besonders bei meinem Betreuer, Herrn Univ.-Prof. Dr. Friedrich Propst, für seine Hilfe und seine unermüdliche Geduld bedanken. Er hat mich immer unterstützt und ermutigt, auch wenn ich manchmal in Sackgassen geraten bin. Danke für die vielen interessanten Diskussionen zu spannenden Details wie auch zum großen Ganzen.

Weiters möchte ich mich bei Frau Andrea Schwarzová bedanken, die mich durch den administrativen Dschungel begleitet hat.

Nicht zuletzt gilt auch meiner Freundin Dipl.-Ing. (FH) Eva Seidl mein Dank für das aufmerksame Korrekturlesen und das professionelle Layout in InDesign.

Einleitung

Das Thema *»Synonyme Codons«* erwies sich als weitaus vielfältiger als zu Beginn gedacht. Es spannt einen Bogen von der Grundfrage nach dem Leben, über Astronomie, Codierungstheorien, physikalische und chemische Untersuchungsmethoden bis hin zum Hauptthema, der Molekularbiologie.

Die Sicht auf synonyme Codons hat sich im Laufe der letzten zwanzig Jahre vollständig umgekehrt und bildet heute die Grundlage zum Verständnis vieler genetisch bedingter Krankheiten sowie zur erfolgreichen Entwicklung von Therapeutika und Impfstoffen.

Zielsetzung und Methoden

Da in vielen klassischen Lehrbüchern wenig über synonyme Codons zu finden ist, setzte ich mir zum Ziel, einen möglichst umfassenden Überblick darüber zu erstellen, welche Auswirkungen die unterschiedliche Verwendung synonymer Codons haben kann und welche zum Teil dramatischen Auswirkungen synonyme Mutationen haben können. Bis vor wenigen Jahren wurden synonyme Mutationen vernachlässigt und zum Teil nicht mit in Datenbanken aufgenommen, obwohl die Daten dazu aus Sequenzierungen von Genen bereits vorhanden waren.

Durch ausführliche Literaturrecherche – vorwiegend in wissenschaftlichen Veröffentlichungen sowie in DNA- bzw. Proteindatenbanken – wurden Sachverhalte gesammelt und verglichen.

Nach einem kurzen geschichtlichen Abriss und Darstellung der bekannten Grundlagen folgt eine allgemeine Betrachtung zu möglichen Codierungen, von denen einige aber aus verschiedenen Gründen für die Entstehung von Leben unwahrscheinlich bzw. ungeeignet scheinen.

In tiefergehenden ausführlichen Darstellungen, die durch viele eigene Grafiken und Zeichnungen ergänzt wurden, werden neuere Erkenntnisse zu Krankheiten und Konzepte, die neuen Forschungsbereichen zu Grunde liegen, aufbereitet und verständlich gemacht.

Zusammenfassung

Nachdem Friedrich Miescher bereits 1869 das Nuclein (die DNA) isoliert hatte und damals schon vermutete, dass die genetische Information in der DNA für die Synthese von Proteinen in codierter Form vorliegen muss und nachdem Albrecht Kossel die vier Nucleinbasen um 1910 entdeckt hatte, nahm die Molekularbiologie mit der Aufklärung der dreidimensionalen Helix-Struktur der DNA durch Rosalind Franklin, Francis Crick und James Watson im Jahre 1953 einen großen Aufschwung.

Nachdem 1961 von Crick und Brenner nachgewiesen wurde, dass der Code aus Tripletts besteht, wurden die 64 Codons, die jeweils einer der 20 proteinogenen Aminosäuren bzw. einem Stopp-Codon zugeordnet werden konnten, bis zum Jahre 1966 vollständig identifiziert. Spätestens ab diesem Zeitpunkt war daher bekannt, dass der genetische Code redundant sein muss, dass also den meisten Aminosäuren mehr als ein Codon zugeordnet ist. Diese Codons, die jeweils die selbe Aminosäure codieren, werden als synonyme Codons bezeichnet.

Da eine Codierung essentiell ist, um die Informationen über das Leben weiterzugeben, folgt eine allgemeine Betrachtung über mögliche Codierungen, die aber zeigt, dass viele dieser Möglichkeiten denkbar, aber für Leben nicht geeignet erscheinen. Einige dieser theoretischen Codierungsmöglichkeiten werden inzwischen von Forschungsgruppen aufgegriffen.

Das Human-Genome-Projekt, das im Jahre 2000 einen ersten Abschluss fand und die folgende verbesserte und verbilligte Möglichkeit, DNA vollständig und automatisch zu sequenzieren, erlaubte es, die zugrunde liegenden Mutationen vieler genetischer Krankheiten in den Genen genau zu bestimmen und zu lokalisieren.

Obwohl synonyme Mutationen keine Änderung der Aminosäuresequenz des betroffenen Proteins verursachen, weswegen sie ursprünglich nicht weiter beachtet wurden, können sie die Funktion der DNA oder RNA auf verschiedene Weise beeinträchtigen und sind der Grund vieler Krankheiten mit zum Teil erheblichem Krankheitswert. Eine Auswahl an synonymen Mutationen und deren Gegenüberstellung mit klinischen Symptomen zeigt die Vielfalt der Möglichkeiten auf.

Abschließend werden biotechnologische Methoden zur Herstellung lebenswichtiger Enzyme oder Hormone wie beispielsweise Insulin dargestellt sowie die Herstellung von mRNA-Impfstoffen, die erstmals während der Covid-19-Pandemie zwischen Ende 2019 und Mitte 2023 eingesetzt wurden.

Summary

Friedrich Miescher had isolated the ‘nuclein’ (DNA) in 1869 and already presumed that the genetic information instructing the synthesis of proteins must be encoded and Albrecht Kossel had discovered the four nucleic bases around the year 1910. Since the three-dimensional helical structure of DNA was deciphered by Rosalind Franklin, Francis Crick and James Watson in 1953, molecular biology has taken off.

After Crick and Brenner demonstrated in 1961 that the code consisted of triplets, the 64 codons, each of which could be assigned to one of the 20 proteinogenic amino acids or a stop codon, were fully identified by 1966. From this point onwards it was known that the genetic code is redundant and that most amino acids are assigned more than one codon. These codons, which each encode the same amino acid, are called synonyms.

Since coding is essential to pass on the information about life, I present general considerations about encoding possibilities that show that many of these possibilities seem conceivable, but not suitable for life. Some of these potential expansions of genetic encoding are now being taken up by research groups.

The Human Genome Project which was completed for the first time in 2000 and the subsequent improved and cheaper ability to sequence DNA completely and automatically made it possible to precisely determine and locate the underlying mutations in the genes of many genetic diseases.

Despite having no effect on the amino acid sequence of the respective protein, this being the reason for them to be neglected for a long time, synonymous mutations can affect DNA or RNA in various ways and are the cause of many diseases, some of which have significant disease impact. I present a selection of synonymous mutations resulting in clinical symptoms to demonstrate the variety of possibilities.

Finally, biotechnological methods for producing vital enzymes or hormones, such as insulin, are presented, as well as the production of mRNA vaccines, which were first administered during the Covid-19-pandemic.

Inhalt

1 Aufbau von DNA und RNA

1.1 Entdeckung der Struktur der DNA

Bereits im Jahr 1869 – 84 Jahre vor den Veröffentlichungen von Rosalind Franklin, Maurice Wilkins, James Watson und Francis Crick über die Struktur der DNA – beschäftigte sich der Basler Mediziner *Friedrich Miescher* mit der Suche nach der Erbsubstanz (Dahm, 2005). Er isolierte Eiterzellen aus Verbänden Kranker, die ihm die *Tübinger Chirurgische Klinik* überließ. In aufwändigen Verfahren extrahierte er daraus die Zellkerne, in der ursprünglichen Annahme, es handle sich um Proteine. Langwierige Versuche, die Substanz der Zellkerne mit Wasser, Alkohol, Säuren und Basen auszuwaschen und letztendlich die Maßnahme, noch vorhandene Proteine mit Pepsin zu verdauen, führte ihn zu einer Substanz, die er *Nuclein* nannte.

In seiner 1871 veröffentlichten Schrift *»Ueber die chemische Zusammensetzung der Eiterzellen«* ist Folgendes zu lesen:

> **Ich griff daher zu einem Mittel, dessen energische Eiweiss lösende Wirkung auch schon Anwendung in der Chemie der Albuminkörper gefunden 1), zu pepsinhaltigen Flüssigkeiten. [...] Das Sediment bestand lediglich aus isolirten Kernen, ohne irgend eine Spur von Protoplasmaresten. [...] Wir haben vielmehr hier Körper sui generis, mit keiner jetzt bekannten Gruppe vergleichbar. [...] Die Erkenntnis der Beziehungen zwischen Kernstoffen, Eiweissstoffen und ihren nächsten Umsatzprodukten wird allmählich den Vorhang lüften helfen, der die inneren Vorgänge des Zellenwachsthums noch so gänzlich verhüllt. (Miescher, 1871)**

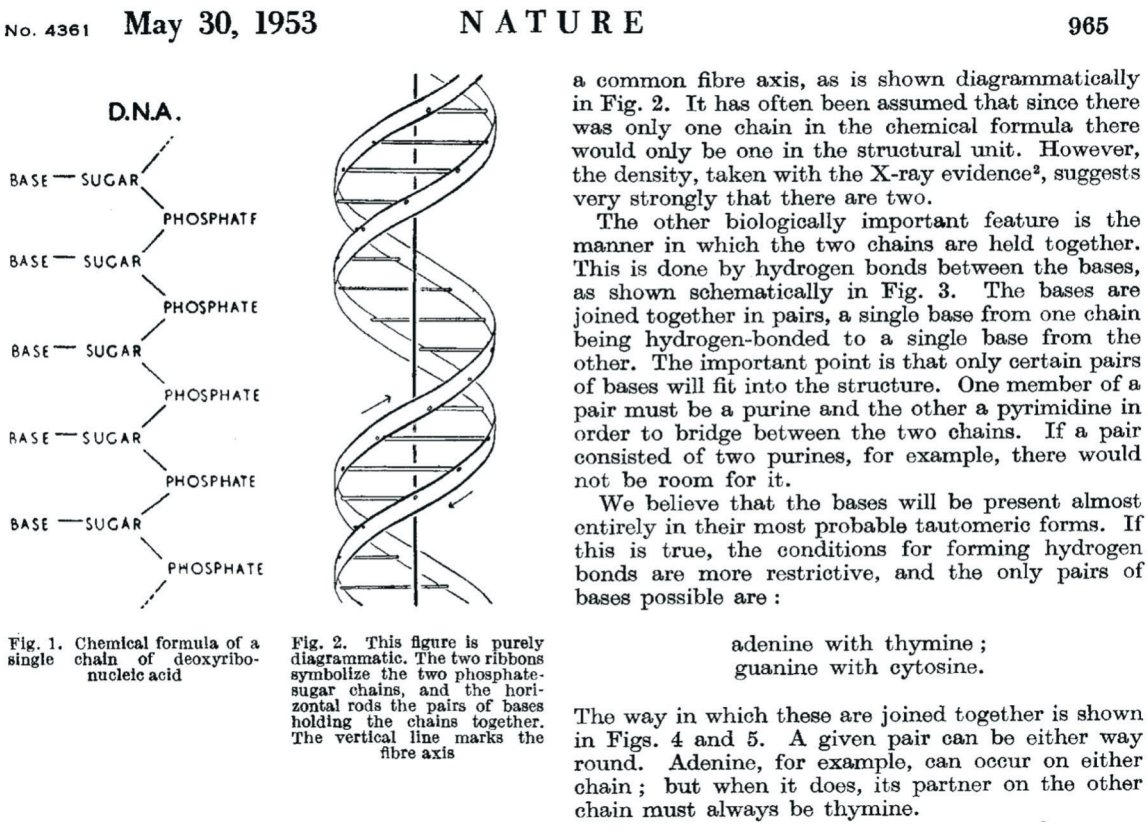
Diese Erkenntnis, dass das Nuclein kein Eiweißkörper ist, war einer der ersten wichtigen Schritte auf der Suche nach der Erbsubstanz.

1904 bewies der deutsche Mediziner Theodor Bovery, dass die Chromosomen Träger der Erbinformation sind (Graw, 2020, S. 3). Seine Untersuchungen von Chromosomen mittels gewöhnlicher Lichtmikroskope lieferten nicht die notwendige Auflösung, um die genaue Struktur der DNA darzustellen. Die Auflösung von Lichtmikroskopen ist durch die Wellenlängen des sichtbaren Lichtes begrenzt (λ ~ 400–780 nm). Sofern nicht spezielle Tricks angewendet werden, kann nur eine Auflösung von bis zu 200 nm erreicht werden. Erst die Einführung röntgenkristallografischer Aufnahmen steigerte die Auflösung bis in den Zehntelnanometer-Bereich (10^{-10} m).

Rosalind Franklin, Franklin Gosling und Maurice Wilkins arbeiteten am King's College in London mittels röntgenkristallografischer Aufnahmen an der DNA eines Thymusextraktes. Watson und Crick arbeiteten gleichzeitig daran, ein Modell der DNA zu entwickeln. Sie waren davon ausgegangen, dass die Phosphorgruppen im Inneren der Helix liegen, die Basen jedoch an der Außenseite. Nachdem Wilkins unveröffentlichte Aufnahmen von Rosalind Franklin ohne deren Genehmigung an Crick weitergegeben hatte, überarbeitete dieser sein fehlerhaftes Modell (Nickelsen, 2017, S. 55).

Im Mai 1953 veröffentlichten Watson und Crick in der Fachzeitschrift Nature das berühmte Paper über die Struktur der DNA (Watson & Crick, 1953 a), (siehe Abbildung 1.1).

Genetical Implications of the Structure of Deoxyribonucleic Acid



No. 4361 May 30, 1953 NATURE 965

D.N.A.

BASE — SUGAR — PHOSPHATE

Fig. 1. Chemical formula of a single chain of deoxyribonucleic acid

Fig. 2. This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

a common fibre axis, as is shown diagrammatically in Fig. 2. It has often been assumed that since there was only one chain in the chemical formula there would only be one in the structural unit. However, the density, taken with the X-ray evidence[2], suggests very strongly that there are two.

The other biologically important feature is the manner in which the two chains are held together. This is done by hydrogen bonds between the bases, as shown schematically in Fig. 3. The bases are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other. The important point is that only certain pairs of bases will fit into the structure. One member of a pair must be a purine and the other a pyrimidine in order to bridge between the two chains. If a pair consisted of two purines, for example, there would not be room for it.

We believe that the bases will be present almost entirely in their most probable tautomeric forms. If this is true, the conditions for forming hydrogen bonds are more restrictive, and the only pairs of bases possible are :

adenine with thymine ;
guanine with cytosine.

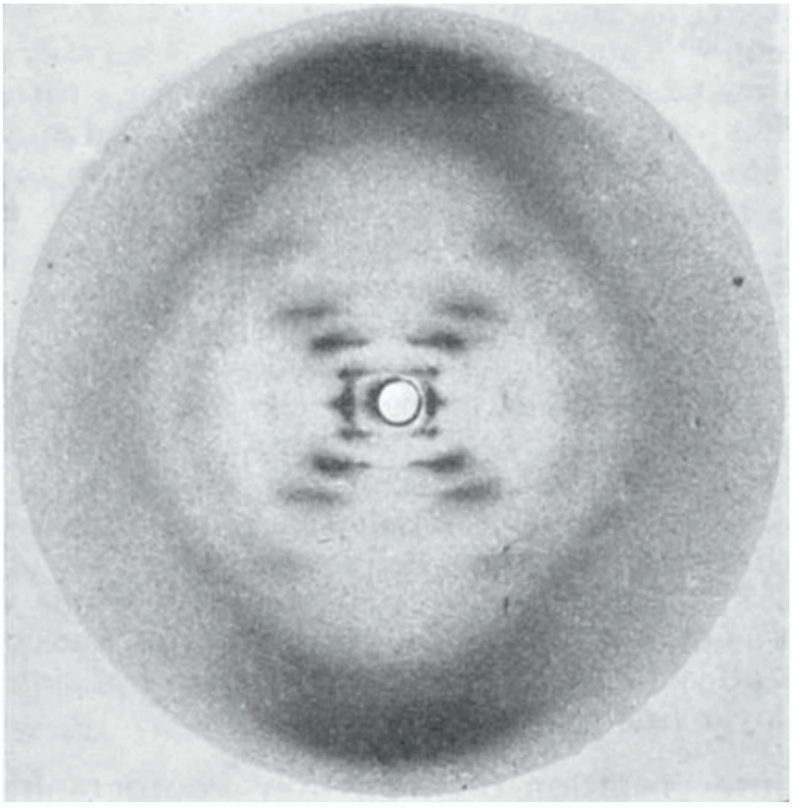
The way in which these are joined together is shown in Figs. 4 and 5. A given pair can be either way round. Adenine, for example, can occur on either chain ; but when it does, its partner on the other chain must always be thymine.

Abb. 1.1 DNA -Doppelhelix (Watson & Crick, 1953 a).

1962 bekamen Watson, Crick und Wilkins den *Nobelpreis für Physiologie oder Medizin*. Rosalind Franklin war bereits 1958 im Alter von 37 Jahren an einem Karzinom verstorben und konnte nicht mehr für den Nobelpreis vorgeschlagen werden (Nickelsen, 2017, S. 67), da dieser nicht posthum vergeben werden darf. Rosalind Franklin wurde in der Nobel Lecture von Francis Crick mit keinem Wort erwähnt. In dieser Veröffentlichung (Watson & Crick, 1953 a) wurden die korrekte Anordnung der Basen *nach innen* und das Verhältnis der Basen zueinander beschrieben, woraus sie auf die fixe Paarung von Adenin mit Thymin und von Guanin mit Cytosin schlossen. Sie folgerten:

> **It follows that in a long molecule many different permutations are possible, and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information (Watson & Crick, 1953 a).**

In der vorhergehenden Ausgabe von Nature vom April 1953 veröffentlichten Rosalind Franklin und Franklin Gosling die molekulare Konfiguration von *Natriumthymonucleat* (Natriumsalz der DNA aus Thymusextrakt), (Franklin & Gosling, 1953). Darin ist auch die in Abbildung 1.2 gezeigte berühmte Fotografie der Röntgenstrukturanalyse zu sehen, aus welcher Franklin und Gosling korrekt auf die doppelsträngige Konfiguration der Helix geschlossen hatten. Sie hatten bereits dort dargelegt, dass die Basen ins Innere der Doppelhelix zeigen müssen und die Phosphatgruppen nach außen.



Sodium deoxyribose nucleate from calf thymus. Structure *B*

Abb. 1.2 Röntgenstrukturanalyse der DNA eines Kalbsthymus (Franklin & Gosling, 1953).

Franklin und Gosling geben hier auch den daraus berechneten Abstand der Basen von 0,34 nm sowie eine Ganghöhe von 3,4 nm an. Dies entspricht auch der Gesamtlänge eines komplementären Basenpaares (inklusive Backbone) aus je einer Purinbase und einer Pyrimidinbase. Da Purin und Pyrimidin sehr unterschiedliche Größen haben, ergäbe eine Paarung zweier Purinbasen oder zweier Pyrimidinbasen stark unterschiedliche Längen. Der Abstand der beiden Stränge der Doppelhelix wäre also – je nach Paarung – einmal länger, einmal kürzer, wodurch der DNA-Doppelstrang stark verzogen würde und sich nicht mehr gleichmäßig um die (gedachte) Achse winden könnte (Alberts et al., 2021, S. 194; Watson & Crick, 1953 a).

Ebenfalls in der Aprilausgabe 1953 (Watson & Crick, 1953 b) hatten Watson und Crick bereits gemutmaßt, *[... dass die Basenpaarung einen möglichen Kopiermechanismus für genetisches Material darstellen könne]:*

> **[...] it has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material [...].**

Durch den fixen Abstand der beiden DNA-Stränge bilden sich zwei verschieden große Furchen aus: die *große Furche* (major groove) und die *kleine Furche* (minor groove). Diese Tatsache hatte bereits Rosalind Franklin in ihrem Paper dargelegt (Franklin & Gosling, 1953).

1.2 Aufbau von DNA und RNA

Die DNA (Desoxyribonucleinsäure) besteht aus langen Ketten kovalent verknüpfter Zucker und Phosphatgruppen. Weiters ist ein Kohlenstoffatom des Zuckers (in der Nomenklatur trägt es die Nummer 1, siehe Abbildung 1.3) kovalent mit einer sogenannten Base verknüpft. Zwei solcher Ketten sind jeweils über nicht-kovalente Bindungen zwischen den Basen miteinander verbunden. Der Zucker ist ein ringförmiger Fünferzucker, eine Ribose (siehe Abbildung 1.3), die bei der DNA, im Gegensatz zur RNA, als Desoxyribose vorliegt. Hier fehlt am C2-Atom eine OH-Gruppe. Der Ring liegt als *Furanoseform* vor, das bedeutet, dass sich nur 4 C-Atome im Ring befinden, während sich das 5. C-Atom außerhalb des Rings befindet (Wollrab, 2009, S. 800).

Bei Zucker (Ribose) und Basen werden die Atome gemäß spezieller Vorschriften der IUPAC *(International Union of Pure and Applied Chemistry)* durchnummeriert. Dies geschieht, um anzugeben, an welchem Atom sich eine Seitenkette oder an welchem Atom sich eine Verbindung zu einem weiteren Molekül befindet. Kommen nun Ribose oder Desoxyribose und Basen gemeinsam in der DNA oder RNA vor, so kämen einzelne Zahlen der Nummerierung doppelt vor. Es wird nun für die einzelnen Bestandteile die Nummerierung nach der Nomenklatur der IUPAC beibehalten, um jedoch Zucker und Basen eindeutig auseinanderhalten zu können, bekommen die Basen den Vorrang. Die Ringatome im Zucker werden daher alle mit einem ‘ versehen, also 1‘, 2‘ etc. (IUPAC online, blue book).

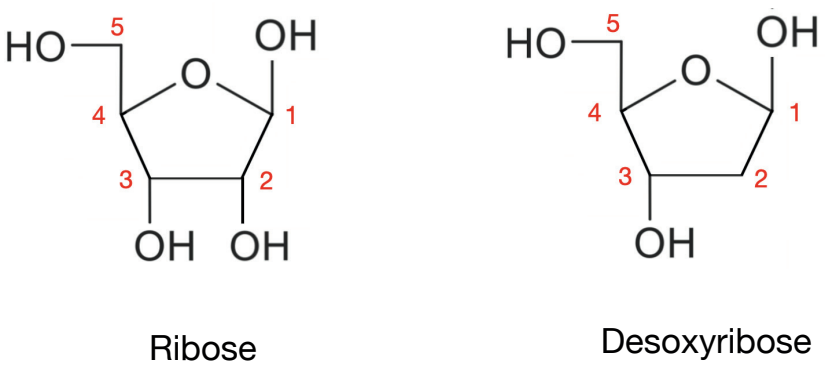


Abb. 1.3 Ribose (RNA) und Desoxyribose (DNA) unterscheiden sich durch eine fehlende OH-Gruppe am 2. C-Atom (Nummerierung gemäß IUPAC), (Buttlar et al., 2020, adaptiert).

Das 5‘-Ende bzw. das 3‘-Ende der DNA und RNA werden gemäß der Nummerierung der Desoxyribose/Ribose bezeichnet. Da eine Verlängerung immer nur am 3‘-Ende erfolgen kann, läuft die Synthese immer vom 5‘-Ende in Richtung 3‘-Ende ab. Dabei wird das 3‘C-Atom der Ribose/Desoxyribose über zwei aufeinanderfolgende, kovalente Esterbindungen einer Phosphatgruppe PO_4^{3-} (Phosphodiester) mit dem 5‘C-Atom der nächsten Ribose/Desoxyribose verknüpft. Dieses Gerüst aus Zucker und Phosphat bildet das Rückgrat (den *Backbone*) der DNA/RNA.

Bei kovalenten Bindungen handelt es sich um feste Bindungen zwischen zwei Atomen. Hierbei teilen sich zwei Atome äußere Elektronen und bilden so ein oder mehrere gemeinsame Elektronenpaare aus.

1.3 Basen und deren Paarung in DNA und RNA

Sowohl bei Eukaryoten, als auch bei Bakterien und den meisten Archaeen lagern sich je zwei antiparallele Stränge der DNA als Doppelstrang aneinander. Zwischen je zwei Zuckern dieser Stränge befindet sich jeweils ein Paar der zueinander komplementären Stickstoffbasen (Christen 2016, S. 87ff). Die Basen Adenin (A) und Thymin (T) [Uracil (U) in der RNA] sind über je zwei Wasserstoffbrücken miteinander verbunden, Cytosin und Guanin über drei Wasserstoffbrücken (Watson-Crick-Paare).

Abb. 1.4 Symbolische Darstellung der Paarung komplementärer Stickstoffbasen

Purin ist das Grundmolekül der Nucleobasen Adenin und Guanin, *Pyrimidin* ist das Grundmolekül der Nucleobasen Cytosin und Thymin (Uracil in der RNA). Sowohl Purin (Abb. 1.5) als auch Pyrimidin (Abb. 1.6) sind heterozyklische Verbindungen aus einem oder mehreren Ringen aus Kohlenstoffatomen mit einem oder mehreren Fremdatomen, in diesem Fall Stickstoff (N), (Berg et al., 2018, S. 257). Purin (Abb. 1.5) ist eine heterozyklische Verbindung aus einem Sechserring und einem Fünferring mit je zwei Stickstoffatomen in jedem Ring.

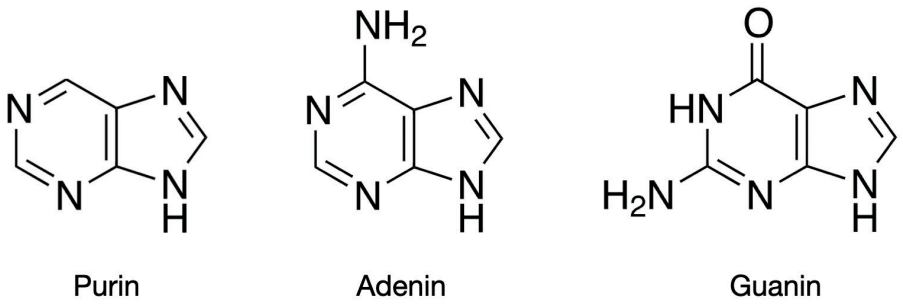


Abb. 1.5 Grundmolekül Purin, sowie die beiden davon abgeleiteten Purinbasen Adenin und Guanin (Wikipedia).

Pyrimidin (Abb. 1.6) ist ein einfacher heterocyklischer Sechserring mit zwei Stickstoffatomen, der als Grundmolekül für die Nucleobasen Cytosin und Thymin (Uracil in der RNA) dient (Berg et al., 2018, S. 259).

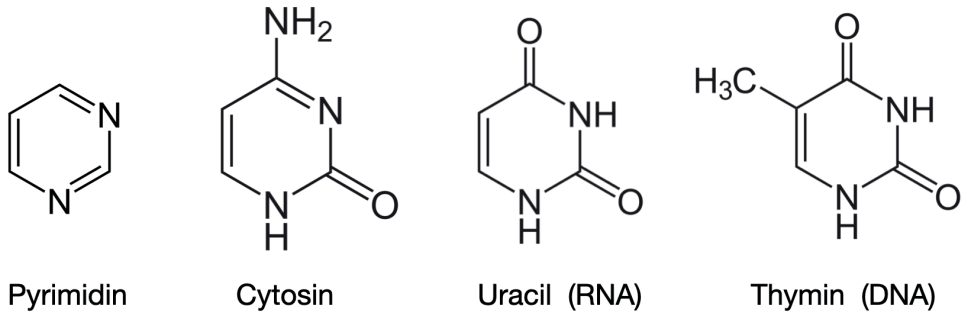


Abb. 1.6 Grundmolekül Pyrimidin sowie die abgeleiteten Pyrimidinbasen Cytosin, Uracil und Thymin (Wikipedia).

Zwischen den beiden Strängen bildet sich jeweils ein Paar aus je einer Purinbase und einer Pyrimidinbase (Christen 2016, S. 88), (siehe Abbildung 1.7). Die komplementären Basen binden aneinander über nicht-kovalente Bindungen, sogenannte Wasserstoffbrückenbindungen, die die Doppelhelix stabilisieren (Watson & Crick, 1953 a). Diese wird zusätzlich durch Van-der-Waals-Kräfte zwischen den aufeinandergestapelten Basen stabilisiert (Graw, 2020, S. 29).

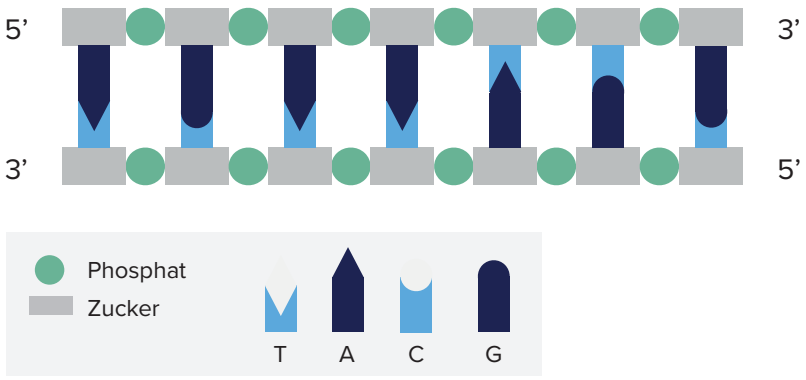


Abb. 1.7 DNA-Doppelstrang mit komplementären Basen (in linearer Darstellung).

Nicht-kovalente Wechselwirkungen sind keine echten chemischen Bindungen sondern sie beruhen auf elektrostatischen Anziehungen. Wasserstoffbrücken bilden sich zwischen einem an ein Sauerstoff- oder Stickstoffatom gebundenes Wasserstoffatom (O-H, N-H) eines Moleküls und einem Sauerstoff- oder Stickstoffatom in einer C=O- bzw. C=N-Konfiguration eines anderen Moleküls aus. Aufgrund der geringen Elektronegativität des Wasserstoffs wandern die Bindungselektronen einer O-H oder einer N-H Bindung näher an das Sauerstoff- oder Stickstoffatom, wodurch sich am Wasserstoff eine positive Teilladung ausbildet. Diese wird elektrostatisch vom anderen Bindungspartner, der ein Sauerstoff- oder Stickstoffatom in C=O- bzw. C=N-Konfiguration mit jeweils einem freien Elektronenpaar am Sauerstoff- oder Stickstoffatom enthält, angezogen. Die Bindung über Wasserstoffbrücken ist wesentlich schwächer als eine kovalente Bindung, denn zwei DNA-Stränge – wie die Doppelhelix – oder ein hybrider DNA-RNA-Doppelstrang müssen sich im Zuge der Replikation bzw. Transkription voneinander lösen können und dürfen daher keine kovalenten chemischen Bindungen eingehen. Dies ist beispielsweise bei den Wasserstoffbrücken zwischen den Basenpaaren der Fall (Buttlar et al., 2020, S. 8).

Bei Van-der-Waals-Kräften kommt es durch Ungleichverteilung der Elektronen kurzzeitig zu einem Dipol. Diese können beim nächstliegenden Atom ebenfalls einen Dipol induzieren, wodurch es zu einer schwachen elektrostatischen Anziehung kommt (Berg et al., 2018, S. 12). Übereinanderliegende Lagen der DNA-Doppelhelix werden durch Van-der-Waals-Kräfte zusätzlich stabilisiert. Van-der-Waals-Kräfte haben nur eine sehr geringe Reichweite.

1.4 N-glykosidische Bindung

Eine Base (A, T, G, C) wird über eine N-glykosidische Bindung kovalent an das 1‘C-Atom des Zuckers (Ribose oder Desoxyribose) gebunden. Hierbei wird eine OH-Gruppe des Zuckermoleküls mit einem Stickstoff einer Purin- oder Pyrimidinbase unter Abspaltung von Wasser verbunden. Abbildung 1.8 zeigt eine N-glykosidische Bindung zwischen einer Desoxyribose und Cytosin zu Desoxycytidin, einem Nucleosid. Mehrere Nucleotide bei der Biosynthese von DNA und RNA verknüpft.

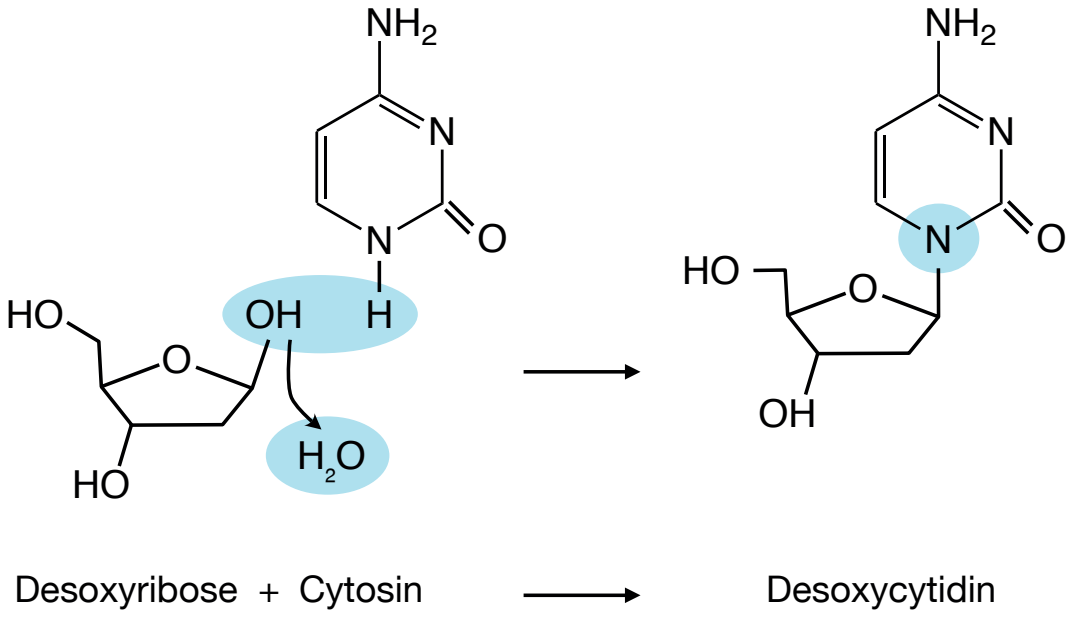


Abb. 1.8 N-glykosidische Bindung zwischen einem Zucker (einer Desoxyribose) und einer Base (Cytosin) zu einem Nucleosid (Desoxycytidin) unter Abspaltung von Wasser (Buttlar et al., 2020, S. 8, Wikipedia, adaptiert).

1.5 Unterschiede von DNA und RNA

- Die RNA hat im Backbone eine Ribose (einen zyklischen Fünferzucker als Furanose).
- Die DNA hat im Backbone eine Desoxyribose (fehlende OH-Gruppe am C2-Atom des Zuckers gegenüber der RNA). Dadurch ist die DNA stabiler (Chemie.de; Buttlar et al., 2020, S. 12).
- Die RNA hat statt des Thymins (T) *Uracil (U)* als zweite Pyrimidinbase (Buttlar et al., 2020, S.12).
- Die DNA ist ein Informationsspeicher (Buttlar et al., 2020, S.12).
- Die RNA liegt nur in manchen Viren als Informationsspeicher vor (Buttlar et al., 2020, S.12).
- Die RNA (mRNA, tRNA) ist Informationsvermittler (Alberts et al. 2021, S. 252).

- Die RNA hat auch regulatorische, katalytische und strukturelle Aufgaben (Alberts et al., 2021, S. 252) z. B. in den Ribosomen.
- Die DNA liegt in Eukaryoten als doppelsträngige Doppelhelix vor, die in Chromosomen organisiert ist. Sie befindet sich im Zellkern.
- Die DNA liegt in Prokaryoten meist als zirkulär geschlossener Doppelstrang vor, selten als linearer Doppelstrang, wie bei Borrelien. Da Prokaryoten keinen Zellkern besitzen, befindet sich die DNA im Cytoplasma.
- Die RNA kommt als Einzelstrang in mRNA *(messenger RNA)*, tRNA *(transfer RNA)* und rRNA *(ribosomale RNA)* vor. Es kommen intramolekulare Wasserstoffbrücken vor (siehe Kapitel 1.6).
- Die RNA kann in manchen Viren als Doppelstrang vorliegen.
- Die rRNA (ribosomale RNA) kommt in Ribosomen vor und ist eine nichtcodierende RNA mit katalytischer Funktion.
- Die mtDNA *(mitochondriale DNA)* und ctDNA *(Chloroplasten-DNA)* sind meist zirkuläre Doppelstrang-DNAs.
- circRNA *(zirkuläre RNA),* also geschlossene Einzelstrang-RNA kommt in allen Spezies vor (Zhou et al., 2020), ist aber noch weitgehend unerforscht.

1.6 Organisation der DNA

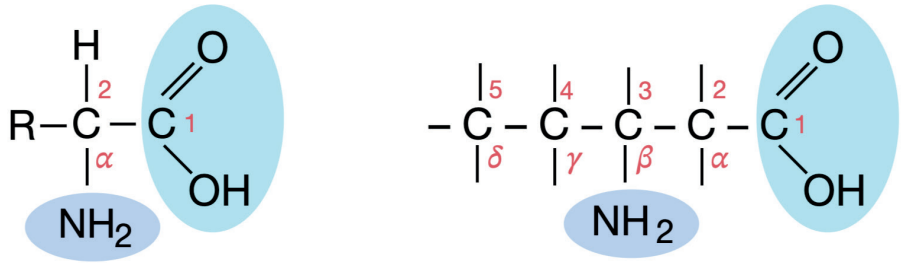
In Eukaryoten – also Lebewesen mit einem Zellkern – liegt die DNA als Doppelhelix vor. Die DNA ist in Chromosomen organisiert, wobei die Anzahl der Chromosomen je nach Spezies unterschiedlich ist. Menschen haben 2 mal 23, also 46 Chromosomen, der Haushund hingegen 78, um nur zwei Beispiele zu nennen (Graw 2020, S. 269).

In Bakterien (Prokaryoten) liegt die DNA im sogenannten Bakterienchromosom als doppelsträngiges, meist ringförmig geschlossenes Molekül vor, das mangels Zellkerns frei im Cytoplasma schwimmt. Es kommen zusätzlich extrachromosomale, kleinere, ebenfalls doppelsträngige DNA-Moleküle vor *(Plasmide),* (Graw 2020, S. 133), (siehe Kapitel 5.2, Chromosomenmutationen).

Die Chromosomen – z. B. die 23 Chromosomenpaare des Menschen – unterscheiden sich in ihrer Größe, d.h. in der Länge des DNA-Doppelstrangs, der dem jeweiligen Chromosom zu Grunde liegt. Der DNA-Doppelstrang des größten menschlichen Chromosoms, Chromosom 1, hat in etwa die Länge von 247 Millionen Basenpaaren *(Megabasenpaare, Mbp),* (Wikipedia, Chromosom und ensembl-Datenbank). Die chromosomale DNA enthält die Gene, das sind Abschnitte auf der DNA, die die Information für den Aufbau von Makromolekülen, Proteinen oder RNA-Molekülen enthalten. Außerdem sind in Genen jene Informationen enthalten, unter welchen Bedingungen diese synthetisiert werden. Die Anzahl der Gene in einem Chromosom variiert zwischen 100 und 2000. Die Gesamtzahl aller Gene variiert von Spezies zu Spezies, beim Menschen sind es etwa 21000 (Berg et al., 2018, S. 127), (siehe Kapitel 4.2).

1.7 Aminosäuren, Peptide und Proteine

Aminosäuren bestehen aus einer Kohlenstoffkette mit mindestens zwei C-Atomen, mindestens einer Carboxylgruppe –COOH, mindestens einer Aminogruppe –NH2, einem -H-Atom und einem Rest. Der Rest kann im einfachsten Fall aus einem Wasserstoffatom (–H) bestehen (wie bei Glycin) oder aus längeren, komplexen Seitenketten, die auch Verzweigungen oder Ringsysteme enthalten können. Abbildung 1.9 zeigt den prinzipiellen Aufbau und die Bezeichnung von Aminosäuren.



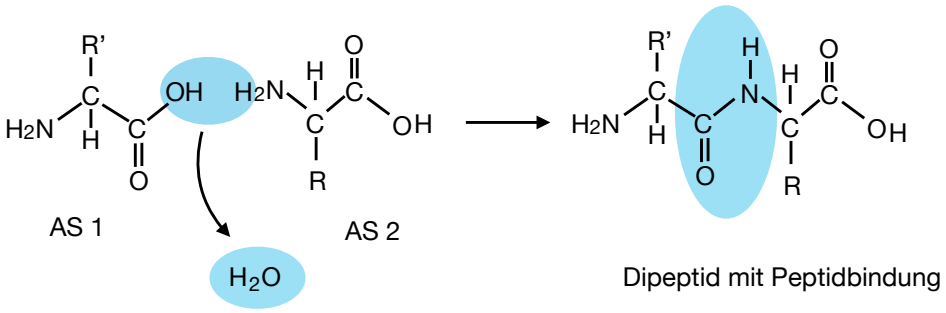
α-Aminosäure β-Aminosäure

α, β, γ	Nummerierung der C-Atome in Aminosäuren	–NH_2	Aminogruppe
1, 2, 3	alternative Nummerierung der C-Atome in Aminosäuren	–COOH	Carboxylgruppe
R	Seitenkette		

Abb. 1.9 Darstellung von Aminosäuren mit unterschiedlicher Nummerierung (Chemiezauber, online, adaptiert).

Die Nummerierung der Kohlenstoffatome erfolgt entweder von der Carboxlygruppe aus (C1) oder mit griechischen Buchstaben von jenem C-Atom, das direkt nach der Carboxylgruppe folgt (Cα). Jene Aminosäuren, die über den genetischen Code codiert werden, nennt man *kanonische* oder *proteinogene Aminosäuren*. Bei diesen handelt es sich ausschließlich um α-Aminosäuren, welche die Aminogruppe am α-Kohlenstoffatom (C2) haben. Diese bilden die Bausteine für Proteine, Peptide und Enzyme. Im menschlichen Körper kommen aber auch andere Aminosäuren vor, die die Aminogruppe an einem anderen als dem Cα-Atom haben.

Aminosäuren werden mittels einer Peptidbindung miteinander verknüpft (siehe Abbildung 1.10). Dabei verbindet sich die Aminogruppe (–NH_2) der einen Aminosäure mit der Carboxylgruppe (–COOH) der vorangegangenen Aminosäure unter Abspaltung von Wasser.



AS1	Aminosäure 1
AS2	Aminosäure 2
R/R'	Seitenketten unterschiedlicher Aminosäuren

Abb. 1.10 Mittels Peptidbindung verknüpfte Aminosäuren (Wikipedia, Peptidbindung, adaptiert).

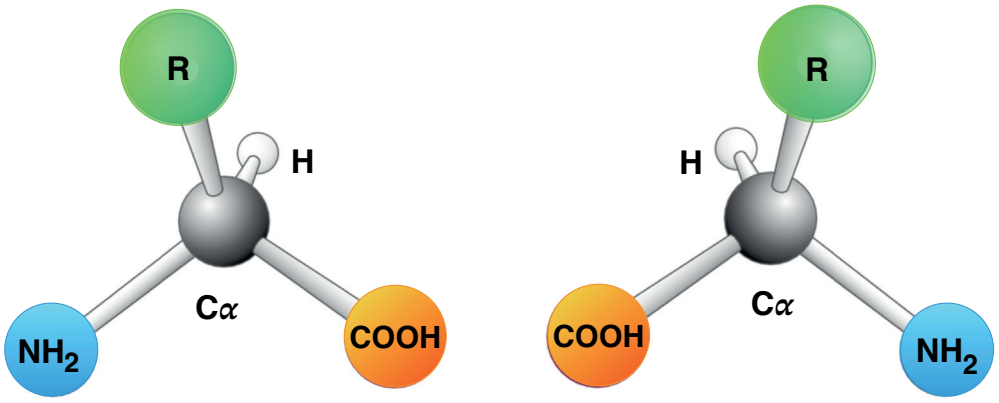
Peptide sind unverzweigte Ketten von wenigen miteinander über Peptidbindungen verknüpften Aminosäuren. Ketten mit bis zu 50 Aminosäuren werden als *Polypeptide* bezeichnet, noch längere Ketten werden als Proteine bezeichnet. Jedes Protein hat eine andere Aufgabe. Es gibt beispielsweise Strukturproteine, Enzyme, Transportkanäle, Signalmoleküle, Rezeptoren, Antikörper oder auch Faltungshelfer.

Menschliche Peptide bzw. Proteine haben sehr unterschiedliche Längen. Es gibt nur wenige Peptide im menschlichen Körper, die aus nur zwei Aminosäuren zusammengesetzt sind, wie z. B. Carnosin. Dies ist ein in Muskeln und im Gehirn vorkommendes Dipeptid, das aus Alanin und Histidin zusammengesetzt ist (Gressner et al., 2013, S. 306). Titin (Connectin) – ein Strukturprotein des Muskels – hingegen ist das größte bekannte Protein mit ca. 38 000 Aminosäuren Länge (Graw, 2020, S. 327 und S. 967).

Alle kanonischen Aminosäuren – außer Glycin – sind optisch aktiv, das heißt, sie drehen die Schwingungsebene (Polarisationsebene) von hindurchgehendem polarisiertem Licht. Optische Aktivität tritt auf, wenn ein *asymmetrisches C-Atom existiert,* das bedeutet ein C-Atom mit vier verschiedenen Liganden (Resten, R). Bei den kanonischen Aminosäuren ist dieses asymmetrische C-Atom das Cα-Atom. Glycin enthält zwei identische Liganden (zwei H-Atome) und ist daher als einziges nicht optisch aktiv.

Die Bezeichnung D-Aminosäure und L-Aminosäure geht auf eine räumliche Darstellung von Emil Fischer (Wikipedia, Emil Fischer) – die sogenannte Fischer-Projektion – zurück. Aus dieser kann man aber nicht auf die Drehrichtung der Polarisationsebene schließen. Hierbei bedeuten L laevus (lat.) – links, D dexter (lat.) – rechts.

Abbildung 1.11 zeigt die räumliche Asymmetrie der beiden möglichen Konfigurationen einer Aminosäure, die durch Drehung nicht zur Deckung zu bringen sind. Sie werden auch als *Enantiomere* oder *Stereoisomere* bezeichnet. Das asymmetrische C-Atom *(Chiralitätszentrum)* befindet sich in der Mitte eines gedachten Tetraeders, an dessen Eckpunkten sich die vier verschiedenen Liganden befinden.



NH_2	Aminogruppe	H	Wasserstoff	R	Rest (Seitenkette)
COOH	Carboxylgruppe	Cα	asymmetrisches C-Atom		

Abb. 1.11 Stereoisomere einer Aminosäure (Berg et al., 2018, S. 35 adaptiert).

In Lebewesen kommen fast ausschließlich L-Aminosäuren vor. Geringe Mengen von D-Aminosäuren wurden laut Fujii and Saito (2004) in Zähnen, Knochen und im *Beta-Amyloid-Protein* von Alzheimer-Plaques entdeckt. In manchen Geweben – wie im Dentin der Zähne – findet im Laufe des Lebens eine gewisse *Razemisierung* von L- zu D-Aminosäuren statt. Dadurch kann man auf das Sterbealter (nicht das Sterbedatum) einer Person schließen (Ritz-Timme et al., 2002).

> **[...] fand sich ein enger Zusammenhang zwischen dem Razemisierungsgrad von Asparaginsäure und dem Dentinalter.**

Die Seitenketten der kanonischen Aminosäuren haben unterschiedliche Strukturen und unterschiedliche chemische Eigenschaften (Siehe Anhang). Sie können polar oder unpolar, hydrophil oder hydrophob, sauer oder basisch sein. Die Seitenketten können aliphatisch sein oder auch Ringe enthalten, wie Tyrosin oder Phenylalanin. Für die korrekte Faltung der Proteine sind vor allem Wasserstoffbrücken, elektrostatische Wechselwirkungen und hydrophobe Wechselwirkungen verantwortlich.

Es können sich aber auch kovalente Schwefelbrücken (Disulfidbrücken) zwischen zwei Cysteinresten ausbilden. Diese stellen echte kovalente Bidungen dar. Disulfidbrücken spielen zum Beispiel eine wichtige Rolle bei der dreidimensionalen Stabilisierung von Insulin. (Berg et al., 2018, S. 482).

1.8 Transkription

Um von der in der DNA gespeicherten Information zu Proteinen zu gelangen, sind zwei Schritte erforderlich: Transkription und Translation.

Die Weitergabe der Information von der DNA über die RNA zum Protein wurde als

zentrales Dogma der Molekularbiologie

bezeichnet, das auf Crick (1958) zurückgeht. Dieses Dogma musste 1970 durch die Entdeckung der reversen Transkriptase, die RNA in DNA umschreiben kann, ergänzt werden. Howard Temin und David Baltimore bekamen dafür 1975 den Nobelpreis (DocCheck Flexikon, Reverse Transkriptase).

1.8.1 Ablesen und Übertragen der Information auf die mRNA

Zuerst muss aus der DNA im Zellkern ein Abschnitt, der der Länge eines Gens entspricht, in ein Transportmolekül, die sogenannte prä-Messenger-RNA (prä-mRNA), übertragen werden. Bei Eukaryoten enthält dieses Primärtranskript Bereiche, die nicht für Proteine codieren und entfernt werden müssen. Außerdem müssen zu Beginn eine Kappe und am Ende der mRNA noch ein Poly-A-Schwanz angehängt werden (Christen et al., 2016, S. 111). Die Ablesung der Information aus der DNA und die daraus folgende Synthese der mRNA wird als Transkription bezeichnet. Die mRNA dient ausschließlich der *Übermittlung des Bauplans für ein Protein, der in der DNA festgelegt ist.* Es wird jeweils nur ein kleiner Teil der DNA eines Chromosoms abgelesen, nämlich nur die Länge eines einzigen Gens. Die 23 Chromosomenpaare enthalten ca. 20 000 Gene (siehe Kapitel 3, Human Genom Project). Da sowohl Chromosomen als auch Gene unterschiedliche Längen haben, hat ein Chromosom ungefähr 100 bis einige tausend Gene.

Um die Information aus der DNA abzulesen und eine mRNA zu synthetisieren, wird ein Enzym, eine sogenannte *Polymerase* benötigt, die den Zusammenbau der RNA katalysiert. Polymerasen sind spezifische Enzyme, die jeweils einen oder nur sehr wenige Vorgänge katalysieren.

Für den ersten Schritt der Transkription ist eine DNA-abhängige RNA-Polymerase nötig, die RNA-Polymerase II (Pol II). Diese muss den Startpunkt *(Promotor)* suchen, ab welchem die DNA in RNA übersetzt werden soll *(Initiation)*. Da die DNA als gewundene Doppelhelix vorliegt, kann die Information nicht direkt abgelesen werden. Ein spezifisches Enzym aus der *Familie der Helicasen* muss daher die DNA-Doppelhelix zuerst für ein kurzes Stück von ca. 15 Basen Länge entspiralisieren (Nordheim et al., 2018, S. 309).

In diesem Teil liegt die DNA dann in Form von zwei Einzelsträngen vor, von denen der Matrizenstrang als Vorlage dient. Beginnend beim Startpunkt werden nun Triphosphate der vier Basen

A, C, G und U (Adenosintriphosphat, Cytosintriphosphat, Guanosintriphosphat und Uridintriphosphat), die sich im Kernplasma befinden, mit Hilfe der DNA-abhängigen RNA-Polymerase II aneinandergereiht. Die Ribonucleosidtriphosphate bestehen aus der jeweiligen Base, einem Zucker (Ribose) und drei Phosphatresten. Beim Verknüpfen werden die beiden endständigen Phosphatreste als Pyrophosphat abgespalten (Berg et al., 2018, S. 146). Die RNA-Polymerase synthetisiert durch Verknüpfung der Nucleosidtriphosphate eine zum Matrizenstrang komplementäre prä-mRNA.

Abb. 1.12 zeigt eine teilweise entspiralisierte DNA-Doppelhelix. DIe RNA-Polymerase II läuft an der DNA entlang und verknüpft die Ribonucleosidtriphosphate mit der wachsenden Kette. Dies wird als *Elongation* bezeichnet. Die RNA-Polymerase II katalysiert die Phosphodiesterbindungen (Alberts et al., 2021, S. 253). Es bildet sich vorübergehend ein ca. neun Basenpaare langes DNA-RNA-Hybrid, ein sogenanntes Heteroduplex (Graw, 2020, S. 78).

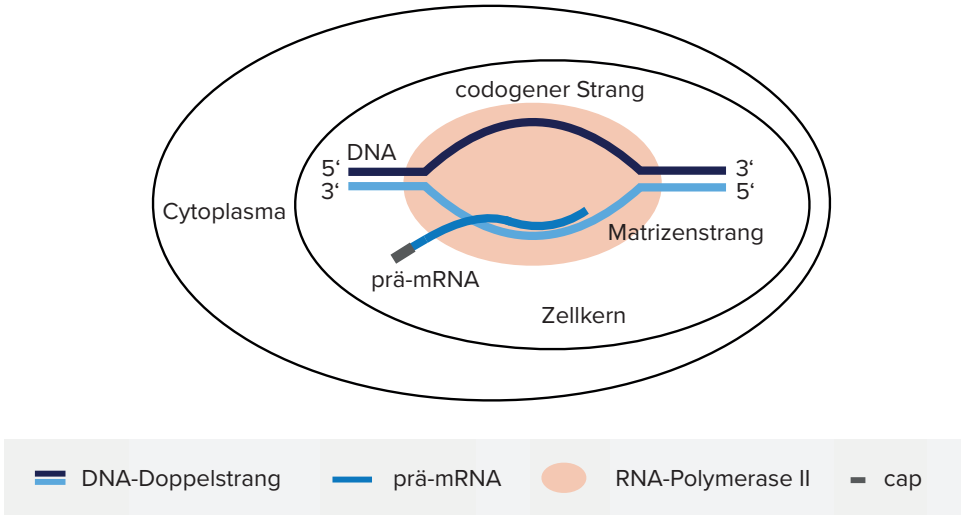


Abb. 1.12 Transkription in Eukaryoten, Initiation und Elongation.

Die Basen der neu entstehenden RNA sind komplementär zu jenen Basen der als Matrize dienenden DNA. In der durch die Helicase geöffneten Blase transkribiert die RNA-Polymerase II sukzessive die einzelnen Nucleotide. Vor der RNA-Polymerase II entspiralisiert die Helicase den DNA-Doppelstrang, dahinter wird dieser wieder geschlossen. Es werden ca. 50 Ribonucleotide pro Sekunde angefügt (Christen et al., 2016, S. 111). Eine spezielle Sequenz, der Terminator, ist das Signal für die Polymerase II, die Transkription zu beenden (Termination).

RNA-Polymerasen machen bei der Transkription im Gegensatz zu den DNA-Polymerasen relativ viele Fehler ($\sim 10^{-4}$). Da die mRNA als Zwischenprodukt aber deutlich kürzer lebt als die DNA, und da die RNA außerdem kein dauerhafter Informationsspeicher ist, sind Fehler hierbei weniger problematisch. Fehlerhafte mRNA-Transkripte und fehlerhafte Proteine, die daraus translatiert wurden, werden schnell wieder abgebaut. Da von einem DNA-Abschnitt viele mRNAs transkribiert werden, können die fehlerhaften mRNAs und Proteine vernichtet/abgebaut werden, ohne dass ein zu großer Mangel bei der Synthese der Proteine entsteht (Alberts et al., 2021, S. 254).

Die Transkription besteht aus drei Schritten:

a) Initiation
b) Elongation
c) Termination

Die prä-mRNA wird bereits während der Entstehung am 5‘-Ende mit einer Kappe (cap/capping) versehen, die dem Schutz gegen Nucleasen (abbauende Enzyme) und nach der Prozessierung zum besseren Transport aus dem Zellkern ins Cytoplasma dient. Diese Kappe besteht aus einem modifizierten Guanosin, dem 7-Methylguanosin, das über eine Triphosphatbrücke an die 5‘-Position der Ribose gebunden wird (Alberts et al., 2021, S. 261f).

1.8.2 Exons – Introns – Spleißen

Diese durch Transkription entstandene prä-mRNA besteht aus sich abwechselnden *Exons* und *Introns*. *Exons* (expressed regions) sind jene Bereiche, die als Vorlage für die Synthese der Aminosäurenkette dienen. Sie enthalten den codierenden Anteil der prä-mRNA. *Introns* (intragenic regions) sind Abschnitte unterschiedlicher Länge zwischen den Exons, die nichts zur Codierung beitragen, aber wichtige Informationen zur weiteren Verarbeitung, zum Beispiel dem *Spleißen* haben. Im Allgemeinen sind die Introns deutlich länger als die Exons.

Gilbert (1978) schreibt von einer neuen Entdeckung, dass die DNA, die in eine Aminosäuresequenz übersetzt werden soll, von Bereichen unterbrochen ist, die nichts zur Codierung beitragen und die herausgeschnitten werden. Gilbert schlägt den Namen *Introns* für diese Bereiche vor.

> **[...] which I suggest we call introns (for intragenic regions) – alternating with regions which will be expressed – exons.**

Gilbert (1978) sieht den Vorteil in den *Mosaikgenen*, darin, dass durch das Spleißen das Muster schneller verändert werden kann, als die selten eintretenden Punktmutationen es vermögen. Dadurch wird die Evolution wesentlich schneller vorangetrieben. Beim *Spleißen* der prä-mRNA werden die Introns herausgeschnitten und die verbleibenden Exons wieder zusammengefügt, sodass eine *reife mRNA* entsteht. Nach dem Herausspleißen der Introns aus der prä-mRNA wird auch das 3‘-Ende mit einem Schutz aus 20 bis 200 Adeninnucleotiden versehen (Polyadenylierung, Poly-A-Schwanz). Der Poly-A-Schwanz dient der Stabilisierung der mRNA und ist sowohl für den Transport der mRNA ins Cytoplasma als auch für die Initiation der darauffolgenden Translation im Cytoplasma notwendig (Passmore & Coller, 2022). Erst wenn das Spleißen beendet ist und beide Enden mit einem Schutz versehen sind – Kappe und Poly-A-Schwanz – kann die nun als *reife mRNA* bezeichnete RNA durch die Kernporen aktiv ins Cytosol transportiert werden, wo die Translation (Synthese der Poplypeptide) stattfindet.

Abb. 1.13 zeigt – ausgehend von einer prä-mRNA – einige der Möglichkeiten, wie die Exons nach dem Spleißen zusammengesetzt sein können (alternatives Spleißen). Erst nach dem Spleißen wird der Poly-A-Schwanz angefügt.

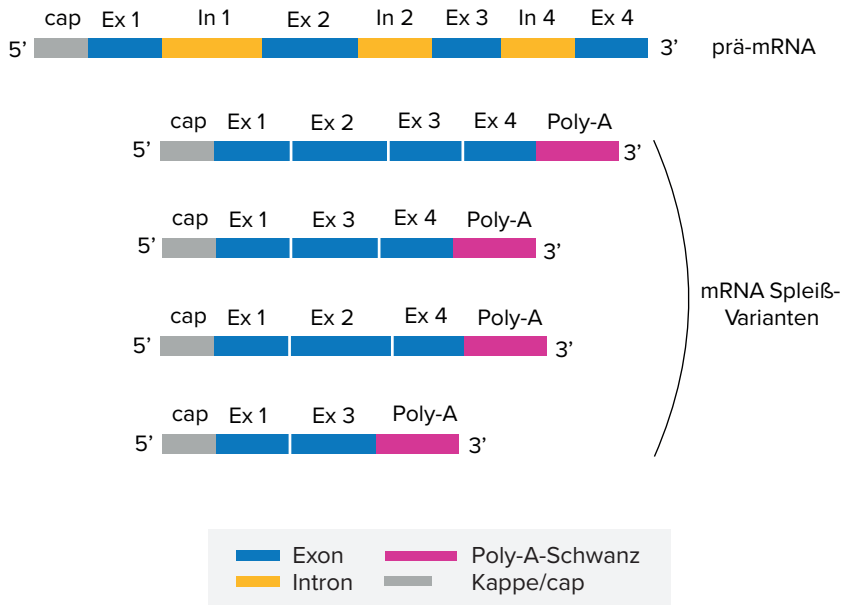


Abb. 1.13 Mögliche Spleißvarianten beim alternativen Spleißen einer prä-mRNA in Eukaryoten.

Die kombinatorischen Möglichkeiten, die Exons miteinander zu verbinden, sind – besonders bei Genen mit sehr vielen Exons – weitaus höher, als die wirklich genutzte Anzahl. Dennoch trägt der Vorgang des Spleißens zusammen mit der Möglichkeit der baukastenartigen Neuzusammensetzung der Exons dazu bei, dass sich aus den beim Human Genome Project gefundenen 20 000 Genen mRNAs für mehr als 100 000 verschiedene Proteine zusammensetzen lassen (siehe Kapitel 4). Die 1941 aufgestellte *Ein-Gen-ein-Enzym-Hypothese von Beadle und Tatum* musste also modifiziert werden (Graw, 2020, S. 70).

Ein Intron kann sich an beliebiger Stelle befinden, nicht nur exakt zwischen zwei Codons, sondern auch in der Mitte desselben. Befindet sich ein Intron nach der dritten Base eines Codons, also zwischen zwei Codons, so wird es als Phase-0-Intron bezeichnet. Analog dazu gibt es Phase-1-Introns nach der ersten Base und Phase-2-Introns nach der zweiten Base. Wird nun ein Intron oder mehrere Introns gemeinsam mit einem oder mehreren Exons herausgespleißt, so muss der im nächsten verbleibenden Exon folgende Teil das ursprüngliche Codon zu einem Triplett ergänzen (siehe Abbildung 1.13). Ist dies nicht der Fall, so ergibt sich eine fehlerhafte Zusammensetzung mit mehr oder weniger als drei Basen. In Abbildung 1.13 ist dargestellt, dass nach dem Spleißen eine Rasterverschiebung auftreten kann. Es ergibt sich daher bald ein vorzeitiges Stopp-Codon und ein verkürztes unbrauchbares Protein.

Ein sauberes Spleißprodukt ergibt sich beim alternativen Spleißen nur dann, wenn das erste und das letzte Intron, die gemeinsam mit Exons herausgespleißt werden sollen, die selbe Phase haben. Dadurch wird die Anzahl möglicher Spleißprodukte reduziert.

Nguyen et al. (2006) untersuchten die Phasenverteilung von Introns und ermittelten die Verteilung von Phase-0- : Phase-1- : Phase-2-Introns. Obwohl die Fragestellung eine andere war, zeigt sich bei den Ergebnissen, dass für alle Spezies die Phase-0-Introns die Häufigsten sind und die Phase-2-Introns die Seltensten. Es kommen jeweils alle drei Phasen vor, wobei die Häufigkeit der Phase-2-Introns in keinem Fall unter 20 Prozent liegt. Man kann also nicht a priori davon ausgehen kann, dass alle Introns dieselbe Phase haben.

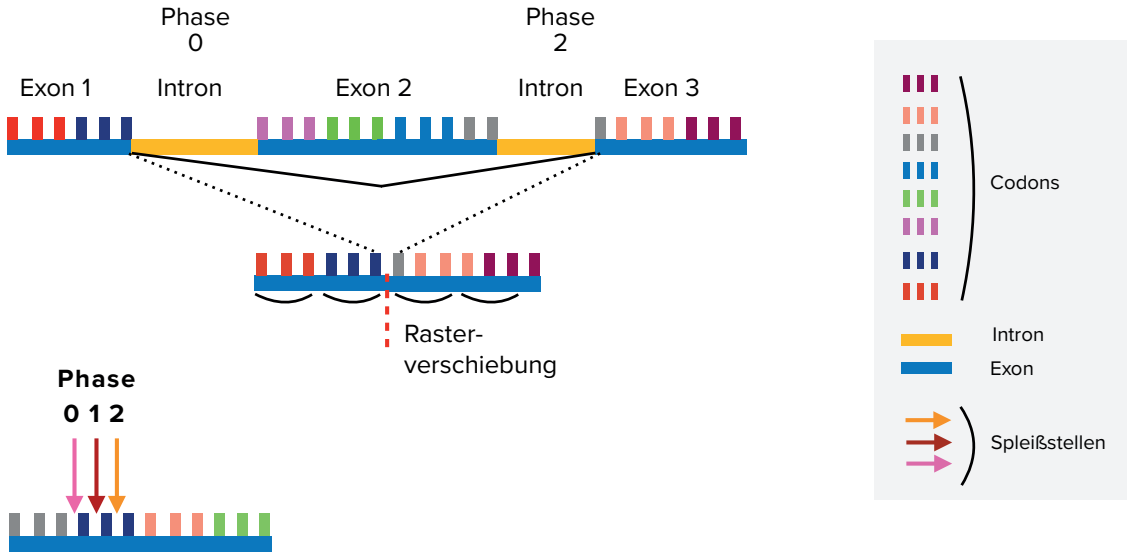


Abb. 1.14 Mögliche Positionen, an denen sich Introns befinden können und mögliche Rasterverschiebung beim Herausspleißen von Introns verschiedener Phase.

Abb. 1.15 zeigt die Situation nach Beendigung der Transkription und des Spleißens bei Eukaryoten. Die Blase im DNA-Doppelstrang ist wieder vollständig geschlossen. Das 5‘-Ende der mRNA ist mit einer Kappe aus 7-Methylguanosin versehen, das 3‘-Ende mit einem Poly-A-Schwanz. Die so geschützte reife mRNA befindet sich immer noch im Zellkern.

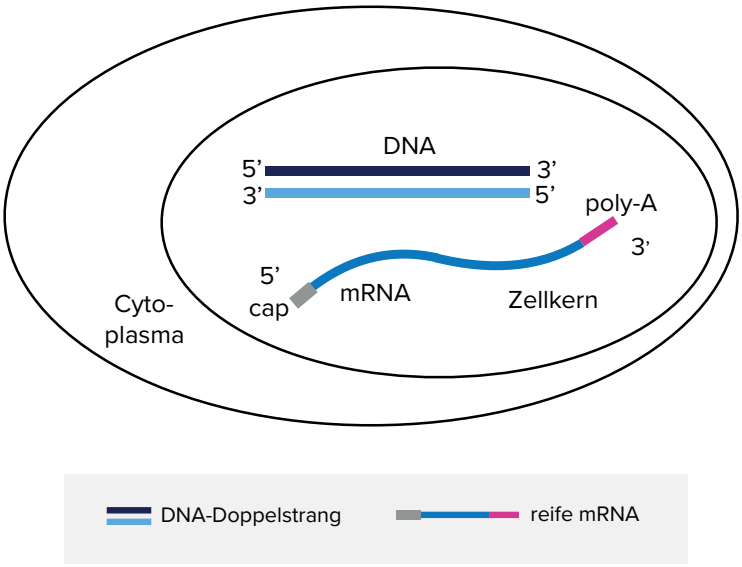


Abb. 1.15 Abgeschlossene Transkription nach Spleißen.

Sobald nach erfolgreichem Capping, Spleißen und Anfügen des Poly-A-Schwanzes eine *reife mRNA* erzeugt wurde, wird diese *aktiv* durch *Kernporen* mit Hilfe von Kernporenproteinen aus dem Zellkern in das Cytoplasma transportiert, wo die Translation (Synthese der Proteine) stattfinden kann (Alberts et al., 2021, S. 558). Bei der Translation wird die in der mRNA codiert vorliegende Information decodiert und zur Synthese von Proteinen verwendet. Dies geschieht mit Hilfe von *Ribosomen*, die eine Art Proteinproduktionsstätte darstellen und die in allen Eukaryoten, Prokaryoten, Archaeen sowie in einigen Zellorganellen wie in Mitochondrien und Chloroplasten vorkommen.

Die reife mRNA enthält zwischen Kappe und Poly-A-Schwanz nur mehr die nahtlos aneinandergefügten Exons des Gens. Diese umfassen den codierenden Bereich sowie die 5‘- und 3‘-untranslatierten Regionen (5‘-UTR, 3‘-UTR).

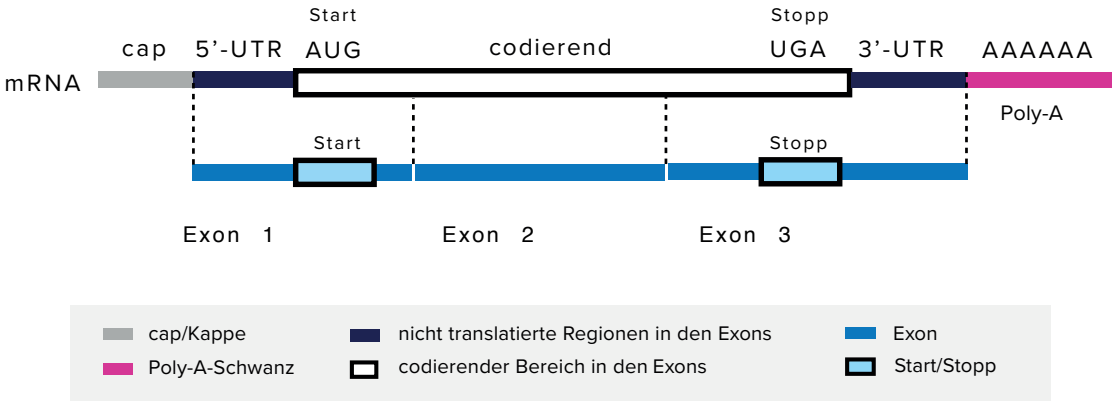


Abb. 1.16 Einteilung einer reifen mRNA in Exons, bestehend aus codierendem Bereich sowie den untranslatierten Bereichen 5‘-UTR und 3‘-UTR, sowie der Kappe und dem Poly-A-Schwanz.

Nachdem alle Introns entfernt wurden, verbleibt nahe den ursprünglichen Spleißstellen – ca. 20 bis 24 Nucleotide (nt) in 5‘-Richtung entfernt – jeweils ein mehrteiliger Proteinkomplex EJC (Exon-Junction-Complex). Der erste Exon-Junction-Komplex befindet sich im ersten Exon, der letzte im vorletzten Exon jeweils knapp oberhalb der darauf folgenden Nahtstelle zum nächsten Exon. Nach einem regulären Stopp-Codon gibt es keinen EJC mehr. Der EJC wird mit der reifen mRNA aus dem Kern ins Cytoplasma exportiert, wo dann die Translation stattfindet. Die Stoßstelle zwischen den Exons wird als Exon-Exon-Junction (EEJ) bezeichnet (siehe Abbildung 1.17).

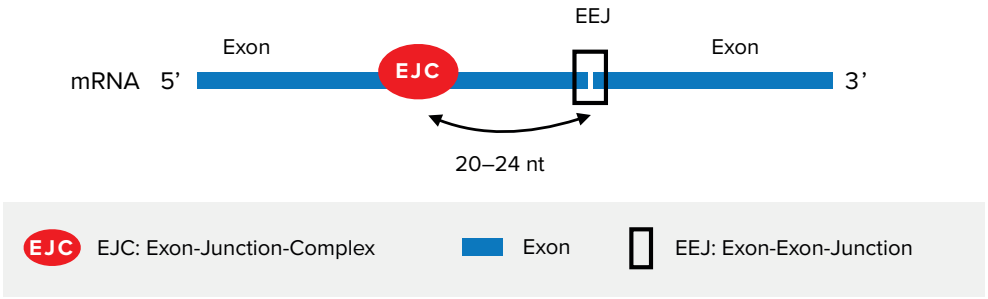


Abb. 1.17 Anordnung der Exon-Exon-Junction (EEJ) und des Exon-Junction-Komplexes (EJC) der reifen mRNA vor der ersten Translationsrunde.

1.9 Aufbau und Funktion von Ribosomen

Ribosomen haben eine Größe von ~ 20–25 nm und bestehen jeweils aus einer *großen* sowie einer *kleinen Untereinheit*. Die Gesamtgröße der Ribosomen und auch die Größe der Untereinheiten unterscheiden sich jedoch zwischen Eukaryoten und Prokaryoten. Die Ribosomen von Prokaryoten sind etwas kleiner als die der Eukaryoten, sie können aber alle den gemeinsamen genetischen Code entziffern (siehe Kapitel 3).

Geringfügige Abweichungen im Code gibt es bei einigen wenigen Zellorganellen, nämlich bei den Mitochondrien und den Chloroplasten, die gemäß der Endosymbiontentheorie in die Zellen von Eukaryoten und Prokaryoten eingewandert sind (Heinrich et al., 2014, S. 157). Die beiden Untereinheiten (UE) der Ribosomen bestehen jeweils aus Proteinen und ribosomalen RNAs (rRNAs) – je nach Domäne – aus 1 bis 3 unterschiedlichen rRNAs und 21 bis 50 verschiedenen Proteinen (Heinrich et al., 2014, S. 604). Die Größe der rRNAs werden in Svedberg (S) angegeben, einer Einheit eines Sedimentationskoeffizienten für die Größe von Teilchen in einer Ultrazentrifuge. Laut Heinrich et al. (2014, S. 582) setzen sich die Ribosomen und ihre Untereinheiten (UE) folgendermaßen zusammen:

Eukaryoten:	Große UE	3 rRNAs:	28 S, 5 S und 5,8 S	Proteine:	49
	Kleine UE	1 rRNA:	18 S	Proteine:	33
Prokaryoten:	Große UE	2 rRNAs:	23 S, 5 S	Proteine:	34
	Kleine UE	1 rRNA:	16 S	Proteine:	21

Bei den ribosomalen RNAs (rRNAs) handelt es sich um *nichtcodierende RNAs,* die im Nucleolus (Kernkörperchen) von der DNA-abhängigen RNA-Polymerase I von der ribosomalen DNA (rDNA) abgelesen und transkribiert werden. Diese rRNAs werden aber *nicht* in Proteine übersetzt. Die rRNA verbleibt im Nucleolus, wo sie auf die ribosomalen Proteine, die im Cytoplasma synthetisiert werden, wartet. Hier werden dann die beiden ribosomalen Untereinheiten einzeln zusammengesetzt. Die rRNAs haben nicht nur *strukturelle* sondern auch *katalytische Aufgaben* bei der Proteinsynthese. Die 28S-rRNA der großen Untereinheit katalysiert beispielsweise die Bildung der Peptidbindung bei der Translation (Heinrich et al., 2014, S. 580 und S. 604).

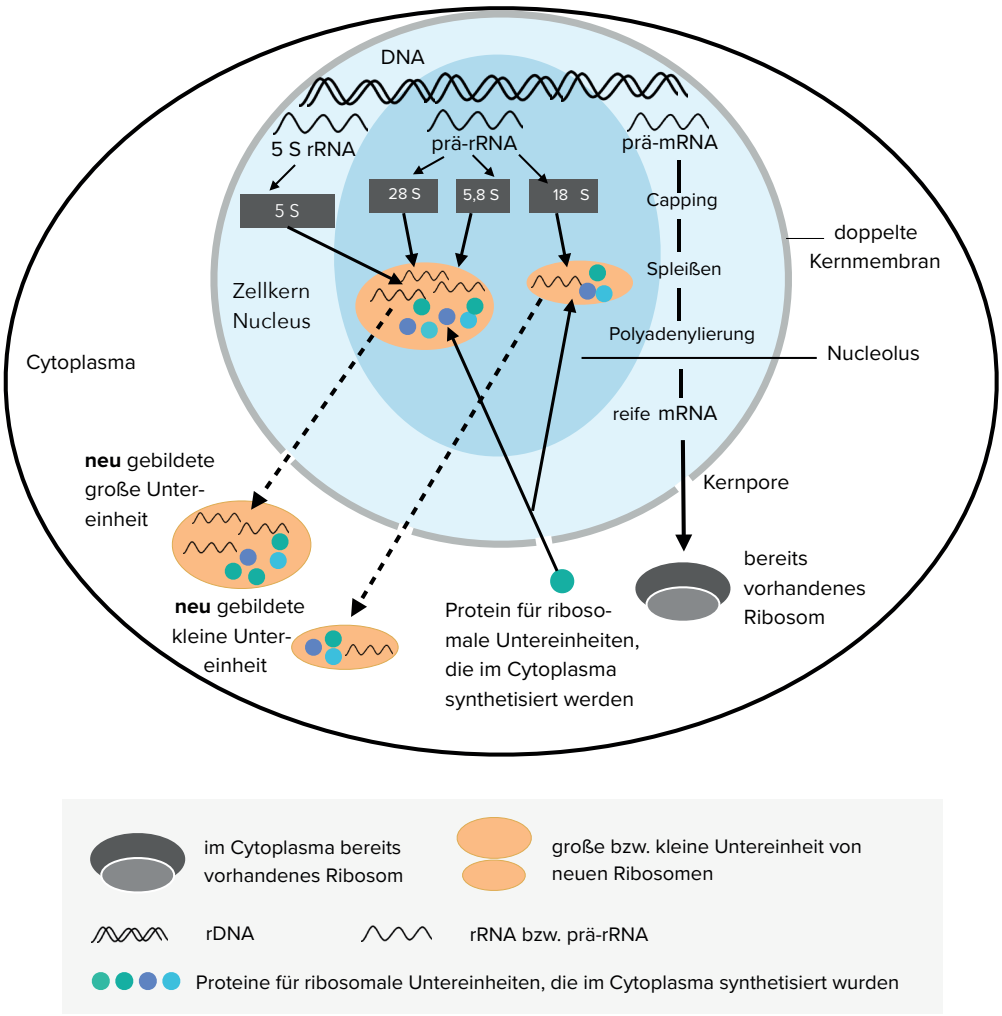


Abb. 1.18 Synthese und Zusammenbau der Untereinheiten eines Ribosoms aus rRNAs und Proteinen in Eukaryoten.

Abb. 1.18 zeigt den Zusammenbau neuer Ribosomen in Eukaryoten im Nucleolus (im Kernkörperchen). Hierbei handelt es sich um eine verdichtete Struktur innerhalb der Zellkerns (Nucleus), die aber *nicht durch eine Membran abgetrennt ist.* Nach dem Transport der reifen mRNA durch die Kernporen ins Cytoplasma wird sie dort durch Ribosomen in Proteine translatiert, darunter auch in jene Proteine, die für den Zusammenbau der neuen Ribosomen benötigt werden.

Diese müssen dann ebenfalls durch die Kernporen zurück in den Nucleolus transportiert werden, wo sie gemeinsam mit den rRNAs zu den ribosomalen Untereinheiten zusammengesetzt werden. Die fertigen ribosomalen Untereinheiten werden danach wieder zurück ins Cytoplasma transportiert. Im Cytoplasma vereinigen sich je eine große und eine kleine Untereinheit zur Proteinsynthese. Die unterschiedlichen Proteine und die mRNAs sind so groß, dass sie sich nicht mehr durch Diffusion zwischen Cytoplasma und Kern bewegen können. Die benötigten *Kernporen* sind komplizierte Strukturen, die aus ungefähr 30 verschiedenen Proteinen bestehen und die nicht nur große Moleküle, sondern auch Makromolekule wie Proteine und Ribonucleotidproteinkomplexe *aktiv* zwischen den Kompartimenten hin und her transportieren können (Alberts et al., 2021, S. 555).

1.10 Translation

Die Translation ist der letzte Schritt der Proteinbiosynthese, bei dem mit Hilfe der Ribosomen die einzelnen Aminosäuren aneinandergefügt werden. Damit ein Ribosom dies tun kann, wird ein Transportmolekül (Transfer-RNA, tRNA) benötigt, welches die Aminosäuren zu den Ribosomen transportiert. Sowohl die Ribosomen als auch die tRNA, befinden sich im Cytoplasma. Während die mRNA die Information für die Abfolge der Aminosäuren übermittelt, transportiert die tRNA diese physisch. Die tRNA ist ebenso wie die rRNA eine nichtcodierende RNA (Graw, 2020, S. 86; Heinrich et al., 2014, S. 147).

Die tRNA selbst wird bei Eukaryoten – wie alle anderen RNAs – von der DNA in Genen abgelesen und transkribiert. Die Transkription geschieht – wie bei der 5S-rRNA – im Zellkern durch die RNA-Polymerase III. Die humane tRNA wird in nahezu 600 Genen, die sich auf unterschiedlichen Chromosomen befinden, codiert (Graw, 2020, S. 96). Zusätzlich gibt es noch einige mitochondriale tRNA-Gene. Bei der Transkription entsteht eine prä-tRNA, die auch Introns enthält, die – im Unterschied zu den Spleißvorgängen bei der mRNA – durch eine Endonuclease herausgeschnitten werden. Die Länge der tRNA wird von verschiedenen Autor*innen unterschiedlich angegeben, beispielsweise mit 65 bis 110 Nucleotiden bei Heinrich et al., 2014 (S. 147).

Die tRNA hat unterschiedliche Bereiche, von denen einige hochkonserviert sind. Nach der Transkription erfolgen auch bei der tRNA noch diverse Basenmodifikationen, beispielsweise in Dihydrouridin, Pseudouridin (Ψ) sowie das Anfügen einer CCA-Sequenz am 3‘-Ende des Akzeptorstammes.

1.10.1 Aufbau der tRNA

Die reife tRNA ist zu einer sogenannten Kleeblattstruktur gefaltet (siehe Abb. 1.19), die drei fixe Schleifen und eine variable Schleife aufweist. Die fixen Schleifen sind die Dihydrouridinschleife, die TΨC-Pseudouracilschleife und die Anticodonschleife. Die Bereiche zwischen den Schleifen – als Arme oder Stamm bezeichnet – werden durch intramolekulare Basenpaarungen (Wasser-

stoffbrücken) stabilisiert. Die variable Schleife hat unterschiedliche Länge und keinen Arm und ist daher ohne intramolekulare Basenpaarungen.

Die *Anticodonschleife* hat in den Positionen *34, 35 und 36* einen Bereich aus drei Basen, das sogenannte Anticodon, welches komplementär zu dem Basentriplett des entsprechenden Codons der mRNA ist. Damit paaren sich die Basen des Anticodons der tRNA mit den Basen der mRNA, während die tRNA die A-, P- und E-Stellen des Ribosoms durchläuft (siehe Translation).

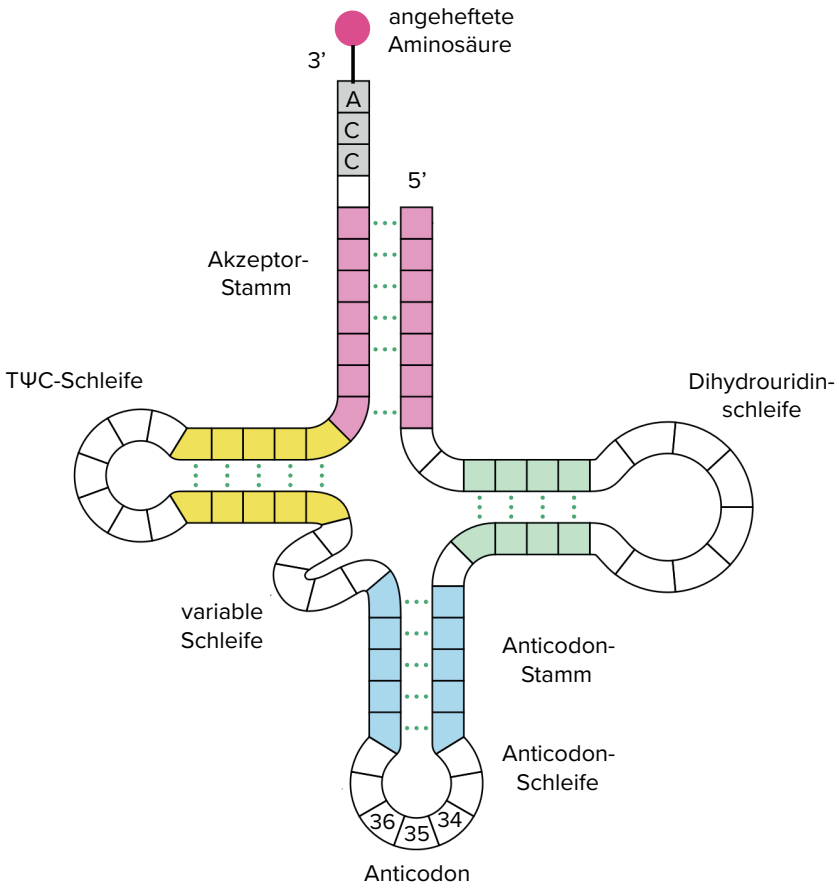


Abb. 1.19 Aufbau einer tRNA (nach Berg et al., 2018, S. 1063, adaptiert).

Am 3‘-Ende des Akzeptorstammes befindet sich die Sequenz CCA, welche bei allen tRNAs identisch ist. Das letzte Nucleosid dieser Sequenz – ein Adenosin – bindet an jene Aminosäure, die dem Anticodon der tRNA entspricht.

Um die tRNA mit ihrer spezifischen Aminosäure zu beladen, wird ein Enzym, die Aminoacyl-tRNA-Synthetase (aa-tRNA-Synthetase), mit der entsprechenden Aminosäure beladen. Danach wird an einer weiteren Bindungsstelle der Aminoacyl-tRNA-Synthetase eine unbeladene tRNA gebunden. Diese tRNA wird anhand ihres Anticodons, der Dihydrouridinschleife, oder anhand von speziellen Elementen wie beispielsweise der Base 73 im Akzeptorstamm erkannt (Wikipedia, tRNA; Gomez & Ibba, 2020). Die bereits an der Aminoacyl-tRNA-Synthetase gebundene

Aminosäure wird nun auf die tRNA übertragen und diese so beladene tRNA wird freigesetzt. Die Synthetase steht nun wieder für weitere Beladungen zur Verfügung (siehe Kapitel 7.4.1).

1.10.2 Proteinsynthese mit Hilfe von Ribosomen

Die Proteinsynthese an den Ribosomen beginnt in der Regel beim ersten AUG-Triplett im 5'-Bereich der mRNA. Dieses wird als Start-Codon für die Translation und als Beginn des offenen Leserahmens der von der mRNA codierten Aminosäureabfolge interpretiert. Beim *ersten Durchgang* des Ribosoms *(pioneer round)* entlang der mRNA synthetisiert dieses das Protein und entfernt dabei gleichzeitig *alle Exon-Junction-Komplexe* (siehe Abbildung 1.20). Bei jedem weiteren Durchgang eines Ribosoms enthält die mRNA keine EJCs mehr und ist quasi sauber (Schaaf & Zschocke, 2018, S. 19).

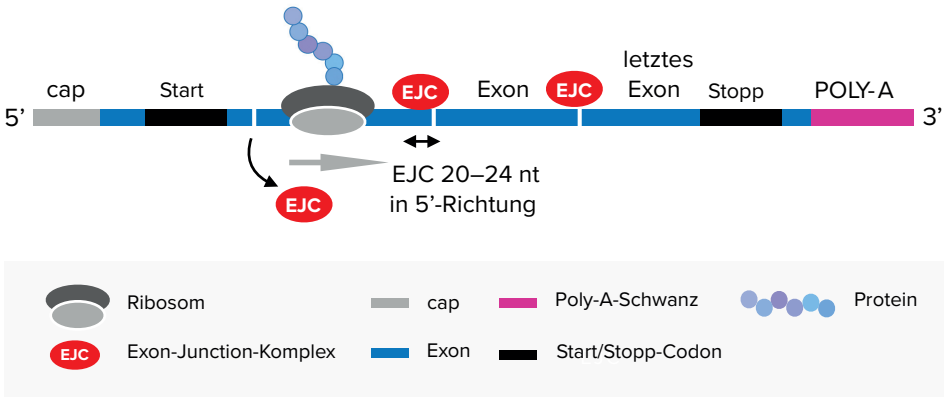


Abb. 1.20 Erster Durchgang des Ribosoms bei der Translation. Alle EJCs werden entfernt.

Erreicht das Ribosom das reguläre Stopp-Codon kurz vor dem 3'-Ende, so steht hierfür keine tRNA zur Verfügung. An der A-Stelle des Ribosoms lagern sich statt einer mit einer Aminosäure beladenen tRNA Freisetzungsfaktoren an, wodurch das Protein an der P-Stelle freigesetzt wird und das Ribosom in die kleine und die große Untereinheit dissoziiert (siehe Abbildungen 1.22i und 1.22j). Die mRNA kann nun mehrfach von weiteren Ribosomen abgelesen werden. Nach einer gewissen Zeit wird die mRNA abgebaut (Degradation).

Bakterien-RNAs haben eine mittlere Lebensdauer von ca. drei Minuten, Eukaryoten-RNAs bis zu einigen Stunden. Im 3'-UTR-Bereich ist festgelegt, wie groß die mittlere Lebensdauer der mRNA ist (Alberts et al., 2021, S. 266). Schlautmann und Gehring (2020) betonen in Ihrem Review die Wichtigkeit des EJC.

> **Specifically, the EJC's core components and its associated proteins regulate different steps of gene expression, including pre-mRNA splicing, mRNA export, translation, and nonsense-mediated mRNA decay (NMD).**

Abbildung 1.21 zeigt die Legende für die Abbildungen 1.22a–j.

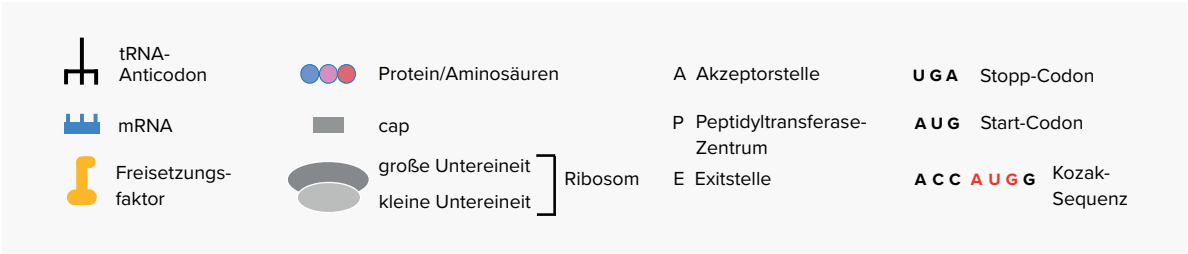


Abb. 1.21 Legende für die Abbildungen 1.22a–j.

Die Methionin-Aminoacyl-tRNA-Synthetase bindet die tRNA für die Aminosäure Methionin und auch jene für die Initiator-tRNA$^{met}_{i}$. Die beiden tRNAs unterscheiden sich jedoch und sie haben auch verschiedene Funktionen, obwohl die Synthetase die selbe ist. Die Initiator-tRNA$^{met}_{i}$ bindet als einzige direkt an die P-Stelle in der kleinen Untereinheit des noch nicht geschlossenen Ribosoms; die Methionyl-tRNA bindet – so wie alle anderen tRNAs – an die Akzeptorstelle des bereits geschlossenen Ribosoms. Wichtig für den Beginn der Translation ist ein *korrekter Startpunkt,* von dem aus die Ribosomen die Codons in Dreierschritten (Tripletts) ablesen können (siehe auch Kapitel 3).

Die Ribosomen besitzen drei Stellen, an denen die tRNA binden kann:

- die Akzeptorstelle (A),
- das Peptidyltransferasezentrum (P),
- die Exitstelle (E), (Heinrich et al., 2014, S. 606; Alberts et al., 2021, S. 276).

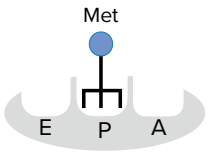


Abb. 1.22a Initiation

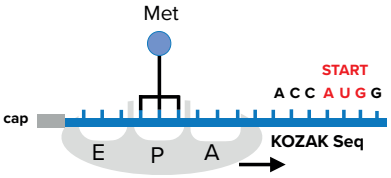


Abb. 1.22b Suche nach dem Startpunkt

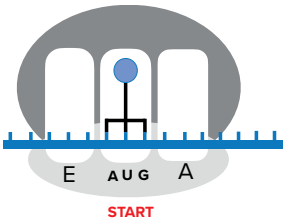


Abb. 1.22c Vervollständigen des Ribosoms

Die mit Methionin beladene Initiator-tRNA bindet gemeinsam mit einem Initiationskomplex an die P-Stelle der kleinen Untereinheit des Ribosoms (Abbildung 1.22a). Der Initiationskomplex wurde aus Gründen der Übersichtlichkeit nicht eingezeichnet.

Die kleine Untereinheit des Ribosoms bindet dann meist an die Kappe der mRNA. Im Allgemeinen wandert dieser Komplex entlang der mRNA in 3‘-Richtung, bis er auf die Kozak-Sequenz trifft. Kozak (1999) beschreibt die Basenfolge ACCAUGG als optimale Erkennungssequenz in Eu-

karyoten, in welche das Start-Codon AUG eingebettet ist (Abbildung 1.22b). Danach verbindet sich die große Untereinheit des Ribosoms mit der kleinen Untereinheit und bildet so das fertige Ribosom, wobei der Initiationskomplex entfernt wird (Abbildung 1.22c). Die Translation beginnt also in der Regel beim ersten AUG. Bei den Eukaryoten befinden sich zwischen der Kappe (cap) und dem Start-Codon noch kurze untranslatierte Sequenzen, die für Regulationen zuständig sind (Schaaf & Zschocke, 2018, S. 10). Das *Start-Codon AUG* ist sowohl bei Eukaryoten als auch bei Prokaryoten das Codon für die Aminosäure Methionin. Dieses bildet daher immer die erste Aminosäure der Proteinkette. Hier beginnt das Leseraster (Berg et al., 2018, S. 152).

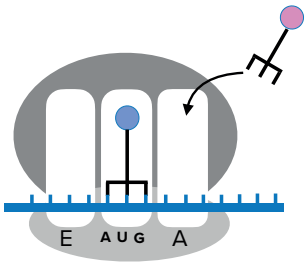


Abb. 1.22d Beginn der Elongation

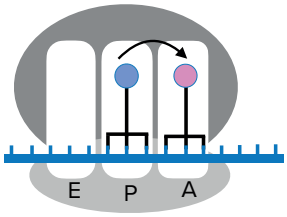


Abb. 1.22e Bindung einer beladenen tRNA an der A-Stelle

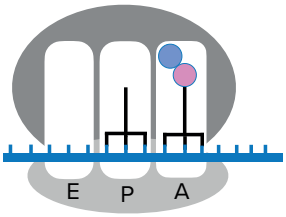


Abb. 1.22f Übertragung der Aminosäure mit Ausbildung einer Peptidbindung

Danach beginnt die Elongation (Abb. 1.22d), wobei eine tRNA, deren Anticodon komplementär zum nächstfolgenden Codon der mRNA ist, an der A-Stelle (Akzeptorstelle) des Ribosoms bindet (Abb. 1.22e). Die erste Aminosäure (Methionin), die mit der tRNA an der P-Stelle verbunden ist, wird unter Ausbildung einer Peptidbindung auf die nächste Aminosäure jener tRNA an der A-Stelle übertragen (Abbildung 1.22f). Die tRNA an der A-Stelle ist dann mit zwei Aminosäuren verbunden. Die Initiator-tRNA ist die einzige, die direkt an der P-Stelle bindet, alle weiteren tRNAs binden an die A-Stelle.

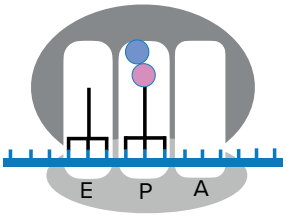


Abb. 1.22g Translokation des Ribosoms um ein Codon

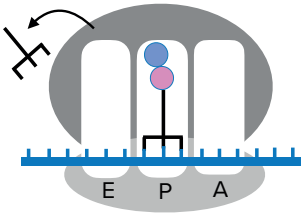


Abb. 1.22h Freisetzen der unbeladenen tRNA

Das Ribosom wandert nun um drei Stellen (ein Codon) in 3‘-Richtung weiter (Translokation), wodurch sich die nun nicht mehr beladene tRNA in der Exit-Stelle (E) befindet (Abb. 1.22g) und freigesetzt wird (Abb. 1.22h). An die nun freie A-Stelle (A) kann die nächste beladene tRNA binden und die ausgestoßene unbeladene tRNA wird weiterverwendet.

Bei der Elongation in Eukaryoten werden ca. drei bis fünf Aminosäuren pro Sekunde angefügt, bei Prokaryoten ca. 20 (Heinrich et al., 2014, S. 603). Dieser Zyklus wird so lange fortgesetzt, bis das Ribosom zu einem der drei Stopp-Codons (UAA, UAG, UGA) kommt. Da zu diesen keine tRNAs zur Verfügung stehen, bindet statt dessen ein Freisetzungsfaktor (Release-Faktor) an die A-Stelle (Abbildung 1.22i). Dadurch wird die Translation beendet (Termination) und das fertige Protein, die tRNA, die mRNA, der Freisetzungsfaktor und die beiden Untereinheiten des Ribosoms freigesetzt (Abbildung 1.22j). Nachdem die große und die kleine Untereinheit des Ribosoms sich voneinander getrennt haben, stehen diese für eine weitere Translation zur Verfügung (Berg et al., 2018, S. 152). Welcher Releasefaktor zum Einsatz kommt, hängt davon ab, welches Stopp-Codon vorhanden ist (Berg et al., 2018, S. 1080; siehe Kapitel 7.4.3).

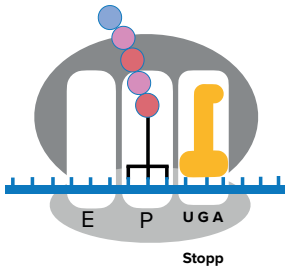


Abb. 1.22i Bindung eines Freisetzungsfaktors an die A-Stelle beim Erreichen eines Stopp-Codons.

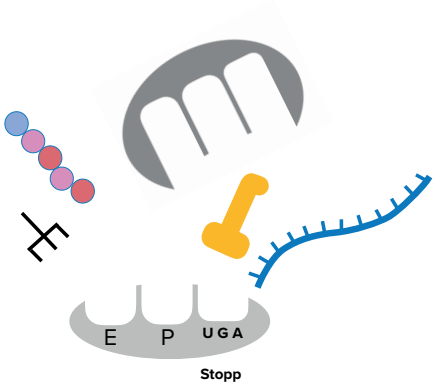


Abb. 1.22j Freisetzung des Proteins, der tRNA, des Freisetzungsfaktors, der mRNA und der beiden Untereinheiten des Ribosoms.

An einer mRNA befinden sich mehrere Ribosomen hintereinander, die je eine Aminosäurekette synthetisieren (siehe Abbildung 1.23). Dies wird als *Polysom* oder *Polyribosom* bezeichnet. Es können sich, je nach Länge der mRNA, bis zu 50 Ribosomen gleichzeitig auf einer mRNA befinden (Schaaf & Zschocke, 2018, S. 23).

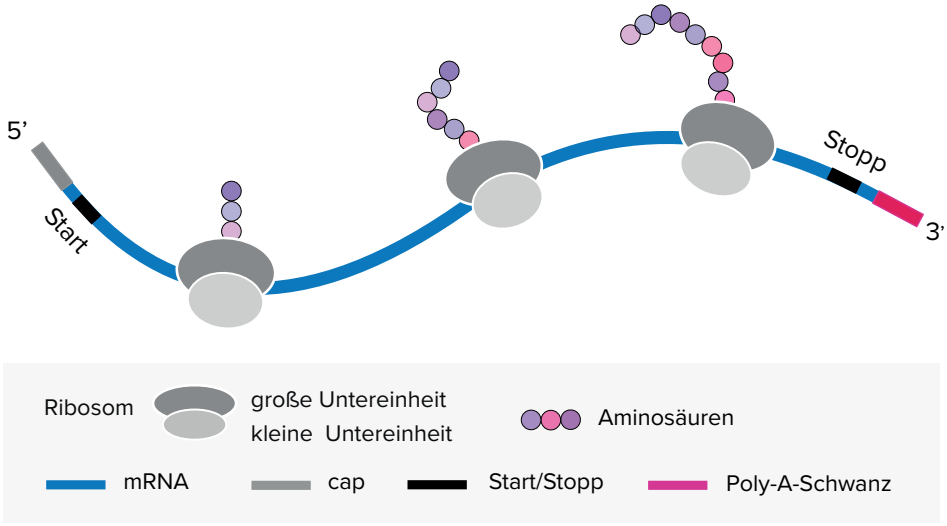


Abb. 1.23 Polysom mit mehreren Ribosomen auf der mRNA.

Ebenso wie die Transkription, hat auch die Translation drei Stufen:

- die Initiation,
- die Elongation,
- die Termination.

Da Prokaryoten keinen Zellkern haben, finden die Transkription – Umschreiben der DNA in RNA – und die Translation – Synthese von Proteinen – beide im Cytoplasma statt. Die beiden Vorgänge sind nicht – wie bei den Eukaryoten – durch Membranen getrennt und können daher sofort hintereinander stattfinden, ohne dass der erste Schritt, die Transkription, schon beendet sein muss. Bei den Prokaryoten findet keine Prozessierung der prä-mRNA statt, das heißt es werden keine Kappen und keine Poly-A Schwänze an die prä-mRNA angehängt. Zudem haben Prokaryoten keine Introns, die erst nach vollständiger Synthese der prä-mRNA herausgeschnitten werden müssten. Ein weiterer Unterschied bei der Translation bei Prokaryoten besteht darin, dass beim Start der Translation ein Formyl-Methionin verwendet wird. Prokaryoten können außerdem mehrere Gene, die hintereinander in einem Operon organisiert sind, gemeinsam transkribieren und translatieren.

2 Codierung

Hier soll der Frage nachgegangen werden, wie die genetische Information codiert ist und ob alle Lebewesen dieselbe oder eine sehr ähnliche Codierung aufweisen. Es wird weiters untersucht, welche unterschiedlichen Codierungen und Speichermöglichkeiten für genetische Informationen – zumindest theoretisch – denkbar sind und ob sie sinnvoll oder weniger sinnvoll erscheinen, auch wenn es – derzeit – nicht nachweisbar ist, ob solche alternativen Codierungen jemals existiert haben, oder auf anderen habitablen Planeten existieren.

2.1 Codierung versus Verschlüsselung

Einleitend zu diesem Kapitel wird der Unterschied zwischen *Codierung* und *Verschlüsselung* dargelegt, zumal diese beiden Ausdrücke leider in mehreren klassischen Lehrbüchern der Biochemie und Molekularbiologie nicht unterschieden werden bzw. abwechselnd oder sogar synonym verwendet werden.

Ursprünglich wurden Informationen ausschließlich mündlich weitergegeben, beispielsweise die Information, wie aus dem Lauf der Sterne festzustellen sei, wann das Getreide ausgesät werden muss. Diese Information war *allen Mitgliedern einer Gruppe* bekannt. Sie musste sorgfältig gehütet und zuverlässig von Generation zu Generation weitergegeben werden. Im Zuge des Handels wurde es jedoch nötig, festzuhalten, wieviel gekauft oder verkauft wurde. Es entwickelten sich einfache Zählmethoden (Codierungen), wie beispielsweise die Verwendung von Tonkügelchen mit eingeritzten Symbolen, als Beweis für die jeweilige Menge abgelieferten Getreides. (Maier, 2020, ARD online).

Das Festhalten verschiedener Ereignisse auf Stein oder Ton in Form einer Schrift oder von Symbolen, war ein wichtiger Schritt. Es war nicht mehr nötig, dass alle Menschen die gesamte Information hatten, sondern es genügte, zu wissen, wie die Information codiert war. Diese In-

formationen durften *nicht geheim* sein, sondern musste *offen* und für jeden lesbar sein. Jedermann sollte den Code kennen. Man musste nicht mehr die gesamte Information von Mensch zu Mensch weitergeben, sondern es genügte, die Codierung weiterzugeben. Die *Decodierung* entspricht dem Lesen und Verstehen dieser Texte.

Wichtig ist hier, dass es *verschiedene Codierungssyteme* für ein und dieselbe Information geben kann, dass diese Systeme aber alle offen zugänglich und ineinander umwandelbar sind, damit möglichst viele Menschen Zugang zu dieser Information haben können.

Beispiele für verschiedene Codierungssysteme sind verschiedene Schriftzeichen für den selben Buchstaben oder Pictogramme statt Schriftzeichen. Als einfache Beispiele seien folgende genannt:

S (latein), σ (griechisch), س (arabisch)
Schrift *Ausgang* versus Pictogramm

Es gibt mehrere Beispiele in der Geschichte, wo das Schriftzeichensystem in einem Land mehrfach geändert wurde. So wurde beispielsweise in der Türkei unter Atatürk, die arabische Schrift gegen lateinische Schriftzeichen ausgetauscht, ohne dass sich die Sprache oder der Inhalt der Information geändert hätte, oder in Aserbaidschan, wo einander arabische, lateinische, kyrillische und ab 1991 wieder lateinische Schriftzeichen abgewechselt haben (Wikipedia: Aserbaidschanische Sprache, Mustafa Kemal Atatürk).

Weitere bekannte Beispiele für eine *Codierung* sind alle Schriften, der ASCII-Code (American Standard Code für Information Interchange) oder auch der Binärcode, der in Computersystemen Verwendung findet und der QR-Code (Quick-response-Code), um nur einige zu nennen. Abbildung 2.1 zeigt, wie eine Nachricht mittels einer Codierung übermittelt werden kann.

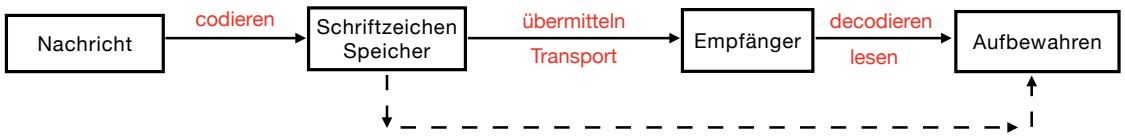


Abb. 2.1 Übermittlung von Informationen (Adaptiert nach Shannon & Weaver, 1976, S. 17).

Ein Code besteht im einfachsten Fall aus einer Tabelle, die Zeichen anderen Zeichen zuordnet. So kann man beispielsweise den Code der DNA mittels einer Tabelle in den Code der RNA umcodieren, wie nachstehende Tabelle zeigt. In der Zelle geschieht dies bei der Transkription.

DNA	→	RNA	DNA	→	RNA
T (Thymin)		U (Uracil)	G (Guanin)		G (Guanin)
A (Adenin)		A (Adenin)	C (Cytosin)		C (Cytosin)

Eine *Verschlüsselung* wird im Gegensatz zur Codierung immer dann verwendet, wenn eine Information *geheim* bleiben sollte/musste und daher nicht für jedermann zugänglich sein sollte, beispielsweise im Krieg. Zur Verschlüsselung von Nachrichten (Kryptographie) wird ein Schlüssel verwendet, der nur einer einzigen Person oder einigen wenigen bekannt ist, und der unter allen Umständen *geheim* bleiben muss. Eine Verschlüsselung kann mittels eines Buches oder mittels eines mehr oder weniger komplizierten Algorithmus erfolgen.

Allgemein gilt:

Codierung	≒	*Decodierung*	*(mittels öffentlicher Codetabelle)*
Verschlüsselung	≒	*Entschlüsselung*	*(mittels geheimen Schlüssels)*

Die Information in der DNA bzw. RNA stellt eine *Codierung* und keine Verschlüsselung dar, denn *jedes Ribosom – egal welcher Spezies* – sollte den DNA- bzw. RNA-Code *verstehen und decodieren* können – sonst könnten genetische Informationen von Viren oder Bakterien im Zuge einer Infektion nicht von Ribosomen anderer Organismen abgelesen werden. Eine Verschlüsselung würde in diesem Fall keinen Sinn ergeben. Eine Codierung darf nicht nur deshalb als Verschlüsselung bezeichnet werden, weil wir sie – noch – nicht verstehen. Eine Verschlüsselung darf nicht nur deshalb als Codierung bezeichnet werden, weil der gewählte Schlüssel zu primitiv ist und leicht geknackt werden kann.

2.2 Kombinatorische Möglichkeiten der Codierung

2.2.1 Annahmen

- Es gibt eine begrenzte Anzahl an unterschiedlichen Symbolen/Ziffern/Zeichen, die hier der Einfachheit halber als *Basen* bezeichnet werden, auch wenn sie chemisch anders aussehen könnten, als die bekannten Basen.
- Es gibt eine unterschiedliche Anzahl von Stellen, in welcher die Zeichen jeweils – zumindest virtuell – zu Zeichengruppen (Codons) bzw. zu einzelnen Informationspaketen oder Informationseinheiten zusammengefasst werden.
- Zur Umsetzung der einzelnen Codons (Informationspakete) in Aminosäureketten (Proteine bzw. Peptide) wird zusätzlich eine Translationsvorrichtung (Übersetzungsvorrichtung – analog zu den Ribosomen) benötigt.
- Jedes Codon steht für genau *eine* Aminosäure, in die es übersetzt wird.
- Die (Un-) Möglichkeit der genauen Rückübersetzung wird hierbei außer Acht gelassen
- Die Bögen in den Abbildungen stellen das Leseraster bzw. die Codonlänge dar und werden zur besseren Lesbarkeit eingezeichnet. Sie stellen *kein Abstandszeichen* dar.
- Durch die Länge der Peptide/Proteine können sich Codons bzw. Aminosäuren wiederholen.

2.2.2 Kombinatorische Überlegungen für Codons

Es wird von der Überlegung ausgegangen, dass die Codierung der genetischen Information (in der DNA bzw. RNA) durch einzelne Zeichen bzw. Zeichengruppen in quasi »aufgefädelter« Reihenfolge vorliegt. Dies stellt eine Art *physischen Hauptspeicher* dar. Dass die DNA in Form einer *doppelt* gewundenen Struktur als Doppelhelix vorliegt, wird bei den Überlegungen nicht zwingend vorausgesetzt. Es wird angenommen, dass sowohl Einfachstränge als auch Doppelstränge vorliegen können. Es werden hier die DNA und die RNA gleichgesetzt, da die Unterschiede gering sind und sie sich sehr einfach umcodieren – transkribieren – lassen (Ersetzen von Thymin (T) durch Uracil (U) oder umgekehrt).

2.2.3 Unärer Code (eine Base)

Ein *Unärsystem* mit nur einem einzigen Zeichen, könnte zwar theoretisch in der Biologie verwendet werden, dafür müssten aber die verschiedenen Aminosäuren mit unterschiedlich langen Codons codiert werden. Die maximale Länge entspräche der maximalen Anzahl zu codierender Aminosäuren eventuell mit je einem zusätzlichen Start- und Stopp-Zeichen. Es gibt hierbei nur ein einziges Zeichen (Base), das unterschiedlich oft pro Codon vorkommt. Diese Notation kommt als so genannte *Bierdeckelnotation* vor, bei der jedem konsumierten Glas Bier ein Strich zugeordnet wird. Diese einfache Notation wird manchmal auch in Gefängnissen oder beim Militär verwendet, um Tage bis zur Entlassung zu zählen. Eine solche Notation wurde auch beim Kerbholz (Kerbstock) verwendet, um Schulden zu notieren. Jede Kerbe im Holz bedeutet eine bestimmte Summe. (wikipedia: Kerbholz, Strichliste, Unärsystem), (Weule, 1915).

Oft wurde das Kerbholz der Länge nach gespalten und sowohl der Schuldner, als auch der Gläubiger bekamen je eine Hälfte, um das Ganze fälschungssicher zu machen. (Kraus, 2021, S. 23). Dies erinnert an die doppelsträngige DNA, die es erlaubt, Fehler zu erkennen und zu reparieren, da eine Sicherheitskopie vorhanden ist.

Das große Problem bei dieser Art der Notation ist, dass die Codons eine unterschiedliche Länge aufweisen. Sollen nun mehrere Dinge oder Vorgänge, wie beispielsweise Schulden bei verschiedenen Händlern, codiert werden, so musste entweder jedesmal ein eigenes Kerbholz verwendet werden, oder es müssten auf ein und demselben Kerbholz Trennzeichen markiert werden, um die verschiedenen Vorgänge auseinanderhalten zu können.

Bei der DNA/RNA, bei der die Codons hintereinander aufgefädelt sind, kann im Falle eines Unärcodes nicht erkannt werden, wo ein Codon zu Ende ist und wo ein Neues anfängt. Jeder Aminosäure würde ein Codon mit einer *unterschiedlichen Anzahl* an identischen Basen zugeordnet. Diese Art der Codierung erlaubt aber nur dann eine *eindeutige* Zuordnung, wenn Abstandsmarkierungen – das bedeutet Endemarkierungen nach jedem Codon – vorhanden sind.

Derartige Abstandsmarkierungen werden auch als Komma bezeichnet. Abbildung 2.2 zeigt eine solche Codierung mit gedachter Abstandsmarkierung.

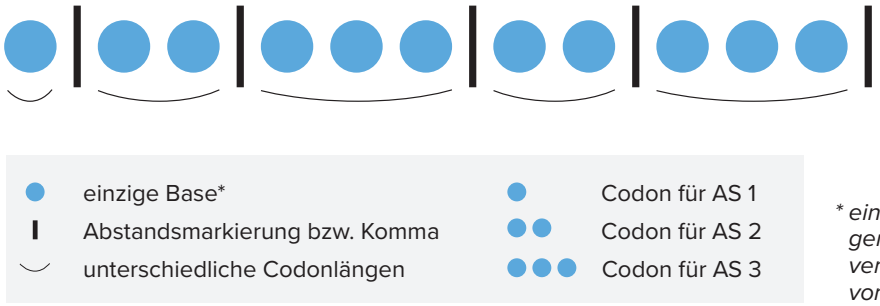


Abb. 2.2 Einfachstrang mit Unärcode und Abstandsmarkierung/Komma, mit unterschiedlicher Codonlänge.

Das Stopp-Codon würde das *Ende der gesamten Aminosäurekette* markieren, im Gegensatz zur Abstandsmarkierung bzw. einem Komma, das nur das Ende eines Codons für eine bestimmte Aminosäure darstellt. Genau genommen handelt es sich um einen Code mit zwei Zeichen, die allerdings unterschiedliche Funktion haben, ein Zeichen für die Codierung, ein Zeichen als Codonende bzw. Abstand oder Komma.

Für n Aminosäuren benötigt man beim Unärcode n Codons mit Längen von 1 bis n. Der Vorteil eines solchen Codes besteht darin, dass der Code *beliebig verlängert* werden kann, das heißt, es gibt keine Obergrenze für die Anzahl möglicher Codons bzw. für die Anzahl codierbarer Aminosäuren. Für die 20 proteinogenen Aminosäuren benötigt man daher Codonlängen von 1 bis 20.

2.2.4 Dualer Code (2 Basen)

Beim dualen Code stehen zwei unterschiedliche Basen zur Verfügung. Hat der Code eine Länge von einem Zeichen, so können nur zwei unterschiedliche Aminosäuren codiert werden, hat der Code eine Länge von beispielsweise 2 Zeichen, so können maximal $2^2 = 4$ verschiedene Aminosäuren codiert werden. Ab zwei verschieden Basen ist es möglich, mit einem *fixen Leseraster* zu arbeiten. Je nach Länge eines Codons – die Länge entspricht der Weite des Leserasters – können unterschiedlich viele Aminosäuren codiert werden. Es können im Prinzip beliebig viele Aminosäuren codiert werden, jedoch muss die Mindest-Codonlänge an die Zahl der zu codierenden Aminosäuren angepasst werden. Die Maximalzahl möglicher Codons beträgt B^n, wobei B die Anzahl der zur Verfügung stehender Basen ist und n die Anzahl der Stellen (Codonlänge).

2.2.4.1 Einfachstrang, dualer Code, Codonlänge 1

Abbildung 2.3 zeigt einen Einfachstrang eines dualen Codes mit der Codonlänge 1. Es können hierbei mit zwei Basen nur zwei Aminosäuren codiert werden, was für leistungsfähige Proteine zu wenig sein dürfte.

Abb. 2.3 Dualer Code mit 2 Basen. Die Codonlänge beträgt 1, daher entspricht jede Base einer Aminosäure.

2.2.4.2 Einfachstrang, dualer Code, Codonlänge 2

Mit zwei Basen kann man jedoch die Codonlänge erhöhen, wobei die Codonlänge fix ist, sodass bei einer Länge von zwei Basen mit Codonlänge 2 bereits die Codierung von vier Aminosäuren möglich ist. Abbildung 2.4 zeigt einen solchen Einfachstrang mit zwei Basen, also einem dualen Code. Hier beträgt die Codonlänge 2.

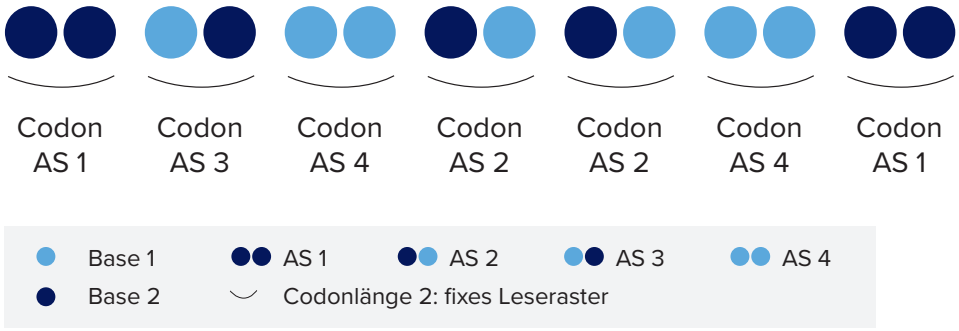


Abb. 2.4 Dualer Code mit 2 Basen und fixer Codonlänge 2.

Stehen *mindestens zwei* unterschiedliche Basen zur Verfügung, so ist eine konstante Codonlänge mit *fixen Leseraster* möglich.

Die *minimal* erforderliche *Codonlänge ist* davon abhängig, wieviele Aminosäuren mindestens codiert werden müssen. Darüber hinaus sind keine Abstandsmarkierungen (Codon-Ende-Markierungen) mehr nötig. Der Code kann daher *kommafrei* sein.

2.2.4.3 Einfachstrang, dualer Code, Codonlänge 3

Geht man nun zu größeren Codonlängen über, so zeigt sich, dass die Anzahl der möglichen Codons und damit der codierbaren Aminosäuren schnell ansteigt. Abbildung 2.5 zeigt einen Einfachstrang mit zwei verschiedenen Basen und einer Codonlänge von drei. Es können hiermit bereits 2^3=8 verschiedene Aminosäuren codiert werden. Die Darstellungen von Einzelsträngen mit zwei Basen und größeren Codonlängen sowie fixem Leseraster, können sinngemäß fortgeführt werden.

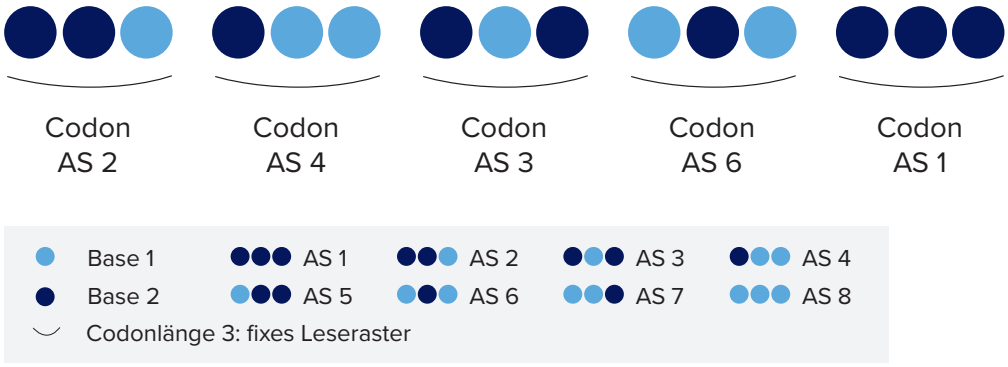
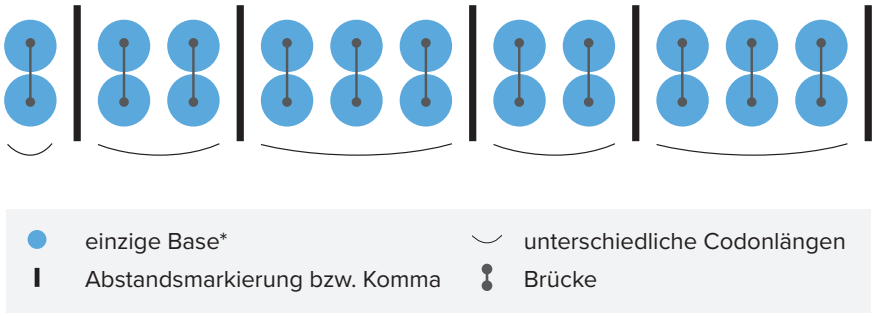


Abb. 2.5 Dualer Code mit 2 Basen und fixer Codonlänge 3.

2.2.5 Doppelstränge

Wie erwähnt erscheint die Verwendung von Doppelsträngen nicht zwangsläufig notwendig, jedoch sind Doppelstränge durch Brücken verbunden und daher meist stabiler. Darüber hinaus erhöhen Doppelstränge die Sicherheit, denn in einem Einzelstrang kann eine Mutation nicht erkannt werden. In einem Doppelstrang kann eine Mutation in einem der beiden Stränge nicht nur erkannt sondern auch repariert werden. Hier soll dargestellt werden, welche unterschiedlichen theoretischen Möglichkeiten sich bei *zusätzlicher* Einführung von *Doppelsträngen* ergeben.

2.2.5.1 Doppelstrang beim Quasi-Unärcode



** eine einzige, allen Codons gemeinsame Base, die in verschiedenen Codons mehrfach vorkommen kann*

Abb. 2.6 Doppelstrang mit identischen Strängen, die nicht komplementär sind, mit Quasi-Unärcode und unterschiedlicher Codonlänge.

Um einen Doppelstrang zu erhalten, müssen beim unären Code, bei dem die einzelnen Basen identisch sind, diese jeweils *mit sich selbst* Brücken ausbilden, wie beispielsweise Wasserstoff-Brücken (H-Brücken). Alle Brücken müssen daher identisch sein.

Abb. 2.6 zeigt die Ausbildung eines solchen theoretischen Doppelstranges mit nur einer Base, mit zwei identischen Strängen und mit gedachten Abstandsmarkierungen bzw. Kommas. Die Brücken sowie die ganzen Stränge sind jeweils identisch. Ob dadurch eine Fehlpaarung wahrscheinlicher wird als bei verschiedenen komplementären Basen, kann nicht abgeschätzt werden.

Da die beiden Ketten identisch wären, gäbe es auch keinen Unterschied zwischen codogenem und nicht-codogenem Strang. Beim codogenen bzw. nicht-codogenen Strang handelt es sich um die beiden komplementären Stränge der DNA-Doppelhelix, die in unterschiedlichen Richtungen abgelesen werden. Unärcode ist zwar ebenso, wie bei allen anderen Codes, ein solcher Doppelstrang denkbar, doch tritt auch hier, ebenso wie beim Einfachstrang, das Problem der unterschiedlichen Codonlängen auf.

2.2.5.2 Doppelstrang beim Dualcode

Bei der Einführung von Doppelsträngen mit zwei Basen sind zwei Szenarien denkbar:

Szenario 1: Jede Base bindet mit sich selbst über Wasserstoffbrücken, das bedeutet, die beiden Stränge sind, wie beim Unärcode, identisch. Abbildung 2.7 zeigt einen Doppelstrang mit zwei Basen und Codonlänge zwei. Da hier Base 1 mit Base 1 bzw. Base 2 mit Base 2 paart, muss es zwei unterschiedliche Arten der Brückenbildungen geben. Es kann sich in beiden Fällen um einfache oder mehrfache Wasserstoffbrücken handeln, die die beiden Stränge verbinden.

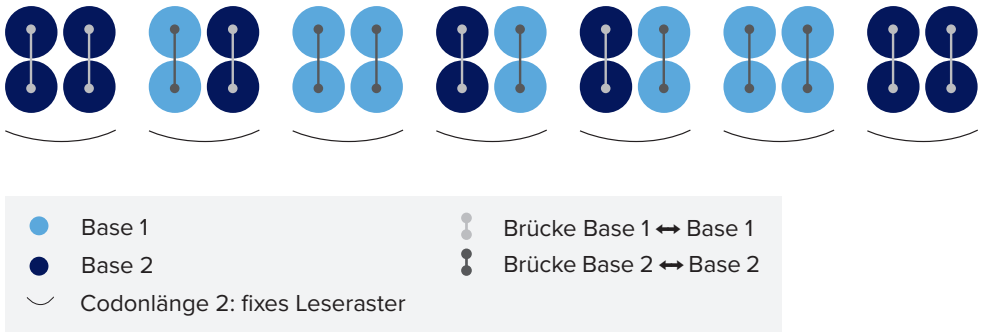


Abb. 2.7 Doppelstrang mit zwei identischen Strängen, die nicht komplementär sind, mit 2 Basen und fixer Codonlänge 2.

Bei Doppelsträngen bilden sich immer Brücken aus, die beispielsweise Wasserstoffbrücken sein können. Je nach Art der beteiligten Basen können die Art und/oder Anzahl der Brücken zwischen den identischen Basenpaaren unterschiedlich sein.

Szenario 2. Jede Base bindet mit der anderen Base über eine oder mehrere Wasserstoffbrücken. Dies bedeutet, dass die beiden Stränge zueinander *komplementär* sind. Abstandsmarkierungen/Kommas sind auch hier – wegen des fixen Leserasters – nicht nötig.

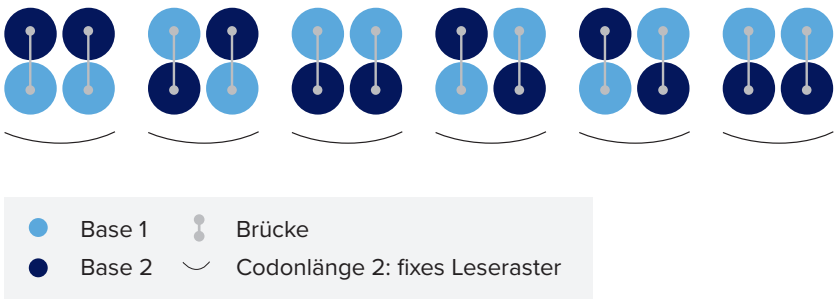


Abb. 2.8 Doppelstrang mit einem dualen Code (2 Basen) mit fixer Codonlänge von 2. Komplementäre Basen paaren miteinander.

Hier paart immer Base 1 mit Base 2, sodass die Brücken identisch sind. Da ab zwei verschiedenen Basen *konstante Codonlängen* bzw. ein fixes Leseraster verwendet werden kann, entfällt die Notwendigkeit für flexible Translationsvorrichtungen wie Ribosomen, die unterschiedliche Codonlängen verarbeiten können.

2.2.5.3 Doppelstrang beim dreibasigen Code

Bei einem Code mit drei Basen können sich *keine fixen Basenpaare* zusammenschließen, denn dazu müsste die Anzahl der Basen *geradzahlig* sein. Bei einem dreibasigen Code gibt es wieder zwei Möglichkeiten. Einerseits können identische Basen miteinander paaren, was zwei identische Stränge ergibt, andererseits gibt es die theoretische Möglichkeit einer *gemischten Kombination*. Für einen Anteil (geradzahlige Basenzahl) erfolgt eine komplementäre Basenpaarung, für die übrig gebliebene Base – es handelt sich ja um eine ungeradzahlige Anzahl von Basen – erfolgt eine Paarung mit sich selbst (siehe Abb. 2.9).

Diese Betrachtungen gelten analog für alle Codes mit einer ungeraden Anzahl von Basen.

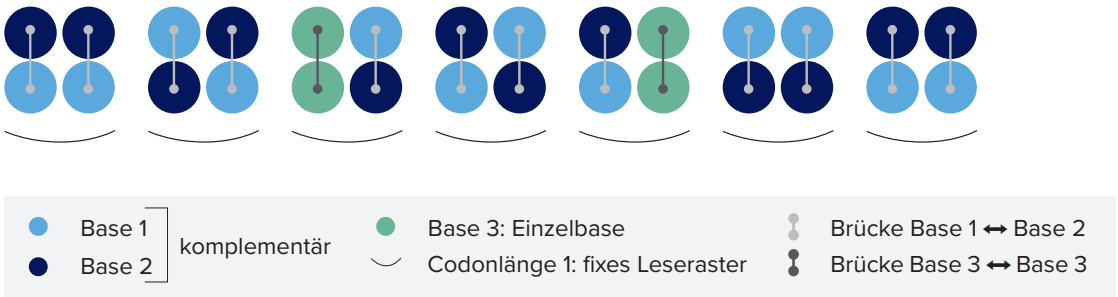


Abb. 2.9 Gemischte Kombination von komplementären und nicht komplementären Basen im Doppelstrang bei ungeradzahliger Basenzahl.

2.3 Anzahl der Codierungsmöglichkeiten

Eine Zahl mit der Basis B (Anzahl der Basen) mit n Stellen (Länge des Codons) kann maximal den Wert B^n annehmen (Anzahl der codierbaren Möglichkeiten). Es sind auch bei mehr als einer Base Codons mit *unterschiedlicher Länge* denkbar. In diesem Fall wäre wieder ein Komma notwendig oder es müsste sich um einen Code wie den Huffman-Code handeln, bei dem keine Kommas notwendig sind und bei dem es dennoch nicht zu Verwechslungen kommen kann (Ziegenbalg et al., 2016, S. 281–287, Huffman-Kodierung, Wikipedia).

Aus den obigen Betrachtungen ergeben sich theoretisch folgende kombinatorische Möglichkeiten für Codons mit fixer Länge, die also ein fixes Leseraster aufweisen können. Dies ist, wie ausgeführt, erst ab mindestens zwei verschiedenen Basen möglich.

2.3.1 Anzahl der Codierungsmöglichkeiten beim dualen Code

a) 2 Basen 1 Stelle pro Codon → 2^1 = 2 Möglichkeiten
b) 2 Basen 2 Stellen pro Codon → 2^2 = 4 Möglichkeiten
c) 2 Basen 3 Stellen pro Codon → 2^3 = 8 Möglichkeiten
d) 2 Basen 4 Stellen pro Codon → 2^4 = 16 Möglichkeiten
e) 2 Basen 5 Stellen pro Codon → 2^5 = *32 Möglichkeiten*
f) 2 Basen 6 Stellen pro Codon → 2^6 = 64 Möglichkeiten

Da es heute 20 proteinogene Aminosäuren gibt, wäre ein zweibasiger fünfstelliger Code ausreichend.

Zu a) 2 Basen – 1 Stelle: Kommas sind nicht nötig, da hier ein *fixes Leseraster* mit Abstand 1 eingehalten wird. Die Anzahl möglicher Kombinationen/ Aminosäuren ist mit 2 aber für heutiges Leben viel zu gering.

Zu b) 2 Basen – 2 Stellen: Vier mögliche Codons für vier Aminosäuren erscheint auch hier gering, obwohl es Theorien gibt, wonach es ursprünglich nur 4 Aminosäuren gegeben haben soll. (Francis, 2013).

Zu c) 2 Basen – 3 Stellen: Dies würde einem einfachen Triplettcode mit 8 Möglichkeiten entsprechen. Gemäß der Annahme, dass es ursprünglich nur 4 oder 6 Aminosäuren gegeben habe, wäre diese Möglichkeit durchaus plausibel. Der Triplettcode könnte im Zuge der Evolution um zwei weitere Basen erweitert worden sein, sodass sich die 20 proteinogenen Aminosäuren codieren lassen, ohne aber die Ribosomen als Translationsvorrichtungen wesentlich ändern zu müssen.

2.3.2 Anzahl der Codierungsmöglichkeiten beim ternären Code

Die Überlegungen werden analog weitergeführt:

a) 3 Basen 1 Stelle pro Codon → 3^1 = 3 Möglichkeiten
b) 3 Basen 2 Stellen pro Codon → 3^2 = 9 Möglichkeiten
c) 3 Basen 3 Stellen pro Codon → 3^3 = *27 Möglichkeiten*
d) 3 Basen 4 Stellen pro Codon → 3^4 = 81 Möglichkeiten

Ein dreibasiger dreistelliger Code wäre zur Codierung von den 20 kanonischen Aminosäuren ausreichend. Bei einem solchen Code können wieder nur jeweils zwei identische Basen zwischen den beiden Ketten miteinander paaren, die dritte Base müsste mit sich selbst paaren (ungeradzahlige Anzahl von Basen), sofern es sich um einen Doppelstrang handelt.

Ein vierstelliger dreibasiger Code hätte sehr viele überzählige Codons, sofern wir von 20 zu codierenden Aminosäuren ausgehen. Dieser Code hätte eine sehr hohe Redundanz, sofern alle überzähligen Codons belegt wären.

2.3.2.1 Morsecode

Einen bekannten *Code* stellt das *Morsealphabet* dar. Dieser Code wurde erstmals 1844 von dem Amerikaner Samuel Morse in seinem Telegrafen eingesetzt, um Zeichen elektrisch übertragen zu können. (wikipedia, Morsecode) Der Morsecode enthält, je nach Interpretation, mindestens drei Zeichen:

- langer Ton (—)
- kurzer Ton (•)
- stumm (_)

Der Morsecode ist ein Code mit *variabler Länge* und die einzelnen Zeichen treten mit unterschiedlicher Häufigkeit auf. Diese unterschiedlichen Häufigkeiten sind typisch für natürliche Sprache. So ist beispielsweise *e* der häufigste Buchstabe im Deutschen, *n* der zweithäufigste und *q* der seltenste Buchstabe. Soll bei der Übertragung Platz und Zeit gespart werden, so bietet sich ein Code an, der die Buchstaben der Häufigkeit nach ordnet und für den häufigsten Buchstaben den kürzesten Code auswählt. Ein solcher Code ist beispielsweise der genannte Morsecode, der einem *e* einen Punkt (kurzen Ton) • und einem seltenen Buchstaben wie beispielsweise dem *q* einen langen Code mit vier Zeichen zuordnet (– – • –). Die Häufigkeiten der Buchstaben des Morsecodes bezogen sich damals auf die englische Sprache.

Der Morsecode wäre dann nicht eindeutig, wenn die Zeichenfolge ohne Abstand übertragen würde. Aus der Zeichenfolge •• kann man nicht entnehmen, ob es sich zweimal um den Buch-

staben *e* oder einmal den Buchstaben *i* handelt. Es sind daher spezielle Buchstaben-Ende-Codes (Kommas) nötig, die anzeigen, wann ein Buchstabe zu Ende ist. Der Vorteil eines solchen Kommacodes ist, dass bei einer fehlerhaften Übertragung eines Buchstabens oder auch des Kommas nur ein oder maximal zwei Buchstaben fehlerhaft interpretiert werden. Danach synchronisiert sich der Code wieder, sofern in kurzem Abstand vom ersten Fehler kein weiterer auftritt. Für natürliche Sprachen ist dies von Vorteil, da ein Text auch mit einigen wenigen Fehlern verstanden werden kann, sofern es sich nicht um Zahlen handelt. Ein solcher Code wäre in der Molekularbiologie, wie bereits dargelegt, möglicherweise problematisch, da er sehr unterschiedliche Codonlängen enthält.

Das Stummzeichen ist beim Morsecode wegen der variablen Codonlängen notwendig und es ist immer das *letzte Zeichen* eines jeden Morse-Buchstabens (Morse-Codons). Dies entspricht der Abstandsmarkierung bzw. einem Komma. Es handelt sich also um einen *Kommacode*. (Mittelbach, 1992, S. 204).

Ein Buchstabe der menschlichen Sprache, die in Morsecode übersetzt werden soll, hat mindestens ein Zeichen (kurzer Ton und ein Komma oder langer Ton und ein Komma), maximal aber 6 Zeichen. (Die Ziffern 1 bis 10 werden durch 5 Zeichen und ein Komma codiert.) Das Ende eines Wortes müsste wegen der hohen Redundanz der Sprache nicht unbedingt dargestellt werden, das heißt das Wort könnte genauso beendet werden wie ein Buchstabe. Dies entspricht sehr alten lateinischen und griechischen Inschriften, bei denen ursprünglich kein Abstand zwischen den Wörtern gemacht wurde. (wikipedia: scriptura continua). Die Unterscheidung zwischen Buchstabenende (kurze Pause) und Wortende (lange Pause) wird jedoch beim Morsen eingehalten. Mitunter wird der Morsecode auch als *binärer* Code bezeichnet (die Pause wird nicht als Zeichen angesehen) oder als *quartärer* Code, wobei zwei unterschiedliche Pausenlängen (stumm) als Buchstabenende (kurze Pause) und als Wortende (lange Pause) angesehen werden. Aufgrund der Redundanz der Sprache wäre die Unterscheidung der Pausenlängen jedoch nicht zwingend nötig.

2.3.3 Anzahl der Codierungsmöglichkeiten beim quartären Code

Dieser Code hat wieder eine gerade Anzahl von Basen, sodass bei einem Doppelstrang jeweils zwei gleiche oder zwei unterschiedliche d.h. komplementäre Basen miteinander Paare über Wasserstoffbrücken bilden können.

a) 4 Basen 1 Stelle pro Codon → 4^1 = 4 Möglichkeiten
b) 4 Basen 2 Stellen pro Codon → 4^2 = 16 Möglichkeiten
c) 4 Basen 3 Stellen pro Codon → 4^3 = *64 (entspricht unserem bekannten Code)*
d) 4 Basen 4 Stellen pro Codon → 4^4 = 256 Möglichkeiten

Zu a) 4 Basen – 1 Stelle: Diese Anzahl ist sehr gering, da die heutige Anzahl von 20 Aminosäuren nicht abgedeckt wäre.

Zu b) 4 – Basen – 2 Stellen: Diese Möglichkeit deckt ebenfalls unsere heutige Anzahl von Aminosäuren nicht ab.

Zu c) 4 Basen – 3 Stellen: Hierbei handelt es sich um den heute bekannten und von allen Lebensformen verwendeten Triplettcode (drei Stellen).

Zu d) 4 Basen – 4 Stellen: Diese Anzahl von 256 Möglichkeiten ist sehr groß. Da nur 20 Codons nötig sind, wären sehr viele Codons entweder nicht belegt oder mehrfach – ca. zehnfach – zu belegen.

2.3.4 Anzahl der Codierungsmöglichkeiten beim fünfbasigen Code

a) 5 Basen 1 Stelle → 5^1 = 5 Möglichkeiten
b) 5 Basen 2 Stellen → *5^2 = 25 Möglichkeiten*
c) 5 Basen 3 Stellen → 5^3 = 125 Möglichkeiten

Ein fünfbasiger zweistelliger Code würde ebenfalls ausreichen um unsere heute bekannten 20 kanonischen Aminosäuren zu codieren.

Allgemein gilt: Bei allen Codons mit einer Länge ≥ 2 (Stellenzahl/Codonlänge) oder mit einer Basenzahl ≥ 2 kann mit *fixen Codonlängen* gearbeitet werden, sodass keine Abstandsmarkierungen/Kommas mehr nötig sind und mit einem *fixem Leseraster* abgelesen werden kann. Ein solcher Code mit fixem Leseraster wird als *kommafrei* bezeichnet.

Beim genetischen Code gibt es kein Stellenwertsystem und es müssen keine Rechenoperationen wie Addition, Subtraktion, Multiplikation oder Division durchgeführt werden. Obwohl die Position der einzelnen Basen innerhalb eines Codons von Bedeutung ist, muss keine bestimmte Reihenfolge der Codons untereinander (wie etwa beim Zählen) eingehalten werden.

2.4 Experimentelle Erweiterung um zusätzliche Basen

2.4.1 Anzahl der Codierungsmöglichkeiten beim sechsbasigen Code

a) 6 Basen 1 Stelle → 6^1 = 6 Möglichkeiten
b) 6 Basen 2 Stellen → 6^2 = 36 Möglichkeiten
c) 6 Basen 3 Stellen → 6^3 = 216 Möglichkeiten
d) 6 Basen 4 Stellen → 6^4 = 1296 Möglichkeiten

Die experimentelle Erweiterung des bekannten genetischen Triplettcodes um zwei weitere Basen (ein Basenpaar) ergäbe 216 – 64 = 152 zusätzliche Codierungsmöglichkeiten. Bei gleichzeitiger experimenteller Erhöhung der Codonlänge, steigt die Zahl der Möglichkeiten weiter stark an (Zhang et al., 2017; Dien et al., 2018; Kapitel 7.4.3; Erweiterung des genetischen Codes, Sanofi, online).

2.4.2 Anzahl der Codierungsmöglichkeiten beim achtbasigen Code (Hachimoji-Code)

Eine weitere, oft genutzte Möglichkeit der experimentellen Biologie, stellt der Hachimoji-Code, ein Achtercode, dar. Hachimoji kommt aus dem Japanischen und bedeutet acht Buchstaben. Dieser Code enthält zusätzlich zu den bekannten, bereits vorhandenen 2 Basenpaaren zwei weitere Basenpaare, wobei auch hier der Triplettcode beibehalten wird (Hoshika et al., 2019). Auch in diesem Fall würde mit einer Erhöhung der Codonlänge die Zahl der Möglichkeiten sehr schnell ansteigen. In der experimentellen Biologie werden sogar bis zu 12 Basen verwendet (Eberlein et al., 2020).

a) 8 Basen 1 Stelle → 8^1 = 8 Möglichkeiten
b) 8 Basen 2 Stellen → 8^2 = 64 Möglichkeiten
c) 8 Basen 3 Stellen → 8^3 = 512 Möglichkeiten
d) 8 Basen 4 Stellen → 8^4 = 4096 Möglichkeiten

2.5 Redundanter Code – nicht redundanter Code. Ist der Code eindeutig?

Die Übersetzung von einem Codon in eine Aminosäure ist eindeutig – die Rückübersetzung von einer Aminosäure in ein Codon ist nicht eindeutig. Dies bedeutet, dass der Code nicht bijektiv (ein-eindeutig) ist. Ein ein-eindeutiger umkehrbarer Code in welchem die Anzahl der möglichen Codons exakt der geforderten Anzahl der zu codierenden 20 Aminosäuren entspricht, kann aber nicht erzeugt werden. Es bleiben also, unabhängig vom gewählten Code, immer Codons übrig. Diese übrigen Codons können als Zweitcode für eine bereits codierte Aminosäure doppelt belegt werden und so zu einer gewissen Redundanz beitragen, oder sie bleiben unbelegt. Abbildung 2.10 zeigt symbolisch die Unmöglichkeit der exakten Rückübersetzung von der Aminosäure zum Codon.

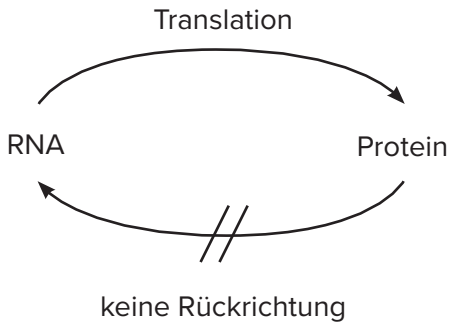


Abb. 2.10 Die Übersetzung von RNA in Aminosäuren funktioniert nur in eine Richtung.

Aus den obigen Überlegungen ergibt sich folgende Beobachtung: Es gibt *keinen Code* mit konstanter Codonlänge, der die *genaue Anzahl* unserer heutigen 20 proteinogenen Aminosäuren codieren könnte, ohne dass Codons übrig bleiben.

Es stellt sich auch die Frage, was passiert, wenn den überzähligen Codons – es stehen mehr Codons zur Verfügung, als für die Codierung aller Aminosäuren nötig ist – nichts zugeordnet wird. Ein Code, der nur die absolut notwendige Anzahl von Aminosäuren erzeugt, wäre auch insofern unflexibel, als im Laufe der Evolution neue Aminosäuren hinzugekommen sind, die Codierung sich aber nur sehr wenig bis gar nicht geändert hat. Darüberhinaus bildet *keines der möglichen Codierungssysteme* in eineindeutiger Weise alle 20 proteinogenen Aminosäuren und mindestens ein Stopp-Codon ab, ohne überzählige Codons zu haben. Beim heutigen genetischen Code codiert jedes Codon für eine Aminosäure bzw. für ein Stopp-Codon. Codons, die für die gleiche Aminosäure stehen, nennt man *synonyme Codons* (Schaaf & Zschocke, 2018, S. 39).

Ein Code, bei dem mehreren Elementen (Codons) jeweils ein Element des Zielraumes (Aminosäuren) zugeordnet wird, nennt man *surjektiv*. Im Falle des genetischen Codes entspricht dies dem redundanten Code (siehe Abb. 2.11). Ein ein-eindeutiger Code wäre ein solcher, bei dem einem Codon genau eine Aminosäure zugeordnet wird (siehe Abb. 2.12) Ein solcher Code existiert aber nicht in der Natur, da es mehr Codons als kanonische Aminosäuren gibt. Ein solcher Code wäre reversibel (bijektiv), das heißt, man könnte von der Aminosäure in eindeutiger Weise auf den Code rückschließen. Bezogen auf den genetischen Code ist ein solcher als fiktiv anzusehen.

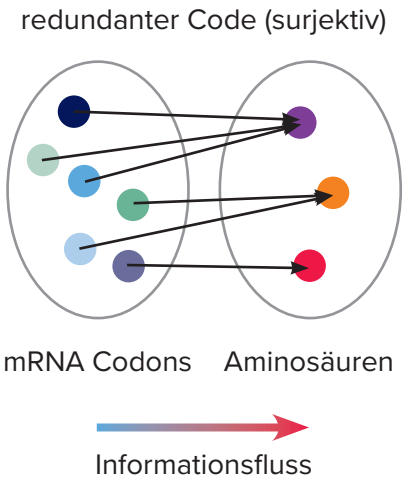


Abb. 2.11 Darstellung eines redundanten Codes, bei dem mehreren Codons die selbe Aminosäure zugeordnet wird (surjektiver Code). (Adaptiert nach Mathematikseiten der Uni Bielefeld, online).

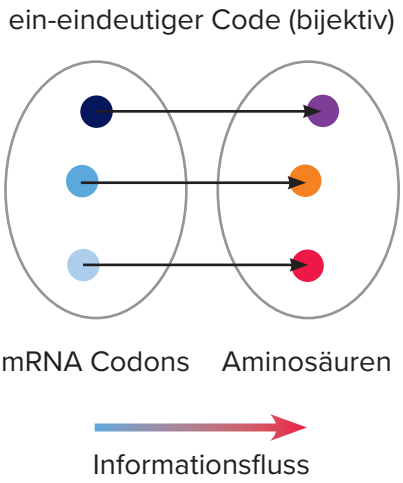


Abb. 2.12 Darstellung eines fiktiven ein-eindeutigen Codes, bei dem einem Codon genau eine Aminosäure zugeordnet wird (bijektiver Code). (Adaptiert nach Mathematikseiten der Uni Bielefeld, online).

Abbildung 2.13 zeigt die mathematische Abbildung der Codons auf die zugehörigen Aminosäuren im Falle eines Codes, bei welchem einzelne Codons nicht abgebildet werden und in der Zielmenge (Menge der Aminosäuren) leer bleiben.

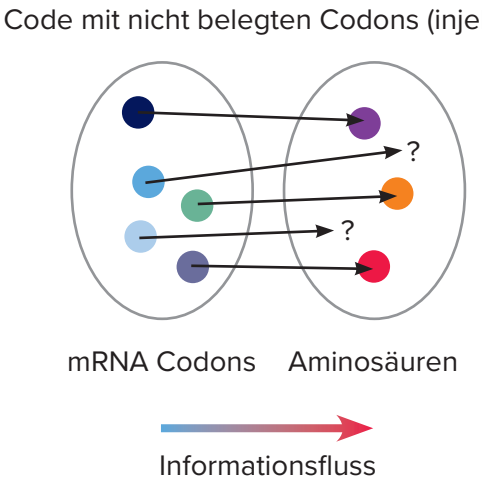


Abb. 2.13 Darstellung eines Codes mit nicht belegten Codons. (Adaptiert nach Mathematikseiten der Uni Bielefeld, online).

In den Abbildungen 2.11 bis 2.13 wurde der Informationsfluss von Codons (nicht von Basen) zu Aminosäuren dargestellt.

Code mit unterschiedlicher Länge (variable-length code): Soll ein Code möglichst sparsam bezüglich der Gesamtlänge über die gesamte Nachricht sein, so empfiehlt sich immer dann ein

Code mit variabler Länge, wenn die einzelnen Codewörter mit sehr unterschiedlicher Häufigkeit auftreten. Dies ist, wie beschrieben, beim Morsecode, beim Shannon-Fano-Code oder beim Huffman-Code der Fall.

Code mit fixer Länge (fixed-length code): Ein Code mit fixer Länge kann zwar prinzipiell immer verwendet werden, er ist aber dann besonders von Vorteil, wenn alle Codewörter mit der selben Häufigkeit auftreten.

Beim heute bekannten *genetischen Code* handelt es sich um einen *kommafreien Code fixer Länge* (mit fixem Leseraster). Da aber dieverschiedenen Codons mit deutlich unterschiedlicher Häufigkeit auftreten (siehe Abb. 3.2), ist theoretisch ein platzsparenderer Code denkbar. Abgesehen von den seltenen Stopp-Codons variieren die Häufigkeiten zwischen den Codons nahezu um den Faktor 10. Die Häufigkeitsverteilung der Codons variiert zusätzlich von Spezies zu Spezies und von Art zu Art sehr stark. Eine weitere Eigenschaft des genetischen Codes ist, dass er überlappungsfrei ist, abgesehen von wenigen Ausnahmen bei einigen speziellen Viren (Crick et al., 1957).

Die nachfolgende Zusammenstellung zeigt die möglichen Kombinationen von Codes mit fixer bzw. variabler Länge sowie mit oder ohne Komma und deren Vor- und Nachteile.

	Code fixer Länge		**Code variabler Länge**	
	Komma	kein Komma	Komma	kein Komma
Nachteil	Unnötig langer Code, da nach jedem Codon ein Komma eingefügt wird. Dieses ist nicht nötig, da bekannt ist, wann das nächste Zeichen anfängt.*	Bei fehlerhafter Übertragung ist der Text nach dem Fehler nicht mehr zu verwenden, und stellt Unsinn dar.	Die Zeichenfolge wird durch die Kommas länger.	Je nach Art der Codierung kann sich, nicht nur bei fehlerhafter Übertragung, Unsinn ergeben.
Vorteil	Bei fehlerhafter Übertragung synchronisiert sich der Code nach dem nächsten Komma wieder von selbst. (Der Fehler bleibt aber bestehen).	Einsparung des Kommas, wodurch der Code sparsamer ist (z. B. genetischer Code).	Durch statistisch angepasste Länge der Codons gemäß der Häufigkeit des Auftretens, ergibt sich oft dennoch eine Ersparnis in der Gesamtlänge. (Z. B. Morsecode).	Es gibt spezielle Codes, bei denen auch ohne Komma keine Verwechslung auftreten kann, z. B. beim *Huffman-Code*. Diese Codes sind besonders platzsparend und der Code ist selbstsynchronisierend.

** Ein solcher Code könnte beim genetischen Code mit 4 Basen und einer Wortlänge von 3 (Triplettcode) gebildet werden, wenn eine Base als Komma reserviert ist. Es blieben in diesem Fall drei Basen zur Codierung übrig.*

Drei Basen mit drei Stellen (Codonlänge 3) ergeben die Möglichkeit $3^3 = 27$ verschiedene Aminosäuren zu codieren, was für den heute bekannten genetischen Code mit 20 proteinogenen Aminosäuren ausreichend wäre. Das Komma wäre dann an der vierten Stelle durch die vierte Base repräsentiert, die aber selbst nicht zur Codierung zählt. (Bresch & Hausmann, 1972, S. 243).

Die Redundanz des Codes würde durch die geringere Anzahl an Möglichkeiten zwar verringert, es ergibt sich jedoch durch das Komma der Vorteil, dass sich bei fehlerhaften Codons oder auch bei fehlerhaftem Komma der Code von selbst wieder synchronisiert. Ab dem nächsten folgenden korrekten Komma werden die Codons wieder korrekt abgelesen. (Es muss dabei natürlich auch in Betracht gezogen werden, dass auch das Komma fehlerhaft übertragen werden kann.)

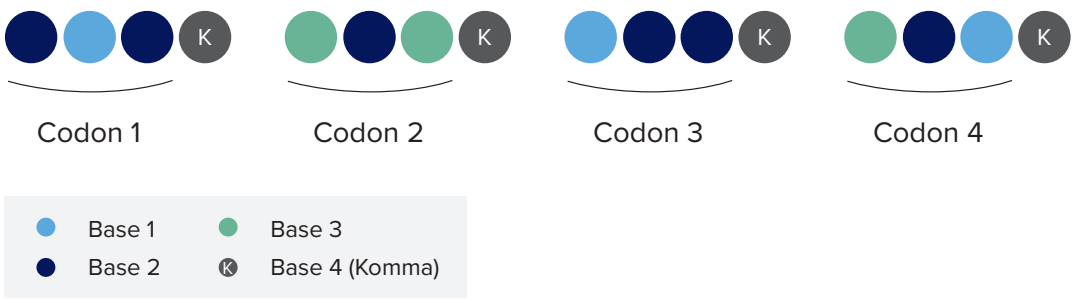


Abb. 2.14 Triplettcode mit 4 Basen, wobei die vierte Base als Komma verwendet wird. Solche Überlegungen gehen auch auf Crick et al. (1961) zurück.

Zusammmenfassend lässt sich sagen, dass Codes mit fixer Codonlänge immer eine *Maximalzahl* an codierbaren Möglichkeiten aufweisen. Ein Kommacode mit unterschiedlichen Codonlängen kann die codierbaren Möglichkeiten durch weiteres Verlängern beliebig ausweiten, ohne dass der schon vorhandene Anteil geändert werden muss. Der Sonderfall eines Kommacodes mit fixer Codonlänge hat ebenso, wie der kommafreie Code mit fixer Codonlänge, nur eine begrenzte Anzahl von Möglichkeiten.

Bei der Bezeichnung *selbstsynchronisierend* ist Vorsicht geboten, denn in der theoretischen Informatik werden solche Codes als *selbstsynchronisierend* bezeichnet, die eine eindeutige Decodierung erlauben, wenn *keine Fehler bei der Übertragung* auftreten. Dazu gehören Codes mit fixer Länge und Kommacodes, egal ob diese eine fixe oder eine variable Länge aufweisen. Bei Kommacodes synchronisiert sich das Ableseraster nach einem Fehler wieder, der Fehler in der Mitte bleibt aber bestehen. Es gibt weiterhin spezielle *Codes mit variabler Länge*, die kein Komma benötigen, welche aber eindeutig sind. Dabei handelt es sich beispielsweise um den Huffman-Code oder den Shannon-Fano-Code.

Bei einem kommafreien Code konstanter Länge muss man prinzipiell von vorne (vom Start-Codon aus) anfangen zu zählen. Es ist nicht möglich, an einer beliebigen Stelle in der Mitte zu be-

ginnen, da man nicht weiß, an welcher Stelle des Rasters man sich befindet (Crick et al., 1961). Im Gegensatz dazu, kann man bei Sprache spätestens nach dem Wortende oder allenfalls nach dem Satzende, wieder sinnvolle Inhalte ablesen.

Vergleichbar ist die Speicherung des genetischen Codes (Code mit fixer Länge) mit dem Computer *pdp-8* der *Digital Equipment Corporation*, der Anfang der 1970er-Jahre benutzt wurde. Auf den Magnetbändern, auf denen die Daten gespeichert wurden, gab es damals noch kein Filesystem. Um einen bestimmten Datensatz zu finden, musste das Band jedesmal ganz zurückgepult werden und die seitlich vorhandene Markierung (clock) wurde abgezählt, bis man zum gewünschten Ort des Datensatzes kam.

Doig (1997) berechnete die minimale Codonlänge für den genetischen Code mit 20 Aminosäuren und einem Stopp-Codon bei einem vierbasigen Code unter Verwendung der Huffman-Codierung. Diese weist eine variable Codonlänge auf und ist daher platzsparender als eine Codierung mit fixer Codonlänge. Er errechnet dafür eine mittlere Codonlänge von 2,11, was insgesamt eine Ersparnis gegenüber dem vierbasigen Code mit fixer Länge von 3 darstellt. Nicht berücksichtigt wurde dabei, dass sich die Codonverwendung (codon bias) und die Häufigkeiten einzelner Aminosäuren von Spezies zu Spezies unterscheiden und daher der Code speziestypisch ausfallen müsste. Damit wäre der Code für Viren, Bakterien und sogar für verschiedene Eukaryoten unterschiedlich. Weiters geht bei dieser Codierung die Redundanz verloren, die aber für die Faltung von Proteinen wichtig ist (siehe Kapitel 3.4), (Ziegenbalg et al., 2016, S. 281–287; Huffman-Kodierung, Wikipedia; Shannon-Fano-Kodierung, Wikipedia; Borys, 2011, S. 197).

Bei den in diesem Kapitel dargestellten Varianten handelt es sich um rein theoretische Überlegungen, was eine mögliche genetische Codierung betrifft.

Erste Lebensformen entstanden vor etwa 3,8 Milliarden Jahren auf der Erde aus organischen Molekülen (Geiger, 2009, S. 196). Um die den lebenden Systemen zu Grunde liegende Information (Bauplan) von Generation zu Generation weitergeben zu können, muss sich früh eine Speichermöglichkeit mit einer Codierung gebildet haben. Vermutlich sind alle Spezies aus dieser Urform hervorgegangen, was im Rahmen der Evolution die Einheitlichkeit des Codes hervorgerufen hat.

3 Der heute bekannte genetische Code

3.1 Allgemeines

Das Wichtigste für eigenständiges Leben ist ein Konstruktionsplan – eine Information –, wie das Leben von Generation zu Generation weitergegeben werden kann. Der Plan – die Information – ist lediglich eine Vorschrift, kann selbst aber nichts. Diese Bauvorschrift ist mit einem Code für die einzelnen Aminosäuren versehen, aus denen sich Proteine zusammensetzen. Dies ist vergleichbar mit einer schriftlichen Anweisung, in welcher *Reihenfolge* die Aminosäuren zu einem Protein zusammenzufügen sind. Es bedarf einer Anweisung, die – wie in Büchern – die Information von Generation zu Generation weitergibt.

Nachdem die allgemeine Struktur der DNA bereits seit 1953 bekannt war, veröffentlichten Crick et al. (1961) einige allgemeine Erkenntnisse über den genetischen Code als solchen. Aus Experimenten schlossen sie, dass der Code aus jeweils drei Basen oder einem Vielfachen davon für eine Aminosäure besteht, von einem fixen Startpunkt aus gelesen werden muss, nicht überlappend ist und kein Komma zur Separation der einzelnen Codewörter besitzt. Eine weitere wichtige Erkenntnis war, dass der Code degeneriert sein muss, dass also eine Aminosäure durch mehrere synonyme Codons codiert werden kann. Wäre dies nicht der Fall, so müssten 64 (mögliche Codons) minus 20 (Aminosäuren) 44 Nonsense-Codons ergeben, was sie aufgrund von Mutationsexperimenten ausschlossen.

Marshall Nirenberg und sein Mitarbeiter Heinrich Matthaei hatten eine Methode entwickelt, kurze RNA-Nucleotide ohne Vorlage zu synthetisieren (Alberts et al., 2021, S. 269), (Nirenberg & Matthaei, 1961). Mit Hilfe von Experimenten mit synthetischer poly-U-RNA und radioaktiv markierten Aminosäuren konnten sie das erste Codon decodieren. Für UUU ergab sich die Aminosäure Phenylalanin.

Nirenberg und Leder veröffentlichten 1964 noch einmal zusammenfassend die durchgeführten Experimente, die sie mit Hilfe der synthetischen RNA-Oligonucleotide poly-U, poly-A und poly-C und jeweils einer radioaktiv markierten Aminosäure durchgeführt hatten. (Nirenberg & Leder, 1964). Weitere Experimente mit gemischten RNA-Oligonucleotiden brachten weitere Erkenntnisse über Codons, aber es dauerte noch bis 1966, bis alle Codons bekannt waren.

Die Anweisung für jede Aminosäure besteht aus jeweils drei Buchstaben, die ohne Abstände aneinandergereiht werden, wie dies im antiken Griechisch der Fall war. In Kapitel 2 wurden mehrere theoretisch denkbare Möglichkeiten erörtert, wie eine solche Codierung aussehen könnte. Der heute bekannte *Genetische Code* wird – mit geringfügigen Ausnahmen – von allen Domänen des Lebens (Eukaryoten, Prokaryoten, Archaeen) sowie Viren – verwendet (Schaaf & Zschocke, 2018, S. 20). Es soll an dieser Stelle nicht auf die Diskussion eingegangen werden, ob Archaeen eine eigene Domäne sind bzw. ob Viren als Lebewesen zu betrachten seien.

Es ist wichtig, dass alle Domänen und Viren einen weitgehend identischen Code verwenden, denn ansonsten könnten beispielsweise Viren oder Bakterien nicht Eukaryoten befallen bzw. Viren könnten nicht Bakterien befallen und deren Reproduktionssystem in parasitärer Weise missbrauchen. Der Code für jene Aminosäuren, aus denen sich Proteine zusammensetzen, soll im Einzelnen dargelegt werden. Proteine, die ohne Verwendung des genetischen Codes und ohne Ribosomen über andere Wege synthetisiert werden – meist handelt es sich um sehr kurze Polypeptide –, sind nicht Gegenstand dieser Betrachtung.

Aminosäuren, die mit Hilfe von Ribosomen und des genetischen Codes in Proteine übersetzt werden, werden als *kanonische Aminosäuren* oder *proteinogene Aminosäuren* bezeichnet. Der bekannte Genetische Code besteht aus 4 unterschiedlichen Zeichen, den Basen Adenin (A), Uracil (U) bzw. Thymin (T) in der DNA, Guanin (G) und Cytosin (C), die jeweils in Dreiergruppen (Tripletts) zusammengefasst sind. Dadurch ergeben sich $4^3 = 64$ verschiedene Codierungsmöglichkeiten, die aus der Codonsonne abgelesen werden können. Ein Triplett, welches für eine Aminosäure codiert, wird als *Codon* bezeichnet (Christen 2016, S. 87ff)..

3.2 Codonsonne – Synonyme Codons

Die Codonsonne (Abb. 3.1) gibt die Reihenfolge innerhalb der Basentripletts an, die für je eine der 20 kanonischen bzw. proteinogene Aminosäuren stehen. Die Reihenfolge ist von innen vom 5‘-Ende der DNA/RNA nach außen zum 3‘- Ende zu lesen. (Die Bezeichnung 5‘ bzw. 3‘ bezieht sich auf die Nummerierung der Zucker in der DNA). Die Codonsonne verwendet üblicherweise den Code für die RNA, mit Uracil (U) als Base. Soll der Code für die DNA abgelesen werden, muss Uracil (U) durch Thymin (T) ersetzt werden.

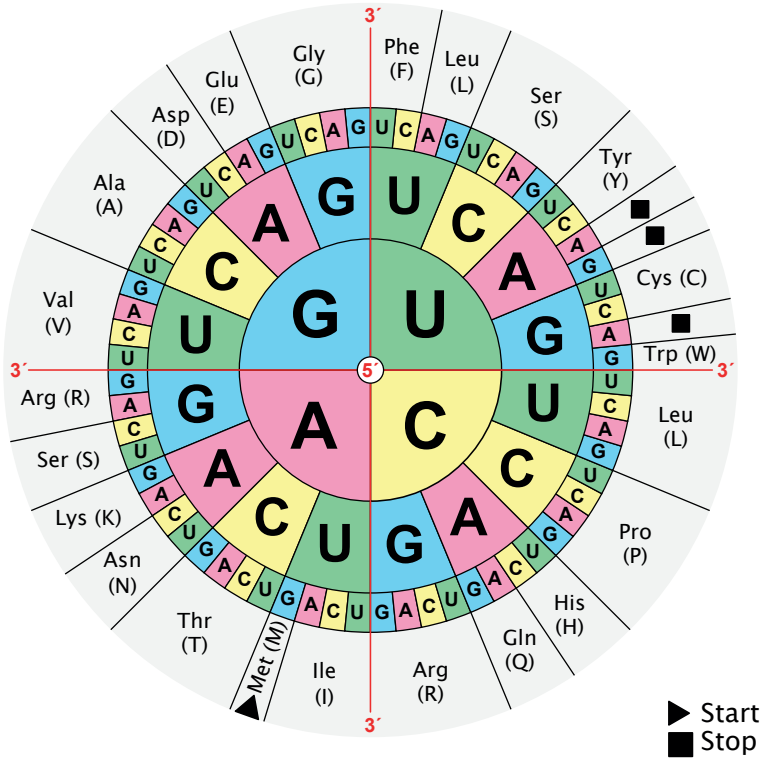


Abb. 3.1 Codonsonne (Wikipedia).

Im äußeren Ring wird die codierte Aminosäure sowohl im Dreibuchstabencode als auch im Einbuchstabencode vermerkt (Liste der Aminosäuren im Anhang).

Das Start-Codon hat denselben Code, wie die Aminosäure Methionin, wohingegen es drei unterschiedliche Stopp-Codons gibt. In diesem Fall würden also 20 verschiedene Codons genügen, um alle 20 kanonischen Aminosäuren zu codieren. Es werden also nur 20 von 64 verschiedenen Codierungsmöglichkeiten benötigt, daher bleiben theoretisch 64 – 20 = 44 Codons übrig, die nicht notwendig sind, um die erforderliche Anzahl von Aminosäuren zu codieren. Zieht man die drei Stopp-Codons ab, so verbleiben 61 Codons für 20 Aminosäuren. Da alle Codons belegt sind, kann man nun einen *Mittelwert* dafür berechnen, wieviele Codons pro Aminosäure im Mittel zur Verfügung stehen. Im Durchschnitt sind es 61 : 20 = 3,05 ≈ 3 Codons pro Aminosäure. Die Verteilung der Anzahl von Codons für dieselbe Aminosäure ist jedoch sehr unterschiedlich. Wie aus der Codonsonne ersichtlich, stehen für Methionin (Met) und Tryptophan (Trp) jeweils nur ein einziges Codon zur Verfügung, für Leucin, Serin und Arginin aber jeweils sechs.

Gibt es für eine Aminosäure mehrere unterschiedliche Codons, so werden diese als *Synonyme Codons* bezeichnet.

3.3 Unterschiedliche Häufigkeiten synonymer Codons

Die Häufigkeit der Verwendung der unterschiedlichen synonymen Codons ist von Spezies zu Spezies und sogar von Gen zu Gen unterschiedlich und wird als *Codon bias* bezeichnet (Graw, 2020, S. 96). Eine Darstellung der Häufigkeit der Verwendung synonymer Codons ist in der Codonsonne nicht möglich. Selten verwendete Codons werden als *seltene Codons* oder *rare codons* bezeichnet, häufig verwendete als *häufige Codons* bzw. *frequent codons*. Diese Angabe muss für jede Spezies und für jedes Gen separat erfolgen.

Abbildung 2.2 zeigt die Zuordnung der einzelnen Aminosäuren im *Menschen* zu ihren möglichen Codons. Jede Spalte steht für *eine* der 20 proteinogenen Aminosäuren – in alphabetischer Reihenfolge – wobei die synonymen Codons der jeweiligen Aminosäure in dieser Spalte übereinander angeordnet sind (Anzahl der möglichen Codons pro 1000 Aminosäuren, gemittelt über alle Gene, aufgetragen gegen die einzelnen Aminosäuren). Die Codons einer Spalte codieren für *dieselbe Aminosäure*. Da es zu fast allen Aminosäuren – außer Methionin (Met) und Tryptophan (Trp) – mehr als ein Codon gibt, wird der genetische Code als *redundant* oder *degeneriert* bezeichnet. Die Werte wurden aus der NIH Genbank entnommen (siehe Anhang).

Es galt bis vor ca. 10 Jahren als gesichert, dass synonyme Codons bei Untersuchungen bezüglich Mutationen unberücksichtigt bleiben können, da sie für die selbe Aminosäure codieren. Viele klassische Lehrbücher (wie z. B. Christen 2016) erwähnen nur, dass synonyme Codons für die selbe Aminosäure codieren, Probleme die sich bei Genexpression, Faltungen und gar Krebs ergeben können, werden nicht erwähnt. In neueren Lehrbüchern (Graw 2020, S. 117) wird bereits darauf hingewiesen, dass synonyme Codons *funktionell nicht gleichwertig* sind, weil dies beispielsweise die Translationsrate beeinflussen können.

Häufigkeit der synonymen Codons im Menschen (H. sapiens), pro 1000 Aminosäuren

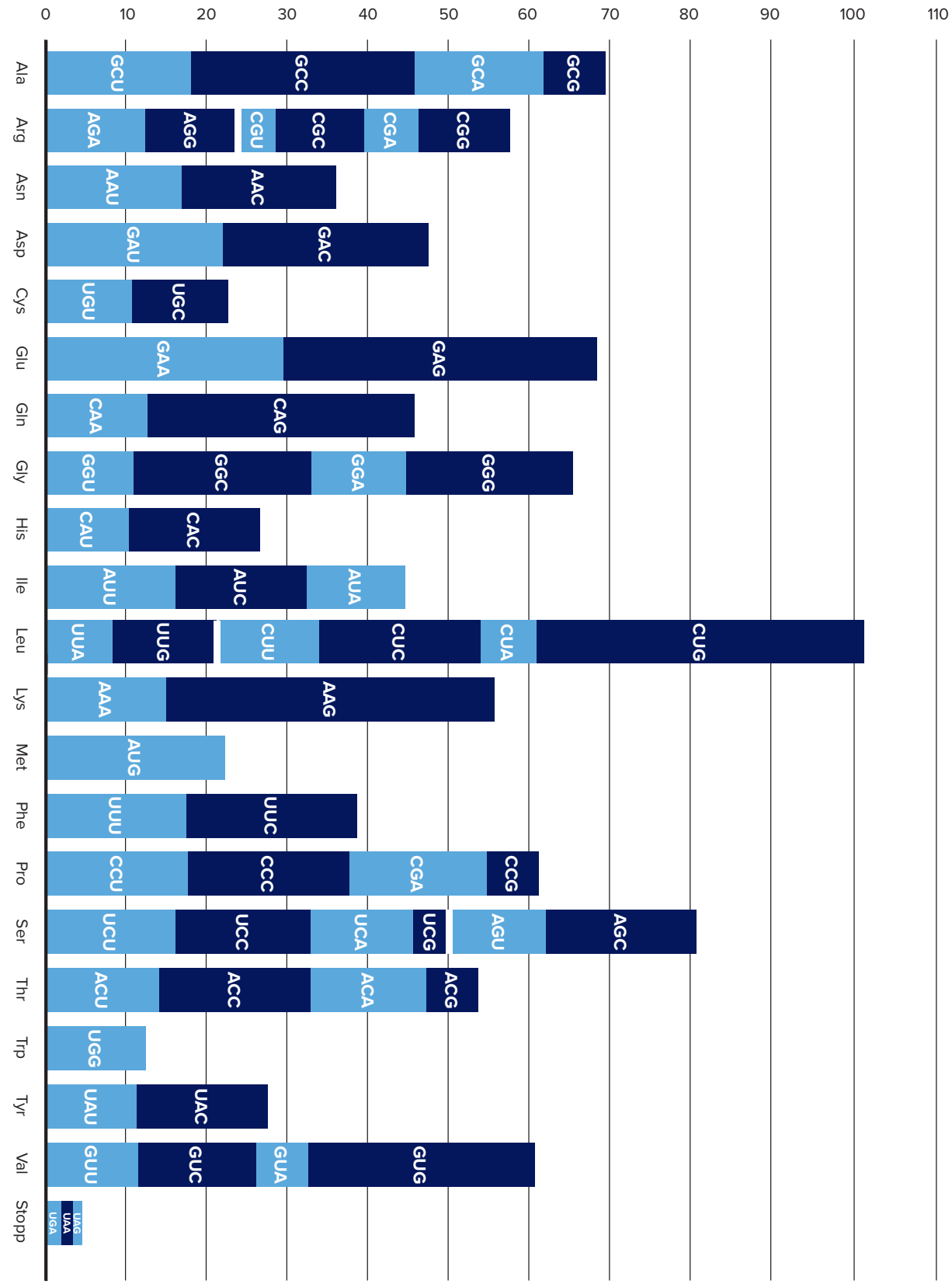


Abb. 3.2 Verteilung der synonymen Codons beim Menschen (homo sapiens) in Abhängigkeit von der jeweiligen proteinogenen Aminosäure – in alphabetischer Reihenfolge – pro 1000 Aminosäuren, gemittelt über alle Gene. Werte aus der NIH Genbank – im Anhang.

3.4 Was bedeutet Redundanz – Code im Code

Im *Computerlexikon (online)* kann man lesen:

> **Redundanz bedeutet, dass etwas öfter vorhanden ist, als es eigentlich gebraucht wird. Bei Daten bedeutet das, dass mehr Daten vorhanden oder übertragen werden, als eigentlich notwendig. Dies dient zur Kontrolle, ob die Daten korrekt angekommen sind.**

Und im *digitalen Wörterbuch der deutschen Sprache* findet man:

> **Redundanz: Vorhandensein von überflüssigen, für die Information nicht notwendigen Elementen in einer Nachricht**

Diese Definitionen legen nahe, dass man einen Teil der Codierung weglassen kann, ohne dass *Missverständnisse bzw. Probleme* auftreten können. Überträgt man diese Definitionen auf die Genetik, so bedeutet dies, dass die synonymen Codons unberücksichtigt bleiben können, da sie für die selbe Aminosäure codieren und daher keine zusätzliche Information liefern. Daher wurden oft Mutationen, die zu synonymen Codons führten, in Gendatenbanken für Krebsmutationen nicht vermerkt.

Redundanz bedeutet aber auch, dass eine gewisse Fehlertoleranz existiert. Diese ist für die meisten Codons (außer für Tryptophan und Methionin) in einem gewissen Maße gegeben. Besonders hoch ist die Redundanz für Serin, Arginin und Leucin, die jeweils durch sechs verschiedene Codons repräsentiert werden. Die Redundanz gilt aber nicht für alle Basenpositionen gleichermaßen – sie ist für die dritte Position im Triplett am höchsten (siehe Wobble-Theorie, Kapitel 3.5) und für die erste Position am geringsten (Saier, Jr, Milton H, 2019).

Als Vorteil des redundanten Codes wurde früher argumentiert, (Saier, Jr., Milton, H. 2019)

> **„... that the redundant code decreases the deleterious consequences of random point mutations“**
>
> **[...dass der redundante Code die schädlichen Auswirkungen von zufälligen Punktmutationen verringert]**

Berücksichtigt man neuere Erkenntnisse, so zeigt sich, dass synonyme Codons zwar die selbe Aminosäure im Protein ergeben, dass diese synonymen Codons dennoch nicht beliebig austauschbar sind, da es noch mindestens eine weitere Ebene der Information gibt (Code im Code). (Komar, 2021; Liu, 2020). In dieser zweiten Ebene sind die sogenannten synonymen Codons

nicht mehr redundant, da ein Codon eine eigene, nicht austauschbare Bedeutung bzw. Auswirkung haben kann.

3.5 Wobble-Hypothese

Francis Crick hat 1966 die Wobble-Hypothese unter folgendem Titel veröffentlicht:

> **Codon-Anticodon Pairing: The Wobble Hypothesis**
>
> **It is suggested that while the standard base pairs may be used rather strictly in the first two positions of the triplet, there may be some wobble in the pairing of the third base. This hypothesis is explored systematically, and it is shown that such a wobble could explain the general nature of the degeneracy of the genetic code (Crick, 1966).**

Während der Translation bindet das Anticodon der tRNA innerhalb des Ribosoms an die mRNA. Die Basen eines Codons der mRNA Adenin (A), Guanin (G), Cytosin(C) und Uracil (U) paaren mit dem Anticodon der tRNA (A mit U und G mit C). A bildet mit U zwei Wasserstoffbrücken aus, C bildet mit G jeweils drei Wasserstoffbrücken (Watson-Crick-Paarung). Die Paarung der beiden ersten Basen mit dem Anticodon ist sehr strikt, die dritte Base der mRNA kann aber in manchen Fällen mit einer unüblichen Base des Anticodons paaren.

Dadurch dass die Anticodonschleife quasi verbogen ist, ist die Paarung an der dritten Position des mRNA-Codons nicht so genau. Die Abstände zwischen der dritten Base der mRNA und der Base des Anticodons haben einen anderen Abstand, als bei der Standardpaarung. An Position 34 des Anticodons, der Wobble-Position, können sich daher auch andere Basen mit der dritten Base der mRNA paaren (Buttlar et al., 2020, S. 117).

Abbildung 3.3 zeigt einen Ausschnitt der tRNA, den Anticodonstamm und die Anticodonschleife, (siehe auch Kapitel 1, tRNA) und die möglichen Basenpaarungen (bei Eukaryoten) zwischen dem Codon der mRNA und dem Anticodon der tRNA.

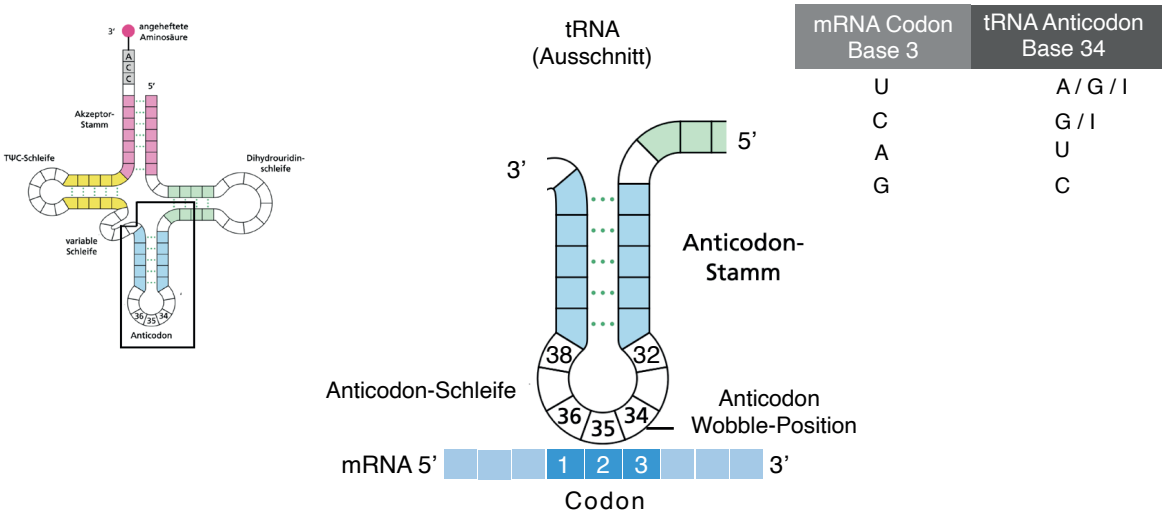


Abb. 3.3 Mögliche Paarungen der Wobble-Position des Anticodons mit der mRNA (Berg et al., 2018, S. 1063; Alberts 2022, S. 360 adaptiert).

Durch posttranskriptionelle Modifikationen erhält die tRNA verschiedene modifizierte Basen, wie beispielsweise Pseudouridin oder Inosin. Inosin, ein Nucleosid aus einer Purinbase und einer Ribose, entsteht durch posttranskriptionelle Desaminierung von Adenosin und kommt manchmal an der Wobble-Position der tRNA vor. Inosin kann mit Uracil bzw. Cytosin der mRNA paaren, da es aber ausschließlich mit U oder C paart, betrifft dies ausschließlich synonyme Codons. INN im Anticodon paart sowohl mit NNU als auch mit NNC. Diese Codons sind immer synonym, wie aus der Codonsonne zu entnehmen ist. N bezeichnet gemäß IUPAC eine beliebige Base (siehe Anhang).

Durch die Möglichkeit für ein und dieselbe tRNA, an der Wobble-Position unterschiedliche Basenpaarungen einzugehen, ist nicht für jedes der 61 Codons eine eigene tRNA nötig (Berg et al., 2018, S. 1065). Eine Modifikation in Position 32 des Anticodons von U zu C kann die Codonerkennung beeinflussen, obwohl sich diese Position nicht mehr direkt im Anticodon befindet, wohl aber noch innerhalb der Anticodonschleife (Agris et al., 2017, siehe Abb. 3.3). Olejniczak und Uhlenbeck (2006) stellten fest, dass die Wasserstoffbrückenbindung zwischen den Nucleotiden 32 und 38 die Affinität der tRNA zur A-Stelle des Ribosoms erheblich beeinflusst.

Einen Sonderfall für unübliche Basenpaarung zwischen Codon und Anticodon stellt das Überlesen von regulären Stopp-Codons dar (NTC, normal termination codon). Dabrowski et al. (2015) untersuchten in einem Review unter anderem das Überlesen *(translational readthrough)* von regulären Stopp-Codons in Eukaryoten. Dies ist von unterschiedlichen Faktoren abhängig, insbesondere vom Stopp-Codon selbst und von den das Codon umgebenden Sequenzen.

Erscheint ein Stopp-Codon in der A-Stelle des Ribosoms, so steht keine passende tRNA zur Verfügung und es kommen statt dessen zwei eukaryotische Releasefaktoren (Freisetzungsfaktoren) eRF1 und eRF3 zum Einsatz. Wenn am 3'-Ende der codierenden Sequenz der mRNA

das Stopp-Codon in die A-Stelle des Ribosoms gelangt, kann sich in seltenen Fällen eine near cognate-tRNA *statt der Releasefaktoren* an diese Stelle im Ribosom binden und das Ribosom translatiert weiter. Hier kann es zur Ausbildung von zum Teil unüblichen Basenpaarungen kommen, die auch die Wobble-Position betreffen können. Das führt zum Einbau einer Aminosäure und Fortsetzung der Translation über das Stopp-Codon hinaus bis zum nächsten Stopp-Codon (Williams et al., 2004; Dabrowski et al., 2015). Da hier keine Mutation in der DNA vorliegt, kommt dies nicht bei jedem Translationsvorgang des Ribosoms vor, da das Stopp-Codon weiterhin vorhanden ist. Dieses wird nur mit einer bestimmten geringen Häufigkeit überlesen. Mit derselben geringen Häufigkeit wie das Überlesen, werden fehlerhafte Proteine gebildet.

Unmittelbar nach dem regulären Stopp-Codon gibt es meist weitere Stopp-Codons, die eine übermäßige Verlängerung des Proteins durch Überlesen des regulären Stopp-Codons verhindern (Williams et al., 2004; Dabrowski et al., 2015). Laut Dabrowski et al. (2015) werden die verschiedenen Stopp-Codons mit unterschiedlicher Häufigkeit überlesen. Aus der Codonsonne (siehe Abbildung 3.1) ist Folgendes zu entnehmen:

- Es gibt drei Stopp-Codons: UGA, UAA und UAG.
- UAA und UAG unterscheiden sich nur in der dritten Position voneinander (in der Wobble-Position).
- UGA unterscheidet sich in der zweiten Position von UAA bzw. in der zweiten und dritten Position von UAG.

	UGA Stopp	**UAA Stopp**	**UAG Stopp**
3. Base Wobble-Position	**UGG Trp (h) UGC Cys (h) UGU Cys (h)**	UAC Tyr UAU Tyr	UAC Tyr UAU Tyr
1. Base	**AGA Arg (h)** **CGA Arg (h)** GGA Gly	**GAA Glu (h)** AAA Lys CAA Gln	**GAG Gly (h)** AAG Lys CAG Gln
2. Base	UCA Ser UUA Leu	UCA Ser UUA Leu	UCG Ser UUG Leu **UGG Trp (h)**

(h) mismatch kommt laut Dabrowski et al. (2015) häufiger vor

Die weiteren Einflüsse auf eine Fehlpaarung zwischen einer tRNA und einem Stopp-Codon, wie beispielsweise der Einfluss von einzelnen Basen oder kurzen Sequenzen vor oder nach dem eigentlichen Codon, sollen hier nicht weiter betrachtet werden, da es sich hierbei nicht um eine Wobble-Paarung handelt.

4 Das Humangenomprojekt

1990 wurde das öffentlich finanzierte *Human Genome Project* (HGP) von Luigi Cavall-Sforza gegründet, mit dem Ziel das menschliche Genom komplett zu sequenzieren. Unter Leitung von James Watson beteiligten sich 40 Länder an dem Projekt. 2001 war das menschliche Genom weitgehend sequenziert, verstanden waren die Gene aber noch keineswegs (Wikipedia – Humangenomprojekt).

Schon früh gab es Streit zwischen Watson und der damaligen Leiterin des National Institute of Health (NIH), *Bernadine Healy*, da diese die Gensequenzen patentieren lassen wollte, James Watson aber nicht. (Wikipedia – Humangenomprojekt). Watson verließ daraufhin das Projekt.

Nach Jones et al. (2018) wurden 1996 die *Bermuda Prinzipien vom HGP (Human Genome Project)* angenommen und 1998 formal implementiert, wonach genetische Daten täglich veröffentlicht werden sollten, um Patentierungen zu verhindern. Durch die Veröffentlichungen – noch vor der Publikation wissenschaftlicher Artikel – gehören diese Daten zur *public domain*. Das NIH (National Institute of Health) und der *UK nonprofit charity the Wellcome Trust* trugen 90 Prozent der Kosten. Sonstige mitarbeitende Institutionen – vor allem auch aus Europa – mussten sich diesen Prinzipien unterwerfen. Es handelt sich bei diesen Prinzipien um keinen völkerrechtlich verbindlichen Vertrag, sodass es einzelnen Personen oder Institutionen weiterhin freisteht, Patente anzumelden. Allerdings gehören solche Institute dann nicht mehr zum Consortium.

Es gibt inzwischen unzählige freie Datenbanken in europäischen und außereuropäischen Ländern, die u.a. DNA-Sequenzen, RNA-Sequenzen, Proteinsequenzen, Sequenzen von Krebsmutationen und speziell auch von synonymen Mutationen bereitstellen. So baute beispielsweise die Universität Freiburg als eine der ersten ab 2019 eine spezielle Datenbank für synonyme Krebsmutationen auf und stellte sie online zur Verfügung (SynMICdb Synonymous Mutations In Cancer Database).

4.1 Preisverfall

Das ursprüngliche *Human Genome Project* war sehr kostenintensiv und nahm eine lange Zeit in Anspruch. Es konnte nur dadurch bewältigt werden, dass sich 40 Länder sowohl an der Arbeit als auch an den Kosten beteiligt haben.

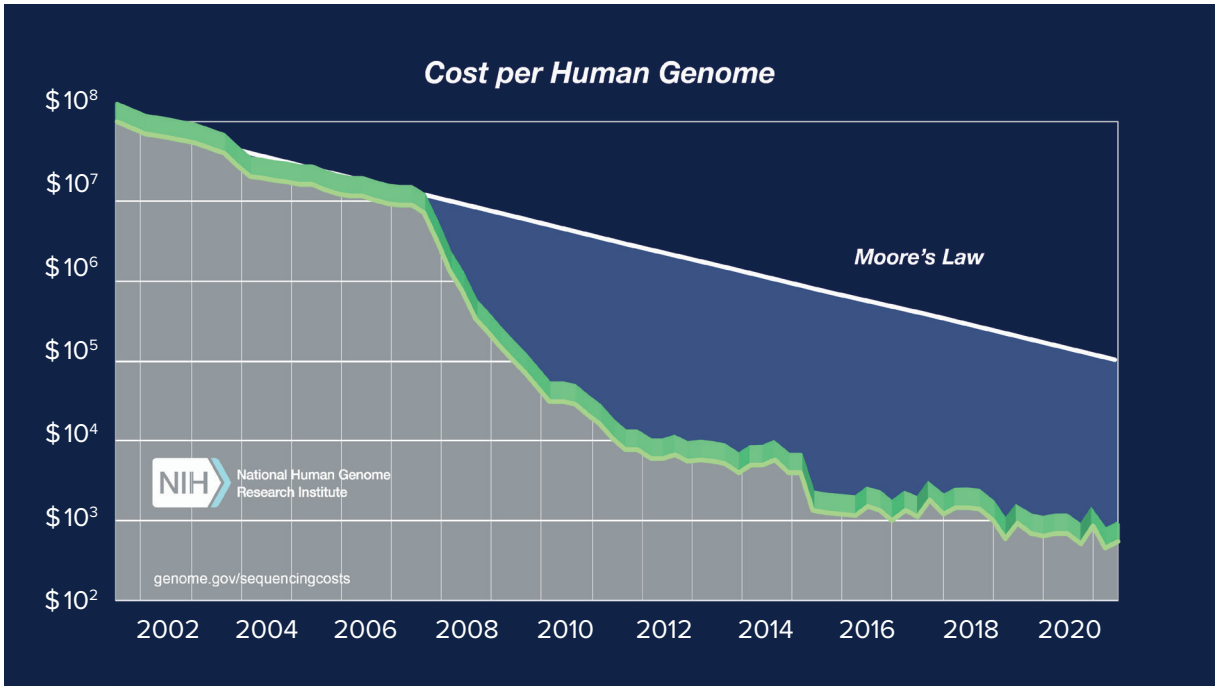


Abb. 4.1 Kosten der Sequenzierung eines menschlichen Genoms von 2001 bis 2021 (Wetterstrand 2023, National Human Genome Research Institute NIH, online, adaptiert).

Die Kosten für die Sequenzierung eines kompletten Genoms sind zwischenzeitlich stark gesunken. Bis 2001 kostete eine Genomsequenzierung etwa 300 Millionen US$ – bis 2021 ist der Preis auf unter 1000 US$ gesunken (siehe Abb. 4.1, Wetterstrand 2023, NIH). Dies ist auf schnellere und automatisierte Sequenzierungsverfahren wie beispielsweise Next Generation Sequencing zurückzuführen (DNA Sequencing Costs, Wetterstrand, 2023 online).

Laut NIH (Wetterstrand, 2023, online) wurde von 2001 bis 2007 Sanger-Sequenzierung verwendet, danach second generation bzw. next generation sequencing.

4.2 Nachfolgeprojekte

Das 2003 gestartete Nachfolgeprojekt *The ENCODE* Project (ENCyclopedia Of DNA Elements) hatte zum Ziel, alle funktionellen Elemente des menschlichen Genoms zu identifizieren. (Feingold et al., 2004) Der komplette Katalog sollte alle proteincodierenden und nicht-codierenden Teile der DNA enthalten, sowie regulatorische Elemente der Transkription.

Weitere Projekte, wie das 1000 Genom-Projekt folgten.

Eine wesentliche Erkenntnis aus all diesen Projekten war, dass es nur etwa 21 000 protein-codierende menschliche Gene gibt und ca. 3×10^9 Nucleobasen (Berg et al., 2018, S. 127). Ursprünglich war die Zahl der Gene auf etwa 100 000 geschätzt worden (Lander, 2011).

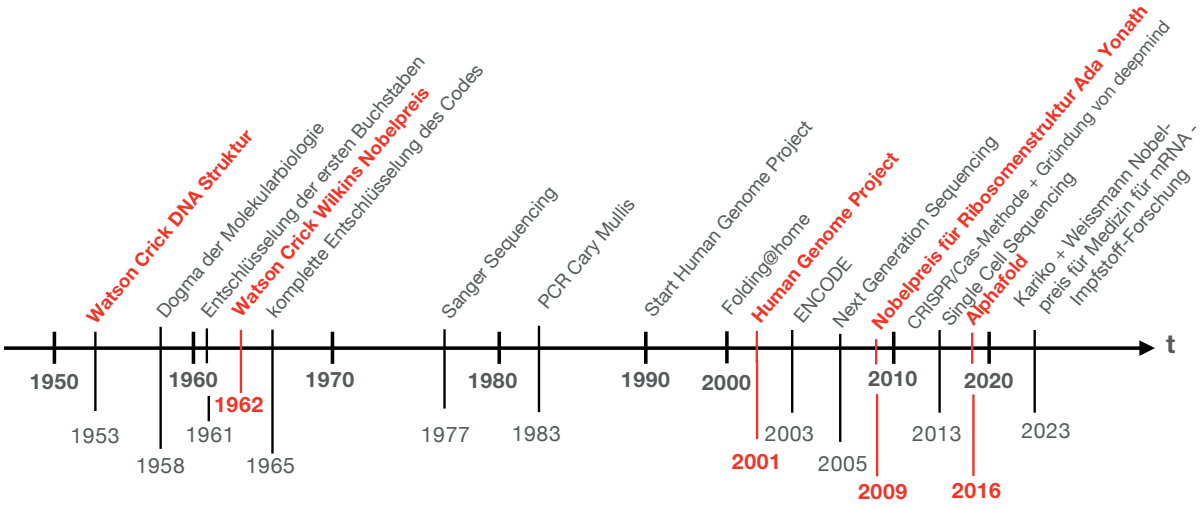


Abb. 4.2 Zeitleiste wichtiger molekulargenetischer Entdeckungen und Entwicklungen.

5 Mutationen

Laut Lexikon der Biologie (Spektrum, online) ist eine Mutation eine Veränderung des Erbguts. Mutationen können spontan auftreten oder werden durch äußere Einflüsse hervorgerufen. Mutationen, die in Keimzellen Fehler erzeugen, können von Generation zu Generation weitergegeben werden. Mutationen kommen in allen Lebewesen vor. Obwohl sie für ein einzelnes Individuum schädlich sein können, sind sie für die Evolution und das Überleben einer Spezies wichtig.

Ohne Mutationen wäre die Erde heute noch von den ersten Urlebewesen bevölkert. Durch Mutationen bilden sich gewisse Variationen innerhalb einer Population aus. Wenn sich die Umweltbedingungen ändern – höhere oder niedrigere Temperatur, Missernten – so sind immer einige Individuen mit Variationen im Erbgut vorhanden, die besser an diese neuen Bedingungen angepasst sind und daher einen Selektionsvorteil haben.

5.1 Ursachen von Mutationen

Vor jeder Zellteilung muss die Erbinformation, die bei eukaryotischen Zellen in Form der DNA im Zellkern niedergelegt ist, vollständig kopiert werden, damit jede der beiden Tochterzellen die vollständige Erbinformation erhält. Dabei wird von der DNA-Polymerase ungefähr eine von 10^5 Basen inkorrekt eingebaut. Durch Reparaturvorgänge (3‘ → 5‘ *exonucleolytic proof reading*) sinkt die Fehlerrate auf 10^{-7} und durch eine zusätzliche Reparatur von Basenfehlpaarungen *(mismatch repair)* wird die Fehlerrate bis auf 10^{-10} weiter verringert (Alberts et al., 2022, S. 260).

Im Folgenden sei, ohne Anspruch auf Vollständigkeit, ein Auszug von wichtigen Substanzen und Einflüssen genannt, die Mutationen induzieren können.

- Elektromagnetische Strahlung wie Licht, UV-Strahlung, Röntgenstrahlung oder γ-Strahlung können vielfältige Schäden, wie beispielsweise Hautkrebs, erzeugen.
- Bomben-Fallout, Kernkraftwerksunfälle und oberirdische Kernwaffentests (USA, Frankreich, China etc.) aus den 1960er-Jahren lieferten α-, β-, γ-Strahlung und radioaktive Elemente wie ^{137}Cs, ^{121}J, ^{90}Sr, ^{239}Pu, die weiter zerfallen.
- Die Erdkruste liefert α-, β- und γ-Strahlung aus dem Untergrund. Granit aus dem Waldviertel, uranhaltige Mineralien wie Schwerspat (BaSO4) und Thorium (häufiges radioaktives Element der Erdkruste) liefern Belastungen.
- Phosphatdünger enthält oft Uran, das weiter zerfällt.
- Kohle enthält die beiden α-Strahler Thorium (Th) und Uran (U), die beim Verbrennen freigesetzt werden.
- Baustoffe und alte Fliesen aus den 1970er-Jahren enthalten teilweise ^{40}K (β^+-, β^--Strahler) und Thorium und geben gasförmiges Radon ab.
- Hohe Radonbelastung aus Zerfällen von Uran aus der Erdkruste gibt es in den Stollen von Bad Gastein.
- Medikamente wie Thalidomid sind fruchtschädigend (Conterganskandal in den 1960er-Jahren).
- Viele Chemikalien, Chemotherapien sowie Schadstoffe aus dem Rauch von Zigaretten sind mutagen.

5.2 Chromosomenmutationen

Chromosomenmutationen sind Änderungen in Zahl, Form oder Struktur der Chromosomen (Graw, 2020, S. 497; Nordheim et al., 2018, S. 250).

5.2.1 Numerische Abweichungen

Bei *Polyploidie* kommt ein kompletter mehrfacher Chromosomensatz – dreifacher (Triploidie), vierfacher (Tetraploidie) – vor. Dies ist beim Menschen nicht mit dem Leben vereinbar (Nordheim et al., 2018, S. 250). Mehrfache Chromosomensätze kommen jedoch häufig bei Pflanzen, wie beispielsweise dem Weizen, vor.

Anoiploidie (Genommutation) ist die *Abweichung der Anzahl einzelner Chromosomen* von der *Norm*. Diese entstehen bei der Meiose oder Mitose (Schaaf & Zschocke, 2018, S. 43).

Einzelne Chromosomen können, statt wie üblich zweifach, dreifach vorliegen. Die bekannteste Störung beim Menschen ist die Trisomie 21, bei der das Chromosom 21 dreifach vorliegt. Bei Trisomien sind nur Individuen mit Trisomie 21 (Down-Syndrom), 13 (Pätau-Syndrom) oder 18 (Ed-

wards-Syndrom) lebensfähig. Bei Pflanzen, wie beispielsweise dem Stechapfel, kommen Trisomien regelmäßig vor (Spektrum, Lexikon der Biologie, Graw, 2020, S. 743).

Monosomien, bei denen ein Chromosom nur *einfach* vorhanden ist, sind beim Menschen, außer bei den Geschlechtschromosomen, nicht mit dem Leben vereinbar. Bei den Geschlechtschromosomen kann das X-Chromosom nur einfach vorhanden sein (Turner-Syndrom), ohne dass ein zweites X-Chromosom oder ein Y-Chromosom vorhanden wäre. Eine Monosomie mit nur einem Y-Chromosom – ohne X-Chromosom – ist nicht möglich. Das Y-Chromosom ist mit ca. 64 proteincodierenden Genen sehr klein und ihm fehlen viele lebenswichtige Gene, die nur im X-Chromosom, das ca. 852 proteincodierende Gene besitzt, vorhanden sind (Graw, 2020, S. 773).

Beim Menschen kommen größere Abweichungen bei der *Anzahl der Chromosomen* nur bei den Geschlechtschromosomen vor, wobei jene Individuen lebensfähig sind, die zumindest ein X-Chromosom besitzen.

5.2.2 Strukturelle Chromosomenmutationen

Bei den strukturellen Veränderungen können *Verluste* (Deletionen) oder *Einfügungen* (Insertionen) von *Teilen eines Chromosoms* vorkommen. Dies darf aber nicht mit den Deletionen und Insertionen von einzelnen Basen im Gen verwechselt werden. Weiters kommen Inversionen vor, bei denen ein Teil eines Chromosoms umgekehrt eingebaut wird, sowie Translokationen, bei denen abgebrochene Stücke eines Chromosoms an ein anderes angeheftet werden (Graw 2020, S. 497). Bei Duplikationen werden einzelne Bereiche eines Chromosoms mehrfach eingebaut.

5.3 Genmutationen

Ein Gen ist ein Abschnitt auf einem Chromosom, der für ein einzelnes Protein oder – im Falle von alternativem Splicing – für mehrere nahe verwandte Proteine codiert (siehe Kapitel 1.6). Laut Berg et al. (2018) besitzen Menschen zwischen 20 000 und 25 000 Gene. Es können Duplikationen, Deletionen, Insertionen oder Inversionen einzelner *Genabschnitte* auftreten.

Mutationen, die sich auf einzelne Gene beziehen, betreffen aber oft nur Änderungen in der Basensequenz, wobei einzelne Basen ersetzt (Substitution), entfernt (Deletion) oder eingefügt (Insertion) werden können. Wird nur *eine einzelne Base* in einem Gen verändert, so handelt es sich um eine *Punktmutation*. Solche Änderungen kommen in der DNA von Eukaryoten, Prokaryoten (Bakterien bzw. Archaeen) sowie in DNA bzw. RNA von Viren vor.

Im Folgenden werden verschiedene Arten von Punktmutationen in codierenden Bereichen von Genen (Exons) betrachtet.

5.3.1 Insertion

Bei einer Insertion wird eine zusätzliche Base eingefügt, wodurch sich das *Leseraster* um eine Base *verschiebt*. Nach der Einfügung dieser zusätzlichen Base, werden alle folgenden Tripletts verändert und damit auch alle folgenden Aminosäuren, die durch die Translation angefügt werden. Das Protein hat daher ab dem Punkt der Insertion eine andere Aminosäureabfolge, wie in Abbildung 5.1 dargestellt wird. Die Translation wird solange weitergeführt, bis durch das veränderte Leseraster ein *falsches* Stopp-Codon auftritt (vorzeitiges Stopp-Codon, PTC Premature Termination-Codon) und die Synthese abbricht. In der Mehrheit der Fälle wird das verkürzte Protein unbrauchbar oder zumindest funktionseingeschränkt sein.

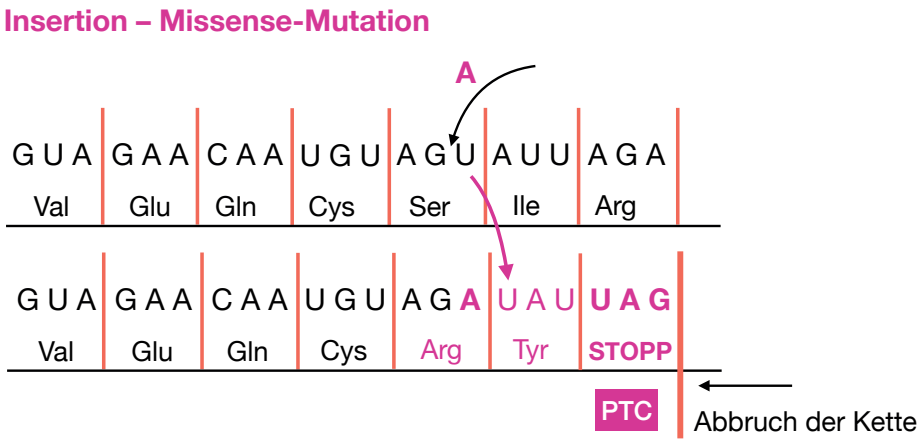


Abb. 5.1 Insertion einer Base mit Leserasterverschiebung und vorzeitigem Stopp-Codon (PTC).

Es besteht aber auch die Möglichkeit, dass bei der Insertion direkt ein vorzeitiges Stopp-Codon (PTC) entsteht und die Kette sofort abbricht (siehe Abbildung 5.2).

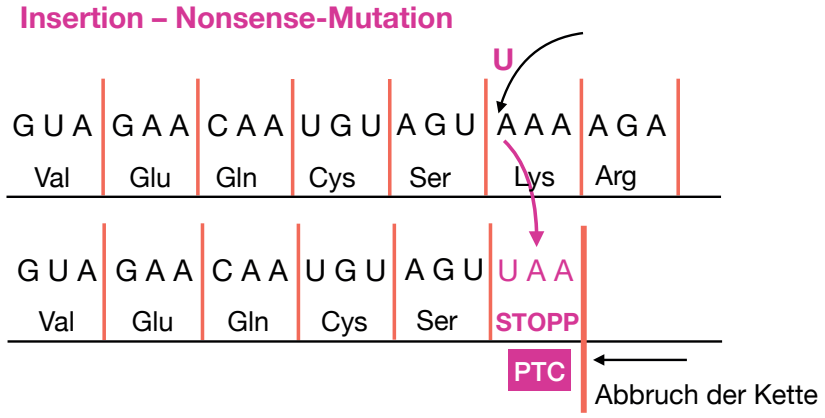


Abb. 5.2 Insertion einer Base mit direkter Ausbildung eines Stopp-Codons (PTC).

Bei der Insertion von zwei Basen ergibt sich ebenfalls eine Leserasterverschiebung, die in mehreren veränderten Aminosäuren und in der Regel in einem vorzeitigen Stopp-Codon endet.

5.3.2 Deletion

Bei einer Deletion geht durch eine Mutation eine Base verloren. Dadurch kann es direkt zur Ausbildung eines Stopp-Codons und einem Kettenabbruch kommen (siehe Abbildung 5.3).

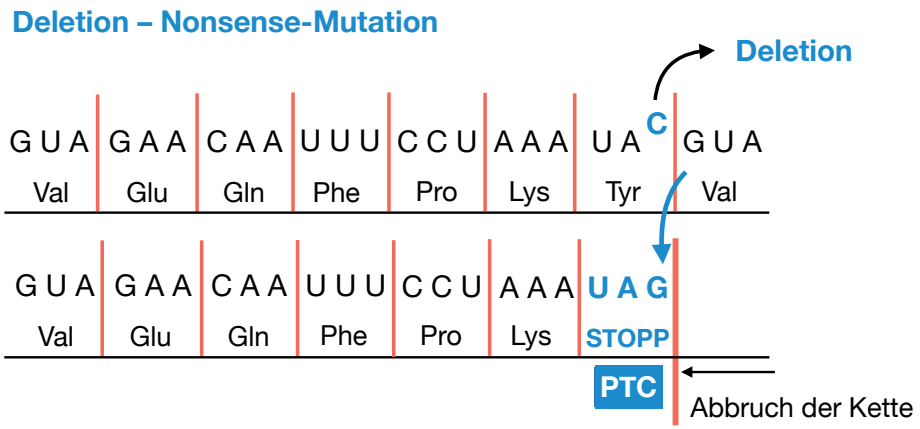


Abb. 5.3 Nonsense Mutation durch Deletion einer Base mit direkter Ausbildung eines Stopp-Codons (PTC).

Im zweiten Fall kommt es durch die entstehende Rasterverschiebung zu einigen veränderten Aminosäuren (veränderte Aminosäurenabfolge) gefolgt von einem vorzeitigen Stopp-Codon (PTC), an welchem die Translation abbricht (siehe Abbildung 5.4). Auch hier ist das entstehende Protein meist unbrauchbar oder zumindest funktionseingeschränkt.

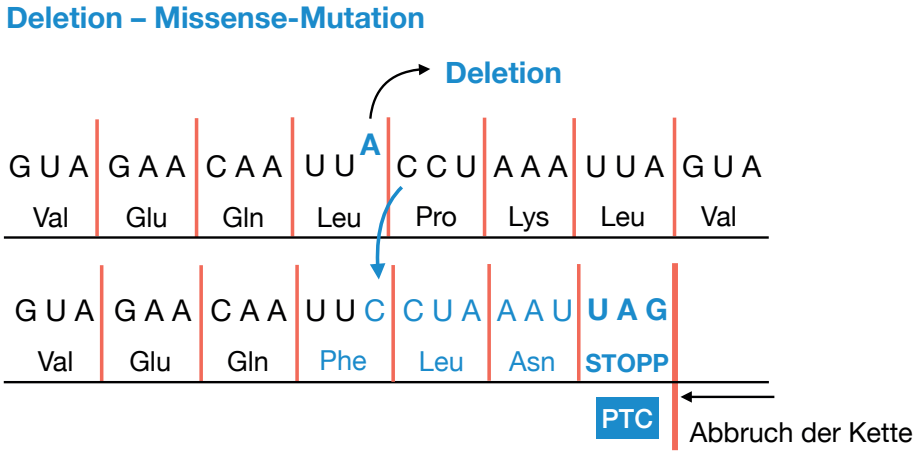


Abb. 5.4 Deletion einer Base mit Leserasterverschiebung und vorzeitigem Stopp-Codon (PTC).

Ausnahmen stellen hier Insertionen oder Deletionen *von drei Basen* oder einem Vielfachen davon dar. Werden drei Basen, oder ein Vielfaches davon, eingefügt (Insertion) oder entfernt (Deletion), so kommt es darauf an, wo die Einfügung/Entfernung stattfindet. Grundsätzlich wird das Leseraster zwar beibehalten, man muss hier jedoch unterscheiden, ob die Deletion bzw. die Insertion genau im Raster erfolgt oder nicht. Erfolgt die Insertion genau zwischen zwei Tripletts oder betrifft die Deletion jeweils ein oder mehrere ganze Tripletts, so verlängert oder verkürzt sich das Protein jeweils um eine oder mehrere Aminosäuren (Schaaf & Zschocke, 2018, S. 52; Wegner et al., 2016, S. 118). Erfolgt die Deletion jedoch so, dass drei nebeneinanderliegende Basen in zwei verschiedenen (nebeneinanderliegenden) Tripletts entfernt werden, so ergibt sich zusätzlich eine Verschiebung, wodurch nicht nur eine ganze Aminosäure entfernt wird, sondern es wird zusätzlich eine Aminosäure verändert (In-Frame-Insertion, DocCheck Flexikon; In-Frame-Deletion, DocCheck Flexikon).

Dasselbe gilt bei der Einfügung (Insertion) von drei aufeinanderfolgenden Basen (oder einem Vielfachen davon). Werden ein oder mehrere Tripletts (im Raster) eingefügt, besitzt das Protein eine oder mehrere zusätzliche Aminosäuren. Werden die drei Basen (oder ein Vielfaches davon) aber *nicht im Raster* eingefügt, wird das Protein zwar auch um eine oder mehrere Aminosäuren verlängert, es wird aber *zusätzlich eine Aminosäure verändert*.

5.3.3 Substitution

Wird bei einer Mutation eine Base durch eine andere *ersetzt*, so entsteht ein anderes Triplett, *ohne* dass eine *Leserasterverschiebung* auftritt. Es ergeben sich mehrere Möglichkeiten:

Das neue Codon mit der ausgetauschten Base codiert für eine *andere Aminosäure*. Diese Mutation wird als *Missense-Mutation* bezeichnet. Das entstehende Protein ist oft unbrauchbar oder funktionseingeschränkt (siehe Abbildung 5.5).

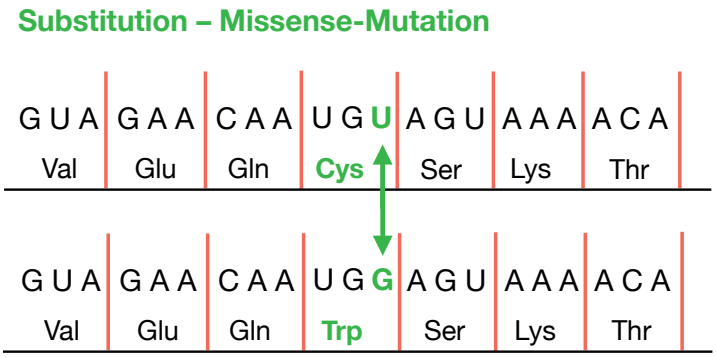


Abb. 5.5 Missense-Mutation durch Austausch einer Base, wobei sich eine andere Aminosäure ergibt.

Ein Beispiel für eine solche Missense-Mutation ist die *Sichelzellenanämie*. Hier gibt es in einer Untereinheit des Hämoglobins eine Punktmutation, die zu einem Basenaustausch von G**A**G nach G**T**G (Glu nach Val) führt (Schaaf and Zschocke, 2018, S. 287). Durch diese Mutation kann das Hämoglobin weniger Sauerstoff binden, da es auskristallisiert, wodurch sich die Erythrozyten sichelartig verbiegen und gegenseitig verhaken (Graw, 2020, S. 597). Tritt diese Mutation heterozygot auf, das heißt, sie wird nur von einem Elternteil vererbt, so ist die Auswirkung weniger dramatisch. Diese Mutation hat jedoch den Vorteil, dass sie gegen Malaria schützt, weil sich die Erreger im betroffenen Blut nicht so gut vermehren können. Somit haben heterozygote Personen in Gegenden, in denen Malaria endemisch auftritt, einen erheblichen Selektionsvorteil.

Handelt es sich dagegen bei dem neuen durch eine Substitution geänderten Codon um ein Stopp-Codon, so wird die Proteinsynthese an diesem vorzeitigen Stopp-Codon (PTC) abgebrochen. Diese Mutation wird als *Nonsense-Mutation* bezeichnet. Das entstehende Protein ist verkürzt und meistens unbrauchbar (siehe Abbildung 5.6). In den meisten Fällen ist das verkürzte Protein instabil (Kuzmiak & Maquat, 2006).

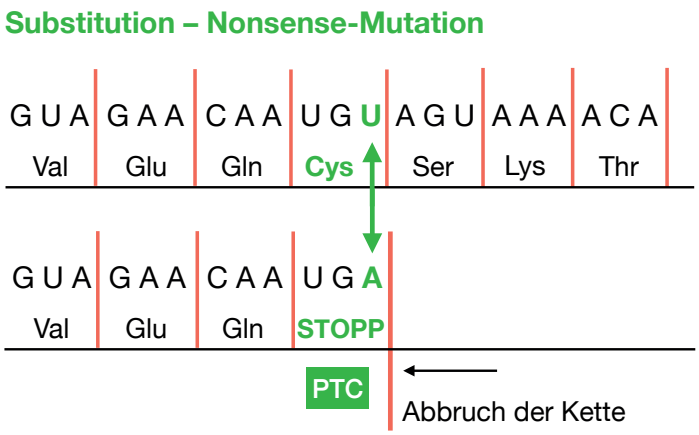


Abb. 5.6 Nonsense-Mutation durch direkte Umwandlung eines Codons in ein vorzeitiges Stopp-Codon (PTC).

Bei ca. 11 % aller genetischen Defekte tritt ein *vorzeitiges Stopp-Codon* auf, wodurch das entstehende Protein verkürzt wird. Dies tritt beispielsweise bei manchen Formen der Duchenne-Muskeldystrophie auf (Angulski et al., 2023). Die Gabe von Aminoglycosiden kann ein *Readthrough* (Überlesen des vorzeitigen Stopp-Codons) bewirken und so ein zumindest teilweise funktionsfähiges Protein wiederherstellen (Baradaran-Heravi et al., 2021). Allgemein werden Medikamente, die das Readthrough erhöhen, als TRIDs *(translational read-through inducing drugs)* bezeichnet. Dies soll hier aber nicht weiter behandelt werden.

Ist das neue Codon mit der substituierten Base unterschiedlich zum ursprünglichen Codon, codiert aber für *die selbe Aminosäure*, so wird diese Mutation als *synonyme Mutation* bezeichnet. Wie in Abbildung 5.7 dargestellt ist die Abfolge der Aminosäuren identisch.

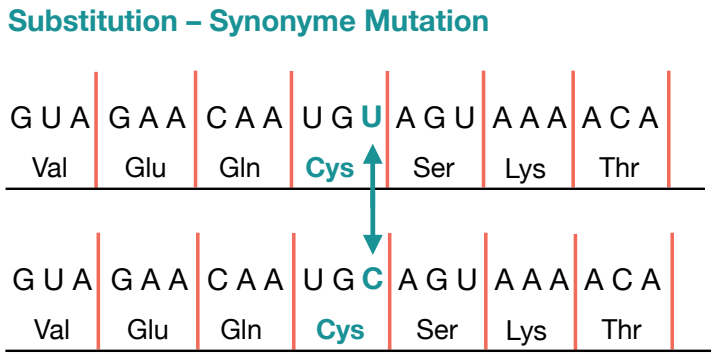


Abb. 5.7 Synonyme Mutation durch Substitution einer Base.

Bis vor wenigen Jahren wurden die synonymen Mutationen nicht beachtet und als harmlos angesehen. Sie wurden vielfach auch als *stille Mutationen (silent mutations)* bezeichnet. Es können jedoch aufgrund synonymer Mutationen fehlerhafte Faltungen, Über- oder Unterexpression von Genen sowie Spleißfehler auftreten, die zu Erbkrankheiten und Krebs führen können (siehe Kapitel 6). Eine besondere Art der Substitution ist jene, bei der ein Stopp-Codon durch ein synonymes Stopp-Codon ersetzt wird (siehe Abbildung 5.8).

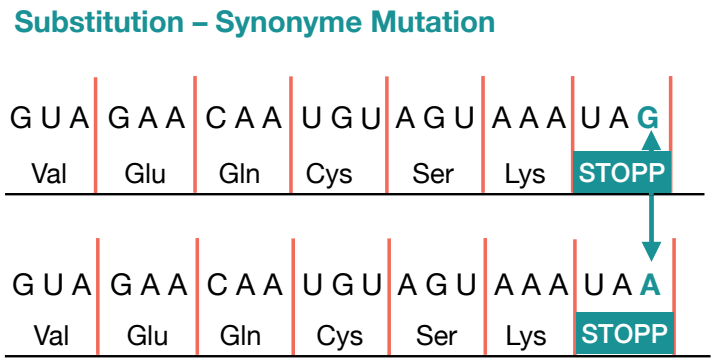


Abb. 5.8 Synonyme Mutation durch Substitution einer Base, wobei aus einem Stopp-Codon ein anderes Stopp-Codon entsteht.

Wird bei einer Missense-Substitution eine Aminosäure durch eine solche mit ähnlicher Seitenkette bzw. ähnlichen Eigenschaften ersetzt, so hat dies manchmal keine oder nur geringe Auswirkungen. Die Seitenketten der Aminosäuren können polar, unpolar/hydrophob, basisch oder sauer sein (siehe Anhang) – ebenso, wenn der Ort der Mutation unkritisch für das Funktionieren des Proteins ist. Eine solche Mutation wird als *neutrale Mutation* bezeichnet.

5.3.4 Nonstopp-Mutation

Eine Nonstopp-Mutation stellt keine eigene Art der Punktmutation dar, sondern sie kann bei allen drei Arten derselben, nämlich bei Deletionen, Insertionen und Substitutionen vorkommen.

Wenn die Deletion bzw. Insertion nahe dem 3‘-Ende vorkommen, muss nicht zwangsläufig ein vorzeitiges Stopp-Codon entstehen. Durch die Leserasterverschiebung wird das vorhandene Stopp-Codon nicht mehr erkannt und es wird danach weiter translatiert (Nonstopp-Mutation). Die nachfolgenden Basen der nichttranslatierten Bereiche und unter Umständen auch des Poly-A-Schwanzes werden dann als Sense-Codons interpretiert. Es gibt aber in dem 3‘- UTR untranslatierten Bereich zusätzliche *in-frame Stopp-Codons*, bis zu denen die Translation des Proteins weiter geführt werden kann. Falls diese aufgrund der Rasterverschiebung nicht erkannt werden, wird die Translation bis in den Poly-A-Schwanz weitergeführt (Shibata et al., 2015).

> **»When read-through of a stop codon occurs, the translation continues to the next in-frame stop codon or to the poly(A) tail.«**

In Abbildung 5.9 (Nonstopp-Mutation) wird exemplarisch die Variante einer Substitution im Stopp-Codon dargestellt. Hierbei wird in einem Stopp-Codon eine Base ersetzt (Substitution), sodass ein Codon für eine Aminosäure entsteht. Dadurch ist das Stopp-Codon quasi verschwunden und das Ribosom translatiert weiter. Nonstopp-Mutationen werden oft auch als Readthrough-Mutationen bezeichnet, obwohl streng genommen das Stopp-Codon nicht überlesen, sondern verändert wird.

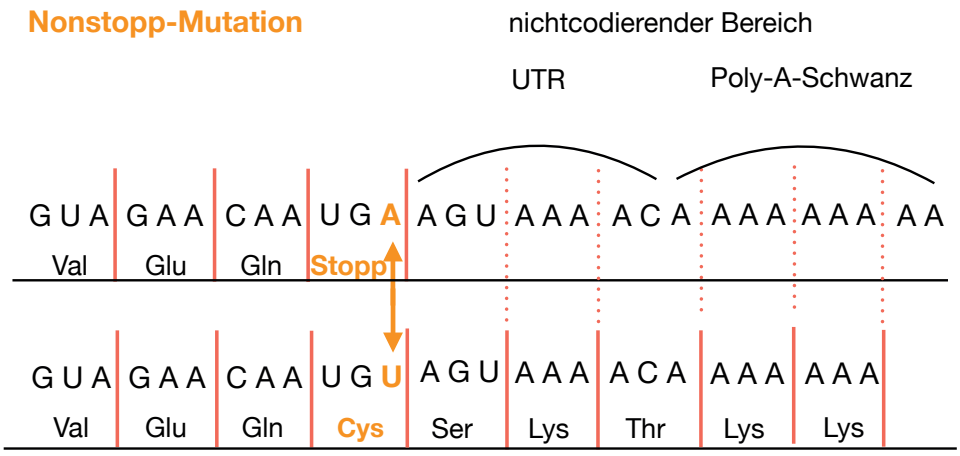


Abb. 5.9 Nonstopp-Mutation durch Substitution einer Base im regulären Stopp-Codon, wodurch dieses in ein Sense-Codon umgewandelt wird. Die Translation wird bis in den untranslatierten Bereich und unter Umständen bis in den Poly-A-Schwanz weitergeführt.

Zusammenfassend wird in Abbildung 5.10 eine Übersicht aller Punktmutationen dargestellt.

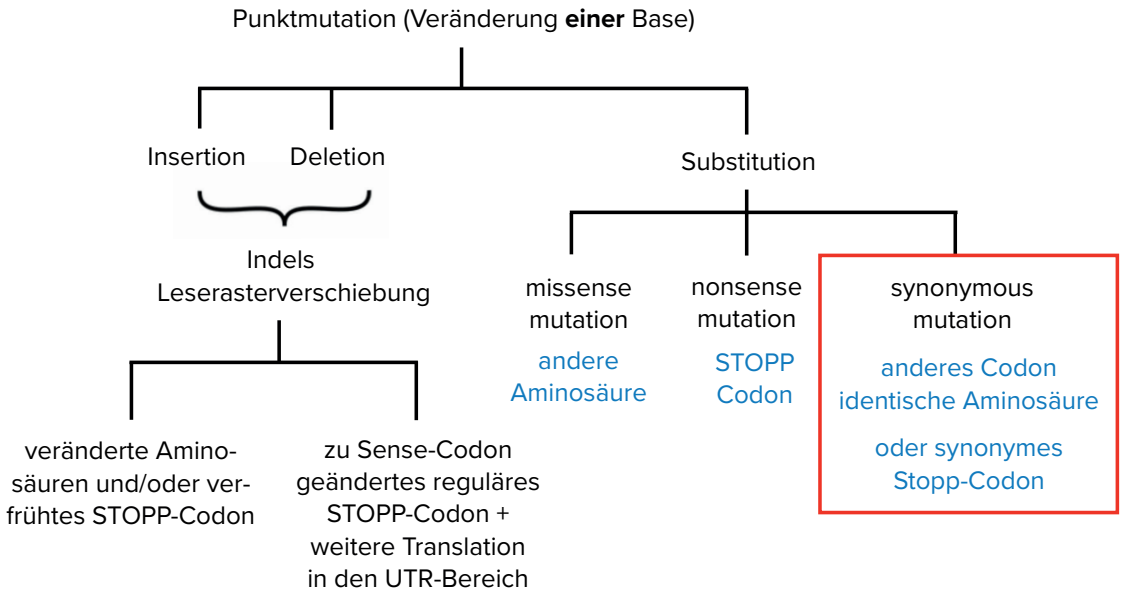


Abb. 5.10 Übersicht über die Arten der Punktmutationen.

Oft werden in der Literatur auch solche Mutationen, die aufgrund einer Deletion oder Insertion entstanden sind, wobei es durch die entstandene Rasterverschiebung in der Folge zu einem vorzeitigen Stopp-Codon kommt, als Nonsense-Mutationen bezeichnet.

5.4 Synonyme Mutationen

Bei synonymen Mutationen handelt es sich um solche Mutationen, bei denen das mutierte Triplett für dieselbe Aminosäure codiert, wie das ursprüngliche Triplett. Wie bereits in Abbildung 5.7 dargestellt, kann die Substitution einer einzelnen Base zu einer synonymen Mutation führen. Wie aus der Codonsonne in Abbildung 5.11 ersichtlich, wird dabei im Allgemeinen die dritte Base ausgetauscht (Wobble-Theorie, siehe Kapitel 3).

Vorzeitige Stopp-Codons können nur bei direkter Umwandlung eines Codons in ein Stopp-Codon (Nonsense-Mutation) oder bei Rasterverschiebungen, die durch Insertionen bzw. Deletionen hervorgerufen wurden, auftreten. Bei einer synonymen Mutation wird jedoch jeweils ein Codon für eine Aminosäure in ein Codon für *dieselbe Aminosäure* umgewandelt. Es kann dabei *in keinem Fall* ein vorzeitiges Stopp-Codon auftreten.

Die einzelnen Aminosäuren haben unterschiedlich viele synonyme Codons. Nur zwei Aminosäuren werden durch jeweils nur ein einziges Codon codiert, nämlich Methionin (Met), das auch als Start-Codon dient, und Tryptophan (Trp).

Neun Aminosäuren (Phe, Tyr, Cys, His, Gln, Asn, Lys, Asp, Glu) werden durch jeweils zwei unterschiedliche Codons codiert, wobei diese sich nur in der dritten Base unterscheiden. Der im Zuge einer synonymen Punktmutation erfolgende Austausch von einer Purinbase (A oder G) mit der anderen Purinbase (G oder A) oder von einer Pyrimidinbase (C oder T) mit der anderen Pyrimidinbase (T oder C) ist deshalb bevorzugt, da sich durch die jeweils nahezu identische Größe der jeweiligen Basen die Geometrie nur wenig verändert. Diese Art der Mutation wird als *Transition* bezeichnet.

Wird bei der Mutation eine Purinbase durch eine Pyrimidinbase ersetzt oder umgekehrt, so wird durch die sehr unterschiedlichen Größen der Basen die Geometrie der DNA verändert, was den Einbau erschwert. Diese Art der Mutation wird als *Transversion* bezeichnet.

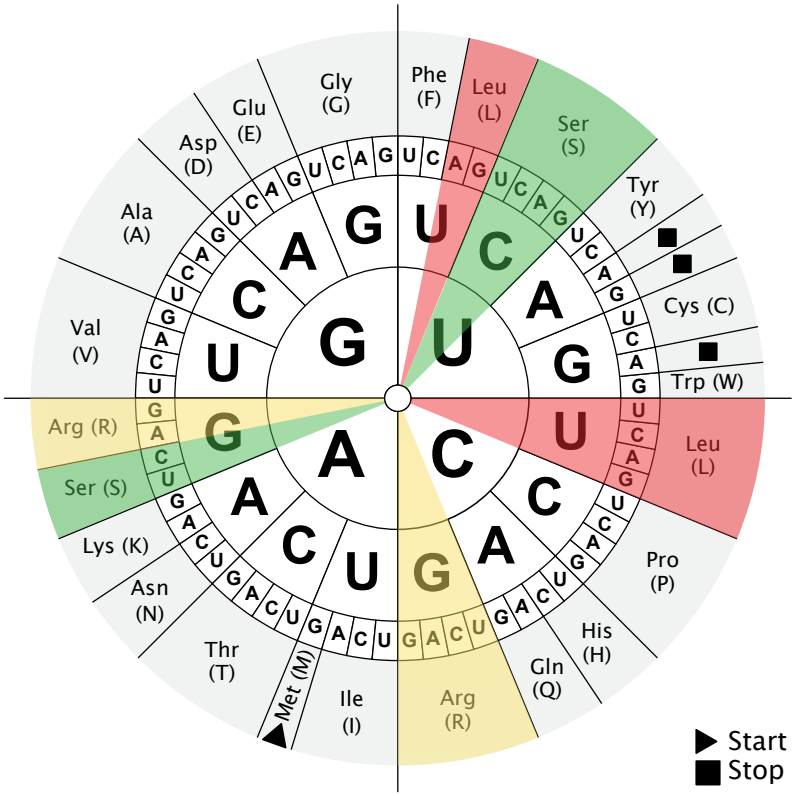


Abb. 5.11 Codonsonne mit Markierungen jener drei Aminosäuren, die 6 verschiedene Codons haben. (Arginin-Arg, Serin-Ser, Leucin-Leu), (Grundform Wikipedia, adaptiert).

Die am häufigsten verwendete Codonsonne ist so angeordnet, dass jeweils zwei Purinbasen (A, G) und zwei Pyrimidinbasen (U, C) in einem *Block* nebeneinanderliegen. Bei Aminosäuren, die nur zwei synonyme Codons haben, handelt es sich bei der dritten Base (im äußeren Ring) jeweils um die selbe Art von Basen, nämlich zwei Purinbasen (A, G) oder zwei Pyrimidinbasen (U, C).

Aminosäuren mit jeweils 4 verschiedenen synonymen Codons (Pro, Thr, Val, Ala, Gly) unterscheiden sich nur in der dritten Base, wobei diese dann, gemäß der *Wobble-Theorie*, beliebig ist (siehe Kapitel 3.5). Die Aminosäure Isoleucin (Ile) stellt insofern eine Ausnahme dar, als sie *drei* verschiedene synonyme Codons besitzt, die sich aber im selben Block befinden.

Alle diese synonymen Codons, die innerhalb eines Blocks liegen, zeichnen sich dadurch aus, dass sich diese *nur in der dritten Base* unterscheiden. Mutationen von einem der synonymen Codons zu einem anderen – sofern diese im selben Block liegen – erfordern also nur den Austausch (Substitution) einer einzigen Base, nämlich der dritten Base.

Drei Aminosäuren (Arginin, Serin und Leucin) besitzen je einen Zweierblock und einen Viererblock synonymer Codons (siehe Abbildung 5.11), also insgesamt je sechs verschiedene synonyme Codons. Hier gibt es daher außer dem bereits beschriebenen möglichen Austausch der dritten Base (Wobble-Theorie) innerhalb eines Blocks noch *blockübergreifende* Möglichkeiten synonymer Mutationen.

Im Folgenden werden die Möglichkeiten von Mutationen dargestellt, die den Übergang zwischen den synonymen Codons des Zweierblocks und des Viererblocks der jeweiligen Aminosäure zeigen. Die drei verschiedenen Stopp-Codons befinden sich ebenfalls in zwei verschiedenen Blöcken und sollen in diesem Zusammenhang wie die Codons einer Aminosäure behandelt werden. Zuerst werden alle möglichen blockübergreifenden Mutationen von *Leucin* und *Arginin* dargestellt und verglichen (siehe Abbildungen 5.12 und 5.13).

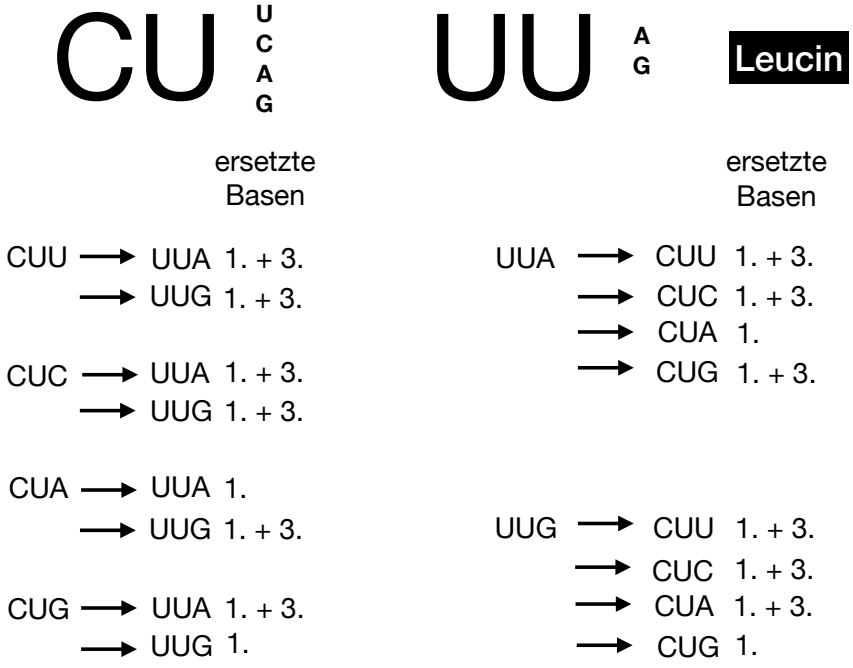


Abb. 5.12 Darstellung der möglichen blockübergreifenden synonymen Mutationen für die Aminosäure Leucin.

Bei den *blockübergreifenden* synonymen Codons bei *Leucin* fällt auf, dass nur vier von ihnen durch eine Punktmutation ineinander übergeführt werden können, wobei die erste Base ersetzt werden muss. Bei 12 der synonymen Mutationen bedarf es jeweils einer doppelten Mutation. wobei Substitutionen in der ersten und dritten Base notwendig sind.

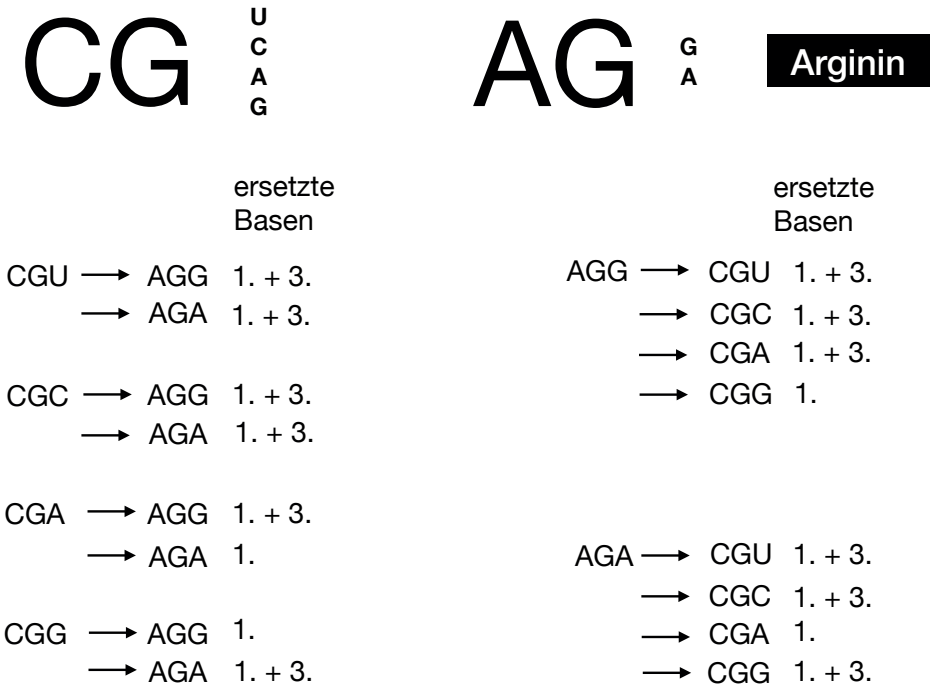


Abb. 5.13 Darstellung der möglichen blockübergreifenden synonymen Mutationen für die Aminosäure Arginin.

Bei Arginin ist die Situation analog zu jener bei Leucin. Auch hier gibt es vier mögliche Punktmutationen, wobei ebenfalls die erste Base ersetzt werden muss, sowie ebenfalls zwölf doppelte Mutationen, wobei auch die erste und die dritte Base ersetzt werden müssen. Bei Arginin und Leucin ist jeweils die mittlere Base in den beiden Blöcken identisch. Blockübergreifend gibt es Punktmutationen und Doppelmutationen. Wie in Abbildung 5.14 dargestellt, ergibt sich im Gegensatz zu den Aminosäuren Leucin und Arginin bei *Serin* eine etwas andere Situation.

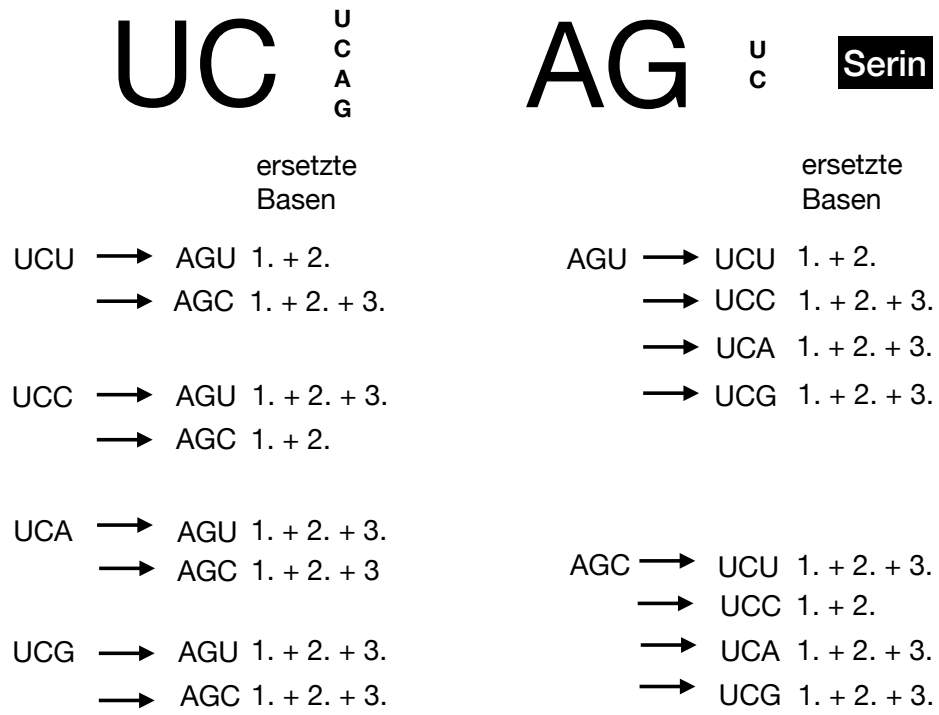


Abb. 5.14 Darstellung der möglichen blockübergreifenden synonymen Mutationen für die Aminosäure Serin.

Während bei Arginin und Leucin jeweils vier blockübergreifende Punktmutationen und zwölf blockübergreifende Doppelmutationen vorkommen können, sind bei Serin *keine blockübergreifenden Punktmutationen* möglich, da die erste und die zweite Base in den beiden Blöcken unterschiedlich sind, wie auch Inouye et al. (2020) feststellen. Von den insgesamt 16 möglichen synonymen Mutation sind vier Doppelmutationen und zwölf Dreifachmutationen.

Abbildung 5.15 zeigt eine mögliche blockübergreifende Doppelmutation direkt und über eine *Zwischenstufe* am Beispiel von Leucin. Sofern die beiden Basen nacheinander ersetzt werden, ergeben sich zwei verschiedene Wege, wobei das eine Zwischenprodukt eine weitere synonyme Variante von Leucin ist, das andere Zwischenprodukt für den alternativen Weg aber Phenylalanin. Bei Dreifachmutationen ergeben sich mehrere mögliche Wege, da hier zwei Zwischenstufen erforderlich sind.

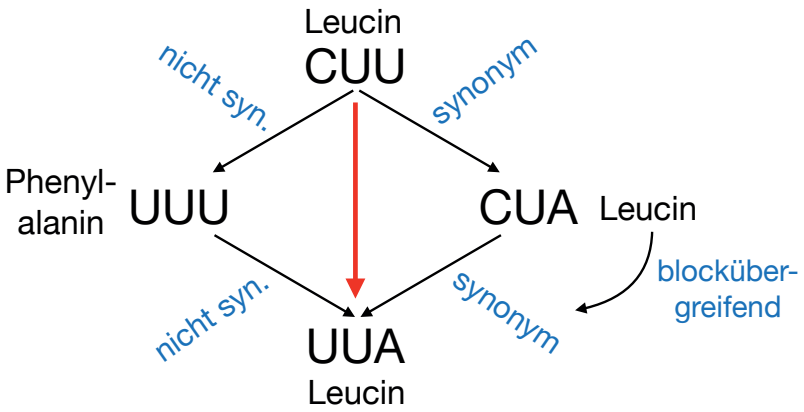


Abb. 5.15 Zwei Wege einer blockübergreifenden Doppelmutation von Leucin CUU zu Leucin UUA über zwei Punktmutationen (Belinky et al., 2019, adaptiert).

In Abbildung 5.16 werden die möglichen synonymen Mutationen zwischen den drei Stopp-Codons UAA, UAG und UGA dargestellt. Viele Autor*innen sehen die Übergänge zwischen den verschiedenen Stopp-Codons jedoch nicht als Mutationen an. Es wurden hier nur direkte synonyme Mutationen berücksichtigt.

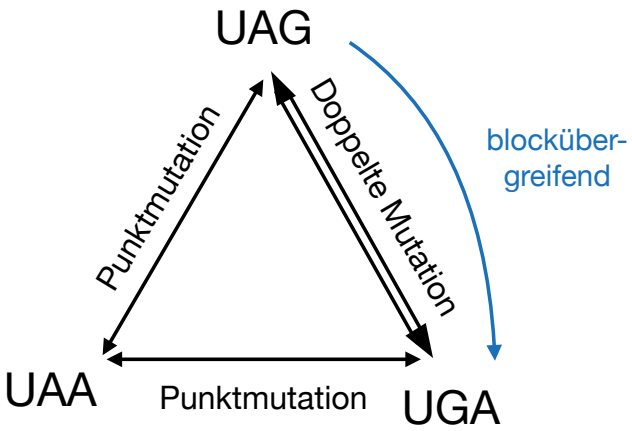


Abb. 5.16 Synonyme Mutationen zwischen allen drei Stopp-Codons.

Es sind auch Umwege über mehrere Mutationen und über andere Codons für Aminosäuren denkbar. Zwei einzelne Mutationen UAG → UGA müssen über eine Zwischenstufe führen, die entweder ein Sense-Codon für eine Aminosäure darstellt oder über das verbleibende UAA-Stopp-Codon. Bei einem Sense-Codon in der Zwischenstufe, würde das Stopp-Codon zwischenzeitlich verschwinden und die Translation in den 3'-UTR-Bereich fortgeführt werden.

6 Auswirkungen synonymer Codons auf zelluläre Prozesse

»the sound of silence«

Für die 20 kanonischen Aminosäuren stehen 61 unterschiedliche Codons zur Verfügung. Abgesehen von Methionin und Tryptophan mit jeweils nur einem einzigen Codon, gibt es für alle anderen Aminosäuren zwischen zwei und sechs unterschiedliche Codons (siehe Kapitel 3 und 5). Zwei Codons werden *synonym* genannt, wenn sie für die selbe Aminosäure codieren. Wird ein Codon durch eine Mutation in ein synonymes Codon umgewandelt, so wird diese Mutation als *synonyme Mutation* bezeichnet.

Die ursprüngliche Bezeichnung für synonyme Mutationen als *stille bzw. stumme Mutationen* rührte daher, dass diese als harmlos und daher als nicht berücksichtigenswert eingestuft wurden, da sie keine Änderung der Reihenfolge der Aminosäuren im Protein hervorrufen. Es wurde daher auch angenommen, dass sie *keinerlei Auswirkungen auf die Funktionsfähigkeit des Proteins haben*, weder positive noch negative. Im Englischen werden diese oft auch als *silent mutations* bezeichnet. Bei der Bezeichnung *sound of silence* wurde von vielen Autor*innen der Molekularbiologie der Titel des Songs *Sound of Silence* von Simon and Garfunkel aufgegriffen, den diese 1966 komponierten (Wikipedia).

Trotz der verbreiteten Ansicht, dass stille bzw. synonyme Mutationen nicht wichtig seien, mahnten Richard & Beckmann bereits 1995 zur Vorsicht, was die angebliche Wirkungslosigkeit synonymer Mutationen betrifft und wiesen darauf hin, dass eine von ihnen identifizierte synonyme Mutation Gly ➔ Gly (GGC Glycin ➔ GGT Glycin) bei einem Patienten mit einer Muskeldystrophie zu einer neuen 5‘-Spleißstelle (Donor-Spleißstelle) führt, somit die Deletion der letzten 44 Basen in Exon 16 hervorruft und damit Krankheitswert hat.

> **Our demonstration of the pathogenic character of a mutation substituting one glycine codon for another *underscores the need to interpret such apparently ‚neutral' mutations with caution*. Until one assesses their impact on mRNA processing and stability or even codon usage bias, it may be difficult to ascertain the neutrality or non-pathogenicity of apparent polymorphisms. (Richard & Beckmann, 1995)**

Neuere Untersuchungen bestätigen, dass synonyme Mutationen sehr oft dramatische Auswirkungen haben können und daher gar nicht still sind. Ursprünglich erfolgte die Einteilung von Krankheiten bzw. von Erbkrankheiten ausschließlich aufgrund klinischer Symptome. Seit es die Möglichkeit gibt, ganze Gene oder auch Genome kostengünstig zu sequenzieren, konnte festgestellt werden, dass klinisch identische oder sehr ähnlich aussehende Krankheiten durch sehr unterschiedliche Mutationen – mitunter auch in verschiedenen Genen oder auch auf verschiedenen Chromosomen – hervorgerufen werden können. Bei vielen bekannten genetischen Defekten wurden seither zusätzliche, bisher unbekannte Mutationen entdeckt. Darunter befinden sich oft auch bisher vernachlässigte synonyme Mutationen.

Im Folgenden werden nur synonyme Mutationen in codierenden Bereichen von Genen (Exons) betrachtet. Diese weisen zwar eine Änderung in der Basensequenz auf, sie rufen aber *keine Änderung der Aminosäurenabfolge* hervor. Introns sind *nichtcodierende* Bereiche von Genen, in denen Mutationen jedoch auch Auswirkungen – beispielsweise auf Spleißstellen – haben können. Diese sollen hier aber nicht Gegenstand der Betrachtung sein. Es werden daher *exemplarisch* einige synonyme Mutationen und deren Auswirkungen auf molekulargenetischer Ebene sowie deren klinische Auswirkungen dargestellt.

6.1 Epigenetik

Epigenetik bezeichnet „stabile Veränderungen der Regulation der Genexpression, die über Zellteilungen hinweg aufrechterhalten, also vererbt werden, ohne dass dabei die DNA-Sequenz verändert wird" (Graw, 2020, S. 415). Zum Unterschied von allen anderen Beispielen in diesem Kapitel, bei denen es sich um Prozesse auf der Ebene der RNA-Transkripte handelt, geht es bei der Epigenetik um die Auswirkungen synonymer Mutationen auf DNA-Ebene. Auch hier zeigt es sich, dass die Redundanz des Codes nichts Entbehrliches beinhaltet, sondern sie ermöglicht, in der Nucleotidsequenz zusätzlich zur reinen Bauplaninformation weitere Instruktionen unterzubringen, die z. B. der Regulation der Transkription dienen.

Die ursprüngliche Annahme war, dass die Redundanz lediglich dafür sorgt, dass aufgrund der Wobble-Theorie Mutationen in der dritten Base nicht ins Gewicht fallen, da die synonymen Codons für die selbe Aminosäure codieren. Es stellte sich jedoch heraus, dass aufgrund der Redundanz des Codes ein zusätzliches Informationsreservoir zur Verfügung steht, welches vor allem in der Epigenetik und bei der Faltung von Proteinen ausgenutzt wird.

Da nicht jedes Protein in jedem Gewebe benötigt wird bzw. manche Enzyme nur zu manchen Zeiten nachgeliefert werden müssen, ist es nötig, dass die entsprechenden Gene – je nach Bedarf – aus- bzw. eingeschaltet werden können. Dies kann beispielsweise durch Anlagern von Methylgruppen an Cytosin in der DNA erfolgen – oder durch Methylierung, Phosphorylierung oder Acetylierung von Histonen (Graw, 2020, S. 269). Die Cytosinmethylierung als Genexpressionsregulation findet nicht in allen Spezies statt. Manche Spezies, wie der Mensch, die Hausmaus (Mus musculus) und der Zebrafisch (Danio rerio) verwenden die Cytosinmethylierung zur Regulation der Genexpression, andere, wie beispielsweise E. coli und die Fruchtfliege (Drosophila melanogaster) nicht (Branciamore et al., 2010).

Hier soll nur die Methylierung von Cytosin in CpG-Dinucleotiden der DNA betrachtet werden, da nur in diesem Fall synonyme Codons eine Rolle spielen.

Der genetische Code der DNA ist die unterste Ebene der Codierung und stellt eine Abbildungsvorschrift von der DNA zum Protein dar. Der epigenetische Code, der das Ein- bzw. Ausschalten von Genen reguliert, stellt eine weitere, dem genetischen Code überlagerte Codierung dar, die allerdings in die Regulierung eingreift und daher eher als Durchführungsvorschrift zu verstehen ist, die besagt, ob die Transkription ausgeführt werden soll oder nicht. Bei der Methylierung von Cytosin (methyliert ja/nein) handelt es sich im Gegensatz zum DNA-Code um einen *binären Code* (Bierhoff, 2024).

Die Überlagerung der beiden Codes ist nur möglich, weil der genetische Code redundant ist und somit geeignete synonyme Codons *ausgewählt und verwendet* werden können, die es erlauben, den binären Methylierungscode zu überlagern, ohne die Aminosäuresequenz des Proteins zu ändern.

6.1.1 Anordnung von Dinucleotiden

Da es in der DNA genau vier verschiedene Basen (A, T, G, C) gibt, ergeben sich 16 verschiedene Möglichkeiten zur Anordnung von Dinucleotiden. Dinucleotide sind nebeneinanderliegende Basen auf einem DNA-Strang und dürfen nicht mit der Basenpaarung aus komplementären Strängen verwechselt werden.

Nimmt man eine Gleichverteilung an, so würde im menschlichen Genom alle Dinucleotide, also auch CpG-Dinucleotide etwa 6,3% (1/16) ausmachen. Bei manchen Organismen sind diese Dinucleotide über die gesamte DNA statistisch gleich verteilt, wie z. B. bei E. coli, beim Menschen gilt diese Gleichverteilung aber nicht.

In einer bestimmten DNA-Sequenz – z. B. dem Genom von E. coli oder dem Genom des Menschen – werden die Positionen aller CpG-Nucleotide lokalisiert. Die Suche erfolgt nach dem Sliding-Window-Prinzip. Es ist anzunehmen, dass diese Dinucleotide ursprünglich ungefähr gleich häufig auftraten wie alle anderen. CG-Dinucleotide (meist als CpG-Dinucleotide – Cytosin-phos-

phatidyl-Guanosin – bezeichnet) finden sich im menschlichen Genom statistisch weniger häufig als alle anderen Dinucleotide.

Eine mögliche epigenetische Markierung der DNA besteht in der Methylierung von Cytosin in der 5'-Position zu Methyl-Cytosin (^{m}C, ^{met}C, m^5C), wobei die Methylgruppen (–CH3) durch Methyltransferasen übertragen werden. Cytosin wird bevorzugt dann methyliert, wenn danach ein Guanin folgt. Die Methylierung ist prinzipiell durch Demethylasen wieder rückgängig zu machen. Es handelt sich also hierbei nicht um eine Mutation, die in der DNA festgeschrieben wird, sondern um eine *reversible Strukturänderung*. Diese Änderung (Methylierung) kann beispielsweise durch toxische Substanzen, Stress, Hunger oder auch bestimmte Ernährung hervorgerufen werden. In manchen Fällen kann diese Struktur auch von einer Generation zur nächsten weitergegeben werden (Entringer et al., 2016).

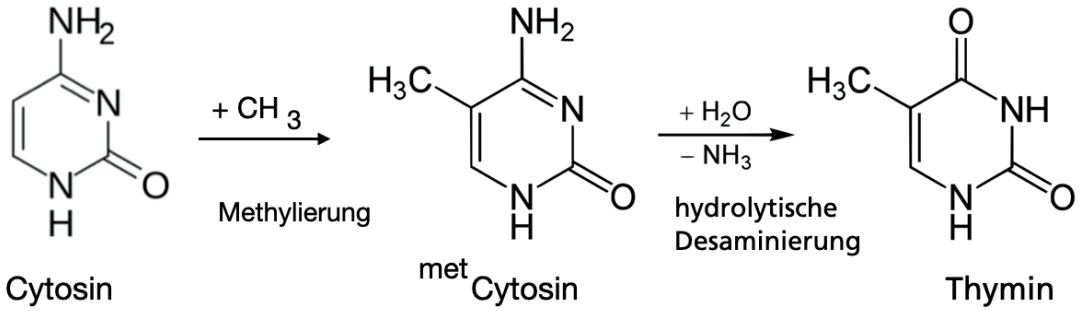


Abb. 6.1 Umwandlung von Cytosin durch Methylierung in Methyl-Cytosin (^{met}C) und danach durch hydrolytische Desaminierung in Thymin (Yikrazuul Wikipedia, gemeinfrei, Beschriftung verändert).

Methyliertes Cytosin wird häufig durch spontane hydrolytische Desaminierung in Thymin umgewandelt (siehe Abbildung 6.1). Wird diese Mutation von CpG zu TpG nicht von etwaigen Reparatursystemen korrigiert, kommt es im Laufe der Evolution zu der beobachteten verminderten Häufigkeit von CpG-Dinucleotiden.

Hütt & Dehnert (2016) haben die DNA-Sequenz der menschlichen Thymidylat-Synthase, die bei der DNA-Reparatur eine Rolle spielt, auf Häufigkeiten von Codons, die CG enthalten, untersucht.

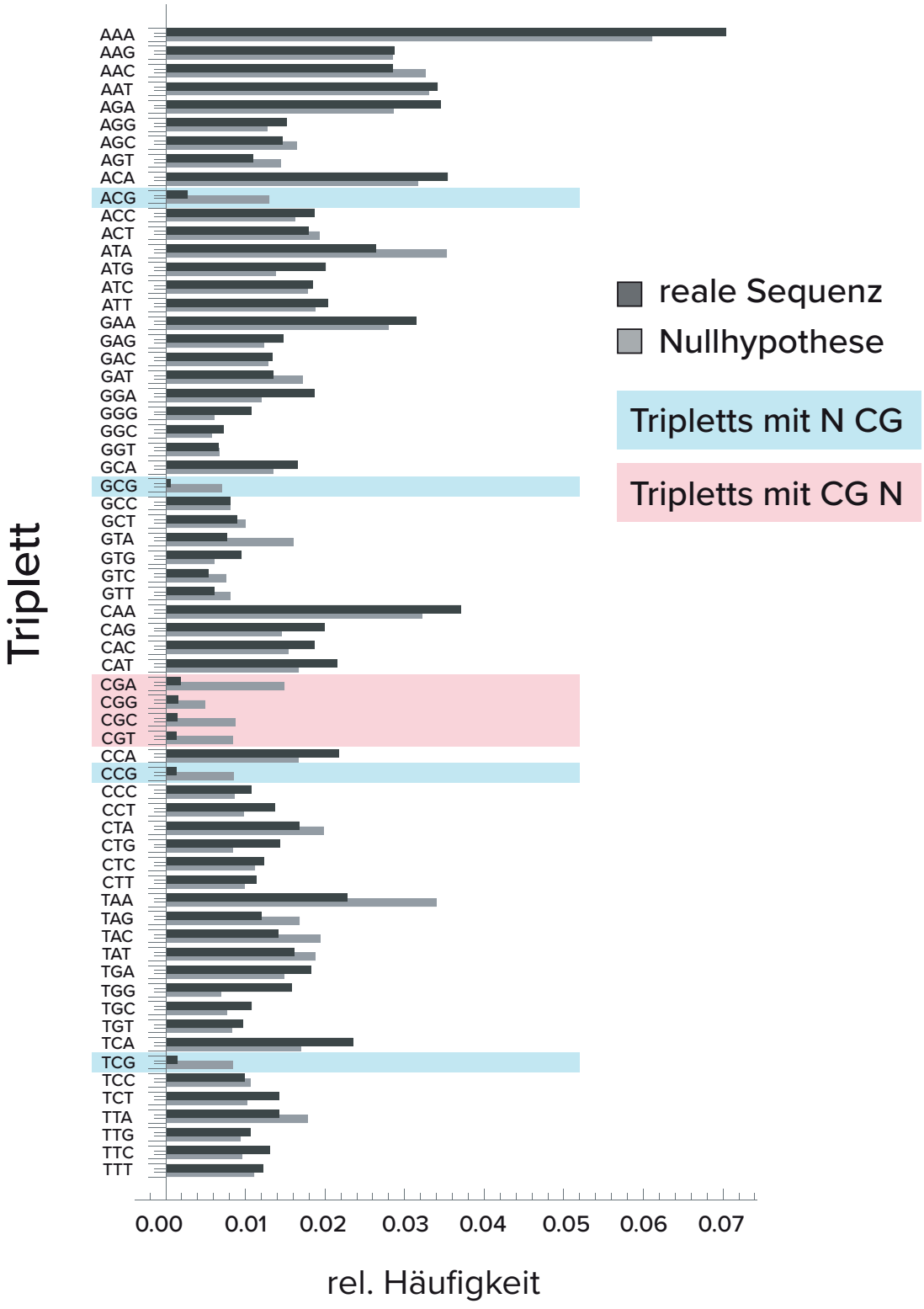


Abb. 6.2 Analyse der Tripletthäufigkeiten der menschlichen Thymidylat-Synthase (Hütt & Dehnert, 2016, S. 95, Abb. 2.34b, adaptiert).

In Abbildung 6.3 zeigt sich deutlich, dass CpG-Dinucleotide stark unterrepräsentiert sind. In Abbildung 6.2 ist zu beobachten, dass Codons mit CGN (rosa unterlegt) bzw. NCG (türkis unterlegt) weniger häufig vorkommen. CG-Sequenzen, welche codonübergreifend sind (NNC – GNN), wurden bei dieser Untersuchung nicht berücksichtigt. N kann laut IUPAC A, T, C oder G bedeuten (siehe Anhang: Bezeichnung der Nucleobasen laut IUPAC).

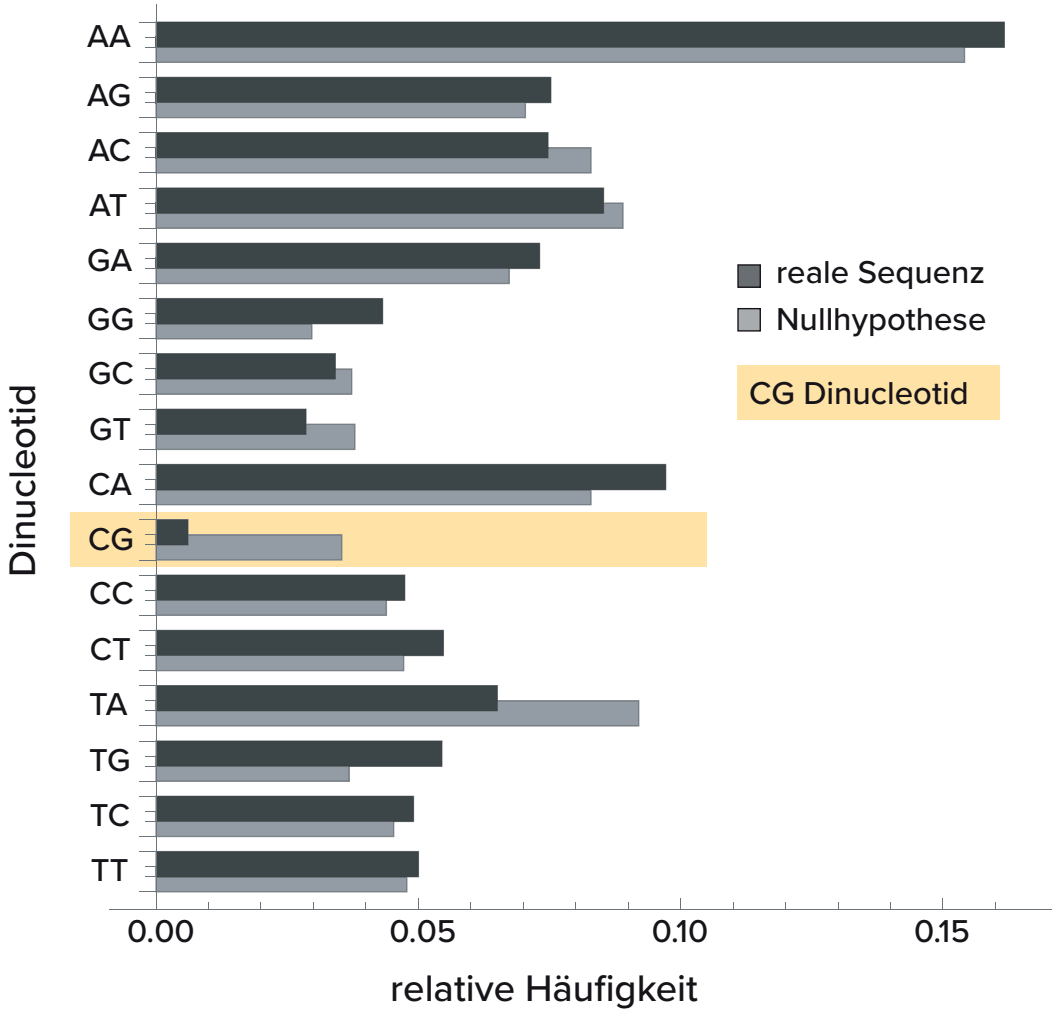


Abb. 6.3 Analyse der Dinucleotidhäufigkeiten der menschlichen Thymidylat-Synthase (Abb. 2.34a aus Hütt & Dehnert,2016, S. 95, adaptiert).

CpG-Dinucleotide kommen also statistisch gesehen weniger häufig vor als andere Dinucleotide, was auf die Desaminierung von ^{met}C zurückzuführen ist. Auffallend ist hier, dass die Häufigkeit der CpG-Dinucleotide wesentlich geringer ist als die Häufigkeit der GC-Dinucleotide, obwohl die Nullhypothese den selben Wert für beide ergibt. Aus der extrem geringen Anzahl der CpG-Dinucleotide lässt sich vermuten, dass diese durch Methylierung des Cytosins mit anschließender Desaminierung zumindest teilweise verschwunden sind. Untersuchungen zeigen jedoch, dass die *verbleibenden* CpG-Dinucleotide beim Menschen sehr ungleich verteilt sind. Takai & Jones (2002) kommen bei ihrer Untersuchung von den Chromosomen 21 bzw. 22 des Menschen ebenfalls auf eine auffallend niedrige CpG-Anzahl. Bei E. coli ist die Verteilung der

Dinucleotide allerdings relativ gleichförmig, was auf die Abwesenheit von CpG-Methylierung bei E. coli zurückzuführen ist.

Für Methyl-Cytosin gibt es zwei gegenläufige Effekte:

1) Es reguliert die Genexpression und ist daher wichtig zum Ein- und Ausschalten eines Gens.
2) Es ist mutationsanfällig (^{met}C → T) und verschwindet, wodurch die Genregulation gestört werden kann.

Lander et al. (2001) gingen von einem gemessenen GC-Gehalt von 42 % aus, der unter dem bei Gleichverteilung anzunehmenden Wert von 50 % liegt. Aus dem gemessenen Wert von 42 % ergibt sich jeweils ein Gehalt von 21 % (0,21) für G und für C. Der Gehalt von CpG-Dinucleotiden beträgt daher 0,21 x 0,21 = 0,04 (4 %). Es müssten sich also 4 % CpG-Dinucleotide in einem Gen finden lassen. Die Anzahl der experimentell ermittelten CpG-Dinucleotide ergab aber nur einen Wert von ca. einem Fünftel dieser berechneten Anzahl, was auf die Desaminierung des ^{met}C und die Umwandlung in T in den CpG-Dinucleotiden zurückzuführen ist. Aus den CpG-Dinucleotiden wurden also durch Desaminierung CT-Dinucleotide. Die meisten CpG-Dinucleotide waren also durch diese Umwandlung verschwunden und der GC-Gehalt wurde durch diese Umwandlung ebenfalls gesenkt.

6.1.2 CpG-Inseln

Obwohl aufgrund der ^{met}C → T-Mutation in der DNA CpG-Dinucleotide weniger häufig vorhanden sind als andere Dinucleotide, ergeben genauere Untersuchungen eine statistisch signifikante *Ungleichverteilung* der verbleibenden CpG-Dinucleotide.

Sowohl Gardiner-Garden & Frommer (1987) als auch Takai & Jones (2022) wiesen nach, dass CpG-Inseln am 5‘-Ende eines Gens gehäuft vorkommen. Die Methylierung zu Beginn eines Gens bestimmt, ob dieses abgelesen werden kann oder ob es stillgelegt wird, weil das hergestellte Protein in einem bestimmten Gewebe beispielsweise nicht benötigt wird und daher nicht exprimiert werden soll.

Die statistische Suche nach CpG-Inseln wird oft dazu verwendet, einzugrenzen, wo sich ein Gen befindet bzw. wo ein solches beginnt (Hütt & Dehnert, 2016, S. 94). CpG-Inseln stellen hochkonservierte Bereiche dar, die wichtige Funktionen ausüben und in welchem ^{met}C → T-Mutationen unterdrückt sind (Hütt & Dehnert, 2016, S. 95). Kurreck et al. (2022, S. 989) gehen davon aus, dass sich im menschlichen Genom ca. 30 Millionen CpGs befinden, die in Zukunft analysiert werden können.

6.1.3 Mutationen von Cytosin in CpG-Dinucleotiden

Der Wiener Kinderarzt Andreas Rett beschrieb 1966 eine neurologische Entwicklungsstörung bei Mädchen, bei der die betroffenen Kinder nach wenigen Monaten bereits erworbene Fähigkeiten wieder verlernen. Es folgen stereotype Handbwegungen und Verlust der Sprache sowie vermindertes Schädelwachstum (Graw, 2020, S. 866).

Adrian Bird, ein britischer Biochemiker und Molekularpathologe, untersuchte epigenetische Prozesse, zu denen auch die DNA-Methylierung gehört und entdeckte dabei die CpG-Inseln. Er arbeitete zeitweilig auch am *Research Institute of Molecular Pathology* in Wien. 1999 entschlüsselte er jene Mutationenen im MeCP2- Protein, das spezifisch an methyliertes Cytosin in CpG-Inseln bindet, die das Rett-Syndrom verursachen. Hierdurch wurde ersichtlich, wie wichtig korrekte Methylierung für die Entwicklung neuronaler Prozesse ist (Research Institute of Molecular Pathology, online; Amir et al., 1999).

Das Gen für das MeCP2-Protein befindet sich am langen Arm des X-Chromosoms. Bis auf wenige Ausnahmen sind nur Frauen vom Rett-Syndrom betroffen. Durch die statistische Stilllegung jeweils eines X-Chromosoms bei Frauen steht meist ein Teil des nicht betroffenen Gens zur Verfügung. Bei männlichen Embryonen steht – mangels eines zweiten X-Chromosoms – kein nicht betroffenes Gen zur Verfügung, sodass diese bereits in einem frühen Stadium der Entwicklung sterben.

Seit 2024 liegt der europäischen Arzneimittelbehörde EMA eine Anmeldung für ein Medikament zur Behandlung des Rett-Syndroms vor (EMA online):

> **Adeno-associated viral vector serotype 9 containing the human MECP2 gene, an intron encoding a miRNA generating sequence, and complementary miRNA binding sites**

Wird Cytosin methyliert um ein Gen abzuschalten, so bildet sich Methylcytosin (^{met}C), das durch hydrolytische Desaminierung zu Thymin mutieren kann (siehe Abbildung 6.1). Es ist jedoch wichtig, dass Gene – je nach Erfordernis – ein- bzw. ausgeschaltet werden können. Würde in den Promotorbereichen jedes Cytosin eines CpG-Dinucleotids zu Thymin mutieren, so ginge diese Möglichkeit der Regulation des Gens verloren.

Branciamore et al. (2010) untersuchten Mutationen in CpG-Inseln und deren Einfluss in proteincodierenden Bereichen sowie in Hox-Genen, indem sie das methylierte Genom von Homo sapiens, der Hausmaus (Mus musculus), dem Zebrafisch (Danio rerio) und Drosophila melanogaster verglichen.

Hox-Gene codieren für Transkriptionsionsfaktoren, die die Identität der Segmente der kranio-kaudalen Körperachse (von Kopf bis Fuß) determinieren; das heißt, jedem Segment werden eine bestimmte Funktion und bestimmte Organe zugewiesen. Beim Menschen gibt es 39 Hox-Gene, die in 4 Clustern organisiert sind. Bei der Fruchtfliege gibt es 11 Hox-Gene (Branciamore et al., 2010). Die zeitlich und räumlich genau aufeinander abgestimmte Expression einzelner Hox-Gene, die letztlich durch die Methylierung der CpG-Inseln reguliert wird, sorgt für die korrekte Reihenfolge und Identität der Segmente – beim Menschen die Abfolge von Kopf, Hals, Armen und Wirbeln bis hin zu den Beinen. Bei der Fruchtfliege sorgen die Hox-Gene für die korrekte Reihenfolge von Kopf, Antennen und Beinen. Bei diesen ist eine Störung bekannt, die statt der Antennen Beine ausbildet (Antennapedia; Graw, 2020, S. 496). Beim Menschen wird eine Art der Polydaktylie (Mehrfingerigkeit) durch eine Mutation in einem Hox-Gen verursacht (Max Planck Institut für molekulare Genetik, online).

Branciamore et al. (2010) untersuchten CpG-Dinucleotide in den vierfach degenerierten Codons – also jene, die bei Mutationen in der dritten Base (Wobble-Position) nicht zu einer Änderung der Aminosäuresequenz führen, die aber die epigenetische Markierung zerstören könnten.

Es ergeben sich drei theoretische Möglichkeiten für Mutationen von CpG-Dinucleotiden, in denen das Cytosin zu Methyl-Cytosin methyliert wird (N ist gemäß IUPAC eine beliebige Base).

1) **CG**N: Das CpG-Dinucleotid befindet sich in der *ersten und zweiten* Stelle eines Codons.
2) N**CG**: Das CpG-Dinucleotid befindet sich in der *zweiten und dritten* Stelle eines Codons.
3) NN**C – G**NN: Das CpG-Dinucleotid ist *codonübergreifend*.

Im DNA-Doppelstrang ist die Sequenz CG palindromisch, das bedeutet, dass durch die komplementäre Paarung von C und G die Reihenfolge trotz umgekehrter Leserichtung von nicht-codogenem Strang und codogenem Strang gewahrt bleibt (siehe Abbildung 6.4).

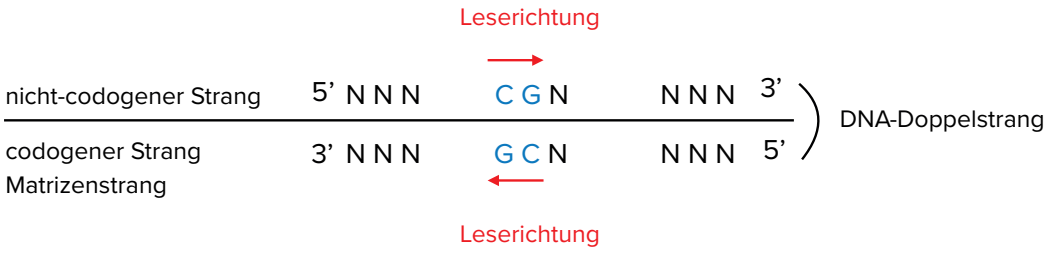


Abb. 6.4 Darstellung der palindromischen Sequenz CG, die in beiden Leserichtungen identisch ist.

Fall 1: **CG**N. Die CpG-Sequenz kommt in der ersten und zweiten Stelle eines Codons vor. Abbildung 6.5 zeigt die möglichen Mutationen von methyliertem Cytosin im nicht-codogenen Strang und im codogenen Strang.

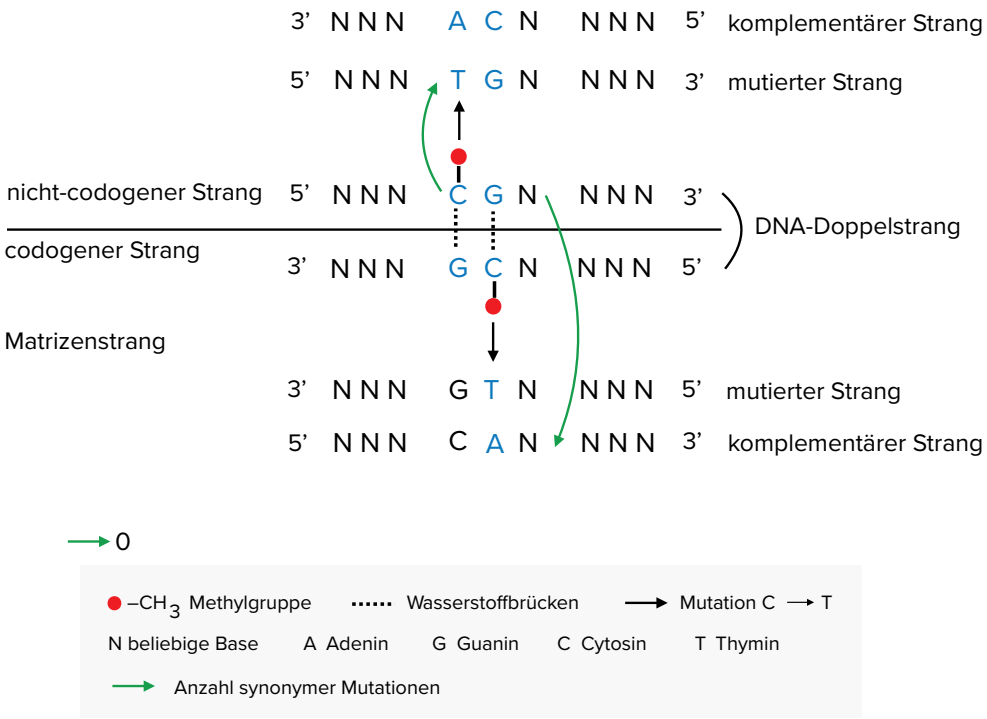


Abb. 6.5 Synonyme Mutationen von methyliertem Cytosin bei einem CpG-Dinucleotid an erster und zweiter Stelle eines Codons.

Im *nicht-codogenen Strang* wird CGN (Arg) durch eine Mutation nach Methylierung ^{met}C → T zu TGN (**CG**N → TGN). Hierbei ergeben sich keine synonymen Codons, da sich CGN und TGN in unterschiedlichen Quadranten der Codonsonne befinden und Arginin (CGN) nur vierfach redundant ist.

Da die Sequenz palindromisch ist, wird im *codogenen Strang* (Matrizenstrang) das C ebenfalls methyliert und kann daher zu T mutieren. Aus NCG (Pro Ser Ala Thr) wird hier durch die Mutation ^{met}C → T die veränderte Sequenz NTG. Nach Replikation der DNA entsteht daraus die komplementäre Sequenz CAN (His bzw. Gln). Hier ergeben sich ebenfalls keine synonymen Codons.

Fall 2: N**CG**. Hier kommt die CpG-Sequenz in der zweiten und dritten Position eines Codons vor (siehe Abbildung 6.6).

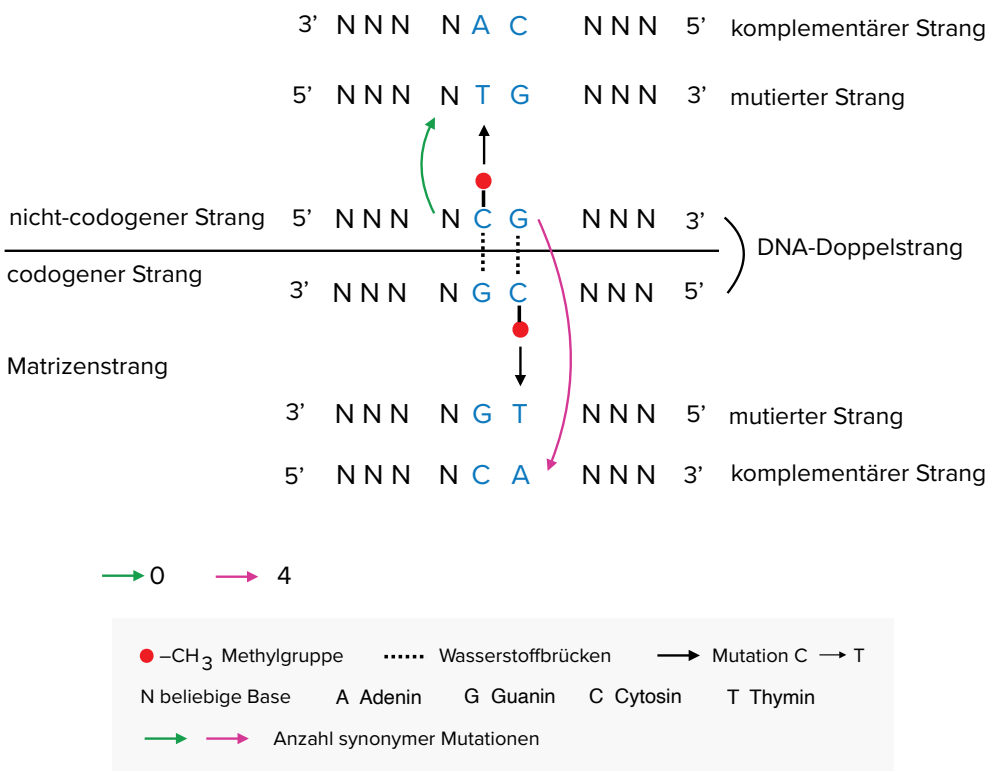


Abb. 6.6 Synonyme Mutationen von methyliertem Cytosin bei einem CpG-Dinucleotid an zweiter und dritter Stelle eines Codons.

Im nicht-codogenen Strang wird NCG (Pro, Ser, Ala, Thr) durch eine Mutation nach Methylierung ^{met}C ➔ T zu NTG (Leu, Val Met) (NCG ➔ NTG).

Da die Sequenz palindromisch ist, wird im *codogenen Strang* (Matrizenstrang) das C ebenfalls methyliert und kann daher zu T mutieren.

Unabhängig davon, welche Base N ist, ist die C ➔ T Mutation in jedem Fall synonym. Aus den Codons für Pro (CCG), Ser (TCG), Ala (GCG) und Thr (ACG) werden durch die Mutation synonyme Codons für Pro (CCA), Ser (TCA), Ala (GCA) und Thr (ACA).

Fall 3: NN**C – G**NN. Im dritten Fall kommt die CpG-Sequenz in der dritten Position eines Codons und in der ersten Position des nachfolgenden Codons vor. Die CpG-Sequenz ist also *codonübergreifend*.

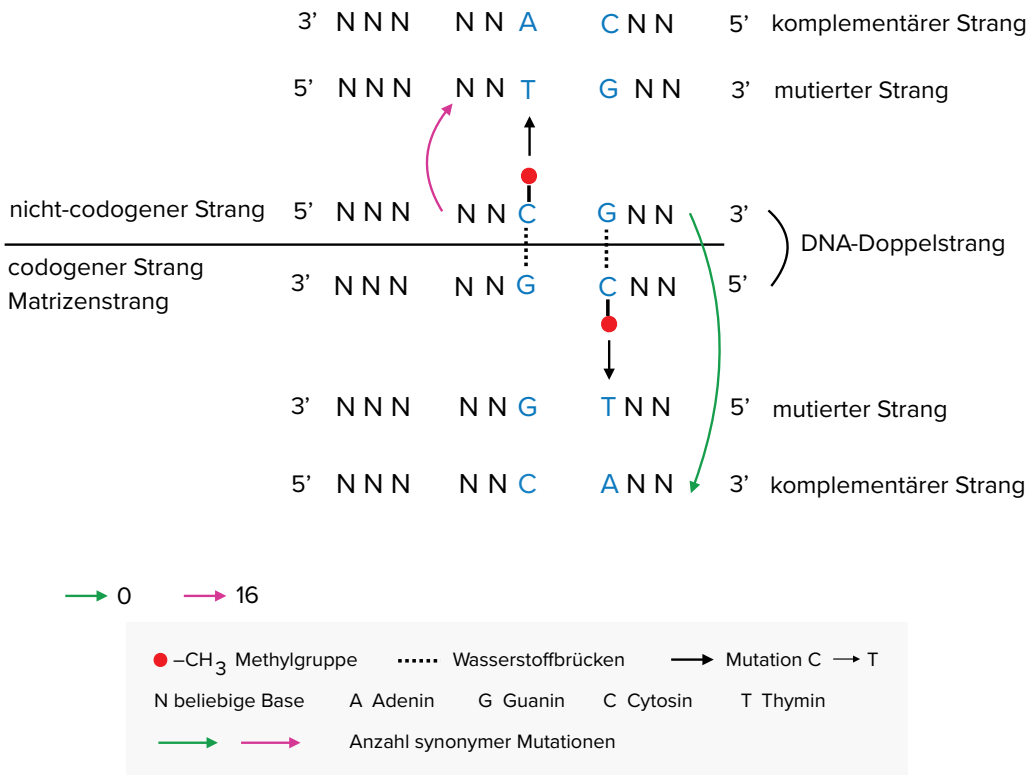


Abb. 6.7 Synonyme Mutationen von methyliertem Cytosin bei einem codonübergreifenden CpG-Dinucleotid.

Erstes Codon: Das erste Codon NNC (Arg, His, Pro, Leu, Cys, Tyr, Ser, Phe, Gly, Asp, Ala, Val, Ser, Asn, Thr, Ile) wird durch Methylierung des C und anschließende Mutation des ^{met}C → T zu NNT (Arg, His, Pro, Leu, Cys, Tyr, Ser, Phe, Gly, Asp, Ala, Val, Ser, Asn, Thr, Ile) (NNC → NNT). Hier gibt es 16 verschiedene Möglichkeiten einer synonymen Mutation. Wenn hingegen das C im codogenen Strang von der Mutation zu T betroffen ist, gibt es keine synonymen Mutationen.

Zweites Codon: Das zweite Codon im nicht-codogenen Strang GNN enthält kein Cytosin und wird daher nicht methyliert.

Nach Replikation ergibt sich im komplementären Strang zum codogenen Strang ANN (GNN → ANN). Da sich diese Codons in unterschiedlichen Quadranten der Codonsonne befinden und der Quadrant GNN keine sechsfach redundanten Codons enthält, gibt es keine Möglichkeit für synonyme Mutationen.

Zusammenfassend lässt sich sagen, dass bei Mutationen aufgrund von Methylierung des Cytosins ^{met}C ➔ T nur in den Fällen 2 und 3 synonyme Mutationen vorkommen können.

Wie hier dargelegt, sind CpG-Dinucleotide, sofern das C methyliert ist, besonders anfällig für C ➔ T Mutationen. Weiters führen solche Mutationen in codierenden Regionen des Genoms sehr oft zu synonymen Codons, die jedoch die Aminosäuresequenz des Proteins nicht verändern würden und möglicherweise auch die Funktion des Proteins nicht beeinträchtigen würden. Tatsächlich fanden Branciamore et al. (2010) heraus, dass in Spezies, die ihr Genom methylieren können, CpG-Dinucleotide aufgrund spontaner Mutationen CpG ➔ ^{met}CpG ➔ TpG unterrepräsentiert sind. Im komplementären Strang entsteht bei Replikation aus T ein A, wodurch es im Laufe der Evolution zu einer fünffachen Reduktion der Häufigkeit von CpG-Dinucleotiden gekommen ist. Hierbei handelt es sich um die Gesamthäufigkeit der CpG-Dinucleotide, denn deren Verteilung entlang des Genoms ist sehr unterschiedlich.

Untersucht man für Proteine codierende Abschnitte humaner DNA, so findet man, dass die Häufigkeit der CpG-Dinucleotide auf ein Fünftel reduziert ist. Dies entspricht der Beobachtung, dass eine C ➔ T Mutation in vielen Codons zu synonymen Codons führt, wobei sich die Aminosäuresequenz nicht ändert. Falls sich dabei die Funktion des Proteins nicht ändert, gibt es keinen Selektionsdruck, die CpG-Dinucleotide zu erhalten. Branciamore et al. (2010) stellten aber fest, dass in Genen, die für wichtige Proteine der Embryonalentwicklung codieren (wie z. B. für Hox-Proteine), CpG-Dinucleotide genauso häufig zu finden sind, wie es dem statistischen Mittelwert entspricht. Daher vermuten die Autor*innen einen Selektionsdruck zugunsten der Erhaltung von CpG-Dinucleotiden. Sie vermuten weiters, dass bei diesen Genen die Expression während der Embryonalentwicklung über die Methylierung des Cytosin in CpG-Dinucleotiden geregelt wird. C ➔ T Mutationen würden die korrekte Entwicklung des Organismus stören.

Park et al. untersuchten 2016 mehr als 31000 menschliche proteincodierende Sequenzen. Die Daten entnahmen sie der CCDS Datenbank (siehe Anhang). Sie fanden heraus, dass die Codons GCG (Ala), CCG (Pro), UCG (Ser) und ACG (Thr) in den initialen 50 Codons signifikant häufiger vorkommen und dass es ich bei diesen Codons um *minor codons* handelt. (*Minor codons* entspricht der Bezeichnung *seltene Codons* bzw. *rare codons*). Die Autor*innen nehmen an, dass die Konzentrationen von tRNA-Molekülen für alle Codons gleich sind. Basierend auf dieser Annahme postulieren die Autor*innen, dass die tRNA-Konzentrationen pro Codon für seltene Codons (wie die genannten Codons für die Aminosäuren Ala, Pro, Ser und Thr) vergleichsweise hoch sind. Wenn man zusätzlich annimmt, dass die Translationsgeschwindigkeit mit der tRNA-Konzentration pro Codon korreliert, könnte die Anhäufung seltener Codons in den initialen 50 Codons zu einer Beschleunigung der Translation dieser mRNAs führen. Eine analoge Untersuchung synonymer Codons aller anderen Aminosäuren, selten oder häufig, ergab allerdings keine signifikante Anhäufung der entsprechenden seltenen Codons. Park et al. (2016) argumentieren in diesem Fall konträr zur *codon ramp* von Tuller et al. (2010).

Branciamore et al. (2010) sehen jedoch CpG-Dinucleotide, die in Promotorbereichen von menschlichen Genen gehäuft vorkommen, als evolutionär konservierte Dinucleotide in CpG-Inseln an, die epigenetische Funktionen zur Ein- bzw. Ausschaltung von Genen besitzen (siehe Kapitel 6.1). Bemerkenswert ist, dass Park et al. (2016) einen Mittelwert über alle 31000 menschlichen Gene bildeten und eine Häufung der seltenen N **CG**-Codons feststellen konnten, während Branciamore et al. (2010) nur einzelne Gene betrachteten. Die Möglichkeit, dass es sich bei der Häufung von N **CG**-Codons innerhalb der ersten 50 Codons um epigenetische Effekte handeln könnte, wird nicht erwähnt/erwogen, obwohl Park et al. (2016) bemerken, dass es sich bei allen vier bevorzugten Codons (p-codons, prefered codons) um solche mit CG-Nucleotiden handelt.

6.2 Spleißfehler

Wie bereits in Kapitel 1.8.2 dargelegt, besteht die prä-mRNA von Eukaryoten aus abwechselnd angeordneten Exons und Introns. In den Exons befinden sich Bereiche, die für Proteine codieren sowie regulatorische Bereiche. Die Introns bestehen aus nicht-codierenden Bereichen, die während oder nach der Transkription aus der prä-mRNA herausgespleißt werden.

An der 5‘-Spleißstelle – am Ende eines Exons und am Beginn des folgenden Introns – befinden sich bestimmte *Konsensussequenzen*, welche eine Spleißstelle für das Spleißosom – einem RNA-Protein-Komplex, der das Spleißen katalysiert (DocCheck Flexikon Spliceosom/Spleißosom) – markieren. An der 3‘-Spleißstelle – am Ende des Introns und am Beginn des folgenden Exons – befinden sich ebenfalls *Konsensussequenzen*. Diese sind Teile des Spleißsignals und werden von den Spleißosomen erkannt (Berg et al., 2018, S. 154). Eine Konsensussequenz ist die am häufigsten vorkommende Nucleotidsequenz an einer bestimmten wichtigen Position, wie beispielsweise an einer Spleißstelle.

Ma et al. (2015) untersuchten Spleißstellen in menschlichen Genen und fanden, dass die Häufigkeiten der Konsensusseqenzen zu Beginn und am Ende eines Exons deutlich geringer ausfallen als in Introns. Die Konsensussequenzen in den Introns lagen bei über 99 Prozent, was zeigt, dass diese hoch konserviert sind. Jene der Exons hatten deutlich geringere Häufigkeiten, wodurch sich auch eine größere Variabilität zwischen verschiedenen Spleißstellen zeigt. In Abbildung 6.8 sind die Untersuchungsergebnisse über die Häufigkeiten der Konsensussequenzen von Ma et al. (2015) zusammengefasst.

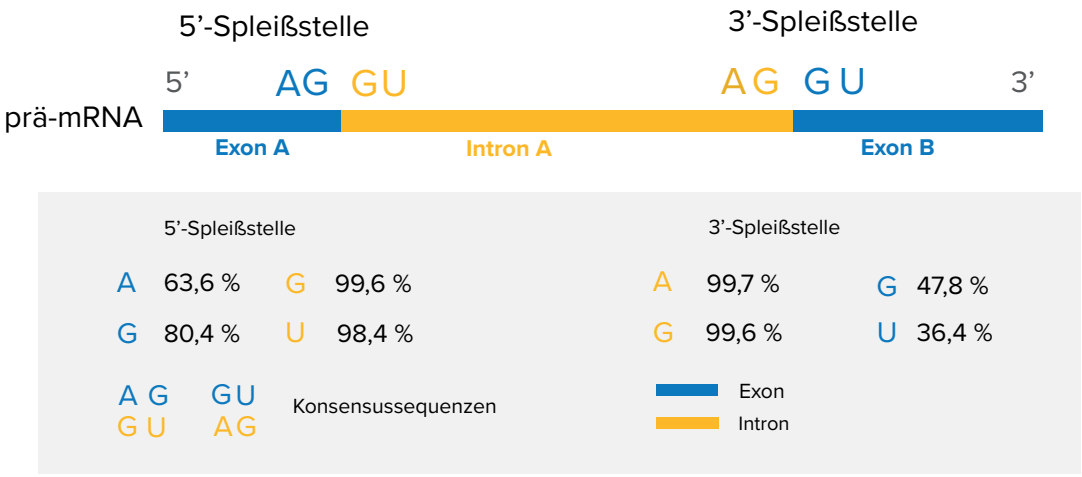


Abb. 6.8 Häufigkeiten der Konsensussequenzen an Spleißstellen von menschlichen Genen. Zahlen aus Ma et al. (2015).

6.2.1 Regulärer Spleißvorgang

In Abbildung 6.9 ist die Situation vor und nach einem regulären Spleißvorgang dargestellt. Das Intron, welches sich zwischen der 5‘-Spleißstelle und der nachfolgenden 3‘-Spleißstelle befindet, wird vom Spleißosom herausgespleißt (hier exemplarisch Intron A). Danach schließen die Exons aneinander an.

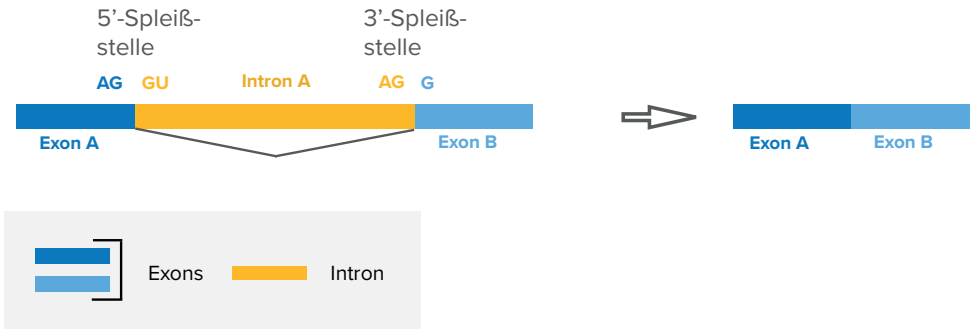


Abb. 6.9 Reguläres Spleißen: Entfernung eines Introns zwischen den Konsensussequenzen der 5‘-Spleißstelle und der 3‘-Spleißstelle (Berg et al., 2018, S. 155, adaptiert).

Hier ist zu bemerken, dass in manchen Veröffentlichungen die Bezeichnung *Donorspleißstelle* anstelle von 5‘-Spleißstelle sowie *Akzeptorspleißstelle* anstelle von 3‘-Spleißstelle verwendet wird.

6.2.2 Mutation der 5‘-Spleißstelle mit Exon-Skipping

Spleißmutationen – Mutationen, die das Spleißen beeinträchtigen – können sowohl in Exons als auch in Introns vorkommen. Da synonyme Mutationen nur in codierenden Bereichen – also in Exons – vorkommen, werden hier nur Mutationen in Exons betrachtet.

Befindet sich eine synonyme Mutation im ersten oder im letzten Triplett eines Exons, so werden die 3‘- bzw. die 5‘-Spleißstelle verändert (siehe Abbildung 6.10). Die Spleißstellen werden vom Spleißosom nicht mehr erkannt und daher kann es fälschlicherweise zum Herausspleißen zweier Introns mit dem dazwischen liegenden Exon kommen (Exon-Skipping – Überspringen eines Exons).

Diaz et al. (2022) beschreiben einen Fall von congenitaler (angeborener) Hyperinsulinämie (Neugeborenendiabetes) bei einem nepalesischen Patienten. Aufgrund eines Gendefekts im ABCC8-Gen wurden große Mengen Insulin ausgeschüttet und es kam bei dem Patienten kurz nach der Geburt zu einer massiven Hypoglykämie (Unterzuckerung). Die sonst übliche Gabe des *Medikaments Diazoxid*, das normalerweise die Insulinausschüttung verringert, zeigte keinerlei Wirkung.

Genetische Untersuchungen zeigten eine bis zu diesem Zeitpunkt noch unbekannte synonyme Mutation des ABCC8-Gens an der Wobble-Position im letzten Triplett von Exon 38 (GAG Alanin ➝ GAA Alanin). Das ABCC8-Gen codiert für eine Untereinheit eines ATP-sensitiven Kaliumkanals. Bei diesem Defekt handelt es sich nicht um eine Autoimmunkrankheit wie bei Diabetes Typ I (Schaaf & Zschocke, 2018, S. 338). Die Eltern waren miteinander verwandt (Cousin/Cousine) und trugen beide die selbe heterozygote (gemischterbige) Mutation. Beim Patienten trat die Mutation homozygot (reinerbig) auf. Weitere Überprüfungen zeigten, dass das aus 21 Nucleotiden bestehende Exon 38 aus der prä-mRNA herausgespleißt wird und in der mRNA fehlt (Exon Skipping), (siehe Abbildung 6.10).

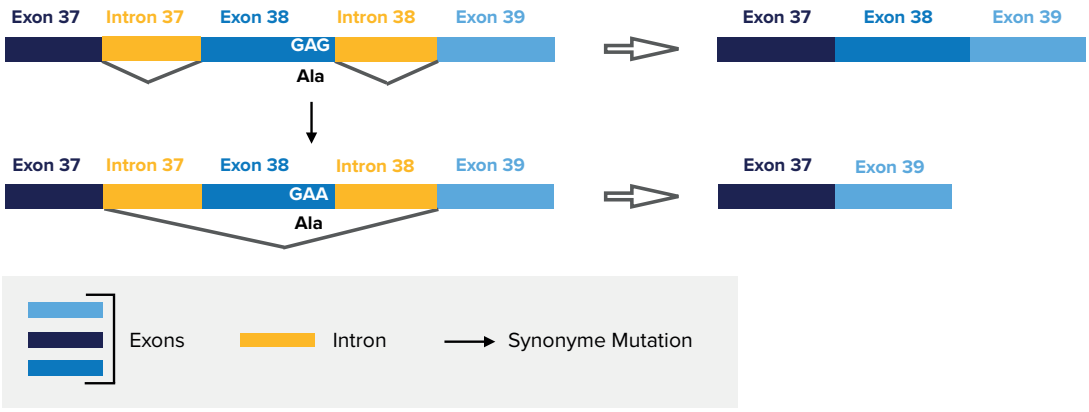


Abb. 6.10 Eine synonyme Mutation Ala (GAG) ➝ Ala (GAA) in der Konsensussequenz (im letzten Triplett) von Exon 38 bewirkt das Überspringen des Exons 38 des ABCC8-Gens.

Diese Mutation verändert nach Diaz et al. (2022) die übliche Konsensussequenz der 5‘-Spleißstelle des Exons 38 von AG in AA, sodass diese vom Spleißosom nicht mehr erkannt wird. Auch hier handelt es sich um die nicht sehr hoch konservierte Konsensussequenz am Ende des des Exons. Durch Vergleiche mit ähnlichen Fällen und auch durch den Vergleich mit den Eltern des Patienten, bei denen beiden diese Mutation nachgewiesen wurde, erschien es naheliegend, dass diese synonyme Mutation die Hyperinsulämie hervorruft. Wegen des Nichtansprechens auf das Medikament Diazoxid wurde in der Folge der größte Teil des Pankreas operativ entfernt.

6.2.3 Aktivierung einer kryptischen Spleißstelle mit partieller Intron-Retention

Kryptische Spleißstellen werden laut Roca et al. (2003) nur dann verwendet, wenn die natürliche Spleißstelle durch eine Mutation zerstört ist. Kryptische Spleißstellen weisen eine der Konsensussequenz ähnliche Sequenz auf, werden aber normalerweise nicht erkannt, es sei denn, es gibt Mutationen an anderer Stelle im Gen oder sehr oft auch nahe der authentischen Spleißstelle. Die kryptischen Spleißstellen können sich im Exon oder auch im Intron befinden.

Yu et al. (2023) führten am Wuhan Children’s Hospital eine Studie an einer einjährigen Patientin durch, die von dilatativer Cardiomyopathie (DCM – Dilated Cardiomyopathy) betroffen war. Hierbei ist meist der linke Ventrikel des Herzens vergrößert, was zu einer Minderleistung des Herzens führt. Das TNNI3-Gen, das sich auf dem langen Arm des Chromosoms 19 befindet, codiert für Troponin I, ein Protein, welches ausschließlich im Herzmuskel vorkommt und eine wichtige Rolle bei der Kontraktion des Herzmuskels spielt. Eine Mutation im letzten Codon von Exon 2 (Ala GCG → Ala GCA) zerstört die vorhandene Spleißstelle, sodass diese nicht mehr erkannt wird. Dies führt zur Verwendung einer kryptischen Spleißstelle, was zu einer Retention von 45 Nucleotiden (nt) von Intron 2 führt (siehe Abbildung 6.11).

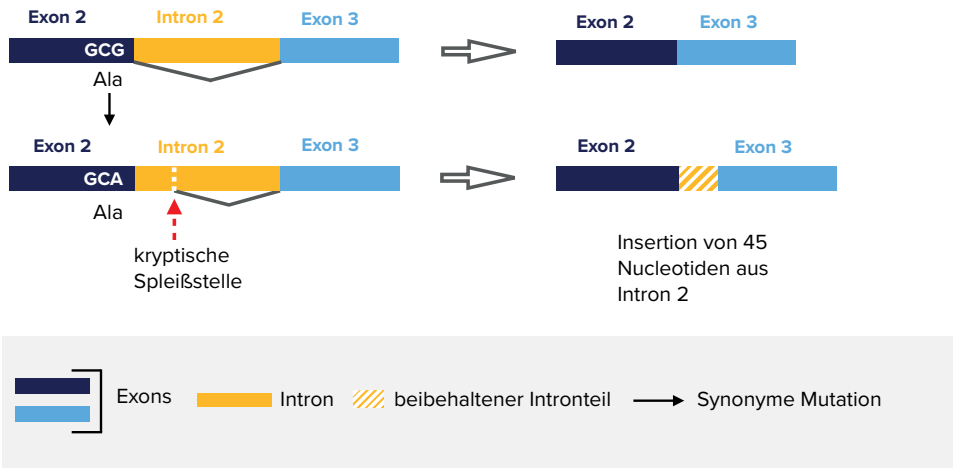


Abb. 6.11 Partielle Intron-Retention aufgrund einer synonymen Mutation von Ala (GCG) zu Ala (GCA), die eine vorhandene 5‘-Spleißstelle verändert, woraufhin diese nicht mehr erkannt wird. Stattdessen wird eine kryptische Spleißstelle verwendet (Yu et al., 2023, adaptiert).

Im vorliegenden Fall fällt auf, dass die Nucleotidsequenz am Ende des Wild-Typ Exons 2 (GCG) nicht der Konsensussequenz NAG (siehe Abbildung 6.8) entspricht. Wie aus der Tabelle der Abb. 6.8 ersichtlich sind am Ende der Exons die Konsensussequenzen nicht so hoch konserviert und variieren daher stärker als in den Introns. Zu einem gewissen Prozentsatz können auch andere Sequenzen auftreten. Laut Zusammenstellung nach Ma et al. (2015) tritt C an vorletzter Stelle im Exon mit einer Häufigkeit von 10,8 % auf. Die 45 zusätzlichen Nucleotide des ursprünglichen Introns 2 bis zur neuen kryptischen Spleißstelle werden beibehalten und nach dem Spleißen als Codons interpretiert, wodurch weiter translatiert wird und zusätzliche Aminosäuren eingefügt werden.

Wie aus der Datenbank *Homo sapiens troponin I3, cardiac type (TNNI3), mRNA* (siehe Anhang) ersichtlich (siehe Abbildung 6.12) werden aufgrund der Intron-Retention 5 zusätzliche Aminosäuren eingefügt. Danach ergibt sich ein vorzeitiges Stopp-Codon (TAA). Das Stopp-Codon erscheint, noch bevor alle zusätzlichen 45 Nucleotide aus Intron 2 translatiert werden konnten, wodurch die Translation vor Erreichen des Exons 3 abbricht. Abbildung 6.12 zeigt einen Ausschnitt der Sequenz von Exon 2 und 3 sowie von Intron 2 des *TNNI3-Gens*.

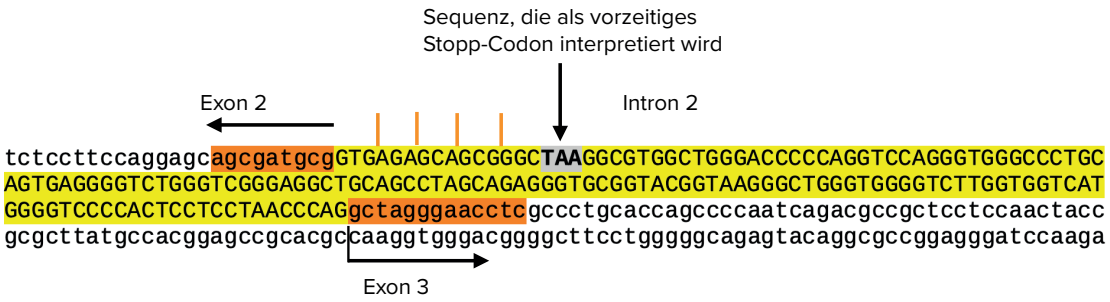


Abb. 6.12 Ausschnitt der nicht mutierten Sequenz des TNNI3-Gens. Das Ende von Exon 2 und der Beginn von Exon 3 sind rot unterlegt, Intron 2 gelb. Durch eine synonyme Mutation am Ende von Exon 2 – Ala (GCG ➔ GCA) – wird das Intron teilweise beibehalten und weiter translatiert. Nach 5 zusätzlichen Aminosäuren erscheint ein vorzeitiges Stopp-Codon (aus Datenbank: Homo sapiens troponin I3, cardiac type (TNNI3), mRNA umformatiert).

Dieses Beispiel zeigt, dass durch eine synonyme Mutation in einer 5'-Spleißstelle nicht nur ein Exon herausgespleißt werden kann (Exon-Skipping), sondern dass stattdessen auch ein Teil eines Introns beibehalten (Intron-Retention) und in die mRNA integriert werden kann.

6.2.4 Entstehung einer neuen 5‘-Spleißstelle mit Teilverlust eines Exons

Beim *von-Willebrand-Syndrom* (VWS, VWD von Willebrand disease) handelt es sich um eine monogene Krankheit, die von einer oder mehreren Mutationen in einem einzigen Gen, das sich auf Chromosom 12 befindet, verursacht wird. Bei dieser Krankheit treten vermehrt Hämatome, verstärkte Blutungen bei minimalen Verletzungen und verlängerte Blutungen bei chirurgischen Eingriffen auf. Diese sogenannten Gerinnungsstörungen werden durch unterschiedliche Mutationen im *von-Willebrand-Gen* verursacht, die einen Mangel am *von-Willebrand-Faktor* (vWF) hervorrufen (Schaaf & Zschocke, 2018, S. 283). Der von-Willebrand-Faktor (vWF) ist ein Glykoprotein, welches die Verbindung zwischen verletzter Gefäßwand und Thrombozyten vermittelt (DokCheckFlexikon).

Daidone et al. (2011) untersuchten das von-Willebrand-Gen einer 15 Jahre alten Patientin, die von milden Blutungsneigungen nach geringen Verletzungen, häufigen Hämatomen sowie von Nasenbluten betroffen war. Ihre Mutter und Großmutter waren ebenfalls von Blutungsneigungen betroffen. Der Vergleich des von-Willebrand-Gens der Patientin mit jenem der Großmutter, des Großvaters und deren Abkömmlingen ergab, dass bei allen Personen mit klinischen Symptomen die selbe synonyme Mutation vorhanden war. Die genaue Sequenzierung ergab eine synonyme Mutation eines Glycin-Codons: GGC → GGT in Exon 41.

Beim Wildtyp (Glycin-Codon: GGC in Exon 41) wurde Intron 41 korrekt herausgespleißt, beim mutierten Glycin GGT in Exon 41 wurde die veränderte Sequenz GGT als neue 5‘-Spleißstelle interpretiert. Die alte 5‘-Spleißstelle im Intron 41 wurde aufgrund der Mutation nicht mehr erkannt, da diese eine neue stärkere 5‘-Spleißstelle im Exon 41 erzeugt hat. Gemeinsam mit Intron 41 wurden die letzten 27 Nucleotide (nt) des Exons 41 herausgespleißt. Dem Transkript fehlen 27 Nucleotide, dem Protein fehlen daher 9 Aminosäuren. Abbildung 6.13 zeigt den Spleißvorgang vor und nach der Mutation.

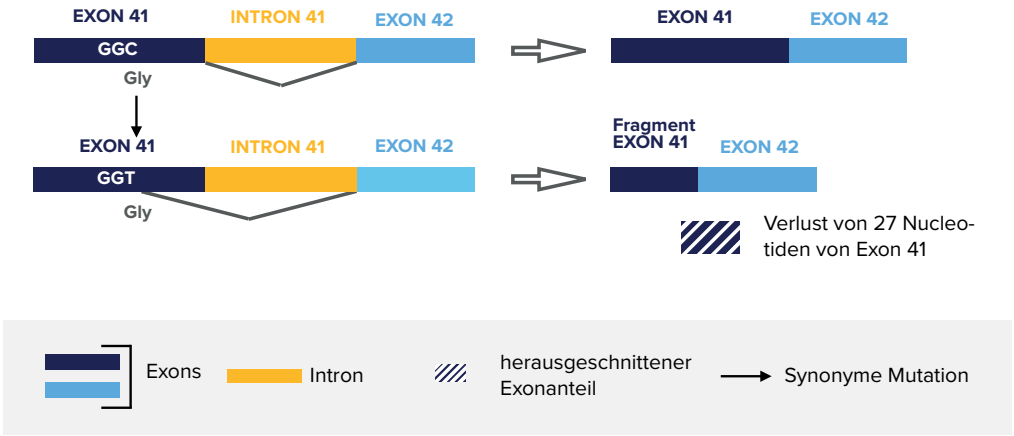


Abb. 6.13 Eine synonyme Mutation in Exon 41 des von-Willebrand-Gens erzeugt eine neue 5'-Spleißstelle, die einen Verlust von 27 Nucleotiden bewirkt (Daidone et al., 2011, verändert).

Ma et al. (2022) beschreiben die *gezielte Suche* nach Mutationen bei einem sieben Jahre alten Patienten mit *X-linked Hypophosphatemie* mittels Whole-Exom-Sequenzierung. Bei diesem Defekt ist das PHEX-Gen (*PHEX* – phosphate regulating endopeptidase X-linked) betroffen. Der Defekt führt zu einem niedrigen Phosphatspiegel, wodurch sich unter anderem eine Rachitis mit Deformationen der Knochen und der Zähne ausbildet. Der Patient zeigte verkürztes Längenwachstum sowie eine verminderte Knochendichte und verbreiterte Epiphysenfugen im Röntgenbild. Es wurden sowohl der Patient als auch seine Eltern und seine Schwester getestet. Die Ergebnisse der Whole-Exom-Sequenzierung wurden mit den bis dahin bereits bekannten 720 Einträgen zu PHEX-Varianten verglichen, die in der *Human Gene Mutation Database* (HGMD) verzeichnet waren.

Die neu identifizierte synonyme Mutation (Arg CGC → Arg CGT) in Exon 14, die bei den Familienmitgliedern nicht vorhanden war, führte zur Entstehung einer neuen 5'-Spleißstelle in Exon 14 und einem Verlust von 58 Nucleotiden von Exon 14. Es bleibt dann nur mehr ein Fragment dieses Exons erhalten (siehe Abbildung 6.14). Auch hier entstand durch die Mutation eine neue hoch konservierte Konsensussequenz (GT), die als Intron interpretiert wurde.

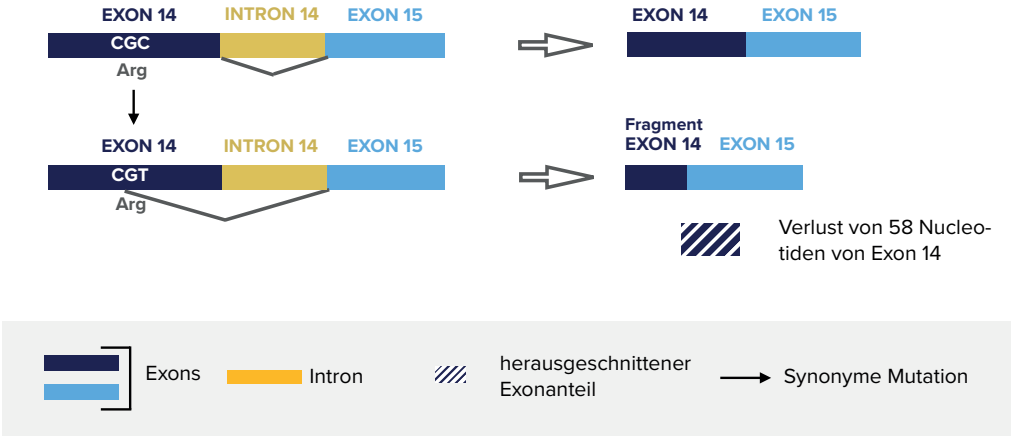


Abb. 6.14 Durch eine synonyme Mutation in Exon 14 entsteht eine neue 5'-Spleißstelle. Dadurch erfolgt ein Verlust von 58 Nucleotiden am Ende von Exon 14 (in Anlehnung an Ma et al., 2022).

Die Gemeinsamkeit der beiden Fälle (Ma et al., 2022 und Daidone et al., 2011) besteht darin, dass die Mutationen jeweils eine neue Basenfolge GT in den beiden letzten Basen eines Tripletts erzeugen, die als neue 5'-Spleißstelle missinterpretiert wird. Der Unterschied besteht jedoch darin, dass es sich um unterschiedliche Tripletts für unterschiedliche Aminosäuren handelt: bei Ma et al. (2022) CGC → CGT (Arginin), bei Daidone et al. (2011) GGC → GGT (Glycin).

6.2.5 Exonic Splice Enhancers (ESE) und Exonic Splice Silencers (ESS)

Exonische Spleißverstärker (exonic splice enhancer, ESE) sind kurze Sequenzen, die das Splicing verstärken, während exonische Spleißunterdrücker (exonic splice silencer, ESS) dieses hemmen. Zwar gibt es auch in Introns Spleißverstärker und Spleißunterdrücker, diese sind aber nicht Gegenstand dieser Betrachtung. Wird die Sequenz eines exonic splice enhancer durch eine Mutation, zum Beispiel eine synonyme Mutation, verändert, wird das entsprechende Exon nicht mehr inkludiert sondern übersprungen.

Collin et al. (2018) untersuchten eine Hörbehinderung einer niederländischen Familie, wobei elf Familienmitglieder klinische Symptome zeigten. Diese Hörminderung bestand in einer Absenkung der Hörkurve bei mittleren Frequenzen von 500–2000 Hertz.

Durch Sequenzanalysen konnte bei mehreren Familienmitgliedern eine bisher unbekannte heterozygote synonyme Mutation des TECTA-Gens in Exon 16 festgestellt werden. Die synonyme Mutation (Leu CTG → Leu CTA), stellte sich als ursächlich für die Hörminderung heraus. Es wurde nachgewiesen, dass diese Mutation *den Verlust eines ESE* verursacht, wodurch in der mRNA für das TECTA-Protein die Codons für 37 Aminosäuren (das gesamte Exon 16) fehlten.

Das TECTA-Gen codiert für α-Tectorin, eines der beiden Glycoproteine der Tektorialmembran, einer gallertartigen Struktur, welche die Haarzellen in der Cochlea (Hörschnecke im Innenohr) bedeckt. Der Schall verursacht eine Bewegung der Tektorialmembran und stimuliert so den Hörnerv über die Haarzellen (DocCheck Flexikon, Tektorialmembran, Cortiorgan).

6.3 mRNA-Stabilität

Die mRNA-Struktur bzw. deren Faltung hat entscheidenden Einfluss auf deren Stabilität. Durch intramolekulare Basenpaarungen bildet die mRNA Haarnadelstrukturen (stem loops) aus, die Einfluss auf die Stabilität bzw. den Abbau der mRNA haben. Darüber hinaus haben die Haarnadelstrukturen auch einen Einfluss auf die Geschwindigkeit der Proteinsynthese und damit auf die Konzentration des codierten Proteins in der Zelle (Berg et al., 2018, S. 138).

6.3.1 Veränderte mRNA-Struktur verändert die Stabilität

Nackley et al. (2006) untersuchten Auswirkungen verschiedener synonymer Mutationen auf die Catechol-O-Methyltransferase (COMT). Diese ist an der der Inaktivierung von Neurotransmittern durch Übertragung einer Methylgruppe beteiligt. Sie stellten fest, dass die unterschiedlichen synonymen Codons eine veränderte mRNA-Struktur bewirken, und dass die stabilste Struktur mit dem geringsten Proteinlevel, reduzierter Enzymaktivität und größerer Schmerzempfindlichkeit assoziiert ist (Nackley et al., 2006; Sauna & Kimchi-Sarfaty, 2011). Zusätzlich hängt die Faltung der mRNA und dadurch die Proteinexpression davon ab, an welcher Stelle sich der synonyme Austausch befindet.

Nackley et al. (2006) fanden zwei synonyme Mutationen rs 4818: Leu (CUC) → Leu (CUG) sowie rs4633: His (CAC) → His (CAT). Dadurch, dass eine Base im synonymen Codon verändert ist, verändert sich auch die intramolekulare Basenpaarung innerhalb der Haarnadelschleifen der mRNA, was zu einer veränderten Struktur derselben führt.

Die Catechol-O-Methyltransferase spielt unter anderem eine wichtige Rolle bei der Neurotransmission der Schmerzleitung. Der Spiegel des Enzyms wird durch synonyme Mutationen beeinflusst und beeinträchtigt die Schmerzempfindung der Patienten erheblich. Die von Nackley et al. untersuchten Varianten zeigten stark unterschiedliche Strukturen der Haarnadelschleifen in der mRNA, die die Expression des Enzyms beeinflussen.

	rs 4818: Leu CUC → CUG	rs 4633: His CAC → CAT
RNA Haarnadel/Stabilität	↑ höher	↓ niedriger
Proteinlevel	↓ niedriger	↑ höher
Schmerzempfindlichkeit	↑ höher	↓ niedriger

Zudem stellten sie fest, dass die synonymen Mutationen weit größeren Einfluss auf das Schmerzgeschehen haben, als die nicht-synonymen Mutationen (Nackley et al., 2006; Shabalina et al., 2013). Auch hier zeigt sich die von Shabalina et al. betonte Erkenntnis, dass die mRNA zusätzlich zum Bauplan noch weitere Informationen enthält, die nur aufgrund der Redundanz des genetischen Codes untergebracht werden können.

Auch Tsao et al. (2011) fanden, dass eine höhere COMT-Aktivität mit einer geringeren Schmerzempfindlichkeit einhergeht. Sie stellten fest, dass durch die Mutation rs4633 zwei Watson-Crick-Paare verloren gehen und sich dadurch der einsträngige Bereich um zwei Nucleotide vergrößert, was die Stabilität verringert.

Andere Autor*innen – wie etwa Vetterlein et al. (2023) – sehen in ihrer Metaanalyse eine eher inkonsistente Evidenz, was den Zusammenhang zwischen COMT-mRNA-Stabilität, Proteinspiegel und Schmerzempfindlichkeit betrifft, möglicherweise aufgrund von Heterogenität der untersuchten Gruppen und der Heterogenität verschiedener Schmerzarten.

6.3.2 Kombination mehrerer synonymer Mutationen

Duan et al. (2003) beschreiben ein Beispiel für die Verringerung der mRNA-Stabilität und der Beeinflussung der Proteinsynthese durch eine synonyme Mutation im DRD2-Gen, das für einen Dopaminrezeptor codiert.

Die synonyme Mutation C957T (rs6277: 957C ➔ T, Pro319Pro), (siehe SNpedia online), veränderte die mRNA-Faltung und führte zu verminderter Stabilität und Translation und veränderte die Dopamin-induzierte Hochregulierung der DRD2-Expression. Bei der C957T-Mutation handelt es sich um einen Übergang von einem häufigen Codon auf ein seltenes Codon. Die Halbwertszeit der mRNA wurde durch Infektion von infizierten CHO-Zellen (Chinese hamster ovary), die mit Actinomycin behandelt wurden, um die Transkription zu verhindern, gemessen. Diese sank von 8 Stunden beim Wildtyp auf 4 Stunden bei der C957T-Mutation.

Eine zweite Variante G1101A wies diesen Effekt nicht auf, annulierte aber die Auswirkung der genannten C957T-Variante. Dies zeigte, dass die Kombination mehrerer synonymer Varianten vollkommen andere Auswirkungen haben können als die beiden Varianten einzeln (Duan et al., 2003; Shabalina et al., 2013).

6.3.3 micro-RNA

Micro-RNAs (miRNA) sind eine Klasse kurzer, nicht-codierender RNA-Stücke, die aus der DNA, wo sie codiert sind, transkribiert, nicht aber translatiert werden (Salzmann & Weidhaas, 2022). Die transkribierte pre-miRNA bildet Haarnadelschleifen, wird dann ins Cytoplasma transportiert, zerschnitten und zu reifer miRNA weiterverarbeitet, die ungefähr 22 Nucleotide lang ist. Die reife miRNA wird an einen RISC-Protein-Komplex gebunden (RNA-induced silencing complex)

und sucht mit diesem zur miRNA komplementäre Sequenzen in der mRNA und bindet an diese (Graw, 2020, S. 380). miRNAs können so die Genexpression in manchen Genen sehr schnell regulieren, indem sie die Translation hemmen (Salzmann & Weidhaas, 2022). Andererseits kann die Bindung der miRNA an die mRNA deren Abbau einleiten und damit die Stabilität der mRNA beeinflussen. Laut Berg et al. (2018, S. 1134) wurden bisher 700 verschiedene miRNAs identifiziert.

Morbus Crohn ist eine entzündliche Darmkrankheit die durch Bakterien im Darmepithel hervorgerufen wird. Rhodes (2011) beschreibt, dass Morbus Crohn sehr oft mit persistierenden E.-coli-Bakterien in der Darmschleimhaut assoziiert ist. Es sind viele Mutationen für Morbus Crohn bekannt, darunter eine synonyme Mutation im IRGM-Gen (Immunity-related GTPase family M protein), das für die Autophagie und damit Beseitigung intrazellulärer Bakterien wichtig ist (Salzmann & Weidhaas, 2022). Brest et al. (2011) konnten zeigen, dass die micro-RNA miR-196 bei Patienten mit Morbus Crohn in befallenen Bereichen des Darmepithels überexprimiert ist, wodurch die c.313C-IRGM-Variante herunterreguliert wird, nicht aber die mutierte Risikovariante c.313T. Brest et al. (2011) zeigten auch, dass die miRNA-196 eine geringere Bindungsaktivität zur mutierten c.313T-Variante des IRGM-Gens hat als zur c.313C-Variante. Dadurch, dass die synonyme Variante c.313T die miRNA-Bindungsstelle verändert, vermuten sie, dass ein Verlust der Regulation des IRGM-Proteins eintritt, welches die Autophagie beeinflusst. Dies führt zu einer veränderten antibakteriellen Aktivität und einer abnormen Persistenz der Bakterien in der Schleimhaut. Bei der Mutation handelt es sich um eine Änderung des Leucin-Codons von CTG zu TTG, wobei zu bemerken ist, dass diese Mutation blockübergreifend ist und nicht die Wobble-Position betrifft.

6.4 Faltung und Struktur von Proteinen

In Eukaryoten beginnt laut Netzer & Hartl (1997) die Faltung noch während der Translation (cotranslational), bevor das naszierende Protein freigesetzt wird, wobei Chaperone die Faltung unterstützen und Aggregation verhindern. Laut Netzer und Hartl falten viele Proteine aufgrund sequenzieller cotranslationaler Faltung korrekt, während sie aber bei heterologer Expression in E. coli Fehlfaltungen aufweisen.

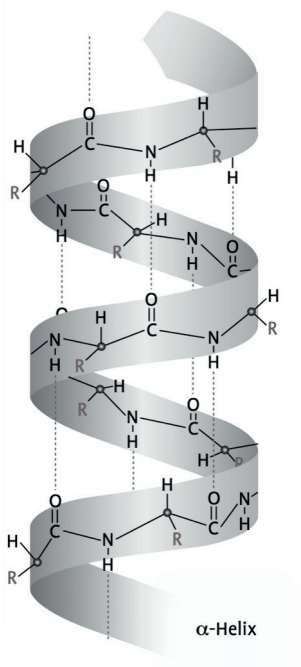
6.4.1 Strukturen von Proteinen

Proteine haben vielfältige Aufgaben als Hormone, Enzyme oder Strukturproteine. Sie sind aus 20 verschiedenen kanonischen Aminosäuren zusammengesetzt und bestehen aus bis zu 34 000 (Titin) Aminosäuren. Eine Proteindomäne ist die kleinste stabile Struktureinheit und enthält zwischen 40 und 350 Aminosäuren (Alberts et al., 2021, S. 142). Sie besteht aus regelmäßig angeordneten Strukturen (α-Helix und β-Faltblatt), die durch lockere Strukturen wie β-turns (Schleifen, Kehren) oder random coils miteinander verbunden sind.

Bei der Translation wird am Ribosom eine lineare Kette aus Aminosäuren gebildet, die durch Peptidbindungen miteinander verknüpft werden (siehe Kapitel 1.7). Die durch die DNA vorgegebene Reihenfolge der Aminosäuren wird als *Primärstruktur* bezeichnet.

6.4.2 Sekundärstruktur

α-*Helices* sind – anders als bei der DNA-Doppelhelix – einsträngige, gewundene Strukturen. β-*Faltblätter bilden* gefaltete Strukturen, ähnlich denen eines gefalteten Blatts Papier. Diese beiden Strukturelemente bilden im Wesentlichen die *Sekundärstruktur* eines Proteins. Die β-turn-loops oder random-coil-Regionen stellen keine durch Wasserstoffbrückenbindungen stabilisierten Bereiche dar und weisen keine erkennbare Struktur auf. Sie treten zwischen den regelmäßig angeordneten Bereichen der α-*Helices und der* β-*Faltblatt-Strukturen auf, wenn diese die Richtung ändern.*



Eine α-Helix wird durch Wasserstoffbrückenbindungen (gepunktete Verbindungen) zwischen einer N-H-Gruppe und einer 4 Aminosäuren entfernten C=O-Gruppe stabilisiert, wobei die meisten Helices rechtsgängig sind (Alberts et al., 2021, S. 137). Die Seitenketten (R) der Aminosäuren weisen nach außen (siehe Abbildung 6.15).

Diese Faltungen sind nur aufgrund der freien Drehbarkeit der Einfachbindungen im Backbone eines Proteins möglich (siehe Diederwinkel). Die erste Beschreibung der Helix in Polypetidketten geht auf Linus Pauling zurück (Pauling et al., 1951).

Abb. 6.15 Räumliche Darstellung einer α-Helix, stabilisiert durch Wasserstoffbrücken (gepunktete Verbindungen). R sind unterschiedliche Seitenketten der Aminosäuren (Reinard, 2021, S. 28).

Bei der β-Faltblattstruktur werden nebeneinanderliegende lineare Aminosäureketten durch Wasserstoffbrücken zwischen Abschnitten der Proteinkette gebildet, die nicht unmittelbar benachbart sein müssen, wobei diese Abschnitte parallel (gleiche Ausrichtung) oder antiparallel (entgegengesetzte Ausrichtung) orientiert sein können. Quervernetzungen können durch van-der-Waals-Kräfte, ionische Wechselwirkungen oder auch durch kovalente Schwefelbrücken erfolgen (Berg et al., 2018, S. 48–49).

Abbildung 6.16 zeigt die Struktur eines β-Faltblatts mit antiparallel ausgerichteten Ketten (Reinard, 2021, S. 28). β-Faltblatt-Strukturen sind für die Zugfestigkeit von Seide verantwortlich (Alberts et al., 2021, S. 141).

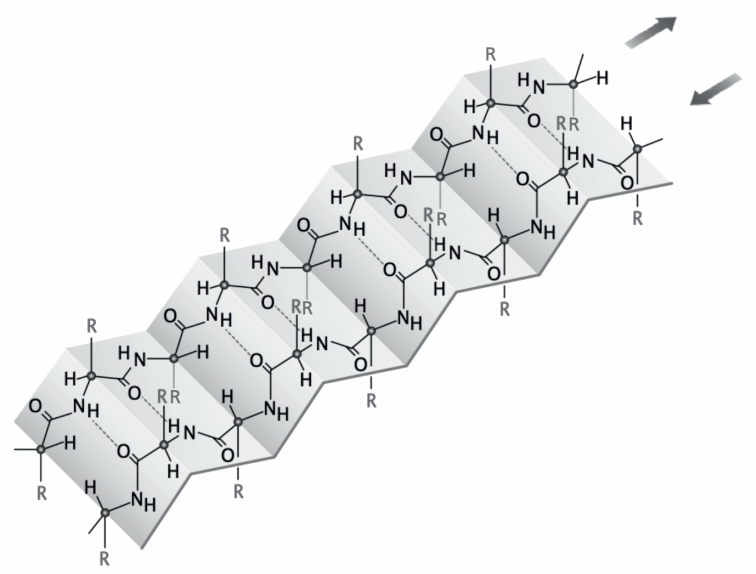


Abb. 6.16 Struktur eines β-Faltblatts mit antiparallel ausgerichteten Aminosäureketten, stabilisiert durch Wasserstoffbrücken (gepunktete Verbindungen). R sind unterschiedliche Seitenketten der Aminosäuren (adaptiert nach Reinard, 2021, S. 28).

Die weitere räumliche Anordnung der Sekundärstrukturen wird als *Tertiärstruktur* bezeichnet. Abbildung 6.17 zeigt die Tertiärstruktur einer Ribonuclease II von E. coli mit β-Faltblättern und α-Helices. Die RNase spaltet Phosphodiesterbindungen in RNA hydrolytisch (Wikipedia).

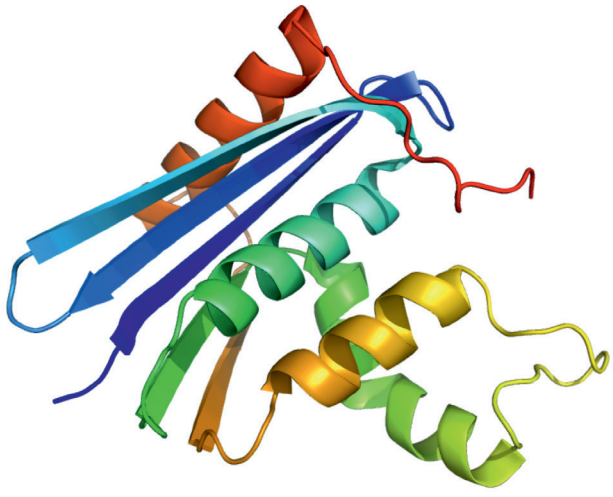


Abb. 6.17 Tertiärstruktur einer RNAse (Ribonuclease II von E. coli). Antiparallele β-Faltblätter (blau, grün, türkis) – α-Helices (rot, gold, blaugrün, grün), (Wikipedia, Vossman).

Lagern sich mehrere Polypeptidketten mit Tertiärstruktur zu einem größeren Komplex zusammen, so bilden diese die *Quartärstruktur* (Buttlar et al., 2020, S. 17).

Nur die genaue Ausbildung dieser Faltungsstrukturen stellt die Funktionsfähigkeit eines Proteins sicher. Ist ein Protein falsch gefaltet, wird es meist im Zuge einer Qualitätskontrolle durch Proteasen abgebaut (Buttlar et al., 2020, S. 131). Ein falsch gefaltetes Protein kann aber auch Aggregate (Plaques) bilden, an denen die Proteasen nicht mehr angreifen können und die daher nicht mehr abgebaut werden können. Solche aus fehlgefalteten Proteinen bestehenden Plaques können letal sein, wie sich an Krankheiten wie BSE (Bovine spongiforme Enzephalopathie), Kuru, der Creutzfeldt-Jakob-Krankheit, Chorea Huntington oder auch an der meist genetisch bedingten tödlichen familiären Schlaflosigkeit zeigt. Alzheimer und Parkinson zeigen ebenfalls durch Plaques hervorgerufene, fortschreitende neurologische Störungen, die sich aber erst innerhalb eines längeren Zeitraumes von mehreren Jahren ausbilden (Buttlar et al., 2020, S. 131).

6.4.3 Vorgang der Faltung

Die Faltung eines Proteins kann bereits während der Translation beginnen (cotranslational), nachdem erst ein Teil des Proteins durch das Ribosom synthetisiert ist. Dies geschieht sehr oft bereits im Exit-Tunnel des Ribosoms (Fedyukina & Cavagnero, 2011). Dies bedeutet, dass die Faltung am aminoterminalen Ende, das als erstes synthetisiert wird, beginnt.

In den 1950er-Jahren führte der norwegisch-amerikanische Biochemiker Christian Anfinsen Experimente zur Struktur und Faltung der Ribonuclease-A aus Rinderpankreas durch. Er denaturierte das Enzym mit Hilfe von Harnstoff und Mercaptoethanol (ein Reduktionsmittel) und destabilisierte dadurch die H-Brücken sowie die vier Schwefelbrücken, sodass danach ein ungefaltetes lineares Protein vorlag. Er re-oxidierte die Lösung und konnte beobachten, dass sich das Enzym von selbst wieder in die aktive Form zurückfaltete, sofern die Lösung bei 23 bis 24 °C gelassen wurde und die Harnstoffkonzentration durch Verdünnung oder Dialyse wieder verringert wurde. Er schloss daraus, dass die Information zur Faltung in der Reihenfolge der Aminosäuren festgelegt ist. Weiters beobachtete er, dass die volle Aktivität erst erreicht wurde, nachdem alle vier Schwefelbrücken wieder korrekt gebildet waren (Anfinsen et al., 1961; Berg et al., 2018, S. 60). Anfinsen postulierte, dass jedes Protein eine eindeutige physiologische Faltung besitzt, die durch die minimale Energie gekennzeichnet ist. Bei der Renaturierung werden lokal Zwischenstufen eingenommen, die lokalen Minima entsprechen, jedoch wieder neu arrangiert werden, bis das absolute Minimum der freien Energie erreicht ist (Anfinsen, 1972).

Während sich kurze Proteine spontan korrekt falten können, sind für längere Proteine Hilfsmittel nötig, die als Chaperone (Anstandsdamen) bezeichnet werden. Chaperone sind Proteine, die sich an die wachsenden Proteinketten anlagern um bei der energetisch günstigsten Faltung behilflich zu sein, oder aber Kammern, in die das ungefaltete oder nur teilweise gefaltete Protein vorübergehend eingesperrt wird, um sich ohne weitere Einflüsse korrekt falten zu können (Alberts et al., 2021, S. 134–135).

6.4.4 Big-Data: Vorhersage dreidimensionaler Strukturen aus der Aminosäuresequenz

Da die Funktionsfähigkeit von Proteinen/Enzymen von der dreidimensionalen Struktur abhängt, war und ist es wichtig, diese genau zu bestimmen. Ursprünglich konnten solche Strukturen nur in aufwändigen Verfahren experimentell mittels Kryoelektronenmikroskopie, Röntgenstrukturanalyse oder NMR bestimmt werden. Neuere Versuche, die dreidimensionale Struktur zu bestimmen, gehen allerdings wieder von der Idee Anfinsens aus, diese alleine aus der Aminosäuresequenz vorherzusagen.

2010 wurde die Firma *Deepmind* gegründet, die Programme zur künstlichen Intelligenz entwickelte. 2014 übernahm Google die Firma. 2016 begann Deepmind mit dem Programm *AlphaFold* Proteinstrukturen alleine aus der Aminosäuresequenz vorherzusagen. Als Basis verwendet AlphaFold experimentell ermittelte Strukturen von Domänen (Teilbereichen) verschiedener Proteine, mit denen AlphaFold trainiert wurde. Aufgrund der immer umfangreicheren – beispielsweise durch NMR, Röntgenstrukturanalyse oder Kryoelektronenmikroskopie experimentell gewonnen – Daten und der größer werdenden Leistungsfähigkeit der Computer, kommen die Vorhersageergebnisse inzwischen sehr nahe an die rein experimentell bestimmten heran. Alle Ergebnisse sind öffentlich zugänglich (Wikipedia, AlphaFold). 2021 wurde *AlphaFold 2* vorgestellt, das als Open-Source-Programm öffentlich zugänglich ist und das bereits 16-mal schneller ist als AlphaFold 1 (Callaway, 2021). Die AlphaFold-Datenbank hat inzwischen (Stand: 2023) 350 000 Proteinstrukturen (siehe Deepmind-Homepage). Im Mai 2024 wurde die weiter verbesserte Version AlphaFold 3 vorgestellt (siehe Online-Zugang).

Es scheint sich jetzt zu erfüllen, was Anfinsen am 11. Dezember 1972 in seiner Rede zum Nobelpreis sagte, die 1973 veröffentlicht wurde (Anfinsen 1973):

> **Empirical considerations of the large amount of data now available on correlations between sequence and three dimensional structure [...], together with an increasing sophistication in the theoretical treatment of the energetics of polypeptide chain folding [...] are beginning to make more realistic the idea of the a priori prediction of protein conformation.**

6.4.5 Fehlfaltung aufgrund von synonymen Mutationen

Bartoszewski et al. (2010) untersuchten die häufigste Mutation des *cystic fibrosis transmembrane conductance regulator (CFTR)* Gens. Hierbei handelt es sich um ein Membranprotein, das einen Chloridkanal bildet, wodurch normalerweise ein Ausgleich in der Salzkonzentration erfolgt. Mutationen führen zu cystischer Fibrose (Mukoviszidose), die durch einen zähen Schleim vor allem in der Lunge gekennzeichnet ist. Die häufigste Mutation besteht in der Deletion von

drei Nucleotiden, die zwei verschiedene Codons betreffen. Dadurch kommt es zum Verlust der Aminosäure Phenylalanin an Position 508 sowie zu einer *synonymen Änderung* des Codons für die Aminosäure 507 von Ile ATC zu Ile ATT (siehe Abbildung 6.18).

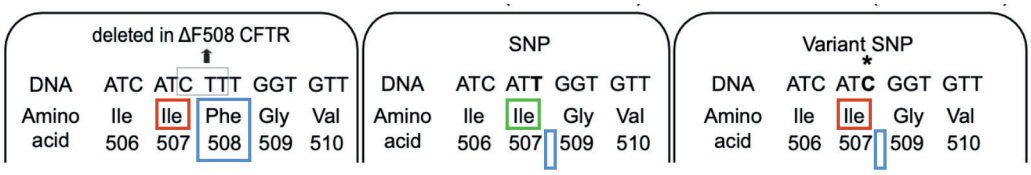


Abb. 6.18 a,b,c Deletion von 3 Nucleotiden (Bartoszewski et al., 2010, adaptiert).

Bartoszewski et al. (2010) zeigten, dass es zu einer Änderung der Sekundärstruktur der mRNA sowie zu einer verminderten Translationsrate kommt (siehe auch Kapitel 6.3). Da das CFTR-Protein vor allem co-translational faltet, kommt es dadurch außerdem zu einer Fehlfaltung des Proteins.

Falls es durch eine Single-Nucleotid-Variante zu einer synonymen Variante des Phenylalanin-Codons (TTC) kommt, dann ergibt die Deletion der drei Nucleotide zwar ebenso einen Verlust der Aminosäure 508 (Phenylalanin), das Codon für Isoleucin bleibt aber erhalten. Dadurch ist die Faltung der mRNA nahezu identisch und das um eine Aminosäure verkürzte Protein voll funktionsfähig. Die Fehlfaltung der mRNA und des Proteins kommt also in diesem Fall nicht durch die Deletion von drei Nucleotiden und daher der Deletion einer Aminosäure zustande, sondern durch die Ausbildung eines synonymen Codons für die Aminosäure Isoleucin. Weiters stellten Bartoszewski et al. (2010) fest, dass sich die Halbwertszeit der fehlgefalteten mRNA kaum von jener des Wildtyps unterscheidet.

Das MDR1-Gen codiert für das P-glycoprotein (P-gp), ein Membranprotein, das dem Schutz der Zelle vor fremden toxischen Substanzen dient. Zytotoxische Substanzen werden durch das P-gp unter ATP-Verbrauch *aktiv* aus der Zelle herausgepumpt, wie beispielsweise Neurotoxine, die aus Gehirnzellen über die Blut-Hirn-Schranke wieder ausgeschleust werden. Durch diesen Mechanismus ergeben sich aber auch Resistenzen gegen viele Wirkstoffe (Berg et al., 2018, S. 443). Werden Medikamente ausgeschleust, so sinkt deren Spiegel und deren Bioverfügbarkeit. Dies geschieht beispielsweise bei Zytostatika, die gegen Krebs wirken sollen und die durch die Ausschleusung nicht in ausreichender Menge zur Verfügung stehen. Es wird daher versucht, durch Inhibitoren die Wirkung des P-gp abzuschwächen, um Medikamente wie Zytostatika nicht daran zu hindern, ihren Bestimmungsort in den Zellen – beispielsweise in Tumorzellen – zu erreichen.

Kimchi-Sarfaty et al. (2007) untersuchten synonyme und nicht-synonyme Polymorphismen des MDR1-Gens, um festzustellen, ob diese die Aktivität des P-gp beeinträchtigen. Die wichtigsten der betrachteten Polymorphismen waren unter anderem eine synonyme Änderung eines Glycin-Codons (GGC zu GGT) in Exon 12 (Mutation C1236T) sowie eine ebenfalls synonyme Ände-

rung eines Isoleucin-Codons (ATC zu ATT) in Exon 26 (Mutation C3435T). Die Experimente der Autor*innen legen nahe, dass es sich bei der synonymen Mutation C3435T (Ile → Ile) um die wichtigste Mutation handelt.

Diese synonyme Mutationen kann sowohl heterozygot als auch homozygot vorkommen und es konnten dabei keine Unterschiede beim Export von Drogen oder von toxischen Substanzen durch das P-gp festgestellt werden. Kimchi-Sarfaty et al. (2007) stellten jedoch eine erhöhte Sensitivität gegenüber den beiden Inhibitoren Cyclosporin-A und Verapamil fest. Bei Cyclosporin-A handelt es sich um ein zyklisches Peptid, das als Immunsupressivum wirkt und zusätzlich das Ausschleusen von Drogen durch das P-Glycoprotein hemmt. Verapamil ist ein Calciumantagonist, der ebenfalls ein Inhibitor von P-gp ist. Die Autor*innen vermuteten, dass durch die synonyme Mutation C3435T eine Veränderung im Faltungsprozess des P-gp stattgefunden hat. Dies konnten sie mit Hilfe der Protease Trypsin nachweisen, die den Wildtyp des P-gp abbaut, die C3435T-Variante aber nur minimal, was nahelegt, dass die Protease aufgrund der veränderten Faltung von P-gp nur mehr eine eingeschränkte Möglichkeit hat, die entsprechenden Peptidbindungen zu spalten. Die Spaltung der Mutante erforderte eine deutlich erhöhte Trypsinkonzentration. Ein weiteres Indiz für die veränderte Struktur der C3435T-Mutante war die veränderte Bindung der Mutante an einen konformationssensitiven Antikörper gegen P-gp.

Weiters stellten Kimchi-Sarfaty et al. (2007) fest, dass durch die synonyme C3435T-Mutation das häufige Isoleucin-Codon ATC (47 % relative Häufigkeit) durch das weniger häufige synonyme Codon ATT (35 %) ersetzt wird. Um die Vermutung, dass die relative Häufigkeit des Isoleucin-Codons an dieser Stelle die Faltung von P-gp beeinflusst, weiter zu untermauern, wurden in-vitro-Tests mit einer weiteren synonymen Mutation (C3435A) durchgeführt. Diese resultiert in einem Isoleucin-Codon (ATA), das mit 18 % eine noch weiter reduzierte relative Häufigkeit aufweist. Die Experimente zeigten, dass die weitere Reduktion der relativen Häufigkeit des Isoleucin-Codons an dieser Stelle die Sensitivität von P-gp noch weiter reduziert. Die Autor*innen kommen zum Schluss, dass bei Änderung von häufigen zu seltenen Isoleucin-Codons an dieser Stelle das Timing der cotranslationalen Faltung und damit die Struktur von P-gp verändert wird.

Rosenberg et al. (2022) untersuchten, ob die dreidimensionale Struktur eines Polypeptids nicht nur von der Reihenfolge der Aminosäuren abhängig ist, sondern im speziellen auch von der Verwendung bestimmter synonymer Codons. Sie fanden, dass die Diederwinkel zwischen bestimmten Aminosäuren nicht nur von der gegenseitigen sterischen Hinderung der Seitenketten dieser Aminosäuren abhängen, sondern auch davon, welche der synonymen Codons für diese Aminosäuren in der codierenden Sequenz für das Protein verwendet wurden.

Wie in Kapitel 1.7 dargelegt, besteht ein Protein bzw. eine Polypeptidkette aus aufeinanderfolgenden Aminosäuren, die jeweils mittels einer Peptidbindung verknüpft sind. Da Peptidbindungen partiellen Doppelbindungscharakter haben, sind diese nicht frei drehbar. Im Gegensatz dazu sind die Bindungen zwischen dem Cα-Atom und der Carbonylgruppe bzw. der Aminogruppe jeweils Einfachbindungen, um die die benachbarten Peptidbindungen frei rotieren kön-

nen. Unter *Diederwinkel* versteht man die Winkel, um die die Einfachbindungen drehbar sind (Berg et al., 2018, S. 45), (siehe Abbildung 6.19).

ϕ ist der Winkel zwischen dem Cα-Atom und N, ψ der Winkel zwischen dem Cα-Atom und der Carbonylgruppe (C=O). Welche Winkel ϕ und ψ annehmen können hängt von der gegenseitigen sterischen Hinderung der Seitenketten ab (Berg et al., 2018, S. 46–47). Die Torsionswinkel werden auch als Diederwinkel oder *Dihedralwinkel* (dihedral angle) bezeichnet, und geben allgemein die Winkel zwischen mathematischen Ebenen an.

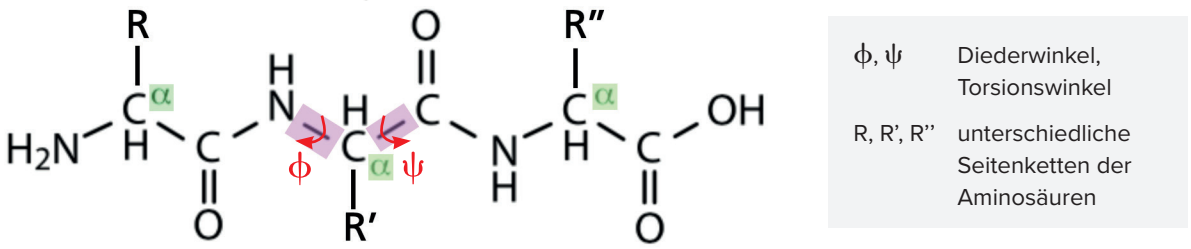


Abb. 6.19 Darstellung der Torsionswinkel einer Polypeptidkette (nach Berg et al., 2018, S. 47, adaptiert).

Die Ergebnisse von Rosenberg et al. werden allerdings angezweifelt. Akeju & Cope (2022) und González-Delgado et al. (2022) weisen auf methodologische Mängel in der Arbeit von Rosenberg et al. hin.

Da reife mRNA im Gegensatz zur prä-mRNA nur wenige nichtcodierende Bereiche hat, die zur Regulierung benutzt werden, vermuteten Orešič & Shalloway (1998), dass die unbenutzte Information, die aus synonymen Codons resultiert, wie ein übergeordneter zweiter Code dazu verwendet werden könnte, eine korrekte cotranslationale Proteinfaltung zu fördern. Sie stellten speziesspezifische Korrelationen zwischen der relativen Häufigkeit synonymer Codons und der Sekundärstruktur von Proteinen fest, indem sie 35 humane und 31 Proteine von E. coli verglichen. Da synonyme Codons die Struktur stromaufwärts oder stromabwärts beeinflussen können, wurden Aminosäuren mit einem Abstand von bis zu zehn Positionen in beide Richtungen miteinbezogen. Bei humanen Proteinen ergab sich *eine Häufung des Codons GAU* (gegenüber dem synonymen Codon GAC) für Asparaginsäure am N-terminalen Ende *einer* α*-Helix*. Bei E. coli fand sich dagegen eine *Häufung des Codons AAU für Asparagin* (gegenüber AAC) am C-terminalen Ende eines β-Faltblattbereiches. Die Ergebnisse zeigten zumindest für einige Codons eine Korrelation zwischen synonymen Codons und der Sekundärstruktur des Proteins. Daraus schlossen Orešič & Shalloway (1998), dass nur sehr wenige humane Proteine in E. coli korrekt gefaltet werden können.

6.5 Genauigkeit und Effizienz der Translation

Die Genauigkeit, mit der Ribosomen ein Protein anhand der mRNA synthetisieren, beläuft sich im Schnitt auf etwa einen Fehler pro 10 000 Aminosäuren (Berg et. al, 2018, S.1061). Decodierungsfehler können beispielsweise durch Stress, Mutationen, Viren bzw. Bakterien oder Ribosom-Sliding hervorgerufen werden oder aber durch Verwendung unterschiedlicher synonymer Codons.

6.5.1 Ribosom-Sliding

Tritt in der Mitte eines Gens eine Poly-A-Sequenz von mindestens drei AAA-Codons auf, so tritt ein Sliding des Ribosoms auf. Smith et al. (2022) fanden Hinweise, dass das Ribosom in einzelnen Schritten ein bis drei Nucleotide rückwärts wandert (Ribosom-Sliding). Wandert es ein oder zwei Schritte rückwärts, so geht das Leseraster verloren und die Translation endet in einem vorzeitigen Stopp-Codon (Koutmou et al., 2015). Bei drei Rückwärtsschritten bleibt zwar das Leseraster erhalten, es ergibt sich aber der Einbau einer zusätzlichen Aminosäure. Die Poly-A-Regionen formen im Ribosom-mRNA-Kanal eine Struktur, die den Verlust des Leserasters bewirkt (Loss of frame). Die Rückwärtsbewegung ist laut Smith et al. (2022) sehr langsam, wodurch Kollisionen mit den nachfolgenden Ribosomen entstehen und die Translationsgeschwindigkeit erheblich erniedrigt wird.

Es kommt, wie bei einer normalen Frame-Shift-Mutation, zu einem vorzeitigen Stopp-Codon und es wird ein *Nonsense-Mediated-Decay* in Gang gesetzt. Im Gegensatz dazu verlangsamen sich die Ribosomen bei einer Folge von synonymen AAG-Lys-Codons zwar, sie gleiten aber nicht und produzieren daher ein korrektes Protein (Koutmou et al., 2015). Durch eine synonyme Mutation, welche ein AAG-Lys-Codon in ein AAA-Lys-Codon verwandelt, können unter Umständen mehrere AAA-Lys-Codons in Folge entstehen und daraus folgend ein Ribosom-Sliding. Bioinformatische Untersuchungen zeigen, dass aufeinanderfolgende AAA-Sequenzen in codierenden Bereichen von Genen unterrepräsentiert sind (Koutmou et al., 2015; Arthur et al., 2015). Ein ähnlicher Vorgang wird in Kapitel 7.3 beschrieben, wo bei aufeinanderfolgenden N1-Methyl-Pseudouridin-Nucleotiden (Slippery-Sequenzen) das Ribosom die mRNA entlang gleitet und das Raster verliert.

6.5.2 Translationsgeschwindigkeit

Frühere Arbeiten über die Verwendung synonymer Codons haben sich an der Häufigkeit, mit der die Codons im Genom auftreten, orientiert. Es wurde angenommen, dass die Häufigkeit eines Codons mit der Translationseffizienz und der Translationsgeschwindigkeit des betreffenden Codons korreliert. Neuere Arbeiten berücksichtigen jedoch für die Zuordnung der Translationsgeschwindigkeit eines Codons die Verfügbarkeit der entsprechenden tRNAs.

Es wird hierbei angenommen, dass Codons, für die es in der Zelle eine hohe Konzentration von tRNAs mit dem entsprechenden Anticodon gibt, schneller translatiert werden als solche, für die die Konzentrationen der passenden tRNAs niedrig sind. Es hat sich in vielen Fällen gezeigt, dass die Konzentration einer tRNA proportional zur Anzahl der für diese tRNA codierenden Gene ist. Die Feststellung der Konzentration einer tRNA erfolgt hier also indirekt. Die Expression der tRNA-Gene ist jedoch in multizellulären Organismen wie dem Menschen von Zelltyp zu Zelltyp verschieden, sodass eine direkte Bestimmung der Konzentrationen aller tRNAs die überlegene Methode ist.

Tuller et al. (2010) untersuchten die Verteilung synonymer Codons entlang der codierenden Sequenz von Genen bei verschiedenen Spezies aller drei Domänen des Lebens. Die Verfügbarkeit der Codon-entsprechenden tRNAs wurde aus der Anzahl der tRNA-Gene ermittelt. Sie stellten eine Korrelation der Anzahl der Genkopien für eine tRNA, die das selbe Codon decodieren können, mit der Verfügbarkeit der tRNA in saccharomyces cerevisiae (Bäckerhefe) fest. Tuller et al. (2010) verwendeten den von dos Reis et al. (2004) eingeführten *tRNA-Adaption-Index* (tAI), wobei in den tAI auch die relative zelluläre Konzentration mit einberechnet wurde. Die einzelnen Faktoren, die in die Definition mit einbezogen wurden, wurden zusätzlich gewichtet.

Der tRNA-Adaption-Index (tAI) gibt die Effizienz der Translation an und wird üblicherweise als geometrischer Mittelwert über das gesamte Genom oder auch nur über einzelne Gene bestimmt.

$$\bar{g} = \sqrt[n]{g_1 * g_2 * g_3 * \cdots * g_n}$$

wobei g_i der tAI eines Codons und $\bar{g}$ der geometrische Mittelwert der tAIs aller Codons ist.

Tuller et al. (2010) wiesen darauf hin, dass innerhalb eines ganzen Genoms die Verteilung relativ gleichförmig ist und dass es daher wichtig ist, einzelne Gene zu untersuchen. Sie gingen dann noch einen Schritt weiter und bezogen den tAI nicht auf das gesamte Genom oder auf ein Gen, sondern sie legten wesentlich kürzere Sequenzabschnitte zugrunde. Sie verwendeten daher einen lokalen tAI (local tAI), der jedes Codon entlang der Sequenz berücksichtigt, das heißt, sie berechneten nicht das geometrische Mittel der tAI-Werte der Codons, die in einem Gen vorkommen, sondern sie ermittelten das geometrische Mittel der tAI-Werte von kurzen Abschnitten von 10 bis 20 Codons Länge entlang der Gensequenz.

Sabi et al. (2010) haben aufgrund dieser Vorgaben ein Berechnungswerkzeug für den tAI programmiert. Unter Verwendung des ermittelten lokalen tAI, stellten Tuller et al. (2010) weiters fest, dass innerhalb der ersten 30 bis 50 Codons der mRNA eine evolutionär konservierte *codon ramp* existiert, die für eine langsame Translation zu Begin derselben verantwortlich ist. Verschiedene Spezies haben jedoch jeweils einen unterschiedlichen lokalen tAI. Bei der codon

ramp handelt es sich um einen Bereich, in dem bevorzugt Codons mit niedrigen tAI-Werten Verwendung finden, wodurch die Translation verlangsamt wird und die Dichte der Ribosomen zunimmt. Die Ribosomendichte innerhalb der ersten 30 bis 50 Codons verhält sich also umgekehrt proportional zur Translationsgeschwindigkeit und damit zur Translationseffizienz.

Es stellen sich hier gegensätzliche Effekte (trade off) ein:

- Die Verringerung der Translationseffizienz durch langsame Translation aufgrund verlangsamter Geschwindigkeit der Ribosomen wird hervorgerufen durch Verwendung von Codons mit niedrigen tAI-Werten.
- Die Dichte aufeinanderfolgender Ribosomen ist innerhalb der ersten 30 bis 50 Codons größer – durch die verringerte Geschwindigkeit im Anfangsbereich der mRNA wird aber eine Kollision der Ribosomen verhindert (positionsabhängige ribosomale Dichte/position-dependent ribosomal density).
- Ribosomen, die bei langsamen Codons mit niedrigen tAI-Werten jenseits der Initialphase kollidieren, bleiben längere Zeit auf einem Transkript erhalten und verhindern so die Produktion des Proteins. Tuller et al. (2010) nehmen daher an, dass durch die codon ramp eine solche Kollision der Ribosomen jenseits der Initialphase vermieden wird.

Tuller et al. (2010) ermittelten für Prokaryoten eine mittlere Länge der *codon ramp* von 24 Codons, für Eukaryoten eine mittlere Länge von 35 Codons. Obwohl seltene Codons aufgrund ihrer langsameren Translationsgeschwindigkeit als nicht optimal für die Genexpression angesehen wurden, enthalten die mRNAs vieler Proteine an bestimmten Stellen codierender Sequenzen große Cluster seltener synonymer Codons (conserved rare codon clusters, CRCC).

Chaney et al. (2017) stellten daher die Frage, ob die seltenen synonymen Codons an bestimmten spezifischen Positionen innerhalb homologer Proteine konserviert sind, obwohl der *codon usage bias* zwischen verschiedenen Spezies variiert. Obwohl gezeigt wurde, dass eine Änderung der relativen Codon-Seltenheit mit der lokalen Translationsrate korreliert, fanden Chaney et al. (2017), dass diese ebenso mit der cotranslationalen Faltung sowie dem tAI korreliert.

Im Gegensatz zu Tuller et. al (2010), die seltene Codons am Anfang der codierenden Sequenz untersuchten, analysierten Chaney et al. (2017) Cluster innerhalb von Proteindomänen. Chaney et al. (2017) konnten für viele Eukaryoten, Bakterien und Archaeen zeigen, dass die Positionen von Clustern seltener Codons bei einigen Familien homologer Proteine konserviert sind und dass damit die *Position wichtiger ist, als die Identität* der Codons. Daher gehen Chaney et al. (2017) von einer funktionellen Rolle der Cluster aus.

Außerdem stellten die Autor*innen fest, dass die meisten der konservierten Cluster sich eher in konservierten Proteindomänen befinden, und nicht dazwischen. Wie bereits in Kapitel 6.4.1 dargelegt, bestehen Proteine meist aus mehreren funktionellen Domänen, die durch unstrukturier-

te Bereiche miteinander verbunden sind (Alberts et al., 2021, S. 142). Die einzelnen Domänen falten sich schon während der Entstehung unabhängig voneinander (cotranslationale Faltung).

Diese sogenannten Domänen haben spezifische Teilfunktionen, die zusammen die Gesamtfunktion des Proteins bestimmen. Es war daher naheliegend, anzunehmen, dass Cluster seltener Codons in Verbindungssequenzen zwischen Domänen liegen. Das würde die Translation in diesen Bereichen verlangsamen und der bereits synthetisierten Domäne Zeit geben, sich korrekt zu falten, bevor weitere Teile des Proteins synthetisiert werden. Chaney et al. (2017) zeigten aber, dass die Cluster der seltenen Codons innerhalb von Domänen liegen. Dies bedeutet, dass im Laufe der Evolution nicht nur die Aminosäuresequenzen homologer Proteine beibehalten werden mussten, um die Funktion dieser Proteine zu gewährleisten, sondern auch, dass Aminosäuren an bestimmten Positionen im Protein in der mRNA durch seltene und nicht durch häufige synonyme Codons codiert werden müssen. Es geht also um die Frage, ob es im Laufe der Evolution einen Selektionsdruck gegeben hat, der bewirkt hat, dass Aminosäuren an bestimmten Positionen im Protein durch seltene Codons codiert werden.

Kirchner et al. (2017) untersuchten die durch einen synonymen Polymorphismus veränderte Funktion eines Proteins (Cystic Fibrosis Transmembrane Conductance Regulator, CFTR), welches in der Zellmembran integriert ist und einen Anionenkanal, in diesem Fall einen Chloridkanal, bildet. Die Dysfunktion dieses Kanals vermindert den Transport von Cl^--Ionen vom Zellinneren nach außen und ist die Ursache von *Cystischer Fibrose (Mukoviszidose)*. Die Autor*innen untersuchten einen synonymen Polymorphismus T2562G, bei dem das Codon ACU der Aminosäure (Threonin) durch das synonyme Codon ACG ersetzt wird. Der Austausch des Codons ACU durch das synonyme Codon ACG verändert die *lokale Translationsgeschwindigkeit*, wodurch die Stabilität und Funktion des CFTR-Proteins beeinträchtigt wird. Das veränderte Codon paart mit einem Anticodon einer tRNA, die nur in geringer Menge zur Verfügung steht und die besonders selten in humanen Bronchialepithelzellen vorkommt – im Gegensatz zu anderen Geweben. Daraus leiten die Autor*innen einen gewebespezifischen Effekt ab.

Der Wildtyp ACU mit der tRNA mit dem Anticodon AGU hat eine hohe Translationsgeschwindigkeit. Das mutierte Codon ACG mit der tRNA mit dem Anticodon CGU hat eine niedrige Translationsgeschwindigkeit, da die entsprechende tRNA nur in geringer Konzentration vorhanden ist. Kirchner et et al. (2017) nehmen an, dass der T2562G-Polymorphismus, sofern er gemeinsam mit Missense-Mutation vorkommt, das Potential hat, die Schwere der Krankheit zu modifizieren.

Rauscher et al. (2021) beschreiben einen positiven epistatischen Effekt zwischen dem T2562G-Polymorphismus und einer in der Nähe auftretenden Missense-Mutation. Die Autor*innen stellen fest, dass dieser die Folgen der Missense-Mutation weitestgehend aufhebt, obwohl beide einzeln schädliche Effekte – wie fehlerhafte Faltung und verminderte Funktion des Kanals – haben.

7 Erweiterte Codierung und Verwendung synonymer Codons in der synthetischen Biologie

7.1 Codon-Optimierung

Viele Proteine wie Insulin, manche Wachstumsfaktoren, Hormone, Antikörper oder Gerinnungsfaktoren werden heute synthetisch hergestellt. Früher wurden viele dieser Medikamente aus Schlachtabfällen (Insulin beispielsweise aus dem Pankreas von Schweinen oder Rindern) oder auch aus menschlichem Blutplasma hergestellt (Geschichte der Diabetologie, Wikipedia). Wachstumshormone wurden sogar aus den Hypophysen (Hirnanhangdrüsen) Verstorbener extrahiert. Die Gefahr einer Kontamination mit Erregern (z. B. Hepatitis C), Prionen oder allergischer Reaktionen auf artfremdes Eiweiß bei nicht ausreichender Reinigung, war relativ groß. Es werden daher zunehmend andere Methoden verwendet (Ein Medizinskandal vor Gericht, NZZ online, 2008).

Einfache Moleküle können unter Umständen vollständig chemisch synthetisiert werden. Bei komplizierten Strukturen, etwa Hormonen, Antikörpern oder Impfstoffen wird zunehmend eine Produktion mittels heterologer Genexpression gewählt. Dies ist jedoch noch nicht bei allen Substanzen möglich. Die heterologe Genexpression geschieht meist durch *Einschleusen von humanen Genen* in andere Organismen wie E. coli oder Hefe *(saccharomyces cerevisiae)*. In Bioreaktoren können relativ einfach und kostengünstig große Mengen des gewünschten menschlichen Proteins im industriellen Maßstab produziert werden.

18 der 20 kanonischen Aminosäuren werden durch mehr als ein synonymes Codon codiert. Ausnahmen sind Methionin und Tryptophan, die beide nur durch jeweils ein Codon codiert werden. Für jede Spezies lässt sich bestimmen, welche der synonymen Codons einer bestimmten

Aminosäure häufiger oder seltener verwendet werden *(codon bias)*. Dieser Codon-Bias ist spezies-spezifisch und unterscheidet sich mitunter erheblich vom Codon-Bias anderer Spezies Insbesondere unterscheidet sich der Codon-Bias zwischen humanen und bakteriellen Zellen wie E. coli oder der Hefe, die häufig in Bioreaktoren verwendet werden.

Um die Ausbeute eines biotechnologisch herzustellenden menschlichen Proteins zu erhöhen, werden Codons, die im Zielorganismus (Hefe, E. coli etc.) selten vorkommen, gegen solche synonyme Codons ausgetauscht, die im Zielorganismus häufig sind und daher eine effiziente Expression gewährleisten. Dieser Vorgang wird als *Codon-Optimierung (codon optimization/optimisation)* bezeichnet. Obwohl die Ausbeute erhöht wird, kann dieses Verfahren die Sicherheit und Wirksamkeit der hergestellten Proteine und Impfstoffe stark beeinträchtigen (Mauro & Chappell, 2014).

Es gibt inzwischen zahlreiche Firmen, die Codon-Optimierungen für verbesserte Expression in Zielorganismen anbieten. Diese Optimierungen erfolgen ausschließlich durch Computeralgorithmen. *Eurofins Scientific* mit Sitz in Luxemburg bietet Optimierungen mittels der optimisation software *GENEius* an, oder auch *Thermofisher scientific* mit *Invitrogen GeneArt Gene Synthesis*, um nur zwei zu nennen. Es können in allen Fällen Sequenzen direkt über die Webseite hochgeladen werden mit Angabe eines Zielorganismus.

Mauro & Chappell bemängelten 2014, dass die verwendeten Programme sehr unterschiedliche Optimierungsstrategien anwenden, wie die ausschließliche Verwendung der häufigsten Codons oder das ausschließliche Ersetzen der seltensten Codons. Inzwischen können jedoch bei den meisten Firmen, die Optimierungssequenzen anbieten, vom Auftraggeber einige Parameter bezüglich der zu ersetzenden Codons ausgewählt werden.

7.2 Codon-Harmonisierung

Da bei der Codon-Optimierung, bei der *alle seltenen Codons durch häufige ersetzt werden*, Probleme auftreten können, wurde ein neuer Ansatz entwickelt: die sogenannte *Codon-Harmonisierung*. Laut Angov et al. (2008) werden hierbei synonyme Substitutionen vorgenommen, sodass die Abfolge von seltenen und häufigen Codons des Ursprungsorganismus – zumindest in manchen Bereichen – direkt auf den Zielorganismus übertragen wird (siehe Abbildung 7.1). Dadurch versucht man, die Produktionsbedingungen des Proteins im Zielorganismus an jene des Ursprungsorganismus anzupassen.

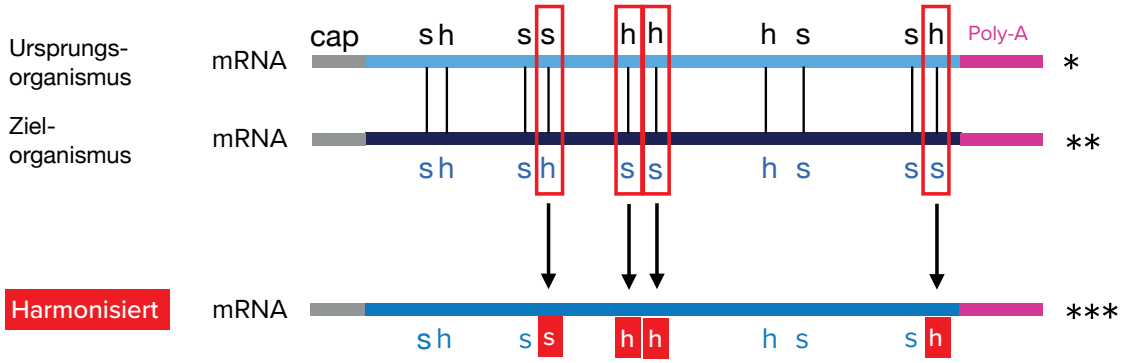


Abb. 7.1 Darstellung der Codon-Harmonisierung für heterologe Genexpression.

Nach der Identifizierung von Clustern von Codons, welche langsam translatiert werden, und einzelnen Codons, die eine Pause in der Translation bewirken, wurden die Codons im Gen durch synonyme Codons ersetzt, welche im Zielorganismus (hier E. coli) ebenfalls langsam translatiert werden bzw. eine Pause in der Translation hervorrufen.

Es zeigte sich, dass häufig verwendete synonyme Codons mit höher geordneten Protein-Domänen wie α-Helix und β-Faltblatt assoziiert sind. Seltene Codons kommen vor allem bei den Verbindungsstücken *(loops)* zwischen den α-Helix und β-Faltblatt-Bereichen vor und sie codieren außerdem für Aminosäuren mit sperrigen hydrophoben Seitenketten. Dies zeigt deutlich, dass solche Codons bevorzugt an solchen Stellen vorkommen, an denen sich die Struktur ändert und die wichtig für die korrekte Faltung des Proteins sind.

Wright et al. veröffentlichten 2021 einen weiter entwickelten Algorithmus namens CHARMING (für *Codon HARMonizING*), um mRNA-Sequenzen für heterologe Genexpression zu optimieren. Dieser Algorithmus definiert den Codon-Bias genauer und auch die angenommene Konzentration der tRNA für die einzelnen Codons wird berücksichtigt. Je nach spezifischer Anwendung können unterschiedliche Harmonisierungsstrategien vorteilhafter sein. So kann beispielsweise der Codon-Bias über das gesamte Gen oder nur über ein Sliding-Window anpassbarer Länge bestimmt

werden. Die Beurteilung, ob ein Codon bezüglich der Translationseffizienz als optimal anzusehen ist, kann ebenfalls anhand unterschiedlicher Kriterien beurteilt werden:

- Der *CAI (Codon Adaption Index)* ermittelt jene Codons, die mit der größten Häufigkeit in hoch exprimierten Genen auftreten (Sharp & Li, 1987).
- Der *tAI (tRNA Adaptation Index)* ermittelt die Häufigkeit der Codons basierend auf den Konzentrationen der tRNAs mit den jeweiligen Anticodons (dos Reis, 2004).
- Der *%MinMax-Index* ermittelt die häufigsten Codons bezogen auf alle *open reading frames* des gesamten Genoms eines Organismus (Rodriguez et al., 2017).

Ein weiteres wichtiges Unterscheidungskriterium bezüglich der Harmonisierung ist die Einteilung synonymer Codons in

- *absolut häufige Codons* – wobei mit der Häufigkeit aller Codons verglichen wird und
- *relativ häufige Codons* – wobei nur mit der Häufigkeit der synonymen Codons für eine einzige Aminosäure verglichen wird.

Der von Wright et al. (2021) entwickelte CHARMING-Algorithmus harmonisiert eine Sequenz in einem iterativen Prozess. Vor Beginn der Berechnungen müssen verschiedene Parameter angegeben werden, wie die zu harmonisierende Sequenz, welche Art der Codon-Harmonisierung verwendet werden soll, die Größe des Sliding-Windows sowie die Codon-Usage-Tabellen des Ursprungs- und Zielorganismus. Außerdem kann auf Wunsch des Auftraggebers auch ein Set verschiedener Harmonisierungen berechnet werden.

7.3 mRNA-Impfstoffe

Im Dezember 2019 wurde eine neuartige Lungenkrankheit in China gemeldet. 2020 wurde das Virus als ein Beta-Coronavirus identifiziert, wie SARS *(Severe acute respiratory syndrome coronavirus)* und MERS *(Middle East respiratory syndrome coronavirus)*, (Robert-Koch-Institut, Steckbrief, online). Die Sequenz des neuen Virus, das in der Folge als SARS-CoV-2 bezeichnet wurde *(Severe acute respiratory syndrome coronavirus type 2)*, war bereits wenige Wochen nach Ausbruch der Krankheit öffentlich zugänglich. Hierbei handelt es sich um ein membranumhülltes einsträngiges RNA-Virus mit positiver Leserichtung *(+ssRNA)*. Aus der Hülle des SARS-CoV-Virus ragen spikeförmige Glykoproteine *(Spike-Proteine)* nach außen, welche die rezeptorbindende Domäne tragen (DocCheck Flexikon, online).

Als Impfstoff wird ausschließlich eine synthetisch hergestellte, 4284-Nucleotid-lange RNA verwendet, deren Sequenz mit jener des *kompletten Spike-Proteins* des SARS-CoV-2-Virus über-

einstimmt. Da es sich nur um einen kleinen Teil des gesamten Virus-RNA Genoms handelt, kann dieser Impfstoff keine Infektion hervorrufen.

Der mRNA-Impfstoff wird in Lipid-Nanopartikel (LNP) verpackt, um in die Zelle eingeschleust werden zu können, wo er, wie eine normale mRNA, durch Ribosomen in das Spike-Protein translatiert wird. Damit die Impfstoff-RNA von der Zelle nicht als fremd erkannt wird, werden eine Kappe (cap), eine 5‘-UTR, eine 3‘-UTR (untranslatierte Region) sowie ein Poly-A-Schwanz angehängt (Hildt, 2022; Patentanmeldung WO2022223617 (A1)).

Die Biochemikerin Katalin Karikó und der Immunologe Drew Weissman bekamen 2023 den *Nobelpreis für Medizin oder Physiologie* für ihre Forschungen, die mRNA so zu modifizieren, dass sie vom menschlichen Immunsystem nicht als fremd erkannt wird und daher nicht sofort abgebaut wird. Dies erreichten sie dadurch, dass sie alle Uridin-Nucleotide durch N1-Methyl-Pseudouridin-Nucleotide (m1Ψ) ersetzten. Dadurch kann eine ausreichende Menge des Spike-Proteins, beispielsweise des Spike-Protein des Sars-CoV-2-Virus, von den zellulären Ribosomen erzeugt werden, um das Immunsystem zur Bildung von Antikörpern gegen das Spike-Protein anzuregen (Sahin et al., 2014). Abbildung 7.2 zeigt die Umwandlung von Uridin in N1-Methyl-Pseudouridin.

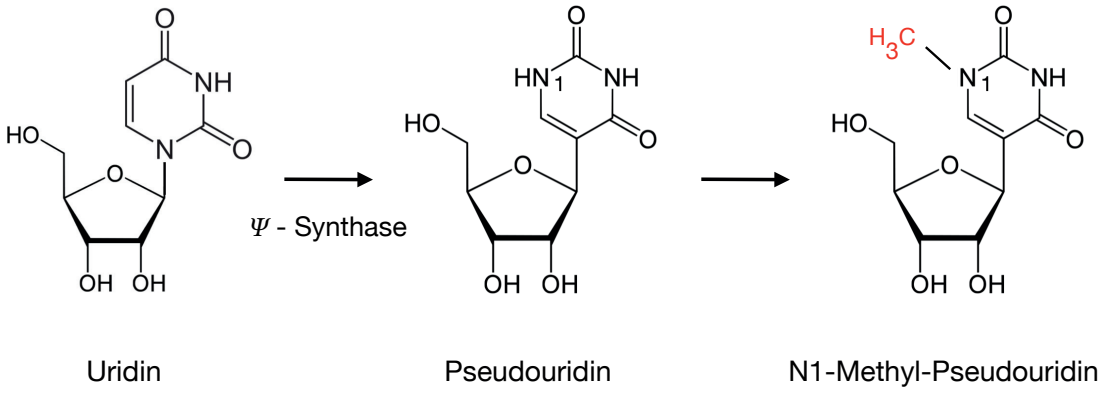


Abb. 7.2 Umwandlung von Uridin in N1-Methyl-Pseudouridin (Morais et al., 2021, modifiziert)

Die modifizierten Codons der RNA, bei welchen U durch m1Ψ ersetzt wird, codieren jeweils für die selbe Aminosäure wie das ursprüngliche Codon. In der internationalen Patentanmeldung WO2022223617 (A1) von Pfizer (siehe espacenet) wird angegeben, dass die Impfstoff-RNA *ausschließlich* m1Ψ statt Uridin enthält. Außerdem wurde eine Codon-Optimierung vorgenommen.

> **The RNA does not contain any uridines; instead of uridine the modified N1-methylpseudouridine is used in RNA synthesis.**

Zudem kommt am Ende der mRNA jeweils ein doppeltes Stopp-Codon **ΨGA ΨGA** statt eines UAA zum Einsatz, um ein versehentliches *read-through* zu verhindern (Berthub online, Morais et al., 2021).

Die Modifikation der in der mRNA vorkommenden Uridine zu m1Ψ hat allerdings auch Nachteile. So wurde eine +1 Rasterverschiebung beobachtet. Dadurch ergeben sich bei Verwendung von m1Ψ vorzeitige Stopp-Codons und damit verkürzte Proteine (Tursi & Weiner 2023; Mulroney et al. 2024). Die Rasterverschiebungen werden durch sogenannte *Slippery-Sequenzen* in der mRNA hervorgerufen, das sind drei aufeinanderfolgende N1-Methyl-Pseudouridin-Nucleotide m1Ψ m1Ψ m1Ψ, an denen das Ribosom entlang gleitet und das Raster verliert (Mulroney et al. 2024). Dies ähnelt dem in Kapitel 6.5.1 beschriebenen Ribosom-Sliding entlang einer Sequenz aus mehreren AAA-Codons.

Tursi & Weiner (2023) stellen fest, dass bei m1Ψ-Modifikationen in acht Prozent der Fälle diese +1 Rasterverschiebung vorkommt, woraus falsch gefaltete, verkürzte oder anderweitig veränderte Proteine resultieren können, die ebenfalls eine *eigene Immunantwort* hervorrufen können. Dies kommt aber bei anderen untersuchten Modifikationen wie z. B. bei 5-Methyl-Cytidin nicht vor. Da es mehrere Slippery-Sequenzen in der Spike-Protein-RNA gibt, ergeben sich auch mehrere falsche und daher verkürzte Proteine.

Mulroney et al. (2024) ersetzten in den Slippery-Sequenzen (m1Ψ m1Ψ m1Ψ) einzelne Codons durch synonyme Codons. Dadurch gab es keine drei aufeinanderfolgenden m1Ψ-Pseudouridin-Nucleotide mehr, wodurch das Gleiten des Ribosoms verhindert wurde und keine verkürzten Proteine mehr auftraten. Dabei kann die m1Ψ m1Ψ m1Ψ Sequenz innerhalb eines Codons auftreten, was Phenylalanin codiert. Diese Sequenz könnte beispielsweise durch m1Ψ m1Ψ C, das ebenfalls für Phenylalanin codiert, ersetzt werden. Sofern die drei m1Ψ zu verschiedenen Codons gehören (N m1Ψ m1Ψ – m1Ψ N N) genügt es eines der Codons durch ein synonymes zu ersetzen.

Zhang et al. (2023) haben einen neuen Algorithmus zum Optimieren einer mRNA-Sequenz für Impfungen gegen Sars-CoV-2 entwickelt. Wichtigste Herausforderungen sind die geringe Effizienz, die Instabilität der mRNA und die Notwendigkeit, die zelluläre Immunantwort der Zellen, die den mRNA-Impfstoff aufgenommen haben, zu umgehen, und die optimale Codon-Verwendung, um eine erhöhte Expression zu erzielen. Das Ziel ist es, eine starke Immunantwort des Immunsystems des menschlichen Organismus gegen das Spike-Protein hervorzurufen. Hierbei sollte keine zelluläre Immunantwort der Zellen, die den RNA-Impfstoff aufgenommen haben, auftreten. Diese Zellen könnten den mRNA-Impfstoff als fremd erkennen und abbauen. Die Impfstoff-mRNA muss das Cytoplasma der Zellen nicht nur unbeschadet erreichen, sondern auch lange genug überleben, um genügend Antigen für eine Immunantwort zu bilden (Blakney, 2023).

Der von Zhang et al. (2023) entwickelte Algorithmus namens *linear design* ermöglichte eine Berechnung der optimalen Codonverteilung für das Spike-Protein von Sars-CoV-2 innerhalb von elf Minuten. Außerdem konnte der Antikörpertiter bei Mäusen auf das 128-fache erhöht werden, was durch Tests bewiesen wurde.

Zhang et al. verwendeten *lattice parsing*, ein Verfahren, das auch in der Linguistik angewendet wird und wobei die lineare Anordnung in einzelne Abschnitte zerlegt wird. Mit dieser Methode werden solche Sequenzen gesucht, die die größte grammatikalische Korrektheit aufweisen. In einem ersten Schritt wird auf die *Minimum-Free-Energy-Change (MFE)* und den *Codon-Adaption-Index (CAI)* optimiert. Je geringer die beiden sind, desto stabiler ist die mRNA (Blakney, 2023; Zhang et al., 2023).

Die Wildtyp-RNA enthält nur lineare Bereiche, während die optimierte RNA viele doppelsträngige Bereiche enthält *(Loops, Hairpins)*, wodurch sie von Nucleasen weniger angreifbar wird und daher stabiler ist. Insgesamt muss eine Balance zwischen Stabilität und Codon-Optimalität gefunden werden, da diese gegenläufig sein können. Die bisherigen proprietären Programme der verschiedenen Unternehmen bezogen sich nur auf die größere Expression, berücksichtigten aber die mRNA-Stabilität nicht (Blakney, 2023).

Zhang et al. testeten die aufgrund ihrer Algorithmen berechneten Impfstoffe in vivo nur an Mäusen. Blakney (2023) bemängelt, dass die herausragenden Ergebnisse jedoch nicht direkt auf den Menschen übertragbar seien, da das angeborene Immunsystem der Mäuse, im Gegensatz zu jenem von Menschen, schwächer ist, was die Reaktion auf fremde RNA betrifft. Außerdem seien weder die 5‘-UTR noch die 3‘-UTR-Bereiche bei der Modifizierung berücksichtigt worden.

7.4 Umprogrammieren des genetischen Codes

Künstliche Veränderungen des genetischen Codes und Einbau nicht-kanonischer Aminosäuren in Proteine eröffnen neue Möglichkeiten, Proteine mit bisher nicht bekannten Eigenschaften zu synthetisieren, die bisher im Laufe der Evolution noch nicht aufgetreten sind. Hierzu werden verschiedene Ansätze gewählt wie das Umprogrammieren von Stopp-Codons oder das Umprogrammieren von einigen synonymen Codons einer der 20 kanonischen Aminosäuren, dem Erhöhen der Anzahl der zwei vorhandenen Basenpaare durch Synthese zusätzlicher neuer Basenpaare oder dem Erweitern der Codonlänge von drei (Triplett-Code) auf vier (Quadruplett-Code) Basen.

7.4.1 Beladung von tRNAs mit einer Aminosäure

Unabhängig davon, welcher Ansatz gewählt wird, muss eine tRNA mit der entsprechenden nicht-kanonischen Aminosäure (ncAA, non canonical Amino Acid) beladen werden, ob es sich nun um eine natürlich vorkommende nicht-kanonische oder eine synthetische Aminosäure handelt. Zur Beladung einer tRNA mit einer Aminosäure ist ein Enzym namens Aminoacyl-tRNA-Synthetase (aa-tRNA-Synthetase) nötig. Dieses Enzym hat mehrere wichtige Bindungsstellen, eine spezifische Bindungsstelle für eine der 20 kanonischen Aminosäuren und eine zur Bindung einer unbeladenen tRNA mit dem entsprechenden Anticodon. Für jede kanonische Aminosäure gibt es genau eine aa-tRNA-Synthetase, also insgesamt 20.

An der zweiten Bindungsstelle der aa-tRNA-Synthetase bindet eine tRNA mit einem bestimmten Anticodon, welches jener Aminosäure entspricht, die an der ersten Bindungssstelle bindet. Diese Bindungsstelle für die tRNA ist aber weniger spezifisch, wodurch mehrere tRNAs gebunden werden können, die entsprechende Anticodons für alle synonymen Codons für eine bestimmte Aminosäure haben (Isoakzeptoren). Abbildung 7.3 zeigt die aa-tRNA-Sythetase ohne und mit Bindung an Ihre Aminosäure und die dazugehörige tRNA.

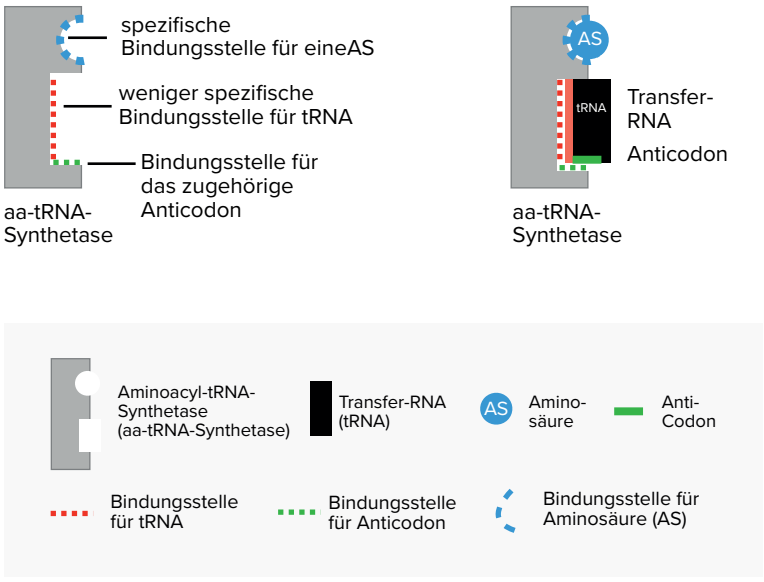


Abb. 7.3 Aminoacyl-tRNA-Synthetase mit Bindungsstellen, ohne und mit Beladung durch eine Aminosäure und eine tRNA.

Es gibt nicht für jedes Codon eine eigene tRNA sondern es sind aufgrund der Wobble-Paarung weniger als die möglichen 61 (Anzahl der Codons abzüglich dreier Stopp-Codons). Die Anzahl der tRNAs mit unterschiedlichen Anticodons variiert von Spezies zu Spezies, so gibt es beim Menschen tRNAs mit Anticodons für 48 der 61 möglichen Codons (Alberts et al., 2021, S. 272).

Da die Zahl der tRNAs zwischen 20 (Anzahl der kanonischen Aminosäuren) und 61 (Anzahl der Codons für die kanonischen Aminosäuren) liegt, gibt es einige tRNAs, die mit mehreren Codons paaren können (Wobble-Effekt) und einige Aminosäuren, für die mehrere tRNAs zur Verfügung stehen (Alberts et al., 2021, S. 272).

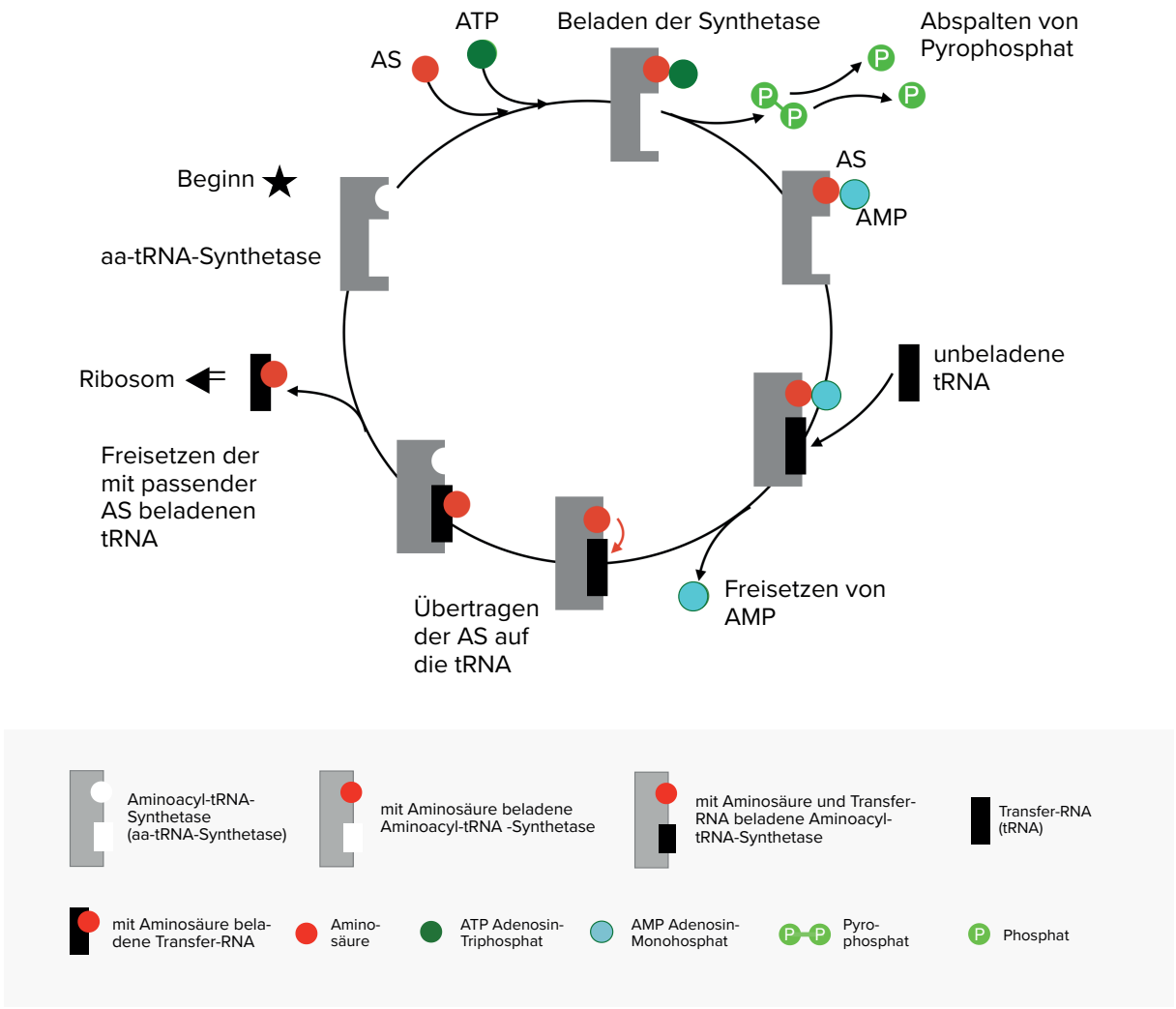


Abb. 7.4 Beladung einer Transfer-RNA mit einer Aminosäure mit Hilfe einer aa-tRNA-Synthetase (DocCheck Flexikon, Aminoacyl-tRNA-Synthetase, adaptiert).

Die aa-tRNA-Synthetase bindet zuerst an ihre spezifische Aminosäure (AS) sowie an ein Adenosintriphosphat (ATP) als Energielieferant. Beim Aktivieren der AS geht diese unter Abspaltung eines Pyrophosphats (P-P) eine Säureanhydridbindung mit der aa-tRNA-Synthetase ein. Im nächsten Schritt wird an die zweite Bindungsstelle der aa-tRNA-Synthetase eine unbeladene tRNA gebunden, deren Anticodon komplementär zu einem der synonymen Codons für die bereits gebundene Aminosäure ist. Die bereits an die aa-tRNA-Synthetase gebundene Aminosäure wird unter Abspaltung des verbliebenen Adenosinmonophosphats (AMP) auf das 3‘-Ende der tRNA übertragen, wo sie eine kovalente Esterbindung eingeht. Die mit ihrer Aminosäure beladene tRNA wird freigesetzt und kann nun am Ribosom zur Verlängerung der Aminosäurekette verwendet werden (siehe Kapitel 1.10.2). Die aa-tRNA-Synthetase kann nun erneut an eine

weitere spezifische Aminosäure binden und eine weitere entsprechende tRNA mit dieser beladen (DocCheck Flexikon, Aminoacyl-tRNA-Synthetase), (Alberts et al., 2021, S. 272).

7.4.2 Beladung von tRNAs mit einer nicht-kanonischen Aminosäure

Um nun nicht-kanonische Aminosäuren in ein Protein einbauen zu können, muss zuerst eine aa-tRNA-Synthetase an die neue Aminosäure und die dazugehörige tRNA angepasst werden. Es muss ein orthogonales System aus einer ncAA, einer angepassten aa-tRNA-Synthetase und einer tRNA erzeugt werden.

Orthogonal bedeutet, dass die neue (orthogonale) tRNA nicht von den ursprünglich vorhandenen aa-tRNA-Synthetasen erkannt werden darf, dass die vorhandenen tRNAs nicht durch die neue Synthetase beladen werden dürfen (Hahn et al., 2013) und dass die neue, angepasste aa-tRNA-Synthetase ausschließlich die gewünschte ncAA, nicht aber eine der kanonischen Aminosäuren akzeptiert. Dies wird dadurch erreicht, dass im aktiven Zentrum von Synthetasen viele Mutationen erzeugt werden, aus denen durch Selektion passende Synthetasen ausgewählt werden.

7.4.3 Umprogrammieren von Stopp-Codons oder Sense-Codons

Wenn das Ribosom im Rahmen der Proteinsynthese an das Stopp-Codon in der mRNA gelangt, wird das fertige Protein mit Hilfe von Releasefaktoren freigesetzt, die statt einer tRNA an die A-Stelle am Ribosom binden. Bei Prokaryoten kommen dabei für die drei Stopp-Codons zwei verschiedene Releasefaktoren zum Einsatz, Releasefaktor 1 (RF1) und Releasefaktor 2 (RF2). Für Prokaryoten gilt (Buttlar et al., 2020, S. 129):

Stopp-Codon		RF1	RF2
TGA	opal		X
TAA	ochre	X	X
TAG	amber	X	

Soll nun ein Stopp-Codon als Codon für eine ncAA innerhalb einer Protein-codierenden Sequenz verwendet werden, so muss auch der jeweilige Releasefaktor (RF1 bzw. RF2) geändert bzw. außer Kraft gesetzt werden, weil der Releasefaktor bei jedem Vorkommen des Stopp-Codons in der codierenden Sequenz zum Abbruch der Proteinsynthese an dieser Stelle und zur Freisetzung eines unvollständigen Proteinfragments führen würde.

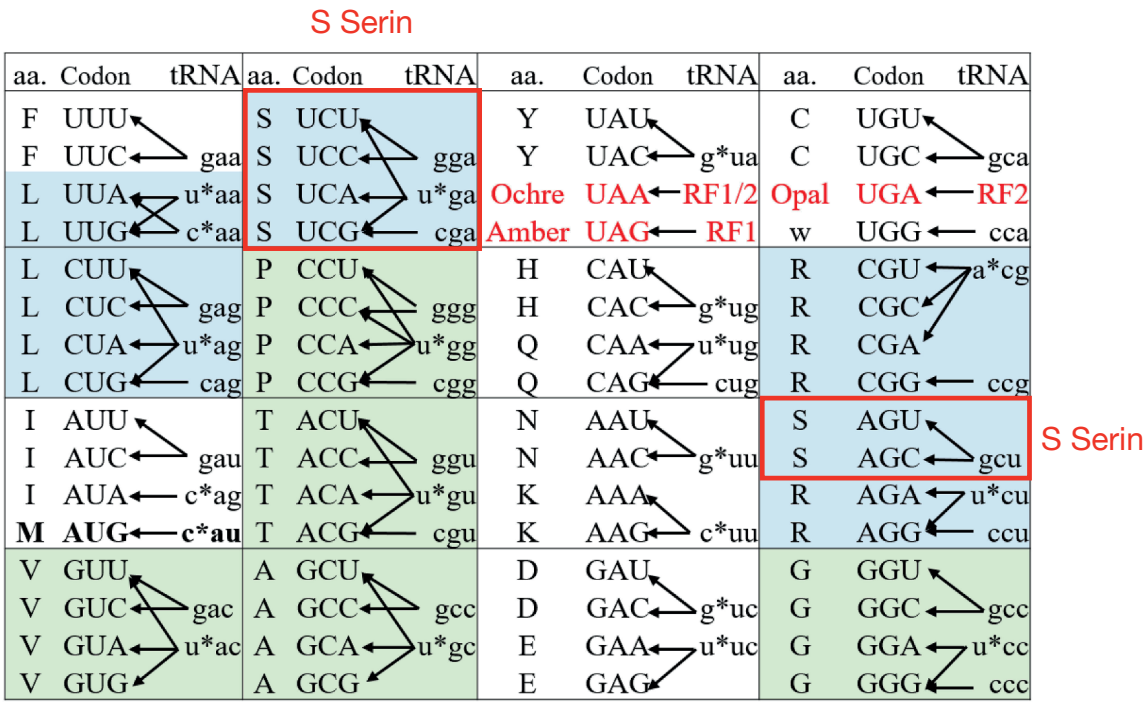
Aus der Tabelle ergibt sich, dass bei Prokaryoten wie E. coli zwei Stopp-Codons, von jeweils nur von einem der beiden Releasefaktoren erkannt werden (TGA, TAG). Nur diese können für ncAA umprogrammiert werden. Für das Stopp-Codon TAA müssten beide Releasefaktoren außer Kraft gesetzt werden. Dann würde aber kein Releasefaktor für die reguläre Beendigung der Proteinsynthese am Ende der codierenden Sequenz mehr zur Verfügung stehen. Das TAG-

Stopp-Codon ist das am seltensten vorkommende Stopp-Codon in E. coli und wird daher oft zur Einführung von ncAA verwendet (Cui et al., 2020).

Fredens et al. (2019) ersetzten in allen Genen eines bestimmten E.-coli-Stammes, deren codierende Sequenz mit einem TAG-Stopp-Codon endet, dieses durch ein anderes Stopp-Codon. Das war notwendig, weil sonst in Abwesenheit von RF1 die Synthese der von diesen Genen codierten Proteine nicht korrekt am Stopp-Codon beendet würde. Dies wurde durch die Totalsynthese des E.-coli-Genoms (4 Mbp) erreicht (Fredens et al., 2019).

In dem von Robertson et al. (2021) verwendeten E.-coli-Stamm wurden durch Totalsynthese des Genoms nicht nur alle TAG-Stopp-Codons durch andere Stopp-Codons ersetzt, sondern es wurden auch zwei der synonymen Serin-Codons aus dem Viererblock (TCG und TCA) durch andere der verbleibenden synonymen Serin-Codons ersetzt. Das Gen für RF1 wurde ebenfalls entfernt. Auch die Gene für $tRNA^{Ser}_{CGA}$ und für $tRNA^{Ser}_{UGA}$ wurden entfernt. Es ist in diesem Fall darauf zu achten, dass die Gene für alle tRNAs, die TCG und TCA decodieren können, entfernt werden. Darüber hinaus müssen aber für die verbleibenden synonymen Codons noch tRNAs zur Verfügung stehen. Es ist daher nicht möglich, beliebige synonyme Codons zur Reprogrammierung auszuwählen.

Aus Abbildung 7.5 (Cui et al., 2020) ist ersichtlich, welche der synonymen Codons beispielsweise für Serin zum Umprogrammieren in Frage kommen können. Es sind in diesem Schema sowohl die Codons als auch die zugehörigen tRNAs eingetragen. Daraus ergibt sich, dass die Serin-Codons TCA und/oder TCG zum Umprogrammieren geeignet sind. Der Stern (*) in der Tabelle zeigt die Änderung des Nucleotids der tRNA an Position 34 an und entspricht somit der Wobble-Position.



The genetic code and its decoding tRNA isoacceptors of *E. coli*

Abb. 7.5 Darstellung aller synonymen Codons sowie der tRNA-Isoakzeptoren von E. coli (Cui et al., 2020, adaptiert). Sterne (*) zeigen die Änderung des jeweiligen Nucleotids der tRNA an Position 34 an und entpsrechen somit der Wobble-Position.

Nachdem die beiden Serin-Codons TCG und TCA sowie das Stopp-Codon TAG frei sind, können diese neuen ncAAs zugeordnet werden, indem durch Mutation und Selektion neue orthogonale Synthetasen (o-aa-tRNA-Synthetasen) geschaffen werden, welche zu diesen ncAA passen, und es werden auch neue o-tRNAs (orthogonale tRNAs) geschaffen, welche ausschließleich an die o-aa-tRNA-Synthetase binden können.

Es wurde also ein orthogonales System aus ncAA, tRNA und aa-tRNAA-Synthetase erzeugt, wobei die o-aa-tRNAA-Synthetase (orthogonale aa-tRNA-Synthetase) nicht an die Serin-tRNAs für die verbleibenden synonymen Codons binden kann und umgekehrt die ursprünglichen aa-tRNA-Synthetasen weder an die neuen ncAAs noch an die o-tRNA binden können.

Werden übrigens in einem E.-coli-Genom das Gen für einen Releasefaktor für ein Stopp-Codon sowie die Gene für die $tRNA^{Ser}_{CGA}$ und für die $tRNA^{Ser}_{UGA}$ entfernt, so kann ein Ribosom diese Codons nicht mehr erkennen. Das Ribosom bleibt beim ersten Codon, welches es nicht mehr erkennen kann, stehen, sodass eine Infektion von diesem veränderten E.-coli-Stamm beispielsweise mit einem T6-Phagen (ein Enterobacteria Phage, der E. coli infiziert) oder mit anderen Phagen keinerlei Auswirkung mehr hat. Der mutierte E.-coli-Stamm ist gegen Phagen-Infektionen resistent.

Für die biotechnologische Synthese von Proteinen mit völlig neuen Eigenschaften wird es von Vorteil sein, viele unterschiedliche ncAA einbauen zu können. Ein großes Problem besteht darin, dass nur sehr wenige Codons befreit werden können, sodass die Möglichkeiten, mehrere unterschiedliche ncAA gleichzeitig einbauen zu können, begrenzt sind. Eine Forschungsrichtung forscht daran, den genetischen Code dahingehend zu erweitern, dass Codons mit vier Basen Länge (Quadruplett-Codons) verwendet werden, welche aber ausschließlich die vier bekannten Basen (A, T, G, C) verwenden. Dadurch ergeben sich $4^4 = 256$ unterschiedliche Codons, die zur Verfügung stehen.

Hierzu müssen wieder orthogonale Syteme aus o-tRNAs mit dazugehörigen ncAAs, o-aa-tRNAA-Synthetasen, geschaffen werden. Weiters müssen die Ribosomen derart angepasst werden, dass diese einen Quadruplett-Code lesen können (Chen & Schindlinger, 2010; Neumann et al., 2010; Guo & Niu, 2022). Orthogonale Ribosomen müssen eine veränderte Ribosom-Bindungsstelle (RBS, ribosome binding site) erkennen können, also jenen Bereich der mRNA, an den das Ribosom bei der Translationsinitiation bindet (DocCheck Flexikon, ribosomale Bindungsstelle) und umgekehrt dürfen nur mRNAs mit der mutierten RBS von den orthogonalen Ribosomen translatiert werden. In diesem Fall muss aber – wegen des neuen fixen Leserasters von vier Basen Länge – die gesamte DNA neu synthetisiert werden.

Weitere Möglichkeiten der synthetischen Biologie bestehen darin, zusätzliche Basenpaare einzuführen wie beispielsweise beim Hachimoji-Code (siehe Kapitel 2). Dieser hat insgesamt acht Nucleobasen, die vier bekannten Basen (A, T, C, G) sowie vier zusätzliche Basen (Z, P, S, B). P und B sind Purin-Analoge, Z und S Pyrimidin-Analoge, sodass P mit Z und B mit S Watson-Crick-ähnliche Basenpaare bilden können. Die Codierung mittels Tripletts wird aber beibehalten (Hoshika et al., 2019). Dadurch können $8^3 = 512$ unterschiedliche Codons (Tripletts) gebildet werden (zusätzliche Basen siehe Anhang, Hoshika et al., 2019).

Ein ähnlicher Ansatz wird von Zhang et al. (2017) und Dien et al. (2018) verfolgt, die ein zusätzliches nicht-kanonisches Basenpaar verwenden (zusammen mit den vier kanonischen Basen ergeben sich $6^3 = 216$ mögliche Triplett-Codons). Sie konnten zeigen, dass das zusätzliche Basenpaar in lebenden Zellen bei der Replikation der DNA, der Transkription und der Translation am Ribosom toleriert wird.

8 Ausblick

Durch den großen Erfolg der mRNA-Impfstoffe während der Covid-19-Pandemie hat die Forschung einen massiven Aufschwung erlebt. Viele Firmen forschen daran, diese Technik an andere Virusinfektionen wie beispielsweise an die Influenza zu adaptieren. Florian Krammer, namhafter Vakzinologe an der Universität Wien, entwickelt mit seinem Team einen universellen Influenzaimpfstoff. Genau wie der Virologe Christian Drosten von der Berliner Charité sieht auch er in der Vogelgrippe das Potential für eine neue Pandemie (MedUni Wien online, derStandard online).

Die übergroße Menge an Kunststoffabfällen in der Umwelt hat Forscher dazu veranlasst, nach effizienten Abbaumöglichkeiten zu suchen. Es wurden bereits mehrere natürlich vorkommende Enzyme isoliert, die in geringem Maße PET (Polyethylene-Terephthalat) angreifen können. Durch gezielte Mutationen wird die Effizienz und die Eigenschaften der Enzyme, wie beispielsweise bessere Temperaturstabilität, verbessert (Richter et al., 2023). Die weitere Forschung erstreckt sich auf unterschiedliche Kunstoffsorten.

Das Konzept der Click-Chemie erlaubt es, Synthesen modular und mit hohen Ausbeuten durchzuführen, ohne eine aufwändige Totalsynthese durchführen zu müssen. So können beispielsweise Additionen an C–C-Mehrfachbindungen durchgeführt werden oder es kann durch eine Cycloaddition ein Fünferring mit drei N-Atomen (Triazolring) hergestellt werden (Kunz, 2009; Ochoa & Milam, 2020). Es ist bereits ein erstes Tumormedikament, das mittels Click-Chemie hergestellt wurde, in einer ersten vorklinischen Studie (Srinivasan et al., 2023).

Die Kombination des Umprogrammierens von synonymen Codons oder Stopp-Codons um gezielt ncAA in Polypeptide einzubauen, welche funktionelle Gruppen aus der Click-Chemie tragen, eröffnet weitere Möglichkeiten (Saleh et al., 2019).

Um Pflanzen resilienter gegen Stress zu machen, werden zunehmend biotechnologische Methoden eingesetzt, wie RNA-Interferenz oder epigenetische Modifikatonen (Giudice et al., 2021).

Durch bessere Codon-Harmonisierung (siehe Kapitel 7), wird es möglich sein, humane Enzyme/Hormone mittels heterologer Genexpression herzustellen, die weniger Nebenwirkungen haben als bisher.

Die Erweiterung des genetischen Codes durch Einführen neuer komplementärer Basenpaare erlaubt die Erhöhung der Codierungsmöglichkeiten (siehe Kapitel 7). Durch Synthetisierung neuer orthogonaler Systeme können neue Arzneimittel entwickelt werden, die andere als die 20 codogenen Aminosäuren, in Proteine einbauen können. Die Firma Sanofi mit Sitz in Paris und Frankfurt am Main, arbeitet bereits an solch neuen Medikamenten, die mittels heterologer Genexpression in E. coli hergestellt werden sollen (Sanofi online).

Literaturverzeichnis

Agris, P. F., Eruysal, E., Narendran, A., Väre, V. Y. P., Vangaveti, S. & Ranganathan, S. (2017). Celebrating wobble decoding: Half a century and still much is new. RNA Biology, 15(4–5), 537–553. https://doi.org/10.1080/15476286.2017.1356562

Alberts, B., Bray, D., Hopkin, K., Johnson, A. D., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2021). Lehrbuch der Molekularen Zellbiologie. Fünfte Auflage. Weinheim: Wiley-VCH. (Kindle Ausgabe)

Alberts, B., Heald R., Johnson, A. D., Morgan, D., Raff, M., Roberts, K. & Walter, P. (2022). Molecular biology of the cell. In W.W. Norton & Company, ISBN 978039884821

Akeju, O. J. & Cope, A. L. (2022). Re-examining supposed correlations between synonymous codon usage and protein bond angles inE. coli. bioRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2022.10.26.513858

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U. & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nature Genetics, 23(2), 185–188. https://doi.org/10.1038/13810

Anfinsen, C. B. & National Institutes of Health. (1972). STUDIES ON THE PRINCIPLES THAT GOVERN THE FOLDING OF PROTEIN CHAINS. In Nobel Lecture. https://www.nobelprize.org/uploads/2018/06/anfinsen-lecture.pdf

Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. Science, 181(4096), 223–230. https://doi.org/10.1126/science.181.4096.223

Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. (1961). THE KINETICS OF FORMATION OF NATIVE RIBONUCLEASE DURING OXIDATION OF THE REDUCED POLYPEPTIDE CHAIN. Proceedings of the National Academy of Sciences of the United States of America, 47(9), 1309–1314. https://doi.org/10.1073/pnas.47.9.1309

Angov, E., Hillier, C. J., Kincaid, R. L. & Lyon, J. A. (2008). Heterologous Protein Expression Is Enhanced by Harmonizing the Codon Usage Frequencies of the Target Gene with those of the Expression Host. PLOS ONE, 3(5), e2189. https://doi.org/10.1371/journal.pone.0002189

Angulski, A. B. B., Hosny, N., Cohen, H., Martin, A. A., Hahn, D., Bauer, J. & Metzger, J. M. (2023). Duchenne muscular dystrophy: disease mechanism and therapeutic strategies. Frontiers in Physiology, 14. https://doi.org/10.3389/fphys.2023.1183101

Arthur, L., Pavlovic-Djuranovic, S., Koutmou, K. S., Green, R., Szczęsny, P. & Djuranović, S. (2015). Translational control by lysine-encoding A-rich sequences. Science Advances, 1(6). https://doi.org/10.1126/sciadv.1500154

Baradaran-Heravi, A., Balgi, A. D., Hosseini-Farahabadi, S., Choi, K., Has, C. & Roberge, M. (2021). Effect of small molecule eRF3 degraders on premature termination codon readthrough. Nucleic Acids Research, 49(7), 3692–3708. https://doi.org/10.1093/nar/gkab194

Bartoszewski, R. A., Jablonsky, M., Bartoszewska, S., Stevenson, L., Dai, Q., Kappes, J., Collawn, J. F. & Bebok, Z. (2010). A Synonymous Single Nucleotide Polymorphism in ΔF508 CFTR Alters the Secondary Structure of the mRNA and the Expression of the Mutant Protein. Journal Of Biological Chemistry/ The Journal Of Biological Chemistry, 285(37), 28741–28748. https://doi.org/10.1074/jbc.m110.154575

Belinky, F., Sela, I., Rogozin, I. B. & Koonin, E. V. (2019). Crossing fitness valleys via double substitutions within codons. BMC Biology, 17(1). https://doi.org/10.1186/s12915-019-0727-4

Berg, J. M., Tymoczko, J. L., Gatto, G. J. & Stryer, L. (2018). Stryer Biochemie. (8. Auflage 2018.) In Springer eBooks. https://doi.org/10.1007/978-3-662-54620-8

Bierhoff, H. (2024). Genetik, Epigenetik und Umweltfaktoren der Lebenserwartung – Welche Rolle spielt Nature-versus-Nurture beim Altern? Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz, 67(5), 521–527. https://doi.org/10.1007/s00103-024-03873-x

Blakney, A. K. (2023). A tool for optimizing messenger RNA sequence. Nature, 621(7978), 262–264. https://doi.org/10.1038/d41586-023-01745-z

Branciamore, S., Chen, Z. X., Riggs, A. D. & Rodin, S. N. (2010). CpG island clusters and pro-epigenetic selection for CpGs in protein-coding exons of HOX and other transcription factors. Proceedings of the National Academy of Sciences of the United States of America, 107(35), 15485–15490. https://doi.org/10.1073/pnas.1010506107

Bresch, C. & Hausmann, M. R. (1972). Klassische und molekulare Genetik. In Springer eBooks. https://doi.org/10.1007/978-3-642-87168-9

Brest, P., Lapaquette, P., Souidi, M., Lebrigand, K., Cesaro, A., Vouret-Craviari, V., Mari, B., Barbry, P., Mosnier, J., Hébuterne, X., Harel-Bellan, A., Mograbi, B., Darfeuille-Michaud, A. & Hofman, P. (2011). A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nature Genetics, 43(3), 242–245. https://doi.org/10.1038/ng.762

Borys, T. (2011). Codierung und Kryptologie. In Vieweg+Teubner eBooks. https://doi.org/10.1007/978-3-8348-8252-3

Buttlar, J., Klein, C., Bruch, A., Fachinger, A., Funk, J., Hawer, H. & Kuijpers, A. (2020). Tutorium Genetik (1. Auflage). In Springer eBooks. https://doi.org/10.1007/978-3-662-56067-9

Callaway, E. (2021). DeepMind's AI for protein structure is coming to the masses. Nature. https://doi.org/10.1038/d41586-021-01968-y

Chaney, J. L., Steele, A., Carmichael, R., Rodríguez, A., Specht, A. T., Ngo, K. Y., Li, J., Emrich, S. J. & Clark, P. L. (2017). Widespread position-specific conservation of synonymous rare codons within coding sequences. PLOS Computational Biology, 13(5), e1005531. https://doi.org/10.1371/journal.pcbi.1005531

Chen, I. A. & Schindlinger, M. (2010). Quadruplet codons: One small step for a ribosome, one giant leap for proteins. BioEssays, 32(8), 650–654. https://doi.org/10.1002/bies.201000051

Christen, P., Jaussi, R. & Benoit, R. (2016). Biochemie und Molekularbiologie. In Springer eBooks. https://doi.org/10.1007/978-3-662-46430-4

Collin, R. W., De Heer, A. R., Oostrik, J., Pauw, R., Plantinga, R., Huygen, P., Admiraal, R., De Brouwer, A. P., Strom, T. M., Cremers, C. & Kremer, H. (2008). Mid-frequency DFNA8/12 hearing loss caused by a synonymous TECTA mutation that affects an exonic splice enhancer. European Journal of Human Genetics, 16(12), 1430–1436. https://doi.org/10.1038/ejhg.2008.110

Crick, F., Barnett, L., Brenner, S. & Watts-Tobin, R. J. (1961). General Nature of the Genetic Code for Proteins. Nature, 192(4809), 1227–1232. https://doi.org/10.1038/1921227a0

Crick, F. (1966). Codon—anticodon pairing: The wobble hypothesis. Journal of Molecular Biology, 19(2), 548–555. https://doi.org/10.1016/s0022-2836(66)80022-0

Crick, F., (1958) "On protein synthesis". Wellcome Collection. https://wellcomecollection.org/works/z3d5fnyg

Crick, F. H. C., Griffith, J. S. & Orgel, L. E. (1957). CODES WITHOUT COMMAS. Proceedings Of The National Academy Of Sciences, 43(5), 416–421. https://doi.org/10.1073/pnas.43.5.416

Cui, Z., Johnston, W. A. & Alexandrov, K. (2020). Cell-Free Approach for Non-canonical Amino Acids Incorporation Into Polypeptides. Frontiers in Bioengineering And Biotechnology, 8. https://doi.org/10.3389/fbioe.2020.01031

Dabrowski, M., Bukowy-Bieryllo, Z. & Zietkiewicz, E. (2015). Translational readthrough potential of natural termination codons in eucaryotes – The impact of RNA sequence. RNA Biology, 12(9), 950–958. https://doi.org/10.1080/15476286.2015.1068497

Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. Developmental Biology, 278(2), 274–288. https://doi.org/10.1016/j.ydbio.2004.11.028

Daidone, V., Gallinaro, L., Cattini, M. G., Pontara, E., Bertomoro, A., Pagnan, A. & Casonato, A. (2011). An apparently silent nucleotide substitution (c.7056C>T) in the von Willebrand factor gene is responsible for type 1 von Willebrand disease. Haematologica, 96(6), 881–887. https://doi.org/10.3324/haematol.2010.036848

DeBenedictis, E. A., Carver, G. D., Chung, C. Z., Söll, D. & Badran, A. H. (2021). Multiplex suppression of four quadruplet codons via tRNA directed evolution. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-25948-y

Diaz, J. V. R., Jin, Y., Garber, K. B., Cossen, K., Li, Y., Jin, P., Li, H. & Ham, J. N. (2022). A homozygous exonic variant leading to exon skipping in ABCC8 as the cause of severe congenital hyperinsulinism. American Journal of Medical Genetics. Part A, 188(8), 2429–2433. https://doi.org/10.1002/ajmg.a.62843

Dien, V. T., Holcomb, M., Feldman, A. W., Fischer, E. C., Dwyer, T. J. & Romesberg, F. E. (2018). Progress Toward a Semi-Synthetic Organism with an Unrestricted Expanded Genetic Alphabet. Journal Of The American Chemical Society, 140(47), 16115–16123. https://doi.org/10.1021/jacs.8b08416

Doig, A. J. (1997). Improving the Efficiency of the Genetic Code by Varying the Codon Length—The Perfect Genetic Code. Journal Of Theoretical Biology, 188(3), 355–360. https://doi.org/10.1006/jtbi.1997.0489

dos Reis, M. D. (2004). Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Research, 32(17), 5036–5044. https://doi.org/10.1093/nar/gkh834

Duan, J., Wainwright, M. S., Comeron, J. M., Saitou, N., Sanders, A. R., Gelernter, J. & Gejman, P. V. (2003). Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. Human Molecular Genetics Online/Human Molecular Genetics, 12(3), 205–216. https://doi.org/10.1093/hmg/ddg055

Eberlein, L., Beierlein, F. R., Van Eikema Hommes, N. J. R., Radadiya, A., Heil, J., Benner, S. A., Clark, T., Kast, S. M. & Richards, N. G. J. (2020). Tautomeric equilibria of nucleobases in the Hachimoji expanded genetic Alphabet. Journal of Chemical Theory and Computation, 16(4), 2766–2777. https://doi.org/10.1021/acs.jctc.9b01079

Entringer, S., Buß, C. & Heim, C. (2016). Frühe Stresserfahrungen und Krankheitsvulnerabilität. Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz, 59(10), 1255–1261. https://doi.org/10.1007/s00103-016-2436-2

Fedyukina, D. V. & Cavagnero, S. (2011). Protein folding at the exit tunnel. Annual Review of Biophysics, 40(1), 337–359. https://doi.org/10.1146/annurev-biophys-042910-155338

Feingold, E.A. et al. (2004). The ENCODE (ENCyclopedia of DNA Elements) project. (2004). Science (New York, N.Y.), 306(5696), 636–640. https://doi.org/10.1126/science.1105136

Francis, B. R. (2013). Evolution of the Genetic Code by Incorporation of Amino Acids that Improved or Changed Protein Function. Journal Of Molecular Evolution, 77(4), 134–158. https://doi.org/10.1007/s00239-013-9567-y

Franklin, R. & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. Nature, 171(4356), 740–741. https://doi.org/10.1038/171740a0

Fredens, J., Wang, K., De La Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., Chia, T., Schmied, W. H., Dunkelmann, D. L., Beránek, V., Uttamapinant, C., Llamazares, A. G., Elliott, T. S. & Chin, J. W. (2019). Total synthesis of Escherichia coli with a recoded genome. Nature, 569(7757), 514–518. https://doi.org/10.1038/s41586-019-1192-5

Fujii, N. & Saito, T. (2004) 'Homochirality and life', Chemical record, 4(5), pp. 267–278. doi:10.1002/tcr.20020.

Gardiner-Garden, M. & Frommer, M. (1987). CpG Islands in vertebrate genomes. Journal of Molecular Biology, 196(2), 261–282. https://doi.org/10.1016/0022-2836(87)90689-9

Geiger H. (2009). Astrobiologie. vdf Hochschulverlag AG an der ETH Zürich. https://doi.org/10.3218/3699-2

Gilbert, W. (1978). Why genes in pieces? Nature, 271(5645), 501. https://doi.org/10.1038/271501a0

Giudice, G., Moffa, L., Varotto, S., Cardone, M. F., Bergamini, C., De Lorenzis, G., Velasco, R., Nerva, L. & Chitarra, W. (2021). Novel and emerging biotechnological crop protection approaches. Plant Biotechnology Journal. https://doi.org/10.1111/pbi.13605

González-Delgado, J., Mier, P., Bernadó, P., Neuvial, P. & Cortés, J. (2022). Statistical tests to detect differences between codon-specific Ramachandran plots. bioRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2022.11.29.518303

Graw, J. (2020). Genetik. In Springer eBooks. https://doi.org/10.1007/978-3-662-60909-5

Gressner, A. & Arndt, T. (2013). Lexikon der Medizinischen Laboratoriumsdiagnostik. In Springer eBooks. https://doi.org/10.1007/978-3-642-12921-6

Guo, J. & Niu, W. (2022). Genetic code expansion through quadruplet Codon decoding. Journal Of Molecular Biology/Journal Of Molecular Biology, 434(8), 167346. https://doi.org/10.1016/j.jmb.2021.167346

Hahn, L., Heitmüller, S., Lammers, C. & Neumann, H. (2013). Erweiterung des Genetischen Codes: Eine Methode mit Potenzial. Georg-August Universität Göttingen, Institut für Mikrobiologie und Genetik, Göttingen, https://analyticalscience.wiley.com/content/article-do/erweiterung-des-genetischen-codes-eine-methode-mit-potenzial

Heinrich, P., Müller, M., Graeve, L., Löffler, G. & Petrides, P. E. (2014). Löffler/Petrides Biochemie und Pathobiochemie (9., vollständig überarbeitete Auflage). In Springer-Lehrbuch. https://doi.org/10.1007/978-3-642-17972-3

Hildt, E. (2022). Übersicht über die in der EU zugelassenen COVID-19-Impfstoffe – von der Technologie über die klinische Prüfung zur Zulassung. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz, 65(12), 1237–1243. https://doi.org/10.1007/s00103-022-03600-4

Hoshika, S., Leal, N. A., Kim, M., Kim, M., Karalkar, N. B., Kim, H., Bates, A. M., Watkins, N. E., SantaLucia, H. A., Meyer, A. J., DasGupta, S., Piccirilli, J. A., Ellington, A. D., SantaLucia, J., Georgiadis, M. M. & Benner, S. A. (2019). Hachimoji DNA and RNA: A genetic system with eight building blocks. Science, 363(6429), 884–887. https://doi.org/10.1126/science.aat0971

Hütt, M.-T. & Dehnert, M. (2016). Methoden der Bioinformatik: Eine Einführung zur Anwendung in Biologie und Medizin In Springer eBooks. https://doi.org/10.1007/978-3-662-46150-1

Inouye, M., Takino, R., Ishida, Y., Inouye, K. (2020). Evolution of the genetic code; Evidence from serine codon use disparity in Escherichia coli. Proceedings of the National Academy of Sciences – PNAS, 117(46), 28572–28575. https://doi.org/10.1073/pnas.2014567117

Jones, K. M., Ankeny, R. A. & Cook-Deegan, R. (2018). The Bermuda Triangle: The pragmatics, policies, and principles for Data sharing in the History of the Human Genome Project. Journal of the History of Biology, 51(4), 693–805. https://doi.org/10.1007/s10739-018-9538-7

Kimchi-Sarfaty, C., Oh, J. M., Kim, I., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V. & Gottesman, M. M. (2007). A "Silent" Polymorphism in the MDR 1 Gene Changes Substrate Specificity. Science, 315(5811), 525–528. https://doi.org/10.1126/science.1135308

Kirchner, S., Cai, Z., Rauscher, R., Kastelic, N., Anding, M., Czech, A., Kleizen, B., Ostedgaard, L. S., Braakman, I., Sheppard, D. N. & Ignatova, Z. (2017). Alteration of protein function by a silent polymorphism linked to tRNA abundance. PLoS Biology, 15(5), e2000779. https://doi.org/10.1371/journal.pbio.2000779

Komar, A. A. (2021). A Code Within a Code: How Codons Fine-Tune Protein Folding in the Cell. Biochemistry, 86(8), 976–991. https://doi.org/10.1134/s0006297921080083

Koutmou, K. S., Schuller, A. P., Brunelle, J. L., Radhakrishnan, A., Djuranovic, S. & Green, R. (2015). Ribosomes slide on lysine-encoding homopolymeric A stretches. eLife, 4. https://doi.org/10.7554/elife.05534

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. Gene, 234(2), 187–208. https://doi.org/10.1016/s0378-1119(99)00210-3

Kraus, M. H. (2021). Eins, zwei, viele. In Springer eBooks. https://doi.org/10.1007/978-3-662-63154-6

Kunz, D. (2009). Klick-Chemie. Synthesen, die gelingen. Chemie in Unserer Zeit, 43(4), 224–230. https://doi.org/10.1002/ciuz.200900475

Kurreck, J., Engels, J. W. & Lottspeich, F. (2022). Bioanalytik (4th ed. 2022). In Springer eBooks. https://doi.org/10.1007/978-3-662-61707-6

Kuzmiak, H. & Maquat, L. E. (2006). Applying nonsense-mediated mRNA decay research to the clinic: progress and challenges. Trends in Molecular Medicine, 12(7), 306–316. https://doi.org/10.1016/j.molmed.2006.05.005

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., Frazier, M. (2001). Initial sequencing and analysis of the human genome. Nature, 409(6822), 860–921. https://doi.org/10.1038/35057062

Liu, Y. (2020). A code within the genetic code: codon usage regulates co-translational protein folding. Cell Communication and Signaling, 18(1), 145–145. https://doi.org/10.1186/s12964-020-00642-6

Ma, X., Pang, Q., Zhang, Q., Jiang, Y., Wang, O., Li, M., Xing, X. & Xia, W. (2022). A Novel Synonymous Variant of PHEX in a Patient with X-Linked Hypophosphatemia. Calcified Tissue International, 111(6), 634–640. https://doi.org/10.1007/s00223-022-01003-w

Ma, S. L., Vega-Warner, V., Gillies, C., Sampson, M. G., Kher, V., Sethi, S. K. & Otto, E. A. (2015). Whole Exome Sequencing Reveals Novel PHEX Splice Site Mutations in Patients with Hypophosphatemic Rickets. PloS One, 10(6), e0130729. https://doi.org/10.1371/journal.pone.0130729

Maier, Y. (2020), Von der Lehmkugel zum Tintenfass, ARD alpha (online https://www.ardalpha.de/wissen/geschichte/kulturgeschichte/keilschrift-schrift-ursprung-mesopotamien-100.html

Mauro, V. P. & Chappell, S. A. (2014). A critical analysis of codon optimization in human therapeutics. Trends in Molecular Medicine, 20(11), 604–613. https://doi.org/10.1016/j.molmed.2014.09.003

Mauro, V. P. (2018). Codon Optimization in the Production of Recombinant Biotherapeutics: Potential Risks and Considerations. BioDrugs: Clinical Immunotherapeutics, Biopharmaceuticals, and Gene Therapy, 32(1), 69–81. https://dx.doi.org/10.1016/j.molmed.2014.09.003

Miescher, F. (1871). Ueber die chemische Zusammensetzung der Eiterzellen. Publiziert in Hoppe-Seyler's medizinisch-chemischen Untersuchungen, Heft 4, 1871, S. 441–460. https://static-content.springer.com/esm/art%3A10.1007%2Fs00439-007-0433-0/MediaObjects/439_2007_433_MOESM2_ESM.pdf

Morais, P., Adachi, H. & Yu, Y. (2021). The critical contribution of pseudouridine to mRNA COVID-19 vaccines. Frontiers in Cell and Developmental Biology, 9. https://doi.org/10.3389/fcell.2021.789427

Müller-Esterl, W. (2018). Biochemie. In Springer eBooks. https://doi.org/10.1007/978-3-662-54851-6

Mulroney, T. E., Pöyry, T., Yam-Puc, J. C., Rust, M., Harvey, R. F., Kalmar, L., Horner, E., Booth, L., Ferreira, A. P., Stoneley, M., Sawarkar, R., Mentzer, A. J., Lilley, K. S., Smales, C. M., Von der Haar, T., Turtle, L., Dunachie, S., Klenerman, P., Thaventhiran, J. E. D. & Willis, A. E. (2023). N1-methylpseudouridylation of mRNA causes +1 ribosomal frameshifting. Nature, 625(7993), 189–194. https://doi.org/10.1038/s41586-023-06800-3

Nackley, A. G., Shabalina, S. A., Tchivileva, I. E., Satterfield, K., Korchynskyi, O., Makarov, S. S., Maixner, W. & Diatchenko, L. (2006). Human Catechol-O-Methyltransferase Haplotypes Modulate Protein Expression by Altering mRNA Secondary Structure. Science, 314(5807), 1930–1933. https://doi.org/10.1126/science.1131262

Netzer, W. J. & Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature, 388(6640), 343–349. https://doi.org/10.1038/41024

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. (2010). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature, 464(7287), 441–444. https://doi.org/10.1038/nature08817

Nguyen, H. D., Yoshihama, M. & Kenmochi, N. (2006). Phase distribution of spliceosomal introns: implications for intron origin. BMC Evolutionary Biology, 6(1). https://doi.org/10.1186/1471-2148-6-69

Nickelsen, K. (2017). Die Entdeckung der Doppelhelix. In Klassische Texte der Wissenschaft. https://doi.org/10.1007/978-3-662-47150-0

Nirenberg, M. & Leder, P. (1964). RNA Codewords and Protein Synthesis. Science, 145(3639), 1399–1407. https://doi.org/10.1126/science.145.3639.1399

Nirenberg, M. W. & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proceedings Of The National Academy Of Sciences Of The United States Of America, 47(10), 1588–1602. https://doi.org/10.1073/pnas.47.10.1588

Nordheim, A., Knippers, R., Dröge, P., Meister, G., Schiebel, E., Vingron, M. & Walter, J. (2018). Molekulare Genetik (11., unveränderte Auflage.). Georg Thieme Verlag. https://doi.org/10.1055/b-006-149922

Ochoa, S. & Milam, V. T. (2020). Modified Nucleic Acids: Expanding the Capabilities of Functional Oligonucleotides. Molecules, 25(20), 4659. https://doi.org/10.3390/molecules25204659

Olejniczak, M. & Uhlenbeck, O. C. (2006). tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. Biochimie, 88(8), 943–950. https://doi.org/10.1016/j.biochi.2006.06.005

Orešič, M. & Shalloway, D. (1998). Specific correlations between relative synonymous codon usage and protein secondary structure. Journal of Molecular Biology, 281(1), 31–48. https://doi.org/10.1006/jmbi.1998.1921

Park, J., Kwon, M., Yamaguchi, Y., Firestein, B. L., Park, J., Yun, J., Yang, J. & Inouye, M. (2016). Preferential use of minor codons in the translation initiation region of human genes. Human Genetics, 136(1), 67–74. https://doi.org/10.1007/s00439-016-1735-x

Passmore, L. A. & Coller, J. (2022). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. Nature Reviews. Molecular Cell Biology, 23(2), 93–106. https://doi.org/10.1038/s41580-021-00417-y

Pauling, L., Corey, R. B. & Branson, H. R. (1951). The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. Proceedings Of The National Academy Of Sciences Of The United States Of America, 37(4), 205–211. https://doi.org/10.1073/pnas.37.4.205

Rauscher, R., Bampi, G. B., Guevara-Ferrer, M., Santos, L. A., Joshi, D., Mark, D., Strug, L. J., Rommens, J. M., Ballmann, M., Sorscher, E. J., Oliver, K. E. & Ignatova, Z. (2021). Positive epistasis between disease-causing missense mutations and silent polymorphism with effect on mRNA translation velocity. Proceedings Of The National Academy Of Sciences Of The United States Of America, 118(4). https://doi.org/10.1073/pnas.2010612118

Reinard, T. (2021). Molekularbiologische Methoden 2.0. Uni-Taschenbücher GmbH Verlag. 3. aktualisierte Auflage. https://doi.org/10.36198/9783838587950

Richter, P. K., Blázquez-Sánchez, P., Zhao, Z., Engelberger, F., Wiebeler, C., Künze, G., Frank, R., Krinke, D., Frezzotti, E., Lihanova, Y., Falkenstein, P., Matysik, J., Zimmermann, W., Sträter, N. & Sonnendecker, C. (2023). Structure and function of the metagenomic plastic-degrading polyester hydrolase PHL7 bound to its product. Nature Communications, 14(1). https://doi.org/10.1038/s41467-023-37415-x

Richard, I. & Beckmann, J. S. (1995). How neutral are synonymous codon mutations? Nature Genetics, 10(3), 259. https://doi.org/10.1038/ng0795-259

Ritz-Timme, S., Rochholz, G., Stammert, R. & Ritz, H. (2002). Biochemische Altersschätzung Zur Frage genetischer und soziokultureller (ethnischer) Einflüsse auf die Razemisierung von Asparaginsäure in Dentin. Rechtsmedizin, 12(4), 203–206. https://doi.org/10.1007/s00194-002-0152-8

Robertson, W. E., Funke, L. F. H., De La Torre, D., Fredens, J., Elliott, T. S., Spinck, M., Christova, Y., Cervettini, D., Böge, F. L., Liu, K. C., Buse, S., Maslen, S., Salmond, G. P. C. & Chin, J. W. (2021). Sense codon reassignment enables viral resistance and encoded polymer synthesis. Science, 372(6546), 1057–1062. https://doi.org/10.1126/science.abg3029

Roca, X., Sachidanandam, R. & Krainer, A. R. (2003). Intrinsic differences between authentic and cryptic 5' splice sites. Nucleic Acids Research, 31(21), 6321–6333. https://doi.org/10.1093/nar/gkg830

Rodriguez, A., Wright, G., Emrich, S. & Clark, P. L. (2017). %MinMax: A versatile tool for calculating and comparing synonymous codon usage and its impact on protein folding. Protein Science, 27(1), 356–362. https://doi.org/10.1002/pro.3336

Rosenberg, A. A., Marx, A. & Bronstein, A. M. (2022). Codon-specific Ramachandran plots show amino acid backbone conformation depends on identity of the translated codon. Nature Communications, 13(1), 2815–2815. https://doi.org/10.1038/s41467-022-30390-9

Gomez, M. A. R. & Ibba, M. (2020). Aminoacyl-tRNA synthetases. RNA, 26(8), 910–936. https://doi.org/10.1261/rna.071720.119

Sabi, R. & Tuller, T. (2014). Modelling the Efficiency of Codon–tRNA Interactions Based on Codon Usage Bias. DNA Research, 21(5), 511–526. https://doi.org/10.1093/dnares/dsu017

Sahin, U., Karikó, K. & Türeci, Ö. (2014). mRNA-based therapeutics — developing a new class of drugs. Nature Reviews. Drug Discover/Nature Reviews. Drug Discovery, 13(10), 759–780. https://doi.org/10.1038/nrd4278

Saier, M. H. (2019). Understanding the Genetic Code. Journal Of Bacteriology, 201(15). https://doi.org/10.1128/jb.00091-19

Saleh, A. M., Wilding, K. M., Calve, S., Bundy, B. C. & Kinzer-Ursem, T. L. (2019). Non-canonical amino acid labeling in proteomics and biotechnology. Journal Of Biological Engineering, 13(1). https://doi.org/10.1186/s13036-019-0166-3

Salzman, D. W., & Weidhaas, J. B. (2011). miRNAs in the spotlight: Making "silent" mutations speak up. Nature Medicine, 17(8), 934–935. https://doi.org/10.1038/nm0811-934

Sauna, Z. E. & Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. Nature Reviews. Genetics, 12(10), 683–691. https://www.nature.com/articles/nrg3051

Schaaf, C. P. & Zschocke, J. (2018). Basiswissen Humangenetik. (3. überarbeitete und erweiterte Auflage). In Springer-Lehrbuch. https://doi.org/10.1007/978-3-662-56147-8

Schlautmann, L. P. & Gehring, N. H. (2020). A Day in the Life of the Exon Junction Complex. Biomolecules, 10(6), 866. https://doi.org/10.3390/biom10060866

Shabalina, S. A., Spiridonov, N. A. & Kashina, A. (2013). Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity.

Shannon, C. E. & Weaver, W. (1976). Mathematische Grundlagen der Informationstheorie. München Wien: Oldenbourg.

Shibata, Ohoka, N., Sugaki, Y., Onodera, C., Inoue, M., Sakuraba, Y., Takakura, D., Hashii, N., Kawasaki, N., Gondo, Y. & Naito, M. (2015). Degradation of Stop Codon Read-through Mutant Proteins via the Ubiquitin-Proteasome System Causes Hereditary Disorders. The Journal of Biological Chemistry, 290(47), 28428–28437. https://doi.org/10.1074/jbc.M115.670901

Smith, T. J., Tardu, M., Khatri, H. R. & Koutmou, K. S. (2022). mRNA and tRNA modification states influence ribosome speed and frame maintenance during poly(lysine) peptide synthesis. The Journal of Biological Chemistry, 298(6), 102039–102039. https://doi.org/10.1016/j.jbc.2022.102039

Srinivasan, S., Yee, N. A., Zakharian, M., Alečković, M., Mahmoodi, A., Nguyen, T. & Oneto, J. M. M. (2023). SQ3370, the first clinical click chemistry-activated cancer therapeutic, shows safety in humans and translatability across species. bioRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2023.03.28.534654

Takai, D. & Jones, P. A. (2002). Comprehensive Analysis of CpG Islands in Human Chromosomes 21 and 22. Proceedings of the National Academy of Sciences – PNAS, 99(6), 3740–3745. https://doi.org/10.1073/pnas.052410099

Tsao, D., Shabalina, S. A., Gauthier, J., Dokholyan, N. V. & Diatchenko, L. (2011). Disruptive mRNA folding increases translational efficiency of catechol-O-methyltransferase variant. Nucleic Acids Research, 39(14), 6201–6212. https://doi.org/10.1093/nar/gkr165

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I. & Pilpel, Y. (2010). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. Cell, 141(2), 344–354. https://doi.org/10.1016/j.cell.2010.03.031

Tursi, N. J. & Weiner, D. B. (2023). Modified messenger-RNA components alter the encoded protein. Nature (London), 625(7993), 37–38. https://doi.org/10.1038/d41586-023-03787-9

Vetterlein, Annabel, et al. “Are Catechol-O-Methyltransferase Gene Polymorphisms Genetic Markers for Pain Sensitivity after All? – A Review and Meta-Analysis.” Neuroscience and Biobehavioral Reviews, vol. 148, 2023, pp. 105112–105112, https://doi.org/10.1016/j.neubiorev.2023.105112

Watson, J. D. & Crick, F. H. C. (1953 a). Genetical Implications of the Structure of Deoxyribonucleic Acid. Nature (London), 171(4361), 964–967. https://doi.org/10.1038/171964b0

Watson, J. D. & Crick, F. H. C. (1953 b). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature (London), 171(4356), 737–738. https://doi.org/10.1038/171737a0

Wegner, R.-D., Trimborn, M., Stumm, M. & Wieacker, P. F. (2016). Humangenetische Grundlagen für Gynäkologen: Fachgebundene genetische Beratung im Überblick. Berlin; Boston: De Gruyter.

Weule, K. (1915). Vom Kerbstock zum Alphabet: Urformen der Schrift. Sechste Auflage. Stuttgart: Franckh.

Williams, I., Richardson, J., Starkey, A. & Stansfield, I. (2004). Genome-wide prediction of stop codon readthrough during translation in the yeast Saccharomyces cerevisiae. Nucleic Acids Research, 32(22), 6605–6616. https://doi.org/10.1093/nar/gkh1004

Wollrab, A. (2009). Organische Chemie: Eine Einführung für Lehramts- und Nebenfachstudenten. Springer.

Wright, G., Rodriguez, A., Li, J., Milenkovic, T., Emrich, S. J. & Clark, P. L. (2021). CHARMING: Harmonizing synonymous codon usage to replicate a desired codon usage pattern. Protein Science, 31(1), 221–231. https://doi.org/10.1002/pro.4223

Yu, T., Yan, F., Xu, Y., Hunag, Y., Gong, H., Zhao, P., Sun, D., Zhang, Y., Zhang, F. & He, X. (2023). Identification of a novel TNNI3 synonymous variant causing intron retention in autosomal recessive dilated cardiomyopathy. Gene, 856, 147102. https://doi.org/10.1016/j.gene.2022.147102

Zhang, H., Zhang, L., Lin, A., Xu, C., Li, Z., Liu, K., Liu, B., Ma, X., Zhao, F., Jiang, H., Chen, C., Shen, H., Li, H., Mathews, D. H., Zhang, Y. & Huang, L. (2023). Algorithm for optimized mRNA design improves stability and immunogenicity. Nature, 621(7978), 396–403. https://doi.org/10.1038/s41586-023-06127-z

Zhang, Y., Ptacin, J. L., Fischer, E. C., Aerni, H. R., Caffaro, C. E., Jose, K. S., Feldman, A. W., Turner, C. R. & Romesberg, F. E. (2017). A semi-synthetic organism that stores and retrieves increased genetic information. Nature, 551(7682), 644–647. https://doi.org/10.1038/nature24659

Zhou, W., Cai, Z., Liu, J., Wang, D., Ju, H. & Xu, R. (2020). Circular RNA: metabolism, functions and interactions with proteins. Molecular Cancer, 19(1). https://doi.org/10.1186/s12943-020-01286-3

Ziegenbalg, J., Ziegenbalg, O. & Ziegenbalg, B. (2016). Algorithmen von Hammurapi bis Gödel. In Springer eBooks. https://doi.org/10.1007/978-3-658-12363-5

Abbildungsverzeichnis

Sofern nicht anders vermerkt, handelt es sich um eigene Darstellungen.

Onlinequellenverzeichnis

Adenin	https://commons.wikimedia.org/wiki/File:Adenin.svg. Ersteller NEUROtiker gemeinfrei
AlphaFold	https://de.wikipedia.org/wiki/AlphaFold
AlphaFold 3	https://golgi.sandbox.google.com/about https://blog.google/technology/ai/google-deepmind-isomorphic-alphafold-3-ai-model/#life-molecules
Aminoacyl-tRNA-Synthetase	https://flexikon.doccheck.com/de/Aminoacyl-tRNA-Synthetase
Aminoglykosid-Antibiotika	https://de.wikipedia.org/wiki/Aminoglykosid-Antibiotika
Aminosäuren	https://commons.wikimedia.org/wiki/File:Overview_proteinogenic_amino_acids-DE.svg Die natürlich vorkommenden 20 proteinogenen Standard-Aminosäuren, gruppiert nach physikalisch-chemischen Eigenschaften. Ersteller: sponk, gemeinfrei
Aminosäuren Nummerierung	https://chemiezauber.de/inhalt/basic-5-kl-10-kohlenstoffverbindungen-organische-chemie-2/aminosaeuren-und-eiweisse-proteine/aminosaeuren-und-peptide/550-nomenklatur-namensgebung
Aneuploidie	https://www.spektrum.de/lexikon/biologie/aneuploidie/3473)
Aserbaidschanische Sprache	https://de.wikipedia.org/wiki/Aserbaidschanische_Sprache
Atatürk, Mustafa Kemal	https://de.wikipedia.org/wiki/Mustafa_Kemal_Atatürk#Schriften
Basenpaare	https://de.wikipedia.org/wiki/Basenpaar
Berthub	https://berthub.eu/articles/posts/german-reverse-engineering-source-code-of-the-biontech-pfizer-vaccine/
Bird, Adrian	https://www.imp.ac.at/achievements/research-milestones/adrian-bird-dna-methylation
Chromosom	https://de.wikipedia.org/wiki/Chromosom
Codonsonne	https://de.wikipedia.org/wiki/Code-Sonne#/media/Datei:Aminoacids_table.svg von User:Mouagip, gemeinfrei
Coronavirus	https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html Robert Koch Institut, Steckbrief Neuartiges Coronavirus
Corti-Organ	https://flexikon.doccheck.com/de/Corti-Organ
Desaminierung von Cytosin	https://de.wikipedia.org/wiki/Cytosin#/media/Datei:DesaminierungCtoU.png
Desaminierung von Methyl-Cytosin	https://de.wikipedia.org/wiki/5-Methylcytosin
Diabetologie, Geschichte der	https://de.wikipedia.org/wiki/Geschichte_der_Diabetologie

DNA Sequencing Cost https://www.genome.gov/sequencingcostsdata
Kosten der DNA Sequenzierung laut NIH
Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)

Drosten, Christian https://www.derstandard.at/story/3000000226442/vogelgrippevirus-in-usa-k246n-nte-n228chste-pandemie-ausl246sen

EMA https://www.ema.europa.eu/en/homepage
European Medicines Agency (seit März 2019 in Amsterdam/Niederlande)

Eurofins Scientific https://eurofinsgenomics.eu/media/892634/an_geneius_compared-to-competitors.pdf
Eurofins' adaption and optimisation software "GENEius" in comparison to other optimisation algorithms

Exom https://flexikon.doccheck.com/de/Exon

Fischer-Projektion https://de.wikipedia.org/wiki/Fischer-Projektion

Gelbe Liste Gentamicin. https://www.gelbe-liste.de/wirkstoffe/Gentamicin_82

Glycosidische Bindung https://de.wikipedia.org/wiki/Glycosidische_Bindung

Guanin https://commons.wikimedia.org/wiki/File:Guanin.svg
Ersteller NEUROtiker gemeinfrei (Basenpaare).

Huffman-Codierung https://de.wikipedia.org/wiki/Huffman-Kodierung

Humangenomprojekt https://de.wikipedia.org/wiki/Humangenomprojekt

Influenzaimpfstoff https://www.meduniwien.ac.at/web/ueber-uns/news/2024/news-im-februar-2024/florian-krammer-uebernimmt-professur-fuer-infektionsmedizin-an-der-meduni-wien/

In-Frame-Deletion https://flexikon.doccheck.com/de/In-frame-Deletion

In-Frame-Insertion https://flexikon.doccheck.com/de/In-frame-Insertion

Intronphasen https://de.wikipedia.org/wiki/Intron#Intronphasen

IUPAC https://iupac.qmul.ac.uk/BlueBook/PDF/BlueBookV2.pdf
IUPAC Nomenclature of Organic Chemistry. IUPAC Recommendations and Preferred Names 2013. Prepared for publication by Henri A. Favre and Warren H. Powell, Royal Society of Chemistry, ISBN 978-0-85404-182-4.

Kerbholz https://de.wikipedia.org/wiki/Kerbholz

Lexikon der Biologie Spektrum. https://www.spektrum.de/alias/lexikon/lexikon-der-biologie/574856

MDR1-Defekt https://de.wikipedia.org/wiki/MDR1-Defekt

Medizinskandal vor Gericht, NZZ https://www.nzz.ch/ein_medizin-skandal_vor_gericht-ld.463125

Morsecode https://de.wikipedia.org/wiki/Morsecode

Mutation https://www.spektrum.de/lexikon/biologie/mutation/44508

Patente https://worldwide.espacenet.com/publicationDetails/originalDocument?CC=WO&NR=2022223617A1&KC=A1&FT=D&ND=3&date=20221027&DB=en.worldwide.espacenet.com&locale=en_EP
espacenet: Datenbank des europäischen Patentamtes
WO2022223617 (A1) mRNA-Impfstoff von Pfizer

Peptidbindung https://de.wikipedia.org/wiki/Peptidbindung

P-Glykoprotein https://flexikon.doccheck.com/de/P-Glykoprotein

Polydaktylie https://www.molgen.mpg.de/92271/research_report_331001?c=90687

Präfixcode https://de.wikipedia.org/wiki/Präfixcode

Proteindomänen https://www.spektrum.de/lexikon/biologie/protein-module/54158
Lexikon der Biologie online

Purin https://commons.wikimedia.org/wiki/File:Purines.svg
Ersteller Charlesy, gemeinfrei

Ramachandran, G. N.	https://de.wikipedia.org/wiki/G._N._Ramachandran
Random Coil	https://de.wikipedia.org/wiki/Random_Coil
redundant	Computerlexikon: http://www.computerlexikon.com/was-ist-redundant Digitales Wörterbuch der Deutschen Sprache: https://www.dwds.de/wb/Redundanz
Rett-Syndrom, Gentherapie	https://www.ema.europa.eu/en/medicines/human/orphan-designations/eu-3-23-2884
reverse Transkriptase	https://flexikon.doccheck.com/de/Reverse_Transkriptase
Ribonukleinsäure	https://www.chemie.de/lexikon/Ribonukleins%C3%A4ure.html
ribosomale Bindungsstelle	https://flexikon.doccheck.com/de/Ribosomale_Bindungsstelle
RNase	https://commons.wikimedia.org/wiki/File:RNase_H_fix.png#/media/Datei:RNase_H_fix.png. Vossmann. https://de.wikipedia.org/wiki/Ribonukleasen_H
Röntgenstrukturanalyse	https://www.spektrum.de/lexikon/biochemie/roentgenstrukturanalyse/5466 Lexikon der Biochemie, Spektrum
Sanofi	https://www.sanofi.de/de/magazin/wissenschaft/erweitertes-genetisches-alphabet-synthorine. Erweitertes genetisches Alphabet (2020)
scriptura continua/scriptio continua	https://de.wikipedia.org/wiki/Scriptio_continua
Shannon-Fano-Kodierung	https://de.wikipedia.org/wiki/Shannon-Fano-Kodierung
SNpedia RS6277	https://www.snpedia.com/index.php/Rs6277 (Mutation C957T)
Sound of Silence, The	https://de.wikipedia.org/wiki/The_Sound_of_Silence
Spikeprotein	https://flexikon.doccheck.com/de/Spikeprotein
Spliceosom/Spleißosom	https://flexikon.doccheck.com/de/Spliceosom
Strichliste	https://de.wikipedia.org/wiki/Strichliste
surjektiv, bijektiv	https://www.math.uni-bielefeld.de/~einsteigen/rematheb/apps/injektiv_surjektiv.php Mathematikseiten der UNI-Bielefeld
Tektorialmembran	https://flexikon.doccheck.com/de/Tektorialmembran
tRNA	https://de.wikipedia.org/wiki/TRNA
Unärsystem	https://de.wikipedia.org/wiki/Un%C3%A4rsystem
von-Willebrand-Faktor	https://flexikon.doccheck.com/de/Von-Willebrand-Faktor

Datenbanken

CCDS Datenbank
https://www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi

Database Commons – a catalog of worldwide biological databases
https://ngdc.cncb.ac.cn/databasecommons/database/id/6693
A catalog of worldwide biological databases (ab 2015). Es handelt sich um eine *übergeordnete Datenbank, die Datenbanken* auflistet, welche nach verschiedenen Kriterien unterteilt sind.

deepmind
https://deepmind.com/blog/article/putting-the-power-of-alphafold-into-the-worlds-hands

ensembl
http://www.ensembl.org/index.html
Gene Homo sapiens: http://www.ensembl.org/Homo_sapiens/Info/Annotation
Chromosom 1: http://www.ensembl.org/Homo_sapiens/Location/Chromosome?r=1
Chromosom 21: http://www.ensembl.org/Homo_sapiens/Location/Chromosome?r=21

Homo sapiens troponin I3, cardiac type (TNNI3), mRNA
https://www.ncbi.nlm.nih.gov/nuccore/1653962375
Homo sapiens TNNI3 gene, GenBank: X90780.1

Human Gene Mutation Database (HGMD)
https://www.hgmd.cf.ac.uk/ac/index.php
»The Human Gene Mutation Database (HGMD®) represents an attempt to collate all known (published) gene lesions responsible for human inherited disease and is maintained in Cardiff by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips, K. Evans, S. Heywood, M.J. Hayden, M.M. Chapman, M.E Mort, L. Azevedo and D.S. Millar.«

NIH National Institutes of Health
Pubmed
https://pubmed.ncbi.nlm.nih.gov

OMIM® An Online Catalog of Human Genes and Genetic Disorders
https://www.omim.org/

SynMICdb *Synonymous Mutations in Cancer database*
http://synmicdb.dkfz.de/rsynmicdb/
2019 gegründete Datenbank der Universität Freiburg/Breisgau
»SynMICdb aims to facilitate the study of synonymous mutations in cancer. These mutations alter the nucleotide sequence of the gene and the mRNA, but not the amino acid sequence of the respective protein. Synonymous mutations can change mRNA structure, stability, splicing, translation efficiency or impact cis-regulatory sequences. These mutations can be recurrent and conserved and mirror many aspects of cancer driving mutations.«

The NHGRI-EBI
Catalog of human genome-wide association studies
https://www.ebi.ac.uk/gwas

Übergeordnete Datenbank
Database Commons
https://ngdc.cncb.ac.cn/databasecommons/database/id/6693
Diese Datenbank listet andere Datenbanken auf:
(nach verschiedenen Kriterien unterteilt) Data Type: DNA, RNA, Protein:
Database Category: Raw bio-data, Gene, genome and annotation ...

Abkürzungen

A	Adenin
aa-tRNA-Synthetase	Aminoacyl-tRNA-Synthetase
AS/AA	Aminosäure/Amino Acid
C	Cytosin
CAI	Codon Adaption Index
Covid-19	CoronaVirus Disease of 2019
DAACP	D-Amino Acid Containing Peptide
DNA	Desoxyribonucleinsäure
EEJ	Exon-Exon-Junction
EJC	Exon-Junction-Complex
EMA	European Medicines Agency
ENCODE	Encyclopedia Of DNA Elements
ESE	exonic splice enhancer
ESS	exonic splice silencer
G	Guanin
HGP	Human Genome Project
IGSR	The International Genome Sample Resource – supporting open human variation data
IUPAC	International Union of Pure and Applied Chemistry (seit 1919) Normungsinstitut für die Bezeichnung chemischer Formeln mit Sitz in Zürich
local tAI	lokaler tRNA-Adaption-Index
MERS	Middle East respiratory syndrome coronavirus
mRNA	messenger RNA
m1Ψ	N1-Methyl-Pseudouridin
ncAA	non canonical Amino Acid
nt	Nucleotid (Base + Zucker + Phosphat)
NHGRI	National Human Genome Research Institute
NIH	National Institute of Health
NMD	nonsense mediated decay
NMR	Nuclear Magnetic Resonance
NTC	normal termination codon
ORF	open reading frame
PHEX -Gen	phosphate regulating endopeptidase X-linked
prä-mRNA	prä-Messenger-RNA
PTC	premature termination codon (vorzeitiges Stopp-Codon)
RNA	Ribonukleinsäure
rRNA	ribosomale RNA
SARS	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus type 2
T	Thymin
tAI	tRNA-Adaption-Index
tRNA	transfer RNA
U	Uracil
UTR	untranslated region (untranslatierter Bereich)
WIPO	Wold Intellectual Property Organisation, Sitz in Genf (CH)
WHO	World Health Organization, Sitz in Genf (CH)
vWS	von-Willebrand-Syndrom
vWF	von-Willebrand-Faktor
WES	Whole Exom Sequenzierung
Ψ	Pseudouridin

Einheiten

Basen-(Paare):

bp	Basenpaare
kbp	kilo Basenpaare (103 Basenpaare)
kb	kilo-Basen (für Einzelstrang)
Mbp	MegaBasenpaare (106 Basenpaare)
Mb	Mega-Basen für Einzelstrang
Gbp	Gigabasenpaare (109 Basenpaare)
Gb	Giga-Basen für Einzelstrang

Größen (Längen):

Å Ångström	10^{-10} m (alte Einheit) 1 Å = 0,1 nm
Mikrometer µm/µ	10^{-6} m
Nanometer nm	10^{-9} m
Picometer pm	10^{-12} m
Svedberg	S (Sv) nicht linearer Sedimentationskoeffizient für die Größe von Teilchen in einer Ultrazentrifuge mit CsCl (Gleichgewichtszentrifugation z.B. von Ribosomen). [10^{-13} s]

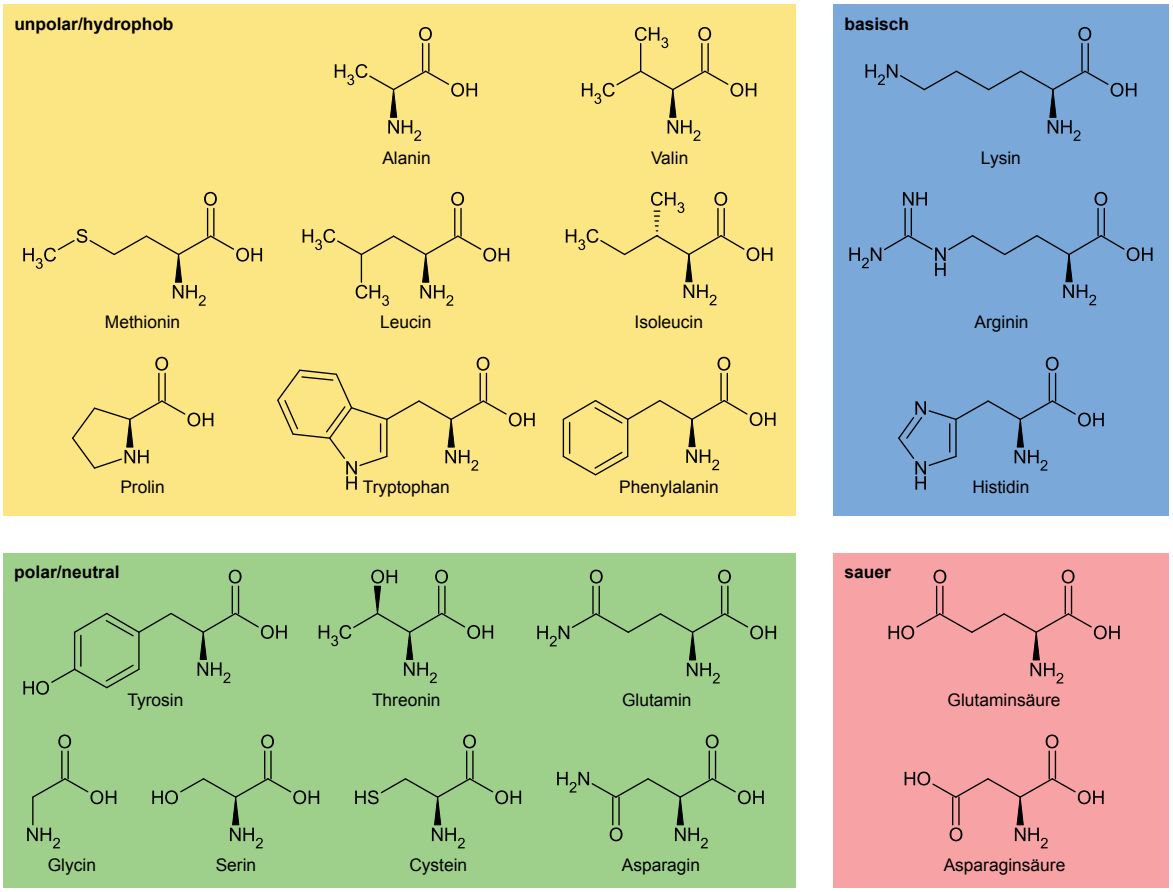
Anhänge

Aminosäuren

Liste der 20 kanonischen Aminosäuren mit Einbuchstabencode und Dreibuchstabencode in alphabetischer Reihenfolge:

Aminosäure	Dreibuchstabencode	Einbuchstabencode
Alanin	Ala	A
Arginin	Arg	R
Asparagin	Asn	N
Asparaginsäure	Asp	D
Cystein	Cys	C
Glutamin	Gln	Q
Glutaminsäure	Glu	E
Glycin	Gly	G
Histidin	His	H
Isoleucin	Ile	I
Leucin	Leu	L
Lysin	Lys	K
Methionin (Start)	Met	M
Phenylalanin	Phe	F
Prolin	Pro	P
Serin	Ser	S
Threonin	Thr	T
Tryptophan	Trp	W
Tyrosin	Tyr	Y
Valin	Val	V

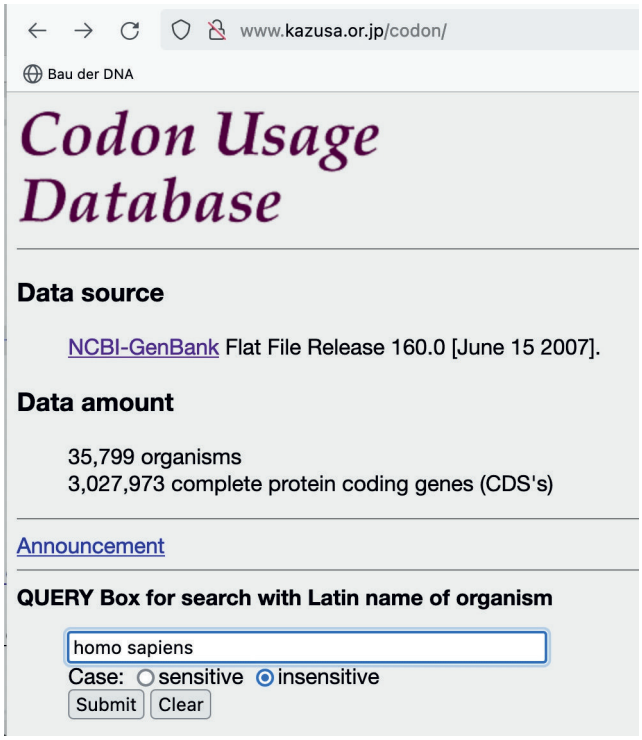
Die natürlich vorkommenden 20 proteinogenen (kanonischen) Standard-Aminosäuren, gruppiert nach physikalisch-chemischen Eigenschaften. (Wikipedia, https://de.wikipedia.org/wiki/Aminosäuren, Ersteller: sponk, gemeinfrei).

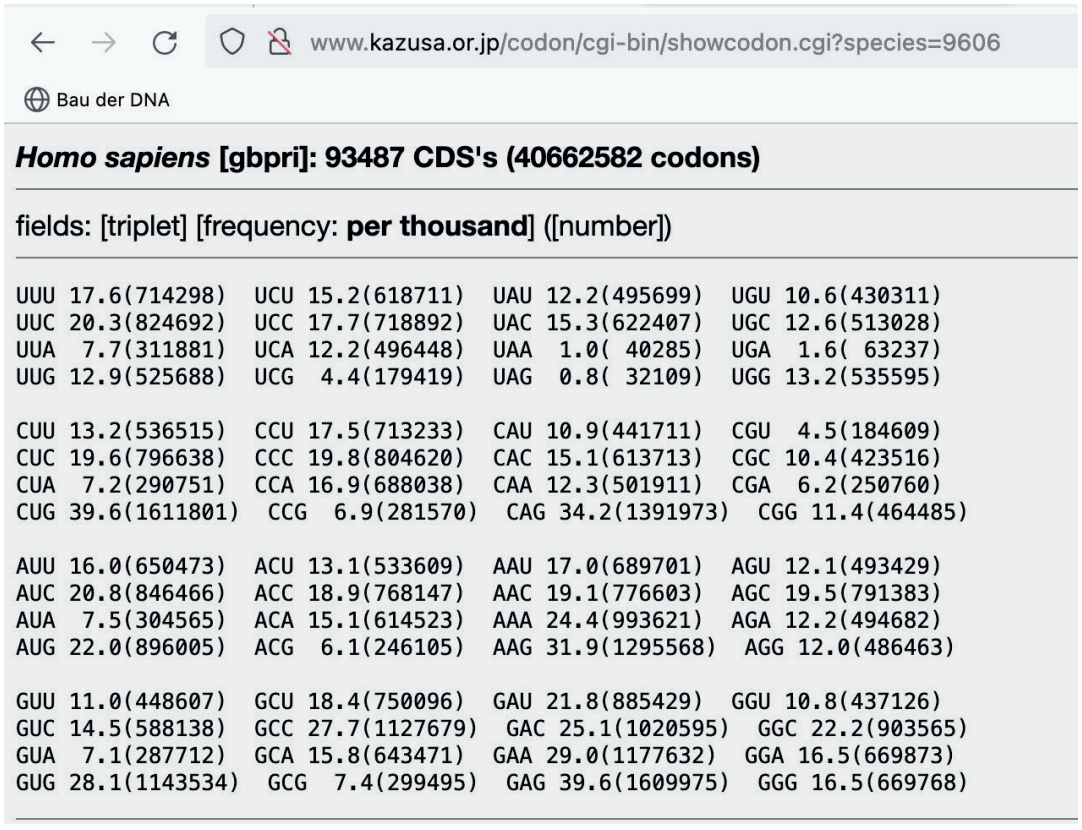


Codon Usage Tabelle

http://www.kazusa.or.jp/codon/

http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=9606





Nucleobasen

Bezeichnung laut UPAC

https://iubmb.qmul.ac.uk/misc/naseq.html

Summary of single-letter code recommendationsNomenclature for Incompletely Specified Bases in Nucleic Acid Sequences

Recommendations 1984

World Wide Web version Prepared by G. P. Moss

School of Physical and Chemical Sciences, Queen Mary University of London,

Mile End Road, London, E1 4NS, UK

Symbol	Meaning	Origin of designation
G	G	Guanine
A	A	Adenine
T	T	Thymine
C	C	Cytosine
R	G or A	puRine
Y	T or C	pYrimidine
M	A or C	aMino
K	G or T	Keto
S	G or C	Strong interaction (3 H bonds)
W	A or T	Weak interaction (2 H bonds)
H	A or C or T	not-G, H follows G in the alphabet
B	G or T or C	not-A, B follows A
V	G or C or A	not-T (not-U), V follows U
D	G or A or T	not-C, D follows C
N	G or A or T or C	aNy

Hachimoji-Basen

RNA- und DNA-Basen des Hachimoji-Codes (Hoshika et al., 2019).

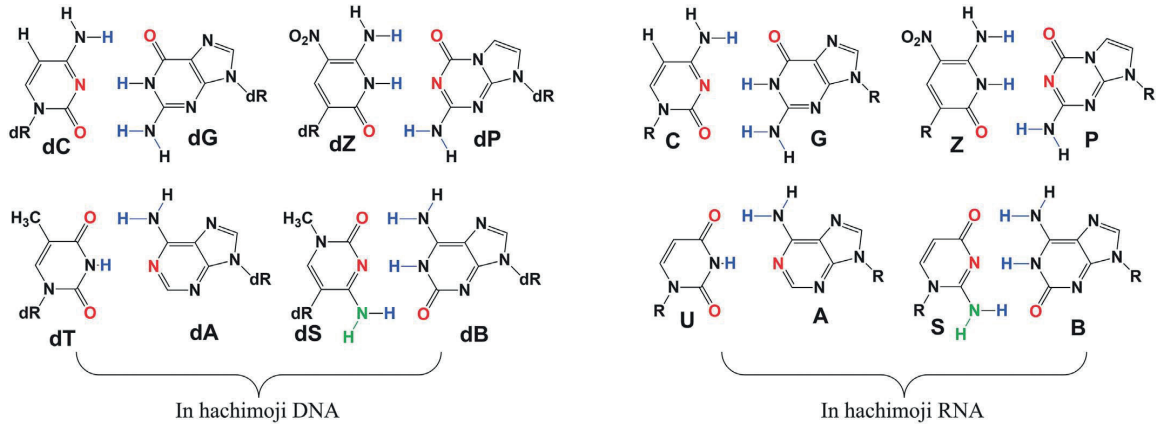


Fig. 1. The eight nucleotides of hachimoji DNA and hachimoji RNA are designed to form four size- and hydrogen bond-complementary pairs. Hydrogen bond donor atoms involved in pairing are blue; hydrogen bond acceptor atoms are red. The left two pairs in each set are formed from the four standard nucleotides (note missing hydrogen-bonding group in the A:T pair, a peculiarity of standard terran DNA and RNA). The right two pairs in each set are formed from the four new nonstandard nucleotides. Notice the absence of electron density in the minor groove of S, which has a NH (green) moiety.