

Evolution der Moleküle.

Von der Evolution im Reagenzglas zur Erzeugung von maßgeschneiderten Molekülen.

Peter Schuster. Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17,
1090 Wien und Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA.
E-Mail: pks@tbi.univie.ac.at.

Evolution wird als zentrales Fachgebiet der Biologie verstanden und dennoch gehen viele entscheidende Impulse zu einem tieferen Verstehen evolutionärer Prozesse und ihrer Zusammenhänge von Physik und Chemie aus. Insbesondere seit der Mitte des zwanzigsten Jahrhunderts gab es auf drei Gebieten spektakuläre Fortschritte: (i) bei der Erforschung der Chemie der Lebensprozesse durch die Molekularbiologie, (ii) bei der Physik der Beobachtung und ihrer analytischen Aufbereitung durch Spektroskopie und verfeinerte Trennmethoden, und (iii) bei der Datenerhebung, Datenauswertung und Theorienbildung durch den Einsatz elektronischer Rechner. Die gegenwärtige Biologie befindet sich in einem gewaltigen Umstrukturierungsprozess: Wissen, welches die belebte Welt auf der Basis der molekularen Wissenschaften interpretiert, wird in einem bisher nie dagewesenem Umfang erarbeitet, gespeichert und zur Lösung mannigfaltiger Probleme genutzt. Diese Entwicklung hat auch vor der Evolutionsbiologie nicht halt gemacht. Molekulare Einsichten stellen unser Wissen über die Vererbung auf eine völlig neue Basis. Eine Fülle von epigenetischen Phänomenen findet eine einfache Erklärung auf molekularer Grundlage. Die Regulation der Genaktivitäten verläuft bei höheren Organismen auf Wegen, die vor zwanzig Jahren noch unbekannt waren. Durch die chemisch-physikalische Analyse von Evolutionsprozessen können wir heute erahnen, was das Leben vor der anorganischen Natur auszeichnet. Evolution im Darwinschen Sinnen kann außerhalb von Zellen studiert, analysiert und zum Zweck des Designs von maßgeschneiderten Molekülen nutzbar gemacht werden. Beim gegenwärtigen Stand der Wissenschaft ergänzen evolutionäres und rationales Design einander in idealer Weise, da unser heutiges Wissen um Strukturen und Funktionen von Biomolekülen noch in vielen Bereichen lückenhaft ist.

Prolog

Evolution ist heute in der Gesellschaft zum Schlagwort geworden. Das Paradigma der steten Veränderung hat die frühneuzeitliche Vorstellung von einer harmonischen und im Gleichgewicht befindlichen Welt abgelöst. Der Wechsel von einem statischen zu einem dynamischen Weltbild begann in spektakulärer Form in der Biologie: Die bisher als konstant angesehenen, durch einen Schöpfungsakt oder anders entstandenen Arten wurden als das Resultat einer Momentaufnahme eines dauerndem Wandel unterworfenen Prozesses erkannt, wie schon durch den Titel des Jahrhundertwerkes *„On the Origin of Species“* von Charles Darwin zum Ausdruck kommt.¹ Die biologische Evolution ist aber keineswegs bloße Dynamik, denn die

Entwicklung des irdischen Lebens ist durch eine Zunahme der Komplexität der Biosphäre und durch das Entstehen immer komplexer strukturierter oder höher geordneter Organismen charakterisiert. Das Verständnis für Dynamik verbreitete sich nur allmählich: Alfred Wegeners frühe Arbeit über die Veränderung der Erdkruste² wurde von seinen Zeitgenossen aus der Geologie abgelehnt und die Bestätigung für die Richtigkeit seiner Vermutungen erfolgte erst mehr als ein halbes Jahrhundert später. Am Anfang der aus der Himmelsmechanik entstandenen klassischen Physik steht mit der Kinematik auch das Denken in zeitlichen Abläufen im Zentrum aber hier handelt es sich um Formen *ewiger* Bewegungen: Klassische Oszillationen oder Planetenbahnen sind periodisch und ebenso wie elastische Stöße umkehrbar oder reversibel.^a

Die systematische Untersuchung irreversibler Prozesse begann erst mit der Entwicklung der Thermodynamik in der zweiten Hälfte des neunzehnten Jahrhunderts. Alle natürlichen Vorgänge sind irreversibel und können nicht umgekehrt werden, ohne andere Veränderungen in der Welt zu hinterlassen. Diese fundamentale Einsicht verdanken wir unter anderen drei Physikern, dem Deutschen Rudolf Clausius, dem Schotten James Clerk Maxwell und dem Österreicher Ludwig Boltzmann. Zu Anfang schien die Thermodynamik in Form ihres zweiten Hauptsatzes einer Vorstellung der biologischen Evolution zu widersprechen. Die beiden Hauptsätze wurden von Clausius auf die knappe Form gebracht: 1. Die Energie der Welt ist konstant und 2. die Entropie der Welt strebt einem Maximum zu. Entropie ist ein Maß für Unordnung und die Bedingung der steten Zunahme der Entropie ist ein Ausdruck für die Irreversibilität aller natürlichen Vorgänge. Wie sollten da Ordnung und Struktur entstehen können, wenn die Zunahme der Unordnung ein Naturgesetz ist? Schon für Boltzmann, einem glühenden Verehrer Darwins, waren der zweite Hauptsatz und die biologische Evolution vereinbar aber die systematische Untersuchung von Ordnungsentstehung blieb der zweiten Hälfte des zwanzigsten Jahrhunderts vorbehalten. Aus einzelnen Studien über spontane Musterbildung entstand eine physikalische Theorie der Selbstorganisation und die Antwort auf die oben aufgeworfene Frage entpuppte sich als nahezu trivial: Ordnung kann nur in einem Teil der Welt entstehen und die Zunahme an Unordnung im Rest der Welt muss die Ordnungszunahme in diesem Teil überkompensieren. Nichtgleichgewichtsstrukturen oder dissipative Strukturen können nur in Systemen mit Energie- und/oder Materialfluss existieren

^a Ein reversibler Vorgang ist dadurch gekennzeichnet, dass eine Umkehr der Zeit – in den dynamischen Gleichungen wird t durch $-t$ ersetzt – zu einem in der Realität möglichen Prozess wird. Ein illustratives Beispiel für einen reversiblen Prozess ist die Bewegung eines Pendels: Abgesehen von der sogenannten Phase, welche durch den Anfangszeitpunkt der Bewegung festgelegt wird, sind die Bewegungen für t und $-t$ identisch. Als Beispiel für einen irreversiblen Vorgang kann der Temperatenausgleich dienen. Eine an einem Ende glühende und am anderen Ende kalte Eisenstange wird nach einiger Zeit dieselbe Temperatur an beiden Enden aufweisen. Der in der Zeit umgekehrte Vorgang, dass eine Eisenstange an einem Ende zu glühen beginnt, wogegen das andere Ende kalt wird, wurde bis jetzt niemals beobachtet.

und ein solcher ist auf der Erde immer gegeben, da die Sonnenstrahlung mit einer Temperatur von 5800 K höherwertige Energie besitzt als die von der Erde abgegebene Strahlung mit einer Temperatur von 255 K.^b Zwei Konsequenzen des eben geschilderten, aus den Gesetzen der Thermodynamik abgeleiteten Sachverhalts sind dessen ungeachtet von weltbildhafter Dimension: (i) Die biologische Evolution kann nicht in alle Ewigkeit weitergehen, und (ii) die biologische Evolution ist ein lokales Phänomen und daher auf einen Teil der Welt beschränkt.

Die ebenfalls in der zweiten Hälfte des zwanzigsten Jahrhunderts stattfindende, spektakuläre Entwicklung der Molekularbiologie spannt eine solide Brücke zwischen Chemie und Biologie. Evolutionsvorgänge einschließlich des Darwinschen Prinzips und anderer Mechanismen können nunmehr mit den rigorosen Methoden von Mathematik und Physik in molekularer Auflösung studiert und analysiert werden. Technische Fortschritte gestatten es, die ursprünglich in Sinne einer Komplexitätsreduktion an einzelnen oder einigen wenigen Biomolekülen durchgeführten Untersuchungen in der Systembiologie in Richtung auf eine molekulare Beschreibung ganzer Zellen oder gar ganzer Organismen auszudehnen.

Von der Evolution der Moleküle soll in diesem Kapitel die Rede sein. Wir werden einfache molekulare Systeme kennenlernen, welche ungeachtet ihrer Einfachheit zur Evolution durch Variation und Selektion in Darwinschen Sinne befähigt sind. Molekulare Modellsysteme können experimentell realisiert und theoretisch analysiert werden kann. Evolution im Reagenzglas wird zur Lösung von schwierigen Aufgaben in der Biotechnologie herangezogen und einige auf Evolutionsprinzipien aufbauende technische Verfahren sind aus dem Laboralltag nicht mehr wegzudenken. Beim heutigen Stand des Wissens lässt sich die molekulare Theorie auf einfache biologische Entitäten wie Viren oder Bakterien übertragen und es werden neue Einblicke in die Grundlagen der biologischen Evolution gewonnen und neue Pharmaka auf des Basis der erweiterten Wissens entwickelt..

Darwins Selektionsprinzip 1859 und heute

In der gesellschaftlichen Diskussion wird häufig Charles Darwin mit dem Evolutionsgedanken in einem Zug genannt. In der Tat war es das Verdienst des Genies Darwins, aus einer Fülle von Beobachtungen die Essenz der evolutionären Veränderungen herausgefunden zu haben: Erfolg in der Biologie wird stets umgesetzt und erfolgreichere Varianten haben eine größere Zahl von Nachkommen. Da Eigenschaften und strukturelle Merkmale vererbbar sind, kommt das Erbmaterial, das zu mehr Erfolg führt, in den kommenden Generationen häufiger vor. Das

^b Das Ausnutzen einer Temperaturdifferenz zur Arbeitsleitung ist das Prinzip der Dampfmaschine. Auf der Erde wird die genannte Differenz in der Strahlungstemperatur der von der Sonne kommenden und der in den Weltraum abgestrahlten Energie durch die Pflanzen in der Photosynthese der für chemische Synthesen genutzt.

Darwinsches Prinzip baut auf drei Voraussetzungen auf: (i) Vermehrung mit Vererbung – die Nachkommen gleichen ihren Vorfahren in vielen ihrer Merkmalen, (ii) Variation – der Vermehrungsprozess ist als Kopiervorgang nicht vollkommen und daher unterscheiden sich die Kopien mitunter von den Vorlagen – und (iii) Selektion – die Populationen sind zwangsläufig endlich und dies hat zur Konsequenz, dass die weniger erfolgreicheren Varianten mit der Zeit aussterben.

Das Erfolgsrezept der Darwinschen Theorie besteht in seiner Einfachheit und in seinem hohen Abstraktionsgrad. Für das Eintreten der natürlichen Auslese ist es unwesentlich, nach welchen Mechanismen vererbt wird und wie die Varianten entstehen. Worauf es einzig ankommt, ist die Tatsache, dass es Vererbung gibt und dass Varianten gebildet werden. In der Tat waren Darwins Gedanken zur Vererbung schlichtweg falsch und von der Natur der Mutationen gab es bis zur Mitte des vorigen Jahrhunderts überhaupt keine brauchbaren Vorstellungen. Ein weiterer wichtiger Gesichtspunkt für die vielfältige Anwendbarkeit des Darwinschen Prinzips besteht darin, dass auch die Natur der sich vermehrenden Einheiten keine Rolle spielt. Ob wir eine Population von Molekülen, Zellen, Organismen oder anderen Einheiten betrachten, ist unerheblich, wenn die drei genannten Voraussetzungen erfüllt sind, wird Selektion eintreten. Andererseits macht die Darwinsche Theorie auch keine Aussagen darüber, was man tun muss, um das zu erwartende oder zu befürchtende Ergebnis zu beeinflussen, ausgenommen die einsichtige Maßnahme, die Zahlen unerwünschter Nachkommen zu reduzieren oder diese ganz zu eliminieren, worin das Grundprinzip jeder Tier- oder Pflanzenzüchtung besteht.

Was hat sich an den Evolutionsvorstellungen seit den mehr als 150 Jahren nach dem Erscheinen der Darwinschen ‚*Origin of Species*‘ geändert? Schon zur Lebenszeit Darwins machte der Augustinermönch Gregor Mendel die bahnbrechende Entdeckung von Regeln der Vererbung bei höheren Organismen:³ Die genetische Information der beiden Elternteile wird in kleine Pakete – heute nennen wir sie Gene – zerlegt und in den Nachkommen neu kombiniert, wobei die Auswahl der Genvariante – väterlich oder mütterlich – zufällig ist. Die *Rekombination* der elterlichen Gene – Allele genannt – ergibt das Genom des männlichen oder weiblichen Nachkommen. Nach den Mendelschen Vorstellungen ist nur die Kombination der beiden Allele für die ererbten Eigenschaften maßgeblich und abgesehen vom geschlechtsspezifischen Erbmateriale spielt es keine Rolle, welches der beiden Allele vom Vater und welches von der Mutter kommt. Die Bedeutung der Mendelschen Erkenntnisse wurde den Evolutionsbiologen vorerst nicht bewusst. Erst bei der sogenannten Wiederentdeckung der Arbeiten Mendels um 1900 wurde der Begriff des *Gen*s eingeführt und es gelang dem Mathematiker Ronald Fisher zusammen mit seinen Kollegen J.B.S. Haldane und Sewall Wright die Darwinsche Evolutionstheorie und die Mendelsche Vererbungslehre im theoretische Gebäude der Populationsgenetik zu vereinigen.

Der zentrale Begriff des Gens – ursprünglich erfunden als eine Abstraktion gewonnen aus empirischen Resultaten – fand fürs erste in der Molekularbiologie eine zufriedenstellende Interpretation, welche auch in der Lage war die meisten Abweichungen von den Mendelschen Regeln zu erklären. Als Geburtsstunde der Molekularbiologie wird vielfach der korrekte Vorschlag der Doppelhelix-Struktur der Desoxyribonukleinsäure DNA durch James Watson und Francis Crick im Jahre 1952 angesetzt.⁴ Die vorgeschlagene Struktur war in zweifacher Hinsicht bahnbrechend: (i) Die Struktur gab den entscheidenden Hinweis auf das molekulare Prinzip der biologischen Vermehrung (Abbildung 1) und (ii) sie enthüllte den Trick der Natur zwischenmolekulare Kräfte zur digitalen Codierung der genetischen Information. In das durch die Molekülgeometrie der Helix festgelegte Gerüst zweier DNA-Moleküle passen nur die komplementären Nukleotidbuchstabenpaare, $A \Leftrightarrow T(U)$ und $G \Leftrightarrow C$, und das hat zur Konsequenz, dass bei Kenntnis eines der beiden DNA-Moleküle aus der Doppelhelix das zweite eindeutig ergänzt werden kann (Abbildung 1). Die andere Paarungen ausschließende Komplementarität ist eine Form von Digitalisierung der chemischen Wechselwirkungen und bildet letztlich die Grundlage für die digitale Codierung von Information über Strukturen und Funktionen und für die Weitergabe dieser digital codierten Information von Generation zu Generation. Vermehrung, Vererbung und Variation finden ihre heutige Erklärung durch die Kenntnis der molekularen Strukturen der Nukleinsäuren. Wenige Jahre nach der Publikation der Struktur Doppelhelix goss Crick das Wissen der frühen Molekularbiologie in das sogenannte *zentrale Dogma der Molekularbiologie*:^{5,6,7,c} Die genetische Information, die in den Sequenzen der Biopolymeren gespeichert ist, fließt zwischen den Nukleinsäuren und von Ribonukleinsäure zu Protein, und niemals von Protein zur Nukleinsäure (Abbildung 2). Das Dogma ist die Konsequenz einer biochemischen Tatsache: Weder Kopieren, $DNA \rightarrow DNA$, $RNA \rightarrow RNA$, und Umschreiben von Nukleinsäuren, $DNA \rightarrow RNA$, $RNA \rightarrow DNA$, noch das Übersetzen von RNA in Protein ist ohne Katalyse möglich. Während der Lebensentstehung auf der Erde wurden molekulare Maschinen entwickelt, welche Nukleinsäuren kopieren und Nukleinsäuresequenzen in Proteine unter Berücksichtigung des – fast – universellen genetischen Codes übersetzen. Der genetische Codes ordnet drei Nukleotidbuchstaben jeweils eine Aminosäure zu (Abbildung 2). Die molekularen Maschinen sind evolutionshistorisch gewachsen und können wegen ihrer enormen molekularen Komplexität nicht oder nur mehr unwesentlich verändert werden.^d

^c Der Ausdruck *Dogma* ist wie auch Francis Crick später selbst bemerkte für den Sachverhalt der zellulären Proteinsynthese unpassend. Ein Dogma ist ein nicht zu hinterfragender Glaubensinhalt, wogegen das zentrale Dogma einen klar begründeten chemisch-biologischen Sachverhalt darstellt. Eine ins Detail gehende Darstellung des zentralen Dogmas und seiner Geschichte findet sich in Ref.7.

^d Ähnliches gilt auch für die durch die Evolution entstandenen höheren Lebewesen: Auch wenn es für Vögel einen großen Selektionsvorteil bedeuten würde, sechs Gliedmaßen – zwei Beine, zwei Flügel und zwei Hände – zu besitzen, die Komplexität der organismischen Entwicklung macht eine Rückkehr zu früheren Formen mit mehr Gliedmaßen unmöglich.

Dessen ungeachtet ist der genetische Code nicht *völlig eingefroren* sondern unterliegt selbst einer zwar nur marginalen Evolution. Es gibt bei Organellen wie Mitochondrien und einigen primitiven Organismen geringe, idiosynkratische Abweichungen von der Universalität.⁸ In der Natur kommen auch Erweiterungen des genetischen Codes von 20 auf 22 Aminosäuren vor.⁹ Eine überaus interessante Entwicklung auf dem Gebiet der synthetischen Biologie befasst sich mit der Umprogrammierung der ribosomalen Proteinsynthese: Durch die Verwendung von synthetischen Komponenten ist es gelungen durch Abwandlung des genetischen Codes unnatürliche Aminosäuren am Ribosom in Proteine einzuschleusen,¹⁰ und dies eröffnet der Biotechnologie völlig neue Wege der Proteinsynthese.

Etwa gleichzeitig mit dem Wissen um die Prozessierung der genetischen Information in der Zelle entstanden auch die Vorstellungen über die Arbeitsteilung zwischen den Biomolekülen: Nukleinsäuren sind die Informationsträger – im Jargon der Informatik ist die RNA die ‚*Working-Copy*‘ eines Gens und die DNA die ‚*Backup-Copy*‘ des Genoms – und die Proteine übernehmen alle anderen Funktionen von hochspezifischen Katalysatoren metabolischer Reaktionen bis zu Gerüstsubstanzen mit besonderen Eigenschaften. Die einzigen bekannten Ausnahmen waren die in die Proteinsynthese involvierten RNA-Moleküle: die transfer RNA-Moleküle die ribosomalen RNA-Moleküle. Eine Entdeckung um das Jahr 1980 herum warf diese Vorstellung über den Haufen: RNA-Moleküle können bestimmte Reaktionen genau so spezifisch katalysieren wie Proteine (siehe den Abschnitt *Von der Theorie zur Anwendung*).^{11,12} Später wurde gezeigt, dass auch DNA-Moleküle als Katalysatoren wirken können.¹³

Die einzelnen Zellen eines Organismus *exprimieren*^e nur einen Bruchteil aller Gene und eine Grundfrage der Molekulargenetik war und ist es, auf welche Weise geregelt wird, welche Gene exprimiert und welche Gene stillgelegt werden. François Jacob und Jacques Monod fanden schon im Jahre 1960 an Hand eines metabolischen Gens des Bakteriums *Escherichia coli*, dass ein Protein das Gen aus- und einschalten kann.¹⁴ Die für dieses Protein – den sogenannten *lac-Repressor* – kodierende DNA befindet sich an einer anderen Stelle des Genoms aber vor den betreffenden Genen – in diesem Fall vor dem *lac-Operon* – liegt ein Sequenzabschnitt auf der DNA den der Repressor höchst spezifisch erkennt: Ist der Repressor an die DNA gebunden, wird kein Protein synthetisiert. Durch die Entdeckung von Jacob und Monod wurde erstmals die Vorstellung eines logisch konsistenten Synthese- und Regulationssystems der Zelle möglich und die Molekularbiologie konnte eine erste Vorstellung von der Chemie des Lebens bieten. Bis ins kleinste Detail gehende Untersuchungen vornehmlich an Bakterien und Viren enthüllten eine

^e Exprimieren von Genen bedeutet, dass die betreffenden Gene tatsächlich in Proteinstrukturen übersetzt werden. Die Genexpression in einer Zelle hängt von ihrer und von ihrem Zustand insbesondere vom Lebensalter im Zellzyklus ab.

Fülle von weiteren Einzelheiten. Unter anderem wurde gezeigt, dass die Expression eines Gens nicht nur von einem einzigen Proteinmolekül sondern im Allgemeinen von mehreren Molekülen gesteuert wird. Nichtsdestoweniger kristallisierte sich die Vorstellung heraus, dass das Gen ein Stück DNA darstellt, welches für ein Protein codiert und von Sequenzen begleitet wird, die für die Genregulation unentbehrlich sind. Durch die *ein Gen* \Rightarrow *ein Protein*-Hypothese wurde der abstrakte Genbegriff der Genetik fassbar.¹⁵

Der enorme technische Fortschritt bei der Sequenzierung von Nukleinsäuren initiiert in den 1970er Jahren durch die bahnbrechenden Arbeiten von Frederick Sanger¹⁶ und Walter Gilbert¹⁷ und weitergehend bis heute¹⁸ hat eine wahre Revolution in der molekularen Genetik eingeleitet. Waren die Untersuchungen in Biochemie und Molekularbiologie auf ein einziges oder einige wenige Moleküle in Form hochgereinigter Substanzen beschränkt, so konnte man nunmehr daran denken Zellen als Gesamtheit molekular zu erfassen. Es begann mit der Sequenzierung ganzer Genome von Viren mit einigen tausend Nukleotidbuchstaben über Bakterien mit einigen Millionen bis zu den höheren Organismen mit einigen Milliarden Buchstaben – das menschliche Genome ist beispielsweise 3×10^9 Nukleotide lang und auf 23 Chromosomenpaaren gespeichert. Die an Bakterien erhobenen Befunde entsprachen den Vorstellungen: Die Regelung der Genexpression erwies sich zwar komplexer als erwartet, folgt aber den bekannten Prinzipien. Ein typisches Bakterium wie *Escherichia coli* – Stamm K-12 – enthält in seinem Genom, welches auf einer zirkulären DNA mit 4.6×10^6 Nukleotidpaaren gespeichert ist, 4288 Protein codierende Gene, die in 2584 Operons organisiert sind, sieben Operons für ribosomale RNA-Moleküle und 86 Gene für transfer-RNAs. Dementsprechend codiert der ganz überwiegende Teil des *E. coli* Genoms für Protein und die restlichen Gene codieren RNA-Moleküle, die in der zellulären Proteinsynthese benötigt werden. Bis in die 1980er Jahre wurde stillschweigend angenommen, dass das an Bakterien erforschte genetische Regulationssystem auch bei höheren Organismen seine Gültigkeit hat.

Die erste Überraschung war die Tatsache, dass nur ein Bruchteil der DNA höherer Organismen für Proteine codiert– im Fall des Menschen codieren nur etwa 1.5 % der DNA für 20000 Proteingene – und daher vorerst angenommen wurde, dass der Rest funktionslose „junk-DNA“ darstellt.¹⁹ In der Folge änderten zwei Entdeckungen die vorherrschende Meinung über die Regulation der Genexpression bei höheren Organismen: (i) Ein DNA-Abschnitt kann in den verschiedenen Geweben in verschiedene Proteine übersetzt werden^f und (ii) nicht codierende

^f Man spricht von alternativem „Splicing“ der transkribierten RNA. Die von der DNA umgeschriebene messenger-RNA enthält innerhalb der Protein codierenden Sequenz Abschnitt, sogenannte „Introns“, welche vor der Übersetzung in Protein herausgeschnitten werden. Beim alternativen „Splicing“ erfolgt das Herausschneiden unterschiedlich in den verschiedenen Geweben, wodurch verschiedene Proteine gebildet werden. Daher bilden

DNA wird zum größten Teil in RNA transkribiert – umgeschrieben – und die auf diese Weise gebildeten RNA-Moleküle haben entscheidende Funktionen in der zellulären Regulation.²⁰ Alternatives ‚*Splicing*‘ erhöht die Zahl der funktionellen Proteine bei gleichbleibender Zahl der Gene. Die Regulation der Genexpression bei höheren Organismen erwies sich als überaus komplex: In den letzten 15 Jahren wurde eine wahre Fülle von Einzelheiten über die Rolle der RNA in diesem Zusammenhang bekannt und die Reihe der Entdeckungen geht ungebremst weiter. Viele an Bakterien gewonnene Erkenntnisse lassen sich nicht auf höhere Organismen übertragen und dies gilt insbesondere für Genregulationsnetzwerke. Man mag sich fragen, warum erfindet die Natur bei den höheren Organismen ein neu Art der Genregulation? Eine gängige Vermutung geht davon aus, dass ausschließlich auf Operons aufbauende genetische Netzwerke für eine wesentlich größere Zahl von Genen als jene, die bei Bakterien realisiert ist, einen für die Zellen nicht leistbaren *Regulationsoverhead* erfordern würde,²¹ und deshalb ein weiteres Regulationsprinzip notwendig wurde. Andere Hypothesen bringen die RNA-Regulation mit Zelldifferenzierung, Entwicklung und anderen komplexen biologischen Phänomenen in Zusammenhang.²² Der Genbegriff im Sinne von *ein Gen* \Rightarrow *ein Protein* ist obsolet geworden. Heute wird das Gen am besten als ein Abschnitt auf der DNA definiert, der für die Bildung eines bestimmten funktionellen Produktes, Protein oder RNA, verantwortlich ist und dessen Sequenzinformation in codierter Form enthält.

Genetik ist nur eine von mehreren Formen der Vererbung und unter dem Begriff *Epigenetik* wird eine ganze Reihe von Mechanismen des Transfers codierter Information aus früheren Generationen und Umweltfaktoren auf einen Organismus verstanden. Es drängt sich die Frage auf: Haben Molekularbiologie und Molekulargenetik dann überhaupt zu einem Verstehen der Evolution beigetragen können oder haben sie nur Verwirrung gestiftet? Es fällt nicht schwer eine klare Antwort zu geben, denn erst durch die molekularen Einsichten wurden nicht wenige Phänomene entmystifiziert und aus Narrativen wurden quantifizierbare Begriffe. War Epigenetik die längste Zeit im 20. Jahrhundert der *Mistkübel* für unverstandene Beobachtungen bei der Vererbung, so kommt jetzt dadurch Licht ins Dunkel und die molekularen Mechanismen werden erkennbar. Es soll nur ein Beispiel angeführt werden, welches als genomische Prägung – ‚*genomic imprinting*‘ – charakterisiert wird.²³ Seit langem war es bekannt, dass genetische Erkrankungen unterschiedliche Auswirkungen zeigen können, je nachdem ob das defekte Gen vom Vater oder von der Mutter ererbt wurde, obwohl die Mendelsche Genetik einen solchen Unterschied nicht vorsieht, wenn das Gen auf einem *Autosom* liegt. Die Erklärung dieses epigenetischen Phänomens besteht darin, dass an einem Genort entweder das väterliche oder das mütterliche Allel – aber nur eines von beiden – durch chemische Modifikation der DNA

die 20000 Gene auf der DNA eine sehr viel größere Zahl an Proteinen. Diese Zahl erhöht sich noch weiter durch unterschiedliche Modifikation der Proteine nach der Übersetzung.

stillgelegt ist.²⁴ Neben der elternabhängigen Vererbung ist für die genomische Prägung die Reversibilität der Stilllegung charakteristisch. Zum Unterschied davon kann ein Gen, das durch eine Defektmutation stillgelegt wurde, in künftigen Generationen nicht mehr wieder aktiviert werden. Eine wichtige Konsequenz der genomischen Prägung führt dazu, dass es bei Säugetieren keine Parthenogenese gibt, denn an bestimmten Genorten ist das mütterliche Allel stillgelegt und das Fehlen des väterlichen Gens ist letal.

Nahezu das gesamte Wissen über Vererbung und Variation, das zu Darwins Zeiten verfügbar war, ist heute obsolet. Wir stehen vor einer immer noch wachsenden Menge an äußerst komplexen Einzelheiten, die in ein umfassendes Gebäude einzuordnen die vordringliche Aufgabe einer neuen theoretischen Biologie sein muss.²⁵

Darwinsche Selektion so einfach wie möglich

In der Natur arbeitet das Darwinsche Prinzip mit sehr komplexen Einheiten, Zellen oder Organismen. Um die drei Voraussetzungen für das Eintreten Darwinscher natürlicher Auslese – Vermehrung mit Vererbung, Variation und Selektion – zu erfüllen, brauchen die Objekte, auf denen Selektion wirken soll, aber nicht so komplex zu sein. Die Frage nach einem minimalen molekularen und dennoch evolutionsbefähigten System wurde schon in den 1960er Jahren beantwortet.^{26,27} Ein solches Minimalsystem verwendet dieselben Träger genetischer Information – DNA oder RNA – wie die Natur und daher sind auch die für die Codierung verwendeten Buchstaben – A, T (U), G und C – identisch. Das Prinzip der Vermehrung ist in Abbildung 1 skizziert und der Replikationsmechanismus bietet auch eine Erklärung für das Auftreten von Mutanten als Basis der Variationen: Mutationen sind Kopierfehler.⁸ Selektion tritt dann, wie schon betont wurde, von selbst ein, da sie eine unvermeidbare Konsequenz der endlichen Größe von Populationen darstellt. Zur erfolgreichen Implementierung eines Replikationsassays wird noch einiges andere benötigt, aber das sind weder intakte Zellen noch Organismen, denn die drei Voraussetzungen können auch von Molekülen erfüllt werden.

Vermehrung durch eindeutige Ergänzung eines Einzelstranges zum Doppelstrang (Abbildung 1) ist ein einfaches Prinzip und dennoch stößt seine Umsetzung auf eine fundamentale Schwierigkeit: Die Ausbildung von Nukleobasenpaaren und ihre Einpassung in die Doppelhelix der Nukleinsäure ist ein von der Thermodynamik favorisierter Prozess. Das Aufschmelzen einer

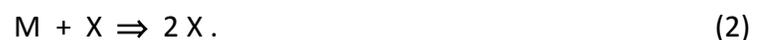
⁸ Kopierfehler bilden weitem nicht den einzigen Mutationsmechanismus: DNA kann beispielsweise auch durch verschiedene äußere Einwirkungen – Chemikalien, energiereiche Strahlung in Form von Röntgenstrahlen oder γ -Strahlen – auf einzelne Nukleobasen modifiziert werden und dann kann bei der nächsten Replikation an dieser Stelle eine Mutation, falls nicht vorher ein Korrekturenzym die Fehlstelle repariert hat .

Doppelhelix und die Trennung in zwei Einzelmoleküle kostet daher Energie und zwar umso mehr Energie je länger die Doppelhelix ist. Die nächstliegende Kopiermethode bestehend aus dem Ergänzen komplementärer Stränge, X_+ und X_- , und dem Aufschmelzen von Doppelsträngen,



kann nicht effizient implementiert werden.^h Der Trick der Natur, der von RNA-Viren verwendet wird und den sich auch Sol Spiegelman²⁰ bei seinen Experimenten zur *in vitro* Evolution von RNA-Bakteriophagen durch das Zugeben eines RNA polymerisierenden Enzyms – einer Replikase -- zunutze gemacht hat, besteht darin, die Separation der Doppelhelix schon während der Synthese durch den Katalysator zu bewerkstelligen.²⁸ Eine einfache Rechnung zeigt, dass die in der obigen Gleichung (1a,b) dargestellte plus-minus Replikation in zwei Phasen erfolgt: (i) Eine rasche Phase, in welcher sich plus- und minus-Strang in eine Art Gleichgewicht setzen, und (ii) eine Wachstumsphase, in welcher das plus-minus-Ensemble gemeinsam wächst. In der molekularbiologischen Labortechnik wird routinemäßig zur Vervielfältigung von DNA eine Reaktion verwendet, die dem komplementären – plus-minus – Mechanismus entspricht. Man bedient sich dabei eines speziellen, einsträngige DNA zur Doppelhelix vervollständigenden Enzyms des Bakteriums *Thermus aquaticus* – Taq-Polymerase – und nimmt den Schritt des Aufschmelzens der Doppelhelix durch Erwärmen der Probe vor.ⁱ Diese Methode – ‚*Polymerase Chain Reaction*‘ (PCR) genannt – stellt die wichtigste Technik zum Nachweis einiger weniger DNA-Moleküle durch Vervielfältigung dar und wird in vielen Bereichen unter anderem auch in der forensischen Pathologie zur Identifizierung eindeutiger menschlicher Spuren angewandt.

Bei der zellulären DNA-Replikation benutzt die Natur eine alternative Vorgangsweise wird: Die Doppelhelix wird mit Hilfe einer komplexen Maschinerie die mehr als 20 Proteine verwendet direkt kopiert, wobei jeder der beiden Stränge zu einer Doppelhelix vervollständigt wird:



Die beiden DNA-Moleküle der nächsten Generation enthalten dementsprechend einen neu synthetisierten und einen parentalen Strang und aus diesem Grund wird diese Form der DNA-Replikation als *semikonservativ* charakterisiert. Die Bruttoreaktionsgleichung (2), $M+X \rightarrow 2X$, symbolisiert einen Vermehrungsvorgang oder eine autokatalytische Reaktion in der Sprache des

^h Mit M bezeichnet wird das für die Synthese des Nukleinsäuremoleküls nötige Material.

ⁱ Temperaturerhöhung führt zum Aufschmelzen von DNA- oder RNA-Doppelhelices. Die notwendige Energie wird dabei in Form von Wärme zugeführt.

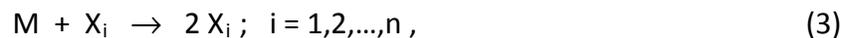
Chemikers. Falls die Menge an verfügbarem Material M praktisch unbegrenzt ist, wächst die Konzentration von X exponentiell mit der Zeit: $x(t) = x(0) e^{kt}$, wobei die Größe k als Malthus-Parameter bezeichnet wird und $x(0)$ die zum Zeitpunkt $t = 0$ vorhandene Menge an X darstellt. In der Wachstumsphase verhält sich ein nach dem Mechanismus (1a,b) replizierendes plus-minus-Ensemble wie ein einziger Autokatalysator, der sich gemäß (2) vermehrt.

In Spiegelmans Experimenten²⁹ wurde die RNA des Bakteriophagen Q β in ein Reaktionsmedium bestehend aus den für die RNA-Synthese notwendigen Bausteinen – A, T, G und U – und einem Phagen spezifischen Replikationsenzym, der Q β -Replikase, in ein für die Reaktion geeignetes Milieu eingebracht, worauf spontan RNA-Synthese einsetzte. Die Geschwindigkeit oder *Rate* der Synthese wurde durch radioaktive Markierung gemessen. Spiegelman hat geeignete Bedingungen für eine Evolution der Moleküle im Reagenzglas dadurch geschaffen, dass er regelmäßig nach einer bestimmten Zeitspanne Δt eine kleine Probe aus der jeweils letzten Reaktionsmischung in neues – unverbrauchtes – Reaktionsmedium überimpfte und auf diese Weise den durch die Synthese eingetretenen Verbrauch kompensierte. Diese Methode wird auch in der Mikrobiologie verwendet und allgemein als ‚*serial transfer*‘ charakterisiert. Spiegelmans beobachtete an Serien von 70 und mehr Überimpfungen machte, dass die RNA-Syntheserate mit der Zeit zunahm und dass diese Zunahmen nicht graduell sondern abrupt und stufenweise erfolgten, und er hat den Verlauf des Experiments auch richtig gedeutet: Durch Fehlerhaftes Kopieren entstehen neue RNA-Moleküle und jene, die sich rascher replizieren, werden in den künftigen Generationen angereichert, das heißt vermehrt auftreten. Schlussendlich werden nur jene Varianten in der Reaktionslösung verbleiben, welche sich unter den gegebenen Bedingungen am raschesten vermehren. Spiegelman hat in seinen Überimpfserien auch getestet, ob die virale RNA Bakterien infizieren kann und dabei gefunden, dass die sich rascher replizierenden Varianten ihre Infektiosität verloren hatten. Die RNA hatte sich an die neue Umgebung des Reaktionsmilieus angepasst und alle Information abgestoßen, welche für die Infektion der Bakterien und während des Phagen-Lebenszyklus in den Zellen notwendig war.

Obwohl die chemische Kinetik der Replikation von Ribonukleinsäure-Molekülen mit Hilfe des Enzyms Q β -Replikase einen durchaus komplexen Vielstufenprozess darstellt,^{30,31} kann die RNA-Synthese durch drei Phasen charakterisiert werden: (i) Die Konzentration der RNA-Moleküle ist sehr viel geringer als jene der Replikase und wir beobachten exponentielles Anwachsen der RNA in der Reaktionsmischung, $x(t) = x(0) e^{kt}$, (ii) die RNA-Moleküle liegen gegenüber den Enzymmolekülen im Überschuss vor, alle Proteinmoleküle sind in die RNA-Produktion involviert, es gibt fast keine freien Enzymmoleküle in der Lösung, die RNA-Syntheserate ist daher konstant und die RNA-Konzentration wächst linear mit der Zeit: $x(t) = x(0) + k't$, (iii) der RNA-Überschuss ist so groß, dass die gebildete RNA am Enzym gebunden bleibt und keine weitere RNA an diesem

Proteinmolekül synthetisiert werden kann, was zur Folge hat, dass die RNA-Synthese durch Produkthemmung zum Stillstand kommt. Für die RNA-Evolution kommt in erster Linie die exponentielle Wachstumsphase mit einem Überschuss an Enzymmolekülen in Frage und dementsprechend sind die Bedingungen der seriellen Überimpfexperimente so zu wählen, dass sich die RNA im Unterschuss befindet. Ein Vergleich mit der Vermehrung von Zellen oder Organismen ist illustrativ: Die komplexen Einheiten machen es sehr schwer die Fitness aus den Eigenschaften abzuleiten und es ist im allgemeinen nur möglich Fitness *a posteriori* – nach Bekanntwerden des Selektionsergebnisses – zu interpretieren, bei den evolvierenden Molekülen wird hingegen Fitness als Funktion von Reaktions- und Gleichgewichtsparametern darstellbar und kann daher auch unabhängig vom Evolutionsexperiment bestimmt werden. Bei Zellen und Organismen ist der Vermehrungsvorgang, $M+X \rightarrow 2X$, einfach durch Teilung oder Fertilisation und Entwicklung vorgegeben, wogegen die Vermehrung der RNA-Moleküle in dem genannten Replikationsassay einem komplexen Mehrstufenmechanismus folgt, der nur unter geeignet gewählten und eingestellten Bedingungen einer einfachen Autokatalyse folgt.

Einfachheit und Universalität des Selektionsprinzips lässt sich am besten mit ein wenig Mathematik illustrieren. Wir brauchen dazu nur den Mechanismus der einfachen Vermehrung (2) auf mehrere Varianten auszudehnen und die einzelne Variante durch einen Index zu spezifizieren: X_1, X_2, \dots, X_n . Für das Eintreten von Selektion ist es noch notwendig, dass die Varianten um eine gemeinsame Ressource – hier ist es das Synthesematerial M – konkurrieren. Die chemischen Reaktionsgleichungen lauten dementsprechend:



und die einzelnen Varianten unterscheiden sich durch die Zahl ihrer Nachkommen gemessen als Fitnesswerte, f_1, f_2, \dots, f_n . In diesen Fitnesswerten ist bereits die Abhängigkeit der Zahl der Nachkommen von den vorhandenen Ressourcen M enthalten, die der Einfachheit halber als konstant angesehen werden: $f_j = k_j [M]$, wobei k_j den tatsächlichen Reaktionsparameter des Vermehrungsprozesses und $[M]$ die Konzentration der vorhandenen Ressourcen darstellt. Die Evolution einer solchen Population lässt sich durch gewöhnliche Differentialgleichungen beschreiben, welche mittels einfacher Regeln aus den kinetischen Reaktionsgleichungen gebildet werden:

$$\frac{dx_i}{dt} = f_i x_i - x_i \sum_{k=1}^n f_k x_k ; i = 1, 2, \dots, n ; \sum_{k=1}^n x_k = 1. \quad (4)$$

Die Interpretation von Gleichung (4) ist einfach und reflektiert klar den gegebenen Sachverhalt: Der sogenannte Differentialquotient auf der linken Seite der Gleichung, dx_i/dt , bringt zum Ausdruck, wie sich die Variable $x_i(t)$ im – unendlich klein gedachten – Zeitpunkt t verändert: Ist

er negativ, $dx_i/dt < 0$, dann wird $x_i(t)$ kleiner, ist er positiv, $dx_i/dt > 0$, dann wächst $x_i(t)$. Auf der rechten Seite der Gleichung stehen zwei Terme, der erste bringt das Wachstum von $x_i(t)$ durch Vermehrung zum Ausdruck: Die Zunahme ist umso größer je mehr Individuen der chemischen Spezies X_i bereits vorhanden sind und sie ist proportional zur Fitness f_i , welche angibt wie viele Nachkommen – im Mittel – von einem Individuum X_i zu erwarten sind. Der negative Term auf der rechten Seite kompensiert eine etwaige Zunahme der Gesamtpopulation durch die Vermehrung. Dies ist erforderlich, da alle realistischen Ressourcen endlich sind und die Population ohne Beschränkung ins Unendliche anwachsen würde. Mit dem genannten Term bleibt die Größe der Population konstant und wir können ohne Einschränkung annehmen, dass sie den Wert $\sum_k x_k = 1$ annimmt – wir haben es mit *relativen* Konzentrationen zu tun und $x_i \cdot 100$ ist der prozentuelle Anteil der Spezies X_i an der Population. Die Differentialgleichung für die Evolution der Population kann ohne großen Aufwand gelöst werden und man erhält:

$$x_i(t) = \frac{x_i(0) e^{f_i t}}{\sum_{k=1}^n x_k(0) e^{f_k t}}; \quad i=1, 2, \dots, n, \quad (5)$$

wobei $x_i(0)$ und $x_k(0)$ die Mengen an X_i oder X_k zum Zeitpunkt $t=0$ bedeuten. Um das Ergebnis zu interpretieren betrachten wir die Langzeitlösung, $\lim t \rightarrow \infty$: Der Nenner in Gleichung (5) besteht aus einer Summe von exponentiell wachsenden Termen. Wenn man lange genug wartet, wird der Summand mit dem größten Exponenten $f_m - f_m = \max\{f_k; k=1, 2, \dots, n\}$ – unbeschadet der Anfangswerte $x_i(0)$ alle anderen Terme übertreffen. Lassen wir die kleineren Summanden alle weg, was im Grenzwert $\lim t \rightarrow \infty$ gerechtfertigt ist, so werden Zähler und Nenner gleich und wir erhalten $\lim_{t \rightarrow \infty} x_m(t) = 1$ oder mit anderen Worten, die Spezies mit der höchsten Fitness, X_m , wird selektiert. Zwei weitere Resultate, die leicht überprüft werden können, illustrieren die Natur des Selektionsprozesses: Erstens, die differentielle Fitness ausgedrückt durch die Gleichung

$$\frac{1}{x_i(t)} \frac{dx_i}{dt} = f_i - \sum_{k=1}^n f_k x_k(t) = f_i - \bar{f}(t) \quad (6)$$

legt fest, welche Spezies X_i im nächsten Zeitabschnitt einen Konzentrationszuwachs zu verzeichnen haben werden – es sind dies jene, deren Fitness über dem Populationsmittel liegt, $f_i > \bar{f}(t)$ – und welche an Menge abnehmen werden – dies sind jene, die $f_i < \bar{f}(t)$ erfüllen. Das zweite Resultat betrifft die zeitliche Abhängigkeit der mittleren Fitness

$$\frac{d\bar{f}}{dt} = \sum_{k=1}^n f_k^2 - \left(\sum_{k=1}^n f_k\right)^2 = \text{var}(f_k) \geq 0. \quad (7)$$

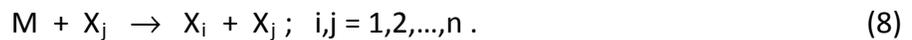
Dieser Ausdruck ist den Statistikern gut bekannt, wird als *Varianz* der Fitnesswerte bezeichnet und ist eine nicht negative Größe *per definitionem*. Die mittlere Fitness der Population kann daher nur wachsen oder konstant bleiben. Eine verschwindende Variation, $\text{var}(f_k)=0$, bedeutet, dass die Population einheitlich ist, das heißt, dass sie nur ein einzige chemische Spezies enthält. Im vorliegenden Fall ist diese Spezies die Variante mit der höchsten Fitness, X_m , und wenn die Population nur mehr diese Spezies enthält, ist Selektion eingetreten. Ein Illustrationsbeispiel für Selektion ist in Abbildung 3 gezeigt: Je nach dem Fitnesswert, kann die Konzentration einer Spezies monoton zunehmen, abnehmen oder durch ein Maximum gehen.

Zur Illustration der Arbeitsweise des Selektionsprinzips kann man sich einen Wettbewerb im Hochspringen vorstellen. Zu Beginn wird die Latte niedrig gelegt und fast alle Athleten können sie überspringen, dann wird das für den Verbleib im Wettbewerb zu überspringende Niveau Schritt für Schritt erhöht und alle Bewerber scheiden erfolglos aus bis schließlich nur mehr der beste Springer überbleibt. Die natürliche Selektion legt die Latte automatisch auf den Mittelwert der Sprunghöhen.

Replikation und Mutation

Variation kommt durch Mutation oder Rekombination zustande, in abstrakter Form war der Mechanismus der Rekombination durch die Mendelschen Vererbungsregeln bekannt aber hinsichtlich der Mutation tappte man bis zu den Anfängen der Molekularbiologie völlig im Dunkel und man behandelte sie daher auch wie einen ‚*Deus ex machina*‘ (Abbildung 3). Als der Vorschlag für die Struktur der DNA von Watson und Crick auf dem Tisch lag, konnte man sogleich nicht nur eine Möglichkeit für die biologische DNA-Synthese sehen sondern auch einen Mechanismus für das Auftreten von Mutanten erraten (Abbildung 1): Mutationen werden als Kopierfehler interpretiert. Dabei kann der Fehler unmittelbar während des Kopiervorganges auftreten wie in der Abbildung angedeutet aber es kann auch ein Schaden an der DNA unabhängig vom Kopiervorgang eintreten, welcher zur Konsequenz hat, dass beim Kopieren eine falscher Nukleotidbuchstabe eingesetzt werden, falls das zelluläre Reparatursystem den Fehler noch nicht ausgebessert hat. Andere relativ häufig vorkommende Fehler sind Auslassungen – *Deletionen* -- und Verdopplungen von Sequenzteilen – *Insertionen*. War man bis zu den Anfängen der Molekularbiologie auf relativ brutale und unspezifische Techniken wie radioaktive Strahlung oder Modifikation mit mutagenen Chemikalien angewiesen, so kann man heute präzise positionsspezifische Änderungen an den DNA-Sequenzen anbringen, welche ortsspezifische Mutagenese – ‚*site-directed mutagenesis*‘ -- genannt werden. Erzeugung und Analyse von gezielten Mutanten ist heute eine Standardmethode und wird zum Beispiel bei systematischen Strukturuntersuchungen an Biomolekülen eingesetzt.

Die Experimentalarbeiten zur zellfreien Evolution von Spiegelman wurden komplementiert durch die etwa gleichzeitige Entwicklung einer kinetischen Theorie der Evolution von Molekülen durch Manfred Eigen und Mitarbeiter.^{21,32,33} Replikation und Mutation werden als parallele chemische Reaktionen modelliert:



Die korrekte Replikation entspricht dem bereits diskutierten Spezialfall mit $i = j$. Die der kinetischen Gleichung (4) entsprechende Differentialgleichung kann gelöst werden, wenn die Lösung auch nicht so einfach ist wie im mutationsfreien Fall. Hier gehen wir nur auf die Ergebnisse im Vergleich zur reinen Selektion ein. Im Replikations-Mutations-System tritt ebenfalls Selektion ein nur das selektierte Objekt ist keine einzelne Spezies sondern eine – oder mehrere – Spezies mit ihren Mutanten. Am besten stellt man sich einen *Clan* in der ursprünglichen Bedeutung als eine schottische Familie bestehend aus dem Familienoberhaupt und allen Familienmitgliedern. Um zum Ausdruck zu bringen, dass eine solche Verteilung von Mutanten das genetische Reservoir einer sich asexuell vermehrenden Art darstellt wurde der Ausdruck *Quasispezies* gewählt.³⁴ Eine Quasispezies ist die stationäre Lösung oder Langzeitlösung der kinetischen Gleichung für Replikation und Mutation. Im allgemeinen Fall^j ist eine Quasispezies um eine Sequenz höchster Fitness – um die Mastersequenz – herum gruppiert, das heißt, sie entspricht wie angedeutet einem Cluster bestehend aus einer fittesten Sequenz umgeben von einer *Mutantenwolke* im Raum der Sequenzen oder Genotypen. Einige Eigenschaften von Quasispezies können direkt durch Kombination der mathematischen Analyse mit empirischen Befunden hergeleitet werden. Zur Illustration geben wir ein einfaches Beispiel an: Wir nehmen an, dass alle Mutationspfade, die aus einer Sequenz von Einzel- oder Punktmutationen bestehen, möglich seien, und daher kann bei konstanter Sequenz- oder Kettenlänge l jede Sequenz von jeder Sequenz durch eine endliche Zahl von Punktmutationen erreicht werden. Dann ergibt die mathematische Analyse, dass alle Sequenzen in der Quasispezies vorkommen und dass ihre stationäre Konzentration durch den Hamming-Abstand $d_H(X_i, X_m)$ von und der Fitnessdifferenz zur Mastersequenz, $f_m - f_i$, bestimmt werden. Häufige

^j Allgemein soll hier unter anderem Abwesenheit von *Neutralität* bedeuten. Unter Neutralität versteht man den Fall, dass zwei oder mehrere Sequenzen dieselben Fitnesswerte besitzen, und daher bedeutet Abwesenheit von Neutralität, dass alle Sequenzen unterschiedliche Fitness aufweisen. Neutralität spielt in der Realität – sowohl bei *in vitro* Evolutionsexperimenten als auch in der Natur – eine wichtige Rolle und sie kann unter anderem auch in die Quasispeziesstheorie problemlos eingebaut werden. Eine erschöpfende Behandlung würde aber den für dieses Kapitel vorgegebenen Rahmen sprengen.

Mutanten sind daher jene, die nur wenige Mutationsschritte von der Mastersequenz entfernt sind und eine geringe Fitnessdifferenz zur Mastersequenz aufweisen.^k

Eine wichtige Größe für die Analyse von Populationen aus replizierenden und mutierenden Individuen ist die mittlere Mutationsrate, p : Sie gibt an, wie groß die Wahrscheinlichkeit ist, dass während des Kopiervorganges ein Fehler beim Einbau eines Nukleotidbausteins auftritt. Für eine Kettenlänge l gelten daher folgende Wahrscheinlichkeiten:

$$(1-p)^l \quad \text{für die vollkommen korrekte Replikation } X_m \rightarrow 2X_m \text{ oder } X_i \rightarrow 2X_i, \text{ und} \quad (9a)$$

$$(1-p)^{l-d_H(X_i, X_m)} p^{d_H(X_i, X_m)} \quad \text{für die Mutation } X_m \rightarrow X_i. \quad (9b)$$

Mit dieser Definition einer Fehlerrate kann man unschwer die Quasispezies als Funktion der Mutationswahrscheinlichkeit untersuchen (Abbildung 3). Beim Übergang zur fehlerfreien Replikation, $\lim p \rightarrow 0$, verschwindet die Mutantenwolke und nur die Mastersequenz bleibt als Sequenz maximaler Fitness über. Mit steigenden Fehlerraten wird der Anteil der Mastersequenz an der Population erwartungsgemäß immer kleiner und gleichzeitig nehmen die Fehlerklassen gemäß ihrem Hammingabstand an Bedeutung zu: zuerst die Mutanten mit einem Fehler, bei höheren Fehlerraten jene mit zwei, dann jene mit drei Fehlern, usw. Interessanterweise gibt es eine kritische Fehlerrate, p_{cr} , bei welcher sich die Natur der Langzeitmutantenverteilung spontan ändert. Aus der stationären Verteilung, der Quasispezies, werden Cluster von Sequenzen, welcher im Sinne eines Irrflugs durch den Sequenzraum driften. Dabei werden die individuellen chemischen Spezies andauernd durch andere abgelöst. Mit anderen Worten es existiert gar keine stationäre Verteilung und Vererbung im herkömmlichen Sinn ist nicht möglich. Die Grenze zwischen Stationarität und Zufallsdrift wird als *Fehlerschranke* bezeichnet. Sie kann näherungsweise leicht berechnet werden:

$$p_{cr} \cong \frac{\ln \sigma_m}{l} \quad \text{oder} \quad l_{max} \cong \frac{\ln \sigma_m}{p} \quad \text{mit} \quad \sigma_m = \frac{f_m}{\bar{f}_{-m}} \quad (10)$$

Die *Superiorität* der Mastersequenz σ_m ist gegeben durch den Quotienten zwischen der Fitness der Mastersequenz und der mittleren Fitness der restlichen Population. Die erste Gleichung

^k Dieser Befund wurde aus kinetischen Differentialgleichungen hergeleitet und gilt dementsprechend für den Grenzwert großer Populationen, $\lim N \rightarrow \infty$. Eine große Population ist eine, in welcher die Zahl der Individuen, N , die Zahl der möglichen Sequenzen, n , bei weitem übertrifft. Für Nukleinsäuren gilt $n=4^l$, wenn l die Sequenz- oder Kettenlänge darstellt. Da 4^l schon bei sehr moderaten Kettenlängen wie $l=100$ eine unvorstellbar große Zahl ist, $4^{100}=1.6 \times 10^{60}$, muss man in der Realität die Quasispezies auf die unmittelbare Umgebung der Mastersequenz im Sequenzraum beschränken, da sonst Konzentrationen von Bruchteilen von Molekülen im Reaktionsvolumen auftreten.

bringt zum Ausdruck, dass die maximale tolerierbare Fehlerrate bei konstanter Sequenzlänge l mit der Sequenzlänge abnimmt oder mit anderen Worten für jede vorgegebene Länge gibt es eine definierte Fehlerrate, die nicht überschritten werden kann. Andererseits definiert die Fehlerschranke eine maximale Kettenlänge der Nukleinsäure für eine vorgegebene Replikationsgenauigkeit. Sollen längeren Moleküle repliziert werden, so benötigt man eine genauer arbeitende Replikationsmaschinerie.

Das hier besprochene Replikations-Mutations-System kann in gleicher Weise als ein denkbar einfachstes, evolutionsbefähigtes biologisches Modell oder als ein Netzwerk chemischer Reaktionen verstanden werden, welches ungeachtet seiner relativen Komplexität sowohl einer experimentellen Implementierung als auch der mathematischen Analyse zugänglich ist. Es ist daher naheliegend an genau diesem System herauszuarbeiten, was die biologische Sichtweise gegenüber Physik und Chemie auszeichnet. Wir fokussieren hier auf den Informationsbegriff wie er üblicherweise in der Biologie verwendet wird und wie er auch der Informatik zugrunde liegt: Information ist genetische Information – oder in menschlichen Gesellschaften genetische Information vermehrt durch kulturelle Information – und sie ist in digitaler Form codierbar, das heißt sie kann in digitaler verschlüsselter Form weitergereicht und anschließend decodiert werden. Digitalisierung bedeutet hier nicht nur das Verwandeln in Symbole oder *Buchstaben* sondern auch problemlos nutzbare Digitalisierung, welche beim Lesen auch so etwas wie näherungsweise Gleichwertigkeit der Buchstaben erfordert.¹ Die chemisch-zwischenmolekulare Wechselwirkung zwischen G und C ist etwa zehnmal stärker als jene zwischen A und T – oder U. Dessen ungeachtet sind GC und AT in der Geometrie der Doppelhelix und in ihrer biochemischen Funktion gleichberechtigt und kommen in der DNA etwa gleich häufig vor, und dies obwohl die Nukleobasenpaare unterschiedlich stark wechselwirken. Die Replikation muss zügig erfolgen können und darf nicht etwa bei den stärkeren Paaren oder bei Clustern stärkerer Paare stehen bleiben. Wäre eine solche näherungsweise Gleichwertigkeit nicht gegeben, würde die Evolution unweigerlich zu einer andauernden Verschiebung des AT/GC Verhältnisses in eine Richtung führen. Ähnliches gilt für die Translation, bei der die einzelnen Tripletts oder *Codons* trotz unterschiedlicher Wechselwirkungsstärke mit vergleichbarer Geschwindigkeit verarbeitet werden müssen. Der für die biologische Information notwendige Übergang von den durch Gleichgewichtskonstanten oder anderen thermodynamisch-kinetischen Parameter abgestuften

¹ Es bedarf einer Erklärung, was unter *näherungsweise Gleichwertigkeit* verstanden werden soll. Denken wir beispielsweise an den Buchstaben Q in unserer lateinischen Alphabet. Er wird nicht wirklich so wie die anderen Buchstaben verwendet, Q kommt fast nur vergemeinschaftet mit dem u als Digraph qu vor. Ausnahmen gibt es fast nur bei Eigennamen und in der Transkription von chinesischen Schriftzeichen. Jedoch bei Aufzählungen a,b,c,... oder als mathematisches Symbol ist q eigenständig, und bei der Ermittlung der Buchstabenzahl wird q wie ein normaler Buchstabe gezählt.

und nach Stärken geordneten Wechselwirkungen zu einem näherungsweise gleichwertigen digitalen System entspricht einem Übergang von der Chemie zur molekularen Genetik.^m

Ein genetisches System kann Information digital speichern aber wie kommt es zu Beginn zur Informationserzeugung? Um diesen nicht trivial erscheinenden Vorgang zu beleuchten, betrachten wir eine Quasispezies, die sich in einer Art Gleichgewicht mit ihrer Umgebung befindet. Nehmen wir an, es kommt zu einer Änderung der Umgebung, welche ihrerseits zu einer Änderung der Fitnesswerte Anlass gibt. Das System kommt wieder ins Gleichgewicht und eine neue Quasispezies wird gebildet, welche sich von ihrem Vorgänger unterscheidet. Die Veränderungen in der Mutantenverteilung bilden letztlich die Veränderungen in der Umwelt ab und in diesem Sinne kann man sagen, dass im Raum der Konzentrationen und Sequenzen Information über die Änderung der Umgebung entstanden ist. Für die Einzelsequenz bedeutet eine Verbesserung der Information über Umgebung eine Erhöhung der Fitness.

Zusammenfassend können wir das Replikations-Mutations-System dadurch charakterisieren, dass die einzelnen Spezies des reinen Selektionssystems durch Cluster von Genotypen ersetzt werden. Diese Cluster sind umso breiter je höher die Mutationsrate ist. Ähnlich dem *stille Post*-Spiel von Kindern gibt es eine höchste Fehlerrate, oberhalb derer keine Information stabil durch der Nachrichtenweitergabe transferiert werden kann. Ein weiterer Unterschied zwischen den beiden Evolutionssystemen besteht darin, dass bei von null verschiedenen Mutationsraten die durch Gleichung (7) ausgedrückte evolutionäre Optimierung ihre Allgemeingültigkeit verliert. Es gibt Anfangsbedingungen, von denen aus die mittlere Fitness nicht zunimmt. Dieser Sachverhalt ist ganz einfach vorstellbar: Wählen wir zu Anfang eine einheitliche Population, die nur aus Mastersequenzen besteht, so kann die mittlere Fitness durch die Erzeugung von Mutanten nur abnehmen. Aus dem mathematisch rigorosen Optimierungstheorem ausgedrückt durch (7) wird eine zweifellos sehr nützliche Optimierungsheuristik die in sehr vielen aber eben nicht in allen Fällen gültig ist.

Von der Theorie zur Anwendung

Darwin widmete sich in seiner ‚*Origin of Species*‘ sehr ausführlich den Ergebnissen der Tier- und Pflanzenzüchtung, da sie die Formenvielfalt demonstrieren, welche sich durch Evolution mit künstlicher Selektion erzeugen lässt. Die bislang beschriebene Vorgangsweise bei der Evolution

^m Man könnte hier einwenden, dass auch in einem binären 0,1- System wie in Computern perfekt codiert werden kann. Dem angesprochenen Problem der chemischen Unterschiede entgeht man aber dennoch nicht, denn bei identischen Paarungen kommt die Verschiedenheit durch die unterschiedlichen Paarwechselwirkungen, 00, 01, 10 und 11, zustande, die durch einen strukturchemischen Trick ausgeräumt werden müssen.

in vitro entsprechen der natürlichen Auslese, da es während des gesamten Experimentes gab keine direkte Intervention des Experimentators im Evolutionsverlauf gab. Es liegt nahe zu fragen, ob man nicht auch entsprechend einer künstlichen Auslese *Molekülzüchtung* betreiben könnte. Mit anderen Worten es soll versucht werden, die Evolution im Reagenzglas zu lenken und für technische Zwecke auszunutzen, um maßgeschneiderte Moleküle für vorgegebene Aufgaben zu erzeugen. Die Schwierigkeit, welcher man dabei begegnet ist der Unterschied zwischen Pflanzensamen und Jungtieren einerseits und Molekülen andererseits: Samen und Tiere kann man von Hand aus verlesen, für Moleküle wird jedoch eine trickreiche physikalische oder chemische Methode benötigt, um Moleküle mit erwünschten Eigenschaften von jenen zu trennen, welche diese nicht oder in nicht ausreichendem Maße aufweisen. Die evolutionäre Biotechnologie baut auf denselben drei Voraussetzungen wie Darwins Theorie auf: Vermehrung mit Vererbung, Variation und Selektion. Vermehrung und Variation – zumeist als Amplifikation und Diversifikation bezeichnet – sind mit dem Repertoire der heute in jedem molekularbiologischen Labor zur Verfügung stehenden Methoden leicht zu realisieren: Als allgemeine Amplifikationsmethode steht PCR auf DNA-Ebene zur Verfügung oder man verwendet spezielle Reaktionen zur direkten Vervielfältigung von RNA. Einen hinreichend diversen genetischen Pool an Sequenzen kann man entweder durch Replikation mit kontrollierter Mutationsrate oder durch *Zufallssynthese* erzeugen.ⁿ Die Selektion wird vom Experimentator gelenkt und erfordert, wie gesagt, Intuition und experimentelles Geschick.

Der Ablauf der Erzeugung von maßgeschneiderten Molekülen durch evolutionäre Strategien ist in Abbildung 4 skizziert. Am Beginn steht die Erzeugung einer ausreichenden genetischen Vielfalt durch Replikation mit erhöhter Mutationsrate oder Zufallssynthese. Der zweite Schritt besteht in der Anwendung gelenkter Selektion, welche die für den vorgegebenen Zweck am besten geeigneten Moleküle auswählt und anschließend entscheidet ein Test, ob das gewünschte Ergebnis bereits erzielt wurde. Falls dies nicht der Fall ist, werden die ausgewählten Moleküle einem weiteren Selektionszyklus bestehend aus Amplifikation, Diversifikation und Selektion unterworfen, um weitere Verbesserungen zu erreichen. Die Selektionszyklen werden weiter fortgesetzt bis entweder das gewünschte Ergebnis erhalten wurde oder keine weiteren Verbesserungen mehr auftreten. Im Allgemeinen bewegt sich die Zahl der notwendigen Selektionszyklen – Spezialfälle ausgenommen – um 30 oder 40 herum.

Eine bahnbrechende Entdeckung wurde in 1980er Jahren durch Thomas Cech³⁵³⁶ und Sidney Altman³⁷ an natürlichen RNA-Molekülen gemacht: RNA kann ebenso wie Protein Reaktionen

ⁿ Zufallssynthese kann man in den meisten Syntheseautomaten bewerkstelligen, indem man keine Vorlage – Template – zugibt. Es werden dann die Bausteine in beliebiger Reihenfolge an die wachsende Polynukleotidkette angehängt. Bei hinreichend langen Molekülen sind alle erhaltenen Einzelsequenzen verschieden. Für eine Probe von 10^{15} Molekülen liegt diese Länge bei $n \approx 25$.

mit höchster Spezifität katalysieren. In der Natur handelt es sich dabei stets um Reaktionen, bei welchen andere RNA-Moleküle prozessiert werden. Ein Ribozym, welches wegen seiner typischen Struktur ‚*Hammerhead*‘ Ribozym (Abbildung 5) bezeichnet wird, wurde in kleinen Pflanzenpathogenen – sogenannten *Viroiden* – entdeckt und erwies sich letztendlich als ein universelles Element in vielen Genomen.³⁸ Das Hammerhead-Ribozym spaltet ein anderes RNA-Molekül höchst spezifisch an einer ganz genau definierten Stelle. Die Wirkung des RNA-Katalysators besteht darin, dass das Substrat durch die Bindung an den Katalysator in eine Geometrie gebracht wird, in welcher die Spaltung der Bindung an der betreffenden Stelle spontan als energetisch bevorzugter Prozess erfolgt. Die natürlichen Funktionen von Hammerhead-Ribozymen sind das Prozessieren von replizierten oder transkribierten RNA-Molekülen.³⁹ Besondere Aufmerksamkeit erregten Ribozyme, als sich herausstellte, dass die Proteinsynthesefabriken der Zelle – die Ribosomen – auch durch ihre RNA-Moleküle katalytisch wirken und dies wird durch den Satz, „*The ribosome is a ribozyme!*“^{40,41}, perfekt ausgedrückt. Ein typisches Merkmal der Katalyse durch Ribozyme ist hohe Spezifität aber im Vergleich zu Proteinen geringere Beschleunigungen der Reaktionen.

Eine weitere wichtige Entdeckung einer Funktion von RNA-Molekülen folgte um die Jahrhundertwende: RNA-Moleküle mit zwei – oder mehreren – hinreichend langlebigen Konformationen^o – sogenannte *Riboswitches* – regeln die zelluläre Synthese einer Reihe von Proteinen die im Metabolismus eine Rolle spielen.^{42,43} Die Tatsache, dass RNA-Sequenzen zwei oder mehrere langlebige Konformationen ausbilden können, wurde schon vor der Entdeckung von Riboswitches in der Natur auf der Grundlage von Strukturberechnungen theoretisch avorausgesagt.^{44,45,46} Sequenzen von RNA-Molekülen, welche eine, zwei oder mehrere langlebige Konformationen ausbilden, können mit Hilfe von Computeralgorithmen designt werden. Zufallssequenzen bilden zumeist eine Vielzahl von größtenteils kurzlebigen Konformationen aus und müssen in ihren Sequenzen modifiziert werden, um Moleküle mit einer dominanten Konformation oder einigen wenigen dominanten Konformationen zu erhalten. Bei natürlichen Molekülen mit definierten dominanten Konformationen wurde die Sequenzen durch die biologische Evolution designt.

Am Beginn der Entwicklung einer evolutionären Biotechnologie stehen die Spiegelmanschen Experimente und der Versuch, die Evolution *in vitro* in vorgegebene Bahnen zu lenken, war naheliegend. Die ersten Experimente gelenkter Evolution hatten die Umprogrammierung von

^o Eine Konformation eines Moleküls entspricht in etwa einem Minimum der freien Energie. Biomoleküle wie Proteine oder Nukleinsäuren haben im allgemeinen eine Vielzahl von Konformationen aber die meisten davon sind kurzlebig, das heißt, sie wandeln sich zu rasch in andere Konformationen um, als dass sie eine wesentliche Rolle spielen können. Die meisten Moleküle haben nur einen langlebigen Zustand, welcher dem globalen Minimum der freien Energie entspricht.

Ribozymen zum Inhalt: Gerald Joyce gelang es spezifische, nur RNA spaltende Ribozyme zu spezifisch nur DNA spaltenden Molekülen zu evolvieren.⁴⁷ In der Sprache der biochemischen Katalyse kann man sagen, die Substratspezifität eines Katalysators wurde durch Evolution *in vitro* verändert. In der Folge wurde eine Fülle von Ribozymen durch gelenkte Evolution hergestellt, welche das katalytische Repertoire der natürlichen Ribozyme gewaltig erweiterten und zeigten, dass die katalytischen Fähigkeiten von Ribozymen keineswegs auf die Prozessierung von RNA und DNA beschränkt sind. Ein eindrucksvolles Beispiel eines artifiziellen Ribozyms, welches die in der Natur nur sehr selten vorkommenden Diels-Alder Reaktion effizient katalysiert.⁴⁸ Mittlerweile wurden auch Ribozyme entwickelt, welche therapeutische Verwendung finden.⁴⁹ Die Erzeugung von maßgeschneiderten Molekülen durch gelenkte Evolution ist keineswegs auf Nucleinsäuren – RNA oder DNA – beschränkt. Beispiele für evolutionäres Proteindesign finden sich in einem Sammelband,⁵⁰ in einem Spezialband einer Zeitschrift⁵¹ und in einem Übersichtsartikel (siehe auch den übernächsten Abschnitt).⁵² Interessanterweise wurden auch Versuche unternommen, das evolutionäre Design auf kleine organisch-chemische Moleküle auszudehnen.⁵³

Ein Beispiel zur Illustration der Vorgangsweise bei der Selektion zur *Züchtung* von RNA-Molekülen, welche für die Bindung an vorgegebene Zielstrukturen – sogenannte ‚Targets‘ – optimiert werden sollen, sei hier beschrieben.^{54,55} Das Verfahren, SELEX – **S**ystematic **E**volution of **L**igands by **EX**ponential **E**nrichment – genannt, ist in Abbildung 6 skizziert. Zur Anreicherung der zu selektierenden Moleküle verwendet man eine chromatographische Säule, auf welcher die Zielmoleküle durch eine chemische Bindung an die stationäre Phase angebunden sind. Eine Lösung mit den RNA-Molekülen wird über die Säule fließen gelassen, die an das Target hinreichend stark bindenden Moleküle bleiben hängen und werden anschließend mit einem anderen Lösungsmittel eluiert und durch Amplifikation, Diversifikation und erneuter Trennung auf der Säule in verbesserter Form isoliert und auf Verbesserung getestet. Mit dem SELEX-Verfahren gelingt es, RNA-Aptamere – so werden Moleküle bezeichnet, die optimal an vorgegebene Targets binden—zu erzeugen, welche Bindungsfestigkeiten aufweisen, die an die stärksten in der Natur bekannten Bindungskonstanten herankommen. Für weitere Einzelheiten sei auf einen Sammelband über Aptamere verwiesen.⁵⁶

Evolutionäres Proteindesign wurde unter anderem eingesetzt, um Enzyme an nicht natürliche Bedingungen anzupassen.⁵⁷ Unter anderem waren zwei Aufgaben zu lösen: (i) Änderung der Thermostabilität und anderer thermodynamischer Eigenschaften⁵⁸, und (ii) Erzielen von hinreichender Löslichkeit und enzymatischer Aktivität in nichtwässrigen Lösungsmitteln.^{59,60,61} Die Natur selbst bietet eindrucksvolle Beispiele unterschiedlicher Temperaturanpassungen von Bakterien und Archebakterien im Temperaturbereich etwa -15°C bis zu 120°C, die nach ihrer optimalen Wachstumstemperatur in Psychro- oder Cryophile (<15°C), Mesophile (15-50°C),

Thermophile (50-80°C) und Hyperthermophile (>80°C) eingeteilt werden. Durch evolutionäres Design gelang es sowohl die thermodynamische Stabilität von Proteinen entscheidend zu erhöhen als auch das Temperaturoptimum der katalytischen Aktivität im oben genannten Temperaturbereich nach Belieben zu verschieben.

Katalytische Antikörper sind eine besonders interessante Variante des Proteindesigns, da sie *de novo* erzeugten Proteinkatalysatoren entsprechen.^{62,63,64} Ein häufig verwendetes Konzept zur Erzeugung von katalytischen Antikörpern ist relativ einfach und verläuft über stabile Analoga von Übergangszuständen:^{p,65} Zuerst wird ein dem Übergangszustand analoges *Hapten* synthetisiert – so bezeichnet man den niedrigmolekularen Bestandteil eines Antigens – und dann wird es mit einem Träger zu einem *Immunogen* verbunden. Gegen das Immunogen werden monoklonale Antikörper in großer Zahl gezüchtet und aus dieser Menge werden durch ein spezielles Selektionsverfahren – *catELISA* genannt⁶⁶ – die Proteine mit der stärksten katalytischen Wirkung selektiert. Man kann die Erzeugung der katalytischen Antikörper auch als eine Parallele zum SELEX-Verfahren sehen, wo bei hier die Produktion der den Aptameren entsprechenden Moleküle dem Immunsystem überlassen wird. Katalytische Antikörper wurden für viele chemische Reaktionen gezüchtet, besonders spektakulär ist dabei – ebenso wie im Fall der Ribozyme – die gelungene Herstellung von Katalysatoren für die Diels-Alder Reaktion.

Abschließend erwähnen wir noch eine neue Strategien der Virusbekämpfung,⁶⁷ welche aus der Theorie der Replikations-Mutationskinetik folgt. Viren sind einem sehr starken Selektionsdruck durch die Abwehrmechanismen ihrer Wirtsorganismen ausgesetzt, welchem durch eine möglichst große Variabilität hervorgerufen durch eine hohe Mutationsrate begegnet wird.⁶⁸ Wie bereits dargelegt, gibt es für die Mutationsraten eine obere Grenze, welche durch die Fehlerschranke vorgegeben ist. In der Tat operieren RNA-Viren und andere Pathogene auf RNA-Basis nur knapp unterhalb der genetischen Fehlerschranke und ein naheliegendes Konzept zur Virusbekämpfung geht davon aus, dass eine künstliche Erhöhung der Mutationsrate eine Elimination der Viruspopulation auslösen wird. Dieses Prinzip der *letalen Mutagenese* führt zum Aussterben der Viruspopulation auf zwei verschiedenen Wegen, und zwar wird (i) durch ein Überschreiten der Fehlerschranke ein Zusammenbruch der viralen Vererbung ausgelöst oder (ii) durch eine Akkumulation letaler Varianten stirbt die Viruspopulation aus.^{69,70,71} Gegen eine ganze Reihe von tierischen und humanpathogenen Viren konnten auf der Basis der letalen Mutagenese neue Medikamente entwickelt werden.

^p Eine chemische Reaktion verläuft von den Reaktanten über Übergangszustände zu den Produkten. Im Allgemeinen sind Übergangszustände instabil, durch Abwandlung der atomaren Zusammensetzung können aber stabile Strukturen synthetisiert werden, welche die gleiche oder eine sehr ähnliche Geometrie wie der Übergangszustand aufweisen. Diese bezeichnet man als Analoge des Übergangszustandes.

Epilog

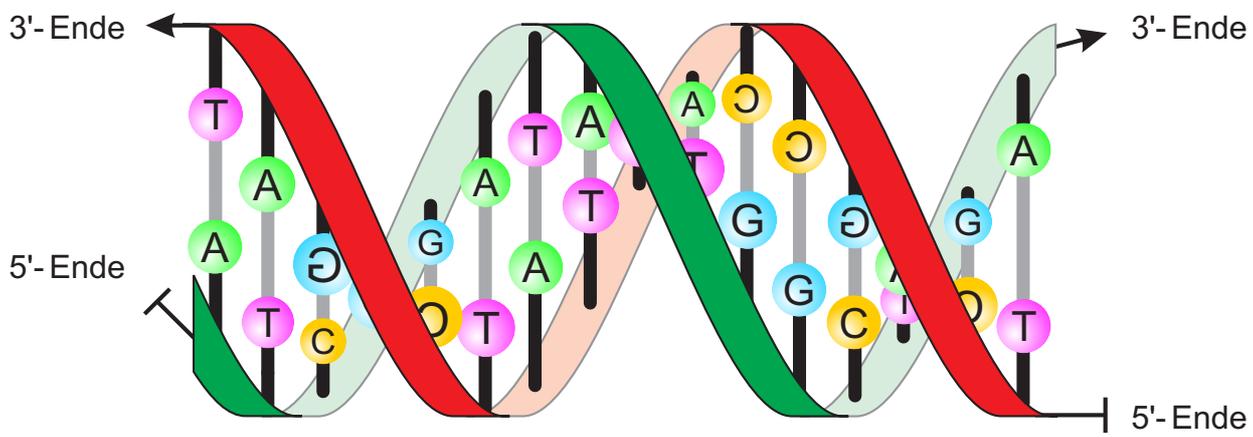
Die Theorie der biologischen Evolution war viele Jahrzehnte lang geprägt von einem heftigen Streit darüber, ob zu Lebzeiten erworbene Eigenschaften vererbt werden können oder nicht. Nicht ganz zutreffend wurden die beiden Konzepte zwei *Vätern* der Evolutionstheorie, Jean-Baptiste de Lamarck und Charles Darwin, zugeschrieben. Mit den heute offenbar werdenden Mechanismen werden diese Auseinandersetzungen obsolet: In der auf der Genetik basierenden Weitergabe von Merkmalen ist für den Lamarck zugeschriebenen Mechanismus kein Platz, aber ein Teil der Eigenschaften von Organismen kommen durch verschiedene epigenetische Prozesse zustande und für diese gibt es durchaus auch die Einflussnahme von Umweltfaktoren auf die Expression von Genen, und wenn sich diese einer Modifikation der DNA im Sinne von Imprinting bedient, gibt es auch eine Weitergabe über ein paar Generationen hinweg. Es ist nicht mehr die Frage, ob es epigenetische Erbfaktoren gibt, offen ist nur mehr ihre relative Bedeutung. Die Mendelschen Regeln sind schließlich auch in guter Näherung gültig.

Auf das mechanistische Minimum reduziert kann Darwinsche Evolution mathematisch analysiert, experimentell implementiert und studiert werden. In dieser Form ist sie keine abstrakte Theorie sondern ein naturwissenschaftliches experimenteller Untersuchung zugängliches Modellsystem, welches in allen Stufen einem Test zugeführt werden kann: Beobachtung – Analyse – Modellierung – Vorhersage – Anwendung. In diesem Sinne wurde Darwinsche Evolution auch zur Problemlösung eingesetzt. Evolutionäres Design von Molekülen ist überall dort angebracht, wo unser Wissen um die biologischen Strukturen und Funktionen noch nicht ausreicht, die Lösungen auf rationalem Weg zu ermitteln. Ist hingegen rationales Design möglich und erfolgreich, dann ist es dem evolutionären Design nicht zuletzt schon aus ökonomischen Gründen überlegen: Rationales Design gestattet es das gewünschte Molekül direkt, das heißt ohne Umwege synthetisieren zu können, wogegen die Evolution einen großen Aufwand an suboptimalen und ungeeigneten Lösungen durchspielen muss. Allerdings hält sich dieser Mehraufwand für ein Gesamtprojekt in Grenzen, da ein einmal in Form einer selektierten Sequenz erhaltenes Molekül auf konventionellem molekularbiologischem Weg direkt vervielfältigt werden kann.

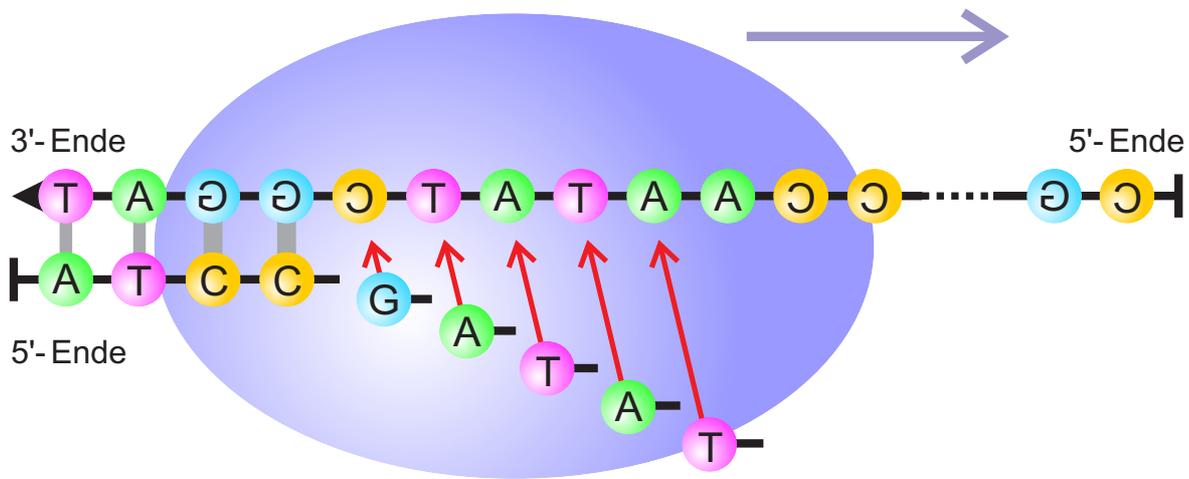
Abbildungsunterschriften:

Abbildung 1: **Genetik und die Struktur der Nukleinsäuren.** Die Struktur der β -DNA als eine von zwei Einzelsträngen gebildete Doppelhelix wurde von Watson und Crick im Jahre 1952 korrekt vorhergesagt.⁹ Der DNA Einzelstrang hat zwei verschiedene Enden, das 5'-Ende und das 3'-Ende. Das obere Bild zeigt die außen liegenden *Rückgrate* der beiden Stränge, welche in entgegengesetzte Richtungen laufen. Die konventionelle Schreibweise der Sequenzen geht vom 5'-Ende (links) zum 3'-Ende (rechts). Zur besseren Kennzeichnung sind die Nukleobasen, Adenin (A), Thymin (T) (ersetzt durch Uracil (U) in RNA), Guanin (G) und Cytosin (C), im gegenläufigen Strang ($3' \leftarrow 5'$) spiegelsymmetrisch geschrieben. Ist ein Strang gegeben, dann kann der zweite eindeutig ergänzt werden. Das mittlere Bild skizziert die Ergänzung des zweiten Stranges während der DNA-Synthese durch eine Einzelstränge komplettierende DNA-Polymerase, beispielsweise die Taq-Polymerase, welche wie alle natürlichen Polymerasen, die Vorlage vom 3' zum 5'-Ende abliest und daher den neuen Strang beginnend am 5'-Ende Buchstabe für Buchstabe aufbaut. Im unteren Bild wird das Entstehen einer Punktmutation durch fehlerhaften Einbau (weiß unterlegt) skizziert.

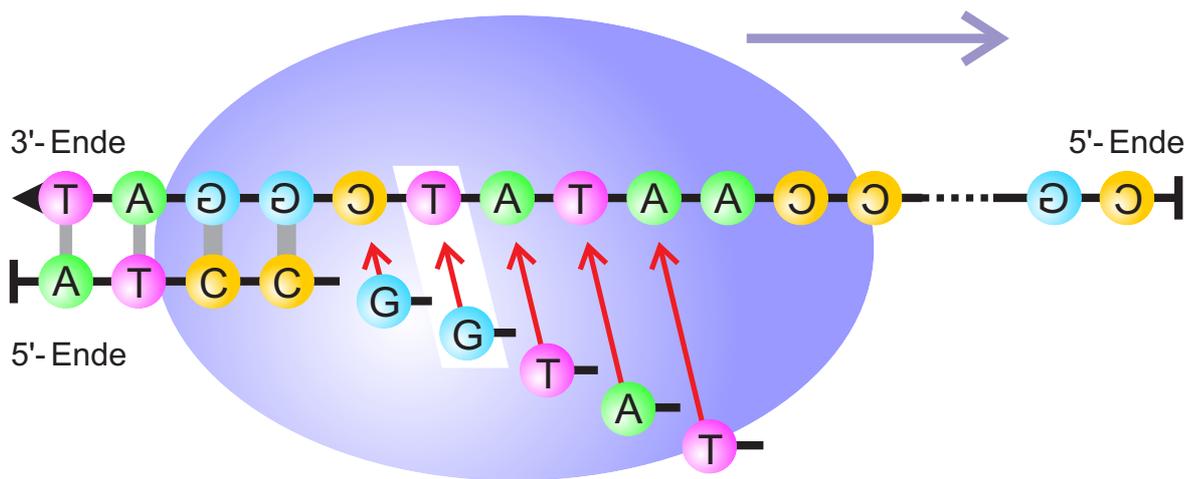
⁹ Die DNA kommt in mehreren Formen vor – die häufigsten sind A-DNA, β -DNA und z-DNA. Die Struktur der DNA in der Zelle ist fast immer die β -Form.



DNA-Doppelhelix



korrekte Replikation



Mutation

- | | | | | | |
|--------|----------|---------|----------|--------|----------|
| Adenin | A | Thymin | T | Uracil | U |
| Guanin | G | Cytosin | C | | |

Abbildung 2: **Der Transfer von Information und das zentrale Dogma in der Molekularbiologie.**

Das zentrale Dogma (linkes Schema) betrifft den Transfer von Sequenzinformation, welche während der Synthese Baustein für Baustein von einem Biopolymermolekül – der Vorlage -- auf ein anderes Molekül – das Syntheseprodukt – übertragen wird. Die genetische Information ist in der Sequenz der Bausteine von DNA- oder RNA-Molekülen gespeichert. Durch die chemische Maschinerie der Zelle (schwarze Pfeile) wird ein Stück DNA entweder repliziert (DNA → DNA) oder in RNA (DNA → RNA) umgeschrieben – transkribiert – und die RNA kann nachfolgend in ein Protein übersetzt werden (RNA → Protein). In der Natur kommen auch andere Informationstransfers vor (rote Pfeile): RNA-Replikation (RNA → RNA) und reverse Transkription (RNA → DNA). Die rechte Seite der Abbildung zeigt die Logik von Transkription und Translation. Bei der Transkription (Bild Mitte) wird ein DNA-Strang (Bild oben; schwarzes Rückgrat) auf RNA (violettes Rückgrat) umgeschrieben. Wie bei der Replikation werden die einzelnen Nukleobasen vom 3'-Ende der DNA-Vorlage her Buchstabe für Buchstabe zu Nukleobasenpaaren ergänzt, wobei T in der DNA dem U in der RNA entspricht. Bei der Translation wird ein Stück RNA Basentriplett für Basetriplett in der Richtung vom 5'-Ende zum 3'-Ende der RNA-Vorlage abgelesen, durch eine transfer-RNA (rot) als Aminosäurerest interpretiert und in die wachsende Polypeptidkette (grün) eingebaut (Die Abkürzungen bedeuten einzelne Aminosäurereste: met = Methionin, ile = Isoleuzin, leu = Leuzin, gly = Glyzin, thr = Threonin, etc.). Beginn und Ende der zu übersetzenden Buchstabensequenz sind durch ein Startcodon (AUG, seltener GUG oder UUG) und ein Stopcodon (UAA, UAG oder UGA; nicht im Bild) gekennzeichnet. Man beachte, dass Replikation, Transkription und reverse Transkription einerseits, und Translation andererseits in verschiedene Richtungen auf den Sequenzen der Nucleinsäuren laufen.

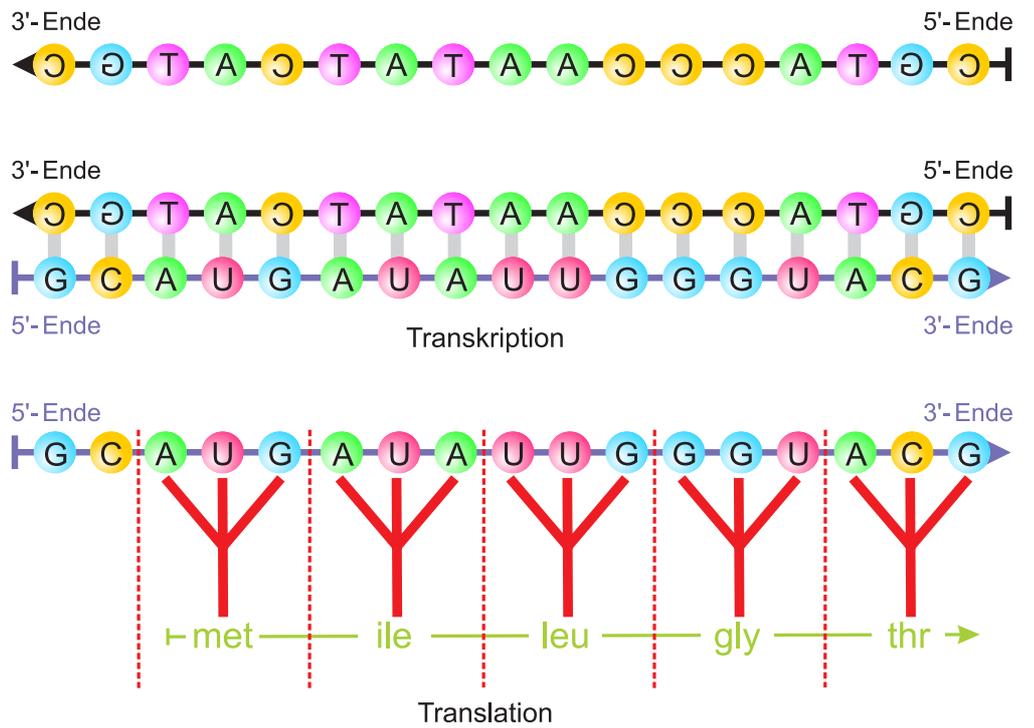
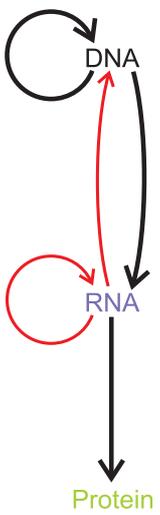


Abbildung 3: **Populationsdynamik bei Evolution durch Replikation und Mutation.** Die obere Abbildung zeigt die Zeitabhängigkeit der relativen Konzentrationen $x_i(t)$ einzelner Spezies X_i in einer durch Replikation evolvierenden Population. Zu Beginn sind drei Varianten X_1 , X_2 und X_3 mit den Fitnesswerten $f_1 = 1$, $f_2 = 2$, $f_3 = 3$ vorhanden, die Anfangswerte wurden so gewählt, dass die zu Beginn fittestete Spezies den geringsten Anteil an der Anfangspopulation hat: $x_1(0) = 0.9$, $x_2(0) = 0.08$ und $x_3(0) = 0.02$. Der Anteil an Spezies X_1 (gelb) in der Population nimmt stetig ab bis sie schließlich ausstirbt, der Fitnesswert der Spezies X_2 (rot) liegt zu Anfang über der mittleren Fitness der Population und daher wächst die Konzentration $x_2(t)$ vorerst geht aber mit steigendem Anteil an der Spezies X_3 (schwarz) mit einem noch höheren Fitnesswert durch ein Maximum um dann abzunehmen, im Zeitraum $0 \leq t < 6$ ist X_3 die fitteste Spezies und ihr Anteil nimmt stetig zu. In diesem Zeitraum strebt die Population in Richtung auf die Selektion von X_3 . Zum Zeitpunkt $t = 6$ wird eine winzig kleine Menge einer noch fitteren Spezies X_4 (blau, $f_4 = 7$) in die Population eingebracht: $x_4 = 0$ für $0 \leq t < 6$ und $x_4(6) = 0.0001$. Entsprechend dem Darwinschen Prinzip wächst die Konzentration von X_4 und dadurch wird schließlich auch Spezies X_3 verdrängt bis sie schließlich ausstirbt. Das Auftreten der Mutation ist nicht Bestandteil des mathematischen Selektionsmodells, die Mutante wird gleich einem ‚*Deus ex machina*‘ in die Population eingebracht. Das untere Bild zeigt die Abhängigkeit einer Quasispezies, einer stationären Mutantenverteilung von der mittleren Mutationsrate p .^r Im Grenzwert $\lim p \rightarrow 0$ verschwinden alle Mutanten aus der Quasispezies und beim Übergang zur fehlerfreien Replikation bleibt im stationären Zustand ausschließlich die fitteste Variante, die *Mastersequenz* X_0 , erhalten. Mit zunehmender Mutationsrate wird der Anteil der Mastersequenz an der Population immer geringer bis letztendlich beim Erreichen der Fehlerschranke bei $p = p_{cr} = 0.04501$ die stationäre Verteilung instabil wird. Gezeigt sind hier die Konzentrationen der einzelnen Fehlerklassen: $X_0 = Y_0$ ist die Mastersequenz (schwarz), Y_1 ist die Summe aller 50 Einfehlermutanten (rot), Y_2 ist die Summe aller 1225 Zweifehlermutanten (gelb), Y_3 ist die Summe aller 19 600 Dreifehlermutanten (grün), usw. Für die Berechnungen wurde eine Sequenzlänge von $l=50$ angenommen, die Fitnesswerte waren: $f_m = f_0 = 10$ für die Mastersequenz und $f_i = 1$ für alle anderen Spezies.

^r Ohne Verlust an Allgemeinheit wurde hier der Einfachheit halbe mit binären Sequenzen an Stelle der natürlichen Vierbuchstabensequenzen gerechnet.

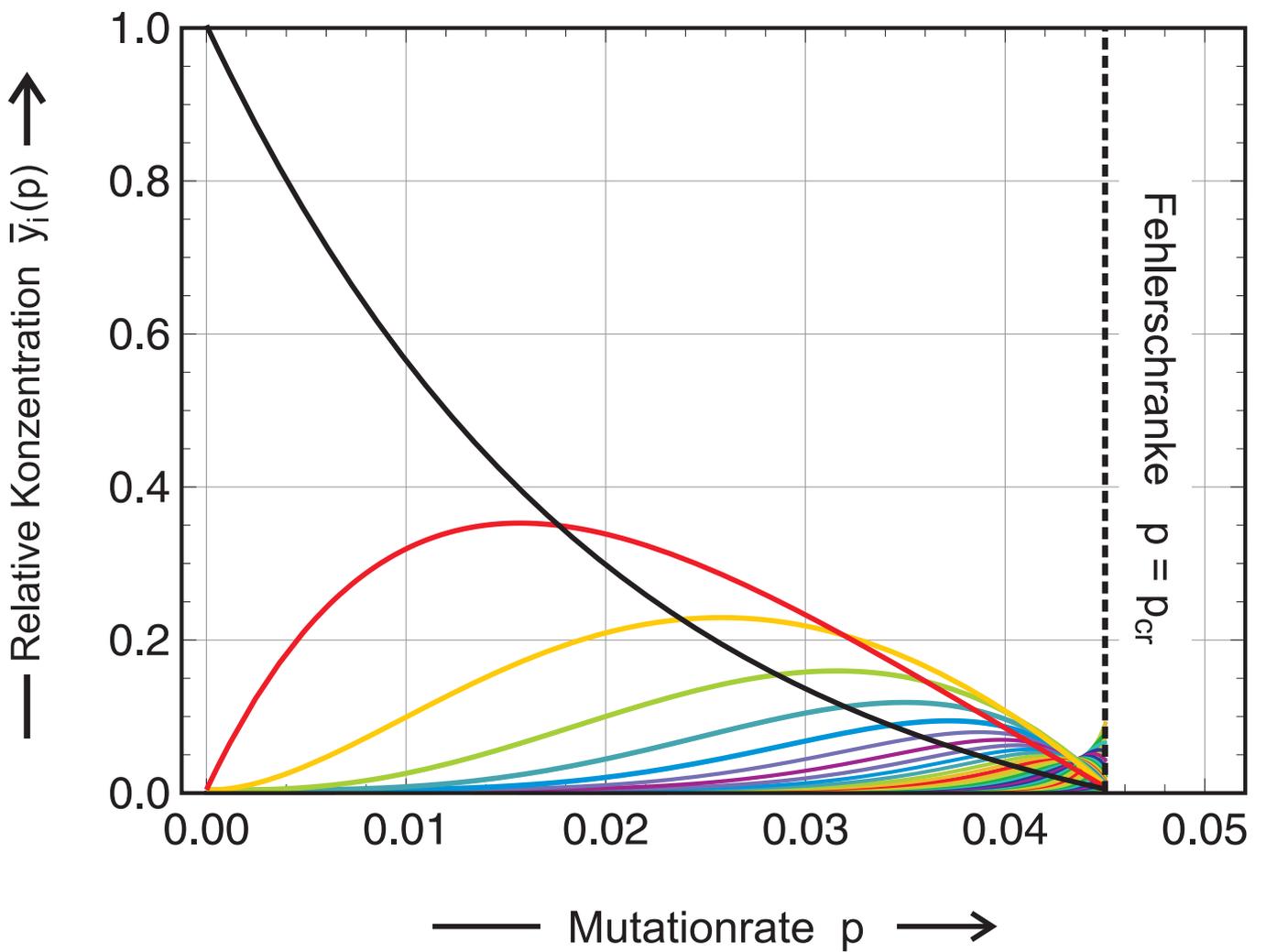
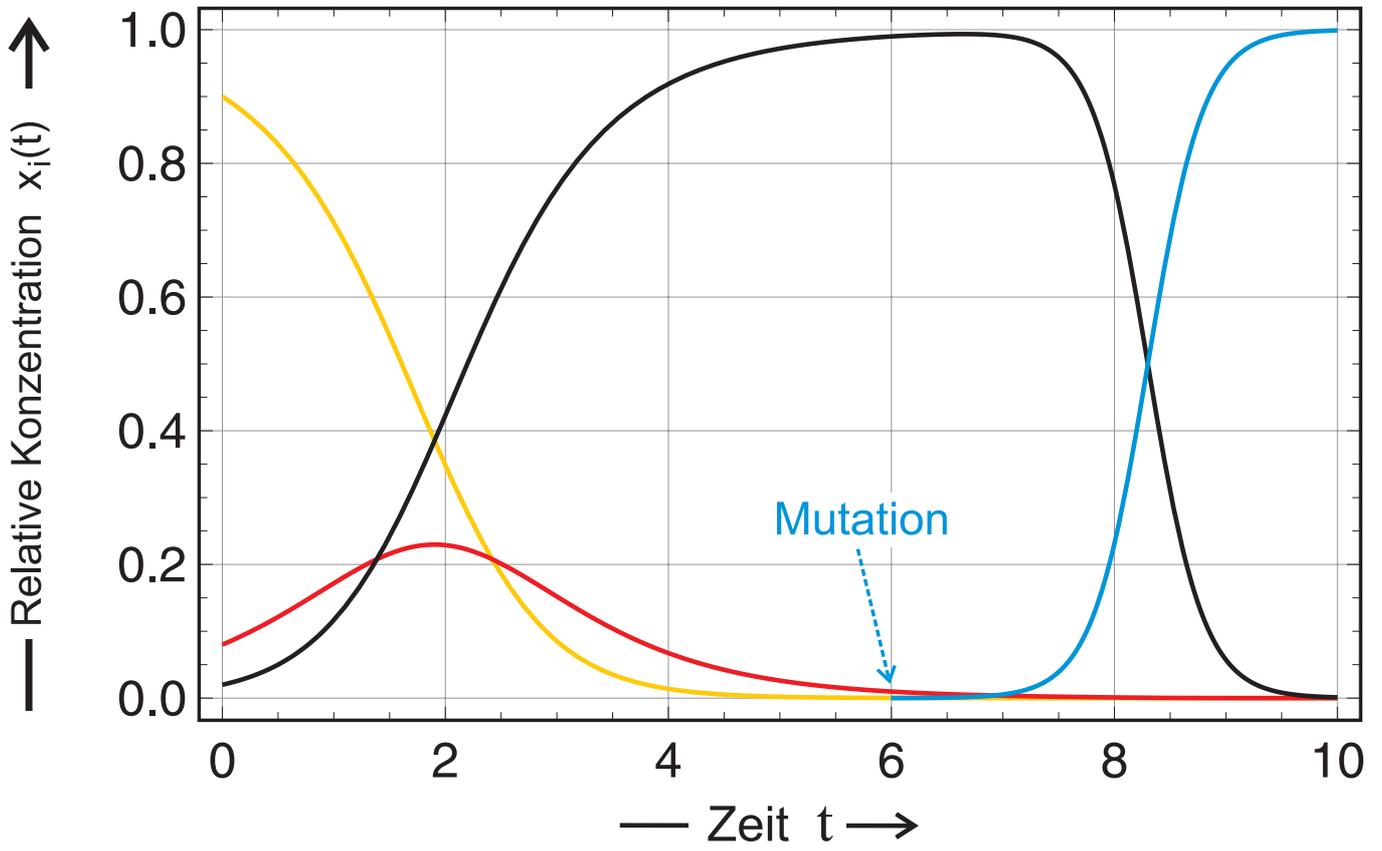


Abbildung 4: **Selektionszyklen in der evolutionären Biotechnologie.** Die gerichtete Evolution von Molekülen folgt dem allgemeinen in der Abbildung gezeigten Protokoll und folgt genau den gleichen Vorbedingungen wie das darwinsche Prinzip: Vermehrung mit Vererbung, Variation und Selektion. Vermehrung und Variation sind synonym mit Amplifikation beispielweise durch PCR und Diversifikation durch Replikation mit erhöhter Mutationsrate oder Zufallssynthese und stellen Standardtechniken der modernen Molekularbiologie dar. Gerichtete Selektion von geeigneten Molekülen stellt eine Herausforderung an das Geschick des Experimentators dar (Bezüglich eines prominenten Beispiels, SELEX, siehe Abbildung 6). Die selektierten Moleküle werden einem Test unterworfen und wenn das gewünschte Resultat vorliegt, ist das Evolutionsexperiment beendet. Die die erwarteten Eigenschaften noch nicht erreicht, wird ein weiterer Selektionszyklus bestehend aus Amplifikation, Diversifikation und Selektion angeschlossen und die Zyklen werden solange fortgesetzt bis entweder das Ziel erreicht wurde oder keine Verbesserung mehr erzielt werden kann.

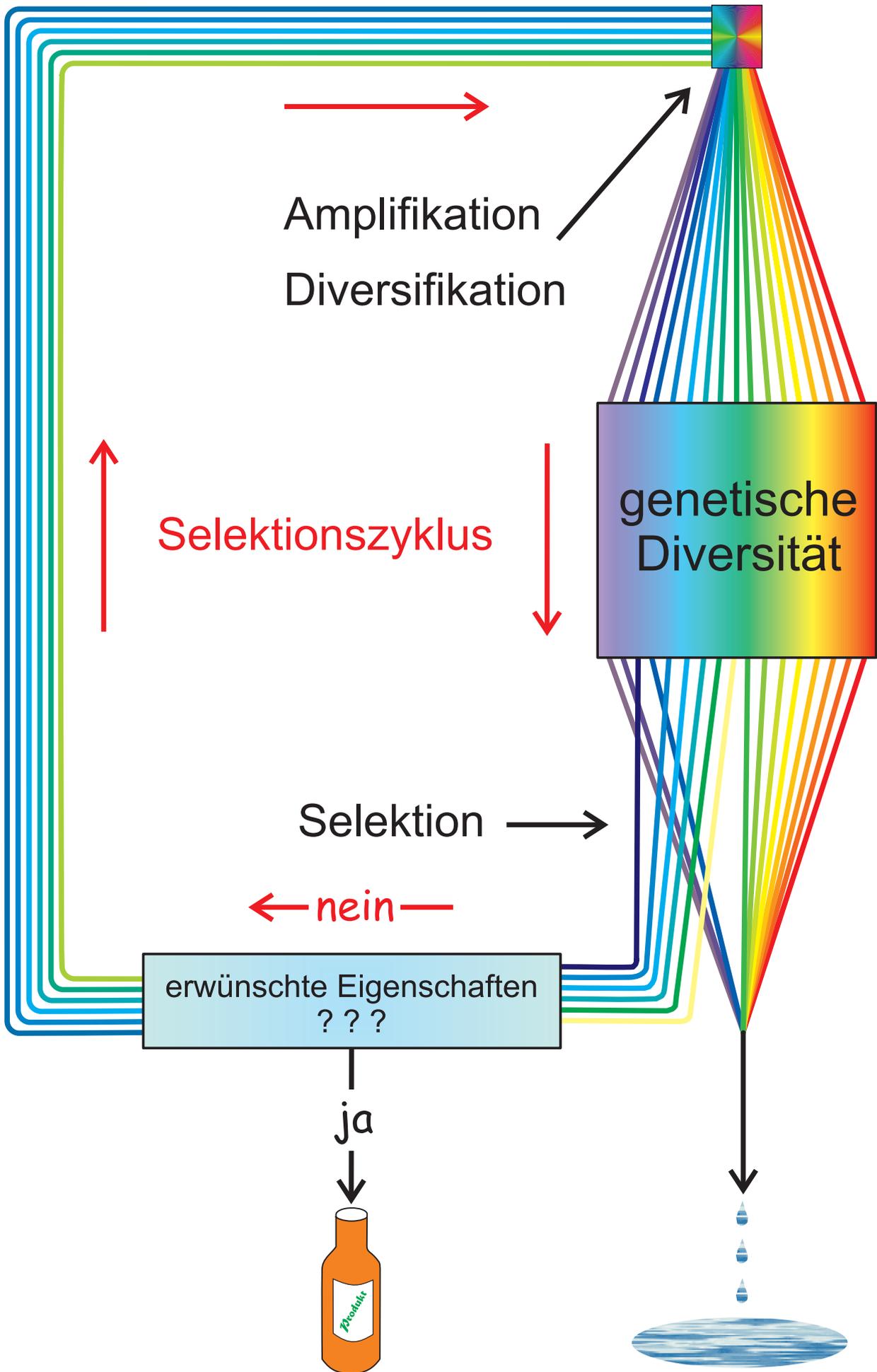
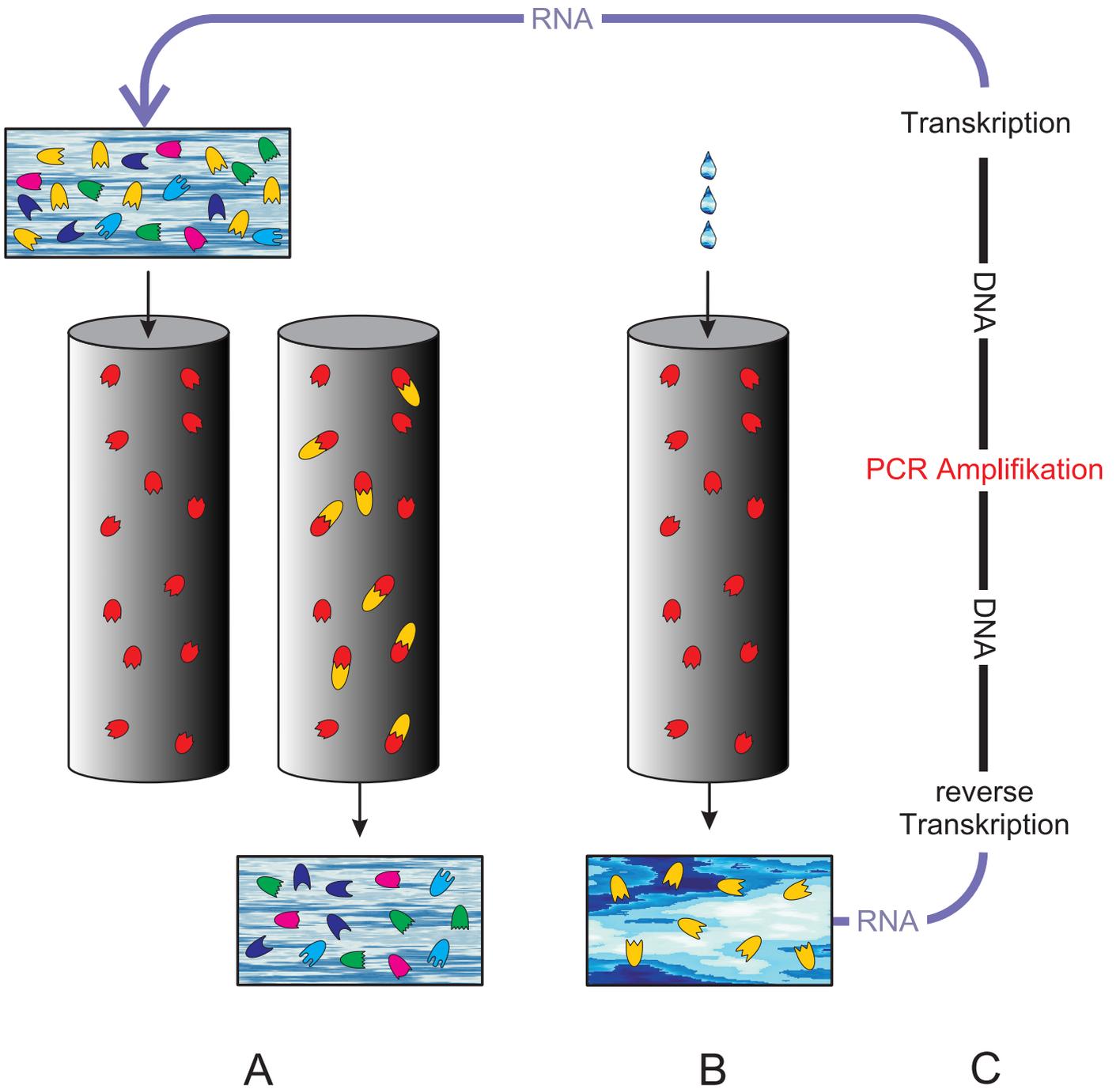


Abbildung 5: **Das ‚Hammerhead‘ Ribozym.** Das Bild zeigt den Enzym-Substratkomplex eines ‚Hammerhead‘-Ribozyms. Das Substrat (rotes Rückgrat) ist hier ein selbständiges RNA-Molekül, welches in seiner Nukleobasensequenz in zwei Strecken komplementär zum katalysierenden Ribozym ist. Dazwischen liegt die hoch spezifische Spaltungsstelle, welche durch die Geometrie des Enzym-Substratkomplexes in eine lokale Konformation gebracht wird, welche das Lösen der Bindung als einen energetisch begünstigten Prozess ablaufen lässt. Der Katalysator (violette Rückgrat) ist sequenzspezifisch auf sein Substrat abgestimmt, wodurch sich die außerordentlich hohe Spezifität erklärt. Das Bild enthüllt noch eine wichtige Eigenschaft von einsträngigen RNA-Molekülen: Die Struktur – hier auf die durch lokale Doppelhelices kreierte Sekundärstruktur beschränkt – wird in einfacher Weise durch die Sequenz codiert, denn die Strukturelemente sind Doppelhelices, die durch Rückfaltung der Polynukleotidkette auf sich selbst gebildet werden und neben den Watson-Crick Nukleobasenpaaren wegen der etwas anderen Struktur der RNA gegenüber der DNA auch noch GU und UG Nukleobasenpaare enthalten können.

Abbildung 6: **Das SELEX-Verfahren zur Erzeugung optimal bindender Moleküle.** SELEX – ,Systematic Evolution of Ligands by EXponential Enrichment – ist, wie schon der Name sagt, eine Methode zur Erzeugung von optimal an Targets bindende Molekülen, für welche der Name *Aptamere* geschaffen wurde. SELEX arbeitet in drei logischen Schritten: (A) Selektion von Aptameren aus eine Lösung, die eine Mischung möglicher Kandidaten enthält, (B) Ablösung der anhaftenden Moleküle durch ein anderes Lösungsmittel und (C) Komplettierung eines Selektionszyklus in Sinne von Abbildung 4. Die Ausgangsmischung der Aptameren ist zumeist das Ergebnis einer Zufallssynthese und weist daher sehr große genetische Diversität auf. Die Chromatographiersäule im Schritt (A) enthält die Targetmoleküle (rot) fest an die stationäre Phase gebunden und daher bleiben in diesem Schritt nur die geeigneten Aptameren (gelb) an der Säule haften. Amplifikation und Diversifikation können zumeist in einem einzigen Schritt durchgeführt werden, da die PCR-Vervielfältigung im Allgemeinen nicht sehr genau ist. Notfalls kann man auch die Mutationsrate künstlich erhöhen. Das Lösungsmittel wird von Zyklus zu Zyklus geändert, so dass das Anheften an die Targetmoleküle immer schwieriger wird. Die Zyklen sind beendet, wenn keine weitere Verbesserung der Bindung beobachtet wird.



Verwendete molekularbiologische Begriffe:

Allel	Gene kommen in mehreren Varianten vor die Allele genannt werden. Fast alle höheren Organismen sind in ihren Phänotypen diploid und haben für ihre Gene mit Ausnahme der Gene auf den Geschlechtschromosomen ein väterliches und ein mütterliches Allel.
Assay	Ein biochemischer oder molekularbiologischer Aufbau eines Experiments zur Erfüllung einer bestimmten Funktion, z.B. ein Replikationsassay dient zum Replizieren von Molekülen.
Autosom	Alle Chromosomen außer den Geschlechtschromosomen – X und Y bei Säugetieren – werden Autosomen genannt.
Bakteriophage	wörtlich übersetzt ein Bakterienfresser sind ein relativ kleine RNA- oder DNA-Viren, welche Bakterien befallen und ein paar Gene codieren, deren Übersetzungsprodukte für den Lebenszyklus des Virus unentbehrlich sind.
Chromosom	Ein Chromosom kann am besten als eine ‚Perlenkette‘ von Genen vorgestellt werden. Eine menschliche Körperzelle hat 22 Paare von Autosomen und ein Paar Geschlechtschromosomen. XY beim Mann und XX bei der Frau.
Code	Der genetische Code ist der Übersetzungsschlüssel von RNA in Protein. Jeweils drei Nukleotide codieren für einen Aminosäurerest. Darüber hinaus gibt es noch Satzzeichen, welche den Beginn und das Ende der Übersetzung markieren.
DNA	Die Desoxyribonukleinsäure ist das Trägermolekül, welches die genetische Information von allen Lebewesen außer einigen Klasse von Viren und anderen kleinen Pathogenen enthält.
Fitness	Ein Begriff der Populationsgenetik, der als Maß für die Zahl der Nachkommen dient, mit denen eine bestimmte Variante in künftigen Generationen vertreten sein wird.
Genom	Das <i>Genom</i> ist die gesamte auf der DNA – oder bei Viren auch RNA – des Organismus gespeicherte genetische Information. In der makroskopische Biologen spricht man zumeist vom <i>Genotyp</i> .
Hammingabstand	Der Hammingabstand zwischen zwei Sequenzen misst die minimale Zahl an Punktmutationen, die notwendig sind, um von einer Sequenz zur anderen zu gelangen.

Konzentration Die Menge an vorhandenem und zugänglichem Material wird in der Chemie als Konzentration bezeichnet. Berechnet wird die Konzentration als die Zahl der Moleküle einer bestimmten Spezies dividiert durch das Volumen der Probe. Um Konzentrationen in Mol/Liter zu erhalten muss noch durch die Loschmidtsche Zahl N_L dividiert werden.

messenger-RNA Die messenger-RNA ist der Träger der Information für die Synthese eines Proteins. Sie wird durch Transkription eines Stückes DNA und nachfolgende Prozessierung gebildet und am Ribosom in Protein übersetzt.

Mitochondrien Mitochondrien sind Zellorganellen in den Zellen aller höheren Organismen, welche die zelluläre Energiegewinnung durch Oxidation mit molekularem Sauerstoff bewerkstelligen.

Nukleobase Eine Nukleobase oder Nukleinbase ist eine der vier Klassen von Bausteinen in der DNA: A = Adenin, T = Thymin, G = Guanin, und C = Cytosin. In der RNA wird T durch U = Uracil vertreten.

Nukleotid Baustein einer Nukleinsäure auch Nukleotidbuchstabe bezeichnet, der aus drei Einheiten, Nukleobase, Desoxyribose (DNA) oder Ribose (RNA), und Phosphatrest besteht.

Nukleotidpaar Element einer Nukleinsäuredoppelhelix, das aus zwei gegenübergestellten Nukleotiden besteht. In die Geometrie der biologischen Nukleinsäuren passen nur bestimmte Paarungen, die Watson-Crick Paare AT und TA sowie GC und CG in der DNA.

Operon Eine Regulationseinheit auf der DNA bestehend aus den Regulatorsequenzen, an welchen die Regulatormoleküle, Aktivatoren und Repressoren, gebunden werden und im allgemeinen mehreren Genen, die für Proteine codieren, welche gemeinsam Funktionen ausüben.

Organellen Organellen sind zelluläre Substrukturen, die eine gemeinsame Funktion ausüben.

Peptid Peptide entstehen durch Verknüpfung einzelner Aminosäuren. Kurze Ketten werden als Oligopeptide, lange als Polypeptide bezeichnet. Proteine sind Polypeptide die zur Ausübung einer Funktion befähigt sind.

Pufferlösung Eine im Allgemeinen wässrige Lösung, in welcher bestimmte Konzentrationen konstant gehalten werden. Als Beispiel sei die Wasserstoffionenkonzentration gemessen als pH-Wert genannt.

- Quasispezies** Eine Quasispezies ist die stationäre Variantenverteilung eines Replikations-Mutationssystems.
- Ribosom** Ribosomen sind die Protein synthetisierenden molekularen Maschinen aller Zellen. Von einigen Autoren werden sie auch den Zellorganellen zugezählt.
- RNA** Ribonukleinsäuren sind in manchen primitiven Organismen wie Viroiden und einigen Klassen von Viren Träger der genetischen Information. In den Zellen aller Organismen übernimmt die RNA als messenger-RNA, transfer-RNA und ribosomaler RNA die tragenden Funktionen in der Proteinsynthese. Darüber hinaus ist sie in den Zellen höherer Organismen ein wichtiges Element der Regulation von Genexpression.
- transfer-RNA** Das Verbindungsstück zwischen der Protein codierenden messenger-RNA und der in Synthese begriffenen Polypeptidkette. Mit dem sogenannten Anticodon bestehend aus drei Nukleotiden liest die transfer-RNA drei Nukleotide der messenger-RNA ab und vermittelt die dem Triplet zugeordnete Aminosäure zum Aufbau der wachsenden Polypeptidkette.
- Viroid** Viroide sind bei Pflanzen pathogene RNA-Moleküle, die sich in ihren pflanzlichen Wirtszellen ohne eigene Genprodukte vermehren. Außerhalb der Wirtszellen überleben Viroide als *nackte* RNA-Moleküle.
- Virus** Viren sind Pathogene, die Bakterien, Pilze, Pflanzen und Tiere befallen. Ihr genetisches Material kann DNA oder RNA sein. Im Allgemeinen codiert ihr genetisches Material für einige virusspezifische Gene, die in den Zellen entscheidende Funktionen in den viralen Lebenszyklen haben.

Literaturzitate:

- ¹ Darwin, Charles. 1859. On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life. John Murray, London.
- ² Wegener, Alfred. 1912. Die Entstehung der Kontinente und Ozeane. *Geologische Rundschau* **3**:276-292.
- ³ Mendel, Gregor. 1866. Versuche über Pflanzenhybriden. *Verhandlungen der Naturforschenden Vereins zu Brünn* **4**:3-47.
- ⁴ Watson, James D., Crick, Francis H.C. 1953. Molecular structure of nucleic acids: A structure for deoxyribonucleic acid. *Nature* **171**:737-738.
- ⁵ Crick, Francis H.C. 1958. On Protein Synthesis. *Symp.Soc.Exp.Biol.* **XII**:139-163.
- ⁶ Crick, Francis H.C. 1960. Central dogma of molecular biology. *Nature* **227**:561-563.
- ⁷ Thieffry, Denis. 1998. Forty years under the central dogma. *Trends in Biochemical Sciences* **23**:312-316.
- ⁸ Osawa, Syozo, Jukes, Thomas H., Watanabe, Kimitsuna, Muto, Akira. 1992. Recent evidence for evolution of the genetic code. *Microbiol.Rev.* **56**:229-264.
- ⁹ Ambrogelly, Alexandre, Palioura, Sotiria, Söll, Dieter. 2007. Natural expansion of the genetic code. *Nature Chem. Biol.* **3**:29-35.
- ¹⁰ Wang, Qian, Parrish, Angela R., Wang, Lei. 2009. Expanding the genetic code for biological studies. *Chemistry & Biology* **16**:323-336.
- ¹¹ Cech, Thomas R. 1987. The chemistry of self-splicing RNA and RNA enzymes. *Science* **236**:1532-1539.
- ¹² Cech, Thomas R. 1990. Self-splicing of group I introns. *Annu. Rev. Biochem.* **59**:543-568.
- ¹³ Breaker, Ronald R. 1997. DNA enzymes. *Nature Biotechnology* **15**:427-431.
- ¹⁴ Jacob, François, Monod, Jacques. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J.Mol.Biol.* **3**:318-356.
- ¹⁵ Davis, Rowland H. 2007. Beadle's progeny: Innocence rewarded, innocence lost. *Journal of Biosciences* **32**:197-205
- ¹⁶ Sanger, Frederick. 2001. The early days of DNA sequences. *Nature Medicine* **7**:267-268.
- ¹⁷ Maxam, Allan, Gilbert Walter. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**:560-564
- ¹⁸ Erika Check Hayden. 2009. Genome sequencing: The third generation. *Nature* **457**:768-769.
- ¹⁹ Ohne, Susumu. 1972. So much "junk" DNA in our genome. In: Ed. H.H. Smith. Evolution of Genetic Systems. *Brookhaven Symposia in Biology* **23**:366-370. Gordon and Breach, New York.
- ²⁰ Mattick, John S. 2003. Challenging the dogma: The hidden layer of non-protein-coding RNAs in complex organisms. *BioEssays* **25**:930-939.
- ²¹ Mattick, John S. 2004. RNA regulation: A new genetics? *Nature Review Genetics* **5**:316-323.
- ²² Mattick, John S. 2010. RNA as the substrate for epigenome-environment interactions. *Bioessays* **32**:548-552.
- ²³ Killian, Keith. 2005. Genomic Imprinting. *Atlas Genet. Cytogenet. Oncol. Haematol.* Retrieved 22.11.1013
URL: <http://AtlasGeneticsOncology.org/GenomImprintID20032.html>
- ²⁴ Jänisch, Rudolf. 1997. DNA methylation and imprinting: Why bother? *Tends in Genetics* **13**:323-329.
- ²⁵ Brenner, Sydney. 1999. Theoretical biology in the third millenium. *Phil. Trans. Roy. Soc. London* **354**:1963-1965.
- ²⁶ Spiegelman, Sol: 1971. An approach to the experimental analysis of precellular evolution. *Quart. Rev. Biophys.* **4**:213-253.

-
- ²⁷ Eigen, Manfred. 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**:465-523.
- ²⁸ Weissmann, Charles. 1974. The making of a phage. *FEBS Letters* **40**:S10-S18.
- ²⁹ Gerald F. Joyce. 2007. Vierzig Jahre Evolution im Reagenzglas. *Angewandte Chemie* **119**:6540-6557.
- ³⁰ Biebricher, Christof K., Eigen, Manfred, Gardiner Jr., William C. 1983. Kinetics of RNA replication. *Biochemistry* **22**:2544-2559
- ³¹ Biebricher, Christof K. 1983. Darwinian selection of self-replicating RNA molecules. *Evolutionary Biology* **16**:1-52.
- ³² Eigen, Manfred, Schuster, Peter. 1978. The hypercycle. A principle of natural self-organization. Part B. The abstract hypercycle. *Naturwissenschaften* **65**:7-41.
- ³³ Eigen, Manfred, McCaskill, John, Schuster Peter. 1989. The molecular quasispecies. *Adv. Chem. Phys.* **75**:149-263.
- ³⁴ Eigen, Manfred, Schuster, Peter. 1977. The hypercycle. A principle of natural self-organization. Part A. The emergence of the hypercycle. *Naturwissenschaften* **64**:541-565.
- ³⁵ Cech, Thomas R., Zaug, Arthur. J., Grabowski, Paula J. 1981. The intervening sequence of the ribosomal RNA precursor is converted into a circular RNA in isolated nuclei of tetrahymena. *Cell* **27**:467-476.
- ³⁶ Kruger, Kelly., Grabowski, Paula J., Zaug, Arthur. J., Sands, Julie., Gottschling, Daniel. E. and Cech, Thomas R. 1982. Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* **31**, 147-157.
- ³⁷ Gurrier-Takada, C., Gardiner, K., March T., Pace, Norman, Altman, Sidney. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**:849-857.
- ³⁸ Hamman, Christian, Luptak, Andrej, Perreault, Jonathan, De la Peña. 2012. The ubiquitous hammerhead ribozyme. *RNA* **18**:871-885.
- ³⁹ Doudna, Jennifer A., Cech. Thomas R. 2002. The chemical repertoire of natural ribozymes. *Nature* **418**:222-228.
- ⁴⁰ Cech, Thomas R. 2000. The ribosome is a ribozyme. *Science* **289**:878-879.
- ⁴¹ Steitz, Thomas A., Moore, Peter B. 2003. RNA, the first macromolecular catalyst: The ribosome is a ribozyme. *Trends in Biochemical Sciences* **28**:411-418.
- ⁴² Roth, Adam, Breaker, Ronald R. 2009. The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* **78**:305-334.
- ⁴³ Breaker Ronald, R. 2012. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**:a003566.
- ⁴⁴ Reidys, Christian, Stadler, Peter F., Schuster, Peter. 1997. Generic properties of combinatorial maps. RNA secondary structures. *Bull. Math. Biol.* **59**:339-397.
- ⁴⁵ Flamm, Christoph, Hofacker, Ivo L., Maurer-Stroh, Sebastian, Stadler, Peter F., Zehl; Martin. 2001. Design of multistable RNA molecules. *RNA* **7**:254-265.
- ⁴⁶ Schuster, Peter. 2006. Prediction of RNA secondary structures. From theory to models and real molecules. *Rep. Prog. Phys.* **69**:1419-1477.
- ⁴⁷ Joyce, Gerald F. 2004. Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.* **73**:791-836.
- ⁴⁸ Serganov, Alexander, Keiper, Sonja, Malinina, Lucy, Tereshko, Valentina, Skripkin, Eugene, Höbartner, Claudia, Polonskaia, Anna, Phan, Ahn Tuân, Wombacher, Richard, Micura, Ronald, Dauter, Zbigniew, Jäschke, Andres, Patel, Dinshaw J. 2005. Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nature Struct. & Mol. Biol.* **12**:218-224.
- ⁴⁹ Muhlbacher, Jérôme, St-Pierre, Patrick, Lafontaine, Daniel A. 2010. Therapeutic applications of ribozymes and riboswitches. *Curr. Op. Pharmacology* **10**:551-556.
-

-
- ⁵⁰ Brakmann, Susanne, Johnson, Kai, Eds. 2002. *Directed evolution of proteins or how to improve enzymes for biocatalysis*. Wiley-VCh Verlag, Weinheim, DE.
- ⁵¹ Arnold, Frances H., Ed. 2001. Evolutionary protein design. *Adv. Protein Chemistry* **55**:1-438.
- ⁵² Jäckel, Christian, Kast, Peter, Hilvert, Donald. 2008. Protein design by directed evolution. *Annu. Rev. Biophys.* **37**:152-173.
- ⁵³ Wrenn, SA, Jarrett, Harbury, Pehr B. 2007. Chemical evolution as a tool for molecular discovery. *Annu. Rev. Biochem.* **76**:331-349.
- ⁵⁴ Tuerk, Craig, Gold, Larry. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to T4 DNA polymerase. *Science* **249**:505-520.
- ⁵⁵ Ellington, Andrew D., Szostak, Jack W. 1990. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**:818-822.
- ⁵⁶ Klussmann, Sven, Ed. 2006. *The aptamer handbook. Functional oligonucleotides and their applications*. Wiley-VCh Verlag, Weinheim, DE.
- ⁵⁷ Liszka, Michael J., Clark, Melind E., Schneider, Elizabeth, Clark, Douglas S. 2012. Nature versus nurture: Developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.* **3**:77-102.
- ⁵⁸ Wintrode, Patrick L., Arnold, Frances H. 2001. Temperature adaptation of enzymes: Lessons from laboratory evolution. *Adv. Protein Chemistry* **55**:161-225.
- ⁵⁹ Zaks, Aleksey, Klibanov, Alexander M. 1988. Enzymatic catalysis in nonaqueous solvents. *J. Biol. Chem.* **263**:3194-3201.
- ⁶⁰ Chen, Keqin, Arnold, Frances H. 1991. Enzyme engineering for nonaqueous solvents: Random mutagenesis to enhance activity of subtilisin E in polar organic media. *Nature Biotechnology* **9**:1073-1077.
- ⁶¹ You, Lingchong, Arnold Frances H. 1996. Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng.* **9**:77-83.
- ⁶² Tramontano, Alfonso, Janda, Kim D., Lerner, Richard A. 1986. Catalytic antibodies. *Science* **234**:1566-1570.
- ⁶³ Pollack, Scott J., Jacobs, Jeffrey W., Schultz, Peter G. 1986. Selective chemical catalysis by an antibody. *Science* **234**:1570-1573.
- ⁶⁴ Shokat, K. M., Schultz, Peter G. 1990. Catalytic antibodies. *Annu. Rev. Immunol.* **8**:335-363.
- ⁶⁵ Xu, Yang, Yamamoto, Noburo, Janda, Kim D. 2002. Catalytic antibodies: Hapten design strategies and screening methods. *Bioorganic & Medicinal Chemistry* **12**:5247-5268.
- ⁶⁶ Tawfik, Dan S., Green, Bernard S., Chap, Rachel, Sela, Michael, Eshnar, Zelig. 1993. *catELISA*: A facile general route to catalytic antibodies. *Proc. Natl. Acad. Sci. USA* **90**:373-377.
- ⁶⁷ Domingo, Esteban, Parrish, Colin R., Holland, John, Eds. 2008. Origin and evolution of viruses. Second edition. Elsevier-AcademicPress, Amsterdam, NL.
- ⁶⁸ Gago, Selena, Elena, Santiago F., Flores, Ricardo, Sanjuán. 2009. Extremely high mutation rate of a hammerhead viroid. *Science* **323**:1308.
- ⁶⁹ Bull, James J., Anceles Myers, Lauren, Lachmann Michael. 2005. Quasispecies made simple. *PLoS Comput. Biol.* **1**:e61.
- ⁷⁰ Bull, James J., Sanjuán Rafael, Wilke, Claus O. 2007. Theory of lethal mutagenesis for viruses. *J. Virology* **81**:2930-2939.
- ⁷¹ Tejero, Hector, Marín, Arturo, Motero, Francisco. 2010. Effect of lethality on the extinction and on the error threshold of quasispecies. *J. Theoret. Biol.* **262**:733-741.
-