

Romantik, Supercodes und die Milchstraßen-DNA

Ein künstlerisches Transanimationsprinzip

Dieser Artikel stellt in relativer Ausführlichkeit dar, wie sich hochauflösende Bilder präzise in Form synthetisierter DNA-Moleküle¹ codieren lassen. Für dieses Thema sind diverse Elemente der Molekularbiologie, Mathematik und Informatik unverzichtbar, doch gruppieren sich diese wichtigen technisch-wissenschaftlichen Aspekte um ein starkes poetisches Motiv.

Die Künstler des Goldenen Zeitalters waren ganz und gar der mimetischen Darstellung der natürlichen Welt, vor allem des menschlichen Körpers, verpflichtet, strebten die Griechen doch nichts Geringeres als dessen „vollkommene Kenntnis“ an. Aus den idealisierten Proportionen des menschlichen Körpers leiteten sie die klassischen Grundlagen der Musik, der Architektur und sogar der Wissenschaft und Mathematik ab. Zur großen Kunsttradition rund um diese „Selbsterkenntnis“ gehörte in vielen historischen Beispielen auch die Suche nach einer besonderen Kontrolle über die flüchtigen Eigenschaften der Vitalität und der Funktionen, die Leben und Tod voneinander scheiden.

Die Suche nach den „Geheimnissen des Lebens“, die die Domäne der Literatur und der Kunstgeschichte ist, wird heute natürlich immer intensiver in den Labors der sogenannten „Life Sciences“ auf der ganzen Welt verfolgt. In ihnen ist – zumindest was gewisse Biomoleküle betrifft – der jahrhundertalte Traum vom „Zum-Leben-Erwecken“ unbelebter Materie nicht länger eine Angelegenheit von Magie, Mythos, Legende oder göttlicher Intervention. Gleichwohl findet man die vielleicht dramatischsten und folgenreichsten Versuche, unbelebte Materie zum Leben zu erwecken, nicht wirklich in der Kunst oder Molekularbiologie, sondern vielmehr im Umfeld neuester wissenschaftlicher Versuche, mit Außerirdischen zu kommunizieren. Mit einer Alchemie aus Raketen, Plaketten auf Raumsonden, leistungsstarken Radaranlagen und in den Weltraum gesendeter Binärbotschaften versucht die Wissenschaft den gesamten Kosmos zu animieren.

1986 stellte ich durch die Verbindung mathematischer Strategien, die früher zum Abfassen von Botschaften für außerirdische Intelligenzen verwendet wurden, mit Standardtechniken der DNA-Synthese mein erstes synthetisches DNA-Molekül her. Dieses Molekül mit dem Titel *Microvenus* entstand als Kunstwerk in Zusammenarbeit mit Dana Boyd und Jon Beckwith an der Harvard Medical School und mit Hatch Echols an der University of California in Berkeley. Es enthielt eine Rastergrafik der altgermanischen Rune für „Leben“² und eine schematische Abbildung äußerer weiblicher Genitalien, die bis dahin aus grafischen Darstellungen in seriösen wissenschaftlichen Botschaften an außerirdische Intelligenzen ausgeschlossen waren. Mit der für das Folgeprojekt, *Riddle of Life*³, synthetisierten DNA setzten wir die molekularen Implikationen der modell-basierten Mitteilungen, die die Nobelpreisträger Max Delbrück und George W. Beadle im Jahr 1958 miteinander ausgetauscht haben, um. Die 1993 in Zusammenarbeit mit Burkhardt Wittigs Labor an der Freien Universität Berlin geschaffene *Riddle of Life*-DNA ist mit Max Delbrücks englischem Text „I am the riddle of life, know me and you will know yourself“ codiert. Sowohl das *Microvenus*- als auch das *Riddle of Life*-Molekül sind seither in die DNA virusähnlicher bakterieller Vektoren, sogenannter Plasmide, eingesetzt und anschließend in Laborkulturen von *E-Coli*-Bakterien kloniert worden.

Mit ihren Texten und einfachen Strichgrafiken in Form von DNA deuten diese Arbeiten die Verwendungsmöglichkeit von DNA als Speichermedium für konventionelle Computerda-

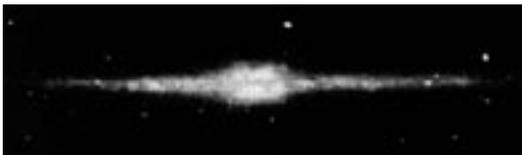
ten an. Die für die Herstellung des *Microvenus*- und *Riddle of Life*-Moleküls verwendeten Codierstrategien wären dafür allerdings ungeeignet und es war dabei auch nicht an eine direkte Kompatibilität mit den organisatorischen und operativen Konventionen von Computer-Datenbanken gedacht. Mein Plan zur Entwicklung einer computerfreundlichen und biochemisch praktikablen Codierstrategie zur Umwandlung einer normalen Computerdatei in eine DNA-Sequenz steht in Übereinstimmung mit dem wissenschaftlich-technischen Ideal einer solchen biologischen Datenbank. Mein Entschluss, ein Bild der Milchstraße in dieser Form darzustellen, hat allerdings hoffnungslos romantische Gründe.

Vor vielen Jahren zeigte mir ein Freund einen Skizzenblock mit Zeichnungen für eine Kindergeschichte über ein Mädchen, das zu seinem Glück eine Maus finden musste, die eine Landkarte der ganzen Welt in ihrem Ohr eingeschrieben hatte. Inspiriert wurde ich auch – obwohl das technisch oder wissenschaftlich belanglos ist – von dem Umstand, dass das „Vergissmeinnicht“ in der botanischen Taxonomie *Mysotis* genannt wird, was im Griechischen „Mausohr“ heißt.

Ein Gen-großes Bild der Milchstraße

Die ersten als Sequenz von DNA-Basen⁴ codierten hochauflösenden Bilddaten sind eine Landkarte der Milchstraße. Das Bild wurde als 3867-mer DNA-Molekül (ein Molekül mit 3867 Basen) codiert. Ein synthetisches DNA-Molekül mit 3867 Basen ist ziemlich groß,⁵ ja, gehört vermutlich zu den größten je hergestellten synthetischen DNA-Molekülen. Es ist großenteils mit vielen in der Natur vorkommenden Genen vergleichbar und übertrifft sogar manche Plasmide.⁶ Im Gegensatz zu den Genen organisierter Lebensformen wird dieses jedoch nicht von den diversen Bausteinen des Zellapparats in signifikante Proteine oder Enzyme übersetzt, sondern ist einzig und allein zur Übersetzung mittels technischer Hilfsmittel gedacht.

Die zur Herstellung der Milchstraßen-DNA-Karte verwendeten Daten wurden ursprünglich von Instrumenten des NASA-Satelliten *Cosmic Background Explorer* (COBE)⁷ gesammelt. Bevor die Ergebnisse der COBE-Experimente in den frühen neunziger Jahren zugänglich gemacht wurden, waren für den Astronomen viele Teile unserer eigenen Galaxie durch intergalaktische Staubwolken verdeckt. Die interferometrischen Infrarotsensoren des COBE-Satelliten lieferten die ersten hochauflösenden Karten der gesamten Galaxie mit bisher nie dagewesenen Bildern von ihrem Kern.⁸ Das Portfolio dieser Entdeckungen bildet vielleicht den wichtigsten Fortschritt in der Kartografie seit Gerhardus Mercator im 16. Jahrhundert.



COBE-Bild der Milchstraße

Die primäre Existenzform des COBE-Bilds war eine von „Ein“- und „Aus“-Zuständen, zuerst in den Halbleitern der Detektoren im Weltraum und später im Halbleiterspeicher der Computer, an die die ursprünglichen Daten gesandt wurden. Die grafische Darstellung oben ist nur *eine* von mehreren Ausprägungsmöglichkeiten für diese Information.

Computer-Codes

Die dem Infrarotbild der Milchstraße – oder der binären Identität *jeder* Computerdatei – entsprechenden „Nullen“ und „Einsen“ lassen sich leicht mit einem der Editoren dar-

stellen, die man beim Kauf eines Computers mit den Softwarepaketen mitgeliefert bekommt. Diese verbreiteten Hilfsprogramme ermöglichen eine rasche und leichte Umwandlung binärer „Bilddaten“ in verschiedene Standardformate. Gängige Bildformate wie JPEG, GIF, TIFF usw. lassen sich entweder als Textdateien aus alphanumerischen Zeichen oder als Bilddateien aus „Pixels“ genannten Bildpunkten betrachten. Ich ermittelte die dem Milchstraßenbild entsprechenden Binärdaten mit einem für Macintosh geschriebenen Ressourceneditor „ResEdit“ (siehe Binäräquivalent des COBE-Milchstraßenbilds, siehe Seite 219).

Da es zum Verrücktwerden wäre, selbst kleinere Computeroperationen mit einem Binär-code auszuführen, sprechen menschliche Programmierer den binären Computerspeicher mit übergeordneten Informationsstrukturen an, die den Binär-code in die 16 hexadezimalen oder Basis-16-Zahlen „0“ bis „F“ kompilieren. Um Zeit zu sparen, habe ich für meine DNA-Codierung des COBE-Milchstraßenbilds ebenfalls ein hexadezimalen oder „Hex“-Äquivalent des Bildes verwendet. Dasselbe ResEdit-Programm, mit dem ich bereits die Binärdaten erhalten hatte, lieferte mir auch das Hex-Äquivalent des Milchstraßenbilds (siehe Hexadezimaläquivalent des COBE-Milchstraßenbilds, siehe Seite 220).

DNA-Zahlen

Interpretiert man die DNA mathematisch, so ist die Umwandlung von digitaler Information in DNA-Sequenzen und umgekehrt ein einfacher Rechenvorgang. Die DNA-Moleküle sind variable Modulverbindungen, die zumindest konzeptuell der mathematischen Struktur von Computerspeichern gleichen. Der augenfälligste Unterschied ist der, dass der DNA-Speicher mit vier ganzzahligen Werten statt mit zwei Binärzahlen arbeitet.

Die vier DNA-„Zahlen“ bestehen aus den vier beweglichen Teilen der DNA-Moleküle, den sogenannten DNA-Basen Cytosin, Thymin, Adenin und Guanin bzw. „C“, „T“, „A“ und „G“. Bei doppelsträngigen DNA-Molekülen sind diese Basen in Paaren angeordnet, die die Mittelsprossen der leiterartigen DNA-Struktur bilden. „C“ verbindet sich immer mit „G“, und „G“ mit „C“; „A“ verbindet sich immer mit „T“, und „T“ mit „A“. Die Natur verwendet Dann-Moleküle mit variablen Basensequenzen auf ähnliche Weise zur Informationsspeicherung, wie der Computer Daten auf Festplatten, Magnetbändern, CDs und Chips speichert.

Die Schaltung der „Ein/Aus“-Zustände bei Halbleitern beruht darauf, dass die Ladung bei „Aus“ geringer ist als bei „Ein“. Die „Aus“-Zustände werden als „0“ dargestellt, und die „Ein“-Zustände als „1“. Quantitätswerte lassen sich auch zur logischen Reihung der DNA-Basen verwenden. Glücklicherweise hat keine der vier Basen dieselbe Größe. So kann man jeder von ihnen einen Wert zuordnen, der ihrem Molekulargewicht entspricht:

Molekulargewichte der DNA-Basen

Molekül	Molekulargewicht	Wert
Cytosin	111.10	0
Thymin	126.11	1
Adenin	135.13	2
Guanin	151.13	3

Mit den oben angegebenen Werten lässt sich jede numerische Datenbank in eine DNA-Sequenz übersetzen. Damit ist ein Schlüssel für die gegenseitige Konvertierung von DNA-, Binär- und Hexadezimalwerten gegeben.

DNA-Binär-Hex-Zahlenschlüssel**C = 00; T = 01; A = 10; G = 11**

CC = 0000 = (0)	AC = 1000 = (8)
CT = 0001 = (1)	AT = 1001 = (9)
CA = 0010 = (2)	AA = 1010 = (A)
CG = 0011 = (3)	AG = 1011 = (B)
TC = 0100 = (4)	GC = 1100 = (C)
TT = 0101 = (5)	GT = 1101 = (D)
TA = 0110 = (6)	GA = 1110 = (E)
TG = 0111 = (7)	GG = 1111 = (F)

Verwendet man entweder für die Hexadezimalwerte „A“ und „C“ oder für die DNA-Basen „A“ und „C“ Pseudonyme (wie z. B. „X“ und „Z“), so lässt sich der eine Code schon auf einem kleinen Personalcomputer mit der „Suchen-und-Ersetzen“-Funktion eines Textprogramms wie MS-Word leicht in den anderen übersetzen. Das Milchstraßenbild kann somit direkt in den folgenden 2936-mer DNA-Code übertragen werden, allerdings mit einem Problem (2936-mer Milchstraßen-DNA [Primärstrang], siehe Seite 222).

Das Problem bei dieser Sequenz hängt mit der Schwärze des Weltalls zusammen. Schwarze Flächen werden im Binär- oder Hex-Code als wiederkehrende „0000“ bzw. „0“ und im DNA-Code als wiederkehrende „CC“-Sequenzen wiedergegeben. Aufgrund der Dominanz schwarzer Flächen im COBE-Milchstraßenbild gibt es in der entsprechenden DNA-Sequenz lange Poly-„C“-Ketten. Die durch die elektromechanische Torsion einer einzelnen Base verursachten Gestaltveränderungen eines einsträngigen DNA-Moleküls⁹ sind zwar kaum messbar, aber die gemeinsame Torsionswirkung vieler identischer Basen kann zur Bildung von Schleifen und Windungen führen, die in einem heterogeneren Molekül nicht auftreten würden. Ein in einer wässrigen Lösung schwebendes DNA-Molekül ist normalerweise äußerst flexibel. In diesem flexiblen Zustand weist die DNA sehr feine Strukturmerkmale auf, die von anderen damit interagierenden Molekülen erkannt werden. Enzyme, die mit heterogener DNA ganz normal reagieren, neigen dazu, an langen Poly-C-Ketten „abzugleiten“ oder sie zu „überspringen“. Überdies sind die üblichen DNA-Sequenzier- oder „Rücklese“-Techniken bei der Sequenzierung langer Poly-C-Ketten unverlässlich. Auch hierbei gibt es ein „Überspring“-Problem, das mit der Art zusammenhängt, wie sich DNA-Moleküle in der sogenannten „Gelelektrophorese“, einem zentralen Bestandteil der gängigen Sequenzieretechniken, durch chemische Gele bewegen.¹⁰ Darum wäre es äußerst schwierig, die 2936-mer Milchstraßen-DNA in dieser Form zu synthetisieren und in den Reproduktionsapparat lebender Zellen zu klonieren, und es wäre gleichermaßen schwierig, sie mit der existierenden Technologie zu sequenzieren oder rückzulesen.

Ironischerweise kommen lange Poly-C-Ketten auch in der Natur vor, aber nur im sogenannten Junk. Die Junk-DNA wird von den „Transkriptions“- und „Translations“-Prozessen¹¹, die am der Betriebsgeschehen funktionierender DNA¹² beteiligt sind, nicht erfasst. Ebenso wie synthetische Poly-C-DNA lassen sich auch Poly-C-Junk-Moleküle nicht leicht sequenzieren und sind daher mit wenigen Ausnahmen in bestehenden Genom-Datenbanken nicht verzeichnet. Aus diesem Grund bringt es auch nichts, in den Archiven der Genom-Forschung nach Homologien der 2936-mer Milchstraßen-DNA zu suchen.

DNA-Supercodes

Nummerische Daten lassen sich also „doch nicht direkt in brauchbare DNA übertragen. Ein brauchbarer DNA-„Speicher“ muss Daten speichern können, ohne dass er biochemisch

problematisch ist. Um einen solchen funktionierenden biologischen Informationsspeicher für die Daten des Milchstraßenbilds – oder für jede andere generische, nichtbiologische Datenbank – zu schaffen, können die korrespondierenden DNA-Sequenzen in Sequenzen einer zweiten Generation umcodiert werden, die 1.) „biochemisch verträglich“ sind, 2.) die nichtbiologischen Daten der ersten Generation sehr getreu wiedergeben und 3.) den Umfang des ursprünglichen Datenbestands nicht wesentlich erhöhen.

Zu diesem Zweck wird hier eine zweite Verschlüsselungsstrategie, der „DNA-Supercode“, vorgestellt, der die Übersetzung einer bestimmten DNA-Sequenz in eine Reihe von Sequenzen zweiter Ordnung oder supercodierter Sequenzen erlaubt, die über verschiedene Zwischenstufen wieder präzise in die Originalsequenz zurückverwandelt werden können. Supercodierte Milchstraßensequenzen behalten sowohl die ursprünglichen COBE-Bilddaten bei als auch das logische System der gestaffelten Werte, mit dem diese in die DNA übersetzt wurde.

Ich experimentierte anfangs mit einer Reihe verschiedener Datenhandlungsstrategien, die auf unterschiedliche Weise mit Zeichenrotation und Doublett- und Triplettsverschlüsselung der Originalsequenz arbeiteten. Immer jedoch gaben die durch die jeweilige Strategie erzeugten Symmetrien eine biochemisch signifikante Symmetrie an die supercodierte DNA-Sequenz weiter, wobei einige dieser Strategien den Umfang des Moleküls beträchtlich erhöhten. Supercodes, die die Daten der ersten Generation lediglich in Faktoren zerlegen, scheinen ein unzuverlässiges Mittel zur Schaffung nichtbiologischer Datenbanken mittels nicht vergrößerter biochemisch geeigneter DNA-Moleküle zu sein. Obwohl jede der untersuchten Sequenzen Aspekte aufwies, die sich einfach synthetisieren und assemblieren ließen, war die Schaffung eines handhabbaren Milchstraßenmoleküls nur mit einem willkürlichen Patchwork dieser Supercodes möglich. Um die ursprünglichen Daten daraus zu decodieren, benötigte man eine „Straßenkarte“ dieses willkürlichen Code-Patchworks. Theoretisch ließe sich zwar eine solche „Straßenkarte“ oder Decodieranleitung mit jeder molekularen Datenbank mitliefern, doch scheint ein solches Szenario nicht realistisch, weil es uneingeschränkte und zeitaufwendige Komplexität zulässt. Die Notwendigkeit zusätzlicher „Straßenkarten“ für die Decodierung würde den ursprünglichen Datenumfang wahrscheinlich signifikant erhöhen. Meiner Meinung nach sind beliebig zusammengewürfelte Anordnungen von faktorisierten Supercodes kein praktikables Mittel zum Aufbau biologischer Speicher für konventionelle Datenbanken.

Da aber andererseits generische Daten selbst nicht „uniform“ sind, ist zu erwarten, dass variable strukturelle Aspekte (auch unerwünschte) ihrer Übersetzung in DNA willkürlich auftreten. Die oben beschriebenen DNA-Supercodestrategien basieren auf kontinuierlichen und regelmäßigen Modifikationen der Originaldaten. Diese können unerwünschte strukturelle Elemente topologisch komplexer gestalten, dafür kann die inhärente Symmetrie uniformer Angleichungen strukturelle Probleme hervorrufen, wo es vorher keine gab. Ein besserer Supercode wäre asymmetrisch und würde eine variable Verschlüsselung zulassen, die sich auf die Lösung verschiedener Problemtypen zuschneiden lässt. Genau auf diese Weise funktionieren gewisse natürliche Operationen des genetischen Codes.

Degenerierter Supercode

Die Natur muss bestimmte Gene im Milieu des evolutionären Wandels erhalten. Der Einfluss der natürlichen Selektion und der genetische Apparat der geschlechtlichen Fortpflanzung sorgen dafür, dass der Kontext natürlicher DNA ständig in Fluss bleibt. Der Natur gelingt es irgendwie, ganz bestimmte Proteine mit einem ständig umgeschriebenen DNA-Code zu reproduzieren. Dazu kann die Basensequenz des betreffenden Gens auf viele verschiedene Arten „reformuliert“ werden. Jede dieser alternativen Sequenzen ist in Zellprodukte oder Proteine übersetzbar, die mit dem Produkt der Originalsequenz

identisch sind. Die Fähigkeit der Natur zu solchen Reformulierungen basiert auf der sogenannten Degeneriertheit des genetischen Codes. Um diese Eigenschaft der Degeneriertheit zu erklären, will ich kurz die grundlegenden Prozesse rekapitulieren, mit denen Information in DNA gespeichert, in RNA transkribiert und in Proteine translatiert werden.

Transkription und Translation

Lange DNA-Moleküle werden letztlich in sämtliche Substanzen übersetzt, aus denen Lebewesen bestehen. In einem von Biologen als „Transkription“ bezeichneten Vorgang wird eine Kopie oder Matrize von einer Seite der DNA-Doppelhelix in einen anderen Nucleinsäuretyp – RNA (Ribonucleinsäure) – umgeschrieben. Mit Hilfe eines Enzyms, der „RNA-Polymerase“, wird die Information des ursprünglichen DNA-Moleküls in eine Vielzahl von RNA-Molekülen, die sogenannte Messenger- oder Boten-RNA (mRNA) kopiert.¹³ mRNA-Moleküle sind bis auf zwei grundsätzliche Ausnahmen strukturell mit DNA-Molekülen identisch: 1.) der Ribose-Zucker der RNA enthält ein zusätzliches Sauerstoffmolekül und 2.) die DNA-Base Thymin („T“) wird in die kleinste mRNA-Base, Uracil („U“), transkribiert.

mRNA-Moleküle sind Zwischenhändler bei dem Prozess, mit dem der ursprüngliche DNA-Code in Proteine translatiert wird. In den Zellen wird die in mRNA-Molekülen gespeicherte Information durch sanduhrförmige Bausteine, die Ribosome, verarbeitet. Diese heften sich an die mRNA-Moleküle an und lesen sie in Dreiergruppen ab. Sodann wird mit Hilfe eines wieder anderen RNA-Moleküls, der Transfer-RNA oder tRNA, eine neue Matrize angefertigt. Diese Matrize der dritten Generation wird nicht mehr in einer Nucleinsäure wie DNA oder RNA geschrieben, sondern die Information des ursprünglichen DNA-Moleküls – und des RNA-Mittlers – wird in ein Protein umgewandelt. Diese neue Matrize entspricht genauso wie die mRNA der Basensequenz des ursprünglichen DNA-Moleküls. In der Regel fügt das Ribosom für jede Dreiergruppe von mRNA-Basen eine Aminosäure hinzu. Für jedes Basentriplett oder Codon des ursprünglichen DNA-Moleküls findet sich ein entsprechendes Codon in der mRNA-Matrize, das durch die Tätigkeit eines Ribosoms und tRNAs in eine Aminosäure translatiert wird.

Es gibt in fast allen Lebewesen nur 20 Aminosäuren.^{14, 15} Aus diesen 20 Aminosäuren erzeugt die Natur so unterschiedliche Dinge wie Tomaten und Menschen. Aminosäuren verbinden sich zu Peptiden und diese wiederum zu Proteinen. Fast alles in der natürlichen biologischen Welt ist aus – oder durch Interaktion mit – Proteinen gemacht. Diese letzte Matrize ist wesentlich größer als die DNA oder mRNA, und diese Kopie wird schlussendlich zum lebenden Organismus selbst.

Man kann sich diese verschiedenartigen Vorgänge leicht wie die Vorgänge in einer Fabrik vorstellen: Die Genom-DNA enthält die ursprünglichen „Zeichnungen“. Die mRNA-Moleküle sind die an die Arbeitsplätze verteilten „Blaupausen“. Ribosome und tRNA Moleküle sind die zellulären „Fließbänder und Fabrikarbeiter“, die den Aufbau langer Aminosäureketten ausführen, welche den ursprünglichen Zeichnungen entsprechen. Dies sind, in Summe, die biologischen Operationen der Transkription und der Translation.

DNA wird in eine mRNA-Kopie gepaust, welche von Zelloperationen bearbeitet wird, die jedes Nucleinsäure-Triplett in eine von 20 Aminosäuren translatiert. Aus vier Einzelbasen lässt sich ein Satz von 64 Triplets bilden. Jedes dieser Triplets steuert die Produktion einer der 20 Aminosäuren. Diese Verbindung der 64 Nucleinsäure-Triplets mit den 20 Aminosäuren – häufig in Form einer rechteckigen Tabelle dargestellt – ist der genetische Code (siehe Seite 225).

Obwohl es in diesem Code 64 Positionen gibt, sind in der Natur mit wenigen Ausnahmen nur 20 Aminosäuren und das Stopp-Codon codiert. Die 64 Positionen des genetischen Codes enthalten aus drei Buchstaben bestehende „Wörter“ für lediglich 20 verschiedene „Bedeutungen“ (mehr, wenn man die drei Stopp-Codons ebenfalls als „Bedeutung“ zählt).

Im genetischen Code gibt es also 44 Codons mehr als zur Codierung dieser 20 genetischen „Bedeutungen“ nötig sind, sodass in den meisten Fällen verschiedene Codons zur Codierung einer bestimmten Aminosäure verwendet werden. Diese Synonymie lässt eine beträchtliche Flexibilität in der Zusammensetzung von DNA-Codes zu, um den Aufbau bestimmter Proteine zu steuern. Dank dieser Flexibilität können selbst kleine Proteine mit einer astronomischen Anzahl alternativer DNA-Sequenzen beschrieben werden. Und genau diese Eigenschaft bezeichnen die Biologen als die „Degeneriertheit“¹⁶ des genetischen Codes. Sollte ein Teil einer bestimmten DNA-Sequenz aus irgendwelchen Gründen problematisch werden, so gibt es zahlreiche andere Wege, diesen Teil umzuschreiben und damit das Problem zu lösen, ohne das zu translatierende Protein zu ändern.

Man beachte, dass die Degeneriertheit des genetischen Codes dazu verwendet werden kann, die Identität bestimmter Translationsprodukte beizubehalten, nicht aber die genaue Identität der DNA-Sequenz, die diese ursprünglich codierte.

Interpretiert man dagegen die DNA-Triplets als Codons für *Zahlen* anstelle von Aminosäuren, so kann man die Degeneriertheit des genetischen Codes zur Schaffung supercodierter DNA-Sequenzen nutzen, aus denen sich eine ursprüngliche DNA-Sequenz exakt wiederherstellen lässt. Ich will nun beschreiben, wie ein solches System zur Supercodierung synthetischer DNA-Moleküle – wie der Milchstraßen-DNA – verwendet werden kann, um biochemisch handhabbare nichtbiologische Datenspeicher zu konstruieren.

Damit dieser Supercode mit einer ähnlichen Variabilität wie der genetische Code arbeitet, werden 64 Triplets auf ähnliche Weise zur Darstellung von 20 Zahlwerten verwendet, wie die Natur damit 20 Aminosäuren darstellt. Dieser Supercode operiert also mit der mathematischen 20er Basis, bei der jedes Triplet zur Bezeichnung einer Vigesimalzahl zwischen „0“ und „J“ (in Dezimalzahlen 0 bis 19) verwendet wird. Auch dabei werden in Imitation der natürlichen Degeneriertheit des genetischen Codes die 20 Zahlwerte analog der natürlichen Verteilung der Aminosäuren auf den 64-er Code abgebildet, allerdings mit vier Ausnahmen: das Triplet „CCC“ wird für die Darstellung von „C“ reserviert, „UUU“ („TTT“) steht für „T“, „AAA“ für „A“ und „GGG“ für „G“.

In der herkömmlichen Schriftsprache wird die Vorkommenshäufigkeit bestimmter alphabetischer Zeichen in einem Text durch die lexikalischen und umgangssprachlichen Charakteristika der betreffenden Sprache reguliert. So ist etwa der Buchstabe „e“ der häufigste Buchstabe im Englischen oder Deutschen, ist das aber nicht unbedingt in jeder anderen Sprache. Ähnlich werden auch Aminosäuren aus DNA-Triplets je nach der artspezifischen Verwendungshäufigkeit übersetzt. Das heißt, je nachdem, von welcher Spezies die DNA stammt, werden gewisse DNA-Triplets mehr oder weniger häufig in Aminosäuren translatiert. Zur Reihung der Triplets für die Zahlenwerte „0“ bis „J“ beschloss ich hier, die normale Translationshäufigkeit von Aminosäuren in den Zellen des Homo sapiens heranzuziehen.

Beim Homo sapiens ist die ungefähre Reihenfolge der Translationshäufigkeit der Aminosäuren und der ihnen logisch zuordenbaren Zahlenwerte auf der Basis 20 (die Quantitäten entsprechen der Vorkommenshäufigkeit) wie aus Tabelle auf Seite 226 ersichtlich (Quelle: Dr. Jeff Spitzner).¹⁷

Diese Zahlenwert/Codon-Zuordnungen könnten als intuitiv falsch erscheinen, weil die höchsten Zahlen den am wenigsten häufigen Codons zugeordnet sind. Im Allgemeinen weisen Computerdaten keine langen Folgen gleicher Werte auf, die dann als hohe Zahlenwerte codiert würden. Folglich würde ein degenerierter Supercode häufiger Daten mit niedrigeren Zahlenwerten darstellen als mit höheren. Da die Zahlenzuordnung auf der Verwendungshäufigkeit beruht, werden also die häufiger verwendeten Codons niedrigeren Zahlenwerten zugeordnet, die im Computercode normalerweise häufiger verwendet werden.

Im folgenden Schlüssel sind die Basis-20-Zahlenäquivalente und die Codons für „C“, „T“, „A“ und „G“ (Konturschrift) in einem Format ausgedrückt, das üblicherweise zur Darstellung des genetischen Codes verwendet wird (siehe Seite 227).

Ein Schlüssel für die Umwandlung der Vigesimal- in Dezimalzahlen wird wie auf Seite 228 abgebildet wiedergegeben.

Bei der Arbeit mit diesen Supercode werden die drei Stopp-Codons, „UAA“ (TAA), „UGA“ (TGA), und „UAG“ (TAG) wie in der Natur zum *Stoppen* oder Terminieren des Translationsvorgangs verwendet, in diesem Fall aber auch, um den Daten-„Leseraster“ eines Moleküls je nach Bedarf zu *starten*. Da die effizienteste Supercodeverwendung das selektive Editieren einer bestimmten Sequenz zuließe, traf ich die willkürliche Entscheidung, zwei der Stopp-Codons zum Ein- und Ausschalten des Supercode-„Editors“ zu verwenden, u. zw. dergestalt dass:

UAA oder TAA (*) = Faktoriere die folgende Sequenz + Lösche dieses Codon
und

UGA oder TGA () = Uneditierte Sequenz folgt + Lösche dieses Codon**

Das dritte Stopp-Codon wird zum Löschen einer eingefügten Sequenz verwendet:

UAG oder TAG (X) = Lösche die folgende Sequenz + Lösche dieses Codon

Als einen Beispielsatz mit dem ersten Codon (TAA) könnte man etwa die Sequenz „AATCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC“ als „TAATCTAAATTTGGTCAACCC“ supercodieren. Man beachte, dass „TAATCTAAATTTGGTCAACCC“ lediglich eine von vielen Sequenzen ist, die mit dem degenerierten Basis-20-Supercode zur exakten Beschreibung der Sequenz „AATCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC“ erzeugt werden kann, wobei:

TAATCTAAATTTGGTCAACCC
= **TAA (zerlege die folgende Sequenz + lösche TAA)**
+ **TCT (2) AAA (Adenin)**
+ **TTT (Thymin)**
+ **GGT (1) CAA (B) CCC (Cytosin) [Basis-20]**
= **(lösche TAA) + 2 Adenin + 1 Thymin + 1B Cytosin**
= **(lösche TAA) + 2 Adenin + 1 Thymin + 31 Cytosin [Basis-10]**
= **AATCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC [DNA]**

Als einen Beispielsatz mit dem zweiten Codon (TGA) könnte man die Sequenz „CGCTAGCTGCGATA“ als „TGACGCTAGCTGCGATA“ supercodieren. In diesem Fall lässt sich nur ein korrekter Supercode-Ausdruck erzeugen: „CGCTAGCTGCGATA=TGACGCTAGCTGCGATA“ wobei:

TGACGCTAGCTGCGATA
= **TGA (uneditierte Sequenz folgt + lösche TGA)**
+ **CGCTAGCTGCGATA**
= **CGCTAGCTGCGATA [DNA]**

Das dritte Stopp-Codon, „UAG“ (TAG), ist für Supercode bestimmt, der bei der Codierung der Milchstraßenbilddaten nicht direkt beteiligt ist. Seine Rolle wird später noch näher erläutert.

Da die Stopp-Codons *sowohl* in der Originalsequenz *als auch* im Supercode nur schwer simultan zu handhaben sind, wurden sämtliche Stopp-Codons der ursprünglichen 2936-mer Sequenz in supercodierte Sequenzen übersetzt, die keine Originalstopps enthalten (Beispiele nachfolgend):

TAA = TAATTTTCCAAA
TGA = TAATTTGGGAAA
TAG = TAATTTGGCAAAGG

Eine andere Option wäre, sowohl „TGA“- als auch „TAA“-Supercoderausdrücke zu verwenden, nämlich so dass:

TAA = TGATATAAAAA

TGA = TAATTTGAGA

TAG = TAATTTAAATGAG

Nach der Supercodierung aller Stopps der Originalsequenz sind sämtliche in der editierten Sequenz direkt vorkommenden Stopp-Codons Supercod-„Anweisungen“, die schließlich im Verlauf der Decodierung gelöscht werden. Derselbe Decodierungsvorgang stellt auch den ursprünglichen Satz von Stopp-Codons und ihre exakte Positionierung in der 2936-mer Milchstraßen-DNA wieder her.

Aufgrund seiner Degeneriertheit könnte man Trillionen verschiedener Supercodesequenzen erstellen, die sich alle in die exakt gleichen Milchstraßenbilddaten decodieren ließen. Viele davon würden die Aktivität natürlicher vorkommender DNA in lebenden Zellen imitieren und wären von natürlicher DNA nicht zu unterscheiden. Somit kann der degenerierte Basis-20-Supercod zur Schaffung von Computerdatenbanken in DNA-Molekülen verwendet werden, die mit den heute verfügbaren Techniken der Molekularbiologie manipuliert und sequenziert werden können.

Neben seiner Rolle als Datenmanipulator ist der degenerierte Basis-20-Supercod auch zum Maßschneiden von Sequenzen verwendbar, und damit zur Minimierung von In-vivo-Translationen supercodierter Datenbanken in unerwünschte Peptide oder Proteine, aber auch zum erleichterten Assemblieren der zahlreichen Produkte der DNA-Synthese (Oligonucleide) in ganze doppelsträngige, gengroße Moleküle wie die Milchstraßen-DNA. Aber zuerst einmal zur Frage der zufälligen Translation:

In-vivo-Translation supercodierter Datenbanken

Die im Lauf der Evolution zur Steuerung der Proteinproduktion entstandenen DNA-Sequenzen sind hoch spezialisiert. Dagegen würden durch die Umwandlung normaler Computerdateien in DNA erzeugte Sequenzen nur sehr wenig oder gar keine biologische Aktivität aufweisen und die Wahrscheinlichkeit, dass aus solchen Datenbanken irgendwelche pathologischen Substanzen (via Transkription und Translation) exprimiert werden könnten, ist extrem gering. Da die Möglichkeit besteht, dass einmal alle generischen, nichtbiologischen Datenbanken als DNA-Sequenzen codiert werden, und da sich die exakten Bedingungen, die zur In-vivo-Translation dieser Sequenzen in pathologische Substanzen führen, nicht absolut vorhersagbar sind, wurden verschiedene Aspekte des Supercodes zur Begrenzung einer In-vivo-Translation supercodierter DNA herangezogen.

Stopp-Codons

Wie oben erwähnt, beginnen alle Supercoderausdrücke mit Stopp-Codons als „Anweisungen“ oder „Einschränkungen“ und sind als solche über die gesamte supercodierte DNA-Sequenz verteilt. Darüber hinaus sind die Leseraster von Supercod Ausdrücken im Gegensatz zu denen natürlicher genetischer Anweisungen nicht immer auf die Niederschrift in Dreiergruppen beschränkt. Da die Stopp-initiierten Anweisungsausdrücke des Supercodes oft von einem natürlichen Leseraster in den nächsten rutschen können, lassen sich Stopp-Codons überall in den multiplen Leserastern supercodierter Sequenzen verteilen.

Start-Codons

So wie Stopp-Codons das Ende einer in Protein zu translatierenden DNA-Sequenz anzeigen, so enthält auch der Anfang einer Sequenz ein spezielles Signal, das die Translation in Gang setzt. Die Rolle dieses speziellen Start-Codons, das den Anfang einer in Protein zu translatierenden Sequenz anzeigt, übernimmt gewöhnlich das „ATG“-Triplet

(Methionincodon).¹⁸ Um die Wahrscheinlichkeit einer In-vivo-Translation weiter zu reduzieren, kann man mit dem Supercode die Start-Codons aus allen sechs Leserastern einer bestimmten Sequenz entfernen. Zur Demonstration dieser Möglichkeit wurden aus der supercodierten Milchstraßen-DNA praktisch sämtliche „ATG“-Codons entfernt.¹⁹

Supercode-Hilfen für die Gen-Assemblierung

Aufgrund der Längenbeschränkungen für die Textbeiträge zu dieser Publikation habe ich Einzelheiten über meine speziellen Pläne zur Assemblierung des Milchstraßen-DNA-Moleküls aus einzelnen Bestandteilen weggelassen. Aus demselben Grund bin ich auch nicht näher auf die Struktur und automatisierte Synthese der DNA eingegangen.

Verschiedene Verfahren zur Assemblierung großer synthetischer DNA-Moleküle mit hunderten oder tausenden von Basenpaaren benötigen den Einsatz einer großen Zahl individuell synthetisierter Oligonucleotide. Bei der Milchstraßen-DNA müssen für die gesamte Sequenz mindestens 45 Einzelteile assembliert werden. Eine der Operationen des DNA-Supercode wurde als Hilfsinstrument für die Reparatur möglicher, bei der Herstellung solch großer Assemblagen auftretender Fehler reserviert. Wie vorhin angeführt, werden zwei der drei Stopp-Codons („TAA“ and „TGA“) als „Anleitungsausdrücke“ verwendet, die die Verschlüsselung (Supercodierung) oder Nichtverschlüsselung einer Sequenz steuern. Das dritte Stopp-Codon, „TAG“, ist dazu bestimmt, Supercodeausdrücke einzuleiten, die beim Decodiervorgang bewusst gelöscht werden sollen. In der nachfolgenden supercodierten Milchstraßen-DNA-Sequenz wurde das „TAG“-Codon dazu benutzt, einmalige Erkennungsstellen (unten fett gedruckt) für Restriktionsenzyme an Positionen einzufügen, die jedes der etwa 100-mer Fragmente der gesamten Sequenz flankieren. Der Supercode wurde auch benutzt, Duplikate der Erkennungsstellen für Restriktionsenzyme „umzuformulieren“ (zu entfernen), um die Anzahl der einmaligen zu erhöhen. Diese einmaligen Erkennungsstellen ermöglichen eine einfache Abtrennung jedes einzelnen Fragments mit nur ein oder zwei Restriktionsenzymen. Das herausgeschnittene fehlerhafte Fragment lässt sich dann von den verbleibenden fehlerfreien Sequenzen mit einer Methode (Gelelektrophorese) trennen, die DNA-Moleküle nach ihrer Größe sortiert. Wenn also in irgendeinem Teil der assemblierten Sequenz ein Fehler gefunden wird, muss man lediglich ein einziges korrigiertes 100-mer Oligonucleotid-Fragment resynthetisieren. Das Reassemblieren des korrigierten „Gens“ würde dann die Zusammenfügung von zwei größeren bereits assemblierten Fragmenten mit einem fehlerkorrigierten Fragment von etwa 100 Basenpaaren Länge erfordern. Die folgende supercodierte Milchstraßen-DNA-Sequenz besteht aus 348 diskreten Supercodeausdrücken, die alle mit „TAG“, „TAA“ oder „TGA“ beginnen, und von zwei 18-mer-„Armen“ flankiert ist, um die Einfügung in bakterielle Vektoren zu ermöglichen (siehe 3867 Milchstraßen-Supercode-DNA, Seite 231).

Der Supercode ist auch für die Installation einmaliger Erkennungsstellen verwendet worden, die jenen Teil der Milchstraßen-DNA flankieren, der die Bilddaten enthält, weil bei jedem Terminus zusätzliche, den Klonierungsvektoren komplementäre Sequenzen eingebaut wurden. Diese Terminationssequenzen enthalten spezielle „palindromische“ Restriktionsstellen, so dass später mit nur einem Enzym (*BpuA I*) selektiv nur die Sequenz mit den Bilddaten aus der DNA des Klonierungsvektors (oder der -vektoren) herausgeschnitten werden kann, in die sie eingesetzt worden ist.

Die Wiederherstellung sichtbarer Bilder

Die COBE-Karte der Milchstraßengalaxie ist nun zu einer Sequenz aus DNA-Basen kompiliert worden. Mit einem DNA-Supercode wurde diese Sequenz so angepasst, dass sie strukturell und biochemisch dem Molekularapparat lebender Zellen gleicht. Nun lässt

sich ein korrespondierender Satz von Oligonucleotiden effizient zu einem DNA-Molekül mit 3867 Basen synthetisieren und assemblieren. Mit Hilfe gebräuchlicher Techniken kann man dieses Molekül in eine von mehreren Zellbibliotheken oder biologischen Trägern einsetzen. Ähnlich kann man die Originalsequenz mit existierenden Instrumenten und Techniken wiederherstellen.

Dazu stehen eine Reihe von DNA-Sequenzierverfahren zur Verfügung. Die avanciertesten heute eingesetzten DNA-Sequenziermaschinen überspielen die Information der DNA-Sequenz direkt in einen Computerspeicher. Die aus dem biologischen Speicher wieder gewonnene Sequenzinformation der Milchstraßen-DNA kann – egal mit welcher Methode – wieder in Textform in den Computer eingegeben und dann rasch mit denselben (oder ähnlichen) Hilfsprogrammen, mit denen die DNA-Sequenz ursprünglich aus den Bilddaten erzeugt wurde, wieder in diese zurückverwandelt werden.

Zuerst kehrt man die Supercodeprotokolle um und konvertiert die supercodierte 3867-mer DNA-Sequenz mit Hilfe der „Ersetzen“-Funktion in die 2936-mer-Sequenz der ersten Generation. Dann kann man mit der „Ersetzen“-Funktion und dem DNA-auf-Hexadezimal-Schlüssel die DNA-Sequenz der ersten Generation in Hexadezimaldaten konvertieren. Von hier aus lässt sich das COBE-Milchstraßenbild mit ResEdit auf einem Macintosh in 10 Schritten wiederherstellen:

1. Kopiere die Hexadezimaldaten als Text.
2. Rufe ResEdit auf (durch Anklicken des Icons).
3. Rufe „CREATE NEW FILE“ auf.
4. Rufe „CREATE NEW RESOURCE“ auf.
5. Wähle „PICT“-Quelle.
6. Rufe eine leere „PICT“-Datei im „(FILENAME)“ Fenster auf.
7. Rufe „OPEN USING HEX EDITOR“ aus der aktiven Datei im „PICTs FROM (FILENAME)“ Fenster auf.
8. Wähle [markiere] alle Daten [Nullen] im „PICT ID = (#)“ Fenster.
9. Füge die Textdatei mit den hexadezimalen Milchstraßendaten ein
10. Schließe das „PICT ID=(#)“ [Hex-Editor] Fenster.

Damit sollte das digitale COBE-Milchstraßenbild automatisch im „PICTs FROM (FILENAME)“ Fenster erscheinen. Der „OPEN RESOURCE EDITOR“ liefert ein Bild in voller Größe.

Das Mausohr

Man weiß, dass lebende Organismen diskrete „biologische Perioden“ ausbilden, die genau mit lokalen Planetenzyklen korrespondieren.²⁰ Diese Perioden beschreiben die Wechselwirkungen des Sonne-Mond-Erde-Systems so präzise, dass man, wenn sich nur die Masse eines dieser Körper schätzen lässt, daraus die Masse der anderen beiden und die Entfernung zwischen allen dreien errechnen kann. Newtons fundamentale Gleichung $F=ma$ (Kraft = Masse x Beschleunigung), die die Bewegung aller Objekte beschreibt, lässt sich zur Beschreibung der Himmelskörperbewegungen in planetarischen Modellen umformen. Kraft wird zur universalen Gravitationskraft, der sogenannten Gravitationskonstante (K).²¹ Masse wird zur Masse der interagierenden Himmelskörper wie Erde-Sonne oder Erde-Mond ($M_1 + M_2$). Beschleunigung entspricht bei einem Objekt in kurvilinear (orbitaler) Bewegung dessen Winkelgeschwindigkeit multipliziert mit dem Radius der Kurve oder der Entfernung zwischen den beschriebenen Objekten (R) dividiert durch die Zeit oder Orbitalperiode (p). Newtons Gesetz der Planetenbewegung wird wie folgt notiert:

$$K (M_1 + M_2) = R^3 / p^2$$

Dieses Gesetz beschreibt eher kreisförmige als elliptische Umlaufbahnen, aber dem Amateur-„Bio-Astronomen“ wird diese Gleichung genügen, um grobe Masse- und Ent-

fernungsschätzungen vorzunehmen, in die biologische Perioden als p aufgenommen werden. Um etwa die Masse der Erde und die Entfernung Erde-Sonne zu bestimmen, würde man erst eine Schätzung der Sonnenmasse als M_1 einführen (es gibt verschiedene Mittel, um die Masse der meisten sichtbaren Sterne zu schätzen) und das 365-Tage-Jahr als p . Sobald die Erdmasse bestimmt ist, könnte eine ähnliche Gleichung für das Erde-Mond-Modell erstellt werden.

Für unsere Zwecke hier genügt das, um darauf hinzuweisen, dass Mäuse und andere lebende Organismen bereits subtile „Karten“ des lokalen Kosmos in sich tragen und dass ein künstliches Gen, das astronomische Informationen enthält, zu einem gewissen Grad redundant sein mag.

Gegenwärtig kennen Biologen mehrere Methoden zur Schaffung rekombinanter oder „transgener“ Mäuse. Bei einer dieser Methoden werden geschwächte Retroviren als Vektoren verwendet. Normale Viren „übernehmen“ den genetischen Apparat infizierter Zellen, um neue Viren zu produzieren. Retroviren treiben diese verdeckte Aktion einen Schritt weiter und fügen ihre Gene tatsächlich in die genomische DNA von ihnen infizierter Zellen ein. Eines dieser Viren, das Molonivirus, wurde genetisch so manipuliert, dass es keine pathologischen Eigenschaften mehr aufweist, aber die Fähigkeit beibehält, seine Wirtszellen zu infizieren – und seine Gene permanent in sie zu integrieren. Mit konventionellen Techniken wird fremde DNA in die DNA des Molonivirus „eingeschnitten“. Heute benutzen Biologen das Molonivirus und andere Retroviren routinemäßig zur Einfügung experimenteller Gene in die Genom-DNA von Labormäusen.

Eine anderen Methode, die sogenannte „Oocyten-Injektion“, operiert mit der mikromechanischen und biologischen Manipulation reifer Eizellen, die dann fertilisiert und chirurgisch in Leihmütter eingepflanzt werden. Der erste Schritt bei dieser Methode erfordert die Entnahme reifer Eizellen (Oocyten) aus dem Ovarium (das in diesem Fall das einer Maus ist). Mit etwa 1 mm Größe können Mauseizellen mit bloßem Auge gesehen werden. Dann wird unter einem Mikroskop reine („fremde“ oder synthetische) DNA mit ganz feinen Glaspipetten, sogenannten Mikropipetten, direkt in die Oocyten injiziert. Ist die Fremd-DNA in eine Oozyte injiziert worden, ist sie irgendwie für immer in eine der zellulären Chromosomen integriert (wie genau dieser Integrationsprozess vor sich geht, wird noch immer nicht ganz verstanden). Die künstlich manipulierte Oocyte wird in vitro-fertilisiert und dann in den Uterus der Mausleihmutter eingepflanzt. Von da an entwickelt sich der transgene Embryo ganz normal weiter. Der Nachwuchs der Leihmaus wird auf das Vorhandensein des neuen Gens untersucht und dann mit traditionellen Tierzuchtmethoden eine reine Linie von Mäusen, die das Gen in sich tragen, gezüchtet.

Anmerkungen

- 1 Desoxyribonucleinsäure (DNA), die in Struktur und Funktion mit der in biologischen Organismen natürlich vorkommenden DNA identisch ist, ist durch chemische Synthese künstlich herstellbar.
- 2 „Microvenus“, *Art Journal*. Spring, 1995
- 3 „'Genetic Art' builds cryptic bridge between two cultures“, *Nature*. No.378. S. 229. 1995
- 4 Die „DNA-Basen“ werden auf S. 238 ff dieses Artikels erklärt.
- 5 Das größte synthetische DNA-Molekül, das ich bis jetzt gefunden habe, ist ein synthetisches Gen mit 7000 DNA-Basen, das die Midland Certified Reagent Co. Molecular Biology Group (3112 Cuthbert Ave., Midland, TX 79701) für eine Biotechnologie-Firma in der Gegend um Boston konstruiert hat (unveröffentlicht).
- 6 Plasmide sind autonome, virusartige Elemente, die selbst ganze Gensammlungen enthalten.
- 7 COBE brachte auch wichtige, in diesem Artikel nicht erörterte Erkenntnisse, unter anderem eine Kartierung des kosmischen Strahlungshintergrunds, die die wissenschaftlichen Theorien über Kosmologie und den „Urknall“ tiefgreifend beeinflusst haben.

- 8 Das in der Milchstraßen-DNA codierte COBE-Bild ist ein „noch nie gesehenes“, nahezu im Infrarotbereich liegendes Bild der Milchstraße. Das Bild wurde aus einem Konglomerat von Daten kompiliert, die durch das „Diffuse Infrared Background Experiment“ (DIRBE), einem von drei verschiedenen wissenschaftlichen COBE-Experimenten, gewonnen wurden. Die Bilddaten wurden mit den durch Flüssighelium gekühlten Detektoren des DIRBE-Experiments mit Unterbrechungen während der ersten sechs Monate im Orbit gesammelt und im April 1990 freigegeben. Es zeigt die Milchstraße in einer Randansicht mit dem galaktischen Nordpol oben, dem Südpol unten und dem Zentrum in der Mitte. Das Bild wurden von verschiedenen Punkten in unserem eigenen Sonnensystem aus gesammelt, das nahe der galaktischen Außenfläche liegt. Es kombiniert mehrere fast im Infrarotbereich gemachte Aufnahmen. Die vorherrschende Lichtquelle in diesem Wellenbereich stammt von Sternen unserer eigenen Galaxie. Obwohl unser Sonnensystem ein Teil der Milchstraße ist, wirkt das Bild wie eine Fernansicht, weil das Licht größtenteils von der Sternenpopulation stammt, die näher beim Zentrum der Galaxie liegt als die Sonne. Ein Bild von der galaktischen Scheibe und den Spiralarmen der Milchstraße ist noch nie gemacht worden, weil die dafür benötigten Aufnahmewinkel Tausende von Lichtjahren entfernt liegen.
- Der COBE-Satellit wurde am 18. November 1989 an Bord der letzten Deltarakete der NASA von der Vandenberg Air Force Station in Kalifornien gestartet. COBE wurde speziell zur Erforschung der „Urknall“-Strahlung entwickelt. (Quelle: NASA COBE gif comments)
- 9 Doppelsträngige DNA-Moleküle werden Strang für Strang synthetisiert.
- 10 Aus noch nicht ganz geklärten Gründen werden Poly-Cs bei der Sequenzierung in konventionellen Acrylamidgels „übersprungen“.
- 11 Die biologischen Prozesse *Transkription* und *Translation* werden weiter unten in diesem Artikel ausführlicher beschrieben.
- 12 Der Umstand, dass „Junk“-DNA nicht der mit der Protein-Translation einhergehenden permanenten chemischen Manipulation unterworfen ist, scheint damit zu tun zu haben, dass „Junk“-DNA konserviert ist, d. h. mit sehr geringer Wahrscheinlichkeit gezielt editiert wird oder Mutationen erlebt, wie sie bei normaler DNA vorkommen.
- 13 Beim biologischen Vorgang der Transkription heftet sich das Enzym RNA-Polymerase an die doppelsträngigen DNA-Moleküle an und formt mit ihr einen sogenannten „Transkriptions-Komplex“. Die RNA-Polymerase bildet einen einzelnen mRNA-Strang aus, der von nur einem Strang der „Mutter“-DNA genommen wird. D. h. es entsteht ein einsträngiges mRNA-Molekül, das eine nur in 5'→3'-Richtung des DNA-Muttermoleküls angefertigte Matrize darstellt.
- 14 Lehninger, 1975
- 15 Neben den 20 gängigen Aminosäuren sind in einigen spezialisierten Proteintypen auch noch andere relativ seltene Aminosäuren isoliert worden, u. a. Hydroxyprolin, Hydroxylysin, Desmosin und Isodesmosin. Einige sehr ungewöhnliche methylierte Aminosäuren sind außerdem in bestimmten Muskelproteinen gefunden worden, u.a. Methylhistidin, Methyllysin und Trimethyllysin. Sie alle sind Derivate von einer gängigen Aminosäure. Über 150 weitere Aminosäuren kommen biologisch entweder allein oder in kombinierter Form vor, jedoch nie in Proteinen.
- 16 Auch Mathematiker benutzen diesen Begriff zur Beschreibung von Problemen, die eine Vielzahl von Lösungen haben. Statt einer einzigen richtigen Lösung, die an bestimmter Punkt oder eine bestimmte Zahl ist, gibt es bei einer *degenerierte* Lösung ein „Plateau“ richtiger Punkte oder Zahlen.
- 17 Man beachte, dass Stopp-Codons die am seltensten translatierten DNA-Triplets sind und damit hinter den 20 Aminosäuren auf den 21. Platz fallen.
- 18 In seltenen Fällen können auch einige andere Codons als Startsignale für die Proteintranslation in manchen Organismen fungieren. Die Codons „CTG“ und „GTG“ können manchmal als Start-Codons verwendet werden.
- 19 Lediglich ein „ATG“-Codon in der Milchstraßen-DNA-Sequenz wurde nicht supercodiert, da es an einer einmaligen Erkennungsstelle für ein Restriktionsenzym liegt. Dieses einzige verbleibende ATG-Codon ist von nahen Stopp-Codons flankiert, so dass nur wenige Aminosäuren translatiert werden können.
- 20 Biologische Analogien wurden für die 365-Tage-Periode des Jahres gefunden, die die Umlaufbahn der Erde um die Sonne beschreiben; für die Monats-Periode mit 27,3 Tagen, die der Periode einer Umlaufbahn des Mondes um die Erde entspricht; für den 24-Stunden-Tageszyklus (die Periode der Rotation der Erde um ihre eigene Achse) sowie für die circadianischen Rhythmen, die in Wirklichkeit jahreszeitliche Abfolgen von Licht und Dunkelheit sind, die je nach der Neigung der Erdachse variieren. Obwohl einige Organismen diese Perioden intensiver als andere manifestieren, ist es wahrscheinlich gerechtfertigt anzunehmen, dass sich astronomische Perioden in der „biologischen Uhr“ aller Lebewesen widerspiegeln.
- 21 6.670×10^{-8} dyne $\text{cm}^2 / \text{gm}^2$