



Anlage 1

(zu § 11 Abs. 1 Satz 2 PatV)

Standards für die Einreichung von Sequenzprotokollen

Dienststelle München

Dienststelle Jena

Informations- und Dienstleistungszentrum Berlin

Postanschrift

80297 München

07738 Jena

10958 Berlin

Telefax

+49 89 2195-2221

+49 3641 40-5690

+49 30 25992-404

Telefon

Zentraler Kundenservice:

+49 89 2195-1000

Zahlungsempfänger: Bundeskasse Halle/DPMA

IBAN: DE84 7000 0000 0070 0010 54, BIC (SWIFT-Code): MARKDEF1700

Anschrift der Bank: Bundesbankfiliale München, Leopoldstr. 234, 80807 München

Internet:

<https://www.dpma.de>

Inhaltsverzeichnis

Definitionen	3
Sequenzprotokoll.....	3
Nucleotidsequenzen.....	4
Zu verwendende Symbole.....	4
Zu verwendendes Format	4
Aminosäuresequenzen.....	4
Zu verwendende Symbole.....	4
Zu verwendendes Format	4
Sonstige verfügbare Angaben im Sequenzprotokoll.....	5
Obligatorische Datenelemente	5
Fakultative Datenelemente.....	5
Angabe von Merkmalen	6
Freier Text.....	6
Nachgereichtes Sequenzprotokoll.....	6
Sequenzprotokoll in maschinenlesbarer Form	6
Zulässige numerische Kennzahlen.....	7
Tabelle 1 Liste der Nucleotide	9
Tabelle 2 Liste der modifizierten Nucleotide.....	9
Tabelle 3 Liste der Aminosäuren.....	10
Tabelle 4 Liste der modifizierten und seltenen Aminosäuren	11
Tabelle 5 Liste der Merkmalschlüssel zu Nucleotidsequenzen	11
Tabelle 6 Liste der Merkmalschlüssel zu Proteinsequenzen	15
Beispiel:.....	17

Definitionen

1. Im Rahmen dieses Standards gelten folgende Definitionen:

- i) unter einem "Sequenzprotokoll" ist ein Teil der Beschreibung der Anmeldung in der eingereichten Fassung oder ein zur Anmeldung nachgereichtes Schriftstück zu verstehen, das die Nucleotid- und/oder Aminosäuresequenzen im Einzelnen offenbart und sonstige verfügbare Angaben enthält.
- ii) In das Protokoll dürfen nur unverzweigte Sequenzen von mindestens 4 Aminosäuren bzw. unverzweigte Sequenzen von mindestens 10 Nucleotiden aufgenommen werden. Verzweigte Sequenzen, Sequenzen mit weniger als 4 genau definierten Nucleotiden oder Aminosäuren sowie Sequenzen mit Nucleotiden oder Aminosäuren, die nicht in Nr. 48, Tabellen 1, 2, 3 und 4 aufgeführt sind, sind ausdrücklich ausgeschlossen.
- iii) Unter "Nucleotiden" sind nur Nucleotide zu verstehen, die mittels der in Nr. 48, Tabelle 1 aufgeführten Symbole wiedergegeben werden können. Modifikationen wie z. B. methylierte Basen können nach der Anleitung in Nr. 48, Tabelle 2 beschrieben werden, sind aber in der Nucleotidsequenz nicht explizit auszuweisen.
- iv) Unter "Aminosäuren" sind die in Nr. 48, Tabelle 3 aufgeführten gängigen L-Aminosäuren aus den natürlich vorkommenden Proteinen zu verstehen. Aminosäuresequenzen, die mindestens eine D-Aminosäure enthalten, fallen nicht unter diese Definition. Aminosäuresequenzen, die posttranslational modifizierte Aminosäuren enthalten, können mittels der in Nr. 48, Tabelle 3 aufgeführten Symbole wie die ursprünglich translatierte Aminosäure beschrieben werden, wobei die modifizierten Positionen wie z. B. Hydroxylierungen oder Glykosylierungen nach der Anleitung in Nr. 48, Tabelle 4 beschrieben werden, diese Modifikationen aber in der Aminosäuresequenz nicht explizit auszuweisen sind. Unter die vorstehende Definition fallen auch Peptide oder Proteine, die anhand der in Nr. 48, Tabelle 3 aufgeführten Symbole sowie einer an anderer Stelle aufgenommenen Beschreibung, die beispielsweise Aufschluss über ungewöhnliche Bindungen, Quervernetzungen (z. B. Disulfidbrücken) und "end caps", Nichtpeptidbindungen usw. gibt, als Sequenz wiedergegeben werden können.
- v) Unter "Sequenzkennzahl" ist eine Zahl zu verstehen, die jeder im Protokoll aufgeführten Sequenz als SEQ ID NO zugewiesen wird.
- vi) Die "numerische Kennzahl" ist eine dreistellige Zahl, die für ein bestimmtes Datenelement steht.
- vii) Unter "sprachneutralem Vokabular" ist ein festes Vokabular zu verstehen, das im Sequenzprotokoll zur Wiedergabe der vom Hersteller einer Sequenzdatenbank vorgeschriebenen wissenschaftlichen Begriffe verwendet wird (dazu gehören wissenschaftliche Namen, nähere Be-

stimmungen und ihre Entsprechungen im festen Vokabular, die Symbole in Nr. 48, Tabellen 1, 2, 3 und 4 und die Merkmalschlüssel in Nr. 48, Tabellen 5 und 6).

- viii) Unter "zuständiger Behörde" ist die Internationale Recherchenbehörde oder die mit der internationalen vorläufigen Prüfung beauftragte Behörde, die die internationale Recherche bzw. die internationale vorläufige Prüfung zu der internationalen Anmeldung durchführt, oder das Bestimmungsamt bzw. ausgewählte Amt zu verstehen, das mit der Bearbeitung der internationalen Anmeldung begonnen hat.

Sequenzprotokoll

2. Das Sequenzprotokoll im Sinne der Nr. 1 i) ist ans Ende der Anmeldung zu setzen, wenn es mit ihr zusammen eingereicht wird. Dieser Teil ist mit "Sequenzprotokoll" oder "Sequence Listing" zu überschreiben, muss mit einer neuen Seite beginnen und sollte gesondert nummeriert werden. Das Sequenzprotokoll ist Bestandteil der Beschreibung; vorbehaltlich Nr. 35 erübrigt es sich deshalb, die Sequenzen in der Beschreibung an anderer Stelle nochmals zu beschreiben.
3. Wird das Sequenzprotokoll im Sinne der Nr. 1 i) nicht zusammen mit der Anmeldung eingereicht, sondern als gesondertes Schriftstück nachgereicht (siehe Nr. 36), so ist es mit der Überschrift "Sequenzprotokoll" oder "Sequence Listing" und einer gesonderten Seitennummerierung zu versehen. Die in der Anmeldung in der eingereichten Fassung gewählte Nummerierung der Sequenzen (siehe Nr. 4) ist auch im nachgereichten Sequenzprotokoll beizubehalten.
4. Jeder Sequenz wird eine eigene Sequenzkennzahl zugeteilt. Die Kennzahlen beginnen mit 1 und setzen sich in aufsteigender Reihenfolge als ganze Zahlen fort. Gibt es zu einer Kennzahl keine Sequenz, so ist am Anfang der auf die SEQ ID NO folgenden Zeile unter der numerischen Kennzahl <400> der Code 000 anzugeben. Unter der numerischen Kennzahl <160> ist die Gesamtzahl der SEQ ID NOs anzugeben, und zwar unabhängig davon, ob im Anschluss an eine SEQ ID NO eine Sequenzkennzahl oder der Code 000 folgt.
5. In der Beschreibung, in den Ansprüchen und in den Zeichnungen der Anmeldung ist auf die im Sequenzprotokoll dargestellten Sequenzen mit "SEQ ID NO", gefolgt von der betreffenden Kennzahl, zu verweisen.
6. Für die Darstellung von Nucleotid- und Aminosäuresequenzen ist zumindest eine der folgenden drei Möglichkeiten zu wählen:
 - i) nur Nucleotidsequenz,
 - ii) nur Aminosäuresequenz,
 - iii) Nucleotidsequenz zusammen mit der entsprechenden Aminosäuresequenz.

Bei einer Offenbarung im unter Ziffer iii) genannten Format muss die Aminosäuresequenz als solche im Sequenzprotokoll mit einer eigenen Sequenzkennzahl gesondert offenbart werden.

Nucleotidsequenzen

Zu verwendende Symbole

7. Nucleotidsequenzen sind nur anhand eines Einzelstrangs in Richtung vom 5'-Ende zum 3'-Ende von links nach rechts wiederzugeben. Die Begriffe 3' und 5' werden in der Sequenz nicht dargestellt.
8. Die Basen einer Nucleotidsequenz sind anhand der einbuchstabigen Codes für Nucleotidsequenzzeichen darzustellen. Es dürfen nur die in Nr. 48, Tabelle 1 aufgeführten Kleinbuchstaben verwendet werden.
9. Modifizierte Basen sind in der Sequenz selbst wie die entsprechenden unmodifizierten Basen oder mit "n" wiederzugeben, wenn die modifizierte Base zu den in Nr. 48, Tabelle 2 aufgeführten gehört; die Modifikation ist im Merkmalsteil des Sequenzprotokolls anhand der Codes in Nr. 48, Tabelle 2 näher zu beschreiben. Diese Codes dürfen in der Beschreibung oder im Merkmalsteil des Sequenzprotokolls, nicht jedoch in der Sequenz selbst verwendet werden (siehe auch Nr. 31). Das Symbol "n" steht immer nur für ein einziges unbekanntes oder modifiziertes Nucleotid.

Zu verwendendes Format

10. Bei Nucleotidsequenzen sind höchstens 60 Basen pro Zeile – mit einem Leerraum zwischen jeder Gruppe von 10 Basen – aufzuführen.
11. Die Basen einer Nucleotidsequenz (einschließlich Intronen) sind jeweils in Zehnergruppen aufzuführen; dies gilt nicht für die nichtcodierenden Teile der Sequenz. Bleiben am Ende nichtcodierender Teile einer Sequenz weniger als 10 Basen übrig, so sind sie zu einer Gruppe zusammenzufassen und durch einen Leerraum von angrenzenden Gruppen zu trennen.
12. Die Basen der codierenden Teile einer Nucleotidsequenz sind als Triplets (Codonen) aufzuführen.
13. Die Zählung der Nucleotidbasen beginnt bei der ersten Base der Sequenz mit 1. Von hier aus ist die gesamte Sequenz in 5'-3'-Richtung fortlaufend durchzuzählen. Am rechten Rand ist jeweils neben der Zeile mit den einbuchstabigen Codes für die Basen die Nummer der letzten Base dieser Zeile anzugeben. Die vorstehend beschriebene Zählweise für Nucleotidsequenzen gilt auch für Nucleotidsequenzen mit ringförmiger Konfiguration, wobei allerdings die Bestimmung des ersten Nucleotids der Sequenz dem Anmelder überlassen bleibt.
14. Eine Nucleotidsequenz, die aus einem oder mehreren nichtbenachbarten Abschnitten einer größeren Sequenz oder aus Abschnitten verschiedener Sequenzen besteht, ist als gesonderte Sequenz mit eigener Sequenzkennzahl zu nummerieren. Sequenzen mit einer oder mehreren Lücken sind als mehrere gesonderte Sequenzen mit eigenen Sequenzkennzahlen zu nummerieren, wobei die Zahl der gesonderten Sequenzen der Zahl der jeweils zusammenhängenden Sequenzdatenreihen entspricht.

Aminosäuresequenzen

Zu verwendende Symbole

15. Die Aminosäuren einer Protein- oder Peptidsequenz sind in Richtung von der Amino- zur Carboxylgruppe von links nach rechts aufzuführen, wobei die Amino- und Carboxylgruppen in der Sequenz nicht darzustellen sind.
16. Die Aminosäuren sind anhand des dreibuchstabigen Codes mit großem Anfangsbuchstaben entsprechend der Liste in Nr. 48, Tabelle 3 darzustellen. Eine Aminosäuresequenz, die einen Leerraum oder interne Terminatorsymbole (z.B. "Ter", "*" oder ".") enthält, ist nicht als eine einzige Aminosäuresequenz, sondern als getrennte Aminosäuresequenzen darzustellen (siehe Nr. 21).
17. Modifizierte und seltene Aminosäuren sind in der Sequenz selbst wie die entsprechenden unmodifizierten Aminosäuren oder mit "Xaa" wiederzugeben, wenn sie zu den in Nr. 48, Tabelle 4 aufgeführten gehören und die Modifikation im Merkmalsteil des Sequenzprotokolls anhand der Codes in Nr. 48, Tabelle 4 näher beschrieben wird. Diese Codes dürfen in der Beschreibung oder im Merkmalsteil des Sequenzprotokolls, nicht jedoch in der Sequenz selbst verwendet werden (siehe auch Nr. 31). Das Symbol "Xaa" steht immer nur für eine einzige unbekannte oder modifizierte Aminosäure.

Zu verwendendes Format

18. Bei Protein- oder Peptidsequenzen sind höchstens 16 Aminosäuren pro Zeile mit einem Leerraum zwischen den einzelnen Aminosäuren aufzuführen.
19. Die den Codonen der codierenden Teile einer Nucleotidsequenz entsprechenden Aminosäuren sind unmittelbar unter den jeweiligen Codonen anzugeben. Wird ein Codon durch ein Intron aufgespalten, so ist das Aminosäuresymbol unter dem Teil des Codons anzugeben, der zwei Nucleotide enthält.
20. Die Zählung der Aminosäuren beginnt bei der ersten Aminosäure der Sequenz mit 1. Fakultativ können die dem reifen Protein vorausgehenden Aminosäuren wie beispielsweise Präsequenzen, Prosequenzen, Präprosequenzen und Signalsequenzen, soweit vorhanden, mit negativem Vorzeichen nummeriert werden, wobei die Rückwärtszählung mit der Aminosäure vor Nummer 1 beginnt. Null (0) wird nicht verwendet, wenn Aminosäuren zur Abgrenzung gegen das reife Protein mit negativem Vorzeichen nummeriert werden. Die Nummern sind im Abstand von jeweils 5 Aminosäuren unter der Sequenz anzugeben. Die vorstehend beschriebene Zählweise für Aminosäuresequenzen gilt auch für Aminosäuresequenzen mit ringförmiger Konfiguration, wobei allerdings die Bestimmung der ersten Aminosäure der Sequenz dem Anmelder überlassen bleibt.
21. Eine Aminosäuresequenz, die aus einem oder mehreren nichtbenachbarten Abschnitten einer größeren Sequenz oder aus Abschnitten verschiedener Sequenzen besteht, ist als gesonderte Sequenz mit eigener Sequenzkennzahl zu nummerieren. Sequenzen

zen mit einer oder mehreren Lücken sind als mehrere gesonderte Sequenzen mit eigenen Sequenzkennzahlen zu nummerieren, wobei die Zahl der gesonderten Sequenzen der Zahl der jeweils zusammenhängenden Sequenzdatenreihen entspricht.

Sonstige verfügbare Angaben im Sequenzprotokoll

22. Die Angaben sind im Sequenzprotokoll in der Reihenfolge anzugeben, in der sie in der Liste der numerischen Kennzahlen der Datenelemente in Nr. 47 aufgeführt sind.
23. Für die Angaben im Sequenzprotokoll sind nur die in Nr. 47 aufgeführten numerischen Kennzahlen, nicht aber die dazugehörigen Beschreibungen zu verwenden. Die Angaben müssen unmittelbar auf die numerische Kennzahl folgen; im Sequenzprotokoll brauchen nur diejenigen numerischen Kennzahlen angegeben zu werden, zu denen auch Angaben vorliegen. Die einzigen beiden Ausnahmen zu dieser Vorschrift bilden die numerischen Kennzahlen <220> und <300>, die für die Rubrik "Merkmal" bzw. "Veröffentlichungsangaben" stehen und mit den Angaben unter den numerischen Kennzahlen <221> bis <223> bzw. <301> bis <313> zusammenhängen. Werden unter diesen numerischen Kennzahlen im Sequenzprotokoll Angaben zu den Merkmalen oder zur Veröffentlichung gemacht, so sollte auch die numerische Kennzahl <220> bzw. <300> aufgeführt, die dazugehörige Rubrik aber nicht ausgefüllt werden. Generell sollte zwischen den numerischen Kennzahlen eine Leerzeile eingefügt werden, wenn sich die an erster oder zweiter Position der numerischen Kennzahl stehende Ziffer ändert. Eine Ausnahme von dieser allgemeinen Regel bildet die numerische Kennzahl <310>, der keine Leerzeile vorausgehen darf. Auch vor jeder Wiederholung der numerischen Kennzahl ist eine Leerzeile einzufügen.

Obligatorische Datenelemente

24. In das Sequenzprotokoll sind außerdem vor der eigentlichen Nucleotid- und/oder Aminosäuresequenz die folgenden in Nr. 47 definierten Angaben (obligatorische Datenelemente) aufzunehmen:

<110>	Name des Anmelders
<120>	Bezeichnung der Erfindung
<160>	Anzahl der SEQ ID NOs
<210>	SEQ ID NO: x
<211>	Länge
<212>	Art
<213>	Organismus
<400>	Sequenz

Ist der Name des Anmelders (numerische Kennzahl <110>) nicht in Buchstaben des lateinischen Alphabets geschrieben, so ist er – im Wege der Transliteration oder der Übersetzung ins Englische – auch in lateinischen Buchstaben anzugeben.

Die Datenelemente mit Ausnahme der Angaben unter den numerischen Kennzahlen <110>, <120> und <160> sind für jede im Sequenzprotokoll aufgeführte Sequenz zu wiederholen. Gibt es zu einer Sequenzkennzahl keine Sequenz, so müssen nur die Datenelemente unter den numerischen Kennzahlen <210> und <400> angegeben werden (siehe Nr. 4 und SEQ ID NO: 4 in dem am Ende dieses Standards enthaltenen Beispiel).

25. In Sequenzprotokolle, die zusammen mit der dazugehörigen Anmeldung oder vor Vergabe einer Anmelde-nummer eingereicht werden, ist neben den unter Nr. 24 genannten Datenelementen auch das Folgende aufzunehmen:

<130>	Bezugsnummer
-------	--------------

26. In Sequenzprotokolle, die auf Aufforderung einer zuständigen Behörde oder nach Vergabe einer Anmelde-nummer eingereicht werden, sind neben den unter Nr. 24 genannten Datenelementen auch die Folgenden aufzunehmen:

<140>	Vorliegende Patentanmeldung
<141>	Anmeldetag der vorliegenden Anmeldung

27. In Sequenzprotokolle, die zu einer Anmeldung eingereicht werden, die die Priorität einer früheren Anmeldung in Anspruch nimmt, sind neben den unter Nr. 24 genannten Datenelementen auch die Folgenden aufzunehmen:

<150>	Frühere Patentanmeldung
<151>	Anmeldetag der früheren Anmeldung

28. Wird in der Sequenz "n", "Xaa" oder eine modifizierte Base oder modifizierte/seltene L-Aminosäure aufgeführt, so müssen die folgenden Datenelemente angegeben werden:

<220>	Merkmal
<221>	Name/Schlüssel
<222>	Lage
<223>	Sonstige Informationen

29. Ist der Organismus (numerische Kennzahl <213>) eine "künstliche Sequenz" oder "unbekannt", so müssen die folgenden Datenelemente angegeben werden:

<220>	Merkmal
<223>	Sonstige Angaben

Fakultative Datenelemente

30. Alle in Nr. 47 definierten Datenelemente, die unter Nr. 24 bis 29 nicht erwähnt sind, sind fakultativ (fakultative Datenelemente).

Angabe von Merkmalen

31. Merkmale, die (unter der numerischen Kennzahl <220>) zu einer Sequenz angegeben werden, sind durch die in Nr. 48, Tabellen 5 und 6¹⁾ aufgeführten "Merkmalschlüssel" zu beschreiben.

Freier Text

32. Unter "freiem Text" ist eine verbale Beschreibung der Eigenschaften der Sequenz ohne Verwendung des sprachneutralen Vokabulars im Sinne der Nr. 1 vii) unter der numerischen Kennzahl <223> (Sonstige Angaben) zu verstehen.
33. Der freie Text sollte sich auf einige kurze, für das Verständnis der Sequenz unbedingt notwendige Begriffe beschränken. Er sollte für jedes Datenelement nicht länger als 4 Zeilen sein und höchstens 65 Buchstaben pro Zeile umfassen. Alle weiteren Angaben sind in den Hauptteil der Beschreibung in der dort verwendeten Sprache aufzunehmen.
34. Freier Text kann in deutscher oder in englischer Sprache abgefasst sein.
35. Enthält das Sequenzprotokoll, das Bestandteil der Beschreibung ist, freien Text, so ist dieser im Hauptteil der Beschreibung in der dort verwendeten Sprache zu wiederholen. Es wird empfohlen, den in der Sprache des Hauptteils der Beschreibung abgefassten freien Text in einen besonderen Abschnitt der Beschreibung mit der Überschrift "Sequenzprotokoll – freier Text" aufzunehmen.

Nachgereichtes Sequenzprotokoll

36. Ein Sequenzprotokoll, das nicht Teil der eingereichten Anmeldung war, sondern nachträglich eingereicht wurde, darf nicht über den Offenbarungsgehalt der in der Anmeldung angegebenen Sequenzen hinausgehen. Dem nachgereichten Sequenzprotokoll muss eine Erklärung, die dies bestätigt, beigefügt sein. Das heißt, dass ein Sequenzprotokoll, das nachträglich eingereicht wurde, nur die Sequenzen beinhalten darf, die auch in der eingereichten Anmeldung aufgeführt sind.
37. Sequenzprotokolle, die nicht zusammen mit der Anmeldung eingereicht worden sind, sind nicht Teil der Offenbarung der Erfindung. Gemäß § 11 Abs. 3 besteht die Möglichkeit, Sequenzprotokolle, die zusammen mit der Anmeldung eingereicht wurden, im Wege der Mängelbeseitigung zu berichtigen.

Sequenzprotokoll in maschinenlesbarer Form

38. Das in der Anmeldung enthaltene Sequenzprotokoll ist zusätzlich auch in maschinenlesbarer Form einzureichen.
39. Das zusätzlich eingereichte maschinenlesbare Sequenzprotokoll muss mit dem geschriebenen Protokoll identisch sein und ist zusammen mit einer Erklärung folgenden Wortlauts einzureichen: "Die in maschinenlesbarer Form aufgezeichneten Angaben sind mit dem geschriebenen Sequenzprotokoll identisch."

40. Das gesamte ausdrückbare Exemplar des Sequenzprotokolls muss in einer einzigen Datei enthalten sein, die nach Möglichkeit auf einer einzigen Diskette oder einem sonstigen vom Deutschen Patent- und Markenamt akzeptierten elektronischen Datenträger aufzeichnet sein soll. Diese Datei ist mittels der IBM²⁾-Codetabellen (Code Page) 437, 932³⁾ oder einer kompatiblen Codetabelle zu codieren. Eine Codetabelle, wie sie z. B. für japanische, chinesische, kyrillische, arabische, griechische oder hebräische Schriftzeichen benötigt wird, gilt als kompatibel, wenn sie das lateinische Alphabet und die arabischen Ziffern denselben Hexadezimalpositionen zuordnet, wie die genannten Codetabellen.
41. Folgende Medientypen und Formatierungen für zusätzlich in maschinenlesbarer Form eingereichte Sequenzprotokolle werden akzeptiert:

Physikalisches Medium	Typ	Formatierung
CD-R	120 mm Recordable Disk	ISO 9660
DVD-R	120 mm DVD-Recordable Disk (4,7 GB)	konform zu ISO 9660 oder OSTA UDF (1.02 oder höher)
DVD+R	120 mm DVD-Recordable Disk (4,7 GB)	konform zu ISO 9660 oder OSTA UDF (1.02 oder höher)

42. Die maschinenlesbare Fassung kann mit beliebigen Mitteln erstellt werden. Sie muss aber den vom Deutschen Patent- und Markenamt angegebenen Formaten entsprechen. Vorzugsweise sollte zum Erstellen eine dafür vorgesehene spezielle Software, wie z. B. PatentIn verwendet werden.
43. Bei Verwendung von physikalischen Datenträgern ist eine Datenkompression erlaubt, sofern die komprimierte Datei in einem selbstextrahierenden Format erstellt worden ist, das sich auf einem vom Deutschen Patent- und Markenamt angegebenen Betriebssystem (MS Windows) selbst dekomprimiert. Ebenso können inhaltlich zusammengehörige Dateien in einem sich nicht selbstextrahierenden Format komprimiert sein, wenn die Archivdatei im Zip-Format der Version vom 13. Juli 1998 vorliegt und weder andere Zip-Archive oder eine Verzeichnisstruktur beinhaltet.
44. Auf dem physikalischen Datenträger ist ein Etikett fest anzubringen, auf dem von Hand oder mit der Maschine in Blockschrift der Name des Anmelders, die Bezeichnung der Erfindung, eine Bezugsnummer, der Zeitpunkt der Aufzeichnung der Daten und das Computerbetriebssystem eingetragen sind.
45. Wird der physikalische Datenträger erst nach dem Tag der Anmeldung eingereicht, so sind auf den Etiketten auch der Anmeldetag und die Anmeldenummer

¹⁾ Diese Tabellen enthalten Auszüge aus den Merkmalstabellen "DDBJ/EMBL/Genbank Feature Table" (Nucleotidsequenzen) und "SWISS PROT Feature Table" (Aminosäuresequenzen).

²⁾ IBM ist eine für die International Business Machine Corporation, Vereinigte Staaten von Amerika, eingetragene Marke.

³⁾ Die genannten Codetabellen gelten für Personal Computer als de facto Standard.

anzugeben. Korrekturen oder Änderungen zum Sequenzprotokoll sind sowohl schriftlich als auch in maschinenlesbarer Form einzureichen.

46. Jede Korrektur der ausgedruckten Version des Sequenzprotokolls, die auf Grund der PCT Regel 13ter.1(a)(i) oder 26.3 eingereicht, jede Verbesserung eines offensichtlichen Fehlers in der ausgedruckten Version, die auf Basis der PCT Regel 91 eingereicht und jeder Zusatz, der nach Artikel 34 PCT in die ausgedruckte Version des Sequenzprotokolls eingebunden wurde, muss zusätzlich in einer verbesserten und mit den Zusätzen versehenen Version des Sequenzprotokolls in maschinenlesbarer Form eingereicht werden.

47. Numerische Kennzahlen

In Sequenzprotokollen, die zu Anmeldungen eingereicht werden, dürfen nur die nachstehenden numeri-

schen Kennzahlen verwendet werden. Die Überschriften der nachstehenden Datenelementrubriken dürfen in den Sequenzprotokollen nicht erscheinen.

Die numerischen Kennzahlen der obligatorischen Datenelemente, d. h. der Datenelemente, die in alle Sequenzprotokolle aufgenommen werden müssen (siehe Nr. 24 dieses Standards: Kennziffern 110, 120, 160, 210, 211, 212, 213 und 400), und die numerischen Kennzahlen der Datenelemente, die in den in diesem Standard genannten Fällen aufgenommen werden müssen (siehe Nr. 25, 26, 27, 28 und 29 dieses Standards: Kennzahlen 130, 140, 141, 150 und 151 sowie 220 bis 223), sind mit dem Buchstaben "O" gekennzeichnet.

Die numerischen Kennzahlen der fakultativen Datenelemente (siehe Nr. 30 dieses Standards) sind mit dem Buchstaben "F" gekennzeichnet.

Zulässige numerische Kennzahlen			
Numerische Kennzahl	Numerische Kennzahl Beschreibung	Obligatorisch (O) oder fakultativ (F)	Bemerkungen
<110>	Name des Anmelders	O	Ist der Name des Anmelders nicht in lateinischen Buchstaben geschrieben, so muss er - im Wege der Transliteration oder der Übersetzung ins Englische - auch in lateinischen Buchstaben angegeben werden
<120>	Bezeichnung der Erfindung	O	
<130>	Bezugsnummer	O in den Fällen nach Nr. 25 Standard	Siehe Nr. 25 Standard
<140>	Vorliegende Patentanmeldung	O in den Fällen nach Nr. 26 Standard	Siehe Nr. 26 Standard; die vorliegende Patentanmeldung ist zu kennzeichnen durch den zweibuchstabigen Code nach dem WIPO-Standard ST. 3, gefolgt von der Anmeldenummer (in dem Format, das von der Behörde für gewerblichen Rechtsschutz verwendet wird, bei der diese Patentanmeldung eingereicht wird) oder - bei internationalen Anmeldungen - von der internationalen Anmeldenummer
<141>	Anmeldetag der vorliegenden Anmeldung	O in den Fällen nach Nr. 26 Standard	Siehe Nr. 26 Standard; das Datum ist entsprechend dem WIPO-Standard ST.2 anzugeben (CCYY MM DD)
<150>	Frühere Patentanmeldung	O in den Fällen nach Nr. 27 Standard	Siehe Nr. 27 Standard; die frühere Patentanmeldung ist zu kennzeichnen durch den zweibuchstabigen Code entsprechend dem WIPO-Standard ST.3, gefolgt von der Anmeldenummer (in dem Format, das von der Behörde für gewerblichen Rechtsschutz verwendet wird, bei der die frühere Patentanmeldung eingereicht wurde) oder - bei internationalen Anmeldungen - von der internationalen Anmeldenummer
<151>	Anmeldetag der früheren Anmeldung	O in den Fällen nach Nr. 27 Standard	Siehe Nr. 27 Standard; das Datum ist entsprechend dem WIPO-Standard ST.2 anzugeben (CCYY MM DD)
<160>	Anzahl der SEQ ID NOs	O	
<170>	Software	F	
<210>	Angaben zu SEQ ID NO: x	O	Anzugeben ist eine ganze Zahl, die die SEQ ID NO darstellt
<211>	Länge	O	Sequenzlänge, ausgedrückt als Anzahl der Basen oder Aminosäuren

Zulässige numerische Kennzahlen			
Numerische Kennzahl	Numerische Kennzahl Beschreibung	Obligatorisch (O) oder fakultativ (F)	Bemerkungen
<212>	Art	O	Art des in SEQ ID NO: x sequenzierten Moleküls, und zwar entweder DNA, RNA oder PRT; enthält eine Nucleotidsequenz sowohl DNA- als auch RNA-Fragmente, so ist "DNA" anzugeben; zusätzlich ist das kombinierte DNA/RNA-Molekül im Merkmalsteil unter <220> bis <223> näher zu beschreiben
<213>	Organismus	O	Gattung/Art (d.h. wissenschaftlicher Name), "künstliche Sequenz" oder "unbekannt"
<220>	Merkmal	O in den Fällen nach Nr. 28 und 29 Standard	Freilassen; siehe Nr. 28 und 29 Standard; Beschreibung biologisch signifikanter Stellen in der Sequenz gemäß SEQ ID NO:x (kann je nach der Zahl der angegebenen Merkmale mehrmals vorkommen)
<221>	Name/Schlüssel	O in den Fällen nach Nr. 28 Standard	Siehe Nr. 28 Standard; es dürfen nur die in Nr. 48, Tabelle 5 oder 6 beschriebenen Schlüssel verwendet werden
<222>	Lage	O in den Fällen nach Nr. 28 Standard	Siehe Nr. 28 Standard; <ul style="list-style-type: none"> - von (Nummer der ersten Base/Aminosäure des Merkmals) - bis (Nummer der letzten Base/Aminosäure des Merkmals) - Basen (Ziffern verweisen auf die Positionen der Basen in einer Nucleotidsequenz) - Aminosäuren (Ziffern verweisen auf die Positionen der Aminosäurereste in einer Aminosäuresequenz) - Angabe, ob sich das Merkmal auf dem zum Strang des Sequenzprotokolls komplementären Strang befindet
<223>	Sonstige Angaben	O in den Fällen nach Nr. 28 und 29 Standard	Siehe Nr. 28 und 29 Standard; sonstige relevante Angaben, wobei sprachneutrales Vokabular oder freier Text (in deutscher oder in englischer Sprache) zu verwenden ist; freier Text ist im Hauptteil der Beschreibung in der dort verwendeten Sprache zu wiederholen (siehe Nr. 35 Standard); enthält die Sequenz einer der in Nr. 48, Tabellen 2 und 4 aufgeführten modifizierten Basen oder modifizierten/seltenen L-Aminosäuren, so ist für diese Base oder Aminosäure das dazugehörige Symbol aus Nr. 48, Tabellen 2 und 4 zu verwenden.
<300>	Veröffentlichungsangaben	F	Freilassen; dieser Abschnitt ist für jede relevante Veröffentlichung zu wiederholen
<301>	Verfasser	F	
<302>	Titel	F	Titel der Veröffentlichung
<303>	Zeitschrift	F	Name der Zeitschrift, in der die Daten veröffentlicht wurden
<304>	Band	F	Band der Zeitschrift, in dem die Daten veröffentlicht wurden
<305>	Heft	F	Nummer des Hefts der Zeitschrift, in dem die Daten veröffentlicht wurden
<306>	Seiten	F	Seiten der Zeitschrift, auf denen die Daten veröffentlicht wurden
<307>	Datum	F	Datum der Zeitschrift, an dem die Daten veröffentlicht wurden; Angabe nach Möglichkeit entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)
<308>	Eingangsnummer in der Datenbank	F	Von der Datenbank zugeteilte Eingangsnummer einschließlich Datenbankbezeichnung
<309>	Datenbank-Eingabedatum	F	Datum der Eingabe in die Datenbank; Angabe entsprechend dem WIPO-Standard ST. 2 (CCYY MM DD)

Zulässige numerische Kennzahlen			
Numerische Kennzahl	Numerische Kennzahl Beschreibung	Obligatorisch (O) oder fakultativ (F)	Bemerkungen
<310>	Dokumentennummer	F	Nummer des Dokuments, nur bei Patentdokumenten; die vollständige Nummer hat nacheinander Folgendes zu enthalten: den zweibuchstabigen Code entsprechend dem WIPO-Standard ST. 3, die Veröffentlichungsnummer entsprechend dem WIPO-Standard ST. 6 und den Code für die Dokumentenart nach dem WIPO-Standard ST. 16
<311>	Anmeldetag	F	Anmeldetag des Dokuments, nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<312>	Veröffentlichungsdatum	F	Datum der Veröffentlichung des Dokuments; nur bei Patentdokumenten; Angabe entsprechend WIPO-Standard ST. 2 (CCYY MM DD)
<313>	Relevante Reste in SEQ ID NO: x von bis	F	
<400>	Sequenz	O	SEQ ID NO: x sollte in der der Sequenz vorausgehenden Zeile hinter der numerischen Kennzahl stehen (siehe Beispiel)

48. Symbole für Nucleotide und Aminosäuren und Merkmalstabellen

Tabelle 1 Liste der Nucleotide		
Symbol	Bedeutung	Ableitung der Bezeichnung
a	a	<u>A</u> denin
g	g	<u>G</u> uanin
c	c	<u>C</u> ytosin
t	t	<u>T</u> hymine
u	u	<u>U</u> racil
r	g oder a	P <u>u</u> rin
y	t/u oder c	P <u>y</u> rimidin
m	a oder c	<u>A</u> mino
k	g oder t/u	<u>K</u> eto
s	g oder c	<u>s</u> tarke Bindungen, 3-H-Brücken
w	a oder t/u	<u>w</u> schwache (e: <u>w</u> weak) Bindungen, 2 H-Brücken
b	g oder c oder t/u	nicht a
d	a oder g oder t/u	nicht c
h	a oder c oder t/u	nicht g
v	a oder g oder c	nicht t, nicht u
n	a oder g oder c oder t/u, unbekannt oder sonstige	beliebig (e: <u>a</u> ny)

Tabelle 2 Liste der modifizierten Nucleotide	
Symbol	Bedeutung
ac4c	4-Acetylcytidin
chm5u	5-(Carboxyhydroxymethyl)uridin
cm	2'-O-Methylcytidin
cmnm5s2u	5-Carboxymethylaminomethyl-2-thiouridin
cmnm5u	5-Carboxymethylaminomethyluridin
d	Dihydrouridin
fm	2'-O-Methylpseudouridin
gal q	beta,D-Galactosylqueuosin
gm	2'-O-Methylguanosin
i	Inosin
i6a	N6-Isopentenyladenosin
m1a	1-Methyladenosin
m1f	1-Methylpseudouridin
m1g	1-Methylguanosin
m1i	1-Methylinosin
m22g	2,2-Dimethylguanosin
m2a	2-Methyladenosin
m2g	2-Methylguanosin
m3c	3-Methylcytidin

Tabelle 2 Liste der modifizierten Nucleotide	
Symbol	Bedeutung
m5c	5-Methylcytidin
m6a	N6-Methyladenosin
m7g	7-Methylguanosin
mam5u	5-Methylaminomethyluridin
mam5s2u	5-Methoxyaminomethyl-2-thiouridin
man q	beta,D-Mannosylqueuosin
mcm5s2u	5-Methoxycarbonylmethyl-2-thiouridin
mcm5u	5-Methoxycarbonylmethyluridin
mo5u	5-Methoxyuridin
ms2i6a	2-Methylthio-N6-isopentenyladenosin
ms2t6a	N-((9-beta-D-Ribofuranosyl-2-methylthio-purin-6-yl)carbamoyl)threonin
mt6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)N-methylcarbamoyl)threonin
mv	Uridin-5-oxeyessigsäuremethylester
o5u	Uridin-5-oxeyessigsäure(v)
osyw	Wybutoxosin
p	Pseudouridin
q	Queuosin
s2c	2-Thiocytidin
s2t	5-Methyl-2-thiouridin
s2u	2-Thiouridin
s4u	4-Thiouridin
t	5-Methyluridin
t6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)carbamoyl)threonin
tm	2'-O-Methyl-5-methyluridin
um	2'-O-Methyluridin
yw	Wybutosin
x	3-(3-Amino-3-carboxypropyl)uridin,(acp3)u

Tabelle 3 Liste der Aminosäuren	
Symbol	Bedeutung
Ala	Alanin
Cys	Cystein
Asp	Asparaginsäure
Glu	Glutaminsäure
Phe	Phenylalanin
Gly	Glycin
His	Histidin
Ile	Isoleucin
Lys	Lysin
Leu	Leucin
Met	Methionin
Asn	Asparagin
Pro	Prolin
Gln	Glutamin
Arg	Arginin
Ser	Serin
Thr	Threonin
Val	Valin
Trp	Tryptophan
Tyr	Tyrosin
Asx	Asp oder Asn
Glx	Glu oder Gln
Xaa	Unbekannt oder sonstige

Tabelle 4 Liste der modifizierten und seltenen Aminosäuren	
Symbol	Bedeutung
Aad	2-Aminoadipinsäure
bAad	3-Aminoadipinsäure
bAla	beta-Alanin, beta-Aminopropionsäure
Abu	2-Aminobuttersäure
4Abu	4-Aminobuttersäure, Piperidinsäure
Acp	6-Aminocapronsäure
Ahe	2-Aminoheptansäure
Aib	2-Aminoisobuttersäure
bAib	3-Aminoisobuttersäure
Apm	2-Aminopimelinsäure
Dbu	2,4-Diaminobuttersäure
Des	Desmosin
Dpm	2,2'-Diaminopimelinsäure
Dpr	2,3-Diaminopropionsäure
EtGly	N-Ethylglycin

Tabelle 4 Liste der modifizierten und seltenen Aminosäuren	
Symbol	Bedeutung
EtAsn	N-Ethylasparagin
Hyl	Hydroxylysin
aHyl	allo-Hydroxylysin
3Hyp	3-Hydroxyprolin
4Hyp	4-Hydroxyprolin
Ide	Isodesmosin
alle	allo-Isoleucin
MeGly	N-Methylglycin, Sarkosin
Melle	N-Methylisoleucin
MeLys	6-N-Methyllysin
MeVal	N-Methylvalin
Nva	Norvalin
Nle	Norleucin
Orn	Ornithin

Tabelle 5 Liste der Merkmalschlüssel zu Nucleotidsequenzen	
Schlüssel	Beschreibung
allele	ein verwandtes Individuum oder ein verwandter Stamm enthält stabile alternative Formen desselben Gens und unterscheidet sich an dieser (und vielleicht an anderer) Stelle von der vorliegenden Sequenz
attenuator	1. Region einer DNA, in der die Beendigung der Transkription reguliert wird und die Expression einiger bakterieller Operons gesteuert wird 2. zwischen dem Promotor und dem ersten Strukturgen liegender Sequenzabschnitt, der eine partielle Beendigung der Transkription bewirkt
C_region	konstante Region der leichten und schweren Immunglobulinketten und der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; enthält je nach Kette ein oder mehrere Exons
CAAT_signal	CAAT-Box; Teil einer konservierten Sequenz, der etwa 75 Basenpaare stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten liegt und an der RNA-Polymerase-Bindung beteiligt sein kann; Konsensussequenz=GG (C oder T) CAATCT
CDS	codierende Sequenz; Sequenz von Nucleotiden, die mit der Sequenz der Aminosäuren in einem Protein übereinstimmt (beinhaltet Stopcodon); Merkmal schließt eine mögliche Translation der Aminosäure ein
konflikt	unabhängige Bestimmungen "derselben" Sequenz unterscheiden sich an dieser Stelle oder in dieser Region voneinander
D-loop	D-Schleife; Region innerhalb der mitochondrialen DNA, in der sich ein kürzeres RNA-Stück mit einem Strang der doppelsträngigen DNA paart und dabei den ursprünglichen Schwesterstrang in dieser Region verdrängt; dient auch zur Beschreibung der Verdrängung einer Region des einen Stranges eines DNA-Doppelstrangs durch eine einzelsträngige Nucleinsäure bei der durch ein recA-Protein ausgelösten Reaktion
D-segment	Diversity-Region der schweren Kette von Immunglobulin und der Beta-Kette eines T-Zell-Rezeptors
enhancer	eine als Cis-Element wirkende Sequenz, die die Aktivität (einiger) eukaryontischer Promotoren verstärkt und in beliebiger Richtung und Position zum Promotor (stromaufwärts oder -abwärts) funktioniert
exon	Region des Genoms, die für einen Teil der gespleißten mRNA codiert; kann 5'UTR, alle CDSs und 3'UTR enthalten

Tabelle 5
Liste der Merkmalschlüssel zu Nucleotidsequenzen

Schlüssel	Beschreibung
GC_signal	GC-Box; eine konservierte, GC-reiche Region stromaufwärts vom Startpunkt der eukaryontischen Transkriptionseinheiten, die in mehreren Kopien und in beiden Richtungen vorkommen kann; Konsensussequenz = GGGCGG
gene	biologisch signifikante Region, codierende Nucleinsäure
iDNA	intervenierende DNA; DNA, die durch verschiedene Arten der Rekombination eliminiert wird
intron	DNA-Abschnitt, der transkribiert, aber beim Zusammenspleißen der ihn umgebenden Sequenzen (Exons) aus dem Transkript wieder herausgeschnitten wird
J_segment	J-Kette (Verbindungskette) zwischen den leichten und den schweren Immunglobulinketten und den Alpha-, Beta- und Gamma-Ketten der T-Zell-Rezeptoren
LTR	lange, sich an den beiden Enden einer gegebenen Sequenz direkt wiederholende Sequenz, wie sie für Retroviren typisch ist
mat_peptide	für ein reifes Peptid oder Protein codierende Sequenz; Sequenz, die für das reife oder endgültige Peptid- oder Proteinprodukt im Anschluss an eine posttranslationale Modifizierung codiert; schließt im Gegensatz zur entsprechenden CDS das Stopcodon nicht ein
misc_binding	Stelle in einer Nucleinsäure, die einen anderen Teil, der nicht durch einen anderen Bindungsschlüssel (primer_bind oder protein_bind) beschrieben werden kann, kovalent oder nicht kovalent bindet
misc_difference	die Merkmalsequenz unterscheidet sich von der im Eintrag und kann nicht durch einen anderen Unterscheidungsschlüssel (conflict, unsure, old_sequence, mutation, variation, allele bzw. modified_base) beschrieben werden
misc_feature	biologisch signifikante Region, die nicht durch einen anderen Merkmalschlüssel beschrieben werden kann; neues oder seltenes Merkmal
misc_recomb	Stelle, an der ein allgemeiner, ortsspezifischer oder replikativer Rekombinationsvorgang stattfindet, bei dem die DNA-Doppelhelix aufgebrochen und wieder zusammengefügt wird, und die nicht durch andere Rekombinationsschlüssel (iDNA oder Virion) oder den betreffenden Herkunftsschlüssel (/insertions_seq, /transposon, /proviral) beschrieben werden kann
misc_RNA	Transkript oder RNA-Produkt, das nicht durch andere RNA-Schlüssel (prim_transcript, precursor_RNA, mRNA, 5'clip, 3'clip, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron polyA_site, rRNA, tRNA, scRNA oder snRNA) beschrieben werden kann
misc_signal	Region, die ein Signal enthält, das die Genfunktion oder -expression steuert oder ändert, und die nicht durch andere Signalschlüssel (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator oder rep_origin) beschrieben werden kann
misc_structure	Sekundär-, Tertiär- oder sonstige Struktur oder Konformation, die nicht durch andere Strukturschlüssel (stem_loop oder D-loop) beschrieben werden kann
modified_base	das angegebene Nucleotid ist ein modifiziertes Nucleotid und ist durch das angegebene Molekül (ausgedrückt durch die modifizierte Base) zu ersetzen
mRNA	messenger-RNA (Boten-RNA); enthält eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon) und eine 3'-nichttranslatierte Region (3'UTR)

Tabelle 5 Liste der Merkmalschlüssel zu Nucleotidsequenzen	
Schlüssel	Beschreibung
mutation	ein verwandter Stamm weist an dieser Stelle eine plötzliche, erhebliche Sequenzveränderung auf
N_region	zusätzliche Nucleotide, die zwischen neu geordnete Immunglobulinabschnitte eingefügt werden
old_sequence	die vorliegende Sequenz stellt eine geänderte Version der früher an dieser Stelle befindlichen Sequenz dar
polyA_signal	Erkennungsregion, die zur Endonuclease-Spaltung eines RNA-Transkripts mit anschließender Polyadenylierung nötig ist; Konsensussequenz: AATAAA
polyA_site	Stelle auf einem RNA-Transkript, an der durch posttranslationale Polyadenylierung Adeninreste eingefügt werden
precursor_RNA	noch nicht gereifte RNA-Spezies; kann eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nicht-translatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nichttranslatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip) enthalten
prim_transcript	primäres (ursprüngliches, nicht prozessiertes) Transkript; enthält eine am 5'-Ende abzuschneidende Region (5'Clip), eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron), eine 3'-nicht-translatierte Region (3'UTR) und eine am 3'-Ende abzuschneidende Region (3'Clip)
primer_bind	nichtkovalente Primer-Bindungsstelle für die Initiierung der Replikation, Transkription oder reversen Transkription; enthält eine oder mehrere Stellen für synthetische Elemente, z.B. PCR-Primerelemente
promoter	Region auf einem DNA-Molekül, die an der Bindung der RNA-Polymerase beteiligt ist, die die Transkription initiiert
protein_bind	nichtkovalente Protein-Bindungsstelle auf der Nucleinsäure
RBS	ribosomale Bindungsstelle
repeat_region	Genomregion mit repetitiven Einheiten
repeat_unit	einzelnes Repeat (repetitive Einheit)
rep_origin	Replikationsursprung; Startpunkt der Duplikation der Nucleinsäure, durch die zwei identische Kopien entstehen
rRNA	reife ribosomale RNA; RNA-Komplex des Ribonucleoprotein-Partikels (Ribosom), der Aminosäuren zu Proteinen zusammenfügt
S_region	Switch-Region der schweren Immunglobulinketten; beteiligt am Umbau der DNA von schweren Ketten, der zur Expression einer anderen Immunglobulin-Klasse aus derselben B-Zelle führt
satellite	viele tandemartig hintereinander geschaltete (identische oder verwandte) Repeats einer kurzen grundlegenden repetitiven Einheit; viele davon unterscheiden sich in der Basenzusammensetzung oder einer anderen Eigenschaft vom Genomdurchschnitt und können so von der Hauptmasse der genomischen DNA abgetrennt werden
scRNA	kleine, cytoplasmische RNA; eines von mehreren kleinen cytoplasmischen RNA-Molekülen im Cytoplasma und (manchmal) im Zellkern eines Eukaryonten
sig_peptide	für ein Signalpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines sekretorischen Proteins codiert; diese Domäne spielt bei der Anheftung des naszierenden Polypeptids an die Membran eine Rolle; Leader-Sequenz
snRNA	kleine Kern-RNA; eine der vielen kleinen RNA-Formen, die nur im Zellkern vorkommen; einige der snRNAs spielen beim Spleißen oder bei anderen RNA-verarbeitenden Reaktionen eine Rolle
source	bezeichnet die biologische Herkunft des genannten Sequenzabschnitts; die Angabe dieses Schlüssels ist obligatorisch; jeder Eintrag muss mindestens einen Herkunftsschlüssel aufweisen, der die gesamte Sequenz umfasst; es dürfen zu jeder Sequenz auch mehrere Herkunftsschlüssel angegeben werden

Tabelle 5
Liste der Merkmalschlüssel zu Nucleotidsequenzen

Schlüssel	Beschreibung
stem_loop	Haarnadelschleife; eine Doppelhelix-Region, die durch Basenpaarung zwischen benachbarten (invertierten) komplementären Sequenzen in einem RNA- oder DNA-Einzelstrang entsteht
STS	Sequence Tagged Site; kurze, nur als Einzelkopie vorkommende DNA-Sequenz, die einen Kartierungspunkt auf dem Genom bezeichnet und durch PCR ermittelt werden kann; eine Region auf dem Genom kann durch Bestimmung der Reihenfolge der STSs kartiert werden
TATA_signal	TATA-Box; Goldberg-Hogness-Box; ein konserviertes AT-reiches Septamer, das sich rund 25 Basenpaare vor dem Startpunkt jeder eukaryontischen RNA-Polymerase-II-Transkriptionseinheit befindet und bei der Positionierung des Enzyms für eine korrekte Initiation eine Rolle spielen kann; Konsensussequenz = TATA (A oder T) A (A oder T)
terminator	DNA-Sequenz, die entweder am Ende des Transkripts oder neben einer Promotor-Region liegt und bewirkt, dass die RNA-Polymerase die Transkription beendet; kann auch die Bindungsstelle eines Repressor-Proteins sein
transit_peptide	für Transitpeptid codierende Sequenz; Sequenz, die für eine N-terminale Domäne eines im Zellkern codierten Organellen-Proteins codiert; diese Domäne ist an der posttranslationalen Einschleusung des Proteins in die Organelle beteiligt
tRNA	reife, transfer-RNA, ein kleines RNA-Molekül (75 - 85 Basen lang), das die Translation einer Nucleinsäure-Sequenz in eine Aminosäure-Sequenz vermittelt
unsure	der Autor kennt die Sequenz in dieser Region nicht genau
V-region	variable Region der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für das variable Aminoende; kann aus V-, D-, N- und J-Abschnitten bestehen
V_segment	variabler Abschnitt der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gamma-Ketten von T-Zell-Rezeptoren; codiert für den Großteil der variablen Region (V_region) und die letzten Aminosäuren des Leader-Peptids
variation	ein verwandter Stamm enthält stabile Mutationen desselben Gens (z.B. RFLPs, Polymorphismen usw.), die sich an dieser (und möglicherweise auch an anderer) Stelle von der vorliegenden Sequenz unterscheiden
3'clip	3'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird
3'UTR	Region am 3'-Ende eines reifen Transkripts (nach dem Stopcodon), die nicht in ein Protein translatiert wird
5'clip	5'-äußerste Region eines Precursor-Transkripts, die bei der Prozessierung abgeschnitten wird
5'UTR	Region am 5'-Ende eines reifen Transkripts (vor dem Initiationscodon), die nicht in ein Protein translatiert wird
-10_signal	Pribnow-Box; konservierte Region rund 10 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten, die bei der Bindung der RNA-Polymerase eine Rolle spielt; Konsensussequenz = TAtAaT
-35_signal	konserviertes Hexamer rund 35 Basenpaare stromaufwärts vom Startpunkt der bakteriellen Transkriptionseinheiten; Konsensussequenz = TTGACa oder TGTTGACA

Tabelle 6 Liste der Merkmalschlüssel zu Proteinsequenzen	
Schlüssel	Beschreibung
CONFLICT	in den einzelnen Unterlagen ist von verschiedenen Sequenzen die Rede
VARIANT	den Angaben der Autoren zufolge gibt es Sequenzvarianten
VARSPLIC	Beschreibung von Sequenzvarianten, die durch alternatives Spleißen entstanden sind
MUTAGEN	experimentell veränderte Stelle
MOD_RES	posttranslationale Modifikation eines Rests
ACETYLATION	N-terminale oder sonstige
AMIDATION	in der Regel am C-Terminus eines reifen aktiven Peptids
BLOCKED	unbestimmte Gruppe, die das N- oder C-terminale Ende blockiert
FORMYLATION	des N-terminalen Methionin
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	von Asparagin; Asparaginsäure, Prolin oder Lysin
METHYLATION	in der Regel von Lysin oder Arginin
PHOSPORYLATION	von Serin, Threonin, Tyrosin, Asparaginsäure oder Histidin
PYRROLIDONE CARBOXYLIC ACID	N-terminales Glutamat, das ein internes cyclisches Lactam gebildet hat
SULFATATION	in der Regel von Tyrosin
LIPID	kovalente Bindung eines Lipidanteils
MYRISTATE	Myristat-Gruppe, die durch eine Amidbindung an den N-terminalen Glycin-Rest der reifen Form eines Proteins oder an einen internen Lysin-Rest gebunden ist
PALMITATE	Palmitat-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest oder durch eine Esterbindung an einen Serin- oder Threonin-Rest gebunden ist
FARNESYL	Farnesyl-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest gebunden ist
GERANYL-GERANYL	Geranyl-geranyl-Gruppe, die durch eine Thioetherbindung an einen Cystein-Rest gebunden ist
GPI-ANCHOR	Glykosyl-phosphatidylinositol-(GPI-)Gruppe, die an die alpha-Carboxylgruppe des C-terminalen Rests der reifen Form eines Proteins gebunden ist
N-ACYL DIGLYCERIDE	N-terminales Cystein der reifen Form eines prokaryontischen Lipoproteins mit einer amidgebundenen Fettsäure und einer Glyceryl-Gruppe, an die durch Esterbindungen zwei Fettsäuren gebunden sind
DISULFID	Disulfidbindung; den "VON"- und den "BIS" -Endpunkt bilden die beiden Reste, die durch eine ketteninterne Disulfidbindung verbunden sind; sind der "VON"- und der "BIS" -Endpunkt identisch, ist die Disulfidbindung kettenübergreifend, und die Art der Quervernetzung ist im Beschreibungsfeld anzugeben
THIOLEST	Thiolesterbindung; den "VON"- und den "BIS" -Endpunkt bilden die beiden Reste, die durch die Thiolesterbindung verbunden sind
THIOETH	Thioetherbindung; den "VON"- und den "BIS" -Endpunkt bilden die beiden Reste, die durch die Thioetherbindung verbunden sind
CARBOHYD	Glykosylierungs-Stelle; die Art des Kohlenhydrats (sofern bekannt) ist im Beschreibungsfeld anzugeben

Tabelle 6 Liste der Merkmalschlüssel zu Proteinsequenzen	
Schlüssel	Beschreibung
METAL	Bindungsstelle für ein Metallion; die Art des Metalls ist im Beschreibungsfeld anzugeben
BINDING	Bindungsstelle für eine beliebige chemische Gruppe (Coenzym, prosthetische Gruppe usw.); die Art der Gruppe ist im Beschreibungsfeld anzugeben
SIGNAL	Bereich einer Signalsequenz (Präpeptid)
TRANSIT	Bereich eines Transit-Peptids (mitochondriales, chloroplastidäres oder für Microbodies)
PROPEP	Bereich eines Propeptids
CHAIN	Bereich einer Polypeptid-Kette im reifen Protein
PEPTIDE	Bereich eines freigesetzten aktiven Peptids
DOMAIN	Bereich einer wichtigen Domäne auf der Sequenz; die Art dieser Domäne ist im Beschreibungsfeld anzugeben
CA_BIND	Bereich einer Calcium-bindenden Region
DNA_BIND	Bereich einer DNA-bindenden Region
NP_BIND	Bereich einer Nucleotidphosphat-bindenden Region; die Art des Nucleotidphosphats ist im Beschreibungsfeld anzugeben
TRANSMEM	Bereich einer Transmembran-Region
ZN_FING	Bereich einer Zink-Finger-Region
SIMILAR	Grad der Ähnlichkeit mit einer anderen Proteinsequenz; im Beschreibungsfeld sind genaue Angaben über diese Sequenz zu machen
REPEAT	Bereich einer internen Sequenzwiederholung
HELIX	Sekundärstruktur: Helices, z. B. Alpha-Helix, 3(10)-Helix oder Pi-Helix
STRAND	Sekundärstruktur: Beta-Strang, z. B. durch Wasserstoff-Brückenbindungen stabilisierter Beta-Strang, oder Rest in einer isolierten Beta-Brücke
TURN	Sekundärstruktur: Schleife, z. B. durch Wasserstoff-Brückenbindungen stabilisierte Schleife (3-, 4- oder 5-Schleife)
ACT_SITE	Aminosäure(n), die bei der Aktivität eines Enzyms mitwirkt (mitwirken)
SITE	irgendeine andere wichtige Stelle auf der Sequenz
INIT_MET	die Sequenz beginnt bekanntermaßen mit einem Start-Methionin
NON_TER	der Rest am Sequenzanfang oder -ende ist nicht der Terminalrest; steht er an der Position 1, so bedeutet das, dass diese nicht der N-Terminus des vollständigen Moleküls ist; steht er an letzter Position, so ist diese Position nicht der C-Terminus des vollständigen Moleküls; für diesen Schlüssel gibt es kein Beschreibungsfeld

Tabelle 6 Liste der Merkmalschlüssel zu Proteinsequenzen	
Schlüssel	Beschreibung
NON_CONS	nicht aufeinander folgende Reste; zeigt an, dass zwei Reste in einer Sequenz nicht aufeinander folgen, sondern dass zwischen ihnen einige nichtsequenzierte Reste liegen
UNSURE	Unsicherheiten in der Sequenz; mit diesem Schlüssel werden Regionen einer Sequenz beschrieben, bei der sich der Autor bezüglich der Sequenzzuweisung nicht sicher ist

Beispiel:

<110> Smith, John; Smithgene Inc.

<120> Beispiel für ein Sequenzprotokoll

<130> 01 - 00001

<140> PCT/EP98 / 00001

<141> 1998-12-31

<150> US 08 / 999,999

<151> 1997-10-15

<160> 4

<170> PatentIn Version 2.0

<210> 1

<211> 389

<212> DNA

<213> Paramecium sp.

<220>

<221> CDS

<222> (279) . . . (389)

<300>

<301> Doe, Richard

<302> Isolation and Characterization of a Gene Encoding a Protease from Paramecium sp.

<303> Journal of Genes

<304> 1

<305> 4

<306> 1-7

<307> 1988-06-31

<308> 123456

<309> 1988-06-31

<400> 1

agctgtagtc attcctgtgt cctcttctct ctgggcttct caccctgcta atcagatctc 60

aggagagtg tctgaccct cctctgcctt tgcagctca caggcaggca ggcaggcagc 120

tgatgtggca attgctggca gtccacagg ctttcagcc aggctaggg tgggtccgc 180

cgcgccgagg cgcccctct cgcgctctc tcgcgctct ctctcgctct cctctcgctc 240

ggacctgatt aggtgagcag gaggaggggg cagttagc atg gtt tca atg ttc agc 296

Met Val Ser Met Phe Ser

1

5

ttg tct ttc aaa tgg cct gga ttt tgt ttg ttt gtt tgt ttg ttc caa 344

Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu Phe Val Cys Leu Phe Gln

10

15

20

tgt ccc aaa gtc ctc ccc tgt cac tca tca ctg cag ccg aat ctt 389

Cys Pro Lys Val Leu Pro Cys His Ser Ser Leu Gln Pro Asn Leu

25

30

35