Deutsches
Patent- und Markenamt

STANDARD FÜR DIE DARSTELLUNG VON NUKLEOTID- UND AMINOSÄURESEQUENZPROTOKOLLEN IN XML (EXTENSIBLE MARKUP LANGUAGE)

nach § 11 Absatz 2 der Patentverordnung

Das Deutsche Patent- und Markenamt hat diesen Standard für die Darstellung von Nukleotid- und Aminosäuresequenzprotokollen in XML nach § 11 Absatz 2 der Patentverordnung vom 1. September 2003 (BGBl. I S. 1702), die zuletzt durch Artikel 1 der Verordnung zur Änderung der Patentverordnung und der Gebrauchsmusterverordnung im Hinblick auf Nukleotid- und Aminosäuresequenzprotokolle vom 14. Juni 2022 (BGBl. I S. 878) geändert worden ist, im Bundesanzeiger bekannt gemacht.

(BAnz AT 28.06.2024 B8)

Dieser Standard entspricht den Empfehlungen des WIPO-Standards ST.26, Version 1.7, für die Darstellung von Nukleotid- und Aminosäuresequenzprotokollen in XML (EXTENSIBLE MARKUP-LANGUAGE).[1]

Er ersetzt die bisherige, gemäß BAnz AT 22.06.2023 B5 bekanntgemachte, zum 1. Juli 2023 in Kraft getretene Fassung des Standards für die Darstellung von Nukleotid- und Aminosäuresequenzprotokollen in XML.

Die neue Version des Standards tritt zum 1. Juli 2024 in Kraft.

Sie gilt für alle Patent- und Gebrauchsmusteranmeldungen, die am oder nach dem 1. Juli 2024 beim DPMA eingereicht werden. Sie gilt auch für an oder nach dem 1. Juli 2024 eingereichte Anmeldungen, die die Priorität einer Anmeldung beanspruchen, die noch vor dem 1. Juli 2024 eingereicht wurde und daher noch ein Sequenzprotokoll nach den bis 30. Juni 2024 geltenden Vorschriften enthält. Die neue Version des Standards gilt auch für Teilanmeldungen, auch wenn die Stammanmeldung vor dem 1. Juli 2024 eingereicht wurde. Für Patent- und Gebrauchsmusteranmeldungen, die vor dem 1. Juli 2024 eingereicht worden sind, ist die bisherige Version des Standards in seiner bis zum 30. Juni 2024 geltenden Fassung anzuwenden.

[1] Soweit in den Anhängen zu diesem Standard auf den „ST.26“, den „WIPO-Standard ST.26“ oder den „Standard ST.26“ verwiesen wird, wird damit der vorliegende Standard bezeichnet.

	Anschrift	Telefon	Telefax
Dienststelle München **Dienststelle Jena** **Informations- und Dienstleistungszentrum Berlin**	Zentrale Postanschrift: 80297 München	Zentraler Kundenservice: +49 89 2195-1000	Zentrale Telefaxnummer: +49 89 2195-2221

Zahlungsempfänger: Bundeskasse
IBAN: DE84 7000 0000 0070 0010 54, BIC (SWIFT-Code): MARKDEF1700
Anschrift der Bank: Bundesbankfiliale München, Leopoldstraße 234, 80807 München

Internet:
https://www.dpma.de

P 2790c/7.24

INHALTSVERZEICHNIS

EINFÜHRUNG 3
DEFINITIONEN 3
GELTUNGSBEREICH 4
REFERENZEN 5
DARSTELLUNG VON SEQUENZEN 5
Nukleotidsequenzen 5
Aminosäuresequenzen 8
Darstellung besonderer Situationen 10
XML-STRUKTUR DES SEQUENZPROTOKOLLS 10
Wurzelelement 11
Allgemeiner Informationsteil 11
Sequenzdatenteil 15
Merkmaltabelle 17
Merkmalschlüssel 17
Obligatorische Merkmalschlüssel 17
Lage des Merkmals 18
Merkmal-Qualifier 20
Obligatorische Merkmal-Qualifier 20
Qualifier-Elemente 20
Freitext 23
Codierende Sequenzen 24
Varianten 25

ANHÄNGE

Anhang I – Kontrolliertes Vokabular
Anhang II – Dokumententypdefinition (DTD) für Sequenzprotokolle
Anhang III – Beispielexemplar eines Sequenzprotokolls (XML-Datei)
Anhang IV – In einer XML-Instanz eines Sequenzprotokolls zu verwendender ausgewählter Zeichensatz aus dem Unicodeblock Basis-Lateinisch
Anhang V – Zusätzliche Anforderungen für den Datenaustausch (nur für Ämter für geistiges Eigentum)
Anhang VI – Leitfaden mit illustrierten Beispielen
Anlage – Sequenzen des Leitfadens in XML
Anhang VII – Empfehlung für die Überführung eines Sequenzprotokolls von ST.25 nach ST.26: potenzielles Hinzufügen oder Streichen von Gegenständen

EINFÜHRUNG

1. Mit dem vorliegenden Standard wird festgelegt, welche in einer Patentanmeldung offenbarten Nukleotid- und Aminosäuresequenzen in ein Sequenzprotokoll aufzunehmen sind, wie diese Offenbarungen darzustellen sind und die Dokumenttypdefinition (DTD), die für ein Sequenzprotokoll im Format XML (eXtensible Markup Language) zu verwenden ist.

2. Dieser Standard dient folgenden Zwecken:

(a) Er soll Anmeldern ermöglichen, für eine Patentanmeldung ein einziges Sequenzprotokoll zu erstellen, das sowohl für internationale als auch für nationale oder regionale Verfahren geeignet ist.

(b) Er soll die Genauigkeit und Qualität der Sequenzdarstellungen erhöhen, damit sie im Interesse der Anmelder, der Öffentlichkeit und der Prüfer leichter verbreitet werden können.

(c) Er soll die Recherche von Sequenzdaten erleichtern.

(d) Er soll ermöglichen, dass Sequenzdaten in elektronischer Form ausgetauscht und in Computerdatenbanken gespeichert werden.

DEFINITIONEN

3. Für die Zwecke dieses Standards bezeichnet der Ausdruck

(a) "Aminosäure" jede beliebige Aminosäure, die unter Verwendung eines der in Anhang I aufgeführten Symbole (siehe Abschnitt 3, Tabelle 3) dargestellt werden kann. Zu solchen Aminosäuren gehören u. a. D-Aminosäuren und Aminosäuren, die modifizierte oder synthetische Seitenketten enthalten. Unter Aminosäuren sind nicht modifizierte L-Aminosäuren zu verstehen, sofern sie nicht in der Merkmaltabelle gemäß Absatz 30 als modifiziert näher beschrieben werden. Für die Zwecke dieses Standards wird ein Peptidnukleinsäure-Rest (PNA-Rest) nicht als Aminosäure, sondern, wie in Absatz 3(g)(i)(2) ausgeführt, als Nukleotid aufgefasst.

(b) "kontrolliertes Vokabular" die in diesem Standard verwendete Terminologie, die bei der Beschreibung der Merkmale einer Sequenz, d. h. in Annotationen zu Regionen oder Stellen von Interesse, gemäß Anhang I obligatorisch zu verwenden ist.

(c) "Aufzählung der Reste" die Offenbarung einer Sequenz in einer Patentanmeldung, indem jeder Rest der Sequenz der Reihe nach aufgeführt wird, wobei

(i) der Rest durch einen Namen, eine Abkürzung, ein Symbol oder eine Struktur (z. B. HHHHHHQ oder HisHisHisHisHisHisGln) dargestellt wird oder

(ii) Mehrfachreste durch eine Kurzformel dargestellt (z. B. His_6Gln) werden.

(d) "absichtlich übersprungene Sequenz" bzw. "leere Sequenz" einen Platzhalter, der dazu dient, eine Konsistenz mit der Offenbarung der Anmeldung und der Nummerierung der Sequenzen im Sequenzprotokoll auch dann aufrechtzuerhalten, wenn z. B. eine Sequenz aus der Offenbarung entfernt wird, um zu vermeiden, dass die Sequenzen in der Offenbarung und im Sequenzprotokoll neu nummeriert werden müssen.

(e) "modifizierte Aminosäure" jede beliebige in Absatz 3(a) beschriebene Aminosäure mit Ausnahme von L-Alanin, L-Arginin, L-Asparagin, L-Asparaginsäure, L-Cystein, L-Glutamin, L-Glutaminsäure, L-Glycin, L-Histidin, L-Isoleucin, L-Leucin, L-Lysin, L-Methionin, L-Phenylalanin, L-Prolin, L-Pyrrolysin, L-Serin, L-Selenocystein, L-Threonin, L-Tryptophan, L-Tyrosin und L-Valin.

(f) "modifiziertes Nukleotid" jedes beliebige in Absatz 3(g) beschriebene Nukleotid mit Ausnahme von Desoxyadenosin-5'-Monophosphat, Desoxyguanosin-5'-Monophosphat, Desoxycytidin-5'-Monophosphat, Desoxythymidin-5'-Monophosphat, Adenosin-5'-Monophosphat, Guanosin-5'-Monophosphat, Cytidin-5'-Monophosphat und Uridin-5'-Monophosphat.

(g) "Nukleotid" jedes beliebige Nukleotid oder Nukleotid-Analogon, das unter Verwendung eines der in Anhang I aufgeführten Symbole (siehe Abschnitt 1, Tabelle 1) dargestellt werden kann, wobei das Nukleotid oder Nukleotid-Analogon enthält:

(i) eine Rückgrateinheit, bestehend aus:

(1) 2'-Desoxyribose-5'-Monophosphat (die Rückgrateinheit eines Desoxyribonukleotids) oder Ribose-5'-Monophosphat (die Rückgrateinheit eines Ribonukleotids) oder

(2) einem Analogon eines 2'-Desoxyribose-5'-Monophosphats oder Ribose-5'-Monophosphats, das als Rückgrat eines Nukleinsäure-Analogons eine Anordnung von Nukleinbasen bedingt, die deren Anordnung in Nukleinsäuren mit einem 2'-Desoxyribose-5'-Monophosphat- oder Ribose-5'-Monophosphat-Rückgrat nachahmt, wobei das Nukleinsäure-Analogon in der Lage ist, eine Basenpaarung mit einer komplementären Nukleinsäure herzustellen. Beispiele für Rückgrateinheiten sind Aminosäuren wie in Peptidnukleinsäuren, Glykolmoleküle wie in Glykolnukleinsäuren, Threofuranosyl-Zuckermoleküle wie in Threose-Nukleinsäuren, Morpholinringe und Phosphordiamidat-Gruppen wie in Morpholinos und Cyclohexenyl-Moleküle wie in Cyclohexenyl-Nukleinsäuren

und

(ii) die Rückgrateinheit liegt entweder

(1) in Bindung an eine Nukleinbase vor, bei der es sich auch um eine modifizierte oder synthetische Purin- oder Pyrimidin-Nukleinbase handeln kann, oder

(2) ohne Purin- oder Pyrimidin-Nukleinbase vor, sofern das Nukleotid Bestandteil einer Nukleotidsequenz ist, die als "AP-Stelle" oder "abasische Stelle" bezeichnet wird.

(h) "Rest" ein einzelnes Nukleotid oder eine einzelne Aminosäure oder deren jeweilige Analoga in einer Sequenz.

(i) "Sequenzkennzahl" eine eindeutige Nummer (Ganzzahl), die jeder im Protokoll aufgeführten Sequenz zugewiesen ist.

(j) "Sequenzprotokoll" einen Teil der Beschreibung der Patentanmeldung in der eingereichten Fassung oder ein nach dem Anmeldetag eingereichtes Dokument, das die offenbarten Nukleotid- und/oder Aminosäuresequenz(en) und nähere Beschreibungen gemäß diesem Standard enthält.

(k) "spezifisch definiert" alle anderen als die in Anhang I mit dem Symbol "n" dargestellten Nukleotide und alle anderen als die in Anhang I mit dem Symbol "X" dargestellten Aminosäuren (siehe Abschnitt 1, Tabelle 1 bzw. Abschnitt 3, Tabelle 3).

(l) "unbekanntes" Nukleotid oder "unbekannte" Aminosäure den Umstand, dass ein einzelnes Nukleotid oder eine einzelne Aminosäure vorliegt, deren Identität jedoch nicht bekannt ist oder nicht offenbart wird.

(m) "Sequenzvariante" eine Nukleotid- oder Aminosäuresequenz, die einen oder mehrere Unterschiede zur Primärsequenz aufweist. Mögliche Unterschiede sind alternative Reste (siehe Absätze 15 und 27), modifizierte Reste (siehe Absätze 3(g), 3(h), 16 und 29), Deletionen, Insertionen und Substitutionen. Siehe die Absätze 93 bis 95.

(n) "Freitext" ein Format für den Wert bestimmter Qualifier, der in Form eines beschreibenden Textes oder in einem anderen (in Anhang I aufgeführten) festgelegten Format angegeben wird. Siehe Absatz 85.

(o) "sprachabhängiger Freitext" den Freitextwert bestimmter Qualifier, der zu Zwecken internationaler, nationaler oder regionaler Verfahren möglicherweise übersetzt werden muss.[2] Siehe Absatz 87.

4. Für die Zwecke dieses Standards bedeuten die Wörter oder Wortgruppen:

(a) "kann": eine fakultative oder zulässige Vorgehensweise, die jedoch nicht obligatorisch ist.

(b) "muss": eine Anforderung des Standards, deren Missachtung zur Nichterfüllung des Standards führt.

(c) "darf nicht": ein im Standard enthaltenes Verbot.

(d) "sollte": eine nachdrücklich empfohlene, jedoch nicht obligatorische Vorgehensweise;.

(e) "sollte nicht" eine nachdrücklich nicht empfohlene, jedoch nicht verbotene Vorgehensweise.

GELTUNGSBEREICH

5. Mit diesem Standard werden die Anforderungen an die Darstellung von Nukleotid- und Aminosäuresequenzprotokollen der in Patentanmeldungen offenbarten Sequenzen festgelegt.

[2] Nach § 11a Absatz 3 der Patentverordnung ist sprachabhängiger Freitext in deutscher Sprache abzufassen. Er kann zusätzlich auch in englischer Sprache angegeben werden. Dies gilt unabhängig davon, dass in diesem Standard Beispiele sprachabhängigen Freitexts in englischer Sprache gefasst sind. Beispiele sprachabhängigen Freitext in deutscher Sprache sind an zahlreichen Stellen in Klammern ergänzt.

6. Ein Sequenzprotokoll, das diesem Standard entspricht (im Folgenden: "Sequenzprotokoll"), enthält einen allgemeinen Informationsteil und einen Sequenzdatenteil. Das Sequenzprotokoll muss in einer einzigen Datei im XML-Format unter Verwendung der in Anhang II dargestellten Dokumenttypdefinition (DTD) dargestellt werden. Die bibliografischen Angaben im allgemeinen Informationsteil dienen lediglich der Zuordnung des Sequenzprotokolls zu der Patentanmeldung, für die es eingereicht wird. Der Sequenzdatenteil besteht aus einem oder mehreren Sequenzdatenelementen, die jeweils Informationen über eine Sequenz enthalten. Die Sequenzdatenelemente enthalten diverse Merkmalschlüssel und zugehörige Qualifier, die auf den Spezifikationen der International Nucleotide Sequence Database Collaboration (INSDC) und des UniProt-Konsortiums basieren.

7. Für die Zwecke dieses Standards ist unter einer Sequenz, die in ein Sequenzprotokoll aufgenommen werden muss, jede Sequenz zu verstehen, die an einer beliebigen Stelle in einer Anmeldung durch Aufzählung der Reste offenbart wird und dargestellt werden kann als:

(a) unverzweigte Sequenz oder lineare Region einer verzweigten Sequenz aus zehn oder mehr spezifisch definierten Nukleotiden, wobei benachbarte Nukleotide verbunden sind durch

(i) eine 3'-5'- (oder 5'-3'-)Phosphodiesterbindung oder

(ii) eine beliebige chemische Bindung, die zu einer Anordnung benachbarter Nukleinbasen führt, mit der die Anordnung der Nukleinbasen in natürlich vorkommenden Nukleinsäuren nachgeahmt wird, oder

(b) unverzweigte Sequenz oder lineare Region einer verzweigten Sequenz, die vier oder mehr spezifisch definierte Aminosäuren enthält, die ein einziges Peptid-Rückgrat bilden, d. h., benachbarte Aminosäuren werden durch Peptidbindungen zusammengehalten.

8. Ein Sequenzprotokoll darf keine mit einer eigenen Sequenzkennzahl versehenen Sequenzen enthalten, die aus weniger als zehn spezifisch definierten Nukleotiden oder weniger als vier spezifisch definierten Aminosäuren bestehen.

REFERENZEN

9. Für diesen Standard sind die folgenden Standards und Ressourcen relevant:

International Nucleotide Sequence Database Collaboration (INSDC)	http://www.insdc.org/;
Internationale Norm ISO 639-1:2002	Codes für Sprachennamen – Teil 1: Alpha-2-Code;
UniProt-Konsortium	http://www.uniprot.org/;
W3C XML 1.0	http://www.w3.org/;
WIPO-Standard ST.2	Standard manner for designating calendar dates by using Gregorian calendar;
WIPO-Standard ST.3	Recommended standard on two-letter codes for the representation of states, other entities and intergovernmental organizations;
WIPO-Standard ST.16	Recommended standard code for the identification of different kinds of patent documents;
WIPO-Standard ST.25	Standard for the presentation of nucleotide and amino acid sequence listings in patent applications.

DARSTELLUNG VON SEQUENZEN

10. Jede Sequenz, die von Absatz 7 erfasst wird, muss mit einer separaten Sequenzkennzahl versehen werden; dies gilt auch für Sequenzen, die zu einer Region einer längeren Sequenz identisch sind. Die Sequenzkennzahlen müssen mit der Zahl 1 beginnen und sich fortlaufend in ganzzahligen Schritten erhöhen. Wenn für eine Sequenzkennzahl keine Sequenz vorhanden ist, d. h. eine absichtlich übersprungene Sequenz vorliegt, muss anstelle einer Sequenz "000" verwendet werden (siehe Absatz 58). Die Gesamtzahl der Sequenzen muss im Sequenzprotokoll angegeben werden und sie muss der Gesamtzahl der Sequenzkennzahlen entsprechen, unabhängig davon, ob im Anschluss an eine Sequenzkennzahl eine Sequenz oder "000" folgt.

Nukleotidsequenzen

11. Nukleotidsequenzen müssen anhand eines Einzelstrangs in Richtung vom 5'-Ende zum 3'-Ende von links nach rechts dargestellt werden, oder in einer Richtung von links nach rechts, mit der die 5'-3'-Richtung nachgeahmt wird. Die Bezeichnungen 5' und 3' oder ähnliche Bezeichnungen dürfen nicht in die Sequenz aufgenommen werden. Eine doppelsträngige Nukleotidsequenz, die durch Aufzählung der Reste beider Stränge offenbart wird, muss wie folgt dargestellt werden:

(a) als eine einzelne Sequenz oder als zwei separate Sequenzen mit jeweils eigener Sequenzkennzahl, wenn die beiden gesonderten Stränge vollständig komplementär zueinander sind, oder

(b) als zwei separate Sequenzen mit jeweils eigener Sequenzkennzahl, wenn die beiden Stränge nicht vollständig komplementär zueinander sind.

12. Für die Zwecke dieses Standards hat das erste in der Sequenz dargestellte Nukleotid die Restenummer 1. Im Falle ringförmiger Nukleotidsequenzen muss der Anmelder auswählen, welchem Nukleotid die Restenummer 1 zugewiesen wird. Die Nummerierung erfolgt fortlaufend über die gesamte Sequenz in 5'-3'-Richtung oder in einer Richtung, mit der die 5'-3'-Richtung nachgeahmt wird. Die letzte Restenummermuss der Anzahl der Nukleotide in der Sequenz entsprechen.

13. Alle Nukleotide in einer Sequenz müssen unter Verwendung eines der in Anhang I aufgeführten Symbole (siehe Abschnitt 1, Tabelle 1) dargestellt werden. Es dürfen nur Kleinbuchstaben verwendet werden. Jedes zur Darstellung eines Nukleotids verwendete Symbol steht für nur einen Rest.

14. Das Symbol "t" wird in DNA als Thymin und in RNA als Uracil ausgelegt. Uracil in DNA oder Thymin in RNA gilt als modifiziertes Nukleotid und muss in der Merkmaltabelle gemäß Absatz 19 näher beschrieben werden.

15. Wenn ein Mehrdeutigkeitssymbol (das zwei oder mehr alternative Nukleotide repräsentiert) angebracht ist, sollte das restriktivste Symbol verwendet werden, wie in Anhang I (Abschnitt 1, Tabelle 1) aufgeführt. Wenn beispielsweise ein Nukleotid in einer bestimmten Position "a" oder "g" sein könnte, sollte "r" statt "n" verwendet werden. Das Symbol "n" wird als eines der Symbole "a", "c", "g" oder "t/u" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Das Symbol "n" darf für nichts anderes als ein Nukleotid verwendet werden. Ein einzelnes modifiziertes oder "unbekanntes" Nukleotid kann durch das Symbol "n" in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle dargestellt werden, wie in den Absätzen 16, 17, 21 oder 93–96 beschrieben. Zur Darstellung von Sequenzvarianten, d. h. Alternativen, Deletionen, Insertionen oder Substitutionen, siehe die Absätze 93 bis 100.

16. Modifizierte Nukleotide sollten in der Sequenz nach Möglichkeit als die entsprechenden nicht modifizierten Nukleotide dargestellt werden, d. h., "a", "c", "g" oder "t". Ein modifiziertes Nukleotid in einer Sequenz, das durch kein anderes Symbol in Anhang I (siehe Abschnitt 1, Tabelle 1) dargestellt werden kann, d. h. ein "sonstiges" ("other"), z. B. nicht natürlich vorkommendes Nukleotid, muss durch das Symbol "n" dargestellt werden. Das Symbol "n" steht für nur einen Rest.

17. Ein modifiziertes Nukleotid muss in der Merkmaltabelle (siehe Absatz 60 ff.) unter Verwendung des Merkmalschlüssels "modified_base" und des obligatorischen Qualifiers "mod_base" in Verbindung mit einer einzigen Abkürzung aus Anhang I (siehe Abschnitt 2, Tabelle 2) als Qualifier-Wert näher beschrieben werden; wenn die Abkürzung "OTHER" verwendet wird, muss der vollständige, ungekürzte Name des modifizierten Nukleotids als Wert in einem "note"-Qualifier angegeben werden. Für die Aufzählung alternativer modifizierter Nukleotide kann der Qualifier-Wert "OTHER" in Verbindung mit einem weiteren "note"-Qualifier verwendet werden (siehe Absätze 97 und 98). Die oben genannten in Anhang I angegebenen Abkürzungen (oder vollständigen Namen) (siehe Abschnitt 2, Tabelle 2) dürfen in der Sequenz selbst nicht verwendet werden.

18. Eine Nukleotidsequenz, die eine oder mehrere Regionen aufeinander folgender modifizierter Nukleotide mit derselben Rückgrateinheit enthält (siehe Absatz 3(g)(i)(2)), muss in der Merkmaltabelle gemäß Absatz 17 näher beschrieben werden. Die modifizierten Nukleotide jeder dieser Regionen können in einem einzigen INSDFeature-Element gemeinsam beschrieben werden, wie in Absatz 22 vorgesehen. Der restriktivste, ungekürzte chemische Name, der alle modifizierten Nukleotide im Bereich umfasst, oder eine Liste der chemischen Namen aller Nukleotide im betreffenden Bereich muss als Wert des "note"-Qualifiers angegeben werden. Beispielsweise kann eine Glykolnukleinsäuresequenz, die als "a", "c", "g" oder "t" beschriebene Nukleinbasen enthält, im "note"-Qualifier als "2,3-dihydroxypropyl nucleosides" ("2,3-Dihydroxypropyl-Nukleoside") angegeben werden. Alternativ kann die gleiche Sequenz im "note"-Qualifier als "2,3-dihydroxypropyladenine, 2,3-dihydroxypropylthymine, 2,3-dihydroxypropylguanine or 2,3-dihydroxypropylcytosine" ("2,3-Dihydroxypropyladenin, 2,3-Dihydroxypropylthymin, 2,3-Dihydroxypropylguanin oder 2,3-Dihydroxypropylcytosin") angegeben werden. Enthält ein einzelnes modifiziertes Nukleotid in der Region eine zusätzliche Modifikation, so muss auch das modifizierte Nukleotid in der Merkmaltabelle gemäß Absatz 17 näher beschrieben werden.

19. Uracil in DNA oder Thymin in RNA gelten als modifizierte Nukleotide. Sie müssen in der Sequenz als "t" dargestellt und in der Merkmaltabelle näher beschrieben werden, hierzu ist der Merkmalschlüssel "modified_base", der Qualifier "mod_base" mit dem Wert "OTHER" und der Qualifier "note" mit dem Wert "uracil" ("Uracil") bzw. "thymine" ("Thymin") zu verwenden.

20. Die folgenden Beispiele veranschaulichen die Darstellung modifizierter Nukleotide gemäß den Absätzen 16–18:

Beispiel 1: Modifiziertes Nukleotid unter Verwendung einer Abkürzung aus Anhang I (siehe Abschnitt 2, Tabelle 2)

```
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>15</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>i</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Beispiel 2: Modifiziertes Nukleotid unter Verwendung von "OTHER" aus Anhang I (siehe Abschnitt 2, Tabelle 2)

```
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>4</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>xanthine</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Beispiel 3: Eine aus von Absatz 3(g)(i)(2) erfassten modifizierten Nukleotiden bestehende Nukleotidsequenz mit zwei einzelnen Nukleotiden, die eine weitere Modifikation beinhalten

```
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>1..954</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>2,3-dihydroxypropyl nucleosides</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>439</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>i</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>684</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>xanthine</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

21. Jedes "unbekannte" Nukleotid muss in der Sequenz durch das Symbol "n" dargestellt werden. Ein "unbekanntes" Nukleotid sollte in der Merkmaltabelle (siehe Absatz 60 ff.) unter Verwendung des Merkmalschlüssels "unsure" näher beschrieben werden. Das Symbol "n" steht für nur einen Rest.

22. Eine Region, die eine bekannte Zahl an zusammenhängenden "a"-, "c"-, "g"-, "t"- oder "n"-Resten enthält, für die dieselbe Beschreibung gilt, kann gemeinsam durch ein einziges INSDFeature-Element mit der Syntax "x..y" als Lagedeskriptor im Element `INSDFeature_location` angegeben werden (siehe Absätze 64 bis 71). Zur Darstellung von Sequenzvarianten, d. h. Alternativen, Deletionen, Insertionen oder Substitutionen, siehe Absätze 93 bis 100.

23. Das folgende Beispiel veranschaulicht die Darstellung einer Region modifizierter Nukleotide, für die dieselbe Beschreibung gilt, gemäß dem obigen Absatz 22:

```
<INSDFeature>
	<INSDFeature_key>modified_base</INSDFeature_key>
	<INSDFeature_location>358..485</INSDFeature_location>
	<INSDFeature_quals>
		<INSDQualifier>
			<INSDQualifier_name>mod_base</INSDQualifier_name>
			<INSDQualifier_value>OTHER</INSDQualifier_value>
		</INSDQualifier>
		<INSDQualifier>
			<INSDQualifier_name>note</INSDQualifier_name>
			<INSDQualifier_value>isoguanine</INSDQualifier_value>
		</INSDQualifier>
	</INSDFeature_quals>
</INSDFeature>
```

Aminosäuresequenzen

24. Die Aminosäuren in einer Aminosäuresequenz müssen in Amino-Carboxy-Richtung von links nach rechts dargestellt werden. Die Amino- und Carboxylgruppen dürfen in der Sequenz nicht dargestellt werden.

25. Für den Zweck dieses Standards hat die erste Aminosäure in der Sequenz die Restenummer 1, dies gilt insbesondere auch für dem reifen Protein vorangehende Aminosäuren, z. B. Präsequenzen, Prosequenzen, Prä-Prosequenzen und Signalsequenzen. Im Falle einer ringförmigen Aminosäuresequenz, bei der der Ring ausschließlich aus Aminosäureresten besteht, die durch Peptidbindungen verbunden sind, d. h. die keine Amino- und Carboxy-Termini enthält, muss der Anmelder auswählen, welcher Aminosäure die Restenummer 1 zugewiesen wird. Die Nummerierung verläuft fortlaufend über die gesamte Sequenz in Amino-Carboxy-Richtung.

26. Alle Aminosäuren in einer Sequenz müssen unter Verwendung eines der in Anhang I aufgeführten Symbole (siehe Abschnitt 3, Tabelle 3) dargestellt werden. Es dürfen nur Großbuchstaben verwendet werden. Jedes zur Darstellung einer Aminosäure verwendete Symbol steht für nur einen Rest.

27. Wenn ein Mehrdeutigkeitssymbol (das zwei oder mehr alternative Aminosäuren repräsentiert) angebracht ist, sollte das restriktivste Symbol verwendet werden, wie in Anhang I (Abschnitt 3, Tabelle 3) aufgeführt. Wenn beispielsweise eine Aminosäure in einer bestimmten Position Asparaginsäure oder Asparagin sein könnte, sollte "B" statt "X" verwendet werden. Das Symbol "X" wird als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Das Symbol "X" darf für nichts anderes als eine Aminosäure verwendet werden. Eine einzelne modifizierte oder "unbekannte" Aminosäure kann durch das Symbol "X" in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle dargestellt werden, wie z. B. in den Absätzen 29, 30, 32 oder 93–98 beschrieben. Zur Darstellung von Sequenzvarianten, d. h. Alternativen, Deletionen, Insertionen oder Substitutionen, siehe die Absätze 93 bis 100.

28. Offenbarte Aminosäuresequenzen, die durch interne Terminatorsymbole getrennt sind, z. B. durch "Ter", Sternchen "*", Punkt "." oder Leerzeichen, müssen als separate Sequenzen für jede Aminosäuresequenz angegeben werden, die mindestens vier spezifisch definierte Aminosäuren enthält und von Absatz 7 erfasst wird. Jeder solchen separaten Sequenz muss eine eigene Sequenzkennzahl zugewiesen werden. Terminatorsymbole und Leerzeichen dürfen nicht in den Sequenzen eines Sequenzprotokolls enthalten sein (siehe Absatz 57).

29. Modifizierte Aminosäuren, einschließlich D-Aminosäuren, sollten in der Sequenz nach Möglichkeit als die entsprechenden nicht modifizierten Aminosäuren dargestellt werden. Eine modifizierte Aminosäure in einer Sequenz, die durch kein anderes Symbol in Anhang I (siehe Abschnitt 3, Tabelle 3) dargestellt werden kann, d. h. eine "sonstige" ("other") Aminosäure, muss durch das Symbol "X" dargestellt werden. Das Symbol "X" steht für nur einen Rest.

30. Eine modifizierte Aminosäure muss in der Merkmaltabelle näher beschrieben werden (siehe Absatz 60 ff.). Gegebenenfalls sollten die Merkmalschlüssel "CARBOHYD" oder "LIPID" zusammen mit dem Qualifier "note" verwendet werden. Der Merkmalschlüssel "MOD_RES" sollte für andere posttranslational modifizierte Aminosäuren in Verbindung mit dem Qualifier "note" verwendet werden; andernfalls sollte der Merkmalschlüssel "SITE" in Verbindung mit dem Qualifier "note" verwendet werden. Der Wert für den Qualifier "note" muss entweder eine in Anhang I aufgeführte Abkürzung (siehe Abschnitt 4, Tabelle 4) sein oder der vollständige, ungekürzte Name der modifizierten Aminosäure. Die in der oben genannten in Tabelle 4 angegebenen Abkürzungen oder die vollständigen, ungekürzten Namen dürfen in der Sequenz selbst nicht verwendet werden.

31. Die folgenden Beispiele veranschaulichen die Darstellung modifizierter Aminosäuren gemäß dem obigen Absatz 30:

Beispiel 1: Posttranslational modifizierte Aminosäure

```
<INSDFeature>
    <INSDFeature_key>MOD_RES</INSDFeature_key>
    <INSDFeature_location>3</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>3Hyp</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Beispiel 2: Nicht posttranslational modifizierte Aminosäure

```
<INSDFeature>
    <INSDFeature_key>SITE</INSDFeature_key>
    <INSDFeature_location>3</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>Orn</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Beispiel 3: D-Aminosäure

```
<INSDFeature>
    <INSDFeature_key>SITE</INSDFeature_key>
    <INSDFeature_location>9</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>D-Arginine</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

32. Jede "unbekannte" Aminosäure muss in der Sequenz durch das Symbol "X" dargestellt werden. Eine "unbekannte", als "X" bezeichnete Aminosäure muss in der Merkmaltabelle (siehe Absatz 60 ff.) unter Verwendung des Merkmalschlüssels "UNSURE" und gegebenfalls mit dem Qualifier "note" näher beschrieben werden. Das Symbol "X" steht für nur einen Rest.

33. Das folgende Beispiel veranschaulicht die Darstellung einer "unbekannten" Aminosäure gemäß dem obigen Absatz 32:

```
<INSDFeature>
    <INSDFeature_key>UNSURE</INSDFeature_key>
    <INSDFeature_location>3</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>A or V</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

34. Eine Region, die eine bekannte Zahl an zusammenhängenden "X"-Resten enthält, für die dieselbe Beschreibung gilt, kann gemeinsam mittels der Syntax "x..y" als Lagedeskriptor im Element INSDFeature_location angegeben werden (siehe Absätze 64 bis 70). Zur Darstellung von Sequenzvarianten, d. h. Alternativen, Deletionen, Insertionen oder Substitutionen, siehe Absätze 93 bis 100.

Darstellung besonderer Situationen

35. Eine durch Aufzählung der Reste offenbarte Sequenz, die als eine einzige kontinuierliche Sequenz aus einem oder mehreren nicht zusammenhängenden Segmenten einer größeren Sequenz oder aus verschiedenen Sequenzen entstammenden Segmenten aufgebaut ist, muss in das Sequenzprotokoll aufgenommen und mit einer eigenen Sequenzkennzahl versehen werden.

36. Eine Sequenz mit aus spezifisch definierten Resten bestehenden Regionen, die durch eine oder mehrere aus zusammenhängenden "n"- oder "X"-Resten bestehende Regionen voneinander getrennt sind (siehe Absatz 15 bzw. 27), wobei die genaue Zahl dieser "n"- oder "X"-Reste für jede solche Region offenbart wird, muss als einzelne Sequenz in das Sequenzprotokoll aufgenommen und mit einer eigenen Sequenzkennzahl versehen werden.

37. Eine Sequenz mit aus spezifisch definierten Resten bestehenden Regionen, die durch eine oder mehrere aus einer unbekannten oder nicht offenbarten Zahl an Resten bestehende Lücken voneinander getrennt sind, darf im Sequenzprotokoll nicht als eine einzelne Sequenz dargestellt werden. Jede aus spezifisch definierten Resten bestehende Region, die von Absatz 7 erfasst wird, muss als separate Sequenz in das Sequenzprotokoll aufgenommen und mit einer eigenen Sequenzkennzahl versehen werden.

XML-STRUKTUR DES SEQUENZPROTOKOLLS

38. Gemäß Absatz 6 setzt sich die XML-Instanz einer Sequenzprotokolldatei gemäß diesem Standard aus folgenden Bestandteilen zusammen:

(a) einem allgemeinen Informationsteil mit Angaben zu der Patentanmeldung, auf die sich das Sequenzprotokoll bezieht; und

(b) einem Sequenzdatenteil, der ein oder mehrere Sequenzdatenelemente umfasst, die jeweils Informationen über eine Sequenz enthalten.

Anhang III enthält ein Beispiel für ein Sequenzprotokoll.

39. Das Sequenzprotokoll muss im Format XML 1.0 unter Verwendung der in Anhang II wiedergegebenen "Dokumenttypdefinition (DTD) für Sequenzprotokolle" dargestellt werden.

(a) Die erste Zeile der XML-Instanz muss die XML-Deklaration enthalten:

```
<?xml version="1.0" encoding="UTF-8"?>.
```

(b) Die zweite Zeile der XML-Instanz muss eine Dokumenttyp-Deklaration (DOCTYPE) enthalten:

```
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">.
```

40. Das gesamte elektronische Sequenzprotokoll muss in einer einzigen Datei enthalten sein. Die Datei muss mit Unicode UTF-8 codiert werden. Dabei gelten die folgenden Einschränkungen:

(a) Die in den Elementen `ApplicantName`, `InventorName` und `InventionTitle` des allgemeinen Informationsteils enthaltenen Angaben und der Wert `NonEnglishQualifier_value` des Sequenzdatenteils können aus allen gültigen Unicode-Zeichen bestehen, die in der XML 1.0-Spezifikation angegeben sind, mit Ausnahme der Unicode-Codepoints 0000-001F und 007F-009F. Die reservierten Zeichen ", &, ', < und > (Unicode-Codepoints 0022, 0026, 0027, 003C bzw. 003E) müssen gemäß Absatz 41 ersetzt werden und

(b) die Informationen, die in allen sonstigen Elementen und Attributen des allgemeinen Informationsteils und in allen sonstigen Elementen und Attributen des Sequenzdatenteils enthalten sind, müssen aus druckbaren Zeichen (einschließlich des Leerzeichens) aus dem Unicodeblock Basis-Lateinisch bestehen (d. h, sie sind beschränkt auf die Unicode-Codepoints 0020 bis einschließlich 007E – siehe Anhang IV). Die reservierten Zeichen ", &, ', < und > (Unicode-Codepoints 0022, 0026, 0027, 003C bzw. 003E) müssen gemäß Absatz 41 ersetzt werden.

41. In einer XML-Instanz eines Sequenzprotokolls dürfen keine numerischen Zeichenreferenzen[3] verwendet werden, und die folgenden reservierten Zeichen müssen durch die entsprechenden vordefinierten Entitäten ersetzt werden, wenn sie im Wert eines Attributs oder im Inhalt eines Elements verwendet werden:

Reserviertes Zeichen	Vordefinierte Entitäten
<	<
>	>
&	&
"	"
'	'

[3] Eine numerische Zeichenreferenz verweist auf ein Zeichen anhand seines Codepoints im universellen Zeichensatz (Unicode) unter Verwendung des folgenden Formats: "&#nnnn;" oder "&#xhhhh;", wobei "nnnn" für den Codepoint in Dezimal- und "hhhh" für den Codepoint in Hexadezimal-Darstellung steht.

Ein Beispiel ist in Absatz 71 aufgeführt. Als Zeichenreferenzen sind ausschließlich die in diesem Absatz festgelegten vordefinierten Entitäten zulässig.

42. Alle obligatorischen Elemente müssen ausgefüllt werden (außer für eine absichtlich übersprungene Sequenz, wie in Absatz 58 beschrieben). Fakultative Elemente, für die kein Inhalt verfügbar ist, sollten in der XML-Instanz nicht erscheinen (es sei denn, sie dienen nach Absatz 97 dazu, im Wert für den Qualifier "replace" die Deletion einer Sequenz darzustellen).

Wurzelelement

43. Das Wurzelelement einer XML-Instanz gemäß diesem Standard ist das Element `ST26SequenceListing` mit den folgenden Attributen:

Attribut	Beschreibung	obligatorisch/fakultativ
`dtdVersion`	Version der DTD, mit der die Datei erstellt wurde, im Format "V#_#" z. B. "V1_3"	obligatorisch
`fileName`	Name der Sequenzprotokolldatei	fakultativ
`softwareName`	Name der Software, mit der die Datei generiert wurde	fakultativ
`softwareVersion`	Version der Software, mit der die Datei generiert wurde	fakultativ
`productionDate`	Erstelldatum der Sequenzprotokolldatei (Format "JJJJ-MM-TT")	fakultativ
`originalFreeTextLanguageCode`	Der Sprachcode (siehe Verweis auf ISO 639-1:2002 in Absatz 9) für die Originalsprache, in der die Werte der sprachabhängigen Freitext-Qualifier abgefasst wurden	fakultativ
`nonEnglishFreeTextLanguageCode`	Der Sprachcode (siehe Verweis auf ISO 639-1:2002 in Absatz 9) für die Elemente `NonEnglishQualifier_value`	obligatorisch, wenn das Element `NonEnglishQualifier_value` im Sequenzprotokoll vorkommt

44. Das folgende Beispiel zeigt das Wurzelelement `ST26SequenceListing` einer XML-Instanz und seine Attribute gemäß Absatz 43:

```
<ST26SequenceListing dtdVersion="V1_3" fileName="US11-405455-SEQL.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="de" nonEnglishFreeTextLanguageCode="fr">
  {...}*
</ST26SequenceListing>

*{...} steht für den allgemeinen Informationsteil und den Sequenzdatenteil, die in
diesem Beispiel nicht enthalten sind.
```

Allgemeiner Informationsteil

45. Die Elemente des Allgemeinen Informationsteils beziehen sich wie folgt auf Informationen zur Patentanmeldung:

Element	Beschreibung	obligatorisch/ fakultativ
ApplicationIdentification Das Element ApplicationIdentification setzt sich zusammen aus:	Kennung der Anmeldung, für die das Sequenzprotokoll eingereicht wird	obligatorisch, wenn zu einem beliebigen Zeitpunkt nach der Zuweisung des Aktenzeichens ein Sequenzprotokoll eingereicht wird
IPOfficeCode	ST.3-Code des Anmeldeamts.	obligatorisch
ApplicationNumberText	Das vom Anmeldeamt vergebene Aktenzeichen der Anmeldung (z. B. PCT/IB2013/099999)	obligatorisch
FilingDate	Der Anmeldetag der Patentanmeldung, für die das Sequenzprotokoll eingereicht wird (ST.2-Format "JJJJ-MM-TT", wobei für das Kalenderjahr 4 Stellen, für den Kalendermonat 2 Stellen und für den Tag innerhalb des Kalendermonats ebenfalls 2 Stellen verwendet werden, z. B. 2015-01-31)	obligatorisch, wenn das Sequenzprotokoll zu einem beliebigen Zeitpunkt nach der Zuordnung eines Anmeldetags eingereicht wird
ApplicantFileReference	Eine einzige eindeutige Kennung, die vom Anmelder zur Identifizierung einer bestimmten Anmeldung zugewiesen und unter Verwendung der in Absatz 40(b) genannten Zeichen eingegeben wird	obligatorisch, wenn zu einem beliebigen Zeitpunkt vor der Zuweisung des Aktenzeichens ein Sequenzprotokoll eingereicht wird, ansonsten fakultativ
EarliestPriorityApplicationIdentification	Kennung der frühesten Prioritätsanmeldung (enthält auch IPOfficeCode, ApplicationNumberText und FilingDate, siehe ApplicationIdentification oben)	obligatorisch, wenn Priorität beansprucht wird
ApplicantName	Name des erstgenannten Anmelders unter Verwendung der in Absatz 40(a) vorgeschriebenen Zeichen. Dieses Element enthält das obligatorische Attribut languageCode, wie in Absatz 47 festgelegt.	obligatorisch
ApplicantNameLatin	Wird ApplicantName unter Verwendung anderer als der in Absatz 40(b) genannten Zeichen eingegeben, so muss zusätzlich eine Übersetzung bzw. Transliteration des Namens des erstgenannten Anmelders unter Verwendung der in Absatz 40(b) vorgeschriebenen Zeichen eingegeben werden.	obligatorisch, wenn ApplicantName nicht lateinische Zeichen enthält

Element	Beschreibung	obligatorisch/ fakultativ
`InventorName`	Name des erstgenannten Erfinders unter Verwendung der in Absatz 40(a) vorgeschriebenen Zeichen. Dieses Element enthält das obligatorische Attribut `languageCode`, wie in Absatz 47 festgelegt.	fakultativ
`InventorNameLatin`	Wird `InventorName` in anderen als den in Absatz 40(b) genannten Zeichen eingegeben, so kann auch eine Übersetzung oder Transliteration des Namens des erstgenannten Erfinders unter Verwendung der in Absatz 40(b) vorgeschriebenen Zeichen eingegeben werden.	fakultativ
`InventionTitle`	Angabe der Bezeichnung der Erfindung unter Verwendung der in Absatz 40(a) vorgeschriebenen Zeichen in der Sprache der Anmeldung. Übersetzungen der Bezeichnung der Erfindung in weitere Sprachen können mit den in Absatz 40(a) vorgeschriebenen Zeichen unter Verwendung zusätzlicher `InventionTitle`-Elemente eingegeben werden. Dieses Element enthält das obligatorische Attribut `languageCode`, wie in Absatz 48 festgelegt. Die Bezeichnung der Erfindung sollte aus zwei bis sieben Wörtern bestehen.	obligatorisch in der Sprache der Anmeldung, fakultativ für weitere Sprachen
`SequenceTotalQuantity`	Die Gesamtzahl aller Sequenzen im Sequenzprotokoll einschließlich absichtlich übersprungener Sequenzen (auch als „leere Sequenzen“ bezeichnet) (siehe Abschnitt 10).	obligatorisch

46. Die folgenden Beispiele veranschaulichen die Darstellung in dem allgemeinen Informationsteil zum Sequenzprotokoll gemäß Absatz 45:

Beispiel 1: Sequenzprotokoll, das vor der Zuweisung des Aktenzeichens der Anmeldung und Zuordnung des Anmeldetags eingereicht wird

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="Invention_SEQL.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="ja">
    <ApplicantFileReference>AB123</ApplicantFileReference>
    <EarliestPriorityApplicationIdentification>
        <IPOfficeCode>IB</IPOfficeCode>
        <ApplicationNumberText>PCT/IB2013/099999</ApplicationNumberText>
        <FilingDate>2014-07-10</FilingDate>
    </EarliestPriorityApplicationIdentification>
    <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
    <InventorName languageCode="en">Keiko Nakamura</InventorName>
    <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
    <SequenceTotalQuantity>9</SequenceTotalQuantity>
    <SequenceData sequenceIDNumber="1"> {...}* </SequenceData>
    <SequenceData sequenceIDNumber="2"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="3"> {...} </SequenceData>
```

```
    <SequenceData sequenceIDNumber="4"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="5"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="6"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="7"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="8"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>

*{...} steht für relevante Informationen für jede Sequenz, die in diesem Beispiel
nicht aufgeführt sind.
```

Beispiel 2:Sequenzprotokoll, das nach der Zuweisung des Aktenzeichens der Anmeldung und der Zuordnung des Anmeldetags eingereicht wird

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="1_3" fileName="Invention_SEQL.xml" softwareName="WIPO
Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="ja">
     <ApplicationIdentification>
        <IPOfficeCode>US</IPOfficeCode>
        <ApplicationNumberText>14/999,999</ApplicationNumberText>
        <FilingDate>2015-01-05</FilingDate>
    </ApplicationIdentification>
    <ApplicantFileReference>AB123</ApplicantFileReference>
    <EarliestPriorityApplicationIdentification>
        <IPOfficeCode>IB</IPOfficeCode>
        <ApplicationNumberText>PCT/IB2014/099999</ApplicationNumberText>
        <FilingDate>2014-07-10</FilingDate>
    </EarliestPriorityApplicationIdentification>
    <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
    <InventorName languageCode="en">Keiko Nakamura</InventorName>
    <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
    <SequenceTotalQuantity>9</SequenceTotalQuantity>
    <SequenceData sequenceIDNumber="1"> {...}* </SequenceData>
    <SequenceData sequenceIDNumber="2"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="3"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="4"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="5"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="6"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="7"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="8"> {...} </SequenceData>
    <SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>*

{...} steht für relevante Informationen für jede Sequenz, die in diesem Beispiel nicht
aufgeführt sind.
```

47. Der Name des Anmelders und fakultativ der Name des Erfinders müssen im Element `ApplicantName` bzw. `InventorName` angegeben werden, wie sie in der Sprache, in der die Anmeldung eingereicht wird, allgemein verwendet werden. Der zutreffende Sprachcode (siehe Verweis auf ISO 639-1:2002 in Absatz 9) muss für jedes Element im Attribut `languageCode` angegeben werden. Enthält der angegebene Name des Anmelders andere Zeichen als die des lateinischen Alphabets gemäß Absatz 40(b), so ist im Element `ApplicantNameLatin` zusätzlich eine Transliteration oder Übersetzung des Namens des Anmelders in Zeichen des lateinischen Alphabets anzugeben. Enthält der angegebene Name des Erfinders andere Zeichen als die des lateinischen Alphabets, so kann im Element `InventorNameLatin` zusätzlich eine Transliteration oder Übersetzung des Namens des Erfinders in Zeichen des lateinischen Alphabets angegeben werden.

48. Die Bezeichnung der Erfindung muss im Element `InventionTitle` in der Sprache der Anmeldung angegeben werden und kann unter Verwendung mehrerer `InventionTitle`-Elemente auch in weiteren Sprachen angegeben werden (siehe Tabelle in Absatz 45).[4] Im Attribut `languageCode` des Elements muss der zutreffende Sprachcode (siehe Verweis auf ISO 639-1:2002 in Absatz 9) angegeben werden.

49. Das folgende Beispiel veranschaulicht die Darstellung der Namen der Personen und der Bezeichnung der Erfindung gemäß den vorstehenden Absätzen 47 und 48:

Beispiel: Der Name des Anmelders und der Name des Erfinders werden jeweils in japanischer und lateinischer Schrift dargestellt, und die Bezeichnung der Erfindung wird in japanischer, englischer und französischer Sprache dargestellt

```
<ApplicantName languageCode="ja">出願製薬株式会社</ApplicantName>
<ApplicantNameLatin>Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantNameLatin>
<InventorName languageCode="ja">特許 太郎</InventorName>
<InventorNameLatin>Taro Tokkyo</InventorNameLatin>
<InventionTitle languageCode="ja">efgタンパク質をコードするマウスabcd-1遺伝子
</InventionTitle>
<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg
protein</InventionTitle>
<InventionTitle languageCode="fr">Gène abcd-1 de Mus musculus pour protéine
efg</InventionTitle>
```

Sequenzdatenteil

50. Der Sequenzdatenteil muss aus einem oder mehreren `SequenceData`-Elementen bestehen, die jeweils Informationen über eine Sequenz enthalten.

51. Jedes `SequenceData`-Element muss mit dem obligatorischen Attribut `sequenceIDNumber` versehen sein, in dem die Sequenzkennzahl (siehe Absatz 10) für die jeweilige Sequenz enthalten ist. Zum Beispiel:

```
<SequenceData sequenceIDNumber="1">
```

52. Das `SequenceData`-Element muss das abhängige Element `INSDSeq` enthalten, das wie folgt aus weiteren abhängigen Elementen besteht:

Element	Beschreibung	obligatorisch/nicht enthalten	
		Sequenzen	absichtlich übersprungene Sequenzen
`INSDSeq_length`	Länge der Sequenz	obligatorisch	obligatorisch ohne Wert
`INSDSeq_moltype`	Molekültyp	obligatorisch	obligatorisch ohne Wert
`INSDSeq_division`	Hinweis darauf, dass eine Sequenz mit einer Patentanmeldung in Verbindung steht	obligatorisch mit dem Wert "PAT"	obligatorisch ohne Wert
`INSDSeq_feature-table`	Liste der Annotationen der Sequenz	obligatorisch	darf NICHT enthalten sein
`INSDSeq_sequence`	Sequenz	obligatorisch	obligatorisch mit dem Wert "000"

53. Im Element `INSDSeq_length` muss die Anzahl der Nukleotide oder Aminosäuren der im Element `INSDSeq_sequence` enthaltenen Sequenz offenbart werden. Zum Beispiel:

```
<INSDSeq_length>8</INSDSeq_length>
```

[4] Nach § 11a Absatz 2 der Patentverordnung ist die Bezeichnung der Erfindung im Sequenzprotokoll in deutscher Sprache anzugeben. Sie kann zusätzlich auch in weiteren Sprachen angegeben werden.

54. Im Element `INSDSeq_moltype` muss der Typ des dargestellten Moleküls offenbart werden. Bei Nukleotidsequenzen, einschließlich Sequenzen aus Nukleotid-Analoga, muss als Molekültyp DNA oder RNA angegeben werden. Bei Aminosäuresequenzen muss als Molekültyp AA angegeben werden. (Dieses Element ist zu unterscheiden von dem in den Absätzen 55 und 84 beschriebenen Qualifier "mol_type"). Zum Beispiel:

```
<INSDSeq_moltype>AA</INSDSeq_moltype>
```

55. Bei einer Nukleotidsequenz, die sowohl DNA- als auch RNA-Segmente mit einem oder mehreren Nukleotiden enthält, muss als Molekültyp DNA angegeben werden. Das kombinierte DNA/RNA-Molekül muss in der Merkmaltabelle näher beschrieben werden. Dabei sind der Merkmalschlüssel "source", der obligatorische Qualifier "organism" mit dem Wert "synthetic construct" ("synthetisches Konstrukt") und der obligatorische Qualifier "mol_type" mit dem Wert "other DNA" zu verwenden. Jedes DNA- und RNA-Segment des kombinierten DNA/RNA-Moleküls muss mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note" näher beschrieben werden, wobei der Qualifier angibt, ob es sich bei dem Segment um DNA oder RNA handelt.

56. Das folgende Beispiel veranschaulicht die Beschreibung einer Nukleotidsequenz, die, wie in Absatz 55 beschrieben, sowohl DNA- als auch RNA-Segmente enthält:

```
<INSDSeq>
    <INSDSeq_length>120</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
        <INSDFeature>
            <INSDFeature_key>source</INSDFeature_key>
            <INSDFeature_location>1..120</INSDFeature_location>
            <INSDFeature_quals>
                <INSDQualifier>
                    <INSDQualifier_name>organism</INSDQualifier_name>
                    <INSDQualifier_value>synthetic construct</INSDQualifier_value>
                </INSDQualifier>
                <INSDQualifier>
                    <INSDQualifier_name>mol_type</INSDQualifier_name>
                    <INSDQualifier_value>other DNA</INSDQualifier_value>
                </INSDQualifier>
            </INSDFeature_quals>
        </INSDFeature>
        <INSDFeature>
            <INSDFeature_key>misc_feature</INSDFeature_key>
            <INSDFeature_location>1..60</INSDFeature_location>
            <INSDFeature_quals>
                <INSDQualifier>
                    <INSDQualifier_name>note</INSDQualifier_name>
                    <INSDQualifier_value>DNA</INSDQualifier_value>
                </INSDQualifier>
            </INSDFeature_quals>
        </INSDFeature>
        <INSDFeature>
            <INSDFeature_key>misc_feature</INSDFeature_key>
            <INSDFeature_location>61..120</INSDFeature_location>
            <INSDFeature_quals>
                <INSDQualifier>
                    <INSDQualifier_name>note</INSDQualifier_name>
                    <INSDQualifier_value>RNA</INSDQualifier_value>
                </INSDQualifier>
            </INSDFeature_quals>
        </INSDFeature>
    </INSDSeq_feature-table>
<INSDSeq_sequence>cgacccacgcgtccgaggaaccaaccatcacgtttgaggacttcgtgaaggaattggataatacccgt
ccctaccaaaatggcgagcgccgactcattgctcctcgtaccgtcgagcggc</INSDSeq_sequence>
</INSDSeq>
```

57. Im Element `INSDSeq_sequence` muss die Sequenz offenbart werden. In die Sequenz dürfen nur die jeweils zutreffenden, in Anhang I (siehe Abschnitt 1, Tabelle 1 und Abschnitt 3, Tabelle 3) aufgeführten Symbole aufgenommen werden. Die Sequenz darf keine Zahlen, Satzzeichen oder Whitespaces enthalten.

58. Eine absichtlich übersprungene Sequenz muss in das Sequenzprotokoll aufgenommen und wie folgt dargestellt werden:

(a) das Element `SequenceData` und sein Attribut `sequenceIDNumber`, für das die Sequenzkennzahl der übersprungenen Sequenz als Wert angegeben wird;

(b) die Elemente `INSDSeq_length`, `INSDSeq_moltype` und `INSDSeq_division`, jedoch ohne Angabe eines Werts;

(c) das Element `INSDSeq_feature-table` darf nicht enthalten sein und

(d) das Element `INSDSeq_sequence` mit der Zeichenkette "000" als Wert.

59. Das folgende Beispiel veranschaulicht die Darstellung einer absichtlich übersprungenen Sequenz gemäß dem obigen Absatz 58:

```
<SequenceData sequenceIDNumber="3">
    <INSDSeq>
        <INSDSeq_length/>
        <INSDSeq_moltype/>
        <INSDSeq_division/>
        <INSDSeq_sequence>000</INSDSeq_sequence>
    </INSDSeq>
</SequenceData>
```

Merkmaltabelle

60. Die Merkmaltabelle enthält Informationen über die Lage und Funktionen verschiedener Regionen innerhalb einer bestimmten Sequenz. Eine Merkmaltabelle ist für jede Sequenz erforderlich. Eine Ausnahme bilden lediglich absichtlich übersprungene Sequenzen, für die keine Merkmaltabelle angegeben werden darf. Die Merkmaltabelle ist im Element `INSDSeq_feature-table` enthalten, das aus einem oder mehreren `INSDFeature`-Elementen besteht.

61. Jedes `INSDFeature`-Element beschreibt ein Merkmal und besteht wie folgt aus abhängigen Elementen:

Element	Beschreibung	obligatorisch/fakultativ
`INSDFeature_key`	Ein Wort oder eine Abkürzung zur Angabe eines Merkmals	obligatorisch
`INSDFeature_location`	Region der Sequenz, die dem Merkmal entspricht	obligatorisch
`INSDFeature_quals`	Qualifier mit zusätzlichen Informationen zu einem Merkmal	obligatorisch, wenn für den Merkmalschlüssel (z. B. "source") ein oder mehrere Qualifier erforderlich sind, andernfalls fakultativ

Merkmalschlüssel

62. In Anhang I sind die nach diesem Standard ausschließlich zu verwendenden Merkmalschlüssel (feature keys) mitsamt den zugehörigen, ebenfalls ausschließlich zu verwendenden Qualifiern aufgelistet. Dort ist auch angegeben, ob diese Qualifier obligatorisch oder fakultativ sind. In Anhang I sind in Abschnitt 5 die ausschließlich zu verwendenden Merkmalschlüssel für Nukleotidsequenzen und in Abschnitt 7 die ausschließlich zu verwendenden Merkmalschlüssel für Aminosäuresequenzen aufgelistet.

Obligatorische Merkmalschlüssel

63. Der Merkmalschlüssel "source" ist für alle Nukleotidsequenzen und für alle Aminosäuresequenzen obligatorisch, mit Ausnahme absichtlich übersprungener Sequenzen. Jede Sequenz muss mit einem einzelnen Merkmalschlüssel "source" versehen sein, der die gesamte Sequenz umfasst. Wenn eine Sequenz mehreren Herkunftsquellen entstammt, können diese in der Merkmaltabelle näher beschrieben werden. Hierzu werden für Nukleotidsequenzen der Merkmalschlüssel "misc_feature" und der Qualifier "note" sowie für Aminosäuresequenzen der Merkmalschlüssel "REGION" und der Qualifier "note" verwendet.

Lage des Merkmals

64. Das obligatorische Element `INSDFeature_location` muss mindestens einen Lagedeskriptor enthalten, der eine Stelle oder Region definiert, die einem Merkmal der im Element `INSDSeq_sequence` angegebenen Sequenz entspricht. Aminosäuresequenzen müssen im obligatorischen Element `INSDFeature_location` exakt einen Lagedeskriptor enthalten. Nukleotidsequenzen dürfen im obligatorischen Element `INSDFeature_location` mehr als einen Lagedeskriptor aufweisen, wenn diese in Verbindung mit einem oder mehreren Lageoperatoren verwendet werden (siehe Absätze 67 bis 70).

65. Der Lagedeskriptor kann eine einzelne Restenummer sein, eine Region, die ein zusammenhängendes Intervall von Restenummern begrenzt, oder eine Stelle oder Region, die über den angegebenen Rest oder das angegebene Resteintervall hinausgeht. Der Lagedeskriptor darf keine Nummerierung für Reste, die über den Bereich der Sequenz im `INSDSeq_sequence`-Element hinausgehen, enthalten. Nur bei Nukleotidsequenzen kann ein Lagedeskriptor eine Stelle zwischen zwei benachbarten Restenummern sein. Wenn ein Merkmal diskontinuierlichen Stellen oder Regionen einer Nukleotidsequenz entspricht, müssen mehrere Lagedeskriptoren in Verbindung mit einem Lageoperator verwendet werden (siehe Absätze 67 bis 70).

66. In der folgenden Tabelle wird die Syntax für jede Art des Lagedeskriptors angegeben. Dabei gilt: x und y sind als positive Ganzzahlen angegebene Restenummern, die nicht größer sind als die Länge der im `INSDSeq_sequence`-Element angegebenen Sequenz; und x ist kleiner als y.

(a) Lagedeskriptoren für Nukleotid- und Aminosäuresequenzen:

Art des Lagedeskriptors	Syntax	Beschreibung
Einzelne Restenummer	`x`	Verweist auf einen einzelnen Rest in der Sequenz
Restenummern, die ein Sequenzintervall begrenzen	`x..y`	Verweist auf einen kontinuierlichen Bereich von Resten, der durch den ersten und letzten dieser Reste begrenzt wird und diese mit einschließt.

Reste vor der ersten oder nach der letzten angegebenen Restenummer	`<x`
	`>x`
	`<x..y`
	`x..>y`
	`<x..>y`
		Verweist auf eine Region, die einen bestimmten Rest oder ein bestimmtes Resteintervall enthält und über einen bestimmten Rest hinausgeht. Die Symbole "<" und ">" können in Verbindung mit einer einzelnen Restenummer oder mit der ersten und letzten Restenummer eines Resteintervalls verwendet werden, um anzuzeigen, dass ein Merkmal über die angegebene Restenummer hinausgeht.

(b) Lagedeskriptoren nur für Nukleotidsequenzen:

Art des Lagedeskriptors	Syntax	Beschreibung
Eine Stelle zwischen zwei benachbarten Nukleotiden	`x^y`	Verweist auf eine Stelle zwischen zwei benachbarten Nukleotiden, z. B. die Stelle einer endonukleolytischen Spaltung. Die Positionsnummern für benachbarte Nukleotide werden durch ein Caret (^) getrennt. Die zulässigen Formate für diesen Deskriptor sind x^x+1 (z. B. 55^56) oder, für zirkuläre Nukleotide, x^1, wobei "x" die volle Länge des Moleküls ist, d. h. 1000^1 steht für ringförmige Moleküle der Länge 1000.

(c) Lagedeskriptoren nur für Aminosäuresequenzen:

Art des Lagedeskriptors	Syntax	Beschreibung
Restenummern, die durch eine ketteninterne Quervernetzung zusammengeführt werden	`x..y`	Verweist auf Aminosäuren, die durch eine ketteninterne Quervernetzung zusammengeführt werden, wenn der Syntaxausdruck mit einem Merkmal verwendet wird, das auf eine solche Quervernetzung hinweist, z. B. "CROSSLNK" oder "DISULFID".

67. Das Element `INSDFeature_location` von Nukleotidsequenzen kann einen oder mehrere Lageoperatoren enthalten. Ein Lageoperator ist ein Präfix entweder für einen Lagedeskriptor oder für eine Kombination von Lagedeskriptoren, die ein einzelnes, aber diskontinuierliches Merkmal beschreiben. Er gibt an, wo sich die dem Merkmal entsprechende Lage in der angegebenen Sequenz befindet oder wie das Merkmal aufgebaut ist. Unten folgt eine Liste der Lageoperatoren nebst Definitionen. Lageoperatoren können nur für Nukleotide verwendet werden.

Syntax für Lage	Beschreibung der Lage
`join(Lage,Lage,…,Lage)`	Die angegebenen Lagen werden zu einer zusammenhängenden Sequenz verbunden (von einem Ende zum anderen).
`order(Lage,Lage,…,Lage)`	Die Elemente befinden sich in der angegebenen Reihenfolge, aber es wird nichts darüber ausgesagt, ob das Verbinden dieser Elemente sinnvoll ist.
`complement(Lage)`	Zeigt an, dass sich das Merkmal auf dem Strang befindet, der komplementär zu dem durch den Lagedeskriptor angegebenen Sequenzintervall ist, wenn es in 5'-3'-Richtung oder in der Richtung gelesen wird, die die 5'-3'-Richtung nachahmt.

68. Die Lageoperatoren "join" und "order" setzen voraus, dass mindestens zwei kommagetrennte Lagedeskriptoren vorhanden sind. Lagedeskriptoren, die sich auf Stellen zwischen zwei benachbarten Resten beziehen, d. h. x^y, dürfen nicht innerhalb eines „join"- oder „order"-Lageoperators verwendet werden. Mit der Verwendung des Lageoperators "join" wird impliziert, dass die durch die Lagedeskriptoren beschriebenen Reste durch biologische Prozesse physisch in Kontakt gebracht werden (z. B. Exons, die zum Merkmal eines Codierungsabschnitts beitragen).

69. Der Lageoperator "complement" kann innerhalb derselben Lage in Kombination mit entweder "join" oder "order" verwendet werden. "join" und "order" dürfen innerhalb derselben Lage nicht kombiniert werden.

70. Die folgenden Beispiele veranschaulichen Lagen von Merkmalen, wie in den Absätzen 64 bis 69 beschrieben:

(a) Lagen für Nukleotid- und Aminosäuresequenzen:

Beispiel für eine Lage	Beschreibung
`467`	Verweist auf Rest 467 in der Sequenz.
`340..565`	Verweist auf einen kontinuierlichen Bereich von Resten, der durch die Reste 340 und 565 begrenzt wird und diese mit einschließt.
`<1`	Verweist auf eine Lage des Merkmals vor dem ersten Rest.
`<345..500`	Gibt an, dass der genaue untere Begrenzungspunkt eines Merkmals unbekannt ist. Die Lage beginnt bei einem Rest vor 345 und reicht bis einschließlich Rest 500.
`<1..888`	Zeigt an, dass das Merkmal vor dem ersten sequenzierten Rest beginnt und bis einschließlich Rest 888 reicht.
`1..>888`	Zeigt an, dass das Merkmal am ersten sequenzierten Rest beginnt und über Rest 888 hinausreicht.
`<1..>888`	Zeigt an, dass das Merkmal vor dem ersten sequenzierten Rest beginnt und über Rest 888 hinausreicht.

(b) Lagen nur für Nukleotidsequenzen:

Beispiel für eine Lage	Beschreibung
`123^124`	Verweist auf eine Stelle zwischen den Resten 123 und 124.
`join(12..78,134..202)`	Gibt an, dass die Regionen 12 bis 78 und 134 bis 202 zu einer zusammenhängenden Sequenz verbunden werden sollten.
`complement(34..126)`	Beginnt bei dem Nukleotid, das komplementär zu Nukleotid 126 ist, und endet bei dem Nukleotid, das komplementär zu Nukleotid 34 ist (das Merkmal befindet sich auf dem Strang, der komplementär zu dem dargestellten Strang ist).
`complement(join(2691..4571, 4918..5163))`	Verbindet die Nukleotide 2691 bis 4571 und 4918 bis 5163 und komplementiert anschließend die verbundenen Segmente (das Merkmal befindet sich auf dem Strang, der komplementär zu dem

	dargestellten Strang ist).
`join(complement(4918..5163), complement(2691..4571))`	Komplementiert die Regionen 4918 bis 5163 und 2691 bis 4571 und verbindet anschließend die komplementierten Segmente (das Merkmal befindet sich auf dem Strang, der komplementär zu dem dargestellten Strang ist).

(c) Lagen nur für Aminosäuresequenzen:

Beispiel für eine Lage	Beschreibung
`340..565`	Zeigt an, dass die Aminosäuren an den Positionen 340 und 565 durch eine ketteninterne Quervernetzung zusammengeführt werden, wenn der Syntaxausdruck mit einem Merkmal verwendet wird, das auf eine solche Quervernetzung hinweist, z. B. "CROSSLNK" oder "DISULFID".

71. In einer XML-Instanz eines Sequenzprotokolls müssen die Zeichen "<" und ">" in einem Lagedeskriptor durch die entsprechenden vordefinierten Entitäten ersetzt werden (siehe Absatz 41). Zum Beispiel:

Lage des Merkmals "<1":
```
<INSDFeature_location>&lt;1</INSDFeature_location>
```

Lage des Merkmals "1..>888":
```
<INSDFeature_location>1..&gt;888</INSDFeature_location>
```

Merkmal-Qualifier

72. Qualifier dienen dazu, über den Merkmalschlüssel und die Lage des Merkmals hinaus weitere Informationen über Merkmale bereitzustellen. Für Qualifier-Werte gibt es drei Formate. Diese entsprechen den verschiedenen Arten von Informationen, die durch Qualifier dargestellt werden:

(a) Freitext (siehe Absätze 85 bis 87);

(b) kontrolliertes Vokabular oder Zahlenwerte (z. B. eine Nummer oder ein Datum); und

(c) Sequenzen.

73. Die für die Merkmalschlüssel von Nukleotidsequenzen ausschließlich zu verwendenden Qualifier und deren festgelegte Wertformate, falls vorhanden, sind in Anhang I, Abschnitt 6 im Einzelnen aufgeführt. Die für die Merkmalschlüssel von Aminosäuresequenzen ausschließlich zu verwendenden Qualifier und deren festgelegten Wertformate, falls vorhanden, sind in Abschnitt 8 im Einzelnen aufgeführt.

74. Jede von Absatz 7 erfasste Sequenz, die als Qualifier-Wert angegeben wird, muss gesondert in das Sequenzprotokoll aufgenommen und mit einer eigenen Sequenzkennzahl versehen werden (siehe Absatz 10).

Obligatorische Merkmal-Qualifier

75. Ein obligatorischer Merkmalschlüssel, nämlich "source" für Nukleotidsequenzen und Aminosäuresequenzen, erfordert zwei obligatorische Qualifier: "organism" und "mol_type". Auch einige fakultative Merkmalschlüssel bedürfen obligatorischer Qualifier.

Qualifier-Elemente

76. Das Element `INSDFeature_quals` enthält ein oder mehrere `INSDQualifier`-Elemente. Jedes `INSDQualifier`-Element stellt einen einzelnen Qualifier dar und besteht wie folgt aus drei abhängigen Elementen und einem fakultativen Attribut:

Element/Attribut	Beschreibung	obligatorisch/fakultativ
`INSDQualifier_name`	Name des Qualifiers (siehe Anhang I, Abschnitte 6 und 8)	obligatorisch
`INSDQualifier_value`	Wert des Qualifiers, sofern vorhanden, im festgelegten Format (siehe Anhang I, Abschnitte 6 und 8) und bestehend aus den in Absatz 40(b) vorgegebenen	obligatorisch, wenn angegeben (siehe Absatz 87 und Anhang I, Abschnitte 6 und 8)

	Zeichen	
NonEnglishQualifier_value	Wert des Qualifiers, sofern vorhanden, im festgelegten Format (siehe Anhang I, Abschnitte 6 und 8) und bestehend aus den in Absatz 40(a) vorgegebenen Zeichen	obligatorisch, wenn angegeben (siehe Absatz 87 und Anhang I, Abschnitte 6 und 8)
id	Ein Qualifier mit einem Wert im Format sprachabhängiger Freitext, der eindeutig über das fakultative XML-Attribut ‘id’ im Element INSDQualifier angegeben werden kann (siehe Absatz 87(d)). Der Wert des Attributs ‘id’ muss das Format ‘q’ gefolgt von einer positiven Ganzzahl haben. Der Wert des Attributs ‘id’ muss für ein INSDQualifier-Element eindeutig sein, d. h. der Attributwert darf in einer Sequenzprotokolldatei nur einmal verwendet werden.	fakultativ

77. Der Organismus-Qualifier, d. h. "organism" für Nukleotidsequenzen (siehe Anhang I, Abschnitt 6) und "organism" für Aminosäuresequenzen (siehe Anhang I, Abschnitt 8), muss die Herkunft der Sequenz, d. h. einen einzelnen Organismus oder Ursprung, offenbaren. Die Bezeichnungen von Organismen sollten einer Taxonomiedatenbank entstammen.

78. Wenn die Sequenz natürlich vorkommt und der Herkunftsorganismus einen lateinischen Gattungs- und Artnamen hat, muss diese Bezeichnung als Qualifier-Wert verwendet werden. Der bevorzugte englische gebräuchliche Name kann für Nukleotidsequenzen und für Aminosäuresequenzen mittels des Qualifiers "note" angegeben werden, darf jedoch nicht als Wert des Organismus-Qualifiers verwendet werden.

79. Die folgenden Beispiele veranschaulichen die Darstellung des Herkunftsorganismus einer Sequenz, wie in den Absätzen 77 und 78 beschrieben:

Beispiel 1: Herkunft einer Nukleotidsequenz

```
<INSDSeq_feature-table>
    <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..5164</INSDFeature_location>
        <INSDFeature_quals>
            <INSDQualifier>
                <INSDQualifier_name>organism</INSDQualifier_name>
                <INSDQualifier_value>Solanum lycopersicum</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier>
                <INSDQualifier_name>note</INSDQualifier_name>
                <INSDQualifier_value>common name: tomato</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier>
                <INSDQualifier_name>mol_type</INSDQualifier_name>
                <INSDQualifier_value>genomic DNA</INSDQualifier_value>
            </INSDQualifier>
        </INSDFeature_quals>
    </INSDFeature>
</INSDSeq_feature-table>
```

Beispiel 2: Herkunft einer Aminosäuresequenz

```
<INSDSeq_feature-table>
    <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..174</INSDFeature_location>
        <INSDFeature_quals>
            <INSDQualifier>
                <INSDQualifier_name>organism</INSDQualifier_name>
```

```
                    <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
                </INSDQualifier>
                <INSDQualifier>
                    <INSDQualifier_name>mol_type</INSDQualifier_name>
                    <INSDQualifier_value>protein</INSDQualifier_value>
                </INSDQualifier>
            </INSDFeature_quals>
        </INSDFeature>
    </INSDSeq_feature-table>
```

80. Wenn die Sequenz natürlich vorkommt und der Herkunftsorganismus eine bekannte lateinische Gattung ist, aber die Spezies nicht spezifiziert oder nicht identifiziert ist, muss als Wert des Organismus-Qualifiers die lateinische Gattung gefolgt von "sp." eingetragen werden. Zum Beispiel:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Bacillus sp.</INSDQualifier_value>
```

81. Wenn die Sequenz natürlich vorkommt, aber der lateinische Gattungs- und Artname des Organismus unbekannt ist, muss als Wert des Organismus-Qualifiers "unidentified" ("unidentifiziert") eingetragen werden. Alle bekannten taxonomischen Informationen sollten für Nukleotidsequenzen und für Aminosäuresequenzen im Qualifier "note" angegeben werden. Zum Beispiel:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>unidentified</INSDQualifier_value>
<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>bacterium B8</INSDQualifier_value>
```

82. Wenn die Sequenz natürlich vorkommt und der Herkunftsorganismus, wie z. B. ein Virus, keinen lateinischen Gattungs- und Artnamen hat, muss als Wert des Organismus-Qualifiers ein anderer akzeptierter wissenschaftlicher Name (z. B. "Canine adenovirus type 2" ("Canines Adenovirus Typ 2")) verwendet werden. Zum Beispiel:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Canine adenovirus type 2</INSDQualifier_value>
```

83. Wenn die Sequenz nicht natürlich vorkommt, muss der Wert des Organismus-Qualifiers als "synthetic construct" ("synthetisches Konstrukt") angegeben werden. Weitere Informationen über die Art und Weise, wie die Sequenz erzeugt wurde, können für Nukleotidsequenzen für Aminosäuresequenzen mittels des Qualifiers "note" angegeben werden. Zum Beispiel:

```
<INSDSeq_feature-table>
    <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..40</INSDFeature_location>
        <INSDFeature_quals>
            <INSDQualifier>
                <INSDQualifier_name>organism</INSDQualifier_name>
                <INSDQualifier_value>synthetic construct</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier>
                <INSDQualifier_name>mol_type</INSDQualifier_name>
                <INSDQualifier_value>protein</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier>
                <INSDQualifier_name>note</INSDQualifier_name>
                <INSDQualifier_value>synthetic peptide used as assay for
antibodies</INSDQualifier_value>
            </INSDQualifier>
        </INSDFeature_quals>
    </INSDFeature>
</INSDSeq_feature-table>
```

84. Der Molekültyp-Qualifier, d. h. "mol_type" für Nukleotidsequenzen (siehe Anhang I, Abschnitt 6) und "mol_type" für Aminosäuresequenzen (siehe Anhang I, Abschnitt 8) muss den Typ des in der Sequenz dargestellten Moleküls offenbaren. Dieser Qualifier ist zu unterscheiden von dem Element `INSDSeq_moltype`, das in Absatz 54 erläutert wird:

(a) Der Wert für den Qualifier "mol_type" für eine Nukleotidsequenz muss einer der folgenden sein: "genomic DNA", "genomic RNA", "mRNA", "tRNA", "rRNA", "other RNA", "other DNA", "transcribed RNA", "viral cRNA", "unassigned DNA" oder "unassigned RNA". Wenn die Sequenz nicht natürlich vorkommt, d. h. der Wert des "organism"-Qualifiers "synthetic construct" ("synthetisches Konstrukt") ist, muss der Wert des Qualifiers "mol_type" entweder "other RNA" oder "other DNA" sein.

(b) Für Aminosäuresequenzen ist der Wert des "mol_type"-Qualifiers "protein".

Freitext

85. Freitext ist, wie in Absatz 3(n) erläutert, ein Format für den Wert bestimmter Qualifier, der in Form eines beschreibenden Textes oder in einem anderen (in Anhang I aufgeführten) festgelegten Format angegeben wird.

86. Die Verwendung von Freitext muss auf einige kurze Begriffe beschränkt werden, die für das Verständnis einer Eigenschaft der Sequenz unerlässlich sind. Der Freitext darf, außer beim Qualifier “translation”, pro Qualifier 1000 Zeichen nicht überschreiten.

87. Sprachabhängiger Freitext ist, wie in Absatz 3(o) erläutert, der Freitextwert bestimmter Qualifier, der zu Zwecken internationaler, nationaler oder regionaler Verfahren möglicherweise übersetzt werden muss.[5] Für Nukleotidsequenzen sind die Qualifier, deren Werte als sprachabhängiger Freitext angegeben werden, in Anhang I, Abschnitt 6, Tabelle 5 aufgeführt. Für Aminosäuresequenzen sind die Qualifier, deren Werte als sprachabhängiger Freitext angegeben werden, in Anhang I, Abschnitt 8, Tabelle 6 aufgeführt.

(a) Sprachabhängiger Freitext muss im Element `INSDQualifier_value` auf Englisch oder im Element `NonEnglishQualifier_value` in einer anderen Sprache als Englisch oder entsprechend in beiden Elementen angegeben werden. Zu beachten ist, dass keine Übersetzung erforderlich ist, wenn der Name eines Organismus ein lateinischer Gattungs- und Artname ist. Anderen Sprachen als Englisch entstammende Fachtermini und Eigennamen, die international verwendet werden, gelten für die Zwecke des Werts des `INSDQualifier_value`-Elements als Englisch (z. B. "in vitro" oder "in vivo").

(b) Wenn das Element `NonEnglishQualifier_value` in einem Sequenzprotokoll vorkommt, muss der entsprechende Sprachcode (siehe Verweis auf ISO 639-1:2002 in Absatz 9) im Attribut `nonEnglishFreeTextLanguageCode` im Wurzelelement angegeben werden (siehe Absatz 43). Die Werte in den `NonEnglishQualifier_value`-Elementen ein und desselben Sequenzprotokolls müssen alle in der Sprache gehalten sein, die im Attribut `nonEnglishFreeTextLanguageCode` angegeben ist. Das Element `NonEnglishQualifier_value` ist nur für Qualifier zulässig, für deren Wert das Format sprachabhängiger Freitext vorgesehen ist.

(c) Wenn für einen einzelnen Qualifier sowohl das Element `NonEnglishQualifier_value` als auch das Element `INSDQualifier_value` vorhanden sind, müssen die in beiden Elementen enthaltenen Informationen gleichwertig sein. Eine der folgenden Bedingungen muss erfüllt sein: `NonEnglishQualifier_value` enthält eine Übersetzung des Werts von `INSDQualifier_value` oder `INSDQualifier_value` enthält eine Übersetzung des Werts von `NonEnglishQualifier_value` oder beide Elemente enthalten eine Übersetzung des Qualifier-Werts aus der Sprache, die im Attribut `originalFreeTextLanguageCode` angegeben ist (siehe Absatz 43).

(d) Bei Qualifiern mit dem Wertformat “sprachabhängiger Freitext” kann das Element `INSDQualifier` ein fakultatives Attribut `id` enthalten. Der Wert dieses Attributs muss das Format "q" gefolgt von einer positiven Ganzzahl haben, z. B. "q23", und muss für ein `INSDQualifier`-Element eindeutig sein, d. h. der Attributwert darf in einer Sequenzprotokolldatei nur einmal verwendet werden.

88. Die folgenden Beispiele veranschaulichen die Darstellung sprachabhängiger Freitexte, wie in Absatz 87 erläutert.

Beispiel 1: Sprachabhängiger Freitext in einem `INSDQualifier_value`-Element:

```
<INSDFeature>
      <INSDFeature_key>regulatory</INSDFeature_key>
      <INSDFeature_location>1..60</INSDFeature_location>
      <INSDFeature_quals>
        <INSDQualifier id="q1">
               <INSDQualifier_name>function</INSDQualifier_name>
               <INSDQualifier_value>binds to regulatory protein Est3</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_quals>
</INSDFeature>
```

Beispiel 2: Sprachabhängiger Freitext in einem `INSDQualifier_value`-Element und einem `NonEnglishQualifier_Value`-Element:

```
<INSDFeature>
```

[5] Nach § 11a Absatz 3 der Patentverordnung ist sprachabhängiger Freitext in deutscher Sprache abzufassen. Er kann zusätzlich auch in englischer Sprache angegeben werden.

```
        <INSDFeature_key>ACT_SITE</INSDFeature_key>
        <INSDFeature_location>51..64</INSDFeature_location>
        <INSDFeature_quals>
            <INSDQualifier id="q45">
                <INSDQualifier_name>note</INSDQualifier_name>
                <INSDQualifier_value>cleaves carbohydrate chain</INSDQualifier_value>
                <NonEnglishQualifier_Value>spaltet die Kohlenhydratkette
                </NonEnglishQualifier_value>
            </INSDQualifier>
        </INSDFeature_quals>
</INSDFeature>
```

Beispiel 3: Sprachabhängiger Freitext in einem `NonEnglishQualifier_value`-Element:

```
<INSDFeature>
        <INSDFeature_key>ACT_SITE</INSDFeature_key>
        <INSDFeature_location>51..64</INSDFeature_location>
        <INSDFeature_quals>
            <INSDQualifier id="q1034">
                <INSDQualifier_name>note</INSDQualifier_name>
                <NonEnglishQualifier_Value>spaltet die Kohlenhydratkette
                </NonEnglishQualifier_value>
            </INSDQualifier>
        </INSDFeature_quals>
</INSDFeature>
```

Codierende Sequenzen

89. Der Merkmalschlüssel "CDS" kann verwendet werden, um codierende Sequenzen zu identifizieren, d. h. Sequenzen von Nukleotiden, die mit der Sequenz von Aminosäuren in einem Protein einschließlich Stopcodon korrespondieren. Die Angabe der Lage des Merkmals "CDS" im obligatorischen Element `INSDFeature_location` muss auch das Stopcodon enthalten.

90. In Verbindung mit dem Merkmalschlüssel "CDS" können die Qualifier "transl_table" und "translation" verwendet werden (siehe Anhang I). Wenn der Qualifier "transl_table" nicht verwendet wird, wird die Verwendung der Tabelle für den Standardcode (siehe Anhang I, Abschnitt 9, Tabelle 7) angenommen.

91. Zur Identifizierung eines Codons, das entweder für Pyrrolysin oder für Selenocystein codiert, muss der Qualifier "transl_except" zusammen mit dem Merkmalschlüssel "CDS" und dem Qualifier "translation" verwendet werden.

92. Eine von Absatz 7 erfasste Aminosäuresequenz, die von der codierenden Sequenz codiert und in einem "translation"-Qualifier offenbart ist, muss in das Sequenzprotokoll aufgenommen und mit einer eigenen Sequenzkennzahl versehen werden. Die der Aminosäuresequenz zugewiesene Sequenzkennzahl muss als Wert im Qualifier "protein_id" in Verbindung mit dem Merkmalschlüssel "CDS" angegeben werden. Der Qualifier "organism" des Merkmalschlüssels "source" für die Aminosäuresequenz muss mit dem Qualifier seiner codierenden Sequenz identisch sein. Zum Beispiel:

```
          <INSDFeature>
              <INSDFeature_key>CDS</INSDFeature_key>
              <INSDFeature_location>1..507</INSDFeature_location>
              <INSDFeature_quals>
                  <INSDQualifier>
                      <INSDQualifier_name>transl_table</INSDQualifier_name>
                      <INSDQualifier_value>11</INSDQualifier_value>
                  </INSDQualifier>
                  <INSDQualifier>
                      <INSDQualifier_name>translation</INSDQualifier_name>
                      <INSDQualifier_value>MLVHLERTTIMFDFSSLINLPLIWGLLIAIAVLLYILMDGFDLGIGILL
                      PFAPSDKCRDHMISSIAPFWDGNETWLVLGGGGLFAAFPLAYSILMPAFYIPIIIMLLGLIVRGVSFEFR
                      FKAEGKYRRLWDYAFHFGSLGAAFCQGMILGAFIHGVEVNGRNFSGGQLM
                      </INSDQualifier_value>
                  </INSDQualifier>
                  <INSDQualifier>
                  <INSDQualifier_name>protein_id</INSDQualifier_name>
                  <INSDQualifier_value>89</INSDQualifier_value>
                  </INSDQualifier>
```

```
            </INSDFeature_quals>
        </INSDFeature>
```

Varianten

93. Eine Primärsequenz und jede Variante dieser Sequenz, die jeweils durch eine Aufzählung ihrer Reste offenbart und von Absatz 7 erfasst werden, müssen jeweils in das Sequenzprotokoll aufgenommen und mit einer eigenen Sequenzkennzahl versehen werden.

94. Jede Sequenzvariante, die an einer oder mehreren Positionen als einzelne Sequenz mit durch Aufzählung dargestellten alternativen Resten offenbart wird, muss in das Sequenzprotokoll aufgenommen werden und sollte als eine einzelne Sequenz dargestellt werden, in der die durch Aufzählung dargestellten alternativen Reste durch das restriktivste Mehrdeutigkeitssymbol dargestellt werden (siehe Absätze 15 und 27).

95. Jede Sequenzvariante, die nur durch Verweis auf Deletion(en), Insertion(en) oder Substitution(en) in einer im Sequenzprotokoll enthaltenen Primärsequenz offenbart wird, sollte in das Sequenzprotokoll aufgenommen werden. Eine solche in das Sequenzprotokoll aufgenommene Sequenzvariante

(a) kann durch eine Annotation zur Primärsequenz dargestellt werden, wenn sie Variationen an einer einzigen oder mehreren unterschiedlichen Stellen enthält und diese Variationen unabhängig voneinander auftreten;

(b) sollte als separate Sequenz dargestellt und mit einer eigenen Sequenzkennzahl versehen werden, wenn sie Variationen an mehreren unterschiedlichen Stellen enthält und diese Variationen in wechselseitiger Abhängigkeit auftreten;

(c) muss als separate Sequenz dargestellt und mit einer eigenen Sequenzkennzahl versehen werden, wenn sie eine inserierte oder substituierte Sequenz aus mehr als 1000 Resten enthält (siehe Absatz 86).

96. Die folgende Tabelle zeigt die korrekte Verwendung von Merkmalschlüsseln und Qualifiern für Nukleinsäure- und Aminosäure-Sequenzvarianten:

Art der Sequenz	Merkmalschlüssel	Qualifier	Verwendung
Nukleinsäure	variation	replace oder note	Natürlich vorkommende Mutationen und Polymorphismen, z. B. Allele, RFLP.
Nukleinsäure	misc_difference	replace oder note	Variabilität, die künstlich eingeführt wird, z. B. durch genetische Manipulation oder chemische Synthese.
Aminosäure	VAR_SEQ	note	Variante, die durch alternatives Spleißen, alternative Promotoren, alternative Initiierung und ribosomale Leserasterverschiebung erzeugt wird.
Aminosäure	VARIANT	note	Jede Art von Variante, für die VAR_SEQ nicht anwendbar ist.

97. Die Annotation einer Sequenz für eine bestimmte Variante muss einen Merkmalschlüssel und einen Qualifier, wie in der Tabelle oben angegeben, sowie die Lage des Merkmals enthalten. Der Wert für den Qualifier "replace" darf nur ein einzelnes alternatives Nukleotid oder eine einzelne alternative Nukleotidsequenz sein, wobei die Symbole aus Abschnitt 1, Tabelle 1 zu verwenden sind, oder er muss leer sein. Eine Aufzählung alternativer Reste kann als Wert des Qualifiers "note" angegeben werden. Insbesondere muss eine Aufzählung alternativer Aminosäuren im Qualifier "note" als Wert angegeben werden, wenn in einer Sequenz "X" verwendet wird, und etwas anderes als „eines der Symbole 'A', 'R', 'N', 'D', 'C', 'Q', 'E', 'G', 'H', 'I', 'L', 'K', 'M', 'F', 'P', 'O', 'S', 'U', 'T', 'W', 'Y' oder 'V'" darstellt (siehe Absatz 27). Eine Deletion muss durch einen leeren Qualifier-Wert für den Qualifier "replace" oder durch den Hinweis im Qualifier "note", dass der Rest gelöscht werden kann, dargestellt werden. Ein oder mehrere inserierte oder substituierte Reste müssen im Qualifier "replace" oder "note" angegeben werden. Der Wert der Qualifier "replace" und "note" wird im Freitextformat angegeben und darf, wie in Absatz 86 festgelegt, 1000 Zeichen nicht überschreiten. In Absatz 100 ist geregelt, welche von Absatz 7 erfassten Sequenzen in einem Qualifier-Wert als Insertion oder Substitution angegeben werden.

98. Gegebenenfalls sollten die in Anhang I aufgeführten Symbole (siehe Abschnitte 1 bis 4 und dort Tabellen 1 bis 4) zur Darstellung von Restevarianten verwendet werden. Wenn es sich bei der Restevariante um einen modifizierten Rest handelt, der nicht in den Tabellen 2 oder 4 des Anhangs I aufgeführt ist, muss als Wert des Qualifiers "note" der vollständige, ungekürzte Name des modifizierten Rests angegeben werden. Modifizierte Reste müssen in der Merkmaltabelle gemäß Absatz 17 oder 30 näher beschrieben werden.

99. Die folgenden Beispiele veranschaulichen die Darstellung von Varianten, wie oben in den Absätzen 95 bis 98 beschrieben:

Beispiel 1: Merkmalschlüssel "misc_difference" für durch Aufzählung dargestellte alternative Nukleotide.
Das "n" an Position 53 der Sequenz kann eines von fünf alternativen Nukleotiden sein.

```
<INSDFeature>
    <INSDFeature_key>misc_difference</INSDFeature_key>
    <INSDFeature_location>53</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>w, cmnm5s2u, mam5u, mcm5s2u, or
p</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>53</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
         </INSDQualifier>
         <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>cmnm5s2u, mam5u, mcm5s2u, or p</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Beispiel 2: Merkmalschlüssel "misc_difference" für eine Deletion in einer Nukleotidsequenz.
Das Nukleotid an Position 413 der Sequenz wird deletiert.

```
<INSDFeature>
    <INSDFeature_key>misc_difference</INSDFeature_key>
    <INSDFeature_location>413</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>replace</INSDQualifier_name>
            <INSDQualifier_value></INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Beispiel 3: Merkmalschlüssel "misc_difference" für eine Insertion in einer Nukleotidsequenz.
Die Sequenz "atgccaaatat" wird zwischen den Positionen 100 und 101 der Primärsequenz inseriert.

```
<INSDFeature>
    <INSDFeature_key>misc_difference</INSDFeature_key>
    <INSDFeature_location>100^101</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>replace</INSDQualifier_name>
            <INSDQualifier_value>atgccaaatat</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Beispiel 4: Merkmalschlüssel "variation" für eine Substitution in einer Nukleotidsequenz.
Ein Cytosin substituiert das Nukleotid, das an Position 413 der Sequenz angegeben ist.

```
<INSDFeature>
    <INSDFeature_key>variation</INSDFeature_key>
    <INSDFeature_location>413</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>replace</INSDQualifier_name>
            <INSDQualifier_value>c</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Beispiel 5: Merkmalschlüssel "VARIANT" für eine Substitution in einer Aminosäuresequenz.
Die an Position 100 der Sequenz angegebene Aminosäure kann substituiert werden durch `I`, `A`, `F`, `Y`, `aIle`, `MeIle` oder `Nle`.

```
<INSDFeature>
    <INSDFeature_key>VARIANT</INSDFeature_key>
    <INSDFeature_location>100</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>I, A, F, Y, aIle, MeIle, or Nle
          </INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
    <INSDFeature_key>MOD_RES</INSDFeature_key>
    <INSDFeature_location>100</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>aIle, MeIle, or Nle</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Beispiel 6: Merkmalschlüssel "VARIANT" für eine Substitution in einer Aminosäuresequenz.
Die an Position 100 der Sequenz angegebene Aminosäure kann durch eine beliebige Aminosäure mit Ausnahme von Lys, Arg oder His substituiert werden.

```
<INSDFeature>
    <INSDFeature_key>VARIANT</INSDFeature_key>
    <INSDFeature_location>100</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>not K, R, or H</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

100. Eine von Absatz 7 erfasste Sequenz, die als Annotation zu einer Primärsequenz in einem Qualifier-Wert als Insertion oder Substitution angegeben wird, muss ebenfalls in das Sequenzprotokoll aufgenommen und mit einer eigenen Sequenzkennzahl versehen werden.

[Anhang I folgt]

ANHANG I

KONTROLLIERTES VOKABULAR

INHALTSVERZEICHNIS

ABSCHNITT 1: LISTE DER NUKLEOTIDE 2
ABSCHNITT 2: LISTE DER MODIFIZIERTEN NUKLEOTIDE 2
ABSCHNITT 3: LISTE DER AMINOSÄUREN 4
ABSCHNITT 4: LISTE DER MODIFIZIERTEN AMINOSÄUREN 5
ABSCHNITT 5: MERKMALSCHLÜSSEL FÜR NUKLEOTIDSEQUENZEN 6
ABSCHNITT 6: QUALIFIER FÜR NUKLEOTIDSEQUENZEN 23
ABSCHNITT 7: MERKMALSCHLÜSSEL FÜR AMINOSÄURESEQUENZEN 47
ABSCHNITT 8: QUALIFIER FÜR AMINOSÄURESEQUENZEN 54
ABSCHNITT 9: TABELLEN FÜR DEN GENETISCHEN CODE 56

ABSCHNITT 1: LISTE DER NUKLEOTIDE

In Tabelle 1 sind die in Sequenzprotokollen zu verwendenden Symbole für Nukleotidbasen aufgeführt. Sofern nicht näher beschrieben, wird das Symbol "t" in DNA als Thymin und in RNA als Uracil ausgelegt. Wenn ein Mehrdeutigkeitssymbol (das zwei oder mehr alternative Basen darstellt) angebracht ist, sollte das restriktivste Symbol verwendet werden. Wenn beispielsweise eine Base in einer bestimmten Position "a" oder "g" sein könnte, sollte "r" statt "n" verwendet werden. Das Symbol "n" wird als "a oder c oder g oder t/u" ausgelegt, wenn es ohne weitere Beschreibung verwendet wird.

Tabelle 1: Liste der Nukleotidsymbole

Symbol	Definition
a	Adenin
c	Cytosin
g	Guanin
t	Thymin in DNA/Uracil in RNA (t/u)
m	a oder c
r	a oder g
w	a oder t/u
s	c oder g
y	c oder t/u
k	g oder t/u
v	a oder c oder g; nicht t/u
h	a oder c oder t/u; nicht g
d	a oder g oder t/u; nicht c
b	c oder g oder t/u; nicht a
n	a oder c oder g oder t/u; "unbekanntes“ ("unknown") oder „sonstiges“ ("other")

ABSCHNITT 2: LISTE DER MODIFIZIERTEN NUKLEOTIDE

Die in Tabelle 2 aufgeführten Abkürzungen sind die einzigen zulässigen Werte für den Qualifier "mod_base". Wenn ein spezifisches modifiziertes Nukleotid in der Tabelle unten nicht aufgeführt ist, muss als Wert die Abkürzung "OTHER" verwendet werden. Wenn die Abkürzung "OTHER" verwendet wird, muss der vollständige, ungekürzte Name der modifizierten Base in einem "note"-Qualifier angegeben werden. Die in Tabelle 2 angegebenen Abkürzungen dürfen in der Sequenz selbst nicht verwendet werden.

Tabelle 2: Liste der modifizierten Nukleotide

Abkürzung	Definition
ac4c	4-Acetylcytidin
chm5u	5-(Carboxyhydroxymethyl)uridin
cm	2'-O-Methylcytidin
cmnm5s2u	5-Carboxymethylaminomethyl-2-thiouridin
cmnm5u	5-Carboxymethylaminomethyluridin
dhu	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

Abkürzung	Definition
m3c	3-Methylcytidin
m4c	N4-Methylcytosin
m5c	5-Methylcytidin
m6a	N6-Methyladenosin
m7g	7-Methylguanosin
mam5u	5-Methylaminomethyluridin
mam5s2u	5-Methylaminomethyl-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-methylthiopurin-6-yl)carbamoyl)threonin
mt6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)N-methyl-carbamoyl)threonin
mv	Uridin-5-oxyessigsäuremethylester
o5u	Uridin-5-oxyessigsäure (v)
osyw	Wybutoxosin
p	Pseudouridin
q	Queuosin
s2c	2-Thiocytidin
s2t	5-Methyl-2-thiouridin
s2u	2-Thiouridin
s4u	4-Thiouridin
m5u	5-Methyluridin
t6a	N-((9-beta-D-Ribofuranosylpurin-6-yl)carbamoyl)threonin
tm	2'-O-Methyl-5-methyluridin
um	2'-O-Methyluridin
yw	Wybutosine
x	3-(3-Amino-3-carboxypropyl)uridin, (acp3)u
OTHER	("note"-Qualifier erforderlich)

ABSCHNITT 3: LISTE DER AMINOSÄUREN

In Tabelle 3 sind die in Sequenzprotokollen zu verwendenden Symbole für Aminosäuren aufgeführt. Wenn ein Mehrdeutigkeitssymbol (das zwei oder mehr alternative Aminosäuren darstellt) angebracht ist, sollte das restriktivste Symbol verwendet werden. Wenn beispielsweise eine Aminosäure in einer bestimmten Position Asparaginsäure oder Asparagin sein könnte, sollte "B" statt "X" verwendet werden. Das Symbol "X" wird als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, wenn es ohne weitere Beschreibung verwendet wird.

Tabelle 3: Liste der Aminosäurensymbole

Symbol	Definition
A	Alanin
R	Arginin
N	Asparagin
D	Asparaginsäure (Aspartat)
C	Cystein
Q	Glutamin
E	Glutaminsäure (Glutamat)
G	Glycin
H	Histidin
I	Isoleucin
L	Leucin
K	Lysin
M	Methionin
F	Phenylalanin
P	Prolin
O	Pyrrolysin
S	Serin
U	Selenocystein
T	Threonin
W	Tryptophan
Y	Tyrosin
V	Valin
B	Asparaginsäure oder Asparagin
Z	Glutamin oder Glutaminsäure
J	Leucin oder Isoleucin
X	A oder R oder N oder D oder C oder Q oder E oder G oder H oder I oder L oder K oder M oder F oder P oder O oder S oder U oder T oder W oder Y oder V; "unbekannte" ("unknown") oder "sonstige" („other“)

ABSCHNITT 4: LISTE DER MODIFIZIERTEN AMINOSÄUREN

In Tabelle 4 sind die Abkürzungen aufgeführt, die für modifizierte Aminosäuren im obligatorischen Qualifier "note" für die Merkmalschlüssel "MOD_RES" oder "SITE" als einzige zulässig sind. Der Wert für den Qualifier "note" muss entweder eine Abkürzung aus dieser Tabelle sein oder der vollständige, ungekürzte Name der modifizierten Aminosäure. Die in dieser Tabelle angegebenen Abkürzungen (oder vollständigen Namen) dürfen nicht in der Sequenz selbst verwendet werden.

Tabelle 4: Liste der modifizierten Aminosäuren

Abkürzung	Modifizierte Aminosäure
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-Diaminopimelinsäure
EtGly	N-Ethylglycin
EtAsn	N-Ethylasparagin
Hyl	Hydroxylysin
aHyl	allo-Hydroxylysin
3Hyp	3-Hydroxyprolin
4Hyp	4-Hydroxyprolin
Ide	Isodesmosin
aIle	allo-Isoleucin
MeGly	N-Methylglycin, Sarcosin
MeIle	N-Methylisoleucin
MeLys	6-N-Methyllysin
MeVal	N-Methylvalin
Nva	Norvalin
Nle	Norleucin
Orn	Ornithin

ABSCHNITT 5: MERKMALSCHLÜSSEL FÜR NUKLEOTIDSEQUENZEN

Dieser Abschnitt enthält die Liste der zulässigen Merkmalschlüssel für Nukleotidsequenzen. Außerdem werden die obligatorischen und fakultativen Qualifier aufgeführt. Die Merkmalschlüssel werden in alphabetischer Reihenfolge aufgeführt. Sofern unter "Molekülart" nicht anders angegeben, können die Merkmalschlüssel für DNA oder für RNA gleichermaßen verwendet werden. Bestimmte Merkmalschlüssel können neben der angegebenen Art des Organismus auch für künstliche Sequenzen geeignet sein.

Die Namen von Merkmalschlüsseln müssen in der XML-Instanz des Sequenzprotokolls in exakt der Form verwendet werden, in der sie in den folgenden Beschreibungen unter "Merkmalschlüssel" angegeben sind, mit Ausnahme der Merkmalschlüssel "3'UTR" und "5'UTR". Siehe dazu die Anmerkung in der Beschreibung für die Merkmalschlüssel "3'UTR" und "5'UTR".

5.1.	Merkmalschlüssel	C_region
	Definition	konstante Region von leichten und schweren Immunglobulinketten und Alpha-, Beta- und Gammaketten von T-Zell-Rezeptoren; enthält je nach Kette ein oder mehrere Exons
	Fakultative Qualifier	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Art des Organismus	Eukaryoten

5.2.	Merkmalschlüssel	CDS
	Definition	codierende Sequenz; Sequenz von Nukleotiden, die mit der Sequenz der Aminosäuren in einem Protein korrespondiert (Lage beinhaltet Stopcodon); Merkmal kann die konzeptionelle Translation der Aminosäure beinhalten.
	Fakultative Qualifier	allele circular_RNA codon_start EC_number exception function gene gene_synonym map note number operon product protein_id pseudo pseudogene ribosomal_slippage standard_name translation transl_except transl_table trans_splicing

	Anmerkung	Der Qualifier "codon_start" kann als gültigen Wert "1" oder "2" oder "3" annehmen, dieser Wert gibt an, an welchem Offset (Versatz) das erste vollständige Codon eines Codierungsmerkmals relativ zur ersten Base dieses Merkmals gefunden werden kann; "transl_table" definiert die Tabelle für den genetischen Code, wenn es sich nicht um die Tabelle für den Standardcode bzw. den universellen genetischen Code handelt; Ausnahmen vom genetischen Code außerhalb des Bereichs der angegebenen Tabellen werden im Qualifier "transl_except" angegeben; von den Qualifiern "translation", "pseudogene" oder "pseudo" ist in Verbindung mit einem "CDS"-Merkmalschlüssel jeweils nur einer zulässig; bei Verwendung des Qualifiers "translation" ist der Qualifier "protein_id" obligatorisch, wenn das Translationsprodukt vier oder mehr spezifisch definierte Aminosäuren enthält.

5.3.	Merkmalschlüssel	centromere
	Definition	biologisch bedeutsame Region, die als Centromer identifiziert und experimentell charakterisiert wurde
	Fakultative Qualifier	note standard_name
	Anmerkung	Das Merkmal "centromere" beschreibt das Intervall der DNA, das einer Region entspricht, wo Chromatiden zusammengehalten und Kinetochoren ausgebildet werden.

5.4.	Merkmalschlüssel	D-loop
	Definition	Verdrängungs-(displacement-)Schleife; Region innerhalb der mitochondrialen DNA, in der sich ein kurzes Stück RNA mit einem DNA-Strang paart und dabei den ursprünglichen Schwesterstrang in dieser Region verdrängt; wird auch verwendet, um die Verdrängung einer Region eines Strangs eines DNA-Doppelstrangs durch eine einzelsträngige invasive Nukleinsäure bei der durch das RecA-Protein katalysierten Reaktion zu beschreiben
	Fakultative Qualifier	allele gene gene_synonym map note
	Molekülart	DNA

5.5.	Merkmalschlüssel	D_segment
	Definition	Diversitätssegment der schweren Immunglobulinkette und der Betakette des T-Zell-Rezeptors
	Fakultative Qualifier	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Art des Organismus	Eukaryoten

5.6.	Merkmalschlüssel	exon
	Definition	Region des Genoms, die für einen Teil der gespleißten mRNA, rRNA und tRNA codiert; kann 5'UTR, alle CDS und 3'UTR enthalten
	Fakultative Qualifier	allele EC_number function gene gene_synonym map note number product pseudo pseudogene standard_name trans_splicing

5.7.	Merkmalschlüssel	gene
	Definition	biologisch bedeutsame Region, die als Gen identifiziert und benannt wurde
	Fakultative Qualifier	allele function gene gene_synonym map note operon product pseudo pseudogene phenotype standard_name trans_splicing
	Anmerkung	Das Merkmal "gene" beschreibt das Intervall der DNA, das einem genetischen Merkmal oder Phänotyp entspricht; das Merkmal ist definitionsgemäß nicht streng an seine Endpositionen gebunden; es soll eine Region darstellen, in der sich das Gen befindet.

5.8.	Merkmalschlüssel	iDNA
	Definition	intervenierende DNA; DNA, die durch eine der verschiedenen Arten der Rekombination eliminiert wird
	Fakultative Qualifier	allele function gene gene_synonym map note number standard_name
	Molekülart	DNA
	Anmerkung	Z. B. bei der somatischen Prozessierung von Immunglobulingenen.

5.9.	Merkmalschlüssel	intron
	Definition	ein transkribiertes DNA-Segment, das durch das Zusammenspleißen der umgebenden Sequenzen (Exons) aus dem Transkript herausgeschnitten wird
	Fakultative Qualifier	allele

		function gene gene_synonym map note number pseudo pseudogene standard_name trans_splicing

5.10.	Merkmalschlüssel	J_segment
	Definition	Verbindungssegment zwischen leichten und schweren Immunglobulinketten und den Alpha-, Beta- und Gamma-Ketten der T-Zell-Rezeptoren
	Fakultative Qualifier	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Art des Organismus	Eukaryoten

5.11.	Merkmalschlüssel	mat_peptide
	Definition	für ein reifes Peptid oder Protein codierende Sequenz; Sequenz, die für das reife oder endgültige Peptid- oder Proteinprodukt im Anschluss an eine posttranslationale Modifikation codiert; die Lage schließt (im Gegensatz zur entsprechenden CDS) das Stopcodon nicht mit ein
	Fakultative Qualifier	allele EC_number function gene gene_synonym map note product pseudo pseudogene standard_name

5.12.	Merkmalschlüssel	misc_binding
	Definition	Stelle in einer Nukleinsäure, die kovalent oder nicht kovalent eine andere Einheit bindet, die durch keinen anderen Bindungsschlüssel ("primer_bind" oder "protein_bind") beschrieben werden kann
	Obligatorische Qualifier	bound_moiety
	Fakultative Qualifier	allele function gene gene_synonym map note
	Anmerkung	Zu beachten ist, dass für die Beschreibung ribosomaler Bindungsstellen der Merkmalschlüssel "regulatory" und der Qualifier "regulatory_class" mit dem Wert "ribosome_binding_site" verwendet werden müssen.

5.13.	Merkmalschlüssel	misc_difference
	Definition	die Merkmalsequenz unterscheidet sich an dieser Lage von der dargestellten Sequenz und kann nicht durch einen anderen Unterscheidungsschlüssel ("variation" oder "modified_base") beschrieben werden
	Fakultative Qualifier	allele clone compare gene gene_synonym map note phenotype replace standard_name
	Anmerkung	Der Merkmalschlüssel "misc_difference" muss verwendet werden, um künstlich, z. B. durch genetische Manipulation oder chemische Synthese, eingeführte Variabilität zu beschreiben; zur Annotierung einer Deletion, Insertion oder Substitution ist der Qualifier "replace" zu verwenden. Der Merkmalschlüssel "variation" muss verwendet werden, um natürlich vorkommende genetische Variabilität zu beschreiben.

5.14.	Merkmalschlüssel	misc_feature
	Definition	biologisch bedeutsame Region, die nicht durch einen anderen Merkmalschlüssel beschrieben werden kann; neues oder seltenes Merkmal
	Fakultative Qualifier	allele function gene gene_synonym map note number phenotype product pseudo pseudogene standard_name
	Anmerkung	Dieser Schlüssel sollte nicht verwendet werden, wenn eine Region lediglich zu Anmerkungszwecken oder zur Verwendung in der Lage eines anderen Merkmals gekennzeichnet werden soll.

5.15.	Merkmalschlüssel	misc_recomb
	Definition	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 oder Qualifier des "source"-Schlüssels ("proviral") beschrieben werden kann
	Fakultative Qualifier	allele gene gene_synonym map note recombination_class standard_name
	Molekülart	DNA

5.16.	Merkmalschlüssel	misc_RNA
	Definition	jedes Transkript oder RNA-Produkt, das nicht durch andere RNA-Schlüssel ("prim_transcript", "precursor_RNA“, "mRNA“, "5’UTR“, "3’UTR“, "exon“, "CDS“,

		"sig_peptide“, "transit_peptide“, "mat_peptide“, "intron“, "polyA_site“, "ncRNA“, "rRNA“ und "tRNA“) beschrieben werden kann
	Fakultative Qualifier	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
5.17.	Merkmalschlüssel	misc_structure
	Definition	jede sekundäre oder tertiäre Nukleotidstruktur oder Konformation, die nicht durch andere Strukturschlüssel ("stem_loop" und "D-loop") beschrieben werden kann
	Fakultative Qualifier	allele function gene gene_synonym map note standard_name
5.18.	Merkmalschlüssel	mobile_element
	Definition	Region des Genoms, die mobile Elemente enthält
	Obligatorische Qualifier	mobile_element_type
	Fakultative Qualifier	allele function gene gene_synonym map note rpt_family rpt_type standard_name
5.19.	Merkmalschlüssel	modified_base
	Definition	das angezeigte Nukleotid ist ein modifiziertes Nukleotid und sollte durch das (im Wert des Qualifiers "mod_base") angegebene Molekül substituiert werden
	Obligatorische Qualifier	mod_base
	Fakultative Qualifier	allele frequency gene gene_synonym map note
	Anmerkung	Als Werte für den obligatorischen Qualifier "mod_base" dürfen nur die Abkürzungen für modifizierte Basen verwendet werden, die in Abschnitt 2 dieses Anhangs aufgeführt sind.

5.20.	Merkmalschlüssel	mRNA
	Definition	messenger RNA (Boten-RNA); enthält eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon) und eine 3'-nichttranslatierte Region (3'UTR)
	Fakultative Qualifier	allele circular_RNA function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing

5.21.	Merkmalschlüssel	ncRNA
	Definition	ein nicht proteincodierendes Gen, keine ribosomale RNA und Transfer-RNA, dessen funktionales Molekül das RNA-Transkript ist
	Obligatorische Qualifier	ncRNA_class
	Fakultative Qualifier	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
	Anmerkung	Das Merkmal "ncRNA" darf nicht für die Annotation von ribosomaler und Transfer-RNA verwendet werden, hierzu müssen die Merkmalschlüssel "rRNA" bzw. "tRNA" verwendet werden.

5.22.	Merkmalschlüssel	N_region
	Definition	zusätzliche Nukleotide, die zwischen neu geordnete Immunglobulinabschnitte inseriert werden
	Fakultative Qualifier	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Art des Organismus	Eukaryoten

5.23.	Merkmalschlüssel	operon
	Definition	Region, die ein polycistronisches Transkript enthält, das eine Gruppe von Genen umfasst, die unter der Kontrolle derselben regulatorischen Sequenzen/Promotoren stehen und demselben biologischen Reaktionsweg angehören

	Obligatorische Qualifier	operon
	Fakultative Qualifier	allele function map note phenotype pseudo pseudogene standard_name

5.24.	Merkmalschlüssel	oriT
	Definition	Transferursprung; Region eines DNA-Moleküls, in der während des Konjugations- oder Mobilisierungsprozesses der Transfer eingeleitet wird
	Fakultative Qualifier	allele bound_moiety direction gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq standard_name
	Molekülart	DNA
	Anmerkung	Der Merkmalschlüssel "rep_origin" muss zur Beschreibung des Replikationsursprungs verwendet werden; die zulässigen Werte des Qualifiers "direction" sind "left", "right" oder "both", wobei in Verbindung mit dem Merkmalschlüssel "oriT" nur "left" und "right" gültig sind; Transferursprünge können im Chromosom vorliegen; Plasmide können mehrere Transferursprünge enthalten.

5.25.	Merkmalschlüssel	polyA_site
	Definition	Stelle auf einem RNA-Transkript, an der durch posttranslationale Polyadenylierung Adeninreste angefügt werden
	Fakultative Qualifier	allele gene gene_synonym map note
	Art des Organismus	Eukaryoten und eukaryotische Viren

5.26.	Merkmalschlüssel	precursor_RNA
	Definition	jede RNA-Spezies, die noch nicht das reife RNA-Produkt ist; kann ncRNA, rRNA, tRNA, eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron) und eine 3'-nichttranslatierte Region (3'UTR) enthalten
	Fakultative Qualifier	allele function gene gene_synonym map note operon product

		standard_name trans_splicing
	Anmerkung	Verwendung für RNA, die das Ergebnis einer posttranskriptionellen Prozessierung sein kann; wenn bekannt ist, dass die betreffende RNA nicht verarbeitet wurde, ist der Schlüssel "prim_transcript" zu verwenden.

5.27.	Merkmalschlüssel	prim_transcript
	Definition	primäres (ursprüngliches, noch nicht prozessiertes) Transkript; kann ncRNA, rRNA,tRNA, eine 5'-nichttranslatierte Region (5'UTR), codierende Sequenzen (CDS, Exon), intervenierende Sequenzen (Intron) und eine 3'-nichttranslatierte Region (3'UTR) enthalten
	Fakultative Qualifier	allele function gene gene_synonym map note operon standard_name

5.28.	Merkmalschlüssel	primer_bind
	Definition	nicht kovalente Primer-Bindungsstelle für die Initiierung der Replikation, Transkription oder reversen Transkription; enthält eine oder mehrere Stellen für synthetische Elemente, z. B. PCR-Primer-Elemente
	Fakultative Qualifier	allele gene gene_synonym map note standard_name
	Anmerkung	Wird verwendet, um die Stelle einer bestimmten Sequenz zu annotieren, an die ein Primer-Molekül bindet - nicht dazu vorgesehen, die Sequenz des Primer-Moleküls selbst darzustellen; da an PCR-Reaktionen meist Primer-Paare beteiligt sind, kann ein einzelner "primer_bind"-Merkmalschlüssel in Verbindung mit dem Operator "order" (Lage,Lage) mit zwei Lagen verwendet werden, oder es wird ein Paar aus zwei "primer_bind"-Merkmalschlüsseln verwendet.

5.29.	Merkmalschlüssel	propeptide
	Definition	für ein Propeptid codierende Sequenz; codierende Sequenz für die Domäne eines Proproteins, das abgespalten wird, um das reife Protein zu bilden
	Fakultative Qualifier	allele function gene gene_synonym map note product pseudo pseudogene standard_name

5.30.	Merkmalschlüssel	protein_bind
	Definition	nicht kovalente Proteinbindungsstelle auf der Nukleinsäure
	Obligatorische Qualifier	bound_moiety
	Fakultative Qualifier	allele function gene gene_synonym map note operon standard_name
	Anmerkung	Zu beachten ist, dass für die Beschreibung ribosomaler Bindungsstellen der Merkmalschlüssel "regulatory" und der Qualifier "regulatory_class" mit dem Wert "ribosome_binding_site" verwendet werden müssen.

5.31.	Merkmalschlüssel	regulatory
	Definition	jede Region einer Sequenz, die zur Regulation der Transkription, Translation, Replikation oder Chromatin-Struktur beiträgt;
	Obligatorische Qualifier	regulatory_class
	Fakultative Qualifier	allele bound_moiety function gene gene_synonym map note operon phenotype pseudo pseudogene standard_name

5.32.	Merkmalschlüssel	repeat_region
	Definition	Genomregion mit repetitiven Einheiten
	Fakultative Qualifier	allele function gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq satellite standard_name

5.33.	Merkmalschlüssel	rep_origin
	Definition	Replikationsursprung; Startpunkt der Duplikation der Nukleinsäure, durch die zwei identische Kopien entstehen
	Fakultative Qualifier	allele direction function gene gene_synonym map note standard_name
	Anmerkung	Der Qualifier "direction" hat als gültige Werte: "left", "right" oder "both".

5.34.	Merkmalschlüssel	rRNA
	Definition	reife ribosomale RNA; RNA-Komponente des Ribonukleoprotein-Partikels (Ribosom), das Aminosäuren zu Proteinen zusammensetzt
	Fakultative Qualifier	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name
	Anmerkung	Die Größe der rRNA sollte mit dem Qualifier "product" annotiert werden.

5.35.	Merkmalschlüssel	S_region
	Definition	Switch-Region der schweren Immunglobulinketten; beteiligt am Umbau schwerer DNA-Ketten, der zur Expression einer anderen Immunglobulinklasse aus derselben B-Zelle führt
	Fakultative Qualifier	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Art des Organismus	Eukaryoten

5.36.	Merkmalschlüssel	sig_peptide
	Definition	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
	Fakultative Qualifier	allele function gene gene_synonym map note product pseudo pseudogene standard_name

5.37.	Merkmalschlüssel	source
	Definition	bezeichnet die Herkunft der Sequenz; die Angabe dieses Schlüssels ist obligatorisch; jede Sequenz weist einen einzigen "source"-Schlüssel auf, der die gesamte Sequenz umfasst
	Obligatorische Qualifier	organism mol_type
	Fakultative Qualifier	cell_line cell_type chromosome clone clone_lib collected_by collection_date cultivar dev_stage ecotype environmental_sample germline haplogroup haplotype host identified_by isolate isolation_source lab_host lat_lon macronuclear map mating_type note organelle PCR_primers plasmid pop_variant proviral rearranged segment serotype serovar sex strain sub_clone sub_species sub_strain tissue_lib tissue_type variety
	Molekülart	beliebig

5.38.	Merkmalschlüssel	stem_loop
	Definition	Haarnadelschleife; eine Doppelhelix-Region, die durch Basenpaarung zwischen benachbarten (invertierten) komplementären Sequenzen in einem RNA- oder DNA-Einzelstrang entsteht
	Fakultative Qualifier	allele function gene gene_synonym map note operon standard_name

5.39.	Merkmalschlüssel	STS
	Definition	Sequence Tagged Site (STS); kurze Einzelkopie-DNA-Sequenz, die einen Kartierungspunkt auf dem Genom kennzeichnet und durch PCR nachgewiesen werden kann; eine Region des Genoms kann durch Bestimmung einer Reihe von STS kartiert werden
	Fakultative Qualifier	allele gene gene_synonym map note standard_name
	Molekülart	DNA
	Anmerkung	Die Lage der STS schließt Primer im Merkmalschlüssel "primer_bind" oder Primer ein.

5.40.	Merkmalschlüssel	telomere
	Definition	biologisch bedeutsame Region, die als Telomer identifiziert und experimentell charakterisiert wurde
	Fakultative Qualifier	note rpt_type rpt_unit_range rpt_unit_seq standard_name
	Anmerkung	Das Merkmal "telomere" beschreibt das Intervall der DNA, das einer spezifischen Struktur am Ende des linearen eukaryotischen Chromosoms entspricht, die für die Integrität und Aufrechterhaltung des Endes erforderlich ist; diese Region ist im Vergleich zum Rest des Chromosoms einzigartig und stellt das physische Ende des Chromosoms dar.

5.41.	Merkmalschlüssel	tmRNA
	Definition	transfer messenger RNA (Transfer-Boten-RNA); tmRNA fungiert zuerst als tRNA und dann als mRNA, die eine Peptid-Markierung codiert; das Ribosom translatiert diese mRNA-Region der tmRNA und hängt die codierte Peptid-Markierung an das C-Ende des unfertigen Proteins an; diese angehängte Markierung dient später als Signal für den Abbau oder die Proteolyse des Proteins
	Fakultative Qualifier	allele function gene gene_synonym map note product pseudo pseudogene

		standard_name tag_peptide

5.42.	Merkmalschlüssel	transit_peptide
	Definition	für ein 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
	Fakultative Qualifier	allele function gene gene_synonym map note product pseudo pseudogene standard_name

5.43.	Merkmalschlüssel	tRNA
	Definition	reife Transfer-RNA, ein kleines RNA-Molekül (75–85 Basen lang), das die Translation einer Nukleinsäuresequenz in eine Aminosäuresequenz vermittelt
	Fakultative Qualifier	allele circular_RNA anticodon function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing

5.44.	Merkmalschlüssel	unsure
	Definition	eine kleine Region von sequenzierten Basen, in der Regel 10 oder weniger, die nicht zuverlässig identifiziert werden konnte. Eine solche Region kann zugewiesene Basen (a, t, g oder c) oder eine Mischung aus zugewiesenen und nicht zugewiesenen Basen ("n") enthalten.
	Fakultative Qualifier	allele compare gene gene_synonym map note replace
	Anmerkung	Zur Annotierung einer Deletion, Insertion oder Substitution ist der Qualifier "replace" zu verwenden.

5.45.	Merkmalschlüssel	V_region
	Definition	variable Region der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gammaketten von T-Zell-Rezeptoren; codiert für das variable Aminoende; kann aus V_Segmenten, D_Segmenten, N_Regionen und J_Segmenten bestehen
	Fakultative Qualifier	allele

		gene gene_synonym map note product pseudo pseudogene standard_name
	Art des Organismus	Eukaryoten

5.46.	Merkmalschlüssel	V_segment
	Definition	variables Segment der leichten und schweren Immunglobulinketten sowie der Alpha-, Beta- und Gammaketten von T-Zell-Rezeptoren; codiert für den Großteil der variablen Region (V_region) und die letzten Aminosäuren des Leader-Peptids
	Fakultative Qualifier	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Art des Organismus	Eukaryoten

5.47.	Merkmalschlüssel	variation
	Definition	ein verwandter Stamm enthält stabile Mutationen desselben Gens (z. B. RFLPs, Polymorphismen usw.), die sich an dieser (und möglicherweise auch an anderer) Lage von der vorliegenden Sequenz unterscheiden
	Fakultative Qualifier	allele compare frequency gene gene_synonym map note phenotype product replace standard_name
	Anmerkung	Wird verwendet zur Beschreibung von Allelen, RFLPs und anderen natürlich vorkommenden Mutationen und Polymorphismen; zur Annotation einer Deletion, Insertion oder Substitution ist der Qualifier "replace" zu verwenden; eine Variabilität infolge genetischer Manipulation (z. B. ortsgerichteter Mutagenese) muss mit dem Merkmal "misc_difference" beschrieben werden.

5.48.	Merkmalschlüssel	3'UTR
	Definition	1) Region am 3'-Ende eines reifen Transkripts (nach dem Stopcodon), die nicht in ein Protein translatiert wird; 2) Region am 3'-Ende eines RNA-Virus (nach dem letzten Stopcodon), die nicht in ein Protein translatiert wird
	Fakultative Qualifier	allele function gene gene_synonym map

		note standard_name trans_splicing
	Anmerkung	Das Apostroph-Zeichen (einfaches Anführungszeichen oben) hat in XML eine besondere Bedeutung und muss im Wert eines Elements durch "'" ersetzt werden. Daher muss "3'UTR" in der XML-Datei als "3'UTR" dargestellt werden, d. h. als <INSDFeature_key>3'UTR</INSDFeature_key>.

5.49.	Merkmalschlüssel	5'UTR
	Definition	1) Region am 5'-Ende eines reifen Transkripts (vor dem Initiationscodon), die nicht in ein Protein translatiert wird; 2) Region am 5'-Ende eines RNA-Virus (vor dem ersten Initiationscodon), die nicht in ein Protein translatiert wird
	Fakultative Qualifier	allele function gene gene_synonym map note standard_name trans_splicing
	Anmerkung	Das Apostroph-Zeichen (einfaches Anführungszeichen oben) hat in XML eine besondere Bedeutung und muss im Wert eines Elements durch "'" ersetzt werden. Daher muss "5'UTR" in der XML-Datei als "5'UTR" dargestellt werden, d. h. als <INSDFeature_key>5'UTR</INSDFeature_key>.

ABSCHNITT 6: QUALIFIER FÜR NUKLEOTIDSEQUENZEN

Dieser Abschnitt enthält die Liste der Qualifier, die für Merkmale in Nukleotidsequenzen zu verwenden sind. Die Qualifier werden in alphabetischer Reihenfolge aufgeführt.

Wenn als Wertformat "none" angegeben ist, darf weder das Element `INSDQualifier_value` noch das Element `NonEnglishQualifier_value` verwendet werden.

Wenn als Wertformat Freitext angegeben ist, der als sprachabhängig gekennzeichnet ist, muss eines der folgenden Elemente verwendet werden:

1) das Element `INSDQualifier_value`; oder
2) das Element `NonEnglishQualifier_value`; oder
3) sowohl das Element `INSDQualifier_value` als auch das Element `NonEnglishQualifier_value`.[6]

Wenn als Wertformat etwas anderes als "none" angegeben ist, aber nicht als sprachabhängiger gekennzeichneter Freitext , muss das Element `INSDQualifier_value` verwendet werden, und das Element `NonEnglishQualifier_value` darf nicht verwendet werden.

HINWEIS: Jeder „Freitext“-Qualifier-Wert, der für einen Qualifier mit einem sprachabhängigen Freitextformat bereitgestellt wird, kann eine Übersetzung zu Zwecken internationaler, nationaler oder regionaler Verfahren erfordern.[7] Sprachabhängiger Freitext ist für die Werte der in der folgenden Tabelle aufgeführten Qualifier vorgesehen:

Tabelle 5: Liste der Qualifier mit sprachabhängigen Freitextwerten für Nukleotidsequenzen

Abschnitt	Sprachabhängige Freitext-Qualifier
6.3	bound_moiety
6.5	cell_type
6.8	clone
6.9	clone_lib
6.11	collected_by
6.14	cultivar
6.15	dev_stage
6.18	ecotype
6.21	frequency
6.22	function
6.24	gene_synonym
6.26	haplogroup
6.28	host
6.29	identified_by
6.30	isolate
6.31	isolation_source
6.32	lab_host
6.36	mating_type
6.41	note
6.45	organism
6.47	phenotype
6.49	pop_variant
6.50	product
6.66	serotype
6.67	serovar
6.68	sex
6.69	standard_name
6.70	strain
6.71	sub_clone

[6] Nach § 11a Absatz 3 der Patentverordnung ist sprachabhängiger Freitext in deutscher Sprache abzufassen. Er kann zusätzlich auch in englischer Sprache angegeben werden. Es ist danach für sprachabhängigen Freitext entweder nur das Element `NonEnglishQualifier_value` oder sowohl das Element `INSDQualifier_value` als auch das Element `NonEnglishQualifier_value` zu verwenden.

[7] Nach § 11a Absatz 3 der Patentverordnung ist sprachabhängiger Freitext in deutscher Sprache abzufassen. Er kann zusätzlich auch in englischer Sprache angegeben werden.

6.72	sub_species
6.73	sub_strain
6.75	tissue_lib
6.76	tissue_type
6.81	variety

6.1.	Qualifier	allele
	Definition	Name des Allels für das angegebene Gen
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>adh1-1</INSDQualifier_value>
	Anmerkung	Alle genbezogenen Merkmale ("exon", "CDS" usw.) für ein bestimmtes Gen sollten denselben Wert für den Qualifier "allele" aufweisen; der Wert des Qualifiers "allele" muss definitionsgemäß anders sein als der Wert des Qualifiers "gene"; bei Verwendung mit dem Merkmalschlüssel "variation" sollte der Wert des Qualifiers "allele" dem der Variante entsprechen.

6.2.	Qualifier	anticodon
	Definition	Lage des Anticodons der tRNA und der Aminosäure, für die sie codiert
	Obligatorisches Wertformat	(pos:<location>,aa:<amino_acid>,seq:<text>), wobei <location> die Lage des Anticodons, <amino_acid> die dreibuchstabige Abkürzung für die codierte Aminosäure und <text> die Sequenz des Anticodons darstellt
	Beispiel	<INSDQualifier_value>(pos:34..36,aa:Phe,seq:aaa)</INSDQualifier_value> <INSDQualifier_value>(pos:join(5,495..496),aa:Leu,seq:taa)</INSDQualifier_value> <INSDQualifier_value>(pos:complement(4156..4158),aa:Glu,seq:ttg)</INSDQualifier_value>

6.3.	Qualifier	bound_moiety
	Definition	Name des Moleküls/Komplexes, das/der an das angegebene Merkmal binden kann
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>GAL4</INSDQualifier_value>
	Anmerkung	Für die Merkmalschlüssel "misc_binding", "oriT" und "protein_bind" ist nur ein einziger "bound_moiety"-Qualifier zulässig.

6.4.	Qualifier	cell_line
	Definition	Zelllinie, aus der die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>MCF7</INSDQualifier_value>

6.5.	Qualifier	cell_type
	Definition	Zelltyp, aus dem die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>leukocyte</INSDQualifier_value>

6.6.	Qualifier	chromosome
	Definition	Chromosom (z. B. Chromosomennummer), aus dem die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>1</INSDQualifier_value> <INSDQualifier_value>X</INSDQualifier_value>

6.7.	Qualifier	circular_RNA
	Definition	zeigt an, dass die Exons nicht in der richtigen Reihenfolge angeordnet sind oder sich überlappen, weil es sich bei diesem gespleißten RNA-Produkt um eine zirkuläre RNA (circRNA) handelt, die durch "Backsplicing" entstanden ist, z.B. wenn sich im Gen ein Downstream-Exon im RNA-Produkt 5‘ von einem Upstream-Exon befindet
	Wertformat	none
	Anmerkung	sollte für Merkmale wie "CDS", "mRNA", "tRNA" und andere Merkmale verwendet werden, die als Ergebnis eines “Backsplicing”-Ereignisses entstehen. Dieser Qualifier sollte nur verwendet werden, wenn das Spleiß-Ergebnis im "join"-Operator angegeben ist, z.B. join(complement(69611..69724),139856..140087)

6.8.	Qualifier	clone
	Definition	Klon, aus dem die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>lambda-hIL7.3</INSDQualifier_value>
	Anmerkung	Das Merkmal "source" darf nicht mehr als einen "clone"-Qualifier enthalten; wenn die Sequenz aus mehreren Klonen gewonnen wurde, kann diese in der Merkmaltabelle unter Verwendung des Merkmalschlüssels "misc_feature" und eines "note"-Qualifiers zu Angabe der mehreren Klone näher beschrieben werden.

6.9.	Qualifier	clone_lib
	Definition	Klonbibliothek, aus der die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>lambda-hIL7</INSDQualifier_value>

6.10.	Qualifier	codon_start
	Definition	Angabe des Offsets (Versatz), an dem das erste vollständige Codon eines Codierungsmerkmals relativ zur ersten Base dieses Merkmals gefunden werden kann
	Obligatorisches Wertformat	1 oder 2 oder 3
	Beispiel	<INSDQualifier_value>2</INSDQualifier_value>

6.11.	Qualifier	collected_by
	Definition	Name der Personen oder Institute, die die Probe entnommen haben
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>Dan Janzen</INSDQualifier_value>

6.12.	Qualifier	collection_date
	Definition	Datum, an dem die Probe entnommen wurde
	Obligatorisches Wertformat	JJJJ-MM-TT, JJJJ-MM oder JJJJ
	Beispiel	<INSDQualifier_value>1952-10-21</INSDQualifier_value> <INSDQualifier_value>1952-10</INSDQualifier_value> <INSDQualifier_value>1952</INSDQualifier_value>
	Anmerkung	"JJJJ" ist ein vierstelliger Wert, der das Jahr angibt. "MM" ist ein zweistelliger Wert, der den Monat angibt. "TT" ist ein zweistelliger Wert, der den Tag des Monats angibt.

6.13.	Qualifier	compare
	Definition	Details zur Referenz eines vorhandenen öffentlichen INSD-Eintrags, mit dem ein Vergleich erfolgt
	Obligatorisches Wertformat	[accession-number.sequence-version]
	Beispiel	<INSDQualifier_value>AJ634337.1</INSDQualifier_value>
	Anmerkung	Dieser Qualifier kann für die folgenden Merkmale verwendet werden: "misc_difference", "unsure" und "variation". Innerhalb eines einzigen Merkmals sind mehrere "compare"-Qualifier mit unterschiedlichen Inhalten zulässig. Dieser Qualifier ist nicht für die groß angelegte Annotation von Variationen wie SNPs vorgesehen.

6.14.	Qualifier	cultivar
	Definition	Kulturpflanzensorte (cultivated variety), aus der die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>Nipponbare</INSDQualifier_value> <INSDQualifier_value>Tenuifolius</INSDQualifier_value> <INSDQualifier_value>Candy Cane</INSDQualifier_value> <INSDQualifier_value>IR36</INSDQualifier_value>

	Anmerkung	Der Qualifier "cultivar" wird ausschließlich für Produkte der künstlichen Selektion verwendet; für natürliche, benannte Pflanzen- und Pilzvarietäten ist der Qualifier "variety" zu verwenden.

6.15.	Qualifier	dev_stage
	Definition	wenn die Sequenz aus einem Organismus in einem bestimmten Entwicklungsstadium gewonnen wurde, wird es mit diesem Qualifier angegeben
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>fourth instar larva</INSDQualifier_value>

6.16.	Qualifier	direction
	Definition	Richtung der DNA-Replikation
	Obligatorisches Wertformat	"left", "right" oder "both" "left" bedeutet in Richtung auf das 5'-Ende der Sequenz (wie dargestellt) und "right" bedeutet in Richtung auf das 3'-Ende
	Beispiel	<INSDQualifier_value>left</INSDQualifier_value>
	Anmerkung	Die Werte "left", "right" und "both" sind zulässig, wenn der Qualifier "direction" verwendet wird, um den Merkmalschlüssel "rep_origin" zu annotieren. Wird der "direction"-Qualifier jedoch verwendet, um den Merkmalschlüssel "oriT" zu annotieren, sind nur die Werte "left" und "right" zulässig.

6.17.	Qualifier	EC_number
	Definition	Enzyme Commission number (EC-Nummer) für das Enzymprodukt der Sequenz
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>1.1.2.4</INSDQualifier_value> <INSDQualifier_value>1.1.2.-</INSDQualifier_value> <INSDQualifier_value>1.1.2.n</INSDQualifier_value> <INSDQualifier_value>1.1.2.n1</INSDQualifier_value>
	Anmerkung	Die gültigen Werte für EC-Nummern werden in der Liste definiert, die vom Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) herausgegeben wird (veröffentlicht in Enzyme Nomenclature 1992, Academic Press, San Diego, oder einer neueren Ausgabe derselben). Das Format besteht aus einer Zeichenfolge aus vier durch Punkte getrennten Ziffern; bis zu drei Ziffern vom Ende der Zeichenkette her können durch den Bindestrich "-" ersetzt werden, um eine noch nicht spezifizierte Zuordnung anzuzeigen. Wenn noch keine EC-Nummer zugewiesen wurde, können an der letzten Position anstelle einer Ziffer Symbole verwendet werden, die ein "n" enthalten, z. B. "n", "n1" usw. Zu beachten ist, dass solche unvollständigen EC-Nummern vom NC-IUBMB nicht anerkannt werden.

6.18.	Qualifier	ecotype
	Definition	eine Population innerhalb einer bestimmten Spezies mit genetisch bedingten phänotypischen Merkmalen, die die Anpassung an ein lokales Habitat widerspiegeln
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>Columbia</INSDQualifier_value>

	Anmerkung	Beispielsweise eine Population, die als Reaktion auf einen besonders sonnigen Standort haarigere Blätter als üblich entwickelt hat. Der Qualifier "ecotype" wird oft auf genetische Standardbestände von Arabidopsis thaliana angewendet, kann aber auf jeden sessilen Organismus angewendet werden.
6.19.	Qualifier	environmental_sample
	Definition	Kennzeichnet Sequenzen, die durch direkte molekulare Isolierung aus einer großen Umwelt-DNA-Probe (mittels PCR mit oder ohne anschließender Klonierung des Produkts, DGGE oder anderer nicht genannter Methoden) gewonnen wurden, ohne dass der Herkunftsorganismus zuverlässig identifiziert werden konnte. Zu Umweltproben gehören klinische Proben, Darminhalte und andere Sequenzen von anonymen Organismen, die mit einem bestimmten Wirt in Verbindung gebracht werden können. Nicht zu den Umweltproben gehören Endosymbionten, die zuverlässig von einem bestimmten Wirt gewonnen werden können, Organismen aus einer leicht identifizierbaren, aber nicht kultivierten Feldprobe (z. B. viele Cyanobakterien) oder Phytoplasmen, die zuverlässig von kranken Pflanzen gewonnen werden können (auch wenn diese nicht in einer axenischen Kultur angebaut werden können).
	Wertformat	none
	Anmerkung	Wird nur mit dem Merkmalschlüssel "source" verwendet; ein "source"-Merkmalschlüssel, der den Qualifier "environmental_sample" enthält, sollte auch den Qualifier "isolation_source" enthalten; ein "source"-Merkmalschlüssel, der den Qualifier "environmental_sample" enthält, darf nicht den Qualifier "strain" enthalten.
6.20.	Qualifier	exception
	Definition	gibt an, dass die codierende Region nicht nach den biologischen Standardregeln translatiert werden kann
	Obligatorisches Wertformat	einer der folgenden Ausdrücke des kontrollierten Vokabulars: RNA editing rearrangement required for product annotated by transcript or proteomic data
	Beispiel	<INSDQualifier_value>RNA editing</INSDQualifier_value> <INSDQualifier_value>rearrangement required for product</INSDQualifier_value>
	Anmerkung	Nur zur Beschreibung biologischer Mechanismen wie RNA-Editing zu verwenden; der Qualifier "exception" für den Merkmalschlüssel "CDS" kennzeichnet eine Proteintranslation, die sich von der entsprechenden konzeptionellen Translation unterscheidet; darf nicht verwendet werden, wenn der Qualifier "transl_except" angemessen wäre, z. B. bei Abschluss durch ein Stopcodon.
6.21.	Qualifier	frequency
	Definition	Häufigkeit des Auftretens eines Merkmals
	Obligatorisches Wertformat	Freitext, in dem der Anteil der Population, der das Merkmal aufweist, als Bruchteil angegeben wird. sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value>
6.22.	Qualifier	function
	Definition	Funktion, die einer Sequenz zugeordnet ist

	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>essential for recognition of cofactor</INSDQualifier_value>
	Anmerkung	Der Qualifier "function" wird verwendet, wenn die Funktion einer Sequenz nicht aus dem Namen des Gens und/oder Produkts hervorgeht.

6.23.	Qualifier	gene
	Definition	Symbol des Gens, das einer Sequenzregion entspricht
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>ilvE</INSDQualifier_value>
	Anmerkung	Der Qualifier "gene" dient zur Angabe des Gensymbols; der Qualifier "standard_name" dient zur Angabe des vollständigen Namens des Gens.

6.24.	Qualifier	gene_synonym
	Definition	synonymes, ersetztes, veraltetes oder ehemaliges Gensymbol
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>Hox-3.3</INSDQualifier_value> in einem Merkmal mit Hoxc6 als Wert des Qualifiers "gene"
	Anmerkung	Wird verwendet, wenn die Angabe eines Synonyms für das Gensymbol sinnvoll erscheint; bei Verwendung des Qualifiers "gene_synonym" muss stets das primäre Gensymbol in einem „gene“-Qualifier angegeben werden.

6.25.	Qualifier	germline
	Definition	die dargestellte Sequenz wurde nicht im Zuge einer adaptiven Immunantwort somatisch rekombiniert; es handelt sich um die nicht rekombinierte Sequenz, die von der elterlichen Keimbahn ererbt wurde
	Wertformat	none
	Anmerkung	Der Qualifier "germline" darf nicht verwendet werden, um anzuzeigen, dass die Sequenz aus einer Gamet- bzw. Keimzelle stammt; die Qualifier "germline" und "rearranged" dürfen nicht im selben "source"-Merkmalschlüssel verwendet werden; die Qualifier "germline" und "rearranged" dürfen nur für Moleküle verwendet werden, die im Zuge einer adaptiven Immunreaktion eine somatische Rekombination vollziehen können; dabei handelt es sich um die T-Zell-Rezeptor- (TCR) und Immunglobulin-Loci bei Kiefermäulern und den nicht verwandten variablen Lymphozyten-Rezeptor- (VLR) Locus bei Kieferlosen (Neunaugen und Schleimaale); die Qualifier "germline" und "rearranged" sollten nicht außerhalb der Craniota (Schädeltiere) (taxid=89593) verwendet werden.

6.26.	Qualifier	haplogroup
	Definition	Name für eine Gruppe ähnlicher Haplotypen mit einer gemeinsamen Sequenzvariation. Haplogruppen werden häufig verwendet, um die Migration von Populationsgruppen zu verfolgen.
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.

	Beispiel	<INSDQualifier_value>H*</INSDQualifier_value>

6.27.	Qualifier	haplotype
	Definition	Name für einen bestimmten Satz von Allelen, die auf demselben physischen Chromosom miteinander verbunden sind. Wenn keine Rekombination stattfindet, wird jeder Haplotyp als Einheit vererbt und kann verwendet werden, um den Genfluss in Populationen zu verfolgen.
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>Dw3 B5 Cw1 A1</INSDQualifier_value>

6.28.	Qualifier	host
	Definition	natürlicher (nicht im Labor gehaltener) Wirt des Organismus, dem das sequenzierte Molekül entstammt
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>Homo sapiens</INSDQualifier_value> <INSDQualifier_value>Homo sapiens 12 year old girl</INSDQualifier_value> <INSDQualifier_value>Rhizobium NGR234</INSDQualifier_value>

6.29.	Qualifier	identified_by
	Definition	Name des Experten, der die Probe taxonomisch identifiziert hat
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>John Burns</INSDQualifier_value>

6.30.	Qualifier	isolate
	Definition	individuelles Isolat, aus dem die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>Patient #152</INSDQualifier_value> <INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>

6.31.	Qualifier	isolation_source
	Definition	beschreibt die physische, ökologische und/oder lokale geografische Herkunft der biologischen Probe, aus der die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiele	<INSDQualifier_value>rumen isolates from standard pelleted ration-fed steer #67</INSDQualifier_value> <INSDQualifier_value>permanent Antarctic sea ice</INSDQualifier_value> <INSDQualifier_value>denitrifying activated sludge from carbon_limited continuous reactor</INSDQualifier_value>

	Anmerkung	Wird nur im Zusammenhang mit dem Merkmalschlüssel "source" verwendet; ein "source"-Merkmalschlüssel, der den Qualifier "environmental_sample" enthält, sollte auch den Qualifier "isolation_source" enthalten.
6.32.	Qualifier	lab_host
	Definition	wissenschaftlicher Name des Laborwirts, der zur Vermehrung des Herkunftsorganismus verwendet wurde, aus dem das sequenzierte Molekül gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>Gallus gallus</INSDQualifier_value> <INSDQualifier_value>Gallus gallus embryo</INSDQualifier_value> <INSDQualifier_value>Escherichia coli strain DH5 alpha</INSDQualifier_value> <INSDQualifier_value>Homo sapiens HeLa cells</INSDQualifier_value>
	Anmerkung	Sofern bekannt, sollte der vollständige binäre wissenschaftliche Name des Wirtsorganismus verwendet werden; auch ergänzende Informationen zum Wirt sind möglich.
6.33.	Qualifier	lat_lon
	Definition	die geografischen Koordinaten des Orts, an dem die Probe entnommen wurde
	Obligatorisches Wertformat	Freitext - Breiten- und Längengrad im Format „d[d.dddd] N|S d[dd.dddd] W|E"
	Beispiel	<INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value> <INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value>
6.34.	Qualifier	macronuclear
	Definition	wenn es sich bei der dargestellten Sequenz um DNA handelt und diese einem Organismus entstammt, der eine chromosomale Differenzierung zwischen makronuklearen und mikronuklearen Stadien durchläuft, zeigt dieser Qualifier an, dass die Sequenz aus makronuklearer DNA stammt
	Wertformat	none
6.35.	Qualifier	map
	Definition	Position des Merkmals auf der Genomkarte
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>8q12-q13</INSDQualifier_value>
6.36.	Qualifier	mating_type
	Definition	Paarungstyp des Organismus, aus dem die Sequenz gewonnen wurde; der Qualifier "mating_type" wird für Prokaryoten verwendet und für Eukaryoten, die eine Meiose ohne geschlechtsdimorphe Gameten durchlaufen
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiele	<INSDQualifier_value>MAT-1</INSDQualifier_value> <INSDQualifier_value>plus</INSDQualifier_value> <INSDQualifier_value>-</INSDQualifier_value> <INSDQualifier_value>odd</INSDQualifier_value>

		<INSDQualifier_value>even</INSDQualifier_value>
	Anmerkung	Die Werte "male" (männlich) und "female" (weiblich) für den Qualifier "mating_type" sind für Prokaryoten gültig, nicht aber für Eurokaryoten; Näheres ist dem Eintrag für den Qualifier "sex" zu entnehmen.

6.37.	Qualifier	mobile_element_type
	Definition	Typ und Name oder Kennung des mobilen Elements, das im übergeordneten Merkmal beschrieben wird
	Obligatorisches Wertformat	<mobile_element_type>[:<mobile_element_name>] wobei <mobile_element_type> einer der folgenden Werte ist: transposon retrotransposon integron insertion sequence non-LTR retrotransposon SINE MITE LINE other
	Beispiel	<INSDQualifier_value>transposon:Tnp9</INSDQualifier_value>
	Anmerkung	Der Qualifier "mobile_element_type" ist nur mit dem Merkmalschlüssel "mobile_element" zulässig. Der Merkmalschlüssel "mobile_element" sollte verwendet werden, um sowohl die Elemente darzustellen, die derzeit mobil sind, als auch die Elemente, die in der Vergangenheit mobil waren. Der Wert "other" für <mobile_element_type> setzt das Vorhandensein von <mobile_element_name> voraus.

6.38.	Qualifier	mod_base
	Definition	Abkürzung für eine modifizierte Nukleotidbase
	Obligatorisches Wertformat	Abkürzung für modifizierte Basen aus diesem Anhang, Abschnitt 2
	Beispiel	<INSDQualifier_value>m5c</INSDQualifier_value> <INSDQualifier_value>OTHER</INSDQualifier_value>
	Anmerkung	Spezifische modifizierte Nukleotide, die in Abschnitt 2 dieses Anhangs nicht aufgeführt sind, werden durch die Qualifier "mod_base" und "note" annotiert, wobei "mod_base" als Wert "OTHER" und "note" als Wert den vollständigen Namen der modifizierten Base enthält.

6.39.	Qualifier	mol_type
	Definition	Molekültyp der Sequenz
	Obligatorisches Wertformat	einer der folgenden Werte: genomic DNA genomic RNA mRNA tRNA rRNA other RNA other DNA transcribed RNA viral cRNA unassigned DNA unassigned RNA
	Beispiel	<INSDQualifier_value>genomic DNA</INSDQualifier_value> <INSDQualifier_value>other RNA</INSDQualifier_value>
	Anmerkung	Der Qualifier "mol_type" ist für den Merkmalschlüssel "source" obligatorisch; der Wert "genomic DNA" impliziert nicht, dass es sich um ein Zellkernmolekül handelt (so müssen z. B. Organell- und Plasmid-DNA mit dem Wert "genomic DNA" beschrieben werden); ribosomale RNA-Gene müssen mit dem Wert "genomic DNA" beschrieben werden; "rRNA" darf nur verwendet werden, wenn das ribosomale RNA-Molekül selbst sequenziert wurde; die Werte "other RNA" und "other DNA" müssen für synthetische Moleküle verwendet werden; die Werte "unassigned DNA" bzw. "unassigned RNA" müssen verwendet werden, wenn das In-vivo-Molekül nicht bekannt ist.
6.40.	Qualifier	ncRNA_class
	Definition	eine strukturierte Beschreibung der Klassifizierung der nicht codierenden RNA, die mit dem übergeordneten Merkmalschlüssel "ncRNA" beschrieben wird
	Obligatorisches Wertformat	TYPE wobei TYPE einer der folgenden Ausdrücke des kontrollierten Vokabulars ist: antisense_RNA autocatalytically_spliced_intron circRNA ribozyme hammerhead_ribozyme lncRNA RNase_P_RNA RNase_MRP_RNA telomerase_RNA guide_RNA sgRNA rasiRNA scRNA scaRNA siRNA pre_miRNA miRNA piRNA snoRNA snRNA SRP_RNA vault_RNA Y_RNA other
	Beispiel	<INSDQualifier_value>autocatalytically_spliced_intron </INSDQualifier_value> <INSDQualifier_value>siRNA</INSDQualifier_value> <INSDQualifier_value>scRNA</INSDQualifier_value> <INSDQualifier_value>other</INSDQualifier_value>
	Anmerkung	Spezifische ncRNA-Typen, die noch nicht im kontrollierten Vokabular für den Wert des Qualifiers "ncRNA_class" enthalten sind, müssen mit "other" angegeben und mit

		einem "note"-Qualifier versehen werden, in dem die Klassifizierung der neuen nicht codierenden RNA (ncRNA_class) kurz beschrieben wird.
6.41.	Qualifier	note
	Definition	Anmerkungen oder Zusatzinformationen jeder Art
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>A comment about the feature</INSDQualifier_value>
6.42.	Qualifier	number
	Definition	eine Zahl, die die Reihenfolge der genetischen Elemente (z. B. Exons oder Introns) in 5'-3'-Richtung angibt
	Obligatorisches Wertformat	Freitext (ohne Whitespaces)
	Beispiel	<INSDQualifier_value>4</INSDQualifier_value> <INSDQualifier_value>6B</INSDQualifier_value>
	Anmerkung	Der Text darf nur aus ganzen Zahlen, Buchstaben oder einer Kombination aus ganzen Zahlen und/oder Buchstaben bestehen, die als Datenwert ohne Whitespaces dargestellt werden; für zusätzliche Angaben sollte der Qualifier "standard_name" verwendet werden. Beispiel: "number"-Qualifier mit dem Wert "2A" und "standard_name"-Qualifier mit dem Wert "long" ("lang").
6.43.	Qualifier	operon
	Definition	Name der Gruppe zusammenhängender Gene, die in ein einzelnes Transkript transkribiert wurden, zu dem dieses Merkmal gehört
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>lac</INSDQualifier_value>
6.44.	Qualifier	organelle
	Definition	Typ der membrangebundenen intrazellulären Struktur, aus der die Sequenz gewonnen wurde
	Obligatorisches Wertformat	einer der folgenden Ausdrücke des kontrollierten Vokabulars: chromatophore hydrogenosome mitochondrion nucleomorph plastid mitochondrion:kinetoplast plastid:chloroplast plastid:apicoplast plastid:chromoplast plastid:cyanelle plastid:leucoplast plastid:proplastid
	Beispiele	<INSDQualifier_value>chromatophore</INSDQualifier_value> <INSDQualifier_value>hydrogenosome</INSDQualifier_value> <INSDQualifier_value>mitochondrion</INSDQualifier_value> <INSDQualifier_value>nucleomorph</INSDQualifier_value> <INSDQualifier_value>plastid</INSDQualifier_value> <INSDQualifier_value>mitochondrion:kinetoplast</INSDQualifier_value> <INSDQualifier_value>plastid:chloroplast</INSDQualifier_value>

<INSDQualifier_value>plastid:apicoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:chromoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:cyanelle</INSDQualifier_value>
<INSDQualifier_value>plastid:leucoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:proplastid</INSDQualifier_value>

6.45.	Qualifier	organism
	Definition	wissenschaftlicher Name des Organismus, aus dem das sequenzierte genetische Material entnommen wurde (falls bekannt), oder die verfügbaren taxonomischen Informationen (falls der Organismus nicht klassifiziert ist); oder ein Hinweis darauf, dass es sich bei der Sequenz um ein synthetisches Konstrukt handelt
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>

6.46.	Qualifier	PCR_primers
	Definition	PCR-Primer, die zur Amplifizierung der Sequenz verwendet wurden. Alle Primer, die für eine einzige PCR-Reaktion verwendet wurden, sollten in einem einzigen "PCR_primers"-Qualifier aufgeführt werden. Wenn an einer einzigen PCR-Reaktion mehrere Vorwärts- oder Rückwärtsprimer beteiligt sind, ergeben sich mehrere Sätze von Werten für fwd_name/fwd_seq bzw. rev_name/rev_seq.
	Obligatorisches Wertformat	[fwd_name: XXX1,]fwd_seq: xxxxx1,[fwd_name: XXX2,]fwd_seq: xxxxx2, [rev_name: YYY1,]rev_seq: yyyyy1,[rev_name: YYY2,]rev_seq: yyyyy2
	Beispiel	<INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt,rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value> <INSDQualifier_value>fwd_name: hoge1, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<i>gtgtatcttact</INSDQualifier_value> <INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt, fwd_name: CO1P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value>
	Anmerkung	"fwd_seq" und "rev_seq" sind beide obligatorisch; "fwd_name" und "rev_name" sind beide fakultativ. Beide Sequenzen müssen in 5'-3'-Richtung dargestellt werden. Die Sequenzen müssen in den Symbolen aus Abschnitt 1 dieses Anhangs angegeben werden, mit Ausnahme der modifizierten Basen, die in spitze Klammern ("<" und ">") eingeschlossen sein müssen. Da die spitzen Klammern "<" und ">" in XML reservierte Zeichen sind, müssen sie durch "<" und ">" ersetzt werden.

6.47.	Qualifier	phenotype
	Definition	Phänotyp, der durch das Merkmal bestimmt wird, wobei der Ausdruck Phänotyp ein physikalisches, biochemisches oder verhaltensbezogenes Merkmal oder einen entsprechenden Satz von Merkmalen bezeichnet
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>erythromycin resistance</INSDQualifier_value>

6.48.	Qualifier	plasmid
	Definition	Name des natürlich vorkommenden Plasmids, aus dem die Sequenz gewonnen wurde, wobei der Ausdruck Plasmid eine unabhängig replizierende genetische Einheit bezeichnet, die nicht durch die Qualifier "chromosome" oder "segment" beschrieben werden kann

	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>pC589</INSDQualifier_value>

6.49.	Qualifier	pop_variant
	Definition	Name der Subpopulation oder des Phänotyps der Probe, aus der die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>pop1</INSDQualifier_value> <INSDQualifier_value>Bear Paw</INSDQualifier_value>

6.50.	Qualifier	product
	Definition	Name des Produkts, das mit dem Merkmal verbunden ist, z. B. die mRNA des Merkmals "mRNA", das Polypeptid des Merkmals "CDS", das reife Peptid des Merkmals "mat_peptide" usw.
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>trypsinogen</INSDQualifier_value> (wenn der Qualifier im Merkmal "CDS" aufgeführt wird) Beispiel <INSDQualifier_value>trypsin</INSDQualifier_value> (wenn der Qualifier im Merkmal "mat_peptide" aufgeführt wird) Beispiel <INSDQualifier_value>XYZ neural-specific transcript</INSDQualifier_value> (wenn der Qualifier im Merkmal "mRNA" aufgeführt wird)

6.51.	Qualifier	protein_id
	Definition	Sequenzkennzahl des Proteins, im Sequenzprotokoll eine ganze Zahl zur Kennzeichnung der Proteinsequenz, die durch die im "CDS"-Merkmalschlüssel und im "translation"-Qualifier angegebene codierende Sequenz codiert wird
	Obligatorisches Wertformat	eine ganze Zahl größer Null
	Beispiel	<INSDQualifier_value>89</INSDQualifier_value>

6.52.	Qualifier	proviral
	Definition	dieser Qualifier kennzeichnet eine Sequenz, die aus einem Virus oder Phagen gewonnen wurde, der in das Genom eines anderen Organismus integriert ist
	Wertformat	none

6.53.	Qualifier	pseudo
	Definition	gibt an, dass es sich bei diesem Merkmal um eine funktionslose Version des im Merkmalschlüssel genannten Elements handelt
	Wertformat	none
	Anmerkung	Der Qualifier "pseudo" sollte für funktionslose Gene verwendet werden, die nicht formell als Pseudogene beschrieben werden, z. B., wenn die CDS aus anderen Gründen als Pseudogenisierungs-Ereignissen nicht translatiert wird. Solche anderen Gründe können auch Sequenzierungs- oder Assemblierungsfehler sein. Zur Annotierung von

Pseudogenen muss der Qualifier "pseudogene" verwendet werden, der die Art ("TYPE") des Pseudogens angibt.

6.54.	Qualifier	pseudogene
	Definition	gibt an, dass es sich bei diesem Merkmal um ein Pseudogen des im Merkmalschlüssel genannten Elements handelt
	Obligatorisches Wertformat	TYPE wobei TYPE einer der folgenden Ausdrücke des kontrollierten Vokabulars ist: processed unprocessed unitary allelic unknown
	Beispiel	<INSDQualifier_value>processed</INSDQualifier_value> <INSDQualifier_value>unprocessed</INSDQualifier_value> <INSDQualifier_value>unitary</INSDQualifier_value> <INSDQualifier_value>allelic</INSDQualifier_value> <INSDQualifier_value>unknown</INSDQualifier_value>
	Anmerkung	Definition der Werte für "TYPE": "processed" - Das Pseudogen ist durch eine reverse Transkription von mRNA in cDNA und eine anschließende Reintegration in das Genom entstanden. Daher hat es jede Intron/Exon-Struktur verloren und könnte einen Pseudo-PolyA-Schwanz haben. "unprocessed" - Das Pseudogen ist aus einer Kopie des Elterngens durch Duplikation entstanden, gefolgt von der Ansammlung zufälliger Mutationen. Zu den Veränderungen im Vergleich zum funktionalen homologen Gen gehören Insertionen, Deletionen, vorzeitige Stopcodons, Leserasterverschiebungen und ein höherer Anteil an nicht-synonymen gegenüber synonymen Substitutionen. "unitary" - Das Pseudogen stammt nicht von einem Elterngen ab. Es ist das ursprüngliche Gen, das in einigen Arten funktional ist, in anderen Arten oder Stämmen jedoch in irgendeiner Weise (Indel, Mutation, Rekombination) daran gehindert wird. "allelic" - Ein (unitäres) Pseudogen, das in der Population stabil ist, dort aber vor allem auch ein funktionales alternatives Allel hat. Somit kann ein Stamm das Gen und ein anderer das Pseudogen aufweisen. MHC-Haplotypen haben allelische Pseudogene. "unknown" - Der Einreichende kennt das Ereignis der Pseudogenisierung nicht.

6.55.	Qualifier	rearranged
	Definition	die im Eintrag dargestellte Sequenz wurde im Zuge einer adaptiven Immunantwort somatisch rekombiniert; es handelt sich nicht um die nicht rekombinierte Sequenz, die von der elterlichen Keimbahn ererbt wurde
	Wertformat	none
	Anmerkung	Der Qualifier "rearranged" darf nicht verwendet werden, um Rekombinationen von Chromosomen anzuzeigen, die nicht Bestandteil einer adaptiven Immunreaktion sind; die Qualifier "germline" und "rearranged" dürfen nicht im selben "source"-Merkmalschlüssel verwendet werden; die Qualifier "germline" und "rearranged" dürfen nur für Moleküle verwendet werden, die im Zuge einer adaptiven Immunreaktion eine somatische Rekombination vollziehen können; dabei handelt es sich um die T-Zell-Rezeptor- (TCR) und Immunglobulin-Loci bei Kiefermäulern und den nicht verwandten variablen Lymphozyten-Rezeptor- (VLR) Locus bei Kieferlosen (Neunaugen und Schleimaale); die Qualifier "germline" und "rearranged" sollten nicht außerhalb der Craniota (Schädeltiere) (taxid=89593) verwendet werden.

6.56.	Qualifier	recombination_class
	Definition	eine strukturierte Beschreibung der Klassifizierung von Hotspot-Regionen der Rekombination innerhalb einer Sequenz

	Obligatorisches Wertformat	TYPE wobei TYPE einer der folgenden Ausdrücke des kontrollierten Vokabulars ist: meiotic mitotic non_allelic_homologous chromosome_breakpoint other
	Beispiel	<INSDQualifier_value>meiotic</INSDQualifier_value> <INSDQualifier_value>chromosome_breakpoint</INSDQualifier_value>
	Anmerkung	Spezifische Rekombinationsklassen, die noch nicht im kontrollierten Vokabular für den Wert des Qualifiers "recombination_class" enthalten sind, müssen mit "other" angegeben und mit einem "note"-Qualifier versehen werden, in dem die Klassifizierung der neuen Rekombinationsklasse(recombination_class) kurz beschrieben wird.
6.57.	Qualifier	regulatory_class
	Definition	eine strukturierte Beschreibung der Klassifizierung der in einer Sequenz enthaltenen regulatorischen Elemente mit Bezug auf Transkription, Translation, Replikation und Chromatinstruktur
	Obligatorisches Wertformat	TYPE wobei TYPE einer der folgenden Ausdrücke des kontrollierten Vokabulars ist: attenuator CAAT_signal DNase_I_hypersensitive_site enhancer enhancer_blocking_element GC_signal imprinting_control_region insulator locus_control_region matrix_attachment_region minus_35_signal minus_10_signal polyA_signal_sequence promoter recoding_stimulatory_region recombination_enhancer replication_regulatory_region response_element ribosome_binding_site riboswitch silencer TATA_box terminator transcriptional_cis_regulatory_region uORF other
	Beispiel	<INSDQualifier_value>promoter</INSDQualifier_value> <INSDQualifier_value>enhancer</INSDQualifier_value> <INSDQualifier_value>ribosome_binding_site</INSDQualifier_value>
	Anmerkung	Spezifische regulatorische Klassen, die noch nicht im kontrollierten Vokabular für den Wert des Qualifiers "regulatory_class" enthalten sind, müssen mit "other" angegeben und mit einem "note"-Qualifier versehen werden, in dem die Klassifizierung der neuen regulatorischen Klasse (regulatory_class) kurz beschrieben wird.
6.58.	Qualifier	replace
	Definition	gibt an, dass die an der Lage des Merkmals identifizierte Sequenz durch die im Wert des Qualifiers angezeigte Sequenz ersetzt wird; wenn der Qualifier keine Sequenz (d. h. keinen Wert) enthält, zeigt dies eine Deletion an

	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>a</INSDQualifier_value> <INSDQualifier_value></INSDQualifier_value> - für eine Deletion

6.59.	Qualifier	ribosomal_slippage
	Definition	während der Proteintranslation können bestimmte Sequenzen Ribosomen dazu bringen, zu einem alternativen Leseraster zu wechseln (Frameshift), und zwar durch einen Mechanismus, der als „ribosomal Slippage" bekannt ist
	Wertformat	none
	Anmerkung	Bei der Angabe der Lage des Merkmals "CDS" muss ein "join"-Operator wie z. B. [join(486..1784,1787..4810)] verwendet werden, um anzuzeigen, an welcher Stelle der Frameshift auftritt.

6.60.	Qualifier	rpt_family
	Definition	Typ der repetitiven Sequenz, z. B. "Alu" oder "Kpn"
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>Alu</INSDQualifier_value>

6.61.	Qualifier	rpt_type
	Definition	Struktur und Verteilung der repetitiven Sequenz
	Obligatorisches Wertformat	einer der folgenden Ausdrücke des kontrollierten Vokabulars: tandem direct inverted flanking nested terminal dispersed long_terminal_repeat non_ltr_retrotransposon_polymeric_tract centromeric_repeat telomeric_repeat x_element_combinatorial_repeat y_prime_element other
	Beispiel	<INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value>
	Anmerkung	Definition der Werte: "tandem" - eine repetitive Sequenz, die benachbart und in gleicher Orientierung zu einer anderen vorliegt; "direct" - eine repetitive Sequenz, die nicht immer benachbart, aber in gleicher Orientierung vorliegt; "inverted" – ein repetitives Paar, das in reverser Orientierung auf demselben Molekül vorliegt; "flanking" - eine repetitive Sequenz, die außerhalb der Sequenz vorliegt, für die sie funktionale Bedeutung hat (z. B. an den Zielorten für die Insertion eines Transposons); "nested" - eine repetitive Sequenz, die durch die Insertion eines anderen Elements unterbrochen wird; "dispersed" - eine repetitive Sequenz, die im gesamten Genom verteilt vorliegt; "terminal" - eine repetitive Sequenz an den Enden und innerhalb der Sequenz, für die sie funktionale Bedeutung hat (z. B. Transposon-LTRs);

"long_terminal_repeat" – eine Sequenz, die direkt an beiden Enden einer definierten Sequenz wiederholt wird, von der Art, wie sie in der Regel bei Retroviren zu finden ist;
"non_ltr_retrotransposon_polymeric_tract" - ein Polymertrakt, z. B. poly(dA), innerhalb eines Retrotransposons ohne LTR;
"centromeric_repeat" - eine repetitive Region im modularen Zentromer;
"telomeric_repeat" - eine repetitive Region im Telomer;
"x_element_combinatorial_repeat" - eine repetitive Region zwischen dem X-Element und dem Telomer oder dem benachbarten Y'-Element (kombinatorisches Repetitiv);
"y_prime_element" - eine repetitive Region, die an Repetitive im Telomer oder kombinatorische Repetitive des X-Elements angrenzt, entweder als einzelne Kopie oder als Tandem-Repetitiv von zwei bis vier Kopien;
"other" - ein Repetitiv mit wichtigen Attributen, die nicht durch andere Werte beschrieben werden können.

6.62.	Qualifier	rpt_unit_range
	Definition	Lage einer repetitiven Einheit, ausgedrückt als Bereich
	Obligatorisches Wertformat	<base_range> - wobei <base_range> die erste und letzte Base (durch zwei Punkte getrennt) einer repetitiven Einheit wiedergibt
	Beispiel	<INSDQualifier_value>202..245</INSDQualifier_value>
	Anmerkung	Dient zur Angabe des Basenbereichs der Sequenz, der eine repetitive Einheit innerhalb der durch die Merkmalschlüssel "OriT" und "repeat_region" spezifizierten Region darstellt.

6.63.	Qualifier	rpt_unit_seq
	Definition	Identität einer repetitiven Sequenz
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>aagggc</INSDQualifier_value> <INSDQualifier_value>ag(5)tg(8)</INSDQualifier_value> <INSDQualifier_value>(AAAGA)6(AAAA)1(AAAGA)12</INSDQualifier_value>
	Anmerkung:	Dient zur Angabe der Zeichenabfolge der Sequenz, die eine repetitive Einheit innerhalb der durch die Merkmalschlüssel "OriT" und "repeat_region" spezifizierten Region darstellt.

6.64.	Qualifier	satellite
	Definition	Kennung für einen Satelliten-DNA-Marker, bestehend aus zahlreichen (identischen oder verbundenen) Tandem-Repetitiven einer kurzen repetitiven Grundeinheit
	Obligatorisches Wertformat	<satellite_type>[:<class>][<identifier>] - wobei <satellite_type> einer der folgenden Werte ist: satellite microsatellite minisatellite
	Beispiel	<INSDQualifier_value>satellite: S1a</INSDQualifier_value> <INSDQualifier_value>satellite: alpha</INSDQualifier_value> <INSDQualifier_value>satellite: gamma III</INSDQualifier_value> <INSDQualifier_value>microsatellite: DC130</INSDQualifier_value>
	Anmerkung	Viele Satelliten weisen eine Basenzusammensetzung oder andere Eigenschaften auf, die sich von denen des restlichen Genoms unterscheiden und anhand derer sie identifiziert werden können.

6.65.	Qualifier	segment
	Definition	Name des sequenzierten Viren- oder Phagensegments
	Obligatorisches Wertformat	Freitext
	Beispiel	<INSDQualifier_value>6</INSDQualifier_value>

6.66.	Qualifier	serotype
	Definition	serologische Varietät einer Spezies, die durch ihre antigenen Eigenschaften gekennzeichnet ist
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>B1</INSDQualifier_value>
	Anmerkung	Wird nur mit dem Merkmalschlüssel "source" verwendet; der Bakteriologische Code empfiehlt die Verwendung des Ausdrucks "serovar" anstelle von "serotype" für Prokaryoten; siehe den International Code of Nomenclature of Bacteria (1990 Revision) Anhang 10.B "Infraspecific Terms".

6.67.	Qualifier	serovar
	Definition	serologische Varietät einer Spezies (in der Regel prokariotisch), die durch ihre antigenen Eigenschaften gekennzeichnet ist
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>O157:H7</INSDQualifier_value>
	Anmerkung	Wird nur mit dem Merkmalschlüssel "source" verwendet; der Bakteriologische Code empfiehlt die Verwendung des Ausdrucks "serovar" anstelle von "serotype" für Prokaryoten; siehe den International Code of Nomenclature of Bacteria (1990 Revision) Anhang 10.B "Infraspecific Terms".

6.68.	Qualifier	sex
	Definition	Geschlecht des Organismus, aus dem die Sequenz gewonnen wurde; der Qualifier "sex" wird für eukaryotische Organismen verwendet, die eine Meiose durchlaufen und sexuell dimorphe Gameten haben
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiele	<INSDQualifier_value>female</INSDQualifier_value> <INSDQualifier_Value>male</INSDQualifier_Value> <INSDQualifier_Value>hermaphrodite</INSDQualifier_Value> <INSDQualifier_value>unisexual</INSDQualifier_value> <INSDQualifier_value>bisexual</INSDQualifier_value> <INSDQualifier_value>asexual</INSDQualifier_value> <INSDQualifier_value>monoecious</INSDQualifier_value> [oder monecious] <INSDQualifier_value>dioecious</INSDQualifier_value> [oder diecious]
	Anmerkung	Der Qualifier "sex" sollte (anstelle des Qualifiers "mating_type") bei Metazoa, Embryophyta, Rhodophyta und Phaeophyceae verwendet werden; der Qualifier "mating_type" sollte (anstelle des Qualifiers "sex") bei Bakterien, Archaea und Pilzen verwendet werden; bei Viren sollten weder der Qualifier "sex" noch der Qualifier "mating_type" verwendet werden; außerhalb der oben aufgeführten Taxa

		sollte der Qualifier "mating_type" verwendet werden, es sei denn, der Wert des Qualifiers wird aus dem in den obigen Beispielen angegebenen Vokabular übernommen.
6.69.	Qualifier	standard_name
	Definition	anerkannter Standardname für dieses Merkmal
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden
	Beispiel	<INSDQualifier_value>dotted</INSDQualifier_value>
	Anmerkung	Zur Angabe des vollständigen Gennamens ist der Qualifier "standard_name" zu verwenden, aber zur Angabe des Gensymbols ist der Qualifier "gene" zu verwenden (im obigen Beispiel ist der Wert des Qualifiers „gene“ Dt).
6.70.	Qualifier	strain
	Definition	Stamm, aus dem die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>BALB/c</INSDQualifier_value>
	Anmerkung	Einträge für ein Merkmal, die einen "strain"-Qualifier enthalten, dürfen keinen "environmental_sample"-Qualifier enthalten.
6.71.	Qualifier	sub_clone
	Definition	Subklon, aus dem die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>lambda-hIL7.20g</INSDQualifier_value>
	Anmerkung	Das Merkmal "source" darf nicht mehr als einen "subclone"-Qualifier enthalten; wenn die Sequenz aus mehreren Subklonen gewonnen wurde, können diese unter Verwendung des Merkmalschlüssels "misc_feature" und eines "note"-Qualifiers näher beschrieben werden.
6.72.	Qualifier	sub_species
	Definition	Name der Unterart des Organismus, aus der die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>lactis</INSDQualifier_value>
6.73.	Qualifier	sub_strain
	Definition	Name oder Kennung eines genetisch oder anderweitig modifizierten Stammes, aus dem die Sequenz gewonnen wurde, abgeleitet von einem Elternstamm (anzugeben im Qualifier "strain"). Unterstamm (sub_strain), aus dem die Sequenz gewonnen wurde

Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
Beispiel	<INSDQualifier_value>abis</INSDQualifier_value>
Anmerkung	Im Merkmal "source" muss dieser Qualifier mit einem "strain"-Qualifier einhergehen; wenn der Elternstamm nicht angegeben wird, sollte der modifizierte Stamm im Qualifier "strain" und nicht im Qualifier "sub_strain" annotiert werden. Beispiel: entweder der Qualifier "strain" mit dem Wert "K-12" und der Qualifier "sub_strain" mit dem Wert "MG1655" oder nur der Qualifier "strain" mit dem Wert "MG1655".

6.74.	Qualifier	tag_peptide
	Definition	Lage der Basen, an der das Polypeptid für den Proteolyse-Tag der tmRNA codiert wird, und sein Terminationscodon
	Obligatorisches Wertformat	<base_range> - wobei <base_range> die erste und letzte Base (durch zwei Punkte getrennt) der Lage des Proteolyse-Tags wiedergibt
	Beispiel	<INSDQualifier_value>90..122</INSDQualifier_value>
	Anmerkung	Es wird empfohlen, die Aminosäuresequenz, die dem Qualifier "tag_peptide" entspricht, durch Beschreibung des partiellen 5‘ CDS-Merkmals zu annotieren, z. B. "CDS“ mit der Lage <90..122.

6.75.	Qualifier	tissue_lib
	Definition	Gewebebibliothek, aus der die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>tissue library 772</INSDQualifier_value>

6.76.	Qualifier	tissue_type
	Definition	Gewebetyp, aus dem die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>liver</INSDQualifier_value>

6.77.	Qualifier	transl_except
	Definition	translationale Ausnahme: einzelnes Codon, dessen Translation nicht mit dem genetischen Code übereinstimmt, der durch die Qualifier "organism" oder "transl_table" definiert ist
	Obligatorisches Wertformat	(pos:<location>,aa:<amino_acid>), wobei <amino_acid> für die dreibuchstabige Abkürzung der Aminosäure steht, die vom Codon an der Position "base_range" codiert wird
	Beispiel	<INSDQualifier_value>(pos:213..215,aa:Trp)</INSDQualifier_value> <INSDQualifier_value>(pos:462..464,aa:OTHER)</INSDQualifier_value> <INSDQualifier_value>(pos:1017,aa:TERM)</INSDQualifier_value> <INSDQualifier_value>(pos:2000..2001,aa:TERM)</INSDQualifier_value>
	Anmerkung	Wenn die Aminosäure nicht zu den in Abschnitt 3 dieses Anhangs aufgeführten spezifischen Aminosäuren gehört, ist "OTHER" als <amino_acid> zu verwenden und der Namen der ungewöhnlichen Aminosäure in einem „NOTE"-Qualifier anzugeben; für die modifizierte Aminosäure Selenocystein ist die dreibuchstabige Abkürzung "Sec" (einbuchstabiges Symbol "U" in der Aminosäuresequenz) als <amino_acid> zu verwenden; für die modifizierte Aminosäure Pyrrolysin ist die dreibuchstabige Abkürzung "Pyl" (einbuchstabiges Symbol "O" in der Aminosäuresequenz) als <amino _acid> zu verwenden; die Lage unvollständiger Terminationscodons, bei denen das TAA-Stopcodon durch Hinzufügen von 3’-A-Resten zur mRNA komplementiert wird, wird entweder durch eine einzelne "base_position" oder durch einen Bereich "base_range" angegeben (siehe das dritte und vierte Beispiel oben), und zwar in Verbindung mit einem "note"-Qualifier", der die entsprechende Information enthält: "stop codon completed by the addition of 3’ A residues to the mRNA" ("Stopcodon durch Hinzufügen von 3’-A-Resten zur mRNA komplementiert").

6.78.	Qualifier	transl_table
	Definition	Festlegung der Tabelle für den genetischen Code, die verwendet wird, wenn es sich nicht um die Tabelle für den Standardcode bzw. den universellen genetischen Code handelt. Die verwendeten Tabellen sind in diesem Anhang beschrieben
	Obligatorisches Wertformat	<integer>, wobei <integer> die der Tabelle für den genetischen Code zugewiesene ganzzahlige Nummer ist
	Beispiel	<INSDQualifier_value>3</INSDQualifier_value> - Hier soll der mitochondriale Code von Hefen verwendet werden.
	Anmerkung	Wenn der Qualifier "transl_table" nicht zur weiteren Annotierung eines "CDS"-Merkmalschlüssels verwendet wird, dann werden die codierenden Sequenzen mit dem Standardcode (d. h. dem universellen genetischen Code) translatiert. Ausnahmen vom genetischen Code, die nicht in den Bereich der angegebenen Tabellen fallen, werden im Qualifier "transl_except" angegeben.

6.79.	Qualifier	trans_splicing
	Definition	zeigt an, dass Exons von zwei RNA-Molekülen in intermolekularer Reaktion zur Bildung reifer RNA ligiert sind
	Wertformat	none
	Anmerkung	Sollte in Verbindung mit Merkmalen wie "CDS", "mRNA" und anderen verwendet werden, die als Produkt eines Transspleißereignisses erzeugt werden. Dieser Qualifier darf nur verwendet werden, wenn bei der Angabe der Lage des Merkmals das Spleißereignis im Operator "join" angezeigt wird, z. B. als "join(complement(69611..69724),139856..140087)".

6.80.	Qualifier	translation
	Definition	eine einbuchstabig abgekürzte Aminosäuresequenz, abgeleitet vom genetischen Standardcode (oder universellen Code) oder einer im Qualifier "transl_table" angegebenen Tabelle und im Qualifier "transl_except" als Ausnahme angegeben
	Obligatorisches Wertformat	zusammenhängende Zeichenfolge aus einbuchstabigen Abkürzungen für Aminosäuren aus Abschnitt 3 dieses Anhangs, für Aminosäure-Ausnahmen ist "X" zu verwenden
	Beispiel	<INSDQualifier_value>MASTFPPWYRGCASTPSLKGLIMCTW</INSDQualifier_value>
	Anmerkung	Darf nur mit dem Merkmal "CDS" verwendet werden; muss mit dem Qualifier "protein_id" versehen werden, wenn das Ergebnis der Translation vier oder mehr spezifisch definierte Aminosäuren enthält; zur Festlegung und Kennzeichnung der Tabellen für den genetischen Code siehe "transl_table"; von den Qualifiern "translation", "pseudo" und "pseudogene" darf nur einer zur weiteren Annotierung eines "CDS"-Merkmals verwendet werden.

6.81.	Qualifier	variety
	Definition	Varietät (= Varietas, ein formaler Rang nach Linné) des Organismus, aus dem die Sequenz gewonnen wurde
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>insularis</INSDQualifier_value>
	Anmerkung	Für Kulturpflanzensorten, also Produkte künstlicher Selektion, ist der Qualifier "cultivar" zu verwenden; andere als Pflanzen- und Pilzvarietäten sollten mit einem "note"-Qualifier versehen werden, z. B. mit dem Wert <INSDQualifier_value>breed:Cukorova</INSDQualifier_value>

ABSCHNITT 7: MERKMALSCHLÜSSEL FÜR AMINOSÄURESEQUENZEN

Dieser Abschnitt enthält die Liste der zulässigen Merkmalschlüssel für Aminosäuresequenzen. Die Merkmalschlüssel werden in alphabetischer Reihenfolge aufgeführt.

7.1.	Merkmalschlüssel	ACT_SITE
	Definition	Aminosäure(n), die bei der Aktivität eines Enzyms mitwirkt (mitwirken)
	Fakultative Qualifier	note
	Anmerkung	Jeder Aminosäurerest des aktiven Zentrums muss separat mit dem Merkmalschlüssel "ACT_SITE" annotiert werden. Die entsprechende Restenummer des Aminosäurerests muss als Lagedeskriptor im Element für die Lage des Merkmals ("INSDFeature_location") angegeben werden.

7.2.	Merkmalschlüssel	BINDING
	Definition	Bindungsstelle für eine beliebige chemische Gruppe (Coenzym, prosthetische Gruppe usw.). Die chemische Beschaffenheit der Gruppe ist im "note"-Qualifier anzugeben.
	Obligatorische Qualifier	note
	Anmerkung	Beispiele für Werte des "note"-Qualifiers: "Heme (covalent)" ("Häm (kovalent)") und "Chloride" ("Chlorid"). Sofern geboten, sollten die Merkmalschlüssel "CA_BIND", "DNA_BIND", "METAL" und "NP_BIND" anstelle von "BINDING" verwendet werden.

7.3.	Merkmalschlüssel	CA_BIND
	Definition	Bereich einer Calcium-bindenden Region
	Fakultative Qualifier	note

7.4.	Merkmalschlüssel	CARBOHYD
	Definition	Glykosylierungsstelle
	Obligatorische Qualifier	note
	Anmerkung	Dieser Schlüssel beschreibt das Auftreten der Anhaftung eines Glykans (Mono- oder Polysaccharid) an einen Proteinrest. Die Art der Bindung (C-, N- oder O-Bindung) mit dem Protein wird im "note"-Qualifier angegeben. Wenn die Art des endständigen reduzierenden Zuckers bekannt ist, wird seine Abkürzung in Klammern angezeigt. Drei auf die Abkürzung folgende Punkte "..." zeigen eine Verlängerung der Kohlenhydratkette an. Wenn keine Punkte folgen, bedeutet dies umgekehrt, dass ein Monosaccharid gebunden ist. Beispiele für im "note"-Qualifier verwendete Werte: "N-linked (GlcNAc...)" ("N-verknüpftes GlcNAc"); "O-linked (GlcNAc)" ("O- verknüpftes GlcNAc"); "O-linked (Glc...)" ("O- verknüpftes Glc... "); "C-linked (Man) partial" ("C- verknüpftes Man partiell"); "O-linked (Ara...)" ("O- verknüpftes Ara...").

7.5.	Merkmalschlüssel	CHAIN
	Definition	Bereich einer Polypeptidkette im reifen Protein
	Fakultative Qualifier	note

7.6.	Merkmalschlüssel	COILED
	Definition	Bereich einer Coiled-Coil-(Doppelwendel) Region
	Fakultative Qualifier	note

7.7.	Merkmalschlüssel	COMPBIAS
	Definition	Bereich einer Region mit lokalen Häufungen (compositional bias)
	Fakultative Qualifier	note

7.8.	Merkmalschlüssel	CONFLICT
	Definition	verschiedene Quellen melden voneinander abweichende Sequenzen
	Fakultative Qualifier	note
	Anmerkung	Beispiele für Werte des "note"-Qualifiers: "Missing" ("fehlend"); "K -> Q"; "GSDSE -> RIRLR"; "V -> A".

7.9.	Merkmalschlüssel	CROSSLNK
	Definition	posttranslational gebildete Aminosäurebindungen
	Obligatorische Qualifier	note
	Anmerkung	Kovalente Bindungen verschiedener Typen, die zwischen zwei Proteinen (kettenübergreifende Quervernetzung bzw. "interchain cross-link") oder zwischen zwei Teilen desselben Proteins (ketteninterne Quervernetzung bzw. "intrachain cross-link") gebildet werden; ausgenommen sind Quervernetzungen, die durch Disulfidbindungen gebildet werden und für die der Merkmalschlüssel "DISULFID" zu verwenden ist. Bei einer kettenübergreifenden Quervernetzung besteht der Lagedeskriptor im Element für die Lage des Merkmals ("INSDFeature_location") aus der Restenummer der Aminosäure, die mit dem anderen Protein quervernetzt ist. Bei einer ketteninternen Quervernetzung besteht der Lagedeskriptor im Element für die Lage des Merkmals ("INSDFeature_location") aus den Restenummern der quervernetzten Aminosäuren im Format "x..y", z. B. "42..50". Der "note"-Qualifier gibt die Art der Quervernetzung an, zumindest den Namen des Konjugats und die Identität der beiden beteiligten Aminosäuren. Beispiele für Werte des "note"-Qualifiers: "Isoglutamyl cysteine thioester (Cys-Gln)" ("Isoglutamyl-Cystein-Thioester (Cys-Gln)"); "Beta-methyllanthionine (Cys-Thr)" ("Beta-Methyllanthionin (Cys-Thr)") und "Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin)" ("Glycyl-Lysin-Isopeptid (Lys-Gly) (kettenübergreifend mit G-Cter in Ubiquitin)")

7.10.	Merkmalschlüssel	DISULFID
	Definition	Disulfidbindung
	Obligatorische Qualifier	note
	Anmerkung	Bei einer kettenübergreifenden Disulfidbindung ist der Lagedeskriptor im Element für die Lage des Merkmals ("INSDFeature_location") die Restenummer der Cysteins, das an das andere Protein gebunden ist. Bei einer ketteninternen Quervernetzung besteht der Lagedeskriptor im Element für die Lage des Merkmals ("INSDFeature_location") aus den Restenummern der gebundenen Cysteine im Format "x..y", z. B. "42..50". Für kettenübergreifende Disulfidbindungen gibt der "note"-Qualifier die Art der Quervernetzung an, indem er das andere Protein identifiziert, z. B. "Interchain (between A and B chains)" ("Kettenübergreifend (zwischen A- und B-Kette)").

7.11.	Merkmalschlüssel	DNA_BIND
	Definition	Bereich einer DNA-bindenden Region
	Obligatorische Qualifier	note
	Anmerkung	Die Art der DNA-bindenden Region wird im "note"-Qualifier angegeben. Beispiele für Werte des "note"-Qualifiers: "Homeobox" ("Homöobox") und "MYB 2" ("Myb 2")

7.12.	Merkmalschlüssel	DOMAIN
	Definition	Bereich einer Domäne, die als eine spezifische Kombination von Sekundärstrukturen definiert ist, die in einer charakteristischen dreidimensionalen Struktur oder Faltung organisiert sind
	Obligatorische Qualifier	note
	Anmerkung	Die Art der Domäne wird im "note"-Qualifier angegeben. Wenn mehrere Kopien einer Domäne vorliegen, werden die Domänen nummeriert. Beispiele für Werte des "note"-Qualifiers: "Ras-GAP" und "Cadherin 1".

7.13.	Merkmalschlüssel	HELIX
	Definition	Sekundärstruktur: Helices, z. B. Alpha-Helix, 3(10)-Helix oder Pi-Helix
	Fakultative Qualifier	note
	Anmerkung	Dieses Merkmal wird nur für Proteine verwendet, deren tertiäre Struktur bekannt ist. Es werden nur drei Arten von Sekundärstrukturen angegeben: Helices (Schlüssel "HELIX"), Beta-Stränge (Schlüssel "STRAND") und Schleifen (Schlüssel "TURN"). Reste, die nicht in einer dieser Klassen angegeben sind, befinden sich in einer „loop"- oder „random-coil"-Struktur.

7.14.	Merkmalschlüssel	INIT_MET
	Definition	Start-Methionin
	Fakultative Qualifier	note
	Anmerkung	Der Lagedeskriptor im Element für die Lage des Merkmals ("INSDFeature_location") ist "1". Dieser Merkmalschlüssel zeigt an, dass das N-terminale Methionin abgespalten ist. Wenn das Initiator-Methionin nicht abgespalten ist, wird dieses Merkmal nicht verwendet.

7.15.	Merkmalschlüssel	INTRAMEM
	Definition	Bereich einer Region, der innerhalb einer Membran verbleibt und diese nicht durchquert
	Fakultative Qualifier	note

7.16.	Merkmalschlüssel	LIPID
	Definition	kovalente Bindung einer Lipideinheit
	Obligatorische Qualifier	note
	Anmerkung	Die chemische Beschaffenheit der gebundenen Lipideinheit wird im "note"-Qualifier angegeben, wobei mindestens der Name der lipidierten Aminosäure angegeben wird. Beispiele für Werte des "note"-Qualifiers: "N-myristoyl glycine" ("N-Myristoylglycin"); "GPI-anchor amidated serine" ("GPI-verankertes amidiertes Serin") und "S-diacylglycerol cysteine" ("S-Diacylglycerolcystein").

7.17.	Merkmalschlüssel	METAL
	Definition	Bindungsstelle für ein Metallion
	Obligatorische Qualifier	note

	Anmerkung	Der "note"-Qualifier gibt die Art des Metalls an. Beispiele für Werte des "note"-Qualifiers: "Iron (heme axial ligand)" ("Eisen (Hämo-Achsenligand)") und "Copper" ("Kupfer").
7.18.	Merkmalschlüssel	MOD_RES
	Definition	posttranslationale Modifikation eines Rests
	Obligatorische Qualifier	note
	Anmerkung	Die chemische Beschaffenheit des modifizierten Rests wird im "note"-Qualifier angegeben, dabei wird mindestens der Name der posttranslational modifizierten Aminosäure eingetragen. Wenn die modifizierte Aminosäure in Abschnitt 4 dieses Anhangs aufgeführt ist, kann anstelle des vollständigen Namens die Abkürzung verwendet werden. Beispiele für Werte des "note"-Qualifiers: "N-acetylalanine" ("N-Acetylalanin"), "3-Hyp" ("3-Hyp") und "MeLys" ("MeLys") oder "N-6-methyllysine" ("N-6-Methyllysin").
7.19.	Merkmalschlüssel	MOTIF
	Definition	kurzes (bis zu 20 Aminosäuren) biologisch bedeutsames Sequenzmotiv
	Fakultative Qualifier	note
7.20.	Merkmalschlüssel	MUTAGEN
	Definition	Stelle, die durch Mutagenese experimentell verändert wurde
	Fakultative Qualifier	note
7.21.	Merkmalschlüssel	NON_STD
	Definition	nicht standardmäßige Aminosäure
	Fakultative Qualifier	note
	Anmerkung	Dieser Schlüssel beschreibt nur das Auftreten der nicht standardmäßigen Aminosäuren Selenocystein (U) und Pyrrolysin (O) in der Aminosäuresequenz.
7.22.	Merkmalschlüssel	NON_TER
	Definition	der Rest am Sequenzanfang oder -ende ist nicht der Terminalrest
	Fakultative Qualifier	note
	Anmerkung	Bezogen auf Position 1 bedeutet dies, dass die erste Position nicht der N-Terminus des vollständigen Moleküls ist. Bezogen auf die letzte Position bedeutet dies, dass diese Position nicht der C-Terminus des vollständigen Moleküls ist.
7.23.	Merkmalschlüssel	NP_BIND
	Definition	Bereich einer Nukleotidphosphat-bindenden Region
	Obligatorische Qualifier	note
	Anmerkung	Die Art des Nukleotidphosphats wird im "note"-Qualifier angegeben. Beispiele für Werte des "note"-Qualifiers: "ATP" und "FAD".

7.24.	Merkmalschlüssel	PEPTIDE
	Definition	Bereich eines freigesetzten aktiven Peptids
	Fakultative Qualifier	note

7.25.	Merkmalschlüssel	PROPEP
	Definition	Bereich eines Propeptids
	Fakultative Qualifier	note

7.26.	Merkmalschlüssel	REGION
	Definition	Bereich einer bedeutsamen Region in der Sequenz
	Fakultative Qualifier	note

7.27.	Merkmalschlüssel	REPEAT
	Definition	Bereich einer internen Sequenzwiederholung
	Fakultative Qualifier	note

7.28.	Merkmalschlüssel	SIGNAL
	Definition	Bereich einer Signalsequenz (Präpeptid)
	Fakultative Qualifier	note

7.29.	Merkmalschlüssel	SITE
	Definition	Jede einzelne bedeutsame Aminosäurestelle in der Sequenz, die nicht durch einen anderen Merkmalschlüssel definiert ist. Der Schlüssel kann auch für eine Aminosäurebindung verwendet werden, die durch die Positionen der beiden benachbarten Aminosäuren dargestellt wird.
	Mandatory qualifier	note
	Anmerkung	Wenn "SITE" zur Annotierung einer modifizierten Aminosäure verwendet wird, muss der Wert für den "note"-Qualifier entweder eine in Abschnitt 4 dieses Anhangs aufgeführte Abkürzung oder der vollständige, ungekürzte Name der modifizierten Aminosäure sein.

7.30.	Merkmalschlüssel	source
	Definition	bezeichnet die Herkunft der Sequenz; die Angabe dieses Schlüssels ist obligatorisch; jede Sequenz weist ein einziges "source"-Merkmal auf, das die gesamte Sequenz umfasst
	Obligatorische Qualifier	mol_type organism
	Fakultative Qualifier	note

7.31.	Merkmalschlüssel	STRAND
	Definition	Sekundärstruktur: Beta-Strang, z. B. durch Wasserstoff-Brückenbindungen stabilisierter Beta-Strang oder Rest in einer isolierten Beta-Brücke

	Fakultative Qualifier	note
	Anmerkung	Dieses Merkmal wird nur für Proteine verwendet, deren tertiäre Struktur bekannt ist. Es werden nur drei Arten von Sekundärstrukturen angegeben: Helices (Schlüssel "HELIX"), Beta-Stränge (Schlüssel "STRAND") und Schleifen (Schlüssel "TURN"). Reste, die nicht in einer dieser Klassen angegeben sind, befinden sich in einer "loop"- oder "random-coil"-Struktur.

7.32.	Merkmalschlüssel	TOPO_DOM
	Definition	topologische Domäne
	Fakultative Qualifier	note

7.33.	Merkmalschlüssel	TRANSMEM
	Definition	Bereich einer Transmembran-Region
	Fakultative Qualifier	note

7.34.	Merkmalschlüssel	TRANSIT
	Definition	Bereich eines Transitpeptids (Mitrochondrion, Chloroplast, Thylakoid, Cyanelle, Peroxisom usw.)
	Fakultative Qualifier	note

7.35.	Merkmalschlüssel	TURN
	Definition	Sekundärstruktur-Schleife, z. B. durch Wasserstoff-Brückenbindung stabilisierte Schleife (3-, 4- oder 5-Schleife)
	Fakultative Qualifier	note
	Anmerkung	Dieses Merkmal wird nur für Proteine verwendet, deren tertiäre Struktur bekannt ist. Es werden nur drei Arten von Sekundärstrukturen angegeben: Helices (Schlüssel "HELIX"), Beta-Stränge (Schlüssel "STRAND") und Schleifen (Schlüssel "TURN"). Reste, die nicht in einer dieser Klassen angegeben sind, befinden sich in einer "loop"- oder "random-coil"-Struktur.

7.36.	Merkmalschlüssel	UNSURE
	Definition	Unsicherheiten in der Sequenz
	Fakultative Qualifier	note
	Anmerkung	Dieser Schlüssel wird verwendet, um Regionen einer Aminosäuresequenz zu beschreiben, bei der sich die Autoren bezüglich der Sequenzdarstellung unsicher sind.

7.37.	Merkmalschlüssel	VARIANT
	Definition	Den Angaben der Autoren zufolge gibt es Sequenzvarianten
	Fakultative Qualifier	note

7.38.	Merkmalschlüssel	VAR_SEQ
	Definition	Beschreibung von Sequenzvarianten, die durch alternatives Spleißen, alternative Promotoren, alternative Initiierung und ribosomale Leserasterverschiebung erzeugt werden
	Fakultative Qualifier	note

7.39.	Merkmalschlüssel	ZN_FING
	Definition	Bereich einer Zink-Finger-Region
	Obligatorische Qualifier	note
	Anmerkung	Die Art des Zink-Fingers wird im "note"-Qualifier angegeben. Zum Beispiel: "GATA-type" ("GATA-Typ") und "NR C4-type" ("NR C4-Typ").

ABSCHNITT 8: QUALIFIER FÜR AMINOSÄURESEQUENZEN

Dieser Abschnitt enthält die Liste der zulässigen Qualifier, die für Aminosäuresequenzen zu verwenden sind.

Wenn als Wertformat Freitext angegeben ist, der als sprachabhängig gekennzeichnet ist, muss eines der folgenden Elemente verwendet werden:

1) das Element `INSDQualifier_value`; oder
2) das Element `NonEnglishQualifier_value`; oder
3) sowohl das Element `INSDQualifier_value` als auch das Element `NonEnglishQualifier_value`.[8]

Wenn als Wertformat nicht als sprachabhängig gekennzeichneter Freitext angegeben ist, muss das Element "INSDQualifier_value" verwendet werden, und das Element "NonEnglishQualifier_value" darf nicht verwendet werden.

HINWEIS: Jeder Qualifier-Wert, der für einen Qualifier mit einem sprachabhängigen Freitextformat bereitgestellt wird, kann eine Übersetzung zu Zwecken internationaler, nationaler oder regionaler Verfahren erfordern.[9] Sprachabhängiger Freitext" ist für die Werte der in der folgenden Tabelle aufgeführten Qualifier vorgesehen:

Tabelle 6: Liste der Qualifier mit sprachabhängigen Freitextwerten für Aminosäuresequenzen

Abschnitt	Sprachabhängige Freitext Qualifier
8.2	note
8.3	organism

8.1.	Qualifier	mol_type
	Definition	In-vivo-Molekültyp der Sequenz
	Obligatorisches Wertformat	protein
	Beispiel	<INSDQualifier_value>protein</INSDQualifier_value>
	Anmerkung	Der Qualifier "mol_type" ist für den "source"-Merkmalschlüssel obligatorisch.

8.2.	Qualifier	note
	Definition	Anmerkungen oder Zusatzinformationen jeder Art
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.
	Beispiel	<INSDQualifier_value>Heme (covalent)</INSDQualifier_value>
	Anmerkung	Der Qualifier "note" ist für folgende Merkmalschlüssel obligatorisch: BINDING, CARBOHYD, CROSSLNK, DISULFID, DNA_BIND, DOMAIN, LIPID, METAL, MOD_RES, NP_BIND; SITE und ZN_FING

8.3.	Qualifier	organism
	Definition	wissenschaftlicher Name des Organismus, dem das Peptid entstammt
	Obligatorisches Wertformat	Freitext sprachabhängig: Dieser Wert muss möglicherweise zu Zwecken internationaler/nationaler/regionaler Verfahren übersetzt werden.

[8] Nach § 11a Absatz 3 der Patentverordnung ist sprachabhängiger Freitext in deutscher Sprache abzufassen. Er kann zusätzlich auch in englischer Sprache angegeben werden. Es ist danach für sprachabhängigen Freitext entweder nur das Element `NonEnglishQualifier_value` oder sowohl das Element `INSDQualifier_value` als auch das Element `NonEnglishQualifier_value` zu verwenden.

[9] Nach § 11a Absatz 3 der Patentverordnung ist sprachabhängiger Freitext in deutscher Sprache abzufassen. Er kann zusätzlich auch in englischer Sprache angegeben werden.

Beispiel	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
Anmerkung	Der Qualifier "organism" ist für den "source"-Merkmalschlüssel obligatorisch.

ABSCHNITT 9: TABELLEN FÜR DEN GENETISCHEN CODE

In Tabelle 7 sind die Tabellen für den genetischen Code wiedergegeben, die für die Translation codierender Sequenzen zu verwenden sind. Der Wert für den Qualifier "trans_table" ist die der jeweiligen Tabelle für den genetischen Code zugewiesene Nummer. Wenn ein "CDS"-Merkmal mit einem "translation"-Qualifier, nicht aber mit einem "transl_table"-Qualifier beschrieben wird, wird standardmäßig der Standardcode mit der Nummer 1 für die Übersetzung verwendet. (Hinweis: Die Nummern 7, 8, 15 und 17 bis 20 fehlen in Tabelle 7, da es keine Tabellen für den genetischen Code mit diesen Nummern mehr gibt.)

Tabelle 7: Tabellen für den genetischen Code

1 – Standardcode
AAs = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG Starts = ---M---------------M---------------M---------------------------- Base1 = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG Base2 = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG Base3 = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
2 - Mitochondrialer Code von Wirbeltieren
AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSS**VVVVAAAADDEEGGGG Starts = --------------------------------MMMM---------------M------------ Base1 = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG Base2 = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG Base3 = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
3 - Mitochondrialer Code von Hefen
AAs = FFLLSSSSYY**CCWWTTTTPPPPHHQQRRRRIIMMTTTTNNKKSSRRVVVVAAAADDEEGGGG Starts = ----------------------------------MM---------------M------------ Base1 = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG Base2 = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG Base3 = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
4 - Mitrochondrialer Code von Schimmelpilzen, Protozoen, Hohltieren und Code von Mykoplasmen/Spiroplasmen
AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG Starts = --MM---------------M------------MMMM---------------M------------ Base1 = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG Base2 = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG Base3 = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
5 - Mitochondrialer Code von Wirbellosen
AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSSSSVVVVAAAADDEEGGGG Starts = ---M----------------------------MMMM---------------M------------ Base1 = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG Base2 = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG Base3 = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
6 - Nukleärer Code von Ciliophora, Dasycladaceen und Hexamita
AAs = FFLLSSSSYYQQCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG Starts = ----------------------------------M---------------------------- Base1 = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG Base2 = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG Base3 = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG

9 - Mitochondrialer Code von Stachelhäutern und Plattwürmern

```
  AAs  =  FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts =  --------------------------------M---------------M------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

10 - Mitochondrialer Code von Euploten

```
  AAs  =  FFLLSSSSYY**CCCWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  --------------------------------M----------------------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

11 - Code von Bakterien, Archaeen und Pflanzenplastiden

```
  AAs  =  FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ---M---------------M------------MMMM---------------M------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

12 - Alternativer nukleärer Code von Hefen

```
  AAs  =  FFLLSSSSYY**CC*WLLLSPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  -------------------M---------------M----------------------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

13 - Mitochondrialer Code von Seescheiden

```
  AAs  =  FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSSGGVVVVAAAADDEEGGGG
Starts =  ---M------------------------------MM---------------M------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

14 - Alternativer mitochondrialer Code von Plattwürmern

```
  AAs  =  FFLLSSSSYYY*CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts =  --------------------------------M----------------------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

16 - Mitochondrialer Code von Chlorophyceae

```
  AAs  =  FFLLSSSSYY*LCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  --------------------------------M----------------------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

21 - Mitochondrialer Code von Saugwürmern

```
  AAs  =  FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts =  ----------------------------------M---------------M------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

22 - Mitochondrialer Code von Scenedesmus obliquus

```
  AAs  =  FFLLSS*SYY*LCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  --------------------------------M---------------------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

23 - Mitochondrialer Code von Thraustochytrium

```
  AAs  =  FF*LSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  --------------------------------M--M---------------M------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

24 - Mitochondrialer Code von Flügelkiemern

```
  AAs  =  FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSSKVVVVAAAADDEEGGGG
Starts =  ---M---------------M---------------M---------------M------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

25 - Code der Candidate Division SR1 und von Gracilibakterien

```
  AAs  = FFLLSSSSYY**CCGWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M------------------------------M---------------M------------
Base1  = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

26 - Nukleärer Code von Pachysolen tannophilus

```
  AAs  = FFLLSSSSYY**CC*WLLLAPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = -------------------M---------------M----------------------------
Base1  = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

27 - Nukleärer Code von Karyorelictea

```
  AAs  =  FFLLSSSSYYQQCCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  --------------*--------------------M----------------------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

28 - Nukleärer Code von Condylostoma

```
  AAs  =  FFLLSSSSYYQQCCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------**--*--------------------M----------------------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

29 - Nukleärer Code von Mesodinium

```
  AAs  =  FFLLSSSSYYYYCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------------------------------M----------------------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

30 - Nukleärer Code von Peritrichia

```
  AAs  =  FFLLSSSSYYEECC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------------------------------M----------------------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

31 - Nukleärer Code von Blastocrithidia

```
  AAs  =  FFLLSSSSYYEECCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------**----------------------M----------------------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

33 - Mitochondrialer UAA-Tyr-Code von Cephalodiscidae

```
  AAs  =  FFLLSSSSYYY*CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSSKVVVVAAAADDEEGGGG
Starts =  ---M---------------M---------------M---------------M------------
Base1  =  tttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

[Anhang II folgt]

ANHANG II

DOKUMENTTYPDEFINITION (DTD) FÜR SEQUENZPROTOKOLLE

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Annex II of WIPO Standard ST.26, Document Type Definition (DTD) for Sequence Listing

This entity may be identified by the PUBLIC identifier:
**************************************************************************************
****
PUBLIC "-//WIPO//DTD SEQUENCE LISTING 1.3//EN" "ST26SequenceListing_V1_3.dtd"
**************************************************************************************
****
* PUBLIC DTD URL

* https://www.wipo.int/standards/dtd/ST26SequenceListing_V1_3.dtd
**************************************************************************

* Revision of Annex II to WIPO Standard ST.26 was approved by the Committee on WIPO
* Standards (CWS) at its tenth session.

**************************************************************************
* CONTACTS
**************************************************************************
*
* xml.standards@wipo.int
*
*
**************************************************************************
* NOTES
**************************************************************************

* The sequence data part is a subset of the complete INSDC DTD V.1.5 that only covers
* the requirements of WIPO Standard ST.26.

**************************************************************************
* REVISION HISTORY
**************************************************************************
2022-11-25: Comment related to filename approved at CWS/10 (no update to version number)
2021-11-05: Revised Version 1.3 approved at CWS/9 (small edits to the comments)
2020-05-20: Version 1.3 approved at CWS/8.
Changes:
- Optional originalFreeTextLanguageCode attribute added to <ST26SequenceListing> to allow
applicants to indicate the language of the free text in the original sequence listing.
- Optional nonEnglishFreeTextLanguageCode attribute added to <ST26SequenceListing> to allow
applicants to indicate the language of the free text provided in the element
<NonEnglishQualifier_value>.
- Optional id attribute added to INSDQualifier to facilitate comparison of language-
dependent qualifier values between sequence listings.
- Optional element <NonEnglishQualifier_value> added to element <INSDQualifier> to allow
applicants to type language-dependent qualifiers in a non-English Language with the
characters set forth in paragraph 40(a) of the ST.26 main body document.


2018-10-19: Version 1.2 approved at CWS/6.
Changes:
```

```
<INSDQualifier*> changed to <INSDQualifier+> for alignment with business needs and advice
from NCBI (an INSDFeature_quals element (if present) should have one or more INSDQualifier
elements).

2017-06-02: Version 1.1 approved at the CWS/5
Changes:
Comments added to <INSDSeq_length>, <INSDSeq_division> and <INSDSeq_sequence> to clarify
the reason of the differences between the INSDC DTD v.1.5 and ST26 Sequence Listing DTD
V1_1.
2016-03-24: Version 1.0 adopted at the CWS/4Bis
2014-03-11: Final draft for adoption.
********************************************************************************

ST26SequenceListing
********************************************************************************
* ROOT ELEMENT
********************************************************************************
-->
<!ELEMENT ST26SequenceListing ((ApplicantFileReference | (ApplicationIdentification,
ApplicantFileReference?)), EarliestPriorityApplicationIdentification?, (ApplicantName,
ApplicantNameLatin?)?, (InventorName, InventorNameLatin?)?, InventionTitle+,
SequenceTotalQuantity, SequenceData+)>
<!--The elements ApplicantName and InventorName are optional in this DTD to facilitate
the conversion between various encoding schemes-->
<!--originalFreeTextLanguageCode:
The language code (see reference in paragraph 9 to ISO 639-1:2002) for the single original
language in which the language-dependent free text qualifiers (NonEnglishQualifier_value)
were prepared.
-->
<!--nonEnglishFreeTextLanguageCode:
The language code (see reference in paragraph 9 to ISO 639-1:2002) for the language in
which the language-dependent free text qualifiers (NonEnglishQualifier_value) currently
correspond.
-->
<!--fileName:
By default the file name will be set to the value provided for the project name in WIPO
Sequence. If the value is identical to the actual ST.26 XML filename, it should be noted
that Offices may enforce their requirements for the filename used which may restrict which
characters are allowable for submitted electronic files. It is also acceptable for the
value of the filename attribute and the actual file name to be different. Please refer to
the WIPO Sequence and ST.26 Knowledge Base for further details on Offices' naming
conventions for electronic files
--->
<!ATTLIST ST26SequenceListing
               dtdVersion CDATA #REQUIRED
               fileName CDATA #IMPLIED
               softwareName CDATA #IMPLIED
               softwareVersion CDATA #IMPLIED
               productionDate CDATA #IMPLIED
               originalFreeTextLanguageCode CDATA #IMPLIED
               nonEnglishFreeTextLanguageCode CDATA #IMPLIED
>

<!--ApplicantFileReference
Applicant's or agent's file reference, mandatory if application identification not
provided.
-->
<!ELEMENT ApplicantFileReference (#PCDATA)>
<!--ApplicationIdentification
Application identification for which the sequence listing is submitted, when available.
-->
```

```
<!ELEMENT ApplicationIdentification (IPOfficeCode, ApplicationNumberText, FilingDate?)>
<!--EarliestPriorityApplicationIdentification
Identification of the earliest priority application, which contains IPOfficeCode,
ApplicationNumberText and FilingDate elements.

-->
<!ELEMENT EarliestPriorityApplicationIdentification (IPOfficeCode, ApplicationNumberText,
FilingDate?)>
<!--ApplicantName
The name of the first mentioned applicant in characters set forth in paragraph 40(a) of the
ST.26 main body document.
-->
<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of
names of languages - Part 1: Alpha-2
-->
<!ELEMENT ApplicantName (#PCDATA)>
<!ATTLIST ApplicantName
               languageCode CDATA #REQUIRED
>
<!--ApplicantNameLatin
Where ApplicantName is typed in characters other than those as set forth in paragraph
40(b), a translation or transliteration of the name of the first mentioned applicant must
also be typed in characters as set forth in paragraph 40(b) of the ST.26 main body
document.
-->
<!ELEMENT ApplicantNameLatin (#PCDATA)>
<!--InventorName
Name of the first mentioned inventor typed in the characters as set forth in paragraph
40(a).-->
<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of
names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventorName (#PCDATA)>
<!ATTLIST InventorName
               languageCode CDATA #REQUIRED
>
<!--InventorNameLatin
Where InventorName is typed in characters other than those as set forth in paragraph 40(b),
a translation or transliteration of the first mentioned inventor may also be typed in
characters as set forth in paragraph 40(b).
-->
<!ELEMENT InventorNameLatin (#PCDATA)>
<!--InventionTitle
Title of the invention typed in the characters as set forth in paragraph 40(a) in the
language of filing. A translation of the title of the invention into additional languages
may be typed in the characters as set forth in paragraph 40(a) using additional
InventionTitle elements. The title of invention should be between two to seven words.
-->
<!--languageCode: Appropriate language code from ISO 639-1 - Codes
for the representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventionTitle (#PCDATA)>
<!ATTLIST InventionTitle
               languageCode CDATA #REQUIRED
>
<!--SequenceTotalQuantity
Indicates the total number of sequences in the document.
Its purpose is to be quickly accessible for automatic processing.
-->
<!ELEMENT SequenceTotalQuantity (#PCDATA)>
<!--SequenceData
```

```
Data for individual Sequence.
For intentionally skipped sequences see the ST.26 main body document.
-->
<!ELEMENT SequenceData (INSDSeq)>
<!ATTLIST SequenceData
               sequenceIDNumber CDATA #REQUIRED
>
<!--IPOfficeCode
ST.3 code. For example, if the application identification is PCT/IB2013/099999, then
IPOfficeCode value will be "IB" for the International Bureau of WIPO.
-->
<!ELEMENT IPOfficeCode (#PCDATA)>
<!--ApplicationNumberText
The application identification as provided by the office of filing (e.g. PCT/IB2013/099999)
-->
<!ELEMENT ApplicationNumberText (#PCDATA)>
<!--FilingDate
The date of filing of the patent application for which the sequence listing is submitted in
ST.2 format "CCYY-MM-DD", using a 4-digit calendar year, a 2-digit calendar month and a 2-
digit day within the calendar month, e.g., 2015-01-31. For details, please see paragraphs 7
(a) and 11 of WIPO Standard ST.2.
-->
<!ELEMENT FilingDate (#PCDATA)>
<!--*********************************************************************************
* INSD Part
*********************************************************************************

The purpose of the INSD part of this DTD is to define a customized DTD for sequence
listings to support the work of IP offices while facilitating the data exchange with the
public repositories.

The INSD part is subset of the INSD DTD v1.5 and as such can only be used to generate an
XML instance as it will not support the complete INSD structure.

This part is based on:

The International Nucleotide Sequence Database (INSD) collaboration.

INSDSeq provides the elements of a sequence as presented in the GenBank/EMBL/DDBJ-style
flatfile formats. Not all elements are used here.
-->
<!--INSDSeq
Sequence data. Changed INSD V1.5 DTD elements, INSDSeq_division and INSDSeq_sequence from
optional to mandatory per business requirements.
-->
<!ELEMENT INSDSeq (INSDSeq_length, INSDSeq_moltype, INSDSeq_division, INSDSeq_other-
seqids?, INSDSeq_feature-table?, INSDSeq_sequence)>
<!--INSDSeq_length
The length of the sequence. INSDSeq_length allows only integer.
-->
<!ELEMENT INSDSeq_length (#PCDATA)>
<!--INSDSeq_moltype
Admissible values: DNA, RNA, AA
-->
<!ELEMENT INSDSeq_moltype (#PCDATA)>
<!--INSDSeq_division
Indication that a sequence is related to a patent application. Must be populated with the
value PAT.
-->
<!ELEMENT INSDSeq_division (#PCDATA)>
<!--INSDSeq_other-seqids
```

```
In the context of data exchange with database providers, the IPOs should populate for each
sequence the element INSDSeq_other-seqids with one INSDSeqid containing a reference to the
corresponding published patent and the sequence identification.
-->
<!ELEMENT INSDSeq_other-seqids (INSDSeqid?)>
<!--INSDSeq_feature-table
Information on the location and roles of various regions within a particular sequence.
Whenever the element INSDSeq_feature-table is used, it must contain at least one feature.
-->
<!ELEMENT INSDSeq_feature-table (INSDFeature+)>
<!--INSDSeq_sequence
The residues of the sequence. The sequence must not contain numbers, punctuation or
whitespace characters.
-->
<!ELEMENT INSDSeq_sequence (#PCDATA)>
<!--INSDSeqid
Intended for the use of IPOs in data exchange only.

Format:
pat|{office code}|{publication number}|{document kind code}|{Sequence identification
number}

where office code is the code of the IP office publishing the patent document, publication
number is the publication number of the application or patent, document kind code is the
letter codes to distinguish patent documents as defined in ST.16 and Sequence
identification number is the number of the sequence in that application or patent

Example:
pat|WO|2013999999|A1|123456

This represents the 123456th sequence from WO patent publication No. 2013999999 (A1)
-->
<!ELEMENT INSDSeqid (#PCDATA)>
<!--INSDFeature
Description of one feature.
-->
<!ELEMENT INSDFeature (INSDFeature_key, INSDFeature_location, INSDFeature_quals?)>
<!--INSDFeature_key
A word or abbreviation indicating a feature.
-->
<!ELEMENT INSDFeature_key (#PCDATA)>
<!--INSDFeature_location
Region of the presented sequence which corresponds to the feature.
-->
<!ELEMENT INSDFeature_location (#PCDATA)>
<!--INSDFeature_quals
List of qualifiers containing auxiliary information about a feature.
-->
<!ELEMENT INSDFeature_quals (INSDQualifier+)>
<!--INSDQualifier
Additional information about a feature.
For coding sequences and variants see the ST.26 main body document.
-->
<!--id
Unique identifier for the INSDQualifier to facilitate comparison of versions of a sequence
listing specifically having language-dependent qualifier values in different languages.
-->
<!ELEMENT INSDQualifier (INSDQualifier_name, INSDQualifier_value?,
NonEnglishQualifier_value?)>
<!ATTLIST INSDQualifier
                    id ID #IMPLIED
```

```
>
<!--INSDQualifier_name
Name of the qualifier.
-->
<!ELEMENT INSDQualifier_name (#PCDATA)>
<!--INSDQualifier_value
Value of the qualifier. Where the qualifier is language-dependent its value must be in the
English language and typed with the characters set forth in paragraph 40 (b).
-->
<!ELEMENT INSDQualifier_value (#PCDATA)>
<!--NonEnglishQualifier_value
Value of a language-dependent qualifier in a language that is not English and typed with
the characters set forth in paragraph 40 (a). The language is indicated with the attribute
nonEnglishFreeTextLanguageCode.
-->
<!ELEMENT NonEnglishQualifier_value (#PCDATA)>
```

[Anhang III folgt]

ANHANG III

BEISPIELEXEMPLAR EINES SEQUENZPROTOKOLLS (XML-Datei)

Der Anhang III kann abgerufen werden unter: https://www.wipo.int/standards/en/xml_material/st26/st26-annex-iii-sequence-listing-specimen.xml

[Anhang IV folgt]

ANHANG IV

IN EINER XML-INSTANZ EINES SEQUENZPROTOKOLLS ZU VERWENDENDER AUSGEWÄHLTER ZEICHENSATZ AUS DEM UNICODEBLOCK BASIS-LATEINISCH

Das Et-Zeichen (0026) ist nur als Teil einer vordefinierten Entität zulässig. Das Anführungszeichen (0022), der Apostroph (0027), das Kleiner-als-Zeichen (003C) und das Größer-als-Zeichen (003E) müssen durch ihre vordefinierten Entitäten dargestellt werden. Außerdem muss das Et-Zeichen (0026) durch seine vordefinierte Entität dargestellt werden, wenn es als solches im Wert eines Attributs oder Inhalt eines Elements verwendet wird.

Unicode-Codepoint	Zeichen	Name (Beschreibung)
0020		SPACE (Leerzeichen)
0021	!	EXCLAMATION MARK (Ausrufezeichen)
0022	"	QUOTATION MARK (Anführungszeichen)
0023	#	NUMBER SIGN (Doppelkreuz)
0024	$	DOLLAR SIGN (Dollarzeichen)
0025	%	PERCENT SIGN (Prozentzeichen)
0026	&	AMPERSAND (Et-Zeichen)
0027	'	APOSTROPHE (Apostroph)
0028	(	LEFT PARENTHESIS (runde Klammer links)
0029	)	RIGHT PARENTHESIS (runde Klammer rechts)
002A	*	ASTERISK (Sternchen)
002B	+	PLUS SIGN (Pluszeichen)
002C	,	COMMA (Komma)
002D	-	HYPHEN-MINUS (Bindestrich-Minus)
002E	.	FULL STOP (Punkt)
002F	/	SOLIDUS (Schrägstrich)
0030	0	DIGIT ZERO (Ziffer Null)
0031	1	DIGIT ONE (Ziffer Eins)
0032	2	DIGIT TWO (Ziffer Zwei)
0033	3	DIGIT THREE (Ziffer Drei)
0034	4	DIGIT FOUR (Ziffer Vier)
0035	5	DIGIT FIVE (Ziffer Fünf)
0036	6	DIGIT SIX (Ziffer Sechs)
0037	7	DIGIT SEVEN (Ziffer Sieben)
0038	8	DIGIT EIGHT (Ziffer Acht)
0039	9	DIGIT NINE (Ziffer Neun)
003A	:	COLON (Doppelpunkt)
003B	;	SEMICOLON (Semikolon)
003C	<	LESS-THAN-SIGN (Kleiner-als-Zeichen)
003D	=	EQUALS SIGN (Gleichheitszeichen)
003E	>	GREATER-THAN-SIGN (Größer-als-Zeichen)
003F	?	QUESTION MARK (Fragezeichen)
0040	@	COMMERCIAL AT (At-Zeichen)
0041	A	LATIN CAPITAL LETTER A (Lateinischer Großbuchstabe A)
0042	B	LATIN CAPITAL LETTER B (Lateinischer Großbuchstabe B)
0043	C	LATIN CAPITAL LETTER C (Lateinischer Großbuchstabe C)
0044	D	LATIN CAPITAL LETTER D (Lateinischer Großbuchstabe D)
0045	E	LATIN CAPITAL LETTER E (Lateinischer Großbuchstabe E)
0046	F	LATIN CAPITAL LETTER F (Lateinischer Großbuchstabe F)
0047	G	LATIN CAPITAL LETTER G (Lateinischer Großbuchstabe G)
0048	H	LATIN CAPITAL LETTER H (Lateinischer Großbuchstabe H)

0049	I	LATIN CAPITAL LETTER I (Lateinischer Großbuchstabe I)
004A	J	LATIN CAPITAL LETTER J (Lateinischer Großbuchstabe J)
004B	K	LATIN CAPITAL LETTER K (Lateinischer Großbuchstabe K)
004C	L	LATIN CAPITAL LETTER L (Lateinischer Großbuchstabe L)
004D	M	LATIN CAPITAL LETTER M (Lateinischer Großbuchstabe M)
004E	N	LATIN CAPITAL LETTER N (Lateinischer Großbuchstabe N)
004F	O	LATIN CAPITAL LETTER O (Lateinischer Großbuchstabe O)
0050	P	LATIN CAPITAL LETTER P (Lateinischer Großbuchstabe P)
0051	Q	LATIN CAPITAL LETTER Q (Lateinischer Großbuchstabe Q)
0052	R	LATIN CAPITAL LETTER R (Lateinischer Großbuchstabe R)
0053	S	LATIN CAPITAL LETTER S (Lateinischer Großbuchstabe S)
0054	T	LATIN CAPITAL LETTER T (Lateinischer Großbuchstabe T)
0055	U	LATIN CAPITAL LETTER U (Lateinischer Großbuchstabe U)
0056	V	LATIN CAPITAL LETTER V (Lateinischer Großbuchstabe V)
0057	W	LATIN CAPITAL LETTER W (Lateinischer Großbuchstabe W)
0058	X	LATIN CAPITAL LETTER X (Lateinischer Großbuchstabe X)
0059	Y	LATIN CAPITAL LETTER Y (Lateinischer Großbuchstabe Y)
005A	Z	LATIN CAPITAL LETTER Z (Lateinischer Großbuchstabe Z)
005B	[	LEFT SQUARE BRACKET (Eckige Klammer links)
005C	\	REVERSE SOLIDUS (Umgekehrter Schrägstrich)
005D	]	RIGHT SQUARE BRACKET (Eckige Klammer rechts)
005E	^	CIRCUMFLEX ACCENT (Zirkumflex)
005F	_	LOW LINE (Unterstrich)
0060	`	GRAVE ACCENT (Gravis)
0061	a	LATIN SMALL LETTER A (Lateinischer Kleinbuchstabe a)
0062	b	LATIN SMALL LETTER B (Lateinischer Kleinbuchstabe b)
0063	c	LATIN SMALL LETTER C (Lateinischer Kleinbuchstabe c)
0064	d	LATIN SMALL LETTER D (Lateinischer Kleinbuchstabe d)
0065	e	LATIN SMALL LETTER E (Lateinischer Kleinbuchstabe e)
0066	f	LATIN SMALL LETTER F (Lateinischer Kleinbuchstabe f)
0067	g	LATIN SMALL LETTER G (Lateinischer Kleinbuchstabe g)
0068	h	LATIN SMALL LETTER H (Lateinischer Kleinbuchstabe h)
0069	i	LATIN SMALL LETTER I (Lateinischer Kleinbuchstabe i)
006A	j	LATIN SMALL LETTER J (Lateinischer Kleinbuchstabe j)
006B	k	LATIN SMALL LETTER K (Lateinischer Kleinbuchstabe k)
006C	l	LATIN SMALL LETTER L (Lateinischer Kleinbuchstabe l)
006D	m	LATIN SMALL LETTER M (Lateinischer Kleinbuchstabe m)
006E	n	LATIN SMALL LETTER N (Lateinischer Kleinbuchstabe n)
006F	o	LATIN SMALL LETTER O (Lateinischer Kleinbuchstabe o)
0070	p	LATIN SMALL LETTER P (Lateinischer Kleinbuchstabe p)
0071	q	LATIN SMALL LETTER Q (Lateinischer Kleinbuchstabe q)
0072	r	LATIN SMALL LETTER R (Lateinischer Kleinbuchstabe r)
0073	s	LATIN SMALL LETTER S (Lateinischer Kleinbuchstabe s)
0074	t	LATIN SMALL LETTER T (Lateinischer Kleinbuchstabe t)
0075	u	LATIN SMALL LETTER U (Lateinischer Kleinbuchstabe u)
0076	v	LATIN SMALL LETTER V (Lateinischer Kleinbuchstabe v)
0077	w	LATIN SMALL LETTER W (Lateinischer Kleinbuchstabe w)
0078	x	LATIN SMALL LETTER X (Lateinischer Kleinbuchstabe x)
0079	y	LATIN SMALL LETTER Y (Lateinischer Kleinbuchstabe y)
007A	z	LATIN SMALL LETTER Z (Lateinischer Kleinbuchstabe z)
007B	{	LEFT CURLY BRACKET (Geschweifte Klammer links)
007C	|	VERTICAL LINE (Senkrechter Strich)
007D	}	RIGHT CURLY BRACKET (Geschweifte Klammer rechts)
007E	~	TILDE (Tilde)

[Anhang V folgt]

ANHANG V

ZUSÄTZLICHE ANFORDERUNGEN FÜR DEN DATENAUSTAUSCH (NUR FÜR ÄMTER FÜR GEISTIGES EIGENTUM)

Im Rahmen des Datenaustauschs mit Datenbankanbietern (INSD-Mitgliedern) sollten die Ämter für geistiges Eigentum für jede Sequenz das Element `INSDSeq_other-seqids` mit einer `INSDSeqid` befüllen, die auf das entsprechende veröffentlichte Patent und die Sequenzkennzahl verweist, und dabei folgendes Format verwenden:

pat|{Code des Amts}|{Publikationsnummer}|{Schriftarten-Code}|{Sequenzkennzahl}.

Dabei ist der Code des Amts der gemäß ST.3 angegebene Code des Amts für geistiges Eigentum, das das Patentdokument veröffentlicht hat; der Schriftarten-Code ist der Code für die Identifizierung verschiedener Arten von Patentdokumenten nach ST.16; die Publikationsnummer ist die Publikationsnummer der Anmeldung oder des Patents und die Sequenzkennzahl ist die Kennzahl der Sequenz in besagter Anmeldung oder besagtem Patent.

Beispiel:

pat|WO|2013999999|A1|123456

Übertragung in eine valide XML-Instanz:

```
<INSDSeq_other-seqids>
  <INSDSeqid>pat|WO|2013999999|A1|123456</INSDSeqid>
</INSDSeq_other-seqids>
```

Dabei steht "123456" für die 123456teSequenz aus der WO-Publikation mit der Veröffentlichungsnummer 2013999999 (A1).

[Anhang VI folgt]

ANHANG VI

LEITFADEN MIT ILLUSTRIERTEN BEISPIELEN

INHALTSVERZEICHNIS

EINFÜHRUNG……1
INDEX DER BEISPIELE……7
BEISPIELE……10
ANLAGE……82

EINFÜHRUNG

Dieser Standard soll u. a. "Anmeldern ermöglichen, für eine Patentanmeldung ein einziges Sequenzprotokoll zu erstellen, das sowohl für internationale als auch für nationale oder regionale Verfahren geeignet ist". Zwecks Verwirklichung dieses Ziels soll mit diesem Leitfaden bei allen Anmeldern und Ämtern für geistiges Eigentum (IPOs) Klarheit und Einvernehmen über die Anforderungen geschaffen werden, die für die Aufnahme und Darstellung offenbarter Sequenzen gelten.

Bestandteile dieses Leitfadens sind die vorliegende Einführung, ein Index der Beispiele, Beispiele für offenbarte Sequenzen und eine Anlage mit einem Sequenzprotokoll im XML-Format, das die Sequenzen aus den Beispielen enthält. In dieser Einführung werden bestimmte Konzepte und die Terminologie erläutert, die in diesem Leitfaden im Weiteren verwendet werden. Die Beispiele veranschaulichen die Anforderungen, die in bestimmten Absätzen des Standards niedergelegt sind. Zu jedem Beispiel wird die Nummer des jeweils wichtigsten Absatzes angeführt. Einige Beispiele dienen auch der Illustration anderer Absätze, wobei die entsprechenden Querverweise am Ende der jeweiligen Beispiele aufgeführt sind. Dem Index sind die Seitenzahlen für die Beispiele und alle angegebenen Querverweise zu entnehmen. Jede Sequenz in einem Beispiel, die in ein Sequenzprotokoll aufgenommen werden muss oder kann, wurde mit einer Sequenzkennzahl (SEQ ID NO) versehen und wird in der Anlage zu diesem Dokument im XML-Format angezeigt.

Für alle Beispiele gilt, dass erläuternde Informationen zu einer Sequenz jeweils als Gesamtheit der Offenbarung in Bezug auf diese Sequenz zu verstehen sind. Bei den Antworten werden nur die im jeweiligen Beispiel explizit beschriebenen Informationen berücksichtigt.

Der vorliegende Leitfaden dient als Hilfestellung für die Erstellung eines Sequenzprotokolls, das am Anmeldetag der Patentanmeldung eingereicht wird. Bei Sequenzprotokollen, die nach dem Anmeldetag der Patentanmeldung eingereicht werden sollen, muss berücksichtigt werden, ob die bereitgestellten Informationen vom Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten. Daher ist es möglich, dass diese Leitlinien auf ein Sequenzprotokoll, das nach dem Anmeldetag der Patentanmeldung eingereicht wurde, nicht anwendbar sind.

Erstellung eines Sequenzprotokolls

Bei Erstellung eines Sequenzprotokolls für eine Patentanmeldung sind folgende Fragen zu berücksichtigen:

1. Ist die Aufnahme einer bestimmten offenbarten Sequenz nach ST.26 Absatz 7 erforderlich?

2. Wenn die Aufnahme einer bestimmten offenbarten Sequenz nach ST.26 nicht erforderlich ist, wäre die Aufnahme nach ST.26 dennoch zulässig?

3. Wenn die Aufnahme einer bestimmten offenbarten Sequenz nach ST.26 erforderlich oder zulässig ist, wie sollte diese Sequenz dann im Sequenzprotokoll dargestellt werden?

In Bezug auf die erste Frage verlangt ST.26 Absatz 7 (mit bestimmten Einschränkungen) die Aufnahme einer in einer Patentanmeldung durch Aufzählung der Reste offenbarten Sequenz, wenn die Sequenz zehn oder mehr spezifisch definierte Nukleotide oder vier oder mehr spezifisch definierte Aminosäuren enthält.

In Bezug auf die zweite Frage verbietet ST.26 Absatz 8 die Aufnahme von Sequenzen mit weniger als zehn spezifisch definierten Nukleotiden oder weniger als vier spezifisch definierten Aminosäuren.

Um die beiden Fragen zu beantworten, muss klar sein, was genau unter "Aufzählung der Reste" und "spezifisch definiert" zu verstehen ist.

In Bezug auf die dritte Frage veranschaulichen die in diesem Dokument dargestellten Sequenzoffenbarungen eine Vielzahl von Szenarien. Ergänzt werden diese durch eine umfassende Erörterung der bevorzugten Darstellungsweise für jede einzelne Sequenz oder, wenn eine Sequenz mehrere Variationen enthält, für die "umfassendste Sequenz" im Sinne dieses Standards. Da es nicht möglich ist, jedes erdenkliche Sequenzszenario abzuhandeln, wird in diesem Leitfaden angestrebt, die Begründung für den Umgang mit jedem Beispiel und die Auslegung der ST.26-Bestimmungen so darzulegen, dass sie auf andere, nicht in den Beispielen enthaltene Sequenzszenarien übertragen werden können.

Aufzählung der Reste

In ST.26 Absatz 3(c) ist die "Aufzählung der Reste" definiert als die Offenbarung einer Sequenz in einer Patentanmeldung, indem jeder Rest der Sequenz der Reihe nach aufgeführt wird, wobei (i) der Rest durch einen Namen, eine Abkürzung, ein Symbol oder eine Struktur dargestellt wird oder (ii) Mehrfachreste durch eine Kurzformel dargestellt werden. Eine Sequenz sollte in einer Patentanmeldung durch "Aufzählung der Reste" unter Verwendung der konventionellen Symbole offenbart werden, d. h. der Nukleotidsymbole gemäß ST.26 Anhang I, Abschnitt 1, Tabelle 1 (d. h. der Symbole in Kleinbuchstaben oder ihren gleichwertigen Entsprechungen in Großbuchstaben[1]) und der Aminosäuresymbole gemäß ST.26. Anhang I, Abschnitt 3, Tabelle 3 (d. h. Symbole in Großbuchstaben oder ihren gleichwertigen Entsprechungen in Kleinbuchstaben[10]) .Im Folgenden werden diese Nukleotid- und Aminosäuresymbole als konventionelle Symbole bezeichnet. Darstellungen von Nukleotiden und Aminosäuren, die von den Darstellungen in diesen Tabellen abweichen, werden im Folgenden als "nicht konventionell" bezeichnet.

Wird eine Darstellung eines Rests als gleichwertig zu einem konventionellen Symbol oder einer konventionellen Abkürzung (z. B. "Z_1" bedeutet "A") oder zu einer spezifischen Sequenz konventioneller Symbole (z. B. "Z_1" bedeutet "agga") offenbart, so wird die Sequenz so interpretiert, als ob sie mit den gleichwertigen konventionellen Symbolen oder Abkürzungen offenbart worden wäre. Auf dieser Grundlage wird festgestellt, ob die Aufnahme in das Sequenzprotokoll nach ST.26 Absatz 7 erforderlich oder nach Absatz 8 verboten ist. Wenn ein nicht konventionelles Nukleotidsymbol als Mehrdeutigkeitssymbol verwendet wird (z. B. "X_1" = Inosin oder Pseudouridin), aber keinem der konventionellen Mehrdeutigkeitssymbole in Abschnitt 1, Tabelle 1 ("m", "r", "w", "s", "y", "k", "v", "h", "d", "b" oder "n") entspricht, dann wird der Rest als "n"-Rest interpretiert. Auf dieser Grundlage wird festgestellt, ob die Aufnahme in das Sequenzprotokoll nach ST.26 Absatz 7 erforderlich oder nach Absatz 8 verboten ist. Wenn ein nicht konventionelles Aminosäuresymbol als Mehrdeutigkeitssymbol verwendet wird (z. B. "Z_1" bedeutet "A", "G", "S" oder "T"), aber keinem der konventionellen Mehrdeutigkeitssymbole in Abschnitt 3, Tabelle 3 ("B", "Z", "J" oder "X") entspricht, dann wird der Rest als "X"-Rest interpretiert. Auf dieser Grundlage wird festgestellt, ob die Aufnahme der Sequenz in das Sequenzprotokoll nach ST.26 Absatz 7 erforderlich oder nach Absatz 8 verboten ist.

Es sollte darauf geachtet werden, Sequenzen unter Verwendung konventioneller Symbole zu offenbaren. Wenn Sequenzen jedoch in anderer Weise offenbart werden, kann es notwendig sein, die Offenbarung heranzuziehen, um die Bedeutung der nicht konventionellen Darstellung zu bestimmen.

Selbst wenn ein konventionelles Symbol verwendet wird, muss die Erläuterung zur Sequenz in der Offenbarung herangezogen werden, um zu bestätigen, dass das Symbol in konventioneller Weise verwendet wird. Bei einer nicht konventionellen Verwendung des Symbols ist diese Erläuterung heranzuziehen, um festzustellen, ob die Aufnahme der Sequenz in das Sequenzprotokoll nach ST.26 Absatz 7 erforderlich oder nach Absatz 8 verboten ist.

Spezifisch definiert

Nach ST.26 Absatz 3(k) bezeichnet der Ausdruck "spezifisch definiert" jedes in Anhang I aufgeführte Nukleotid, das nicht durch das Symbol "n" dargestellt wird, und jede in Anhang I aufgeführte Aminosäure, die nicht durch das Symbol "X" dargestellt wird, wobei "n" und "X" in konventioneller Weise verwendet werden, wie in Abschnitt 1, Tabelle 1 ("a oder c oder g oder t/u; 'unbekanntes' ('unknown') oder 'sonstiges' ('other')") bzw. Abschnitt 3, Tabelle 3 ("A oder R oder N oder D oder C oder Q oder E oder G oder H oder I oder L oder K oder M oder F oder P oder O oder S oder U oder T oder W oder Y oder V; 'unbekannte' ('unknown') oder 'sonstige' ('other')") beschrieben. Auf Grundlage der obigen Ausführungen über konventionelle Symbole oder nicht konventionelle Symbole oder Abkürzungen und deren Verwendung in konventioneller oder nicht konventioneller Weise wird festgestellt, ob ein Nukleotid oder eine Aminosäure "spezifisch definiert" ist.

10 HINWEIS: Während in einer Offenbarung im Zuge einer Anmeldung Nukleotide oder Aminosäuren mit Symbolen aus Klein- oder Großbuchstaben dargestellt werden dürfen, dürfen für eine Sequenz, die in einem Sequenzprotokoll enthalten ist, zur Darstellung einer Nukleotidsequenz nur Kleinbuchstaben (siehe ST.26, Absatz 13) und zur Darstellung einer Aminosäuresequenz nur Großbuchstaben verwendet werden (siehe ST.26 Absatz 26).

Umfassendste Sequenz

Wird eine Sequenz, die die in Absatz 7 genannten Anforderungen erfüllt, in einer Anmeldung nur einmal durch Aufzählung der Reste offenbart, aber in mehrfachen Ausführungsformen unterschiedlich beschrieben (z. B., indem "X" in einer Ausführungsform an einer oder mehreren Stellen jede beliebige Aminosäure, in weiteren Ausführungsformen jedoch nur eine begrenzte Anzahl von Aminosäuren darstellt), muss nach ST.26 nur die Sequenz in das Sequenzprotokoll aufgenommen werden, die durch Aufzählung der Reste offenbart wird. Wenn eine solche Sequenz mehrere Mehrdeutigkeitssymbole "n" oder "X" enthält, werden diese gemäß den Absätzen 15 und 27 als jedes beliebige Nukleotid bzw. jede beliebige Aminosäure ausgelegt, sofern keine weitere Annotation erfolgt. Als einzelne Sequenz muss folglich die umfassendste offenbarte Sequenz aufgenommen werden. Die umfassendste Sequenz ist diejenige einzelne Sequenz, deren Restevarianten durch die restriktivsten Mehrdeutigkeitssymbole dargestellt werden, die die meisten offenbarten Ausführungsformen enthalten. Wenn eine Sequenz nur einmal durch Aufzählung der Reste offenbart wird, ihre Länge jedoch aufgrund einer unterschiedlichen Anzahl von Kopien variieren kann, gilt entsprechend die längste Ausführungsform als die umfassendste Sequenz. Beispielsweise gilt für eine Sequenz mit einer sich wiederholenden Region, die von zwei bis fünf Kopien variieren kann, die Ausführungsform mit fünf Kopien der sich wiederholenden Region als die umfassendste Sequenz und sollte in das Sequenzprotokoll aufgenommen werden. Sofern praktikabel, wird allerdings die Aufnahme zusätzlicher spezifischer Sequenzen nachdrücklich empfohlen, z. B. solche zusätzlichen Ausführungsformen, die einen wesentlichen Teil der Erfindung darstellen. Die Aufnahme der zusätzlichen Sequenzen ermöglicht eine gründlichere Recherche und informiert die Öffentlichkeit über den Gegenstand, für den ein Patent erstrebt wird.

Verwendung der Mehrdeutigkeitssymbole

Korrekte Verwendung des Mehrdeutigkeitssymbols "n" in einem Sequenzprotokoll

Das Symbol "n"

a. darf für nichts anderes als ein einzelnes Nukleotid verwendet werden;

b. wird als "a", "c", "g" oder "t/u" ausgelegt, wenn es nicht in Verbindung mit einer näheren Beschreibung verwendet wird;

c. sollte verwendet werden, um eines der folgenden Nukleotide zusammen mit einer näheren Beschreibung darzustellen:

    i. modifiziertes Nukleotid, z. B. natürliches, synthetisches oder nicht natürlich vorkommendes Nukleotid, das nicht durch ein anderes Symbol in Anhang I (siehe Abschnitt 1, Tabelle 1) dargestellt werden kann;

    ii. "unbekanntes" ("unknown") Nukleotid, d. h. nicht bestimmt, nicht offenbart oder unsicher;

    iii. eine abasische (basenfreie) Stelle; oder

d. darf verwendet werden, um eine Sequenzvariante darzustellen, d. h. Alternativen, Deletionen, Insertionen oder Substitutionen, wenn "n" das restriktivste Mehrdeutigkeitssymbol ist.

Korrekte Verwendung des Mehrdeutigkeitssymbols "X" in einem Sequenzprotokoll

Das Symbol "X"

a. darf für nichts anderes als eine einzelne Aminosäure verwendet werden;

b. wird als "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, wenn es nicht in Verbindung mit einer näheren Beschreibung verwendet wird;

c. sollte verwendet werden, um eine der folgenden Aminosäuren zusammen mit einer näheren Beschreibung darzustellen:

    i. modifizierte Aminosäure, z. B. natürliche, synthetische oder nicht natürlich vorkommende Aminosäure, die nicht durch ein anderes Symbol in Anhang I (siehe Abschnitt 3, Tabelle 3) dargestellt werden kann;

ii. "unbekannte" ("unknown") Aminosäure, d. h. nicht bestimmt, nicht offenbart oder unsicher; oder

d. darf verwendet werden, um eine Sequenzvariante darzustellen, d. h. Alternativen, Deletionen, Insertionen oder Substitutionen, wobei "X" das restriktivste Mehrdeutigkeitssymbol ist.

Annotation modifizierter Reste

Dieser Standard verlangt, dass "modifizierte" Reste für Nukleotide nach Absatz 17 und für Aminosäuren nach Absatz 30 annotiert werden.

Nach ST.26 Absatz 3(e) bezeichnet der Ausdruck "modifizierte Aminosäure" jede beliebige in Absatz 3(a) beschriebene Aminosäure mit Ausnahme von L-Alanin, L-Arginin, L-Asparagin, L-Asparaginsäure, L-Cystein, L-Glutamin, L-Glutaminsäure, L-Glycin, L-Histidin, L-Isoleucin, L-Leucin, L-Lysin, L-Methionin, L-Phenylalanin, L-Prolin, L-Pyrrolysin, L-Serin, L-Selenocystein, L-Threonin, L-Tryptophan, L-Tyrosin und L-Valin. Entsprechend bezeichnet der Ausdruck "modifiziertes Nukleotid" gemäß dem Standard jedes beliebige in Absatz 3(g) beschriebene Nukleotid mit Ausnahme von Desoxyadenosin-5'-Monophosphat, Desoxyguanosin-5'-Monophosphat, Desoxycytidin-5'-Monophosphat, Desoxythymidin-5'-Monophosphat, Adenosin-5'-Monophosphat, Guanosin-5'-Monophosphat, Cytidin-5'-Monophosphat und Uridin-5'-Monophosphat (ST.26 Absatz 3(f)).

Aus den vorgenannten Definitionen ergibt sich, dass Modifikationen an den Nukleinbasen oder dem Zucker-Phosphat-Rückgrat einer Nukleinsäure sowie Modifikationen der Aminosäure-Seitenketten von Aminosäuren oder des Peptid-Rückgrats eines Peptids zur Entstehung von einem oder mehreren "modifizierten Nukleotiden" bzw. einer oder mehreren "modifizierten Aminosäuren" führen. Daher müssen solche Nukleotide und Aminosäuren annotiert werden. Beispiele für Rückgrat-Modifikationen sind Nukleotidanaloga wie Peptidnukleinsäuren (PNA) und Glykolnukleinsäuren (GNA) sowie D-Aminosäuren.

Es ist zu beachten, dass die Modifikation einer terminalen Aminosäure eines Peptids oder eines terminalen Nukleotids einer Nukleinsäure nicht notwendigerweise zur Entstehung einer "modifizierten Aminosäure" oder eines "modifizierten Nukleotids" führt. In diesem Fall ist durch eine genauere Betrachtung der terminalen Modifikation festzustellen, ob die Modifikation die chemische Struktur des Rests so verändert, dass der Rest nicht mehr unter die in den Absätzen 3(e) und 3(f) beschriebenen Ausnahmen fällt. So gilt beispielsweise ein Peptid, dessen C-terminales Ende über eine normale Amidbindung an eine Struktur gebunden ist (z. B. als Teil einer verzweigten Sequenz, wie Peptid 2 in Beispiel 7(b) -3), nicht als "modifizierter Rest" und muss daher nicht annotiert werden. Entsprechend gilt ein Peptid, dessen N-terminaler Rest in einer Amidbindung an Biotin gebunden ist, ebenfalls nicht als "modifizierter Rest" und muss daher nicht annotiert werden. In beiden Szenarien ist die Struktur des an der C-terminalen oder N-terminalen Bindung beteiligten Restes gegenüber den konventionellen Aminosäuren, die in Absatz 3(e) des Standards aufgeführt werden, nicht verändert.

Im Gegensatz dazu gelten terminale Modifikationen, die die chemische Struktur des Restes verändern, als "modifizierte Reste" und müssen annotiert werden. Die Methylierung des C-Terminus in Beispiel 3(c)-1 etwa ändert die chemische Struktur des terminalen Rests durchaus, da die Methylgruppe die Hydroxylgruppe ersetzt, die üblicherweise an der Alpha-Carboxylgruppe zu finden ist. Daher muss dieses methylierte Lysin als "modifizierter Rest" annotiert werden.

Es obliegt dem Anmelder, jede Modifikation eines terminalen Rests innerhalb einer durch Aufzählung dargestellten Sequenz daraufhin zu prüfen, ob die Struktur des terminalen Rests verändert ist oder nicht. Wenn sich die Struktur des modifizierten Rests von den in den Absätzen 3(e) und 3(f) beschriebenen konventionellen Aminosäuren bzw. Nukleotiden unterscheidet, muss die Modifikation annotiert werden.

Des Weiteren wird empfohlen, dass Anmelder ihre Sequenzprotokolle grundsätzlich mit so vielen Informationen versehen, wie es zu einer möglichst genauen Darstellung ihrer Offenbarungen geboten ist. Daher sollte eine Modifikation vorzugsweise auch dann aufgenommen werden, wenn sie nicht annotiert werden muss.

Allerdings muss eine Annotation von Varianten einer durch Aufzählung dargestellten Primärsequenz den Anforderungen von ST.26 Absätze 93–100 entsprechen. Modifikationen, die als Varianten einer durch Aufzählung dargestellten Sequenz offenbart werden, müssen womöglich nicht in das Sequenzprotokoll aufgenommen werden. Die Annotation von Varianten ist in ST.26 Absätze 93–95 definiert.

Darstellung modifizierter Reste

ST.26 gibt vor, dass modifizierte Nukleotide und Aminosäuren im Sequenzprotokoll nach Möglichkeit als die entsprechenden nicht modifizierten Reste dargestellt werden sollten (siehe Absätze 16 und 29). Die Formulierung "sollten" beschreibt eine

nachdrücklich empfohlene, jedoch nicht obligatorische Vorgehensweise (siehe Absatz 4(d)). Es liegt im Ermessen des Anmelders, ob ein modifizierter Rest durch den entsprechenden nicht modifizierten Rest oder die Variable "n" oder "X" dargestellt wird.

Als allgemeine Faustregel gilt: Wenn ein Rest durch Addition einer funktionellen Gruppe, z. B. durch Methylierung oder Acetylierung, modifiziert wird und die Struktur des nicht modifizierten Rests im Wesentlichen unverändert bleibt, wird die Darstellung durch den nicht modifizierten Rest empfohlen. Zum Beispiel sollte ein methyliertes Adenosin im Sequenzprotokoll vorzugsweise durch "a" dargestellt werden. Wenn sich der modifizierte Rest jedoch in seiner Struktur von allen nicht modifizierten Resten unterscheidet, wird die Darstellung durch ein "n" oder "X" empfohlen. Norleucin beispielsweise ist ein Isomer von Leucin, und seine Seitenkette ist eine lineare Struktur mit vier Kohlenstoffatomen. Leucin hat ebenfalls eine Seitenkette aus vier Kohlenstoffatomen, die jedoch am zweiten Kohlenstoff verzweigt ist. Daher handelt es sich bei Norleucin nicht einfach um das Ergebnis einer Modifikation, die einem Leucin hinzugefügt wurde, sondern um eine völlig andere (wenn auch verwandte) Struktur. Daher wird empfohlen, Norleucin in einem Sequenzprotokoll durch ein "X" darzustellen.

Ein Nukleotid ist "spezifisch definiert", wenn es durch ein anderes Symbol als "n" dargestellt wird, und eine Aminosäure ist "spezifisch definiert", wenn sie durch ein anderes Symbol als "X" dargestellt wird (siehe ST.26 Absatz 3(k)). Daher ist ein 2'-O-Methyladenosin, das durch ein "a" in der Sequenz dargestellt wird, spezifisch definiert, während Norleucin, das durch ein "X" in der Sequenz dargestellt wird, nicht spezifisch definiert ist.

Tabelle A – Konventionelle Nukleotidsymbole und Definitionen

Symbol	Definition
a	Adenin
c	Cytosin
g	Guanin
t	Thymin in DNA Uracil in RNA (t/u)
m	a oder c
r	a oder g
w	a oder t/u
s	c oder g
y	c oder t/u
k	g oder t/u
v	a oder c oder g; nicht t/u
h	a oder c oder t/u; nicht g
d	a oder g oder t/u; nicht c
b	c oder g oder t/u; nicht a
n	a oder c oder g oder t/u; "unbekanntes" ("unknown") oder "sonstiges" ("other")

Tabelle B – Konventionelle Aminosäuresymbole, Dreibuchstabencode und Definitionen

Symbol	Dreibuchstabencode	Definition
A	Ala	Alanin
R	Arg	Arginin
N	Asn	Asparagin
D	Asp	Asparaginsäure (Aspartat)
C	Cys	Cystein
Q	Gln	Glutamin
E	Glu	Glutaminsäure (Glutamat)
G	Gly	Glycin
H	His	Histidin
I	Ile	Isoleucin
L	Leu	Leucin
K	Lys	Lysin
M	Met	Methionin
F	Phe	Phenylalanin
P	Pro	Prolin
O	Pyl	Pyrrolysin
S	Ser	Serin
U	Sec	Selenocystein
T	Thr	Threonin
W	Trp	Tryptophan
Y	Tyr	Tyrosin
V	Val	Valin
B	Asx	Asparaginsäure oder Asparagin
Z	Glx	Glutamin oder Glutaminsäure
J	Xle	Leucin oder Isoleucin
X	Xaa	A oder R oder N oder D oder C oder Q oder E oder G oder H oder I oder L oder K oder M oder F oder P oder O oder S oder U oder T oder W oder Y oder V, "unbekannte" ("unknown") oder "sonstige" ("other")

INDEX DER BEISPIELE

Absatz 3(a) – Definition von "Aminosäure" 10

Beispiel 3(a)-1: D-Aminosäuren 10

Absatz 3(c) – Definition von "Aufzählung der Reste" 11

Beispiel 3(c)-1: Aufzählung von Aminosäuren nach chemischer Struktur 11

Beispiel 3(c)-2: Kurzformel für eine Aminosäuresequenz 13

Absatz 3(g) – Definition von "Nukleotid" 14

Beispiel 3(g)-1: Durch einen C3-Spacer unterbrochene Nukleotidsequenz 14

Beispiel 3(g)-2: Nukleotidsequenz mit alternativen Resten, einschließlich eines C3-Spacers 15

Beispiel 3(g)-3: Abasische Stelle 17

Beispiel 3(g)-4: Nukleinsäure-Analoga 18

Absatz 3(k) – Definition von "spezifisch definiert" 19

Beispiel 3(k)-1: Mehrdeutigkeitssymbole für Nukleotide 19

Beispiel 3(k)-2: Konventionelle und nicht konventionelle Verwendung des Mehrdeutigkeitssymbols "n" 20

Beispiel 3(k)-3: Nicht konventionelle Verwendung des Mehrdeutigkeitssymbols "n" 22

Beispiel 3(k)-4: Mehrdeutigkeitssymbole, die nicht "n" sind, sind "spezifisch definiert" 23

Beispiel 3(k)-5: Nicht konventionelle Verwendung der Mehrdeutigkeitsabkürzung "Xaa" 24

Absatz 7(a) – Nukleotidsequenzen, die in ein Sequenzprotokoll aufgenommen werden müssen 25

Beispiel 7(a)-1: Verzweigte Nukleotidsequenz 25

Beispiel 7(a)-2: Lineare Nukleotidsequenz mit Sekundärstruktur 27

Beispiel 7(a)-3: Nicht konventionelle Verwendung von Mehrdeutigkeitssymbolen für Nukleotide 28

Beispiel 7(a)-4: Nicht konventionelle Verwendung von Mehrdeutigkeitssymbolen für Nukleotide 29

Beispiel 7(a)-5: Nicht konventionelle Nukleotidsymbole 30

Beispiel 7(a)-6: Nicht konventionelle Nukleotidsymbole 31

Beispiel 7(a)-7: Invertierte Nukleotide I 32

Beispiel 7(a)-8: Invertierte Nukleotide II 34

Absatz 7(b) – Aminosäuresequenzen, die in ein Sequenzprotokoll aufgenommen werden müssen 36

Beispiel 7(b)-1: Vier oder mehr spezifisch definierte Aminosäuren 36

Beispiel 7(b)-2: Verzweigte Aminosäuresequenz 37

Beispiel 7(b)-3: Verzweigte Aminosäuresequenz 40

Beispiel 7(b)-4: Zyklisches Peptid, das eine verzweigte Aminosäuresequenz enthält 41

Beispiel 7(b)-5: Zyklisches Peptid, das eine verzweigte Aminosäuresequenz enthält 44

Absatz 11(a) – Doppelsträngige Nukleotidsequenz – vollständig komplementär 45

Beispiel 11(a)-1: Doppelsträngige Nukleotidsequenz – gleiche Länge 45

Absatz 11(b) – Doppelsträngige Nukleotidsequenz – nicht vollständig komplementär 46

Beispiel 11(b)-1: Doppelsträngige Nukleotidsequenz – verschiedene Längen 46

Beispiel 11(b)-2: Doppelsträngige Nukleotidsequenz –Segment ohne Basenpaarung 47

Absatz 12 – Ringförmige Nukleotidsequenz 48

Beispiel 12-1: Ringförmige Nukleotidsequenz 48

Absatz 14 – Auslegung des Symbols "t" als Uracil in RNA 49

Beispiel 14-1: Das Symbol "t" repräsentiert Uracil in RNA 49

Absatz 27 – Das restriktivste Mehrdeutigkeitssymbol für Aminosäuren sollte verwendet werden 51

Beispiel 27-1: Kurzformel für eine Aminosäuresequenz 51

Beispiel 27-2: Kurzformel – weniger als vier spezifisch definierte Aminosäuren 52

Beispiel 27-3: Kurzformel – vier oder mehr spezifisch definierte Aminosäuren 53

Absatz 28 – Aminosäuresequenzen, die durch interne Terminatorsymbole getrennt sind 54

Beispiel 28-1: Codierende Nukleotidsequenz und codierte Aminosäuresequenz 54

Absatz 29 – Darstellung einer "sonstigen" ("other") Aminosäure 56

Beispiel 29-1: Das restriktivste Mehrdeutigkeitssymbol für eine "sonstige" ("other") Aminosäure 56

Beispiel 29-2: Verwendung der entsprechenden nicht modifizierten Aminosäure 57

Absatz 30 – Annotation einer modifizierten Aminosäure 59

Beispiel 30-1: Merkmalschlüssel "CARBOHYD" 59

Beispiel 30-2: Posttranslational modifizierte Aminosäuren 60

Absatz 36 – Sequenzen, die Regionen mit einer genauen Zahl an zusammenhängenden "n"- oder "X"-Resten enthalten 61

Beispiel 36-1: Sequenz mit einer Region, die eine bekannte Zahl an "X"-Resten enthält, dargestellt als eine einzelne Sequenz 61

Beispiel 36-2: Sequenz mit mehreren Regionen, die eine bekannte Zahl an "X"-Resten enthalten, dargestellt als eine einzelne Sequenz 62

Beispiel 36-3: Sequenz mit mehreren Regionen, die eine bekannte Zahl an "X"-Resten enthalten, dargestellt als eine einzelne Sequenz 63

Absatz 37 – Sequenzen, die Regionen mit einer unbekannten Zahl an zusammenhängenden "n"- oder "X"-Resten enthalten 64

Beispiel 37-1: Sequenz mit Regionen, die eine unbekannte Zahl an "X"-Resten enthalten, darf nicht als eine einzelne Sequenz dargestellt werden 64

Beispiel 37-2: Sequenz mit Regionen, die eine unbekannte Zahl an "X"-Resten enthalten, darf nicht als eine einzelne Sequenz dargestellt werden 65

Absatz 55– Eine Nukleotidsequenz, die sowohl DNA- als auch RNA-Segmente enthält 66

Beispiel 55-1: Kombiniertes DNA/RNA-Molekül 66

Absatz 89 – Der Merkmalschlüssel "CDS" **67**

Beispiel 89-1: Codierende Nukleotidsequenz und codierte Aminosäuresequenz **67**

Beispiel 89-2: Lage des Merkmals geht über die offenbarte Sequenz hinaus **68**

Absatz 92 – Aminosäuresequenz, die durch eine codierende Sequenz codiert ist **70**

Beispiel 92-1: Aminosäuresequenz, die durch eine codierende Sequenz mit Introns codiert wird **70**

Absatz 93 – Primärsequenz und eine Variante, die jeweils durch Aufzählung der Reste dargestellt werden **72**

Beispiel 93-1: Darstellung der durch Aufzählung dargestellten Varianten **72**

Beispiel 93-2: Darstellung der durch Aufzählung dargestellten Varianten **73**

Beispiel 93-3: Darstellung einer Konsensussequenz **74**

Absatz 94 – Sequenzvariante, die als eine einzelne Sequenz mit durch Aufzählung dargestellten alternativen Resten offenbart wird **76**

Beispiel 94-1: Darstellung einer einzelnen Sequenz mit durch Aufzählung dargestellten alternativen Aminosäuren **76**

Beispiel 94-2: Darstellung einer einzelnen Sequenz mit durch Aufzählung dargestellten alternativen Aminosäuren, bei denen es sich um modifizierte Aminosäuren handeln kann **77**

Absatz 95(a) – Sequenzvariante, die nur durch Verweis auf eine Primärsequenz mit mehreren unabhängigen Variationen offenbart wird **78**

Beispiel 95(a)-1: Darstellung einer Sequenzvariante durch Annotation der Primärsequenz **78**

Absatz 95(b) – Sequenzvariante, die nur durch Verweis auf eine Primärsequenz mit mehreren voneinander abhängigen Variationen offenbart wird **80**

Beispiel 95(b)-1: Darstellung einzelner Sequenzvarianten mit mehreren voneinander abhängigen Variationen **80**

BEISPIELE

Absatz 3(a) – Definition von "Aminosäure"

Beispiel 3(a)-1: D-Aminosäuren

In einer Patentanmeldung wird folgende Sequenz beschrieben:

Cyclo (D-Ala-D-Glu-Lys-Nle-Gly-D-Met-D-Nle)

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In Absatz 3(a) des Standards sind "Aminosäuren" definiert als u.a. "D-Aminosäuren" und Aminosäuren, die modifizierte oder synthetische Seitenketten enthalten. Basierend auf dieser Definition enthält das durch Aufzählung dargestellte Peptid fünf Aminosäuren, die spezifisch definiert sind (D-Ala, D-Glu, Lys, Gly und D-Met). Daher muss die Sequenz nach ST.26 Absatz 7(b) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Nach Absatz 29 sollten D-Aminosäuren in der Sequenz als die entsprechenden nicht modifizierten L-Aminosäuren dargestellt werden. Ferner muss jede modifizierte Aminosäure, die nicht durch ein anderes in Anhang I, Abschnitt 3, Tabelle 3 aufgeführtes Symbol dargestellt werden kann, durch das Symbol "X" dargestellt werden.

In diesem Beispiel enthält die Sequenz drei D-Aminosäuren, die durch eine in Anhang I, Abschnitt 3, Tabelle 3 aufgeführte nicht modifizierte L-Aminosäure dargestellt werden können, sowie eine L-Aminosäure (Nle) und eine D-Aminosäure (D-Nle), die durch das Symbol "X" dargestellt werden müssen.

Aus Absatz 25 geht hervor: Wenn eine Aminosäuresequenz ringförmig ist und der Ring ausschließlich aus Aminosäureresten besteht, die durch Peptidbindungen verbunden sind, muss der Anmelder auswählen, welcher Aminosäure die Restenummer 1 zugewiesen wird. Dementsprechend kann die Sequenz wie folgt dargestellt werden:

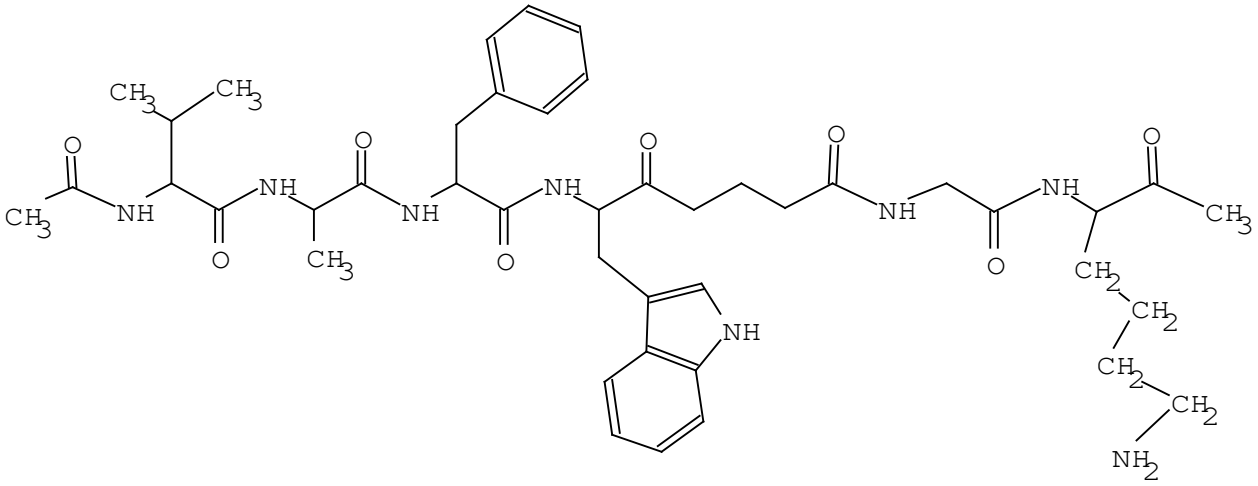
AEKXGMX (SEQ ID NO: 1)

oder auf andere Weise, indem einer beliebigen anderen Aminosäure in der Sequenz die Restenummer 1 zugewiesen wird. Für jede D-Aminosäure muss ein Merkmalschlüssel "SITE" und ein Qualifier "note" mit dem vollständigen, ungekürzten Namen der D-Aminosäure als Qualifier-Wert angegeben werden, z. B. "D-alanine" ("D-Alanin") und "D-norleucine" ("D-Norleucin"). Erforderlich sind des Weiteren ein Merkmalschlüssel "SITE" und ein Qualifier "note" mit der Abkürzung für L-Norleucin als Qualifier-Wert, d. h. "Nle", wie in Anhang I, Abschnitt 4, Tabelle 4 festgelegt. Außerdem sollte durch einen Merkmalschlüssel "REGION" und einen Qualifier "note" angezeigt werden, dass das Peptid ringförmig ist.

Relevante Absätze im Standard ST.26: 3(a), 7(b), 25, 26, 29, 30 und 31

Absatz 3(c) – Definition von "Aufzählung der Reste"

Beispiel 3(c)-1: Aufzählung von Aminosäuren nach chemischer Struktur



Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

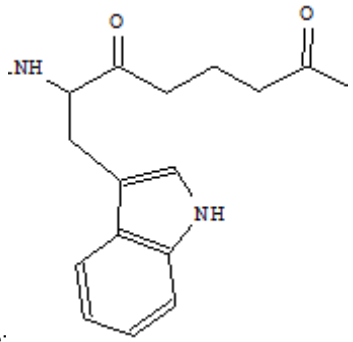
JA

Das durch Aufzählung dargestellte Peptid mit der abgebildeten Struktur enthält mindestens vier spezifisch definierte Aminosäuren. Daher muss die Sequenz in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz kann wie folgt dargestellt werden:

VAFXGK (SEQ ID NO: 2)



Dabei steht "X" für eine "sonstige" ("other") modifizierte Aminosäure: . Diese erfordert einen Merkmalschlüssel "SITE" mit dem Qualifier "note". Im Qualifier "note" wird der vollständige, ungekürzte Name des modifizierten Tryptophans an Position 4 des durch Aufzählung dargestellten Peptids angegeben, z. B. "6-amino-7-(1H-indol-3-yl)-5-oxoheptanoic acid" ("6-Amino-7-(1H-Indol-3-yl)-5-Oxoheptansäure"). Durch die Methylierung des C-Terminus ändert sich die chemische Struktur des terminalen Lysins, da das -OH am terminalen Ende durch $-CH_3$ ersetzt wird. Aufgrund dieser Änderung in der Struktur gilt das Lysin innerhalb der Sequenz als "modifizierte Aminosäure". Entsprechend sind ein "SITE"-Merkmalschlüssel und ein "note"-Qualifier erforderlich, um die Methylierung des C-Terminus anzuzeigen. Valin hingegen gilt nicht als "modifizierte Aminosäure", da die Addition der Acetylgruppe mit einer konventionellen Peptidbindung einhergeht. Die Acetylierung ändert die Struktur des Valins nicht. Entsprechend sollte die Acetylierung des N-Terminus durch einen zusätzlichen Merkmalschlüssel "SITE" und einen "note"-Qualifier angezeigt werden.

Alternativ kann die Sequenz wie folgt dargestellt werden:

VAFW (SEQ ID NO: 3)

Ein Merkmalschlüssel "SITE" mit dem Qualifier "note" ist erforderlich, um die Modifizierung von Tryptophan an Position 4 des durch Aufzählung dargestellten Peptids mit dem folgenden Wert anzuzeigen: "C-terminus linked via a glutaraldehyde bridge to dipeptide GK" ("C-Terminus über eine Glutaraldehydbrücke an Dipeptid-GK gebunden"). Darüber hinaus sollte ein zusätzlicher "SITE"-Merkmalschlüssel an Position 1 und ein "note"-Qualifier verwendet werden, um die Acetylierung des N-Terminus anzuzeigen.

Relevante Absätze im Standard ST.26: 3(c), 7(b), 29, 30 und 31

Beispiel 3(c)-2: Kurzformel für eine Aminosäuresequenz

$(G_4z)_n$

Wobei G = Glycin, z = jede beliebige Aminosäure und die Variable n kann jede ganze Zahl sein.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der Offenbarung wird angezeigt, dass es sich bei "n" um "jede ganze Zahl" handeln kann; daher ist die umfassendste Ausführungsform von "n" unbestimmt. Da "n" unbestimmt ist, kann das in der Formel angegebene Peptid nicht auf eine bestimmte Länge erweitert werden, sodass die nicht erweiterte Formel berücksichtigt werden muss.

Das durch Aufzählung dargestellte Peptid in der nicht erweiterten Formel ("n" = 1) enthält vier spezifisch definierte Aminosäuren, von denen jede Gly ist, und das Symbol "z". Konventionell ist "Z" das Symbol für "Glutamin oder Glutaminsäure", doch in dem Beispiel ist "z" als "jede beliebige Aminosäure" definiert. Gemäß ST.26 wird eine Aminosäure, die nicht spezifisch definiert ist, durch "X" dargestellt. Auf der Grundlage dieser Analyse enthält das durch Aufzählung dargestellte Peptid, d. h. GGGGX, vier Glycinreste, die durch Aufzählung dargestellt und spezifisch definiert sind. Daher ist nach ST.26 Absatz 7(b) die Aufnahme der Sequenz in ein Sequenzprotokoll erforderlich.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

In der Sequenz wird "z" als nicht konventionelles Symbol verwendet, dessen Definition der Offenbarung zu entnehmen ist (siehe Einführung in dieses Dokument). Da "z" als jede beliebige Aminosäure definiert ist, wird zur Darstellung dieser Aminosäure das konventionelle Symbol "X" verwendet. Daher muss die Sequenz als einzelne Sequenz dargestellt werden:

GGGGX (SEQ ID NO: 4)

und sollte annotiert werden mit dem Merkmalschlüssel "REGION", der Lage des Merkmals ">5" (entspricht >5) und einem "note"-Qualifier mit dem Wert "The entire sequence of amino acids 1-5 can be repeated one or more times." ("Die gesamte Sequenz der Aminosäuren 1-5 kann einmal oder mehrmals wiederholt werden.")

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" in diesem Beispiel "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") annotiert werden.

Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 3(c), 7(b) und 27

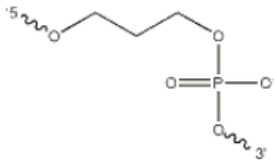
Absatz 3(g) – Definition von "Nukleotid"

Beispiel 3(g)-1: Durch einen C3-Spacer unterbrochene Nukleotidsequenz

In einer Patentanmeldung wird folgende Sequenz beschrieben:

atgcatgcatgcncggcatgcatgc

Dabei ist n = ein C3-Spacer mit folgender Struktur:



Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die durch Aufzählung dargestellte Sequenz enthält zwei Segmente spezifisch definierter Nukleotide, die durch einen C3-Spacer getrennt sind.

Der C3-Spacer ist kein Nukleotid gemäß Absatz 3(g); das konventionelle Symbol "n" wird auf nicht konventionelle Weise verwendet (siehe Einführung in dieses Dokument). Folglich ist jedes der beiden Segmente eine separate Nukleotidsequenz. Da beide Segmente mehr als zehn spezifisch definierte Nukleotide enthalten, müssen beide in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Jedes Segment muss als separate Sequenz in ein Sequenzprotokoll aufgenommen werden, wobei jedes Segment mit einer eigenen Sequenzkennzahl versehen wird:

atgcatgcatgc (SEQ ID NO.: 5)

cggcatgcatgc (SEQ ID NO.: 6)

Das Cytosin in jedem Segment, das an den C3-Spacer angeheftet ist, sollte in einer Merkmaltabelle unter Verwendung des Merkmalschlüssels "misc_feature" und eines "note"-Qualifiers näher beschrieben werden. Aus dem Wert für den "note"-Qualifier mit dem Format "Freitext" sollte hervorgehen, dass ein Spacer vorliegt, der mit einer anderen Nukleinsäure verbunden ist, und dieser Spacer sollte durch seinen vollständigen, ungekürzten chemischen Namen oder durch seinen gebräuchlichen Namen, z. B. C3-Spacer, kenntlich gemacht werden.

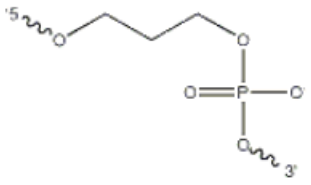
Relevante Absätze im Standard ST.26: 3(g), 7(a) und 15

Beispiel 3(g)-2: Nukleotidsequenz mit alternativen Resten, einschließlich eines C3-Spacers

In einer Patentanmeldung wird folgende Sequenz beschrieben:

atgcatgcatgcncggcatgcatgc

Dabei ist n = c, a, g oder ein C3-Spacer mit folgender Struktur:



Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der durch Aufzählung dargestellten Sequenz, die durch die Variable "n" unterbrochen wird, liegen 24 spezifisch definierte Reste vor. Um festzustellen, ob das "n" auf konventionelle oder nicht konventionelle Weise verwendet wird, muss die Erläuterung zur Sequenz in der Offenbarung herangezogen werden (siehe die Einführung in dieses Dokument).

Aus der Offenbarung geht hervor, dass n = c, a, g oder ein C3-Spacer ist. Das "n" ist ein konventionelles Symbol, das auf nicht konventionelle Weise verwendet wird, da es laut Erläuterung auch für einen C3-Spacer verwendet wird, was nicht der Definition eines Nukleotids entspricht. Des Weiteren steht das Symbol "n" der Erläuterung nach für "c", "a" oder "g"; daher ist nach ST.26 die Aufnahme der Sequenz aus 25 Nukleotiden in ein Sequenzprotokoll erforderlich. Da zwei durch einen C3-Spacer getrennte Segmente zwei Sequenzen darstellen, die sich von der aus 25 Nukleotiden bestehenden Sequenz unterscheiden, können auch diese beiden Sequenzen aus je 12 Nukleotiden in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Aus dem Beispiel geht hervor, dass "n = c, a, g oder ein C3-Spacer" ist. Wie oben erwähnt, ist ein C3-Spacer kein Nukleotid. Laut Absatz 15 darf das Symbol "n" für nichts anderes als ein Nukleotid verwendet werden; daher kann das Symbol "n" in einem Sequenzprotokoll keinen C3-Spacer darstellen.

Des Weiteren sollte laut Absatz 15 das restriktivste Symbol verwendet werden, wenn ein Mehrdeutigkeitssymbol angebracht ist. Gemäß Anhang I, Abschnitt 1, Tabelle 1 steht das Symbol "v" für "a oder c oder g" und ist damit restriktiver als "n".

Wenn die Variable "n" im Beispiel für c, a oder g steht, ist die durch Aufzählung der Reste dargestellte einzelne Sequenz, die die meisten offenbarten Ausführungsformen enthält und daher als die umfassendste Sequenz (siehe Einführung in dieses Dokument) in ein Sequenzprotokoll aufgenommen werden muss, wie folgt:

atgcatgcatgcvcggcatgcatgc (SEQ ID NO: 7)

Es wird nachdrücklich empfohlen, weitere Sequenzen aufzunehmen, die für die Offenbarung oder die Ansprüche der Erfindung wesentlich sind, wie in der Einführung in dieses Dokument erläutert.

Wenn es sich bei der Variable "n" im Beispiel um einen C3-Spacer handelt, kann die Sequenz als zwei separate Segmente spezifisch definierter Nukleotide auf beiden Seiten der Variablen "n" betrachtet werden, d. h. atgcatgcatgc (SEQ ID NO: 8); und cggcatgcatgc (SEQ ID NO: 9). Wenn diese beiden Sequenzen für die Offenbarung oder die Ansprüche wesentlich sind, sollten sie ebenfalls in das Sequenzprotokoll aufgenommen werden, wobei jede Sequenz mit einer eigenen Sequenzkennzahl zu versehen ist.

Das Cytosin in jedem Segment, das an den C3-Spacer angeheftet ist, sollte in einer Merkmaltabelle unter Verwendung des Merkmalschlüssels "misc_feature" und eines "note"-Qualifiers näher beschrieben werden. Im "note"-Qualifier mit dem Format "Freitext" sollte angegeben werden, dass ein Spacer vorliegt, der mit einer anderen Nukleinsäure verbunden ist, und dieser Spacer sollte durch seinen vollständigen, ungekürzten chemischen Namen oder durch seinen gebräuchlichen Namen, z. B. C3-Spacer, kenntlich gemacht werden.

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 3(g), 7(a) und 15

Beispiel 3(g)-3: Abasische Stelle

In einer Patentanmeldung wird folgende Sequenz beschrieben:

gagcattgac-AP-taaggct

Dabei ist AP eine abasische Stelle.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die spezifisch definierten Reste der durch Aufzählung dargestellten Sequenz werden durch eine abasische Stelle unterbrochen. Auf der 5'-Seite der abasischen Stelle befinden sich zehn Nukleotide, und auf der 3'-Seite der abasischen Stelle befinden sich sieben Nukleotide. Laut Absatz 3(g)(ii)(2) ist eine abasische Stelle als "Nukleotid" definiert, wenn sie Teil einer Nukleotidsequenz ist. Bei der Beurteilung, ob und wie die Sequenz in ein Sequenzprotokoll aufgenommen wird, wird die in diesem Beispiel aufgeführte abasische Stelle folglich als "Nukleotid" betrachtet. Dementsprechend sind die Reste auf beiden Seiten der abasischen Stelle Teil einer einzigen durch Aufzählung dargestellten Sequenz mit insgesamt 18 Nukleotiden, von denen 17 spezifisch definiert sind. Daher muss die Sequenz nach ST.26 Absatz 7(a) als einzelne Sequenz in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

gagcattgacntaaggct (SEQ ID NO: 10)

Die abasische Stelle muss durch ein "n" dargestellt und in einer Merkmaltabelle näher beschrieben werden. Die bevorzugte Methode der Annotation ist der Merkmalschlüssel "modified_base" und der obligatorische Qualifier "mod_base" mit dem Wert "OTHER". Des Weiteren muss die modifizierte Base durch einen "note"-Qualifier als abasische Stelle beschrieben werden.

Relevante Absätze im Standard ST.26: 3(g), 7(a) und 17

Beispiel 3(g)-4: Nukleinsäure-Analoga

Eine Patentanmeldung offenbart die folgende Glykolnukleinsäure-(GNA)-Sequenz:

PO_4-tagttcattgactaaggctccccattgact-OH

Dabei ahmt das linke Ende der Sequenz das 5'-Ende einer DNA-Sequenz nach.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA – Die einzelnen Reste, die eine GNA-Sequenz darstellen, werden nach ST.26 Absatz 3(g)(i)(2) als Nukleotide betrachtet. Dementsprechend hat die Sequenz mehr als zehn durch Aufzählung dargestellte und "spezifisch definierte" Nukleotide und muss in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

GNA-Sequenzen verfügen nicht über ein 5'- und ein 3'-Ende, sondern über ein 3'- und ein 2'-Ende. Das 3'-Ende, das üblicherweise mit einer terminalen Phosphatgruppe dargestellt wird, entspricht dem 5'-Ende von DNA oder RNA. (Beachten Sie, dass andere Nukleinsäure-Analoga möglicherweise auf andere Weise dem 5'-Ende und dem 3'-Ende von DNA und RNA entsprechen.) Gemäß Absatz 11 muss die Sequenz "in einer Richtung von links nach rechts, mit der die 5'-3'-Richtung nachgeahmt wird", in ein Sequenzprotokoll aufgenommen werden. Daher muss sie wie folgt in ein Sequenzprotokoll aufgenommen werden:

tagttcattgactaaggctccccattgact (SEQ ID NO: 11)

Die Sequenz muss in einer Merkmaltabelle mit dem Merkmalschlüssel "modified_base" und dem obligatorischen Qualifier "mod_base" mit der Abkürzung "OTHER" beschrieben werden. In einem "note"-Qualifier muss der vollständige, ungekürzte Name der modifizierten Nukleotide angegeben werden, z. B. "glycol nucleid acids" ("Glykolnukleinsäuren") oder "2,3-dihydroxypropyl nucleosides" ("2,3-Dihydroxypropyl-Nukleoside"). Um die gesamte Sequenz als GNA zu beschreiben, kann ein einzelnes "INSDFeature"-Element verwendet werden, wobei die Lage des Merkmals ("INSDFeature_location") den Bereich "1..30" enthält.

Relevante Absätze im Standard ST.26: 3(d), **3(g)**, 7(a), 11, 16, 18, 65 und 66

Absatz 3(k) – Definition von "spezifisch definiert"

Beispiel 3(k)-1: Mehrdeutigkeitssymbole für Nukleotide

5’ NNG KNG KNG K 3’

N und K sind IUPAC-IUB-Codes für Mehrdeutigkeit

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

NEIN

Die IUPAC-IUB-Codes für Mehrdeutigkeit entsprechen der Liste der Nukleotidsymbole, die in Anhang I, Abschnitt 1, Tabelle 1 definiert ist. Nach Absatz 3(k) sind spezifisch definierte Nukleotide alle anderen als die in Anhang I mit dem Symbol "n" dargestellten Nukleotide. Folglich sind "K" und "G" spezifisch definierte Nukleotide, und "N" ist kein spezifisch definiertes Nukleotid.

Die durch Aufzählung dargestellte Sequenz enthält keine zehn oder mehr spezifisch definierten Nukleotide und muss daher nach ST.26 Absatz 7(a) nicht in ein Sequenzprotokoll aufgenommen werden.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

NEIN

Nach Absatz 8 darf ein Sequenzprotokoll keine Sequenzen enthalten, die aus weniger als zehn spezifisch definierten Nukleotiden bestehen. Die durch Aufzählung dargestellte Sequenz enthält keine zehn oder mehr spezifisch definierten Nukleotide und darf daher nicht in ein Sequenzprotokoll aufgenommen werden.

Relevante Absätze im Standard ST.26: 3(k), 7(a), 8 und 13

Beispiel 3(k)-2: Konventionelle und nicht konventionelle Verwendung des Mehrdeutigkeitssymbols "n"

In einer Anmeldung wird die folgende künstliche Sequenz offenbart: 5'-AATGCCGGAN-3'. In der Offenbarung wird des Weiteren angegeben:

(i) in einer Ausführungsform steht N für jedes beliebige Nukleotid;

(ii) in einer Ausführungsform ist N fakultativ, steht aber vorzugsweise für G;

(iii) in einer Ausführungsform steht N für K;

(iv) in einer Ausführungsform steht N für C.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

NEIN

In der durch Aufzählung dargestellten Sequenz liegen neun spezifisch definierte Nukleotide und ein "N" vor. Um festzustellen, ob das Symbol "N" auf konventionelle Weise verwendet wird, muss die Erläuterung zur Sequenz in der Offenbarung herangezogen werden (siehe Einführung in dieses Dokument).

Die Betrachtung der offenbarten Ausführungsformen (i) bis (iv) der durch Aufzählung dargestellten Sequenz zeigt, dass die umfassendste Ausführungsform von "N" "jedes beliebige Nukleotid" ist. In der umfassendsten Ausführungsform wird "N" in der durch Aufzählung dargestellten Sequenz auf konventionelle Weise verwendet.

In bestimmten Ausführungsformen wird "N" als spezifisch definierter Rest beschrieben (in Teil (iv) steht N für C). Allerdings wird bei der Beurteilung, ob eine Sequenz in ein Sequenzprotokoll aufgenommen werden muss, nur die umfassendste Ausführungsform (d. h. "N steht für jedes beliebige Nukleotid") berücksichtigt. Die durch Aufzählung dargestellte Sequenz, die beurteilt werden muss, ist folglich 5'-AATGCCGAN-3'.

Dieser Analyse zufolge enthält die durch Aufzählung dargestellte Sequenz, das heißt AATGCCGAN, keine zehn spezifisch definierten Nukleotide. Aus diesem Grund ist nach ST.26 Absatz 7(a) nicht erforderlich, dass die Sequenz in ein Sequenzprotokoll aufgenommen wird, obwohl "n" in einigen Ausführungsformen auch als spezifisches Nukleotid definiert ist.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

NEIN

Die Sequenz "AATGCCGGAN" darf nicht in ein Sequenzprotokoll aufgenommen werden.

Eine beschriebene alternative Sequenz kann jedoch in ein Sequenzprotokoll aufgenommen werden, wenn das "N" durch ein spezifisch definiertes Nukleotid ersetzt wird.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Aufnahme von Sequenzen, mit denen Ausführungsformen dargestellt werden, die ein wesentlicher Bestandteil der Erfindung sind, wird **nachdrücklich** empfohlen. Die Aufnahme dieser zusätzlichen Sequenzen ermöglicht eine gründlichere Recherche und informiert die Öffentlichkeit über den Gegenstand, für den ein Patent beantragt wird.

Für das obige Beispiel wird dringend empfohlen, dass die folgenden drei zusätzlichen Sequenzen mit einer jeweils eigenen Sequenzkennzahl in das Sequenzprotokoll aufgenommen werden:

aatgccggag (SEQ ID NO: 12)

aatgccggak (SEQ ID NO: 13)

aatgccggac (SEQ ID NO: 14)

Wenn weniger als alle drei der oben genannten Sequenzen aufgenommen werden, sollte das Nukleotid, das "n" ersetzt, zwecks Beschreibung der Alternativen annotiert werden. Wenn zum Beispiel nur die oben genannte Sequenz mit der SEQ ID NO 12 in das Sequenzprotokoll aufgenommen wird, sollte der Merkmalschlüssel "misc_difference" mit der Lage des Merkmals "10" in Verbindung mit zwei "replace"-Qualifiern verwendet werden, wobei der Wert für den einen "k" und den anderen "c" wäre.

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer

Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geisitges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 3(k), 7(a), 8 und 13

Beispiel 3(k)-3: Nicht konventionelle Verwendung des Mehrdeutigkeitssymbols "n"

In einer Anmeldung wird die folgende Sequenz offenbart: 5'-aatgttggan-3'

Dabei steht n für c.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Gemäß Absatz 3(k) sind spezifisch definierte Nukleotide alle anderen als die in Anhang I, Abschnitt 1, Tabelle 1 mit dem Symbol "n" dargestellten Nukleotide.

In diesem Beispiel wird "n" auf nicht konventionelle Weise verwendet, um ausschließlich "c" darzustellen. In der Offenbarung ist nicht angegeben, dass "n" in konventioneller Weise verwendet wird, um ein beliebiges Nukleotid darzustellen. Daher muss die Sequenz so interpretiert werden, als wäre das gleichwertige konventionelle Symbol, d. h. "c", in der Sequenz verwendet worden (siehe Einführung in dieses Dokument). Die durch Aufzählung dargestellte Sequenz, die berücksichtigt werden muss, ist folglich:

5'-aatgttggac-3'

Diese Sequenz enthält zehn spezifisch definierte Nukleotide und muss daher nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden: aatgttggac (SEQ ID NO: 15)

Relevante Absätze im Standard ST.26: 3(k) und 7(a)

Beispiel 3(k)-4: Mehrdeutigkeitssymbole, die nicht "n" sind, sind "spezifisch definiert"

In einer Patentanmeldung wird folgende Sequenz beschrieben:

5’ NNG KNG KNG KAG VCR 3’

Dabei sind N, K, V und R IUPAC-IUB-Codes für Mehrdeutigkeit

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die IUPAC-IUB-Codes für Mehrdeutigkeit entsprechen der Liste der Nukleotidsymbole, die in Anhang I, Abschnitt 1, Tabelle 1 definiert ist. Gemäß Absatz 3(k) sind spezifisch definierte Nukleotide alle anderen als die in Anhang I, Abschnitt 1, Tabelle 1 mit dem Symbol "n" dargestellten Nukleotide. Folglich sind "K", "V" und "R" "spezifisch definierte" Nukleotide.

Diese Sequenz enthält elf durch Aufzählung dargestellte und "spezifisch definierte" Nukleotide und muss daher nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

nngkngkngkagvcr (SEQ ID NO: 16)

Relevante Absätze im Standard ST.26: 3(k), 7(a) und 15

Beispiel 3(k)-5: Nicht konventionelle Verwendung der Mehrdeutigkeitsabkürzung "Xaa"

In einer Patentanmeldung wird folgende Sequenz beschrieben:

Xaa-Tyr-Glu-Xaa-Xaa-Xaa-Leu

Xaa an Position 1 ist jede beliebige Aminosäure, Xaa an Position 4 ist Lys, Xaa an Position 5 ist Gly und Xaa an Position 6 ist Leucin oder Isoleucin.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das in der Formel durch Aufzählung dargestellte Peptid enthält an den Positionen 2, 3 und 7 drei spezifisch definierte Aminosäuren. Die erste Aminosäure wird durch eine konventionelle Abkürzung dargestellt, nämlich Xaa, die für jede beliebige Aminosäure steht. Die Aminosäuren an der 4., 5. und 6. Position werden jedoch durch konventionelle Abkürzungen dargestellt, die auf nicht konventionelle Weise verwendet werden (siehe Einführung in dieses Dokument). Daher wird die Erläuterung zur Sequenz in der Offenbarung herangezogen, um die Definition von "Xaa" an diesen Positionen zu bestimmen. Da "Xaa" in den Positionen 4–6 als spezifische Aminosäure angegeben ist, muss die Sequenz so interpretiert werden, als ob die gleichwertigen konventionellen Abkürzungen, d. h. Lys, Gly und (Leu oder Ile), in der Sequenz verwendet worden wären. Folglich enthält die Sequenz vier oder mehr spezifisch definierte Aminosäuren und muss nach ST.26 Absatz 7(b) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

In der Sequenz wird die konventionelle Abkürzung "Xaa" in nicht konventioneller Weise verwendet. Daher muss die Erläuterung zur Sequenz in der Offenbarung herangezogen werden, um die Definition von "Xaa" an den Positionen 4, 5 und 6 zu bestimmen. In der Erläuterung wird angegeben, dass "Xaa" in Position 4 für ein Lysin, in Position 5 für ein Glycin und in Position 6 für ein Leucin oder Isoleucin steht. Die konventionellen Symbole für diese Aminosäuren sind K, G und J. Daher sollte die Sequenz in folgender Form in das Sequenzprotokoll aufgenommen werden.

XYEKGJL (SEQ ID NO: 17)

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" an Position 1 der SEQ ID NO: 17 "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") versehen werden.

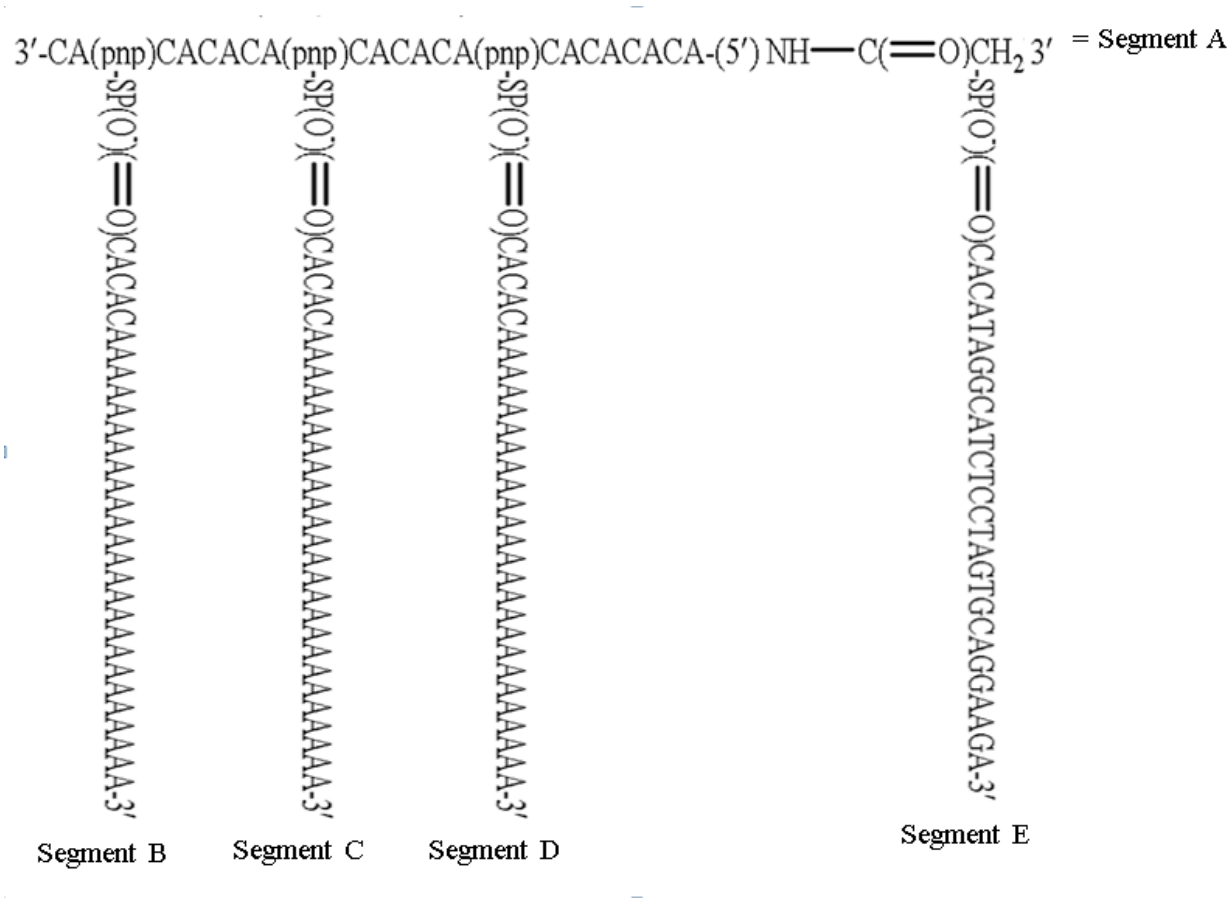
Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

Relevante Absätze im Standard ST.26: 3(k), 7(b), 26 und 27

Absatz 7(a) – Nukleotidsequenzen, die in ein Sequenzprotokoll aufgenommen werden müssen

Beispiel 7(a)-1: Verzweigte Nukleotidsequenz

Die Beschreibung offenbart die folgende verzweigte Nukleotidsequenz:



Dabei ist "pnp" eine Bindung oder ein Monomer, das eine Bromoacetylamino-Funktionalität enthält;

3’-CA(pnp)CACACA(pnp)CACACA(pnp)CACACACA-(5’)NH—C(=O)CH_2 3’ ist Segment A;

SP(O^-)(=O)CACACAAAAAAAAAAAAAAAAAAAAAA 3’ ist Segment B, C und D und

SP(O^-)(=O)CACATAGGCATCTCCTAGTGCAGGAAGA 3’ ist Segment E.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA – die vier vertikalen Segmente B–E müssen in ein Sequenzprotokoll aufgenommen werden

NEIN – das horizontale Segment A darf nicht in ein Sequenzprotokoll aufgenommen werden

Die obige Abbildung ist ein Beispiel für eine verzweigte "kombinierte" Nukleinsäuresequenz, die fünf lineare Segmente enthält: das horizontale Segment A und die vier vertikalen Segmente B bis E.

Nach Absatz 7(a) muss eine lineare Region einer verzweigten Nukleotidsequenz aus zehn oder mehr spezifisch definierten Nukleotiden, wobei benachbarte Nukleotide in 3’-5’-Richtung verbunden sind, in ein Sequenzprotokoll aufgenommen werden.

Die vier vertikalen Segmente B bis E enthalten jeweils mehr als zehn spezifisch definierte Nukleotide, wobei benachbarte Nukleotide in 3'-5'-Richtung miteinander verbunden sind und daher jeweils in ein Sequenzprotokoll aufgenommen werden müssen.

Im horizontalen Segment A sind die linearen Regionen der Nukleotidsequenz durch die Nicht-Nukleotideinheit "pnp" miteinander verbunden, und jede dieser verbundenen linearen Regionen enthält weniger als zehn spezifisch definierte Nukleotide. Da also keine Region des Segments A zehn oder mehr spezifisch definierte Nukleotide enthält, wobei benachbarte Nukleotide in 3'-5'-Richtung verbunden sind, ist nach ST.26 Absatz 7(a) keine Aufnahme in ein Sequenzprotokoll erforderlich.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

NEIN

Nach Absatz 8 darf ein Sequenzprotokoll keine Sequenzen enthalten, die aus weniger als zehn spezifisch definierten Nukleotiden bestehen.

Keine Region von Segment A enthält zehn oder mehr spezifisch definierte Nukleotide, wobei benachbarte Nukleotide in 3'-5'-Richtung miteinander verbunden sind; daher darf Segment A nicht als separate Sequenz mit einer eigenen Sequenzkennzahl in ein Sequenzprotokoll aufgenommen werden.

Die Segmente B, C, D und E können jedoch annotiert werden, um anzuzeigen, dass sie mit Segment A verbunden sind.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Segmente B, C und D sind identisch und müssen in einem Sequenzprotokoll als eine einzelne Sequenz aufgeführt werden:

cacacaaaaaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO: 18)

Das erste "c" in der Sequenz sollte mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note" näher beschrieben werden, ein Beispiel für den Qualifier-Wert wäre: "This sequence is one of four branches of a branched polynucleotide." ("Diese Sequenz ist eine der vier Verzweigungen eines verzweigten Polynukleotids.")

Segment E muss in einem Sequenzprotokoll als einzelne Sequenz aufgeführt werden:

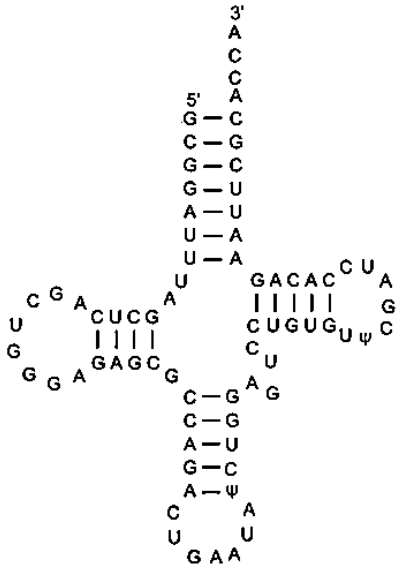
cacataggcatctcctagtgcaggaaga (SEQ ID NO: 19)

Das erste "c" in der Sequenz sollte mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note" näher beschrieben werden, ein Beispiel für den Qualifier-Wert wäre: "This sequence is one of four branches of a branched polynucleotide." ("Diese Sequenz ist eine der vier Verzweigungen eines verzweigten Polynukleotids.")

Relevante Absätze im Standard ST.26: 7(a), 8, 11, 13 und 17

Beispiel 7(a)-2: Lineare Nukleotidsequenz mit Sekundärstruktur

In einer Patentanmeldung wird folgende Sequenz beschrieben:



Dabei steht Ψ für Pseudouridin.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die Nukleotidsequenz enthält 73 durch Aufzählung dargestellte und spezifisch definierte Nukleotide. Somit enthält diese Beispielsequenz zehn oder mehr "spezifisch definierte" Nukleotide und muss daher nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Aus der Offenbarung geht hervor, dass "Ψ" mit Pseudouridin gleichzusetzen ist. Das einzige konventionelle Symbol, das zur Darstellung von Pseudouridin verwendet werden kann, ist "n"; daher ist das "Ψ" ein nicht konventionelles Symbol, das zur Darstellung des konventionellen Symbols "n" verwendet wird (siehe Einführung in dieses Dokument). Dementsprechend muss die Sequenz so interpretiert werden, als ob sie anstelle der beiden "Ψ"-Symbole zwei "n"-Symbole hätte.

Das Symbol "u" darf in einem Sequenzprotokoll nicht verwendet werden, um Uracil in einem RNA-Molekül darzustellen. Nach Absatz 14 wird das Symbol "t" in RNA als Uracil ausgelegt. Die Sequenz muss wie folgt aufgenommen werden:

gcggatttagctcagctgggagagcgccagactgaatanctggagtcctgtgtncgatccacagaattcgcacca (SEQ ID NO: 20)

Der obligatorische Qualifier "mol_type" des obligatorischen Merkmalschlüssels "source" hat den Wert "tRNA". Mit dem Merkmalschlüssel „tRNA" und entsprechenden Qualifiern können weitere Informationen bereitgestellt werden.

Die "n"-Reste sind in einer Merkmaltabelle unter Verwendung des Merkmalschlüssels "modified_base" und des obligatorischen Qualifiers "mod_base" mit der Abkürzung "p" für Pseudouridin als Qualifier-Wert näher zu beschreiben (siehe Anhang I, Tabelle 2).

Relevante Absätze im Standard ST.26: 7(a), 11, 13, 14, 17, 62, 84 und Anhang I, Abschnitte 2 und 5, Merkmalschlüssel 5.43

Beispiel 7(a)-3: Nicht konventionelle Verwendung von Mehrdeutigkeitssymbolen für Nukleotide

In einer Patentanmeldung wird folgende Sequenz beschrieben:

5’ GATC-MDR-MDR-MDR-MDR-GTAC 3’

Aus der Erläuterung zur Sequenz in der Offenbarung geht ferner hervor: "Ein 'DR-Element' besteht aus der Sequenz 5’ ATCAGCCAT 3’. Ein mutiertes DR-Element bzw. MDR ist ein DR-Element, bei dem die mittleren 5 Nukleotide, CAGCC, zu TTTTT mutiert sind."

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der durch Aufzählung dargestellten Sequenz wird das Symbol "MDR" verwendet. Wenn unklar ist, ob ein in einer Sequenz verwendetes Symbol ein konventionelles Symbol, d. h. ein Symbol gemäß Anhang I, Abschnitt 3, Tabelle 3, oder ein nicht konventionelles Symbol sein soll, muss zur Klärung die Erläuterung zur Sequenz in der Offenbarung herangezogen werden (siehe Einführung in dieses Dokument). Gemäß Tabelle 3 könnte "MDR" als drei konventionelle Symbole (m = a oder c, d = a oder g oder t/u, r = g oder a) oder als Abkürzung, d. h. als Kurznotation für eine andere Struktur interpretiert werden.

Aus der Offenbarung geht hervor, dass ein MDR-Element mit 5’ ATTTTTTAT 3’ gleichzusetzen ist. Folglich handelt es sich bei den Buchstaben "MDR" um konventionelle Symbole, die auf nicht konventionelle Weise verwendet werden; daher muss die Sequenz so interpretiert werden, als ob sie mit den gleichwertigen konventionellen Symbolen offenbart worden wäre. Dementsprechend lautet die durch Aufzählung dargestellte Sequenz, deren Aufnahme in ein Sequenzprotokoll geprüft wird:

5’ GATC ATTTTTTAT ATTTTTTAT ATTTTTTAT ATTTTTTAT GTAC 3’

Diese durch Aufzählung dargestellte Sequenz enthält 44 spezifisch definierte Nukleotide und muss daher nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

gatcatttttttatatttttttatatttttttatatttttttatgtac (SEQ ID NO: 21)

Relevante Absätze im Standard ST.26: 7(a) und 13

Beispiel 7(a)-4: Nicht konventionelle Verwendung von Mehrdeutigkeitssymbolen für Nukleotide

In einer Patentanmeldung wird folgende Sequenz beschrieben:

5’ ATTC-N-N-N-N-GTAC 3’

Aus der Erläuterung zur Sequenz in der Offenbarung geht ferner hervor, dass "N" aus der Sequenz 5’ ATACGCACT 3’ besteht.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der durch Aufzählung dargestellten Sequenz wird das Symbol "N" verwendet. Um festzustellen, ob das "N" auf konventionelle oder nicht konventionelle Weise verwendet wird, muss die Erläuterung zur Sequenz in der Offenbarung herangezogen werden (siehe Einführung in dieses Dokument).

Aus der Offenbarung geht hervor, dass "N" mit 5’ ATACGCACT 3’ gleichzusetzen ist. Somit ist das "N" ein konventionelles Symbol, das auf nicht konventionelle Weise verwendet wird. Dementsprechend muss die Sequenz so interpretiert werden, als ob sie mit den gleichwertigen konventionellen Symbolen offenbart worden wäre:

5’ ATTC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3’

Diese Sequenz enthält 44 spezifisch definierte Nukleotide und muss daher nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

attcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO: 22)

Relevante Absätze im Standard ST.26: 7(a) und 13

Beispiel 7(a)-5: Nicht konventionelle Nukleotidsymbole

In einer Patentanmeldung wird folgende Sequenz beschrieben:

5’ GATC-β-β-β-β-GTAC 3’

Aus der Erläuterung zur Sequenz in der Offenbarung geht ferner hervor, dass "β" aus der Sequenz 5’ ATACGCACT 3’ besteht.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der durch Aufzählung dargestellten Sequenz wird das nicht konventionelle Symbol "β" verwendet. Daher muss die Erläuterung zur Sequenz in der Offenbarung herangezogen werden, um die Definition von "β" zu bestimmen (siehe Einführung in dieses Dokument).

Aus der Offenbarung geht hervor, dass "β" mit 5’ ATACGCACT 3’ gleichzusetzen ist. Das "β" ist also ein nicht konventionelles Symbol, das eine Sequenz von neun spezifisch definierten, konventionellen Symbolen darstellt. Dementsprechend muss die Sequenz so interpretiert werden, als ob sie mit den gleichwertigen konventionellen Symbolen offenbart worden wäre:

5’ GATC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3’

Diese durch Aufzählung dargestellte Sequenz enthält 44 spezifisch definierte Nukleotide und muss daher nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

gatcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO: 23)

Relevante Absätze im Standard ST.26: 7(a) und 13

Beispiel 7(a)-6: Nicht konventionelle Nukleotidsymbole

In einer Patentanmeldung wird folgende Sequenz beschrieben:

5’ GATC-β-β-β-β-GTAC 3’

Aus der Erläuterung zur Sequenz in der Offenbarung geht weiter hervor, dass "β" mit Adenin, Inosin oder Pseudouridin gleichzusetzen ist.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

NEIN

In der durch Aufzählung dargestellten Sequenz wird das nicht konventionelle Symbol "β" verwendet. Daher muss die Erläuterung zur Sequenz in der Offenbarung herangezogen werden, um die Definition von "β" zu bestimmen (siehe Einführung in dieses Dokument).

Aus der Offenbarung geht hervor, dass "β" mit Adenin, Inosin oder Pseudouridin gleichzusetzen ist. Das einzige konventionelle Symbol, das zur Darstellung von "Adenin, Inosin oder Pseudouridin" verwendet werden kann, ist "n"; daher ist das "β" ein nicht konventionelles Symbol, das zur Darstellung des konventionellen Symbols "n" verwendet wird. Dementsprechend muss die Sequenz so interpretiert werden, als ob sie anstelle der vier "β"-Symbole vier "n"-Symbole hätte (unten als "N" angezeigt):

5’ GATC-N-N-N-N-GTAC 3’

Die durch Aufzählung dargestellte Sequenz enthält nur acht spezifisch definierte Nukleotide und muss daher nach ST.26 Absatz 7(a) nicht in ein Sequenzprotokoll aufgenommen werden.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

NEIN

Die durch Aufzählung dargestellte Sequenz 5’GATC-N-N-N-N-GTAC 3’ darf nicht in ein Sequenzprotokoll aufgenommen werden.

Allerdings darf eine offenbarte alternative Sequenz in ein Sequenzprotokoll aufgenommen werden, wenn mindestens zwei der "n"-Symbole durch Adenin ersetzt werden, sodass eine Sequenz mit mindestens zehn oder mehr spezifisch definierten Nukleotiden entsteht.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Eine mögliche zulässige Darstellung ist:

gatcaaaagtac (SEQ ID NO: 24)

Im obigen Beispiel sollten die vier Adenin-Nukleotide, die die β-Symbole ersetzen, annotiert werden, um darauf hinzuweisen, dass diese Positionen durch Inosin oder Pseudouridin substituiert werden könnten.

Dazu sollte der Merkmalschlüssel "misc_difference" mit der Lage des Merkmals 5-8 und einem "note"-Qualifier mit z. B. folgendem Qualifier-Wert verwendet werden: "A nucleotide in any of positions 5-8 may be replaced with inosine or pseudouridine." ("Ein Nukleotid an einer der Positionen 5-8 kann durch Inosin oder Pseudouridin ersetzt werden.") Da es sich bei diesen Alternativen um modifizierte Nukleotide handelt, wäre der Merkmalschlüssel "modified_base" in Verbindung mit dem Qualifier "mod_base" erforderlich. Als Wert für den Qualifier "mod_base" kann "OTHER" und als Wert für den zugehörigen "note"-Qualifier "i or p" ("i oder p") eingesetzt werden.

Auch andere Permutationen sind möglich.

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss,

ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 7(a), 8, 13 und 17

Beispiel 7(a)-7: Invertierte Nukleotide I

Eine Patentanmeldung offenbart die folgende einzelsträngige DNA-Sequenz:

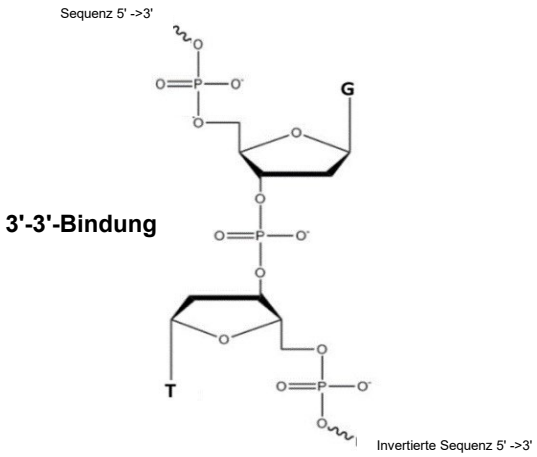
5' attgactaagtgttccccattgact 5'

wobei sich die Direktionalität der Sequenz innerhalb des Strangs aufgrund einer reversen 3'-3'-Bindung zwischen den Resten 12 und 13 ändert. Der unterstrichene Abschnitt der Sequenz hat eine 3'-5'-Orientierung von links nach rechts.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die ersten 12 Reste sind in der üblichen 5'-3'-Orientierung dargestellt. Das "g" an Position 12 ist über eine reverse 3'-3'-Bindung an das "t" an Position 13 gebunden:



Der Rest des Moleküls, dargestellt in den Positionen 13 bis 25, weist die entgegengesetzte Orientierung – 3'-5' – auf. Gemäß ST.26 Absatz 11 müssen Nukleotidsequenzen in Richtung vom 5'-Ende zum 3'-Ende von links nach rechts dargestellt werden.

Um dieses Molekül demzufolge in dem Sequenzprotokoll korrekt darzustellen, muss es anhand von zwei Sequenzen dargestellt werden: einer Sequenz für die Positionen 1 bis 12 und einer zweiten für die Positionen 13 bis 25. Jeder Abschnitt der Sequenz enthält zehn oder mehr "spezifisch definierte" Nukleotide und muss daher nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Positionen 1-12 müssen wie folgt in das Sequenzprotokoll aufgenommen werden:

attgactaagtg (SEQ ID NO.:99)

Position 12 sollte in einer Merkmaltabelle mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note" mit einem Wert beschrieben werden, der angibt, dass der Rest über eine 3'-3'-Phosphodiester-Bindung an ein Thymidin-3'-Monophosphat an eine invertierte Nukleotidsequenz gebunden ist.

Die Positionen 13-25 müssen wie folgt in das Sequenzprotokoll aufgenommen werden:

tcagttacccctt (SEQ ID NO.: 100)

Es ist zu beachten, dass diese Sequenz eine verglichen mit der Darstellung in der ursprünglichen Offenbarung umgekehrte Orientierung aufweist, sodass die Sequenz jetzt von links nach rechts vom 5'-Ende zum 3'-Ende orientiert ist. Position 13 sollte in einer Merkmaltabelle mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note" mit einem Wert beschrieben werden, der angibt, dass der Rest über eine 3'-3'-Phosphodiester-Bindung an ein Guanosin-3'-Monophosphat an eine invertierte Nukleotidsequenz gebunden ist.

Relevante Absätze im Standard ST.26: Absätze 7(a), 11

Beispiel 7(a)-8: Invertierte Nukleotide II

Eine Patentanmeldung offenbart die folgende DNA-Sequenz:

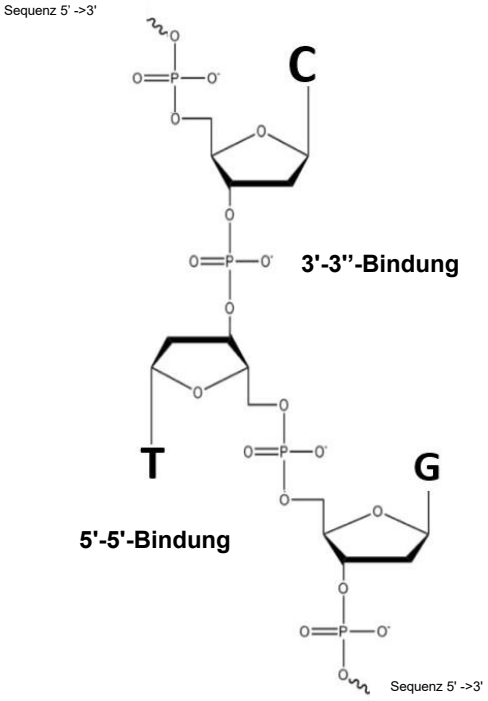
5'-attgactaagtgc**t**gcccattgact-3'

wobei der unterstrichene Thymin-Rest (Position 14) ein invertiertes Nukleotid ist, das über eine 3'-3'-Phosphodiester-Bindung an das Cytosin und über eine 5'-5'-Phosphodiester-Bindung an das Guanin gebunden ist.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das invertierte Thymidin an Position 14 unterbricht die 5'-3'-Direktionalität der Sequenz durch Einführen einer 3'-3'-Bindung zwischen den Resten 13 and 14 und einer 5'-5'-Bindung zwischen den Resten 14 und 15:



Das invertierte Thymidin verbindet den ersten Abschnitt der Sequenz, die Reste 1-13, mit dem zweiten Abschnitt der Sequenz, den Resten 15-25. Jeder Abschnitt der Sequenz enthält mindestens zehn durch Aufzählung dargestellte und spezifisch definierte Nukleotide. Daher muss jeder Abschnitt der Sequenz nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden. Das einzelne Thymidin, das die beiden Abschnitte der Sequenz verbindet, kann nicht als separate Sequenz in das Sequenzprotokoll aufgenommen werden, da es keine Sequenz aus mindestens 10 spezifisch definierten Nukleotiden ist.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Das invertierte Thymidin an Position 14 unterbricht die 5'-3'-Direktionalität der Sequenz durch Einführen einer 3'-3'-Bindung zwischen den Resten 13 and 14 und einer 5'-5'-Bindung zwischen den Resten 14 und 15. Gemäß ST.26 Absatz 11 müssen Nukleotidsequenzen in Richtung vom 5'-Ende zum 3'-Ende von links nach rechts dargestellt werden. Um dieses Molekül demzufolge in dem Sequenzprotokoll korrekt darzustellen, muss es anhand von zwei Sequenzen dargestellt werden: einer ersten Sequenz für die Positionen 1 bis 13 und einer zweiten Sequenz für die Positionen 15 bis 25.

Die Positionen 1-13 müssen wie folgt in das Sequenzprotokoll aufgenommen werden:

attgactaagtgc (SEQ ID NO: 101)

Position 13 sollte in einer Merkmaltabelle mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note" mit einem Wert beschrieben werden, der angibt, dass der Rest über eine 3'-3'-Phosphodiester-Bindung an ein Thymidin-3'-Monophosphat an eine andere Sequenz gebunden ist, das wiederum über eine 5'-5'-Phosphodiester-Bindung an eine andere Sequenz gebunden ist.

Die Positionen 15-25 müssen wie folgt in das Sequenzprotokoll aufgenommen werden:

gcccattgact (SEQ ID NO: 102)

Position 1 sollte in einer Merkmaltabelle mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note" mit einem Wert beschrieben werden, der angibt, dass der Rest über eine 5'-5'-Phosphodiester-Bindung an ein Thymidin-3'-Monophosphat an eine andere Sequenz gebunden ist, das wiederum über eine 3'-3'-Phosphodiester-Bindung an eine andere Sequenz gebunden ist.

Relevante Absätze im Standard ST.26: Absätze 7(a), 11

Absatz 7(b) – Aminosäuresequenzen, die in ein Sequenzprotokoll aufgenommen werden müssen

Beispiel 7(b)-1: Vier oder mehr spezifisch definierte Aminosäuren

XXXXXXXXDXXXXXXXXXXFXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAXXXXXXXXXXXXXXXXXXXXGXXXXX

Wobei X = jede beliebige Aminosäure

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das durch Aufzählung dargestellte Peptid enthält vier spezifisch definierte Aminosäuren. Das Symbol "X" wird konventionell verwendet, um die verbleibenden Aminosäuren als jede beliebige Aminosäure darzustellen (siehe Einführung in dieses Dokument).

Da vier spezifisch definierte Aminosäuren vorliegen, d. h. Asp, Phe, Ala und Gly, muss die Sequenz nach ST.26 Absatz 7(b) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz muss wie folgt dargestellt werden:

XXXXXXXXDXXXXXXXXXXFXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAXXXXXXXXXXXXXXXXXXXXGXXXXX (SEQ ID NO: 25)

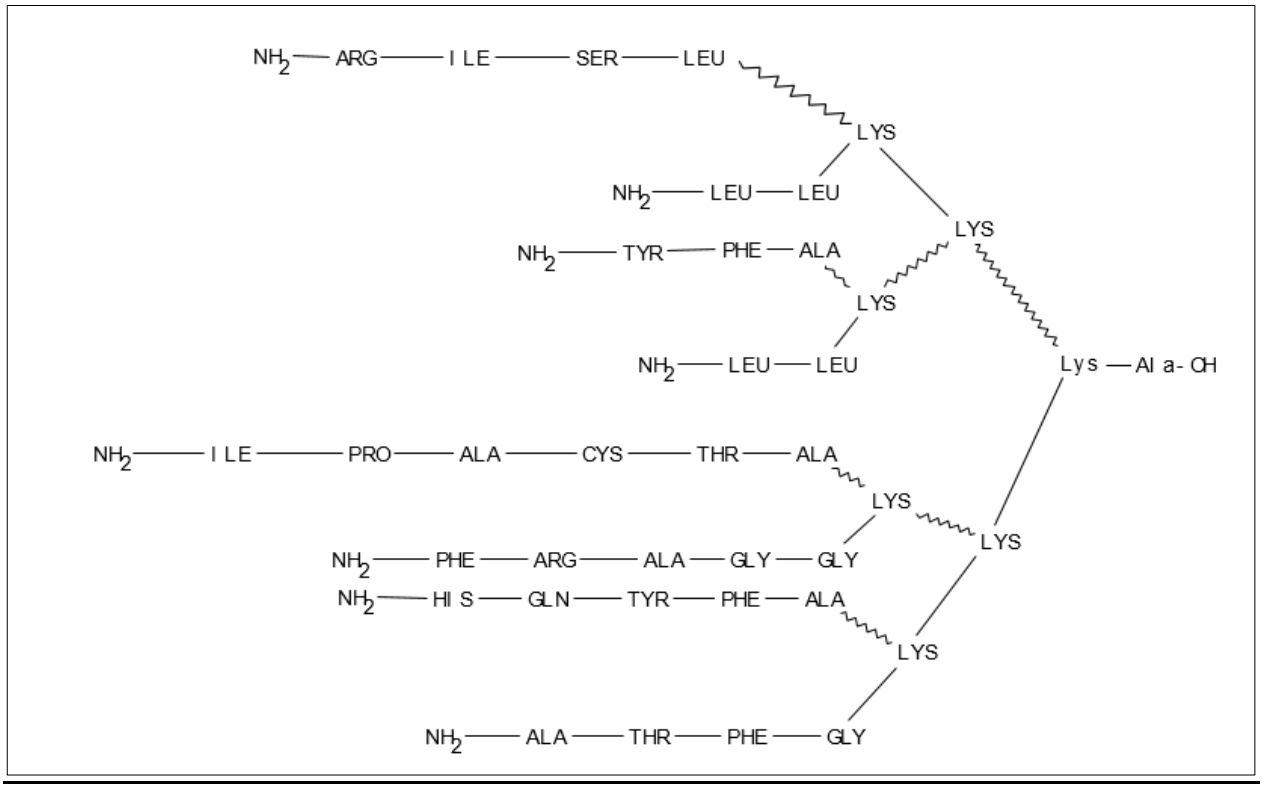
Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" in der SEQ ID NO 25 "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") versehen werden.

Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

Relevante Absätze im Standard ST.26: 7(b), 8 und 27

Beispiel 7(b)-2: Verzweigte Aminosäuresequenz

In der Anmeldung wird eine verzweigte Sequenz beschrieben, bei der ausgehend von den Lysinresten als Gerüst acht Zweige ausgebildet sind, an denen mehrere lineare Peptidketten befestigt sind. Als dibasische Aminosäure weist Lysin zwei Stellen für Peptidbindungen auf. Das Peptid wird wie folgt abgebildet:

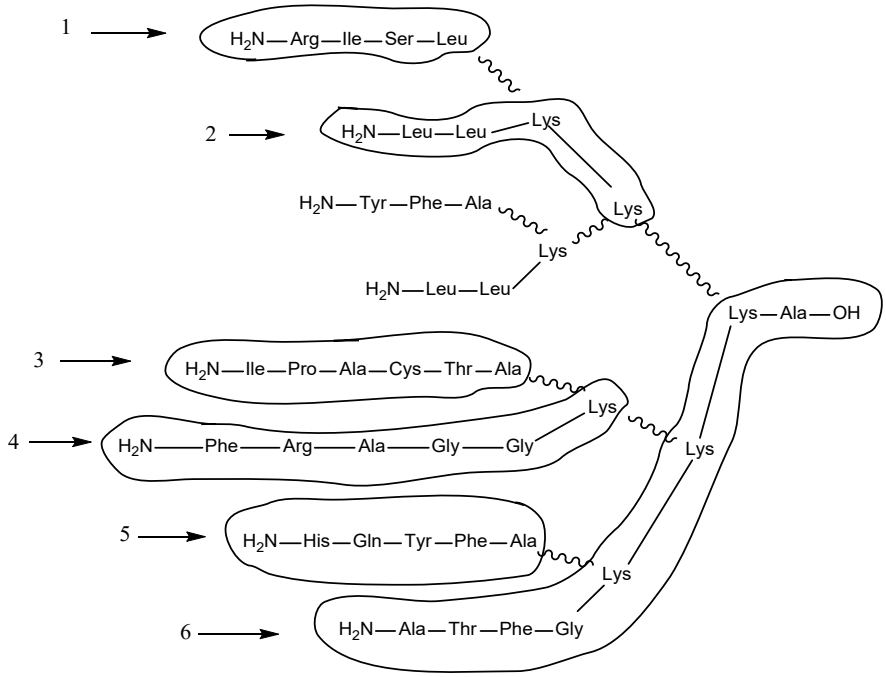


In dem oben dargestellten verzweigten Peptid bestehen die Bindungen zwischen Lysin und einer anderen, durch ——— dargestellten Aminosäure aus einer Amidbindung zwischen dem terminalen Amin des Lysins und dem Carboxylende der gebundenen Aminosäure. Die durch ∿∿∿ dargestellten Bindungen stellen Amidbindungen zwischen dem Amin der Seitenkette des Lysins und dem Carboxylende der gebundenen Aminosäure dar.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Im obigen Beispiel wird eine verzweigte Sequenz offenbart, bei der die Lysinreste als Gerüst dienen. Nach Absatz 7(b) muss die unverzweigte oder lineare Region der Sequenz in ein Sequenzprotokoll aufgenommen werden, wenn sie vier oder mehr spezifisch definierte Aminosäuren enthält. Im Folgenden sind die linearen Regionen des verzweigten Peptids eingekreist, die vier oder mehr spezifisch definierte Aminosäuren haben:



Nach ST.26 Absatz 7(b) ist die Aufnahme der oben abgebildeten Peptide 1–6 in ein Sequenzprotokoll erforderlich.

Folgende Peptide müssen nicht in das Sequenzprotokoll aufgenommen werden:

YFA

LLK

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

NEIN

Nach Absatz 8 darf ein Sequenzprotokoll keine Sequenzen enthalten, die aus weniger als vier spezifisch definierten Aminosäuren bestehen.

Die Peptide YFA und LLK enthalten jeweils nur drei spezifisch definierte Aminosäuren und dürfen daher nicht als separate Sequenzen mit eigenen Sequenzkennzahlen in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Peptide 1–6 müssen mit separaten Sequenzkennzahlen dargestellt werden:

RISL (SEQ ID NO: 26)

LLKK (SEQ ID NO: 27)

IPACTA (SEQ ID NO: 28)

FRAGGK (SEQ ID NO: 29)

HQYFA (SEQ ID NO: 30)

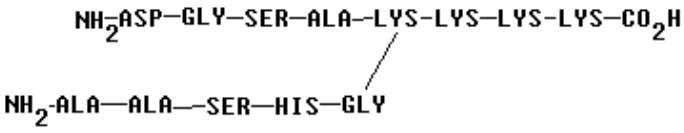
ATFGKKKA (SEQ ID NO: 31)

Die verzweigte Struktur kann mit dem Merkmalschlüssel "SITE" und dem obligatorischen Qualifier "note"annotiert werden, der z. B. folgenden Wert erhält: "This sequence is one part of a branched amino acid sequence" ("Diese Sequenz ist ein Teil einer verzweigten Aminosäuresequenz"). Nach ST.26 Absatz 30 müssen die Sequenzen mit den SEQ ID NO 27, 29 und 31 für jedes Lysin annotiert werden, um anzuzeigen, dass es sich um eine modifizierte Aminosäure handelt. Dabei ist der Merkmalschlüssel "SITE" in Verbindung mit dem Qualifier "note" zu verwenden, um zu beschreiben, dass die Seitenkette des Lysins über eine Amidbindung mit einer anderen Sequenz verbunden ist. Die Sequenzen mit den SEQ ID NO 26, 28 und 30 sollten jeweils annotiert werden, um anzuzeigen, dass die C-terminale Aminosäure mit einer anderen Sequenz verbunden ist. Dabei ist der Merkmalschlüssel "SITE" in Verbindung mit dem Qualifier "note" zu verwenden.

Relevante Absätze im Standard ST.26: 7(b), 8, 26, 29, 30 und 31

Beispiel 7(b)-3: Verzweigte Aminosäuresequenz

Es liegt ein Peptid mit der folgenden Sequenz vor:

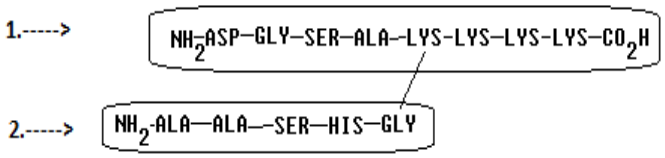


Die Verbindung zwischen dem terminalen Glycinrest in der unteren Sequenz und dem Lysin in der oberen Sequenz entsteht durch eine Amidbindung zwischen dem Carboxy-Terminus des Glycins und der aminoterminalen Seitenkette des Lysins.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die unverzweigte oder lineare Region einer Sequenz muss in ein Sequenzprotokoll aufgenommen werden, wenn sie vier oder mehr spezifisch definierte Aminosäuren enthält. Die linearen Regionen des verzweigten Peptids, die mehr als vier spezifisch definierte Aminosäuren haben, sind im obigen Beispiel:



Nach ST.26 Absatz 7(b) ist die Aufnahme der Sequenzen 1 und 2 in ein Sequenzprotokoll erforderlich.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenzen 1 und 2 müssen mit separaten Sequenzkennzahlen dargestellt werden:

DGSAKKKK (SEQ ID NO: 32)

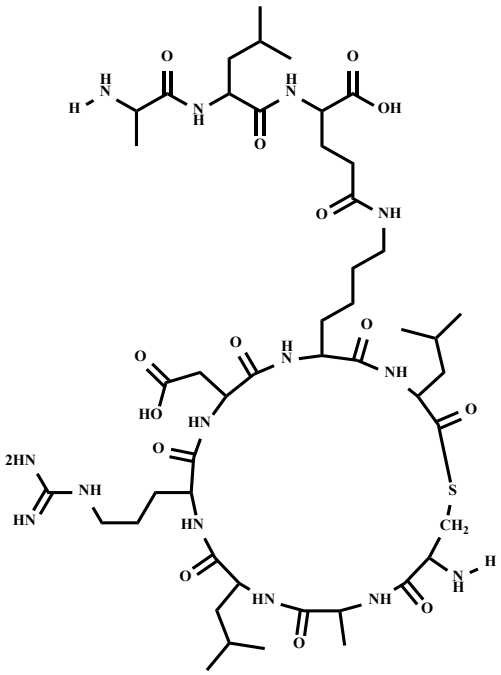
AASHG (SEQ ID NO: 33)

Die Sequenz DGSAKKKK muss annotiert werden, um anzuzeigen, dass es sich bei dem Lysin in Position 5 um eine modifizierte Aminosäure handelt. Dabei ist der Merkmalschlüssel "SITE" in Verbindung mit dem Qualifier "note" zu verwenden, um zu beschreiben, dass die Seitenkette des Lysins über eine Amidbindung mit einer anderen Sequenz verbunden ist. Die Sequenz AASHG sollte annotiert werden, um anzuzeigen, dass das Glycin in Position 5 mit einer anderen Sequenz verbunden ist. Dabei ist der Merkmalschlüssel "SITE" in Verbindung mit dem Qualifier "note" zu verwenden.

Relevante Absätze im Standard ST.26: 7(b), 26, 29, 30 und 31

Beispiel 7(b)-4: Zyklisches Peptid, das eine verzweigte Aminosäuresequenz enthält

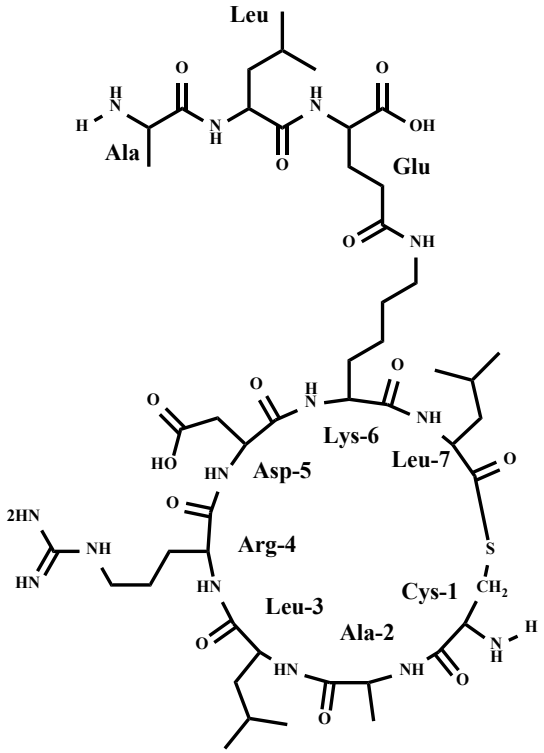
In einer Patentanmeldung wird folgende Struktur offenbart:



Das Cystein und Leucin in der ringförmigen Struktur sind durch die Seitenkette des Cys und den Carboxy-Terminus des Leu miteinander verbunden.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

Bei der abgebildeten Struktur handelt es sich um eine verzweigte ringförmige Aminosäuresequenz, die folgende Aminosäuren enthält:



Da die Seitenkette des Cys und der Carboxy-Terminus des Leu an der Cyclisierung beteiligt sind, befindet sich der N-Terminus des zyklischen Peptids an Cys-1.

JA – die zyklische Region des Peptids

Nach ST.26 Absatz 7(b) muss eine lineare Region einer verzweigten Sequenz, die vier oder mehr spezifisch definierte Aminosäuren enthält, die ein einziges Peptid-Rückgrat bilden, in ein Sequenzprotokoll aufgenommen werden. Im obigen Beispiel enthält die zyklische Region des verzweigten Peptids mehr als vier Aminosäuren und muss daher in ein Sequenzprotokoll aufgenommen werden.

NEIN – der Tripeptid-Zweig des Peptids

Der Tripeptid-Zweig Ala-Leu-Glu muss nicht im Sequenzprotokoll aufgeführt werden.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

NEIN

Nach Absatz 8 darf ein Sequenzprotokoll keine Sequenzen enthalten, die aus weniger als vier spezifisch definierten Aminosäuren bestehen.

Der Tripeptid-Zweig enthält nur drei spezifisch definierte Aminosäuren und darf daher nicht als separate Sequenz mit eigener Sequenzkennzahl in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Dieses Beispiel dient zur Veranschaulichung eines zyklischen Peptids, dessen Ring jedoch nicht ausschließlich aus Aminosäureresten in Peptidbindungen besteht, wie in Absatz 25 beschrieben. Da die Cyclisierung der Aminosäuresequenz durch die Seitenkette des Cysteins (Cys) und den Carboxy-Terminus des Leucins (Leu) erfolgt, muss dem Cystein innerhalb der zyklischen Region des Peptids die Position 1 zugewiesen werden. Dementsprechend muss die Sequenz wie folgt dargestellt werden:

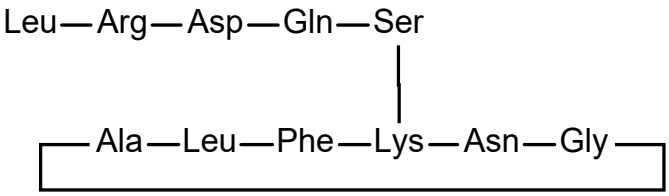
CALRDKL (SEQ ID NO: 90)

Wie in der obigen Abbildung gezeigt, wird die Aminosäuresequenz durch eine Thioesterbindung zwischen der Cystein-Seitenkette und dem Carboxy-Terminus des Leucins cyclisiert. Zur Beschreibung des modifizierten Cysteins, das die ketteninterne Quervernetzung mit dem Leucin bildet, muss der Merkmalschlüssel "SITE" verwendet werden. Das Element für die Lage des Merkmals ("INSDFeature_location") enthält die Restenummern der quervernetzten Aminosäuren im Format "x..y", d. h. "1..7". Im obligatorischen Qualifier "note" sollte die Art der Bindung angegeben werden, z. B. "cysteine leucine thioester (Cys-Leu)" ("Cystein-Leucin-Thioester (Cys-Leu)"), um anzuzeigen, dass Cys-1 und Leu-7 über eine Thioesterbindung verbunden sind. Des Weiteren muss das Lysin in Position 6 annotiert werden, um anzuzeigen, dass es modifiziert wurde. Dabei ist der Merkmalschlüssel "SITE" in Verbindung mit dem obligatorischen Qualifier "note" zu verwenden, wobei im Qualifier-Wert beschrieben wird, dass die Lysin-Seitenkette das Tripeptid ALE bindet.

Relevante Absätze im Standard ST.26: 7(b), 8, 25, 26, 29, 30, 31, 66(c) und 70

Beispiel 7(b)-5: Zyklisches Peptid, das eine verzweigte Aminosäuresequenz enthält

In einer Patentanmeldung wird das folgende verzweigte zyklische Peptid offenbart:

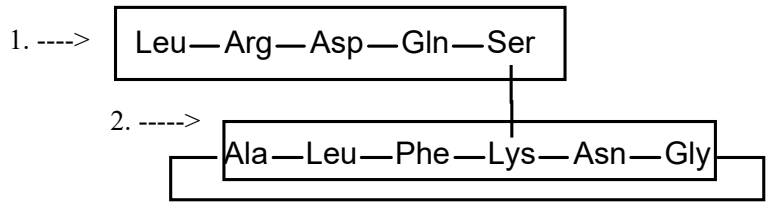


Ser und Lys sind durch eine Amidbindung zwischen dem Carboxy-Terminus des Serins und dem Amin in der Seitenkette des Lys verbunden.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Nach Absatz 7(b) muss jede Sequenz, die vier oder mehr spezifisch definierte Aminosäuren enthält und als lineare Region einer verzweigten Sequenz dargestellt werden kann, in ein Sequenzprotokoll aufgenommen werden. Im obigen Beispiel enthält das Peptid eine zyklische Region, in der die Aminosäuren durch Peptidbindungen verbunden sind, und eine verzweigte Region, die an die Seitenkette des Lys in der zyklischen Region bindet. Die Regionen dieses verzweigten Peptids, die als linear dargestellt werden können und vier oder mehr spezifisch definierte Aminosäuren enthalten, sind:



Nach ST.26 ist die Aufnahme der Sequenzen 1 und 2 dieses verzweigten zyklischen Peptids in ein Sequenzprotokoll erforderlich. Dabei müssen beide Sequenzen eine eigene Sequenzkennzahl erhalten.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Sequenz 1 muss wie folgt dargestellt werden:

LRDQS (SEQ ID NO: 91)

Die Sequenz 1 kann mit dem Merkmalschlüssel "SITE" in Verbindung mit dem Qualifier "note" annotiert werden, um zu beschreiben, dass das Serin in Position 5 über eine Amidbindung zwischen Ser und einer Seitenkette eines Lys in der anderen Sequenz mit dieser anderen Sequenz verbunden ist.

Sequenz 2 ist ein zyklisches Peptid. Nach Absatz 25 muss der Anmelder, wenn eine Aminosäuresequenz ringförmig ist und keinen Amino- und Carboxy-Terminus enthält, auswählen, welchem Aminosäurerest die Positionsnummer 1 zugewiesen wird. Dementsprechend kann die Sequenz wie folgt dargestellt werden:

ALFKNG (SEQ ID NO: 92)

Alternativ kann einer beliebigen anderen Aminosäure in der Sequenz die Restenummer 1 zugewiesen werden. Die Sequenz ALFKNG muss mit dem Merkmalschlüssel "SITE" in Verbindung mit dem Qualifier "note" näher beschrieben werden, um anzuzeigen, dass die Seitenkette des Lys in der Resteposition 4 über eine Amidbindung mit einer anderen Sequenz verbunden ist. Diese Seitenkettenbindung modifiziert das Lys, und nach ST.26 Absatz 30 muss eine modifizierte Aminosäure in der Merkmaltabelle näher beschrieben werden. Außerdem sollte durch einen Merkmalschlüssel "REGION" und einen Qualifier "note" angezeigt werden, dass das Peptid ALFKNG ringförmig ist.

Relevante Absätze im Standard ST.26: 7(b), 25, 26, 30 und 31

Absatz 11(a) – Doppelsträngige Nukleotidsequenz – vollständig komplementär

Beispiel 11(a)-1: Doppelsträngige Nukleotidsequenz – gleiche Länge

In einer Patentanmeldung wird die folgende doppelsträngige DNA-Sequenz beschrieben:

3’-CCGGTTAACGCTA-5’

5’-GGCCAATTGCGAT-3’

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Jede durch Aufzählung dargestellte Nukleotidsequenz weist mehr als zehn spezifisch definierte Nukleotide auf. Mindestens ein Strang muss in das Sequenzprotokoll aufgenommen werden, da die beiden Stränge dieser doppelsträngigen Nukleotidsequenz vollständig komplementär sind.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

JA

Zwar muss nur die Sequenz eines Strangs in das Sequenzprotokoll aufgenommen werden, doch es dürfen auch die Sequenzen beider Stränge mit jeweils eigener Sequenzkennzahl aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die doppelsträngige DNA-Sequenz muss entweder als eine einzelne Sequenz oder als zwei separate Sequenzen dargestellt werden. Jede in das Sequenzprotokoll aufgenommene Sequenz muss in 5’-3’-Richtung dargestellt und mit einer eigenen Sequenzkennzahl versehen werden.

atcgcaattggcc (oberer Strang) (SEQ ID NO: 34)

und/oder

ggccaattgcgat (unterer Strang) (SEQ ID NO: 35)

Relevante Absätze im Standard ST.26: 7(a), **11(a)** und 13

Absatz 11(b) – Doppelsträngige Nukleotidsequenz – nicht vollständig komplementär

Beispiel 11(b)-1: Doppelsträngige Nukleotidsequenz – verschiedene Längen

Eine Patentanmeldung enthält die folgende Zeichnung und Beschriftung:

```
5′-tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc-3′
                      ||||||||||||||||
                      gggtaactgantccgc
```

Die humane Genpromotor-Region ABC1 (oberer Strang) wird von einer PNA-Sonde (unterer Strang) gebunden, wobei "n" in der PNA-Sonde eine universelle PNA-Base ist, die aus der Gruppe bestehend aus 5-Nitroindol und 3-Nitroindol ausgewählt wurde.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA – die Promotor-Region ABC1 (oberer Strang)

Der obere Strang weist mehr als zehn durch Aufzählung darstellte und "spezifisch definierte" Nukleotide auf und muss in ein Sequenzprotokoll aufgenommen werden.

JA – die PNA-Sonde (unterer Strang)

Der untere Strang muss ebenfalls in das Sequenzprotokoll aufgenommen werden, und zwar mit einer eigenen Sequenzkennzahl, da die beiden Stränge nicht vollständig komplementär sind. Die einzelnen Reste, die eine PNA bzw. Peptidnukleinsäure ("peptide nucleic acid") bilden, werden gemäß ST.26 Absatz 3(g) als Nukleotide betrachtet. Folglich weist der untere Strang mehr als zehn durch Aufzählung dargestellte und "spezifisch definierte" Nukleotide auf und muss in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Der obere Strang muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc (SEQ ID NO: 36)

Der untere Strang ist eine Peptidnukleinsäure und weist daher kein 3'- und 5'-Ende auf. Gemäß Absatz 11 muss er "in einer Richtung von links nach rechts, mit der die 5'-3'-Richtung nachgeahmt wird" in ein Sequenzprotokoll aufgenommen werden. Daher muss er wie folgt in ein Sequenzprotokoll aufgenommen werden:

cgcctnagtcaatggg (SEQ ID NO: 37)

Der Qualifier "organism" des Merkmalschlüssels "source" muss den Wert „synthetic construct" ("synthetisches Konstrukt") und der obligatorische Qualifier „mol_type" den Wert „other DNA" erhalten. Der untere Strang muss in einer Merkmaltabelle mit dem Merkmalschlüssel "modified_base" und dem obligatorischen Qualifier "mod_base" mit der Abkürzung "OTHER" beschrieben werden. In einem "note"-Qualifier muss der vollständige, ungekürzte Name der modifizierten Nukleotide angegeben werden, z. B. „N-(2-aminoethyl) glycine nucleosides" ("N-(2-Aminoethyl)glycin-Nukleoside").

Der "n"-Rest muss in einer Merkmaltabelle mit dem Merkmalschlüssel "modified_base" und dem obligatorischen Qualifier "mod_base" mit der Abkürzung "OTHER" näher beschrieben werden. In einem "note"-Qualifier muss der vollständige, ungekürzte Name des modifizierten Nukleotids angegeben werden: "N-(2-aminoethyl) glycin 5-nitroindole or N-(2-aminoethyl) glycin 3-nitroindole" ("(N-(2-Aminoethyl)glycin-5-Nitroindol oder N-(2-Aminoethyl)glycin-3-Nitroindol").

Relevante Absätze im Standard ST.26: 3(g), 7(a), **11(b)**, 17 und 18

Beispiel 11(b)-2: Doppelsträngige Nukleotidsequenz –Segment ohne Basenpaarung

In einer Patentanmeldung wird die folgende doppelsträngige DNA-Sequenz beschrieben:

```
3’-CCGGTTAGCTTATACGCTAGGGCTA-5’
   |||||||      ||||||||||||
5’-GGCCAATATGGCTTGCGATCCCGAT-3’
```

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Jeder Strang der durch Aufzählung dargestellten doppelsträngigen Nukleotidsequenz weist mehr als zehn spezifisch definierte Nukleotide auf. Beide Stränge müssen in das Sequenzprotokoll aufgenommen werden, und zwar jeweils mit einer eigenen Sequenzkennzahl, da die beiden Stränge nicht vollständig komplementär sind.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz jedes Strangs muss in 5’-3’-Richtung dargestellt und mit einer eigenen Sequenzkennzahl versehen werden.

atcgggatcgcatattcgattggcc (oberer Strang) (SEQ ID NO: 38)

und

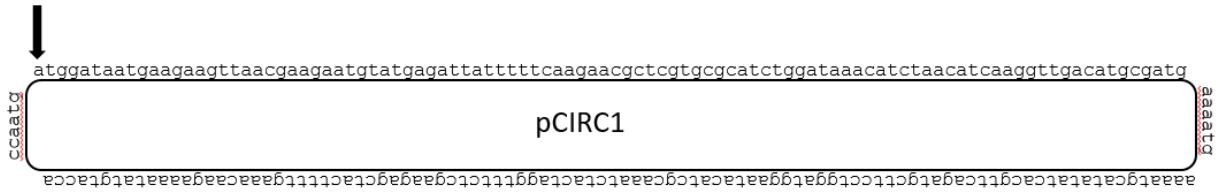
ggccaatatggcttgcgatcccgat (unterer Strang) (SEQ ID NO: 39)

Relevante Absätze im Standard ST.26: 7(a), **11(b)** und 13

Absatz 12 – Ringförmige Nukleotidsequenz

Beispiel 12-1: Ringförmige Nukleotidsequenz

Eine Patentanmeldung enthält die folgende Abbildung, die die DNA-Sequenz des Plasmids pCIRC1 offenbart:



Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die durch Aufzählung dargestellte Nukleotidsequenz weist mehr als zehn spezifisch definierte Nukleotide auf. Daher muss die Sequenz gemäß ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Gemäß ST.26 Absatz 12 muss der Anmelder im Falle ringförmiger Nukleotidsequenzen auswählen, welchem Nukleotid die Restenummer 1 zugewiesen wird. Für die Zwecke dieses Beispiels wird der per Pfeil identifizierte "a"-Rest als Position 1 verwendet. Es kann jedoch jeder Rest als Position 1 verwendet werden. Mit dem per Pfeil angegebenen Rest als Position 1 sollte die Sequenz in folgender Form in das Sequenzprotokoll aufgenommen werden:

atggataatgaagaagttaacgaagaatgtatgagattatttttcaagaacgctcgtgcgcatctggataaacatctaacatcaaggttgacatgcgatgaaaatgaaaatgcatatatcacgttcagatgcttcctggatggaatacatcgcaaatctactaggtttctcgaagagctacttttgaaacaagaaaatatgtaccaccaatg (SEQ ID NO: 98)

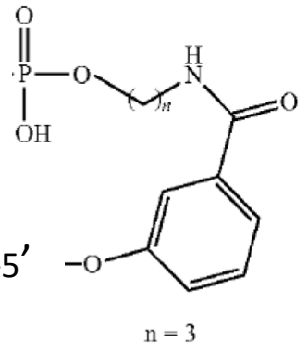
Die Sequenz sollte weiterhin anhand des Merkmalschlüssels "misc_feature" mit der Lage "212^1" näher beschrieben werden. Diese gibt an, dass der letzte Rest in der Sequenz, Position 212, an Rest 1 gebunden ist. In einem "note"-Qualifier muss der Wert angegeben werden, dass das Molekül ringförmig ist.

Relevante Absätze im Standard ST.26: 7(a), **12** und Anhang I, Abschnitt 5, Merkmalschlüssel 5.15

Absatz 14 – Auslegung des Symbols "t" als Uracil in RNA

Beispiel 14-1: Das Symbol "t" repräsentiert Uracil in RNA

Segment A: ccugucgt-3’



Segment B: uaguuguagaggccugucct-5’

n = 3

In einer Patentanmeldung wird folgende Verbindung beschrieben:

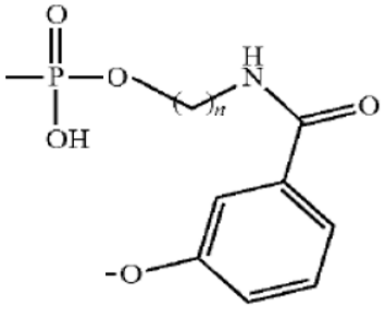
Dabei sind Segment A und Segment B RNA-Sequenzen.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA – Segment B

NEIN – Segment A

Die durch Aufzählung dargestellte Sequenz enthält zwei Segmente spezifisch definierter Nukleotide, die durch die folgende "Linker"-Struktur getrennt sind.



n = 3

Die Linker-Struktur ist kein Nukleotid gemäß Absatz 3(g); daher muss jedes Segment als separate Sequenz betrachtet werden. Segment B enthält mehr als zehn spezifisch definierte Nukleotide und muss somit nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden. Segment A enthält nur acht spezifisch definierte Nukleotide und muss daher nicht in ein Sequenzprotokoll aufgenommen werden.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

NEIN

Segment A enthält weniger als zehn spezifisch definierte Nukleotide und darf somit nach ST.26 Absatz 8 nicht in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Segment B ist ein RNA-Molekül; daher muss das Element "INSDSeq_moltype" gleich "RNA" sein. Das Symbol "u" darf in einem Sequenzprotokoll nicht verwendet werden, um Uracil in einem RNA-Molekül darzustellen. Nach Absatz 14 wird das Symbol "t" in RNA als Uracil ausgelegt. Somit muss Segment B wie folgt in das Sequenzprotokoll aufgenommen werden:

tcctgtccggagatgttgat (SEQ ID NO: 40)

Thymin in RNA gilt als modifiziertes Nukleotid, d. h. modifiziertes Uracil, und muss in der Sequenz als "t" dargestellt und in einer Merkmaltabelle näher beschrieben werden. Dementsprechend muss das Thymin an Position 1 mit dem Merkmalschlüssel "modified_base", dem Qualifier "mod_base" mit "OTHER" als Qualifier-Wert und einem "note"-Qualifier mit "thymine" ("Thymin") als Qualifier-Wert näher beschrieben werden.

Das Thymin, d. h. das modifizierte Uracil, in Position 1 sollte außerdem in einer Merkmaltabelle mit dem Merkmalschlüssel "misc_feature" und einem "note"-Qualifier mit einem Wert wie "The 5' phosphate of the thymidine is attached through the linker 3-hydroxybenzamido-N-propyl-3-phosphate to another nucleotide sequence." (Das 5'-Phosphat des Thymidins ist über den Linker 3-Hydroxybenzamid-N-propyl-3-phosphat an eine andere Nukleotidsequenz gebunden.") näher beschrieben werden. Sofern praktikabel, kann die andere Sequenz direkt als Wert im "note"-Qualifier angegeben werden.

Relevante Absätze im Standard ST.26: 3(g), 7(a), 8, 13, **14**, 19 und 54

Absatz 27 – Das restriktivste Mehrdeutigkeitssymbol sollte verwendet werden

Beispiel 27-1: Kurzformel für eine Aminosäuresequenz

$(GGGz)_2$

Dabei ist z jede beliebige Aminosäure.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die Sequenz wird als Formel offenbart. $(GGGz)_2$ ist einfach eine verkürzte Darstellung der Sequenz GGGzGGGz. Üblicherweise wird eine Sequenz zunächst erweitert, und danach wird die Definition einer beliebigen Variablen, d. h. "z", festgelegt.

In der Sequenz wird das nicht konventionelle Symbol "z" verwendet. Die Definition von "z" ist der Erläuterung zur Sequenz in der Offenbarung zu entnehmen. Dort wird dieses Symbol als jede beliebige Aminosäure definiert (siehe Einführung in dieses Dokument). Das Beispiel enthält keine Einschränkung für "z", z. B., dass es bei jedem Vorkommen die gleiche Aminosäure darstellen muss.

Das Peptid im Beispiel weist acht durch Aufzählung dargestellte Aminosäuren auf, von denen sechs spezifisch definierte Glycinreste und die übrigen beiden die "z"-Variable sind, die in der Sequenz mit dem konventionellen Symbol "X" dargestellt werden sollte. Nach ST.26 Absatz 7(b) muss die Sequenz als einzelne Sequenz mit einer einzigen Sequenzkennzahl in das Sequenzprotokoll aufgenommen werden.

Beachten Sie, dass die Sequenz von Absatz 7(b) erfasst wird, obwohl die durch Aufzählung dargestellten und spezifisch definierten Reste nicht zusammenhängend sind.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

In der Sequenz wird das nicht konventionelle Symbol "z" verwendet, das laut der Offenbarung jede beliebige Aminosäure ist. Das konventionelle Symbol für die Darstellung einer beliebigen Aminosäure ist "X". Daher muss die Sequenz als einzelne erweiterte Sequenz wie folgt dargestellt werden:

GGGXGGGX (SEQ ID NO: 41)

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" in diesem Beispiel "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") versehen werden.

Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

Des Weiteren wird in dem Beispiel nicht offenbart, dass "z" in beiden Positionen der erweiterten Sequenz die gleiche Aminosäure ist. Wenn "z" jedoch in beiden Positionen als die gleiche Aminosäure offenbart wird, sollten ein Merkmalschlüssel "VARIANT" und ein "note"-Qualifier angegeben werden, aus dem hervorgeht, dass "X" in Position 4 und 8 jede beliebige Aminosäure sein kann, solange es in beiden Positionen die gleiche ist.

Relevante Absätze im Standard ST.26: 3(c), 7(b) und **27**

Beispiel 27-2: Kurzformel – weniger als vier spezifisch definierte Aminosäuren

Ein Peptid mit der Formel $(Gly\text{-}Gly\text{-}Gly\text{-}z)_n$

In der Offenbarung heißt es weiter, dass z jede beliebige Aminosäure ist und

(i) die Variable n eine beliebige Länge darstellt oder

(ii) die Variable n für 2–100, vorzugsweise 3 steht.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

NEIN

Die Betrachtung der beiden offenbarten Ausführungsformen (i) und (ii) des durch Aufzählung dargestellten Peptids der Formel zeigt, dass "n" "eine beliebige Länge" sein kann; daher ist die umfassendste Ausführungsform von "n" unbestimmt. Da "n" unbestimmt ist, kann das in der Formel angegebene Peptid nicht auf eine bestimmte Länge erweitert werden, sodass die nicht erweiterte Formel berücksichtigt werden muss.

Das durch Aufzählung dargestellte Peptid in der nicht erweiterten Formel ("n" = 1) enthält drei spezifisch definierte Aminosäuren, von denen jede Gly ist, und das Symbol "z". Konventionell ist "Z" das Symbol für "Glutamin oder Glutaminsäure", doch in dem Beispiel ist "z" als "jede beliebige Aminosäure" definiert (siehe Einführung in dieses Dokument). Gemäß ST.26 wird eine Aminosäure, die nicht spezifisch definiert ist, durch "X" dargestellt. Dieser Analyse zufolge enthält das durch Aufzählung dargestellte Peptid, d. h. GGGX, keine vier spezifisch definierten Aminosäuren. Aus diesem Grund ist nach ST.26, Absatz 7(b) keine Aufnahme erforderlich, obwohl "n" in einigen Ausführungsformen auch als spezifischer numerischer Wert definiert ist.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

JA

Im Beispiel wird ein spezifischer numerischer Wert für die Variable "n" angegeben, d. h. ein unterer Grenzwert von 2, ein oberer Grenzwert von 100 und ein exakter Wert von 3. Jede Sequenz, die mindestens vier spezifisch definierte Aminosäuren enthält, kann in das Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Bevorzugt wird eine Sequenz mit 100 Kopien von GGGX (SEQ ID NO: 42). In einer weiteren Annotierung sollte darauf hingewiesen werden, dass bis zu 98 Kopien von GGGX deletiert werden könnten. Die Aufnahme von weiteren spezifischen Ausführungsformen, die ein wesentlicher Bestandteil der Erfindung sind, wird nachdrücklich empfohlen.

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" in der SEQ ID NO 42 "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") versehen werden.

Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 3(c), 7(b), 26 und **27**

Beispiel 27-3: Kurzformel – vier oder mehr spezifisch definierte Aminosäuren

Ein Peptid mit der Formel $(Gly\text{-}Gly\text{-}Gly\text{-}z)_n$

Dabei ist z jede beliebige Aminosäure, und die Variable n ist 2–100, vorzugsweise 3.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das durch Aufzählung dargestellte Peptid der Formel enthält drei spezifisch definierte Aminosäuren, von denen jede Gly ist, und das Symbol "z". Konventionell ist "Z" das Symbol für "Glutamin oder Glutaminsäure", doch in der Beschreibung zu diesem Beispiel ist "z" als "jede beliebige Aminosäure" definiert (siehe Einführung in dieses Dokument). Gemäß ST.26 wird eine Aminosäure, die nicht spezifisch definiert ist, durch "X" dargestellt. Dieser Analyse zufolge enthält das durch Aufzählung dargestellte repetitive Peptid keine vier spezifisch definierten Aminosäuren. Im Beispiel wird ein spezifischer numerischer Wert für die Variable "n" angegeben, d. h. ein unterer Grenzwert von 2 und ein oberer Grenzwert von 100. Folglich offenbart das Beispiel ein Peptid mit mindestens sechs spezifisch definierten Aminosäuren in der Sequenz GGGzGGGz, die nach ST.26 in ein Sequenzprotokoll aufgenommen werden muss.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Da "z" für jede beliebige Aminosäure steht, wird für die Darstellung der vierten und achten Aminosäure das konventionelle Symbol "X" verwendet.

Nach ST.26 ist die Aufnahme in ein Sequenzprotokoll nur für die einzelne Sequenz erforderlich, die durch Aufzählung der Reste dargestellt wurde. Daher muss mindestens eine Sequenz, die 2, 3 oder 100 Kopien von GGGX enthält, in das Sequenzprotokoll aufgenommen werden. Bevorzugt wird jedoch die umfassendste Sequenz mit 100 Kopien von GGGX (SEQ ID NO: 42) (siehe Einführung in dieses Dokument). Im letzteren Fall sollte in einer weiteren Annotierung darauf hingewiesen werden, dass bis zu 98 Kopien von GGGX deletiert werden könnten. Die Aufnahme von zwei zusätzlichen Sequenzen mit 2 bzw. 3 Kopien von GGGX (SEQ ID NO: 44-45) wird dringend empfohlen.

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" diesem Beispiel "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") versehen werden.

Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

Des Weiteren wird in dem Beispiel nicht offenbart, dass die Variable "z" an beiden Positionen der erweiterten Sequenz die gleiche Aminosäure ist. Wenn "z" jedoch in allen Positionen als die gleiche Aminosäure offenbart wird, sollten ein Merkmalschlüssel "VARIANT" und ein "note"-Qualifier angegeben werden, aus dem hervorgeht, dass "X" in allen Positionen jede beliebige Aminosäure sein kann, solange es überall die gleiche ist.

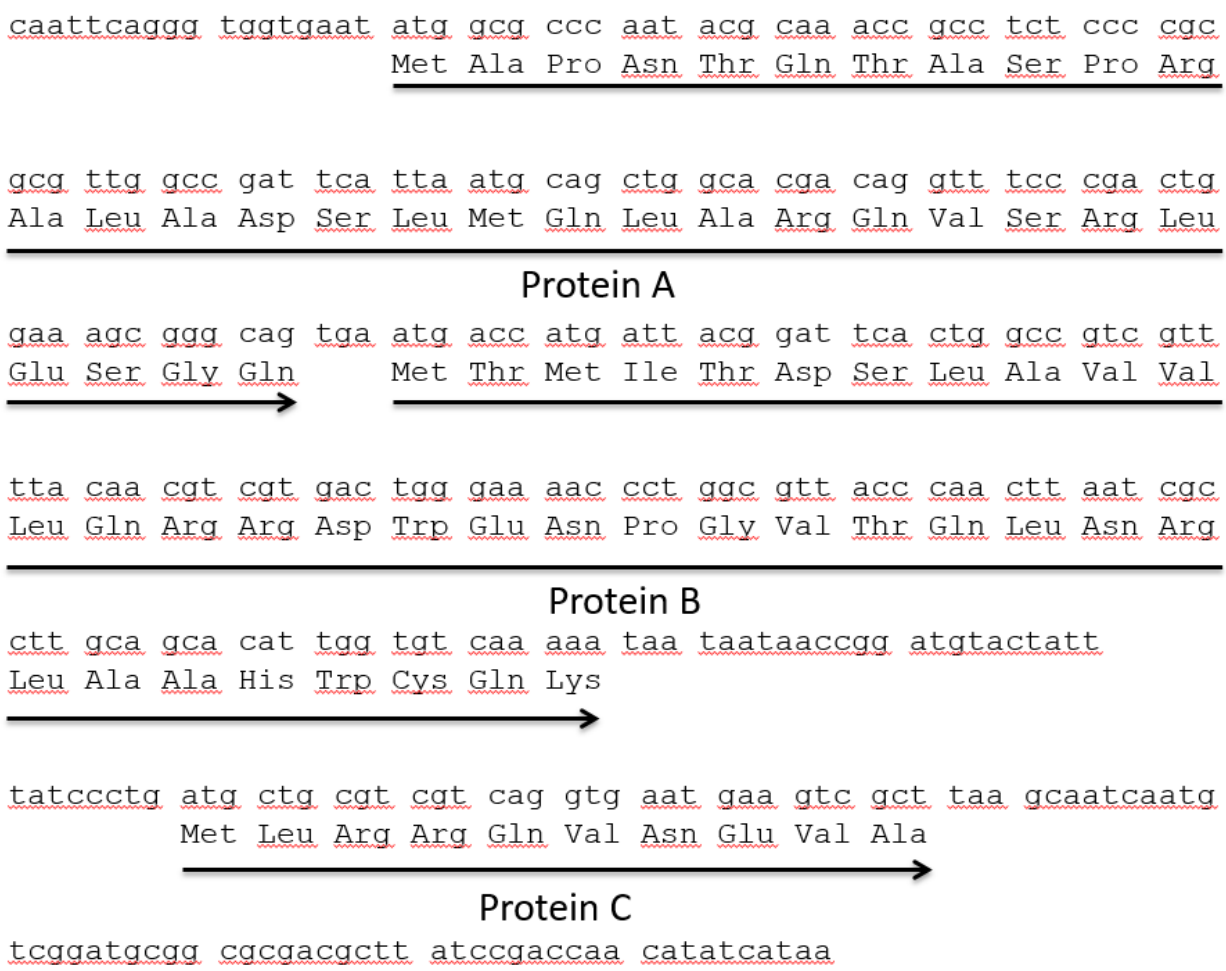
ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 3(c), 7(b), 26 und **27**

Absatz 28 – Aminosäuresequenzen, die durch interne Terminatorsymbole getrennt sind

Beispiel 28-1: Codierende Nukleotidsequenz und codierte Aminosäuresequenz

In einer Patentanmeldung werden folgende Sequenzen beschrieben:



Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der Anmeldung wird eine Nukleotidsequenz mit Terminationscodons beschrieben, die für drei unterschiedene Aminosäuresequenzen codiert.

Die durch Aufzählung dargestellte Nukleotidsequenz enthält mehr als zehn spezifisch definierte Nukleotide und muss als einzelne Sequenz in ein Sequenzprotokoll aufgenommen werden.

Codierte Aminosäuresequenzen, die durch ein internes Terminatorsymbol wie ein Leerzeichen getrennt sind, müssen nach Absatz 28 als separate Sequenzen aufgenommen werden. Da "Protein A", "Protein B" und "Protein C" jeweils vier oder mehr spezifisch definierte Aminosäuren enthalten, müssen sie nach ST.26 Absatz 7(b) alle in ein Sequenzprotokoll aufgenommen und mit einer jeweils eigenen Sequenzkennzahl versehen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Nukleotidsequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

caattcagggtggtgaatatggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggccaggcaggtgagcaggctggaaagcg ggcagtgaatgaccatgattacggattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacattggtgtcaaa aataataataaccggatgtactatttatccctgatgctgcgtcgtcaggtgaatgaagtcgcttaagcaatcaatgtcggatgcggcgcgacgcttatccgaccaacatat cataa (SEQ ID NO: 46)

Die Nukleotidsequenz sollte mit einem "CDS"-Merkmalschlüssel für jedes der drei Proteine näher beschrieben werden, und mit dem Element "INSDFeature_location" muss die Lage jeder codierenden Sequenz, einschließlich des Stopcodons, angegeben werden. Darüber hinaus sollte für jeden "CDS"-Merkmalschlüssel der Qualifier "translation" mit der Aminosäuresequenz des Proteins als Qualifier-Wert angegeben werden. Die Tabelle für den genetischen Code für die Translation (siehe Anhang I, Abschnitt 9, Tabelle 7) wird in der Anmeldung nicht offenbart. Wenn die Tabelle für den Standardcode gilt, ist der Qualifier "transl_table" nicht erforderlich; wenn jedoch eine andere Tabelle für den genetischen Code gilt, muss für den Qualifier "transl_table" der entsprechende Qualifier-Wert aus Tabelle 7 angegeben werden. Schließlich muss der Qualifier "protein_id" mit dem Qualifier-Wert angegeben werden, aus dem die Sequenzkennzahl jeder der translatierten Aminosäuresequenzen hervorgeht.

Die Aminosäuresequenzen müssen als separate Sequenzen aufgenommen werden, wobei jede mit einer eigenen Sequenzkennzahl versehen wird:

MAPNTQTASPRALADSLMQLARQVSRLESGQ (SEQ ID NO: 47)

MTMITDSLAVVLQRRDWENPGVTQLNRLAAHWCQK (DEQ ID NO: 48)

MLRRQVNEVA (SEQ ID NO: 49)

HINWEIS: Das Beispiel 90-1, "Aminosäuresequenz, die durch eine codierende Sequenz mit Introns codiert wird", veranschaulicht die Darstellung einer translatierten Aminosäuresequenz als einzelne Sequenz.

Relevante Absätze im Standard ST.26: 7, 26, **28**, 57, 89–92

Absatz 29 – Darstellung einer "sonstigen" ("other") Aminosäure

Beispiel 29-1: Das restriktivste Mehrdeutigkeitssymbol für eine "sonstige" ("other") Aminosäure

In einer Patentanmeldung wird folgende Sequenz beschrieben:

Ala-Hse-X_1-X_2-X_3-X_4-Tyr-Leu-Gly-Ser

Wobei X_1 = Ala oder Gly,

X_2 = Ala oder Gly,

X_3 = Ala oder Gly,

X_4 = Ala oder Gly und

Hse = Homoserin

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das durch Aufzählung dargestellte Peptid enthält fünf spezifisch definierte Aminosäuren.Das Symbol "X" wird konventionell verwendet, um alternativ zwei Aminosäuren darzustellen (siehe Einführung in dieses Dokument).

Da fünf spezifisch definierte Aminosäuren vorliegen, d. h. Ala, Tyr, Leu, Gly und Ser, muss die Sequenz nach ST.26 Absatz 7(b) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Nach Absatz 29 muss jede "sonstige" ("other") Aminosäure durch das Symbol "X" dargestellt werden. In diesem Beispiel enthält die Sequenz in Position 2 die Aminosäure Hse, die in Anhang I, Abschnitt 3, Tabelle 3 nicht aufgeführt wird. Folglich handelt es sich bei Hse um eine "sonstige" ("other") Aminosäure, die durch das Symbol "X" dargestellt werden muss.

X_1-X_4 sind Variantenpositionen, von denen jede A oder G sein kann. Das restriktivste Mehrdeutigkeitssymbol für die Alternativen A oder G ist "X". Daher kann die Sequenz wie folgt dargestellt werden:

AXXXXXYLGS (SEQ ID NO: 50)

Es wird nachdrücklich empfohlen, alle spezifischen Sequenzen aufzunehmen, die für die Offenbarung oder die Ansprüche der Erfindung wesentlich sind, wie in der Einführung in dieses Dokument erläutert.

Da die Aminosäure Hse nicht in Anhang I, Abschnitt 4, Tabelle 4 enthalten ist, müssen nach ST.26 Absatz 30 ein Merkmalschlüssel "SITE" und ein "note"-Qualifier mit dem vollständigen, ungekürzten Namen von Homoserin angegeben werden.

Da X_1-X_4 eine Alternative von nur zwei Aminosäuren darstellt, ist nach Absatz 27 eine nähere Beschreibung erforderlich. In Absatz 96 wird angegeben, dass der Merkmalschlüssel "VARIANT" mit dem Qualifier "note" und dem Qualifier-Wert "A or G" verwendet werden soll. Da diese Positionen nebeneinander liegen und die gleiche Beschreibung haben, können sie nach ST.26 Absatz 34 mit der Syntax "3..6" als Lagedeskriptor im Element "INSDFeature_location" gemeinsam beschrieben werden.

Relevante Absätze im Standard ST.26: 3(a), 7(b), 25–27, **29**, 30, 34, 66, 70, 71 und 96–97

Beispiel 29-2: Verwendung der entsprechenden nicht modifizierten Aminosäure

In einer Patentanmeldung wird folgende Sequenz beschrieben:

Ala-Hyl-Tyr-Leu-Gly-Ser-Nle-Val-Ser-5ALA

Dabei ist Hyl = Hydroxylysin (posttranslationale Modifikation von Lysin), Nle = Norleucin und 5ALA = δ-Aminolävulinsäure

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das durch Aufzählung dargestellte Peptid enthält mehr als vier spezifisch definierte Aminosäuren, daher muss die Sequenz in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Bei dem Hydroxylysin an Position 2, dem Norleucin an Position 7 und der δ-Aminolävulinsäure an Position 10 handelt es sich jeweils um "modifizierte Aminosäuren". Zunächst muss jede modifizierte Aminosäure daraufhin überprüft werden, ob sie in der Sequenz durch die entsprechende nicht modifizierte Aminosäure oder durch die Variable "X" dargestellt werden sollte. Laut Absatz 29 sollten modifizierte Aminosäuren "in der Sequenz nach Möglichkeit als die entsprechenden nicht modifizierten Aminosäuren dargestellt werden".

Es liegt im Ermessen des Anmelders, ob eine modifizierte Aminosäure durch den entsprechenden nicht modifizierten Rest oder die Variable "X" dargestellt wird. Dabei sollten allerdings die folgenden Vorgaben berücksichtigt werden: Wenn eine Aminosäure durch Addition einer funktionellen Gruppe, z. B. durch Methylierung oder Acetylierung, modifiziert wird und die Grundstruktur der entsprechenden nicht modifizierten Aminosäure im Wesentlichen unverändert bleibt, wird eine Darstellung durch die nicht modifizierte Aminosäure empfohlen. Wenn sich die modifizierte Aminosäure in ihrer Struktur stark von der entsprechenden nicht modifizierten Aminosäure unterscheidet, wird eine Darstellung durch "X" empfohlen.

Die Struktur von Hydroxylysin ist mit derjenigen von Lysin nahezu identisch, lediglich der dritte Kohlenstoff in der Seitenkette wird durch eine Hydroxylgruppe modifiziert. Da die Grundstruktur des entsprechenden nicht modifizierten Lysinrests unverändert bleibt, sollte das Hydroxylysin in der Sequenz durch Lysin ("K") und nicht durch "X" dargestellt werden.

Norleucin ist ein Isomer von Leucin. Die R-Gruppe von Leucin ist eine Kette aus vier Kohlenstoffatomen, die am zweiten Kohlenstoff verzweigt ist. Norleucin hat ebenfalls eine Seitenkette aus vier Kohlenstoffatomen, doch diese ist linear und nicht verzweigt. Daher handelt es sich bei Norleucin nicht einfach um das Ergebnis einer Modifikation, die einem Leucin hinzugefügt wurde, sondern um eine völlig andere (wenn auch verwandte) Struktur. Daher wird empfohlen, Norleucin in einem Sequenzprotokoll durch ein "X" darzustellen.

δ-Aminolävulinsäure weist keine strukturelle Ähnlichkeit mit einer der in Anhang I, Tabelle 3 aufgeführten Aminosäuren auf. Daher wird empfohlen, δ-Aminolävulinsäure in einem Sequenzprotokoll durch ein "X" darzustellen.

Entsprechend sollte die Sequenz in folgender Form in ein Sequenzprotokoll aufgenommen werden:

AKYLGSXVSX (SEQ ID NO: 51)

Gemäß Absatz 30 muss jede modifizierte Aminosäure näher beschrieben werden.

Hydroxylysin ist eine posttranslationale Modifikation von Lysin. Daher muss es mit dem Merkmalschlüssel "MOD_RES" in Verbindung mit einem Qualifier "note", aus dem die Modifikation hervorgeht, beschrieben werden. Es ist zu beachten, dass "Hydroxylysin" in der "Liste der modifizierten Aminosäuren" in Anhang I, Abschnitt 4, Tabelle 4 aufgeführt wird. Daher kann der Wert des "note"-Qualifiers die Abkürzung "Hyl" anstelle des vollständigen, ungekürzten Namens "Hydroxylysin" enthalten.

Norleucin ist kein posttranslational modifizierter Rest. Daher muss es mit dem Merkmalschlüssel "SITE" in Verbindung mit einem Qualifier "note", aus dem die Modifikation hervorgeht, beschrieben werden. Es ist zu beachten, dass "Norleucin" ebenfalls in Anhang I, Abschnitt 4, Tabelle 4 aufgeführt wird. Daher kann der Wert des "note"-Qualifiers die Abkürzung "Nle" anstelle des vollständigen, ungekürzten Namens "Norleucin" enthalten.

Auch δ-Aminolävulinsäure ist kein posttranslational modifizierter Rest. Daher muss es mit dem Merkmalschlüssel "SITE" in Verbindung mit einem Qualifier "note", aus dem die Modifikation hervorgeht, beschrieben werden. δ-Aminolävulinsäure ist in Anhang I, Abschnitt 4, Tabelle 4 nicht aufgeführt; daher muss der Qualifier "note" den vollständigen, ungekürzten Namen des modifizierten Rests, “δ-aminolevulinic acid” ("δ-Aminolävulinsäure"), enthalten.

Relevante Absätze im Standard ST.26: 3(a), 3(e), 7(b), 29 und 30

Absatz 30 – Annotation einer modifizierten Aminosäure

Beispiel 30-1 – Merkmalschlüssel "CARBOHYD"

In einer Patentanmeldung wird ein Polypeptid mit einer spezifisch modifizierten Aminosäure beschrieben, die eine glykosylierte Seitenkette enthält, dadurch gekennzeichnet, dass Cys in den Positionen 4 und 15 des Polypeptids eine Disulfidbindung bildet, wie in folgender Sequenz dargestellt:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Asn(asialyloligosaccharide)-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das durch Aufzählung dargestellte Peptid enthält 17 spezifisch definierte Aminosäuren. Es liegen 16 natürliche Aminosäuren vor, wobei die neunte (Asparagin) glykosyliert ist. Daher muss die Sequenz gemäß ST.26 Absatz 7(b) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Laut ST.26 Absatz 29 sollten modifizierte Aminosäuren in der Sequenz nach Möglichkeit als die entsprechenden nicht modifizierten Aminosäuren dargestellt werden.

Daher muss die Sequenz in folgender Form in ein Sequenzprotokoll aufgenommen werden:

LEYCLKRWNETISHCAW (SEQ ID NO: 52)

Die modifizierte Aminosäure muss näher beschrieben werden. Der Merkmalschlüssel "CARBOHYD" sollte zusammen mit dem (obligatorischen) Qualifier "note" verwendet werden, um das Anheften einer Zuckerkette (eines Sialyloligosaccharids) an Asparagin in Position 9 anzuzeigen. Der Qualifier "note" beschreibt die Art der Bindung, z. B. "N-linked" ("N-Bindung"). Der Lagedeskriptor im Element für die Lage des Merkmals ("INSDFeature_location") ist die Restenummer des modifizierten Asparagins.

Darüber hinaus besteht eine Disulfidbindung zwischen den beiden Cys-Resten. Daher sollte zur Beschreibung einer ketteninternen Quervernetzung der Merkmalschlüssel "DISULFID" verwendet werden. Das Element für die Lage des Merkmals ("INSDFeature_location") enthält die Restenummern der verbundenen Cys-Reste im Format "x..y", also "4..15". Der obligatorische "note"-Qualifier sollte die ketteninterne Disulfidbindung beschreiben.

Relevante Absätze im Standard ST.26: 3(a), 7(b), 26, 29, **30**, 66(c), 70 und Anhang I, Abschnitt 7, Merkmalschlüssel 7.4

Beispiel 30-2 – Posttranslational modifizierte Aminosäuren

In einer Patentanmeldung wird das folgende Polypeptid beschrieben:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Glu-Thr-Ile-Ser-His

Dabei wird das Arg an Position 7 posttranslational citrulliniert.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das durch Aufzählung dargestellte Peptid enthält 13 spezifisch definierte Aminosäuren. Daher muss die Sequenz gemäß ST.26 Absatz 7(b) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Laut ST.26 Absatz 29 sollten modifizierte Aminosäuren in der Sequenz nach Möglichkeit als die entsprechenden nicht modifizierten Aminosäuren dargestellt werden.

Daher sollte die Sequenz in folgender Form in ein Sequenzprotokoll aufgenommen werden:

LEYCLKRWETISH (SEQ ID NO: 97)

Dabei wird das Symbol "R" zur Darstellung des Arginin an Position 7 verwendet.

Es muss näher beschrieben werden, dass das Arginin an Position 7 zu Citrullin modifiziert werden kann. Die Citrullinierung von Arginin ist eine posttranslationale Modifikation. Daher sollte der Merkmalschlüssel "MOD_RES" mit dem obligatorischen "note"-Qualifier verwendet werden, um anzugeben, dass das Arginin zu Citrullin deiminiert werden kann. Der Lagedeskriptor im Element für die Lage des Merkmals ("INSDFeature_location") ist die Restenummer des modifizierten Arginins.

Relevante Absätze im Standard ST.26: 3(a), 7(b), **30** und Anhang I, Abschnitt 7, Merkmalschlüssel 7.18

Absatz 36 – Sequenzen, die Regionen mit einer genauen Zahl an zusammenhängenden "n"- oder "X"-Resten enthalten

Beispiel 36-1: Sequenz mit einer Region, die eine bekannte Zahl an "X"-Resten enthält, dargestellt als eine einzelne Sequenz

LL-100-KYMR

Dabei spiegelt das "-100-" zwischen den Aminosäuren Leucin und Lysin eine Region aus 100 Aminosäuren in der Sequenz wider.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Eine Sequenz, die mindestens vier spezifisch definierte Aminosäuren enthält, die durch eine oder mehrere Regionen aus einer definierten Anzahl von "X"-Resten voneinander getrennt sind, muss nach ST.26 Absatz 36 aufgenommen werden.

In der offenbarten Sequenz wird ein nicht konventionelles Symbol verwendet, nämlich "-100-". Die Definition von "-100-" ist der Erläuterung zur Sequenz in der Offenbarung zu entnehmen. Dort wird dieses Symbol als 100 Aminosäuren zwischen Leucin und Lysin definiert (siehe Einführung in dieses Dokument). Folglich handelt es sich bei "-100-" um eine definierte Region von "X"-Resten. Da sechs der 106 Aminosäuren in der Sequenz spezifisch definiert sind, muss die Sequenz nach ST.26 Absatz 7(b) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Das nicht konventionelle Symbol "-100-" wird als 100 "X"-Reste dargestellt (da jedes Symbol, das zur Darstellung einer Aminosäure verwendet wird, nur einem Rest entspricht). Daher muss eine einzelne Sequenz mit einer Länge von 106 Aminosäuren, die zwischen LL und KYMR 100 "X"-Reste enthält, in ein Sequenzprotokoll aufgenommen werden (SEQ ID NO: 53).

Diese Sequenz enthält 100 "X"-Variablen zwischen LL und KYMR. Als Standardwert für "X" ohne weitere Annotation ist in ST.26 eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" vorgesehen (Absatz 27). Wenn für diese 100 "X"-Variablen ein anderer Wert als der Standardwert definiert ist, muss für jede "X"-Variable eine ordnungsgemäße Annotation vorgenommen werden.

Relevante Absätze im Standard ST.26: 7(b), 26, 27 und **36**

Beispiel 36-2: Sequenz mit mehreren Regionen, die eine bekannte Zahl an "X"-Resten enthalten, dargestellt als eine einzelne Sequenz

Lys-z_2-Lys-z_m-Lys-z_3-Lys-z_n-Lys-z_2-Lys

Dabei ist z jede beliebige Aminosäure, m=20, n=19-20, z_2 bedeutet, dass die Lysinpaare durch zwei beliebige Aminosäuren getrennt werden, und z_3 bedeutet, dass die Lysinpaare durch drei beliebige Aminosäuren getrennt werden.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der offenbarten Sequenz wird ein nicht konventionelles Symbol verwendet, nämlich "z". Daher muss die Offenbarung herangezogen werden, um seine Definition zu bestimmen; dort ist "z" als jede beliebige Aminosäure definiert (siehe Einführung in dieses Dokument). Das konventionelle Symbol für die Darstellung einer beliebigen Aminosäure ist "X". Unter Berücksichtigung der vorliegenden "X"-Variablen enthält das Peptid sechs durch Aufzählung dargestellte und spezifisch definierte Lysinreste und muss folglich in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

In der Sequenz wird "z" als nicht konventionelles Symbol verwendet, dessen Definition der Offenbarung zu entnehmen ist. Da "z" dort als jede beliebige Aminosäure definiert ist, ist das konventionelle Symbol "X".

Die bevorzugte und umfassendste Darstellungsform ist (siehe Einführung in dieses Dokument):

KXXKXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 54)

Dabei ist z_n gleich 20 "X", mit einer näheren Beschreibung, dass die "X"-Variable, die der Position 30 entspricht, deletiert werden kann.

Alternativ oder zusätzlich zur oben aufgeführten Darstellung kann die Sequenz wie folgt dargestellt werden:

KXXKXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 55)

Dabei ist z_n gleich 19 "X", mit einer näheren Beschreibung, dass zwischen den Positionen 29 und 30 eine "X"-Variable inseriert werden kann.

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" in der SEQ ID NO 54 und SEQ ID NO 55 "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") versehen werden.

Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

Relevante Absätze im Standard ST.26: 26, 27 und **36**

Beispiel 36-3: Sequenz mit mehreren Regionen, die eine bekannte Zahl an "X"-Resten enthalten, dargestellt als eine einzelne Sequenz

K-z_2-K-z_m-K-z_3-K-z_n-K-z_2-K

Dabei ist z jede beliebige Aminosäure und m=15-25, vorzugsweise 20-22, n=15-25, vorzugsweise 19-20, z_2 bedeutet, dass die Lysinpaare durch zwei beliebige Aminosäuren getrennt werden, und z_3 bedeutet, dass die Lysinpaare durch drei beliebige Aminosäuren getrennt werden.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der Sequenz in diesem Beispiel wird ein nicht konventionelles Symbol verwendet, nämlich "z". Daher muss die zugehörige Offenbarung herangezogen werden, um die Definition von "z" zu bestimmen (siehe Einführung in dieses Dokument). In der Offenbarung wird dieses Symbol als jede beliebige Aminosäure definiert. Das konventionelle Symbol, das zur Darstellung eines als „jede beliebige Aminosäure“ definierten Rests verwendet wird, ist "X". Unter Berücksichtigung der vorliegenden "X"-Reste enthält das Peptid sechs durch Aufzählung dargestellte und spezifisch definierte Lysinreste und muss folglich in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

In der Sequenz wird "z" als nicht konventionelles Symbol verwendet, dessen Definition der Offenbarung zu entnehmen ist. Da "z" dort als jede beliebige Aminosäure definiert ist, ist das konventionelle Symbol "X". Die bevorzugte und umfassendste Darstellungsform ist:

KXXKXXXXXXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 56)

(Dabei ist m=25 und n=25), mit einer näheren Beschreibung, dass in jeder der "z"$_m$- oder "z"$_n$-Regionen bis zu zehn "X"-Reste deletiert werden können.

Es wird nachdrücklich empfohlen, alle spezifischen Sequenzen aufzunehmen, die für die Offenbarung oder die Ansprüche der Erfindung wesentlich sind, wie in der Einführung in dieses Dokument erläutert.

Alternativ kann die Sequenz wie folgt dargestellt werden:

KXXKXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 57)

(Dabei ist m=15 und n=15), mit einer näheren Beschreibung, dass in jeder der "z"$_m$- oder "z"$_n$-Regionen bis zu zehn "X"-Reste inseriert werden können.

Als weitere Alternativen können beliebige oder alle möglichen Varianten aufgenommen werden.

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" in der SEQ ID NO 56 und SEQ ID NO 57 "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") versehen werden.

Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 27 und **36**

Absatz 37 – Sequenzen, die Regionen mit einer unbekannten Zahl an zusammenhängenden "n"- oder "X"-Resten enthalten

Beispiel 37-1: Sequenz mit Regionen, die eine unbekannte Zahl an "X"-Resten enthalten, darf nicht als eine einzelne Sequenz dargestellt werden

Gly-Gly----Gly-Gly-Xaa-Xaa

Dabei steht das Symbol ---- für eine undefinierte Lücke innerhalb der Sequenz, Xaa steht für jede beliebige Aminosäure, und die Glycin- und Xaa-Reste sind durch Peptidbindungen miteinander verbunden.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

NEIN

ST.26 Absatz 37 verbietet die Aufnahme von Sequenzen, die eine undefinierte Lücke enthalten; daher ist die Aufnahme der gesamten Sequenz nicht erforderlich.

Eine Region einer Sequenz, die an eine undefinierte Lücke angrenzt und vier oder mehr spezifisch definierte Aminosäuren enthält, muss nach ST.26 Absatz 37 aufgenommen werden. Im obigen Beispiel muss keine der beiden an die undefinierte Lücke angrenzenden Regionen aufgenommen werden, da sie jeweils nur zwei spezifisch definierte Aminosäuren enthalten.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

NEIN – nicht die gesamte Sequenz

NEIN – keine Region der Sequenz

Die Aufnahme der gesamten Sequenz ist nach ST.26 Absatz 37 nicht zulässig.

Nach ST.26 Absatz 8 darf keine der an die undefinierte Lücke angrenzenden Regionen aufgenommen werden, da beide Regionen jeweils nur zwei spezifisch definierte Aminosäuren enthalten.

Relevante Absätze im Standard ST.26: 7(b), 8, 26 und **37**

Beispiel 37-2: Sequenz mit Regionen, die eine unbekannte Zahl an "X"-Resten enthalten, darf nicht als eine einzelne Sequenz dargestellt werden

Gly-Gly----Gly-Gly-Ala-Gly-Xaa-Xaa

Dabei steht das Symbol ---- für eine undefinierte Lücke innerhalb der Sequenz, Xaa steht für jede beliebige Aminosäure, und die Glycin- und Xaa-Reste sind durch Peptidbindungen miteinander verbunden.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

NEIN – nicht die gesamte Sequenz

JA – eine Region der Sequenz

Eine Sequenz, die eine undefinierte Lücke enthält, darf nach ST.26 Absatz 37 nicht aufgenommen werden, doch eine Region einer Sequenz, die an eine undefinierte Lücke angrenzt und vier oder mehr spezifisch definierte Aminosäuren enthält, muss aufgenommen werden.

Im obigen Beispiel ist die Aufnahme der gesamten Sequenz, die eine undefinierte Lücke enthält, und die Aufnahme der an die undefinierte Lücke angrenzenden Gly-Gly -Region, die nur zwei spezifisch definierte Aminosäuren enthält, nach ST.26 nicht erforderlich (bzw. verboten). Die an die undefinierte Lücke angrenzende Region Gly-Gly-Ala-Gly-Xaa-Xaa muss jedoch nach ST.26 aufgenommen werden, da sie mindestens vier spezifisch definierte Aminosäuren enthält.

Frage 2: Ist die Aufnahme der Sequenz(en) nach ST.26 zulässig?

NEIN – nicht die gesamte Sequenz und nicht die Gly-Gly-Region

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Region der Sequenz, die an die undefinierte Lücke angrenzt und vier spezifisch definierte Aminosäuren enthält, muss wie folgt dargestellt werden:

GGAGXX (SEQ ID NO: 58)

Die Sequenz sollte annotiert werden, um anzuzeigen, dass sie Teil einer größeren Sequenz ist, die eine undefinierte Lücke enthält. Dafür ist der Merkmalschlüssel "SITE", die Lage des Merkmals "1" und der Qualifier "note" mit einem Wert wie z. B. dem Folgenden zu verwenden: "This residue is linked N-terminally to a peptide having an N-terminal Gly-Gly and a gap of undefined length." ("Dieser Rest ist über eine N-Terminus-Bindung an ein Peptid mit N-terminalem Gly-Gly und einer Lücke von nicht definierter Länge gebunden.")

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" in der SEQ ID NO 58 "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") versehen werden.

Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

Relevante Absätze im Standard ST.26: 7(b), 8, 26, 27 und **37**

Absatz 55 – Eine Nukleotidsequenz, die sowohl DNA- als auch RNA-Segmente enthält

Beispiel 55-1: Kombiniertes DNA/RNA-Molekül

In einer Patentanmeldung wird folgende Oligonukleotidsequenz beschrieben:

AGACCTTcggagucuccuguugaacagauagucaaaguagauC

Dabei stellen die Großbuchstaben DNA-Reste und die Kleinbuchstaben RNA-Reste dar.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die offenbarte Sequenz weist mehr als zehn durch Aufzählung dargestellte und spezifisch definierte Nukleotide auf und muss folglich in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Nukleotidsequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

agaccttcggagtctcctgttgaacagatagtcaaagtagatc (SEQ ID NO: 93)

Beachten Sie, dass die Uracil-Nukleotide im Sequenzprotokoll durch das Symbol "t" dargestellt werden müssen.

In ST.26 Absatz 55 ist zwingend vorgeschrieben, dass eine Nukleotidsequenz, die sowohl DNA- als auch RNA-Segmente enthält, als Molekültyp "DNA" angegeben und mit dem Merkmalschlüssel "source" näher beschrieben werden muss. Dabei ist der obligatorische Qualifier "organism" mit dem Wert "synthetic construct" ("synthetisches Konstrukt") und der obligatorische Qualifier "mol_type" mit dem Wert "other DNA" zu verwenden. Zusätzlich muss jedes Segment der Sequenz mit dem Merkmalschlüssel "misc_feature", der die Lage des Segments enthält, und dem Qualifier "note", der angibt, ob es sich um DNA oder RNA handelt, näher beschrieben werden. Die offenbarte Sequenz enthält zwei DNA-Segmente (Nukleotidpositionen 1-7 und 43) und ein RNA-Segment (Nukleotidpositionen 8-42).

Relevante Absätze im Standard ST.26: 7, 14, **55**-56 und 83

Absatz 89 – Der Merkmalschlüssel "CDS"

Beispiel 89-1: Codierende Nukleotidsequenz und codierte Aminosäuresequenz

In einer Patentanmeldung wird folgende Nukleotidsequenz und deren Translation beschrieben:

```
atg acc gga aat aaa cct gaa acc gat gtt tac gaa att tta tga

Met Thr Gly Asn Lys Pro Glu Thr Asp Val Tyr Glu Ile Leu STOP
```

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die durch Aufzählung dargestellte Nukleotidsequenz enthält mehr als zehn spezifisch definierte Nukleotide.

Die durch Aufzählung dargestellte Aminosäuresequenz enthält mehr als vier spezifisch definierte Aminosäuren.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Nukleotidsequenz muss wie folgt dargestellt werden:

atgaccggaaataaacctgaaaccgatgtttacgaaattttatga (SEQ ID NO: 59)

Die Nukleotidsequenz sollte mit dem Merkmalschlüssel "CDS" näher beschrieben werden, und das Element "INSDFeature_location" muss die gesamte Sequenz einschließlich des Stopcodons (d. h. Position 1 bis einschließlich 45) angeben. Zusätzlich sollte der Qualifier "translation" mit dem Qualifier-Wert "MTGNKPETDVYEIL" angegeben werden. Die Tabelle für den genetischen Code für die Translation (siehe Anhang I, Abschnitt 9, Tabelle 7) wird in der Anmeldung nicht offenbart. Wenn die Tabelle für den Standardcode gilt, ist der Qualifier "transl_table" nicht erforderlich; wenn jedoch eine andere Tabelle für den genetischen Code gilt, muss für den Qualifier "transl_table" der entsprechende Qualifier-Wert aus Tabelle 7 angegeben werden. Schließlich muss der Qualifier "protein_id" mit dem Qualifier-Wert angegeben werden, aus dem die Sequenzkennzahl der translatierten Aminosäuresequenz hervorgeht.

Die Aminosäuresequenz muss separat mit einer eigenen Sequenzkennzahl mit einbuchstabigen Codes wie folgt dargestellt werden:

MTGNKPETDVYEIL (SEQ ID NO: 60)

Das "STOP"-Symbol nach der durch Aufzählung dargestellten Aminosäuresequenz darf nicht in deren Darstellung im Sequenzprotokoll aufgenommen werden.

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 7(a), 7(b), 26, 28, 89, 90 und 92

Beispiel 89-2: Lage des Merkmals geht über die offenbarte Sequenz hinaus

Eine Patentanmeldung enthält die folgende Abbildung, die eine partielle codierende Sequenz und deren translatierte Aminosäuresequenz offenlegt:

```
cat cac gca gca gaa tgt gga ttt tgt cct caa caa tgg caa gtt cta     48
His His Ala Ala Glu Cys Gly Phe Cys Pro Gln Gln Trp Gln Val Leu
1               5                   10                  15

cgt ggg agt ctg tgc att tgt gag ggt cca gct gaa gga tgg ttc ata     96
Arg Gly Ser Leu Cys Ile Cys Glu Gly Pro Ala Glu Gly Trp Phe Ile
            20                  25                  30

tca aga tgt tgg tta tgg tgt ggg cct caa gtc caa ggc ttt atc ttt    144
Ser Arg Cys Trp Leu Trp Cys Gly Pro Gln Val Gln Gly Phe Ile Phe
        35                  40                  45

gga gaa ggc aag gaa gga ggc ggt gac aga cgg gct gaa gcg agc cct    192
Gly Glu Gly Lys Glu Gly Gly Gly Asp Arg Arg Ala Glu Ala Ser Pro
    50                  55                  60

cag gag ttt tgg gaa tgc act tgg                                    216
Gln Glu Phe Trp Glu Cys Thr Trp
65                  70
```

Abbildung 1 – partielle codierende Sequenz des *ITCH1*-Gens des Homo sapiens, die für die Aminosäuren 20 bis einschließlich 91 des 442 Aminosäuren langen ITCH1-Proteins codiert.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der Anmeldung werden eine Nukleotidsequenz und ihre translatierte Aminosäuresequenz offenbart. Die durch Aufzählung dargestellte Nukleotidsequenz enthält mehr als zehn spezifisch definierte Nukleotide und muss in ein Sequenzprotokoll aufgenommen werden.

Die Aminosäuresequenz enthält mehr als vier spezifisch definierte Aminosäuren und muss daher ebenfalls als separate Sequenz mit eigener Sequenzkennzahl in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Nukleotidsequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

catcacgcagcagaatgtggattttgtcctcaacaatggcaagttctacgtgggagtctgtgcatttgtgagggtccagctgaaggatggttcatatcaagatgttggt
tatggtgtgggcctcaagtccaaggctttatctttggagaaggcaaggaaggaggcggtgacagacgggctgaagcgagccctcaggagttttgggaatgcact
tgg (SEQ ID NO: 94)

Die Nukleotidsequenz sollte mit einem "CDS"-Merkmalschlüssel näher beschrieben werden. Das Element "INSDFeature_location" muss die Lage des "CDS"-Merkmals in der Sequenz einschließlich des Stopcodons enthalten.

In der Abbildung ist eine partielle codierende Sequenz dargestellt, die weder das Start- noch das Stopcodon umfasst. Aus der Beschreibung der Sequenz geht jedoch hervor, dass sich das Startcodon vor dem Nukleotid in Position 1 und das Stopcodon hinter dem letzten Nukleotid in Position 216 befindet.

In ST.26 ist zwingend vorgeschrieben, dass der Lagedeskriptor keine Nummerierung für Reste enthalten darf, die außerhalb des im Element "INSDSeq_sequence" angegebenen Sequenzbereichs liegen. Folglich darf der Lagedeskriptor für den Merkmalschlüssel "CDS" im obigen Beispiel keine Positionsnummern außerhalb des Bereichs von 1 bis 216 enthalten. Die Lage des Stopcodons muss im Element "INSDFeature_location" mit dem Symbol ">" dargestellt werden, um anzuzeigen, dass sich das Stopcodon hinter Position 216 befindet. Entsprechend kann das Symbol "<" verwendet werden, um anzuzeigen, dass die Position des Startcodons vor Position 1 liegt. Folglich sollte der Lagedeskriptor für den "CDS"-Merkmalschlüssel wie folgt angegeben werden:

<1..>216

Beachten Sie, dass "<" und ">" reservierte Zeichen sind und in der XML-Instanz des Sequenzprotokolls durch "<" bzw. ">" ersetzt werden müssen.

Der Qualifier "translation" sollte mit der Aminosäuresequenz des Proteins als Qualifier-Wert angegeben werden. Aus der Abbildung geht nicht hervor, welche Tabelle für den genetischen Code für die Translation (siehe Anhang I, Abschnitt 9, Tabelle 7) gilt. Wenn die Tabelle für den Standardcode gilt, ist der Qualifier "transl_table" nicht erforderlich; wenn jedoch eine andere Tabelle für den genetischen Code gilt, muss für den Qualifier "transl_table" der entsprechende Qualifier-Wert aus ST.26 Anhang I Tabelle 7 angegeben werden. Schließlich muss im "CDS"-Merkmal der Qualifier "protein_id" mit dem Qualifier-Wert angegeben werden, aus dem die Sequenzkennzahl der translatierten Aminosäuresequenz hervorgeht.

Die translatierte Aminosäuresequenz muss als separate Sequenz mit eigener Sequenzkennzahl aufgenommen werden:

HHAAECGFCPQQWQVLRGSLCICEGPAEGWFISRCWLWCGPQVQGFIFGEGKEGGGDRRAEASPQEFWECTW
(SEQ ID NO: 95)

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 7, 41, 65, 66, 70, 71, 89 und 92

Absatz 92 – Aminosäuresequenz, die durch eine codierende Sequenz codiert wird

Beispiel 92-1: Aminosäuresequenz, die durch eine codierende Sequenz mit Introns codiert wird

Eine Patentanmeldung enthält die folgende Abbildung, die eine codierende Sequenz und deren Translation offenlegt:

```
atg aag act ttc gca gcc ttg ctt tcc gct gtc act ctc gcg ctc tcg
Met Lys Thr Phe Ala Ala Leu Leu Ser Ala Val Thr Leu Ala Leu Ser

gtg cgc gcc cag gcg gct gtc tgg agt caa t gtaagtgccg ctgcttttca
Val Arg Ala Gln Ala Ala Val Trp Ser Gln

ttgatacgag actctacgcc gagctgacgt gctaccgtat ag gt  ggc ggt aca
                                               Cys Gly Gly Thr

ccg ggt tgg acg ggc gag acc act tgc gtt gct ggt tcg gtt tgt acc
Pro Gly Trp Thr Gly Glu Thr Thr Cys Val Ala Gly Ser Val Cys Thr

tcc ttg agc tca gtgagcgact ttcaatccgt cgtcattgct cctcatgtat
Ser Leu Ser Ser

tgacgattgg ccttcatag tca tac tct caa tgc gtt ccg ggc tcc gca acg
                    Ser Tyr Ser Gln Cys Val Pro Gly Ser Ala Thr

tcc agc gct ccg gcg gcc ccc tca gcg aca act tca ggc ccc gca cct
Ser Ser Ala Pro Ala Ala Pro Ser Ala Thr Thr Ser Gly Pro Ala Pro

acg gac gga acg tgc tcg gcc agc ggg gca tgg ccg cca ttg acc tga
Thr Asp Gly Thr Cys Ser Ala Ser Gly Ala Trp Pro Pro Leu Thr Ter
```

Abbildung 1 – Die in Fettdruck dargestellten Nukleotide sind Intronregionen.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

In der Anmeldung werden eine Nukleotidsequenz und ihre translatierte Aminosäuresequenz offenbart. Die durch Aufzählung dargestellte Nukleotidsequenz enthält mehr als zehn spezifisch definierte Nukleotide und muss als einzelne Sequenz in ein Sequenzprotokoll aufgenommen werden.

Die Nukleotidsequenz enthält codierende Sequenzen (Exons), die durch nicht codierende Sequenzen (Introns) getrennt sind. Die Abbildung zeigt die Translation der Nukleotidsequenz als drei nicht zusammenhängende Aminosäuresequenzen. Aus der Beschriftung der Abbildung geht hervor, dass es sich bei den in Fettdruck dargestellten Regionen von Nukleotiden um Intronsequenzen handelt, die vor der Translation in ein Protein aus einem RNA-Transkript gespleißt werden. Daher sind die drei Aminosäuresequenzen eigentlich eine einzige, zusammenhängende, durch Aufzählung dargestellte Sequenz, die mehr als vier spezifisch definierte Aminosäuren enthält und als eine einzelne Sequenz in ein Sequenzprotokoll aufgenommen werden muss.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Nukleotidsequenz muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

atgaagactttcgcagccttgctttccgctgtcactctcgcgctctcggtgcgcgcccaggcggctgtctggagtcaatgtaagtgccgctgcttttcattgatacgaga
ctctacgccgagctgacgtgctaccgtataggtggcggtacaccgggttggacgggcgagaccacttgcgttgctggttcggtttgtacctccttgagctcagtgag
cgactttcaatccgtcgtcattgctcctcatgtattgacgattggccttcatagtcatactctcaatgcgttccgggctccgcaacgtccagcgctccggcggccccctc
agcgacaacttcaggccccgcacctacggacggaacgtgctcggccagcggggcatggccgccattgacctga (SEQ ID NO: 75)

Die Nukleotidsequenz sollte mit einem "CDS"-Merkmalschlüssel näher beschrieben werden, und mit dem Element "INSDFeature_location" muss die Lage der codierenden Sequenz, einschließlich des durch "Ter" angezeigten Stopcodons, angegeben werden. Bei der Angabe der Lage des Merkmals "CDS" ("INSDFeature_location") muss der Lageoperator "join" verwendet werden, um anzuzeigen, dass die von den angegebenen Lagen codierten Translationsprodukte verbunden sind und ein einziges, zusammenhängendes Polypeptid bilden. Dabei ist das Format "join(x1..y1,x2..y2,x3..y3)" zu verwenden, z. B. "join(1..79,142..212,272..400)". Zusätzlich sollte der Qualifier "translation" mit der Aminosäuresequenz des Proteins als Qualifier-Wert angegeben werden. (Beachten Sie, dass das Terminatorsymbol "Ter" in der letzten Position der Sequenz nicht in die Aminosäuresequenz aufgenommen werden darf.) Die Tabelle für den genetischen Code für die Translation (siehe Anhang I, Abschnitt 9, Tabelle 7) wird in der Anmeldung nicht offenbart. Wenn die Tabelle für den Standardcode gilt, ist der Qualifier "transl_table" nicht erforderlich; wenn jedoch eine andere Tabelle für den genetischen Code gilt, muss für den Qualifier "transl_table" der entsprechende Qualifier-Wert aus Tabelle 7 angegeben werden. Schließlich muss der Qualifier "protein_id" mit dem Qualifier-Wert angegeben werden, aus dem die Sequenzkennzahl der translatierten Aminosäuresequenz hervorgeht.

Die Aminosäuresequenz muss als eine einzelne Sequenz aufgenommen werden:

MKTFAALLSAVTLALSVRAQAAVWSQCGGTPGWTGETTCVAGSVCTSLSSSYSQCVPGSATSSAPAAPSATTSGPAPT
DGTCSASGAWPPLT (SEQ ID NO: 76)

Relevante Absätze im Standard ST.26: 7, 26, 28, 57, 67 und 89–92

Absatz 93 – Primärsequenz und eine Variante, die jeweils durch Aufzählung der Reste dargestellt werden

Beispiel 93-1: Darstellung der durch Aufzählung dargestellten Varianten

Die Beschreibung enthält die folgende Sequenzalignierung:

```
D. melanogaster      ACATTGAATCTCATACCACTTT
D. virilis           ...-..G...C..--.G.....
D. simulans          GT..G.CG..GT..SGT.G...
```

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Es ist in der Fachwelt üblich, in einem Sequenzabgleich durch Punkte anzuzeigen, dass diese Position die gleiche ist wie die darüber liegende Position. Daher werden die "Punkte" in den Sequenzen für *D. virilis* und *D. simulans* als durch Aufzählung dargestellte und spezifisch definierte Nukleotide angesehen. Sie dienen lediglich der verkürzten Darstellung des Umstands, dass sich an einer bestimmten Position dasselbe Nukleotid befindet wie in *D. melanogaster*. Außerdem wird bei Sequenzabgleichen zur Maximierung des Sequenzabgleichs häufig das Symbol "**-**" verwendet, um das Fehlen eines Rests anzuzeigen.

Dementsprechend enthalten die Nukleotidsequenzen von *D. melanogaster* und *D. simulans* 22 durch Aufzählung dargestellte und spezifisch definierte Nukleotide, die Nukleotidsequenz von *D. virilis* hingegen 19. Daher muss nach ST.26 Absatz 7(a) jede Sequenz mit separaten Sequenzkennzahlen in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Sequenz für *Drosophila melanogaster* muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

acattgaatctcataccacttt (SEQ ID NO: 61)

Die Sequenz für *Drosophila virilis* muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

acatggatcccacgacttt (SEQ ID NO: 62)

Die Sequenz für *Drosophila simulans* muss wie folgt in ein Sequenzprotokoll aufgenommen werden:

gtatggcgtcgtatsgtagttt (SEQ ID NO: 63)

Relevante Absätze im Standard ST.26: 7(a), 13 und 93

Beispiel 93-2: Darstellung der durch Aufzählung dargestellten Varianten

Die Beschreibung enthält die folgende Tabelle eines Peptids und dessen funktioneller Varianten. Eine Leerstelle in der Tabelle unten zeigt an, dass die Aminosäure in der Variante die gleiche ist wie die entsprechende Aminosäure in der "Sequenz", und ein "-" zeigt an, dass die entsprechende Aminosäure der "Sequenz" in der Variante deletiert wurde.

Position	1	2	3	4	5	6	7	8	9
Sequenz	A	V	L	T	Y	L	R	G	E
Variante 1									A
Variante 2			P			P			
Variante 3			A	I	G	Y			
Variante 4							-		

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Wie angegeben, zeigt eine Leerstelle in dieser Tabelle an, dass die Aminosäure in der Variante die gleiche ist wie die entsprechende Aminosäure in der "Sequenz". Folglich werden die Aminosäuren der Sequenzvarianten durch Aufzählung dargestellt und spezifisch definiert.

Da die vier Sequenzvarianten jeweils mehr als vier durch Aufzählung dargestellte und spezifisch definierte Aminosäuren enthalten, müssen sie nach ST.26 Absatz 7(b) jeweils mit separaten Sequenzkennzahlen in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

AVLTYLRGE (SEQ ID NO: 77)

AVLTYLRGA (SEQ ID NO: 78)

AVPTYPRGE (SEQ ID NO: 79)

AVAIGYRGE (SEQ ID NO: 80)

AVLTYLGE (SEQ ID NO: 81)

Relevante Absätze im Standard ST.26: 7(b), 26 und 93

Beispiel 93-3: Darstellung einer Konsensussequenz

Eine Patentanmeldung enthält Abbildung 1 mit dem folgenden mehrfachen Sequenzabgleich.

```
Consensus                 LEGnEQFINAakIIRHPkYnrkTlnNDImLIK
Homo sapiens              LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK
Pongo abelii              LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK
Papio Anubis              LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK
Rhinopithecus roxellana   LEGTEQFINAAKIIRHPNYNRITLDNDILLIK
Pan paniscus              LEGNEQFINAAKIIRHPKYNRITLNNDIMLIK
Rhinopithecus bieti       LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK
Rhinopithecus roxellana   LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK
```

Die Großbuchstaben in der Konsensussequenz stellen konservierte Aminosäurereste dar, und die Kleinbuchstaben "n", "a", "k", "r", "l" und "m" stehen für die vorherrschenden Aminosäurereste in den alignierten Sequenzen.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die Kleinbuchstaben in der Konsensussequenz stellen jeweils einen einzelnen Aminosäurerest dar. Folglich enthalten die Konsensussequenz sowie jede der übrigen sieben Sequenzen in Abbildung 1 mindestens vier spezifisch definierte Aminosäuren. Nach ST.26 Absatz 7(b) müssen alle acht Sequenzen in das Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die Kleinbuchstaben in der Konsensussequenz werden als Mehrdeutigkeitssymbole verwendet. Sie stellen jeweils die vorherrschende Aminosäure unter den möglichen Varianten für eine spezifische Position dar. Daher sind die Kleinbuchstaben "n", "a", "k", "r", "l" und "m" konventionelle Symbole, die auf nicht konventionelle Weise verwendet werden, sodass bei der Darstellung der Konsensussequenz anstelle dieser Kleinbuchstaben ein Mehrdeutigkeitssymbol verwendet werden muss.

Dabei sollte das restriktivste Mehrdeutigkeitssymbol verwendet werden. Für die meisten Positionen in der Konsensussequenz ist "X" das restriktivste Mehrdeutigkeitssymbol; doch das restriktivste Mehrdeutigkeitssymbol für "D" oder "N" in den Positionen 20 und 25 ist "B". Die Konsensussequenz sollte wie folgt in ein Sequenzprotokoll aufgenommen werden:

LEGXEQFINAXXIIRHPXYBXXTXBNDIXLIK (SEQ ID NO: 82)

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Daher muss jedes "X" in der Konsensussequenz in einer Merkmaltabelle mit dem Merkmalschlüssel "VARIANT" und dem Qualifier "note" näher beschrieben werden, um die möglichen Varianten für jede Position anzugeben.

Die übrigen sieben Sequenzen müssen wie folgt in ein Sequenzprotokoll aufgenommen werden:

LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK (SEQ ID NO: 83)

LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK (SEQ ID NO: 84)

LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK (SEQ ID NO: 85)

LEGTEQFINAAKIIRHPNYNRITLDNDILLIK (SEQ ID NO: 86)

LEGNEQFINAAKIIRHPKYNRITLNNDIMLIK (SEQ ID NO: 87)

LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 88)

LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 89)

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss,

ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 7(b), 26, 27, 93 und 97

Absatz 94 – Sequenzvariante, die als eine einzelne Sequenz mit durch Aufzählung dargestellten alternativen Resten offenbart wird

Beispiel 94-1: Darstellung einer einzelnen Sequenz mit durch Aufzählung dargestellten alternativen Aminosäuren

In einer Patentanmeldung wird Anspruch auf ein Peptid mit folgender Sequenz erhoben:

(i) Gly-Gly-Gly-[Leu or Ile]-Ala-Thr-[Ser or Thr]

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die Sequenz enthält vier spezifisch definierte Aminosäuren und muss somit nach ST.26 Absatz 7(b) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

In Anhang I, Abschnitt 3, Tabelle 3 ist das Mehrdeutigkeitssymbol "J" als Isoleucin oder Leucin definiert. Die bevorzugte Darstellung der Sequenz ist somit wie folgt:

GGGJATX (SEQ ID NO.: 64)

Dabei ist eine nähere Beschreibung in einer Merkmaltabelle mit dem Merkmalschlüssel "VARIANT" und dem Qualifier "note" erforderlich, um anzuzeigen, dass das "X" Serin oder Threonin ist.

Alternativ kann die Sequenz beispielsweise wie folgt dargestellt werden:

GGGLATS (SEQ ID NO: 65)

Dabei ist eine nähere Beschreibung in einer Merkmaltabelle mit dem Merkmalschlüssel "VARIANT" und dem Qualifier "note" erforderlich, um anzuzeigen, dass L durch I und S durch T ersetzt werden kann.

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 7(b), 8, 26, 27, 94 und 97

Beispiel 94-2 – Darstellung einer einzelnen Sequenz mit durch Aufzählung dargestellten alternativen Aminosäuren, bei denen es sich um modifizierte Aminosäuren handeln kann

In einer Patentanmeldung wird das folgende Polypeptid beschrieben:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Xaa-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

Dabei kann Xaa für Ile, Ala, Phe, Tyr, alle, Melle oder Nle stehen.

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das durch Aufzählung dargestellte Peptid enthält 16 spezifisch definierte Aminosäuren. Daher muss die Sequenz gemäß ST.26 Absatz 7(b) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Das restriktivste Mehrdeutigkeitssymbol, das Ile, Ala, Phe, Tyr, alle, Melle oder Nle umfassen kann, ist "X". Daher muss die Sequenz in folgender Form in ein Sequenzprotokoll aufgenommen werden:

LEYCLKRWXETISHCAW (SEQ ID NO: 96)

Nach ST.26 Absatz 30 muss eine modifizierte Aminosäure in der Merkmaltabelle näher beschrieben werden. Allerdings verlangt Absatz 30 nicht, dass zur Beschreibung modifizierter Aminosäuren ein spezifischer Merkmalschlüssel verwendet wird. Während Absatz 30 die Verwendung der Merkmalschlüssel "CARBOHYD", "LIPID", "MOD_RES" und "SITE" beschreibt, sind diese Merkmalschlüssel eher für Szenarien geeignet, in denen die modifizierte Aminosäure nicht in einer Liste von Alternativen für eine bestimmte Lage enthalten ist. In diesem Beispiel genügt der Merkmalschlüssel "VARIANT" den Anforderungen von Absatz 30, da er die Aufnahme aller Alternativen für die Variantenstelle zulässt. Somit sollte der Merkmalschlüssel "VARIANT" mit einem "note"-Qualifier und dem Qualifier-Wert "Ile, Ala, Phe, Tyr, alle, Melle, or Nle" ("Ile, Ala, Phe, Tyr, alle, Melle oder Nle") verwendet werden, um die Variantenstelle an Position 9 zu beschreiben. Es kann ein zweiter Merkmalschlüssel wie z. B. "SITE" mit einem "note"-Qualifier verwendet werden, um die an Position 9 vorliegenden modifizierten Aminosäuren näher zu beschreiben.

Relevante Absätze im Standard ST.26: 3(a), 7(b), 27, 30, **94**, 96 und Anhang I, Abschnitt 4, Tabelle 4

Absatz 95(a) – Sequenzvariante, die nur durch Verweis auf eine Primärsequenz mit mehreren unabhängigen Variationen offenbart wird

Beispiel 95(a)-1: Darstellung einer Sequenzvariante durch Annotation der Primärsequenz

In einer Anmeldung ist die folgende Offenbarung enthalten:

"Peptidfragment 1 ist Gly-Leu-Pro-Xaa-Arg-Ile-Cys, wobei Xaa jede beliebige Aminosäure … sein kann.

In einer anderen Ausführungsform ist das Peptidfragment 1 Gly-Leu-Pro-Xaa-Arg-Ile-Cys, wobei Xaa Val, Thr, oder Asp… sein kann.

In einer anderen Ausführungsform ist das Peptidfragment 1 Gly-Leu-Pro-Xaa-Arg-Ile-Cys, wobei Xaa Val sein kann."

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Das Peptidfragment 1 enthält in jeder der drei offenbarten Ausführungsformen mindestens sechs spezifisch definierte Aminosäuren; daher muss die Sequenz in ein Sequenzprotokoll aufgenommen werden, wie in ST.26 Absatz 7(b) gefordert.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

In diesem Beispiel wird die durch Aufzählung dargestellte Sequenz von Peptidfragment 1 dreimal offenbart, nämlich in drei verschiedenen Ausführungsformen mit jeweils einer alternativen Beschreibung von Xaa. Das restriktivste Mehrdeutigkeitssymbol für die Xaa-Position ist in diesem Beispiel "X".

Nach ST.26 muss die offenbarte durch Aufzählung dargestellte Sequenz nur einmal aufgenommen werden. In der umfassendsten der drei Ausführungsformen ist Xaa jede beliebige Aminosäure (siehe Einführung in dieses Dokument). Daher muss die Sequenz wie folgt in das Sequenzprotokoll aufgenommen werden:

GLPXRIC (SEQ ID NO: 66)

Laut Absatz 27 wird das Symbol "X" als eines der Symbole "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" oder "V" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da "X" in der SEQ ID NO 66 "jede beliebige Aminosäure" darstellt, muss es mit dem Merkmalschlüssel "VARIANT" und einem "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") versehen werden.

Sofern praktikabel, sollte jedes "X" einzeln annotiert werden. Allerdings können eine Region aus zusammenhängenden "X"-Resten oder mehrfache in der Sequenz verteilte "X"-Reste mit dem Merkmalschlüssel "VARIANT" gemeinsam beschrieben werden, wobei ein Lagedeskriptor mit der Syntax "x..y" (mit x und y als den Positionen des ersten und letzten "X"-Rests) und ein "note"-Qualifier mit dem Wert "X can be any amino acid" ("X kann jede beliebige Aminosäure sein") zu verwenden sind.

Es wird nachdrücklich empfohlen, weitere Sequenzen aufzunehmen, die für die Offenbarung oder die Ansprüche der Erfindung wesentlich sind, wie in der Einführung in dieses Dokument erläutert.

Für das obige Beispiel wird dringend empfohlen, dass die folgenden drei zusätzlichen Sequenzen mit einer jeweils eigenen Sequenzkennzahl in das Sequenzprotokoll aufgenommen werden:

GLPVRIC (SEQ ID NO: 67)

GLPTRIC (SEQ ID NO: 68)

GLPDRIC (SEQ ID NO: 69)

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss,

ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 7(b), 26, 27 und **95(a)**

Absatz 95(b) – Sequenzvariante, die nur durch Verweis auf eine Primärsequenz mit mehreren voneinander abhängigen Variationen offenbart wird

Beispiel 95(b)-1: Darstellung einzelner Sequenzvarianten mit mehreren voneinander abhängigen Variationen

In einer Patentanmeldung wird folgende Konsensussequenz beschrieben:

cgaatg**n**$_1$cccactacgaatg**n**$_2$cacgaatg**n**$_3$cccaca

Dabei können n_1, n_2 und n_3 a, t, g oder c sein.

Mehrere Sequenzvarianten werden wie folgt offenbart:

Wenn n_1 a ist, dann sind n_2 und n_3 t, g oder c;

Wenn n_1 t ist, dann sind n_2 und n_3 a, g oder c;

Wenn n_1 g ist, dann sind n_2 und n_3 t, a oder c;

Wenn n_1 c ist, dann sind n_2 und n_3 t, g oder a;

Frage 1: Ist die Aufnahme der Sequenz(en) nach ST.26 erforderlich?

JA

Die Sequenz enthält mehr als zehn durch Aufzählung dargestellte und "spezifisch definierte" Nukleotide und muss daher nach ST.26 Absatz 7(a) in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sollte(n) die Sequenz(en) im Sequenzprotokoll dargestellt werden?

Die durch Aufzählung dargestellte Sequenz enthält mehr als zehn spezifisch definierte Nukleotide und drei "n"-Reste. Die offenbarte, durch Aufzählung dargestellte Sequenz muss nach ST.26 aufgenommen werden, und wenn ein Mehrdeutigkeitssymbol angebracht ist, sollte das restriktivste Symbol verwendet werden. In diesem Beispiel können n_1, n_2 und n_3 a, t, g oder c sein, sodass "n" das restriktivste Mehrdeutigkeitssymbol ist. Daher muss die Sequenz wie folgt in das Sequenzprotokoll aufgenommen werden:

`cgaatgncccactacgaatgncacgaatgncccaca` (SEQ ID NO: 70)

Laut ST.26 Absatz 15 wird das Symbol "n" als eines der Symbole "a", "c", "g" oder "t/u" ausgelegt, es sei denn, es wird in Verbindung mit einer näheren Beschreibung in der Merkmaltabelle verwendet. Da in dieser Sequenz die Werte aller "n"-Reste dem Standard "a", "c", "g" oder "t" entsprechen, ist keine weitere Annotation erforderlich.

Die durch Aufzählung dargestellte Sequenz enthält an drei unterschiedlichen Positionen Variationen, die in wechselseitiger Abhängigkeit voneinander auftreten. Die Aufnahme zusätzlicher Sequenzen, mit denen zusätzliche Ausführungsformen dargestellt werden, die wesentlicher Bestandteil der Erfindung sind, wird **nachdrücklich** empfohlen, wie in der Einführung in dieses Dokument erläutert. Daher sollten die zusätzlichen Ausführungsformen nach ST.26 Absatz 95(b) in einem Sequenzprotokoll als vier separate Sequenzen mit jeweils eigener Sequenzkennzahl aufgeführt werden:

`cgaatgacccactacgaatgbcacgaatgbcccaca` (SEQ ID NO: 71)

`cgaatgtcccactacgaatgvcacgaatgvcccaca` (SEQ ID NO: 72)

`cgaatggcccactacgaatghcacgaatghcccaca` (SEQ ID NO: 73)

`cgaatgccccactacgaatgdcacgaatgdcccaca` (SEQ ID NO: 74)

(Beachten Sie, dass b = t, g oder c; v = a, g oder c; h = t, a oder c und d = t, g oder a; siehe Anhang I, Abschnitt 1, Tabelle 1.)

Nach ST.26 Absatz 15 muss das restriktivste Symbol verwendet werden, um variable Positionen darzustellen. Folglich dürfen n_2 und n_3 in der Sequenz nicht durch "n" dargestellt werden.

ACHTUNG: Die oben angegebene bevorzugte Darstellung der Sequenz bezieht sich auf ein Sequenzprotokoll, das am Anmeldetag der Patentanmeldung eingereicht wird. Wenn ein Sequenzprotokoll nach dem Anmeldetag einer Patentanmeldung eingereicht wird, ist diese Darstellung womöglich nicht anwendbar, da berücksichtigt werden muss, ob die bereitgestellten Informationen von einem Amt für geistiges Eigentum als zur ursprünglichen Offenbarung hinzugefügter Gegenstand angesehen werden könnten.

Relevante Absätze im Standard ST.26: 7(a), 15 und 95(b)

[Anlage zum Anhang VI von ST.26 folgt]

ANLAGE

SEQUENZEN DES LEITFADENS IN XML

Die Anlage ist verfügbar unter:
https://www.wipo.int/standards/en/xml_material/st26/st26-annex-vi-appendix-guidance-document-sequences_v1_7.xml

[Anhang VII folgt]

ANHANG VII

EMPFEHLUNG FÜR DIE ÜBERFÜHRUNG EINES SEQUENZPROTOKOLLS VON ST.25[11] NACH ST.26: MÖGLICHERWEISE HINZUGEFÜGTER ODER GESTRICHENER GEGENSTAND

Einführung

Die WIPO-Standards ST.25 und ST.26 unterscheiden sich hinsichtlich der Anforderungen an die Darstellung von Nukleotid- und Aminosäuresequenzen. Daraus ergab sich die Frage, ob der Standard ST.26 die Hinzufügung oder Streichung eines Gegenstands in einem Sequenzprotokoll nach sich ziehen würde, das nach dem Standard ST.26 als Teil einer internationalen Anmeldung eingereicht wird, die unter Umständen nicht durch eine Anmeldung gestützt wird, aus der ein Prioritätsanspruch abgeleitet wird.

Gegenstand des Dokuments

In diesem Dokument werden die obligatorischen Anforderungen nach ST.26 und deren mögliche Folgen behandelt. Dieses Dokument geht nicht auf alle erdenklichen Szenarien ein; wenn nicht klar ist, auf welche Weise in einem Sequenzprotokoll nach ST.25 enthaltene Informationen in einem Sequenzprotokoll nach ST.26 darzustellen sind, können diese Informationen jederzeit in die Beschreibung der Anmeldung aufgenommen werden, um die Streichung von Gegenständen zu vermeiden.

Empfehlungen zur möglichen Hinzufügung oder Streichung von Gegenständen

Die Prüfung der in diesem Dokument beschriebenen Sachverhalte ergab, dass die Überführung von ST.25 nach ST.26 als solche nicht zwangsläufig zur Hinzufügung oder Streichung von Gegenständen führen sollte, insbesondere dann nicht, wenn das Sequenzprotokoll nach ST.25 vollständig mit dem Standard ST.25 konform war. Dennoch sollte der Anmelder bei bestimmten Szenarien Vorsicht walten lassen. In diesem Anhang werden Empfehlungen gegeben, um die Hinzufügung oder Streichung von Gegenständen zu vermeiden.

Szenario 1

In ST.25 werden für verschiedene Datentypen numerische Kennzahlen verwendet, z. B. <110> für den Namen des Anmelders. Die zur Auszeichnung der Daten verwendeten Elemente und ihre Attribute sind in ST.26 mit englischen Ausdrücken benannt.

Empfehlung:

Da die englischsprachigen Ausdrücke in ST.26 lediglich die Art des Dateninhalts beschreiben; stellt die Verwendung der entsprechenden Elementnamen und -attribute keinen hinzugefügten Gegenstand dar.

Szenario 2

ST.26 verlangt ausdrücklich die Aufnahme von: a) verzweigten Sequenzen, b) Sequenzen mit D-Aminosäuren, c) Nukleotid-Analoga und d) Sequenzen mit abasischen Stellen. In ST.25 ist nicht klar geregelt, ob solche Sequenzen aufgenommen werden müssen oder nicht aufgenommen werden dürfen.

Empfehlung:

Die in der Anmeldung enthaltene Offenbarung sollte ausreichen, um diese Sequenzen in einem Sequenzprotokoll nach ST.26 darzustellen, auch wenn sie im Sequenzprotokoll nach ST.25 womöglich nicht enthalten waren. Bei bestimmten Arten von Informationen, die nach ST.26 erforderlich sind, muss darauf geachtet werden, dass kein weiterer Gegenstand hinzugefügt wird, der über die Offenbarung hinausgeht, siehe z. B. die nachfolgenden Ausführungen (in Szenario 4) über den Qualifier "mol_type" für Nukleotidsequenzen.

Szenario 3

Nach ST.26 werden Sequenzen mit weniger als zehn spezifisch definierten Nukleotiden (ohne "n") und weniger als vier spezifisch definierten Aminosäuren (ohne "X") ausgeschlossen.

Empfehlung:

Die ausgeschlossenen Sequenzen können in den Hauptteil der Anmeldung aufgenommen werden, sofern sie darin noch nicht enthalten sind.

Szenario 4

Sowohl für Nukleotidsequenzen als auch für Aminosäuresequenzen hat ST.26 den obligatorischen Merkmalschlüssel "source" mit zwei obligatorischen Qualifiern, von denen einer "mol_type" ist. In ST.25 gibt es einen entsprechenden Merkmalschlüssel für

[11] Soweit in diesem Anhang auf den „ST.25“, den „WIPO-Standard ST.25“ oder den „Standard ST.25“verwiesen wird, werden damit die Empfehlungen des WIPO-Standards ST.25 für die Darstellung von Nukleotid- und Aminosäuresequenzprotokollen in Patentanmeldungen bezeichnet, die in § 11 und Anlage 1 der Patentverordnung in der bis 30. Juni 2022 geltenden Fassung umgesetzt sind.

Nukleotidsequenzen (der selten verwendet wird) ohne entsprechende Qualifier und es gibt keinen entsprechenden Merkmalschlüssel für Aminosäuresequenzen.

Nukleotidsequenzen

ST.26 – Merkmalschlüssel 5.37 "source"; obligatorischer Qualifier 6.39 "mol_type" (siehe ST.26 Absatz 75)

Qualifier	Wert
mol_type	genomic DNA
	genomic RNA
	mRNA
	tRNA
	rRNA
	other DNA (für synthetische Moleküle)
	other RNA (für synthetische Moleküle)
	transcribed RNA
	viral cRNA
	unassigned DNA (wenn das In-vivo-Molekül unbekannt ist)
	unassigned RNA (wenn das In-vivo-Molekül unbekannt ist)

Aminosäuresequenzen

ST.26 – Merkmalschlüssel 7.30 "source"; obligatorischer Qualifier 8.1 "mol_type" (siehe ST.26 Absatz 75)

Qualifier	Wert
mol_type	protein

Empfehlung:

Zu beachten ist lediglich das kontrollierte Vokabular, dem die Werte des Qualifiers "mol_type" für Nukleotidsequenzen zu entnehmen sind. Einige der oben aufgeführten Werte, die zur Auswahl stehen, werden in der Offenbarung möglicherweise nicht ausreichend gestützt. Die Hinzufügung von Gegenständen kann jedoch vermieden werden, indem für eine bestimmte Sequenz der allgemeinste Wert verwendet wird, z. B. "other DNA" und "other RNA" für ein synthetisches Molekül und "unassigned DNA" und "unassigned RNA" für ein In-vivo-Molekül.

Szenario 5

Wenn eine Sequenz "Xaa" enthält, müssen nach ST.25 weitere Informationen zu diesem Rest in das Feld <223> aufgenommen werden, das mit den Feldern <221> (Name/Schlüssel des Merkmals) und <222> (Lage des Merkmals) einhergeht. In ST.25 ist kein Standardwert für "Xaa" (wie "X" in ST.26) vorgesehen. In ST.26 hingegen ist ein solcher Standardwert vorgesehen, sodass weitere Informationen nicht in jedem Fall erforderlich sind. Zwei der am häufigsten verwendeten Annotationen in Peptidsequenzen sind "jede beliebige Aminosäure" oder "jede natürlich vorkommende Aminosäure" für die Variable "Xaa" bzw. "X". Diese Ausdrucksweise könnte so ausgelegt werden, dass sie auch andere Aminosäuren enthält als diejenigen, die in den Aminosäuretabellen in ST.25 oder ST.26 aufgeführt sind. Der ST.26-Standardwert für "X" ohne weitere Annotation ist jede beliebige der 22 in Anhang I aufgeführten Aminosäuren (siehe Abschnitt 3, Tabelle 3). Dieser ST.26-Standardwert kann selbst einen hinzugefügten oder gestrichenen Gegenstand darstellen und daher den Umfang einer Patentanmeldung beim Übergang von ST.25 zu ST.26 beeinträchtigen.

Empfehlungen:

(a) Wenn das Sequenzprotokoll nach ST.25 einen Merkmalsnamen im Feld <221>, eine Xaa entsprechende Lage des Merkmals im Feld <222> und sonstige Informationen zu Xaa im Feld <223> enthält und der in <221> angegebene Merkmalsname auch ein geeigneter Merkmalschlüssel in ST.26 ist, z. B. "SITE", "VARIANT" oder "UNSURE", sollte der in ST.26 vorgesehene Merkmalschlüssel verwendet werden. Um die mögliche Streichung von Gegenständen zu vermeiden, müssen außerdem die im Feld <223> angegebenen Informationen in einen begleitenden "note"-Qualifier aufgenommen werden.

(b) Wenn das Sequenzprotokoll nach ST.25 einen Merkmalsnamen im Feld <221>, eine Xaa entsprechende Lage des Merkmals im Feld <222> und sonstige Informationen zu Xaa im Feld <223> enthält und der in <221> angegebene Merkmalsname in ST.26 nicht als Merkmalschlüssel vorgesehen ist, sollte je nach Eignung der ST.26-Merkmalschlüssel "SITE" oder "REGION" verwendet werden. Um die mögliche Streichung von Gegenständen zu vermeiden, müssen außerdem die im Feld <223> angegebenen Informationen sowie der im Feld <221> angegebene nicht

geeignete Merkmalsname in einen begleitenden "note"-Qualifier aufgenommen werden. Ein Beispiel: In einem ST.25-Protokoll wurde ein Merkmalsname verwendet, der weder in ST.25 noch in ST.26 vorgesehen ist. Er wurde im Feld <221> als "Variable" angegeben und im Feld <223> mit der sonstigen Information versehen: „Xaa ist jede beliebige Aminosäure". In diesem Beispiel wäre der Wert des "note"-Qualifiers nach ST.26: "Variable – Xaa is any amino acid" ("Variable –Xaa ist jede beliebige Aminosäure").

(c) Wenn das Sequenzprotokoll nach ST.25 kein Feld mit der Kennzahl <221>, <222> oder <223> enthält, das Xaa entspricht, oder wenn zwar die Felder <221> und <222> für Xaa ausgefüllt, aber keine Informationen in einem entsprechenden Feld <223> vorhanden sind (keines der beiden Szenarien ist mit ST.25 konform, kommt aber dennoch vor), sollten alle Informationen, die im Hauptteil der Anmeldung zur Beschreibung von "Xaa" enthalten sind, in den Qualifier "note" nach ST.26 zusammen mit einem geeigneten Merkmalschlüssel, z. B. "SITE", "REGION" oder "UNSURE", sowie der Lage aufgenommen werden.

Szenario 6

Nach ST.25 wird Uracil in der Sequenz durch "u" und Thymin durch "t" dargestellt. Nach ST.26 werden Uracil und Thymin in der Sequenz durch "t" und ohne weitere Annotation dargestellt; "t" steht für Uracil in RNA und für Thymin in DNA.

Empfehlungen:

(a) Eine DNA-Sequenz, die Uracil enthält, wird nach ST.26 als modifiziertes Nukleotid betrachtet. Das Uracil muss als "t" dargestellt und mit dem Merkmalschlüssel "modified_base" näher beschrieben werden. Dabei sind der Qualifier "mod_base" mit "OTHER" als Qualifier-Wert und der Qualifier "note" mit "uracil" ("Uracil") als Qualifier-Wert zu verwenden. Diese Annotation nach ST.26 gilt nicht als hinzugefügter Gegenstand, wenn die DNA-Sequenz nach ST.25 ein "u" enthielt.

(b) Eine RNA-Sequenz, die Thymin enthält, wird nach ST.26 als modifiziertes Nukleotid betrachtet. Das Thymin muss als "t" dargestellt und mit dem Merkmalschlüssel "modified_base" näher beschrieben werden. Dabei sind der Qualifier "mod_base" mit "OTHER" als Qualifier-Wert und der Qualifier "note" mit "thymine" ("Thymin") als Qualifier-Wert zu verwenden. Diese Annotation nach ST.26 gilt nicht als hinzugefügter Gegenstand, wenn die RNA-Sequenz nach ST.25 ein "t" enthielt.

Szenario 7

Sowohl nach ST.25 als auch nach ST.26 müssen modifizierte Nukleotide oder Aminosäuren mit einer näheren Beschreibung versehen werden. Nach ST.26 kann die Identität eines modifizierten Nukleotids ggf. mit einer Abkürzung aus Anhang I, Abschnitt 2, Tabelle 2 angegeben werden. Andernfalls muss der vollständige, ungekürzte Name des modifizierten Nukleotids angegeben werden. Entsprechend kann die Identität einer modifizierten Aminosäure ggf. mit einer Abkürzung aus Anhang I, Abschnitt 4, Tabelle 4 angegeben werden. Andernfalls muss der vollständige, ungekürzte Name der modifizierten Aminosäure angegeben werden. Im Gegensatz dazu erfordert ST.25, wenn ein modifzierter Rest in keiner Tabelle von ST.25 enthalten ist, nicht die Verwendung des vollständigen, ungekürzten Namens, und nicht selten wird stattdessen eine Abkürzung verwendet.

Empfehlungen:

(a) Wurde sowohl in der Anmeldung als auch im Sequenzprotokoll nach ST.25 für ein modifiziertes Nukleotid oder eine modifizierte Aminosäure nur ein abgekürzter Name verwendet, der nicht in Anhang I, Abschnitt 2, Tabelle 2 oder Anhang I, Abschnitt 4, Tabelle 4 enthalten ist, und verweist der abgekürzte Name in der Fachwelt nur auf ein spezifisches modifiziertes Nukleotid oder eine spezifische modifizierte Aminosäure, so würde die Verwendung des vollständigen, ungekürzten Namens an sich keinen hinzugefügten Gegenstand darstellen.

(b) Wurde sowohl in der Anmeldung als auch im Sequenzprotokoll nach ST.25 für ein modifiziertes Nukleotid oder eine modifizierte Aminosäure nur ein abgekürzter Name verwendet, der nicht in Anhang I, Abschnitt 2, Tabelle 2 oder Anhang I, Abschnitt 4, Tabelle 4 enthalten ist (und wurde die chemische Struktur in der Anmeldung nicht dargestellt), und verweist der abgekürzte Name in der Fachwelt nicht auf ein spezifisches modifiziertes Nukleotid oder eine spezifische modifizierte Aminosäure, d. h., ist die Abkürzung in der Fachwelt entweder überhaupt nicht bekannt oder könnte sie unter Umständen mehrere verschiedene modifizierte Nukleotide oder modifizierte Aminosäuren darstellen, dann ist Konformität mit ST.26 nicht möglich, ohne dass ein Gegenstand hinzugefügt wird. Natürlich sind in einem solchen Fall die Prioritätsanmeldung und das Sequenzprotokoll selbst vage. Um die mögliche Streichung von Gegenständen zu vermeiden, sollte der abgekürzte Name aus dem Sequenzprotokoll nach ST.25 zusätzlich zu dem vollständigen, ungekürzten Namen des modifizierten Nukleotids oder der modifizierten Aminosäure als Wert in einen "note"- Qualifier nach ST.26 aufgenommen werden. Dem vollständigen, ungekürzten Namen des modifizierten Nukleotids oder der modifizierten Aminosäure, der in einem Sequenzprotokoll nach ST.26 erforderlich ist, wird die Priorität aus der früheren Anmeldung nicht gewährt. Um zukünftige Probleme zu vermeiden, ist darauf zu achten, dass das ursprüngliche Sequenzprotokoll (nach ST.25) und die Offenbarung der Anmeldung den ungekürzten Namen enthalten.

Szenario 8

ST.25 enthält eine Reihe von Merkmalschlüsseln, die in ST.26 nicht enthalten sind. Daher müssen Anmelder darauf achten, die in diesen ST.25-Merkmalschlüsseln enthaltenen Informationen in einer mit ST.26 kompatiblen Weise zu erfassen, ohne dass Gegenstände hinzugefügt oder gestrichen werden.

Empfehlungen:

Der folgenden Tabelle ist zu entnehmen, wie die in einem früheren ST.25-Merkmalschlüssel enthaltenen Informationen in Konformität mit ST.26 aufgenommen werden können, ohne dass Gegenstände hinzugefügt oder gestrichen werden. Die Merkmalschlüssel mit den Nummern 1–23 beziehen sich auf Nukleotidsequenzen, und diejenigen mit den Nummern 24–43 beziehen sich auf Aminosäuresequenzen.

Nr.	Merkmalschlüssel in ST.25 <221>	Äquivalent in ST.26		
		Merkmalschlüssel	Qualifier	Qualifier-Wert
1	allele	misc_feature	allele	Wert aus <223>
2	attenuator	regulatory[12]	regulatory_class[12]	"attenuator"
			note *(wenn <223> vorhanden)*	Wert aus <223>
3	CAAT_signal	regulatory[12]	regulatory_class[12]	"CAAT_signal"
			note *(wenn <223> vorh.)*	Wert aus <223>
4	conflict	misc_feature	note	"conflict" und Wert aus <223>
5	enhancer	regulatory[12]	regulatory_class[12]	"enhancer"
			note *(wenn <223> vorh.)*	Wert aus <223>
6	GC_signal	regulatory[12]	regulatory_class[12]	"GC_signal"
			note *(wenn <223> vorhanden)*	Wert aus <223>
7	LTR	mobile_element[12]	rpt_type[12]	"long_terminal_repeat"
			note *(wenn <223> vorh.)*	Wert aus <223>
8	misc_signal	regulatory[12]	regulatory_class[12]	"other"
			note *(wenn <223> vorh.)*	Wert aus <223>
9	mutation	variation	note	"mutation" und Wert aus <223>
10	old_sequence	misc_feature	note	"old_sequence" und Wert aus <223>
11	polyA_signal	regulatory[12]	regulatory_class[12]	"polyA_signal_sequence"
			note *(wenn <223> vorh.)*	Wert aus <223>
12	promoter	regulatory[12]	regulatory_class[12]	"promoter"
			note *(wenn <223> vorh.)*	Wert aus <223>
13	RBS	regulatory[12]	regulatory_class[12]	"ribosome_binding_site"
			note *(wenn <223> vorh.)*	Wert aus <223>
14	repeat_unit (a), wenn repeat_region nicht verwendet wird	misc_feature	note	"repeat_unit" und Wert aus <223>
	repeat_unit (b) wenn repeat_region verwendet wird	repeat_region	rpt_unit_range	erster Rest..letzter Rest
			note *(wenn <223> vorhanden)*	Wert aus <223>
15	satellite	repeat_region	satellite	"satellite" (oder "microsatellite" oder "minisatellite" – falls unterstützt)
			note *(wenn <223> vorh.)*	Wert aus <223>
16	scRNA	ncRNA[12]	ncRNA_class[12]	"scRNA"
			note *(wenn <223> vorh.)*	Wert aus <223>
17	snRNA	ncRNA[12]	ncRNA_class[12]	"snRNA"
			note *(wenn <223> vorh.)*	Wert aus <223>
18	TATA_signal	regulatory[12]	regulatory_class[12]	"TATA_box"[13]
			note	TATA_signal und Wert aus <223> (falls <223> vorhanden)
19	terminator	regulatory[12]	regulatory_class[12]	"terminator"
			note *(wenn <223> vorh.)*	Wert aus <223>
20	3'clip	misc_feature	note	"3'clip" und Wert aus <223>
21	5'clip	misc_feature	note	"5'clip" und Wert aus <223>
22	-10_signal	regulatory[12]	regulatory_class[12]	"minus_10_signal"
			note *(wenn <223> vorh.)*	Wert aus <223>
23	-35_signal	regulatory[12]	regulatory_class[12]	"minus_35_signal"

			note *(wenn <223> vorh.)*	Wert aus <223>

Nr.	Merkmalschlüssel in ST.25 <221>	Äquivalent in ST.26		
		Merkmalschlüssel	**Qualifier**	**Qualifier-Wert**
24	NON_CONS	Dieses Merkmal bezieht sich auf eine aus einer unbekannten Anzahl von Resten bestehenden Lücke in einer einzelnen Sequenz, was sowohl in ST.25 (Absatz 22) als auch in ST.26 (Absatz 37) verboten ist. Folglich muss jede aus spezifisch definierten Resten bestehende Region, die von ST.26 Absatz 7 erfasst wird, im Sequenzprotokoll als separate Sequenz aufgeführt und mit einer eigenen Sequenzkennzahl versehen werden. Um die Hinzufügung oder Streichung von Gegenständen zu vermeiden, muss jede solche Sequenz mit dem Hinweis versehen werden, dass sie Teil einer größeren Sequenz ist, die eine undefinierte Lücke enthält.		
		REGION	note	Beschreibung
		Beschreibung – wo und woran die Sequenz gebunden ist, z. B.: "This residue is linked N-terminally to a peptide having an N-terminal Gly-Gly and a gap of undefined length." ("Dieser Rest ist über eine N-Terminus-Bindung an ein Peptid mit N-terminalem Gly-Gly und einer Lücke von nicht definierter Länge gebunden. ")		
25	SIMILAR	REGION	note	"SIMILAR" und Wert aus <223>, falls vorhanden
26	THIOETH	CROSSLNK	note	"THIOETH" und Wert aus <223>, falls vorhanden
		Weitere Informationen zur Angabe der Lage finden Sie in ST.26 Anhang I, Anmerkung zum Merkmalschlüssel CROSSLNK		
27	THIOLEST	CROSSLNK	note	"THIOLEST" und Wert aus <223>, falls vorhanden
		Weitere Informationen zur Angabe der Lage finden Sie in ST.26 Anhang I, Anmerkung zum Merkmalschlüssel "CROSSLNK"		
28	VARSPLIC	Wird in Szenario 13 erläutert		
29	ACETYLATION	MOD_RES	note	"ACETYLATION" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "MOD_RES" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
30	AMIDATION	MOD_RES	note	"AMIDATION" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "MOD_RES" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
31	BLOCKED	MOD_RES	note	"BLOCKED" und Wert aus <223>, falls vorhanden

[12] Nach ST.26 muss ggf. ein bestimmtes ST.25-Merkmal, z. B. "TATA_signal", durch einen Merkmalschlüssel/Qualifier/Qualifier-Wert mit breiterer Bedeutung ersetzt werden, z. B. "regulatory/regulatory_class/TATA_box".

[13] Damit das Hinzufügen von Gegenständen nicht zu einem teilweisen Prioritätsverlust führt, empfiehlt es sich, den etwas eingeschränkteren Begriff "TATA_signal" in einen "note"-Qualifier aufzunehmen, wie in vorstehender Tabelle dargestellt (Punkt 18). Wenn der Anmelder in seltenen Fällen die Verwendung des "TATA_box"-Werts für den "regulatory_class"-Qualifier für ungeeignet hält, kann anstelle des "TATA_box"-Werts der Wert "other" ("sonstige") verwendet werden. In diesem Fall muss der Begriff "TATA_signal" in einen "note"-Qualifier des Merkmalschlüssels "regulatory" aufgenommen werden.

			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "MOD_RES" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
32	FORMYLATION	MOD_RES	note	"FORMYLATION" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "MOD_RES" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)

Nr.	Merkmalschlüssel in ST.25 <221>	Äquivalent in ST.26		
		Merkmalschlüssel	Qualifier	Qualifier-Wert
33	GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	MOD_RES	note	"GAMMA-CARBOXYLGLUTAMIC ACID HYDROXYLATION" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "MOD_RES" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
34	METHYLATION	MOD_RES	note	"METHYLATION" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "MOD_RES" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
35	PHOSPHORYLATION	MOD_RES	note	"PHOSPHORYLATION" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "MOD_RES" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
36	PYRROLIDONE CARBOXYLIC ACID	MOD_RES	note	"PYRROLIDON CARBOXYLIC ACID" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "MOD_RES" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
37	SULFATATION	MOD_RES	note	"SULFATATION" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "MOD_RES" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
38	MYRISTATE	LIPID	note	"MYRISTATE" und Wert aus <223>, falls vorhanden

			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "LIPID" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
39	PALMITATE	LIPID	note	"PALMITATE" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "LIPID" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
40	FARNESYL	LIPID	note	"FARNESYL" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "LIPID" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
41	GERANYL-GERANYL	LIPID	note	"GERANYL-GERANYL" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "LIPID" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
Nr.	**Merkmalschlüssel in ST.25 <221>**	**Äquivalent in ST.26**		
		Merkmalschlüssel	**Qualifier**	**Qualifier-Wert**
42	GPI-ANCHOR	LIPID	note	"GPI-ANCHOR" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "LIPID" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)
43	N-ACYL DIGLYCERIDE	LIPID	note	"N-ACYL DIGLYCERIDE" und Wert aus <223>, falls vorhanden
			note	Nach ST.26 Anhang I Anmerkung zu Merkmalschlüssel "LIPID" erforderliche Informationen, sofern möglich (ohne hinzugefügte Gegenstände)

Szenario 9

Für bestimmte Merkmalschlüssel, die sowohl in ST.25 als auch in ST.26 für Nukleotid- und Aminosäuresequenzen vorhanden sind, gibt es in ST.26 obligatorische Qualifier, wie unten angegeben. Auch den Merkmalschlüssel "modified_base" für Nukleotidsequenzen gibt es sowohl in ST.25 als auch in ST.26; hierzu enthält Szenario 7 jedoch bestimmte Empfehlungen. In ST.25 gab es keine Qualifier, sondern das Freitextfeld <223>. Wenn die im Feld <223> von ST.25 enthaltenen Informationen als Wert für den obligatorischen Qualifier in ST.26 geeignet sind, sollten sie als solcher eingetragen werden. Wenn das Feld <223> von ST.25 entweder nicht ausgefüllt wurde oder Informationen enthält, die nicht als Wert für den obligatorischen Qualifier in ST.26 geeignet sind, müssen Anmelder darauf achten, die im Merkmalschlüssel bzw. Feld <223> von ST.25 enthaltenen Informationen in einer Weise zu erfassen, die mit ST.26 konform ist, ohne dass ein Gegenstand hinzugefügt oder gestrichen wird.

Nukleotidsequenzen[14]

Merkmalschlüssel	Obligatorischer Qualifier
5.12 – misc_binding	6.3 – bound_moiety
5.30 – protein_bind	6.3 – bound_moiety

Empfehlungen:

(a) Wenn das Feld <223> von ST.25 fehlt oder ungeeignet ist und in der Beschreibung der Anmeldung der Name des Moleküls/Komplexes, das/der an die Lage des Merkmals der Nukleinsäure binden kann, offenbart wird, sollte dieser Name in den Qualifier "bound_moiety" aufgenommen werden.

(i) Alle im Feld <223> von ST.25 enthaltenen Informationen, die nicht für die Aufnahme in den Qualifier "bound_moiety" geeignet sind, sollten in einen entsprechenden fakultativen Qualifier des Merkmalschlüssels, z. B. "note", eingetragen werden.

(b) Wenn das Feld <223> von ST.25 fehlt oder ungeeignet ist und in der Beschreibung der Anmeldung der Name des Moleküls/Komplexes, das/der an die Lage des Merkmals der Nukleinsäure binden kann, nicht offenbart wird, sollte anstelle der Merkmalschlüssel "misc_binding" oder "protein_bind" der Merkmalschlüssel "misc_feature" in Verbindung mit dem Qualifier "note" verwendet werden.

(i) Wenn das Feld <223> von ST.25 fehlte, sollte der Name des Merkmalschlüssels von ST.25 als Wert des Qualifiers "note" eingesetzt werden.

(ii) Wenn das Feld <223> von ST.25 ungeeignete Informationen enthielt, sollten der Name des Merkmalschlüssels von ST.25 und die Informationen aus Feld <223> als Wert des Qualifiers "note" eingesetzt werden.

Aminosäuresequenzen[15]

Merkmalschlüssel	Obligatorischer Qualifier
7.2 – BINDING	8.2 – note
7.4 – CARBOHYD	8.2 – note
7.10 – DISULFID	8.2 – note
7.11 – DNA_BIND	8.2 – note
7.12 – DOMAIN	8.2 – note
7.16 – LIPID	8.2 – note
7.17 – METAL	8.2 – note
7.18 – MOD_RES	8.2 – note
7.23 – NP_BIND	8.2 – note
7.29 – SITE	8.2 – note
7.39 – ZN_FING	8.2 – note

Empfehlungen:

(a) Wenn das Feld <223> von ST.25 fehlt oder ungeeignet ist und in der Beschreibung der Anmeldung die für den obligatorischen Qualifier erforderlichen spezifischen Informationen offenbart wurden, dann sollten diese Informationen in den obligatorischen Qualifier "note" aufgenommen werden.

(i) Alle im Feld <223> von ST.25 enthaltenen Informationen, die nicht für die Aufnahme in den obligatorischen Qualifier "note" geeignet sind (siehe Definition des Merkmalschlüssels und Anmerkung dazu), sollten in einen zweiten "note"-Qualifier eingetragen werden.

(b) Wenn das Feld <223> von ST.25 fehlt oder ungeeignet ist und in der Beschreibung der Anmeldung die für den obligatorischen Qualifier erforderlichen spezifischen Informationen nicht offenbart wurden, sollte stattdessen der ST.26-

[14] Die numerischen Verweise in der folgenden Tabelle beziehen sich auf die Merkmalschlüssel und die Qualifier von ST.26 Anhang I, Kontrolliertes Vokabular.

[15] Die numerischen Verweise in der folgenden Tabelle beziehen sich auf die Merkmalschlüssel und die Qualifier von ST.26 Anhang I, Kontrolliertes Vokabular.

Merkmalschlüssel "SITE" (für eine Aminosäure) oder "REGION" (für einen Bereich von Aminosäuren) in Verbindung mit dem Qualifier "note" verwendet werden.

(i) Wenn das Feld <223> von ST.25 fehlt, sollte der Name des Merkmalschlüssels von ST.25 als Wert des Qualifiers "note" eingesetzt werden.

(ii) Wenn das Feld <223> von ST.25 ungeeignete Informationen enthielt, sollten der Name des Merkmalschlüssels von ST.25 und die Informationen aus Feld <223> als Wert des Qualifiers "note" eingesetzt werden.

Szenario 10

Zu jedem spezifischen Merkmalschlüssel von ST.25 gehört ein Feld <222>, in dem die Lage des Merkmals angegeben werden kann. Für die meisten Merkmale ist nach ST.25 jedoch keine Angabe der Lage des Merkmals erforderlich, und das Format dieser Angabe ist nicht standardisiert. Darüber hinaus gibt es in ST.25 keine Lageoperatoren wie z. B. "join". In ST.26 gibt es standardisierte Lagedeskriptoren und -operatoren, und jedes Merkmal muss mindestens einen Lagedeskriptor enthalten. („CDS“-Merkmale sind ein Sonderfall und werden in Szenario 11 erläutert.)

Empfehlungen:

(a) Wenn das Sequenzprotokoll nach ST.25 ein Feld <222> aufwies, sollte der direkte Import oder der Import in das ST.26-Format nicht zu Bedenken im Hinblick auf hinzugefügte Gegenstände führen.

(b) Wenn das Sequenzprotokoll nach ST.25 kein Feld <222> aufwies, die Angabe der Lage jedoch in der Beschreibung der Anmeldung enthalten war, sollte der direkte Import oder der Import in das ST.26-Format nicht zu Bedenken im Hinblick auf hinzugefügte Gegenstände führen.

(c) Wenn weder das Sequenzprotokoll nach ST.25 noch die Beschreibung der Anmeldung Angaben zur Lage enthalten, wird davon ausgegangen, dass das Merkmal für die gesamte Sequenz gilt. (Die Angabe einer Lage, die weniger als die gesamte Sequenz umfasst, ohne dass dies durch die Beschreibung der Anmeldung gestützt wird, würde wahrscheinlich einen hinzugefügten/gestrichenen Gegenstand darstellen.) Um zukünftige Probleme zu vermeiden, ist darauf zu achten, dass das ursprüngliche Sequenzprotokoll (nach ST.25) und die Offenbarung der Anmeldung möglichst umfassende Angaben zur Lage enthalten.

Szenario 11

Im Sequenzprotokoll nach ST.25 wurde eine codierende Sequenz, die für ein einzelnes, zusammenhängendes Polypeptid codierte, aber durch eine oder mehrere nicht codierende Sequenz(en), z. B. Introns, unterbrochen war, als mehrere separate "CDS"-Merkmale dargestellt, wie unten gezeigt:

<220>
<221> CDS
<222> (1)..(571)

<220>
<221> CDS
<222> (639)..(859)

Im Gegensatz dazu verfügt ST.26 über den Lageoperator „join", der angibt, dass die an den angegebenen Lagen codierten Polypeptide verbunden sind und ein einzelnes, zusammenhängende Polypeptid bilden. (Hinweis: Sowohl ST.25 als auch ST.26 verlangen, dass das Stopcodon in die Lage des Merkmals "CDS" einbezogen wird.)

Empfehlungen:

(a) Wenn aus dem Sequenzprotokoll nach ST.25 oder der Beschreibung der Anmeldung eindeutig hervorgeht, dass die Polypeptidsequenzen, die von mehreren separaten "CDS"-Merkmalen codiert werden, ein einzelnes, zusammenhängendes Polypeptid bilden, muss eine in einem einzelnen "CDS"-Merkmal enthaltene codierende Sequenz, die von einem Intron unterbrochen wird, wie folgt mit dem Lageoperator "join" dargestellt werden, so dass keine Gegenstände hinzugefügt werden:

<INSDFeature_key>CDS</INSDFeature_key>
<INSDFeature_location>join(1..571,639..859)</INSDFeature_location>

(b) Wenn aus dem Sequenzprotokoll nach ST.25 oder der Beschreibung der Anmeldung nicht hervorgeht, dass die Polypeptidsequenzen, die von den beiden separaten "CDS"-Merkmalen codiert werden, ein einzelnes,

zusammenhängendes Polypeptid bilden, würde die Verwendung des Lageoperators "join" wahrscheinlich einen zusätzlichen Gegenstand darstellen.

Szenario 12

Nach ST.25 müssen die Merkmalsnamen aus Tabelle 5 oder 6 stammen. Nach den US-amerikanischen Vorschriften hingegen waren diese Merkmalsnamen empfohlen, aber nicht vorgeschrieben. Daher kann eine Sequenz in einem Sequenzprotokoll nach ST.25 (gemäß US-Vorschriften) einen "benutzerdefinierten" Namen für einen Merkmalschlüssel enthalten, für den es in ST.26 keine Entsprechung gibt. Möglich ist auch, dass für das Feld <221> kein Merkmalsname angegeben wurde oder dass das Feld <221> nicht vorhanden ist. Diese Szenarien können in ähnlicher Weise behandelt werden.

Empfehlung:

Der "benutzerdefinierte" Name eines Merkmalschlüssels aus ST.25 kann in einem Sequenzprotokoll nach ST.26 wie folgt dargestellt werden, ohne dass Gegenstände hinzugefügt werden:

Art	**Merkmalschlüssel in ST.25 <221>**	**Potenzielles Äquivalent in ST.26**		
		Merkmalschlüssel	**Qualifier**	**Qualifier-Wert**
Nukleinsäure	"benutzerdefinierter" Merkmalschlüssel	misc_feature	note	Name des "benutzerdefinierten" Merkmalschlüssels und Wert aus <223>, falls vorhanden
Aminosäure	"benutzerdefinierter" Merkmalschlüssel	SITE oder REGION	note	Name des "benutzerdefinierten" Merkmalschlüssels und Wert aus <223>, falls vorhanden

Szenario 13

ST.25 enthält den Merkmalschlüssel "VARSPLIC", definiert als "Beschreibung von Sequenzvarianten, die durch alternatives Spleißen entstanden sind". In ST.26 wurde "VARSPLIC" durch den breiter gefassten Merkmalschlüssel "VAR_SEQ" ersetzt. Dieser dient laut Definition zur "Beschreibung von Sequenzvarianten, die durch alternatives Spleißen, alternative Promotoren, alternative Initiierung und ribosomale Leserasterverschiebung erzeugt werden". Daher sollte in Sequenzprotokollen nach ST.26 "VAR_SEQ" nicht ohne nähere Erläuterung als Ersatz für "VARSPLIC" verwendet werden.

Empfehlung:

In ST.26 sollte das Merkmal "VAR_SEQ" in Verbindung mit dem Qualifier "note" verwendet werden. Der Qualifier-Wert sollte eine Erläuterung der enger gefassten Bedeutung nach ST.25 enthalten, z. B. "sequence variant produced by alternative splicing" ("Sequenzvariante durch alternatives Spleißen entstanden"). Etwaige zusätzliche Informationen, die nach ST.25 in dem zugehörigen Feld <223> enthalten sind, sollten ebenfalls in den Qualifier "note" aufgenommen werden.

Szenario 14

Wenn die Sequenz künstlicher Herkunft war, muss nach ST.25 im Feld <213> für den Organismus „künstliche Sequenz" eingetragen werden. In ST.26 ist für den Merkmalschlüssel "source" der Qualifier "organism" vorgeschrieben, dessen Wert als „synthetic construct" ("synthetisches Konstrukt") und nicht als „künstliche Sequenz" angegeben werden muss.

Empfehlung:

Als Wert für den Qualifier "organism" muss nach ST.26 "synthetic construct" ("synthetisches Konstrukt") angegeben werden. Um zu verhindern, dass Gegenstände gestrichen werden, sollten alle erläuternden Informationen, die im erforderlichen Feld <223> von ST.25 enthalten sind, in einen "note"-Qualifier (zum Merkmalschlüssel "source") aufgenommen werden.

Szenario 15

Wenn der wissenschaftliche Name des Organismus, dem die Sequenz entstammt, unbekannt ist, muss im entsprechenden Feld <213> von ST.25 „unbekannt" eingetragen werden. In ST.26 ist für den Merkmalschlüssel "source" der Qualifier "organism" vorgeschrieben, dessen Wert als "unidentified" ("unidentifiziert") und nicht als "unbekannt" angegeben werden muss.

Empfehlung:

Als Wert für den Qualifier "organism" muss nach ST.26 "unidentified" ("unidentifiziert") angegeben werden. Um zu verhindern, dass Gegenstände gestrichen werden, sollten alle erläuternden Informationen, die im erforderlichen Feld <223> von ST.25 enthalten sind, in einen "note"-Qualifier (zum Merkmalschlüssel "source") aufgenommen werden.

Szenario 16

Nach ST.25 können die dem reifen Protein vorausgehenden Aminosäuren wie beispielsweise Präsequenzen, Prosequenzen, Prä-Prosequenzen und Signalsequenzen, mit negativem Vorzeichen nummeriert werden, wobei die Rückwärtszählung mit der Aminosäure vor Nummer 1 beginnt. ST.26 lässt keine Zahlen mit negativen Vorzeichen in der Lage des Merkmals zu.

Empfehlungen:

(a) Wenn im Sequenzprotokoll nach ST.25 ein oder mehrere Merkmale in einem <221>-Feld und einem zugehörigen <222>-Feld mit negativer und/oder positiver Nummerierung, z. B. "PROPEP" und/oder "CHAIN", dargestellt waren, sollte im Sequenzprotokoll nach ST.26 der entsprechende Merkmalschlüssel, z. B. "PROPEP" und/oder "CHAIN", verwendet werden. Sofern in einem <223>-Feld Informationen enthalten waren, können diese als Qualifier-Wert für einen "note"-Qualifier verwendet werden.

(b) Wenn im Sequenzprotokoll nach ST.25 kein Merkmal in einem <221>-Feld und einem zugehörigen <222>-Feld dargestellt war, sondern die entsprechenden Informationen zur negativen und/oder positiven Nummerierung in der Beschreibung der Anmeldung enthalten waren, dann sollte im Sequenzprotokoll nach ST.26 der entsprechende Merkmalschlüssel, z. B. "PROPEP" und/oder "CHAIN", verwendet werden. Andernfalls kann der Merkmalschlüssel "REGION" verwendet werden. Informationen aus der Beschreibung der Anmeldung können als Qualifier-Wert für einen "note"-Qualifier verwendet werden.

(c) Wenn weder das Sequenzprotokoll nach ST.25 noch die Beschreibung der Anmeldung Informationen enthält, die die negative und/oder positive Nummerierung erklären, dann sollte im Sequenzprotokoll nach ST.26 der Merkmalschlüssel "REGION" verwendet werden, um die Streichung von Gegenständen auszuschließen. Die Lage des Merkmals erstreckt sich dann über die Region, die im Sequenzprotokoll nach ST.25 mit negativen Vorzeichen nummeriert ist. Außerdem sollte in einem "note"-Qualifier angezeigt werden, dass die Aminosäuresequenz im nach ST.25 erstellten Sequenzprotokoll der Anmeldung, aus der ein Prioritätsanspruch abgeleitet wird, mit negativen Vorzeichen nummeriert wurde.

Szenario 17

ST.25 enthält in den Feldern <300> bis <313> Veröffentlichungsangaben. Solche Angaben sind nach ST.26 nicht vorgesehen.

Empfehlung:

Die in den Feldern <300> bis <313> von ST.25 enthaltenen Informationen sollten, sofern nicht bereits darin enthalten, in den Hauptteil der Anmeldung aufgenommen werden.

Szenario 18

ST.25 bietet keine standardisierte Möglichkeit, um anzugeben, dass eine CDS-Region einer Nukleotidsequenz mithilfe einer anderen Tabelle für den genetischen Code als der Tabelle für den Standardcode translatiert werden soll. Im Gegensatz dazu gibt es in ST.26 den Qualifier "transl_table", der mit dem Merkmalschlüssel "CDS" verwendet werden kann, um anzuzeigen, dass die Region unter Verwendung einer alternativen Tabelle für den genetischen Code translatiert werden soll. Wenn der Qualifier "transl_table" nicht verwendet wird, wird angenommen, dass die Tabelle für den genetischen Standardcode verwendet wird.

Empfehlungen:

(a) Wenn aus dem Sequenzprotokoll nach ST.25 oder der Beschreibung der Anmeldung eindeutig hervorgeht, dass eine CDS-Region mit einer alternativen Tabelle für den genetischen Code translatiert werden soll, muss der Qualifier "transl_table" mit der entsprechenden Nummer der Tabelle für den genetischen Code als Qualifier-Wert verwendet werden. Die Nichtverwendung des Qualifiers "transl_table" würde wahrscheinlich einen hinzugefügten Gegenstand darstellen, da in diesem Fall davon ausgegangen würde, dass die Tabelle für den Standardcode verwendet wird. Wenn die Angaben zu der alternativen Tabelle für den genetischen Code aus dem Sequenzprotokoll nach ST.25 oder aus der Beschreibung der Anmeldung nicht in das Sequenzprotokoll nach ST.26 aufgenommen werden, würde dies wahrscheinlich einen gestrichenen Gegenstand darstellen.

(b) Wenn aus dem Sequenzprotokoll nach ST.25 und der Beschreibung der Anmeldung nicht hervorgeht, dass eine CDS-Region mit einer alternativen Tabelle für den genetischen Code translatiert werden soll, sollte der Qualifier "transl_table" nicht oder ausschließlich mit dem Qualifier-Wert "1" für die Tabelle für den Standardcode verwendet werden. Die Verwendung des Qualifiers "transl_table" mit einem anderen Qualifier-Wert als "1" würde wahrscheinlich hinzugefügte oder gestrichene Gegenstände darstellen.

Szenario 19

In ST.25 ist keine standardisierte Methode zur Angabe der Lage eines Merkmals vorgesehen, insbesondere eines Merkmals an einer Stelle oder in einer Region, die über einen spezifischen Rest oder ein spezifisches Resteintervall hinausgeht, z. B. eine CDS-Region einer Nukleotidsequenz, die über ein oder beide Enden einer offenbarten Sequenz hinausreicht. Im Gegensatz dazu bietet der Lagedeskriptor (in "INSDFeature_location") in ST.26 eine standardisierte Möglichkeit, die Lage einer solchen Stelle oder Region durch die Symbole "<" oder ">" anzuzeigen. Zum Beispiel muss die Lage des Merkmals "CDS" auch das Stopcodon enthalten, indem sie z. B. mit "1..>321" angegeben wird, selbst wenn das Stopcodon nicht Bestandteil der offenbarten Sequenz selbst ist.

Empfehlungen:

(a) Wenn im Sequenzprotokoll nach ST.25 nicht ausdrücklich darauf hingewiesen wird, dass die Lage eines Merkmals über die Sequenz hinausgeht, eine solche Lage jedoch aus der Offenbarung oder der Sequenz selbst ersichtlich ist, z. B. aus dem nicht in der Sequenz enthaltenen Stopcodon eines "CDS"-Merkmals, können die Symbole "<" oder ">" im Sequenzprotokoll nach ST.26 verwendet werden, ohne dass Gegenstände hinzugefügt werden.

(b) Wenn im Sequenzprotokoll nach ST.25 nicht ausdrücklich darauf hingewiesen wird, dass die Lage eines Merkmals über die Sequenz hinausgeht, und eine solche Lage weder aus der Offenbarung noch aus der Sequenz ersichtlich ist, kann ST.26 womöglich nicht eingehalten werden, ohne dass Gegenstände hinzugefügt werden. In diesem Fall dürften die Prioritätsanmeldung und das Sequenzprotokoll selbst unvollständig sein. Unter diesen Umständen wird der Lagebeschreibung des Merkmals im Sequenzprotokoll nach ST.26 nicht die Priorität aus der früheren Anmeldung gewährt. Es ist darauf zu achten, dass das ursprüngliche Sequenzprotokoll (nach ST.25) und die Offenbarung der Anmeldung vollständige Angaben zum Merkmal enthalten.

Szenario 20

Nach ST.25 Anhang I ist, wenn eine Nukleotidsequenz DNA- und RNA-Fragmente enthält, als Wert im Feld <212> "DNA" anzugeben und das kombinierte DNA/RNA-Molekül im Merkmalsteil <220> bis <223> näher zu beschreiben. Die genaue Art der näheren Beschreibung ist jedoch nicht klar, und diese Anforderung wird nicht routinemäßig befolgt. ST.26 Absatz 55 verlangt, dass jedes DNA- und RNA-Segment (aus Gründen der internen Konsistenz wird in ST.26 der Ausdruck "Segment" statt "Fragment" verwendet) des kombinierten DNA/RNA-Moleküls mit dem Merkmalschlüssel "misc_feature", der die Lage des Segments enthält, und dem Qualifier "note", der angibt, ob es sich um DNA oder RNA handelt, näher beschrieben wird.

Empfehlungen:

(a) Wenn im Sequenzprotokoll nach ST.25 die DNA- und RNA-Segmente in einem oder mehreren Merkmalen beschrieben wurden, indem im Feld <221> der Merkmalschlüssel „misc_feature", im Feld <222> die jeweils zutreffende Lage und im Feld <223> angegeben wurde, bei welchen Segmenten es sich um DNA oder RNA handelt, dann sollte die Übernahme dieser Informationen in das ST.26-Format nicht zu Bedenken im Hinblick auf hinzugefügte Gegenstände führen, wenn für jedes einzelne DNA- und RNA-Segment der Merkmalschlüssel „misc_feature" verwendet wird.

(b) Wenn im Sequenzprotokoll nach ST.25 die DNA- und RNA-Segmente in einem oder mehreren Merkmalen beschrieben wurden, indem im Feld <221> ein anderer Merkmalschlüssel als "misc_feature", im Feld <222> die jeweils zutreffende Lage und im Feld <223> angegeben wurde, bei welchen Segmenten es sich um DNA oder RNA handelt, dann dürfte die Übernahme dieser Informationen in das ST.26-Format nicht zu Bedenken im Hinblick auf hinzugefügte oder gestrichene Gegenstände führen, wenn für jedes einzelne DNA- und RNA-Segment der Merkmalschlüssel "misc_feature" in Verbindung mit einem zusätzlichen "note"-Qualifier verwendet wird, in den der ursprüngliche Merkmalschlüssel aus dem Feld <221> als Qualifier-Wert eingetragen wird.

(c) Wenn im Sequenzprotokoll nach ST.25 die Identität (DNA oder RNA) und die Lage jedes Segments in einem <223>-Feld angegeben ist, das nicht mit einem <221>- und <222>-Feld verknüpft ist, und in diesem Feld z. B. die Erklärung für eine künstliche Sequenz enthalten ist, dann sollte die Übernahme dieser Informationen in das ST.26-Format nicht zu Bedenken im Hinblick auf hinzugefügte Gegenstände führen, wenn für jedes einzelne DNA- und RNA-Segment der Merkmalschlüssel "misc_feature" verwendet wird.

(d) Wenn im Sequenzprotokoll nach ST.25 das Molekül in einem Merkmal unter Verwendung des Merkmalschlüssels "misc_feature" im Feld <221> und der Angabe im Feld <223>, dass es sich um ein kombiniertes DNA/RNA-Molekül handelt, beschrieben wird, aber keine Angaben zur Lage jedes einzelnen Segments gemacht werden, und

(i) wenn in der Beschreibung die Lage jedes einzelnen DNA- und RNA-Segments angegeben ist, dann sollte die Übernahme dieser Informationen in das ST.26-Format nicht zu Bedenken im Hinblick auf hinzugefügte Gegenstände führen, wenn für jedes einzelne DNA- und RNA-Segment der Merkmalschlüssel "misc_feature" verwendet wird;

(ii) wenn die Beschreibung keine Angaben zur Lage jedes einzelnen DNA- und RNA-Segments enthält, kann ST.26 möglicherweise nicht eingehalten werden, ohne dass zusätzliche Gegenstände hinzugefügt werden. In diesem Fall dürften die Prioritätsanmeldung und das Sequenzprotokoll selbst unvollständig sein. Unter diesen Umständen wird keiner Lagebeschreibung der Merkmale im Sequenzprotokoll nach ST.26 die Priorität aus der früheren Anmeldung gewährt. Es ist darauf zu achten, dass das ursprüngliche Sequenzprotokoll (nach ST.25) und die Offenbarung der Anmeldung vollständige Angaben zum Merkmal enthalten.

(e) Wenn im Sequenzprotokoll nach ST.25 das Molekül in einem Merkmal unter Verwendung eines anderen Merkmalschlüssels als "misc_feature" im Feld <221> und der Angabe im Feld <223>, dass es sich um ein kombiniertes DNA/RNA-Molekül handelt, beschrieben wird, aber keine Angaben zur Lage jedes einzelnen Segments gemacht werden, und

(i) wenn in der Beschreibung die Lage jedes einzelnen DNA- und RNA-Segments angegeben ist, dann sollte die Übernahme dieser Informationen in das ST.26-Format nicht zur Berücksichtigung hinzugefügter oder gestrichener Gegenstände führen, wenn für jedes einzelne DNA- und RNA-Segment der Merkmalschlüssel "misc_feature" mit einem zusätzlichen "note"-Qualifier verwendet wird, in den der ursprüngliche Merkmalschlüssel aus dem Feld <221> als Wert eingetragen wird;

(ii) wenn die Beschreibung keine Angaben zur Lage jedes einzelnen DNA- und RNA-Segments enthält, kann ST.26 möglicherweise nicht eingehalten werden, ohne dass zusätzliche Gegenstände hinzugefügt werden. In diesem Fall dürften die Prioritätsanmeldung und das Sequenzprotokoll selbst unvollständig sein. Unter diesen Umständen wird keiner Lagebeschreibung der Merkmale im Sequenzprotokoll nach ST.26 die Priorität aus der früheren Anmeldung gewährt. Es ist darauf zu achten, dass das ursprüngliche Sequenzprotokoll (nach ST.25) und die Offenbarung der Anmeldung vollständige Angaben zum Merkmal enthalten.

(f) Wenn im Sequenzprotokoll nach ST.25, z. B. durch die Angabe einer künstlichen Sequenz in einem <223>-Feld, festgestellt wurde, dass es sich bei dem Molekül um ein kombiniertes DNA/RNA-Molekül handelt, aber keine Merkmalschlüssel oder Angaben zur Lage jedes einzelnen Segments enthalten sind, und

(i) wenn in der Beschreibung die Lage jedes einzelnen DNA- und RNA-Segments angegeben ist, dann sollte die Übernahme dieser Informationen in das ST.26-Format nicht zu Bedenken im Hinblick auf hinzugefügte Gegenstände führen, wenn für jedes einzelne DNA- und RNA-Segment der Merkmalschlüssel "misc_feature" verwendet wird;

(ii) wenn die Beschreibung keine Angaben zur Lage jedes einzelnen DNA- und RNA-Segments enthält, kann ST.26 möglicherweise nicht eingehalten werden, ohne dass zusätzliche Gegenstände hinzugefügt werden. In diesem Fall dürften die Prioritätsanmeldung und das Sequenzprotokoll selbst unvollständig sein. Unter diesen Umständen wird keiner Lagebeschreibung der Merkmale im Sequenzprotokoll nach ST.26 die Priorität aus der früheren Anmeldung gewährt. Es ist darauf zu achten, dass das ursprüngliche Sequenzprotokoll (nach ST.25) und die Offenbarung der Anmeldung vollständige Angaben zum Merkmal enthalten.

[Ende des Anhangs VII und des Standards]