

■ WIPO-Standard ST.26 WEITERFÜHRENDER KURS

Webinar-Schulung

Hinweis: Dies ist ein weiterführendes Kursmodul insb. für Teilnehmer die bereits am am ST.26-Grundlagenkurs (Basics) teilgenommen haben. (Deutsch am 7.9.2021)

*Den ST.26-Grundlagenkurs finden Sie unter
https://www.wipo.int/meetings/en/details.jsp?meeting_id=65768*

Die heutigen Themen

- Merkmalschlüssel und Qualifier: die Merkmaltabelle
- Lage von Merkmalen: Werte-Formate, Descriptoren und Operatoren
- Merkmalschlüssel und Qualifier:
Qualifier—Werte Werte: Freitext, Kontrolliertes Vokabular, Sequenzen
Merkmalschlüssel: source, CDS, modified_base
- Fragen? -
- Nukleotidanaloga, modifizierte Aminosäuren, D-Aminosäuren, verzweigte Sequenzen
- Sequenzvarianten
- Sonderfälle: Uracil in DNA, Thymin in RNA, DNA/RNA-Hybride
- Leitfaden – Anhang VI: Basisfragen, Definitionen, Beispiele und Übungen
- Fragen? -

Merkmaltabelle

Elemente für Merkmalschlüssel, Lage und Qualifier

Merkmaltabelle

60. Die Merkmaltabelle enthält Informationen über die Lage und Funktionen verschiedener Abschnitte 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
Merkmalschlüssel	<code>INSDFeature_key</code>	Ein Wort oder eine Abkürzung zur Angabe eines Merkmals	obligatorisch
Lage	<code>INSDFeature_location</code>	Region der Sequenz, die dem Merkmal entspricht	obligatorisch
Qualifier	<code>INSDFeature_qual</code>	Qualifier mit zusätzlichen Informationen zu einem Merkmal	obligatorisch, wenn für den Merkmalschlüssel (z. B. "source") ein oder mehrere Qualifier erforderlich sind, andernfalls fakultativ

Die Merkmaltabelle

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_qual	Qualifier mit zusätzlichen Informationen zu einem Merkmal	obligatorisch, wenn für den Merkmalschlüssel (z. B. "source") ein oder mehrere Qualifier erforderlich sind, andernfalls fakultativ

```

<SequenceData sequenceIDNumber="17">
- <INSDSeq>
  <INSDSeq_length>7</INSDSeq_length>
  <INSDSeq_moltype>AA</INSDSeq_moltype>
  <INSDSeq_division>PAT</INSDSeq_division>
- <INSDSeq_feature-table>
  - <INSDFeature>
    <INSDFeature_key>SOURCE</INSDFeature_key>
    <INSDFeature_location>1..7</INSDFeature_location>
  - <INSDFeature_qual>
    - <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_qual>
  </INSDFeature>
  - <INSDFeature>
    <INSDFeature_key>VARIANT</INSDFeature_key>
    <INSDFeature_location>1</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier>
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>X can be any amino acid</INSDQualifier_value>
    </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>
</INSDSeq_feature-table>
  <INSDSeq_sequence>XYEKGJL</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
  
```

Formate für die **Lage** in einer Sequenz

INSDFeature_location	Region der Sequenz, die dem Merkmal entspricht	obligatorisch
----------------------	--	---------------

- das obligatorische Element erfordert einen **Lagedescriptor**
- wird verwendet zur Identifizierung der Lage (Stelle, Position oder Region) eines Merkmals in einer Sequenz

Formate für Lage

Lagedeskriptoren für **alle** Molekülarten

- ST.26 hat obligatorische Anforderungen an das Format von Lagedeskriptoren
- folgende Lagedeskriptoren können für **Nukleotid- und Aminosäuresequenzen** verwendet werden:

(a) Lagedeskriptor für Nukleotid- und Aminosäuresequenzen:

Art des Lagedeskriptors	Syntax	Beschreibung
Einzelne Restnummer	x	Verweist auf einen einzelnen Rest in der Sequenz
Restnummern, 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 Restnummer	<x >x <x..y x..>y <x..>y	Verweist auf einen Abschnitt, der einen bestimmten Rest oder ein bestimmtes Restintervall enthält und über einen bestimmten Rest hinausgeht. Die Symbole "<" und ">" können in Verbindung mit einem einzelnen Rest oder mit dem ersten und letzten eines Intervalls von Resten verwendet werden, um anzuzeigen, dass ein Merkmal über die angegebene Restnummer hinausgeht.

Formate für Lage

Beispiele

Lagedescriptoren für **alle** Molekülarten

70. Die folgenden Beispiele veranschaulichen Merkmallagen, 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 Merkmallage 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.

Formate für Lage

Lagedeskriptoren **ausschließlich** für **Nukleotid**sequenzen

- Das Lagedescriptorformat x^y kann nur für DNA- und RNA-Sequenzen verwendet werden

(b) Lagedeskriptoren nur für Nukleotidsequenzen:

WIPO-Standard ST.26, Absatz 66(b)

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.

- Bei Verwendung des Lageformats x^y müssen x und y **benachbarte Reste** sein

(b) Lagen nur für Nukleotidsequenzen:

Beispiel für eine Lage	Beschreibung
123 ¹²⁴	Verweist auf eine Stelle zwischen den Resten 123 und 124.

Formate für Lage

Lagedeskriptoren – Sonderfälle für Aminosäuresequenzen

- Der Lagedeskriptor $x..y$ gibt bei Verwendung mit einem Merkmalschlüssel "CROSSLNK" oder "DISULFID" eine ketteninterne Quervernetzung zwischen den beiden angegebenen Resten an

(c) Lagedeskriptoren nur für Aminosäuresequenzen:

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

(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 Ausdruck mit einem Merkmal verwendet wird, das auf eine solche Quervernetzung hinweist, z. B. "CROSSLNK" oder "DISULFID".

Formate für Lagen

Lageoperatoren für Nukleotidsequenzen

- Für die Verwendung in **DNA- und RNA**-Sequenzen stehen **drei Lageoperatoren** zur Verfügung: **"join"**, **"order"** und **"complement"**
- Die Verwendung des Lageoperators "join" impliziert, dass die durch die Lagedeskriptoren beschriebenen Nukleotidreste durch biologische Prozesse physikalisch in Kontakt gebracht werden (ST.26, Absatz 68)
- Lagen, die "join" und "order" verwenden, müssen mindestens zwei durch Komma getrennte Lagedeskriptoren aufweisen

Syntax für Lage	Beschreibung der Lage
<code>join(location, location, ..., location)</code>	Die angegebenen Lagen werden zu einer zusammenhängenden Sequenz verbunden (von einem Ende zum anderen).
<code>order(location, location, ..., location)</code>	Die Elemente werden in der angegebenen Reihenfolge gefunden, aber es wird nichts darüber ausgesagt, ob das Verbinden dieser Elemente sinnvoll ist.
<code>complement(location)</code>	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.

Formate für Lagen

Lageoperatoren für Nukleotidsequenzen

Beispiele Absatz 70 (b)

(b) Lagen nur für Nukleotidsequenzen:

Beispiel für eine Lage	Beschreibung
<code>123^124</code>	Verweist auf eine Stelle zwischen den Resten 123 und 124.
<code>join(12..78,134..202)</code>	Gibt an, dass die Regionen 12 bis 78 und 134 bis 202 zu einer zusammenhängenden Sequenz verbunden werden sollten.
<code>complement(34..126)</code>	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).
<code>complement(join(2691..4571, 4918..5163))</code>	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).
<code>join(complement(4918..5163), complement(2691..4571))</code>	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).

Lageoperatoren für Nukleotidsequenzen

Nukleotidsequenzen: Merkmalschlüssel "CDS"

- Der Merkmalschlüssel "CDS" kann den Lageoperator "join" verwenden, um nicht zusammenhängende Segmente einer Sequenz zu einer einzigen codierenden Region zu verbinden

join(location1,location2)

- Die Merkmalschlüssel "CDS" kann den Operator "complement" verwenden, um anzugeben, dass sich das Merkmal auf dem Strang befindet, der zu der von dem Lagedeskriptor genannten Sequenz komplementär ist

complement(location)

Merkmalschlüssel und Qualifier

Merkmalschlüssel + ihre Qualifier

Teil des kontrollierten Vokabulars

ANHANG I

Merkmalschlüssel

62. In Anhang I sind die nach diesem Standard ausschließlich zu verwendenden Merkmalschlüssel mitsamt den zugehörigen, ebenfalls ausschließlich zu verwendenden zugehörigen 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.

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	25
ABSCHNITT 7: MERKMALSCHLÜSSEL FÜR AMINOSÄURESEQUENZEN	50
ABSCHNITT 8: QUALIFIER FÜR AMINOSÄURESEQUENZEN	57
ABSCHNITT 9: TABELLEN FÜR DEN GENETISCHEN CODE	58

Häufig verwendete Merkmalschlüssel

Nukleotid-Sequenzen:

- CDS
- modified_base
- gene
- variation

insg. **49 M / 81 Q** – für Nukleotid-SEQ

Aminosäure-Sequenzen:

insg. **39 M / 3 Q**

für AA-SEQ:

- mol_type
- note
- organism

Merkmalschlüssel	Beschreibung	Obligatorische Qualifier?
SITE	gibt eine wichtige einzelne Aminosäurestelle an	ein obligatorischer Qualifier NOTE muss die Stelle beschreiben
REGION	gibt eine wichtige Region an	keine; NOTE ist fakultativ
BINDING	gibt die Bindungsstelle für eine chemische Gruppe an	ein obligatorischer Qualifier NOTE muss den Namen der chemischen Gruppe enthalten
UNSURE	beschreibt Regionen von Unsicherheiten in der Sequenz	keine; NOTE ist fakultativ

Qualifier-Werte

- Qualifier stellen **weitere Informationen** über Merkmale bereit und definieren Merkmale näher (Merkmalskennzeichner)
- Qualifier bestehen aus einem Qualifier-**Namen** und oft einem zugeordneten Qualifier-**„Wert“**

```

<INSDQualifier>
  <INSDQualifier_name>transl_table</INSDQualifier_name>
  <INSDQualifier_value>12</INSDQualifier_value>
</INSDQualifier>

```

Name

Wert

- Jeder Merkmalschlüssel enthält eine **Liste von** für dieses Merkmal zulässigen **Qualifiern**. Einige Merkmalschlüssel haben **obligatorische** Qualifier.

- Qualifier werden in Kleinbuchstaben angegeben

Qualifier-Werte

für Nukleotidsequenzen

- Merkmalschlüssel "misc_binding":
 - ein obligatorischer Qualifier "bound_moiety"
 - sechs fakultative Qualifier:

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 "Molekülbereich" nicht anders angegeben, können die Merkmalschlüssel für DNA oder für RNA gleichermaßen verwendet werden. Bestimmte Merkmalschlüssel können neben dem angegebenen Geltungsbereich (des Organismus) auch für künstliche Sequenzen geeignet sein.

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

Beachten Sie, dass für die Beschreibung ribosomaler Bindungsstellen der Merkmalschlüssel "regulatory" und der Qualifier "regulatory_class" mit dem Wert "ribosome_binding_site" verwendet werden müssen.

Qualifier-Werte

für Nukleotidsequenzen

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.

- obligatorischer Qualifier "bound_moiety"

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 möglicher weise zu Zwecken 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.

6.23.	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.

Formate für Qualifier-Werte

für Qualifier-Werte gibt es **drei** Format-Arten
(nach Art der dargestellten Informationen):

- (a) **Freitext** (sprachabhängig oder unabhängig; s. Absätze 85 bis 88);
- (b) **kontrolliertes Vokabular** (definiert, zB Abkürzung einer modifizierte Base oder Zahlenwerte, Datum oder Nummer)
- (c) **Sequenzen** (zB bei PCR-Primern; PCR_primers)

Merkmal-Qualifier

72. Qualifier (Kennzeichner) dienen dazu, über den Merkmalschlüssel und die Merkmallage 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.

=> einige Qualifier haben **KEINEN** Wert

Qualifier-Werte

(a) Formatart "Freitext" (free text)

- Viele Qualifier haben ein „Freitext“-Wertformat
- ST.26, Absatz 3(n) definiert „Freitext“ als "...ein Format für den Wert bestimmter Qualifier, der in Form eines **beschreibenden Textes** oder in einem anderen (in Anhang I aufgeführten) Format angegeben ist“
- "Freitext"-Qualifier-Werte sind auf **1000 Zeichen** begrenzt (ST.26, Absatz 86)
- Eine Teilmenge von Qualifiern mit „Freitext“-Wertformaten ist **"language-dependent"** (sprachabhängig)
- **„sprachabhängiger Freitext“** bestimmter Qualifier muss für nationale, regionale oder internationale Zwecke **möglicherweise übersetzt** werden." ST.26, Absatz 3(o)

Qualifier-Werte

Formatart – „sprachabhängiger Freitext“

- Wie kann festgestellt werden, ob ein Qualifier mit einem Freitext-Wertformat **sprachabhängig** ist?
- ST.26, **Anhang I, Abschnitt 8, Tabelle 6** listet alle **Aminosäuresequenz-**Qualifier mit einem sprachabhängigen Freitext-Wertformat auf
- ST.26, Anhang I, **Abschnitt 6, Tabelle 5** listet alle **Nukleotidsequenz-**Qualifier mit einem sprachabhängigen Freitext-Wertformat auf

Tabelle 5: Liste der **Qualifierwerte für Nukleotidsequenzen** mit sprachabhängigen Freitextwerten

Abschnitt	Sprachabhängige Freitextwerte
6.3	bound_moiety
6.5	cell_type
6.7	clone
6.8	clone_lib
6.10	collected_by
6.13	cultivar
6.14	dev_stage
6.17	ecotype
6.21	function

Qualifier-Werte

Formatart – „sprachabhängiger Freitext“

- „**sprachabhängige Freitext**“ - Qualifier-Werte können im XML-Sequenzprotokoll in **zwei Sprachen** angegeben werden:
zB **Englisch** und eine weitere **nicht-englische** Sprache (ST.26, Absatz 87)
- Ein sprachabhängiger Qualifier-Wert muss in **Englisch** in einem Element **"INSDQualifier_value"** angegeben werden
- Ein sprachabhängiger Qualifier-Wert in einer anderen Sprache als Englisch muss in einem Element **"NonEnglishQualifier_value"** angegeben werden
- Ein Element **"NonEnglishQualifier_value"** ist **NUR** für Qualifier mit **sprachabhängigem** Freitext-Wertformat zulässig (ST.26, Absatz 87(b))

Qualifier-Werte

Formatart – „sprachabhängiger Freitext“

```

<SequenceData sequenceIDNumber="2">
  <INSDSeq>
    <INSDSeq_length>29</INSDSeq_length>
    <INSDSeq_moltype>AA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>SOURCE</INSDFeature_key>
        <INSDFeature_location>1..29</INSDFeature_location>
        <INSDFeature_qual>
          <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 id="q1">
            <INSDQualifier_name>NOTE</INSDQualifier_name>
            <INSDQualifier_value>Synthetic peptide antigen fragment</INSDQualifier_value>
            <NonEnglishQualifier_value>Synthetisches Peptidantigenfragment</NonEnglishQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>GSLSDVRKDVEKRIDKALEAFKNKMDKEK</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="3">

```

Englisch

NonEnglish

FAQ — Advanced Modul

F6: Müssen die Sequenzprotokolle **immer in englischer Sprache** erstellt werden? Kann ich ein Sequenzprotokoll in Deutsch als Hauptsprache haben? Oder kann nur Englisch die Hauptsprache sein?

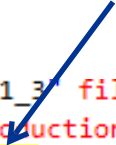
A6: **Der Wert für jeden sprachabhängigen Qualifier in einem Sequenzprotokoll muss in derselben Sprache eingegeben werden.** Diese Sprache kann Englisch oder eine andere Sprache, wie z. B. Deutsch, sein, je nachdem, welche Sprache(n) von dem jeweiligen Amt für diesen Zweck akzeptiert wird (werden). **Der Standard erlaubt auch, dass ein Sequenzprotokoll alle sprachabhängigen Qualifier Werte in Englisch und zusätzlich in einer nicht-englischen Sprache sind.** Ja, es ist also möglich, ein Sequenzprotokoll zu erstellen, in dem alle sprachabhängigen Freitext-Qualifierwerte nur in Deutsch sind .

Qualifier-Werte

Formatart – „sprachabhängiger Freitext“

- Wenn ein XML-Sequenzprotokoll nicht-englische Qualifier-Werte enthält:
 1. das Wurzelement des Sequenzprotokoll-XML muss ein Attribut "nonEnglishFreeTextLanguageCode" mit einer entsprechenden **zweibuchstabigen Sprachcode**-Abkürzung enthalten (ST.26, Absätze 43, 7(b));

```
<ST26SequenceListing dtdVersion="V1_3" fileName="st26-annex-iii-sequence-listing-specimen.xml" softwareName="WIPO  
Sequence" softwareVersion="1.0" productionDate="2022-01-01" originalFreeTextLanguageCode="de"  
nonEnglishFreeTextLanguageCode="de">
```



2. ALLE sprachabhängigen Qualifier im Sequenzprotokoll müssen Werte in der Sprache aufweisen, die im Attribut "nonEnglishFreeTextLanguageCode" angegeben ist
3. Wenn sowohl "NonEnglishQualifier_value" und "INSDQualifier_value" für einen einzelnen Qualifier vorhanden sind, müssen die **in den beiden Elementen enthaltenen Angaben gleich sein** (ST.26, Absatz 87(c)).

Qualifier-Werte

Formatart – „sprachabhängiger Freitext“

- INSDQualifier **"id"**-Attribut – was ist das?
- ST.26, Absatz 87(d) legt fest: *"Für sprachabhängige Qualifier kann das Element "INSDQualifier" ein fakultatives Attribut "id" enthalten. Der Wert dieses Attributs muss das Format "q" haben, gefolgt von einer positiven ganzen Zahl, z.B. "q23", und muss für ein Element "INSDQualifier" eindeutig sein, d.h. der Attributwert darf nur einmal in einer Sequenzprotokolldatei verwendet werden."*

```
<INSDQualifier id="q2">
  <INSDQualifier_name>note</INSDQualifier_name>
  <INSDQualifier_value>common name: tomato</INSDQualifier_value>
  <NonEnglishQualifier_value>gemeinsamen Namen: Tomate</NonEnglishQualifier_value>
</INSDQualifier>
```

- Identifizieren eindeutig Qualifier-Werte, die für den Export in eine XLIFF- Datei durch WIPO Sequence möglicherweise übersetzt werden müssen;
- Werden **automatisch** in einer von WIPO Sequence generierten XML hinzugefügt.

Qualifier-Werte

Formatart – „sprachabhängiger Freitext“

- Das Attribut "originalFreeTextLanguageCode"

```
<ST26SequenceListing dtdVersion="V1_3" fileName="st26-annex-iii-sequence-listing-specimen.xml" softwareName="WIPO  
Sequence" softwareVersion="1.0" productionDate="2022-01-01" originalFreeTextLanguageCode="de"  
nonEnglishFreeTextLanguageCode="de">
```

- ST.26, Absatz 43 definiert das Attribut "originalFreeTextLanguageCode" als den "Sprachcode ... für die einzelne Originalsprache, in der die sprachabhängigen Freitext-Qualifier erstellt wurden.
- Dieses Attribut ist FAKULTATIV

Qualifier-Werte

(b) Formatart „Kontrolliertes Vokabular“

- Qualifier mit **vordefinierten Wertoptionen**
- Beispiel: **"codon_start"** – Werte können "1", "2" oder "3" sein

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>

Qualifier-Werte

Formatart „Kontrolliertes Vokabular“

- Beispiel: `"rpt_type"` mit begrenztem Satz an Wertoptionen

6.61. Qualifier	rpt_type
Definition	Struktur und Verteilung der repetitiven Sequenz
obligatorisches Wertformat	einer der folgenden Ausdrücke aus dem kontrollierten Vokabular: <div style="border: 1px solid red; padding: 5px;"> <pre>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</pre> </div>
Beispiel	<pre><INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value></pre>

Qualifier-Werte

Formatart „Kontrolliertes Vokabular“

- Für Qualifier mit vordefinierten Wertoptionen sind in **WIPO Sequence** alle zulässigen Werte in einer vorab befüllten Dropdown-Liste zu finden:

▼ QUALIFIERS

Add qualifier

Qualifier Name *

rpt_type

Qualifier Value

- 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

▼ SEQUENCE

tgccgatatt	atccataagc	atgggcgtct	ggaagtgcog	ctgacggccc	agaagggcgt
cttactgtta	caagaagagt	atcatgacgc	gcatcatagc	cacttagacc	cgcacttatg
gttggcgcog	caaaatatca	ttgcctcgt	gcaaggattg	gataaacata	ctgctgagtt

180

Qualifier-Werte

Formatart „Kontrolliertes Vokabular“

- Qualifier mit einem **definierten** Wertformat
- Beispiel: "anticodon" – Wert muss im Format

"(pos:<location>,aa:<amino_acid>,seq:<text>)" vorliegen
>(pos:34..36,aa:Phe,seq:aaa)</

6.2.	Qualifier	anticodon
	Definition	Stelle 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>

Qualifier-Werte

(c) Formatart - Sequenzen

- Qualifier, bei denen der Wert eine **Sequenz** ist
- Beispiel: "translation" – Wert muss eine Sequenz mit **einbuchstabigen Aminosäureabkürzungen** sein

6.80. Qualifier	translation
Definition	eine einbuchstabig abgekürzte Aminosäuresequenz, abgeleitet vom genetischen Standardcode (oder universellen Code) oder einer im Qualifier "transl_table" angegebenen Tabelle entnommen 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>MASTFPPWYRGCSTPSLKGLIMCTW</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.

Qualifier-Werte

Formatart „Sequenzen“

- Qualifier-Wert "replace" kann ein einzelner Nukleotidrest, eine Sequenz von Resten, oder leer sein

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	<pre><INSDQualifier_value>a</INSDQualifier_value> <INSDQualifier_value><</INSDQualifier_value> - für eine Deletion</pre>

"Empty value"

- Ein leerer Wert für "replace" bedeutet eine Deletion des im entsprechenden Merkmal angegebenen Rests

Qualifier-Werte

Formatart KEIN Wert

- Qualifier mit KEINEM Wert

- Beispiele:

“environmental_sample”, “germline”, “macronuclear”, **“proviral”**

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

- WIPO Sequence erlaubt nicht das Hinzufügen eines Wertes für Qualifier, die keinen Wert haben
- Diese Qualifier dürfen kein leeres Element "INSDQualifier_value" aufweisen

Merkmalschlüssel + Qualifier

Merkmalschlüssel "source"

- Merkmalschlüssel sind **fakultativ**, mit Ausnahme des Merkmals **"source"**, das für jede Sequenz **obligatorisch** ist
- Jeder Merkmalschlüssel enthält eine **Liste von Qualifiern**, die zur näheren Beschreibung des Merkmals verwendet werden können.

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

Merkmalschlüssel + Qualifier

Merkmalschlüssel "source"

- die meisten Qualifier sind fakultativ; einige Merkmalschlüssel haben jedoch obligatorische Qualifier
- für „source“ sind die Qualifier „mol_type“ und „organism“ obligatorisch

„organism“

77. Der Organismus-Qualifier, d. h. "organism" für Nukleotidsequenzen (siehe Anhang I, Abschnitt 6) und 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.

„mol_type“

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. Diese Qualifier sind zu unterscheiden von 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" 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".

Merkmalschlüssel + Qualifier

- Einige Merkmalschlüssel haben zusätzliche Beschränkungen
 - Bereich Organismus; z.B. "C_region" ist auf Eukaryoten beschränkt
 - Molekülbereich; z.B. ist "D-Loop" auf DNA-Sequenzen beschränkt

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
Molekülbereich	DNA

Merkmalschlüssel + Qualifier

Nukleotidsequenzen: Merkmalschlüssel "CDS"

- Der Merkmalschlüssel "CDS" kann zur **Identifizierung kodierender Sequenzen** verwendet werden.
- Die Lage des Merkmals "CDS" muss das Stopcodon enthalten. (ST.26, Absatz 89)
- *Für "CDS" gibt es KEINE obligatorischen Qualifier - außer das Translationsprodukt hat 4 oder mehr Aminosäuren*
- Zu den **häufig** verwendeten **Qualifiern (22)** gehören:

"pseudo"

"translation"

"transl_table"

"pseudogene"

"transl_except"

"codon_start"

Merkmalschlüssel + Qualifier

Nukleotidsequenzen: Merkmalschlüssel "CDS"

- Eine **Aminosäuresequenz** mit **4** oder mehr **spezifisch definierten** Aminosäuren, die von einer Sequenz kodiert und in einem Qualifier "**translation**" offenbart wird, muss als **separate Sequenz** in das Sequenzprotokoll aufgenommen werden und eine **eigene Kennzahl** zugewiesen bekommen.
- Die der Aminosäuresequenz zugewiesene Sequenzkennzahl muss als der **Wert im Qualifier "protein_id"** innerhalb des Merkmalschlüssels "CDS" angegeben werden.
- Der Qualifier „**organism**“ des Merkmalschlüssels „**source**“ für die Aminosäuresequenz muss mit dem ihrer codierenden Sequenz identisch sein. (ST.26, Absatz 92)

Merkmalschlüssel + Qualifier

Nukleotidsequenzen: Merkmalschlüssel "CDS"

- Ein Merkmal "CDS" kann **nur EINEN** der folgenden **Qualifier** haben:

Qualifier	Beschreibung	Wert / Wertformat
pseudo	das Merkmal CDS betrifft ein nicht-funktionales Gen, keine Translation, aber kein Pseudogen ist (zB ein Sequenzierungsfehler)	keiner
pseudogene	das Merkmal CDS betrifft ein „echtes“ Pseudogen, hat keine Translation	prozessiert nicht prozessiert unitär allelisch unbekannt
translation	gibt die Aminosäuresequenz an, die von der Translation des CDS abgeleitet ist	einbuchstabile Abkürzungen von Aminosäuren

Merkmalschlüssel + Qualifier

Nukleotidsequenzen: Merkmalschlüssel "CDS"

- Qualifier, die die **translatierte Sequenz** definieren

Qualifier	Beschreibung	Wert
transl_table	gibt die genetische Codetabelle an, die zur Translation des CDS verwendet wird; Standard ist "1-Standard Code"	eine Nummer, die einer Translationstabelle in Anhang I, Abschnitt 9 entspricht
transl_except	gibt die Translation eines Codons an, das nicht mit dem in "transl_table" definierten genetischen Code konform ist	(pos:<location>, aa:<amino_acid>)
codon_start	gibt das Leseraster des CDS in Bezug zur ersten Base an	1, 2 oder 3

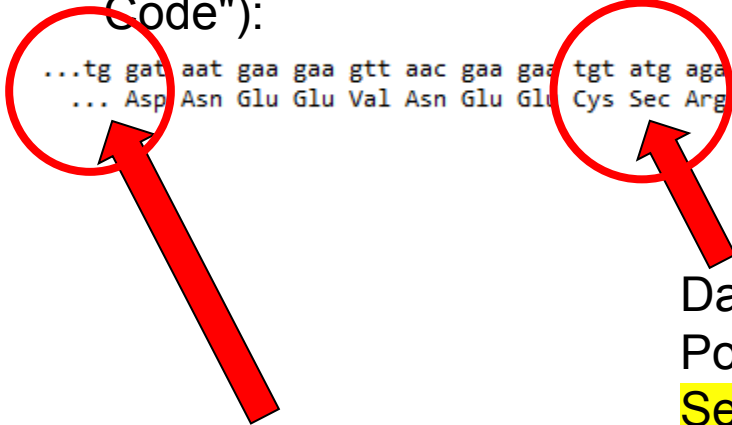
Merkmalschlüssel + Qualifier

Nukleotidsequenzen: Merkmalschlüssel "CDS"

Beispiel:

- SEQ ID NO:1 ist ein Fragment einer codierenden Sequenz aus einem mitochondrialen Hefegen (genetische Codetabelle "3-Yeast Mitochondrial Code"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa 80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```



Das Codon „atg“ an den Positionen 30-32 codiert Selenocystein (Sec)

Die Sequenz beginnt mit einem partiellen Codon

- Welche Informationen sollten in einem Merkmal CDS enthalten sein, um diese Sequenz exakt darzustellen?

Nukleotidsequenzen: Merkmalschlüssel "CDS"

Welche Informationen sollten im Merkmal CDS enthalten sein, um diese Sequenz exakt darzustellen?

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa 80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- **Nukleotid-Sequenz**, im Sequenzprotokoll darzustellen:

tgataatga agaagttaac gaagaatgta tgagattatt ttcaagaac gctcgtcatc taacatcaag gttgacataa

- **Lage** des Merkmals CDS: < 1..80

"<" gibt an, dass der codierende Bereich **vor** der Position 1 beginnt

Die Lage enthält das **Stopcodon** in den Positionen 78-80

- Qualifier "**codon_start**" mit Wert "**3**"

Wert "**3**" gibt an, dass das erste vollständige Codon an der 3. Position innerhalb der Lage beginnt

Nukleotidsequenzen: Merkmalschlüssel "CDS"

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- Merkmalschlüssel "CDS"
- Merkmallage: < 1..80
- Qualifier "codon_start" mit Wert "3"
- Qualifier "transl_table" mit dem Wert "3"

Die genetischen Codetabellen in Anhang 1, Abschnitt 9 werden zur Bestimmung des Wertes des Qualifiers "transl_table" verwendet; „3“ = Mitochond. Code der Hefen

- Qualifier "transl_except" mit dem Wert "(pos:30..32,aa:Sec)"

Der Qualifier "transl_except" setzt die Aminosäure Selenocystein ein

- Qualifier "translation" mit dem Wert "DNEEVNEECURLFFKNARHTTSRLT"

- eine separate Proteinsequenz für die Translation - ohne Stoppcodonen

Qualifier „mod_base“

- Merkmalschlüssel "modified_base" und sein obligatorischer Qualifier "mod_base" sind zur Beschreibung eines **modifizierten Nukleotids** (ST.26, Absatz 16) zu verwenden.
- Ein "modifiziertes Nukleotid" ist **jedes andere** Nukleotid als:
 - Desoxy-[a, g, c, oder t] 3'-monophosphat *ST.26, Absatz 3(f)*
 - [a, g, c, oder u] 3'-monophosphat
- Ein "modifiziertes Nukleotid" ist - **wenn möglich** - durch das entsprechende **unmodifizierte** Nukleotid (in Anhang 1, Abschnitt 1, Tabelle 1) darzustellen.

Beispiel: „2'-O-Methylcytidin" kann in der Sequenz durch "c" wiedergegeben werden, mit dem Wert der Abkürzung aus Tabelle 2, hier „cm“.

Tabelle 2: Liste der modifizierten Nukleotide

Abkürzung	Definition / Modifiziertes Nukleotid
ac4c	4-Acetylcytidin
chm5u	5-(Carboxyhydroxymethyl)uridin
cm	2'-O-Methylcytidin

Nukleotidsequenzen: Merkmalschlüssel "modified_base"

47

Qualifier „mod_base“- Definition *gelistet* (in Tabelle 2)

- Der **Wert** des obligatorischen Qualifiers "**mod_base**" muss aus Tabelle 2, (Anhang I, Abschnitt 2) ausgewählt werden, wenn Nukleotid dort gelistet ist.

Tabelle 2: Liste der modifizierten Nukleotide

Abkürzung	Definition / Modifiziertes Nukleotid
ac4c	4-Acetylcytidin
chm5u	5-(Carboxyhydroxymethyl)uridin
cm	2'-O-Methylcytidin

- Beispiel für ein modifiziertes Nukleotid **ohne Entsprechung zu Standardnukleotiden**: „Queuosin“

- in der **Sequenz durch "n"** wiederzugeben, mit dem Wert der Abkürzung „q“.

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-Ribofuranosyl)purin-6-y
tm	2'-O-Methyl-5-methyluridin
um	2'-O-Methyluridin
yw	Wybutosine
x	3-(3-Amino-3-carboxypropyl)uridin, (
OTHER	("note"-Qualifier erforderlich)

Nukleotidsequenzen: Merkmalschlüssel "modified_base"

48

Qualifier „mod_base“- Definition *gelistet*

- Beispiel: Nukleotidsequenz mit **Inosin** in Position 15

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

Tabelle 2: Liste der modifizierten I

Abkürzung	
ac4c	4-Acety
chm5u	5-(Carb
cm	2'-O-Me
cmnm5s2u	5-Carb
cmnm5u	5-Carb
dhu	Dihydr
fm	2'-O-Me
gal q	beta-D-
gm	2'-O-Me
i	Inosin
i6a	N6-Isor

- "Inosine" ist in Anhang I, Abschnitt 2, Tabelle 2 mit dem Wert der Abkürzung "i" aufgeführt

Nukleotidsequenzen: Merkmalschlüssel "modified_base"

49

Qualifier „mod_base“- Definition **nicht** gelistet

- bestimmtes modifiziertes Nukleotid **ohne Entsprechung** mit Standardnukleotiden + **nicht gelistet** in Anhang I, Tabelle 2, Abschnitt 1:

- als Wert muss die Abkürzung „**OTHER**“ verwendet werden!

- wird die Abkürzung „**OTHER**“ verwendet, muss in einem zusätzlichen Qualifier **"note"** der **vollständige, nicht abgekürzte Namen** des modifizierten Restes angegeben werden

q	Queuosin
s2c	2-Thiocytidin
s2t	5-Methyl-2-thiouridin
s2u	2-Thiouridin
s4u	4-Thiouridin
m5u	5-Methyluridin
t6a	N-((9-beta-D-Ribofuranosyl)purin
tm	2'-O-Methyl-5-methyluridin
um	2'-O-Methyluridin
yw	Wybutosine
x	3-(3-Amino-3-carboxypropyl)uridin
OTHER	("note"-Qualifier erforderlich)

Nukleotidsequenzen: Merkmalschlüssel "modified_base"

50

Qualifier „**mod_base**“- Definition **nicht** gelistet

Beispiel:

- Nukleotidsequenz mit **Xanthin** in Position 22

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>22</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>
```

- "**Xanthine**" ist in Anhang I, Abschnitt 2, Tabelle 2 NICHT aufgeführt; daher muss der **Wert** für den obligatorischen Qualifier "mod_base" "**OTHER**" sein und ein zusätzlicher Qualifier "note" muss als **Wert** "**xanthine**" enthalten

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

Definition 3(g)(ii)(2)!

- "modified_base" kann zur Beschreibung einer **abasischen Stelle** verwendet werden:

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>11</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>abasic site</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

- Die **abasische Stelle** wird in der Sequenz durch ein **"n"** dargestellt, mit dem Merkmalschlüssel "modified_base" und Qualifier „**mod_base**“ mit dem Wert **"OTHER"** und dem zusätzlichen Qualifier **"note"** mit dem Wert **"abasic site"**

Fragen?

Neue Molekülarten: Nukleotidanaloga

- ein "**modifiziertes Nukleotid**" ist **jedes andere** Nukleotid als:
 - Desoxy-[a, g, c, oder t] 3'- Monophosphat
 - [a, g, c, oder u] 3'- Monophosphat
- Modifikationen betreffen i.d.R. die **einzelne Base**
- „**Nukleotidanaloga**“ sind in der Definition „Nukleotid“ unter Absatz 3(g)(2) umfasst und betreffen insbesondere die Struktur der **Rückgrateinheit**
- Nukleinsäuresequenzen, die ein oder mehrere Nukleotidanaloga enthalten, müssen den Regeln von ST.26 entsprechen
 - übliche Nukleotidanaloga sind **Peptidnukleinsäuren** (PNAs), **Glykolnukleinsäuren** (GNAs), **Phosphorothioate** (PTO)-Oligos (ein Sauerstoff-Atom der PO-Gruppe ist durch ein Schwefelatom ersetzt („S-Oligos“); Morpholinos (PMOs, phosphorodiamidate morpholino oligo;)

Neue Molekülarten: Nukleotidanaloga

Beispiel: Glykolnukleinsäure (GNA)

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

PO₄-tagttcattgactaaggctccccattgact-OH

wobei das PO₄-Ende der Sequenz das 5'-Ende einer DNA-Sequenz nachahmt.

- ✓ Diese Sequenz muss in ein Sequenzprotokoll **aufgenommen** werden
- ✓ Das **PO₄-Ende** ahmt das 5'-Ende nach, so dass die Sequenz in der angezeigten Richtung wiedergegeben werden muss
- ✓ Die gesamte Sequenz muss mit dem Merkmalschlüssel **"modified_base"**, einem Qualifier **"mod_base"** mit dem Wert **"OTHER"**, und einem Qualifier **"note"**, der den **vollständigen, nicht abgekürzten Namen** der modifizierten Nukleotiden enthält, z.B. **"glycol nucleic acids"**, annotiert werden

Neue Molekülarten: Nukleotidanaloga

P04-tagttcattgactaaggctccccattgact-OH

(GNA):

Qualifier "mod_base" mit Wert "OTHER"

```

- <INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>1..30</INSDFeature_location>
  - <INSDFeature_qual>
    - <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 (glycol nucleic acids)</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>tagttcattgactaaggctccccattgact</INSDSeq_sequence>
</INSDSeq>
  
```

Qualifier "note", mit vollständiger Bezeichnung

Hinweis: Eine ausführliche Erläuterung dieses Beispiels finden Sie im WIPO-Standard ST.26, Anhang VI, Beispiel 3(g)-4.

Definition modifizierte Aminosäure:

(ST.26, Definition, Absatz 3(e))

- **"modifizierte Aminosäure"** ist jede andere Aminosäure als:

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	L-Valin	

- Eine **"modifizierte Aminosäure"** - insbesondere **D-Aminosäure** - ist nach Möglichkeit durch die **entsprechende unmodifizierte** Aminosäure (in Anhang I, Abschnitt 3 Tabelle 3) darzustellen.

Beispiel: „**Hydroxylysin**“ kann in der Sequenz durch **"K"** dargestellt werden, mit dem Wert der Abkürzung aus Tabelle 4, hier „**Hyl**“

- „**Ornithin**“ ist durch **"X"** darzustellen - warum?

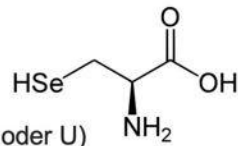
■ Ornithin kann keiner Standardaminosäure zugeordnet werden

es müssen die Format-Anforderungen der INSDC (International Nucleotide Sequence Database Collaboration) beachtet werden

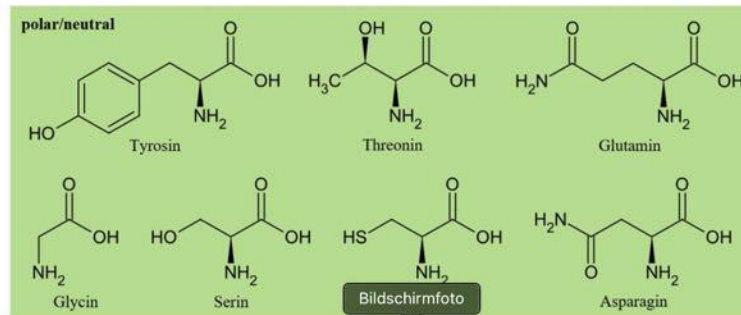
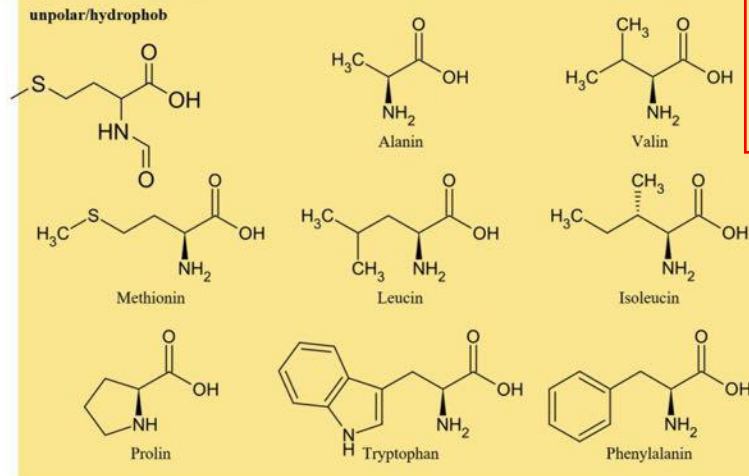
Proteinogene Aminosäuren

L-Aminosäuren, die durch *Recodierung* des genetischen Materials in Proteine eingebaut werden; insgesamt 23 proteinogene AS

- Selenocystein
- Pyrrolysin
- N-formylmethionin gelten als 21.-23. natürliche AS, da sie über ein eigenes Codon in der DNA kodiert werden

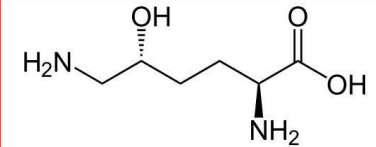


N-formyl-methionine



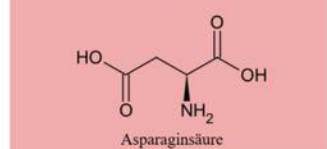
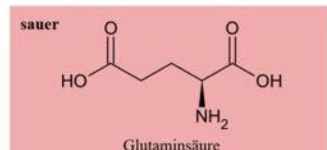
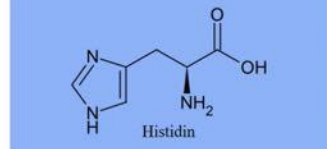
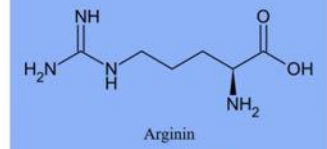
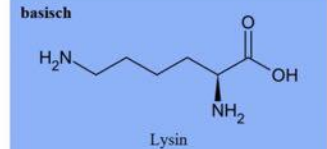
Hydroxylysin

Strukturformel



Allgemeines

Name	Hydroxylysin
basisch	<chem>NC(C(N)C)C(N)C(=O)O</chem> Lysin
	<chem>NC(=N)N(C)C(N)C(=O)O</chem> Arginin
	<chem>C1=CN=C(C=C1)C(N)C(=O)O</chem> Histidin



Bildschirmfoto

Merkmalschlüssel und Qualifier

Aminosäuresequenzen: **Modifizierte** Aminosäuren

- Mehrere Merkmalschlüssel können zur Angabe einer "modifizierten Aminosäure" verwendet werden:

Merkmalschlüssel	Beschreibung
SITE	gibt eine nicht posttranslational modifizierte Aminosäure an
MOD_RES	gibt eine posttranslational modifizierte Aminosäure an
CARBOHYD	gibt eine glykosylierte Aminosäure an
LIPID	gibt die kovalente Bindung eines Lipidanteils an eine Aminosäure an

- Jeder der oben genannten Merkmalschlüssel muss einen **obligatorischen Qualifier** „note“ mit einem Wert, der die Modifizierung beschreibt, enthalten.

Neue Molekülarten: D-Aminosäuren

die funktionellen Gruppen – Carboxygruppe (–COOH) und Aminogruppe (–NH₂) liegen **α-ständig in D-Konfiguration** vor = Spiegelbildisomere der L-Aminosäuren.

L

- Aminosäuresequenzen, die eine oder mehrere D-Aminosäuren enthalten, unterliegen den Regeln von ST.26
- **D-Aminosäuren** sind in der **Definition einer "Aminosäure"** nach ST.26, Absatz 3(a) **enthalten**:

"Unter "Aminosäure" ist jede Aminosäure zu verstehen, die mit einem der in Anhang I (siehe Abschnitt 3, Tabelle 3) aufgeführten Symbole beschrieben werden kann. Zu diesen Aminosäuren gehören unter anderem D-Aminosäuren und Aminosäuren mit modifizierten oder synthetischen Seitenketten."

- D-Aminosäuren müssen in der Sequenz wenn möglich **als** das **entsprechende unmodifizierte** Aminosäuresymbol **L** ausgewiesen werden
- müssen in der Merkmalstabelle als eine **modifizierte Aminosäure** beschrieben werden

Neue Molekülarten: D-Aminosäuren

Beispiel

Eine Patentanmeldung beschreibt die folgende Sequenz:

D-Ala-D-Glu-Lys-Leu-Gly-D-Met

- ✓ Diese Sequenz muss in ein Sequenzprotokoll aufgenommen werden
- ✓ Ist darzustellen als: **AGKLG M**
- ✓ **Annotation:** Alanin in Position 1, Glutaminsäure in Position 2 und Methionin in Position 6 müssen jeweils mit einem Merkmalschlüssel "SITE" und einem Qualifier „note“ mit dem Wert des vollständigen, nicht abgekürzten Namen der entsprechenden Aminosäure beschrieben werden

Neue Molekülarten: D-Aminosäuren

```

- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>1</INSDFeature_location>
  - <INSDFeature_quals>
    - <INSDQualifier id="q4">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>D-alanine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>2</INSDFeature_location>
  - <INSDFeature_quals>
    - <INSDQualifier id="q5">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>D-glutamic acid</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>6</INSDFeature_location>
  - <INSDFeature_quals>
    - <INSDQualifier id="q6">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>D-methionine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>AEKLG</INSDSeq_sequence>

```

Merkmalschlüssel "SITE"

Qualifier „note“
(sprachabhängig, daher
„id“ Format „q“)

Wert des
vollständigen,
ungekürzten
Namen der mod.
Aminosäure

Hinweis: Eine ausführliche Erläuterung eines ähnlichen Beispiels finden Sie im WIPO-Standard ST.26, Anhang VI, Beispiel 3(a)-1.

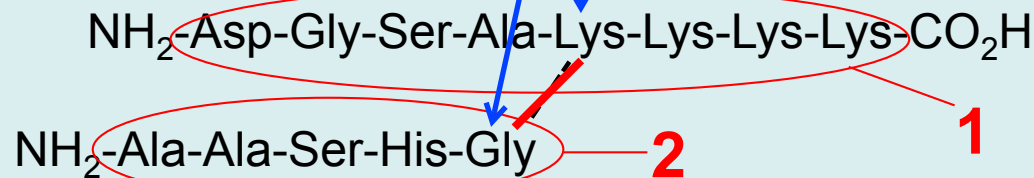
Neue Molekülarten - *verzweigte* Sequenzen:

- Verzweigte Nukleinsäuresequenzen und verzweigte Aminosäuresequenzen unterliegen den **Regeln von ST.26**
- **Lineare Regionen** von verzweigten Sequenzen müssen in ein Sequenzprotokoll **aufgenommen werden, wenn sie 10 oder mehr spezifisch definierte Nukleotide oder 4 oder mehr spezifisch definierte Aminosäuren** enthalten (ST.26, **Absatz 7**)
- Jede lineare Region einer verzweigten Sequenz, welche die Mindestlängenanforderung erfüllt, muss als eine separate Sequenz mit ihrer **eigenen SEQ ID-Kennzahl** aufgenommen werden
- Die Anzahl der **spezifisch definierten Reste** in **jeder einzelnen linearen Region** muss berücksichtigt werden, nicht die Gesamtzahl der spezifisch definierten Reste in der Struktur

Neue Molekülararten - verzweigte Sequenzen:

Position 5

eine Patentanmeldung beschreibt die folgende Sequenzstruktur:



wobei — eine Amidbindung zwischen dem C-Terminus des Glycins und der Seitenkette des Lysins angibt

- Beide linearen Regionen enthalten ≥ 4 spezifisch definierte Aminosäuren, so dass beide in einem Sequenzprotokoll enthalten sein müssen
- Jede lineare Region muss als eine separate Sequenz mit einer eigenen SEQ ID-Kennzahl enthalten sein
- Beide Sequenzen sollten „annotiert“ werden, um die Lage und Art der Amidbindung anzugeben:

Neue Molekulararten - verzweigte Sequenzen:

Lage und Art der Amidbindung

1

```

- <INSDFeature>
  <INSDFeature_key> SITE</INSDFeature_key>
  <INSDFeature_location> 5</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q4">
      <INSDQualifier_name> NOTE</INSDQualifier_name>
      <INSDQualifier_value> Lysine side chain is amide bonded to the carboxy
        terminus of the glycine in SEQ ID NO:2</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence> DGSAK</INSDSeq_sequence>

```

2

```

- <INSDFeature>
  <INSDFeature_key> SITE</INSDFeature_key>
  <INSDFeature_location> 5</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q11">
      <INSDQualifier_name> NOTE</INSDQualifier_name>
      <INSDQualifier_value> glycine carboxy terminus is amide bonded to the side
        chain of the lysine in SEQ ID NO:1, position 5</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence> AASHG</INSDSeq_sequence>

```

Hinweis: Eine ausführliche Erläuterung eines ähnlichen Beispiels finden Sie im WIPO-Standard ST.26, Anhang VI, Beispiel 7(b)-3.

Sequenzvarianten

- ST.26 Absatz 3(m) definiert eine "**Sequenzvariante**" als 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.
- ST.26, **Absätze 93-95** regeln die Darstellung der **Varianten**.

Absatz 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.“*

Wird jede Variante separat gezählt, muss jede Variante ihre eigene SEQ ID-Kennzahl haben!

Sequenzvarianten

Absatz 93

Eine Patentanmeldung enthält eine Abbildung mit dem folgenden multiplen Sequenzalignment:

¶ 93-3: Darstellung einer Konsensussequenz

Eine Patentanmeldung enthält Abbildung 1 mit der folgenden mehrfachen Sequenzalignierung.

<i>Consensus</i>	LEG n EQFINA ak IRHP k nr k T ln ND l MLIK
<i>Homo sapiens</i>	LEG N EQFINA AK IRHP Q YDR K T LN ND I MLIK
<i>Pongo abelii</i>	LEG N EQFINA AK IRHP Q YDR K T VN ND I MLIK
<i>Papio Anubis</i>	LEG T EQFINA AK IRHP D YDR K T LN ND I LLIK
<i>Rhinopithecus roxellana</i>	LEG T EQFINA AK IRHP N YNRI T LD ND I LLIK
<i>Pan paniscus</i>	LEG N EQFINA AK IRHP K YNRI T LN ND I MLIK
<i>Rhinopithecus bieti</i>	LEG N EQFINA TK IRHP K YNGN T LN ND I MLIK
<i>Rhinopithecus roxellana</i>	LEG N EQFINA TQ IRHP K YNGN T LN ND I MLIK

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.

Jede der aufgezählten Sequenzen muss im Sequenzprotokoll als eine **separate Sequenz** mit einer eigenen **SEQ ID-Kennzahl** aufgeführt sein.

Hinweis: Eine ausführliche Erläuterung eines ähnlichen Beispiels finden Sie im WIPO-Standard ST.26, Anhang VI, Beispiel 93-3.

ST.26, Absätze 15 und 27 besagen:

„Wenn ein **Mehrdeutigkeitssymbol** (das zwei oder mehr **alternative Nucleinsäuren bzw. Aminosäuren repräsentiert**) erforderlich ist, sollte das **restriktivste Symbol** verwendet werden.“

Beispiele

Wenn in einer Nucleotidsequenz eine Position **"a or c"** sein kann, soll das Mehrdeutigkeitssymbol **"m"** anstelle von "n" verwendet werden.

Wenn in einer Aminosäuresequenz eine Position **"L or I"** sein kann, verwenden Sie das Mehrdeutigkeitssymbol **"J"** anstelle von "X".

Zur Erinnerung **"n"** und **"X"** haben **Standardwerte**, so dass jedes Mal, wenn "n" oder "X" nicht als Standardwert verwendet wird, eine Annotation erforderlich ist

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; "unknown" oder "other"

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; "unknown" oder "other"

Sequenzvarianten

Absatz 94

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

Beispiel

Eine Patentanmeldung offenbart ein Peptid der Sequenz:

Gly-Gly-Gly-[Leu or Ile]-Ala-Thr-[Ser or Thr]

Da das Peptidfragment in der offenbarten Ausführungsform **mindestens sechs spezifisch definierte Aminosäuren** umfasst, **muss** die Sequenz in ein Sequenzprotokoll aufgenommen werden, wie in ST.26 Absatz 7(b) gefordert.

Kann als **eine Sequenz** in das Sequenzprotokoll aufgenommen werden

[Leu or Ile] ist durch das restriktivste Mehrdeutigkeitssymbol **"J"** darzustellen
 [Ser or Thr] ist durch das Symbol **"X"** mit dem Merkmalschlüssel **"VARIANT"** und Qualifier **"note"** anzugeben, mit der dass **X Serin oder Threonin** ist

Die bevorzugte Darstellung der Sequenz ist **GGGJATX**

Deletion(en), Insertion(en) oder Substitution(en)

Absatz 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 eingefügte (inserierte) oder ersetzte (substituierte) Sequenz aus mehr als 1000 Resten enthält (siehe Absatz 86).

Deletion(en), Insertion(en) oder **Substitution(en)**

Beispiel zu **Absatz 95(a)** (kann)

Eine Anmeldung enthält die folgende Offenbarung:

Peptidfragment 1: **Gly-Leu-Pro-Xaa-Arg-Ile-Cys**

(primäre Sequenz)

wobei Xaa = beliebige Aminosäure

* * *

- in anderer Ausführungsform ist Xaa = Val, Thr oder Asp

* * *

- in weiteren Ausführungsform ist Xaa = Val

*Die primäre Sequenz, Gly-Leu-Pro-Xaa-Arg-Ile-Cys, enthält eine oder mehrere **Variationen an einer einzigen Stelle** und das Auftreten dieser Variationen ist **unabhängig**.*

Deletion(en), Insertion(en) oder Substitution(en)

Beispiel zu Absatz 95(a)

- Die primäre Sequenz muss als umfassendste Ausführungsform in das Sequenzprotokoll aufgenommen werden, da das Peptidfragment 4 oder mehr spezifisch definierte Aminosäuren umfasst.

Varianten können durch Annotation der primären Sequenz dargestellt werden

- Die umfassendste Ausführungsform, die im Sequenzprotokoll enthalten sein muss, wäre gemäß Beispiel die Sequenz mit der Variation „Xaa“, die „jede beliebige Aminosäure“ sein kann

- Die Sequenz muss als **GLPXRIC** dargestellt werden und benötigt den Merkmalschlüssel "VARIANT" in Position 4 und einen Qualifier "note", der angibt, dass Xaa "eine beliebige Aminosäure" ist

- Obwohl nicht erforderlich, wird empfohlen, dass die konkreten Varianten als separate Sequenzen in das Sequenzprotokoll aufgenommen werden:

GLPVRIC

GLPIRIC

GLPDRIC

Hinweis: Eine ausführliche Erläuterung dieses Beispiels finden Sie im WIPO-Standard ST.26, Anhang VI, Beispiel 95(a)-1.

FAQ — Advanced Modul

F13: Wenn Sie in Ihrem **Variantenbeispiel** die drei spezifischen Varianten in das Sequenzprotokoll aufnehmen, sollte dann ein Merkmal oder ein **Qualifier** aufgenommen werden, der angibt, dass die Varianten mit der **primären** generischen **Sequenz verwandt sind**?

A13: Obwohl dies **nicht erforderlich** ist, wird den Antragstellern immer empfohlen, **so viel Sequenz wie möglich in ihr Sequenzprotokoll aufzunehmen**. Eine einfache Möglichkeit, eine bestimmte Variante als eine bestimmte Variante als mit einer primären generischen Sequenz verwandt zu identifizieren, ist die Aufnahme **eines "note"-Qualifizierers in das "source"-Merkmal** aufzunehmen, das die Beziehung zwischen den Sequenzen beschreibt.

Deletion(en), Insertion(en) oder Substitution(en)

Beispiel zu Absatz 95(b)

Eine Patentanmeldung beschreibt die folgende Konsensussequenz:

aatg n1 cccacgaatg n2 cac

wobei n_1 und n_2 a, t, g, oder c sein können.

Mehrere Sequenzvarianten werden wie folgt offenbart:

wenn n_1 a ist, dann ist n_2 t, g, oder c
wenn n_1 t ist, dann ist n_2 a, g, oder c
wenn n_1 g ist, dann ist n_2 t, a, oder c
wenn n_1 c ist, dann ist n_2 t, g, oder a

Die Konsensussequenz enthält Variationen an mehreren verschiedenen Stellen, und das Auftreten dieser Variationen ist voneinander abhängig

Deletion(en), Insertion(en) oder Substitution(en)

Beispiel zu Absatz 95(b)

- Die **Konsensussequenz** muss in das Sequenzprotokoll aufgenommen werden, da das Nukleinsäurefragment 10 oder mehr spezifisch definierte Nukleotide umfasst. Varianten sind als separate Sequenzen darzustellen
- Die **umfassendste Ausführungsform**, die im Sequenzprotokoll enthalten sein muss, wäre die Version mit beiden „n“-Variationen, die jeweils "a, t, g, oder c" sein können
- Die Sequenz muss als **aatgncccacgaatgncac** dargestellt werden
- ST.26 Absatz 15 besagt, dass "das Symbol **"n"** als eines der Symbole "a", "c", "g" oder "t/u" verstanden wird, es sei denn, es wird mit einer weiteren Beschreibung in der Merkmalstabelle verwendet.
- Da der Wert jedes "n"-Restes in dieser Sequenz dem **Standardwert** "a", "c", "g" oder "t" entspricht, ist hierfür kein Merkmalschlüssel bzw. Annotation erforderlich.

Sequenzvarianten

Absatz 95

Deletion(en), Insertion(en) oder Substitution(en)

Beispiel zu Absatz 95(b)

Sequenz **aatg**ncccacgaatg**n**cac

- Die durch Aufzählung ihrer Reste dargestellte Sequenz enthält an zwei unterschiedlichen Positionen Variationen, die in wechselseitiger Abhängigkeit voneinander auftreten.

- Es ist zwar **nicht erforderlich**, aber die **Aufnahme zusätzlicher Sequenzen**, mit denen weitere Ausführungsformen dargestellt werden, die wesentlicher Bestandteil der Erfindung sind, wird **nachdrücklich empfohlen**.

- Die zusätzlichen Ausführungsformen sollten nach ST.26 Absatz 95(b) als vier separate Sequenzen mit jeweils eigener Sequenzkennzahl aufgeführt werden:

wenn n1 = „a“ dann n2 = (t, g, c) = „b“
 wenn n1 = „t“ dann n2 = (a, g, c) = „v“
 wenn n1 = „g“ dann n2 = (t, a, c) = „h“
 wenn n1 = „c“ dann n2 = (t, g, a) = „d“



aatg	a	cccacgaatg	b	cac
aatg	t	cccacgaatg	v	cac
aatg	g	cccacgaatg	h	cac
aatg	c	cccacgaatg	d	cac

- Nach ST.26 Absatz 15 muss das **restriktivste Symbol** verwendet werden, um variable Positionen darzustellen. Folglich darf n2 in der Sequenz nicht durch "n" dargestellt werden.

Deletion(en), Insertion(en) oder Substitution(en)

Beispiel zu **Absatz 95(c)**

Eine Anmeldung enthält die folgende Offenbarung:

... -Met-Gly-Leu-Pro-Arg-Xaa-Arg-Ile-Cys-Lys- ...

wobei **Xaa Gly** ist oder eine **Insertion** der Sequenz
Cys-Tyr-Ile-Lys-Ser-(1000 Aminosäuren)-Leu-Thr-Pro-Lys

- Die Variante, bei der Xaa = eine Insertion von mehr als 1000 Resten ist, muss als eine **separate Sequenz** mit einer **eigenen SEQ ID-Kennzahl** in das Sequenzprotokoll aufgenommen werden.
- Die Variante, bei der **Xaa = Gly** ist, wird ebenfalls als eine separate Sequenz mit einer eigenen SEQ ID-Kennzahl in das Sequenzprotokoll aufgenommen.

...-MGLPRGRICK-...

=> *am Ende des Webinars ein paar Beispielübungen...*

Sequenzvarianten

Korrekte Annotation

- Der für die Annotation einer Variante **korrekte Merkmalschlüssel** hängt vom **Molekültyp** und der **Art der Variation** ab:

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.

Sonderfälle:
DNA/RNA-Hybridmoleküle
Uracil in DNA
Thymin in RNA

Sonderfälle

Uracil in DNA / Thymin in RNA

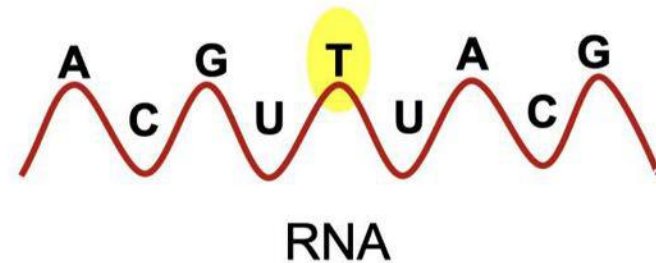
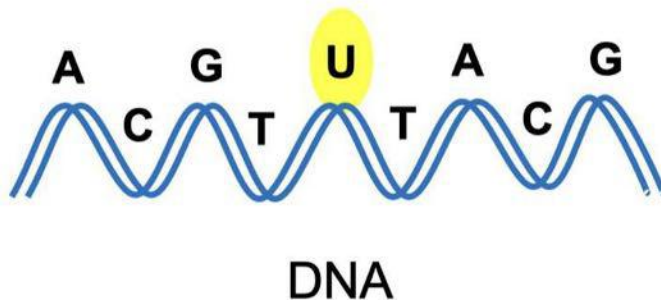
Zur Erinnerung: Das Symbol "u" für Uracil ist in ST.26 nicht zulässig.

- In DNA steht das Symbol "t" für Thymin In
- in RNA steht das Symbol "t" für Uracil
- folgende Szenarien kommen in Betracht:
 1. ein DNA-Molekül mit der Nukleobase Uracil oder ein RNA-Molekül mit der Nukleobase Thymin;
 2. ein DNA / RNA - Hybridmolekül

Sonderfälle

Uracil in DNA / Thymin in RNA

- weist eine Sequenz ein **DNA-Grundgerüst** mit einer Nukleobase **Uracil** auf **ODER** eine Sequenz mit **RNA-Grundgerüst** eine Nukleobase **Thymin**, ist es als ein **"modifiziertes Nukleotid"** zu beschreiben



es gilt ST.26, Absatz 14 + 19

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

19: ...hierzu ist der Merkmalschlüssel "modified_base", der Qualifier "mod_base" mit dem Wert "OTHER" und der Qualifier "note" mit dem Wert "uracil" bzw. "thymine" zu verwenden

Sonderfälle

Uracil in DNA / Thymin in RNA

Eine Anmeldung offenbart die folgende RNA-Sequenz:

5' -cguccacgugtccgaggua-3'

- Alle Uracilreste müssen durch das Symbol "t" dargestellt werden

cg~~t~~cccacg~~t~~tccgaggta

- das vorhandene „Thymin“ in Position 12 muss als **modifiziertes Nukleotid** annotiert werden
- Merkmalschlüssel "**modified_base**" in Position "12"
- Qualifier "**mod_base**" mit Wert "OTHER"
- Qualifier "**note**" mit dem Wert "thymine"

```
<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length>20</INSDSeq_length>
    <INSDSeq_moltype>RNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..20</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>other RNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q7">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
      <INSDFeature>
        <INSDFeature_key>modified_base</INSDFeature_key>
        <INSDFeature_location>12</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q8">
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>thymine</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>cgctcccacgtgtccgaggta</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
```

" modified_base "
in Position " 12"

" mod_base "
Wert " OTHER"

" note "
Wert " thymine"

Sonderfälle

DNA/RNA-Hybridsequenz

Nukleotidsequenz, die sowohl DNA- als RNA-Segmente eines oder mehrerer Nukleotide enthält:

Eine Anmeldung offenbart die folgende DNA/RNA-Hybridsequenz:

5' -ACCTGCcgucccacguguccgagguaGCATTA-3'

wobei Symbole in Großbuchstaben den DNA-Teil und Symbole in Kleinbuchstaben den RNA-Teil darstellen.

- ST.26, Absatz 55 legt fest: "... der Molekültyp muss als DNA angegeben werden. Das kombinierte DNA/RNA-Molekül muss in der Merkmalstabelle mit dem Merkmalschlüssel "source" und dem obligatorischen Qualifier "organism" mit dem Wert "synthetic construct" und dem obligatorischen Qualifier "mol_type" mit dem Wert "other DNA" näher beschrieben werden."
- Molekültyp = "DNA"
- Qualifier Organismusname = "synthetic construct"
- Qualifier "mol_type" = "other DNA"

das kombinierte DNA/RNA-Molekül muss mit dem Merkmalschlüssel **"source"** und dem obligatorischen Qualifier **"organism"** mit dem Wert **"synthetic construct"**, dem obligatorischen Qualifier **"mol_type"** mit Wert **"other DNA"** näher beschrieben werden

Molekültyp = "DNA"

```
<SequenceData sequenceIDNumber="4">
  <INSDSeq>
    <INSDSeq_length>32</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..32</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>other DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q10">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
  </INSDSeq>
</SequenceData>
```

Sonderfälle

DNA/RNA-Hybridsequenz

- Die Reste 1-6 und 27-32 sind **DNA**, die Reste 7-26 sind **RNA**

Eine Anmeldung offenbart die folgende DNA/RNA-Hybridsequenz:

5' - **ACCTGC**cguccacguguccgaggua**GCATTA** - 3'

wobei Symbole in Großbuchstaben den DNA-Teil und Symbole in Kleinbuchstaben den RNA-Teil darstellen.

- da alle Uracilreste durch das Symbol **"t"** dargestellt werden, muss die Sequenz im Sequenzprotokoll wiedergegeben werden als:

acctgccgtcccacgtgtccgaggtagcatta

- Zu beachten:

-> Identifizierung der **einzelnen** DNA- und RNA-**Segmente** (hier drei)

Sonderfälle

DNA/RNA-Hybridsequenz

ST.26, Absatz 55: Jedes DNA- und RNA-Segment (...) muss mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note" näher beschrieben werden

Eine Anmeldung offenbart die folgende DNA/RNA-Hybridsequenz:

5' - **ACCTGC**cguccacguguccgaggua**GCATTA** - 3'

wobei Symbole in Großbuchstaben den DNA-Teil und Symbole in Kleinbuchstaben den RNA-Teil darstellen.

der Qualifier "note" gibt an, ob das Segment DNA oder RNA ist, hier: drei Segmente = drei Merkmalschlüssel "misc_feature"

Segment 1, Reste 1-6:

Merkmalschlüssel "misc_feature" mit Lage "1..6" - Qualifier "note" mit Wert "DNA"

Segment 2, Reste 7-26:

Merkmalschlüssel "misc_feature" mit Lage "7..26" - Qualifier "note" mit Wert "RNA"

Segment 3, Reste 27-32

Merkmalschlüssel "misc_feature" mit Lage "27..32" - Qualifier "note" mit Wert "DNA"

F7: Warum ist es in diesem Beispiel einer **hybriden DNA/RNA-Sequenz** erforderlich „**misc_feature**“ als **DNA** anzugeben, wenn der „**mol_type**“ **DNA** ist ?

A7: Wenn eine Anmeldung eine DNA/RNA-**Hybridsequenz** offenbart, muss die Sequenz in das Sequenzprotokoll mit dem **Molekültyp "DNA"** aufgenommen werden, und der Wert für den **obligatorischen Qualifier "mol_type"** des **Merkmalschlüssels "source"** ist **"other DNA"**.

ST.26 **Absatz 55** schreibt ausdrücklich vor, dass "**jedes DNA- und RNA-Segment des kombinierten DNA/RNA-Moleküls mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note", der angibt, ob es sich bei dem Segment um DNA oder RNA handelt, weiter beschrieben werden muss**". So müssen bei einem DNA/RNA-Hybridmolekül die DNA-Segmente mit einem "**misc_feature**" angegeben werden, obwohl der Molekültyp DNA ist. Dadurch wird eine vollständige und **genaue Beschreibung aller Segmente** des Moleküls gewährleistet

Sonderfälle

DNA/RNA-Hybridsequenz

ST.26, Absatz 55: *Jedes DNA- und RNA-Segment (...) muss mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note" näher beschrieben werden*

```

<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>1..6</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q11">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>DNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>7..26</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q12">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>RNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>27..32</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q13">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>DNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>acctgccgtccccacgtgtccgaggtagcatta</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
</ST26SequenceListing>

```

DNA-Segment 1

RNA-Segment 2

DNA-Segment 3

F8: Müssen im Beispiel der hybriden DNA/RNA-Sequenz (Merkmalsposition 1-6 und 7-26 und 27-32) die Merkmale im generierten XML-Sequenzprotokoll in **dieser Reihenfolge** erscheinen? Oder können die drei Merkmalschlüssel in **beliebiger Reihenfolge** erscheinen, solange alle Positionen berücksichtigt werden?

A8: In ST.26 ist **nicht vorgeschrieben**, dass die Merkmalschlüssel in einer bestimmten Reihenfolge erscheinen müssen. Für ein DNA/RNA-Hybridmolekül muss **jedes DNA- und RNA-Segment** mit dem Merkmalschlüssel **„misc_feature“** berücksichtigt werden (d. h. jeder Rest muss in einem „misc_feature“ feature key enthalten sein), aber **eine bestimmte Reihenfolge ist nicht erforderlich**. Aus **Gründen der Klarheit und des besseren Verständnisses** wird jedoch **empfohlen**, die Merkmale nach Möglichkeit in **sequenzieller Reihenfolge** aufzunehmen

Anleitung und Handlungsempfehlungen zur
Darstellung von Sequenzen mit
modifizierten Resten, neuen Molekülarten
und Sequenzvarianten

Leitfaden - Anhang VI

ANHANG VI

LEITFADEN MIT ILLUSTRierten BEISPIELEN

Version 1.4

*Vom WIPO-Normenausschuss (Committee on WIPO Standards, CWS) auf seiner 8. Tagung am 4. Dezember 2020
genehmigte Revision*

INHALTSVERZEICHNIS

EINFÜHRUNG.....	<u>3.26.vi.2</u>
INDEX DER BEISPIELE.....	<u>3.26.vi.7</u>
BEISPIELE.....	<u>3.26.vi.17</u>
ANLAGE.....	<u>3.26.vi.74</u>

EINFÜHRUNG

Dieser Standard soll u. a. "Anmeldern ermöglichen, für eine Patentanmeldung ein einzelnes/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 offenbarer Sequenzen gelten.

Kriterien für die Aufnahme von Sequenzen (in einer Patentanmeldung) in ein Sequenzprotokoll

- erfordert die Berücksichtigung der folgenden drei ST.26-Fragen:
 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 dieser Sequenz zulässig (möglich)?
 3. Wenn die Aufnahme der offenbarten Sequenz nach ST.26 erforderlich oder zulässig ist, wie sollte diese Sequenz dann dargestellt werden?

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

ST.26 Absatz 7 verlangt (mit weiteren Spezifizierungen) die Aufnahme einer in einer beliebigen Stelle in der Patentanmeldung offenbaren Sequenz unter Aufzählung der Reste, wenn die Sequenz zehn oder mehr spezifisch definierte Nukleotide oder vier oder mehr spezifisch definierte Aminosäuren enthält.

was bedeutet „Aufzählung der Reste“ (enumeration of its residues)

gemäß ST.26 Absatz 3(c) bedeutet "Aufzählung der Reste" die Offenbarung einer Sequenz in einer Patentanmeldung, indem jeder Rest der Sequenz der Reihe nach aufgeführt wird;

(i) der Rest wird durch einen Namen, eine Abkürzung, ein Symbol oder eine Struktur dargestellt oder

(ii) Mehrfachreste werden durch eine Kurzformel dargestellt

die "Aufzählung der Reste" sollte unter Verwendung der Standard-Symbole offenbart werden, d.h. der Symbole gemäß ST.26 Anhang I, Abschnitt 1, Tabelle 1 bzw. Abschnitt 3, Tabelle 3 - andere als die in diesen Tabellen aufgeführten Symbole sind "nicht konventionell".

- Wenn ein **nicht konventionelles Nukleotidsymbol** als **Mehrdeutigkeitssymbol** verwendet wird (z. B. X1 = 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, 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. Z1 bedeutet „A“, „G“, „S“ oder „T“), aber keinem der konventionellen Mehrdeutigkeitssymbole in Abschnitt 3, Tabelle 3 („B“, „Z“, „J“ oder „X“) entspricht, wird der Rest als „X“-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.

wie sollen **modifizierte Reste** dargestellt werden

- ST.26 gibt an, dass modifizierte Nukleotide und Aminosäuren im Sequenzprotokoll nach Möglichkeit **als entsprechender unveränderter Rest** dargestellt werden sollten (siehe Absatz 16 und 29).
- hierbei ist zu beachten, dass es sich bei dieser **Empfehlung um ein "sollte"** handelt, d.h. um ein **"nachdrücklich empfohlene"**, jedoch nicht zwingende Vorgehensweise" (siehe Absatz 4 Buchstabe d)).
- Es liegt im **Ermessen des Anmelders** zu entscheiden, ob ein veränderter Rest durch den entsprechenden unveränderten Rest oder die Variablen "n" oder "X" dargestellt werden soll.
- allgemeine **Faustregel**: Wenn ein Rest durch Hinzufügen einer Komponente, wie z. B. Methylierung oder Acetylierung, modifiziert wird und die Struktur des nicht-modifizierten Restes im Wesentlichen unverändert bleibt, ist die Darstellung des nicht-modifizierten Rests zu empfehlen.

beispielsweise sollte ein **methyliertes Adenosin** im Sequenzprotokoll vorzugsweise durch **"a"** dargestellt werden.
- **unterscheidet sich der modifizierte Rest jedoch strukturell von einem unmodifizierten Rest, so wird ein "n" oder ein "X" empfohlen.**

F15: Warum gelten **Ile und Leu** als ähnlich genug für die Kategorie "J", Ser und Thr dagegen nicht?

A15: ST.26 wurde entwickelt, um sicherzustellen, dass die Sequenzdaten in einem Format vorliegen, das mit den Anforderungen der **INSDC (International Nucleotide Sequence Database Collaboration)** konform ist. Daher sind die in ST.26 zulässigen Aminosäuresymbole die von der INSDC definierten Aminosäuresymbole. **Die INSDC hat kein Symbol für "Ser oder Thr", aber ein Symbol (J) für "Leucin oder Isoleucin".**

Anhang VI

Beispiele und Übungen

Beispiel (93-2)

Sequenzvarianten

Darstellung der durch Aufzählung wiedergegebenen Varianten

Die Beschreibung enthält die folgende Tabelle eines Peptids und dessen funktioneller Varianten.

Position	1	2	3	4	5	6	7	8	9
Sequence	A	V	L	T	Y	L	R	G	E
Variant 1									A
Variant 2			P			P			
Variant 3			A	I	G	Y			
Variant 4							-		

Ein **Leerraum** in der Tabelle gibt an, dass eine Aminosäure in der Variante mit der entsprechenden Aminosäure in der "Sequence" **übereinstimmt**, und ein "-" bedeutet die Deletion der entsprechenden Aminosäure in der "Sequence".

- Welche **Absätze aus ST.26** treffen auf die folgende Offenbarung zu?
- hierzu stellen wir uns die **ST.26-Fragen**:

Position	1	2	3	4	5	6	7	8	9
Sequence	A	V	L	T	Y	L	R	G	E
Variant 1									A
Variant 2			P			P			
Variant 3			A	I	G	Y			
Variant 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 mit der entsprechenden Aminosäure in der "Sequenz" **identisch** ist. Folglich werden die Aminosäuren der Sequenzvarianten aufgezählt und spezifisch definiert.
- Da die vier Sequenzvarianten jeweils mehr als vier aufgezählte und spezifisch definierte Aminosäuren enthalten, müssen sie nach ST.26 **Absatz 7(b)** alle mit einer eigenen Sequenzkennzahl in ein Sequenzprotokoll aufgenommen werden.

Frage 2 erübrigt sich damit.

Beispielübung

Sequenzvarianten

Position	1	2	3	4	5	6	7	8	9
Sequence	A	V	L	T	Y	L	R	G	E
Variant 1									A
Variant 2			P			P			
Variant 3			A	I	G	Y			
Variant 4							-		

AVLTYLRGE (SEQ ID NO: 1)
 AVLTYLRGA (SEQ ID NO: 2)
 AVPTYPRGE (SEQ ID NO: 3)
 AVAIGYRGE (SEQ ID NO: 4)
 AVLTYLGE (SEQ ID NO: 5)

Frage 3: Wie sollten die Sequenzen im Sequenzprotokoll dargestellt werden?

Gemäß **Absatz 26** müssen alle Aminosäuren in einer Sequenz 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.

Da es sich hier um eine Form von **Sequenz-Varianten** handelt, ist **Absatz 93** relevant: Warum?

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.

Absatz 94 und 95 betreffen Sequenzvarianten über eine Primärsequenz offenbaren, die alternative Reste enthält oder darauf verweist (zB auch Deletionen und Insertionen)

Relevante Absätze im Standard ST.26: 7(b), 26 und 93

Verwendung der entsprechenden unmodifizierten Aminosäure (neues Beispiel 29-2):

In einer Patentanmeldung wird die folgende Sequenz beschrieben:

Ala-**Hyl**-Tyr-Leu-Gly-Ser-**Nle**-Val-Ser-**5ALA**

Wobei

Hyl = Hydroxylysin (posttranslationale Modifikation von Lysin),
Nle = Norleucin und
5ALA = δ -Aminolävulinsäure

Frage 1: Muss ST.26 die Sequenz(en) enthalten?

JA

- Das aufgezählte Peptid enthält **mehr als vier spezifisch definierte Aminosäuren**; daher muss die Sequenz in ein Sequenzprotokoll aufgenommen werden.

Frage 3: Wie sind die Sequenz(en) darzustellen?

Verwendung der entsprechenden unmodifizierten Aminosäure (neues Beispiel 29-2):

Ala-**Hyl**-Tyr-Leu-Gly-Ser-**Nle**-Val-Ser-**5ALA**

Wobei

Hyl = Hydroxylysin (posttranslationale Modifikation von Lysin),
Nle = Norleucin und
5ALA = δ -Aminolävulinsäure

Frage 3: Wie sind die Sequenz(en) darzustellen?

- Hydroxylysin in Position 2, Norleucin in Position 7 und die δ -Aminolävulinsäure in Position 10 sind "**modifizierte Aminosäuren**".
- Für jede modifizierte Aminosäure muss festgestellt werden, ob sie **durch die entsprechende unmodifizierte Aminosäure** oder durch die **Variable "X"** in der Sequenz dargestellt werden soll.
- hierzu expliziert **Absatz 29**, dass modifizierte Aminosäuren "in der Sequenz **nach Möglichkeit** als die entsprechenden unmodifizierten Aminosäuren dargestellt werden sollten"!

Verwendung der entsprechenden unmodifizierten Aminosäure (neues Beispiel 29-2):

Ala-**Hyl**-Tyr-Leu-Gly-Ser-**Nle**-Val-Ser-**5ALA**

Wobei

Hyl = Hydroxylysin (posttranslationale Modifikation von Lysin),

Nle = Norleucin und

5ALA = δ -Aminolävulinsäure

Frage: Wie werden die modifizierten Aminosäuren dargestellt?

hier liegt es im **Ermessen des Anmelders** zu entscheiden, ob eine modifizierte Aminosäure durch den entsprechenden **unveränderten** Rest oder die Variablen **"X"** dargestellt wird.

Folgende Hinweise sollten beachtet werden:

Wenn eine Aminosäure durch **Hinzufügung einer chemischen Gruppe**, z.B. Methylierung oder Acetylierung, modifiziert wird und die **Grundstruktur** der entsprechenden unmodifizierten Aminosäure im Wesentlichen **unverändert** ist, wird die Darstellung durch die **unmodifizierte Aminosäure** empfohlen. Wenn sich die modifizierte Aminosäure **strukturell stark** von der entsprechenden unmodifizierten Aminosäure **unterscheidet**, wird die Darstellung durch ein **"X"** empfohlen.

Verwendung der entsprechenden unmodifizierten Aminosäure (neues Beispiel 29-2):

Ala-**Hyl**-Tyr-Leu-Gly-Ser-**Nle**-Val-Ser-**5ALA**

Wobei

Hyl = Hydroxylysin (posttranslationale Modifikation von Lysin),
Nle = Norleucin und
5ALA = δ -Aminolävulinsäure

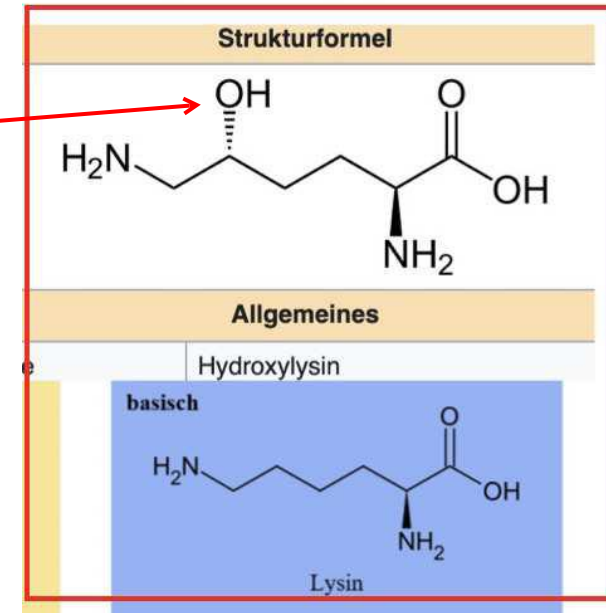
Frage: Wie werden die modifizierten Aminosäuren dargestellt?

Hydroxylysin

Die Struktur ist nahezu identisch mit der von Lysin - der dritte Kohlenstoff des Restes ist durch eine **Hydroxylgruppe** modifiziert

da die Grundstruktur des entsprechenden unmodifizierten Lysinrestes intakt ist, sollte Hydroxylysin in der Sequenz **durch Lysin ("K")** und nicht durch „X“ dargestellt werden

Hydroxylysin



Verwendung der entsprechenden unmodifizierten Aminosäure (neues Beispiel 29-2):

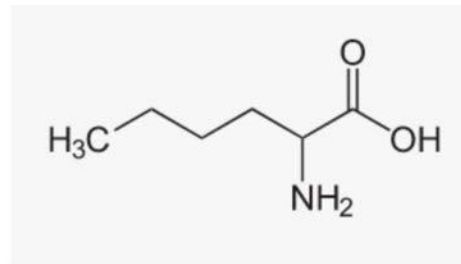
Ala-**Hyl**-Tyr-Leu-Gly-Ser-**Nle**-Val-Ser-**5ALA**

Frage: Wie werden die modifizierten Aminosäuren dargestellt?

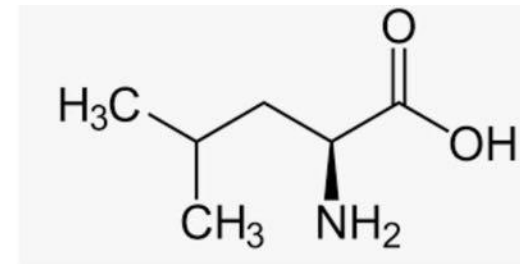
Norleucin

Norleucin ist ein Isomer von Leucin.

Die R-Gruppe von Leucin ist eine 4-er Kohlenstoff-Kette, die am zweiten Kohlenstoffatom verzweigt ist.



Norleucin – Wikipedia



Leucin – Wikipedia

Norleucin hat ebenfalls eine R-Gruppe mit 4 Kohlenstoffatomen, diese ist aber **linear und nicht verzweigt**.

Norleucin erscheint nicht als Ergebnis einer einfachen Modifikation wie sie bei Leucin hinzugefügt wurde, sondern ist als eine andere Struktur einzustufen

es wird empfohlen, Norleucin im Sequenzprotokoll mit "X" zu kennzeichnen.

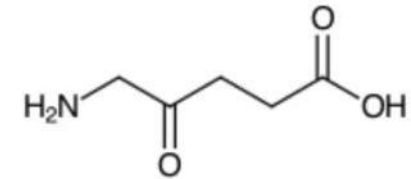
Verwendung der entsprechenden unmodifizierten Aminosäure (neues Beispiel 29-2):

Ala-**Hyl**-Tyr-Leu-Gly-Ser-**Nle**-Val-Ser-**5ALA**

Frage: Wie werden die modifizierten Aminosäuren dargestellt ?

δ -Aminolävulinsäure (5-Aminolävulinsäure)

δ -Aminolävulinsäure ist keiner der in Anhang I, Tabelle 3 aufgeführten Aminosäuren strukturell ähnlich. Daher wird auch hier empfohlen, 5-Ala mit einem „X“ zu kennzeichnen.



5-Aminolävulinsäure

■ wir können die **ST.26-Frage NO. 3** beantworten:

AKYLGXVSX

■ die erforderliche **weitere Kennzeichnung** der modifizierten Aminosäuren wird nun durch **Absatz 30** geregelt:

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

Verwendung der entsprechenden unmodifizierten Aminosäure (neues Beispiel 29-2):

Ala-**Hyl**-Tyr-Leu-Gly-Ser-**Nle**-Val-Ser-**5ALA**

Frage: Wie werden die modifizierten Aminosäuren annotiert ?

Hydroxylysin ist eine **posttranslationale Modifikation** von Lysin.

Daher muss sie mit dem Merkmalsschlüssel **"MOD_RES"** zusammen mit einem Qualifier **"note"**, der die Modifikation beschreibt, annotiert werden.

"Hydroxylysin" ist auch in Anhang 1, Abschnitt 4, Tabelle 4, "Liste der modifizierten Aminosäuren" aufgeführt. Daher kann der Wert des Qualifiers „note“ die **Abkürzung "Hyl"** anstelle des vollständigen, nicht abgekürzten Namens "Hydroxylysin" enthalten.

Norleucin ist **kein posttranslational** modifizierter Rest.

Daher muss es mit dem Merkmalsschlüssel **"SITE"** zusammen mit einem Qualifier **"note"** beschrieben werden, der die Modifikation beschreibt "Norleucin" ist auch in Anhang 1, Abschnitt 4, Tabelle 4 aufgeführt. Daher kann der Wert des Qualifiers "note" die **Abkürzung "Nle"** anstelle des vollständigen, nicht abgekürzten "Norleucin" enthalten.

δ-Aminolävulinsäure ist ebenfalls **kein posttranslational** modifizierter Rest, daher muss sie mit dem Merkmalsschlüssel **"SITE"** zusammen mit einem Qualifier **"note"** beschrieben werden, der die Modifikation beschreibt. δ-Aminolävulinsäure ist **nicht in Anhang 1**, Abschnitt 4, Tabelle 4 aufgeführt, daher muss der Wert des Qualifiers **"note" den vollständigen**, nicht abgekürzten Namen des modifizierten Restes, "δ-Aminolävulinsäure", enthalten.

Darstellung der **Variabilität** der **Sequenzlänge**

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

Beispiel einer offenbarten Sequenz

"P[RK]TN-X(2-4)-MTF-X(1-3)-SQNCE-X(0-1)-I[DE]"

Konsensus-Sequenz basierend auf natürlichen Variationen, mit konservierten und längenvariablen Bereichen

Frage 1: **Muss ST.26 die Sequenz(en) aufnehmen?**

Ja

es sind mindestens 14 **spezifisch definierte** Aminosäuren

P-[R/K]-T-N-X(2-4)-M-T-F-X(1-3)-S-Q-N-C-E-X(0-1)-I-[D/E]

Frage 3: **Wie sind die Sequenz(en) darzustellen?**

Vorschlag einer Herangehensweise:

Darstellung der Variabilität der Sequenzlänge

P-[R/K]-T-N-X(2-4)-M-T-F-X(1-3)-S-Q-N-C-E-X(0-1)-I-[D/E]

1. Feststellung der längsten Variante

Bei einer variablen Anzahl von X-Resten wird empfohlen, die maximale Anzahl von Resten in der Sequenz anzugeben („most encompassing sequence“) und zu **annotieren**, dass ein oder mehrere **Reste fehlen können**, dh mit Merkmalschlüssel **"VARIANT"** mit einem **"note"**-Qualifier

PRTN-XX— MFT-X--- SQNCE—-ID.
PKTN-XXXX-MFT- XXXtSQNCE-X-ID

17 AA (eine mögliche kürzeste Variante
22 AA (eine mögliche längste Variante

2. Feststellung der variablen Aminosäuren

Da **keine modifizierten Reste** vorhanden sind (nur natürliche Variationen) kann der Merkmalschlüssel **"VARIANT"** zusammen mit einem **"note"**-Qualifier verwendet werden, um die Variationen anzuzeigen.

Als **restriktivstes Symbol** ist „X“ zu verwenden, da die variablen Positionen, hier [R oder K] an Position 2 und [D oder E] an letzter Position, nicht durch (andere) Mehrdeutigkeitssymbole dargestellt werden können – weil **„X“ hier nicht als Standardwert** (zB "W", "Y" oder "V") verwendet wird, muss dies über den Qualifier **„note“** annotiert werden.

Darstellung der Variabilität der Sequenzlänge

P-[R/K]-T-N-X(2-4)-M-T-F-X(1-3)-S-Q-N-C-E-X(0-1)-I-[D/E]

3. Daraus ergeben sich zwei zu bevorzugende Darstellungsmöglichkeiten, entweder

A) Jeweils „X“ an Position 2 und letzter: **P-X-TNXXXMTFXXXSQNCEXI-X**; und Kennzeichnung jeder Position mit dem Merkmalschlüssel **"VARIANT"** und dem Qualifier „note“ mit dem Wert "R oder K" bzw. "D oder E", iVm den Annotationen zur Länge (s.u.)

ODER

B) wenn eine Aminosäure häufiger an einer Position auftritt, zB „R“ an Position 2 und „D“ an Position 21: **P-R-TNXXXMTFXXXSQNCEXI-D**;

- wobei jede Position mit Merkmalschlüssel **"VARIANT"** und Qualifier "note" die Information anzeigt, dass „R durch K“ und „D durch E“ ersetzt werden kann.

Möglichkeiten zur Annotation über die drei Bereiche mit Längen-Variationen

- 1. Merkmalschlüssel **'VARIANT'** an Position **5..8**, und ein 'note'-Qualifier mit dem Wert "ein oder zwei X-Reste können fehlen" ;
- 2. der Merkmalschlüssel **"VARIANT"** an Position **13..15** und ein Qualifier "note" mit dem Wert "ein oder zwei X-Reste können fehlen"
- 3. der Merkmalschlüssel **"VARIANT"** an Position **19** und ein "note"-Qualifffier mit dem Wert "Rest kann fehlen“

siehe hierzu auch Anhang VI (Guidance Document), Beispiel 36-3)

Vielen Dank für ihre
Aufmerksamkeit

noch Fragen?

Danksagung:

Emma Francis und Young-Woo Yun (Woody) (WIPO)

Katja Brabec (DPMA)

und allen die dies möglichst gemacht haben

Informationen / Links:

Webinar: WIPO ST.26: Einführung (07. September 2021):

mit Video https://www.wipo.int/meetings/en/details.jsp?meeting_id=65768

Meeting documents:

https://www.wipo.int/edocs/mdocs/mdocs/de/wipo_webinar_standards_2021_5/wipo_webinar_standards_2021_5_presentation.pdf

Allgemeine Fragen, Fragen zur WIPO Sequence Suite, PCT-Anmeldungen, Nationale/regio-nale Anmeldungen; <https://www.wipo.int/standards/de/sequence/faq.html>

Webinar: WIPO ST.26: Advanced:

https://www.wipo.int/meetings/en/details.jsp?meeting_id=62851