

WIPO-Standard ST.26 **WEITERFÜHRENDER KURS**

Webinar-Schulung

Hinweis: Dies ist ein weiterführendes Kursmodul und es wird angenommen, dass Teilnehmer bereits am ST.26-Grundlagenkurs (Basics) teilgenommen haben.

*Den ST.26-Grundlagenkurs finden Sie unter
https://www.wipo.int/meetings/en/details.jsp?meeting_id=62848*

Die heutigen Themen

- Häufig verwendete Merkmalschlüssel und Qualifier
- Formate für Merkmallagen
- Formate für Qualifier-Werte und nicht-englischsprachige Qualifier-Werte
- Sonderfälle - Uracil in DNA bzw. Thymin in RNA; DNA/RNA-Hybridmoleküle
- Nukleotidanaloga, D-Aminosäuren und verzweigte Sequenzen
- Sequenzvarianten

Häufig verwendete Merkmalschlüssel und Qualifier

Merkmalschlüssel und Qualifier

- Merkmalschlüssel können verwendet werden, um einen oder mehrere Reste einer durch die Lage identifizierte Sequenz näher zu beschreiben
 - Merkmalschlüssel für Nukleotidsequenzen sind in Anhang I, Abschnitt 5, aufgeführt
 - Merkmalschlüssel für Nukleotidsequenzen sind in Kleinbuchstaben angegeben; z.B. "misc_binding"
 - Merkmalschlüssel für Aminosäuresequenzen sind in Anhang I, Abschnitt 7, aufgeführt
 - Merkmalschlüssel für Aminosäuresequenzen sind in Großbuchstaben angegeben; z.B. "REGION"

- Qualifier können zur näheren Beschreibung von Merkmalen verwendet werden
 - Qualifier für Nukleotidsequenzen sind in Anhang I, Abschnitt 6, aufgeführt
 - Qualifier für Nukleotidsequenzen sind in Kleinbuchstaben angegeben; z.B. "allele"
 - Qualifier für Aminosäuresequenzen sind in Anhang I, Abschnitt 8, aufgeführt
 - Qualifier für Aminosäuresequenzen sind in Großbuchstaben angegeben; z.B. "NOTE"

Merkmalschlüssel und Qualifier

```

<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_quals>
    - <INSDQualifier>
      <INSDQualifier_name>ORGANISM</INSDQualifier_name>
      <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
    </INSDQualifier>
    - <INSDQualifier>
      <INSDQualifier_name>MOL_TYPE</INSDQualifier_name>
      <INSDQualifier_value>protein</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
- <INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>1</INSDFeature_location>
- <INSDFeature_quals>
  - <INSDQualifier>
    <INSDQualifier_name>NOTE</INSDQualifier_name>
    <INSDQualifier_value>X can be any amino acid</INSDQualifier_value>
  </INSDQualifier>
</INSDFeature_quals>
</INSDFeature>
</INSDSeq_feature-table>
  <INSDSeq_sequence>XYEKGJL</INSDSeq_sequence>
</INSDSeq>
</SequenceData>

```

Merkmalschlüssel und Qualifier

- Merkmalschlüssel sind fakultativ, mit Ausnahme des Merkmals "source"/"SOURCE", das für jede Sequenz obligatorisch ist.
- Jeder Merkmalschlüssel enthält eine Liste von Qualifiern, die zur näheren Beschreibung des Merkmals verwendet werden können. Die meisten Qualifier sind fakultativ; einige Merkmalschlüssel haben jedoch obligatorische Qualifier.

5.31. Feature Key	regulatory
Definition	any region of a sequence that functions in the regulation of transcription, translation, replication or chromatin structure;
Mandatory qualifiers	regulatory_class
Optional qualifiers	allele bound_moiety function gene gene_synonym map note operon phenotype pseudo pseudogene standard_name

- Die Qualifier "mol_type"/"MOL_TYPE" und "organism"/"ORGANISM" sind für das Merkmal "source"/"SOURCE" obligatorisch.

Merkmalschlüssel und Qualifier

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

5.4. Feature Key	D-loop
Definition	displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA by a single stranded invader in the reaction catalyzed by RecA protein
Optional qualifiers	allele gene gene_synonym map note
Molecule scope	DNA

Merkmalschlüssel und Qualifier

Nukleotidsequenzen: Merkmalschlüssel "modified_base"

- Der 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
 - [a, g, c, oder u] 3'-monophosphat (ST.26, Absatz 3(f))
- Ein "modifiziertes Nukleotid" ist nach Möglichkeit durch das entsprechende unmodifizierte Nukleotid (in Anhang 1, Abschnitt 1, Tabelle 1) darzustellen. Andernfalls kann es durch "n" wiedergegeben werden. Beispiel: "2'-O-methylcytidine" ist in der Sequenz durch "c" wiederzugeben. "Queuosine" ist durch "n" wiederzugeben. Das Symbol "n" steht immer nur für einen Rest.
- Der Wert des obligatorischen Qualifiers "mod_base" muss aus den Werten in Anhang I, Abschnitt 2, Tabelle 2 ausgewählt werden. Wenn der Wert "other" ist, dann muss ein zusätzlicher Qualifier "note" den vollständigen, nicht abgekürzten Namen des modifizierten Restes enthalten.

Merkmalschlüssel und Qualifier

Nukleotidsequenzen: Merkmalschlüssel "modified_base"

- 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>
```

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

Merkmalschlüssel und Qualifier

Nukleotidsequenzen: Merkmalschlüssel "modified_base"

- Beispiel: Nukleotidsequenz mit Xanthin in Position 22

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>22</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>xanthine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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 in dem Wert "xanthine" enthalten sein.

Merkmalschlüssel und Qualifier

Nukleotidsequenzen: Merkmalschlüssel "modified_base"

- "modified_base" kann auch 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 kann in der Sequenz durch ein "n" dargestellt und mithilfe eines Merkmalschlüssels "modified_base" sowie eines Qualifiers "mod_base" mit dem Wert "OTHER" und einem zusätzlichen Qualifier "note" mit dem Wert "abasic site" näher beschrieben werden.

Merkmalschlüssel und 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 den Merkmalschlüssel "CDS" gibt es KEINE obligatorischen Qualifier
- Zu den häufig verwendeten Qualifiern gehören:

“pseudo”

“pseudogene”

“translation”

“transl_table”

“codon_start”

“transl_except”

“protein_id”

Merkmalschlüssel und Qualifier

Nukleotidsequenzen: Merkmalschlüssel "CDS"

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

Qualifier	Beschreibung	Wert
pseudo	gibt an, dass das Merkmal CDS nicht-funktional ist und keine Translation hat, aber kein Pseudogen ist	keiner
pseudogene	gibt an, dass das Merkmal CDS ein Pseudogen ist und keine Translation hat	prozessiert nicht prozessiert unitär allelisch unbekannt
Translation	gibt die Aminosäuresequenz an, die von der Translation des CDS abgeleitet ist	einbuchstabile Abkürzungen von Aminosäuren

Merkmalschlüssel und Qualifier

Nukleotidsequenzen: Merkmalschlüssel "CDS"

- Eine Aminosäuresequenz mit 4 oder mehr spezifisch definierten Aminosäuren, die von einer kodierenden Sequenz kodiert und in einem Qualifier "translation" offenbart wird, muss als eine separate Sequenz in das Sequenzprotokoll aufgenommen und ihr eine eigene Sequenzkennzahl zugewiesen werden.
- 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 und Qualifier

Nukleotidsequenzen: Merkmalschlüssel "CDS"

- Qualifier, die die translatierte Sequenz modifizieren können:

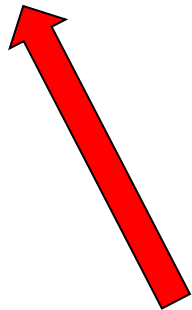
Qualifier	Beschreibung	Wert
transl_table	gibt die genetische Codetabelle an, die zur Translation des CDS verwendet wird; Standard ist "1-Standard Code"	eine Nummer, die einer Translationstabelle in Anhang I, Abschnitt 9 entspricht
transl_except	gibt die Translation eines Codons an, das nicht mit dem 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 und 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 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?

Merkmalschlüssel und 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 *
```

- Sequenz, im Sequenzprotokoll dargestellt als:

```
tggataatga agaagttaac gaagaatgta tgagattatt ttcaagaac gctcgtcatc taacatcaag gttgacataa
```

Merkmalschlüssel und 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 *
```

- Merkmalschlüssel "CDS"
- Merkmallage: < 1..80

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

Die Lage enthält das Stopcodon in den Positionen 78-80

Merkmalschlüssel und 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  *
```

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

Der "codon_start"-
Qualifier-Wert "3" gibt an,
dass das erste
vollständige Codon an der
dritten Position innerhalb
der Lage beginnt

Merkmalschlüssel und 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 *
```

- 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.

Merkmalschlüssel und 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 *
```

- Merkmalschlüssel "CDS"
- Merkmallage: < 1..80
- Qualifier "codon_start" mit dem Wert "3"
- Qualifier "transl_table" mit dem Wert "3"
- Qualifier "transl_except" mit dem Wert "(pos:30..32,aa:Sec)"

Der Qualifier
"transl_except" setzt die
Aminosäure Selenocystein
ein

Merkmalschlüssel und 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  *
```

- Merkmalschlüssel "CDS"
- Merkmallage: < 1..80
- Qualifier "codon_start" mit Wert "3"
- Qualifier "transl_table" mit dem Wert "3"
- Qualifier "transl_except" mit dem Wert "(pos:30..32,aa:Sec)"
- Qualifier "translation" mit dem Wert "**DNEEVNEECURLFFKNARHTTSRLT**"

Das Stopcodon wird im
Qualifier "translation" nicht
ausgewiesen!

Merkmalschlüssel und 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 *
```

- Merkmalschlüssel "CDS"
- Merkmallage: < 1..80
- Qualifier "codon_start" mit Wert "3"
- Qualifier "transl_table" mit dem Wert "3"
- Qualifier "transl_except" mit dem Wert "(pos:30..32,aa:Sec)"
- Qualifier "translation" mit dem Wert "**DNEEVNEECURLFFKNARHTTSRLT**"
- eine separate Proteinsequenz für die Translation

Merkmalschlüssel und 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 *
```

- Merkmalschlüssel "CDS"
- Merkmallage: < 1..80
- Qualifier "codon_start" mit Wert "3"
- Qualifier "transl_table" mit dem Wert "3"
- Qualifier "transl_except" mit dem Wert "(pos:30..32,aa:Sec)"
- Qualifier "translation" mit dem Wert "**DNEEVNEECURLFFKNARHTTSRLT**"
- eine separate Proteinsequenz für die Translation
- Qualifier "protein_id" mit der SEQ ID-Kennzahl des translatierten Proteins

Merkmalschlüssel und Qualifier

Nukleotidsequenzen: Merkmalschlüssel "CDS"

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

`join(location1,location2)`

- Die Merkmallage "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

Aminosäuresequenzen: Häufig verwendete Merkmalschlüssel

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

Merkmalschlüssel und Qualifier

Aminosäuresequenzen: Modifizierte Aminosäuren

- Eine "modifizierte Aminosäure" ist jede Aminosäure außer:

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	

(ST.26, Absatz 3(e))

- Eine "modifizierte Aminosäure" ist nach Möglichkeit durch die entsprechende unmodifizierte Aminosäure darzustellen. Andernfalls kann sie durch "X" dargestellt werden. Beispielsweise ist "hydroxylysine" in der Sequenz durch "K" darzustellen. "Ornithine" ist durch "X" darzustellen.

(ST.26, Absatz 29).

- Das Symbol "X" steht immer nur für einen Rest.

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.

Formate für Merkmalagen

Formate für Lagen

Lagedeskriptoren für alle Molekülarten

- Lagedeskriptoren werden zur Identifizierung der Lage eines Merkmals in einer Sequenz verwendet
- ST.26 hat obligatorische Anforderungen an das Format von Lagedeskriptoren
- Die folgenden Formate von Lagedeskriptoren können für Nukleotid- und Aminosäuresequenzen verwendet werden:

Location descriptor type	Syntax	Description
Single residue number	x	Points to a single residue in the sequence.
Residue numbers delimitating a sequence span	x..y	Points to a continuous range of residues bounded by and including the starting and ending residues.
Residues before the first or beyond the last specified residue number	<x >x <x..y x..>y <x..>y	Points to a region including a specified residue or span of residues and extending beyond a specified residue. The '<' and '>' symbols may be used with a single residue or the starting and ending residue numbers of a span of residues to indicate that a feature extends beyond the specified residue number.

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

Formate für Lagen

Lagedescriptoren für alle Molekülarten

Location Example	Description
467	Points to residue 467 in the sequence.
340..565	Points to a continuous range of residues bounded by and including residues 340 and 565.
<1	Points to a feature location before the first residue.
<345..500	Indicates that the exact lower boundary point of a feature is unknown. The location begins at some residue previous to 345 and continues to and includes residue 500.
<1..888	Indicates that the feature starts before the first sequenced residue and continues to and includes residue 888.
1..>888	Indicates that the feature starts at the first sequenced residue and continues beyond residue 888.
<1..>888	Indicates that the feature starts before the first sequenced residue and continues beyond residue 888.

WIPO-Standard ST.26, Absatz 70(a)

Formate für Lagen

Lagedeskriptoren ausschließlich für Nukleotidsequenzen

- Das folgende Lagedeskriptor-Format kann NUR für DNA- und RNA-Sequenzen verwendet werden:

Location descriptor type	Syntax	Description
A site between two adjoining nucleotides	x^y	Points to a site between two adjoining nucleotides, e.g., endonucleolytic cleavage site. The position numbers for the adjacent nucleotides are separated by a caret (^). The permitted formats for this descriptor are x^{x+1} (for example 55 ⁵⁶), or, for circular nucleotides, x^1 , where “x” is the full length of the molecule, i.e. 1000 ¹ for circular molecule with length 1000.

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

Formate für Lagen

Lagedeskriptoren ausschließlich für Nukleotidsequenzen

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

Location example	Description
123 ¹²⁴	Points to a site between residues 123 and 124.
867 ¹	In a circular molecule with 867 residues, points to a site between the residue indicated as position 1 and the residue indicated as position 867

WIPO-Standard ST.26, Absatz 70(b)

Formate für Lagen

Lagedeskriptoren – Sonderfall für Aminosäuresequenzen

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

Location descriptor type	Syntax	Description
Residue numbers joined by an intrachain cross-link	$x..y$	Points to amino acids joined by an intrachain linkage when used with a feature that indicates an intrachain cross-link, such as "CROSSLNK" or "DISULFID".

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

Formate für Lagen

Lagedeskriptoren – Sonderfall: Aminosäuresequenzen

```

<SequenceData sequenceIDNumber="4">
  <INSDSeq>
    <INSDSeq_length>81</INSDSeq_length>
    <INSDSeq_moltype>AA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>SOURCE</INSDFeature_key>
        <INSDFeature_location>1..81</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>MOL_TYPE</INSDQualifier_name>
            <INSDQualifier_value>protein</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q25">
            <INSDQualifier_name>ORGANISM</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
      <INSDFeature>
        <INSDFeature_key>DISULFID</INSDFeature_key>
        <INSDFeature_location>30..50</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier id="q22">
            <INSDQualifier_name>NOTE</INSDQualifier_name>
            <INSDQualifier_value>disulfide bond</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>QKKMIQFFKITHRYYYDIIIEHLCAKYDMNSVISNALFAKLNLMQYTDGLSTHEKIIINTSNPLTFSIVISLQRCVINLGGST</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```


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
- "complement" kann in Kombination mit "join" oder "order" verwendet werden

Location syntax	Location description
<code>join(location, location, ..., location)</code>	The indicated locations are joined (placed end-to-end) to form one contiguous sequence.
<code>order(location, location, ..., location)</code>	The elements are found in the specified order but nothing is implied about whether joining those elements is reasonable.
<code>complement(location)</code>	Indicates that the feature is located on the strand complementary to the sequence span specified by the location descriptor, when read in the 5' to 3' direction or in the direction that mimics the 5' to 3' direction.

Formate für Lagen

Lageoperatoren für Nukleotidsequenzen

Location example	Description
<code>join(12..78,134..202)</code>	Indicates that regions 12 to 78 and 134 to 202 should be joined to form one contiguous sequence.
<code>order(15..228,341..502)</code>	Indicates that regions 15 to 228 and 341 to 502 are present in the specified order
<code>complement(34..126)</code>	Starts at the nucleotide complementary to 126 and finishes at the nucleotide complementary to nucleotide 34 (the feature is on the strand complementary to the presented strand).
<code>complement(join(2691..4571,4918..5163))</code>	Joins nucleotides 2691 to 4571 and 4918 to 5163, then complements the joined segments (the feature is on the strand complementary to the presented strand).
<code>join(complement(4918..5163),complement(2691..4571))</code>	Complements regions 4918 to 5163 and 2691 to 4571, then joins the complemented segments (the feature is on the strand complementary to the presented strand).

WIPO-Standard ST.26, Absatz 70(b)

Fragen?

Formate für Qualifier-Werte

Qualifier-Werte

- Qualifier definieren Merkmale näher

Qualifier-Werte

- Qualifier definieren Merkmale näher
- Qualifier bestehen aus einem Qualifier-Namen und oft einem Qualifier-Wert

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

Qualifier-Werte

- Qualifier definieren Merkmale näher
- Qualifier bestehen aus einem Qualifier-Namen und oft einem Qualifier-Wert

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

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

Qualifier-Werte

- Qualifier definieren Merkmale näher
- Qualifier bestehen aus einem Qualifier-Namen und oft einem Qualifier-Wert

```
<INSDQualifier>  
  <INSDQualifier_name>transl_table</INSDQualifier_name>  
  <INSDQualifier_value>12</INSDQualifier_value>  
</INSDQualifier>
```
- Jeder Merkmalschlüssel enthält eine Liste der Qualifier, die für dieses Merkmal zulässig sind. Einige Merkmalschlüssel haben obligatorische Qualifier.
- Merkmalschlüssel und ihre zulässigen Qualifier sind in ST.26, Anhang I, Abschnitt 5 (Nukleotidsequenzen) und Abschnitt 7 (Aminosäuresequenzen) aufgeführt.

Qualifier-Werte

- Beispiel: der Merkmalschlüssel "misc_binding" hat einen obligatorischen Qualifier "bound_moiety" und 6 fakultative Qualifier:

5.12.	Feature Key	misc_binding
	Definition	site in nucleic acid which covalently or non-covalently binds another moiety that cannot be described by any other binding key (primer_bind or protein_bind)
	Mandatory qualifiers	bound_moiety
	Optional qualifiers	allele function gene gene_synonym map note
	Comment	note that the regulatory feature key and regulatory_class qualifier with the value "ribosome_binding_site" must be used for describing ribosome binding sites

(ST.26, Anhang I, Abschnitt 5.12)

Qualifier-Werte

- Qualifier definieren Merkmale näher
- Qualifier bestehen aus einem Qualifier-Namen und oft einem Qualifier-Wert

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

- Jeder Merkmalschlüssel enthält eine Liste der Qualifier, die für dieses Merkmal zulässig sind. Einige Merkmalschlüssel haben obligatorische Qualifier.
- Merkmalschlüssel und ihre zulässigen Qualifier sind in ST.26, Anhang I, Abschnitt 5 (Nukleotidsequenzen) und Abschnitt 7 (Aminosäuresequenzen) aufgeführt.
- Qualifier für Nukleotidsequenzen sind zusammen mit ihren Beschreibungen in ST.26, Anhang I, Abschnitt 6 (Nukleotidsequenzen) und Abschnitt 8 (Aminosäuresequenzen) aufgeführt.

Qualifier-Werte

6.3.	Qualifier	<code>bound_moiety</code>
	Definition	name of the molecule/complex that may bind to the given feature
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<code><INSDQualifier_value>GAL4</INSDQualifier_value></code>
	Comment	A single <code>bound_moiety</code> qualifier is permitted on the "misc_binding", "oriT" and "protein_bind" features.

6.22.	Qualifier	<code>gene</code>
	Definition	symbol of the gene corresponding to a sequence region
	Mandatory value format	free text
	Example	<code><INSDQualifier_value>ilvE</INSDQualifier_value></code>
	Comment	Use <code>gene</code> qualifier to provide the gene symbol; use <code>standard_name</code> qualifier to provide the full gene name.

(ST.26, Anhang I, Abschnitte 6.3 und 6.22)

Qualifier-Werte

Formatarten

■ Qualifier-Werte haben mehrere Formatarten:

1. Qualifier mit vordefinierten Wertoptionen;
2. Qualifier mit einem definierten Wertformat;
3. Qualifier, bei denen der Wert eine Sequenz ist;
4. Qualifier mit KEINEM Wert;
5. Qualifier mit "free text"-Werten

- eine Teilmenge von "free text"-Qualifier-Werten wird als "language dependent" kategorisiert

Qualifier-Werte

Formatarten - Vordefinierte Wertoptionen

- Qualifier mit vordefinierten Wertoptionen

- Beispiele:

"codon_start" – Werte können "1", "2" oder "3" sein

6.9.	Qualifier	codon_start
	Definition	indicates the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature.
	Mandatory value format	1 or 2 or 3
	Example	<INSDQualifier_value>2</INSDQualifier_value>

Qualifier-Werte

Formatarten – Vordefinierte Wertoptionen

- Beispiele: "rpt_type" hat einen begrenzten Satz an Wertoptionen

6.60. Qualifier	rpt_type
Definition	structure and distribution of repeated sequence
Mandatory value format	One of the following controlled vocabulary terms or phrases: 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
Example	<INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value>
Comment	Definitions of the values: tandem - a repeat that exists adjacent to another in the same orientation; direct - a repeat that exists not always adjacent but is in the same orientation; inverted - a repeat pair occurring in reverse orientation to one another on the same molecule;

Qualifier-Werte

Formatarten – Vordefinierte Wertoptionen

- 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 *	Qualifier Value
rpt_type	<ul style="list-style-type: none"> 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	atgggctct	ggaagtgccg	ctgacggccc	agaagggcgt
cttactgtta	caagaagagt	atcatgacgc	gcatcatagc	cacttagacc	cgcacttatg
gttggcgccg	caaaatatca	ttgcctcgt	gcaaggattg	gataaacata	ctgctgagtt

180

Qualifier-Werte

Formatarten – definiertes Wertformat

- Qualifier mit einem definierten Wertformat

- Beispiel: "anticodon" – Wert muss im Format

"(pos:<location>,aa:<amino_acid>,seq:<text>)" vorliegen

6.2.	Qualifier	anticodon
	Definition	location of the anticodon of tRNA and the amino acid for which it codes
	Mandatory value format	(pos:<location>,aa:<amino_acid>,seq:<text>) where <location> is the position of the anticodon and <amino_acid> is the three letter abbreviation for the amino acid encoded and <text> is the sequence of the anticodon
	Example	<pre><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></pre>

Qualifier-Werte

Formatarten - Sequenzen

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

6.79. Qualifier	translation
Definition	one-letter abbreviated amino acid sequence derived from either the standard (or universal) genetic code or the table as specified in a transl_table qualifier and as determined by an exception in the transl_except qualifier
Mandatory value format	contiguous string of one-letter amino acid abbreviations from Section 3 of this Annex, "X" is to be used for AA exceptions.
Example	<INSDQualifier_value>MASTFPPWYRGCASTPSLKGLIMCTW</INSDQualifier_value>
Comment	to be used with CDS feature only; must be accompanied by protein_id qualifier when the translation product contains four or more specifically defined amino acids; see transl_table for definition and location of genetic code Tables; only one of the qualifiers translation, pseudo and pseudogene are permitted to further annotate a CDS feature.

Qualifier-Werte

Formatarten - Sequenzen

- Qualifier "replace" – Wert kann ein einzelner Nukleotidrest, eine Sequenz von Resten, oder leer sein
- Ein leerer Wert für "replace" bedeutet eine Deletion des im entsprechenden Merkmal angegebenen Rests

6.57. Qualifier	replace
Definition	indicates that the sequence identified in a feature's location is replaced by the sequence shown in the qualifier's value; if no sequence (i.e., no value) is contained within the qualifier, this indicates a deletion
Mandatory value format	free text
Example	<pre><INSDQualifier_value>a</INSDQualifier_value> <INSDQualifier_value></INSDQualifier_value> - for a deletion</pre>

Qualifier-Werte

Formatarten - Sequenzen

- Qualifier "replace" – Wert kann ein einzelner Nukleotidrest, eine Sequenz von Resten, oder leer sein
- Ein leerer Wert für "replace" bedeutet eine Deletion des im entsprechenden Merkmal angegebenen Rests

6.57. Qualifier	replace
Definition	indicates that the sequence identified in a feature's location is replaced by the sequence shown in the qualifier's value; if no sequence (i.e., no value) is contained within the qualifier, this indicates a deletion
Mandatory value format	free text
Example	<pre><INSDQualifier value>a</INSDQualifier value> <INSDQualifier_value></INSDQualifier_value> - for a deletion</pre>

"Empty value"

Qualifier-Werte

Formatarten - KEIN Wert

- Qualifier mit KEINEM Wert

- Beispiele:

“environmental_sample”

“germline”

“macronuclear”

“proviral”

6.51. Qualifier	proviral
Definition	this qualifier is used to flag sequence obtained from a virus or phage that is integrated into the genome of another organism
Value format	none

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

Qualifier-Werte

Formatarten – "Free Text" (Freitext)

- Viele Qualifier haben ein "Free text"-Wertformat
- ST.26, Absatz 3(n) definiert "free text" als "*...eine Art von Wertformat für bestimmte Qualifier, das in Form eines beschreibenden Textausdrucks oder eines anderen angegebenen Formats (siehe Anhang I) dargestellt wird.*"
- "Free text"-Qualifier-Werte sind auf 1000 Zeichen begrenzt (ST.26, Absatz 86)
- Eine Teilmenge von Qualifiern mit "free text"-Wertformaten ist "*language-dependent*"
- "Language-dependent" Freitext "kann für nationale, regionale oder internationale Verfahren eine Übersetzung erfordern." ST.26, Absatz 3(o)

Qualifier-Werte

Formatarten – "Free text"

- Qualifier, die "language-dependent free text"-Werte haben, müssen eventuell übersetzt werden
- Wie stellen Sie fest, ob ein Qualifier mit einem Freitext-Wertformat sprachabhängig ist?
 - ST.26, Anhang I, Abschnitt 6, Tabelle 5 listet alle Nukleotidsequenz-Qualifier mit einem sprachabhängigen Freitext-Wertformat auf
 - ST.26, Anhang I, Abschnitt 8, Tabelle 6 listet alle Aminosäuresequenz-Qualifier mit einem sprachabhängigen Freitext-Wertformat auf
 - Beachten Sie das "Obligatorische Wertformat" in der Beschreibung der Qualifier

6.5. Qualifier	cell_type
Definition	cell type from which the sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>leukocyte</INSDQualifier_value>

Qualifier-Werte

Formatarten – "Free text"

6.20. Qualifier	frequency
Definition	frequency of the occurrence of a feature
Mandatory value format	free text representing the proportion of a population carrying the feature expressed as a fraction
Example	<pre><INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value></pre>
6.21. Qualifier	function
Definition	function attributed to a sequence
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<pre><INSDQualifier_value>essential for recognition of cofactor </INSDQualifier_value></pre>
Comment	The function qualifier is used when the gene name and/or product name do not convey the function attributable to a sequence.

Qualifier-Werte

Formatarten – "Language-Dependent Free Text"

- "Language-dependent free text"-Qualifier-Werte können im Sequenzprotokoll-XML in zwei Sprachen angegeben werden: Englisch und eine weitere nicht-englische Sprache (ST.26, Absatz 87)

Qualifier-Werte

Formatarten – “Language-Dependent Free Text”

- "Language-dependent free text"-Qualifier-Werte können im Sequenzprotokoll-XML in zwei Sprachen angegeben werden: Englisch und eine weitere nicht-englische Sprache (ST.26, Absatz 87)
- Ein sprachabhängiger Qualifier-Wert in Englisch muss in einem Element "INSDQualifier_value" angegeben werden

Qualifier-Werte

Formatarten – "Language-Dependent Free Text"

- "Language-dependent free text"-Qualifier-Werte können im Sequenzprotokoll-XML in zwei Sprachen angegeben werden: Englisch und eine weitere nicht-englische Sprache (ST.26, Absatz 87)
- Ein sprachabhängiger Qualifier-Wert in Englisch muss 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

Qualifier-Werte

Formatarten – "Language-Dependent Free Text"

- "Language-Dependent free Text"-Qualifier-Werte können im Sequenzprotokoll-XML in zwei Sprachen angegeben werden: Englisch und eine weitere nicht-englische Sprache (ST.26, Absatz 87)
- Ein sprachabhängiger Qualifier-Wert in Englisch muss 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

Formatarten – "Language-Dependent Free Text"

```

<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>GSLSDVRKDVVEKRIDKALEAFKNKMDKEK</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="3">

```

Qualifier-Werte

Formatarten – "Language-Dependent Free Text"

- Wenn ein Sequenzprotokoll-XML 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 und 87(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">
```

Qualifier-Werte

Formatarten – "Language-Dependent Free Text"

- Wenn ein Sequenzprotokoll-XML 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 und 87 b));
 2. ALLE sprachabhängigen Qualifier im Sequenzprotokoll müssen Werte in der im Attribut "nonEnglishFreeTextLanguageCode" angegebenen Sprache haben;

Qualifier-Werte

Formatarten – "Language-Dependent Free Text"

- Wenn ein Sequenzprotokoll-XML 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 und 87 b));
 2. ALLE sprachabhängigen Qualifier in dem 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

Formatarten – "Language-Dependent Free Text"

- 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>
```


Qualifier-Werte

Formatarten – "Language-Dependent Free Text"

- INSDQualifier "id"-Attribute:
 - Identifizieren eindeutig Qualifier-Werte, die für den Export in eine XLIFF-Datei durch WIPO Sequence möglicherweise übersetzt werden müssen;
 - Fakultativ;
 - Nur für sprachabhängige Qualifier zulässig;
 - Müssen in einem Sequenzprotokoll eindeutig sein;
 - Werden automatisch in einer von WIPO Sequence generierten XML hinzugefügt.

Qualifier-Werte

Formatarten – "Language-Dependent Free Text"

- 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

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

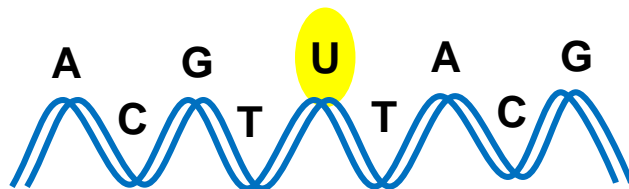
Sonderfälle

- Zur Erinnerung: Das Symbol "u" für Uracil ist in ST.26 nicht zulässig.
- In der DNA steht das Symbol "t" für Thymin
- In der RNA steht das Symbol "u" für Uracil
- Zwei Szenarien kommen in Betracht:
 1. DNA-Molekül mit der Nukleobase Uracil oder RNA-Molekül mit der Nukleobase Thymin;
 2. DNA/RNA-Hybridmolekül

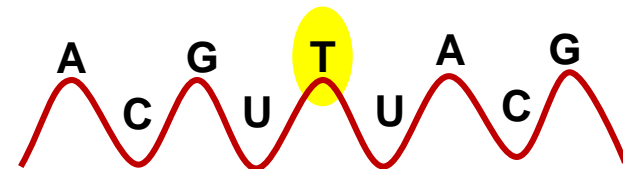
Sonderfälle

Uracil in DNA und Thymin in RNA

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



DNA



RNA

ST.26, Absatz 14 gilt hier:

"14. Das Symbol "t" wird als Thymin in DNA bzw. als Uracil in RNA ausgelegt. Uracil in DNA bzw. Thymin in RNA gelten als modifiziertes Nukleotid und müssen nach der in Absatz 19 angegebenen Merkmalstabelle näher beschrieben werden."

Sonderfälle

DNA- und RNA-Sequenzen

Eine Anmeldung offenbart die folgende RNA-Sequenz:

5' -cgucccacgugtccgaggua - 3'

- Beachten Sie "thymine" in Position 12. Dieser Rest muss als ein modifiziertes Nukleotid annotiert werden.
- ST.26, Absatz 19 legt fest: *Uracil in DNA bzw. Thymin in RNA gelten als modifizierte Nukleotide, müssen in der Sequenz mit dem Symbol "t" dargestellt werden und in der Merkmalstabelle näher beschrieben werden, indem der Merkmalschlüssel "modified_base", der Qualifier "mod_base" mit "OTHER" als Qualifier-Wert und der Qualifier "note" mit "uracil" bzw. "thymine" als Qualifier-Wert verwendet wird.*

• Sonderfälle

• DNA- und RNA-Sequenzen

Eine Anmeldung offenbart die folgende RNA-Sequenz:

5' -cgucccacgugtccgaggua - 3'

- Beachten Sie "thymine" in Position 12. Dieser Rest muss als ein modifiziertes Nukleotid annotiert werden.
- ST.26, Absatz 19 legt fest: *Uracil in DNA bzw. Thymin in RNA gelten als modifizierte Nukleotide, müssen in der Sequenz mit dem Symbol "t" dargestellt werden und in der Merkmalstabelle näher beschrieben werden, indem der Merkmalschlüssel "modified_base", der Qualifier "mod_base" mit "OTHER" als Qualifier-Wert und der Qualifier "note" mit "uracil" bzw. "thymine" als Qualifier-Wert verwendet wird.*

Sonderfälle

DNA- und RNA-Sequenzen

Eine Anmeldung offenbart die folgende RNA-Sequenz:

5' -cguccacgugtccgaggua - 3'

- ✓ Alle Uracilreste müssen durch das Symbol "t" dargestellt werden. Daher muss die Sequenz im Sequenzprotokoll wie folgt wiedergegeben werden:

cg~~t~~ccccacgtgtccgaggta

Sonderfälle

DNA- und RNA-Sequenzen

Eine Anmeldung offenbart die folgende RNA-Sequenz:

5' -cguccacgugtccgaggua -3'

- ✓ Alle Uracilreste müssen durch das Symbol "t" dargestellt werden. Daher muss die Sequenz im Sequenzprotokoll wie folgt wiedergegeben werden:

cg~~t~~cccacg~~t~~gtccgaggta

- ✓ Merkmalschlüssel "modified_base" in Position "12"

Sonderfälle

DNA- und RNA-Sequenzen

Eine Anmeldung offenbart die folgende RNA-Sequenz:

5' -cguccacgugtccgaggua - 3'

- ✓ Alle Uracilreste müssen durch das Symbol "t" dargestellt werden. Daher muss die Sequenz im Sequenzprotokoll wie folgt wiedergegeben werden:

cgtcccacgtgtccgaggta

- ✓ Merkmalschlüssel "modified_base" in Position "12"
- ✓ Qualifier "mod_base" mit Wert "OTHER"

Sonderfälle

DNA- und RNA-Sequenzen

Eine Anmeldung offenbart die folgende RNA-Sequenz:

5' -cguccacgugtccgaggua - 3'

- ✓ Alle Uracilreste müssen durch das Symbol "t" dargestellt werden. Daher muss die Sequenz im Sequenzprotokoll wie folgt wiedergegeben werden:

cgtcccacgtgtccgaggta

- ✓ Merkmalschlüssel "modified_base" in Position "12"
- ✓ Qualifier "mod_base" mit Wert "OTHER"
- ✓ Qualifier "note" mit dem Wert "thymine"

Sonderfälle

DNA- und RNA-Sequenzen

```

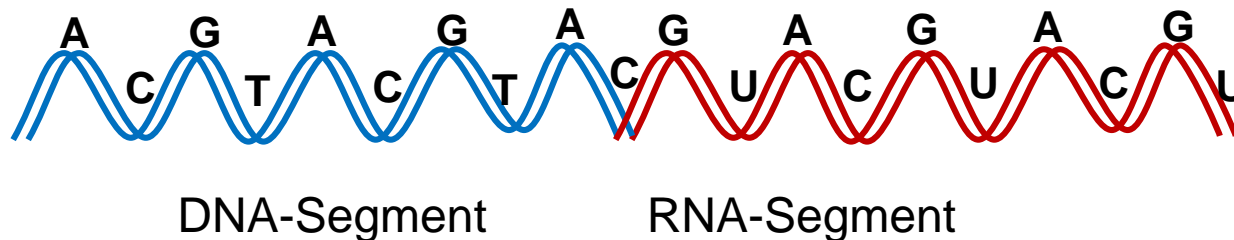
<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length>20</INSDSeq_length>
    <INSDSeq_moltype>RNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..20</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>cggtcccacgtgtccgaggt</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

Sonderfälle

DNA/RNA-Hybridsequenzen

- Wenn eine Sequenz ein Hybridmolekül ist, d.h. ein Teil des Grundgerüsts ist DNA und ein Teil des Grundgerüsts ist RNA:



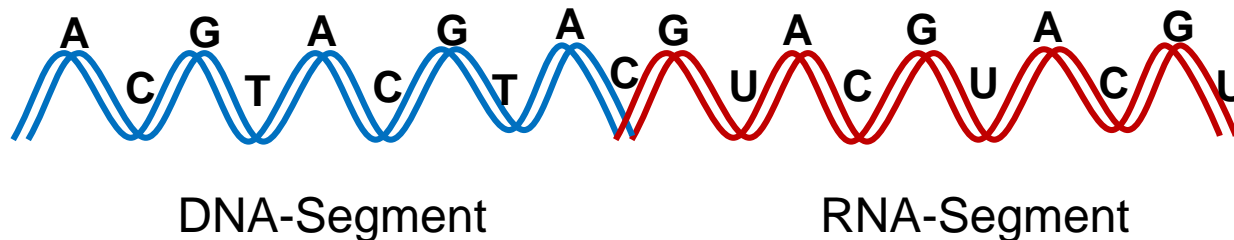
ST.26, Absatz 55 gilt hier:

“55. Bei einer Nukleotidsequenz, die sowohl DNA- als auch RNA-Segmente von einem oder mehreren Nukleotiden enthält, muss die Molekülart 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. Jedes DNA- und RNA-Segment des kombinierten DNA/RNA-Moleküls muss mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note", der angibt, ob es sich bei dem Segment um DNA oder RNA handelt, näher beschrieben werden.”

Sonderfälle

DNA/RNA-Hybridsequenzen

- Wenn eine Sequenz ein Hybridmolekül ist, d.h. ein Teil des Grundgerüsts ist DNA und ein Teil des Grundgerüsts ist RNA:



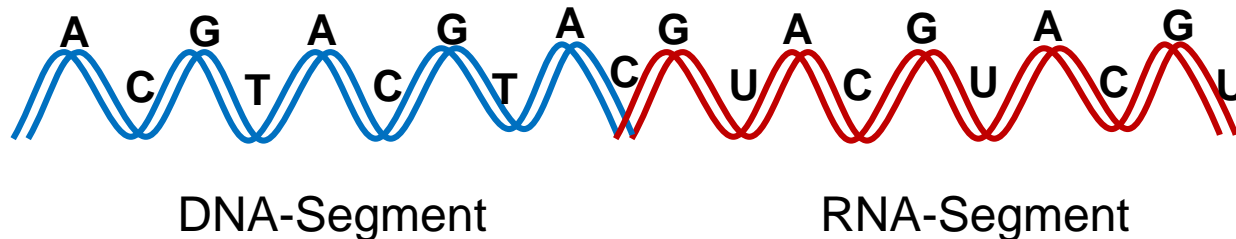
ST.26, Absatz 55 gilt hier:

“55. Bei einer Nukleotidsequenz, die sowohl DNA- als auch RNA-Segmente von einem oder mehreren Nucleotiden enthält, muss die Molekülart 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. Jedes DNA- und RNA-Segment des kombinierten DNA/RNA-Moleküls muss mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note", der angibt, ob es sich bei dem Segment um DNA oder RNA handelt, näher beschrieben werden.”

Sonderfälle

DNA/RNA-Hybridsequenzen

- Wenn eine Sequenz ein Hybridmolekül ist, d.h. ein Teil des Grundgerüsts ist DNA und ein Teil des Grundgerüsts ist RNA:



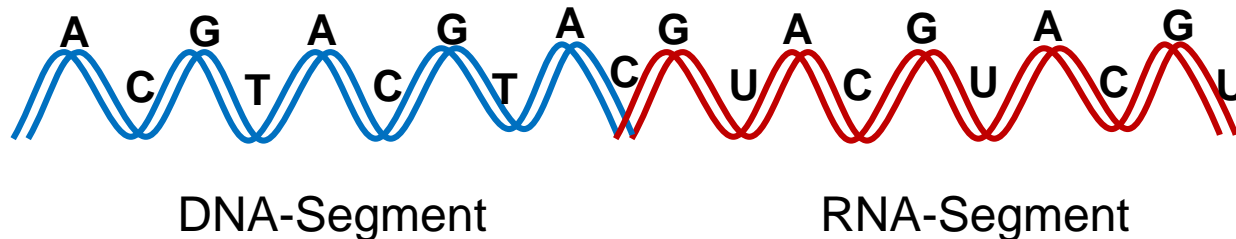
ST.26, Absatz 55 gilt hier:

“55. Bei einer Nukleotidsequenz, die sowohl DNA- als auch RNA-Segmente von einem oder mehreren Nucleotiden enthält, muss die Molekülart 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. Jedes DNA- und RNA-Segment des kombinierten DNA/RNA-Moleküls muss mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note", der angibt, ob es sich bei dem Segment um DNA oder RNA handelt, näher beschrieben werden.”

Sonderfälle

DNA/RNA-Hybridsequenzen

- Wenn eine Sequenz ein Hybridmolekül ist, d.h. ein Teil des Grundgerüsts ist DNA und ein Teil des Grundgerüsts ist RNA:



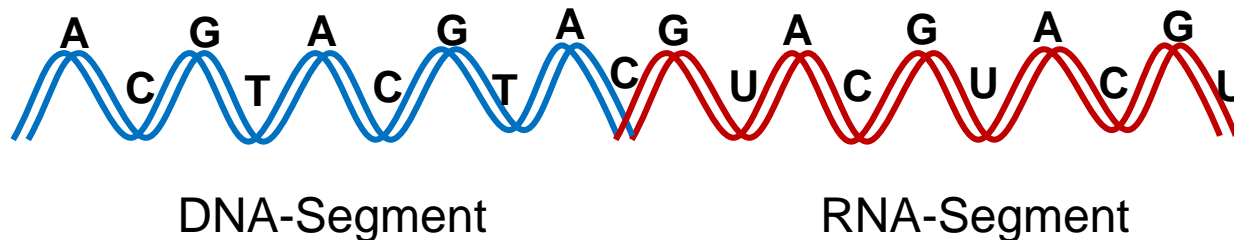
ST.26, Absatz 55 gilt hier:

“55. Bei einer Nukleotidsequenz, die sowohl DNA- als auch RNA-Segmente von einem oder mehreren Nucleotiden enthält, muss die Molekülart 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. Jedes DNA- und RNA-Segment des kombinierten DNA/RNA-Moleküls muss mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note", der angibt, ob es sich bei dem Segment um DNA oder RNA handelt, näher beschrieben werden.“

Sonderfälle

DNA/RNA-Hybridsequenzen

- Wenn eine Sequenz ein Hybridmolekül ist, d.h. ein Teil des Grundgerüsts ist DNA und ein Teil des Grundgerüsts ist RNA:



ST.26, Absatz 55 gilt hier:

“55. Bei einer Nukleotidsequenz, die sowohl DNA- als auch RNA-Segmente von einem oder mehreren Nucleotiden enthält, muss die Molekülart 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. Jedes DNA- und RNA-Segment des kombinierten DNA/RNA-Moleküls muss mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note", der angibt, ob es sich bei dem Segment um DNA oder RNA handelt, näher beschrieben werden.”

Sonderfälle

DNA/RNA-Hybridsequenzen

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.

- Die Reste 1-6 und 27-32 sind DNA, die Reste 7-26 sind RNA
- Zu beachten:
 1. Bezeichnung des Organismus
 2. Molekültyp und "mol_type"
 3. Identifizierung von DNA- und RNA-Segmenten

Sonderfälle

DNA- und RNA-Sequenzen

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"

Sonderfälle

DNA- und RNA-Sequenzen

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.

- ✓ Alle Uracilreste müssen durch das Symbol "t" dargestellt werden. Daher muss die Sequenz im Sequenzprotokoll wie folgt wiedergegeben werden:

acctgccgtcccacgtgtccgaggtagcatta

Sonderfälle

DNA- und RNA-Sequenzen

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: *"Jedes DNA- und RNA-Segment des kombinierten DNA/RNA-Moleküls muss mit dem Merkmalschlüssel "misc_feature" und dem Qualifier "note", der angibt, ob es sich bei dem Segment um DNA oder RNA handelt, näher beschrieben werden."*

✓ Drei Segmente = drei Merkmalschlüssel "misc_feature"

Sonderfälle

DNA- und RNA-Sequenzen

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.

- ✓ 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"

Sonderfälle

DNA- und RNA-Sequenzen

```

<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_quals>
          <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_quals>
      </INSDFeature>
    </INSDSeq_feature-table>
  </INSDSeq>
</SequenceData>

```


Fragen?

Nukleotidanaloga, D-Aminosäuren und verzweigte Sequenzen

Neue Molekülarten

Nukleotidanaloga

- Nukleinsäuresequenzen, die ein oder mehrere Nukleotidanaloga enthalten, müssen den Regeln von ST.26 entsprechen
- Nukleotidanaloga sind in der Definition eines "Nukleotids" unter ST.26, Absatz 3(g)(2) enthalten:

"ein Analog von einem 2'-Desoxyribose-5'-monophosphat oder Ribose-5'-Monophosphat, das bei der Bildung des Grundgerüsts eines Nukleinsäureanalog zu einer Anordnung von Nukleobasen führt, die die Anordnung von Nukleobasen in Nukleinsäuren nachahmt, welche ein 2'-Desoxyribose-5'-monophosphat- oder Ribose-5'-monophosphat-Grundgerüst enthalten, wobei der Nukleinsäureanalog zur Basenpaarung mit einer ergänzenden Nukleinsäure in der Lage ist"

- Gängige Nukleotidanaloga sind Peptidnukleinsäuren (PNAs), Glykolnukleinsäuren (GNAs), Threosenukleinsäuren und Morpholinol
- Müssen in der Richtung von links nach rechts, welche die Richtung von 5' zu 3' nachahmt, wiedergegeben werden. (ST.26, Absatz 11)

Neue Molekülararten

Nukleotidanaloga

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

PO_4 -tagttcattgactaaggctccccattgact-OH

wobei das PO_4 -Ende der Sequenz das 5'-Ende einer DNA-Sequenz nachahmt.

- ✓ Diese Sequenz muss in ein Sequenzprotokoll aufgenommen werden

Neue Moleklarten

Nukleotidanaloga

Eine Patentanmeldung offenbart die folgende Sequenz der Glykolnukleinsure (GNA):

PO_4 -tagttcattgactaaggctccccattgact-OH

wobei das PO_4 -Ende der Sequenz das 5'-Ende einer DNA-Sequenz nachahmt.

- ✓ Diese Sequenz muss in ein Sequenzprotokoll aufgenommen werden
- ✓ Das PO_4 -Ende ahmt das 5'-Ende nach, so dass die Sequenz in der angezeigten Richtung wiedergegeben werden muss.

Neue Molekülarten

Nukleotidanaloga

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

PO_4 -tagttcattgactaaggctccccattgact-OH

wobei das PO_4 -Ende der Sequenz das 5'-Ende einer DNA-Sequenz nachahmt.

- ✓ Diese Sequenz muss in ein Sequenzprotokoll aufgenommen werden
- ✓ Das PO_4 -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 Molekülararten

Nukleotidanaloga

```

- <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>

```

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

Neue Molekülarten

D-Aminosäuren

- 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

Eine Patentanmeldung beschreibt die folgende Sequenz:

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

- ✓ Diese Sequenz muss in ein Sequenzprotokoll aufgenommen werden

Neue Molekülarten

D-Aminosäuren

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: AGKLGGM
- ✓ Das Alanin in Position 1, die Glutaminsäure in Position 2 und das Methionin in Position 6 müssen jeweils mit einem Merkmalschlüssel "SITE" und einem Qualifier "NOTE" mit dem vollständigen, nicht abgekürzten Namen der entsprechenden Aminosäure annotiert werden

Neue Molekulararten

D-Aminosäuren

```

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

```

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

Neue Molekülarten

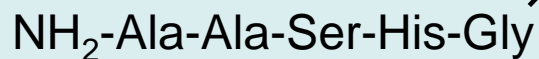
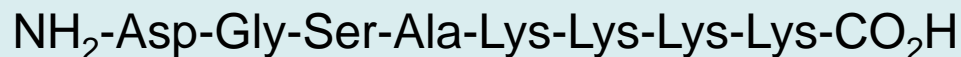
Verzweigte Sequenzen

- 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. (WIPO-Standard 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

Eine Patentanmeldung beschreibt ein Peptid mit der folgenden Sequenz:



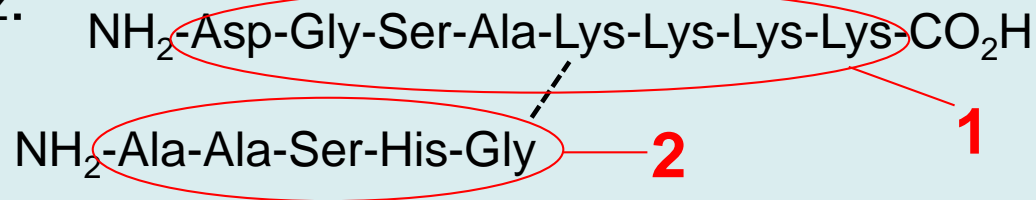
wobei ----- eine Amidbindung zwischen dem C-Terminus des Glycins und der Seitenkette des Lysins angibt

- ✓ Beide linearen Regionen enthalten ≥ 4 spezifisch definierte Aminosäuren, so dass beide in einem Sequenzprotokoll enthalten sein müssen

Neue Molekülarten

Verzweigte Sequenzen

Eine Patentanmeldung beschreibt ein Peptid mit der folgenden Sequenz:



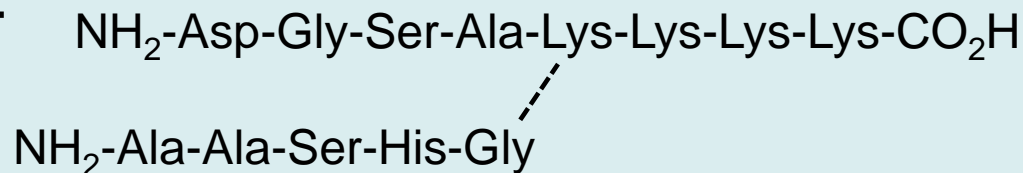
wobei ----- eine Amidbindung zwischen dem C-Terminus des Glycins und der Seitenkette des Lysins angibt

- ✓ 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

Neue Molekülarten

Verzweigte Sequenzen

Eine Patentanmeldung beschreibt ein Peptid mit der folgenden Sequenz:



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 die Art der Amidbindung anzugeben

Neue Molekulararten

Verzweigte Sequenzen

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>DGS AKKKK</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

Sequenzvarianten

- ST.26 Absatz 3(m) definiert eine "**Sequenzvariante**" als

*"eine Nukleotid- oder Aminosäuresequenz, die **einen oder mehrere Unterschiede in Bezug auf eine primäre Sequenz** enthält. Diese Unterschiede können alternative Reste (siehe Absätze 15 und 27), modifizierte Reste (siehe Absätze 3(g), 3(h), 16 und 29), Deletionen, Einfügungen und Substitutionen umfassen. Siehe Absätze 93 bis 95."*

- Die Art und Weise, in der eine Sequenzvariante offenbart wird, bestimmt, wie sie in einem Sequenzprotokoll dargestellt werden muss.
- ST.26, Absätze 93-95 regeln die Darstellung der Varianten.

Sequenzvarianten

Absatz 93

Absatz 93: Eine primäre Sequenz und jede Variante dieser Sequenz, wobei jede durch die Zählung ihrer Reste offenbart wird und in Absatz 7 enthalten ist, muss in das Sequenzprotokoll aufgenommen und ihr eine eigene Sequenzkennzahl zugewiesen werden.



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

Sequenzvarianten

Absatz 93

Eine Patentanmeldung enthält eine Abbildung mit dem folgenden multiplen Sequenzalignment:

Consensus	LEG n EQFINA ak IIRHP k nrk TL n NDI m LIK
Homo sapiens	LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK
Pongo abelii	LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK
Pan paniscus	LEGNEQFINAAKIIRHPKYNRITLNNDIMLIK
Rhinopithecus bieti	LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK
Rhinopithecus roxellana	LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK

Kleinbuchstaben stellen die vorherrschenden Aminosäurereste unter den alignierten Sequenzen dar.

- ✓ Jede der 6 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.

Sequenzvarianten

Korrekte Annotation

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

Type of sequence	Feature Key	Qualifier	Use
Nucleic acid	variation	replace or note	Naturally occurring mutations and polymorphisms, e.g., alleles, RFLPs.
Nucleic acid	misc_difference	replace or note	Variability introduced artificially, e.g., by genetic manipulation or by chemical synthesis.
Amino acid	VAR_SEQ	NOTE	Variant produced by alternative splicing, alternative promoter usage, alternative initiation and ribosomal frameshifting.
Amino acid	VARIANT	NOTE	Any type of variant for which VAR_SEQ is not applicable.

WIPO-Standard ST.26, Absatz 96

Sequenzvarianten

Restriktivstes Mehrdeutigkeitssymbol

- ST.26, Absätze 15 und 27 besagen, dass, wenn ein Mehrdeutigkeitssymbol erforderlich ist, "das restriktivste Symbol zu verwenden ist..."

Was bedeutet das?

Zum Beispiel:

Wenn in einer Nukleotidsequenz eine Position "a or c" sein kann, verwenden Sie das Mehrdeutigkeitssymbol "m" anstelle von "n".

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" für etwas anderes als den Standardwert verwendet wird, eine Annotation erforderlich ist.

Sequenzvarianten

Absatz 94

Absatz 94: Jede Sequenzvariante, die als eine einzelne Sequenz mit gezählten alternativen Resten an einer oder mehreren Positionen offenbart wird, muss in das Sequenzprotokoll aufgenommen werden und ist durch eine einzelne Sequenz darzustellen, wobei die gezählten alternativen Reste durch das restriktivste Mehrdeutigkeitssymbol dargestellt werden (siehe Absätze 15 und 27).



Wenn Varianten NICHT separat gezählt werden, sondern einfach als variable Reste in der primären Sequenz dargestellt werden, dann benötigen diese keine separate SEQ ID-Kennzahl!

Sequenzvarianten

Absatz 94

Eine Patentanmeldung offenbart ein Peptid der Sequenz:

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

- ✓ Kann als einzelne Sequenz in das Sequenzprotokoll aufgenommen werden
- ✓ Die bevorzugte Darstellung der Sequenz ist: GGGJATX
- ✓ [Leu or Ile] ist durch das restriktivste Mehrdeutigkeitssymbol "J" darzustellen
- ✓ [Ser or Thr] ist durch das Symbol "X" zusammen mit dem Merkmalschlüssel "VARIANT" mit einem Qualifier "note", der angibt, dass X Serin oder Threonin ist, darzustellen

Hinweis: Eine ausführliche Erläuterung dieses Beispiels finden Sie im WIPO-Standard ST.26, Anhang VI, Beispiel 94-1.

Sequenzvarianten

Absatz 95

Absatz 95: Jede Sequenzvariante, die nur durch Verweis auf Löschung(en), Einfügung(en) oder Substitution(en) in einer primären Sequenz im Sequenzprotokoll offenbart wird, sollte in das Sequenzprotokoll aufgenommen werden. Wird eine solche Sequenzvariante in das Sequenzprotokoll aufgenommen:

(a) kann sie durch Annotation der primären Sequenz dargestellt werden, wenn sie eine oder mehrere Variationen an einer einzelnen Stelle oder an mehreren verschiedenen Stellen enthält und das Auftreten dieser Variationen unabhängig ist;

(b) sollte sie als eine separate Sequenz dargestellt und ihr eine eigene Sequenzkennzahl zugewiesen werden, wenn sie Variationen an mehreren verschiedenen Stellen enthält und das Auftreten dieser Variationen voneinander abhängig ist; und

(c) muss sie als eine separate Sequenz dargestellt und ihr eine eigene Sequenzkennzahl zugewiesen werden, wenn sie eine eingefügte oder substituierte Sequenz mit mehr als 1000 Resten enthält (siehe Absatz 86).



"Verweis auf Deletion(en), Einfügung(en) oder Substitution(en)" bedeutet, dass die Varianten in Prosa offenbart werden.

Sequenzvarianten

Absatz 95(a)

Eine Anmeldung enthält die folgende Offenbarung:

Peptidfragment 1: Gly-Leu-Pro-Xaa-Arg-Ile-Cys
wobei Xaa eine beliebige Aminosäure ist

* * *

... in einer anderen Ausführungsform ist das Peptidfragment 1:
Gly-Leu-Pro-Xaa-Arg-Ile-Cys, wobei Xaa Val, Thr oder Asp sein kann

* * *

... in einer anderen Ausführungsform ist das Peptidfragment 1:
Gly-Leu-Pro-Xaa-Arg-Ile-Cys, wobei Xaa Val sein kann...

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.

Sequenzvarianten

Absatz 95(a)

- ✓ Die primäre Sequenz muss in das Sequenzprotokoll aufgenommen werden, und Varianten können durch Annotation der primären Sequenz dargestellt werden
- ✓ Die umfassendste Ausführungsform muss die Version sein, die im Sequenzprotokoll enthalten ist – in diesem Beispiel ist es die Version, in der Xaa "eine beliebige Aminosäure" ist
- ✓ 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 folgenden drei 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.

Sequenzvarianten

Absatz 95(b)

Eine Patentanmeldung beschreibt die folgende Konsensussequenz:

aatg n_1 cccacgaatg n_2 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

Sequenzvarianten

Absatz 95(b)

- ✓ Die Konsensussequenz muss in das Sequenzprotokoll aufgenommen werden, und Varianten sind als separate Sequenzen darzustellen
- ✓ Die umfassendste Ausführungsform muss die Version sein, die im Sequenzprotokoll enthalten ist: wobei n_1 und n_2 "a, t, g, oder c sein können"
- ✓ Die Sequenz muss wie folgt dargestellt werden: `aatgncccacgaatgncac`
- ✓ Für "n" ist keine Annotation erforderlich, da "n" bei fehlendem Merkmalschlüssel als "a", "c", "g" oder "t" interpretiert wird (siehe ST.26, Absatz 15)
- ✓ Es ist zwar nicht erforderlich, aber es wird dringend empfohlen, die folgenden vier Varianten als separate Sequenzen in das Sequenzprotokoll aufzunehmen:

<code>aatg<u>a</u>cccacgaatg<u>b</u>cac</code>	(b = t, g, oder c)
<code>aatg<u>t</u>cccacgaatg<u>v</u>cac</code>	(v = a, g, oder c)
<code>aatg<u>g</u>cccacgaatg<u>h</u>cac</code>	(h = t, a, oder c)
<code>aatg<u>c</u>cccacgaatg<u>d</u>cac</code>	(d = t, g, oder a)

Hinweis: Eine ausführliche Erläuterung eines ähnlichen Beispiels finden Sie im WIPO Standard ST.26, Anhang VI, Beispiel 95(b).

Sequenzvarianten

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

Eine Sequenzvariante enthält eine eingefügte oder substituierte Sequenz mit mehr als 1000 Resten

Sequenzvarianten

Absatz 95(c)

- ✓ 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-...

Sequenzvarianten

F: Welcher Absatz trifft auf die folgende Offenbarung zu?

G-L-P-T-R-I-C- [L or I] -A-V- [G or A]

Sequenzvarianten

F: Welcher Absatz trifft auf die folgende Offenbarung zu?

G-L-P-T-R-I-C- [L or I] -A-V- [G or A]

A: Absatz 94:

"Jede Sequenzvariante, die als eine einzelne Sequenz mit gezählten alternativen Resten an einer oder mehreren Positionen offenbart wird, muss in das Sequenzprotokoll aufgenommen werden und ist durch eine einzelne Sequenz darzustellen, wobei die gezählten alternativen Reste durch das restriktivste Mehrdeutigkeitssymbol dargestellt werden."

Sequenzvarianten

F: Welcher Absatz trifft auf die folgende Offenbarung zu?

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".

Sequenzvarianten

F: Welcher Absatz trifft auf die folgende Offenbarung zu?

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".

A: Absatz 93

"Eine primäre Sequenz und jede Variante dieser Sequenz, wobei jede durch die Zählung ihrer Reste offenbart wird und in Absatz 7 enthalten ist, muss in das Sequenzprotokoll aufgenommen und ihr eine eigene Sequenzkennzahl zugewiesen werden."

Sequenzvarianten

F: Welcher Absatz trifft auf die folgende Offenbarung zu?

Eine Abbildung offenbart die folgende Sequenz:

Met-Gly-Ala-Ile-Pro-Asp-Val-Lys-Arg-Ala-Cys-Trp (Sequenz 1)

Die Spezifikation enthält die folgenden Angaben zur Sequenz 1:

... in bestimmten Ausführungsformen ist das Valin in Position 7 der Sequenz 1 durch Alanin ersetzt...

Sequenzvarianten

F: Welcher Absatz trifft auf die folgende Offenbarung zu?

Eine Abbildung offenbart die folgende Sequenz:

Met-Gly-Ala-Ile-Pro-Asp-Val-Lys-Arg-Ala-Cys-Trp (Sequenz 1)

Die Spezifikation enthält die folgenden Angaben zur Sequenz 1:

... in bestimmten Ausführungsformen ist das Valin in Position 7 der Sequenz 1 durch Alanin ersetzt...

A: Absatz 95(a): *"Jede Sequenzvariante, die nur durch Verweis auf Deletion(en), Einfügung(en) oder Substitution(en) in einer primären Sequenz im Sequenzprotokoll offenbart wird, sollte in das Sequenzprotokoll aufgenommen werden. Wird eine solche Sequenzvariante in das Sequenzprotokoll aufgenommen:*

(a) kann sie durch Annotation der primären Sequenz dargestellt werden, wenn sie eine oder mehrere Variationen an einer einzelnen Stelle oder an mehreren verschiedenen Stellen enthält und das Auftreten dieser Variationen unabhängig ist;"

Sequenzvarianten

F: Welcher Absatz trifft auf die folgende Offenbarung zu?

Eine Abbildung offenbart die folgende Sequenz:

Met-Gly-Ala-Ile-Pro-Asp-Val-Lys-Arg-Ala-Cys-Trp (Sequenz 1)

Die Spezifikation enthält die folgenden Angaben zur Sequenz 1:

... wenn das Valin in Position 7 der Sequenz 1 durch Alanin ersetzt wird, dann wird das Alanin in Position 10 durch Valin ersetzt....

Sequenzvarianten

F: Welcher Absatz trifft auf die folgende Offenbarung zu?

Eine Abbildung offenbart die folgende Sequenz:

Met-Gly-Ala-Ile-Pro-Asp-Val-Lys-Arg-Ala-Cys-Trp (Sequenz 1)

Die Spezifikation enthält die folgenden Angaben zur Sequenz 1:

... wenn das Valin in Position 7 der Sequenz 1 durch Alanin ersetzt wird, dann wird das Alanin in Position 10 durch Valin ersetzt....

A: Absatz 95(b) : *Jede Sequenzvariante, die nur durch Verweis auf Deletion(en), Einfügung(en) oder Substitution(en) in einer primären Sequenz im Sequenzprotokoll offenbart wird, sollte in das Sequenzprotokoll aufgenommen werden. Wird eine solche Sequenzvariante in das Sequenzprotokoll aufgenommen:*

(b) sollte sie als eine separate Sequenz dargestellt und ihr eine eigene Sequenzkennzahl zugewiesen werden, wenn sie Variationen an mehreren verschiedenen Positionen enthält und das Auftreten dieser Variationen voneinander abhängig ist;

Fragen?