**ST.26 - ANNEX VII**

RECOMMENDATON FOR THE TRANSFORMATION OF A SEQUENCE LISTING FROM ST.25 TO ST.26:

POTENTIAL ADDED OR DELETED SUBJECT MATTER

Draft proposal presented for consideration at the CWS/6

## Introduction

The requirements for the presentation of nucleotide and amino acid sequences differ between WIPO Standards ST.25 and ST.26. Consequently, the question has been raised as to whether Standard ST.26 would require addition or deletion of any subject matter in a sequence listing submitted as part of an international application under Standard ST.26 that may not be supported by an application from which priority is claimed.

## Scope of the Document

This document addresses the mandatory requirements of ST.26, and any potential consequences of those requirements. This document does not address every possible scenario; if the means of representation in ST.26, of information contained in an ST.25 sequence listing, is not clear, then the information may always be included in the application description to avoid deleted subject matter.

## Recommendations for Potential Added or Deleted Subject Matter

Review of the issues contained in this document demonstrates that conversion from ST.25 to ST.26 by itself should not inherently result in added or deleted subject matter, in particular, where the ST.25 sequence listing was fully compliant with Standard ST.25. However, there are certain scenarios that will require applicant caution. Recommendations have been provided to avoid added or deleted subject matter.

### Scenario 1

ST.25 uses numeric identifiers to tag various types of data, e.g., <110> for Applicant Name. ST.26 uses terms in the English language, as element names and attributes, for data tagging.

#### Recommendation:

The ST.26 terms simply describe the type of data content; therefore, the use of the ST.26 element names and attributes does not constitute added subject matter.

### Scenario 2

ST.26 explicitly requires inclusion of: (a) branched sequences; (b) sequences with D-amino acids; (c) nucleotide analogues; and (d) sequences with abasic sites. Under ST.25, the requirement for inclusion or the prohibition of such sequences is not clear.

#### Recommendation:

The disclosure contained in the application should be sufficient to represent these sequences in an ST.26 sequence listing, when they may not have been included in an ST.25 sequence listing. For certain types of information required by ST.26, care must be taken not to add subject matter beyond that disclosed, e.g., see discussion below (in Scenario 4) on the mol\_type qualifier for nucleotide sequences.

### Scenario 3

ST.26 excludes sequences with less than 10 specifically defined nucleotides (not including “n”) and less than 4 specifically defined amino acids (not including “X”).

#### Recommendation:

The excluded sequences may be included in the application body, where those sequences have not already been included therein.

### Scenario 4

ST.26 has the mandatory feature keys – “source” for all nucleotide sequences and “SOURCE” for all amino acid sequences, each with two mandatory qualifiers. ST.25 has a corresponding feature key for nucleotide sequences (which is rarely used) with no corresponding qualifiers and there is no corresponding feature key for amino acid sequences.

Nucleotide sequences

ST.26 – feature key 5.37 source; mandatory qualifiers 6.44 organism and 6.38 mol\_type (see ST.26 paragraph 75)

|  |  |
| --- | --- |
| **Qualifier** | **Value** |
| mol\_type | genomic DNA |
| genomic RNA |
| mRNA |
| tRNA |
| rRNA |
| other DNA (applies to synthetic molecules) |
| other RNA (applies to synthetic molecules) |
| transcribed RNA |
| viral cRNA |
| unassigned DNA (applies where *in vivo* molecule is unknown) |
| unassigned RNA (applies where *in vivo* molecule is unknown) |

Amino acid sequences

ST.26 – feature key 7.30 SOURCE; mandatory qualifiers 8.3 ORGANISM and 8.1 MOL\_TYPE (*see* ST.26 paragraph 75)

|  |  |
| --- | --- |
| **Qualifier** | **Value** |
| MOL\_TYPE | protein |

#### Recommendation:

The only issue of concern is the controlled vocabulary values associated with the mol\_type qualifier for nucleotide sequences. Some of the value choices listed above may not be sufficiently supported in the disclosure. Added subject matter may be avoided, however, by use of the most generic value for a particular sequence, e.g., “other DNA” and “other RNA” for a synthetic molecule and “unassigned DNA” and “unassigned RNA” for an *in vivo* molecule.

### Scenario 5

Where a sequence includes “Xaa”, ST.25 requires that further information concerning that residue be included in field <223>, which accompanies fields <221> (feature name) and <222> (feature location). ST.25 does not provide a default value for “Xaa” (“X” in ST.26). However, ST.26 does provide such a default value, and therefore, further information is not always required.  Two of the most frequently used annotations in peptide sequences is “any amino acid” or “any naturally occurring amino acid” for variable “Xaa” or “X”. This language could be interpreted to include amino acids other than those listed in the amino acid tables contained in either ST.25 or ST.26. The ST.26 default value for “X” with no further annotation, is any of the 22 individual amino acids listed in Annex I (see Section 3, Table 3).  This ST.26 default value may itself constitute added or deleted subject matter, and therefore, adversely affect the scope of a patent application when transitioning from ST.25 to ST.26.

#### Recommendations:

1. Where the ST.25 sequence listing includes a <221> feature name, <222> feature location corresponding to the Xaa, and <223> further information on Xaa, and the <221> feature name is also an appropriate ST.26 feature key, e.g. SITE, VARIANT, or UNSURE, then the ST.26 feature key should be used. Furthermore, to avoid potential deleted subject matter, the information in field <223> must be included in an accompanying qualifier “NOTE”.
2. Where the ST.25 sequence listing includes a <221> feature name, <222> feature location corresponding to the Xaa, and <223> further information on Xaa, and the <221> feature name is not an ST.26 feature key, then ST.26 feature keys SITE or REGION, as appropriate, should be used. Furthermore, to avoid potential deleted subject matter, the information in field <223>, as well as the inappropriate <221> feature name, must be included in an accompanying qualifier “NOTE”. For example, an ST.25 listing used a feature name that is not in ST.25 or ST.26, <221> Variable, together with further information <223> Xaa is any amino acid. In this example, the value of the ST.26 qualifier NOTE would be “Variable – Xaa is any amino acid”.
3. Where the ST.25 sequence listing provides no <221>, <222>, or <223> field corresponding to the Xaa or where fields <221> and <222> corresponding to the Xaa are included, but no information is included in a corresponding <223> field (neither scenario is compliant with ST.25, but has occurred nonetheless), any information contained in the application body to describe “Xaa” should be included in the ST.26 qualifier “NOTE” together with an appropriate feature key, e.g. SITE, REGION, or UNSURE, and location.

### Scenario 6

In ST.25, uracil is represented in the sequence by “u” and thymine is represented by “t”. In ST.26, uracil and thymine are both represented in the sequence by “t” and without further annotation; “t” represents uracil in RNA and thymine in DNA.

#### Recommendations:

1. Where a DNA sequence contains uracil, ST.26 considers it to be a modified nucleotide, and requires that uracil must be represented as a “t” and be further described using the feature key “modified\_base”, the qualifier “mod\_base” with “OTHER” as the qualifier value and the qualifier “note” with “uracil” as the qualifier value. This ST.26 annotation is not considered added subject matter where the ST.25 DNA sequence contained a “u”.
2. Where an RNA sequence contains thymine, ST.26 considers it to be a modified nucleotide, and requires that thymine must be represented as a “t” and be further described using the feature key “modified\_base”, the qualifier “mod\_base” with “OTHER” as the qualifier value and the qualifier “note” with “thymine” as the qualifier value. This ST.26 annotation is not considered added subject matter where the ST.25 RNA sequence contained a “t”.

### Scenario 7

In both ST.25 and ST.26, modified nucleotides or amino acids must have a further description. In ST.26, the identity of a modified nucleotide may be indicated using an abbreviation from Annex I, Section 2, Table 2, where applicable. Otherwise, the complete unabbreviated name of the modified nucleotide must be indicated. Similarly, the identity of a modified amino acid may be indicated using an abbreviation from Annex I, Section 4, Table 4, where applicable. Otherwise, the complete unabbreviated name of the modified amino acid must be indicated. In contrast, if a modified residue is not contained in an ST.25 table, use of the complete, unabbreviated name is not required, and not infrequently, an abbreviation is used instead.

#### Recommendations:

1. Where only an abbreviated name, which is not in Annex I, Section 2, Table 2 or Section 4, Table 4, was used both in the application and in an ST.25 sequence listing for either a modified nucleotide or a modified amino acid, and the abbreviated name is known in the art to reference only one specific modified nucleotide or modified amino acid, then use of the full, unabbreviated name would not itself constitute added subject matter.
2. Where only an abbreviated name, which is not in Annex I, Section 2, Table 2 or Section 4, Table 4, was used both in the application and in an ST.25 sequence listing for either a modified nucleotide or a modified amino acid (and the application contains no chemical structure), and the abbreviated name is not known in the art to reference one specific modified nucleotide or modified amino acid, i.e., the abbreviation is either not known at all in the art, or could possibly represent multiple different modified nucleotides or modified amino acids, then compliance with ST.26, without introduction of added subject matter, is not possible in this situation. Of course in this case, the priority application and sequence listing are themselves vague. To avoid potential deleted subject matter, the abbreviated name from the ST.25 sequence listing should be placed in an ST.26 “note” or “NOTE” qualifier in addition to the value of the complete unabbreviated name of the modified nucleotide or modified amino acid. The complete unabbreviated name of the modified nucleotide or modified amino acid required in an ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include the unabbreviated name to avoid future issues.

### Scenario 8

ST.25 contains a number of feature keys that are not contained in ST.26. Therefore, applicants must take care to capture the information contained in those ST.25 feature keys in a manner compliant with ST.26 without the introduction of added or deleted subject matter.

#### Recommendations:

The following table provides guidance as to the manner in which the information contained in a former ST.25 feature key may be included in compliance with ST.26 without the introduction of added or deleted subject matter. Numbers 1-23 are feature keys related to nucleotide sequences and numbers 24 – 43 are feature keys related to amino acid sequences.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **ST.25 Feature key <221>** | **ST.26 equivalent** | | |
| **Feature key** | **Qualifier** | **Qualifier value** |
| 1 | allele | misc\_feature | allele | <223> value |
| 2 | attenuator | regulatory[[1]](#footnote-2) | regulatory\_class1 | “attenuator” |
| note *(if <223> present)* | <223> value |
| 3 | CAAT\_signal | regulatory1 | regulatory\_class1 | “CAAT\_signal” |
| note *(if <223> present)* | <223> value |
| 4 | conflict | misc\_feature | note | “conflict” and <223> value |
| 5 | enhancer | regulatory1 | regulatory\_class1 | “enhancer” |
| note *(if <223> present)* | <223> value |
| 6 | GC\_signal | regulatory1 | regulatory\_class1 | “GC\_signal” |
| note *(if <223> present)* | <223> value |
| 7 | LTR | mobile\_element1 | rpt\_type1 | “long\_terminal\_repeat” |
| note *(if <223> present)* | <223> value |
| 8 | misc\_signal | regulatory1 | regulatory\_class1 | “other” |
| note *(if <223> present)* | <223> value |
| 9 | mutation | variation | note | “mutation” and <223> value |
| 10 | old\_sequence | misc\_feature | note | “old\_sequence” and <223> value |
| 11 | polyA\_signal | regulatory1 | regulatory\_class1 | “polyA\_signal\_sequence” |
| note *(if <223> present)* | <223> value |
| 12 | promoter | regulatory1 | regulatory\_class1 | “promoter” |
| note *(if <223> present)* | <223> value |
| 13 | RBS | regulatory1 | regulatory\_class1 | “ribosome\_binding\_site” |
| note *(if <223> present)* | <223> value |
| 14 | repeat\_unit (a) when repeat\_region not used | misc\_feature | note | “repeat\_unit” and <223> value |
| repeat\_unit (b) when repeat\_region used | repeat\_region | rpt\_unit\_range | 1st residue..last residue |
| note *(if <223> present)* | <223> value |
| 15 | satellite | repeat\_region | satellite | “satellite” (or “microsatellite” or  “minisatellite” – if supported) |
| note *(if <223> present)* | <223> value |
| 16 | scRNA | ncRNA1 | ncRNA\_class1 | “scRNA” |
| note *(if <223> present)* | <223> value |
| 17 | snRNA | ncRNA1 | ncRNA\_class1 | “snRNA” |
| note *(if <223> present)* | <223> value |
| 18 | TATA\_signal | regulatory1 | regulatory\_class1 | “TATA\_box” |
| note *(if <223> present)* | <223> value |
| 19 | terminator | regulatory1 | regulatory\_class1 | “terminator” |
| note *(if <223> present)* | <223> value |
| 20 | 3’clip | misc\_feature | note | “3’clip” and <223> value |
| 21 | 5’clip | misc\_feature | note | “5’clip” and <223> value |
| 22 | -10\_signal | regulatory1 | regulatory\_class1 | “minus\_10\_signal” |
| note *(if <223> present)* | <223> value |
| 23 | -35\_signal | regulatory1 | regulatory\_class1 | “minus\_35\_signal” |
| note *(if <223> present)* | <223> value |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 24 | NON\_CONS | This feature relates to a gap of an unknown number of residues in a single sequence, which is prohibited in both ST.25 (paragraph 22) and ST.26 (paragraph 37). Consequently, each region of specifically defined residues that is encompassed by ST.26 paragraph 7 must be included in the sequence listing as a separate sequence and assigned its own sequence identification number. To avoid added/deleted subject matter, each such sequence must be annotated to indicate that it is part of a larger sequence that contains an undefined gap. | | |
| SITE | NOTE | Description |
| Description - as to where and to what the sequence is linked, e.g. this residue is linked N-terminally to a peptide having an N-terminal Gly-Gly and a gap of undefined length. | | |
| 25 | SIMILAR | REGION | NOTE | “SIMILAR” and <223> value if  present |
| 26 | THIOETH | CROSSLNK | NOTE | “THIOETH” and <223> value if present |
| For further location information guidance, see ST.26 Annex I, CROSSLNK Feature Key Comment | | |
| 27 | THIOLEST | CROSSLNK | NOTE | “THIOLEST” and <223> value if  present |
| For further location information guidance, see ST.26 Annex I, CROSSLNK  Feature Key Comment | | |
| 28 | VARSPLIC | Discussed in a Scenario 13 below | | |
| 29 | ACETYLATION | MOD\_RES | NOTE | “ACETYLATION” and <223> value if present |
| NOTE | Information required by ST.26 Annex I MOD\_RES Feature Key Comment, if possible (without added subject matter) |
| 30 | AMIDATION | MOD\_RES | NOTE | “AMIDATION” and <223> value if present |
| NOTE | Information required by ST.26 Annex I MOD\_RES Feature Key Comment, if possible (without added subject matter) |
| 31 | BLOCKED | MOD\_RES | NOTE | “BLOCKED” and <223> value if present |
| NOTE | Information required by ST.26 Annex I MOD\_RES Feature Key Comment, if possible (without added subject matter) |
| 32 | FORMYLATION | MOD\_RES | NOTE | “FORMYLATION” and <223> value if present |
| NOTE | Information required by ST.26 Annex I MOD\_RES Feature Key Comment, if possible (without added subject matter) |
| 33 | GAMMA-  CARBOXYGLUTAMIC  ACID  HYDROXYLATION | MOD\_RES | NOTE | “GAMMA-CARBOXYLGLUTAMIC ACID HYDROXYLATION” and <223> value if present |
| NOTE | Information required by ST.26 Annex I MOD\_RES Feature Key Comment, if possible (without added subject matter) |
| 34 | METHYLATION | MOD\_RES | NOTE | “METHYLATION” and <223> value if present |
| NOTE | Information required by ST.26 Annex I MOD\_RES Feature Key Comment, if possible (without added subject matter) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 35 | PHOSPHORYLATION | MOD\_RES | NOTE | “PHOSPHORYLATION” and <223> value if present |
| NOTE | Information required by ST.26 Annex I MOD\_RES Feature Key Comment, if possible (without added subject matter) |
| 36 | PYRROLIDONE  CARBOXYLIC ACID | MOD\_RES | NOTE | “PYRROLIDONE CARBOXYLIC ACID” and <223> value if present |
| NOTE | Information required by ST.26 Annex I MOD\_RES Feature Key Comment, if possible (without added subject matter) |
| 37 | SULFATATION | MOD\_RES | NOTE | “SULFATATION” and <223> value if present |
| NOTE | Information required by ST.26 Annex I MOD\_RES Feature Key Comment, if possible (without added subject matter) |
| 38 | MYRISTATE | LIPID | NOTE | “MYRISTATE” and <223> value if present |
| NOTE | Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (without added subject matter) |
| 39 | PALMITATE | LIPID | NOTE | “PALMITATE” and <223> value if present |
| NOTE | Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (without added subject matter) |
| 40 | FARNESYL | LIPID | NOTE | “FARNESYL” and <223> value if present |
| NOTE | Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (without added subject matter) |
| 41 | GERANYL-GERANYL | LIPID | NOTE | “GERANYL-GERANYL” and <223> value if present |
| NOTE | Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (without added subject matter) |
| 42 | GPI-ANCHOR | LIPID | NOTE | “GPI-ANCHOR” and <223> value if present |
| NOTE | Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (without added subject matter) |
| 43 | N-ACYL  DIGLYCERIDE | LIPID | NOTE | “N-ACYL DIGLYCERIDE” and <223> value if present |
| NOTE | Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (without added subject matter) |

### Scenario 9

Certain feature keys present in both ST.25 and in ST.26, both for nucleotide sequences and amino acid sequences, have mandatory qualifiers in ST.26, as indicated below. ST.25 did not have any qualifiers, but did have a <223> free text field. When the information contained in an ST.25 <223> field is appropriate as the value for the ST.26 mandatory qualifier, then the information should be included as such. When an ST.25 <223> field has either not been provided or contains information that is not appropriate as the value for the ST.26 mandatory qualifier, then applicants must take care to capture the information contained in the ST.25 feature key/<223> field in a manner compliant with ST.26 without the introduction of added or deleted subject matter.

Nucleotide sequences[[2]](#footnote-3)

|  |  |
| --- | --- |
| **Feature Key** | **Mandatory Qualifier** |
| 5.12 - misc\_binding | 6.3 - bound\_moiety |
| 5.30 - protein\_bind | 6.3 - bound\_moiety |

#### Recommendations:

1. If the ST.25 <223> field is absent or inappropriate, and the application description disclosed the name of the molecule/complex that may bind to the feature location of the nucleic acid, then that name should be included in the qualifier “bound\_moiety”.
2. Any information contained in the ST.25 <223> field that is inappropriate for inclusion in the qualifier “bound\_moiety” should be inserted into an appropriate optional qualifier of the feature key, e.g., “note”.
3. If the ST.25 <223> field is absent or inappropriate, and the application description did not disclose the name of the molecule/complex that may bind to the feature location of the nucleic acid, then the ST.26 feature key “misc\_feature” should be used instead of misc\_binding or protein\_bind, with the qualifier “note”.
4. If the ST.25 <223> field was absent, then the value of the qualifier “note” should be the name of the ST.25 feature key;
5. If the ST.25 <223> field contained inappropriate information, then the value of the qualifier “note” should be the name of the ST.25 feature key and the information from the <223> field.

Amino acid sequences2

|  |  |
| --- | --- |
| **Feature Key** | **Mandatory Qualifier** |
| 7.2 – BINDING | 8.2 – NOTE |
| 7.4 – CARBOHYD | 8.2 – NOTE |
| 7.10 – DISULFID | 8.2 – NOTE |
| 7.11 – DNA\_BIND | 8.2 – NOTE |
| 7.12 – DOMAIN | 8.2 – NOTE |
| 7.16 – LIPID | 8.2 – NOTE |
| 7.17 – METAL | 8.2 – NOTE |
| 7.18 – MOD\_RES | 8.2 – NOTE |
| 7.23 – NP\_BIND | 8.2 – NOTE |
| 7.29 – SITE | 8.2 – NOTE |
| 7.39 – ZN\_FING | 8.2 – NOTE |

#### Recommendations:

1. If the ST.25 <223> field is absent or inappropriate, and the application description disclosed the specific information required in the mandatory qualifier, then that information should be included in the mandatory qualifier “NOTE”.
2. Any information contained in the ST.25 <223> field that is inappropriate for inclusion in the mandatory qualifier “NOTE” (see feature key definition and comment) should be inserted into a second qualifier “NOTE”.
3. If the ST.25 <223> field is absent or inappropriate, and the application description did not disclose the specific information required in the mandatory qualifier, then the ST.26 feature key “SITE” (for one amino acid) or “REGION” (for a range of amino acids) should be used instead, with the qualifier “NOTE”.
4. If the ST.25 <223> field is absent, then the value of the qualifier “NOTE” should be the name of the ST.25 feature key;
5. If the ST.25 <223> field contained inappropriate information, then the value of the qualifier “NOTE” should be the name of the ST.25 feature key and the information from the <223> field.

### Scenario 10

Each specific feature key in ST.25 has a <222> field to indicate a feature location; however, ST.25 does not require an indication of the location for most features and the format of the location information is not standardized. Furthermore,

ST.25 does not have location operators, e.g. “join”. ST.26 has standardized location descriptors and operators and each feature must contain at least one location descriptor. (CDS features are a special case and are discussed below in Scenario 11).

#### Recommendations:

1. If the ST.25 sequence listing had a <222> field, direct importation or importation into ST.26 format should not raise any added subject matter consideration;
2. If the ST.25 sequence listing did not have a <222> field, but location information was contained in the application description, then direct importation or importation into ST.26 format should not raise any added subject matter consideration;
3. If neither the ST.25 sequence listing, nor the application description contained location information, then presumably, the feature applies to the entire sequence. (Indicating a location that is less than the entire sequence without support in the application description would likely constitute added/deleted subject matter.) Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include location information to the extent possible to avoid future issues.

### Scenario 11

In ST.25, a coding sequence that encoded a single, contiguous polypeptide but that was interrupted by one or more non-coding sequence(s), e.g., introns, was indicated as multiple separate CDS features, as illustrated below:

<220>

<221> CDS

<222> (1)..(571)

<220>

<221> CDS

<222> (639)..(859)

In contrast, ST.26 has a join location operator that specifies that the polypeptides encoded by the indicated locations are joined and form a single, contiguous polypeptide. (Note: both ST.25 and ST.26 require that the stop codon be included in the CDS feature location.)

#### Recommendations:

1. If the ST.25 sequence listing or the application description clearly indicated that the polypeptide sequences encoded by the multiple separate CDS features form a single, contiguous polypeptide, then a coding sequence interrupted by an intron in a single CDS feature must be represented with the join location operator, as illustrated below, such that no added subject matter is introduced:

<INSDFeature\_key>CDS</INSDFeature\_key>

<INSDFeature\_location>join(1..571,639..859)</INSDFeature\_location>

1. If the ST.25 sequence listing or the application description did not indicate that the polypeptide sequences encoded by the two separate CDS features form a single, contiguous polypeptide, then use of the join location operator would likely constitute added subject matter.

### Scenario 12

ST.25 specifies that feature names must be one from Table 5 or 6. However, U.S. regulations indicated that these feature names were recommended, but not required. Therefore, a sequence in an ST.25 sequence listing (compliant with U.S. regulations) might have a “custom” feature key name with no corresponding feature key in ST.26. It is also possible that no feature name was provided for the <221> field or the <221> field is absent. These scenarios may be handled in a similar manner.

#### Recommendation:

The “custom” feature key name from ST.25 may be represented in an ST.26 sequence listing with no added subject matter as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | ST.25 Feature Key <221> | Potential ST.26 Equivalent | | |
| Feature key | Qualifier | Qualifier value |
| NA | “Custom” feature key | misc\_feature | note | “custom” feature key  name and <223> value if present |
| AA | “Custom” feature key | SITE or REGION | NOTE | “custom” feature key  name and <223> value  if present |

### Scenario 13

ST.25 contains a feature key “VARSPLIC” defined as “description of sequence variants produced by alternative splicing”. In ST.26, “VARSPLIC” has been replaced with the broader feature key VAR\_SEQ defined as “description of sequence variants produced by alternative splicing, alternative promoter usage, alternative initiation and ribosomal frameshifting”. Therefore, the ST.26 sequence listing should not use “VAR\_SEQ” as a replacement of “VARSPLIC” without a further explanation.

#### Recommendation:

In ST.26 the feature “VAR\_SEQ” should be used with the qualifier “NOTE”, whose value should include an explanation of the ST.25 narrower scope, e.g., “sequence variant produced by alternative splicing”. Any additional information contained in an accompanying ST.25 <223> field should also be included in the qualifier “NOTE”.

### Scenario 14

If the source of a sequence was artificial, the ST.25 <213> Organism field requires the phrase “Artificial Sequence”. In ST.26, the feature key “source” or “SOURCE” requires the qualifier “organism” or “ORGANISM”, whose value must be indicated as “synthetic construct”, rather than “Artificial Sequence”.

#### Recommendation:

The value for the ST.26 qualifier “organism” or “ORGANISM” must be indicated as “synthetic construct”. To avoid potential deleted subject matter, any explanatory information contained in the required ST.25 <223> field should be included in a qualifier “note” or “NOTE” (of the feature key “source” or “SOURCE”).

### Scenario 15

If the scientific name of the source organism of a sequence is unknown, the ST.25 <213> Organism field requires the term “Unknown”. In ST.26, the feature key “source” or “SOURCE” requires the qualifier “organism” or “ORGANISM”, whose value must be indicated as “unidentified”, rather than “Unknown”.

#### Recommendation:

The value for the ST.26 qualifier “organism” or “ORGANISM” must be indicated as “unidentified”. To avoid potential deleted subject matter, any explanatory information contained in the required ST.25 <223> field should be included in a qualifier “note” or “NOTE” (of the feature key “source” or “SOURCE”).

### Scenario 16

ST.25 allows for the enumeration of amino acids to optionally include negative numbers, counting backwards starting with the amino acid next to number 1, for the amino acids preceding the mature protein, for example pre-sequences, pro-sequences, pre-pro-sequences and signal sequences. ST.26 does not allow for negative numbers in the feature location.

#### Recommendations:

1. If the ST.25 sequence listing had a feature or features represented in a <221> and an accompanying <222> field which contained negative and/or positive numbering, e.g. “PROPEP” and/or “CHAIN”, then in the ST.26 sequence listing, the appropriate feature key, e.g., “PROPEP” and/or “CHAIN”, should be used. A qualifier “NOTE” may be used with the information in a <223> field, if any, as the qualifier value;
2. If the ST.25 sequence listing did not have a feature or features represented in a <221> and accompanying <222> field, but information was contained in the application description regarding the negative and/or positive numbering, then in the ST.26 sequence listing, the appropriate feature key, e.g., “PROPEP” and/or “CHAIN”, should be used. Otherwise, the feature key “REGION” may be used. A qualifier “NOTE” may be used with information in the application description, if any, as the qualifier value;
3. If neither the ST.25 sequence listing, nor the application description, contains information explaining the negative and/or positive numbering, then to avoid potential deleted subject matter in the ST.26 sequence listing, the “REGION” feature key should be used, where the feature location spans the negatively numbered region of the ST.25 sequence. Also, a qualifier “NOTE” should be used to indicate that the amino acid sequence was negatively numbered in the ST.25 sequence listing of the application to which priority is claimed.

### Scenario 17

ST.25 provides for publication information in fields <300> to <313>. ST.26 does not provide for inclusion of such information.

#### Recommendation:

The information contained in ST.25 fields <300> to <313> should be inserted into the accompanying application body, if not already contained therein.

### Scenario 18

ST.25 does not provide a standardized way to indicate that a CDS region of a nucleotide sequence was to be translated using a genetic code table other than the standard genetic code table. In contrast, ST.26 has a “transl\_table” qualifier that can be used with the “CDS” feature key to indicate that the region is to be translated using an alternative genetic code table. If the “transl\_table” qualifier is not used, the use of the standard genetic code table is assumed.

#### Recommendations:

1. If the ST.25 sequence listing or the application description clearly indicated that a CDS region is to be translated using an alternative genetic code table, then the “transl\_table” qualifier must be used with the appropriate genetic code table number as the qualifier value. Failure to use the “transl\_table” qualifier would likely constitute added subject matter, as the default “Standard Code” table would be assumed. Failure to include, in the ST.26 sequence listing, the alternative genetic code table information from the ST.25 sequence listing or from the application description would likely constitute deleted subject matter.
2. If the ST.25 sequence listing or the application description did not indicate that a CDS region is to be translated using an alternative genetic code table, then the “transl\_table” qualifier should not be used, or should be used only with the qualifier value “1,” i.e., the Standard Code table. Use of the “transl\_table” qualifier with any qualifier value other than “1” would likely constitute added and deleted subject matter.

### Scenario 19

ST.25 does not provide a standardized way to indicate the location of a feature, in particular, one contained in a site or region that extends beyond a specified residue or span of residues, e.g., a CDS region of a nucleotide sequence that extends beyond one or both ends of a disclosed sequence. In contrast, the ST.26 feature location descriptor provides a standardized way to indicate the location of such a site or region by using the “<“ or “>“ symbols. For example, the “CDS” feature location must include the stop codon, even when the stop codon is not included in the disclosed sequence itself, by indicating the location as e.g., 1..>321.

#### Recommendations:

1. Where the ST.25 sequence listing did not explicitly indicate that the location of a feature extended beyond the sequence, but such a location is either supported by the disclosure or is clear from the sequence itself, e.g., the stop codon of a CDS feature that is not contained in the sequence, then the “<“ or “>“ symbols may be used in the ST.26 sequence listing without addition of subject matter.
2. Where the ST.25 sequence listing did not explicitly indicate that the location of a feature extended beyond the sequence, and such a location is neither supported by the disclosure, nor is clear from the sequence itself, then compliance with ST.26, without introduction of added subject matter, may not be possible in this situation. In this case, the priority application and sequence listing are themselves arguably incomplete. In this situation, the location description of the feature in the ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include complete feature information.

### Scenario 20

ST.25 Appendix I requires that where a nucleotide sequence contains both DNA and RNA fragments, the value in <212> shall be “DNA” and the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section; however, the exact nature of the further description is not clear and this requirement is not routinely followed. ST.26, paragraph 55, requires that each DNA and RNA segment (ST.26 uses “segment” rather than “fragment” for internal consistency) of the combined DNA/RNA molecule must be further described with the feature key “misc\_feature”, which includes the location of the segment, and the qualifier “note”, which indicates whether the segment is DNA or RNA.

#### Recommendations:

1. If the ST.25 sequence listing described the DNA and RNA segments in one or more features using <221> misc\_feature, appropriate locations in <222>, and indications in <223> as to which segments were DNA or RNA, then incorporating that information into ST.26 format, using a misc\_feature for each DNA and RNA segment, should not raise any added subject matter consideration;
2. If the ST.25 sequence listing described the DNA and RNA segments in one or more features using a feature key in <221> other than misc\_feature, appropriate locations in <222>, and indications in <223> identifying which segments are DNA or RNA, then incorporating that information into ST.26 format, using a misc\_feature for each DNA and RNA segment and an additional “note” qualifier with the original <221> feature key as the value, should not raise any added or deleted subject matter consideration;
3. If the ST.25 sequence listing provides the identity (DNA or RNA) and location of each segment in a <223> field that is not associated with a <221> and <222> field, e.g. the explanation for an Artificial Sequence, then incorporating that information into ST.26 format using a misc\_feature for each DNA and RNA segment, should not raise any added subject matter consideration;
4. If the ST.25 sequence listing described the molecule in a feature using a <221> misc\_feature and a <223> noting that the molecule is a combined DNA/RNA molecule, but did not provide location information for each segment, and
   * 1. If the description provided the locations of each DNA and RNA segment, then incorporating that information into ST.26 format using a misc\_feature for each DNA and RNA segment, should not raise any added subject matter consideration;
     2. If the description does not contain the location information of each DNA and RNA segment, then compliance with ST.26, without introduction of added subject matter, may not be possible in this situation. In this case, the priority application and sequence listing are themselves arguably incomplete. In this situation, any location descriptions of the features in the ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include complete feature information.
5. If the ST.25 sequence listing described the molecule in a feature using a feature key in <221> other than misc\_feature and a <223> noting that the molecule is a combined DNA/RNA molecule, but did not provide location information for each segment, and
6. If the description provided the locations of each DNA and RNA segment, then incorporating that information into ST.26 format using a misc\_feature for each DNA and RNA segment and an additional “note” qualifier with the original <221> feature key as the value, should not raise any added or deleted subject matter consideration;
7. If the description does not contain the location information of each DNA and RNA segment, then compliance with ST.26, without introduction of added subject matter, may not be possible in this situation. In this case, the priority application and sequence listing are themselves arguably incomplete. In this situation, any location descriptions of the features in the ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include complete feature information.
8. If the ST.25 sequence listing noted that the molecule is a combined DNA/RNA molecule in a <223> field, e.g. the explanation for an Artificial Sequence, but did not provide any feature key or location information of each segment, and
9. If the description provided the locations of each DNA and RNA segment, then incorporating that information into ST.26 format using a misc\_feature for each DNA and RNA segment, should not raise any added subject matter consideration;
10. If the description does not contain the location information of each DNA and RNA segment, then compliance with ST.26, without introduction of added subject matter, may not be possible in this situation. In this case, the priority application and sequence listing are themselves arguably incomplete. In this situation, any location descriptions of the features in the ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include complete feature information.

[End of Annex VII and of document]

1. ST.26 may require that a specific ST.25 feature, e.g., TATA\_signal, be replaced by a broader feature key/qualifier/value, e.g., regulatory/regulatory\_class/TATA\_box. In such a case, the narrower ST.25 feature will be afforded priority to the earlier application. However, the full breadth of the broader ST.26 feature key/qualifier, e.g., regulatory/regulatory\_class, will not be afforded priority to the earlier application. [↑](#footnote-ref-2)
2. The numeric references in the table below refer to the Feature key and Qualifier numbers of ST.26, Annex I Controlled Vocabulary. [↑](#footnote-ref-3)