C

CWS/5/7 REV.1 ADD.
原文：英文
日期：2017年5月9日

世界知识产权组织标准委员会（CWS）

第五届会议
2017年5月29日至6月2日，日内瓦

关于WIPO标准ST.25向ST.26的过渡规定的建议

秘书处编拟的文件

1. 文件CWS/5/7 Rev.1载有关于从WIPO标准ST.25向ST.26的过渡规定的建议。关于是否应参考国际申请的国际申请日或优先权日确定过渡日期，还是应由申请人作出选择的问题，SEQL工作队暂时决定国际申请日最合适，具体还有待进一步分析，以考虑为转换ST.25序列表而添加或删减事项可能产生的问题，而且要看是否有可用的编著和验证工具能帮助将序列表从ST.25转至ST.26，而无须作添加和删减(见文件CWS/5/7 Rev.1附件第5段至第6段)。

2. SEQL工作队就添加或删减事项的待决问题作了进一步讨论，最后将本文件附件中所转录的讨论结果提交WIPO标准委员会（CWS）第五届会议审议。经过讨论，工作队同意，国际申请日是最合适的过渡参考日期。此外应注意，国际局计划根据附件中所述的结论和建议，在ST.26编著和验证工具中增加功能，帮助申请人把序列表数据从ST.25转至ST.26，而无须作添加和删减。

3. 请CWS注意本文件及其附件的内容，并审议工作队的一致意见，以决定从WIPO标准ST.25向ST.26过渡的参考日期。

[后接附件]

TRANSFORMATION OF A SEQUENCE LISTING FROM ST.25 TO ST.26 AND POTENTIAL ADDED OR DELETED MATTER

Prepared by the United States Patent and Trademark Office (USPTO)

INTRODUCTION

1. In response to joint Circular C. PCT 1485/C. CWS 75 and the request for public comments on Standard ST.26 issued by the USPTO, EPO, and JPO, applicants expressed concern as to whether Standard ST.26 would require addition or deletion of any information in a sequence listing submitted as part of an international application that may not be supported by an earlier ST.25 sequence listing in an application to which priority is claimed. This document addresses the mandatory requirements of ST.26, and any potential consequences of those requirements.

2. This document does not address every 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 matter.

3. 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 matter, in particular, where the ST.25 sequence listing was fully compliant. However, there are certain scenarios that will require applicant caution. Conclusions and recommendations have been provided in italics following each respective scenario to avoid added or deleted matter.

POSSIBLE SCENARIOS AND CORRESPONDING RECOMMENDATIONS

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

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

5. 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 was not clear.

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 information beyond that disclosed, e.g., see discussion below (in item 4) on the mol_type qualifier for nucleotide sequences.

6. ST.26 excludes - sequences <10 nucleotides and <4 AAs.

The excluded sequences may be included in the application description where necessary.

7. ST.26 has the MANDATORY feature keys – “source” for nucleotide sequences and “SOURCE” for 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

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

Qualifier	**Value**
MOL_TYPE -	protein

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 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.*

8. ST.25 did not provide a default value for “Xaa” (“X” in ST.26); however, ST.26 does provide for such a default value. 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 of ST.26 (see Section 3, Table 3). This ST.26 default value may itself constitute added or deleted matter, and therefore, adversely affect the scope of a patent application when transitioning from ST.25 to ST.26.

(a) In ST.25, field <223> is required to provide further information regarding “Xaa.” To avoid potential deleted matter, the information in field <223> should be included in the qualifier “NOTE” together with the feature key “VAR_SEQ” or “VARIANT”, as appropriate.

(b) Where no <223> field was provided or no information was included in a <223> field (neither is compliant with ST.25, but has been used nonetheless), any information contained in the application body to describe "Xaa" should be included in the qualifier "NOTE" together with the feature key "VAR_SEQ" or "VARIANT", as appropriate.

9. 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.

(a) Applicant must be careful where a DNA sequence contains uracil, since in ST.26 that uracil must be represented as a "t" with a further annotation to indicate that the "t" is actually uracil.

(b) Applicant must be careful where an RNA sequence contains thymine, since in ST.26 that thymine must be represented as a "t" with a further annotation to indicate that the "t" in RNA is actually thymine.

10. 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 was not contained in an ST.25 table, use of the complete, unabbreviated name was not required, and not infrequently, an abbreviation would be used instead.

(a) Where only an abbreviated name 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 the Task Force agrees that use of the full, unabbreviated name would not itself constitute added matter.

(b) Where only an abbreviated name 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 matter, is not possible in this situation. Of course in this case, the priority application and sequence listing are themselves vague. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include the unabbreviated name to avoid future issues. To avoid potential deleted matter, the abbreviated name from the ST.25 sequence listing should be placed in an ST.26 "note" or "NOTE" qualifier. 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.*

11. ST.25 contained 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 matter.

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 matter.

No.	*Feature keys in ST.25 that are absent in ST.26*	*Potentially equivalent Feature key/Qualifier/value in ST.26*
1	*<221> allele/<223> value*	*misc_feature/allele/value from <223>*
2	*<221> attenuator/<223> value*	*regulatory*/regulatory_class*/attenuator* → *if needed: /note/value from <223>*
3	*<221> CAAT_signal/<223> value*	*regulatory*/regulatory_class*/CAAT_signal* → *if needed: /note/value from <223>*
4	*<221> conflict/<223> value*	*misc_feature/note/conflict with <223> value*
5	*<221> enhancer/<223> value*	*regulatory*/regulatory_class*/enhancer* → *if needed: /note/value from <223>*
6	*<221> GC_signal/<223> value*	*regulatory*/regulatory_class*/GC_signal* → *if needed: /note/value from <223>*
7	*<221> LTR/<223> value*	*mobile_element*/rpt_type*/long_terminal_repeat* → *if needed: /note/value from <223>*
8	*<221> misc_signal/<223> value*	*regulatory*/regulatory_class*/other* → *if needed: /note/value from <223>*
9	*<221> mutation/<223> value*	*variation/note/mutation with <223> value*
10	*<221> old_sequence/<223> value*	*misc_feature/note/old_sequence with <223> value*
11	*<221> polyA_signal/<223> value*	*regulatory*/regulatory_class*/polyA_signal_sequence* → *if needed: /note/value from <223>*
12	*<221> promoter/<223> value*	*regulatory*/regulatory_class*/promoter* → *if needed: /note/value from <223>*

13	*<221> RBS/<223> value*	*regulatory*/regulatory_class*/ribosome_binding_site* → *if needed: /note/value from <223>*
14	*<221> repeat_unit/<223> value*	*a. misc_feature/note/repeat_unit with <223> value (when repeat_region not used in ST.25)* *b. repeat_region/rpt_unit_range/1st residue..last residue (when repeat_region used in ST.25)* → *if needed: /note/value from <223>*
15	*<221> satellite/<223> value*	*repeat_region/satellite/satellite (or microsatellite or minisatellite – if supported)* → *if needed: /note/value from <223>*
16	*<221> scRNA/<223> value*	*ncRNA*/ncRNA_class*/scRNA* → *if needed: /note/value from <223>*
17	*<221> snRNA/<223> value*	*ncRNA*/ncRNA_class*/snRNA* → *if needed: /note/value from <223>*
18	*<221> TATA_signal/<223> value*	*regulatory*/regulatory_class*/TATA_box* → *if needed: /note/value from <223>*
19	*<221> terminator/<223> value*	*regulatory*/regulatory_class*/terminator* → *if needed: /note/value from <223>*
20	*<221> 3'clip/<223> value*	*misc_feature/note/3'clip with <223> value*
21	*<221> 5'clip/<223> value*	*misc_feature/note/5'clip with <223> value*
22	*<221> -10_signal/<223> value*	*regulatory*/regulatory_class*/minus_10_signal* → *if needed: /note/value from <223>*
23	*<221> -35_signal/<223> value*	*regulatory*/regulatory_class*/minus_35_signal* → *if needed: /note/value from <223>*

**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.*

12. 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, the information should be included as the value for the ST.26 mandatory qualifier. When an ST.25 <223> field has either not been provided or contains information that is not appropriate for inclusion in 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 matter.

Nucleotide sequences

Feature Key	**Mandatory Qualifier**
5.12 - misc_binding	6.3 - bound_moiety
5.30 – protein_bind	6.3 – bound_moiety

(a) If the ST.25 <223> field is absent or inappropriate, and the application description discloses the name of the molecule/complex that may bind to the featured location of the nucleic acid, then that name should be included in the qualifier "bound_moiety".

i. Any information contained 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".

(b) If the ST.25 <223> field is absent or inappropriate, and the application description does not disclose the name of the molecule/complex that may bind to the featured location of the nucleotide, then the ST.26 feature key "misc_feature" should be used instead, with the qualifier "note".

i. 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; and

ii. If the ST.25 <223> field contains 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 sequences

Feature Key	**Mandatory Qualifier**
7.2 – BINDING	8.2 – NOTE
7.4 – CARBOHYD	8.2 – NOTE
7.9 – CROSSLINK	8.2 – NOTE
7.11 – DNA_BIND	8.2 – NOTE
7.12 – DOMAIN	8.2 – NOTE
7.16 – LIPID	8.2 – NOTE
7.23 – NP_BIND	8.2 – NOTE
7.39 – ZN_FING	8.2 – NOTE

(a) If the ST.25 <223> field is absent or inappropriate, and the application description discloses the specific information required in the mandatory qualifier, then that information should be included in the mandatory qualifier "NOTE".

i. 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".

(b) If the ST.25 <223> field is absent or inappropriate, and the application description does 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".

i. 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; and

ii. If the ST.25 <223> field contains 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.

13. Each specific feature key in ST.25 had a <222> field to indicate a feature location; however, ST.25 did not require an indication of the location for most features and the format of the location information was not standardized. Furthermore, ST.25 did 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 item 14)).

(a) If the ST.25 sequence listing had a <222> field, direct importation or importation into ST.26 format should not raise an added matter consideration;

(b) 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 an added matter consideration; and

(c) If neither the ST.25 sequence listing, nor the application description contained location information, then presumably, the feature applies to the entire sequence. 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. (Indicating a location that is less than the entire sequence without support in the application description would likely constitute added/deleted matter.)*

*The Task Force agrees with this presumption and that an indication of the location as the entire sequence does not constitute added matter.

14. 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(d) the stop codon be included in the CDS feature location.)

(a) If the ST.25 sequence listing or the application description clearly indicates 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 matter is introduced:

<INSDFeature_key>CDS</INSDFeature_key>

<INSDFeature_location>join(1..571,639..859)</INSDFeature_location>

(b) If the ST.25 sequence listing or the application description does 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 matter.

15. ST.25 specified 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, an ST.25 sequence listing (compliant with U.S. regulations) might have a “custom feature”* with no corresponding feature key in ST.26.

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

Type	*“Custom Feature” in ST.25*	*Potentially equivalent Feature key/Qualifier/value in ST.26*
NA	*<221> custom feature/<223> value*	*misc_feature/note/custom feature with <223> value*
AA	*<221>custom feature/<223>value*	*SITE or REGION/NOTE/custom feature with <223> value*

**It is also possible that no feature name is provided for the <221> field or the <221> field is absent. These situations may be handled in a similar manner.*

16. ST.25 contained 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.

*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”.*

17. If the source of a sequence was artificial, the ST.25 <213> Organism field required 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”.

The value for the ST.26 qualifier “organism” or “ORGANISM” must be indicated as “synthetic construct”. To avoid potential deleted 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”).

18. If the scientific name of the source organism of a sequence was unknown, the ST.25 <213> Organism field required 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”.

The value for the ST.26 qualifier “organism” or “ORGANISM” must be indicated as “unidentified”. To avoid potential deleted 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”).

19. ST.25 allowed 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.

(a) 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;

(b) 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; and

(c) If neither the ST.25 sequence listing, nor the application description, contained information explaining the negative and/or positive numbering, then to avoid potential deleted 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.

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

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

21. ST.25 did 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.

(a) If the ST.25 sequence listing or the application description clearly indicates 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 matter as the default "Standard Code" table would be assumed; and

(b) If the ST.25 sequence listing or the application description does 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 matter.

[附件和文件完]