

■ Norme ST.26 de l'OMPI PERFECTIONNEMENT

Webinaire de formation

Note : ce module est un module de suivi et présuppose que les participants ont déjà assisté au webinaire sur les Principes fondamentaux de la norme ST.26.

*Le cours sur les Principes fondamentaux de la norme ST.26 est disponible ici :
https://www.wipo.int/meetings/fr/details.jsp?meeting_id=62848*

Points qui seront abordés aujourd'hui

- Clés de caractérisation et qualificateurs fréquemment utilisés
- Formats de l'emplacement de la caractéristique
- Formats des valeurs qualificatives et valeurs qualificatives dans une langue autre que l'anglais
- Cas particuliers : uracile dans de l'ADN ou thymine dans de l'ARN; molécules hybrides d'ADN/ARN
- Analogues nucléotidiques, acides aminés D et séquences ramifiées
- Variantes de séquence

Clés de caractérisation et qualificateurs fréquemment utilisés

Clés de caractérisation et qualificateurs

- Les clés de caractérisation peuvent être utilisées pour décrire plus précisément un ou plusieurs résidus de séquence identifiés au moyen de leur emplacement
 - Les clés de caractérisation pour les séquences de nucléotides sont répertoriées en annexe I, section 5
 - Les clés de caractérisation pour les séquences de nucléotides sont en minuscules; par exemple, “misc_binding”
 - Les clés de caractérisation pour les séquences d’acides aminés sont répertoriées en annexe I, section 7
 - Les clés de caractérisation pour les séquences d’acides aminés sont en majuscules; par exemple, “REGION”

- Des qualificateurs peuvent être utilisés pour décrire plus précisément des caractéristiques
 - Les qualificateurs pour les séquences de nucléotides sont répertoriés en annexe I, section 6
 - Les qualificateurs pour les séquences de nucléotides sont en minuscules; par exemple, “allele”
 - Les qualificateurs pour les séquences d’acides aminés sont répertoriés en annexe I, section 8
 - Les qualificateurs pour les séquences d’acides aminés sont en majuscules; par exemple, “NOTE”

Clés de caractérisation et qualificateurs

```

<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>

```

Clés de caractérisation et qualificateurs

- Les clés de caractérisation sont facultatives, sauf pour la caractéristique “source”/“SOURCE”, qui est obligatoire pour toutes les séquences.
- Chaque clé de caractérisation sera assortie d’une liste de qualificateurs qui peuvent être utilisés pour préciser la description de la caractéristique. La plupart des qualificateurs sont facultatifs; toutefois, certaines clés de caractérisation sont assorties de qualificateurs obligatoires.

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

- Les qualificateurs “mol_type”/“MOL_TYPE” et “organism”/“ORGANISM” sont obligatoires pour la caractéristique “source”/“SOURCE”.

Clés de caractérisation et qualificateurs

- Certaines clés de caractérisation comportent des limitations supplémentaires
 - Pour organisme scope, par exemple, “C_region” se limite à eukaryotes
 - Pour molécule scope, par exemple, “D-loop” se limite aux séquences ADN

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
Molécule scope	DNA

Clés de caractérisation et qualificateurs

Séquences de nucléotides Clé de caractérisation “modified_base”

- La clé de caractérisation “modified_base” et son qualificateur obligatoire “mod_base” doivent être utilisés pour décrire un nucléotide modifié (norme ST.26, paragraphe 16)
- Un “nucléotide modifié” désigne tout nucléotide autre que :
 - deoxy-[a, g, c, or t] 3'-monophosphate
 - [a, g, c, or u] 3'-monophosphate (norme ST.26, paragraphe 3.f)
- Un “nucléotide modifié” devrait être représenté par le nucléotide non modifié correspondant (tableau 1 de l'annexe I, section 1) chaque fois que possible. Autrement, il peut être représenté par “n”. Par exemple, “2'-O-methylcytidine” devrait être représenté par “c” dans la séquence. “Queuosine” devrait être représenté par “n”. Le symbole “n” est l'équivalent d'un seul résidu.
- La valeur du qualificateur obligatoire “mod_base” doit être choisie parmi les valeurs indiquées dans le tableau 2 de l'annexe I, section 2. Si la valeur est “other”, alors un qualificateur supplémentaire “note” doit contenir le nom complet non abrégé du résidu modifié.

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “modified_base”

- Exemple : séquence de nucléotides avec inosine à la 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” est indiqué dans le tableau 2 de l’annexe I, section 2, comme ayant pour abréviation “i”

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “modified_base”

- Exemple : séquence de nucléotides avec xanthine à la 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>

```

- “Xanthine” NE FIGURE PAS dans le tableau 2 de l’annexe I, section 2; aussi la valeur du qualificateur obligatoire “mod_base” doit-elle être “OTHER” et le qualificateur supplémentaire “note” doit apparaître avec pour valeur “xanthine”.

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation

“modified_base”

- “modified_base” peut également être utilisée pour décrire un site abasique :

```

<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>

```

- Le site abasique peut être représenté par un “n” dans la séquence et décrit plus précisément au moyen d’une clé de caractérisation “modified_base” assortie d’un qualificateur “mod_base” ayant pour valeur “OTHER” et d’un qualificateur supplémentaire de type “note” dont la valeur est “site abasique”.

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- La clé de caractérisation “CDS” peut servir à désigner des séquences de codage. L’emplacement de la caractéristique CDS doit comprendre le codon d’arrêt. (Norme ST.26, paragraphe 89)
- Il N’Y A PAS de qualificateurs obligatoires pour la clé de caractérisation “CDS”
- Les qualificateurs fréquemment utilisés sont :

“pseudo”

“pseudogene”

“translation”

“transl_table”

“codon_start”

“transl_except”

“protein_id”

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Une caractéristique “CDS” peut comporter UN seul des qualificateurs suivants :

Qualificateur	Description	Valeur
pseudo	indique que la caractéristique CDS n'est pas fonctionnelle et n'a pas de traduction, mais n'est pas un pseudogène	aucune
pseudogene	indique que la caractéristique CDS est un pseudogène et n'a pas de traduction	processed (traité) unprocessed (non traité) unitary (unitaire) allelic (allélique) unknown (inconnu)
translation	indique que la séquence d'acide aminé provient de la traduction de la CDS	abréviations des acides aminés par une seule lettre

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Une séquence d’acides aminés contenant au moins quatre acides aminés définis de manière spécifique qui est codée selon une séquence de codage et divulguée dans un qualificateur de type “translation” doit figurer dans le listage des séquences et doit disposer de son propre numéro d’identification de séquence.
- Le numéro d’identification de séquence attribué à la séquence d’acides aminés doit figurer dans la valeur du qualificateur “protein_id” associé à la clé de caractérisation “CDS”.
- Le qualificateur “ORGANISM” associé à la clé de caractérisation “SOURCE” de la séquence d’acides aminés doit être identique à celui de sa séquence de codage.
(Norme ST.26, paragraphe 92)

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Qualificateurs pouvant modifier la séquence traduite :

Qualificateur	Description	Valeur
transl_table	indique le tableau du code génétique utilisé pour traduire la caractéristique CDS; par défaut, il s'agit du “1-Code normalisé”	un nombre qui correspond à un tableau de traduction indiqué en annexe I, section 9
transl_except	indique la traduction d'un codon qui n'est pas conforme au code génétique défini dans “transl_table”	(pos:<location>, aa:<amino_acid>)
codon_start	indique le cadre de lecture de la caractéristique CDS relative à la première base	1, 2, ou 3

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Exemple : SEQ ID NO:1 est un fragment d’une séquence de codage provenant d’un gène mitochondrial de levure (tableau du code génétique “3-Code mitochondrial des levures”) :

```
...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 *
```

Le codon aux positions 30 à 32
code la sélénocystéine (Sec)

La séquence commence par
un codon partiel

- Quelles informations devraient figurer dans une caractéristique CDS pour donner une représentation précise de cette séquence?

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Exemple : SEQ ID NO:1 est un fragment d’une séquence de codage provenant d’un gène mitochondrial de levure (tableau du code génétique “3-Code mitochondrial des levures”) :

```
...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 *
```

- Séquence représentée dans le listage des séquences sous la forme suivante :

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

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Exemple : SEQ ID NO:1 est un fragment d'une séquence de codage provenant d'un gène mitochondrial de levure (tableau du code génétique “3-Code mitochondrial des levures”) :

```
...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  *
```

- Clé de caractérisation “CDS”
- Emplacement de la caractéristique : <1.. 80

Le signe “<” indique que la région de codage commence avant la position 1

L'emplacement comprend le codon d'arrêt aux positions 78 à 80

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Exemple : SEQ ID NO:1 est un fragment d’une séquence de codage provenant d’un gène mitochondrial de levure (tableau du code génétique “3-Code mitochondrial des levures”) :

```
...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 *
```

- clé de caractérisation “CDS”
- Emplacement de la caractéristique : <1.. 80
- Qualificateur “codon_start” dont la valeur est “3”

La valeur 3 du qualificateur “codon_start” indique que le premier codon complet commence à la troisième position dans l’emplacement.

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Exemple : SEQ ID NO:1 est un fragment d’une séquence de codage provenant d’un gène mitochondrial de levure (tableau du code génétique “3-Code mitochondrial des levures”) :

```
...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 *
```

- Clé de caractérisation “CDS”
- Emplacement de la caractéristique : <1.. 80
- Qualificateur “codon_start” dont la valeur est “3”
- Qualificateur “transl_table” dont la valeur est “3”

Les tableaux du code génétique en annexe I, section 9, sont utilisés pour déterminer la valeur du qualificateur “transl_table”

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Exemple : SEQ ID NO:1 est un fragment d’une séquence de codage provenant d’un gène mitochondrial de levure (tableau du code génétique “3-Code mitochondrial des levures”) :

```
...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 *
```

- Clé de caractérisation “CDS”
- Emplacement de la caractéristique : <1.. 80
- Qualificateur “codon_start” dont la valeur est “3”
- Qualificateur “transl_table” dont la valeur est “3”
- Qualificateur “transl_except” prenant la valeur “(pos:30..32,aa:Sec)”

Le qualificateur “transl_except” introduira l’acide aminé sélénocystéine

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Exemple : SEQ ID NO:1 est un fragment d’une séquence de codage provenant d’un gène mitochondrial de levure (tableau du code génétique “3-Code mitochondrial des levures”) :

```
...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 *
```

- Clé de caractérisation “CDS”
- Emplacement de la caractéristique : <1.. 80
- Qualificateur “codon_start” dont la valeur est “3”
- Qualificateur “transl_table” dont la valeur est “3”
- Qualificateur “transl_except” prenant la valeur “(pos:30..32,aa:Sec)”
- Qualificateur “translation” dont la valeur est “DNEEVNEECURLFFKNARHTTSRLT”

Le codon d’arrêt n’est pas indiqué dans le qualificateur translation!

Clés de caractérisation et qualificateurs

Séquences de nucléotides : clé de caractérisation “CDS”

- Exemple : SEQ ID NO:1 est un fragment d’une séquence de codage provenant d’un gène mitochondrial de levure (tableau du code génétique “3-Code mitochondrial des levures”) :

```
...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 *
```

- Clé de caractérisation “CDS”
- Emplacement de la caractéristique : <1.. 80
- Qualificateur “codon_start” dont la valeur est “3”
- Qualificateur “transl_table” dont la valeur est “3”
- Qualificateur “transl_except” avec pour valeur “(pos:30.. 32,aa:Sec)”
- Qualificateur “translation” dont la valeur est “DNEEVNEECURLFFKNARHTTSRLT”
- Une séquence de protéine distincte pour la traduction

Clés de caractérisation et qualificateurs

Séquences nucléotidiques : clé de caractérisation “CDS”

- Exemple : SEQ ID NO:1 est un fragment d’une séquence de codage provenant d’un gène mitochondrial de levure (tableau du code génétique “3-Code mitochondrial des levures”) :

```
...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 *
```

- Clé de caractérisation “CDS”
- Emplacement de la caractéristique : <1.. 80
- Qualificateur “codon_start” dont la valeur est “3”
- Qualificateur “transl_table” dont la valeur est “3”
- Qualificateur “transl_except” avec pour valeur “(pos:30.. 32,aa:Sec)”
- Qualificateur “translation” dont la valeur est
“**DNEEVNEECURLFFKNARHTTSRLT**”
- Une séquence de protéine distincte pour la traduction
- Qualificateur “protein_id” comportant le numéro d’identification de séquence de la protéine traduite

Clés de caractérisation et qualificateurs

Séquences nucléotidiques : clé de caractérisation “CDS”

- L'emplacement de la caractéristique “CDS” peut utiliser l'opérateur d'emplacement “join” pour relier des segments discontinus d'une séquence en une seule région de codage.

`join(location1,location2)`

- L'emplacement de la caractéristique “CDS” peut utiliser l'opérateur d'emplacement “complement” pour indiquer que la caractéristique se trouve sur le brin complémentaire à la séquence indiquée par le descripteur d'emplacement

`complement(location)`

Clés de caractérisation et qualificateurs

Séquences d'acides aminés : clés de caractérisation fréquemment utilisées

Clé de caractérisation	Description	Qualificateurs obligatoires?
SITE	indique un site intéressant d'un seul acide aminé	un qualificateur NOTE doit décrire le site
REGION	indique une région d'intérêt	aucun; NOTE est facultatif
BINDING	Indique un site de fixation pour un groupe chimique	un qualificateur NOTE obligatoire doit contenir le nom du groupe chimique
UNSURE	décrit des régions d'incertitude dans la séquence	aucun; NOTE est facultatif

Clés de caractérisation et qualificateurs

Séquences d'acides aminés : acides aminés modifiés

- Un “acide aminé modifié” désigne tout acide aminé différent de :

L-alanine	L-arginine	L-asparagine	L-aspartate
L-cystéine	L-glutamine	L-glutamate	L-glycine
L-histidine	L-isoleucine	L-leucine	L-lysine
L-méthionine	L-phénylalanine	L-proline	L-pyrrolysine
L-sérine	L-sélénocystéine	L-thréonine	L-tryptophane,
	L-tyrosine	L-valine	

(Norme ST.26, paragraphe 3.e)

- Un “acide aminé modifié” devrait être représenté comme l’acide aminé non modifié correspondant, chaque fois que possible. Sinon, il peut être représenté par le symbole “X”. Par exemple, l’“hydroxylysine” devrait être représentée par le symbole “K” dans la séquence. L’“Ornithine” devrait être représentée par “X”. (Norme ST.26, paragraphe 29)
- Le symbole “X” est équivalent à un seul résidu.

Clés de caractérisation et qualificateurs

Séquences d'acides aminés : acides aminés modifiés

- Plusieurs clés de caractérisation peuvent être utilisées pour indiquer un “acide aminé modifié” :

Clé de caractérisation	Description
SITE	indique un acide aminé modifié différemment
MOD_RES	indique un acide aminé modifié après traduction
CARBOHYD	indique un acide aminé ayant fait l'objet d'une glycosylation
LIPID	indique une liaison covalente d'un fragment lipidique avec un acide aminé

- Un qualificateur NOTE obligatoire doit être inclus avec chacune des clés de caractérisation indiquées ci-dessus, assorti d'une valeur qui décrit la modification.

Clés de caractérisation et qualificateurs

Séquences d'acides aminés : acides aminés modifiés

Gly-Ser-N-acetylAla-Ser-Asp-Val-Orn-Lys-Asn-Val-Leu
 1 5 10

où l'alanine à la position 3 est modifiée
 différemment dans une cellule pour devenir
 n-acetyl alanine

asialyloligosaccharide

- Séquence représentée dans le listage des séquences sous la forme :

GSASDVXKNVL

- Clé de caractérisation "MOD_RES" dont l'emplacement est "3" et un qualificateur NOTE dont la valeur est "N-acetylalanine"
- Clé de caractérisation "SITE" dont l'emplacement est "7" et un qualificateur NOTE qui a pour valeur "ornithine"
- Clé de caractérisation "CARBOHYD" dont l'emplacement est "9" et un qualificateur NOTE dont la valeur est "Asn side-chain linked to asialyloligosaccharide" (chaîne latérale Asn reliée à l'asialyloligosaccharide)

Formats de l'emplacement de la caractéristique

Formats d'emplacement

Descripteurs d'emplacement pour tous les types de molécules

- Les descripteurs d'emplacement sont utilisés pour indiquer l'emplacement d'une caractéristique dans une séquence
- La norme ST.26 contient des exigences obligatoires pour le format des descripteurs d'emplacement
- Les formats de descripteurs d'emplacement suivants peuvent être utilisés pour les séquences nucléotidiques et d'acides aminés :

Type de descripteurs d'emplacement	Syntaxe	Description
Numéro de résidu unique	x	Désigne un résidu unique dans la séquence.
Numéros de résidus délimitant un ensemble dans la séquence	x . . y	Désigne une série continue de résidus délimitée par un résidu de début et un résidu de fin, ces deux résidus étant inclus dans la série.
Résidus situés avant le premier ou après le dernier numéro de résidu indiqué	<x >x <x . . y x . . >y <x . . >y	Désigne une région qui comprend un résidu ou une série de résidus indiqués et qui s'étend au-delà d'un résidu indiqué. Les symboles "<" et ">" peuvent être employés à l'égard d'un résidu unique ou des numéros du résidu de début et de fin d'une série de résidus pour signaler qu'une caractéristique s'étend au-delà du numéro de résidu indiqué.

Norme ST.26 de l'OMPI, paragraphe 66.a)

Formats d'emplacement

Descripteurs d'emplacement pour tous les types de molécules

Exemple d'emplacement	Description
467	Désigne le résidu 467 de la séquence.
340..565	Désigne une série continue de résidus dont les bornes sont le 340 et le 565, ces bornes étant incluses dans la série.
<1	Désigne un emplacement de caractéristique situé avant le premier résidu.
<345..500	Indique que le point exact de la borne inférieure d'une caractéristique est inconnu. L'emplacement commence à un résidu situé quelque part avant le 345 et continue jusqu'au résidu 500 inclus.
<1..888	Indique que la caractéristique commence avant le premier résidu de la séquence et continue jusqu'au résidu 888 inclus.
1..>888	Indique que la caractéristique commence au premier résidu de la séquence et continue au-delà du résidu 888.

Norme ST.26 de l'OMPI, paragraphe 70.a)

Formats d'emplacement

Descripteurs d'emplacement exclusivement pour les séquences de nucléotides

- Les formats de descripteurs d'emplacement suivants peuvent être utilisés **UNIQUEMENT** pour les séquences d'ADN et d'ARN :

Type de descripteurs d'emplacement	Syntaxe	Description
Site s'étendant entre deux numéros de résidus adjacents	x^y	Désigne un site entre deux résidus adjacents, par exemple le site d'un clivage endonucléolytique. Les numéros de position des résidus adjacents sont séparés par un caret (^). Les formats autorisés pour ce descripteur sont x^{x+1} (par exemple 55^{56}), ou pour les nucléotides circulaires, x^1 , où "x" est la longueur totale de la molécule, c'est-à-dire 1000^1 pour une molécule circulaire de longueur 1000.

Norme ST.26 de l'OMPI, paragraphe 66.b)

Formats d'emplacement

Descripteurs d'emplacement exclusivement pour les séquences de nucléotides

- Lors de l'utilisation d'un format d'emplacement x^y , x et y doivent être des résidus adjacents

Exemple d'emplacement	Description
123^124	Désigne un site entre les résidus 123 et 124
867^1	Dans une molécule circulaire comportant 867 résidus, désigne un site entre le résidu indiqué à la position 1 et le résidu indiqué à la position 867.

Norme ST.26 de l'OMPI, paragraphe 70.b)

Formats d'emplacement

Descripteurs d'emplacement : cas particulier des séquences d'acides aminés

- Le descripteur d'emplacement $x..y$ décrit une liaison intrachaîne entre deux résidus indiqués lorsqu'il est utilisé avec la clé de caractérisation "CROSSLNK" ou "DISULFID"

Type de descripteurs d'emplacement	Syntaxe	Description
Numéros de résidus reliés par une liaison intrachaîne	$x..y$	Désigne des acides aminés reliés par une liaison intrachaîne lorsqu'utilisés avec une caractéristique indiquant une liaison intrachaîne, telles que "CROSSLNK" ou "DISULFID".

Norme ST.26 de l'OMPI, paragraphe 66.c)

Formats d'emplacement

Descripteurs d'emplacement : cas particulier des séquences d'acides aminés

```

<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>

```

Formats d'emplacement

Opérateurs d'emplacement pour les séquences de nucléotides

- Trois opérateurs d'emplacement peuvent être utilisés dans des séquences d'ADN et d'ARN : “Join,” “order,” et “complement” (jonction, ordonnancement, complément)
- L'utilisation d'opérateurs d'emplacement de jonction “join” implique que les résidus nucléotidiques désignés par les descripteurs d'emplacement sont physiquement mis en contact par des processus biologiques (norme ST.26, paragraphe 68)
- Les opérateurs d'emplacement assurant un rôle de jonction “join” ou d'ordonnancement “order” nécessitent au moins deux descripteurs d'emplacement séparés par des virgules
- L'opérateur d'emplacement “complement” peut être employé en combinaison soit avec “join” soit avec “order”

Syntaxe de l'emplacement	Description de l'emplacement
<code>join(location, location, , location)</code>	Les emplacements indiqués sont joints (placés bout à bout) pour former une seule séquence contiguë.
<code>order(location, location, , location)</code>	Les éléments se trouvent dans l'ordre indiqué mais aucune information ne permet de déterminer s'il est raisonnable de les joindre.
<code>complement(location)</code>	Indique que la caractéristique se trouve sur le brin complémentaire à la série de la séquence indiquée par le descripteur d'emplacement, lorsque la séquence est lue dans le sens 5'-3', ou dans le sens qui reproduit le sens 5'-3'.

Formats d'emplacement

Opérateurs d'emplacement pour les séquences de nucléotides

Exemple d'emplacement	Description
<code>join(12..78,134..202)</code>	Indique que les régions 12 à 78 et 134 à 202 devraient être jointes pour constituer une séquence contiguë
<code>Order(15..228,341..502)</code>	Indique que les régions 15 à 228 et 341 à 502 sont présentes dans l'ordre indiqué.
<code>complement(join(2691..4571,4918..5163))</code>	Joint les nucléotides 2691 à 4571 et 4918 à 5163, puis complète les segments joints (la caractéristique est située sur le brin complémentaire au brin présenté).
<code>join(complement(4918..5163),complement(2691..4571))</code>	Complète les régions 4918 à 5163 et 2691 à 4571, puis joint les segments complétés (la caractéristique est située sur le brin complémentaire au brin présenté).

Norme ST.26 de l'OMPI, paragraphe 70.b)

Questions?

Formats des valeurs des qualificateurs

Valeurs des qualificateurs

- Les qualificateurs définissent les caractéristiques plus précisément

Valeurs des qualificateurs

- Les qualificateurs définissent les caractéristiques plus précisément
- Les qualificateurs se composent d'un nom de qualificateur et, souvent, d'une valeur de qualificateur

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

Valeurs des qualificateurs

- Les qualificateurs définissent les caractéristiques plus précisément
- Les qualificateurs se composent d'un nom de qualificateur et, souvent, d'une valeur de qualificateur

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

- Chaque clé de caractérisation sera assortie d'une liste de qualificateurs qui sont autorisés pour cette caractéristique. Certaines clés de caractérisation ont des qualificateurs obligatoires.

Valeurs des qualificateurs

- Les qualificateurs définissent les caractéristiques plus précisément
- Les qualificateurs se composent d'un nom de qualificateur et, souvent, d'une valeur de qualificateur

```
<INSDQualifier>  
  <INSDQualifier_name>transl_table</INSDQualifier_name>  
  <INSDQualifier_value>12</INSDQualifier_value>  
</INSDQualifier>
```
- Chaque clé de caractérisation sera assortie d'une liste de qualificateurs qui sont autorisés pour cette caractéristique. Certaines clés de caractérisation ont des qualificateurs obligatoires.
- Les clés de caractérisation et leurs qualificateurs autorisés sont indiqués dans l'annexe I, section 5 (séquences de nucléotides) et section 7 (séquences d'acides aminés) de la norme ST.26.

Valeurs des qualificateurs

- Exemple : la clé de caractérisation “misc_binding” a un qualificateur obligatoire, “bound_moiety,” et 6 qualificateurs facultatifs :

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

(Norme ST.26, annexe I, section 5.12)

Valeurs des qualificateurs

- Les qualificateurs définissent les caractéristiques plus précisément
- Les qualificateurs se composent d'un nom de qualificateur et, souvent, d'une valeur de qualificateur

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

- Chaque clé de caractérisation sera assortie d'une liste de qualificateurs qui sont autorisés pour cette caractéristique. Certaines clés de caractérisation ont des qualificateurs obligatoires.
- Les clés de caractérisation et leurs qualificateurs autorisés sont indiqués dans l'annexe I, section 5 (séquences de nucléotides) et section 7 (séquences d'acides aminés) de la norme ST.26.
- Les qualificateurs pour les séquences de nucléotides de concert avec leur description sont indiqués dans l'annexe I, section 6 (séquences de nucléotides) et section 8 (séquences d'acides aminés) de la norme ST.26.

Valeurs des qualificateurs

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.

(Norme ST.26, annexe I, section 6.3 et 6.22)

Valeurs des qualificateurs

Types de format

- Les valeurs des qualificateurs ont plusieurs types de formats :
 1. Qualificateurs avec des choix de valeurs prédéfinies;
 2. Qualificateurs avec un format de valeur défini;
 3. Qualificateurs où la valeur est une séquence;
 4. Qualificateurs avec AUCUNE valeur;
 5. Qualificateurs avec des valeurs de texte libre (“free text”)
 - un sous-ensemble de valeurs de “texte libre” de qualificateurs est classé comme “dépendant de la langue”

Valeurs des qualificateurs

Types de format - Choix de valeurs prédéfinies

- Qualificateurs assortis de choix de valeurs prédéfinies;
- Exemples :

“Codon_start” : les valeurs peuvent être “1”, “2” ou “3”

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>

Valeurs des qualificateurs

Types de format - Choix de valeurs prédéfinies

- Exemples : “rpt_type” est assorti d’un ensemble limité de choix de valeurs

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;

Valeurs des qualificateurs

Types de format - Choix de valeurs prédéfinies

- Pour les qualificateurs assortis de choix de valeurs prédéfinies, le logiciel WIPO Sequence présentera toutes les valeurs autorisées dans un menu déroulant prérempli :

▼ 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	atgggctgt	ggaagtgccg	ctgacggccc	agaagggcgt
cttactgtta	caagaagagt	atcatgacgc	gcatcatagc	cacttagacc	cgcacttatg
gttgccgocg	caaaatatca	ttgcocctgt	gcaaggattg	gataaacata	ctgctgagtt

180

Valeurs des qualificateurs

Types de format - format de valeur défini

- Qualificateurs avec un format de valeur défini
- Exemple : “anticodon” : la valeur doit se présenter sous la forme
“(pos:<location>,aa :<amino_acid>,seq:<text>)”

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>

Valeurs des qualificateurs

Types de format - Séquences

- Qualificateurs où la valeur est une séquence
- Exemple : “translation” : la valeur doit être une séquence utilisant les abréviations d’acides aminés composées d’une lettre

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.

Valeurs des qualificateurs

Types de format - Séquences

- Qualificateur “replace” : la valeur peut être un seul résidu de nucléotide, une séquence de résidus, ou vide
- Une valeur vide pour “replace” signifie la suppression du résidu indiqué dans la caractéristique correspondante

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>

Valeurs des qualificateurs

Types de format - Séquences

- Qualificateur “replace” : la valeur peut être un seul résidu de nucléotide, une séquence de résidus, ou vide
- Une valeur vide pour “replace” signifie la suppression du résidu indiqué dans la caractéristique correspondante

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>

“Valeur vide”

Valeurs des qualificateurs

Types de format - AUCUNE valeur

- Qualificateurs SANS valeur indiquée

- Exemples :

“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

- Le logiciel WIPO Sequence ne permettra pas l'ajout d'une valeur pour les qualificateurs sans valeur indiquée
- Ces qualificateurs doivent comporter un élément “INSDQualifier_value” vide

Valeurs des qualificateurs

Types de format – “Texte libre”

- De nombreux qualificateurs ont un format de valeur de “texte libre”
- La norme ST.26, paragraphe 3.n) définit le “texte libre” comme “...*un format de valeur autorisé pour certains qualificateurs. Il s’agit d’un texte descriptif qui se présente sous forme de segments de phrases ou tout autre format précisé (comme indiqué à l’annexe I)*”.
- Les valeurs de texte libre des qualificateurs sont limitées à 1 000 caractères (norme ST.26, paragraphe 86)
- Un sous-ensemble de qualificateurs assortis de formats de valeur de “texte libre” sont “*dépendants de la langue*”
- Le texte libre “dépendant de la langue” “peut nécessiter une traduction aux fins des procédures internationales, nationales ou régionales”.
Norme ST.26, paragraphe 3.o)

Valeurs des qualificateurs

Types de format – “Texte libre”

- Les qualificateurs qui comportent du “texte libre dépendant de la langue” peuvent nécessiter une traduction
- Comment déterminer si un qualificateur comportant un format de valeur de texte libre est dépendant de la langue?
 - Le tableau 5 de l’annexe I, section 6, de la norme ST.26 répertorie tous les qualificateurs de séquences de nucléotides ayant un format de valeur de texte libre dépendant de la langue
 - Le tableau 6 de l’annexe I, section 8, de la norme ST.26 répertorie tous les qualificateurs de séquences d’acides aminés ayant un format de valeur de texte libre dépendant de la langue
 - Veuillez consulter le “format de valeur obligatoire” dans la description du qualificateur

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>

Valeurs des qualificateurs

Types de format – “Texte libre”

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.

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- Les valeurs des qualificateurs à “texte libre dépendant de la langue” peuvent être données dans deux langues dans le listage des séquences XML : en anglais et une langue autre que l’anglais (norme ST.26, paragraphe 87)

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- Les valeurs des qualificateurs à “texte libre dépendant de la langue” peuvent être données dans deux langues dans le listage des séquences XML : en anglais et une langue autre que l’anglais (norme ST.26, paragraphe 87)
- La valeur dépendant de la langue d’un qualificateur en anglais doit être présentée dans l’élément INSDQualifier_value

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- Les valeurs des qualificateurs à “texte libre dépendant de la langue” peuvent être données dans deux langues dans le listage des séquences XML : en anglais et une langue autre que l’anglais (norme ST.26, paragraphe 87)
- La valeur dépendant de la langue d’un qualificateur en anglais doit être présentée dans l’élément INSDQualifier_value
- La valeur dépendant de la langue d’un qualificateur dans une langue autre que l’anglais doit être présentée dans l’élément NonEnglishQualifier_value

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- Les valeurs des qualificateurs à “texte libre dépendant de la langue” peuvent être données dans deux langues dans le listage des séquences XML : en anglais et une langue autre que l’anglais (norme ST.26, paragraphe 87)
- La valeur dépendant de la langue d’un qualificateur en anglais doit être présentée dans l’élément INSDQualifier_value
- La valeur dépendant de la langue d’un qualificateur dans une langue autre que l’anglais doit être présentée dans l’élément NonEnglishQualifier_value
- Un élément NonEnglishQualifier_value est autorisé **UNIQUEMENT** pour les qualificateurs qui ont un format de valeur de texte libre dépendant de la langue (norme ST.26, paragraphe 87.b))

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

```

<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">

```

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- Lorsqu'un listage des séquences XML contient des valeurs de qualificateurs qui ne sont pas en anglais :

1. l'élément racine du listage des séquences au format XML doit contenir un attribut “nonEnglishFreeTextLanguageCode” composé d'une abréviation de code à deux lettres appropriée (norme ST.26, paragraphes 43 et 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">
```

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- Lorsqu’un listage des séquences XML contient des valeurs de qualificateurs qui ne sont pas en anglais :
 1. l’élément racine du listage des séquences au format XML doit contenir un attribut “nonEnglishFreeTextLanguageCode” composé d’une abréviation de code à deux lettres appropriée (norme ST.26, paragraphes 43 et 87.b));
 2. TOUS les qualificateurs dépendants de la langue figurant dans le listage des séquences doivent présenter des valeurs dans la langue indiquée dans l’attribut “nonEnglishFreeTextLanguageCode”;

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- Lorsqu’un listage des séquences XML contient des valeurs de qualificateurs qui ne sont pas en anglais :
 1. l’élément racine du listage des séquences au format XML doit contenir un attribut “nonEnglishFreeTextLanguageCode” composé d’une abréviation de code à deux lettres appropriée (norme ST.26, paragraphes 43 et 87.b));
 2. TOUS les qualificateurs dépendants de la langue figurant dans le listage des séquences doivent présenter des valeurs dans la langue indiquée dans l’attribut “nonEnglishFreeTextLanguageCode”;
 3. Lorsque NonEnglishQualifier_value et INSDQualifier_value sont tous deux présents pour un seul qualificateur, les informations contenues dans les deux éléments doivent être équivalentes (norme ST.26, paragraphe 87.c)).

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- Attribut INSDQualifier “id” – Qu’est-ce que c’est?
- Le paragraphe 87.d) de la norme ST.26 indique *“Pour les qualificateurs dépendants de la langue, l’élément INSDQualifier_value peut comporter un attribut facultatif id. La valeur de cet attribut doit être dans le format “q” suivi d’un nombre entier positif, par exemple “q23”, et doit être unique pour un élément INSDQualifier_value, c’est-à-dire que la valeur de l’attribut ne doit être utilisée qu’une seule fois dans un fichier de listage des séquences”*.

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

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- Attributs INSDQualifier “id” :
 - Identifient uniquement des valeurs de qualificateurs qui peuvent nécessiter une traduction à des fins d’exportation dans un fichier XLIFF par le logiciel WIPO Sequence;
 - Facultatifs;
 - Uniquement autorisés pour les qualificateurs dépendants de la langue;
 - Doivent être uniques dans un listage de séquences;
 - Ajoutés automatiquement dans un fichier XML généré par WIPO Sequence.

Valeurs des qualificateurs

Types de format – “Texte libre dépendant de la langue”

- 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">
```

- Le paragraphe 43 de la norme ST.26 définit l’attribut “originalFreeTextLanguageCode” comme, “Le code de langue... pour la langue originale unique dans laquelle les qualificateurs de texte libre dépendant de la langue ont été établis”.
- Cet attribut est FACULTATIF.

Cas particuliers :
Molécules hybrides
d'ADN/ARN,
uracile dans de l'ADN et
thymine dans de l'ARN

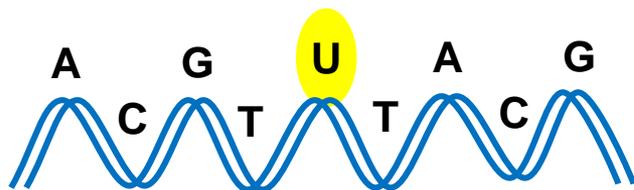
Cas particuliers

- Rappel : le symbole “u” n’est pas autorisé pour l’uracile dans la norme ST.26
- Dans de l’ADN, le symbole “t” représente la thymine
- Dans l’ARN, le symbole “t” représente l’uracile
- Deux scénarios doivent être envisagés :
 1. Une molécule d’ADN avec une base azotée d’uracile ou une molécule d’ARN avec une base azotée de thymine;
 2. Une molécule hybride d’ADN/ARN

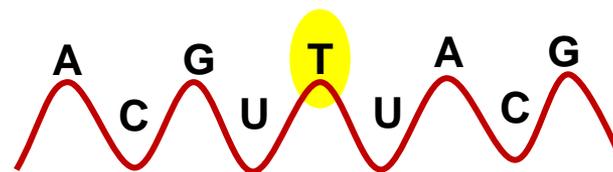
Cas particuliers

Uracile dans de l'ADN et thymine dans de l'ARN

- Si la séquence a un squelette d'ADN avec une base azotée d'uracile, OU un squelette d'ARN avec une base azotée de thymine, la description doit être "nucléotide modifié"



ADN



ARN

Le paragraphe 14 de la norme ST.26 s'applique :

*“14. Le symbole “t” désigne la thymine dans de l'ADN et l'uracile dans de l'ARN. L'uracile dans de l'ADN ou la thymine dans de l'ARN sont considérés comme des **nucléotides modifiés** et doivent être accompagnés d'une description supplémentaire dans le tableau de caractéristiques au sens du paragraphe 19.”*

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence d'ADN suivante :

5' -cgucccacgugtccgaggua-3'

- Observez la “thymine” à la position 12. Ce résidu doit être annoté comme étant un nucléotide modifié.
- Le paragraphe 19 de la norme ST.26 indique : *L'uracile dans de l'ADN ou la thymine dans de l'ARN sont considérés comme des nucléotides modifiés et doivent être représentés dans la séquence par un “t” et être accompagnés d'une description supplémentaire dans le tableau de caractéristiques. Cette description doit comporter la clé de caractérisation “modified_base”, le qualificateur “mod_base” dont la valeur doit être “OTHER”, et un qualificateur de type “note” dont la valeur doit être respectivement “uracil” ou “thymine”.*

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence d'ADN suivante :

5' -cguccacgugtccgaggua-3'

- Observez la “thymine” à la position 12. Ce résidu doit être annoté comme étant un nucléotide modifié.
- Le paragraphe 19 de la norme ST.26 indique : *L'uracile dans de l'ADN ou la thymine dans de l'ARN sont considérés comme des nucléotides modifiés et doivent être représentés dans la séquence par un “t” et être accompagnés d'une description supplémentaire dans le tableau de caractéristiques. Cette description doit comporter la clé de caractérisation “modified_base”, le qualificateur “mod_base” dont la valeur doit être “OTHER”, et un qualificateur de type “note” dont la valeur doit être respectivement “uracil” ou “thymine”.*

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence d'ADN suivante :

5' -cguccacgugtccgaggua-3'

- ✓ Tous les résidus d'uracile doivent être représentés par le symbole "t". C'est pourquoi, la séquence doit être représentée dans le listage des séquences sous la forme :

cg~~t~~cccacgtgtccgaggta

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence d'ADN suivante :

5' -cguccacgugtccgaggua-3'

- ✓ Tous les résidus d'uracile doivent être représentés par le symbole "t". C'est pourquoi, la séquence doit être représentée dans le listage des séquences sous la forme :

cgtcccacgtgttccgaggta

- ✓ Clé de caractérisation "modified_base" avec emplacement "12"

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence d'ADN suivante :

5' -cguccacgugtccgaggua-3'

- ✓ Tous les résidus d'uracile doivent être représentés par le symbole "t". C'est pourquoi, la séquence doit être représentée dans le listage des séquences sous la forme :

cgtcccacgtgtccgaggta

- ✓ Clé de caractérisation "modified_base" avec emplacement "12"
- ✓ Qualificateur "mod_base" avec pour valeur "OTHER"

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence d'ADN suivante :

5' -cguccacgugtccgaggua-3'

- ✓ Tous les résidus d'uracile doivent être représentés par le symbole "t". C'est pourquoi, la séquence doit être représentée dans le listage de séquences sous la forme suivante :

cgtcccacgtgttccgaggta

- ✓ Clé de caractérisation "modified_base" avec emplacement "12"
- ✓ Qualificateur "mod_base" avec pour valeur "OTHER"
- ✓ Qualificateur "note" avec pour valeur "thymine"

Cas particuliers

Séquences d'ADN et d'ARN

```

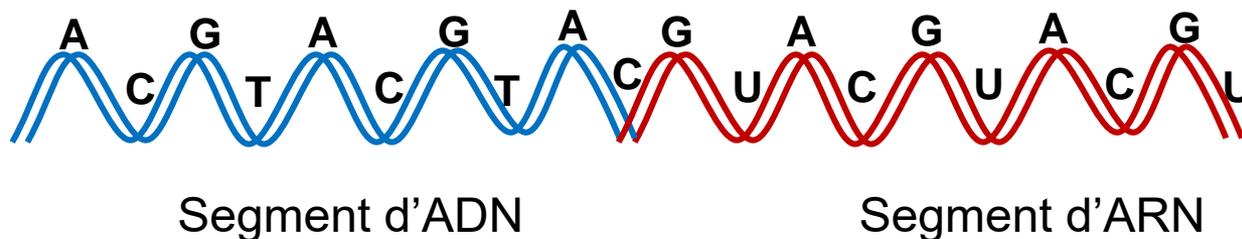
<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length>20</INSDSeq_length>
    <INSDSeq_moltype>RNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..20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

```

Cas particuliers

Séquences hybrides d'ADN/ARN

- Si une séquence est une molécule hybride; par exemple, une partie du squelette est un ADN et une partie du squelette est un ARN :



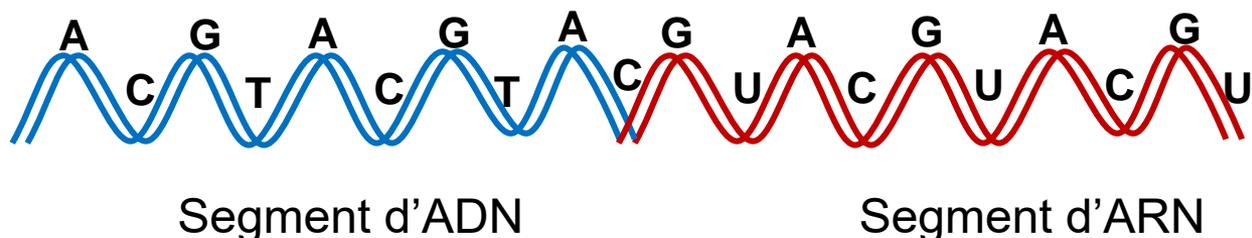
Le paragraphe 55 de la norme ST.26 s'applique :

“55. Pour une séquence de nucléotides qui contient à la fois des segments d'ADN et d'ARN comprenant un ou plusieurs nucléotides, le type de molécule doit prendre la valeur “DNA”. La molécule combinée d'ADN/ARN doit en outre être décrite dans le tableau de caractéristiques à l'aide de la clé de caractérisation “source”, du qualificateur obligatoire “organism”, qui prend la valeur “synthetic construct”, et du qualificateur obligatoire “mol_type”, qui prend la valeur “other DNA”. Chaque segment d'ADN et d'ARN de la molécule combinée d'ADN/ARN doit en outre être décrit par la clé de caractérisation “misc_feature” et par le qualificateur “note”, ce dernier indiquant s'il s'agit d'un segment d'ADN ou d'ARN.”

Cas particuliers

Séquences hybrides d'ADN/ARN

- Si une séquence est une molécule hybride; par exemple, une partie du squelette est un ADN et une partie du squelette est un ARN :



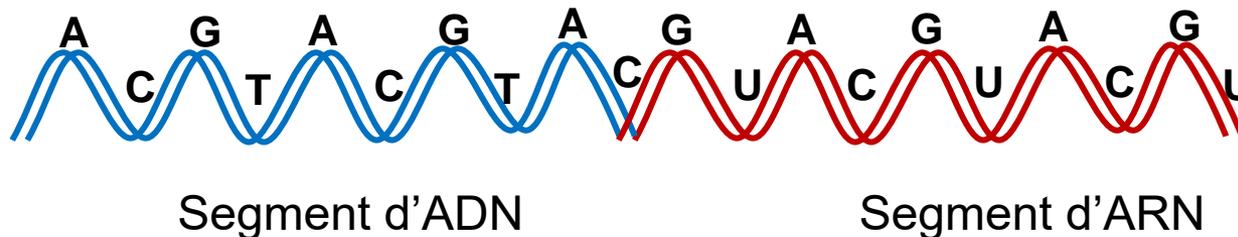
Le paragraphe 55 de la norme ST.26 s'applique :

“55. Pour une séquence de nucléotides qui contient à la fois des segments d'ADN et d'ARN comprenant un ou plusieurs nucléotides, **le type de molécule doit prendre la valeur “DNA”**. La molécule combinée d'ADN/ARN doit en outre être décrite dans le tableau de caractéristiques à l'aide de la clé de caractérisation “source”, du qualificateur obligatoire “organism”, qui prend la valeur “synthetic construct”, et du qualificateur obligatoire “mol_type”, qui prend la valeur “other DNA”. Chaque segment d'ADN et d'ARN de la molécule combinée d'ADN/ARN doit en outre être décrit par la clé de caractérisation “misc_feature” et par le qualificateur “note”, ce dernier indiquant s'il s'agit d'un segment d'ADN ou d'ARN.”

Cas particuliers

Séquences hybrides d'ADN/ARN

- Si une séquence est une molécule hybride; par exemple, une partie du squelette est un ADN et une partie du squelette est un ARN :



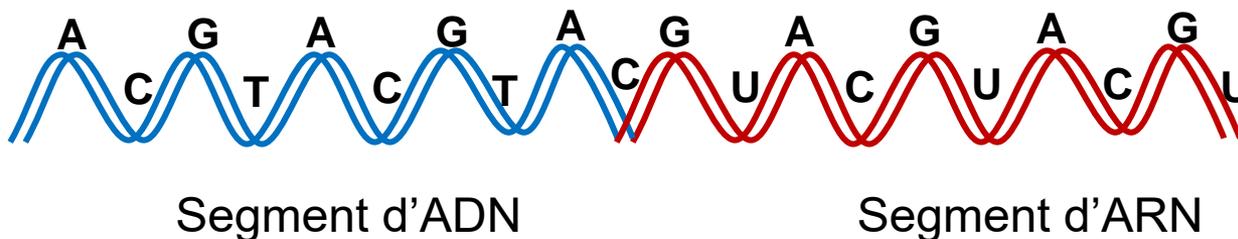
Le paragraphe 55 de la norme ST.26 s'applique :

*“55. Pour une séquence de nucléotides qui contient à la fois des segments d'ADN et d'ARN comprenant un ou plusieurs nucléotides, le type de molécule doit prendre la valeur “DNA”. La molécule combinée d'ADN/ARN doit en outre être décrite dans le tableau de caractéristiques à l'aide de la clé de caractérisation “source”, du qualificateur obligatoire **“organism”, qui prend la valeur “synthetic construct”,** et du qualificateur obligatoire “mol_type”, qui prend la valeur “other DNA”. Chaque segment d'ADN et d'ARN de la molécule combinée d'ADN/ARN doit en outre être décrit par la clé de caractérisation “misc_feature” et par le qualificateur “note”, ce dernier indiquant s'il s'agit d'un segment d'ADN ou d'ARN.”*

Cas particuliers

Séquences hybrides d'ADN/ARN

- Si une séquence est une molécule hybride; par exemple, une partie du squelette est un ADN et une partie du squelette est un ARN :



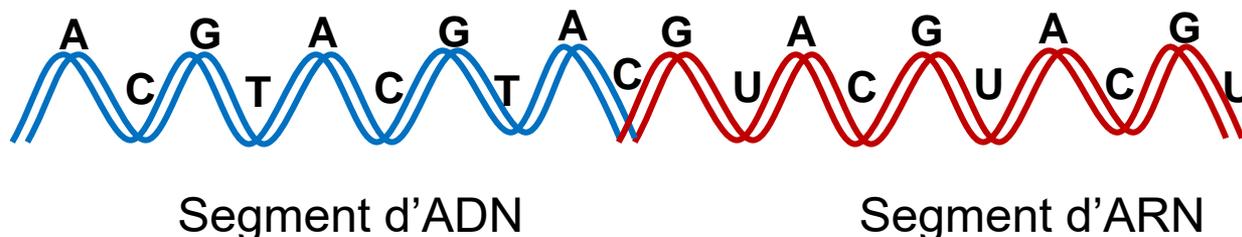
Le paragraphe 55 de la norme ST.26 s'applique :

“55. Pour une séquence de nucléotides qui contient à la fois des segments d'ADN et d'ARN comprenant un ou plusieurs nucléotides, le type de molécule doit prendre la valeur “DNA”. La molécule combinée d'ADN/ARN doit en outre être décrite dans le tableau de caractéristiques à l'aide de la clé de caractérisation “source”, du qualificateur obligatoire “organism”, qui prend la valeur “synthetic construct”, et du qualificateur obligatoire “mol_type”, qui prend la valeur “other DNA”. Chaque segment d'ADN et d'ARN de la molécule combinée d'ADN/ARN doit en outre être décrit par la clé de caractérisation “misc_feature” et par le qualificateur “note”, ce dernier indiquant s'il s'agit d'un segment d'ADN ou d'ARN.”

Cas particuliers

Séquences hybrides d'ADN/ARN

- Si une séquence est une molécule hybride; par exemple, une partie du squelette est un ADN et une partie du squelette est un ARN :



Le paragraphe 55 de la norme ST.26 s'applique :

“55. Pour une séquence de nucléotides qui contient à la fois des segments d'ADN et d'ARN comprenant un ou plusieurs nucléotides, le type de molécule doit prendre la valeur “DNA”. La molécule combinée d'ADN/ARN doit en outre être décrite dans le tableau de caractéristiques à l'aide de la clé de caractérisation “source”, du qualificateur obligatoire “organism”, qui prend la valeur “synthetic construct”, et du qualificateur obligatoire “mol_type”, qui prend la valeur “other DNA”. Chaque segment d'ADN et d'ARN de la molécule combinée d'ADN/ARN doit en outre être décrit par la clé de caractérisation “misc_feature” et par le qualificateur “note”, ce dernier indiquant s'il s'agit d'un segment d'ADN ou d'ARN.”

Cas particuliers

Séquences hybrides d'ADN/ARN

Une demande divulgue la séquence hybride d'ADN/ARN suivante :

5' -ACCTGCcgucccacguguccgagguaGCATTA-3'

où les symboles en majuscules représentent la portion d'ADN et les symboles en minuscules la portion d'ARN.

- Les résidus 1 à 6 et 27 à 32 sont de l'ADN, les résidus 7 à 26 sont de l'ARN

- Ce qu'il faut prendre en compte :
 1. Désignation de l'organisme
 2. Type de molécule et mol_type
 3. Identification des segments d'ADN et d'ARN

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence hybride d'ADN/ARN suivante :

5' -ACCTGCcgucccacguguccgagguaGCATTA-3'

où les symboles en majuscules représentent la portion d'ADN et les symboles en minuscules la portion d'ARN.

- Le paragraphe 55 de la norme ST.26 indique : *“... , le type de molécule doit prendre la valeur “DNA”. La molécule combinée d'ADN/ARN doit en outre être décrite dans le tableau de caractéristiques à l'aide de la clé de caractérisation “source”, du qualificateur obligatoire “organism”, qui prend la valeur “synthetic construct”, et du qualificateur obligatoire “mol_type”, qui prend la valeur “other DNA”.*

- ✓ Type de molécule = “DNA”
- ✓ Nom de l'organisme qualificateur = “synthetic construct”
- ✓ Qualifier mol_type = “other DNA”

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence hybride d'ADN/ARN suivante :

5' -ACCTGCcgucccacguguccgagguaGCATTA-3'

où les symboles en majuscules représentent la portion d'ADN et les symboles en minuscules la portion d'ARN.

- ✓ Tous les résidus d'uracile doivent être représentés par le symbole "t". C'est pourquoi, la séquence doit être représentée dans le listage des séquences sous la forme :

acctgccgtcccacgtgtccgaggtagcatta

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence hybride d'ADN/ARN suivante :

5' -ACCTGCcgucccacguguccgagguaGCATTA-3'

où les symboles en majuscules représentent la portion d'ADN et les symboles en minuscules la portion d'ARN.

- Le paragraphe 55 de la norme ST.26 indique : *“Chaque segment d'ADN et d'ARN de la molécule combinée d'ADN/ARN doit en outre être décrit par la clé de caractérisation “misc_feature” et par le qualificateur “note”, ce dernier indiquant s'il s'agit d'un segment d'ADN ou d'ARN.”*
- ✓ Trois segments = trois clés de caractérisation “misc_feature”

Cas particuliers

Séquences d'ADN et d'ARN

Une demande divulgue la séquence hybride d'ADN/ARN suivante :

5' -ACCTGCcgucccacguguccgagguaGCATTA-3'

où les symboles en majuscules représentent la portion d'ADN et les symboles en minuscules la portion d'ARN.

- ✓ Segment 1, résidus 1 à 6 :
clé de caractérisation "misc_feature" et emplacement "1..6"
Qualificateur "note" qui prend la valeur "DNA"
- ✓ Segment 2, résidus 7 à 26 :
clé de caractérisation "misc_feature" et emplacement "7..26"
Qualificateur "note" qui prend la valeur "RNA"
- ✓ Segment 3, résidus 27 à 32 :
clé de caractérisation "misc_feature" et emplacement "27..32"
Qualificateur "note" qui prend la valeur "DNA"

Cas particuliers

Séquences d'ADN et d'ARN

```

<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>

```

Cas particuliers

Séquences d'ADN et d'ARN

```

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

```

Segment 1 ADN

Segment 2 ARN

Segment 3 ADN

Questions?

Analogues nucléotidiques, acides aminés D et séquences ramifiées

Nouveaux types de molécules

Analogues nucléotidiques

- Les séquences d'acides nucléiques qui contiennent un ou plusieurs analogues nucléotidiques sont soumises aux règles de la norme ST.26
- Les analogues nucléotidiques sont inclus dans la définition d'un "nucléotide" au paragraphe 3.g)2, de la norme ST.26 :

“un analogue du 2' désoxyribose 5' monophosphate ou du ribose 5' monophosphate, qui lorsqu'il constitue le squelette d'un analogue d'acide nucléique, forme une disposition de bases azotées reproduisant celle des acides nucléiques contenant un squelette 2' désoxyribose 5' monophosphate ou ribose 5' monophosphate, l'analogue d'acide nucléique étant capable de former une paire de base avec un acide nucléique complémentaire”

- Les analogues nucléotidiques courants sont les acides nucléiques peptidiques, les acides nucléiques à glycol, les acides nucléiques à thréose et les morpholinos
- Ils doivent être représentés de gauche à droite de manière à reproduire le sens 5'-3'. (Norme ST.26, paragraphe 11)

Nouveaux types de molécules

Analogues nucléotidiques

Une demande de brevet divulgue la séquence d'un acide nucléique à glycol (ANG) suivante :



où l'extrémité gauche de la séquence PO_4 reproduit le côté 5' d'une séquence d'ADN.

✓ Cette séquence doit être incluse dans un listage des séquences

Nouveaux types de molécules

Analogues nucléotidiques

Une demande de brevet divulgue la séquence d'un acide nucléique à glycol (ANG) suivante :

$\text{PO}_4\text{-tagttcattgactaaggctccccattgact-OH}$

où l'extrémité gauche de la séquence PO_4 reproduit le côté 5' d'une séquence d'ADN.

- ✓ Cette séquence doit être incluse dans un listage des séquences
- ✓ où l'extrémité gauche de la séquence PO_4 reproduit le côté 5' de sorte que la séquence doit être représentée dans le sens indiqué

Nouveaux types de molécules

Analogues nucléotidiques

Une demande de brevet divulgue la séquence d'un acide nucléique à glycol (ANG) suivante :



où l'extrémité gauche de la séquence PO_4 reproduit le côté 5' d'une séquence d'ADN.

- ✓ Cette séquence doit être incluse dans un listage des séquences
- ✓ où l'extrémité gauche de la séquence PO_4 reproduit le côté 5' de sorte que la séquence doit être représentée dans le sens indiqué
- ✓ L'intégralité de la séquence doit être annotée au moyen de la clé de caractérisation "modified_base", comporter un qualificateur "mod_base" qui prend la valeur "OTHER" et un qualificateur note qui comprend le nom complet non abrégé des nucléotides modifiés, tels que "glycol nucleic acids" (acides nucléiques à glycol)

Nouveaux types de molécules

Analogues nucléotidiques

```

- <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>

```

Note : une analyse détaillée de cet exemple figure dans la norme ST.26 de l'OMPI, annexe VI, exemple 3.g)-4.

Nouveaux types de molécules

Acides aminés D

- Les séquences d'acides aminés qui contiennent un ou plusieurs acides aminés D sont soumises aux règles de la norme ST.26
- Les acides aminés sont inclus dans la définition d'un "acide aminé" au paragraphe 3.a) de la norme ST.26 :

“acide aminé” désigne tout acide aminé pouvant être représenté à l’aide des symboles indiqués dans l’annexe I (voir section 3, tableau 3). Ces acides aminés comprennent notamment les acides aminés D et les acides aminés contenant des chaînes latérales modifiées ou synthétiques.”

- Les acides aminés D doivent être représentés dans la séquence sous la forme d'un symbole d'acide aminé L non modifié, chaque fois que possible
- doivent être décrits dans le tableau de caractéristiques comme un acide aminé modifié

Nouveaux types de molécules

Acides aminés D

Une demande de brevet décrit la séquence suivante :

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

✓ Cette séquence doit être incluse dans un listage des séquences.

Nouveaux types de molécules

Acides aminés D

Une demande de brevet décrit la séquence suivante :

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

- ✓ Cette séquence doit être incluse dans un listage des séquences.
- ✓ Elle devrait être représentée sous la forme : AGKLGGM
- ✓ L'alanine à la position 1, l'acide glutamique à la position 2 et la méthionine à la position 6 doivent être chacun annotés au moyen d'une clé de caractérisation "SITE" et d'un qualificateur "NOTE" avec le nom complet non abrégé de l'acide aminé correspondant.

Nouveaux types de molécules

Acides aminés D

```

- <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>

```

Note : une analyse détaillée d'un exemple similaire figure dans la norme ST.26 de l'OMPI, annexe VI, exemple 3.a)-1

Nouveaux types de molécules

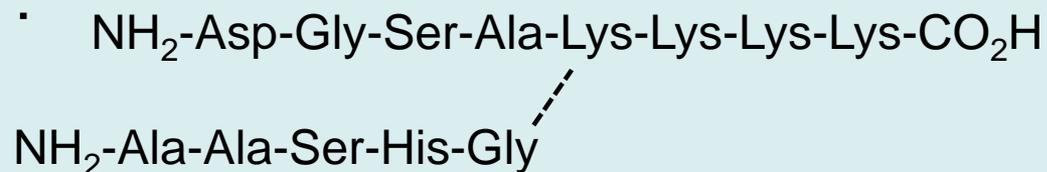
Séquences ramifiées

- Les séquences d'acides nucléiques ramifiées et les séquences d'acides aminés ramifiées sont soumises aux règles de la norme ST.26
- Les régions linéaires de séquences ramifiées contenant au moins 10 nucléotides définis de manière spécifique ou au moins 4 acides aminés définis de manière spécifique doivent être incluses dans un listage de séquences. (Norme ST.26 de l'OMPI, paragraphe 7)
- Chaque région linéaire d'une séquence ramifiée qui satisfait à l'exigence de longueur minimale doit être incluse en tant que séquence distincte et disposer de son propre numéro d'identification de séquence
- Il convient de prendre en compte le nombre de résidus définis de manière spécifique dans chaque région linéaire et non le nombre total de résidus définis de manière spécifique dans la structure

Nouveaux types de molécules

Séquences ramifiées

Une demande de brevet décrit un peptide au moyen de la séquence suivante :



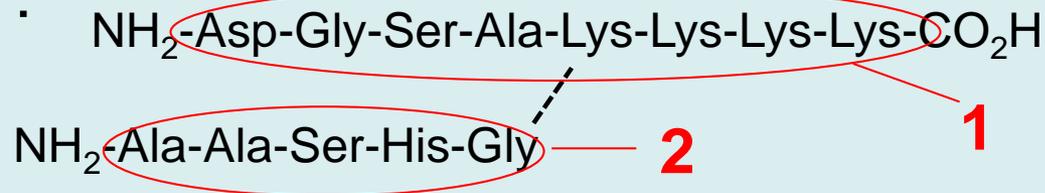
Où ----- indique une liaison amide entre la terminaison carboxyle de la glycine et la chaîne latérale de la lysine

- ✓ Les deux régions linéaires contiennent plus de 4 acides aminés définis de manière spécifique, de sorte que toutes deux doivent être incluses dans le listage des séquences.

Nouveaux types de molécules

Séquences ramifiées

Une demande de brevet décrit un peptide au moyen de la séquence suivante :



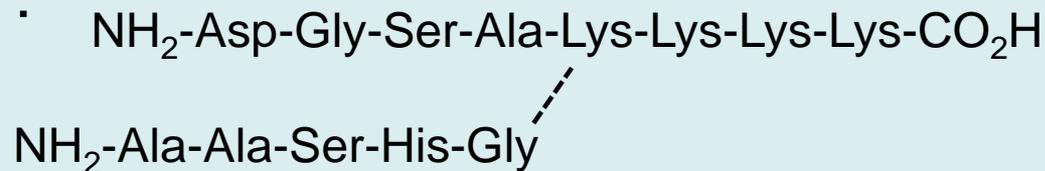
Où ----- indique une liaison amide entre la terminaison carboxyle de la glycine et la chaîne latérale de la lysine

- ✓ Les deux régions linéaires contiennent plus de 4 acides aminés définis de manière spécifique, de sorte que toutes deux doivent être incluses dans le listage des séquences.
- ✓ Chaque région linéaire doit apparaître en tant que séquence distincte et disposer de son propre numéro d'identification de séquence.

Nouveaux types de molécules

Séquences ramifiées

Une demande de brevet décrit un peptide au moyen de la séquence suivante :



Où ----- indique une liaison amide entre la terminaison carboxyle de la glycine et la chaîne latérale de la lysine

- ✓ Les deux régions linéaires contiennent plus de 4 acides aminés définis de manière spécifique, de sorte que toutes deux doivent être incluses dans le listage des séquences.
- ✓ Chaque région linéaire doit apparaître en tant que séquence distincte et disposer de son propre numéro d'identification de séquence.
- ✓ Les deux séquences devraient être annotées pour indiquer l'emplacement et la nature de la liaison aminée.

Nouveaux types de molécules

Séquences ramifiées

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>

```

Note : une analyse détaillée d'un exemple similaire figure dans la norme ST.26 de l'OMPI, annexe VI, exemple 7.b)-3

Variantes de séquence

Variantes de séquence

- Le paragraphe 3.m) de la norme ST.26 définit une “**séquence variante**” comme

*“une séquence de nucléotides ou d’acides aminés qui présente **une ou plusieurs différences par rapport à une séquence primaire**. Ces différences peuvent être des résidus alternatifs (voir les paragraphes 15 et 27), des résidus modifiés (voir les paragraphes 3.g), 3.h), 16 et 29), des suppressions, des adjonctions ou des remplacements. Voir les paragraphes 93 à 95.”*

- La manière dont une séquence variante est divulguée déterminera la façon dont elle doit être représentée dans le listage des séquences.
- Les paragraphes 93 à 95 de la norme ST.26 indiquent comment les variantes doivent être représentées.

Variantes de séquence

Paragraphe 93

Paragraphe 93 : Toute séquence primaire et toute variante de cette séquence, chacune d'elles étant divulguée par énumération de ses résidus et visée par le paragraphe 7, doit figurer dans le listage des séquences et doit disposer de son propre numéro d'identification de séquence.



Si chaque variante est citée séparément, alors chaque variante doit disposer de son propre numéro d'identification de séquence!

Variantes de séquence

Paragraphe 93

Une demande de brevet comprend une figure présentant l'alignement séquentiel multiple suivant :

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

Les lettres minuscules représentent les résidus d'acides aminés prédominants parmi les séquences alignées.

- ✓ Chacune des 6 séquences citées doit apparaître dans le listage des séquences en tant que séquence distincte et disposer de son propre numéro d'identification de séquence.

Note : une analyse détaillée d'un exemple similaire figure dans la norme ST.26 de l'OMPI, annexe VI, exemple 93-3

Variantes de séquence

Annotation appropriée

- La clé de caractérisation pour annoter une variante dépend du type de molécule et de la nature de la variante :

Type de séquence	Clé de caractérisation	Qualificateur	Usage
Acide nucléique	variation	replace or note	Mutations et polymorphismes existant à l'état naturel, p. ex. des allèles ou des polymorphismes de longueur des fragments de restriction
Acide nucléique	misc_difference	replace or note	La variabilité a été créée artificiellement, p. ex. par une manipulation génétique ou une synthèse chimique
Acide aminé	VAR_SEQ	NOTE	La variante a été produite par un épissage alternatif, l'usage de promoteurs alternatifs, une initiation alternative et un déphasage ribosomique
Acide aminé	VARIANT	NOTE	Tout type de variante pour laquelle VAR_SEQ n'est pas applicable

Norme ST.26 de l'OMPI, paragraphe 96

Variantes de séquence

Symbole ambigu le plus restrictif

- Les paragraphes 15 à 27 de la norme ST.26 indiquent que lorsqu'un symbole ambigu est nécessaire, "il faut choisir le symbole le plus restrictif..."

Qu'est-ce que cela veut dire?

Par exemple :

Dans une séquence de nucléotides, si une position peut être "a ou c", il faut employer le symbole "m" au lieu de "n".

Dans une séquence d'acide aminé, si une position peut être "L ou I", il faut employer le symbole "J" au lieu de "X".

Ne pas oublier que "n" et "X" ont des valeurs par défaut, de sorte qu'à chaque fois que l'on utilise "n" ou "X" pour une chose autre que la valeur par défaut, une annotation s'impose.

Variantes de séquence

Paragraphe 94

Paragraphe 94 : Toute séquence variante, divulguée comme une séquence unique avec des résidus alternatifs énumérés à une ou plusieurs positions, doit figurer dans le listage des séquences et devrait être représentée par une séquence unique, les résidus alternatifs énumérés étant représentés par le symbole le plus restrictif (voir les paragraphes 15 et 27).



Si les variantes NE SONT PAS citées séparément, mais sont simplement indiquées en tant que résidus variables dans la séquence primaire, elles n'ont alors pas besoin de disposer d'un numéro d'identification de séquence distinct!

Variantes de séquence

Paragraphe 94

Une demande de brevet divulgue un peptide de la séquence :

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

- ✓ Peut apparaître dans un listage de séquences en tant que séquence unique
- ✓ La représentation privilégiée est la suivante : GGGJATX
- ✓ [Leu or Ile] devrait être représenté par le symbole ambigu le plus restrictif “J”
- ✓ [Ser or Thr] devrait être représenté par le symbole “X” avec une clé de caractérisation “VARIANT” et un qualificateur note pour indiquer que “X” est la sérine ou la thréonine.

Note : une analyse détaillée de cet exemple figure dans la norme ST.26 de l'OMPI, annexe VI, exemple 94-1

Variantes de séquence

Paragraphe 95

Paragraphe 95 : Toute séquence variante, divulguée uniquement par référence à un ou plusieurs suppressions, adjonctions ou remplacements effectués dans une séquence primaire figurant dans le listage des séquences, doit figurer dans le listage des séquences. Si tel est le cas, cette séquence variante :

a) peut être représentée par annotation de la séquence primaire, si elle comporte une ou plusieurs variations à un seul emplacement ou à plusieurs emplacements distincts et que les occurrences de ces variations sont indépendantes;

b) devrait être représentée en tant que séquence distincte et devrait disposer de son propre numéro d'identification de séquence, si elle comporte des variations à plusieurs emplacements distincts et que les occurrences de ces variations sont interdépendantes; et

c) doit être représentée en tant que séquence distincte et doit disposer de son propre numéro d'identification de séquence, si elle comporte une séquence qui a été ajoutée ou remplacée et qui contient plus de 1 000 résidus (voir le paragraphe 86)



“par référence à un ou plusieurs suppressions, adjonctions ou remplacements” signifie que les variantes sont indiquées en prose.

Variantes de séquence

Paragraphe 95.a)

Une demande contient la divulgation suivante :

Fragment peptidique 1 : Gly-Leu-Pro-Xaa-Arg-Ile-Cys
où Xaa peut représenter tout acide aminé

* * *

Dans un autre mode de réalisation, le fragment peptidique 1 est Gly-Leu-Pro-Xaa-Arg-Ile-Cys, où Xaa peut représenter Val, Thr ou Asp...

* * *

dans un autre mode de réalisation, le fragment peptidique 1 est Gly-Leu-Pro-Xaa-Arg-Ile-Cys, où Xaa peut représenter Val...

La séquence primaire, Gly-Leu-Pro-Xaa-Arg-Ile-Cys, comporte une ou plusieurs variations à un seul emplacement et les occurrences de ces variations sont indépendantes

Variantes de séquence

Paragraphe 95.a)

- ✓ La séquence primaire doit apparaître dans le listage des séquences et les variantes peuvent être représentées par une annotation de la séquence primaire
- ✓ Le mode de réalisation le plus englobant doit être la version qui figure dans le listage des séquences; dans cet exemple, il s'agit de la version où Xaa représente "tout acide aminé"
- ✓ La séquence doit être représentée sous la forme GLPXRIC et nécessite la clé de caractérisation "VARIANT" à la position 4 assortie d'un qualificateur "note" indiquant que Xaa représente "tout acide aminé"
- ✓ Bien que cela ne soit pas obligatoire, il est recommandé d'intégrer les trois variantes suivantes dans le listage des séquences en tant que séquences distinctes :

GLPVRIC

GLPTRIC

GLPDRIC

Note : une analyse détaillée de cet exemple figure dans la norme ST.26 de l'OMPI, annexe VI, exemple 95.a)-1.

Variantes de séquence

Paragraphe 95.b)

Une demande de brevet décrit la séquence consensus suivante :

aatg n_1 cccacgaatg n_2 cac

où n_1 et n_2 peuvent être a, t, g, ou c.

Plusieurs variantes de séquences sont divulguées comme suit :

si n_1 est a, alors n_2 est t, g, ou c

si n_1 est t, alors n_2 est a, g, ou c

si n_1 est g, alors n_2 est t, a, ou c

si n_1 est c, alors n_2 est t, g, ou a

La séquence consensus contient des variations à plusieurs endroits distincts et les occurrences de ces variations sont interdépendantes.

Variantes de séquence

Paragraphe 95.b)

- ✓ La séquence consensus doit apparaître dans le listage des séquences et les variantes devraient être représentées en tant que séquence distincte
- ✓ Le mode de réalisation le plus englobant doit être la version qui figure dans le listage des séquences, où n_1 et n_2 “peuvent être a, t, g, ou c”
- ✓ La séquence doit être représentée ainsi : `aatgncccacgaatgncac`
- ✓ Il n’est pas besoin d’annotation pour n, puisque “n” sera interprété comme équivalent à l’un des symboles “a”, “c”, “g” en l’absence de clé de caractérisation (voir le paragraphe 15 de la norme ST.26)
- ✓ Bien que cela ne soit pas obligatoire, il est recommandé d’intégrer les quatre variantes suivantes dans le listage des séquences en tant que séquences distinctes :

`aatgacccacgaatgncac`

(b = t, g, ou c)

`aatgtcccacgaatgvcac`

(v = a, g, ou c)

`aatggcccacgaatgnhcac`

(h = t, a, ou c)

`aatgccccacgaatgndcac`

(d = t, g, ou a)

Note : une analyse détaillée d’un exemple similaire figure dans la norme ST.26 de l’OMPI, annexe VI, exemple 95.b).

Variantes de séquence

Paragraphe 95.c)

Une demande contient la divulgation suivante :

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

où Waa est Gly ou une adjonction de la séquence

Cys-Tyr-Ile-Lys-Ser-(1000 amino acids)-Leu-Thr-Pro-Lys

Une variante de la séquence comporte une séquence ajoutée ou remplacée et qui contient plus de 1 000 résidus.

Variantes de séquence

Paragraphe 95.c)

- ✓ La variante ou Xaa est une adjonction de plus de 1 000 résidus doit apparaître dans le listage des séquences en tant que séquence distincte et disposer de son propre numéro d'identification de séquence.
- ✓ La variante ou Xaa = Gly apparaîtra également dans le listage des séquences en tant que séquence distincte et disposera de son propre numéro d'identification de séquence.

...-MGLPRGRICK-...

Variantes de séquence

Q : Quel paragraphe s'applique à la divulgation suivante?

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

Variantes de séquence

Q : Quel paragraphe s'applique à la divulgation suivante?

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

A : Paragraphe 94 :

“Toute séquence variante, divulguée comme une séquence unique avec des résidus alternatifs énumérés à une ou plusieurs positions, doit figurer dans le listage des séquences et devrait être représentée par une séquence unique, les résidus alternatifs énumérés étant représentés par le symbole le plus restrictif.”

Variantes de séquence

Q : Quel paragraphe s'applique à la divulgation suivante?

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

Dans le tableau, un espace blanc indique qu'un acide aminé dans la variante est le même que l'acide aminé dans la "Séquence" et un "-" indique la suppression de l'acide aminé correspondant dans la "Séquence".

Variantes de séquence

Q : Quel paragraphe s'applique à la divulgation suivante?

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							-		

Dans le tableau, un espace indique qu'un acide aminé dans la variante est le même que l'acide aminé dans la "Séquence" et un "-" indique la suppression de l'acide aminé correspondant dans la "Sequence".

A : Paragraphe 93

“Toute séquence primaire et toute variante de cette séquence, chacune d'elles étant divulguée par énumération de ses résidus et visée par le paragraphe 7, doit figurer dans le listage des séquences et doit disposer de son propre numéro d'identification de séquence.”

Variantes de séquence

Q : Quel paragraphe s'applique à la divulgation suivante?

Une figure représente la séquence suivante :

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

La spécification contient les informations suivantes concernant la Séquence 1 :

... dans certains modes de réalisation, la valine à la position 7 de la séquence 1 est remplacée par l'alanine...

Variantes de séquence

Q : Quel paragraphe s'applique à la divulgation suivante?

Une figure représente la séquence suivante :

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

La spécification contient les informations suivantes concernant la Séquence 1 :

... dans certains modes de réalisation, la valine à la position 7 de la séquence 1 est remplacée par l'alanine...

A : Paragraphe 95.a) : *“Toute séquence variante, divulguée uniquement par référence à un ou plusieurs suppressions, adjonctions ou remplacements effectués dans une séquence primaire figurant dans le listage des séquences, doit figurer dans le listage des séquences. Si tel est le cas, cette séquence variante :*

a) peut être représentée par annotation de la séquence primaire, si elle comporte une ou plusieurs variations à un seul emplacement ou à plusieurs emplacements distincts et que les occurrences de ces variations sont indépendantes,”

Variantes de séquence

Q : Quel paragraphe s'applique à la divulgation suivante?

Une figure représente la séquence suivante :

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

La spécification contient les informations suivantes concernant la Séquence 1 :

... si la valine à la position 7 de la séquence 1 est remplacée par l'alanine, alors l'alanine à la position 10 est remplacée par la valine...

Variantes de séquence

Q : Quel paragraphe s'applique à la divulgation suivante?

Une figure représente la séquence suivante :

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

La spécification contient les informations suivantes concernant la Séquence 1 :

... si la valine à la position 7 de la séquence 1 est remplacée par l'alanine, alors l'alanine à la position 10 est remplacée par la valine...

A : Paragraphe 95.b) : *Toute séquence variante, divulguée uniquement par référence à un ou plusieurs suppressions, adjonctions ou remplacements effectués dans une séquence primaire figurant dans le listage des séquences, doit figurer dans le listage des séquences. Si tel est le cas, cette séquence variante :*

b) devrait être représentée en tant que séquence distincte et devrait disposer de son propre numéro d'identification de séquence, si elle comporte des variations à plusieurs emplacements distincts et que les occurrences de ces variations sont interdépendantes;

Questions?