



OMPI

ORGANISATION MONDIALE
DE LA **PROPRIÉTÉ**
INTELLECTUELLE

Norme ST.26 de l'OMPI PRÉSENTATION

Webinaire de formation

Thèmes traités aujourd'hui

- Qu'est-ce qu'un listage des séquences?
- Bases de données de séquences de l'INSDC
- Pourquoi une nouvelle norme?
- Avantages de la norme ST.26 de l'OMPI
- Différences entre les normes ST.25 et ST.26
- Transition vers la norme ST.26
- Fondamentaux du XML
- Principes de base de la norme ST.26 – Éléments d'un listage des séquences
- Table des matières de la norme ST.26 de l'OMPI
- Présentation de WIPO Sequence

Qu'est-ce qu'un listage des séquences?

Un listage des séquences:

- contient les séquences de nucléotides ou d'acides aminés divulguées dans une demande de brevet et fait partie de la description;
- comprend des informations descriptives sur chaque séquence, appelées annotations;
- est conforme aux exigences de la norme OMPI applicable (ST.25 ou ST.26);
- permet d'effectuer des recherches sur les données de séquence d'une invention:
 - au sein d'un office de propriété intellectuelle
 - dans les bases de données publiques (bases de données de l'INSDC)

Informations générales concernant l'INSDC

- INSDC: Collaboration internationale sur les bases de données de séquences de nucléotides:
 - DDBJ: Banque de données ADN du Japon
 - EMBL-EBI: Institut européen de bioinformatique
 - NCBI: Centre national d'information sur la biotechnologie des États-Unis d'Amérique (GenBank)

- Les offices de propriété intellectuelle ci-dessous alimentent les bases de données de l'INSDC avec des données de séquences de demandes de brevets publiées ou de brevets délivrés:
 - Office européen des brevets
 - Office japonais des brevets
 - Office coréen de la propriété intellectuelle
 - Office des brevets et des marques des États-Unis

- Les bases de données de l'INSDC sont à la disposition du public.

Pourquoi une nouvelle norme ?

- Actuellement, les listages de séquences sont déposés conformément à la norme ST.25 de l'OMPI. Pourtant...
- Le format défini par la norme ST.25 n'est pas conforme aux exigences de l'INSDC, si bien que certaines données se perdent lors de leur saisie dans des bases de données publiques.
- Les règles prévues par la norme ST.25 ne sont pas claires; elles sont interprétées et appliquées de diverses manières par les offices de propriété intellectuelle à travers le monde.
- Certains types de séquences courants aujourd'hui (analogues nucléotidiques, acides aminés D, séquences ramifiées) ne sont pas prévus dans les règles de la norme ST.25 et ne figurent donc pas dans les bases de données consultables.
- Les données ne sont pas structurées. Le format défini par la norme ST.25 est difficile à utiliser pour une validation et un échange de données automatisés.

Avantages de la norme ST.26 de l'OMPI (1)

- Acceptation d'un listage des séquences unique à l'échelle mondiale*
- La norme fait office de guide pour permettre aux offices de propriété intellectuelle de s'accorder sur l'application des règles relatives aux séquences.
- Elle précise quelles informations relatives aux séquences doivent ou peuvent figurer dans un listage des séquences et la manière de les présenter.
- Elle améliore la qualité de la présentation grâce à la structure des listages de séquences au format XML.
- Elle permet d'automatiser davantage la validation des données et de rationaliser le traitement par les offices de propriété intellectuelle.

* excepté pour les traductions requises des qualificatifs de texte libre dépendant de la langue dans la langue de dépôt pour certains offices de propriété intellectuelle, qui pourraient nécessiter des listages de séquences de remplacement.

Avantages de la norme ST.26 de l'OMPI (2)

- Compatibilité des données avec les exigences des fournisseurs des bases de données de l'INSDC (DDBJ, EBI et NCBI). Les annotations des séquences (clés et qualificateurs) seront incluses dans des bases de données consultables par le public.
- Normalisation:
 - des annotations relatives aux caractéristiques
 - des emplacements des caractéristiques
 - des qualificateurs et des valeurs de qualificateurs
 - de la présentation des variantes de séquences
- L'obligation d'inclure des types de séquences supplémentaires (analogues nucléotidiques, acides aminés D, séquences ramifiées) permet d'effectuer des recherches sur un plus grand volume de données de séquences.

Différences entre les normes ST.25 et ST.26

ST.25	ST.26
ASCII .txt avec identifiants numériques	XML avec éléments et attributs
<p><u>Ne doit pas nécessairement</u> inclure:</p> <ul style="list-style-type: none"> - les acides aminés D - les portions linéaires de séquences ramifiées - les analogues nucléotidiques 	<p><u>Doit impérativement</u> inclure:</p> <ul style="list-style-type: none"> - les acides aminés D - les portions linéaires de séquences ramifiées - les analogues nucléotidiques
<p>Annotation des séquences:</p> <ul style="list-style-type: none"> - clés de caractérisation uniquement 	<p>Annotation des séquences:</p> <ul style="list-style-type: none"> - clés de caractérisation et qualificateurs
<p><u>Possibilité</u> d'inclure des séquences:</p> <ul style="list-style-type: none"> - < 10 nucléotides définis de manière spécifique - < 4 acides aminés définis de manière spécifique 	<p><u>Interdiction</u> d'inclure des séquences:</p> <ul style="list-style-type: none"> - < 10 nucléotides définis de manière spécifique - < 4 acides aminés définis de manière spécifique

Différences entre les normes ST.25 et ST.26 dans la partie consacrée aux informations générales

ST.25	ST.26
TOUTES les informations relatives à la revendication de priorité peuvent être indiquées	SEULE la revendication de priorité la plus ancienne peut être indiquée
Les noms de TOUS les déposants et inventeurs peuvent être indiqués	Seul UN demandeur ET éventuellement UN inventeur peuvent être indiqués.
Un seul titre d'invention autorisé	Plusieurs titres d'invention autorisés, chacun dans une langue différente
Les noms des déposants et des inventeurs et les titres des inventions doivent être en caractères latins de base.	Les noms des déposants et des inventeurs et les titres des inventions peuvent être dans n'importe quels caractères Unicode valides accompagnés d'une traduction ou d'une translittération en caractères latins de base.

Différences entre les normes ST.25 et ST.26 dans la partie consacrée aux données des séquences (1)

ST.25	ST.26
Séquences identifiées comme ADN, ARN ou PRT uniquement	Séquences identifiées comme ADN, ARN ou AA avec un qualificateur obligatoire "mol_type" pour décrire plus précisément la molécule.
Noms des organismes : <ul style="list-style-type: none"> - nom de genre ou d'espèce en latin, - nom du virus - "artificial sequence" - "unknown" 	Noms des organismes : <ul style="list-style-type: none"> - nom de genre ou d'espèce en latin, - nom du virus - "synthetic construct" - "unidentified"
Le symbole "u" désigne l'uracile dans les séquences de nucléotides.	Le symbole "t" désigne l'uracile dans les séquences d'ARN et la thymine dans les séquences d'ADN.
Séquences d'acides aminés représentées par des abréviations de trois lettres	Séquences d'acides aminés représentées par des abréviations d'une lettre

Différences entre les normes ST.25 et ST.26 dans la partie consacrée aux données des séquences (2)

ST.25	ST.26
Les variables “n” et “Xaa” doivent faire l’objet d’une définition dans une caractéristique.	La valeur des variables “n” et “X” est établie par défaut, sans définition.
Le format à utiliser pour l’emplacement des caractéristiques n’est pas clairement défini.	Les formats à utiliser pour l’emplacement des caractéristiques sont strictement définis et permettent l’utilisation de “<” et “>” dans tous les types de séquence, et de “^”, “join”, “order” et “complement” dans les séquences de nucléotides.
Séquences en “mode mixte” autorisées. Séquence nucléotidique suivie d’une traduction en acides aminés en dessous.	PAS de “mode mixte”. Les traductions de nucléotides ne sont incluses que dans les qualificatifs “translation”.

Transition vers la norme ST.26

- À sa cinquième session, le CWS a convenu d'une date de transition au **1^{er} janvier 2022**, dite "**date du big-bang**".
- **Tous** les offices de propriété intellectuelle effectueront la transition simultanément aux niveaux national, régional et international (PCT).
- La date de dépôt international sera la date de référence qui détermine si une demande relève des règles relatives aux séquences de la norme ST.25 ou de la norme ST.26, et **NON** la date de priorité.
- Remarque: La norme ST.25 restera en vigueur pour les demandes dont la date de dépôt est antérieure au 1^{er} janvier 2022.

La norme ST.26 de l'OMPI

...que doit-on inclure ?

- séquences de nucléotides:
 - 10 résidus ou plus “spécialement définis” et “énumérés”*
 - comprenant des séquences avec des analogues nucléotidiques tels que des acides nucléiques peptidiques (ANP) et des acides nucléiques à glycol (ANG)
- Séquences d'acides aminés:
 - 10 résidus ou plus “spécialement définis” et “énumérés”
 - comprenant des séquences d'acides aminés D
 - les régions linéaires de séquences ramifiées doivent être incluses dans un listage des séquences.

La norme ST.26 de l'OMPI

... que doit-on inclure?

- Qu'est-ce qu'un nucléotide ou un acide aminé "spécialement défini"?
- "spécialement défini" désigne tout nucléotide différent de ceux qui sont représentés par le symbole "n" et tout acide aminé différent de ceux qui sont représentés par le symbole "X" dans l'annexe I. (Paragraphe 3(m) de la norme ST.26 de l'OMPI)
- seuls les résidus "spécialement définis" comptent pour la longueur minimale requise:
 - 10 nucléotides spécialement définis ou plus; ou
 - 4 acides aminés spécialement définis ou plus.

5'- anctggcaan – 3' seulement 8 nucléotides spécialement définis; ne doit pas être inclus dans un listage de séquence.

5'- agctggcaat – 3' dix nucléotides spécialement définis; doit être inclus dans un listage de séquence.

Norme ST.26 de l'OMPI: document XML

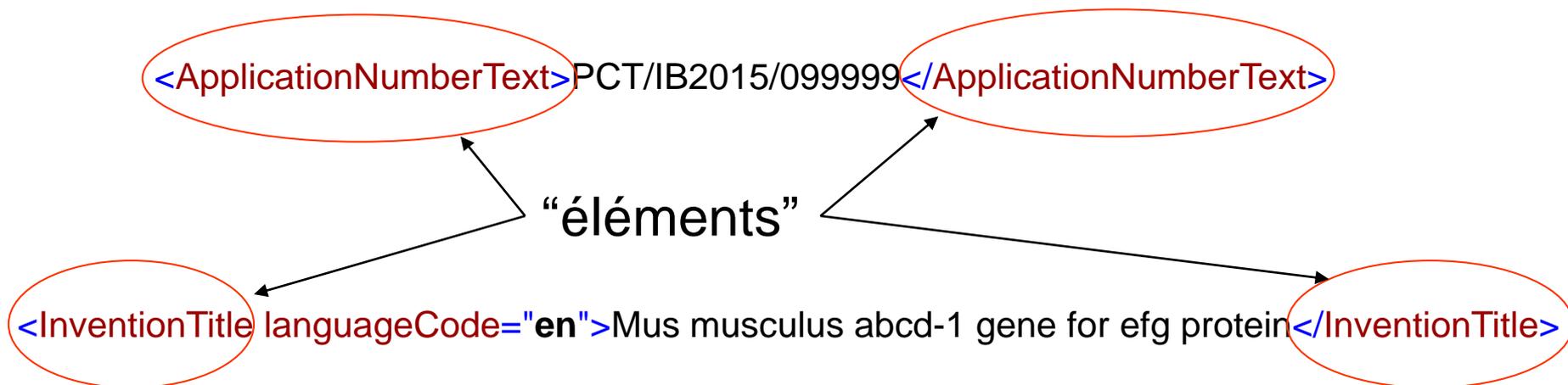
- Doit être présenté dans un fichier unique au format XML 1.0.
- Doit être conforme à la définition de type de document (DTD) présentée dans l'annexe II et aux règles opérationnelles dérivées du contenu de la norme.
- Doit être encodé à l'aide du standard Unicode UTF-8.
- Structure du listage de séquence:
 - déclaration du format XML
 - déclaration de type de document (DOCTYPE)
 - élément racine
 - partie consacrée aux informations générales
 - partie consacrée aux données sur les séquences

Fondamentaux du XML

- XML = *eXtensible Markup Language*
- Les informations sont “étiquetées” au moyen d’éléments et d’attributs descriptifs.
- Système normalisé d’échange de données, déchiffrable par l’homme et par la machine
- DTD = définition de type de document – définit la structure et les éléments et attributs juridiques d’un document XML

Fondamentaux du XML

Éléments, attributs et valeurs (1)



Fondamentaux du XML

Éléments, attributs et valeurs (2)

`<ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>`

“valeurs des éléments”

`<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>`

Fondamentaux du XML

Éléments, attributs et valeurs (3)

```
<ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
```

“attribut”

```
<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
```

“valeur de l’attribut”

Fondamentaux du XML

Caractères réservés

Doivent être remplacés par les entités prédéfinies correspondantes

Reserved Character	Predefined Entities
<	<
>	>
&	&
“	"
'	'

Exemple: l'emplacement de la caractéristique souhaité est “<50..62”

`<INSDFeature_location><50..62</INSDFeature_location>` 

`<INSDFeature_location><50..62</INSDFeature_location>` 

Norme ST.26 de l'OMPI: Exemple

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN" "ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="for WIPO Training" softwareName="WIPO Sequence" softwareVersion="1.1.0-beta4" productionDate="2021-03-24">
  <ApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
    <FilingDate>2015-01-31</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>ABC123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
    <FilingDate>2014-01-31</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantName>
  <InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
  <SequenceTotalQuantity>1</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1">
    <INSDSeq>
      <INSDSeq_length>52</INSDSeq_length>
      <INSDSeq_moltype>DNA</INSDSeq_moltype>
      <INSDSeq_division>PAT</INSDSeq_division>
      <INSDSeq_feature-table>
        <INSDFeature>
          <INSDFeature_key>source</INSDFeature_key>
          <INSDFeature_location>1..52</INSDFeature_location>
          <INSDFeature_qual>
            <INSDQualifier>
              <INSDQualifier_name>mol_type</INSDQualifier_name>
              <INSDQualifier_value>genomic DNA</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier id="q2">
              <INSDQualifier_name>organism</INSDQualifier_name>
              <INSDQualifier_value>Mus musculus</INSDQualifier_value>
            </INSDQualifier>
          </INSDFeature_qual>
        </INSDFeature>
      </INSDSeq_feature-table>
      <INSDSeq_sequence>atgaaattaaacataaaarggatgataaaatgagatttgatataaaaagg</INSDSeq_sequence>
    </INSDSeq>
  </SequenceData>
</ST26SequenceListing>

```

Norme ST.26 de l'OMPI : Composantes (1)

Première ligne: déclaration du format XML

(paragraphe 39 (a) de la norme ST.26)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN" "ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="for WIPO Training" softwareName="WIPO Sequence"
softwareVersion="1.1.0-beta4" productionDate="2021-03-24">
  <ApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
    <FilingDate>2015-01-31</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>ABC123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
    <FilingDate>2014-01-31</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantName>
  <InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
  <SequenceTotalQuantity>1</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1">
```

Norme ST.26 de l'OMPI : Composantes (2)

Deuxième ligne: déclaration de type de document (DOCTYPE)

(paragraphe 39(b) de la norme ST.26)

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN" "ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="for WIPO Training" softwareName="WIPO Sequence"
softwareVersion="1.1.0-beta4" productionDate="2021-03-24">
  <ApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
    <FilingDate>2015-01-31</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>ABC123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
    <FilingDate>2014-01-31</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantName>
  <InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
  <SequenceTotalQuantity>1</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1">

```

Norme ST.26 de l'OMPI : Composantes (3)

Troisième ligne: élément racine
(paragraphe 43 de la norme ST.26)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN" "ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="for WIPO Training" softwareName="WIPO Sequence"
softwareVersion="1.1.0-beta4" productionDate="2021-03-24">
```

```
<ApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
  <FilingDate>2015-01-31</FilingDate>
</ApplicationIdentification>
<ApplicantFileReference>ABC123</ApplicantFileReference>
<EarliestPriorityApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
  <FilingDate>2014-01-31</FilingDate>
</EarliestPriorityApplicationIdentification>
<ApplicantName languageCode="en">Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantName>
<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
<SequenceTotalQuantity>1</SequenceTotalQuantity>
<SequenceData sequenceIDNumber="1">
```

Informations générales
(paragraphe 38(a) et
45 à 49 de la norme ST.26)

Norme ST.26 de l'OMPI : Composantes (4)

```

<SequenceData sequenceIDNumber="1">
  <INSDSeq>
    <INSDSeq_length>52</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..52</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>genomic DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q2">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>Mus musculus</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>atgaaattaaaacataaaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
</ST26SequenceListing>

```

Données des séquences

*(paragraphes 38(b) et
50 à 100 de la norme ST.26)*

Norme ST.26 : Informations générales (1)

- Section sur l'identification de la demande
 - le numéro de la demande, la date de dépôt et le code de l'office de propriété intellectuelle sont obligatoires, s'ils sont connus;
 - autrement, l'identificateur de la demande du déposant suffit.

```
<ApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
  <FilingDate>2015-01-31</FilingDate>
</ApplicationIdentification>
<ApplicantFileReference>ABC123</ApplicantFileReference>
<EarliestPriorityApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
  <FilingDate>2014-01-31</FilingDate>
</EarliestPriorityApplicationIdentification>
```

Norme ST.26 : Informations générales (2)

- Section sur la demande établissant la priorité
 - une seule demande établissant la priorité peut être incluse dans le listage des séquences; il doit s'agir de la demande de priorité la plus ancienne;
 - obligatoire si une priorité est revendiquée.

```
<ApplicationIdentification>  
  <IPOfficeCode>IB</IPOfficeCode>  
  <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>  
  <FilingDate>2015-01-31</FilingDate>  
</ApplicationIdentification>  
<ApplicantFileReference>ABC123</ApplicantFileReference>  
<EarliestPriorityApplicationIdentification>  
  <IPOfficeCode>IB</IPOfficeCode>  
  <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>  
  <FilingDate>2014-01-31</FilingDate>  
</EarliestPriorityApplicationIdentification>
```

Norme ST.26 : Informations générales (3)

- Section sur le nom du déposant et le nom de l'inventeur
 - un seul nom de déposant et un seul nom d'inventeur peuvent figurer dans le listage des séquences; il doit s'agir du premier déposant et du premier inventeur mentionnés;
 - le nom du déposant est obligatoire; le nom de l'inventeur est facultatif;
 - un code de langue pour les noms des déposants et des inventeurs est obligatoire;
 - si le nom du déposant ou de l'inventeur contient des caractères latins de base non Unicode, une translittération ou une traduction en caractères latins de base doit être fournie.

```
<ApplicantName languageCode="ja">出願製薬株式会社</ApplicantName>
<ApplicantNameLatin>Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantNameLatin>
<InventorName languageCode="ja">特許 太郎</InventorName>
<InventorNameLatin>Taro Tokkyo</InventorNameLatin>
```

Norme ST.26 : Informations générales (4)

- Section sur le titre de l'invention
 - il est obligatoire d'indiquer au moins un titre d'invention dans la langue de dépôt;
 - des titres supplémentaires dans d'autres langues peuvent être ajoutés;
 - un code de langue est obligatoire pour chaque titre.

```
<InventionTitle languageCode="en">My spectacular invention</InventionTitle>  
<InventionTitle languageCode="de">Meine spektakuläre Erfindung</InventionTitle>  
<InventionTitle languageCode="lv">Mans iespaidigais izgudrojums</InventionTitle>  
<InventionTitle languageCode="ru">Мое зрелищное изобретение</InventionTitle>
```

Norme ST.26 : Informations générales (5)

- Élément relatif au nombre total des séquences apparaissant dans le listage:
 - obligatoire;
 - le total doit comprendre les séquences omises.

<SequenceTotalQuantity>**6**</SequenceTotalQuantity>

Norme ST.26 : Séquences (1)

- Éléments obligatoires

Le numéro d'identification de la séquence
ou "SEQ ID NO"

La longueur de la séquence

Le type de molécule
(ADN, ARN ou AA)

La division
(toujours "PAT")

```

<SequenceData sequenceIDNumber="1">
  <INSDSeq>
    <INSDSeq_length>52</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..52</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>genomic DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q2">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>Mus musculus</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>atgaaattaaaacataaaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

Norme ST.26 : Séquences (2)

- La clé de caractérisation “source” ou “SOURCE” est obligatoire.

```

<SequenceData sequenceIDNumber="1">
  <INSDSeq>
    <INSDSeq_length>52</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..52</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>genomic DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q2">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>Mus musculus</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>atgaaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

Seule une clé de caractérisation “source” est requise pour chaque séquence.

La clé “source” doit couvrir la séquence tout entière.

Deux qualificateurs obligatoires: “mol_type” et “organism”.

Norme ST.26 : Séquences (3)

Norme ST.26: Données des séquences

```

<SequenceData sequenceIDNumber="1">
  <INSDSeq>
    <INSDSeq_length>52</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..52</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>genomic DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q2">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>Mus musculus</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>atgaaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

Norme ST.26 : Séquences (4)

- Choix de valeurs pour les qualificateurs obligatoires “mol_type” et “MOL_TYPE”:

ADN

genomic DNA

other DNA

unassigned DNA

ARN

genomic RNA

mRNA

tRNA

rRNA

other RNA

transcribed RNA

viral cRNA

unassigned RNA

AA

protein

Norme ST.26 : Séquences (5)

- Choix de valeurs pour les qualificateurs obligatoires “organism” et “ORGANISM”:
 - nom de genre ou d’espèce en latin, p. ex. “Mus musculus”
 - nom de genre suivi de “sp.” e.g., “Mus sp.”
 - nom du virus, p. ex. “Torque teno virus 1”
 - “unidentified”
 - “synthetic construct”
- Les noms communs, comme “mouse”, ne doivent pas être utilisés pour désigner un organisme. Il est possible d’indiquer les noms communs dans le listage des séquences à l’aide d’un qualificateur “note”.

Norme ST.26 : Séquences (6)

Clés de caractérisation et qualificateurs

Outre la clé de caractérisation obligatoire “source” ou “SOURCE”, les déposants peuvent ajouter plusieurs clés de caractérisation facultatives pour décrire plus précisément la séquence:

- des clés de caractérisation différentes pour les séquences de nucléotides et les séquences d’acides aminés;
- chaque clé de caractérisation peut avoir un ou plusieurs qualificateurs facultatifs et un qualificateur obligatoire.

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 de l’OMPI, Annexe I, Sections 5 à 8)

Norme ST.26 : Séquences (7)

- séquences de nucléotides:
 - tous les symboles en lettres minuscules;
 - pas d’espaces ni de numérotation;
 - pas de symbole “u”; le “t” représente l’uracile dans l’ARN
 - le symbole “n” sera considéré comme “équivalent à l’un des symboles ‘a’, ‘c’, ‘g’ ou ‘t/u’”.

Symbol	Nucleotide
a	adenine
c	cytosine
g	guanine
t	thymine in DNA/uracil in RNA (t/u)
m	a or c
r	a or g
w	a or t/u
s	c or g
y	c or t/u
k	g or t/u
v	a or c or g; not t/u
h	a or c or t/u; not g
d	a or g or t/u; not c
b	c or g or t/u; not a
n	a or c or g or t/u; “unknown” or “other”

Norme ST.26 : Séquences (8)

- Séquences d'acides aminés:
 - tous les symboles en lettres majuscules;
 - pas d'espaces ni de numérotation;
 - le symbole “X” sera considéré comme “équivalent à l'un des symboles ‘A’, ‘R’, ‘N’, ‘D’, ‘C’, ‘Q’, ‘E’, ‘G’, ‘H’, ‘I’, ‘L’, ‘K’, ‘M’, ‘F’, ‘P’, ‘O’, ‘S’, ‘U’, ‘T’, ‘W’, ‘Y’, or ‘V’”.

Symbol	Amino acid
A	Alanine
R	Arginine
N	Asparagine
D	Aspartic acid (Aspartate)
C	Cysteine
Q	Glutamine
E	Glutamic acid (Glutamate)
G	Glycine
H	Histidine
I	Isoleucine
L	Leucine
K	Lysine
M	Methionine
F	Phenylalanine
P	Proline
O	Pyrrolysine
S	Serine
U	Selenocysteine
T	Threonine
W	Tryptophan
Y	Tyrosine
V	Valine
B	Aspartic acid or Asparagine
Z	Glutamine or Glutamic acid
J	Leucine or Isoleucine
X	A or R or N or D or C or Q or E or G or H or I or L or K or M or F or P or O or S or U or T or W or Y or V; "unknown" or "other"

Norme ST.26 : Séquences (8)

- Séquences omises: elles permettent à un déposant de supprimer les données de séquences issues d'un listage des séquences sans qu'il soit nécessaire de renuméroter les séquences suivantes.
 - INSDSeq_length, INSDSeq_moltype, INSDSeq_division présents, mais sans valeur;
 - Aucun tableau de caractéristiques ni aucune clé de caractérisation "source";
 - l'élément de la séquence doit avoir pour valeur "000".

```
<SequenceData sequenceIDNumber="7">  
  <INSDSeq>  
    <INSDSeq_length/>  
    <INSDSeq_moltype/>  
    <INSDSeq_division/>  
    <INSDSeq_sequence>000</INSDSeq_sequence>  
  </INSDSeq>  
</SequenceData>
```

Norme ST.26 de l'OMPI: Table des matières

- **Corps du texte** – Exigences en matière d'intégration et de représentation
- **Annexe I** – Vocabulaire contrôlé selon l'INSDC
- **Annexe II** – Définition de type de document (DTD)
- **Annexe III** – Exemple de listage des séquences (fichier XML)
- **Annexe IV** – Sous-ensemble de caractères provenant du tableau de codes des caractères latins de base de la norme Unicode à utiliser dans une instance XML d'un listage des séquences
- **Annexe V** – Prescriptions supplémentaires en matière d'échange de données (uniquement pour les offices de brevets)
- **Annexe VI** – Document d'orientation assorti d'exemples illustratifs
- Appendice à l'annexe VI – Fichier XML contenant tous les exemples de séquences divulgués dans l'annexe VI
- **Annexe VII** – Recommandation concernant la conversion d'un listage des séquences de la norme ST.25 à la norme ST.26

Norme ST.26 de l'OMPI

Corps du texte

Paragraphes	Contenu
1-9	Introduction; Définitions; Portée; Références
10-37	Représentation des séquences
38-49	Structure du listage de séquences en XML
50-71	Partie consacrée aux données sur les séquences; Tableau de caractéristiques; Clés de caractérisation; Clés de caractérisation obligatoires; Emplacement de la caractéristique
72-84	Qualificateurs de caractéristiques; Qualificateurs de caractéristiques obligatoires
85-100	Texte libre; Séquences de codage; Variantes

Annexe I de la norme ST.26 de l'OMPI

Vocabulaire contrôlé

Section	Contenu
1	Liste des nucléotides (symboles à une lettre en minuscule)
2	Liste des nucléotides modifiés
3	Liste des acides aminés (symboles à une lettre en majuscule)
4	Liste des acides aminés modifiés
5	Clés de caractérisation pour les séquences de nucléotides
6	Qualificateurs pour les séquences de nucléotides
7	Clés de caractérisation pour les séquences d'acides aminés (inspirées de UniProt)
8	Qualificateurs pour les séquences d'acides aminés
9	Tableaux du code génétique

Annexe II de la norme ST.26 de l'OMPI

Définition de type de document (DTD)

- Détail de la DTD de la norme ST.26: version 1.3 actuelle
- Partie consacrée aux informations générales
 - informations sur la demande de brevet
- Partie consacrée aux données sur les séquences
 - sous-ensemble de la DTD de l'INSDC
 - un ou plusieurs éléments de données sur les séquences, chacun d'eux contenant des informations sur une séquence

Annexe VI de la norme ST.26 de l'OMPI

Document d'orientation

- Contient 49 exemples concrets de divulgations de séquences et une explication de l'application des règles de la norme ST.26 à chaque exemple;
- Chaque exemple indique:
 1. si l'inclusion de la séquence dans un listage des séquences est prescrite, autorisée ou interdite;
 2. lorsqu'elle est prescrite ou autorisée, la façon de la représenter.
- L'appendice à l'annexe VI est un listage des séquences au format XML conforme à la norme ST.26 qui comprend tous les exemples présentés dans le document d'orientation.

Annexe VII de la norme ST.26 de l'OMPI

Recommandation concernant la conversion d'un listage des séquences de la norme ST.25 à la norme ST.26

- ...ou comment éviter l'ajout de nouveaux éléments.
- Les exigences de la norme ST.26 diffèrent de celles de la norme ST.25: les règles de la norme ST.26 exigent des informations qui n'étaient pas requises par les règles de la norme ST.25.
- La conversion d'un listage des séquences du format ST.25 au format ST.26 nécessitera toujours la participation du déposant.
- La conversion d'un listage des séquences conforme à la norme ST.25 au format ST.26 n'entraînera pas d'ajout de nouveaux éléments si les recommandations de l'annexe VII sont respectées.
- Vingt scénarios de transformation sont présentés avec des recommandations et des exemples.

WIPO Sequence (1)

- Outil bureautique mis au point par l'OMPI pour faciliter la création, la validation et la génération de listages des séquences conformes à la norme ST.26.
- Les États membres ont demandé à l'OMPI de mettre au point cet outil commun à destination de tous les offices et déposants aux niveaux national, régional et international.
- L'utilisation de WIPO Sequence facilite la création de fichiers XML répondant à la norme ST.26 grâce à une interface facile à utiliser: plus besoin de modifier directement un fichier XML.
- Vous pouvez télécharger gratuitement la dernière version à l'adresse suivante :

<https://www.wipo.int/standards/fr/sequence/index.html>

WIPO Sequence (2)

- Sauvegarde des informations relatives aux séquences dans un projet, validation, puis génération d'un listage des séquences au format ST.26.
- Importation de données à partir de listages des séquences ST.26, de projets ST.26, de listages des séquences ST.25, de fichiers aux formats multi-séquence, brut et FASTA.
- Validation des listages de séquences au format XML également.
- Sélection facile des clés de caractérisation, des qualificateurs et des noms d'organismes pertinents dans des menus déroulants.
- Stockage des informations relatives aux déposants et aux inventeurs dans une base de données "Personnes et organisations".
- Prise en charge de l'exportation et de l'importation des fichiers XLIFF utilisés par les traducteurs.

WIPO Sequence: Page d'accueil "Projets"

WIPO | Sequence **PROJECTS** PERSONS & ORGANIZATIONS ORGANISMS HELP PREFERENCES ENGLISH

[NEW PROJECT](#) [IMPORT PROJECT](#) [IMPORT SEQUENCE LISTING](#) [VALIDATE SEQUENCE LISTING](#)

PROJECTS

🔍

Project name :	Applicant file reference	Applicant name	Invention title	Status	Creation Date :
16079428 no source sequence	H0075.70243US00	Ferring B.V.	STABLE LIQUID GONADOTROPIN FORMULATION	invalid	2021-02-01
All features and qualifiers	123abc	Simple Healthkit, Inc.	Compositions and Methods for Treating Cancer	modified	2021-02-01
Ark three letter aa code import from ST.25	abc123	Joe, Smith	Improper 3-letter AA code for import	new	2021-02-26
Beta 4 Test Project	abc123	Stephenson, Adam	Beta 4 Test Project	modified	2021-03-11
DNA RNA with u and t	uspto	Biotech, Inc.	Beta 4 test import ST25	modified	2021-03-15

WIPO Sequence: Détail du projet

Section “Informations générales”

WIPO | Sequence
 FOR WIPO TRAINING
VERIFICATION REPORT
FREE TEXT QUALIFIERS
IMPORT REPORT
DISPLAY THE SEQUENCE LISTING
HELP

 PREFERENCES ENGLISH
 Return to project home

GENERAL INFORMATION SEQUENCES

GENERAL INFORMATION

APPLICATION IDENTIFICATION

Application Identified Before the assignment of the application number Application filed Applicant file reference ABC123	IP Office IB - International Bureau of the World Intellectual Property Organization (WIPO) Application number PCT/IB2015/099999 Filing date 2015-01-30	
--	--	--

PRIORITY IDENTIFICATION

IP Office	Application Number	Filing date	Selected Earliest Priority Application
IB - International Bureau of the World Intellectual Property Organization (WIPO)	PCT/IB2014/111111	2014-01-30	Yes

APPLICANT & INVENTOR

Applicant name Shutsugan Pharmaceuticals Kabushiki Kaisha	Primary applicant	
---	-------------------	--

INVENTION TITLE

Invention title Mus musculus abcd-1 gene for efg protein	Language en - English	
--	-----------------------	--

WIPO Sequence: Détail du projet

Section "Séquences"

WIPO | Sequence
FOR WIPO TRAINING
VERIFICATION REPORT
FREE TEXT QUALIFIERS
IMPORT REPORT
DISPLAY THE SEQUENCE LISTING
HELP
PREFERENCES ENGLI
Return to project home

GENERAL INFORMATION SEQUENCES

SEQUENCE 1

Sequence Number (ID) 1	Molecule Type DNA	
Sequence Name test	Organism Mus musculus	
Length 52		

▼ FEATURES

Add feature

Feature Key	Location	Qualifiers
<u>source</u>	1..52	mol_type = genomic DNA organism = Mus musculus

▼ SEQUENCE

```

atgaaattaa  aacataaaaar  ggatgataaa  atgagatttg  atataaaaaa  gg

```

52

< 1/1 >

↑

La suite?

standards@wipo.int

Séance de questions-réponses

Glossaire: Sigles

- CWS: Comité des normes de l'OMPI
- DDBJ: Banque de données ADN du Japon
- EMBL-EBI: Institut européen de bioinformatique
- OEB: Office européen des brevets
- INSDC: Collaboration internationale sur les bases de données de séquences de nucléotides
- NCBI: Centre national d'information sur la biotechnologie
- OMPI: Organisation Mondiale de la Propriété Intellectuelle