OMPI
ORGANIZACIÓN MUNDIAL DE LA PROPIEDAD INTELECTUAL

Norma ST.26 de la OMPI
INTRODUCCIÓN

Seminarios web de formación

Temas que se abordarán hoy

- ¿Qué es una lista de secuencias?
- Bases de datos de secuencias de la INSDC
- ¿Por qué una nueva norma?
- Ventajas de la Norma ST.26 de la OMPI
- ST.25 frente a ST.26, ¿en qué se diferencian?
- Transición a la Norma ST.26
- Aspectos básicos sobre XML
- Aspectos básicos sobre la Norma ST.26: partes de una lista de secuencias
- Contenido de la Norma ST.26 de la OMPI
- Introducción a WIPO Sequence

¿Qué es una lista de secuencias?

Una lista de secuencias…

- contiene las secuencias de nucleótidos y/o aminoácidos que se divulgan en una solicitud de patente y forma parte de la descripción
- incluye información descriptiva de cada secuencia, conocida como “anotaciones”
- se ajusta a los requisitos de la norma de la OMPI en cuestión (ST.25 o ST.26)
- facilita la búsqueda de datos de secuencia de una invención:
  - En una Oficina de PI
  - En las bases de datos disponibles al público (bases de la INSDC)

Contexto: INSDC

- INSDC (*International Nucleotide Sequence Database Collaboration*):
  - DDBJ: *DNA Databank of Japan*
  - EMBL-EBI: Laboratorio Europeo de Biología Molecular - Instituto Europeo de Bioinformática
  - NCBI: *National Center for Biotechnology Information* (GenBank)
- Ejemplos de Oficinas de PI que envían a las bases de INSDC datos de listas de secuencias relativos a solicitudes publicadas o concedidas:
  - Oficina Europea de Patentes
  - Oficina Japonesa de Patentes
  - Oficina Surcoreana de Propiedad Intelectual
  - Oficina de Patentes y Marcas de los Estados Unidos de América
- Las bases de datos de la INSDC están a disposición del público para efectuar búsquedas

¿Por qué una nueva norma?

- Actualmente las listas de secuencias se presentan en un formato conforme a la Norma ST.25

Sin embargo...

- El formato ST.25 no se ajusta a los requisitos de la INSDC, por lo que parte de la información se pierde al importarla en bases de datos públicas
- Las normas incluidas en la ST.25 no son claras y, por ello, las Oficinas de PI de todo el mundo interpretan y las aplican de manera diferente
- Los tipos de secuencias comunes hoy en día no se contemplan en la Norma ST.25 (análogos de nucleótidos, D-aminoácidos, secuencias ramificadas), por lo que no pueden encontrarse en las bases de datos
- Los datos no están estructurados: es difícil utilizar el formato ST.25 para la validación y el intercambio automático de datos

Ventajas de la Norma ST.26 (1)

- Se acepta una única lista de secuencias en todo el mundo*
- La Norma sirve de orientación para garantizar la coherencia entre las Oficinas de PI respecto de la aplicación de normas relativas a las secuencias
- Aclara qué divulgaciones de secuencias deben o pueden incluirse en una lista de secuencias, y cómo deben representarse
- Mejora de la calidad en lo que respecta a la presentación, gracias a la estructura de las listas de secuencias en formato XML
- Mayor automatización de la validación de datos y simplificación del proceso de tramitación para las Oficinas de PI

- sin perjuicio de las traducciones que puedan exigirse de calificadores de texto libre dependiente del idioma de presentación en ciertas Oficinas de PI, para lo cual quizá sean necesarias listas de secuencias de sustitución

Ventajas de la Norma ST.26 (2)

- Compatibilidad de los datos con los requisitos para proveedores de la base de datos de INSDC (DDBJ, EBI y NCBI); las anotaciones de secuencias (claves de caracterización y calificadores) se incluirán en bases de datos de consulta pública

- Normalización de los siguientes datos:
  - Anotaciones de características
  - Localizaciones de características
  - Calificadores y valores calificadores
  - Presentación de variantes de secuencias

- El requisito relativo a la inclusión de tipos adicionales de secuencias (análogos de nucleótidos, D-aminoácidos, secuencias ramificadas) implica que será posible buscar más datos de secuencias

ST.25 frente a ST.26

Norma ST.25	Norma ST.26
ASCII.txt con identificadores numéricos	XML con elementos y atributos
No se exige que incluya: - D-aminoácidos - Porciones lineales de secuencias ramificadas - Análogos de nucleótidos	Debe incluir: - D aminoácidos - Porciones lineales de secuencias ramificadas - Análogos de nucleótidos
Anotación de secuencias: - Solo claves de caracterización	Anotación de secuencias: - Claves de caracterización y calificadores
Se permite incluir secuencias: - < 10 nucleótidos específicamente definidos - < 4 aminoácidos específicamente definidos	Se prohíben secuencias: - < 10 nucleótidos específicamente definidos - < 4 aminoácidos específicamente definidos

ST.25 frente a ST.26

... Información general

Norma ST.25	Norma ST.26
Se puede incluir TODA la información relativa a solicitudes de prioridad	SOLO puede incluirse la solicitud de prioridad más antigua
Se pueden incluir TODOS los nombres de solicitantes e inventores	SOLO se puede incluir un solicitante Y opcionalmente UN inventor
Se permite añadir un título a la invención	Se puede incluir múltiples títulos de la invención, cada uno en un idioma distinto
Los nombres de los solicitantes o inventores y el título de la invención deben estar en caracteres latinos básicos	Se pueden incluir los nombres de los solicitantes o inventores usando cualquier carácter Unicode válido, junto con una traducción o transliteración en caracteres latinos básicos

ST.25 frente a ST.26

… Datos de secuencia (1)

Norma ST.25	Norma ST.26
Las secuencias se identifican únicamente como DNA, RNA o PRT	Las secuencias se identifican como DNA, RNA o AA, junto con un calificador obligatorio “mol_type” para describir mejor la molécula
Nombre del organismo: - Género/especie en latín - Nombre del virus - “secuencia artificial” (“artificial sequence”) - “desconocido” (“unknown”)	Nombres de los organismos: - Género/especie en latín - Nombre del virus - “constructo sintético” (“synthetic construct”) - “sin identificar” (“unidentified”)
“u” representa el uracilo en secuencias de nucleótidos	“t” representa el uracilo en secuencias de ARN y la timina en secuencias de ADN
Las secuencias de aminoácidos se representan mediante abreviaturas de tres letras	Las secuencias de aminoácidos se representan mediante abreviaturas de una letra

ST.25 frente a ST.26

… Datos de secuencia (2)

Norma ST.25	Norma ST.26
Las variables “n” y “Xaa” deben estar definidas en una caracterización	Se presupone un valor por defecto para las variables “n” y “X” sin definición
El formato de localización de característica no está definido claramente	Formatos de localización de características claramente definidos; se puede usar “<” y “>” en todos los tipos de secuencias, y “^”, “join”, “order” y “complement” en las secuencias de nucleótidos
Se permiten secuencias en modo mixto (“mixed mode”): secuencia de nucleótidos con traducción a aminoácidos más abajo	NO hay “modo mixto”; las traducciones de nucleótidos se incluyen únicamente en calificadores de traducción (“translation”)

Transición a la Norma ST.26

- El CWS/5 acordó fijar el **1 de enero de 2022** como fecha para la **transición generalizada**
- **Todas** las Oficinas de Propiedad Intelectual (OPI) efectuarán una transición simultánea a nivel internacional (PCT), nacional y regional
- La fecha de presentación internacional será la fecha de referencia que determine si en una solicitud se aplica la Norma ST.25 o la Norma ST.26, y NO la fecha de prioridad
  - Nota: La Norma ST.25 seguirá en vigor respecto de solicitudes con una fecha de presentación anterior al 1 de enero de 2022

Norma ST.26 de la OMPI

... ¿Qué debe incluirse?

- Secuencias de nucleótidos:

  - 10 o más residuos “específicamente definidos” y “enumerados”*

  - secuencias con análogos de nucleótidos, como, por ejemplo, ácidos peptidonucleicos (APN) y ácidos nucleicos glicólicos (ANG)

- Secuencias de aminoácidos

  - 4 o más residuos “específicamente definidos” y “enumerados”*

  - secuencias con D-aminoácidos

  - se deben incluir las regiones lineales de secuencias ramificadas en una lista de secuencias

Norma ST.26 de la OMPI

…¿Qué debe incluirse?

- ¿Qué es un nucleótido o aminoácido “específicamente definido”?

- por “específicamente definido” se entenderá todo nucleótido distinto a los representados por el símbolo “n” y todo aminoácido distinto a los representados por el símbolo “X” que se enumeran en el Anexo I.
  (Norma ST.26 de la OMPI, párr. 3.k)).

- únicamente los residuos “específicamente definidos” cuentan en relación con el requisito de longitud mínima:
  - 10 o más nucleótidos específicamente definidos; o
  - 4 o más aminoácidos específicamente definidos

5’- a**n**ctggcaa**n** – 3’	solo 8 nucleótidos específicamente definidos; **no debe** incluirse en una lista de secuencias
5’- agctggcaat – 3’	10 nucleótidos específicamente definidos; **debe incluirse** en una lista de secuencias

Norma ST.26: documento XML

- Debe facilitarse como archivo de formato XML 1.0

- Debe validarse con arreglo a la definición de tipo de documento (DTD) de la Norma ST.26 de la OMPI (Anexo II) y a aquellas normas operativas derivadas del contenido de esta Norma

- Debe cifrarse utilizando el lenguaje Unicode UTF-8

- Estructura de la lista de secuencias:

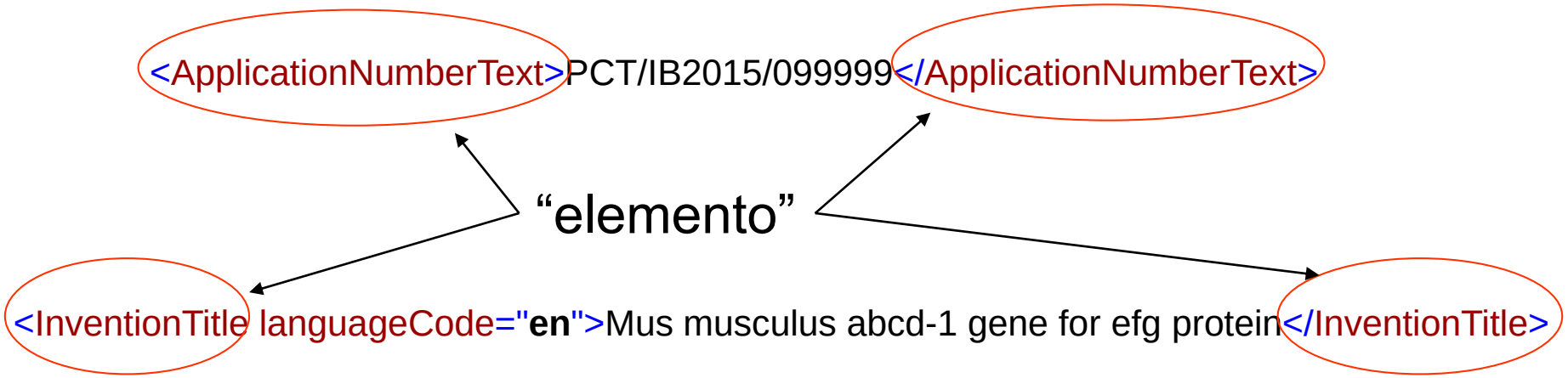
• Declaración XML
• Declaración de tipo de documento (DOCTYPE)
• Elemento raíz
  - Parte de información general
  - Parte de datos de secuencia

Aspectos básicos sobre XML

- XML = lenguaje extensible de marcado
- La información se “etiqueta” mediante el uso de elementos y atributos descriptivos
- Medios normalizados de intercambio de datos legibles por humanos y máquinas
- DTD = Definición de tipo de documento: define la estructura y los atributos y elementos jurídicos de un documento XML

Aspectos básicos sobre XML

Elementos, atributos y valores (1)



Aspectos básicos sobre XML

Elementos, atributos y valores (2)

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

“valor de elemento”

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

Aspectos básicos sobre XML

Elementos, atributos y valores (3)

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

“atributo”

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

“valor de atributo”

Aspectos básicos sobre XML

Caracteres reservados

Deben sustituirse por su entidad predefinida en un valor de elemento.

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

Ejemplo: Localización de característica deseada “<50..62”

<INSDFeature_location><50..62</INSDFeature_location> ✗

<INSDFeature_location><50..62</INSDFeature_location> ✓

Norma ST.26 de la OMPI: Ejemplo


```xml
<?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_quals>
                        <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_quals>
                </INSDFeature>
            </INSDSeq_feature-table>
            <INSDSeq_sequence>atgaaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
        </INSDSeq>
    </SequenceData>
</ST26SequenceListing>
```

Norma ST.26 de la OMPI: Componentes (1)

Línea 1 – declaración XML

(ST.26, párrafo 39.a))

```
<?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">
```

Norma ST.26 de la OMPI: Componentes (2)

Línea 2 – declaración de tipo de documento (DOCTYPE)

(ST.26, párrafo 39.b))

```
<?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">
```

Norma ST.26 de la OMPI: Componentes (3)

Línea 3 – elemento raíz

(ST.26, párrafo 43)

```
<?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">
```

Información general:

(ST.26, párrafos. 38.a), 45 a 49)

Norma ST.26 de la OMPI: Componentes (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_quals>
                    <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_quals>
            </INSDFeature>
        </INSDSeq_feature-table>
        <INSDSeq_sequence>atgaaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
    </INSDSeq>
</SequenceData>
</ST26SequenceListing>
```


Datos de secuencia

(ST.26, párrafos. 38.b), 50 a 100)

Norma ST.26: Información general (1)

- Sección de identificación de la solicitud
  - si se conocen, el número de solicitud, la fecha de presentación y el código de la Oficina de PI son obligatorios;
  - si no, basta con facilitar la referencia del expediente del solicitante

```
<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>
```

Norma ST.26: Información general (2)

- Sección de la solicitud de prioridad

  - en la lista de secuencias solo puede incluirse una solicitud de prioridad, y debe ser la solicitud de prioridad más antigua;

  - es obligatoria en los casos en que se reivindica la prioridad

```
<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>
```

Norma ST.26: Información general (3)

- Sección del nombre del solicitante y nombre del inventor
  - en la lista de secuencias solo puede incluirse un nombre de solicitante y uno de inventor, y deben ser el solicitante y el inventor “principal”;
  - el nombre del solicitante es obligatorio; el del inventor, opcional;
  - es obligatorio introducir un código de idioma para el solicitante y el inventor;
  - si los nombres del solicitante o el inventor contienen caracteres no incluidos en la tabla Unicode de caracteres del alfabeto latino básico, se deberá incluir una transliteración o traducción

```
<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>
```

Norma ST.26: Información general (4)

- Sección del título de la invención
  - se debe incluir al menos un título de invención en el idioma de presentación;
  - se puede incluir más títulos en otros idiomas;
  - se requiere un código de idioma para cada título;

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

Norma ST.26: Información general (5)

- Elemento del número total de todas las secuencias
  - obligatorio;
  - en el total se deben incluir las secuencias ignoradas

```
<SequenceTotalQuantity>6</SequenceTotalQuantity>
```

Norma ST.26: Secuencias (1)

- Elementos obligatorios

“Número de identificación de secuencia” o “SEQ ID NO:”

Longitud de la secuencia

Tipo de molécula (ADN, ARN o AA)

División (siempre “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_quals>
                    <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_quals>
            </INSDFeature>
        </INSDSeq_feature-table>
        <INSDSeq_sequence>atgaaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
    </INSDSeq>
</SequenceData>
```

Norma ST.26: Secuencias (2)

- Característica obligatoria de la fuente (“source” o “SOURCE”)


```
<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_quals>
                    <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_quals>
            </INSDFeature>
        </INSDSeq_feature-table>
        <INSDSeq_sequence>atgaaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
    </INSDSeq>
</SequenceData>
```


Solo se requiere una clave de caracterización de fuente por secuencia

La localización de la fuente debe abarcar a toda la secuencia

Dos calificadores obligatorios: “mol_type” y “organism”

Norma ST.26: Secuencias (3)

Datos de secuencia en formato ST.26:


```xml
<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_quals>
                    <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_quals>
            </INSDFeature>
        </INSDSeq_feature-table>
        <INSDSeq_sequence>atgaaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
    </INSDSeq>
</SequenceData>
```

Norma ST.26: Secuencias (4)

- Calificadores obligatorios “mol_type” y “MOL_TYPE”: alternativas de valor:

DNA	**RNA**	**AA**
genomic DNA	genomic RNA	protein
other DNA	mRNA	
unassigned DNA	tRNA	
	rRNA	
	other RNA	
	transcribed RNA	
	viral cRNA	
	unassigned RNA	

Norma ST.26: Secuencias (5)

- Calificador obligatorio de organismo “organism” y “ORGANISM”: alternativas de valor:
  - Designación de género y especie en latín, p. ej.: “Mus musculus”
  - Designación del género, seguido de “sp.”, p. ej.: “Mus sp.”
  - Nombre de virus, p. ej.: “Torque teno virus 1”
  - “unidentified”
  - “synthetic construct”
- Nombres comunes como “mouse,” no deben usarse como nombres de organismos. Si se quiere, se pueden incluir nombres comunes en una lista de secuencias, pero en un calificador “note”.

Norma ST.26: Secuencias (6)

Claves de caracterización y calificadores

Además de la caracterización obligatoria “source” o “SOURCE”, los solicitantes pueden añadir múltiples características opcionales para describir mejor la secuencia

- claves de caracterización diferentes para secuencias de nucleótidos y de aminoácidos;
- cada característica puede contar con uno o varios calificadores opcionales, así como un calificador obligatorio

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

(ST.26, Anexo I, Secciones 5 a 8)

Norma ST.26: Secuencias (7)

- Secuencias de nucleótidos:
  - todos los símbolos en minúsculas;
  - ni espacios ni numeración;
  - no usar símbolo “u”; la “t” representa el uracilo en ARN
  - “n” tiene un valor por defecto de “a”, “c”, “g” o ‘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”

Norma ST.26: Secuencias (8)

- Secuencias de aminoácidos:
  - todos son símbolos de una letra, en mayúscula;
  - ni espacios ni numeración;
  - “X” tiene un valor por defecto de cualquiera de los símbolos “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” o “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”

Norma ST.26: Secuencias (8)

- Secuencias ignoradas: permiten al solicitante suprimir los datos de secuencia de una lista de secuencias sin tener que volver a numerar las secuencias posteriores.
  - Elementos INSDSeq_length, INSDSeq_moltype, INSDSeq_division, presentes pero sin indicar ningún valor
  - Ningún cuadro de características ni clave de caracterización de fuente
  - El elemento de secuencia debe tener el valor “000”

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

Norma ST.26: Contenido

- **Cuerpo principal** – Requisitos de inclusión o representación
- **Anexo I** – Vocabulario controlado basado en la INSDC
- **Anexo II** – Definiciones de tipo de documento (DTD) conforme a la Norma ST.26
- **Anexo III** – Ejemplo de lista de secuencias (archivo XML) conforme a la Norma ST.26
- **Anexo IV** – Subconjunto de caracteres del alfabeto latino básico para utilización en una instancia XML conforme a la Norma ST.26
- **Anexo V** – Requisitos sobre el intercambio de datos INDS (únicamente para las OPI)
- **Anexo VI** – Documento de orientación con ejemplos
- **Apéndice del Anexo VI** – Archivo XML que incluye todas las secuencias divulgadas como ejemplo en el Anexo VI
- **Anexo VII** – Recomendación para la transformación de la lista de secuencias de la Norma ST.25 con arreglo a la Norma ST.26

Norma ST.26:
Cuerpo principal

Referencia de párrafos	Contenido
1-9	Introducción; Definiciones; Ámbito; Referencias
10-37	Representación de secuencias
38-49	Estructura de la lista de secuencias en formato XML
50-71	Parte de datos de secuencia; Cuadro de características; Claves de caracterización; Claves de caracterización obligatorias; Localización de característica
72-84	Calificadores de caracterización; Calificadores de caracterización obligatorios
85-100	Texto libre; Secuencias codificadoras; Variantes

Norma ST.26 (Anexo I):
Vocabulario controlado

Sección	Contenido
1	Lista de nucleótidos (símbolos de una letra en minúscula)
2	Lista de nucleótidos modificados
3	Lista de aminoácidos (símbolos de una letra en mayúscula)
4	Lista de aminoácidos modificados
5	Claves de caracterización para secuencias de nucleótidos
6	Calificadores para secuencias de nucleótidos
7	Claves de caracterización para secuencias de aminoácidos (adaptadas de UniProt)
8	Calificadores para secuencias de aminoácidos
9	Tablas del código genético

Norma ST.26 (Anexo II):

DTD conforme a la Norma ST.26

- Datos sobre definiciones de tipo de documento (DTD) con arreglo a la Norma ST.26. Versión actual: 1.3

- Parte de información general

  -Elementos relativos a la información de la solicitud de patente

- Parte de datos de secuencia

  -Subconjunto de DTD de la INSDC

  -Uno o varios elementos de datos de secuencia, donde cada elemento contiene información sobre una secuencia

Norma ST.26 (Anexo VI):
Documento de orientación

- Contiene 49 ejemplos reales de divulgaciones de secuencias y una explicación sobre cómo se aplica en cada uno la Norma ST.26;

- En cada ejemplo se analiza lo siguiente:
  1. Si se exige, se permite o se prohíbe incluir la secuencia en una lista de secuencias;
  2. En el caso de que se exija o permita la inclusión de la secuencia en la lista, cómo debe representarse.

- El Apéndice del Anexo VI es una lista de secuencias en formato XML conforme a la Norma ST.26 que incluye los ejemplos presentados en el documento de orientación.

Norma ST.26 (Anexo VII):
Recomendación para la transformación de la lista de secuencias de la Norma ST.25 con arreglo a la Norma ST.26

- …o cómo evitar la adición de nueva materia.
- Los requisitos de la Norma ST.26 difieren de los de la ST.25: en la ST.26 se requiere información que no se exigía en la ST.25
- Para la transformación de una lista de secuencias de la Norma ST.25 a la Norma ST.26 siempre se requiere que el solicitante introduzca información
- La transformación de una lista de secuencias compatible con la Norma ST.25 al formato de la ST.26 no implicará la introducción de nueva materia si se siguen las recomendaciones del Anexo VII
- Presentación de veinte situaciones hipotéticas de transformación con recomendaciones y ejemplos

WIPO Sequence (1)

- Herramienta de escritorio desarrollada por la OMPI para facilitar la autoría, validación y generación de listas de secuencias compatibles con la Norma ST.26

- Los Estados miembros pidieron a la OMPI que desarrollara esta herramienta común para todas las Oficinas y los solicitantes a escala internacional, nacional y regional

- El uso de WIPO Sequence simplifica la creación de archivos en formato XML con arreglo a la Norma ST.26, en un interfaz sencilla, sin necesidad de editar directamente el archivo XML

- Descarga gratuita de la última versión en la dirección:
  https://www.wipo.int/standards/es/sequence/index.html

WIPO Sequence (2)

- La información de secuencias se puede guardar en un proyecto, validar, y después generar en una lista de secuencia en formato ST.26
- Los datos se pueden importar de listas de secuencias en formato ST.26, proyectos en formato ST.26, listas de secuencias en formato ST.25, archivos de formatos de varias secuencias, archivos en formato RAW y archivos en formato FASTA
- Validación de listas de secuencias, también en formato XML
- Se puede seleccionar fácilmente en menús desplegables las claves de caracterización, los calificadores y los nombres de organismos pertinentes
- La información del solicitante y el inventor se puede almacenar en una base de datos de “personas y organizaciones”
- Es posible exportar e importar archivos XLIFF utilizados por los traductores

WIPO Sequence: Página de inicio de proyectos

WIPO | Sequence PROJECTS PERSONS & ORGANIZATIONS ORGANISMS HELP PREFERENCES ENGLISH

NEW PROJECT IMPORT PROJECT IMPORT SEQUENCE LISTING VALIDATE SEQUENCE LISTING

PROJECTS

Search project by name

Project name	Applicant file reference	Applicant name	Invention title	Status	Creation Date
16079428_no_source_seqid1	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: Detalles del proyecto

Sección de información general

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

Add Earliest Priority Application

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

Add Applicant Add Inventor

Applicant name Shutsugan Pharmaceuticals Kabushiki Kaisha Primary applicant

INVENTION TITLE

Add Invention title

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

WIPO Sequence: Detalles del proyecto

Sección de secuencias

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
Sequence Name test
Length 52

Molecule Type DNA
Organism Mus musculus

FEATURES

Add feature

Feature Key	Location	Qualifiers
source	1..52	mol_type = genomic DNA organism = Mus musculus

SEQUENCE

```
atgaaattaa   aacataaaar   ggatgataaa   atgagatttg   atataaaaaa   gg                52
```

1/1

Próximas etapas

standards@wipo.int

Turno de preguntas

Glosario: Siglas

- CWS: Comité de Normas Técnicas de la OMPI
- DDBJ: *DNA Databank of Japan*
- EMBL-EBI: Instituto Europeo de Bioinformática
- OEP: Oficina Europea de Patentes
- INSDC: *International Nucleotide Sequence Database Collaboration*
- OPI: Oficina de Propiedad Intelectual
- NCBI: *National Center for Biotechnology Information*
- OMPI: Organización Mundial de la Propiedad Intelectual