

# **WIPO Standard ST.26**

## **INTRODUÇÃO**

Formação em Webinars

# A ser coberto hoje

- O que é uma listagem sequencial?
- Bases de Dados de Sequências da INSDC
- Por que um novo Padrão?
- Benefícios do WIPO ST.26
- ST.25 vs. ST.26 – quais são as diferenças?
- Transição para o ST.26
- Noções básicas de XML
- Noções básicas do ST.26 – partes de uma listagem sequencial
- Conteúdo do WIPO ST.26
- Introdução ao WIPO Sequence

# O que é uma listagem sequencial?

Uma listagem sequencial...

- Contém sequências de nucleotídeos e/ou aminoácidos revelados em um pedido de patente e faz parte da descrição
- Inclui informações descritivas sobre cada sequência conhecida como anotações
- Está em conformidade com as exigências do WIPO Standard pertinente (ST.25 ou ST.26)
- Permite que os dados sequenciais de uma invenção sejam pesquisáveis:
  - No âmbito de um Instituto de PI
  - Em bases de dados publicamente disponíveis (bases de dados INSDC)

# Histórico - INSDC

- INSDC: International Nucleotide Sequence Database Collaboration (Colaboração Internacional de Bases de Dados de Sequências de Nucleotídeos):
  - DDBJ: Base de dados de DNA do Japão
  - EMBL-EBI: Instituto Europeu de Bioinformática
  - NCBI: Centro Nacional de Informação em Biotecnologia (GenBank)
  
- Os Institutos de PI que enviam dados de sequência de pedidos publicados/emitidos de dados para bases de dados da INSDC incluem:
  - Instituto Europeu de Patentes
  - Instituto Japonês de Patentes
  - Instituto Coreano de Propriedade Intelectual
  - Instituto de Patentes e Marcas dos Estados Unidos
  
- As bases de dados da INSDC são publicamente abertas à pesquisa

# Por que um novo Padrão?

- Atualmente, as listagens sequenciais são processadas em conformidade com o WIPO ST.25

Mas...

- O formato do ST.25 não está em conformidade com as exigências da INSDC, ficando, por conseguinte, os dados perdidos quando inseridos em bases de dados públicas
- As regras do ST.25 não são claras, e os institutos de PI no mundo inteiro interpretam e aplicam as regras de forma diferente
- Os tipos de sequências que são comuns hoje não são cobertos pelas regras do ST.25 (análogos de nucleotídeos, aminoácidos D, sequências ramificadas) e, portanto, não estão presentes em bancos de dados pesquisáveis
- Os dados não estão estruturados – o formato do ST.25 é de uso difícil para a validação automatizada e a troca de dados

# Benefícios do WIPO ST.26 (1)

- Aceitação de uma única listagem sequencial no mundo inteiro\*
  - O padrão serve como orientação para garantir o acordo entre os Institutos de PI sobre pedidos de regras sobre sequências
  - Determina quais revelações de sequências são necessárias ou permitidas para serem incluídas em uma listagem sequencial, e como essas sequências devem ser representadas
  - Melhor qualidade de apresentação devida à estrutura das listagens sequenciais XML
  - Maior automatização da validação de dados e processamento simplificado por parte dos Institutos de PI
- exceto para as traduções necessárias de qualificadores de textos livres dependentes de idioma para o idioma de processamento para determinados Institutos de PI, que podem exigir listagens sequenciais de substituição.

# Benefícios do WIPO ST.26 (2)

- Compatibilidade de dados com os requisitos do servidor do banco de dados da INSDC - DDBJ, EBI e NCBI – Anotações sequenciais (chaves de caracterização e qualificadores) serão incluídos em bases de dados publicamente pesquisáveis
- Padronização de:
  - Anotações de caracterizações
  - Localizações de caracterizações
  - Qualificadores e valores de qualificadores
  - Apresentação de variantes sequenciais
- A exigência de inclusão de tipos adicionais de sequências (análogos de nucleotídeos, aminoácidos D, sequências ramificadas) significa que um número maior de dados de sequências serão pesquisados

# WIPO ST.25 versus ST.26

ST.25	ST.26
ASCII .txt com identificadores numéricos	XML com elementos e atributos
<p><u>Não obrigado</u> a incluir:</p> <ul style="list-style-type: none"> <li>- Aminoácidos D</li> <li>- Porções lineares de sequências ramificadas</li> <li>- Análogos de nucleotídeos</li> </ul>	<p><u>Deve</u> incluir:</p> <ul style="list-style-type: none"> <li>- Aminoácidos D</li> <li>- Porções lineares de sequências ramificadas</li> <li>- Análogos de nucleotídeos</li> </ul>
<p>Anotação de sequências:</p> <ul style="list-style-type: none"> <li>- Somente chaves de caracterização</li> </ul>	<p>Anotação de sequências:</p> <ul style="list-style-type: none"> <li>- Chaves de caracterização e qualificadores</li> </ul>
<p><u>Permitida</u> a inclusão de sequências:</p> <ul style="list-style-type: none"> <li>- &lt; 10 nucleotídeos especificamente definidos</li> <li>- &lt; 4 aminoácidos especificamente definidos</li> </ul>	<p><u>Proibidas</u> as sequências:</p> <ul style="list-style-type: none"> <li>- &lt; 10 nucleotídeos especificamente definidos</li> <li>- &lt; 4 aminoácidos especificamente definidos</li> </ul>



# WIPO ST.25 versus ST.26

## ...para Informação Geral

ST.25	ST.26
TODAS as informações sobre pedidos de prioridade podem ser incluídas.	SOMENTE o pedido de prioridade mais antigo pode ser incluído.
TODOS os nomes de requerentes e de inventores podem ser incluídos.	SOMENTE um requerente E opcionalmente UM inventor podem ser incluídos.
Um título de invenção permitido	Múltiplos títulos de invenção permitidos, cada um num idioma diferente
Os nomes do requerente/inventor e os títulos das invenções devem ser em caracteres latinos básicos.	Os nomes dos requerentes/inventores podem ser incluídos utilizando quaisquer caracteres Unicode válidos juntamente com uma tradução ou transliteração em caracteres latinos básicos

# WIPO ST.25 versus ST.26

## ...para dados de Sequência (1)

ST.25	ST.26
Somente sequências identificadas como DNA, RNA ou PRT	Sequências identificadas como DNA, RNA, ou AA juntamente com um qualificador obrigatório mol_type para descrever melhor a molécula
Nomes de organismos: <ul style="list-style-type: none"> <li>- Gênero/espécie em latim</li> <li>- Nome do vírus</li> <li>- “sequência artificial”</li> <li>- “desconhecido”</li> </ul>	Nomes de organismos: <ul style="list-style-type: none"> <li>- Gênero/espécie em latim</li> <li>- Nome do vírus</li> <li>- “construção sintética”</li> <li>- “não identificado”</li> </ul>
“u” representa o uracil nas sequências de nucleotídeos	“t” representa o uracil nas sequências de RNA e timina nas sequências de DNA
Sequências de aminoácidos representadas por abreviaturas de três letras	Sequências de aminoácidos representadas por abreviaturas de uma letra

# WIPO ST.25 versus ST.26

## ...para dados de Sequência (2)

ST.25	ST.26
As variáveis “n” e “Xaa” devem ter uma definição fornecida numa caracterização	Valor padrão suposto para as variáveis “n” e “X” sem definição
O formato da localização de caracterização não é claramente definido	Formatos de localização de caracterização rigorosamente definidos; permite o uso de “<” e “>” em todos os tipos de sequências, e “^”, “juntar”, “ordenar” e “complemento” em sequências de nucleotídeos
Sequências de “modo misto” permitidas – sequência de nucleotídeos com tradução de aminoácidos mostrada abaixo	NENHUM “modo misto”; as traduções de nucleotídeos estão incluídas apenas nos qualificadores de “tradução”

# Transição para o WIPO ST.26

- CWS/5 concordou com uma data de transição de **1° de janeiro de 2022**, referida como a **data do ‘big-bang’**
- **Todos** os Institutos de Propriedade Intelectual (IPIs) farão a transição simultaneamente, em nível internacional (PCT), nacional e regional.
- A data de depósito internacional (DDI) será a data de referência que determina se um pedido se enquadra nas regras de sequência ST.25 ou ST.26, e **NÃO** a data de prioridade.
- Nota: O ST.25 permanecerá em vigor para pedidos com data de depósito anterior a 1° de janeiro de 2022.

# WIPO ST.26

## ...o que deve ser incluído?

- Sequências de nucleotídeos:
  - 10 ou mais resíduos “especificamente definidos” e “enumerados”\*
  - incluem sequências com análogos de nucleotídeos, tais como ácidos nucleicos peptídicos (ANPs) e ácidos nucleicos glicólicos (ANGs)
- Sequências de aminoácidos:
  - 4 ou mais resíduos “especificamente definidos” e “enumerados”
  - incluem sequências com aminoácidos D
  - regiões lineares de sequências ramificadas devem ser incluídas em uma listagem sequencial

# WIPO ST.26

## ...o que deve ser incluído?

- O que é um nucleotídeo ou aminoácido “especificamente definido”?
- “especificamente definido” significa qualquer nucleotídeo que não seja representado pelo símbolo “n” e qualquer aminoácido que não seja representado pelo símbolo “X”, listados no Anexo I. (WIPO Standard ST.26, pgh. 3(m))
- apenas os resíduos “especificamente definidos” contam para o comprimento mínimo exigido:
  - 10 ou mais nucleotídeos especificamente definidos; ou
  - 4 ou mais aminoácidos especificamente definidos

5'- anctggcaan – 3'      apenas 8 nucleotídeos especificamente definidos; não devem ser incluídos numa listagem sequencial

5'- agctggcaat – 3'      dez nucleotídeos especificamente definidos; devem ser incluídos numa listagem sequencial

# WIPO ST.26: document XML

- Deve ser fornecido como arquivo de formato XML 1.0
- Deve validar contra WIPO ST.26 DTD (Anexo II) e regras comerciais derivadas do conteúdo do Padrão
- Deve ser codificado usando-se o Unicode UTF-8
- Estrutura da listagem sequencial:
  - Declaração XML
  - Declaração do tipo de documento (DOCTYPE)
  - Elemento de raiz
    - Parte de Informação Geral
    - Parte de Dados de Sequência

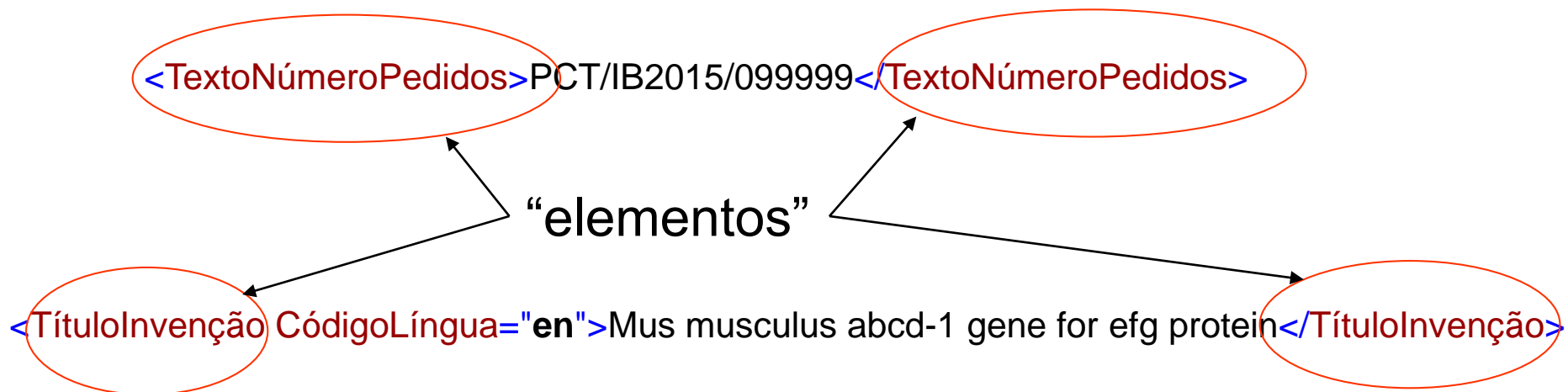
# Noções básicas de XML

- XML = eXtensible Markup Language (linguagem de marcação extensível)
- A informação é ‘rotulada’ utilizando-se elementos descritivos e atributos
- Padronizado significa, relativamente a trocas de dados, que pode ser lido tanto por humanos como por máquinas
- DTD = Document Type Definition (Definição do Tipo de Documento)  
– define a estrutura e os elementos e atributos jurídicos de um documento XML



# Noções básicas de XML

## Elementos, atributos e valores (1)



# Noções básicas de XML

## Elementos, atributos e valores (2)

`<TextoNúmeroPedido>PCT/IB2015/099999</TextoNúmeroPedido>`

“valores de elementos”

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

# Noções básicas de XML

## Elementos, atributos e valores (3)

<TextoNúmeroPedido>PCT/IB2015/099999</TextoNúmeroPedido>

“atributo”

<CódigoLínguaTítuloInvenção="en">Mus musculus abcd-1 gene for efg protein</TítuloInvenção>

“valor de atributo”


# Noções básicas de XML


## Caracteres reservados

Devem ser substituídos por suas entidades pré-definidas em um valor de elemento.

Reserved Character	Predefined Entities
<	&lt;
>	&gt;
&	&amp;
“	&quot;
'	&apos;

Exemplo: A desejada localização de caracterização é “<50..62”

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

`<INSDFeature_location>&lt;50..62</INSDFeature_location>` 

# WIPO ST.26:Exemplo

```

<?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>atgaaattaaaacataaaaarggatgataaaatgagatttgatataaaaagg</INSDSeq_sequence>
    </INSDSeq>
  </SequenceData>
</ST26SequenceListing>

```

# WIPO ST.26: Componentes (1)

Linha 1 – declaração XML  
(ST.26 parágrafo 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">
```

# WIPO ST.26: Componentes (2)

**Linha 2 – declaração do tipo de documento (DOCTYPE)**

*(ST.26 parágrafo 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">

```

# WIPO ST.26: Componentes (3)

**Linha 3 – o element raiz**  
(ST.26 parágrafo 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">
```

**Informação  
Geral:**

(ST.26 parágrafos 38(a),  
45-49)



# WIPO ST.26: 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_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>

```

## Dados de sequência

(ST.26 parágrafos 38(b),  
50-100)

# WIPO ST.26: Informação Geral (1)

- Seção “Identificação do Pedido”
  - o número do pedido, a data do depósito e o código do instituto de PI são obrigatórios, se forem conhecidos;
  - senão, apenas a referência do processo do requerente é suficiente.

```
<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>
```

# WIPO ST.26: Informação Geral (2)

- Seção “Pedido Prioritário”
  - apenas um pedido prioritário pode ser incluído na listagem sequencial, devendo ser o pedido prioritário mais antigo;
  - obrigatória, quando a prioridade é reivindicada.

```
<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>
```

# WIPO ST.26: Informação Geral (3)

- Seção “Requerente e Nome do Inventor”
  - somente um nome de requerente e um nome de inventor podem ser incluídos na listagem sequencial e devem ser o requerente e o inventor “primários”;
  - o nome do requerente é obrigatório; o nome do inventor é opcional;
  - é obrigatório um código linguístico para os nomes dos requerentes e dos inventores;
  - se o nome do requerente e/ou do inventor contiver caracteres latinos básicos não-Unicode, deverá ser incluída uma transliteração ou tradução em caracteres latinos básicos.

```

<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>
  
```

# WIPO ST.26: Informação Geral (4)

- Seção “Título da Invenção”
  - pelo menos um título de invenção no idioma de processamento é obrigatório;
  - títulos adicionais em outros idiomas podem ser incluídos;
  - um código de idioma é obrigatório 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>
```

# WIPO ST.26: Informação Geral (5)

- Elemento “Quantidade Total da Sequência”
  - obrigatório;
  - o total deve incluir as sequências ignoradas

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

# WIPO ST.26: Sequências (1)

- Elementos obrigatórios

O “Número de Identificação Sequencial”  
ou “SEQ ID NO:”

```

<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>atgaaattaaaacataaaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

O comprimento da sequência

O tipo de molécula  
(DNA, RNA ou AA)

A divisão  
(sempre “PAT”)

# WIPO ST.26: Sequências (2)

- “Fonte” obrigatória ou caracterização “FONTE”

```

<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>

```

Uma e única caracterização “fonte” é exigida para cada sequência

A localização “fonte” deve abranger toda a sequência

Dois qualificadores obrigatórios: “mol\_type” e “organism”



# WIPO ST.26: Sequências (3)

## Dados de Sequências ST.26:

```

<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>

```

# WIPO ST.26: Sequências (4)

- Qualificadores obrigatórios “mol\_type” e “MOL\_TYPE” – opções de valores:

## DNA

DNA genômico

outro DNA

DNA não atribuído

## RNA

RNA genômico

mRNA

tRNA

rRNA

outro RNA

RNA transcrito

cRNA viral

RNA não atribuído

## AA

proteína

# WIPO ST.26: Sequências (5)

- Qualificadores obrigatórios de “organismo” e “ORGANISMO” – opções de valor:
  - Nome latino do gênero e da espécie, p. ex., “Mus musculus”
  - Nome de gênero seguido de “sp.”, p. ex., “Mus sp.”
  - Nome de vírus, p. ex., “Torque teno virus 1”
  - “não identificado”
  - “construção sintética”
- Nomes comuns, como “mouse” (ratinho), não devem ser usados como nome de organismo. Caso deseje, os nomes comuns podem ser incluídos na listagem sequencial em um qualificador de nota.

# WIPO ST.26: Sequências (6)

## Chaves de Caracterização e Qualificadores

Além da caracterização “fonte” ou “FONTE” obrigatória, os requerentes podem adicionar várias caracterizações opcionais para descrever mais amplamente a sequência

- Diferentes chaves de caracterização para sequências de nucleotídeos e sequências de aminoácidos;
- Cada caracterização pode ter um ou mais qualificadores opcionais, e pode ter um qualificador obrigatório

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

(WIPO ST.26, Anexo I, Seções 5-8)

# WIPO ST.26: Sequências (7)

- Sequências de nucleotídeos:
  - todos os símbolos em minúsculas;
  - sem espaços e sem numeração;
  - sem símbolos “u”; “t” representa o uracil no RNA
  - “n” tem valor padrão de “qualquer um de ‘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”

# WIPO ST.26: Sequências (8)

- Sequências de aminoácidos:
  - todos com uma única letra, símbolos em maiúsculas;
  - sem espaços, sem numeração;
  - “X” tem valor padrão de “qualquer um de ‘A’, ‘R’, ‘N’, ‘D’, ‘C’, ‘Q’, ‘E’, ‘G’, ‘H’, ‘I’, ‘L’, ‘K’, ‘M’, ‘F’, ‘P’, ‘O’, ‘S’, ‘U’, ‘T’, ‘W’, ‘Y’, ou ‘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”

# WIPO ST.26: Sequências (8)

- Sequências ignoradas: Permitem que um requerente exclua dados de sequência de uma listagem sequencial sem ter de renumerar sequências posteriores.
  - INSDSeq\_length, INSDSeq\_moltype, INSDSeq\_division presente, mas sem nenhum valor;
  - Sem tabela de caracterização e sem caracterização de fonte;
  - O elemento sequencial deve ter o valor “000”

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

# WIPO ST.26: Conteúdo

- **Corpo Principal** – Exigências para inclusão/representação
- **Anexo I** – Vocabulário controlado baseado na INSDC
- **Anexo II** – Documento ST.26 Tipo DTD
- **Anexo III** – Exemplo de arquivo ST.26 XML de Listagem Sequencial
- **Anexo IV** – Subconjunto de Caracteres do Código Latino Básico para Instância XML ST.26
- **Anexo V** – Exigências para Intercâmbio de Dados do INDS (somente IPIs)
- **Anexo VI** – Documento de Orientação com Exemplos
- **Apêndice ao Anexo VI** – Arquivo XML, incluindo todas as revelações de sequências exemplificadas no Anexo VI
- **Anexo VII** – Recomendação para a Transformação de uma Listagem Sequencial de ST.25 para ST.26



# WIPO ST.26:

## Corpo Principal

Referência de parágrafo	Conteúdo
1-9	Introdução; Definições; Escopo; Referências
10-37	Representação de sequências
38-49	Estrutura da listagem sequencial em XML
50-71	Parte de dados de sequência; Tabela de caracterização; Chaves de caracterização; Chaves de caracterização obrigatórias; Localização da caracterização
72-84	Qualificadores de caracterização; Qualificadores de caracterização obrigatória
85-100	Texto livre; Sequências de codificação; Variantes

# WIPO ST.26 Anexo I:

## Vocabulário Controlado

Seção	Conteúdo
1	Lista de Nucleotídeos (letras minúsculas, símbolos de uma única letra)
2	Lista de Nucleotídeos Modificados
3	Lista de Aminoácidos (letras maiúsculas, símbolos de uma única letra)
4	Lista de Aminoácidos Modificados
5	Chaves de Caracterização para Sequências de Nucleotídeos
6	Qualificadores para Sequências de Nucleotídeos
7	Chaves de Caracterização para Aminoácidos (adaptadas a partir de UniProt)
8	Qualificadores para Sequências de Aminoácidos
9	Tabelas de Códigos Genéticos

# WIPO ST.26 Anexo II:

## ST.26 DTD

- Detalhes sobre ST.26 DTD: Versão atual 1.3
  
- Parte de Informação Geral
  - Elementos relacionados com informações sobre pedidos de patentes
  
- Parte de Dados de Sequência
  - Subconjunto do DTD da INSDC
  - Um ou mais elementos de dados sequenciais, em que cada elemento contém informações sobre uma sequência

# WIPO ST.26 Anexo VI: Documento de Orientação

- Contém 49 exemplos reais de revelações e uma explicação de como as regras ST.26 se aplicam a cada exemplo;
- Cada exemplo trata do seguinte:
  1. Se a sequência for necessária, permitida ou proibida de ser incluída numa listagem sequencial;
  2. Se a sequência for necessária ou permitida para ser incluída numa listagem sequencial, como ela deve ser representada.
- O Apêndice ao Anexo VI é uma listagem sequencial ST.26 XML que inclui todos os exemplos apresentados no Documento de Orientação.

# WIPO ST.26 Anexo VII:

## Recomendação para a Transformação de uma Listagem Sequencial de ST.25 para ST.26

- ...ou como evitar a inclusão de novas matérias
- As exigências do ST.26 diferem do ST.25 – As regras do ST.26 exigem informações que não eram requeridas pelas regras do ST.25.
- A transformação de uma listagem sequencial do formato ST.25 para o formato ST.26 sempre exigirá a intervenção do requerente
- A transformação de uma listagem sequencial ST.25 em conformidade com o formato ST.26 não resultará em nova matéria, se as recomendações do Anexo VII forem seguidas
- Vinte cenários de transformação apresentados com recomendações e exemplos

# WIPO Sequence (1)


- Ferramenta desktop desenvolvida pela OMPI para apoiar a criação, a validação e a geração de listagens sequenciais compatíveis com o ST.26
- Os Estados Membros solicitaram à OMPI que criasse essa ferramenta em comum para todos os Institutos e requerentes, em nível internacional, nacional e regional.
- A utilização do WIPO Sequence simplifica a criação de XML ST.26 através de interface de uso muito simples: em nenhum momento é necessário editar diretamente um arquivo XML.
- Baixe a mais recente versão gratuitamente aqui:

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

# WIPO Sequence (2)

- As informações sobre sequências podem ser salvaguardadas no âmbito de um projeto e depois uma listagem sequencial no formato ST.26 poderá ser gerada.
- Os dados podem ser importados a partir de: Listagens sequenciais ST.26, projetos ST.26, listagens sequenciais ST.25, arquivos em formato multi-sequências, arquivos em formato bruto e arquivos em formato FASTA.
- Validação de listagens sequenciais também em formato XML
- Chaves de caracterização relevantes, qualificadores e nomes de organismos podem ser facilmente selecionados a partir de menus suspensos
- As informações do requerente e do inventor podem ser armazenadas numa base de dados “Pessoas e Organizações”.
- Suporta exportação e importação de arquivos XLIFF utilizados por tradutores

# WIPO Sequence: Página inicial de Projetos

 **WIPO** | Sequence

[PROJECTS](#)
[PERSONS & ORGANIZATIONS](#)
[ORGANISMS](#)
[HELP](#)

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

## PROJECTS

Project name ↕	Applicant file reference	Applicant name	Invention title	Status	Creation Date ↕
<a href="#">16079428_no_source_sequence</a>	H0075.70243US00	Ferring B.V.	STABLE LIQUID GONADOTROPIN FORMULATION	invalid	2021-02-01
<a href="#">All features and qualifiers</a>	123abc	Simple Healthkit, Inc.	Compositions and Methods for Treating Cancer	modified	2021-02-01
<a href="#">Ark three letter aa code import from ST.25</a>	abc123	Joe, Smith	Improper 3-letter AA code for import	new	2021-02-26
<a href="#">Beta 4 Test Project</a>	abc123	Stephenson, Adam	Beta 4 Test Project	modified	2021-03-11
<a href="#">DNA RNA with u and t</a>	uspto	Biotech, Inc.	Beta 4 test import ST25	modified	2021-03-15



# WIPO Sequence: Página Detalhada de Projetos

## Seção de Informação Geral

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)	<a href="#">PCT/IB2014/111111</a>	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: Página Detalhada de Projetos

## Seção de Sequências

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 >

↑

E o que mais?

[standards@wipo.int](mailto:standards@wipo.int)

# Sessão de Perguntas e Respostas

# Glossário: Acrônimos

- CWS: Committee on WIPO Standards (Comitê de Normas da OMPI)
- DDBJ: DNA Databank of Japan (Base de Dados de DNA do Japão)
- EMBL-EBI: The European Bioinformatics Institute (Instituto Europeu de Bioinformática)
- EPO: European Patent Office (Instituto Europeu de Patentes)
- INSDC: International Nucleotide Sequence Database Collaboration (Colaboração Internacional do Banco de Dados de Sequências de Nucleotídeos)
- IPO: Intellectual Property Office (Instituto de Propriedade Intelectual)
- NCBI: National Center for Biotechnology Information (Centro Nacional de Informação em Biotecnologia)
- WIPO: World Intellectual Property Organization (Organização Mundial da Propriedade Intelectual)