

# ■ Norma ST.26 de la OMPI

## AVANZADO

Seminario web de formación

*Nota: este es un modulo de continuación, por lo que se presume que los participantes ya han realizado el curso básico sobre la Norma ST.26.*

*El curso básico sobre la Norma ST.26 está disponible en la dirección [https://www.wipo.int/meetings/es/details.jsp?meeting\\_id=62848](https://www.wipo.int/meetings/es/details.jsp?meeting_id=62848)*

# Temas que se abordarán hoy

- Claves y calificadores de caracterización de uso común
- Formatos de localización de característica
- Formatos de valores de calificadores y valores calificadores en idiomas distintos del inglés
- Situaciones especiales: uracilo en ADN y timina en ARN; moléculas híbridas ADN/ARN
- Análogos de nucleótidos, D-aminoácidos y secuencias ramificadas
- Variantes de la secuencia

# Claves y calificadores de caracterización de uso común

# Claves y calificadores de caracterización

- Se pueden usar claves de caracterización para describir con más detalle uno o varios residuos de una secuencia identificada por su localización
  - En la Sección 5 del Anexo I se enumeran las claves de caracterización para las secuencias de nucleótidos
    - Las claves de caracterización para las secuencias de nucleótidos se escriben en minúsculas; por ejemplo, “misc\_binding”
    - En la Sección 7 del Anexo I se enumeran las claves de caracterización para las secuencias de aminoácidos
      - Las claves de caracterización para las secuencias de aminoácidos se escriben en mayúsculas; por ejemplo, “REGION”
  
- Se pueden utilizar calificadores para describir las características con más detalle
  - En la Sección 6 del Anexo I se enumeran los calificadores para las secuencias de nucleótidos
  - Los calificadores para las secuencias de nucleótidos se escriben en minúsculas; por ejemplo, “allele”
  - En la Sección 8 del Anexo I se enumeran los calificadores de las secuencias de aminoácidos
  - Los calificadores para las secuencias de aminoácidos se escriben en mayúsculas; por ejemplo, “NOTE”

# Claves y calificadores de caracterización

```

<SequenceData sequenceIDNumber="17">
- <INSDSeq>
  <INSDSeq_length>7</INSDSeq_length>
  <INSDSeq_moltype>AA</INSDSeq_moltype>
  <INSDSeq_division>PAT</INSDSeq_division>
- <INSDSeq_feature-table>
  - <INSDFeature>
    <INSDFeature_key>SOURCE</INSDFeature_key>
    <INSDFeature_location>1..7</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier>
      <INSDQualifier_name>ORGANISM</INSDQualifier_name>
      <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
    </INSDQualifier>
    - <INSDQualifier>
      <INSDQualifier_name>MOL_TYPE</INSDQualifier_name>
      <INSDQualifier_value>protein</INSDQualifier_value>
    </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>
- <INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>1</INSDFeature_location>
- <INSDFeature_qual>
  - <INSDQualifier>
    <INSDQualifier_name>NOTE</INSDQualifier_name>
    <INSDQualifier_value>X can be any amino acid</INSDQualifier_value>
  </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
  <INSDSeq_sequence>XYEKGJL</INSDSeq_sequence>
</INSDSeq>
</SequenceData>

```

# Claves y calificadores de caracterización

- Las claves de caracterización son opcionales excepto para la característica “source”/“SOURCE”, que es obligatoria en todas las secuencias.
- Toda clave de caracterización tendrá una lista de calificadores que podrán utilizarse para describir con mayor detalle la característica. La mayoría de los calificadores son facultativos; sin embargo, algunas claves de caracterización tienen calificadores obligatorios.

5.31. Feature Key	regulatory
Definition	any region of a sequence that functions in the regulation of transcription, translation, replication or chromatin structure;
Mandatory qualifiers	regulatory_class
Optional qualifiers	allele bound_moiety function gene gene_synonym map note operon phenotype pseudo pseudogene standard_name

- Los calificadores “mol\_type”/“MOL\_TYPE” y “organism”/“ORGANISM” son obligatorios para la característica “source”/“SOURCE”.

# Claves y calificadores de caracterización

- Algunas claves de caracterización tienen otras limitaciones
  - ámbito del organismo; por ejemplo, “C\_region” se limita a los eucariotas
  - ámbito de la molécula; por ejemplo, “D-loop” se limita a las secuencias de ADN

---

5.4.	Feature Key	D-loop
	Definition	displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA by a single stranded invader in the reaction catalyzed by RecA protein
	Optional qualifiers	allele gene gene_synonym map note
	Molecule scope	DNA

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “modified\_base”

- Debe utilizarse la clave de caracterización “modified\_base” y su calificador obligatorio “mod\_base” para describir un nucleótido modificado (ST.26, párrafo 16)
- Un “nucleótido modificado” es cualquier nucleótido que no sea:
  - 3'-monofosfato de desoxi-[a, g, c, o t]
  - 3'-monofosfato de [a, g, c, o u] [ST.26, párrafo 3.f)]
- Un “nucleótido modificado” debe representarse con el correspondiente nucleótido no modificado (Anexo 1, Sección 1, Cuadro 1), siempre que sea posible. De no serlo, puede representarse con “n”. Por ejemplo, “2'-O-methylcytidine” debe representarse con “c” en la secuencia. “Queuosine” debe representarse con “n”. El símbolo “n” equivale a un único residuo.
- El valor del calificador obligatorio “mod\_base” debe seleccionarse entre los valores que figuran en el Cuadro 2 de la Sección 2 del Anexo I. Si el valor es “other”, deberá haber un calificador adicional “note” que contenga el nombre completo y no abreviado del residuo modificado.



# Claves y calificadores de caracterización

Secuencias de nucleótidos: clave de caracterización “modified\_base”

- Ejemplo: secuencia de nucleótidos con inosina en la posición 15

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>15</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

- “Inosine” figura en el Cuadro 2 de la Sección 2 del Anexo I, con la abreviatura “i”

# Claves y calificadores de caracterización

Secuencias de nucleótidos: clave de caracterización “modified\_base”

- Ejemplo: secuencia de nucleótidos con xantina en la posición 22

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>22</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>xanthine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

- “Xanthine” NO figura en el Cuadro 2 de la Sección 2 del Anexo I; por lo tanto, el valor del calificador obligatorio “mod\_base” debe ser “OTHER” y debe incluirse un calificador adicional “note” junto al valor “xanthine”.

# Claves y calificadores de caracterización

Secuencias de nucleótidos: clave de caracterización “modified\_base”

- “modified\_base” también puede utilizarse para describir un sitio abásico:

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>11</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>abasic site</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

- El sitio abásico puede representarse con una “n” en la secuencia, y describirse con más detalle utilizando una clave de caracterización “modified\_base” junto con un calificador “mod\_base” con el valor “OTHER” y un calificador adicional “note” con el valor “abasic site”.

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- La clave de caracterización “CDS” puede utilizarse para identificar secuencias codificadoras. La localización de la característica “CDS” debe incluir el codón de terminación. (ST.26, párrafo 89)
- NO hay calificadores obligatorios para la clave de caracterización “CDS”.
- Los calificadores de uso común son, entre otros:

“pseudo”

“pseudogene”

“translation”

“transl\_table”

“codon\_start”

“transl\_except”

“protein\_id”

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Una característica “CDS” solo puede tener UNO de los siguientes calificadores:

Calificador	Descripción	Valor
“pseudo”	indica que la característica “CDS” no es funcional y no tiene traducción, pero no es un pseudogén	ninguno
“pseudogene”	indica que la característica “CDS” es un pseudogén y no tiene traducción	“processed” “unprocessed” “unitary” “allelic” “unknown”
“translation”	indica la secuencia de aminoácidos derivada de la traducción de la característica “CDS”	abreviaturas de una letra de los aminoácidos

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Una secuencia de aminoácidos que contiene 4 o más aminoácidos específicamente definidos, codificada por una secuencia codificadora y que se divulga en un calificador “translation”, debe incluirse en la lista de secuencias como una secuencia distinta y se le asigna su propio número de identificación de secuencia.
- El identificador de secuencia asignado a la secuencia de aminoácidos deberá figurar como valor del calificador “protein\_id” dentro de la clave de caracterización “CDS”.
- El calificador “ORGANISM” de la clave de caracterización “SOURCE” para la secuencia de aminoácidos deberá ser idéntico al de su secuencia codificadora. (ST.26, párrafo 92)

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Calificadores que pueden modificar la secuencia traducida:

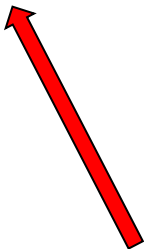
Calificador	Descripción	Valor
“transl_table”	indica la tabla del código genético utilizada para traducir la característica “CDS”; “1 - Código estándar” por defecto	un número que se corresponde con una tabla de traducción en la Sección 9 del Anexo I
“transl_except”	indica la traducción de un codón que no se ajusta al código genético definido en “transl_table”	“(pos:<location>, aa:<amino_acid>)”
“codon_start”	indica el marco de lectura de la característica “CDS” en relación con la primera base	1, 2 o 3

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Ejemplo - “SEQ ID NO:1” es un fragmento de una secuencia codificadora de un gen mitocondrial de levadura (tabla del código genético “3 - Código mitocondrial de levadura”):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```



La secuencia comienza con un codón parcial



El codón en las posiciones 30 a 32 codifica la selenocisteína (Sec)

- ¿Qué información se debería incluir en una característica “CDS” para representar con precisión esta secuencia?



# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Ejemplo - “SEQ ID NO:1” es un fragmento de una secuencia codificadora de un gen mitocondrial de levadura (tabla del código genético “3 - Código mitocondrial de levadura”):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- Secuencia representada en la lista de secuencia como:

tggataatga agaagtaac gaagaatgta tgagattatt ttcaagaac gctcgtcatc taacatcaag gttgacataa

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Ejemplo - “SEQ ID NO:1” es un fragmento de una secuencia codificadora de un gen mitocondrial de levadura (tabla del código genético “3 - Código mitocondrial de levadura”):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- Clave de caracterización “CDS”
- Localización de característica:

<1..80

El símbolo “<” indica que la región de codificación comienza antes de la posición 1

La localización incluye el codón de terminación en las posiciones 78 a 80

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Ejemplo - “SEQ ID NO:1” es un fragmento de una secuencia codificadora de un gen mitocondrial de levadura (tabla del código genético “3 - Código mitocondrial de levadura”):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- Clave de caracterización “CDS”
- Localización de característica: <1..80
- Calificador “codon\_start” con valor “3”

El valor 3 del calificador “codon\_start” indica que el primer codón completo comienza en la tercera posición dentro de la localización

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Ejemplo - “SEQ ID NO:1” es un fragmento de una secuencia codificadora de un gen mitocondrial de levadura (tabla del código genético “3 - Código mitocondrial de levadura”):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- Clave de caracterización “CDS”
- Localización de característica: <1..80
- Calificador “codon\_start” con valor “3”
- Calificador “transl\_table” con valor “3”

Las tablas del código genético de la Sección 9 del Anexo 1 se utilizan para determinar el valor del calificador “transl\_table”.

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Ejemplo - “SEQ ID NO:1” es un fragmento de una secuencia codificadora de un gen mitocondrial de levadura (tabla del código genético “3 - Código mitocondrial de levadura”):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa    80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- Clave de caracterización “CDS”
- Localización de característica: <1..80
- Calificador “codon\_start” con valor “3”
- Calificador “transl\_table” con valor “3”
- Calificador “transl\_except” con valor “(pos:30..32,aa:Sec)”

El calificador “transl\_except” pondrá en el aminoácido selenocisteína

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Ejemplo - “SEQ ID NO:1” es un fragmento de una secuencia codificadora de un gen mitocondrial de levadura (tabla del código genético “3 - Código mitocondrial de levadura”):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- Clave de caracterización “CDS”
- Localización de característica: <1..80
- Calificador “codon\_start” con valor “3”
- Calificador “transl\_table” con valor “3”
- Calificador “transl\_except” con valor “(pos:30..32,aa:Sec)”
- Calificador “translation” con valor “**DNEEVNEECURLFFKNARHTTSRLT**”

¡El codón de terminación no aparece en el calificador “translation”!

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Ejemplo - “SEQ ID NO:1” es un fragmento de una secuencia codificadora de un gen mitocondrial de levadura (tabla del código genético “3 - Código mitocondrial de levadura”):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- Clave de caracterización “CDS”
- Localización de característica: <1..80
- Calificador “codon\_start” con valor “3”
- Calificador “transl\_table” con valor “3”
- Calificador “transl\_except” con valor “(pos:30..32,aa:Sec)”
- Calificador “translation” con valor “**DNEEVNEECURLFFKNARHTTSRLT**”
- Una secuencia proteica distinta para la traducción

# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- Ejemplo - “SEQ ID NO:1” es un fragmento de una secuencia codificadora de un gen mitocondrial de levadura (tabla del código genético “3 - Código mitocondrial de levadura”):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- Clave de caracterización “CDS”
- Localización de característica: <1..80
- Calificador “codon\_start” con valor “3”
- Calificador “transl\_table” con valor “3”
- Calificador “transl\_except” con valor “(pos:30..32,aa:Sec)”
- Calificador “translation” con valor “**DNEEVNEECURLFFKNARHTTSRLT**”
- Una secuencia proteica distinta para la traducción
- Calificador “protein\_id” con el identificador de secuencia “SEQ ID” de la proteína traducida



# Claves y calificadores de caracterización

## Secuencias de nucleótidos: clave de caracterización “CDS”

- En la localización de característica “CDS” se puede utilizar el operador de localización “join” para unir segmentos discontinuos de una secuencia en una única región de codificación

"join"(localización1,localización2)

- En la localización de característica “CDS” se puede utilizar el operador “complement” para indicar que la característica está localizada en la cadena complementaria a la secuencia especificada por el descriptor de localización

"complement"(localización)

# Claves y calificadores de caracterización

Secuencias de aminoácidos: Claves de caracterización de uso común

Clave de caracterización	Descripción	¿Calificadores obligatorios?
SITE	indica un sitio de aminoácido único de interés	el sitio se debe describir mediante un calificador obligatorio "NOTE"
REGION	indica una región de interés	ninguno; el calificador "NOTE" es facultativo
BINDING	indica el sitio de unión de un grupo químico	se debe incluir un calificador obligatorio "NOTE" que contenga el nombre del grupo químico
UNSURE	describe regiones de incertidumbre en la secuencia	ninguno; el calificador "NOTE" es facultativo

# Claves y calificadores de caracterización

## Secuencias de aminoácidos: Aminoácidos modificados

- Un “aminoácido modificado” es cualquier aminoácido que no sea:

L-alanina	L-arginina	L-asparagina
L-ácido aspártico	L-cisteína	L-glutamina
L-ácido glutámico	L-glicina	L-histidina
L-isoleucina	L-leucina	L-lisina
L-metionina	L-fenilalanina	L-prolina
L-pirrolisina	L-serina	L-selenocisteína
L-treonina	L-triptófano	L-tirosina
L-valina		

(ST.26, párrafo 3.e))

- Un “aminoácido modificado” debería representarse como el correspondiente aminoácido no modificado, cuando sea posible. Si no, puede ser representado por el símbolo “X”. Por ejemplo, “hidroxilisina” debería representarse con “K” en la secuencia. “Ornitina” debería representarse con “X”. (ST.26, párrafo 29)
- El símbolo “X” equivale a un único residuo.

# Claves y calificadores de caracterización

## Secuencias de aminoácidos: Aminoácidos modificados

- Se pueden utilizar varias claves de caracterización para indicar un “aminoácido modificado”:

Clave de caracterización	Descripción
“SITE”	indica un aminoácido modificado (modificación no postraducciona)
“MOD_RES”	indica un aminoácido modificado (modificación postraducciona)
“CARBOHYD”	indica un aminoácido glicosilado
“LIPID”	indica el enlace covalente de un fragmento lipídico con un aminoácido

- Se debe incluir un calificador “NOTE” obligatorio para cada una de las claves de caracterización anteriores con un valor que describa la modificación.













# Formatos de localización de característica

# Formatos de localización

## Descriptores de localización para todo tipo de moléculas

- Los descriptores de localización se utilizan para identificar la localización de una característica en una secuencia
- La Norma ST.26 tiene requisitos obligatorios en cuanto al formato de los descriptores de localización
- Se pueden usar los siguientes formatos de descriptores de localización en secuencias de nucleótidos y aminoácidos:

Tipo de descriptor de localización	Sintaxis	Descripción
Número único de residuo	x	Designa un único residuo en la secuencia
Números de residuos que limitan un tramo de secuencia	x..y	Designa una serie continua de residuos delimitada por un residuo de inicio y uno de fin, incluidos éstos.
Residuos antes del primero, o después del último, residuo especificado	<x >x <x..y x..>y <x..>y	Designa una región que incluye un residuo o una serie de residuos especificada y que se extiende más allá del residuo especificado. Los símbolos '<' y '>' podrán utilizarse con un único residuo, o los números de los residuos de inicio y de fin de una serie de residuos para indicar que la característica se extiende más allá del número de residuo especificado.

# Formatos de localización

## Descriptores de localización para todo tipo de moléculas

Ejemplo de localización	Descripción
467	Designa el residuo 467 de la secuencia.
340..565	Designa una serie continua de residuos delimitada por los residuos 340 y 565, incluidos estos.
<1	Designa una localización de característica antes del primer residuo.
<345..500	Indica que se desconoce el punto exacto del límite inferior de una característica. La localización comienza en algún residuo anterior al 345 y continúa hasta el residuo 500, incluido este.
<1..888	Indica que la característica comienza antes del primer residuo de la secuencia y continúa hasta el residuo 888, incluido este.
1..>888	Indica que la característica comienza en el primer residuo de la secuencia y continúa más allá del residuo 888.
<1..>888	Indica que la característica comienza antes del primer residuo de la secuencia y continúa más allá del residuo 888.

*Norma ST.26 de la OMPI, párrafo 70.a)*

# Formatos de localización

## Descriptor de localización exclusivamente para secuencias de nucleótidos

- El siguiente formato de descriptor de localización SOLO puede utilizarse en secuencias de ADN y ARN:

Tipo de descriptor de localización	Sintaxis	Descripción
Un sitio entre dos números de residuo adyacentes	$x^y$	Designa un sitio entre dos residuos adyacentes, por ejemplo, el sitio de una unión endonucleolítica. Los números de posición de los residuos adyacentes están separados por el símbolo (^). Los formatos permitidos para este descriptor son $x^x+1$ (por ejemplo $55^56$ ), o para nucleótidos circulares, $x^1$ , donde "x" es la longitud total de la molécula, es decir, $1000^1$ para una molécula circular con una longitud 1000.

Norma ST.26 de la OMPI, párrafo 66.b)

# Formatos de localización

## Descriptores de localización exclusivamente para secuencias de nucleótidos

- Cuando se utiliza el formato de localización  $x^y$ , x e y deben ser residuos adyacentes

Location example	Description
123 <sup>124</sup>	Points to a site between residues 123 and 124.
867 <sup>1</sup>	In a circular molecule with 867 residues, points to a site between the residue indicated as position 1 and the residue indicated as position 867

*Norma ST.26 de la OMPI, párrafo 70.b)*

# Formatos de localización

## Descriptores de localización: caso especial de secuencias de aminoácidos

- El descriptor de localización “x..y” indica un enlace cruzado entre cadenas en relación con los dos residuos indicados cuando se utiliza una clave de caracterización “CROSSLNK” o “DISULFID”

Tipo de descriptor de localización	Sintaxis	Descripción
Números de residuo unidos por un enlace cruzado entre cadenas	x . . y	Designa aminoácidos unidos por un enlace cruzado entre cadenas cuando se utiliza con una característica que indica un enlace cruzado entre cadenas, como “CROSSLNK” o “DISULFID”

Norma ST.26 de la OMPI, párrafo 66.c)

# Formatos de localización

## Descriptores de localización: caso especial de las secuencias de aminoácidos

```

<SequenceData sequenceIDNumber="4">
  <INSDSeq>
    <INSDSeq_length>81</INSDSeq_length>
    <INSDSeq_moltype>AA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>SOURCE</INSDFeature_key>
        <INSDFeature_location>1..81</INSDFeature_location>
        <INSDFeature_quals>
          <INSDQualifier>
            <INSDQualifier_name>MOL_TYPE</INSDQualifier_name>
            <INSDQualifier_value>protein</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q25">
            <INSDQualifier_name>ORGANISM</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_quals>
      </INSDFeature>
      <INSDFeature>
        <INSDFeature_key>DISULFID</INSDFeature_key>
        <INSDFeature_location>30..50</INSDFeature_location>
        <INSDFeature_quals>
          <INSDQualifier id="q22">
            <INSDQualifier_name>NOTE</INSDQualifier_name>
            <INSDQualifier_value>disulfide bond</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_quals>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>QKKMIQFKKITHRYYYDIEHLCAKYDMNSVISNALFAKLNLMQVTDGLSTHEKIIILNLSNPLTFSIVISLQRCVINLGST</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```



# Formatos de localización

## Operadores de localización para secuencias de nucleótidos

- Hay tres operadores de localización disponibles para su uso en secuencias de ADN y ARN: “join”, “order” y “complement”
- El uso del operador de localización “join” implica que los residuos de nucleótidos designados por los descriptores de localización están puestos en contacto físicamente mediante procesos biológicos (ST.26, párrafo 68)
- Las localizaciones donde se usan “join” y “order” deben contar al menos con dos descriptores de localización separados por coma
- El operador “complement” puede usarse junto con “join” u “orden”.

Sintaxis de localización	Descripción
<code>join(location,location,..., location)</code>	Las localizaciones indicadas están unidas (colocadas extremo con extremo) para formar una secuencia contigua.
<code>order(location,location,...,location)</code>	Los elementos se encuentran en el orden especificado pero ninguna información permite determinar si la unión de esos elementos es razonable.
<code>complement(location)</code>	Indica que la característica está localizada en la cadena complementaria al tramo de la secuencia especificado por el descriptor de localización, cuando se lee en el sentido de 5' a 3' o en el sentido que imite el sentido 5' a 3.

Norma ST.26 de la OMPI, párrafo 67

# Formatos de localización

## Operadores de localización para secuencias de nucleótidos

Ejemplo de localización	Descripción
<code>123^124</code>	Designa un sitio entre los residuos 123 y 124
<code>join(12..78,134..202)</code>	Indica que las regiones 12 a 78 y 134 a 202 deberían unirse para formar una secuencia contigua.
<code>complement(34..126)</code>	Comienza en el nucleótido complementario a 126 y termina en el nucleótido complementario al nucleótido 34 (la característica está en la cadena complementaria a la cadena presentada).
<code>complement(join(2691..4571, 4918..5163))</code>	Une los nucleótidos 2691 a 4571 y 4918 a 5163, y luego complementa los segmentos unidos (la característica está en la cadena complementaria a la cadena presentada).
<code>join(complement(4918..5163), complement(2691..4571))</code>	Complementa las regiones 4918 a 5163 y 2691 a 4571, luego une los segmentos complementados (la característica está en la cadena complementaria a la cadena presentada).

*Norma ST.26 de la OMPI, párrafo 70.b)*

# ¿Preguntas?

# Formatos de valores de los calificadores

# Valores de los calificadores

- Los calificadores definen las características con mayor detalle

# Valores de los calificadores

- Los calificadores definen las características con mayor detalle
- Los calificadores se componen de un nombre y, a menudo, un valor.

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

# Valores de los calificadores

- Los calificadores definen las características con mayor detalle
- Los calificadores están formados por un nombre y, a menudo, un valor.

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

- Cada clave de caracterización contará con una lista de calificadores permitidos para esa característica. Algunas claves de caracterización tienen calificadores obligatorios.

# Valores de los calificadores

- Los calificadores definen las características con mayor detalle
- Los calificadores están formados por un nombre y, a menudo, un valor.

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

- Cada clave de caracterización contará con una lista de calificadores permitidos para esa característica. Algunas claves de caracterización tienen calificadores obligatorios.
- Las claves de caracterización y sus calificadores permitidos se enumeran en la Sección 5 (secuencias de nucleótidos) y la Sección 7 (secuencias de aminoácidos) del Anexo I de la Norma ST.26.



# Valores de los calificadores

- Ejemplo: la clave de caracterización “misc\_binding” tiene un calificador obligatorio, “bound\_moiety” y 6 calificadores opcionales:

---

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, Sección 5.12)

# Valores de los calificadores

- Los calificadores definen las características con mayor detalle
- Los calificadores están formados por un nombre y, a menudo, un valor.

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

- Cada clave de caracterización contará con una lista de calificadores permitidos para esa característica. Algunas claves de caracterización tienen calificadores obligatorios.
- Las claves de caracterización y sus calificadores permitidos se enumeran en la Sección 5 (secuencias de nucleótidos) y la Sección 7 (secuencias de aminoácidos) del Anexo I de la Norma ST.26.
- Los calificadores de las secuencias de nucleótidos, junto con sus descripciones, se enumeran en la Sección 6 (respecto de las secuencias de nucleótidos) y en la Sección 8 (para las secuencias de aminoácidos) del Anexo I de la Norma ST.26.

# Valores de los calificadores

---

6.3. Qualifier	bound_moiety
Definition	name of the molecule/complex that may bind to the given feature
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>GAL4</INSDQualifier_value>
Comment	A single bound_moiety qualifier is permitted on the "misc_binding", "oriT" and "protein_bind" features.

---

6.22. Qualifier	gene
Definition	symbol of the gene corresponding to a sequence region
Mandatory value format	free text
Example	<INSDQualifier_value>ilvE</INSDQualifier_value>
Comment	Use gene qualifier to provide the gene symbol; use standard_name qualifier to provide the full gene name.

(ST.26, Anexo I, Secciones 6.3 y 6.22)

# Valores de los calificadores

## Tipos de formato

- Existen varios tipos de formato para los valores de los calificadores:
  1. Calificadores con opciones de valores predefinidos;
  2. Calificadores con un formato de valor definido;
  3. Calificadores cuyo valor es una secuencia;
  4. Calificadores SIN valor;
  5. Calificadores con valores de “texto libre”
    - un subconjunto de valores de calificador de “texto libre” se clasifica como “dependiente del idioma”

# Valores de los calificadores

## Tipos de formato - Opciones de valores predefinidos

- Calificadores con opciones de valores predefinidos;
- Ejemplos:

“codon\_start” - los valores pueden ser “1”, “2” o “3”

---

6.9.	Qualifier	codon_start
	Definition	indicates the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature.
	Mandatory value format	1 or 2 or 3
	Example	<INSDQualifier_value>2</INSDQualifier_value>

# Valores de los calificadores

## Tipos de formato - Opciones de valores predefinidos

- Ejemplos: en “rpt\_type” hay un conjunto limitado de opciones de valores

6.60. Qualifier	rpt_type
Definition	structure and distribution of repeated sequence
Mandatory value format	One of the following controlled vocabulary terms or phrases: tandem direct inverted flanking nested terminal dispersed long_terminal_repeat non_ltr_retrotransposon_polymeric_tract centromeric_repeat telomeric_repeat x_element_combinatorial_repeat y_prime_element other
Example	<INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value>
Comment	Definitions of the values: tandem - a repeat that exists adjacent to another in the same orientation; direct - a repeat that exists not always adjacent but is in the same orientation; inverted - a repeat pair occurring in reverse orientation to one another on the same molecule;

# Valores de los calificadores

## Tipos de formato - Opciones de valores predefinidos

- En el caso de los calificadores con opciones de valores predefinidos, *WIPO Sequence* presentará todos los valores permitidos en una lista desplegable rellena previamente:

The screenshot displays the 'QUALIFIERS' section of the WIPO Sequence interface. It features a table with columns for 'Qualifier Name' and 'Qualifier Value'. The 'Qualifier Name' column contains the value 'rpt\_type'. The 'Qualifier Value' column is currently open, showing a list of predefined options. The option 'dispersed' is highlighted in blue. Below the 'QUALIFIERS' section, the 'SEQUENCE' section is visible, showing a grid of nucleotide sequences.

Qualifier Name *	Qualifier Value
rpt_type	<ul style="list-style-type: none"> <li>tandem</li> <li>direct</li> <li>inverted</li> <li>flanking</li> <li>nested</li> <li>terminal</li> <li><b>dispersed</b></li> <li>long_terminal_repeat</li> <li>non_itr_retrotransposon_polymeric_tract</li> <li>centromeric_repeat</li> <li>telomeric_repeat</li> <li>x_element_combinatorial_repeat</li> <li>y_prime_element</li> <li>other</li> </ul>

Below the 'QUALIFIERS' section, the 'SEQUENCE' section is visible, showing a grid of nucleotide sequences:

tgccgatatt	attcataaagc	atgggcgtct	ggaagtgcgc	ctgacggccc	agaagggcgt
cttactgtta	caagaagagt	atcatgacgc	goatcatagc	caottagacc	cgcaocttatg
gttggcgcgc	caaaatatca	ttgcctcgt	gcaaggattg	gataaacata	ctgctgagtt

# Valores de los calificadores

## Tipos de formato – Formato de valor definido

- Calificadores con un formato de valor definido
- Ejemplo: “anticodon” – el valor debe tener el formato

“(pos:<location>,aa:<amino\_acid>,seq:<text>)”

6.2. Qualifier	anticodon
Definition	location of the anticodon of tRNA and the amino acid for which it codes
Mandatory value format	(pos:<location>,aa:<amino_acid>,seq:<text>) where <location> is the position of the anticodon and <amino_acid> is the three letter abbreviation for the amino acid encoded and <text> is the sequence of the anticodon
Example	<pre>&lt;INSDQualifier_value&gt;(pos:34..36,aa:Phe,seq:aaa)&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;(pos:join(5,495..496),aa:Leu,seq:taa)&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;(pos:complement(4156..4158),aa:Glu,seq:ttg)&lt;/INSDQualifier_value&gt;</pre>



# Valores de los calificadores

## Tipos de formato - Secuencias

- Calificadores cuyo valor es una secuencia
- Ejemplo: “translation” - el valor debe ser una secuencia en la que se utilice abreviaturas de una letra de aminoácidos

6.79. Qualifier	translation
Definition	one-letter abbreviated amino acid sequence derived from either the standard (or universal) genetic code or the table as specified in a transl_table qualifier and as determined by an exception in the transl_except qualifier
Mandatory value format	contiguous string of one-letter amino acid abbreviations from Section 3 of this Annex, "X" is to be used for AA exceptions.
Example	<INSDQualifier_value>MASTFPPWYRGCSTPSLKGLIMCTW</INSDQualifier_value>
Comment	to be used with CDS feature only; must be accompanied by protein_id qualifier when the translation product contains four or more specifically defined amino acids; see transl_table for definition and location of genetic code Tables; only one of the qualifiers translation, pseudo and pseudogene are permitted to further annotate a CDS feature.

# Valores de los calificadores

## Tipos de formato - Secuencias

- Calificador “replace” – el valor puede ser un único residuo de nucleótido, una secuencia de residuos o estar vacío
- Un valor vacío para “replace” significa que se elimina el residuo indicado en la característica correspondiente

---

6.57. Qualifier	replace
Definition	indicates that the sequence identified in a feature's location is replaced by the sequence shown in the qualifier's value; if no sequence (i.e., no value) is contained within the qualifier, this indicates a deletion
Mandatory value format	free text
Example	<pre>&lt;INSDQualifier_value&gt;a&lt;/INSDQualifier_value&gt;</pre> <pre>&lt;INSDQualifier_value&gt;&lt;/INSDQualifier_value&gt; - for a deletion</pre>

# Valores de los calificadores

## Tipos de formato - Secuencias

- Calificador “replace” – el valor puede ser un único residuo de nucleótido, una secuencia de residuos o estar vacío
- Un valor vacío para “replace” significa que se elimina el residuo indicado en la característica correspondiente

6.57. Qualifier	replace
Definition	indicates that the sequence identified in a feature's location is replaced by the sequence shown in the qualifier's value; if no sequence (i.e., no value) is contained within the qualifier, this indicates a deletion
Mandatory value format	free text
Example	<pre>&lt;INSDQualifier value&gt;a&lt;/INSDQualifier value&gt; &lt;INSDQualifier_value&gt;&lt;/INSDQualifier_value&gt; - for a deletion</pre>

“Valor vacío”

# Valores de los calificadores

## Tipos de formato - SIN valor

- Calificadores SIN valor

- Ejemplos:

“environmental\_sample”    “germline”    “macronuclear”    “proviral”

---

6.51. Qualifier	proviral
Definition	this qualifier is used to flag sequence obtained from a virus or phage that is integrated into the genome of another organism
Value format	none

- *WIPO Sequence* no permitirá añadir un valor para calificadores sin valor

- Estos calificadores no deben tener un elemento “INSDQualifier\_value” vacío

# Valores de los calificadores

## Tipos de formato - “Texto libre”

- Muchos calificadores tienen un formato de valor de “texto libre”.
- En el párrafo 3.n) de la Norma ST.26 se define “texto libre” como un “(...) *tipo de formato de valor para ciertos calificadores, que se presenta en forma de una frase de texto descriptiva u otro formato especificado (como se indica en el Anexo I)*”.
- Los valores de calificador de texto libre se limitan a 1.000 caracteres (párrafo 86 de la Norma ST.26)
- Existe un subconjunto de calificadores con formatos de valor de “texto libre” que son “*dependientes del idioma*”
- El texto libre “dependiente del idioma” puede “requerir traducción para los procedimientos internacionales, nacionales o regionales”.  
[Párrafo 3.o) de la Norma ST.26]

# Valores de los calificadores

## Tipos de formato - “Texto libre”

- Los calificadores que tienen valores de “texto libre dependiente del idioma” pueden requerir traducción
- ¿Cómo se determina si un calificador con un formato de valor de texto libre es dependiente del idioma?
  - En el Cuadro 5 de la Sección 6 del Anexo I de la Norma ST.26 se enumeran todos los calificadores de secuencias de nucleótidos con formato de valor de texto libre dependiente del idioma.
  - En el Cuadro 6 de la Sección 8 del Anexo I de la Norma ST.26 se enumeran todos los calificadores de secuencias de aminoácidos con formato de valor de texto libre dependiente del idioma.
  - Obsérvese el “formato de valor obligatorio” en la descripción del calificador

6.5.	Qualifier	cell_type
	Definition	cell type from which the sequence was obtained
	Mandatory value format	<del>free text</del> Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>leukocyte</INSDQualifier_value>

# Valores de los calificadores

## Tipos de formato - “Texto libre”

6.20.	Qualifier	frequency
	Definition	frequency of the occurrence of a feature
	Mandatory value format	<b>free text</b> representing the proportion of a population carrying the feature expressed as a fraction
	Example	<INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value>
6.21.	Qualifier	function
	Definition	function attributed to a sequence
	Mandatory value format	<b>free text</b> Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>essential for recognition of cofactor </INSDQualifier_value>
	Comment	The function qualifier is used when the gene name and/or product name do not convey the function attributable to a sequence.

# Valores de los calificadores

## Tipos de formato - “Texto libre dependiente del idioma”

- Los valores de los calificadores de “texto libre dependiente del idioma” pueden incluirse en dos idiomas en la lista de secuencias en formato XML: inglés y otro idioma distinto al inglés (ST.26, párrafo 87)



# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

- Los valores de calificador de “texto libre dependiente del idioma” pueden incluirse en dos idiomas en la lista de secuencias en formato XML: inglés y otro idioma distinto al inglés (ST.26, párrafo 87)
- En el elemento “INSDQualifier\_value” se debe proporcionar un valor de calificador dependiente del idioma en inglés

# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

- Los valores de calificador de “texto libre dependiente del idioma” pueden incluirse en dos idiomas en la lista de secuencias en formato XML: inglés y otro idioma distinto al inglés (ST.26, párrafo 87)
- En el elemento “INSDQualifier\_value” se debe proporcionar un valor de calificador dependiente del idioma en inglés
- En el elemento “NonEnglishQualifier\_value” debe proporcionarse un valor de calificador dependiente del idioma en otro idioma distinto del inglés

# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

- Los valores de calificador de “texto libre dependiente del idioma” pueden incluirse en dos idiomas en la lista de secuencias en formato XML: inglés y otro idioma distinto al inglés (ST.26, párrafo 87)
- En el elemento “INSDQualifier\_value” se debe proporcionar un valor de calificador dependiente del idioma en inglés
- En el elemento “NonEnglishQualifier\_value” debe proporcionarse un valor de calificador dependiente del idioma en otro idioma distinto del inglés
- Se permite un elemento “NonEnglishQualifier\_value” ÚNICAMENTE para calificadores con formato de valor de texto libre dependiente del idioma [párrafo 87.b) de la Norma ST.26]

# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

```

<SequenceData sequenceIDNumber="2">
  <INSDSeq>
    <INSDSeq_length>29</INSDSeq_length>
    <INSDSeq_moltype>AA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>SOURCE</INSDFeature_key>
        <INSDFeature_location>1..29</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>ORGANISM</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
          - <INSDQualifier>
            <INSDQualifier_name>MOL_TYPE</INSDQualifier_name>
            <INSDQualifier_value>protein</INSDQualifier_value>
          </INSDQualifier>
          - <INSDQualifier id="q1">
            <INSDQualifier_name>NOTE</INSDQualifier_name>
            <INSDQualifier_value>Synthetic peptide antigen fragment</INSDQualifier_value>
            <NonEnglishQualifier_value>Synthetisches Peptidantigenfragment</NonEnglishQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>GSLSDVRKDV EKRIDKALEAFK NKMDKEK</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="3">

```

# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

- En los casos en que una lista de secuencias XML contiene valores de calificadores en un idioma que no es inglés:

1. El elemento raíz de la lista de secuencias XML debe contener un atributo “nonEnglishFreeTextLanguageCode” con el código de dos letras del idioma correspondiente [párrafos 43 y 87.b) de la Norma ST.26];

```
<ST26SequenceListing dtdVersion="V1_3" fileName="st26-annex-iii-sequence-listing-specimen.xml" softwareName="WIPO  
Sequence" softwareVersion="1.0" productionDate="2022-01-01" originalFreeTextLanguageCode="de"  
nonEnglishFreeTextLanguageCode="de">
```

# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

- En los casos en que una lista de secuencias XML contiene valores de calificadores en un idioma que no es inglés:
  1. El elemento raíz de la lista de secuencias XML debe contener un atributo “nonEnglishFreeTextLanguageCode” con el código de dos letras del idioma correspondiente [párrafos 43 y 87.b) de la Norma ST.26];
  2. TODOS los calificadores dependientes del idioma en de la lista de secuencia deben contener valores en el idioma indicado en el atributo “nonEnglishFreeTextLanguageCode”;

# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

- En los casos en que una lista de secuencias XML contiene valores de calificadores en un idioma que no es inglés:
  1. El elemento raíz de la lista de secuencias XML debe contener un atributo “nonEnglishFreeTextLanguageCode” con el código de dos letras del idioma correspondiente [párrafos 43 y 87.b) de la Norma ST.26];
  2. TODOS los calificadores dependientes del idioma en la lista de secuencia deben contener valores en el idioma indicado en el atributo “nonEnglishFreeTextLanguageCode”;
  3. Cuando hay tanto un elemento “NonEnglishQualifier\_value” como “INSDQualifier\_value” para un único calificador, la información contenida en ambos elementos deberá ser equivalente [párrafo 87.c) de la Norma ST.26].

# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

- ¿Qué es el atributo “id” en INSDQualifier?
- El párrafo 87.d) de la Norma ST.26 dice lo siguiente: *“Para los calificadores dependientes del idioma, el elemento INSDQualifier podrá incluir un atributo id opcional. El valor de este atributo debe estar en formato “q”, seguido de un número entero positivo, por ejemplo, “q23”, y deberá ser único para un elemento INSDQualifier, es decir, el valor del atributo solo deberá utilizarse una vez en un archivo de listas de secuencias.”*

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



# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

- Los atributos “id” en INSDQualifier:
  - identifican de forma exclusiva los valores de los calificadores que pueden requerir una traducción para su exportación en formato XLIFF por parte de *WIPO Sequence*;
  - son facultativos;
  - solo se permiten para calificadores dependientes del idioma;
  - deben ser únicos en una lista de secuencias;
  - se añaden automáticamente en un archivo XML generado por *WIPO Sequence*.

# Valores de los calificadores

## Tipos de formato – “Texto libre dependiente del idioma”

- Atributo “originalFreeTextLanguageCode”

```
<ST26SequenceListing dtdVersion="V1_3" fileName="st26-annex-iii-sequence-listing-specimen.xml" softwareName="WIPO  
Sequence" softwareVersion="1.0" productionDate="2022-01-01" originalFreeTextLanguageCode="de"  
nonEnglishFreeTextLanguageCode="de">
```

- El párrafo 43 de la Norma ST.26 define el atributo “originalFreeTextLanguageCode” como el “código del idioma (...) para el idioma original único en que se prepararon los calificadores de texto libre dependiente del idioma.”
- Este atributo es FACULTATIVO

Situaciones especiales:  
Moléculas híbridas DNA/RNA  
Uracilo en ADN y  
timina en ARN

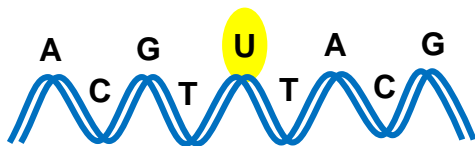
# Situaciones especiales

- Recordatorio: el símbolo “u” de uracilo no se permite en la Norma ST.26
- En el ADN, la “t” es timina
- En el ARN, la “t” es uracilo
- Dos casos que considerar:
  1. Molécula de ADN con nucleobase de uracilo o molécula de ARN con nucleobase de timina;
  2. Molécula híbrida de ADN/ARN

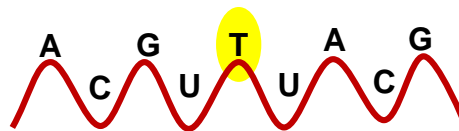
# Situaciones especiales

## Uracilo en el ADN y timina en el ARN

- Si una secuencia tiene un esqueleto de ADN con una nucleobase de uracilo o un esqueleto de ARN con una nucleobase de timina – descripción como “nucleótido modificado”



ADN



ARN

Se aplica el párrafo 14 de la Norma ST.26:

*“14. El símbolo “t” se interpretará como timina en ADN y uracilo en ARN. El uracilo en ADN o la timina en ARN se considerará como un nucleótido modificado y deberá describirse detalladamente en el cuadro de características tal como se prevé en el párrafo 19.”*

# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia de ARN:

5' -cguccacgugtccgaggua-3'

- Observe la “timina” en la posición 12. Este residuo debe anotarse como un nucleótido modificado.
- En el párrafo 19 de la Norma ST.26 se establece lo siguiente: *El uracilo en ADN o la timina en ARN se consideran nucleótidos modificados y deberán representarse en la secuencia por una “t” y describirse con detalle en el cuadro de características utilizando la clave de caracterización “modified\_base”, el calificador “mod\_base” con el valor calificador “OTHER” y el calificador “note” con el valor calificador “uracil” o “thymine”, respectivamente.*

# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia de ARN:

5' -cguccacgugtccgaggua-3'

- Observe la “timina” en la posición 12. Este residuo debe anotarse como un nucleótido modificado.
- El párrafo 19 de la Norma ST.26 establece: “El uracilo en ADN o la timina en ARN se consideran nucleótidos modificados y deberán representarse en la secuencia por una “t” y describirse con detalle en el cuadro de características utilizando la clave de caracterización “*modified\_base*”, el calificador “*mod\_base*” con el valor calificador “OTHER” y el calificador “*note*” con el valor calificador “uracil” o “thymine”, respectivamente.”

# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia de ARN:

5' -cguccccacgugtccgaggua-3'



Todos los residuos de uracilo deben representarse con el símbolo “t”. Por lo tanto, la secuencia debe representarse en la lista de secuencias como:

cgtccccacgtgtccgaggta



# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia de ARN:

**5' -cgucccacgugtccgaggua-3'**

- ✓ Todos los residuos de uracilo deben representarse con el símbolo “t”. Por lo tanto, la secuencia debe representarse en la lista de secuencias como:

**cgtccccacgtgtccgaggta**

- ✓ Clave de caracterización “modified\_base” con localización “12”

# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia de ARN:

**5' -cgucccacgugtccgaggua-3'**

- ✓ Todos los residuos de uracilo deben representarse con el símbolo “t”. Por lo tanto, la secuencia debe representarse en la lista de secuencias como:

**cgtcccacgtgtccgaggta**

- ✓ Clave de caracterización “modified\_base” con localización “12”
- ✓ Calificador “mod\_base” con valor “OTHER”

# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia de ARN:

**5' -cgucccacgugtccgaggua-3'**

- ✓ Todos los residuos de uracilo deben representarse con el símbolo “t”. Por lo tanto, la secuencia debe representarse en la lista de secuencias como:

**cgtcccacgtgtccgaggta**

- ✓ Clave de caracterización “modified\_base” con localización “12”
- ✓ Calificador “mod\_base” con valor “OTHER”
- ✓ Calificador “note” con valor “timina”

# Situaciones especiales

## Secuencias de ADN y ARN

```

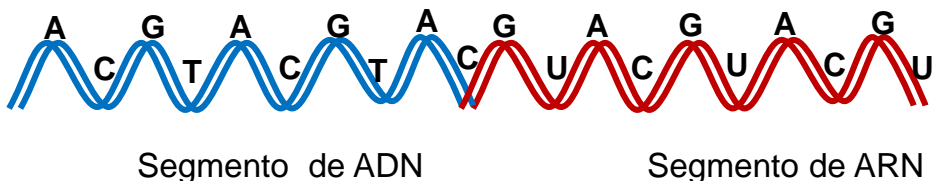
<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length>20</INSDSeq_length>
    <INSDSeq_moltype>RNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..20</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>other RNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q7">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
      <INSDFeature>
        <INSDFeature_key>modified_base</INSDFeature_key>
        <INSDFeature_location>12</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q8">
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>thymine</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>cgctcccacgtgtccgaggta</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

# Situaciones especiales

## Secuencias híbridas de ADN/ARN

- Si una secuencia es una molécula híbrida (es decir, parte del esqueleto es ADN y parte es ARN):



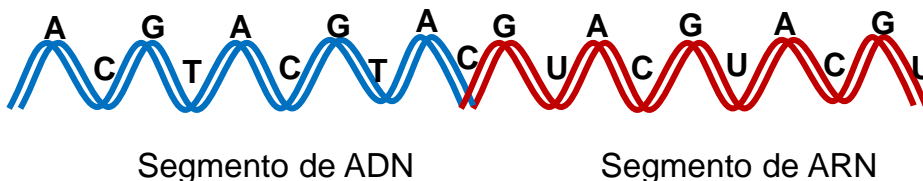
Se aplica el párrafo 55 de la Norma ST.26:

*“55. Para las secuencias de nucleótidos que contienen segmentos de ADN y ARN de uno o más nucleótidos, el tipo de molécula deberá indicarse como ADN. La molécula combinada ADN/ARN deberá describirse con detalle en el cuadro de características, utilizando la clave de caracterización “source” y el calificador obligatorio “organism” con el valor “synthetic construct” y el calificador obligatorio “mol\_type” con el valor “other DNA”. Cada segmento de ADN y ARN de la molécula combinada ADN/ARN deberá describirse con detalle mediante la clave de caracterización “misc\_feature” y el calificador “note”, lo que indica si el segmento es un ADN o un ARN.”*

# Situaciones especiales

## Secuencias híbridas de ADN/ARN

- Si una secuencia es una molécula híbrida (es decir, parte del esqueleto es ADN y parte es ARN):



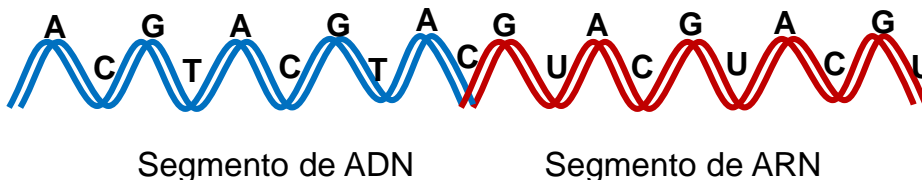
Se aplica el párrafo 55 de la Norma ST.26:

*“55. Para las secuencias de nucleótidos que contienen segmentos de ADN y ARN de uno o más nucleótidos, el tipo de molécula deberá indicarse como ADN. La molécula combinada ADN/ARN deberá describirse con detalle en el cuadro de características, utilizando la clave de caracterización “source” y el calificador obligatorio “organism” con el valor “synthetic construct” y el calificador obligatorio “mol\_type” con el valor “other DNA”. Cada segmento de ADN y ARN de la molécula combinada ADN/ARN deberá describirse con detalle mediante la clave de caracterización “misc\_feature” y el calificador “note”, lo que indica si el segmento es un ADN o un ARN.”*

# Situaciones especiales

## Secuencias híbridas de ADN/ARN

- Si una secuencia es una molécula híbrida (es decir, parte del esqueleto es ADN y parte es ARN):



Se aplica el párrafo 55 de la Norma ST.26:

*“55. Para las secuencias de nucleótidos que contienen segmentos de ADN y ARN de uno o más nucleótidos, el tipo de molécula deberá indicarse como ADN. La molécula combinada ADN/ARN deberá describirse con detalle en el cuadro de características, utilizando la clave de caracterización “source” y el calificador obligatorio “organism” con el valor “synthetic construct” y el calificador obligatorio “mol\_type” con el valor “other DNA”. Cada segmento de ADN y ARN de la molécula combinada ADN/ARN deberá describirse con detalle mediante la clave de caracterización “misc\_feature” y el calificador “note”, lo que indica si el segmento es un ADN o un ARN.”*

# Situaciones especiales

## Secuencias híbridas de ADN/ARN

- Si una secuencia es una molécula híbrida (es decir, parte del esqueleto es ADN y parte es ARN):



Segmento de ADN

Segmento de ARN

Se aplica el párrafo 55 de la Norma ST.26:

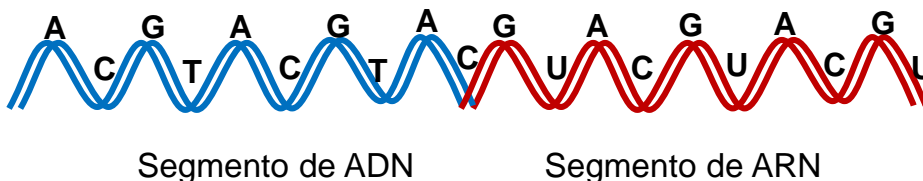
*“55. Para las secuencias de nucleótidos que contienen segmentos de ADN y ARN de uno o más nucleótidos, el tipo de molécula deberá indicarse como ADN. La molécula combinada ADN/ARN deberá describirse con detalle en el cuadro de características, utilizando la clave de caracterización “source” y el calificador obligatorio “organism” con el valor “synthetic construct” y el calificador obligatorio “mol\_type” con el valor “other DNA”. Cada segmento de ADN y ARN de la molécula combinada ADN/ARN deberá describirse con detalle mediante la clave de caracterización “misc\_feature” y el calificador “note”, lo que indica si el segmento es un ADN o un ARN.”*



# Situaciones especiales

## Secuencias híbridas de ADN/ARN

- Si una secuencia es una molécula híbrida (es decir, parte del esqueleto es ADN y parte es ARN):



Se aplica el párrafo 55 de la Norma ST.26:

*“55. Para las secuencias de nucleótidos que contienen segmentos de ADN y ARN de uno o más nucleótidos, el tipo de molécula deberá indicarse como ADN. La molécula combinada ADN/ARN deberá describirse con detalle en el cuadro de características, utilizando la clave de caracterización “source” y el calificador obligatorio “organism” con el valor “synthetic construct” y el calificador obligatorio “mol\_type” con el valor “other DNA”. Cada segmento de ADN y ARN de la molécula combinada ADN/ARN deberá describirse con detalle mediante la clave de caracterización “misc\_feature” y el calificador “note”, lo que indica si el segmento es un ADN o un ARN.”*

# Situaciones especiales

## Secuencias híbridas de ADN/ARN

En una solicitud se divulga la siguiente secuencia híbrida de ADN/ARN:

**5' -ACCTGCcguccacguguccgagguaGCATTA-3'**

donde los símbolos en mayúscula representan una parte de ADN y los símbolos en minúscula representan una parte de ARN

- Los residuos 1-6 y 27-32 son ADN; los residuos 7-26 son ARN
- Por examinar:
  1. Designación del organismo
  2. Tipo de molécula y mol\_type
  3. Identificación de segmentos de ADN y ARN

# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia híbrida ADN/ARN:

5' -ACCTGCcgucccacguguccgagguaGCATTA-3'

donde los símbolos en mayúscula representan la parte de ADN y los símbolos en minúscula la parte de ARN.

- El párrafo 55 de la Norma ST.26 establece lo siguiente: *“(...) el tipo de molécula deberá indicarse como ADN. La molécula combinada ADN/ARN deberá describirse con detalle en el cuadro de características, utilizando la clave de caracterización “source” y el calificador obligatorio “organism” con el valor “synthetic construct” y el calificador obligatorio “mol\_type” con el valor “other DNA”.*
- ✓ Tipo de molécula = “DNA”
- ✓ Nombre del organismo calificador = “synthetic construct”
- ✓ Calificador “mol\_type” = “other DNA”

# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia híbrida ADN/ARN:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

donde los símbolos en mayúscula representan la parte de ADN y los símbolos en minúscula la parte de ARN.

- ✓ Todos los residuos de uracilo deben representarse con el símbolo “t”. Por lo tanto, la secuencia debe representarse en la lista de secuencias como:

**acctgccgtcccacgtgtccgaggtagcatta**

# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia híbrida ADN/ARN:

5' -ACCTGCcgucccaguguccgagguaGCATTA-3'

donde los símbolos en mayúscula representan la parte de ADN y los símbolos en minúscula la parte de ARN.

- El párrafo 55 de la Norma ST.26 establece lo siguiente: *“Cada segmento de ADN y ARN de la molécula combinada ADN/ARN deberá describirse con detalle mediante la clave de caracterización “misc\_feature” y el calificador “note”, lo que indica si el segmento es un ADN o un ARN.”*



Tres segmentos = tres claves de caracterización “misc\_feature”

# Situaciones especiales

## Secuencias de ADN y ARN

En una solicitud se divulga la siguiente secuencia híbrida ADN/ARN:

5' -ACCTGCcguccacguguccgagguaGCATTA-3'

donde los símbolos en mayúscula representan la parte de ADN y los símbolos en minúscula la parte de ARN.

- ✓ Segmento 1, residuos 1-6:  
clave de caracterización “misc\_feature” con localización “1..6”  
Calificador “note” con valor “DNA”
- ✓ Segmento 2, residuos 7-26:  
clave de caracterización “misc\_feature” con localización “7..26”  
Calificador “note” con valor “RNA”
- ✓ Segmento 3, residuos 27-32  
clave de caracterización “misc\_feature” con localización “27..32”  
Calificador “note” con valor “DNA”

# Situaciones especiales

## Secuencias de ADN y ARN

```

<SequenceData sequenceIDNumber="4">
  <INSDSeq>
    <INSDSeq_length>32</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..32</INSDFeature_location>
        <INSDFeature_quals>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>other DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q10">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_quals>
      </INSDFeature>
    </INSDSeq_feature-table>
  </INSDSeq>
</SequenceData>

```





# ¿Preguntas?

# Análogos de nucleótidos, D-aminoácidos y secuencias ramificadas

# Nuevos tipos de moléculas

## Análogos de nucleótidos

- Las secuencias de ácido nucleico que contienen uno o más análogos de nucleótidos están sujetas a lo dispuesto en la Norma ST.26.
- Los análogos de nucleótidos se incluyen en la definición de “nucleótido” conforme a lo dispuesto en el párrafo 3.g)2) de la Norma ST.26:

*“el análogo de un 5’ monofosfato de 2’ desoxirribosa o 5’ monofosfato de ribosa, que al formar el esqueleto de un análogo de ácido nucléico, da lugar a una disposición de las nucleobases que imita la disposición de las nucleobases en ácidos nucleicos que contienen un esqueleto de 5’ monofosfato de 2’ desoxirribosa o 5’ monofosfato de ribosa, en la que el análogo de ácido nucléico es capaz de aparear las bases con un ácido nucleico complementario”*

- Algunos análogos de nucleótidos comunes son los ácidos peptidonucleicos (PNA), los ácidos nucleicos glicólicos (GNA), los ácidos nucleicos treósicos y los morfolinos
- Deben representarse en el sentido de izquierda a derecha que imite el sentido 5’ a 3. (ST.26, párrafo 11)

# Nuevos tipos de moléculas

## Análogos de nucleótidos

En una solicitud de patente se divulga la siguiente secuencia de ácido nucleico glicólico (ANG):

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

donde el extremo final PO4 de la secuencia imita el extremo 5' de una secuencia de ADN.

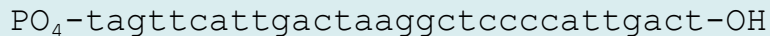


Esta secuencia debe figurar en una lista de secuencias

# Nuevos tipos de moléculas

## Análogos de nucleótidos

En una solicitud de patente se divulga la siguiente secuencia de ácido nucleico glicólico (ANG):



donde el extremo PO<sub>4</sub> de la secuencia imita el extremo 5' de una secuencia de ADN.

- ✓ Esta secuencia debe figurar en una lista de secuencias
- ✓ El extremo PO<sub>4</sub> imita el extremo 5', por lo que la secuencia debe representarse en la dirección mostrada

# Nuevos tipos de moléculas

## Análogos de nucleótidos

En una solicitud de patente se divulga la siguiente secuencia de ácido nucleico glicólico (ANG):

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

donde el extremo PO4 de la secuencia imita el extremo 5' de una secuencia de ADN.

- ✓ Esta secuencia debe figurar en una lista de secuencias
- ✓ El extremo PO4 imita el extremo 5', por lo que la secuencia debe representarse en la dirección mostrada
- ✓ La secuencia completa debe anotarse con la clave de caracterización "modified\_base", un calificador "mod\_base" con el valor "OTHER" y un calificador de nota que incluya el nombre completo, no abreviado, de los nucleótidos modificados, como "ácidos nucleicos glicólicos"

# Nuevos tipos de moléculas

## Análogos de nucleótidos

```

- <INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>1..30</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    - <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>2,3-dihydroxypropyl nucleosides (glycol nucleic acids)</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>tagttcattgactaaggctccccattgact</INSDSeq_sequence>
</INSDSeq>

```

*Nota: se puede encontrar un análisis detallado de este ejemplo en el ejemplo 3.g)-4 del Anexo VI de la Norma ST.26 de la OMPI*

# Nuevos tipos de moléculas

## D-aminoácidos

- Las secuencias de ácido nucleico que contienen uno o más D-aminoácidos están sujetas a lo dispuesto en la Norma ST.26.
- Los D-aminoácidos se incluyen en la definición de “aminoácido” que figura en el párrafo 3.a) de la Norma ST.26:

*“por 'aminoácido' se entenderá todo aminoácido que pueda ser representado mediante cualquiera de los símbolos descritos en el Anexo I (véase el Cuadro 3 de la Sección 3). Quedan comprendidos entre tales aminoácidos, los D-aminoácidos y los aminoácidos que contienen cadenas laterales modificadas o sintéticas.”*

- Los D-aminoácidos deben representarse en la secuencia como el símbolo del L-aminoácido correspondiente no modificado, siempre que sea posible
- deben describirse en el cuadro de características como un aminoácido modificado



# Nuevos tipos de moléculas

## D-aminoácidos

En una solicitud de patente se describe la secuencia siguiente:

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

✓ Esta secuencia debe figurar en una lista de secuencias

# Nuevos tipos de moléculas

## D-aminoácidos

En una solicitud de patente se describe la secuencia siguiente:

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

- ✓ Esta secuencia debe figurar en una lista de secuencias
- ✓ Debe representarse como: AGKLGGM
- ✓ La alanina de la posición 1, el ácido glutámico de la posición 2 y la metionina de la posición 6 deben anotarse cada uno con una clave de caracterización "SITE" y un calificador "NOTE" con el nombre completo no abreviado del aminoácido correspondiente

# Nuevos tipos de moléculas

## D-aminoácidos

```

- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>1</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q4">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>D-alanine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>2</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q5">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>D-glutamic acid</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>6</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q6">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>D-methionine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>AEKLG</INSDSeq_sequence>

```

*Nota: se puede encontrar un análisis detallado de un ejemplo similar en el ejemplo 3.a)-1 del Anexo VI de la Norma ST.26 de la OMPI*

# Nuevos tipos de moléculas

## Secuencias ramificadas

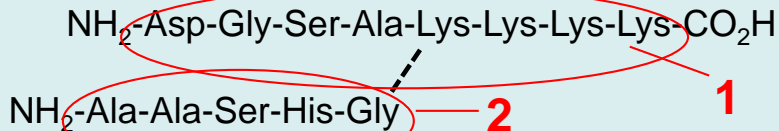
- Las secuencias ramificadas de ácido nucleico y las secuencias de ramificadas de aminoácidos están sujetas a lo dispuesto en la Norma ST.26
- Las regiones lineales de las secuencias ramificadas deben incluirse en una lista de secuencias cuando contienen diez o más nucleótidos específicamente definidos o cuatro o más aminoácidos específicamente definidos. (Norma ST.26 de la OMPI, párrafo 7)
- Cada región lineal de una secuencia ramificada que cumpla el requisito de longitud mínima debe incluirse como una secuencia distinta con su propio identificador de secuencia "SEQ ID"
- Se debe examinar el número de residuos específicamente definidos en cada región lineal individual, no el número total de residuos específicamente definidos en la estructura



# Nuevos tipos de moléculas

## Secuencias ramificadas

En una solicitud de patente se describe un péptido con la siguiente secuencia:



donde ----indica una unión amida entre el carboxi-terminal de la glicina y la cadena lateral de la lisina

- ✓ Ambas regiones lineales contienen  $\geq 4$  aminoácidos específicamente definidos, por lo que se requiere que ambas figuren en una lista de secuencias
- ✓ Cada región lineal debe incluirse como una secuencia distinta con su propio identificador de secuencia "SEQ ID"



# Nuevos tipos de moléculas

## Secuencias ramificadas

1

```

- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>5</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q4">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>Lysine side chain is amide bonded to the carboxy
        terminus of the glycine in SEQ ID NO:2</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>DGSAKKKK</INSDSeq_sequence>

```

2

```

- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>5</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q11">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>glycine carboxy terminus is amide bonded to the side
        chain of the lysine in SEQ ID NO:1, position 5</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>AASHG</INSDSeq_sequence>

```

*Nota: se puede encontrar un análisis detallado de un ejemplo similar en el ejemplo 7.b)3 del Anexo VI de la Norma ST.26.*



# Variantes de secuencias

# Variantes de secuencias

- En el párrafo 3.m) de la Norma ST.26 se define “**variante de secuencia**” como

*“una secuencia de nucleótidos o aminoácidos que contiene una o más diferencias con respecto a una secuencia principal. Estas diferencias pueden incluir residuos alternativos (véanse los párrafos 15 y 27), residuos modificados (véanse los párrafos 3.g), 3.h), 16 y 29), supresiones, inserciones y sustituciones. Véanse los párrafos 93 a 95.”*

- La forma en que se divulga una variante de secuencia determinará cómo debe representarse en una lista de secuencias.
- En los párrafos 93 a 95 de la Norma ST.26 se rige cómo deben representarse las variantes.

# Variantes de secuencias

## Párrafo 93

*Párrafo 93: Toda secuencia principal y toda variante de esa secuencia, cada una de ellas divulgada mediante la enumeración de sus residuos y contemplada en el párrafo 7, deberá incluirse en la lista de secuencias y se le asignará su propio identificador de secuencia.*



Si cada variante se enumera por separado, cada una debe tener su propio identificador de secuencia “SEQ ID”.

# Variantes de secuencias

## Párrafo 93

Una solicitud de patente incluye una figura con el siguiente alineamiento múltiple de secuencias:

Consensus	LEG <b>n</b> EQFINA <b>ak</b> IIRH <b>k</b> Y <b>nrk</b> TL <b>n</b> NDI <b>m</b> LIK
Homo sapiens	LEGNEQFINAAKIIIRHPQYDRKTIINNDIMLIK
Pongo abelii	LEGNEQFINAAKIIIRHPQYDRKTVNNDIMLIK
Pan paniscus	LEGNEQFINAAKIIIRHPKYNRITINNDIMLIK
Rhinopithecus bieti	LEGNEQFINATKIIIRHPKYNGNTIINNDIMLIK
Rhinopithecus roxellana	LEGNEQFINATQIIIRHPKYNGNTIINNDIMLIK

Las letras en minúscula representan los residuos de aminoácidos predominantes entre las secuencias alineadas.

- ✓ Cada una de las seis secuencias enumeradas debe incluirse en la lista de secuencias como una secuencia distinta con su propio identificador SEQ ID.

*Nota: se puede encontrar un análisis detallado de un ejemplo similar en el ejemplo 93-3 del Anexo VI de la Norma ST.26 de la OMPI*

# Variantes de secuencias

## Anotación adecuada

- La clave de caracterización adecuada para anotar una variante depende del tipo de molécula y de la naturaleza de la variación:

Tipo de secuencia	Clave de caracterización	Calificador	Utilización
Acido nucleico	variation	replace o note	Mutaciones y polimorfismos que existen en estado natural, por ejemplo, los alelos o los polimorfismos de longitud de los fragmentos de restricción.
Acido nucleico	misc_difference	replace o note	Variabilidad introducida artificialmente, por ejemplo, mediante manipulación genética o síntesis química.
Aminoácido	VAR_SEQ	NOTE	Variante producida por empalme diferencial, utilización de promotor alternativo, iniciación alternativa y o desplazamiento del marco ribosomal.
Aminoácido	VARIANT	NOTE	Todo tipo de variante para el cual VAR_SEQ no es aplicable.

# Variantes de secuencias

## Símbolo de ambigüedad más restrictivo

- En los párrafos 15 y 27 de la Norma ST.26 se indica que, cuando se requiera un símbolo de ambigüedad, “debería utilizarse el símbolo más restrictivo (...)”

¿Eso qué significa?

Por ejemplo:

En una secuencia de nucleótidos, si una posición puede ser “a o c”, úsese el símbolo de ambigüedad “m” en lugar de “n”.

En una secuencia de aminoácidos, si una posición puede ser “L or I”, úsese el símbolo de ambigüedad “J” en lugar de “X”.

Recuerde que “n” y “X” tienen valores por defecto, por lo que cada vez que se usan “n” o “X” se utilizan para algo distinto al valor por defecto, se requiere una anotación.

# Variantes de secuencias

## Párrafo 94

*Párrafo 94: Toda variante de secuencia, divulgada como una única secuencia con residuos alternativos enumerados en una o más posiciones, se deberá incluir en la lista de secuencias y se debería representar con una única secuencia, en la cual los residuos alternativos enumerados se representan con el símbolo de ambigüedad más restrictivo (véanse los párrafos 15 y 27).*



Si las variantes NO se enumeran por separado, sino que se muestran simplemente como residuos variables en la secuencia principal, no es necesario que tengan un identificador "SEQ ID" distinto.

# Variantes de secuencias

## Párrafo 94

En una solicitud de patente se divulga un péptido de la secuencia:

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

- ✓ Puede incluirse en la lista de secuencias como una única secuencia
- ✓ La representación preferida de la secuencia es: GGGJATX
- ✓ [Leu or Ile] debe representarse con el símbolo de ambigüedad más restrictivo “J”
- ✓ [Ser or Thr] debe representarse con el símbolo “X” junto con la clave de caracterización “VARIANT” y un calificador "NOTE" que indique que X es serina o treonina

*Nota: se puede encontrar un análisis detallado de este ejemplo en el ejemplo 94-1 del Anexo VI de la Norma ST.26 de la OMPI*



# Variantes de secuencias

## Párrafo 95

*Párrafo 95: Toda variante de secuencia, divulgada únicamente por referencia a una o varias supresiones, inserciones o sustituciones en una secuencia principal en la lista de secuencias, debería incluirse en la lista de secuencias. Cuando se indica en la lista de secuencias, tal variante de secuencia:*

*a) podrá representarse por anotación de la secuencia principal, si contiene una o varias variaciones en una única localización o en varias localizaciones distintas y esas variaciones aparecen de forma independiente;*

*b) debería representarse como una secuencia distinta y se le asignará su propio identificador de secuencia, si contiene variaciones en varias localizaciones distintas y esas variaciones aparecen de forma interdependiente; y*

*c) deberá representarse como una secuencia distinta y se le asignará su propio identificador de secuencia, si contiene una secuencia que ha sido insertada o sustituida que contiene más de 1000 residuos (véase el párrafo 86).*



Por “referencia a una o varias supresiones, inserciones o sustituciones” se entiende que las variantes se divulgan en prosa.

# Variantes de secuencias

## Párrafo 95.a)

Una solicitud contiene la divulgación siguiente:

Fragmento de péptido 1: Gly-Leu-Pro-Xaa-Arg-Ile-Cys  
donde Xaa es cualquier aminoácido

\* \* \*

En otra realización, el fragmento de péptido 1 es Gly-Leu-Pro-Xaa-Arg-Ile-Cys, donde Xaa puede ser Val, Thr, o Asp...

\* \* \*

En otra realización, el fragmento de péptido 1 es Gly-Leu-Pro-Xaa-Arg-Ile-Cys donde Xaa puede ser Val...

*La secuencia principal, Gly-Leu-Pro-Xaa-Arg-Ile-Cys, contiene una o varias variaciones en una única localización y esas variaciones aparecen de forma independiente*

# Variantes de secuencias

## Párrafo 95.a)

- ✓ La secuencia principal debe incluirse en la lista de secuencias, y las variantes pueden representarse por anotación de la secuencia principal
- ✓ La realización más abarcadora debe ser la versión incluida en la lista de secuencia; en este ejemplo, la versión en la que Xaa es “cualquier aminoácido”
- ✓ La secuencia debe representarse como GLPXRIC y requiere la clave de característica “VARIANT” en la posición 4 y un calificador “note” que indique que Xaa es “cualquier aminoácido”
- ✓ Aunque no es obligatorio, se recomienda incluir las tres variantes siguientes en la lista de secuencias como secuencias distintas:

GLPVRIC

GLPTRIC

GLPDRIC

*Nota: se puede encontrar un análisis detallado de este ejemplo en el ejemplo 95.a)-1 del Anexo VI de la Norma ST.26 de la OMPI.*

# Variantes de secuencias

## Párrafo 95.b)

En una solicitud de patente se describe la siguiente secuencia de consenso:

aatg $n_1$ cccacgaatg $n_2$ cac

donde  $n_1$  y  $n_2$  pueden ser a, t, g, o c.

Distintas variantes de secuencia se divulgan de la manera siguiente:

si  $n_1$  es a,  $n_2$  es t, g, o c;

si  $n_1$  es t,  $n_2$  es a, g, o c;

si  $n_1$  es g,  $n_2$  es t, a, o c;

si  $n_1$  es c,  $n_2$  es t, g, o a;

*La secuencia de consenso contiene variaciones en muchas localizaciones distintas y esas variaciones aparecen de forma interdependiente.*

# Variantes de secuencias

## Párrafo 95.b)

- ✓ La secuencia de consenso debe incluirse en la lista de secuencias, y las variantes deben representarse como secuencias distintas
- ✓ La realización más abarcadora debe ser la versión incluida en la lista de secuencia: donde n1 y n2 “pueden ser a, t, g, o c”
- ✓ La secuencia debe representarse como: **aatgnccccacgaatgncac**
- ✓ No se requiere una anotación respecto de n, ya que "n" se interpreta como una opción entre "a", "c", "g" o "t" en ausencia de una clave de caracterización (véase el párrafo 15 de la Norma ST.26)
- ✓ Aunque no es obligatorio, se recomienda encarecidamente incluir las cuatro variantes siguientes en la lista de secuencias como secuencias distintas:

**aatgaccccacgaatgbcac** (b = t, g, o c)

**aatgtccccacgaatvcac** (v = a, g, o c)

**aatggccccacgaathcac** (h = t, a, o c)

**aatgcccccacgaatdcac** (d = t, g, o a)

*Nota: se puede encontrar un análisis detallado de un ejemplo similar en el ejemplo 95.b) del Anexo VI de la Norma ST.26 de la OMPI.*

# Variantes de secuencias

## Párrafo 95.c)

Una solicitud contiene la divulgación siguiente:

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

donde Xaa es Gly o una inserción de la secuencia

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

Una variante de secuencia contiene una secuencia que ha sido insertada o sustituida de más de 1.000 residuos

# Variantes de secuencias

## Párrafo 95.c)

- ✓ La variante en la que Xaa = una inserción de más de 1.000 residuos debe incluirse en la lista de secuencias como una secuencia distinta con su propio número identificador de secuencia "SEQ ID".
- ✓ La variante en la que Xaa = Gly también se incluirá en la lista de secuencias como una secuencia distinta con su propio número identificador de secuencia "SEQ ID".

...-MGLPRGRICK-...

# Variantes de secuencias

**P:** ¿Qué párrafo se aplica a la siguiente divulgación?

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



# Variantes de secuencias

**P:** ¿Qué párrafo se aplica a la siguiente divulgación?

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

**R:** Párrafo 94:

*“Toda variante de secuencia, divulgada como una única secuencia con residuos alternativos enumerados en una o más posiciones, se deberá incluir en la lista de secuencias y se debería representar con una única secuencia, en la cual los residuos alternativos enumerados se representan con el símbolo de ambigüedad más restrictivo.”*

# Variantes de secuencias

**P:** ¿Qué párrafo se aplica a la siguiente divulgación?

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

Un espacio en blanco en la tabla indica que un aminoácido en la variante es el mismo que el aminoácido correspondiente en la “Secuencia” y un “-” indica la eliminación del aminoácido correspondiente en la “Secuencia”.

# Variantes de secuencias

**P:** ¿Qué párrafo se aplica a la siguiente divulgación?

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

Un espacio en blanco en la tabla indica que un aminoácido en la variante es el mismo que el aminoácido correspondiente en la “Secuencia” y un “-” indica la eliminación del aminoácido correspondiente en la “Secuencia”.

**R:** Párrafo 93:

*“Toda secuencia principal y toda variante de esa secuencia, cada una de ellas divulgada mediante la enumeración de sus residuos y contemplada en el párrafo 7, deberá incluirse en la lista de secuencias y se le asignará su propio identificador de secuencia.”*

# Variantes de secuencias

**P:** ¿Qué párrafo se aplica a la siguiente divulgación?

En una figura se divulga la secuencia siguiente:

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

La especificación contiene la siguiente información relativa a la Secuencia 1:

... en ciertas realizaciones, la valina en la posición 7 de la secuencia 1 se sustituye por alanina...

# Variantes de secuencias

**P:** ¿Qué párrafo se aplica a la siguiente divulgación?

En una figura se divulga la secuencia siguiente:

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

La especificación contiene la siguiente información relativa a la Secuencia 1:

... en ciertas realizaciones, la valina en la posición 7 de la secuencia 1 se sustituye por alanina...

**R:** Párrafo 95.a): *“Toda variante de secuencia, divulgada únicamente por referencia a una o varias supresiones, inserciones o sustituciones en una secuencia principal en la lista de secuencias, debería incluirse en la lista de secuencias. Cuando se indica en la lista de secuencias, tal variante de secuencia:*

*a) podrá representarse por anotación de la secuencia principal, si contiene una o varias variaciones en una única localización o en varias localizaciones distintas y esas variaciones aparecen de forma independiente;”*

# Variantes de secuencias

**P:** ¿Qué párrafo se aplica a la siguiente divulgación?

En una figura se divulga la secuencia siguiente:

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

La especificación contiene la siguiente información relativa a la Secuencia 1:

... si la valina en la posición 7 de la secuencia 1 se sustituye por alanina, la alanina en la posición 10 se sustituye por valina....

# Variantes de secuencias

**P:** ¿Qué párrafo se aplica a la siguiente divulgación?

En una figura se divulga la secuencia siguiente:

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

La especificación contiene la siguiente información relativa a la Secuencia 1:

... si la valina en la posición 7 de la secuencia 1 se sustituye por alanina, la alanina en la posición 10 se sustituye por valina....

**R:** Párrafo 95.b): *“Toda variante de secuencia, divulgada únicamente por referencia a una o varias supresiones, inserciones o sustituciones en una secuencia principal en la lista de secuencias, debería incluirse en la lista de secuencias. Cuando se indica en la lista de secuencias, tal variante de secuencia:*

*b) debería representarse como una secuencia distinta y se le asignará su propio identificador de secuencia, si contiene variaciones en varias localizaciones distintas y esas variaciones aparecen de forma interdependiente;”*

# ¿Preguntas?