

Comité de Normas Técnicas de la OMPI (CWS)

Novena sesión
Ginebra, 1 a 5 de noviembre de 2021

PROPUESTA DE REVISIÓN DE LA NORMA ST.26 DE LA OMPI

Documento preparado por la Oficina Internacional

INTRODUCCIÓN

1. En la reanudación su cuarta sesión, celebrada en marzo de 2016, el Comité de Normas Técnicas de la OMPI (CWS) aprobó la Norma ST.26 de la OMPI (véase el párrafo 52 del documento CWS/4BIS/16), denominada “Norma recomendada para la presentación de listas de secuencias de nucleótidos y aminoácidos en lenguaje extensible de marcado (XML)”.
2. Las revisiones adicionales de la Norma ST.26 de la OMPI, a saber, las versiones 1.1 y 1.2, fueron aprobadas por el CWS en sus sesiones quinta y sexta, respectivamente (véase el párrafo 41 del documento CWS/5/22 y el párrafo 112 del documento CWS/6/34). En su séptima sesión, celebrada en julio de 2019, el CWS aprobó una nueva revisión de la Norma, la versión 1.3 (véase el párrafo 133 del documento CWS/7/29). Recientemente, en su octava sesión, celebrada en diciembre de 2020, el CWS aprobó varias revisiones sustanciales y de redacción, y las publicó como la versión más reciente de la Norma ST.26, la 1.4 (véase el párrafo 49 del documento CWS/8/24).
3. El Equipo Técnico sobre Listas de Secuencias ha estado trabajando en el marco de la Tarea N.º 44, cuya descripción es:
“Prestar apoyo a la Oficina Internacional facilitando los requisitos de los usuarios y comentarios sobre la herramienta de autoría y validación de la Norma ST.26; prestar apoyo a la Oficina Internacional de la OMPI en la consiguiente revisión de las Instrucciones Administrativas del PCT; y preparar las revisiones necesarias de la Norma ST.26 de la OMPI.”

PROPUESTA DE REVISIÓN DE LA NORMA ST.26 DE LA OMPI

4. El Equipo Técnico sobre Listas de Secuencias ha preparado una propuesta de revisión de la Norma ST.26 de la OMPI para que el CWS la examine y, en su caso, la apruebe. La propuesta figura en los Anexos I a III del presente documento, donde el resaltado amarillo indica el texto nuevo y el resaltado púrpura indica el texto suprimido.
5. Los cambios propuestos en el cuerpo de la Norma ST.26 pueden resumirse como sigue:
 - a) todos los casos de la característica “SOURCE” de las secuencias de aminoácidos se han cambiado por “source” con arreglo a los últimos cambios de UniProt;
 - b) todos los casos de los calificadores “ORGANISM”, “MOL_TYPE” y “NOTE” se han cambiado por “organism”, “mol_type” y “note” con arreglo a los últimos cambios de UniProt;
 - c) todos los casos de “Oficina de patentes” se han cambiado por “OPI”;
 - d) se ha cambiado la posición de la sigla “DTD” para que figure junto a “Definiciones de tipo de documento”;
 - e) en el párrafo 3.g)i)2) se ha sustituido “análogos de nucleótidos” por “componentes del esqueleto”;
 - f) se ha modificado el segmento de código en el párrafo 44 para corregir las comillas utilizadas en softwareVersion, que deben ser: softwareVersion=“1.0”;
 - g) se ha editado el párrafo 46, ejemplos 1 y 2, para cambiar el código nonEnglishFreeTextLanguageCode de “jp” a “ja” y también para corregir la sangría de <FilingDate> a fin de alinearla con <ApplicationNumberText>; y
 - h) se ha editado el párrafo 97 para sustituir la expresión “...un subgrupo de...” por “...un valor distinto de...”.
6. Los cambios propuestos en los Anexos de la Norma ST.26 están resaltados para que el CWS los examine. La lista completa de cambios figura en los Anexos I a III del presente documento:
 - a) todos los casos de la característica “SOURCE” de las secuencias de aminoácidos se han cambiado por “source”, incluida la sección 7 del Anexo I, con arreglo a los últimos cambios de UniProt;
 - b) todos los casos de los calificadores “ORGANISM”, “MOL_TYPE” y “NOTE” se han cambiado por “organism”, “mol_type” y “note”, incluida la sección 8 del Anexo I, con arreglo a los últimos cambios de UniProt;
 - c) la cabecera de la segunda columna de los cuadros 1 y 3 y los cuadros A y B se ha sustituido por “Definición” por motivos de coherencia. También se ha añadido una columna con un código de 3 letras al cuadro B;
 - d) se han añadido calificadores al Anexo I con arreglo a las últimas actualizaciones de la INSDC;
 - e) se ha añadido un nuevo ejemplo 29-2 al Anexo VI para aclarar la inclusión del aminoácido no modificado correspondiente;
 - f) se ha editado la introducción del Anexo VI para justificar las respuestas a las preguntas “¿cuándo debería representarse un residuo modificado por el correspondiente residuo no modificado o por la variable 'n' o 'X'?” y “¿qué modificaciones dan lugar a un 'residuo modificado' y DEBERÁN ser anotadas?”, y se ha añadido texto para explicar la forma de representar las divulgaciones en las que la longitud de la secuencia puede variar en “secuencia más abarcadora”;

- g) se ha revisado el Anexo VI en general para garantizar la coherencia en el uso de “deberá”, “debería”, “preferiblemente” y “podrá” y evitar confusiones, ya que “preferiblemente” no se define en el párrafo 4 de la Norma ST.26;
- h) se ha editado el ejemplo 3.c)-2 del Anexo VI para cambiar “debería” por “deberá”, de modo que la frase siguiente diga:

'Por lo tanto, si se pretende que “X” represente “cualquier aminoácido”, ~~debería~~ deberá anotarse con la clave de caracterización VARIANT y un calificador NOTE con el valor “X puede ser cualquier aminoácido”.';

- i) se ha editado el ejemplo 30-1 del Anexo VI para corregir el último párrafo de la pregunta 3, como sigue:

'Además, hay una unión disulfuro entre los dos residuos Cys. Por lo tanto, se utiliza la clave de caracterización DISULFID para describir un enlace cruzado entre cadenas. El elemento localización de característica son los números de posición de los residuos Cys enlazados en formato “x..y”, es decir, “4..15”. El calificador obligatorio NOTE debería describir el enlace disulfuro entre cadenas. ~~El calificador NOTE no es obligatorio.~~';

- j) se ha incluido el nuevo SEQ ID NO: 51, que forma parte del ejemplo 29-2, en el Apéndice del Anexo VI, y se ha cambiado la numeración de las secuencias subsiguientes para que siga siendo consecutiva; y
- k) se ha editado el Apéndice del Anexo VI, SEQ ID NO: 51, para añadir un calificador NOTE a la clave de caracterización DISULFID del SEQ ID NO: 51:

```
<INSDFeature>
```

```
  <INSDFeature_key>DISULFID</INSDFeature_key>  
  <INSDFeature_location>4..15</INSDFeature_location>
```

```
  <INSDFeature_qualifiers>  
  <INSDQualifier>  
    <INSDQualifier_name>NOTE</INSDQualifier_name>  
    <INSDQualifier_value>intrachain disulfide  
bond</INSDQualifier_value>  
  </INSDQualifier>  
  </INSDFeature_qualifiers>
```

```
</INSDFeature>
```

7. La propuesta de revisión de la Norma ST.26 de la OMPI figura en los Anexos I a III del presente documento para que el CWS la examine:

- El Anexo I contiene la propuesta de modificación de la Norma ST.26 de la OMPI, con las revisiones resaltadas, excepto su Anexo IV, en el que no se propone ninguna modificación (así como los dos componentes siguientes);
- El Anexo II contiene el Anexo III de la Norma ST.26, que se publica en un fichero distinto; y
- El Anexo III contiene el Apéndice del Anexo VI de la Norma ST.26, que se publica en un fichero distinto.

8. *Se invita al CWS a:*

a) tomar nota del contenido del presente documento y de sus Anexos; y

b) estudiar y aprobar las revisiones propuestas en la Norma ST. 26 de la OMPI que figuran en los párrafos 5 y 6 anteriores y se reproducen en los Anexos del presente documento.

[Siguen los Anexos (la propuesta de revisión de la Norma ST.26 de la OMPI)]

NORMA ST.26

NORMA RECOMENDADA PARA LA PRESENTACIÓN DE LISTAS DE SECUENCIAS DE NUCLEÓTIDOS Y
AMINOÁCIDOS MEDIANTE EL LENGUAJE EXTENSIBLE DE MARCADO (XML)

Versión 1.45

Revisión aprobada Propuesta presentada por el Comité Equipo Técnico sobre Listas de Normas Técnicas de Secuencias para que el CWS la OMPI (CWS) examine y apruebe en su octava novena sesión, celebrada el 4 de diciembre de 2020

Nota editorial de la Oficina Internacional

En su quinta sesión, el Comité de Normas Técnicas de la OMPI (CWS) acordó que la transición de la Norma ST.25 a la Norma ST.26 de la OMPI tenga lugar en enero de 2022. Mientras tanto, se deberá seguir usando la Norma ST.25.

La Norma se publica a los fines de la información de las oficinas de propiedad industrial y otras partes interesadas.

ÍNDICE

INTRODUCCIÓN.....	3
DEFINICIONES.....	3
ÁMBITO.....	4
REFERENCIAS.....	5
REPRESENTACIÓN DE SECUENCIAS.....	5
<i>Secuencias de nucleótidos</i>	5
<i>Secuencias de aminoácidos</i>	8
<i>Presentación de casos especiales</i>	9
ESTRUCTURA DE LA LISTA DE SECUENCIAS EN FORMATO XML.....	10
<i>Elemento raíz</i>	11
<i>Parte de información general</i>	11
<i>Parte de datos de secuencia</i>	14
<i>Cuadro de características</i>	16
<i>Claves de caracterización</i>	16
<i>Claves de caracterización obligatorias</i>	16
<i>Localización de característica</i>	17
<i>Calificadores de caracterización</i>	19
<i>Calificadores de caracterización obligatorios</i>	19
<i>Elementos de los calificadores</i>	19
<i>Texto libre</i>	21
<i>Secuencias codificadoras</i>	22
<i>Variantes</i>	23

ANEXOS

[Anexo I](#) - Vocabulario controlado

[Anexo II](#) - Definiciones de tipo de documento (DTD) para las listas de secuencias

[Anexo III](#) - Ejemplo de lista de secuencias (archivo XML)

[Anexo IV](#) - Subconjunto de caracteres del cuadro de códigos de caracteres del alfabeto latino básico de la norma Unicode para su utilización en una instancia XML de una lista de secuencias

[Anexo V](#) - Requisitos adicionales sobre el intercambio de datos (únicamente para las **oficinas de patentes OPI**)

[Anexo VI](#) - Documento de orientación con ejemplos ilustrativos

[Apéndice](#) - Secuencias del documento de orientación en XML

[Anexo VII](#) - Recomendación para la transformación de una lista de secuencias de ST.25 a ST.26: eventual adición o supresión de materia

NORMA ST.26

NORMA RECOMENDADA PARA LA PRESENTACIÓN DE LISTAS DE SECUENCIAS DE NUCLEÓTIDOS Y AMINOÁCIDOS MEDIANTE EL LENGUAJE EXTENSIBLE DE MARCADO (XML)

Versión 1.45

Revisión aprobada Propuesta presentada por el Comité Equipo Técnico sobre Listas de Normas Técnicas de Secuencias para que el CWS la OMPI (CWS) examine y apruebe en su octava novena sesión, celebrada el 4 de diciembre de 2020

INTRODUCCIÓN

1. Esta Norma define la manera de divulgar en una solicitud de patente las secuencias de nucleótidos y aminoácidos que deben figurar en una lista de secuencias, la representación de esas divulgaciones, y la definición de tipo de documento (DTD) cuando las listas de secuencias se presentan en lenguaje extensible de marcado (XML). Se recomienda a las oficinas de propiedad industrial que acepten toda lista de secuencias compatible con esta Norma, que se presente en una solicitud de patente o en relación con una solicitud de patente.

2. La Norma tiene por objetivo:

- a) permitir que el solicitante establezca una única lista de secuencias en una solicitud de patente que sea aceptable a los efectos tanto de los procedimientos internacionales como nacionales o regionales;
- b) mejorar la precisión y calidad de la presentación de las secuencias a fin de facilitar su difusión para beneficio de los solicitantes, el público y los examinadores;
- c) facilitar la búsqueda de datos en las secuencias; y
- d) permitir el intercambio electrónico de datos sobre las secuencias y la introducción de esos datos en bases de datos informatizadas.

DEFINICIONES

3. A los efectos de la presente Norma:

a) 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 aminoácidos deberán interpretarse como L-aminoácidos no modificados a menos que se indique con detalle en el cuadro de características que se trata de aminoácidos modificados tal como se prevé en el párrafo 30. A los fines de la presente norma, un residuo de ácido nucleico péptido (PNA) no se considera aminoácido, pero se considera nucleótido según se describe en el párrafo 3.g)i)2).

b) por "vocabulario controlado" se entenderá la terminología descrita en la presente Norma que deberá utilizarse a la hora de indicar las características de una secuencia, a saber, las anotaciones de regiones o sitios de interés tal como figuran en el Anexo I.

c) por "enumeración de sus residuos" se entenderá la divulgación de una secuencia en una solicitud de patente en la que se enumera, por orden, cada residuo de la secuencia, en la que:

- i) el residuo se representa mediante un nombre, abreviatura, símbolo o estructura (por ejemplo, HHHHHHQ o HisHisHisHisHisHisGln); o
- ii) varios residuos se representan mediante una fórmula abreviada (por ejemplo, His₆Gln).

d) por "secuencia ignorada deliberadamente", o secuencia vacía, se entenderá un espacio reservado para mantener la numeración de las secuencias que figuran en la lista de secuencias a fin de garantizar su coherencia con la numeración de la divulgación, por ejemplo, para no tener que volver a numerar las secuencias contenidas en la divulgación y en la lista de secuencias cuando se suprime una secuencia de la divulgación.

e) por "aminoácido modificado" se entenderá todo aminoácido descrito en el párrafo 3.a) distinto de 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-threonine, L-triptófano, L-tirosina, o L-valina.

f) por "nucleótido modificado" se entenderá todo nucleótido descrito en el párrafo 3.g) distinto de 3'-monofosfato de desoxiadenosina, 3'-monofosfato de desoxiguanosina, 3'-monofosfato de desoxicitidina, 3'-monofosfato de desoxitimidina, 3'-monofosfato de adenosina, 3'-monofosfato de guanosina, 3'-monofosfato de citidina o 3'-monofosfato de uridina.

g) por "nucleótido" se entenderá todo nucleótido o análogo de nucleótido que pueda representarse utilizando cualquiera de los símbolos descritos en el Anexo I (véase el Cuadro 1 de la Sección 1) en el que el nucleótido o análogo de nucleótido contiene:

- i) un componente del esqueleto seleccionado de:

- 1) 5' monofosfato de 2' desoxirribosa (el componente del esqueleto de un desoxirribonucleótido) o 5' monofosfato de ribosa (el componente de la estructura de un ribonucleótido); o
- 2) el análogo de un 5' monofosfato de 2' desoxirribosa o 5' monofosfato de ribosa, que al formar el esqueleto de un análogo de ácido nucleico, 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 nucleico es capaz de aparear las bases con un ácido nucleico complementario; entre los ejemplos de **análogos de nucleótidos** **componentes del esqueleto** figuran los aminoácidos en los ácidos peptidonucleicos, las moléculas de ácido glicólico en los ácidos nucleicos glicólicos, las moléculas de azúcar de treofuranosil en los ácidos nucleicos treósicos, los anillos de morfolina y los grupos de fosfordiamidato en los morfolinós, y las moléculas de ciclohexenilo en los ácidos nucleicos que contienen ciclohexeno.

y

- ii) el componente del esqueleto o bien:
 - 1) está adherido a una nucleobase, incluida una nucleobase modificada o sintética de purina o pirimidina; o bien
 - 2) al carecer de una nucleobase de purina o pirimidina cuando el nucleótido es parte de una secuencia nucleotídica, se menciona como "sitio AP" o "sitio abásico".
- h) por "residuo" se entenderá todo nucleótido o aminoácido individual o sus análogos respectivos en una secuencia.
- i) por "identificador de secuencia" se entenderá un número entero único que se asigna a cada secuencia de la lista de secuencias.
- j) por "lista de secuencias" se entenderá una parte de la descripción de la solicitud de patente tal como fue presentada, o un documento presentado posteriormente a la solicitud, que incluye la(s) secuencia(s) divulgada(s) de nucleótidos y/o aminoácidos junto con toda otra descripción detallada, según estipula la presente Norma.
- k) por "específicamente definido" se entenderá todo nucleótido distinto a los representados por el símbolo "n" y todo aminoácido distinto a los representados por el símbolo "X" que se enumeran en el Anexo I (véanse la Sección 1, Cuadro 1, y la Sección 3, Cuadro 3, respectivamente).
- l) por nucleótido o aminoácido "desconocido" se entenderá la presencia de un único nucleótido o aminoácido cuya identidad es desconocida o no se divulga.
- m) por "variante de secuencia" se entenderá 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.
- n) por "texto libre" se entenderá 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). Véase el párrafo 85.
- o) por "texto libre dependiente del idioma" se entenderá el valor de texto libre (*free text*) de ciertos calificadores que es dependiente del idioma y que podrá requerir traducción para los procedimientos internacionales, nacionales o regionales. Véase el párrafo 87.

4. A los fines de la presente Norma, el término o términos:

- a) "podrá" hace referencia a un enfoque facultativo o autorizado, pero no a un requisito.
- b) "deberá" hace referencia a un requisito de la Norma; la no observancia del requisito, dará lugar al incumplimiento.
- c) "no deberá" hace referencia a una prohibición de la Norma.
- d) "debería" hace referencia a un enfoque al que se insta vivamente, pero no a un requisito.
- e) "no debería" hace referencia a un enfoque que se desaconseja vivamente, pero no a una prohibición.

ÁMBITO

5. La presente Norma establece los requisitos de presentación de las listas de secuencias de nucleótidos y aminoácidos de las secuencias divulgadas en las solicitudes de patente.

6. Una lista de secuencias compatible con la presente Norma (en adelante lista de secuencias) contiene una parte de información general y una parte de datos de secuencia. La lista de secuencias deberá presentarse en un único archivo en formato XML utilizando la definición de tipo de documento (DTD) descrita en el Anexo II. El objetivo de la información bibliográfica contenida en la parte de información general sirve únicamente para vincular la lista de secuencias a la solicitud de patente en el marco de la cual se presenta dicha lista de secuencias. La parte de datos de secuencia está compuesta por uno o más elementos de datos de secuencia, cada uno de los cuales contiene información acerca de una secuencia. Los elementos de datos de secuencia incluyen varias claves de caracterización y los calificadores subsiguientes basados en las especificaciones de la *International Nucleotide Sequence Database Collaboration* (INSDC) y UniProt.

7. A los efectos de la presente Norma, una secuencia deberá figurar en una lista de secuencias cuando se divulga en cualquier parte de una solicitud mediante la enumeración de sus residuos y puede representarse como:

- a) una secuencia no ramificada o una región lineal de una secuencia ramificada que contiene diez o más nucleótidos específicamente definidos, en la que los nucleótidos adyacentes están unidos mediante:
 - i) un enlace fosfodiéster en el sentido 3' a 5' (o 5' a 3'); o
 - ii) cualquier enlace químico que dé lugar a una disposición de nucleobases adyacentes que imite la disposición de las nucleobases en ácidos nucleicos que ocurren naturalmente; o
- b) una secuencia no ramificada o una región lineal de una secuencia ramificada que contiene cuatro o más aminoácidos específicamente definidos, los cuales forman un único esqueleto peptídico, es decir, que los aminoácidos adyacentes están unidos por enlaces peptídicos.

8. Una lista de secuencias no deberá incluir, en calidad de secuencia a la que se ha asignado su propio número de identificación, ninguna secuencia que tenga menos de diez nucleótidos específicamente definidos, o menos de cuatro aminoácidos específicamente definidos.

REFERENCIAS

9. Las siguientes Normas y recursos son referencias pertinentes para la presente Norma:

<i>International Nucleotide Sequence Database Collaboration</i> (INSDC)	http://www.insdc.org/ ;
Norma Internacional ISO 639-1:2002	<i>Codes for the representation of names of languages Part 1: Alpha-2 code</i> ;
UniProt Consortium	http://www.uniprot.org/ ;
W3C XML 1.0	http://www.w3.org/ ;
Norma técnica de la OMPI ST.2	Forma normalizada de designar las fechas según el calendario gregoriano;
Norma técnica de la OMPI ST.3	Códigos normalizados de dos letras, recomendados para la representación de Estados, otras entidades y organizaciones intergubernamentales;
Norma técnica de la OMPI ST.16	Código normalizado para la identificación de los diferentes tipos de documentos de patente;
Norma técnica de la OMPI ST.25	Norma para la presentación de listas de secuencias de nucleótidos y aminoácidos en solicitudes de patente.

REPRESENTACIÓN DE SECUENCIAS

10. Se deberá asignar a cada secuencia comprendida en el párrafo 7 un identificador de secuencia distinto, incluidas las secuencias que sean idénticas a una región de una secuencia más larga. Los identificadores de secuencias deberían comenzar con el número 1, e irán aumentando de forma consecutiva por números enteros. Si a un identificador de secuencia no correspondiese una secuencia, a saber, una secuencia ignorada deliberadamente, se deberá utilizar el código "000" en lugar de la secuencia (véase el párrafo 58). El número total de secuencias deberá indicarse en la lista de secuencias y deberá ser igual al número total de identificadores de secuencias, con independencia de si van seguidos de una secuencia o del código "000".

Secuencias de nucleótidos

11. Toda secuencia de nucleótidos solo deberá representarse mediante una cadena única, en el sentido 5' a 3' y de izquierda a derecha, o en el sentido de izquierda a derecha que imite el sentido 5' a 3'. Los valores 5' y 3' o cualquier otro valor similar no deberán estar incluidos en la secuencia. Toda secuencia de nucleótidos de doble cadena divulgada mediante la enumeración de los residuos de ambas cadenas deberá representarse de la siguiente manera:

- a) una única secuencia o dos secuencias distintas, a las que se les asignará su propio identificador de secuencia, en el que las dos cadenas distintas deberán ser plenamente complementarias entre sí, o
- b) dos secuencias distintas, a las que se les asignará su propio identificador de secuencia, en el que las dos cadenas no son plenamente complementarias entre sí.

12. A los fines de la presente Norma, el primer nucleótido presentado en la secuencia será el residuo de la posición número 1. Cuando las secuencias de nucleótidos tengan una configuración circular, el solicitante deberá elegir el nucleótido del residuo de la posición número 1. La numeración será continua a lo largo de toda la secuencia en el sentido 5' a 3', o en el sentido que imite el sentido 5' a 3'. El último número de posición de los residuos deberá ser igual al número de nucleótidos de la secuencia.

13. Todos los nucleótidos de una secuencia deberán representarse mediante los símbolos descritos en el Anexo I (véase el Cuadro 1 de la Sección 1). Se deberá utilizar únicamente letras minúsculas. Todo símbolo utilizado para representar un nucleótido equivale a un único residuo.

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.

15. Si fuera necesario utilizar un símbolo de ambigüedad (que represente dos o más nucleótidos alternativos), debería utilizarse el símbolo más restrictivo, según figura en el Anexo I (sección 1, Cuadro 1). Por ejemplo, si un nucleótido en una determinada posición pudiera ser "a" o "g", se debería utilizar "r", en vez de "n". El símbolo "n" se interpretará como "a", "c", "g", o "t/u", excepto cuando se utilice acompañado de una descripción detallada en el cuadro de características. El símbolo "n" no deberá utilizarse para representar un elemento distinto a un nucleótido. El símbolo "n" podrá representar un único nucleótido modificado o "unknown", junto con una descripción detallada en el cuadro de características, tal como se prevé en los párrafos 16, 17, 21 o 93 a 96. Para la representación de variantes de secuencias, por ejemplo, alternativas, supresiones, inserciones o sustituciones, véanse los párrafos 93 a 100.

16. Los nucleótidos modificados deberían representarse en la secuencia como los correspondientes nucleótidos no modificados, a saber, "a", "c", "g" o "t" cuando sea posible. Todo nucleótido modificado en una secuencia que no pueda representarse de otra manera por ningún otro símbolo descrito en el Anexo I (véase el Cuadro 1 de la Sección 1), por ejemplo, "other" nucleótido, tal como un nucleótido que no existe en estado natural, deberá representarse mediante el símbolo "n". El símbolo "n" equivale a un único residuo.

17. Los nucleótidos modificados deberán describirse con detalle en el cuadro de características (véanse los párrafos 60 y siguientes) utilizando la clave de caracterización "modified_base" y el calificador obligatorio "mod_base". El valor calificador deberá corresponder a una abreviatura única que figure en el Anexo I (véase el Cuadro 2 de la Sección 2; si la abreviatura es "OTHER", el nombre completo no abreviado del nucleótido modificado deberá indicarse como valor en un calificador "note". Para una lista de nucleótidos modificados alternativos, podrá utilizarse el valor calificador "OTHER" junto con otro calificador "note" (véanse los párrafos 97 y 98). Las abreviaturas (o nombres completos) que figuran en el Anexo I (véase el Cuadro 2 de la Sección 2) a las que se ha referenciado *supra* no deberán utilizarse en la propia secuencia.

18. Una secuencia de nucleótido que contenga uno o más regiones de nucleótidos modificados consecutivos que compartan el mismo componente del esqueleto (véase el párrafo 3.g)i)2)), deberá describirse detalladamente en el cuadro de características según se dispone en el párrafo 17. Los nucleótidos modificados de cada región podrán describirse conjuntamente en un único elemento INSDFeature según lo previsto en el párrafo 22. Deberá proporcionarse el nombre químico no abreviado más restrictivo que comprenda todos los nucleótidos modificados en la serie o una lista de los nombres químicos de todos los nucleótidos, como el valor del calificador "note". Por ejemplo, una secuencia de ácido nucléico glicólico que contenga las nucleobases "a", "c", "g" o "t" podrá describirse en el calificador "note" como "2,3-dihydroxypropyl nucleosides." Por otra parte, la misma secuencia podrá describirse en el calificador "note" como "2,3-dihydroxypropyladenine, 2,3-dihydroxypropylthymine, 2,3-dihydroxypropylguanine, or 2,3-dihydroxypropylcytosine." Cuando un nucleótido modificado en la región incluya una modificación adicional, el nucleótido modificado deberá describirse detalladamente en el cuadro de características según lo previsto en el párrafo 17.

19. 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.

20. Los siguientes ejemplos ilustran la representación de los nucleótidos modificados en la forma prevista en los párrafos 16 a 18:

Ejemplo 1: Nucleótido modificado utilizando una abreviatura que figura en el Anexo I (véase el Cuadro 2 de la Sección 2).

```
<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>
```

Ejemplo 2: Nucleótido modificado utilizando "OTHER" como figura en el Anexo I (véase el Cuadro 2 de la Sección 2).

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>4</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>xanthine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Ejemplo 3: Una secuencia de nucleótido compuesta de nucleótidos modificados comprendidos en el párrafo 3.g)i)2) con dos nucleótidos individuales que incluyen otra modificación

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>1..954</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</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>439</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>684</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>xanthine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

<INSDQualifier>
  <INSDQualifier_name>note</INSDQualifier_name>
  <INSDQualifier_value>xanthine</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

21. Todo nucleótido "unknown" deberá representarse mediante el símbolo "n" en la secuencia. Los nucleótidos "unknown" deberían describirse con detalle en el cuadro de características (véanse los párrafos 60 siguientes) utilizando la clave de caracterización "unsure". El símbolo "n" equivale a un único residuo.

22. Toda región que contiene un número conocido de residuos contiguos "a", "c", "g", "t", o "n" para los cuales se aplica la misma descripción podrá describirse en conjunto utilizando un único elemento INSDFeature con la sintaxis "x.y" como descriptor de localización en el elemento INSDFeature_location (véanse los párrafos 64 a 71). Para la representación de las variantes de secuencia, es decir, alternativas, supresiones, inserciones o sustituciones, véanse los párrafos 93 a 100.

23. El siguiente ejemplo ilustra la representación de una región de nucleótidos modificados para los cuales se aplica la misma descripción en la forma prevista en el párrafo 22:

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>358..485</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>isoguanine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Secuencias de aminoácidos

24. Los aminoácidos de una secuencia de aminoácidos deberán representarse en el sentido del grupo amino al grupo carboxilo, y de izquierda a derecha. Los grupos amino y carboxilo no deberán representarse en la secuencia.

25. A los fines de la presente Norma, el primer aminoácido de la secuencia será el residuo de la posición número 1, incluidos los aminoácidos que preceden a la proteína madura, por ejemplo, las presecuencias, las prosecuencias, y las preprosecuencias, así como las secuencias señal. Cuando una secuencia de aminoácidos tenga una configuración circular y el anillo esté compuesto únicamente por residuos de aminoácidos unidos mediante enlaces peptídicos, es decir, que la secuencia no tenga posiciones amino y carboxilo terminales, el solicitante deberá elegir el aminoácido del residuo de la posición número 1. La numeración será continua a lo largo de toda la secuencia en el sentido de amino a carboxilo.

26. Todos los aminoácidos de una secuencia deberán representarse mediante los símbolos descritos en el Anexo I (véase el Cuadro 3 de la Sección 3). Se deberán utilizar únicamente letras mayúsculas. Todo símbolo utilizado para representar un aminoácido equivale a un único residuo.

27. Si fuese necesario utilizar un símbolo de ambigüedad (que represente dos o más aminoácidos posibles), debería utilizarse el símbolo más restrictivo, según figura en el Cuadro 3 de la Sección 3 del Anexo I. Por ejemplo, si un aminoácido en una posición dada podría ser un ácido aspártico o asparagina, debería utilizarse el símbolo "B" en vez del símbolo "X". El símbolo "X" se interpretará como uno de los símbolos "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", o "V", excepto cuando se utilice en una descripción detallada en el cuadro de características. El símbolo "X" no deberá utilizarse para representar un elemento distinto a un aminoácido. Un único aminoácido modificado o "desconocido" podrá representarse mediante el símbolo "X", junto con una descripción detallada en el cuadro de características, tal como se prevé en los párrafos 29, 30, 32 o 93 a 98. Para la representación de variantes de secuencia, a saber, alternativas, supresiones, inserciones, o sustituciones, véanse los párrafos 93 a 100.

28. Las secuencias de aminoácidos divulgadas separadas por símbolos internos de terminación, representados por ejemplo, por "Ter" o asterisco "*" o punto "." o un espacio en blanco, deberán incluirse como secuencias distintas en el caso de cada secuencia de aminoácidos que contenga al menos cuatro aminoácidos específicamente definidos y esté contemplada en el párrafo 7. A cada secuencia distinta de este tipo deberá asignarse su propio identificador de secuencia. No deberán incluirse símbolos de terminación ni espacios en las secuencias presentadas en una lista de secuencias (véase el párrafo 57).

29. Los aminoácidos modificados, incluidos los D-aminoácidos, deberían representarse en la secuencia como los correspondientes aminoácidos no modificados, cuando sea posible. Todo aminoácido modificado que aparece en una secuencia y que no pueda ser representado por ningún otro símbolo que figure en el Anexo I (véase el Cuadro 3 de la Sección 3), por ejemplo, "other" aminoácido, deberá representarse por el símbolo "X". El símbolo "X" equivale a un único residuo.

30. Los aminoácidos modificados deberán describirse con detalle en el cuadro de características (véanse los párrafos 60 y siguientes). Cuando proceda, deberían utilizarse las claves de caracterización "CARBOHYD" o "LIPID" junto con el calificador "NOTE note". La clave de caracterización "MOD_RES" debería utilizarse para otras modificaciones postraduccionales de los aminoácidos junto con el calificador "NOTE note"; de lo contrario, debería utilizarse la clave de caracterización "SITE" junto con el calificador "NOTE note". El valor del calificador "NOTE note" deberá ser ya sea una abreviatura descrita en el Anexo I (véase el Cuadro 4 de la Sección 4), o el nombre completo no abreviado del aminoácido modificado. Las abreviaturas descritas en el Cuadro 4 que se mencionan *supra* o los nombres completos no abreviados no deberán utilizarse en la propia secuencia.

31. Los siguientes ejemplos ilustran la representación de aminoácidos modificados en la forma prevista en el párrafo

Ejemplo 1: Aminoácido modificado (modificación postraduccional).

```
<INSDFeature_qual>
  <INSDQualifier>
    <INSDQualifier_name>NOTEnote</INSDQualifier_name>
    <INSDQualifier_value>3Hyp</INSDQualifier_value>
  </INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

Ejemplo 2: Aminoácido modificado (modificación no postraduccional).

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>NOTEnote</INSDQualifier_name>
      <INSDQualifier_value>Orn</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Ejemplo 3: D-aminoácido.

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>9</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>NOTEnote</INSDQualifier_name>
      <INSDQualifier_value>D-Arginine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

32. Todo aminoácido “unknown” deberá representarse mediante el símbolo “X” en la secuencia. Todo aminoácido “unknown” designado mediante “X” deberá describirse con detalle en el cuadro de características (véanse los párrafos 60 y siguientes.) mediante la clave de caracterización “UNSURE” y facultativamente el calificador “NOTEnote”. El símbolo “X” es el equivalente de un único residuo.

33. Los siguientes ejemplos ilustran la representación de aminoácidos “unknown” en la forma prevista en el párrafo

```
<INSDFeature>
  <INSDFeature_key>UNSURE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>NOTEnote</INSDQualifier_name>
      <INSDQualifier_value>A or V</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

34. Toda región que contiene un número desconocido de residuos “X” contiguos para los cuales se aplica la misma descripción podrá describirse en conjunto utilizando la sintaxis “x..y” como descriptor de localización en el elemento INSDFeature_location (véanse los párrafos 64 a 70). Para la representación de variantes de secuencia, a saber, supresiones, alternativas, inserciones, o sustituciones, véanse los párrafos 93 a 100.

Presentación de casos especiales

35. Toda secuencia divulgada mediante la enumeración de sus residuos, que se interpreta como una única secuencia contigua a partir de uno o más segmentos no contiguos de una secuencia más grande o de segmentos de diferentes secuencias, deberá figurar en la lista de secuencias y se le deberá asignar su propio identificador de secuencia.

36. Toda secuencia que contenga regiones de residuos específicamente definidos separados por una o más regiones de residuos “n” o “X” contiguos (véanse los párrafos 15 y 27, respectivamente), en la que el número exacto de residuos “n” o “X” de cada región se divulga, deberá figurar en la lista de secuencias de una única secuencia y se le deberá asignar su propio identificador de secuencia.

37. Toda secuencia que contenga regiones de residuos específicamente definidos y separados por uno o más huecos compuestos por un número desconocido o no divulgado de residuos no deberá figurar en la lista de secuencias como una única secuencia. Cada región de residuos específicamente definidos que está comprendida en el párrafo 7 deberá figurar en la lista de secuencias como secuencia distinta y se le deberá asignar su propio identificador de secuencia.

ESTRUCTURA DE LA LISTA DE SECUENCIAS EN FORMATO XML

38. Según lo previsto en el párrafo 6 *supra*, la instancia XML de un archivo que contiene una lista de secuencias compatible con la presente Norma se compone de:

- a) una parte de información general, que contiene la información relativa a la solicitud de patente a la que está asociada la lista de secuencias; y
- b) una parte de datos de secuencia, que contiene uno o más elementos de datos de secuencia, cada uno de los cuales, a su vez contiene información acerca de una secuencia.

En el Anexo III se presenta un ejemplo de una lista de secuencias.

39. La lista de secuencias deberá presentarse en formato XML 1.0 utilizando la DTD presentada en el Anexo II "Definición de tipo de documento (DTD) para listas de secuencias".

- a) La primera línea de la instancia XML deberá contener la declaración XML siguiente:

```
<?xml version="1.0" encoding="UTF-8"?>.
```

- b) La segunda línea de la instancia XML deberá contener una declaración de tipo de documento (DOCTYPE):

```
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1. 3//EN"
"ST26SequenceListing_V1_3.dtd">.
```

40. La lista de secuencias electrónica completa deberá figurar en un solo archivo. El archivo deberá cifrarse utilizando el lenguaje Unicode UTF-8, con las siguientes restricciones:

- a) la información contenida en los elementos ApplicantName, InventorName e InventionTitle de la parte de información general, y el elemento NonEnglishQualifier_value de la parte de datos de secuencia, podrá estar compuesta por cualquier carácter Unicode válido indicado en la especificación XML 1.0, excepto los puntos de código de control Unicode 0000-001F y 007F-009F. Los caracteres reservados ", &, ', <, y > (puntos de código Unicode 0022, 0026, 0027, 003C y 003E respectivamente), deberán sustituirse como se describe en el párrafo 41; y
- b) la información contenida en todos los demás elementos y atributos de la parte de información general y en todos los demás elementos y atributos de la parte de datos de secuencia
 - o deberá estar compuesta por caracteres imprimibles (incluido el carácter de espacio) del cuadro de códigos de caracteres del alfabeto latino básico de la norma Unicode (es decir, limitados a los puntos de código 0020 a 007E – véase el Anexo IV) Los caracteres reservados ", &, ', <, y > (puntos de código Unicode 0022, 0026, 0027, 003C y 003E respectivamente) deberán sustituirse como se describe en el párrafo 41.

41. En una instancia XML de una lista de secuencias, las referencias de caracteres numéricos¹ no deberán utilizarse y los siguientes caracteres reservados deberán sustituirse por las entidades predefinidas correspondientes cuando se utilicen en el valor de un atributo o el contenido de un elemento:

Carácter reservado	Entidades predefinidas
<	<
>	>
&	&
"	"
'	'

Véase el párrafo 71 para consultar un ejemplo. Las únicas referencias de entidades de caracteres permitidas son las entidades predefinidas que se indican en este párrafo.

42. Todos los elementos obligatorios deberán indicarse (excepto en el caso definido en el párrafo 58 para las secuencias ignoradas deliberadamente). Los elementos facultativos para los cuales no existe ningún contenido disponible no deberían aparecer en la instancia XML (salvo lo dispuesto en el párrafo 97 para la representación de una supresión en una secuencia en el valor correspondiente al calificador "replace").

¹ Una referencia de carácter numérico se refiere a un carácter por su conjunto de caracteres universal/punto de código Unicode y utiliza el formato "&#nnnn;" o "&#xhhhh;", donde "nnnn" es el punto de código en forma decimal y "hhhh" es el punto de código en forma hexadecimal.

Elemento raíz

43. El elemento raíz de una instancia XML según lo dispuesto en esta Norma es el elemento ST26SequenceListing, que tiene los siguientes atributos:

Atributo	Descripción	Obligatorio/Facultativo
dtdVersion	Versión de la DTD utilizada para crear este archivo en el formato "V#_#", p. ej. "V1_3".	Obligatorio
fileName	Nombre del archivo que contiene la lista de secuencias.	Facultativo
softwareName	Nombre del programa informático que generó este archivo.	Facultativo
softwareVersion	Versión del programa informático que generó este archivo.	Facultativo
productionDate	Fecha de producción del archivo que contiene la lista de secuencias (programa "AACC-MM-DD").	Facultativo
originalFreeTextLanguageCode	Código del idioma (véase la referencia en el párrafo 9 a la norma ISO 639-1:2002) para el idioma original único en que se prepararon los calificadores de texto libre dependiente del idioma.	Facultativo
nonEnglishFreeTextLanguageCode	Código del idioma (véase la referencia en el párrafo 9 a la norma ISO 639-1:2002) para los elementos NonEnglishQualifier_value.	Obligatorio cuando un elemento NonEnglishQualifier_value está presente en la lista de secuencias

44. El siguiente ejemplo ilustra el elemento raíz ST26SequenceListing, y sus atributos, de una instancia XML en la forma prevista en el párrafo 43 *supra*:

```
<ST26SequenceListing dtdVersion="V1_3" fileName="US11_405455_SEQ1.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="de" nonEnglishFreeTextLanguageCode="fr">
  {...}*
</ST26SequenceListing>
```

*{...} represents the general information part and the sequence data part that have not been included in this example.

Parte de información general

45. Los elementos de la parte de información general se relacionan a la información relativa a la solicitud de patente, de la siguiente manera:

Elemento	Descripción	Obligatorio/Facultativo
ApplicationIdentification	La identificación de la solicitud para la cual se presenta la lista de secuencias	Obligatorio cuando una lista de secuencias se suministra en cualquier momento posterior a la asignación del número de solicitud
La ApplicationIdentification se compone de:		
IPOfficeCode	El código ST.3 de la oficina de presentación	Obligatorio
ApplicationNumberText	La identificación de la solicitud suministrada por la oficina de presentación (por ejemplo, PCT/IB2013/099999)	Obligatorio

Elemento	Descripción	Obligatorio/ Facultativo
FilingDate	La fecha de presentación de la solicitud de patente para la cual se presenta la lista de secuencias (formato ST.2 a "AACC-MM-DD", que utiliza 4 dígitos para representar el año civil, 2 dígitos el mes civil y 2 dígitos el número ordinal de un día dentro del mes civil, por ejemplo, 2015-01-31)	Obligatorio cuando una lista de secuencias se suministra en cualquier momento posterior a la asignación de una fecha de presentación
ApplicantFileReference	Un identificador único asignado por el solicitante para identificar una solicitud específica, escrito en los caracteres descritos en el párrafo 40.b)	Obligatorio cuando una lista de secuencias se suministra en cualquier momento anterior a la asignación del número de solicitud; de lo contrario, Facultativo
EarliestPriorityApplication Identification	La identificación de la solicitud de prioridad más antigua (también contiene IPOfficeCode, ApplicationNumberText y FilingDate, véase ApplicationIdentification <i>supra</i>)	Obligatorio cuando se reivindica la prioridad
ApplicantName	El nombre del primer solicitante mencionado escrito en los caracteres descritos en el párrafo 40.a). Este elemento contiene el atributo obligatorio languageCode descrito en el párrafo 47.	Obligatorio
ApplicantNameLatin	Si se escribe ApplicantName en caracteres distintos a los descritos en el párrafo 40.b), también deberá escribirse la traducción o transliteración del nombre del primer solicitante mencionado en los caracteres descritos en el párrafo 40.b)	Obligatorio cuando ApplicantName contiene caracteres no latinos
InventorName	Nombre del primer inventor mencionado escrito en los caracteres descritos en el párrafo 40.a). Este elemento contiene el atributo obligatorio languageCode descrito en el párrafo 47.	Facultativo
InventorNameLatin	Si InventorName se escribe en caracteres distintos a los descritos en el párrafo 40.b), podrá también escribirse la traducción o transliteración del inventor mencionado en primer lugar en los caracteres descritos en el párrafo 40.b)	Facultativo
InventionTitle	Título de la invención escrita en los caracteres descritos en el párrafo 40.a) en el idioma de presentación. La traducción del título de la invención en otros idiomas podrá escribirse en los caracteres descritos en el párrafo 40.a) utilizando varios elementos InventionTitle. Este elemento contiene el atributo obligatorio languageCode descrito en el párrafo 48. El título de la invención deberá ia contener de preferencia dos a siete palabras.	Obligatorio en el idioma de presentación. Facultativo en los otros idiomas.

Elemento	Descripción	Obligatorio/ Facultativo
SequenceTotalQuantity	El número total de todas las secuencias que figuran en la lista de secuencias, incluidas las secuencias ignoradas deliberadamente (o secuencias vacías) (véase el párrafo 9).	Obligatorio

46. Los siguientes ejemplos ilustran la presentación de la parte de información general de la lista de secuencias en la forma prevista en el párrafo 45 *supra*:

Ejemplo 1: Lista de secuencias presentada antes de la asignación de la identificación de la solicitud y la fecha de presentación.

```
<?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="Invention_SEQ1.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="jp" ja">
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2013/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode="en">Keiko Nakamura</InventorName>
  <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
  <SequenceTotalQuantity>9</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1"> {...} * </SequenceData>
  <SequenceData sequenceIDNumber="2"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="3"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="4"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="5"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="6"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="7"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="8"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>
```

*{...} representa la información pertinente para cada secuencia que no se ha incluido en este ejemplo.

Ejemplo 2: Lista de secuencias presentada después de la asignación de la identificación de la solicitud y fecha de presentación.

```
<?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="Invention_SEQ1.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="jp" ja">
  <ApplicationIdentification>
    <IPOfficeCode>US</IPOfficeCode>
    <ApplicationNumberText>14/999,999</ApplicationNumberText>
    <FilingDate>-2015-01-05</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode="en">Keiko Nakamura</InventorName>
  <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
  <SequenceTotalQuantity>9</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1"> {...} * </SequenceData>
```

```
<SequenceData sequenceIDNumber="2"> {...} </SequenceData>
<SequenceData sequenceIDNumber="3"> {...} </SequenceData>
<SequenceData sequenceIDNumber="4"> {...} </SequenceData>
<SequenceData sequenceIDNumber="5"> {...} </SequenceData>
<SequenceData sequenceIDNumber="6"> {...} </SequenceData>
<SequenceData sequenceIDNumber="7"> {...} </SequenceData>
<SequenceData sequenceIDNumber="8"> {...} </SequenceData>
<SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>*
```

{...} representa la información pertinente para cada secuencia que no se ha incluido en este ejemplo.

47. El nombre del solicitante y, facultativamente, el nombre del inventor deberán indicarse en el elemento `ApplicantName` o `InventorName`, respectivamente, ya que en general se mencionan en el idioma de presentación de la solicitud. El código de idioma adecuado (véase la referencia en el párrafo 9 a la ISO 639-1:2002) deberá indicarse en el atributo `languageCode` de cada elemento. Cuando el nombre del solicitante indicado contenga caracteres distintos a los del alfabeto latino, como se describe en el párrafo 40.b), también deberá indicarse una transliteración o traducción del nombre del solicitante en caracteres del alfabeto latino en el elemento `ApplicantNameLatin`. Cuando el nombre del inventor indicado contenga caracteres distintos a los del alfabeto latino, también podrá indicarse una transliteración o traducción del nombre del inventor en caracteres del alfabeto latino en el elemento `InventorNameLatin`.

48. El título de la invención deberá indicarse en el elemento `InventionTitle` en el idioma de presentación y también podrá indicarse en otros idiomas utilizando varios elementos `InventionTitle` (véase el cuadro del párrafo 45). El código de idioma adecuado (véase la referencia en el párrafo 9 a la ISO 639-1:2002) deberá indicarse en el atributo `languageCode` del elemento.

49. El siguiente ejemplo ilustra la presentación de los nombres y el título de la invención en la forma prevista en los párrafos 47 y 48 *supra*:

Ejemplo: El nombre del solicitante y el nombre del inventor se presentan en caracteres japoneses y latinos, y el título de la invención se presenta en japonés, inglés y francés.

```
<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>
<InventionTitle languageCode="ja">efg タンパク質のためのマウス abcd-1 遺伝子
</InventionTitle>
<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein
</InventionTitle>
<InventionTitle languageCode="fr">Gène abcd-1 de Mus musculus pour protéine efg
</InventionTitle>
```

Parte de datos de secuencia

50. La parte de datos de secuencia deberá constar de uno o más elementos `SequenceData`, y cada elemento contendrá información sobre una sola secuencia.

51. Cada elemento `SequenceData` deberá tener un atributo obligatorio `sequenceIDNumber`, el cual contiene el identificador de secuencia (véase el párrafo 10) de cada secuencia. Por ejemplo:

```
<SequenceData sequenceIDNumber="1">
```

52. El elemento `SequenceData` deberá contener un elemento dependiente `INSDSeq`, que consiste en otros elementos dependientes de la siguiente manera:

Elemento	Descripción	Obligatorio/No incluido	
		Secuencias	Secuencias ignoradas deliberadamente
<code>INSDSeq_length</code>	Longitud de la secuencia	Obligatorio	Obligatorio, sin ningún valor
<code>INSDSeq_moltype</code>	Tipo de molécula	Obligatorio	Obligatorio, sin ningún valor
<code>INSDSeq_division</code>	Indicación de que una secuencia está asociadas a una solicitud de patente	Obligatorio con el valor "PAT"	Obligatorio, sin ningún valor

Elemento	Descripción	Obligatorio/No incluido	
		Secuencias	Secuencias ignoradas deliberadamente
INSDSeq_feature-table	Lista de anotaciones de la secuencia	Obligatorio	NO debe incluirse
INSDSeq_sequence	Secuencia	Obligatorio	Obligatorio con el valor "000"

53. El elemento `INSDSeq_length` deberá divulgar el número de nucleótidos o aminoácidos de la secuencia contenidos en el elemento `INSDSeq_sequence`. Por ejemplo:

```
<INSDSeq_length>8</INSDSeq_length>
```

54. El elemento `INSDSeq_moltype` deberá divulgar el tipo de molécula que se está representando. En el caso de las secuencias de nucleótidos, incluidas las secuencias de análogos de nucleótidos, el tipo de molécula deberá indicarse como ADN o ARN. Para las secuencias de aminoácidos, el tipo de molécula deberá indicarse como AA. (Este elemento es distinto de los calificadores del calificador "mol_type" y "MOL_TYPE" descritos descrito en los párrafos 55 y 84). Por ejemplo:

```
<INSDSeq_moltype>AA</INSDSeq_moltype>
```

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.

56. El siguiente ejemplo ilustra la descripción de una secuencia de nucleótidos que contiene segmentos tanto de ADN como de ARN, en la forma prevista en el párrafo 55 *supra*:

```
<INSDSeq>
<INSDSeq_length>120</INSDSeq_length>
<INSDSeq_moltype>DNA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..120</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>other DNA</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
<INSDFeature>
<INSDFeature_key>misc_feature</INSDFeature_key>
<INSDFeature_location>1..60</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>DNA </INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
<INSDFeature>
<INSDFeature_key>misc_feature</INSDFeature_key>
<INSDFeature_location>61..120</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>RNA </INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
```

```
<INSDSeq_sequence>
cgaccacgctccgaggaaccaaccatcacgtttgaggacttcgtgaaggaattggataatacccgctcctaccaaagt
cgagcgcgcgactcattgctcctcgtaccgctcgagcggc
</INSDSeq_sequence>
</INSDSeq>
```

57. El elemento `INSDSeq_sequence` deberá divulgar la secuencia. Solo deberán incluirse en la secuencia los símbolos adecuados descritos en el Anexo I (véase el Cuadro 1 de la Sección 1 y el Cuadro 3 de la Sección 3). La secuencia no deberá incluir números, signos de puntuación o caracteres en blanco.

58. Toda secuencia ignorada deliberadamente deberá incluirse en la lista de secuencias y representarse de la siguiente manera:

- el elemento `SequenceData` y su atributo `sequenceIDNumber`, indicando como valor el identificador de secuencia de la secuencia ignorada;
- los elementos `INSDSeq_length`, `INSDSeq_moltype`, `INSDSeq_division`, presentes pero sin indicar ningún valor;
- el elemento `INSDSeq_feature-table` no deberá incluirse; y
- el elemento `INSDSeq_sequence`, con el valor "000".

59. El siguiente ejemplo ilustra la representación de una secuencia ignorada deliberadamente en la forma prevista en el párrafo 58 *supra*:

```
<SequenceData sequenceIDNumber="3">
<INSDSeq>
<INSDSeq_length/>
<INSDSeq_moltype/>
<INSDSeq_division/>
<INSDSeq_sequence>000</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
```

Cuadro de características

60. El cuadro de características contiene información sobre la localización y las funciones de varias regiones de una secuencia específica. Se requiere un cuadro de características para cada secuencia, excepto para toda secuencia ignorada deliberadamente, en cuyo caso no deberá figurar. El cuadro de características figura en el elemento `INSDSeq_feature-table`, que consiste en uno o más elementos `INSDFeature`.

61. Cada elemento `INSDFeature` describe una característica, y consiste en elementos dependientes de la siguiente manera:

Elemento	Descripción	Obligatorio/Facultativo
<code>INSDFeature_key</code>	Palabra o abreviatura que indica una característica	Obligatorio
<code>INSDFeature_location</code>	Región de la secuencia que corresponde a la característica	Obligatorio
<code>INSDFeature_qual</code>	Calificador que contiene información auxiliar acerca de la característica	Obligatorio si la clave de caracterización requiere uno o más calificadores, por ejemplo, <code>source</code> ; de lo contrario, Facultativo

Claves de caracterización

62. El Anexo I contiene la lista completa de claves de caracterización que deberán utilizarse en virtud de esta Norma, junto con la lista completa de los calificadores asociados y una indicación sobre el carácter obligatorio o facultativo de esos calificadores. La Sección 5 del Anexo I presenta la lista completa de las claves de caracterización para las secuencias de nucleótidos y la Sección 7 presenta la lista completa de las claves de caracterización para las secuencias de aminoácidos.

Claves de caracterización obligatorias

63. La clave de caracterización "source" es obligatoria para todas las secuencias de nucleótidos y la clave de caracterización "SOURCE" es obligatoria para todas las secuencias de aminoácidos, excepto para toda secuencia ignorada deliberadamente. Cada secuencia deberá tener una única clave de caracterización "source" o "SOURCE" que abarque toda la secuencia. Si una secuencia proviene de varias fuentes, esas fuentes podrán describirse con detalle en el cuadro de características, utilizando la clave de caracterización "misc_feature" y el calificador "note" para las secuencias de nucleótidos, y la clave de caracterización "REGION" y el calificador "NOTE note" para las secuencias de aminoácidos.

Localización de característica

64. El elemento obligatorio `INSDFeature_location` deberá contener al menos un descriptor de localización, que defina un sitio o una región correspondiente a la característica de la secuencia en el elemento `INSDSeq_sequence`. Las secuencias de aminoácidos deberán contener un único descriptor de localización en el elemento `INSDFeature_location` obligatorio. Las secuencias de nucleótidos podrán tener más de un descriptor de localización en elemento `INSDFeature_location` obligatorio cuando se utilice junto con uno o más operador(es) de localización (véanse los párrafos 67a 70).

65. El descriptor de localización puede ser el número de un único residuo, una región que delimite una serie de números de residuos contiguos, o un sitio o región que se extienda más allá del residuo, o de la serie de residuos que se ha especificado. El descriptor de localización no deberá incluir números de residuos fuera de la serie de la secuencia en el elemento `INSDSeq_sequence`. Para las secuencias de nucleótidos únicamente, el descriptor de localización podrá ser un sitio entre dos números de residuo adyacentes. Deberán utilizarse varios descriptores de localización junto con un operador de localización cuando la característica corresponda a sitios o regiones de discontinuos de una secuencia de nucleótidos (véanse los párrafos 67 a 70).

66. La sintaxis de cada tipo descriptor de localización se indica en el cuadro presentado a continuación, donde *x* e *y* son números de residuos, indicados como enteros positivos, no superiores a la longitud de la secuencia en el elemento `INSDSeq_sequence`, y *x* es menor que *y*.

a) Descriptores de localización para secuencias de nucleótidos y de aminoácidos:

Tipo de descriptor de localización	Sintaxis	Descripción
Número único de residuo	<i>x</i>	Designa un único residuo en la secuencia
Números de residuos que limitan un tramo de secuencia	<i>x</i> . <i>y</i>	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	< <i>x</i> > <i>x</i> < <i>x</i> . <i>y</i> <i>x</i> . > <i>y</i> < <i>x</i> . > <i>y</i>	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.

b) Descriptores de localización para secuencias de nucleótidos únicamente:

Tipo de descriptor de localización	Sintaxis	Descripción
Un sitio entre dos números de residuo adyacentes	<i>x</i> ^ <i>y</i>	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 <i>x</i> ^ <i>x</i> + 1 (por ejemplo 55 ^ 56), o para nucleótidos circulares, <i>x</i> ^ 1, donde "x" es la longitud total de la molécula, es decir, 1000 ^ 1 para una molécula circular con una longitud 1000.

c) Descriptores de localización para secuencias de aminoácidos únicamente:

Tipo de descriptor de localización	Sintaxis	Descripción
Números de residuo unidos por un enlace cruzado entre cadenas	<i>x</i> . <i>y</i>	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"

67. El elemento de secuencias de nucleótidos `INSDFeature_location` podrá contener uno o más operadores de localización. Un operador de localización es un prefijo de un descriptor de localización o de una combinación de descriptores de localización que corresponden a una característica única pero discontinua, y especifica el lugar correspondiente a la característica en la secuencia indicada, y la manera de interpretar la característica. A continuación se suministra una lista de operadores de localización con sus respectivas definiciones. Los operadores de localización podrán utilizarse únicamente para los nucleótidos.

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'.

68. Los operadores de localización de unión (join) u orden (order) requieren al menos dos descriptores de localización separados por una coma. Los descriptores de localización relativos a sitios situados entre dos residuos adyacentes, es decir x^y, no deberán utilizarse en una localización de unión u orden. La utilización del operador de localización join implica que los residuos designados por los descriptores de localización están puestos en contacto físicamente mediante procesos biológicos (por ejemplo, los exones que contribuyen a una característica de una región de codificación).

69. El operador de localización "complement" puede utilizarse en combinación ya sea con los operadores "join" u "order" dentro de la misma localización. No deberán utilizarse combinaciones de los operadores "join" y "order" en la misma localización.

70. Los siguientes ejemplos ilustran localizaciones de características en la forma prevista en los párrafos 64 a 69 *supra*:

a) localizaciones para las secuencias de nucleótidos y aminoácidos:

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.

b) localizaciones para secuencias de nucleótidos únicamente:

Ejemplo de localización	Descripción
123^124	Designa un sitio entre los residuos 123 y 124
join(12..78,134..202)	Indica que las regiones 12 a 78 y 134 a 202 deberían unirse para formar una secuencia contigua.
complement(34..126)	Comienza en el nucleótido complementario o a 126 y termina en el nucleótido complementario al nucleótido 34 (la característica está en la cadena complementaria a la cadena presentada).
complement(join(2691..4571, 4918..5163))	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).
join(complement(4918..5163), complement(2691..4571))	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).

c) localizaciones para secuencias de aminoácidos únicamente:

Ejemplo de localización	Descripción
340..565	Indica que los aminoácidos en las posiciones 340 y 565 están unidos por un enlace cruzado entre cadenas cuando se utilizan

	con una característica que indica un enlace cruzado entre cadenas, como "CROSSLNK" o "DISULFID".
--	--

71. En una instancia XML de una lista de secuencias, los caracteres "<" y ">" de un descriptor de localización deberán sustituirse por las entidades predefinidas adecuadas (véase el párrafo 41). Por ejemplo:

```
Feature location "<1":
<INSDFeature_location>&lt;1</INSDFeature_location>

Feature location "1..>888":
<INSDFeature_location>1..&gt;888</INSDFeature_location>
```

Calificadores de caracterización

72. Los calificadores se utilizan para suministrar información acerca de las características que complementa la información indicada por la clave de caracterización y la localización de característica. Existen tres tipos de formatos de valor para representar los diferentes tipos de información comunicada por los calificadores, a saber:

- texto libre (véanse los párrafos 85 a 87);
- vocabulario controlado o valores enumerados (por ejemplo, un número o fecha); y
- secuencias.

73. La Sección 6 del Anexo I contiene una lista completa de calificadores y sus formatos de valor especificados, dado el caso, para cada clave de caracterización de secuencia de nucleótidos, y la Sección 8 contiene la lista completa de calificadores para cada clave de caracterización de secuencia de aminoácidos.

74. Toda secuencia contemplada en el párrafo 7 que se indique como valor calificador deberá incluirse de manera clara en la lista de secuencias y se le deberá asignar su propio identificador de secuencia (véase el párrafo 10).

Calificadores de caracterización obligatorios

75. Una clave de caracterización obligatoria, es decir, "source" para las secuencias de nucleótidos y "SOURCE" para las secuencias de aminoácidos, deberá incluir dos calificadores obligatorios, "organism" y "mol_type" para las secuencias de nucleótidos y "ORGANISM" y "MOL_TYPE" para las secuencias de aminoácidos. Algunas claves de caracterización facultativas también necesitan calificadores obligatorios.

Elementos de los calificadores

76. El elemento INSDFeature_qualis contiene uno o más elementos INSDQualifier. Cada elemento INSDQualifier representa un calificador único y consiste en tres elementos dependientes de la siguiente manera:

Elemento	Descripción	Obligatorio/Facultativo
INSDQualifier_name	Nombre del calificador (véanse las Secciones 6 y 8 del Anexo I)	Obligatorio
INSDQualifier_value	Valor del calificador, dado el caso, en el formato especificado (véanse las Secciones 6 y 8 del Anexo I) y compuesto por los caracteres indicados en el párrafo 40.b).	Obligatorio, cuando se especifica (véanse el párrafo 87 y las Secciones 6 y 8 del Anexo I)
NonEnglishQualifier_value	Valor del calificador, dado el caso, en el formato especificado (véanse las Secciones 6 y 8 del Anexo I) y compuesto por los caracteres indicados en el párrafo 40.a).	Obligatorio, cuando se especifica (véanse el párrafo 87 y las Secciones 6 y 8 del Anexo I).

77. El calificador de organismo, es decir "organism" para las secuencias de nucleótidos (véase la Sección 6 del Anexo I) y "ORGANISM" para las secuencias de aminoácidos (véase la Sección 8 del Anexo I) deberá divulgar la fuente, es decir, un organismo único u origen de la secuencia. Las designaciones de los organismos deberían seleccionarse a partir de una base de datos taxonómica.

78. Si la secuencia existe en estado natural y el organismo fuente tiene una designación de género y especie en latín, esa designación podrá utilizarse como valor calificador. El nombre más usual en inglés podrá especificarse utilizando el calificador "note" para las secuencias de nucleótidos y el calificador "NOTE" para las secuencias de aminoácidos, pero no deberá utilizarse en el valor calificador del organismo.

79. Los siguientes ejemplos ilustran la fuente de la secuencia en la forma prevista en los párrafos 77 y 78 *supra*:

Ejemplo 1: Fuente de una secuencia de nucleótidos.

```
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..5164</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Solanum lycopersicum</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier>
<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>common name: tomato</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>genomic DNA</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
```

Ejemplo 2: Fuente de una secuencia de aminoácidos.

```
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>SOURCEsource</INSDFeature_key>
<INSDFeature_location>1..174</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>ORGANISMorganism</INSDQualifier_name>
<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier>
<INSDQualifier_name>MOL_TYPEmol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
```

80. Si la secuencia existe en estado natural y existe el nombre en latín del género del organismo fuente, pero la especie no se ha especificado o identificado, el valor calificador del organismo deberá indicar el género en latín seguido por "sp.". Por ejemplo:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Bacillus sp.</INSDQualifier_value>
```

81. Si la secuencia existe en estado natural, pero se desconoce la designación en latín del género y la especie del organismo, se deberá indicar "unidentified" como valor calificador del organismo. Toda información taxonómica conocida debería indicarse en el calificador "note" para las secuencias de nucleótidos y en el calificador "NOTEnote" para las secuencias de aminoácidos. Por ejemplo:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>unidentified</INSDQualifier_value>
<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>bacterium B8</INSDQualifier_value>
```

82. Si la secuencia existe en estado natural y el organismo fuente no tiene una designación en latín de género y especie, tal como un virus, entonces deberá utilizarse cualquier otro nombre científico aceptable (por ejemplo, "Canine adenovirus tipo 2") como valor calificador del organismo. Por ejemplo:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Canine adenovirus type 2</INSDQualifier_value>
```

83. Si la secuencia no existe en estado natural, deberá indicarse "synthetic construct" como valor calificador del organismo. Se podrá especificar información adicional sobre la manera en que se generó la secuencia utilizando el calificador "note" para las secuencias de nucleótidos y el calificador "NOTEnote" para las secuencias de aminoácidos. Por ejemplo:

```
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>SOURCEsource</INSDFeature_key>
<INSDFeature_location>1..40</INSDFeature_location>
<INSDFeature_qual>
```



```
<INSDQualifier>
  <INSDQualifier_name>ORGANISMorganism</INSDQualifier_name>
  <INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier>
  <INSDQualifier_name>MOL_TYPEmol_type</INSDQualifier_name>
  <INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier>
  <INSDQualifier_name>NOTEnote</INSDQualifier_name>
  <INSDQualifier_value>synthetic peptide used as assay for
    antibodies</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
```

84. El calificador "mol_type" para las secuencias de nucleótidos (véase la Sección 6 del Anexo I) y el calificador "MOL_TYPEmol_type" para las secuencias de aminoácidos (véase la Sección 8 del Anexo I) deberá divulgar el tipo de molécula representado en la secuencia. Estos calificadores son distintos del elemento INSDSeq_moltype descrito en el párrafo 54:

- a) Para una secuencia de nucleótidos, el valor calificador de "mol_type" deberá ser uno de los siguientes: "genomic DNA", "genomic RNA", "mRNA", "tRNA", "rRNA", "other RNA", "other DNA", "transcribed RNA", "viral cRNA", "unassigned DNA", o "unassigned RNA". Si la secuencia no existe en estado natural, es decir si el valor del calificador "organism" es "synthetic construct", el valor calificador de "mol_type" deberá ser "other RNA" o "other DNA";
- b) Para las secuencias de aminoácidos, el valor calificador de "MOL_TYPEmol_type" es "protein".

Texto libre

85. El texto libre, como se indica en el párrafo 3.n), es un tipo de formato de valor para ciertos calificadores, que se presenta en forma de una frase de texto descriptiva u otro formato específico (como se indica en el Anexo I).

86. La utilización del texto libre deberá limitarse a unos cuantos términos cortos indispensables para entender una característica de la secuencia. Para cada calificador, el texto libre no deberá ser superior a 1000 caracteres.

87. El texto libre dependiente del idioma, como se indica en el párrafo 3.o), es el valor de texto libre de ciertos calificadores que depende del idioma y que podrá exigir traducción para los procedimientos internacionales, nacionales o regionales. Los calificadores para secuencias de nucleótidos con un formato de valor de texto libre dependiente del idioma se identifican en la Sección 6 y el Cuadro 5 del Anexo I. Los calificadores para secuencias de aminoácidos con un formato de valor de texto libre dependiente del idioma se identifican en la Sección 8 y el Cuadro 6 del Anexo I.

a) El texto libre dependiente del idioma deberá presentarse en el elemento INSDQualifier_value en inglés, o en el elemento NonEnglishQualifier_value en un idioma distinto del inglés, o en ambos elementos. Nótese que, si el nombre de un organismo es un nombre latín de género y especie, no se requerirá traducción. Los términos técnicos y los nombres propios procedentes de palabras no inglesas que se utilizan a escala internacional se consideran términos en inglés a efectos del valor del elemento INSDQualifier_value (por ejemplo, *in vitro*, *in vivo*).

b) Si el elemento NonEnglishQualifier_value está presente en una lista de secuencias, el código del idioma correspondiente (véase la referencia en el párrafo 9 a la norma ISO 639-1:2002) deberá indicarse en el atributo nonEnglishFreeTextLanguageCode en el elemento raíz (véase el párrafo 43). Todos los elementos NonEnglishQualifier_value en una única lista de secuencias deberán tener valores en el idioma indicado en el atributo nonEnglishFreeTextLanguageCode. El elemento NonEnglishQualifier_value está únicamente permitido para los calificadores que tienen un formato de valor de texto libre dependiente del idioma.

c) Cuando ambos elementos NonEnglishQualifier_value y INSDQualifier_value estén presentes para un único calificador, la información contenida en los dos elementos deberá ser equivalente. Deberá darse una de las condiciones siguientes: el elemento NonEnglishQualifier_value contiene una traducción del valor del elemento INSDQualifier_value; o INSDQualifier_value contiene una traducción del valor del elemento NonEnglishQualifier_value; o el elemento INSDQualifier_value contiene una traducción del valor del elemento NonEnglishQualifier_value; o ambos elementos contienen una traducción del valor del calificador a partir del idioma especificado en el atributo originalFreeTextLanguageCode (véase el párrafo 43).

d) 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.

88. Los siguientes ejemplos ilustran la presentación del texto libre dependiente del idioma conforme a lo descrito en el párrafo 87.

Ejemplo 1: Texto libre dependiente del idioma en un elemento INSDQualifier_value:

```
<INSDFeature>
  <INSDFeature_key>regulatory</INSDFeature_key>
  <INSDFeature_location>1..60</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q1">
      <INSDQualifier_name>function</INSDQualifier_name>
      <INSDQualifier_value>binds to regulatory protein
Est3</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Ejemplo 2: Texto libre dependiente del idioma en un elemento INSDQualifier_value y un elemento NonEnglishQualifier_value:

```
<INSDFeature>
  <INSDFeature_key>ACT_SITE</INSDFeature_key>
  <INSDFeature_location>51..64</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q45">
      <INSDQualifier_name>NOTE note</INSDQualifier_name>
      <INSDQualifier_value>cleaves carbohydrate chain</INSDQualifier_value>
      <NonEnglishQualifier_value>clive la chaîne glucidique</NonEnglishQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Ejemplo 3: Texto libre dependiente del idioma en un elemento NonEnglishQualifier_value:

```
<INSDFeature>
  <INSDFeature_key>ACT_SITE</INSDFeature_key>
  <INSDFeature_location>51..64</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q1034">
      <INSDQualifier_name>NOTE note</INSDQualifier_name>
      <NonEnglishQualifier_value>clive la chaîne glucidique
    </NonEnglishQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Secuencias codificadoras

89. La clave de caracterización "CDS" podrá utilizarse para identificar secuencias codificadoras, es decir secuencias de nucleótidos que corresponden a la secuencia de aminoácidos en una proteína y el codón de terminación. La localización de la característica "CDS" en el elemento obligatorio INSDFeature_location deberá incluir el codón de terminación.

90. Los calificadores "transl_table" y "translation" podrán utilizarse con la clave de caracterización "CDS" (véase el Anexo I). Cuando el calificador "transl_table" no se utiliza, se asume la utilización del cuadro de códigos normalizados (véase el Cuadro 7 de la Sección 9 del Anexo I).

91. El calificador "trans_except" deberá utilizarse con la clave de caracterización "CDS" y el calificador "translation" para identificar el codón que codifica la pirrolisina o la selenocisteína.

92. Toda secuencia de aminoácidos codificada por la secuencia codificadora y divulgada en un calificador "translation" contemplado en el párrafo 7 deberá ser incluida en la lista de secuencias y tener su propio identificador de secuencia. El identificador de secuencia asignado a la secuencia de aminoácidos deberá figurar como valor del calificador "protein_id" con la clave de caracterización "CDS". El calificador "ORGANISM organism" de la clave de caracterización "SOURCE source" para la secuencia de aminoácidos deberá ser idéntico al de su secuencia codificadora. Por ejemplo:

```
<INSDFeature>
  <INSDFeature_key>CDS</INSDFeature_key>
  <INSDFeature_location>1..507</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>transl_table</INSDQualifier_name>
      <INSDQualifier_value>11</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>translation</INSDQualifier_name>
      <INSDQualifier_value>
```

```
MLVHLERTTIMFDFSSLINLPLIWGLLIAIAVLLYILMDGFDLIGIGILLPFAPSDKCRDHMISSIAPFWDGNETWLVLGGGGGL
FAAFPLAYSILMPAFYIPIIIIMLLGLIVRGVSEFEFRFKAEGKYRRLWDYAFHFHGSGLGAAFCQGMILGAFIHGVEVNGRNFSSGG
QLM
  </INSDQualifier_value>
</INSDQualifier>
<INSDQualifier>
  <INSDQualifier_name>protein_id</INSDQualifier_name>
  <INSDQualifier_value>89</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

Variantes

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.

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).

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 87).

96. El cuadro presentado a continuación indica la utilización adecuada de las claves y calificadores de caracterización para las variantes de secuencias de ácidos nucleicos y aminoácidos:

Tipo de secuencia	Clave de caracterización	Calificador	Utilización
Ácido 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.
Ácido nucleico	misc_difference	replace o note	Variabilidad introducida artificialmente, por ejemplo, mediante manipulación genética o síntesis química.
Aminoácido	VAR_SEQ	NOTE note	Variante producida por empalme diferencial, utilización de promotor alternativo, iniciación alternativa y o desplazamiento del marco ribosomal.
Aminoácido	VARIANT	NOTE note	Todo tipo de variante para el cual VAR_SEQ no es aplicable.

97. La anotación de una secuencia para una variante específica deberá contener una clave de caracterización y un calificador, como se indica en el cuadro *supra*, y la localización de la característica. El valor del calificador "replace" deberá ser exclusivamente un único nucleótido alternativo o una secuencia de nucleótidos en la cual se utilicen los símbolos que se indican en el Cuadro 1 de la Sección 1 o un valor calificador vacío. El valor del calificador "note" o "NOTE" podrá ser una lista de residuos alternativos. En particular, se deberá indicar una lista de aminoácidos alternativos como valor del calificador "NOTE note" cuando se use "X" en una secuencia, pero represente un subgrupo valor distinto de "uno de los símbolos 'A', 'R', 'N', 'D', 'C', 'Q', 'E', 'G', 'H', 'I', 'L', 'K', 'M', 'F', 'P', 'O', 'S', 'U', 'T', 'W', 'Y', o 'V'" (véase el párrafo 27). Toda supresión deberá representarse mediante un valor calificador vacío para el calificador "replace" o una indicación en el calificador "note" o "NOTE" de que se podrá suprimir el residuo. Todo residuo insertado o sustituido deberá indicarse en el calificador "replace", "note" o "NOTE". El valor de los calificadores "replace", "note" y "NOTE" es un texto libre que no deberá ser superior a 1000 caracteres, tal como se prevé en el párrafo 86. Para las secuencias contempladas en el párrafo 6 que se presentan como inserción o sustitución en un valor calificador, véase el párrafo 100.

98. Los símbolos descritos en el Anexo I (véanse los Cuadros 1 a 4 de las Secciones 1 a 4, respectivamente) deberían utilizarse para representar las variantes de residuos, dado el caso. Para el calificador “note” o “NOTE”, cuando la variante de residuo es un residuo modificado que no se describe en los Cuadros 2 o 4 del Anexo I, deberá indicarse el nombre completo no abreviado del residuo modificado como valor calificador. Los residuos modificados deberán ser descritos más ampliamente en el cuadro de características según se indica en el párrafo 17 o 30.

99. Los siguientes ejemplos ilustran la representación de variantes en la forma prevista en los párrafos 95 a 98 *supra*:

Ejemplo 1: Clave de caracterización “misc difference” para nucleótidos alternativos. La “n” en la posición 53 de la secuencia puede ser uno de los cinco nucleótidos alternativos.

```
<INSDFeature>
  <INSDFeature_key>misc difference</INSDFeature_key>
  <INSDFeature_location>53</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>w, cmnm5s2u, mam5u, mcm5s2u, or p</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>53</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>cmnm5s2u, mam5u, mcm5s2u, or p</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Ejemplo 2: Clave de caracterización “misc_difference” para una supresión en una secuencia de nucleótidos. Se ha suprimido el nucleótido en la posición 413.

```
<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>413</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value></INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Ejemplo 3: Clave de caracterización “misc_difference” para una inserción en una secuencia de nucleótidos. La secuencia “atgccaaatat” se ha insertado entre las posiciones 100 y 101 de la secuencia principal.

```
<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>100^101</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value>atgccaaatat</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Ejemplo 4: Clave de caracterización “variation” para una sustitución en una secuencia de nucleótidos. La citosina sustituye al nucleótido indicado en la posición 413 de la secuencia.

```
<INSDFeature>
  <INSDFeature_key>variation</INSDFeature_key>
  <INSDFeature_location>413</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
```

```
<INSDQualifier_value>c</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

Ejemplo 5: Clave de caracterización "VARIANT" para una sustitución en una secuencia de aminoácidos. El aminoácido definido en la posición 100 de la secuencia puede sustituirse por I, A, F, Y, aIle, MeIle o Nle.

```
<INSDFeature>
<INSDFeature_key>VARIANT</INSDFeature_key>
<INSDFeature_location>100</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>NOTE<note></INSDQualifier_name>
<INSDQualifier_value>I, A, F, Y, aIle, MeIle, or Nle
</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
<INSDFeature>
<INSDFeature_key>MOD_RES</INSDFeature_key>
<INSDFeature_location>100</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>NOTE<note></INSDQualifier_name>
<INSDQualifier_value>aIle, MeIle, or Nle</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

Ejemplo 6: Clave de caracterización "VARIANT" para una sustitución en una secuencia de aminoácidos. El aminoácido definido en la posición 100 de la secuencia puede sustituirse por cualquier aminoácido, excepto por Lys, Arg o His.

```
<INSDFeature>
<INSDFeature_key>VARIANT</INSDFeature_key>
<INSDFeature_location>100</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>NOTE<note></INSDQualifier_name>
<INSDQualifier_value>not K, R, or H</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

100. Toda secuencia contemplada en el párrafo 7 que se indique como una inserción o sustitución en un valor calificador para una anotación de secuencia principal también deberá incluirse en la lista de secuencias y se le asignará su propio identificador de secuencia.

[Sigue el Anexo I de la Norma ST.26]

ANEXO I

VOCABULARIO CONTROLADO

Versión 1.4.5

Revisión aprobada Propuesta presentada por el Comité Equipo Técnico sobre Listas de Normas Técnicas de Secuencias para que el CWS la OMPI (CWS) examine y apruebe en su octava novena sesión, celebrada el 4 de diciembre de 2020

ÍNDICE

Revisión aprobada Propuesta presentada por el Comité Equipo Técnico sobre Listas de Normas Técnicas de Secuencias para que el CWS la OMPI (CWS) examine y apruebe en su octava novena sesión, celebrada el 4 de diciembre de 2020	3
INTRODUCCIÓN	3
DEFINICIONES	3
ÁMBITO	4
REFERENCIAS	5
REPRESENTACIÓN DE SECUENCIAS	5
ESTRUCTURA DE LA LISTA DE SECUENCIAS EN FORMATO XML	10
ANEXO I	26
SECCIÓN 1: LISTA DE NUCLEÓTIDOS	31
SECCIÓN 2: LISTA DE NUCLEÓTIDOS MODIFICADOS	31
SECCIÓN 3: LISTA DE AMINOÁCIDOS	33
SECCIÓN 4: LISTA DE AMINOÁCIDOS MODIFICADOS	34
SECCIÓN 5: CLAVES DE CARACTERIZACIÓN (FEATURE KEYS) PARA SECUENCIAS DE NUCLEÓTIDOS	35
5.1. Feature Key C_region	35
5.2. Feature Key CDS	35
5.3. Feature Key centromere	36
5.4. Feature Key D-loop	36
5.5. Feature Key D_segment	36
5.6. Feature Key exon	36
5.7. Feature Key gene	37
5.8. Feature Key iDNA	37
5.9. Feature Key intron	37
5.10. Feature Key J_segment	37
5.11. Feature Key mat_peptide	38
5.12. Feature Key misc_binding	38
5.13. Feature Key misc_difference	38
5.14. Feature Key misc_feature	39
5.15. Feature Key misc_recomb	39
5.16. Feature Key misc_RNA	39
5.17. Feature Key misc_structure	39
5.18. Feature Key mobile_element	40
5.19. Feature Key modified_base	40
5.20. Feature Key mRNA	40
5.21. Feature Key ncRNA	40
5.22. Feature Key N_region	41
5.23. Feature Key operon	41
5.24. Feature Key oriT	41
5.25. Feature Key polyA_site	42

5.26.	Feature Key precursor_RNA	42
5.27.	Feature Key prim_transcript	42
5.28.	Feature Key primer_bind	42
5.29.	Feature Key propeptide	43
5.30.	Feature Key protein_bind	43
5.31.	Feature Key regulatory	43
5.32.	Feature Key repeat_region.....	44
5.33.	Feature Key rep_origin	44
5.34.	Feature Key rRNA	44
5.35.	Feature Key S_region.....	44
5.36.	Feature Key sig_peptide.....	45
5.37.	Feature Key source	45
5.38.	Feature Key stem_loop.....	46
5.39.	Feature Key STS	46
5.40.	Feature Key telomere	46
5.41.	Feature Key tmRNA.....	46
5.42.	Feature Key transit_peptide	47
5.43.	Feature Key tRNA.....	47
5.44.	Feature Key unsure	47
5.45.	Feature Key V_region.....	48
5.46.	Feature Key V_segment.....	48
5.47.	Feature Key variation.....	48
5.48.	Feature Key 3'UTR.....	49
5.49.	Feature Key 5'UTR.....	49
	SECCIÓN 6: CALIFICADORES PARA SECUENCIAS DE NUCLEÓTIDOS	50
6.1.	Qualifier allele	50
6.2.	Qualifier anticodon.....	51
6.3.	Qualifier bound_moiety.....	51
6.4.	Qualifier cell_line	51
6.5.	Qualifier cell_type	51
6.6.	Qualifier chromosome.....	51
6.7.	Qualifier <u>circular_RNA</u>	52
6.8.	<u>Qualifier clone</u>	52
6.9.	Qualifier clone_lib	52
6.10.	Qualifier codon_start.....	52
6.11.	Qualifier collected_by	52
6.12.	Qualifier collection_date	53
6.13.	Qualifier compare	53
6.14.	Qualifier cultivar	53
6.15.	Qualifier dev_stage.....	53
6.16.	Qualifier direction.....	53
6.17.	Qualifier EC_number	54
6.18.	Qualifier ecotype.....	54
6.19.	Qualifier environmental_sample.....	54
6.20.	Qualifier exception	55
6.21.	Qualifier frequency.....	55
6.22.	Qualifier function	55

6.23.	Qualifier gene	55
6.24.	Qualifier gene_synonym	55
6.25.	Qualifier germline.....	56
6.26.	Qualifier haplogroup	56
6.27.	Qualifier haplotype.....	56
6.28.	Qualifier host.....	57
6.29.	Qualifier identified_by	57
6.30.	Qualifier isolate	57
6.31.	Qualifier isolation_source	57
6.32.	Qualifier lab_host.....	58
6.33.	Qualifier lat_lon.....	58
6.34.	Qualifier macronuclear.....	58
6.35.	Qualifier map.....	58
6.36.	Qualifier mating_type.....	59
6.37.	Qualifier mobile_element_type	59
6.38.	Qualifier mod_base.....	59
6.39.	Qualifier mol_type.....	60
6.40.	Qualifier ncRNA_class.....	60
6.41.	Qualifier note	61
6.42.	Qualifier number	61
6.43.	Qualifier operon	61
6.44.	Qualifier organelle.....	61
6.45.	Qualifier organism.....	62
6.46.	Qualifier PCR_primers.....	62
6.47.	Qualifier phenotype.....	62
6.48.	Qualifier plasmid	62
6.49.	Qualifier pop_variant.....	63
6.50.	Qualifier product	63
6.51.	Qualifier protein_id.....	63
6.52.	Qualifier proviral.....	63
6.53.	Qualifier pseudo.....	63
6.54.	Qualifier pseudogene.....	64
6.55.	Qualifier rearranged.....	64
6.56.	Qualifier recombination_class	64
6.57.	Qualifier regulatory_class	65
6.58.	Qualifier replace.....	65
6.59.	Qualifier ribosomal_slippage	66
6.60.	Qualifier rpt_family.....	66
6.61.	Qualifier rpt_type	66
6.62.	Qualifier rpt_unit_range	67
6.63.	Qualifier rpt_unit_seq.....	67
6.64.	Qualifier satellite	67
6.65.	Qualifier segment.....	67
6.66.	Qualifier serotype.....	67
6.67.	Qualifier serovar	68
6.68.	Qualifier sex	68
6.69.	Qualifier standard_name	68

6.70.	Qualifier strain.....	68
6.71.	Qualifier sub_clone	69
6.72.	Qualifier sub_species	69
6.73.	Qualifier sub_strain.....	69
6.74.	Qualifier tag_peptide.....	69
6.75.	Qualifier tissue_lib	70
6.76.	Qualifier tissue_type	70
6.77.	Qualifier transl_except.....	70
6.78.	Qualifier transl_table.....	70
6.79.	Qualifier trans_splicing	71
6.80.	Qualifier translation.....	71
6.81.	Qualifier variety.....	71
	SECCIÓN 7: CLAVES DE CARACTERIZACIÓN PARA SECUENCIAS DE AMINOÁCIDOS	72
7.1.	Feature Key ACT_SITE	72
7.2.	Feature Key BINDING	72
7.3.	Feature Key CA_BIND.....	72
7.4.	Feature Key CARBOHYD.....	72
7.5.	Feature Key CHAIN.....	72
7.6.	Feature Key COILED.....	73
7.7.	Feature Key COMPBIAS	73
7.8.	Feature Key CONFLICT	73
7.9.	Feature Key CROSSLNK	73
7.10.	Feature Key DISULFID.....	73
7.11.	Feature Key DNA_BIND.....	73
7.12.	Feature Key DOMAIN.....	74
7.13.	Feature Key HELIX.....	74
7.14.	Feature Key INIT_MET.....	74
7.15.	Feature Key INTRAMEM.....	74
7.16.	Feature Key LIPID	75
7.17.	Feature Key METAL	75
7.18.	Feature Key MOD_RES	75
7.19.	Feature Key MOTIF.....	75
7.20.	Feature Key MUTAGEN.....	75
7.21.	Feature Key NON_STD.....	75
7.22.	Feature Key NON_TER.....	75
7.23.	Feature Key NP_BIND.....	76
7.24.	Feature Key PEPTIDE.....	76
7.25.	Feature Key PROPEP	76
7.26.	Feature Key REGION.....	76
7.27.	Feature Key REPEAT.....	76
7.28.	Feature Key SIGNAL.....	76
7.29.	Feature Key SITE	76
7.30.	Feature Key SOURCEsource.....	77
7.31.	Feature Key STRAND	77
7.32.	Feature Key TOPO_DOM.....	77
7.33.	Feature Key TRANSMEM.....	77
7.34.	Feature Key TRANSIT.....	77

7.35. Feature Key TURN	77
7.36. Feature Key UNSURE	77
7.37. Feature Key VARIANT.....	78
7.38. Feature Key VAR_SEQ	78
7.39. Feature Key ZN_FING.....	78
SECCIÓN 8: CALIFICADORES PARA SECUENCIAS DE AMINOÁCIDOS	79
8.1. Qualifier MOL_TYPEmol_type.....	79
8.2. Qualifier NOTEnote.....	79
8.3. Qualifier ORGANISMorganism.....	79
SECCIÓN 9: TABLAS DEL CÓDIGO GENÉTICO	80
ANEXO II	84
ANEXO III	89
ANEXO V	90
ANEXO VI	91
DOCUMENTO DE ORIENTACIÓN	91
INTRODUCCIÓN.....	91
ÍNDICE DE EJEMPLOS	97
EJEMPLOS.....	106
APÉNDICE	162
ANEXO VII	163
Revisión aprobada Propuesta presentada por el Comité Equipo Técnico sobre Listas de Normas Técnicas de Secuencias para que el CWS la COMPI (CWS) examine y apruebe en su octava novena sesión, celebrada el 4 de diciembre de 2020.....	163
INTRODUCCIÓN.....	163
ALCANCE DEL DOCUMENTO	163
RECOMENDACIONES RELATIVAS A LA EVENTUAL ADICIÓN O SUPRESIÓN DE MATERIA	163
SECCIÓN 1: LISTA DE NUCLEÓTIDOS	30
SECCIÓN 2: LISTA DE NUCLEÓTIDOS MODIFICADOS.....	31
SECCIÓN 3: LISTA DE AMINOÁCIDOS.....	33
SECCIÓN 4: LISTA DE AMINOÁCIDOS MODIFICADOS.....	34
SECCIÓN 5: CLAVES DE CARACTERIZACIÓN (FEATURE KEYS) PARA SECUENCIAS DE NUCLEÓTIDOS.....	35
SECCIÓN 6: CALIFICADORES PARA SECUENCIAS DE NUCLEÓTIDOS	50
SECCIÓN 7: CLAVES DE CARACTERIZACIÓN PARA SECUENCIAS DE AMINOÁCIDOS	72
SECCIÓN 8: CALIFICADORES PARA SECUENCIAS DE AMINOÁCIDOS	79
SECCIÓN 9: TABLAS DEL CÓDIGO GENÉTICO	80

SECCIÓN 1: LISTA DE NUCLEÓTIDOS

En el cuadro 1 se presentan los **códigos/símbolos** de las bases de nucleótidos que se utilizarán en las listas de secuencias. El símbolo "t" se interpretará como timina en ADN y uracilo en ARN cuando se utilice sin una descripción detallada. Si fuera necesario utilizar un símbolo de ambigüedad (que represente dos o más bases en la alternativa), deberá utilizarse el símbolo más restrictivo. Por ejemplo, si una base en una determinada posición pudiera ser "a" o "g", se utilizará "r" en vez de "n". El símbolo "n" se interpretará como "a o c o g o t/u" cuando se utilice sin una descripción detallada.

Cuadro 1: Lista de nucleótidos

Símbolo	Nombre del nucleótido	Definición
a	adenine	
c	cytosine	
g	guanine	
t	thymine en ADN/uracil en ARN (t/u)	
m	a o c	
r	a o g	
w	a o t/u	
s	c o g	
y	c o t/u	
k	g o t/u	
v	a o c o g; no t/u	
h	a o c or t/u; no g	
d	a o g or t/u; no c	
b	c o g or t/u; no a	
n	a o c o g o t/u; "unknown" o "other"	

SECCIÓN 2: LISTA DE NUCLEÓTIDOS MODIFICADOS

La lista de abreviaturas que se enumeran en el cuadro 2 son los únicos valores permitidos para el calificador "mod_base". Cuando en el cuadro que figura a continuación no figure un nucleótido modificado específico, se utilizará la abreviatura "OTHER" como su valor. Si la abreviatura es "OTHER", el nombre completo no abreviado de la base modificada deberá indicarse como valor en un calificador "note". Las abreviaturas que figuran en el cuadro 2 no deberán utilizarse en la propia secuencia.

Cuadro 2: Lista de nucleótidos modificados

Abreviatura	Nucleótido modificado	Definición
ac4c	4-acetylcytidine	
chm5u	5-(carboxyhydroxymethyl)uridine	
cm	2-O-methylcytidine	
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine	
cmnm5u	5-carboxymethylaminomethyluridine	
dhu	dihydrouridine	
fm	2-O-methylpseudouridine	
gal q	beta-D-galactosylqueuosine	
gm	2-O-methylguanosine	
i	inosine	
i6a	N6-isopentenyladenosine	
m1a	1-methyladenosine	
m1f	1-methylpseudouridine	
m1g	1-methylguanosine	
m1i	1-methylinosine	
m22g	2,2-dimethylguanosine	
m2a	2-methyladenosine	
m2g	2-methylguanosine	
m3c	3-methylcytidine	
m4c	N4-methylcytosine	
m5c	5-methylcytidine	

Abreviatura	Nucleótido modificado Definición
m6a	N6-methyladenosine
m7g	7-methylguanosine
mam5u	5-methylaminomethyluridine
mam5s2u	5-methylaminomethyl-2-thiouridine
man q	beta-D-mannosylqueuosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine
mv	uridine-5-oxoacetic acid-methylester
o5u	uridine-5-oxyacetic acid (v)
osyw	wybutoxosine
p	pseudouridine
q	queuosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
m5u	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	wybutosine
x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u
OTHER	(Requiere calificador "note")

SECCIÓN 3: LISTA DE AMINOÁCIDOS

En el cuadro 3 se presentan los **códigos** **símbolos** de los aminoácidos que se utilizarán en secuencias. Si fuera necesario utilizar un símbolo de ambigüedad (que represente dos o más aminoácidos en la alternativa), se utilizará el símbolo más restrictivo. Por ejemplo, si un aminoácido en una determinada posición pudiera ser ácido aspártico o asparagina, se utilizará el símbolo "B" en vez de "X". El símbolo "X" se interpretará como "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" o "V", cuando se utilice sin una descripción detallada.

Cuadro 3: Lista de **símbolos de los** aminoácidos

Símbolo	Nombre del aminoácido	Definición
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	Pyrolysine	
S	Serine	
U	Selenocysteine	
T	Threonine	
W	Tryptophan	
Y	Tyrosine	
V	Valine	
B	Aspartic acid o Asparagine	
Z	Glutamine o Glutamic acid	
J	Leucine o Isoleucine	
X	A o R o N o D o C o Q o E o G o H o I o L o K o M o F o P o O o S o U o T o W o Y o V; "unknown" o "other"	

SECCIÓN 4: LISTA DE AMINOÁCIDOS MODIFICADOS

En el cuadro 4 se enumeran las únicas abreviaturas permitidas para aminoácidos modificados en el calificador obligatorio "NOTE_{note}" con respecto a las claves de caracterización "MOD_RES" o "SITE". El valor para el calificador "NOTE_{note}" deberá corresponder a una abreviatura del cuadro, cuando proceda, o al nombre completo no abreviado del aminoácido modificado. Las abreviaturas (o nombres completos) que figuran en este cuadro no deberán utilizarse en la propia secuencia.

Cuadro 4: Lista de aminoácidos modificados

Abreviatura	Aminoácido modificado
Aad	2-Aminoadipic acid
bAad	3-Aminoadipic acid
bAla	beta-Alanine, beta-Aminopropionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4-Diaminobutyric acid
Des	Desmosine
Dpm	2,2'-Diaminopimelic acid
Dpr	2,3-Diaminopropionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
Ide	Isodesmosine
alle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
Melle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine
Orn	Ornithine

SECCIÓN 5: CLAVES DE CARACTERIZACIÓN (FEATURE KEYS) PARA SECUENCIAS DE NUCLEÓTIDOS

Esta sección contiene la lista de claves de caracterización permitidas que deberán utilizarse en las secuencias de nucleótidos junto con la lista de calificadores obligatorios y facultativos. Las claves de caracterización se enumeran por orden alfabético. Las claves de caracterización pueden utilizarse para el ADN o el ARN, salvo que se indique otra cosa en "Molecule scope". En algunos casos, puede ser adecuado utilizar determinadas claves de caracterización con secuencias artificiales, además del "organism scope" especificado.

Los nombres de las claves de caracterización deben utilizarse en la instancia XML de la lista de secuencias tal y como aparecen después de "Feature key" en las descripciones que figuran a continuación, a excepción de las claves de caracterización 3'UTR y 5'UTR. Véase "Comment" en la descripción de las claves de caracterización 3'UTR y 5'UTR.

5.1. Feature Key	C_region
Definition	constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; includes one or more exons depending on the particular chain
Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
Organism scope	eukaryotes
5.2. Feature Key	CDS
Definition	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature may include amino acid conceptual translation
Optional qualifiers	allele circular_RNA codon_start EC_number exception function gene gene_synonym map note number operon product protein_id pseudo pseudogene ribosomal_slippage standard_name translation transl_except transl_table trans_splicing
Comment	codon_start qualifier has valid value of 1 or 2 or 3, indicating the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature; transl_table defines the genetic code table used if other than the Standard or universal genetic code table; genetic code exceptions outside the range of the specified tables are reported in transl_except qualifier; only one of the qualifiers translation, pseudogene or pseudo are permitted with a CDS feature key; when the translation qualifier is used, the protein_id qualifier is mandatory if the translation product contains four or more specifically defined amino acids

5.3. Feature Key	centromere
Definition	region of biological interest identified as a centromere and which has been experimentally characterized
Optional qualifiers	note standard_name
Comment	the centromere feature describes the interval of DNA that corresponds to a region where chromatids are held and a kinetochore is formed

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

5.5. Feature Key	D_segment
Definition	Diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain
Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
Organism scope	eukaryotes

5.6. Feature Key	exon
Definition	region of genome that codes for portion of spliced mRNA,rRNA and tRNA; may contain 5'UTR, all CDSs and 3' UTR
Optional qualifiers	allele EC_number function gene gene_synonym map note number product pseudo pseudogene standard_name trans_splicing

5.7. Feature Key	gene
Definition	region of biological interest identified as a gene and for which a name has been assigned
Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene phenotype standard_name trans_splicing
Comment	the gene feature describes the interval of DNA that corresponds to a genetic trait or phenotype; the feature is, by definition, not strictly bound to its positions at the ends; it is meant to represent a region where the gene is located.

5.8. Feature Key	iDNA
Definition	intervening DNA; DNA which is eliminated through any of several kinds of recombination
Optional qualifiers	allele function gene gene_synonym map note number standard_name
Molecule scope	DNA
Comment	e.g., in the somatic processing of immunoglobulin genes.

5.9. Feature Key	intron
Definition	a segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it
Optional qualifiers	allele function gene gene_synonym map note number pseudo pseudogene standard_name trans_splicing

5.10. Feature Key	J_segment
Definition	joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains
Optional qualifiers	allele

	gene
	gene_synonym
	map
	note
	product
	pseudo
	pseudogene
	standard_name
Organism scope	eukaryotes

5.11. Feature Key	mat_peptide
Definition	mature peptide or protein coding sequence; coding sequence for the mature or final peptide or protein product following post-translational modification; the location does not include the stop codon (unlike the corresponding CDS)
Optional qualifiers	allele EC_number function gene gene_synonym map note product pseudo pseudogene standard_name

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

5.13. Feature Key	misc_difference
Definition	featured sequence differs from the presented sequence at this location and cannot be described by any other Difference key (variation, or modified_base)
Optional qualifiers	allele clone compare gene gene_synonym map note phenotype replace standard_name
Comment	the misc_difference feature key must be used to describe variability introduced artificially, e.g., by genetic manipulation or by chemical synthesis; use the

replace qualifier to annotate a deletion, insertion, or substitution. The variation feature key must be used to describe naturally occurring genetic variability.

5.14. Feature Key	misc_feature
Definition	region of biological interest which cannot be described by any other feature key; a new or rare feature
Optional qualifiers	allele function gene gene_synonym map note number phenotype product pseudo pseudogene standard_name
Comment	this key should not be used when the need is merely to mark a region in order to comment on it or to use it in another feature's location

5.15. Feature Key	misc_recomb
Definition	site of any generalized, site-specific or replicative recombination event where there is a breakage and reunion of duplex DNA that cannot be described by other recombination keys or qualifiers of source key (proviral)
Optional qualifiers	allele gene gene_synonym map note recombination_class standard_name
Molecule scope	DNA

5.16. Feature Key	misc_RNA
Definition	any transcript or RNA product that cannot be defined by other RNA keys (prim_transcript, precursor_RNA, mRNA, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, ncrRNA, rRNA and tRNA)
Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing

5.17. Feature Key	misc_structure
Definition	any secondary or tertiary nucleotide structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)

Optional qualifiers	allele function gene gene_synonym map note standard_name
<hr/>	
5.18. Feature Key	mobile_element
Definition	region of genome containing mobile elements
Mandatory qualifiers	mobile_element_type
Optional qualifiers	allele function gene gene_synonym map note rpt_family rpt_type standard_name
<hr/>	
5.19. Feature Key	modified_base
Definition	the indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)
Mandatory qualifiers	mod_base
Optional qualifiers	allele frequency gene gene_synonym map note
Comment	value for the mandatory mod_base qualifier is limited to the restricted vocabulary for modified base abbreviations in Section 2 of this Annex.
<hr/>	
5.20. Feature Key	mRNA
Definition	messenger RNA; includes 5' untranslated region (5'UTR), coding sequences (CDS, exon) and 3' untranslated region (3'UTR)
Optional qualifiers	allele circular_RNA function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
<hr/>	
5.21. .Feature Key	ncRNA
Definition	a non-protein-coding gene, other than ribosomal RNA and transfer RNA, the functional molecule of which is the RNA transcript

Mandatory qualifiers	ncRNA_class
Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
Comment	the ncRNA feature must not be used for ribosomal and transfer RNA annotation, for which the rRNA and tRNA feature keys must be used, respectively

5.22. Feature Key	N_region
Definition	extra nucleotides inserted between rearranged immunoglobulin segments
Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
Organism scope	eukaryotes

5.23. Feature Key	operon
Definition	region containing polycistronic transcript including a cluster of genes that are under the control of the same regulatory sequences/promoter and in the same biological pathway
Mandatory qualifiers	operon
Optional qualifiers	allele function map note phenotype pseudo pseudogene standard_name

5.24. Feature Key	oriT
Definition	origin of transfer; region of a DNA molecule where transfer is initiated during the process of conjugation or mobilization
Optional qualifiers	allele bound_moiety direction gene gene_synonym map note rpt_family rpt_type

	rpt_unit_range rpt_unit_seq standard_name
Molecule scope	DNA
Comment	rep_origin must be used to describe origins of replication; direction qualifier has permitted values left, right, and both, however only left and right are valid when used in conjunction with the oriT feature; origins of transfer can be present in the chromosome; plasmids can contain multiple origins of transfer
<hr/>	
5.25. Feature Key	polyA_site
Definition	site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation
Optional qualifiers	allele gene gene_synonym map note
Organism scope	eukaryotes and eukaryotic viruses
<hr/>	
5.26. Feature Key	precursor_RNA
Definition	any RNA species that is not yet the mature RNA product; may include ncRNA, rRNA, tRNA, 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR)
Optional qualifiers	allele function gene gene_synonym map note operon product standard_name trans_splicing
Comment	used for RNA which may be the result of post-transcriptional processing; if the RNA in question is known not to have been processed, use the prim_transcript key
<hr/>	
5.27. Feature Key	prim_transcript
Definition	primary (initial, unprocessed) transcript; may include ncRNA, rRNA, tRNA, 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR)
Optional qualifiers	allele function gene gene_synonym map note operon standard_name
<hr/>	
5.28. Feature Key	primer_bind
Definition	non-covalent primer binding site for initiation of replication, transcription, or reverse transcription; includes site(s) for synthetic e.g., PCR primer

	elements
Optional qualifiers	allele gene gene_synonym map note standard_name
Comment	used to annotate the site on a given sequence to which a primer molecule binds - not intended to represent the sequence of the primer molecule itself; since PCR reactions most often involve pairs of primers, a single primer_bind key may use the order(location,location) operator with two locations, or a pair of primer_bind keys may be used
<hr/>	
5.29. Feature Key	propeptide
Definition	propeptide coding sequence; coding sequence for the domain of a proprotein that is cleaved to form the mature protein product.
Optional qualifiers	allele function gene gene_synonym map note product pseudo pseudogene standard_name
<hr/>	
5.30. Feature Key	protein_bind
Definition	non-covalent protein binding site on nucleic acid
Mandatory qualifiers	bound_moiety
Optional qualifiers	allele function gene gene_synonym map note operon standard_name
Comment	note that the regulatory feature key and regulatory_class qualifier with the value "ribosome_binding_site" must be used to describe ribosome binding sites
<hr/>	
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

5.32. Feature Key	repeat_region
Definition	region of genome containing repeating units
Optional qualifiers	allele function gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq satellite standard_name

5.33. Feature Key	rep_origin
Definition	origin of replication; starting site for duplication of nucleic acid to give two identical copies
Optional qualifiers	allele direction function gene gene_synonym map note standard_name
Comment	direction qualifier has valid values: left, right, or both

5.34. Feature Key	rRNA
Definition	mature ribosomal RNA; RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins
Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name
Comment	rRNA sizes should be annotated with the product qualifier

5.35. Feature Key	S_region
Definition	switch region of immunoglobulin heavy chains; involved in the rearrangement of heavy chain DNA leading to the expression of a different immunoglobulin class from the same B-cell
Optional qualifiers	allele gene

	gene_synonym map note product pseudo pseudogene standard_name
Organism scope	eukaryotes

5.36. Feature Key	sig_peptide
Definition	signal peptide coding sequence; coding sequence for an N-terminal domain of a secreted protein; this domain is involved in attaching nascent polypeptide to the membrane leader sequence
Optional qualifiers	allele function gene gene_synonym map note product pseudo pseudogene standard_name

5.37. Feature Key	source
Definition	identifies the source of the sequence; this key is mandatory; every sequence will have a single source key spanning the entire sequence
Mandatory qualifiers	organism mol_type
Optional qualifiers	cell_line cell_type chromosome clone clone_lib collected_by collection_date cultivar dev_stage ecotype environmental_sample germline haplogroup haplotype host identified_by isolate isolation_source lab_host lat_lon macronuclear map mating_type note organelle PCR_primers plasmid pop_variant proviral rearranged segment serotype serovar

	sex strain sub_clone sub_species sub_strain tissue_lib tissue_type variety
Molecule scope	any
<hr/>	
5.38. Feature Key	stem_loop
Definition	hairpin; a double-helical region formed by base-pairing between adjacent (inverted) complementary sequences in a single strand of RNA or DNA
Optional qualifiers	allele function gene gene_synonym map note operon standard_name
<hr/>	
5.39. Feature Key	STS
Definition	sequence tagged site; short, single-copy DNA sequence that characterizes a mapping landmark on the genome and can be detected by PCR; a region of the genome can be mapped by determining the order of a series of STSs
Optional qualifiers	allele gene gene_synonym map note standard_name
Molecule scope	DNA
Comment	STS location to include primer(s) in primer_bind key or primers
<hr/>	
5.40. Feature Key	telomere
Definition	region of biological interest identified as a telomere and which has been experimentally characterized
Optional qualifiers	note rpt_type rpt_unit_range rpt_unit_seq standard_name
Comment	the telomere feature describes the interval of DNA that corresponds to a specific structure at the end of the linear eukaryotic chromosome which is required for the integrity and maintenance of the end; this region is unique compared to the rest of the chromosome and represents the physical end of the chromosome
<hr/>	
5.41. Feature Key	tmRNA
Definition	transfer messenger RNA; tmRNA acts as a tRNA first, and then as an mRNA that encodes a peptide tag; the ribosome translates this mRNA region of tmRNA and

attaches the encoded peptide tag to the C-terminus of the unfinished protein; this attached tag targets the protein for destruction or proteolysis

Optional qualifiers
 allele
 function
 gene
 gene_synonym
 map
 note
 product
 pseudo
 pseudogene
 standard_name
 tag_peptide

5.42. Feature Key transit_peptide

Definition transit peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organellar protein; this domain is involved in post-translational import of the protein into the organelle

Optional qualifiers
 allele
 function
 gene
 gene_synonym
 map
 note
 product
 pseudo
 pseudogene
 standard_name

5.43. Feature Key tRNA

Definition mature transfer RNA, a small RNA molecule (75-85 bases long) that mediates the translation of a nucleic acid sequence into an amino acid sequence

Optional qualifiers
 allele
circular_RNA
 anticodon
 function
 gene
 gene_synonym
 map
 note
 operon
 product
 pseudo
 pseudogene
 standard_name
 trans_splicing

5.44. Feature Key unsure

Definition a small region of sequenced bases, generally 10 or fewer in its length, which could not be confidently identified. Such a region might contain called bases (a, t, g, or c), or a mixture of called-bases and uncalled-bases ('n').

Optional qualifiers
 allele
 compare
 gene
 gene_synonym
 map
 note
 replace

Comment use the replace qualifier to annotate a deletion, insertion, or substitution.

5.45. Feature Key	V_region
Definition	variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for the variable amino terminal portion; can be composed of V_segments, D_segments, N_regions, and J_segments
Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
Organism scope	eukaryotes

5.46. Feature Key	V_segment
Definition	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for most of the variable region (V_region) and the last few amino acids of the leader peptide
Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
Organism scope	eukaryotes

5.47. Feature Key	variation
Definition	a related strain contains stable mutations from the same gene (e.g., RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others)
Optional qualifiers	allele compare frequency gene gene_synonym map note phenotype product replace standard_name
Comment	used to describe alleles, RFLP's, and other naturally occurring mutations and polymorphisms; use the replace qualifier to annotate a deletion, insertion, or substitution; variability arising as a result of genetic manipulation (e.g., site directed mutagenesis) must be described with the misc_difference feature

5.48. Feature Key	3'UTR
Definition	<p>1) region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein;</p> <p>2) region at the 3' end of an RNA virus (following the last stop codon) that is not translated into a protein;</p>
Optional qualifiers	<p>allele</p> <p>function</p> <p>gene</p> <p>gene_synonym</p> <p>map</p> <p>note</p> <p>standard_name</p> <p>trans_splicing</p>
Comment	<p>The apostrophe character has special meaning in XML, and must be substituted with "&apos;" in the value of an element. Thus "3'UTR" must be represented as "3&apos;UTR" in the XML file, i.e.,</p> <p><INSDFeature_key>3&apos;UTR</INSDFeature_key>.</p>

5.49. Feature Key	5'UTR
Definition	<p>1) region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein;</p> <p>2) region at the 5' end of an RNA virus (preceding the first initiation codon) that is not translated into a protein;</p>
Optional qualifiers	<p>allele</p> <p>function</p> <p>gene</p> <p>gene_synonym</p> <p>map</p> <p>note</p> <p>standard_name</p> <p>trans_splicing</p>
Comment	<p>The apostrophe character has special meaning in XML, and must be substituted with "&apos;" in the value of an element. Thus "5'UTR" must be represented as "5&apos;UTR" in the XML file, i.e.,</p> <p><INSDFeature_key>5&apos;UTR</INSDFeature_key>.</p>

SECCIÓN 6: CALIFICADORES PARA SECUENCIAS DE NUCLEÓTIDOS

Esta sección contiene la lista de calificadores que deberán utilizarse en las características de las secuencias de nucleótidos. Los calificadores se enumeran por orden alfabético.

Cuando el formato de valor (*value format*) sea “none”, no deberá utilizarse el elemento `INSDQualifier_value` ni tampoco el elemento `NonEnglishQualifier_value`.

Cuando el formato de valor sea texto libre identificado como dependiente del idioma, deberá utilizarse uno de los siguientes elementos:

- 1) el elemento `INSDQualifier_value; 0`
- 2) el elemento `NonEnglishQualifier_value; 0`
- 3) ambos elementos `INSDQualifier_value` y `NonEnglishQualifier_value`.

Cuando el formato de valor sea distinto de “none” pero no se identifique como texto libre dependiente del idioma, deberá utilizarse el elemento `INSDQualifier_value` y no deberá utilizarse el elemento `NonEnglishQualifier_value`.

NOTA: Todo valor calificador (*qualifier value*) que se indique para un calificador con formato de valor “free text” dependiente del idioma podrá requerir traducción para los procedimientos nacionales/regionales. Se considera que los calificadores enumerados en el siguiente cuadro tienen valores de texto libre dependiente del idioma.

Cuadro 5: Lista de valores calificadores para secuencias de nucleótidos con valores de texto libre dependiente del idioma

Sección	Valor de texto libre dependiente del idioma
6.3	bound_moiety
6.5	cell_type
6.78	clone
6.89	clone_lib
6.4011	collected_by
6.4314	cultivar
6.4415	dev_stage
6.4718	ecotype
6.2422	function
6.2324	gene_synonym
6.2526	haplogroup
6.2728	host
6.2829	identified_by
6.2930	isolate
6.3031	isolation_source
6.3432	lab_host
6.3536	mating_type
6.4041	note
6.4445	organism
6.4647	phenotype
6.4849	pop_variant
6.4950	product
6.6566	serotype
6.6667	serovar
6.6768	sex
6.6869	standard_name
6.6970	strain
6.7071	sub_clone
6.7472	sub_species
6.7273	sub_strain
6.7475	tissue_lib
6.7576	tissue_type
6.8081	variety

6.1.	Qualifier	allele
	Definition	name of the allele for the given gene
	Mandatory value format	free text

Example	<INSDQualifier_value>adh1-1</INSDQualifier_value>
Comment	all gene-related features (exon, CDS etc) for a given gene should share the same allele qualifier value; the allele qualifier value must, by definition, be different from the gene qualifier value; when used with the variation feature key, the allele qualifier value should be that of the variant.
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	<INSDQualifier_value>(pos:34..36,aa:Phe,seq:aaa)</INSDQualifier_value> <INSDQualifier_value>(pos:join(5,495..496),aa:Leu,seq:taa)</INSDQualifier_value> > <INSDQualifier_value>(pos:complement(4156..4158),aa:Glu,seq:ttg)</INSDQualifier_value>
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.
6.4. Qualifier	cell_line
Definition	cell line from which the sequence was obtained
Mandatory value format	free text
Example	<INSDQualifier_value>MCF7</INSDQualifier_value>
6.5. Qualifier	cell_type
Definition	cell type from which the sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>leukocyte</INSDQualifier_value>
6.6. Qualifier	chromosome
Definition	chromosome (e.g., Chromosome number) from which the sequence was obtained
Mandatory value format	free text
Example	<INSDQualifier_value>1</INSDQualifier_value>

<INSDQualifier_value>X</INSDQualifier_value>

6.7. Qualifier circular_RNA

Definition indicates that exons are out-of-order or overlapping because this spliced RNA product is a circular RNA (circRNA) created by backsplicing, for example when a downstream exon in the gene is located 5' of an upstream exon in the RNA product

Value format none

Comment should be used on features such as CDS, mRNA, tRNA and other features that are produced as a result of a backsplicing event. This qualifier should be used only when the splice event is indicated in the "join" operator, eg `join(complement(69611..69724),139856..140087)`

6.8. Qualifier clone

Definition clone from which the sequence was obtained

Mandatory value format free text
 Language-dependent: this value may require translation for National/Regional procedures

Example <INSDQualifier_value>lambda-hIL7.3</INSDQualifier_value>

Comment a source feature must not contain more than one clone qualifier; where the sequence was obtained from multiple clones it may be further described in the feature table using the feature key `misc_feature` and a note qualifier to specify the multiple clones.

6.9. Qualifier clone_lib

Definition clone library from which the sequence was obtained

Mandatory value format free text
 Language-dependent: this value may require translation for National/Regional procedures

Example <INSDQualifier_value>lambda-hIL7</INSDQualifier_value>

6.10. 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>

6.11. Qualifier collected_by

Definition name of persons or institute who collected the specimen

Mandatory value format free text
 Language-dependent: this value may require translation for National/Regional procedures

Example <INSDQualifier_value>Dan Janzen</INSDQualifier_value>

6.12. Qualifier	collection_date
Definition	date that the specimen was collected.
Mandatory value format	YYYY-MM-DD, YYYY-MM or YYYY
Example	<INSDQualifier_value>1952-10-21</INSDQualifier_value> <INSDQualifier_value>1952-10</INSDQualifier_value> <INSDQualifier_value>1952</INSDQualifier_value>
Comment	'YYYY' is a four-digit value representing the year. 'MM' is a two-digit value representing the month. 'DD' is a two-digit value representing the day of the month.
6.13. Qualifier	compare
Definition	Reference details of an existing public INSD entry to which a comparison is made
Mandatory value format	[accession-number.sequence-version]
Example	<INSDQualifier_value>AJ634337.1</INSDQualifier_value>
Comment	This qualifier may be used on the following features: misc_difference, unsure, and variation. Multiple compare qualifiers with different contents are allowed within a single feature. This qualifier is not intended for large-scale annotation of variations, such as SNPs.
6.14. Qualifier	cultivar
Definition	cultivar (cultivated variety) of plant from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Nipponbare</INSDQualifier_value> <INSDQualifier_value>Tenuifolius</INSDQualifier_value> <INSDQualifier_value>Candy Cane</INSDQualifier_value> <INSDQualifier_value>IR36</INSDQualifier_value>
Comment	'cultivar' is applied solely to products of artificial selection; use the variety qualifier for natural, named plant and fungal varieties.
6.15. Qualifier	dev_stage
Definition	if the sequence was obtained from an organism in a specific developmental stage, it is specified with this qualifier
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>fourth instar larva</INSDQualifier_value>
6.16. Qualifier	direction
Definition	direction of DNA replication
Mandatory value format	left, right, or both where left indicates toward the 5' end of the sequence (as presented) and right indicates toward the 3' end

Example	<INSDQualifier_value>left</INSDQualifier_value>
Comment	The values left, right, and both are permitted when the direction qualifier is used to annotate a rep_origin feature key. However, only left and right values are permitted when the direction qualifier is used to annotate an oriT feature key.
<hr/>	
6.17. Qualifier	EC_number
Definition	Enzyme Commission number for enzyme product of sequence
Mandatory value format	free text
Example	<INSDQualifier_value>1.1.2.4</INSDQualifier_value> <INSDQualifier_value>1.1.2.-</INSDQualifier_value> <INSDQualifier_value>1.1.2.n</INSDQualifier_value> <INSDQualifier_value>1.1.2.n1</INSDQualifier_value>
Comment	valid values for EC numbers are defined in the list prepared by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) (published in Enzyme Nomenclature 1992, Academic Press, San Diego, or a more recent revision thereof).The format represents a string of four numbers separated by full stops; up to three numbers starting from the end of the string may be replaced by dash "-" to indicate uncertain assignment. Symbols including an "n", e.g., "n", "n1" and so on, may be used in the last position instead of a number where the EC number is awaiting assignment. Please note that such incomplete EC numbers are not approved by NC-IUBMB.
<hr/>	
6.18. Qualifier	ecotype
Definition	a population within a given species displaying genetically based, phenotypic traits that reflect adaptation to a local habitat
Mandatory value Format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Columbia</INSDQualifier_value>
Comment	an example of such a population is one that has adapted hairier than normal leaves as a response to an especially sunny habitat. 'Ecotype' is often applied to standard genetic stocks of Arabidopsis thaliana, but it can be applied to any sessile organism.
<hr/>	
6.19. Qualifier	environmental_sample
Definition	identifies sequences derived by direct molecular isolation from a bulk environmental DNA sample (by PCR with or without subsequent cloning of the product, DGGE, or other anonymous methods) with no reliable identification of the source organism. Environmental samples include clinical samples, gut contents, and other sequences from anonymous organisms that may be associated with a particular host. They do not include endosymbionts that can be reliably recovered from a particular host, organisms from a readily identifiable but uncultured field sample (e.g., many cyanobacteria), or phytoplasmas that can be reliably recovered from diseased plants (even though these cannot be grown in axenic culture)
Value format	none
Comment	used only with the source feature key; source feature keys containing the environmental_sample qualifier should also contain the isolation_source qualifier; a source feature including the environmental_sample qualifier must not include the strain qualifier.

6.20.	Qualifier	exception
	Definition	indicates that the coding region cannot be translated using standard biological rules
	Mandatory value format	One of the following controlled vocabulary phrases: RNA editing rearrangement required for product annotated by transcript or proteomic data
	Example	<INSDQualifier_value>RNA editing</INSDQualifier_value> <INSDQualifier_value>rearrangement required for product</INSDQualifier_value>
	Comment	only to be used to describe biological mechanisms such as RNA editing; protein translation of a CDS with an exception qualifier will be different from the corresponding conceptual translation; must not be used where transl_except qualifier would be adequate, e.g., in case of stop codon completion use.
6.21.	Qualifier	frequency
	Definition	frequency of the occurrence of a feature
	Mandatory value format	free text representing the proportion of a population carrying the feature expressed as a fraction
	Example	<INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value>
6.22.	Qualifier	function
	Definition	function attributed to a sequence
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<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.
6.23.	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.
6.24.	Qualifier	gene_synonym
	Definition	synonymous, replaced, obsolete or former gene symbol
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional

procedures

Example <INSDQualifier_value>Hox-3.3</INSDQualifier_value>
 in a feature where the gene qualifier value is Hoxc6

Comment used where it is helpful to indicate a gene symbol synonym; when the gene_synonym qualifier is used, a primary gene symbol must always be indicated in a gene qualifier

6.25. Qualifier germline

Definition the sequence presented has not undergone somatic rearrangement as part of an adaptive immune response; it is the unrearranged sequence that was inherited from the parental germline

Value format none

Comment germline qualifier must not be used to indicate that the source of the sequence is a gamete or germ cell; germline and rearranged qualifiers must not be used in the same source feature; germline and rearranged qualifiers must only be used for molecules that can undergo somatic rearrangements as part of an adaptive immune response; these are the T-cell receptor (TCR) and immunoglobulin loci in the jawed vertebrates, and the unrelated variable lymphocyte receptor (VLR) locus in the jawless fish (lampreys and hagfish); germline and rearranged qualifiers should not be used outside of the Craniata (taxid=89593)

6.26. Qualifier haplogroup

Definition name for a group of similar haplotypes that share some sequence variation. Haplogroups are often used to track migration of population groups.

Mandatory value format free text
 Language-dependent: this value may require translation for National/Regional procedures

Example <INSDQualifier_value>H*</INSDQualifier_value>

6.27. Qualifier haplotype

Definition name for a specific set of alleles that are linked together on the same physical chromosome. In the absence of recombination, each haplotype is inherited as a unit, and may be used to track gene flow in populations.

Mandatory value format free text

Example <INSDQualifier_value>Dw3 B5 Cw1 A1</INSDQualifier_value>

6.28. Qualifier	host
Definition	natural (as opposed to laboratory) host to the organism from which sequenced molecule was obtained
Mandatory value format	free textLanguage-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Homo sapiens</INSDQualifier_value> <INSDQualifier_value>Homo sapiens 12 year old girl</INSDQualifier_value> <INSDQualifier_value>Rhizobium NGR234</INSDQualifier_value>
6.29. Qualifier	identified_by
Definition	name of the expert who identified the specimen taxonomically
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>John Burns</INSDQualifier_value>
6.30. Qualifier	isolate
Definition	individual isolate from which the sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Patient #152</INSDQualifier_value> <INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>
6.31. Qualifier	isolation_source
Definition	describes the physical, environmental and/or local geographical source of the biological sample from which the sequence was derived
Mandatory-value format	free textLanguage-dependent: this value may require translation for National/Regional procedures
Examples	<INSDQualifier_value>rumen isolates from standard Pelleted ration-fed steer #67</INSDQualifier_value> <INSDQualifier_value>permanent Antarctic sea ice</INSDQualifier_value> <INSDQualifier_value>denitrifying activated sludge from carbon_limited continuous reactor</INSDQualifier_value>
Comment	used only with the source feature key; source feature keys containing an environmental_sample qualifier should also contain an isolation_source qualifier

6.32.	Qualifier	lab_host
	Definition	scientific name of the laboratory host used to propagate the source organism from which the sequenced molecule was obtained
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>Gallus gallus</INSDQualifier_value> <INSDQualifier_value>Gallus gallus embryo</INSDQualifier_value> <INSDQualifier_value>Escherichia coli strain DH5 alpha</INSDQualifier_value> <INSDQualifier_value>Homo sapiens HeLa cells</INSDQualifier_value>
	Comment	the full binomial scientific name of the host organism should be used when known; extra conditional information relating to the host may also be included

6.33.	Qualifier	lat_lon
	Definition	geographical coordinates of the location where the specimen was collected
	Mandatory-value format	free text - degrees latitude and longitude in format "d[d.ddd] N S d[dd.ddd] W E"
	Example	<INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value> <INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value>

6.34.	Qualifier	macronuclear
	Definition	if the sequence shown is DNA and from an organism which undergoes chromosomal differentiation between macronuclear and micronuclear stages, this qualifier is used to denote that the sequence is from macronuclear DNA
	Value format	none

6.35.	Qualifier	map
	Definition	genomic map position of feature
	Mandatory value format	free text
	Example	<INSDQualifier_value>8q12-q13</INSDQualifier_value>

6.36. Qualifier	mating_type
Definition	mating type of the organism from which the sequence was obtained; mating type is used for prokaryotes, and for eukaryotes that undergo meiosis without sexually dimorphic gametes
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Examples	<INSDQualifier_value>MAT-1</INSDQualifier_value> <INSDQualifier_value>plus</INSDQualifier_value> <INSDQualifier_value>-</INSDQualifier_value> <INSDQualifier_value>odd</INSDQualifier_value> <INSDQualifier_value>even</INSDQualifier_value>"
Comment	mating_type qualifier values male and female are valid in the prokaryotes, but not in the eukaryotes; for more information, see the entry for the sex qualifier.
6.37. Qualifier	mobile_element_type
Definition	type and name or identifier of the mobile element which is described by the parent feature
Mandatory value format	<mobile_element_type>[:<mobile_element_name>] where <mobile_element_type> is one of the following: transposon retrotransposon integron insertion sequence non-LTR retrotransposon SINE MITE LINE other
Example	<INSDQualifier_value>transposon:Tnp9</INSDQualifier_value>
Comment	mobile_element_type is permitted on mobile_element feature key only. Mobile element should be used to represent both elements which are currently mobile, and those which were mobile in the past. Value "other" for <mobile_element_type> requires a <mobile_element_name>
6.38. Qualifier	mod_base
Definition	abbreviation for a modified nucleotide base
Mandatory value format	modified base abbreviation chosen from this Annex, Section 2
Example	<INSDQualifier_value>m5c</INSDQualifier_value> <INSDQualifier_value>OTHER</INSDQualifier_value>
Comment	specific modified nucleotides not found in Section 2 of this Annex are annotated by entering OTHER as the value for the mod_base qualifier and including a note qualifier with the full name of the modified base as its value

6.39.	Qualifier	mol_type
	Definition	molecule type of sequence
	Mandatory value format	One chosen from the following: genomic DNA genomic RNA mRNA tRNA rRNA other RNA other DNA transcribed RNA viral CRNA unassigned DNA unassigned RNA
	Example	<INSDQualifier_value>genomic DNA</INSDQualifier_value> <INSDQualifier_value>other RNA</INSDQualifier_value>
	Comment	mol_type qualifier is mandatory on the source feature key; the value "genomic DNA" does not imply that the molecule is nuclear (e.g., organelle and plasmid DNA must be described using "genomic DNA"); ribosomal RNA genes must be described using "genomic DNA"; "rRNA" must only be used if the ribosomal RNA molecule itself has been sequenced; values "other RNA" and "other DNA" must be applied to synthetic molecules, values "unassigned DNA", "unassigned RNA" must be applied where in vivo molecule is unknown.
6.40.	Qualifier	ncRNA_class
	Definition	a structured description of the classification of the non-coding RNA described by the ncrRNA parent key
	Mandatory value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: antisense_RNA autocatalytically_spliced_intron circRNA ribozyme hammerhead_ribozyme lncRNA RNase_P_RNA RNase_MRP_RNA telomerase_RNA guide_RNA sgRNA rasiRNA scrRNA scaRNA sirRNA pre_mirRNA mirRNA pirRNA snRNA snRNA SRP_RNA vault_RNA Y_RNA other
	Example	<INSDQualifier_value>autocatalytically_spliced_intron </INSDQualifier_value> <INSDQualifier_value>sirRNA</INSDQualifier_value> <INSDQualifier_value>scrRNA</INSDQualifier_value> <INSDQualifier_value>other</INSDQualifier_value>
	Comment	specific ncrRNA types not yet in the ncrRNA_class controlled vocabulary must be annotated by entering "other" as the ncrRNA_class qualifier value, and providing a brief explanation of novel ncrRNA_class in a note qualifier

6.41. Qualifier	note
Definition	any comment or additional information
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>A comment about the feature</INSDQualifier_value>
6.42. Qualifier	number
Definition	a number to indicate the order of genetic elements (e.g., exons or introns) in the 5' to 3' direction
Mandatory value format	free text (with no whitespace characters)
Example	<INSDQualifier_value>4</INSDQualifier_value> <INSDQualifier_value>6B</INSDQualifier_value>
Comment	text limited to integers, letters or combination of integers and/or letters represented as a data value that contains no whitespace characters; any additional terms should be included in a standard_name qualifier. Example: a number qualifier with a value of 2A and a standard_name qualifier with a value of "long"
6.43. Qualifier	operon
Definition	name of the group of contiguous genes transcribed into a single transcript to which that feature belongs
Mandatory value format	free text
Example	<INSDQualifier_value>lac</INSDQualifier_value>
6.44. Qualifier	organelle
Definition	type of membrane-bound intracellular structure from which the sequence was obtained
Mandatory value format	One of the following controlled vocabulary terms and phrases: chromatophore hydrogenosome mitochondrion nucleomorph plastid mitochondrion:kinetoplast plastid:chloroplast plastid:apicoplast plastid:chromoplast plastid:cyanelle plastid:leucoplast plastid:proplastid
Examples	<INSDQualifier_value>chromatophore</INSDQualifier_value> <INSDQualifier_value>hydrogenosome</INSDQualifier_value> <INSDQualifier_value>mitochondrion</INSDQualifier_value> <INSDQualifier_value>nucleomorph</INSDQualifier_value> <INSDQualifier_value>plastid</INSDQualifier_value> <INSDQualifier_value>mitochondrion:kinetoplast</INSDQualifier_value> <INSDQualifier_value>plastid:chloroplast</INSDQualifier_value> <INSDQualifier_value>plastid:apicoplast</INSDQualifier_value> <INSDQualifier_value>plastid:chromoplast</INSDQualifier_value>

6.49.	Qualifier	pop_variant
	Definition	name of subpopulation or phenotype of the sample from which the sequence was derived
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>pop1</INSDQualifier_value> <INSDQualifier_value>Bear Paw</INSDQualifier_value>
6.50.	Qualifier	product
	Definition	name of the product associated with the feature, e.g., the mRNA of an mRNA feature, the polypeptide of a CDS, the mature peptide of a mat_peptide, etc.
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>trypsinogen</INSDQualifier_value> (when qualifier appears in CDS feature) <INSDQualifier_value>trypsin</INSDQualifier_value> (when qualifier appears in mat_peptide feature) <INSDQualifier_value>XYZ neural-specific transcript</INSDQualifier_value> (when qualifier appears in mRNA feature)
6.51.	Qualifier	protein_id
	Definition	protein sequence identification number, an integer used in a sequence listing to designate the protein sequence encoded by the coding sequence identified in the corresponding CDS feature key and translation qualifier
	Mandatory value format	an integer greater than zero
	Example	<INSDQualifier_value>89</INSDQualifier_value>
6.52.	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
6.53.	Qualifier	pseudo
	Definition	indicates that this feature is a non-functional version of the element named by the feature key
	Value format	none
	Comment	The qualifier pseudo should be used to describe non-functional genes that are not formally described as pseudogenes, e.g., CDS has no translation due to other reasons than pseudogenization events. Other reasons may include sequencing or assembly errors. In order to annotate pseudogenes the qualifier pseudogene must be used, indicating the TYPE of pseudogene.

6.54. Qualifier	pseudogene
Definition	indicates that this feature is a pseudogene of the element named by the feature key
Mandatory value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: processed unprocessed unitary allelic unknown
Example	<INSDQualifier_value>processed</INSDQualifier_value> <INSDQualifier_value>unprocessed</INSDQualifier_value> <INSDQualifier_value>unitary</INSDQualifier_value> <INSDQualifier_value>allelic</INSDQualifier_value> <INSDQualifier_value>unknown</INSDQualifier_value>
Comment	Definitions of TYPE values: processed - the pseudogene has arisen by reverse transcription of a mRNA into cDNA, followed by reintegration into the genome. Therefore, it has lost any intron/exon structure, and it might have a pseudo-polyA-tail. unprocessed - the pseudogene has arisen from a copy of the parent gene by duplication followed by accumulation of random mutations. The changes, compared to their functional homolog, include insertions, deletions, premature stop codons, frameshifts and a higher proportion of non-synonymous versus synonymous substitutions. unitary - the pseudogene has no parent. It is the original gene, which is functional in some species but disrupted in some way (indels, mutation, recombination) in another species or strain. allelic - a (unitary) pseudogene that is stable in the population but importantly it has a functional alternative allele also in the population. i.e., one strain may have the gene, another strain may have the pseudogene. MHC haplotypes have allelic pseudogenes. unknown - the submitter does not know the method of pseudogenization.
6.55. Qualifier	rearranged
Definition	the sequence presented in the entry has undergone somatic rearrangement as part of an adaptive immune response; it is not the unrearranged sequence that was inherited from the parental germline
Value format	none
Comment	The rearranged qualifier must not be used to annotate chromosome rearrangements that are not involved in an adaptive immune response; germline and rearranged qualifiers must not be used in the same source feature; germline and rearranged qualifiers must only be used for molecules that can undergo somatic rearrangements as part of an adaptive immune response; these are the T-cell receptor (TCR) and immunoglobulin loci in the jawed vertebrates, and the unrelated variable lymphocyte receptor (VLR) locus in the jawless fish (lampreys and hagfish); germline and rearranged qualifiers should not be used outside of the Craniata (taxid=89593)
6.56. Qualifier	recombination_class
Definition	a structured description of the classification of recombination hotspot region within a sequence
Mandatory value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: meiotic mitotic non_allelic_homologous chromosome_breakpoint other

Example	<INSDQualifier_value>meiotic</INSDQualifier_value> <INSDQualifier_value>chromosome_breakpoint</INSDQualifier_value>
Comment	specific recombination classes not yet in the recombination_class controlled vocabulary must be annotated by entering "other" as the recombination_class qualifier value and providing a brief explanation of the novel recombination_class in a note qualifier
<hr/>	
6.57. Qualifier	regulatory_class
Definition	a structured description of the classification of transcriptional, translational, replicational and chromatin structure related regulatory elements in a sequence
Mandatory value format	<p>TYPE</p> <p>where TYPE is one of the following controlled vocabulary terms or phrases:</p> <p>attenuator CAAT_signal DNase_I_hypersensitive_site enhancer enhancer_blocking_element GC_signal imprinting_control_region insulator locus_control_region matrix_attachment_region minus_35_signal minus_10_signal polyA_signal_sequence promoter recoding_stimulatory_region <u>recombination_enhancer</u> replication_regulatory_region response_element ribosome_binding_site riboswitch silencer TATA_box terminator transcriptional_cis_regulatory_region <u>uORE</u> other</p>
Example	<INSDQualifier_value>promoter</INSDQualifier_value> <INSDQualifier_value>enhancer</INSDQualifier_value> <INSDQualifier_value>ribosome_binding_site</INSDQualifier_value>
Comment	specific regulatory classes not yet in the regulatory_class controlled vocabulary must be annotated by entering "other" as the regulatory_class qualifier value and providing a brief explanation of the novel regulatory_class in a note qualifier
<hr/>	
6.58. 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	<INSDQualifier_value>a</INSDQualifier_value> <INSDQualifier_value></INSDQualifier_value> - for a deletion

6.59.	Qualifier	ribosomal_slippage
	Definition	during protein translation, certain sequences can program ribosomes to change to an alternative reading frame by a mechanism known as ribosomal slippage
	Value format	none
	Comment	a join operator, e.g.,: [join(486..1784,1787..4810)] must be used in the CDS feature location to indicate the location of ribosomal_slippage
6.60.	Qualifier	rpt_family
	Definition	type of repeated sequence; "Alu" or "kpn", for example
	Mandatory value format	free text
	Example	<INSDQualifier_value>Alu</INSDQualifier_value>
6.61.	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; flanking - a repeat lying outside the sequence for which it has functional significance (eg. transposon insertion target sites); nested - a repeat that is disrupted by the insertion of another element; dispersed - a repeat that is found dispersed throughout the genome; terminal - a repeat at the ends of and within the sequence for which it has functional significance (eg. transposon LTRs); long_terminal_repeat - a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses; non_ltr_retrotransposon_polymeric_tract - a polymeric tract, such as poly(dA), within a non LTR retrotransposon; centromeric_repeat - a repeat region found within the modular centromere; telomeric_repeat - a repeat region found within the telomere; x_element_combinatorial_repeat - a repeat region located between the X element and the telomere or adjacent Y' element; y_prime_element - a repeat region located adjacent to telomeric repeats or X element combinatorial repeats, either as a single copy or tandem repeat of two to four copies; other - a repeat exhibiting important attributes that cannot be described by other values.

6.62.	Qualifier	rpt_unit_range
	Definition	location of a repeating unit expressed as a range
	Mandatory value format	<base_range> - where <base_range> is the first and last base (separated by two dots) of a repeating unit
	Example	<INSDQualifier_value>202..245</INSDQualifier_value>
	Comment	used to indicate the base range of the sequence that constitutes a repeating unit within the region specified by the feature keys oriT and repeat_region.
6.63.	Qualifier	rpt_unit_seq
	Definition	identity of a repeat sequence
	Mandatory value format	free text
	Example	<INSDQualifier_value>aagggc</INSDQualifier_value> <INSDQualifier_value>ag(5)tg(8)</INSDQualifier_value> <INSDQualifier_value>(AAAGA)6(AAAA)1(AAAGA)12</INSDQualifier_value>
	Comment	used to indicate the literal sequence that constitutes a repeating unit within the region specified by the feature keys oriT and repeat_region
6.64.	Qualifier	satellite
	Definition	identifier for a satellite DNA marker, composed of many tandem repeats (identical or related) of a short basic repeated unit
	Mandatory value format	<satellite_type>[:<class>][<identifier>] - where <satellite_type> is one of the following: satellite microsatellite minisatellite
	Example	<INSDQualifier_value>satellite: S1a</INSDQualifier_value> <INSDQualifier_value>satellite: alpha</INSDQualifier_value> <INSDQualifier_value>satellite: gamma III</INSDQualifier_value> <INSDQualifier_value>microsatellite: DC130</INSDQualifier_value>
	Comment	many satellites have base composition or other properties that differ from those of the rest of the genome that allows them to be identified.
6.65.	Qualifier	segment
	Definition	name of viral or phage segment sequenced
	Mandatory value format	free text
	Example	<INSDQualifier_value>6</INSDQualifier_value>
6.66.	Qualifier	serotype
	Definition	serological variety of a species characterized by its antigenic properties
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>B1</INSDQualifier_value>

Comment	used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for the prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms".
6.67. Qualifier	serovar
Definition	serological variety of a species (usually a prokaryote) characterized by its antigenic properties
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>0157:H7</INSDQualifier_value>
Comment	used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms".
6.68. Qualifier	sex
Definition	sex of the organism from which the sequence was obtained; sex is used for eukaryotic organisms that undergo meiosis and have sexually dimorphic gametes
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Examples	<INSDQualifier_value>female</INSDQualifier_value> <INSDQualifier_value>male</INSDQualifier_value> <INSDQualifier_value>hermaphrodite</INSDQualifier_value> <INSDQualifier_value>unisexual</INSDQualifier_value> <INSDQualifier_value>bisexual</INSDQualifier_value> <INSDQualifier_value>asexual</INSDQualifier_value> <INSDQualifier_value>monoecious</INSDQualifier_value> [or monocious] <INSDQualifier_value>dioecious</INSDQualifier_value> [or diecious]
Comment	The sex qualifier should be used (instead of mating_type qualifier) in the Metazoa, Embryophyta, Rhodophyta & Phaeophyceae; mating_type qualifier should be used (instead of sex qualifier) in the Bacteria, Archaea & Fungi; neither sex nor mating_type qualifiers should be used in the viruses; outside of the taxa listed above, mating_type qualifier should be used unless the value of the qualifier is taken from the vocabulary given in the examples above
6.69. Qualifier	standard_name
Definition	accepted standard name for this feature
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>dotted</INSDQualifier_value>
Comment	use standard_name qualifier to give full gene name, but use gene qualifier to give gene symbol (in the above example gene qualifier value is Dt).
6.70. Qualifier	strain
Definition	strain from which sequence was obtained

Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>BALB/c</INSDQualifier_value>
Comment	feature entries including a strain qualifier must not include the environmental_sample qualifier
<hr/>	
6.71. Qualifier	sub_clone
Definition	sub-clone from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>lambda-hIL7.20g</INSDQualifier_value>
Comment	a source feature must not contain more than one sub_clone qualifier; to indicate that the sequence was obtained from multiple sub_clones, multiple sources may be further described using the feature key "misc_feature" and the qualifier "note"
<hr/>	
6.72. Qualifier	sub_species
Definition	name of sub-species of organism from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>lactis</INSDQualifier_value>
<hr/>	
6.73. Qualifier	sub_strain
Definition	name or identifier of a genetically or otherwise modified strain from which sequence was obtained, derived from a parental strain (which should be annotated in the strain qualifier). sub_strain from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>abis</INSDQualifier_value>
Comment	must be accompanied by a strain qualifier in a source feature; if the parental strain is not given, the modified strain should be annotated in the strain qualifier instead of sub_strain. For example, either a strain qualifier with the value K-12 and a substrain qualifier with the value MG1655 or a strain qualifier with the value MG1655
<hr/>	
6.74. Qualifier	tag_peptide
Definition	base location encoding the polypeptide for proteolysis tag of tmRNA and its termination codon
Mandatory value format	<base_range> - where <base_range> provides the first and last base (separated by two dots) of the location for the proteolysis tag
Example	<INSDQualifier_value>90..122</INSDQualifier_value>
Comment	it is recommended that the amino acid sequence corresponding to the tag_peptide

be annotated by describing a 5' partial CDS feature; e.g., CDS with a location of <90..122

6.75. Qualifier	tissue_lib
Definition	tissue library from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>tissue library 772</INSDQualifier_value>
6.76. Qualifier	tissue_type
Definition	tissue type from which the sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>liver</INSDQualifier_value>
6.77. Qualifier	transl_except
Definition	translational exception: single codon the translation of which does not conform to genetic code defined by organism or transl_table.
Mandatory value format	(pos:<location>,aa:<amino_acid>) where <amino_acid> is the three letter abbreviation for the amino acid coded by the codon at the base_range position
Example	<INSDQualifier_value>(pos:213..215,aa:Trp) </INSDQualifier_value> <INSDQualifier_value>(pos:462..464,aa:OTHER) </INSDQualifier_value> <INSDQualifier_value>(pos:1017,aa:TERM) </INSDQualifier_value> <INSDQualifier_value>(pos:2000..2001,aa:TERM) </INSDQualifier_value>
Comment	if the amino acid is not one of the specific amino acids listed in Section 3 of this Annex, use OTHER as <amino_acid> and provide the name of the unusual amino acid in a note qualifier; for modified amino-acid selenocysteine use three letter abbreviation 'Sec' (one letter symbol 'U' in amino-acid sequence) for <amino_acid>; for modified amino-acid pyrrolysine use three letter abbreviation 'Pyl' (one letter symbol 'O' in amino-acid sequence) for <amino_acid>; for partial termination codons where TAA stop codon is completed by the addition of 3' A residues to the mRNA either a single base_position or a base_range is used for the location, see the third and fourth examples above, in conjunction with a note qualifier indicating 'stop codon completed by the addition of 3' A residues to the mRNA'.
6.78. Qualifier	transl_table
Definition	definition of genetic code table used if other than universal or standard genetic code table. Tables used are described in this Annex
Mandatory value format	<integer> where <integer> is the number assigned to the genetic code table
Example	<INSDQualifier_value>3</INSDQualifier_value> - example where the yeast mitochondrial code is to be used
Comment	if the transl_table qualifier is not used to further annotate a CDS feature key, then the CDS is translated using the Standard Code (i.e. Universal Genetic Code). Genetic code exceptions outside the range of specified tables are reported in transl_except qualifiers.

6.79.	Qualifier	trans_splicing
	Definition	indicates that exons from two RNA molecules are ligated in intermolecular reaction to form mature RNA
	Value format	none
	Comment	should be used on features such as CDS, mRNA and other features that are produced as a result of a trans-splicing event. This qualifier must be used only when the splice event is indicated in the "join" operator, e.g., join(complement(69611..69724),139856..140087) in the feature location

6.80.	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>MASTFPPWYRGCASTPSLKGIMCTW</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.

6.81.	Qualifier	variety
	Definition	variety (= varietas, a formal Linnaean rank) of organism from which sequence was derived.
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>insularis</INSDQualifier_value>
	Comment	use the cultivar qualifier for cultivated plant varieties, i.e., products of artificial selection; varieties other than plant and fungal variatas should be annotated via a note qualifier, e.g., with the value <INSDQualifier_value>breed:Cukorova</INSDQualifier_value>

SECCIÓN 7: CLAVES DE CARACTERIZACIÓN PARA SECUENCIAS DE AMINOÁCIDOS

Esta sección contiene la lista de claves de caracterización permitidas que deberán utilizarse para las secuencias de aminoácidos. Las claves de caracterización se enumeran por orden alfabético.

7.1.	Feature Key	ACT_SITE
	Definition	Amino acid(s) involved in the activity of an enzyme
	Optional qualifiers	NOTE note
	Comment	Each amino acid residue of the active site must be annotated separately with the ACT_SITE feature key. The corresponding amino acid residue number must be provided as the location descriptor in the feature location element.
7.2.	Feature Key	BINDING
	Definition	Binding site for any chemical group (co-enzyme, prosthetic group, etc.). The chemical nature of the group is indicated in the NOTE note qualifier
	Mandatory qualifiers	NOTE note
	Comment	Examples of values for the " NOTE note" qualifier: "Heme (covalent)" and "Chloride." Where appropriate, the features keys CA_BIND, DNA_BIND, METAL, and NP_BIND should be used rather than BINDING.
7.3.	Feature Key	CA_BIND
	Definition	Extent of a calcium-binding region
	Optional qualifiers	NOTE note
7.4.	Feature Key	CARBOHYD
	Definition	Glycosylation site
	Mandatory qualifiers	NOTE note
	Comment	This key describes the occurrence of the attachment of a glycan (mono- or polysaccharide) to a residue of the protein. The type of linkage (C-, N- or O-linked) to the protein is indicated in the " NOTE note" qualifier. If the nature of the reducing terminal sugar is known, its abbreviation is shown between parentheses. If three dots '...' follow the abbreviation this indicates an extension of the carbohydrate chain. Conversely no dots means that a monosaccharide is linked. Examples of values used in the " NOTE note" qualifier: N-linked (GlcNAC...); O-linked (GlcNAC); O-linked (Glc...); C-linked (Man) partial; O-linked (Ara...).
7.5.	Feature Key	CHAIN
	Definition	Extent of a polypeptide chain in the mature protein
	Optional qualifiers	NOTE note

7.6.	Feature Key	COILED
	Definition	Extent of a coiled-coil region
	Optional qualifiers	NOTE note
7.7.	Feature Key	COMPBIAS
	Definition	Extent of a compositionally biased region
	Optional qualifiers	NOTE note
7.8.	Feature Key	CONFLICT
	Definition	Different sources report differing sequences
	Optional qualifiers	NOTE note
	Comment	Examples of values for the " NOTE note" qualifier: Missing; K -> Q; GSDSE -> RIRLR; V -> A.
7.9.	Feature Key	CROSSLNK
	Definition	Post translationally formed amino acid bonds
	Mandatory qualifiers	NOTE note
	Comment	Covalent linkages of various types formed between two proteins (interchain cross-links) or between two parts of the same protein (intrachain cross-links); except for cross-links formed by disulfide bonds, for which the "DISULFID" feature key is to be used. For an interchain cross-link, the location descriptor in the feature location element is the residue number of the amino acid cross-linked to the other protein. For an intrachain cross-link, the location descriptor in the feature location element is the residue numbers of the cross-linked amino acids in "x..x" format, e.g. "42.. 50)". The NOTE note qualifier indicates the nature of the cross-link; at least specifying the name of the conjugate and the identity of the two amino acids involved. Examples of values for the " NOTE note" qualifier: "Isoglutamyl cysteine thioester (Cys-Gln);" "Beta-methylanthionine (Cys-Thr);" and "Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin)"
7.10.	Feature Key	DISULFID
	Definition	Disulfide bond
	Mandatory qualifiers	NOTE note
	Comment	For an interchain disulfide bond, the location descriptor in the feature location element is the residue number of the cysteine linked to the other protein. For an intrachain cross-link, the location descriptor in the feature location element is the residue numbers of the linked cysteines in "x..x" format, e.g. "42.. 50". For interchain disulfide bonds, the NOTE note qualifier indicates the nature of the cross-link, by identifying the other protein, for example, "Interchain (between A and B chains)"
7.11.	Feature Key	DNA_BIND
	Definition	Extent of a DNA-binding region
	Mandatory qualifiers	NOTE note
	Comment	The nature of the DNA-binding region is given in the NOTE note qualifier.

Examples of values for the "NOTEnote" qualifier: "Homeobox" and "Myb 2"

7.12.	Feature Key	DOMAIN
	Definition	Extent of a domain, which is defined as a specific combination of secondary structures organized into a characteristic three-dimensional structure or fold
	Mandatory qualifiers	NOTEnote
	Comment	The domain type is given in the NOTEnote qualifier. Where several copies of a domain are present, the domains are numbered. Examples of values for the "NOTEnote" qualifier: "Ras-GAP" and "Cadherin 1"
7.13.	Feature Key	HELIX
	Definition	Secondary structure: Helices, for example, Alpha-helix; 3(10) helix; or Pi-helix
	Optional qualifiers	NOTEnote
	Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a 'loop' or 'random-coil' structure.
7.14.	Feature Key	INIT_MET
	Definition	Initiator methionine
	Optional qualifiers	NOTEnote
	Comment	The location descriptor in the feature location element is "1". This feature key indicates the N-terminal methionine is cleaved off. This feature is not used when the initiator methionine is not cleaved off.
7.15.	Feature Key	INTRAMEM
	Definition	Extent of a region located in a membrane without crossing it
	Optional qualifiers	NOTEnote

7.16.	Feature Key	LIPID
	Definition	Covalent binding of a lipid moiety
	Mandatory qualifiers	NOTE note
	Comment	The chemical nature of the bound lipid moiety is given in the NOTE note qualifier, indicating at least the name of the lipidated amino acid. Examples of values for the " NOTE note" qualifier: "N-myristoyl glycine"; "GPI-anchor amidated serine" and "S-diacylglycerol cysteine."
7.17.	Feature Key	METAL
	Definition	Binding site for a metal ion.
	Mandatory qualifiers	NOTE note
	Comment	The NOTE note qualifier indicates the nature of the metal. Examples of values for the " NOTE note" qualifier: "Iron (heme axial ligand)" and "Copper".
7.18.	Feature Key	MOD_RES
	Definition	Posttranslational modification of a residue
	Mandatory qualifiers	NOTE note
	Comment	The chemical nature of the modified residue is given in the NOTE note qualifier, indicating at least the name of the post-translationally modified amino acid. If the modified amino acid is listed in Section 4 of this Annex, the abbreviation may be used in place of the the full name. Examples of values for the " NOTE note" qualifier: "N-acetylalanine"; "3-Hyp"; and "MeLys" or "N-6-methyllysine"
7.19.	Feature Key	MOTIF
	Definition	Short (up to 20 amino acids) sequence motif of biological interest
	Optional qualifiers	NOTE note
7.20.	Feature Key	MUTAGEN
	Definition	Site which has been experimentally altered by mutagenesis
	Optional qualifiers	NOTE note
7.21.	Feature Key	NON_STD
	Definition	Non-standard amino acid
	Optional qualifiers	NOTE note
	Comment	This key only describes the occurrence of non-standard amino acids selenocysteine (U) and pyrrolysine (O) in the amino acid sequence.
7.22.	Feature Key	NON_TER
	Definition	The residue at an extremity of the sequence is not the terminal residue
	Optional qualifiers	NOTE note

Comment	If applied to position 1, this means that the first position is not the N-terminus of the complete molecule. If applied to the last position, it means that this position is not the C-terminus of the complete molecule.
7.23. Feature Key	NP_BIND
Definition	Extent of a nucleotide phosphate-binding region
Mandatory qualifiers	NOTE note
Comment	The nature of the nucleotide phosphate is indicated in the NOTE note qualifier. Examples of values for the " NOTE note" qualifier: "ATP" and "FAD".
7.24. Feature Key	PEPTIDE
Definition	Extent of a released active peptide
Optional qualifiers	NOTE note
7.25. Feature Key	PROPEP
Definition	Extent of a propeptide
Optional qualifiers	NOTE note
7.26. Feature Key	REGION
Definition	Extent of a region of interest in the sequence
Optional qualifiers	NOTE note
7.27. Feature Key	REPEAT
Definition	Extent of an internal sequence repetition
Optional qualifiers	NOTE note
7.28. Feature Key	SIGNAL
Definition	Extent of a signal sequence (prepeptide)
Optional qualifiers	NOTE note
7.29. Feature Key	SITE
Definition	Any interesting single amino-acid site on the sequence that is not defined by another feature key. It can also apply to an amino acid bond which is represented by the positions of the two flanking amino acids
Mandatory qualifier	NOTE note
Comment	When SITE is used to annotate a modified amino acid the value for the qualifier " NOTE note" must either be an abbreviation set forth in Section 4 of this Annex, or the complete, unabbreviated name of the modified amino acid.

7.30.	Feature Key	SOURCE source
	Definition	Identifies the source of the sequence; this key is mandatory; every sequence will have a single SOURCE source feature spanning the entire sequence
	Mandatory qualifiers	MOL_TYPE mol_type ORGANISM organism
	Optional qualifiers	NOTE note
7.31.	Feature Key	STRAND
	Definition	Secondary structure: Beta-strand; for example Hydrogen bonded beta-strand or residue in an isolated beta-bridge
	Optional qualifiers	NOTE note
	Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a 'loop' or 'random-coil' structure.
7.32.	Feature Key	TOPO_DOM
	Definition	Topological domain
	Optional qualifiers	NOTE note
7.33.	Feature Key	TRANSMEM
	Definition	Extent of a transmembrane region
	Optional qualifiers	NOTE note
7.34.	Feature Key	TRANSIT
	Definition	Extent of a transit peptide (mitochondrion, chloroplast, thylakoid, cyanelle, peroxisome etc.)
	Optional qualifiers	NOTE note
7.35.	Feature Key	TURN
	Definition	Secondary structure Turns, for example, H-bonded turn (3-turn, 4-turn or 5-turn)
	Optional qualifiers	NOTE note
	Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a 'loop' or 'random-coil' structure.
7.36.	Feature Key	UNSURE
	Definition	Uncertainties in the sequence
	Optional qualifiers	NOTE note

Comment	Used to describe region(s) of an amino acid sequence for which the authors are unsure about the sequence presentation.
<hr/>	
7.37. Feature Key	VARIANT
Definition	Authors report that sequence variants exist
Optional qualifiers	NOTE note
<hr/>	
7.38. Feature Key	VAR_SEQ
Definition	Description of sequence variants produced by alternative splicing, alternative promoter usage, alternative initiation and ribosomal frameshifting
Optional qualifiers	NOTE note
<hr/>	
7.39. Feature Key	ZN_FING
Definition	Extent of a zinc finger region
Mandatory qualifiers	NOTE note
Comment	The type of zinc finger is indicated in the NOTE note qualifier. For example: "GATA-type" and "NR C4-type"

SECCIÓN 8: CALIFICADORES PARA SECUENCIAS DE AMINOÁCIDOS

Esta sección contiene la lista de calificadores permitidos que deberán utilizarse para las secuencias de aminoácidos.

Cuando el formato de valor sea texto libre identificado como dependiente del idioma, deberá utilizarse uno de los elementos siguientes:

- 1) el elemento `INSDQualifier_value; 0`
- 2) el elemento `NonEnglishQualifier_value; 0`
- 3) ambos elementos `INSDQualifier_value` y `NonEnglishQualifier_value`.

Cuando el formato de valor no esté identificado como texto libre dependiente del idioma, deberá utilizarse el elemento `INSDQualifier_value` y no deberá utilizarse el elemento `NonEnglishQualifier_value`.

NOTA: Todo valor calificador (qualifier value) que se indique para un calificador con formato de valor "free text" podrá requerir traducción para los procedimientos nacionales/regionales. Se considera que los calificadores enumerados en el siguiente cuadro tienen valores de texto libre dependiente del idioma.

Cuadro 6: Lista de calificadores para secuencias de aminoácidos con valores de texto libre dependiente del idioma

Sección	Valor de texto libre dependiente del idioma
8.2	<code>NOTE</code> <code>note</code>
8.3	<code>ORGANISM</code> <code>organism</code>

8.1.	Qualifier	<code>MOL_TYPE</code> <code>mol_type</code>
	Definition	In vivo molecule type of sequence
	Mandatory value format	protein
	Example	<INSDQualifier_value>protein</INSDQualifier_value>
	Comment	The " <code>MOL_TYPE</code> <code>mol_type</code> " qualifier is mandatory on the <code>SOURCE</code> <code>source</code> feature key.
8.2.	Qualifier	<code>NOTE</code> <code>note</code>
	Definition	Any comment or additional information
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>Heme (covalent)</INSDQualifier_value>
	Comment	The " <code>NOTE</code> <code>note</code> " qualifier is mandatory for the feature keys: BINDING; CARBOHYD; CROSSLNK; DISULFID; DNA_BIND; DOMAIN; LIPID; METAL; MOD_RES; NP_BIND; <code>SITE</code> and ZN_FING
8.3.	Qualifier	<code>ORGANISM</code> <code>organism</code>
	Definition	Scientific name of the organism that provided the peptide
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
	Comment	The " <code>ORGANISM</code> <code>organism</code> " qualifier is mandatory for the <code>SOURCE</code> <code>source</code> feature key.

2018-10-19: Version 1.2 adopted at the CWS/6
2016-03-24: Version 1.0 adopted at the CWS/4Bis
2014-03-11: Final draft for adoption.

ST26SequenceListing

* ROOT ELEMENT

-->

<!ELEMENT ST26SequenceListing ((ApplicantFileReference | (ApplicationIdentification, ApplicantFileReference?)), EarliestPriorityApplicationIdentification?, (ApplicantName, ApplicantNameLatin?), (InventorName, InventorNameLatin?)?, InventionTitle+, SequenceTotalQuantity, SequenceData+)>

<!--The elements ApplicantName and InventorName are optional in this DTD to facilitate the conversion between various encoding schemes-->

<!--originalFreeTextLanguageCode:

The language code (see reference in paragraph 9 to ISO 639-1:2002) for the single original language in which the language-dependent free text qualifiers (NonEnglishQualifier_value) were prepared.

-->

<!--nonEnglishFreeTextLanguageCode:

The language code (see reference in paragraph 9 to ISO 639-1:2002) for the language in which the language-dependent free text qualifiers (NonEnglishQualifier_value) currently correspond.

-->

<!ATTLIST ST26SequenceListing

 dtdVersion CDATA #REQUIRED

 fileName CDATA #IMPLIED

 softwareName CDATA #IMPLIED

 softwareVersion CDATA #IMPLIED

 productionDate CDATA #IMPLIED>

> originalFreeTextLanguageCode CDATA #IMPLIED

 nonEnglishFreeTextLanguageCode CDATA #IMPLIED

>

<!--ApplicantFileReference

Applicant's or agent's file reference, mandatory if application identification not provided.

-->

<!ELEMENT ApplicantFileReference (#PCDATA)>

<!--ApplicationIdentification

Application identification for which the sequence listing is submitted, when available.

-->

<!ELEMENT ApplicationIdentification (IPOfficeCode, ApplicationNumberText, FilingDate?)>

<!--EarliestPriorityApplicationIdentification

Identification of the earliest priority application, which contains IPOfficeCode, ApplicationNumberText and FilingDate elements.

-->

<!ELEMENT EarliestPriorityApplicationIdentification (IPOfficeCode, ApplicationNumberText, FilingDate?)>

<!--ApplicantName

The name of the first mentioned applicant in characters set forth in paragraph 40 (a) of the ST.26 main body document.

-->

<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of names of languages - Part 1: Alpha-2

-->

<!ELEMENT ApplicantName (#PCDATA)>

<!ATTLIST ApplicantName

 languageCode CDATA #REQUIRED >

<!--ApplicantNameLatin

Where ApplicantName is typed in characters other than those as set forth in paragraph 40 (b), a translation or transliteration of the name of the first mentioned applicant must

also be typed in characters as set forth in paragraph 40 (b) of the ST.26 main body document.

-->

<!ELEMENT ApplicantNameLatin (#PCDATA)>

<!--InventorName
Name of the first mentioned inventor typed in the characters as set forth in paragraph 40 (a).-->

<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of names of languages - Part 1: Alpha-2
-->

<!ELEMENT InventorName (#PCDATA)>

<!ATTLIST InventorName
 languageCode CDATA #REQUIRED >

<!--InventorNameLatin
Where InventorName is typed in characters other than those as set forth in paragraph 40 (b), a translation or transliteration of the first mentioned inventor may also be typed in characters as set forth in paragraph 40 (b).
-->

<!ELEMENT InventorNameLatin (#PCDATA)>

<!--InventionTitle
Title of the invention typed in the characters as set forth in paragraph 40 (a) in the language of filing. A translation of the title of the invention into additional languages may be typed in the characters as set forth in paragraph 40 (a) using additional InventionTitle elements. Preferably The title of invention should be between two to seven words.
-->

<!--languageCode: Appropriate language code from ISO 639-1 - Codes for the representation of names of languages - Part 1: Alpha-2
-->

<!ELEMENT InventionTitle (#PCDATA)>

<!ATTLIST InventionTitle
 languageCode CDATA #REQUIRED >

<!--SequenceTotalQuantity
Indicates the total number of sequences in the document.
Its purpose is to be quickly accessible for automatic processing.
-->

<!ELEMENT SequenceTotalQuantity (#PCDATA)>

<!--SequenceData
Data for individual Sequence.
For intentionally skipped sequences see the ST.26 main body document.
-->

<!ELEMENT SequenceData (INSDSeq)>

<!ATTLIST SequenceData
 sequenceIDNumber CDATA #REQUIRED >

<!--IPOfficeCode
ST.3 code. For example, if the application identification is PCT/IB2013/099999, then IPOfficeCode value will be "IB" for the International Bureau of WIPO.
-->

<!ELEMENT IPOfficeCode (#PCDATA)>

<!--ApplicationNumberText
The application identification as provided by the office of filing (e.g. PCT/IB2013/099999)
-->

<!ELEMENT ApplicationNumberText (#PCDATA)>

<!--FilingDate
The date of filing of the patent application for which the sequence listing is submitted in ST.2 format "CCYY-MM-DD", using a 4-digit calendar year, a 2-digit calendar month and a 2-digit day within the calendar month, e.g., 2015-01-31. For details, please see paragraphs 7 (a) and 11 of WIPO Standard ST.2.
-->

<!ELEMENT FilingDate (#PCDATA)>

<!--*****
* INSD Part

The purpose of the INSD part of this DTD is to define a customized DTD for sequence listings to support the work of IP offices while facilitating the data exchange with the public repositories.

The INSD part is subset of the INSD DTD v1.5 and as such can only be used to generate an XML instance as it will not support the complete INSD structure.

This part is based on:

The International Nucleotide Sequence Database (INSD) collaboration.

INSDSeq provides the elements of a sequence as presented in the GenBank/EMBL/DDBJ-style flatfile formats. Not all elements are used here.

-->

<!--INSDSeq

Sequence data. Changed INSD V1.5 DTD elements, INSDSeq_division and INSDSeq_sequence from optional to mandatory per business requirements.

-->

<!ELEMENT INSDSeq (INSDSeq_length, INSDSeq_moltype, INSDSeq_division, INSDSeq_other-seqids?, INSDSeq_feature-table?, INSDSeq_sequence)>

<!--INSDSeq_length

The length of the sequence. INSDSeq_length allows only integer.

-->

<!ELEMENT INSDSeq_length (#PCDATA)>

<!--INSDSeq_moltype

Admissible values: DNA, RNA, AA

-->

<!ELEMENT INSDSeq_moltype (#PCDATA)>

<!--INSDSeq_division

Indication that a sequence is related to a patent application. Must be populated with the value PAT.

-->

<!ELEMENT INSDSeq_division (#PCDATA)>

<!--INSDSeq_other-seqids

In the context of data exchange with database providers, the **Patent Offices IPOS** should populate for each sequence the element INSDSeq_other-seqids with one INSDSeqid containing a reference to the corresponding published patent and the sequence identification.

-->

<!ELEMENT INSDSeq_other-seqids (INSDSeqid?)>

<!--INSDSeq_feature-table

Information on the location and roles of various regions within a particular sequence. Whenever the element INSDSeq_feature-table is used, it must contain at least one feature.

-->

<!ELEMENT INSDSeq_feature-table (INSDFeature+)>

<!--INSDSeq_sequence

The residues of the sequence. The sequence must not contain numbers, punctuation or whitespace characters.

-->

<!ELEMENT INSDSeq_sequence (#PCDATA)>

<!--INSDSeqid

Intended for the use of **Patent Offices IPOS** in data exchange only.

Format:

pat|{office code}|{publication number}|{document kind code}|{Sequence identification number}

where office code is the code of the IP office publishing the patent document, publication number is the publication number of the application or patent, document kind code is the letter codes to distinguish patent documents as defined in ST.16 and Sequence identification number is the number of the sequence in that application or patent

Example:

```
pat|WO|2013999999|A1|123456
```

This represents the 123456th sequence from WO patent publication No. 2013999999 (A1)

```
-->
<!ELEMENT INSDSeqid (#PCDATA)>

<!--INSDFeature
Description of one feature.
-->
<!ELEMENT INSDFeature (INSDFeature_key, INSDFeature_location, INSDFeature_qual?)>

<!--INSDFeature_key
A word or abbreviation indicating a feature.
-->
<!ELEMENT INSDFeature_key (#PCDATA)>

<!--INSDFeature_location
Region of the presented sequence which corresponds to the feature.
-->
<!ELEMENT INSDFeature_location (#PCDATA)>

<!--INSDFeature_qual
List of qualifiers containing auxiliary information about a feature.
-->
<!ELEMENT INSDFeature_qual (INSDQualifier+)>

<!--INSDQualifier
Additional information about a feature.
For coding sequences and variants see the ST.26 main body document.
-->
<!--id
Unique identifier for the INSDQualifier to facilitate comparison of versions of a
sequence listing specifically having language-dependent qualifier values in different
languages.
-->
<!ELEMENT INSDQualifier (INSDQualifier_name, INSDQualifier_value?,
NonEnglishQualifier_value?)>
<!ATTLIST INSDQualifier
    id ID #IMPLIED
>

<!--INSDQualifier_name
Name of the qualifier.
-->
<!ELEMENT INSDQualifier_name (#PCDATA)>

<!--INSDQualifier_value
Value of the qualifier. Where the qualifier is language-dependent its value must be in
the English language and typed with the characters set forth in paragraph 40 (b).
-->
<!ELEMENT INSDQualifier_value (#PCDATA)>
<!--NonEnglishQualifier_value
Value of a language-dependent qualifier in a language that is not English and typed with
the characters set forth in paragraph 40 (a). The language is indicated with the
attribute nonEnglishFreeTextLanguageCode.
-->
<!ELEMENT NonEnglishQualifier_value (#PCDATA)>
```

[Sigue el Anexo III de la Norma ST.26]

ANEXO III

EJEMPLO DE LISTA DE SECUENCIAS (archivo XML)

Versión 1.3

Revisión aprobada Propuesta presentada por el Comité Equipo Técnico sobre Listas de Normas Técnicas de Secuencias para que el CWS la OMPI (CWS) examine y apruebe en su octava novena sesión, celebrada el 4 de diciembre de 2020

El Anexo III está disponible en: <https://www.wipo.int/standards/en/docs/st26-annex-iii-sequence-listing-specimen.xml>

[Sigue el Anexo V de la Norma ST.26]

Nota editorial: Se ha excluido el Anexo IV del presente documento porque no se propone ningún cambio en la novena sesión del CWS.

ANEXO V

REQUISITOS ADICIONALES SOBRE EL INTERCAMBIO DE DATOS (ÚNICAMENTE PARA LAS OFICINAS DE PATENTES OPI)

Versión 1.4

Revisión aprobada Propuesta presentada por el Comité Equipo Técnico sobre Listas de Normas Técnicas de Secuencias para que el CWS la OMPI (CWS) examine y apruebe en su octava novena sesión, celebrada el 4 de diciembre de 2020

En el contexto del intercambio de datos con los proveedores de bases de datos (miembros de INSD), las Oficinas de patentes propiedad intelectual deberían completar, para cada secuencia, el elemento INSDSeq_other-seqids mediante un INSDSeqid que contenga una referencia a la patente publicada y al identificador de la secuencia correspondientes, en el siguiente formato:

pat{código de oficina}{número de publicación}{código de tipo de documento}{identificador de secuencia}

en que, “código de oficina” es el código de la Oficina de PI que publica el documento de patente como se establece en la Norma Técnica ST.3; “código de tipo de documento” es el código de identificación de diferentes tipos de documentos de patente, como se establece en la Norma ST.16; “número de publicación” es el número de publicación de la solicitud o la patente; e “identificador de secuencia” es el número de la secuencia que figura en la solicitud o la patente.

Ejemplo:

pat|WO|2013999999|A1|123456

Que se traducirá a una instancia XML válida como:

```
<INSDSeq_other-seqids>  
  <INSDSeqid>pat|WO|2013999999|A1|123456</INSDSeqid>  
</INSDSeq_other-seqids>
```

En que “123456” es la secuencia 123456 de la publicación WO n.º 2013999999 (A1).

[Sigue el Anexo VI de la Norma ST.26]

ANEXO VI

DOCUMENTO DE ORIENTACIÓN

Versión 1.45

Revisión aprobada Propuesta presentada por el Comité Equipo Técnico sobre Listas de Normas Técnicas de Secuencias para que el CWS la OMPI (CWS) examine y apruebe en su octava novena sesión, celebrada el 4 de diciembre de 2020

ÍNDICE

| | |
|--------------------------|------------|
| INTRODUCCIÓN | 3.26.vi.2 |
| ÍNDICE DE EJEMPLOS | 3.26.vi.7 |
| EJEMPLOS | 3.26.vi.17 |
| APÉNDICE | 3.26.vi.74 |

INTRODUCCIÓN

En la presente Norma se indica que uno de sus objetivos es “permitir que el solicitante establezca una única lista de secuencias en una solicitud de patente que sea aceptable a los efectos tanto de los procedimientos internacionales como nacionales o regionales”. El propósito del presente documento de orientación es velar por que todos los solicitantes y las oficinas de propiedad intelectual (oficinas de PI) entiendan y acepten los requisitos relativos a la inclusión y la representación de las divulgaciones de secuencias, de manera de alcanzar el objetivo mencionado.

El presente documento de orientación consta de esta introducción, un índice de ejemplos, ejemplos de divulgaciones de secuencias y un apéndice que contiene una lista de secuencias en XML con secuencias que figuran en los ejemplos. En esta introducción se explican algunos conceptos y la terminología utilizada en el resto del presente documento. Los ejemplos ilustran los requisitos de párrafos específicos de la Norma, y cada ejemplo ha sido designado con el número de párrafo pertinente. Algunos ejemplos ilustran además otros párrafos y, al final de cada ejemplo, se indican las referencias a otros ejemplos correspondientes. En el índice se indican los números de página de los ejemplos, así como las referencias a otros ejemplos. A cada una de las secuencias de un ejemplo que debe o puede incluirse en una lista de secuencias se ha asignado un identificador de secuencia (SEQ ID NO) y dichas secuencias figuran en formato XML en el [Apéndice](#) del presente documento.

Para cada ejemplo, toda información explicativa que se presenta con una secuencia ha de ser considerada como la totalidad de la divulgación relativa a esa secuencia. En las respuestas ofrecidas se tiene en cuenta únicamente la información presentada expresamente en el ejemplo.

La orientación que brinda el presente documento está destinada a la preparación de una lista de secuencias que ha de suministrarse en la fecha de presentación de una solicitud de patente. Al preparar una lista de secuencias que haya de suministrarse con posterioridad a la fecha de presentación de una solicitud de patente debe tenerse en cuenta si la oficina de PI podría considerar que en la información suministrada se añade materia a la divulgación original. Por lo tanto, es posible que la orientación brindada en el presente documento no sea aplicable a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente.

Preparación de una lista de secuencias

Al preparar una lista de secuencias para una solicitud de patente es preciso formularse las preguntas siguientes:

1. ¿Exige el párrafo 7 de la Norma ST.26 la inclusión de una secuencia divulgada en particular?
2. Si no se exige la inclusión de una secuencia divulgada en particular, ¿permite la Norma ST.26 la inclusión de esa secuencia?
3. Si la Norma ST.26 exige o permite la inclusión de una secuencia divulgada en particular, ¿cómo debería representarse esa secuencia en la lista de secuencias?

En cuanto a la primera pregunta, el párrafo 7 de la Norma ST.26 (con determinadas restricciones) exige la inclusión de una secuencia divulgada en una solicitud de patente mediante la enumeración de sus residuos, cuando la secuencia contiene diez o más nucleótidos específicamente definidos o cuatro o más aminoácidos específicamente definidos.

En cuanto a la segunda pregunta, el párrafo 8 de la Norma ST.26 prohíbe la inclusión cualquier secuencia que tenga menos de diez nucleótidos específicamente definidos o cuatro aminoácidos específicamente definidos.

Para responder a esas preguntas, es necesario entender claramente el significado de "enumeración de sus residuos" y "específicamente definidos".

En cuanto a la tercera pregunta, el presente documento presenta divulgaciones de secuencias que sirven para ejemplificar distintas hipótesis, junto con un examen completo de los medios de representación preferidos para cada secuencia o, cuando una secuencia contiene múltiples variaciones, la "secuencia más abarcadora", de conformidad con la presente Norma. Puesto que es imposible abordar todas las hipótesis posibles de secuencias inusuales, el presente documento de orientación intenta exponer el razonamiento en el que se funda el enfoque respecto de cada ejemplo, así como la forma de aplicar las disposiciones de la Norma ST.26, de manera que el mismo razonamiento pueda aplicarse a otras hipótesis de secuencias que no estén ejemplificadas.

Enumeración de sus residuos

En el párrafo 3.c) de la Norma ST.26 se define "enumeración de sus residuos" como la divulgación de una secuencia en una solicitud de patente en la que se enumera, por orden, cada residuo de la secuencia, donde i) el residuo se representa mediante un nombre, abreviatura, símbolo o estructura; o ii) varios residuos se representan mediante una fórmula abreviada. Una secuencia debería divulgarse en una solicitud de patente mediante la "enumeración de sus residuos" utilizando símbolos convencionales, que son los símbolos de nucleótidos expuestos en el Cuadro 1 de la Sección 1 del Anexo 1 de la Norma ST.26 (es decir, los símbolos en letra minúscula o sus equivalentes en letra mayúscula¹) y los símbolos de aminoácidos expuestos en el Cuadro 3 de la Sección 3 del Anexo 1 de la Norma ST.26 (es decir, los símbolos en letra mayúscula o sus equivalentes en letra minúscula). Otros símbolos que no sean los que se exponen en dichos cuadros son "no convencionales" (minúscula¹). En lo sucesivo, estos símbolos de nucleótidos y aminoácidos se denominarán símbolos convencionales. Las representaciones de nucleótidos y aminoácidos distintas de las establecidas en estos cuadros se denominan en adelante "no convencionales".

A veces, una secuencia Cuando la representación de un residuo se divulga de una manera que no es la preferida mediante "enumeración de sus residuos" utilizando abreviaturas convencionales o nombres completos (en lugar de símbolos convencionales), expuestos en los Cuadros A y B, más abajo, símbolos convencionales o abreviaturas utilizados de manera no convencional, símbolos o abreviaturas no convencionales, fórmulas/estructuras químicas, o fórmulas abreviadas. Debería procurarse divulgar las secuencias de la manera preferida; sin embargo, cuando las secuencias se divulgan de una manera no es la preferida, podrá ser necesario consultar la explicación de la secuencia en la divulgación para determinar el significado del símbolo o la abreviatura no preferidos que se han utilizado.

Aun cuando se utilizan un símbolo o una abreviatura convencionales, debe consultarse la explicación de la secuencia en la divulgación para confirmar que el símbolo se utiliza de manera convencional. O bien, si el símbolo se utiliza de manera no convencional, la explicación es necesaria para determinar si el párrafo 7 de la Norma ST.26 exige la inclusión en la lista de secuencias o si el párrafo 8 prohíbe la inclusión.

Cuando en la divulgación un símbolo o una abreviatura no convencionales equivalen como equivalente a un símbolo o una abreviatura convencionales (por ejemplo, "convencional (p. ej., "Z1" significa "A")" o a una secuencia específica de símbolos convencionales (por ejemplo, "p. ej., "Z1" significa "agga"), la secuencia se interpreta la secuencia como si se divulgara utilizando el símbolo, los símbolos o la abreviatura o abreviaturas convencionales equivalentes para determinar si el párrafo 7 de la Norma ST.26 exige la requiere su inclusión en la lista de secuencias o si el párrafo 8 prohíbe su inclusión. Cuando se utiliza un símbolo no convencional de nucleótido como símbolo de ambigüedad (por ejemplo, X1 = inosina o pseudouridina), pero no es equivalente a uno de los símbolos convencionales de ambigüedad que figuran en el Cuadro 1 de la Sección 1 (es decir, "m", "r", "w", "s", "y", "k", "v", "h", "d", "b", o "n"), el residuo se interpreta como un residuo "n" para determinar si el párrafo 7 de la Norma ST.26 exige requiere la inclusión de la secuencia en la lista de secuencias o si el párrafo 8 de la Norma ST.26 prohíbe la inclusión. De manera similar, cuando se utiliza un símbolo no convencional de aminoácido como símbolo de ambigüedad (por ejemplo, "Z1" significa "A", "G", "S" o "T"), pero no es equivalente a uno de los símbolos convencionales de ambigüedad que figuran en el Cuadro 3 de la Sección 3 (es decir, B, Z, J o X), el residuo se interpreta como un residuo "X" para determinar si el párrafo 7 de la Norma ST.26 exige requiere la inclusión de la secuencia en la lista de secuencias o si el párrafo 8 de la Norma ST.26 prohíbe la inclusión.

Debe prestarse atención a la hora de divulgar las secuencias utilizando símbolos convencionales; no obstante, cuando las secuencias se divulguen de otro modo, es posible que deba consultarse una explicación en la divulgación para determinar el significado de la representación no convencional.

Cuando se utiliza un símbolo convencional, debe consultarse la explicación de la secuencia en la divulgación para confirmar que el símbolo se utiliza de manera convencional. Si el símbolo se utiliza de manera no convencional, esta

¹ NOTA: Mientras que en la divulgación de una solicitud pueden representarse los nucleótidos o los aminoácidos mediante símbolos en letra tanto minúscula como mayúscula, para una secuencia incluida en una lista de secuencias deberá utilizarse únicamente la letra minúscula para la representación de una secuencia de nucleótidos (véase el párrafo 13 de la Norma ST.26) y únicamente la letra mayúscula para la representación de una secuencia de aminoácidos (véase el párrafo 26 de la Norma ST.26).

explicación es necesaria para determinar si el párrafo 7 de la Norma ST.26 requiere la inclusión de la secuencia en la lista de secuencias o si el párrafo 8 prohíbe la inclusión.

Específicamente definido

En el párrafo 3.k) de la Norma ST.26 se define “específicamente definido” como todo nucleótido distinto a los representados por el símbolo “n” y todo aminoácido distinto a los representados por el símbolo “X” que se enumeran en el Anexo I, utilizándose “n” y “X” de manera convencional, según se describe en el Cuadro 1 de la Sección 1 (es decir, “a o c o g o t/u; ‘unknown’ o ‘other’”) y en el Cuadro 3 de la Sección 3 (es decir, “A o R o N o D o C o Q o E o G o H o I o L o K o M o F o P u o S o U o T o W o Y o V; ‘unknown’ o ‘other’”), respectivamente. La explicación que figura más arriba acerca de los símbolos convencionales o símbolos o abreviaturas no convencionales y su utilización de manera convencional o no convencional se tendrá en cuenta para determinar si un nucleótido o un aminoácido son “específicamente definidos”.

Secuencia más abarcadora

Cuando una secuencia que cumple con los requisitos del párrafo 7 se divulga mediante enumeración de sus residuos una sola vez en una solicitud, pero se describe de forma diferente en múltiples realizaciones de la invención; por ejemplo, en una realización, “X” en uno o más lugares podría ser cualquier aminoácido, pero en otras realizaciones, “X” podría ser solamente un número limitado de aminoácidos, la Norma ST.26 exige la inclusión en una lista de secuencias tan solo de la secuencia única que ha sido enumerada mediante sus residuos. Conforme a los párrafos 15 y 27, cuando esa secuencia contiene múltiples símbolos de ambigüedad “n” o “X”, se interpreta que “n” o “X” representan cualquier nucleótido o aminoácido, respectivamente, en ausencia de anotaciones adicionales. En consecuencia, la secuencia única que se exige incluir es la secuencia más abarcadora divulgada. La secuencia más abarcadora es la secuencia única que tiene variantes de residuos que están representadas por los símbolos de ambigüedad más restrictivos que incluyen en las realizaciones más divulgadas. Del mismo modo, cuando una secuencia se divulga mediante la enumeración de sus residuos una sola vez, pero la longitud de la secuencia puede variar debido a la variación del número de copias, la realización más larga de la secuencia se considera la secuencia más abarcadora. Por ejemplo, considérese una secuencia que contenga una región repetida que puede variar de 2 a 5 copias, como se ha enumerado. La realización con 5 copias de la repetición es la secuencia más abarcadora y debe incluirse en la lista de secuencias. Sin embargo, se alienta encarecidamente a que se incluyan secuencias específicas adicionales, de ser factible, que representen, por ejemplo, realizaciones adicionales que constituyen una parte fundamental de la invención. La inclusión de las secuencias adicionales permite una búsqueda más minuciosa y ofrece información al público acerca de la materia respecto de la cual se solicita una patente.

Utilización del símbolo de ambigüedad

Utilización correcta del símbolo de ambigüedad “n” en una lista de secuencias

El símbolo “n”

- a. no debe utilizarse para representar otras cosa que no sea un nucleótido simple;
- b. se interpretará como “a”, “c”, “g”, o “t/u”, excepto cuando se utilice en una descripción más detallada;
- c. deberá utilizarse para representar cualquiera de los nucleótidos siguientes, junto con una descripción más detallada:
 - i. un nucleótido modificado, por ejemplo, natural, sintético o que no existe en estado natural, que no pueda representarse de otra manera por ningún otro símbolo descrito en el Anexo I (véase el Cuadro 1 de la Sección 1);
 - ii. un nucleótido “unknown”, es decir, no determinado, no divulgado o incierto;
 - iii. un sitio abásico; o
- d. puede utilizarse para representar variaciones de una secuencia, es decir, alternativas, supresiones, inserciones o sustituciones, siendo “n” el símbolo de ambigüedad más restrictivo.

Utilización correcta del símbolo de ambigüedad “X” en una lista de secuencias

El símbolo “X”

- a. no debe utilizarse para representar otras cosa que no sea un aminoácido simple;
- b. se interpretará como “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, o “V”, excepto cuando se utilice en una descripción más detallada;
- c. deberá utilizarse para representar cualquiera de los aminoácidos siguientes, junto con una descripción más detallada:

- i. un aminoácido modificado, por ejemplo, natural, sintético o que no existe en estado natural, que no pueda representarse de otra manera por ningún otro símbolo descrito en el Anexo I (véase el Cuadro 3 de la Sección 3);
 - ii. un aminoácido “unknown”, es decir, no determinado, no divulgado o incierto; o
- d. puede utilizarse para representar variaciones de una secuencia, es decir, alternativas, supresiones, inserciones o sustituciones, siendo “X” el símbolo de ambigüedad más restrictivo.

Anotación de residuos modificados

Esta Norma requiere que los residuos “modificados” se anoten según el párrafo 17 en lo que respecta a los nucleótidos y según el párrafo 30 en lo que respecta a los aminoácidos.

En el párrafo 3.e) de la Norma ST.26 se define “aminoácido modificado” como cualquier aminoácido descrito en el párrafo 3.a) 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 o L-valina. Del mismo modo, la Norma define “nucleótido modificado” como cualquier nucleótido descrito en el párrafo 3.g) que no sea 3'-monofosfato de desoxiadenosina, 3'-monofosfato de desoxiguanosina, 3'-monophosphate de desoxicitidina, 3'-monofosfato de desoxitimidina, 3'-monofosfato de adenosina, 3'-monofosfato de guanosina, 3'-monofosfato de citidina o 3'-monofosfato de uridina (párrafo 3.f) de la Norma ST.26).

De acuerdo con las definiciones anteriores, las modificaciones de las nucleobases o del esqueleto de azúcar-fosfato de un ácido nucleico y las modificaciones de los grupos R del aminoácido o del esqueleto peptídico de un péptido dan como resultado uno o más “nucleótidos modificados” o “aminoácidos modificados”, respectivamente. Por lo tanto, esos nucleótidos y aminoácidos deben ser anotados. Algunos ejemplos de modificaciones del esqueleto son los análogos de los nucleótidos, como los ácidos nucleicos peptídicos (ANP) y los ácidos nucleicos glicólicos (ANG), y los D-aminoácidos.

Nótese que la modificación de un aminoácido terminal de un péptido o de un nucleótido terminal de un ácido nucleico no da lugar necesariamente a un “aminoácido modificado” o a un “nucleótido modificado”. Hay que observar la modificación terminal y determinar si la modificación cambia la estructura química del residuo de manera que este quede fuera de las excepciones establecidas en los párrafos 3.e) y 3.f). Por ejemplo, un péptido en el que el residuo C-terminal está unido a una estructura (como parte de una secuencia ramificada; véase el péptido N.º 2 del ejemplo 7.b)-3) mediante un enlace amídico convencional no se considera un “residuo modificado” y, por tanto, no es necesario anotarlo. Del mismo modo, un péptido en el que el residuo N-terminal está unido por amida a la biotina no se considera un “residuo modificado” y, por tanto, no es necesario anotarlo. En ambos casos, la estructura del residuo implicado en el enlace C-terminal o N-terminal no se modifica con respecto a los aminoácidos convencionales mencionados en el párrafo 3.e) de la Norma.

En cambio, las modificaciones terminales que cambian la estructura química del residuo se consideran “residuos modificados” y deben anotarse. Por ejemplo, la metilación del C-terminal en el ejemplo 3.c)-1 sí cambia la estructura química del residuo terminal, ya que el grupo metilo sustituye al hidroxilo que normalmente se encuentra en el grupo alfa-carboxilo. Por lo tanto, esta lisina metilada debe ser anotada como un “residuo modificado”.

Obsérvese que el solicitante deberá evaluar cada modificación de residuo terminal dentro de una secuencia enumerada y determinar si se ha modificado o no la estructura del residuo terminal. Si la estructura del residuo modificado es diferente de los aminoácidos o nucleótidos convencionales indicados en el párrafo 3.e) y 3.f) de la Norma, la modificación debe ser anotada.

Por último, siempre se recomienda que en las listas de secuencias se incluya toda la información que sea razonable para representar la divulgación con la mayor precisión posible. Por lo tanto, aunque no sea obligatorio anotar una modificación, es preferible incluirla. Sin embargo, la anotación de variantes de una secuencia primaria enumerada debe cumplir los requisitos de los párrafos 93 a 100 de la Norma ST.26. Es posible que no sea necesario incluir en la lista de secuencias las modificaciones que se divulguen como variantes de una secuencia enumerada. La definición de la anotación de variantes figura en los párrafos 93 a 95 de la Norma ST.26.

Representación de los residuos modificados

En la Norma ST.26 se indica que los nucleótidos y aminoácidos modificados deberían representarse en la lista de secuencias, siempre que sea posible, como el correspondiente residuo no modificado (véanse los párrafos 16 y 29).

Obsérvese que esta recomendación es un “debería”, es decir, un “enfoque al que se insta vivamente, pero no un requisito” (véase el párrafo 4.d). Queda a la discreción del solicitante decidir si un residuo modificado será representado por el correspondiente residuo no modificado o por las variables “n” o “X”.

Como regla general, si un residuo es modificado por la adición de una fracción, como la metilación o la acetilación, y la estructura del residuo no modificado no cambia en general, se recomienda la representación mediante el residuo no modificado. Por ejemplo, una adenosina metilada debería representarse preferentemente con “a” en la lista de secuencias. Sin embargo, cuando el residuo modificado es estructuralmente diferente de cualquier residuo no modificado, se recomienda una “n” o una “X”. Por ejemplo, la norleucina es un isómero de la leucina, y su cadena lateral es una estructura lineal de 4 carbonos. La leucina también tiene una cadena lateral de 4 carbonos, pero está ramificada en el segundo carbono. Por lo tanto, la norleucina no es simplemente el resultado de una modificación añadida a la leucina, sino una estructura completamente diferente (aunque relacionada). Por lo tanto, se recomienda que la norleucina se represente con una “X” en la lista de secuencias.

Un nucleótido está “específicamente definido” cuando está representado por cualquier elemento que no sea ‘n’, y un aminoácido está “específicamente definido” cuando está representado por cualquier elemento que no sea ‘X’ (véase el párrafo 3.k) de la Norma ST.26). Por lo tanto, una adenosina de 2’ O-metilo representada por una ‘a’ en la secuencia está específicamente definida, mientras que la norleucina representada por una ‘X’ en la secuencia no está específicamente definida.

Cuadro A – Símbolos y definiciones convencionales de nucleótidos

| Símbolo | Abreviatura | Definición |
|---------|---------------------------------------|------------|
| a | adenine | |
| c | cytosine | |
| g | guanine | |
| t | thymine en DNA
uracil en RNA (t/u) | |
| m | a o c | |
| r | a o g | |
| w | a o t/u | |
| s | c o g | |
| y | c o t/u | |
| k | g o t/u | |
| v | a o c o g; no t/u | |
| h | a o c o t/u; no g | |
| d | a o g o t/u; no c | |
| b | c o g o t/u; no a | |
| n | a o c o g o t/u; “unknown” o “other” | |

Cuadro B - Símbolos, abreviaturas, códigos de tres letras y nombres/definiciones convencionales de aminoácidos

| Símbolo | Abreviatura | Código de 3 letras | Nombre del aminoácido | Definición |
|---------|-------------|--------------------|---------------------------|------------|
| A | Ala | | Alanine | |
| R | Arg | | Arginine | |
| N | Asn | | Asparagine | |
| D | Asp | | Aspartic Acid (Aspartate) | |
| C | Cys | | Cysteine | |
| Q | Gln | | Glutamine | |
| E | Glu | | Glutamic Acid (Glutamate) | |
| G | Gly | | Glycine | |
| H | His | | Histidine | |
| I | Ile | | Isoleucine | |
| L | Leu | | Leucine | |
| K | Lys | | Lysine | |
| M | Met | | Methionine | |
| F | Phe | | Phenylalanine | |

| | | |
|---|-----|---|
| P | Pro | Proline |
| O | Pyl | Pyrrolysine |
| S | Ser | Serine |
| U | Sec | Selenocysteine |
| T | Thr | Threonine |
| W | Trp | Tryptophan |
| Y | Tyr | Tyrosine |
| V | Val | Valine |
| B | Asx | Aspartic Acid o Asparagine |
| Z | Glx | Glutamine o Glutamic Acid |
| J | Xle | Leucine o Isoleucine |
| X | Xaa | A o R o N o D o C o Q o E o G o
H o I o L o K o M o F o P o O o
S o U o T o W o Y o V,
"unknown" o "other" |

ÍNDICE DE EJEMPLOS

Página

Párrafo 3.a) – Definición de “aminoácido”

Ejemplo 3.a)-1: D-aminoácidos16

Referencia a otros ejemplos

Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other”52

Ejemplo 29-2: Uso del correspondiente aminoácido no modificado53

Ejemplo 30-1: Clave de caracterización “CARBOHYD”54

Párrafo 3.c) – Definición de “enumeración de sus residuos”

Ejemplo 3.c)-1: Enumeración de aminoácidos mediante la estructura química17

Ejemplo 3.c)-2: Fórmula abreviada de una secuencia de aminoácidos18

Referencia a otros ejemplos

Ejemplo 27-1: Fórmula abreviada de una secuencia de aminoácidos47

Ejemplo 27-3: Fórmula abreviada – cuatro o más aminoácidos específicamente definidos49

Párrafo 3.f) – Definición de “nucleótido modificado”

Referencia a otros ejemplos

Ejemplo 3.g)-4: Análogos de ácido nucleico22

Párrafo 3.g) – Definición de “nucleótido”

Ejemplo 3.g)-1: Secuencia de nucleótidos interrumpida mediante un espaciador C319

Ejemplo 3.g)-2: Secuencia de nucleótidos con residuos alternativos, incluido un espaciador C320

Ejemplo 3.g)-3: Sitio abásico21

Ejemplo 3.g)-4: Análogos de ácido nucleico22

Referencia a otros ejemplos

Ejemplo 11.b)-1: Secuencia de nucleótidos de doble cadena – longitudes diferentes43

Ejemplo 14-1: El símbolo “t” representa el uracilo en ARN45

Párrafo 3.k) – Definición de “específicamente definido”

Ejemplo 3.k)-1: Símbolos de ambigüedad de nucleótidos23

Ejemplo 3.k)-2: Símbolo de ambigüedad “n” utilizado de manera tanto convencional como no convencional24

Ejemplo 3.k)-3: Símbolo de ambigüedad “n” utilizado de manera no convencional25

Ejemplo 3.k)-4: Los símbolos de ambigüedad distintos de “n” son “específicamente definidos”26

Ejemplo 3.k)-5: Abreviatura de ambigüedad “Xaa” utilizada de manera no convencional27

Párrafo 7 – Secuencias para las que se exige la inclusión en una lista de secuencias

Referencia a otros ejemplos

Ejemplo 28-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada50

Ejemplo 55-1: Molécula combinada ADN/ARN60

Ejemplo 89-2: Una localización de característica se extiende más allá de la secuencia divulgada62

Ejemplo 92-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones64

Párrafo 7.a) – Secuencia de nucleótidos exigida en una lista de secuencias

| | |
|--|--------------------|
| Ejemplo 7.a)-1: Secuencia de nucleótidos ramificada..... | 28 |
| Ejemplo 7.a)-2: Secuencia de nucleótidos lineal que tiene una estructura secundaria | 30 |
| Ejemplo 7.a)-3: Símbolos de ambigüedad de nucleótidos utilizados de manera no convencional | 31 |
| Ejemplo 7.a)-4: Símbolos de ambigüedad de nucleótidos utilizados de manera no convencional | 32 |
| Ejemplo 7.a)-5: Símbolos no convencionales de nucleótidos..... | 33 |
| Ejemplo 7.a)-6: Símbolos no convencionales de nucleótidos..... | 34 |

Referencia a otros ejemplos

| | |
|--|--------------------|
| Ejemplo 3.g)-1: Secuencia de nucleótidos interrumpida por un espaciador C3 | 19 |
| Ejemplo 3.g)-2: Secuencia de nucleótidos con residuos alternativos, incluido un espaciador C3 | 20 |
| Ejemplo 3.g)-3: Sitio abásico..... | 21 |
| Ejemplo 3.g)-4: Análogos de ácido nucleico | 22 |
| Ejemplo 3.k)-1: Símbolos de ambigüedad de nucleótidos | 23 |
| Ejemplo 3.k)-2: Símbolo de ambigüedad “n” utilizado de manera tanto convencional
como no convencional..... | 24 |
| Ejemplo 3.k)-3: Símbolo de ambigüedad “n” utilizado de manera no convencional..... | 25 |
| Ejemplo 3.k)-4: Los símbolos de ambigüedad distintos de “n” son “específicamente definidos” | 26 |
| Ejemplo 11.a)-1: Secuencia de nucleótidos de doble cadena –longitudes iguales | 42 |
| Ejemplo 11.b)-1: Secuencia de nucleótidos de doble cadena –longitudes diferentes | 43 |
| Ejemplo 11.b)-2: Secuencia de nucleótidos de doble cadena – ningún segmento con
apareamiento de bases..... | 44 |
| Ejemplo 14-1: El símbolo “t” representa el uracilo en ARN..... | 45 |
| Ejemplo 89-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada | 61 |
| Ejemplo 93-1: Representación de variantes enumeradas | 66 |
| Ejemplo 95.b)-1: Representación de distintas variantes de secuencias con múltiples variaciones
interdependientes..... | 71 |

Párrafo 7.b) – Secuencias de aminoácidos exigidas en una lista de secuencias

| | |
|--|--------------------|
| Ejemplo 7.b)-1: Cuatro o más aminoácidos específicamente definidos | 35 |
| Ejemplo 7.b)-2: Secuencia de aminoácidos ramificada..... | 36 |
| Ejemplo 7.b)-3: Secuencia de aminoácidos ramificada..... | 38 |
| Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada | 39 |
| Ejemplo 7.b)-5: Péptido cíclico que contiene una secuencia de aminoácidos ramificada | 41 |

Referencia a otros ejemplos

| | |
|---|--------------------|
| Ejemplo 3.a)-1: D-aminoácidos | 16 |
| Ejemplo 3.c)-1: Enumeración de aminoácidos mediante la estructura química | 17 |
| Ejemplo 3.c)-2: Fórmula abreviada de una secuencia de aminoácidos | 18 |
| Ejemplo 3.k)-5: Abreviatura de ambigüedad “Xaa” utilizada de manera no convencional | 27 |

| | |
|--|-----------|
| Ejemplo 27-1: Fórmula abreviada de una secuencia de nucleótidos | 47 |
| Ejemplo 27-3: Fórmula abreviada - cuatro o más aminoácidos específicamente definidos..... | 49 |
| Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido "other" | 52 |
| Ejemplo 29-2: Uso del correspondiente aminoácido no modificado | 53 |
| Ejemplo 30-1: Clave de caracterización "CARBODHYD" | 54 |
| Ejemplo 36-1: Secuencia con una región de un número conocido de residuos "X" representada como secuencia única..... | 55 |
| Ejemplo 37-1: Una secuencia con regiones de un número desconocido de residuos "X" no debe representarse como secuencia única | 58 |
| Ejemplo 37-2: Una secuencia con regiones de un número desconocido de residuos "X" no debe representarse como secuencia única | 59 |
| Ejemplo 89-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada | 61 |
| Ejemplo 93-2: Representación de variantes enumeradas | 67 |
| Ejemplo 93-3: Representación de una secuencia de consenso | 68 |
| Ejemplo 94-1: Representación de una secuencia única con aminoácidos alternativos enumerados .. | 69 |
| Ejemplo 95.a)-1: Representación de una variante de secuencia mediante anotación de la secuencia principal | 70 |

Párrafo 8 – Umbral para la inclusión de secuencias

Referencia a otros ejemplos

| | |
|--|----|
| Ejemplo 3.k)-1: Símbolos de ambigüedad de nucleótidos | 23 |
| Ejemplo 3.k)-2: Símbolo de ambigüedad "n" utilizado de manera tanto convencional como no convencional..... | 24 |
| Ejemplo 7.a)-1: Secuencia de nucleótidos ramificada | 28 |
| Ejemplo 7.a)-6: Símbolos no convencionales de nucleótidos..... | 34 |
| Ejemplo 7.b)-1: Cuatro o más aminoácidos específicamente definidos | 35 |
| Ejemplo 7.b)-2: Secuencia de aminoácidos ramificada | 36 |
| Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada | 39 |
| Ejemplo 14-1: El símbolo "t" representa el uracilo en ARN..... | 45 |
| Ejemplo 37-1: Una secuencia con regiones de un número desconocido de residuos "X" no debe representarse como secuencia única | 58 |
| Ejemplo 37-2: Una secuencia con regiones de un número desconocido de residuos "X" no debe representarse como secuencia única | 59 |
| Ejemplo 94-1: Representación de una secuencia única con aminoácidos alternativos enumerados..... | 69 |

Párrafo 11 – Representación de una secuencia de nucleótidos

Referencia a otros ejemplos

| | |
|---|----|
| Ejemplo 3.g)-4: Análogos de ácido nucleico | 22 |
| Ejemplo 7.a)-1: Secuencia de nucleótidos ramificada | 28 |

Párrafo 11.a) Secuencia de nucleótidos de doble cadena - totalmente complementarias

| | |
|---|-----------|
| Ejemplo 11.a)-1: Secuencia de nucleótidos de doble cadena – longitudes iguales | 42 |
|---|-----------|

Párrafo 11.b) – Secuencia de nucleótidos de doble cadena – no totalmente complementarias

Ejemplo 11.b)-1: Secuencia de nucleótidos de doble cadena – longitudes diferentes43

Ejemplo 11.b)-2: Secuencia de nucleótidos de doble cadena – ningún segmento con apareamiento de bases ..44

Párrafo 13 – Representación de nucleótidos

Referencia a otros ejemplos

Ejemplo 3.k)-2: Símbolo de ambigüedad “n” utilizado de manera tanto convencional como no convencional.....24

Ejemplo 7.a)-1: Secuencia de nucleótidos ramificada28

Ejemplo 14-1: El símbolo “t” representa el uracilo en ARN.....45

Ejemplo 93-1: Representación de variantes enumeradas66

Párrafo 14 – Símbolo “t” interpretado como uracilo en ARN

Ejemplo 14-1: El símbolo “t” representa el uracilo en ARN.....45

Referencia a otros ejemplos

Ejemplo 55-1: Molécula combinada ADN/ARN.....60

Párrafo 15 – Debería utilizarse el símbolo de ambigüedad de nucleótidos más restrictivo

Referencia a otros ejemplos

Ejemplo 3.g)-1: Secuencia de nucleótidos interrumpida por un espaciador C3.....19

Ejemplo 3.g)-2: Secuencia de nucleótidos con residuos alternativos, incluido un espaciador C3.....20

Ejemplo 3.k)-4: Los símbolos de ambigüedad distintos de “n” son “específicamente definidos”26

Ejemplo 95.b)-1: Representación de distintas variantes de secuencias con múltiples variaciones interdependientes71

Párrafo 16 – Representación de un nucleótido modificado

Referencia a otros ejemplos

Ejemplo 3.g)-1: Secuencia de nucleótidos interrumpida por un espaciador C3.....19

Ejemplo 3.g)-4: Análogos de ácido nucleico22

Párrafo 17 – Anotación de un nucleótido modificado

Referencia a otros ejemplos

Ejemplo 3.g)-1: Secuencia de nucleótidos interrumpida por un espaciador C3.....19

Ejemplo 3.g)-3: Sitio abásico.....21

Ejemplo 7.a)-1: Secuencia de nucleótidos ramificada28

Ejemplo 7.a)-2: Secuencia lineal de nucleótidos con una estructura secundaria.....30

Ejemplo 7.a)-6: Símbolos no convencionales de nucleótidos.....34

Párrafo 18 – Anotación de regiones de nucleótidos modificados consecutivos

Referencia a otros ejemplos

Ejemplo 3.g)-4: Análogos de ácido nucleico22

Ejemplo 11.b)-1: Secuencia de nucleótidos de doble cadena – longitudes diferentes43

Párrafo 19 – Anotación de uracilo en ADN o timina en ARN

Referencia a otros ejemplos

Ejemplo 14-1: El símbolo “t” representa el uracilo en ARN.....45

Párrafo 25 – Residuo de la posición número 1 en la secuencia de aminoácidos

Referencia a otros ejemplos

Ejemplo 3.a)-1: D-aminoácidos16

Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada39

Ejemplo 7.b)-5: Péptido cíclico que contiene una secuencia de aminoácidos ramificada41

Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other”52

Párrafo 26 – Representación de aminoácidos

Referencia a otros ejemplos

Ejemplo 7.b)-2: Secuencia de aminoácidos ramificada36

Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada39

Ejemplo 7.b)-5: Péptido cíclico que contiene una secuencia de aminoácidos ramificada41

Ejemplo 36-1: Secuencia con una región de un número conocido de residuos “X” representada como secuencia única.....55

Ejemplo 89-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada61

Ejemplo 92-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones.....64

Ejemplo 93-2: Representación de variantes enumeradas67

Ejemplo 93-3: Representación de una secuencia de consenso68

Párrafo 27 – Debería utilizarse el símbolo de ambigüedad de aminoácidos más restrictivo

Ejemplo 27-1: Fórmula abreviada de una secuencia de aminoácidos.....47

Ejemplo 27-2: Fórmula abreviada – menos de cuatro aminoácidos específicamente definidos.....48

Ejemplo 27-3: Fórmula abreviada - cuatro o más aminoácidos específicamente definidos.....49

Referencia a otros ejemplos

Ejemplo 3.c)-2: Fórmula abreviada de una secuencia de aminoácidos18

Ejemplo 7.b)-1: Cuatro o más aminoácidos específicamente definidos35

Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other”52

Ejemplo 36-1: Secuencia con una región de un número conocido de residuos “X” representada como secuencia única.....55

Ejemplo 36-2: Secuencia con múltiples regiones de un número o una serie conocidos de residuos “X” representada como secuencia única56

Ejemplo 36-3: Secuencia con múltiples regiones de un número o una serie conocidos de residuos “X” representada como secuencia única57

Ejemplo 37-2: Una secuencia con regiones de un número desconocido de residuos “X” no debe representarse como secuencia única.59

Ejemplo 93-3: Representación de una secuencia de consenso68

Ejemplo 94-1: Representación de una secuencia única con aminoácidos alternativos enumerados.....69

Ejemplo 95.a)-1: Representación de una variante de secuencia mediante anotación de la secuencia principal70

Párrafo 28 – Secuencias de aminoácidos separadas por símbolos internos de terminación

Ejemplo 28-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada.....50

Referencia a otros ejemplos

Ejemplo 89-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada61

Ejemplo 92-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones.....64

Párrafo 29 – Representación de un aminoácido “other”

Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other”52

Ejemplo 29-2: Uso del correspondiente aminoácido no modificado53

Referencia a otros ejemplos

Ejemplo 3.a)-1: D-aminoácidos16

Ejemplo 7.b)-2: Secuencia de aminoácidos ramificada36

Ejemplo 7.b)-3: Secuencia de aminoácidos ramificada38

Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada39

Ejemplo 30-1: Clave de caracterización “CARBODHYD”54

Párrafo 30 – Anotación de un aminoácido modificado

Ejemplo 30-1: Clave de caracterización “CARBODHYD”54

Referencia a otros ejemplos

Ejemplo 3.a)-1: D-aminoácidos16

Ejemplo 3.c)-1: Enumeración de aminoácidos mediante la estructura química17

Ejemplo 7.b)-2: Secuencia de aminoácidos ramificada36

Ejemplo 7.b)-3: Secuencia de aminoácidos ramificada38

Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada39

Ejemplo 7.b)-5: Péptido cíclico que contiene una secuencia de aminoácidos ramificada41

Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other”52

Ejemplo 29-2: Uso del correspondiente aminoácido no modificado53

Párrafo 31 – Representación de un D-aminoácido

Referencia a otros ejemplos

Ejemplo 3.a)-1: D-aminoácidos16

Ejemplo 3.c)-1: Enumeración de aminoácidos mediante la estructura química17

Ejemplo 7.b)-2: Secuencia de aminoácidos ramificada36

Ejemplo 7.b)-3: Secuencia de aminoácidos ramificada38

Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada39

Ejemplo 7.b)-5: Péptido cíclico que contiene una secuencia de aminoácidos ramificada41

Párrafo 32 – Anotación de un aminoácido “unknown”

Referencia a otros ejemplos

Ejemplo 3.c)-1: Enumeración de aminoácidos mediante la estructura química17

Párrafo 34 – Anotación de una región contigua de residuos “X”

Referencia a otros ejemplos

Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other”53

Párrafo 36 – Secuencias que contienen regiones de un número exacto de residuos “n” o “X” contiguos

Ejemplo 36-1: Secuencia con una región de un número conocido de residuos “X” representada como secuencia única55

Ejemplo 36-2: Secuencia con múltiples regiones de un número o una serie conocidos de residuos “X” representada como secuencia única56

Ejemplo 36-3: Secuencia con múltiples regiones de un número o una serie conocidos de residuos “X” representada como secuencia única57

Párrafo 37 – Secuencias que contienen regiones de un número desconocido de residuos “n” o “X” contiguos

Ejemplo 37-1: Una secuencia con regiones de un número desconocido de residuos “X” no debe representarse como secuencia única58

Ejemplo 37-2: Una secuencia con regiones de un número desconocido de residuos “X” no debe representarse como secuencia única59

Párrafo 41 – Caracteres reservados

Referencia a otros ejemplos

Ejemplo 89-2: Una localización de característica se extiende más allá de la secuencia divulgada62

Párrafo 54 – El elemento INSDSeq_moltype

Referencia a otros ejemplos

Ejemplo 14-1: El símbolo “t” representa el uracilo en ARN.....45

Párrafo 55 – Secuencia de nucleótidos que contiene segmentos de ADN y ARN

Ejemplo 55-1: Molécula combinada ADN/ARN60

Párrafo 56 – Ejemplo que ilustra una secuencia de nucleótidos que contiene segmentos de ADN y ARN

Referencia a otros ejemplos

Ejemplo 55-1: Molécula combinada ADN/ARN.....60

Párrafo 57 – El elemento INSDSeq_sequence

Referencia a otros ejemplos

Ejemplo 28-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada50

Ejemplo 92-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones.....64

Párrafo 65 – Descriptor de localización

Referencia a otros ejemplos

Ejemplo 3.g)-4: Análogos de ácido nucleico22

Ejemplo 89-2: Una localización de característica se extiende más allá de la secuencia divulgada62

Párrafo 66 – Sintaxis del descriptor de localización

Referencia a otros ejemplos

Ejemplo 3.g)-4: Análogos de ácido nucleico22

Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada39

Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other”52

| | |
|---|----|
| Ejemplo 30-1: Clave de caracterización “CARBODHYD” | 54 |
| Ejemplo 89-2: Una localización de característica se extiende más allá de la secuencia divulgada | 62 |

Párrafo 67 – Operador de localización

Referencia a otros ejemplos

| | |
|--|----|
| Ejemplo 92-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones..... | 64 |
|--|----|

Párrafo 70 – Localizaciones de características

Referencia a otros ejemplos

| | |
|---|----|
| Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada | 39 |
| Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other” | 52 |
| Ejemplo 30-1: Clave de caracterización “CARBODHYD” | 54 |
| Ejemplo 89-2: Una localización de característica se extiende más allá de la secuencia divulgada | 62 |

Párrafo 71 – Representación de los caracteres “<” and “>” en un descriptor de localización

Referencia a otros ejemplos

| | |
|---|----|
| Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other” | 52 |
| Ejemplo 89-2: Una localización de característica se extiende más allá de la secuencia divulgada | 62 |

Párrafo 83 – Ejemplo que ilustra una secuencia de nucleótidos que no ocurre naturalmente

Referencia a otros ejemplos

| | |
|---|----|
| Ejemplo 55-1: Molécula combinada ADN/ARN..... | 60 |
|---|----|

Párrafo 89 – Clave de caracterización “CDS”

| | |
|---|----|
| Ejemplo 89-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada..... | 61 |
|---|----|

| | |
|---|----|
| Ejemplo 89-2: Una localización de característica se extiende más allá de la secuencia divulgada | 62 |
|---|----|

Referencia a otros ejemplos

| | |
|--|----|
| Ejemplo 92-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones..... | 64 |
|--|----|

Párrafo 90 – Los calificadores “transl table” y “translation”

Referencia a otros ejemplos

| | |
|--|----|
| Ejemplo 28-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada..... | 50 |
| Ejemplo 89-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada..... | 61 |
| Ejemplo 92-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones..... | 64 |

Párrafo 92 – Secuencia de aminoácidos codificada por una secuencia codificadora

| | |
|--|----|
| Ejemplo 92-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones..... | 64 |
|--|----|

Referencia a otros ejemplos

| | |
|---|----|
| Ejemplo 28-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada..... | 50 |
| Ejemplo 89-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada..... | 61 |
| Ejemplo 89-2: Una localización de característica se extiende más allá de la secuencia divulgada | 62 |

Párrafo 93 – Secuencia principal y una variante, cada una de ellas enumerada mediante su residuo

| | |
|---|----|
| Ejemplo 93-1: Representación de variantes enumeradas..... | 66 |
|---|----|

| | |
|---|----|
| Ejemplo 93-2: Representación de variantes enumeradas..... | 67 |
|---|----|

| | |
|--|--------------------|
| Ejemplo 93-3: Representación de una secuencia de consenso..... | 68 |
| <u>Párrafo 94 – Variante de secuencia divulgada como secuencia única con residuos alternativos enumerados</u> | |
| Ejemplo 94-1: Representación de una secuencia única con aminoácidos alternativos enumerados..... | 69 |
| <u>Párrafo 95.a) – Variante de secuencia divulgada únicamente mediante referencia a una secuencia principal con múltiples variaciones independientes</u> | |
| Ejemplo 95.a)-1: Representación de una variante de secuencia mediante anotación de la secuencia principal .. | 70 |
| <u>Párrafo 95.b) – Variante de secuencia divulgada únicamente mediante referencia a una secuencia principal con múltiples variaciones interdependientes</u> | |
| Ejemplo 95.b)-1: Representación de distintas variantes de secuencias con múltiples variaciones interdependientes..... | 71 |
| <u>Párrafo 96 – Claves de caracterización y calificadores para una variante de secuencia</u> | |
| <u>Referencia a otros ejemplos</u> | |
| Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other” | 52 |
| <u>Párrafo 97– Anotación de una variante de secuencia</u> | |
| <u>Referencia a otros ejemplos</u> | |
| Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other” | 52 |
| Ejemplo 93-3: Representación de una secuencia de consenso | 68 |
| Ejemplo 94-1: Representación de una secuencia única con aminoácidos alternativos enumerados..... | 69 |

EJEMPLOS

Párrafo 3.a) Definición de "aminoácido"

Ejemplo 3.a)-1: D-aminoácidos

En una solicitud de patente se describe la secuencia siguiente:

Cyclo (D-Ala-D-Glu-Lys-Nle-Gly-D-Met-D-Nle)

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

En la definición del párrafo 3.a) de la Norma se establece que "aminoácido" incluye los "D-aminoácidos" y los aminoácidos que contienen cadenas laterales modificadas o sintéticas. A la luz de esa definición, el péptido enumerado contiene cinco aminoácidos que están específicamente definidos (D-Ala, D-Glu, Lys, Gly y D-Met). Por lo tanto, la secuencia debe incluirse en una lista de secuencias según se exige en el párrafo 7.b) de la Norma ST.26.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

En el párrafo 29 se exige que los D-aminoácidos estén representados en la secuencia como el L-aminoácido correspondiente no modificado. Además, todo aminoácido modificado que no pueda ser representado mediante ningún otro símbolo de los que figuran en el Cuadro 3 de la Sección 3 del Anexo I debe ser representado mediante el símbolo "X".

En este ejemplo, la secuencia contiene tres D-aminoácidos que pueden ser representados por un L-aminoácido no modificado del Cuadro 3 de la Sección 3 del Anexo I, un L-aminoácido (Nle) y un D-aminoácido (D-Nle) que deben ser representados por el símbolo "X".

En el párrafo 25 se indica que cuando las secuencias de aminoácidos tengan una configuración circular y el anillo esté compuesto solo por residuos de aminoácidos unidos por enlaces peptídicos, el solicitante deberá elegir el aminoácido del residuo de la posición número 1. Por lo tanto, la secuencia puede representarse como:

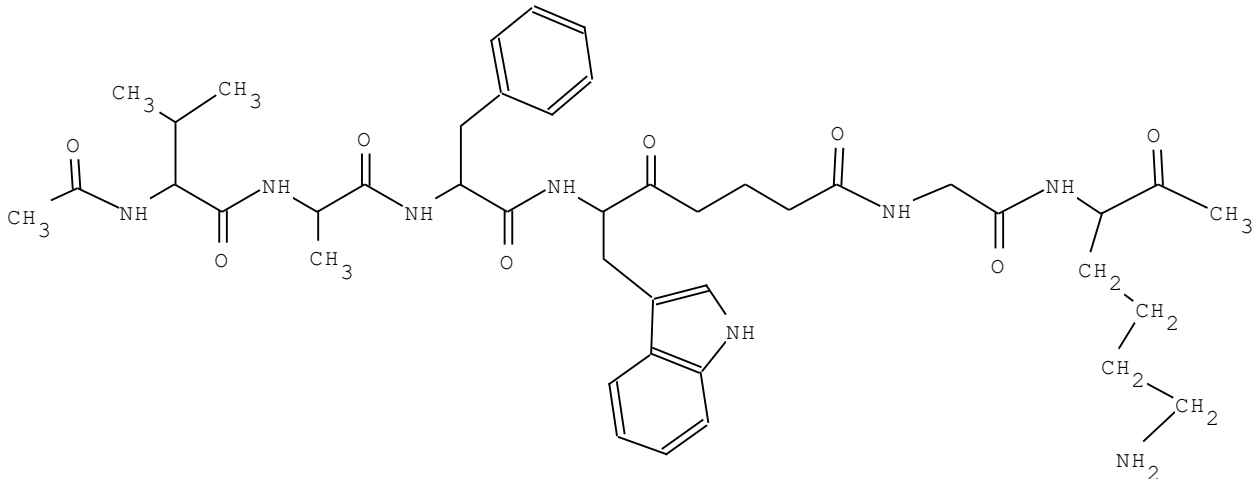
AEKXGMX (SEQ ID NO: 1)

o, de otra forma, con cualquier otro aminoácido en la secuencia del residuo de la posición número 1. Para cada D-aminoácido deberá proporcionarse una clave de caracterización "SITE" y un calificador "NOTEnote" con el nombre completo, no abreviado, del D-aminoácido como valor calificador, por ejemplo, D-alanina y D-norleucina. Además, deberá proporcionarse una clave de caracterización "SITE" y un calificador "NOTEnote" con la abreviatura de L-norleucina como valor calificador, es decir, "Nle", según figura en el Cuadro 4 de la Sección 4 del Anexo I. Por último, deberá proporcionarse una clave de caracterización "REGION" y un calificador "NOTEnote" para indicar que el péptido es circular.

Párrafos pertinentes de la Norma ST.26: 3.a), 7.b), 25, 26, 29, 30 y 31.

Párrafo 3.c) – Definición de “enumeración de sus residuos”

Ejemplo 3.c)-1: Enumeración de aminoácidos mediante la estructura química



Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

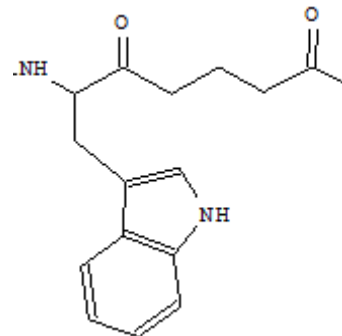
Sí

El péptido enumerado, ilustrado como estructura, contiene como mínimo cuatro aminoácidos específicamente definidos. Por lo tanto, la secuencia deberá incluirse en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia puede ser representada como:

VAFXGK (SEQ ID NO: 2)



donde “X” representa un aminoácido modificado “other” que exige una clave de caracterización “SITE” junto con el calificador “NOTE note”. El calificador “NOTE note” proporciona el nombre completo, no abreviado, del triptófano modificado en la posición 4 del péptido enumerado, por ejemplo, “ácido 6-amino-7-(1H-indol-3-yl)-5-oxo-heptanóico”. Además, se necesitan claves de caracterización adicionales “SITE” y el calificador “NOTE” para indicar la acetilación del N-terminal y la metilación del C-terminal. La metilación del C-terminal cambia la estructura química de la lisina terminal, ya que el -OH del extremo terminal se sustituye por -CH3. Debido a este cambio estructural, la lisina de la secuencia se considera un “aminoácido modificado”. En consecuencia, se requiere la clave de caracterización “SITE” y el calificador “note” para indicar la metilación del C-terminal. Sin embargo, la valina no se considera un “aminoácido modificado”, ya que la adición del grupo acetilo a la valina se realiza con un enlace peptídico convencional. La acetilación no altera la estructura de la valina. En consecuencia, se debe incluir la clave de caracterización “SITE” y el calificador “note” para indicar la acetilación del N-terminal.

Como alternativa, la secuencia puede representarse como:

VAFW (SEQ ID NO: 3)

Se necesita una clave de caracterización “SITE” y un calificador “NOTE note” para indicar la modificación del triptófano en la posición 4 del péptido enumerado, con el valor: “C-terminal unido por medio de un puente de glutaraldehído al dipéptido GK”. Además, se necesita debería incluirse una clave de caracterización adicional “SITE” en la localización 1 y el calificador “NOTE note” para indicar la acetilación del N-terminal.

Párrafos pertinentes de la Norma ST.26: 3.c), 7.b), 29, 30 y 31.

Ejemplo 3.c)-2: Fórmula abreviada de una secuencia de aminoácidos
(G₄Z)_n

donde G= Glicina, z = cualquier aminoácido y la variable n puede ser cualquier número entero.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

En la divulgación se indica que "n" puede ser "cualquier número entero"; por lo tanto, la realización más abarcadora de "n" es indeterminada. Puesto que "n" es indeterminado, el péptido de la fórmula no puede ser expandido a una longitud definida y, por lo tanto, deberá considerarse la fórmula no expandida.

El péptido enumerado en la fórmula no expandida ("n" = 1) provee cuatro aminoácidos específicamente definidos, cada uno de los cuales es Gly, y el símbolo "z". Convencionalmente, "Z" es el símbolo correspondiente a "glutamina o ácido glutámico"; sin embargo, en el ejemplo se define "z" como "cualquier aminoácido". En la Norma ST.26, un aminoácido que no está específicamente definido se representa mediante "X". Sobre la base de ese análisis, el péptido enumerado, es decir, GGGGX, contiene cuatro residuos glicina que están enumerados y específicamente definidos. Así pues, el párrafo 7.b) de la Norma ST.26 exige la inclusión de la secuencia en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia utiliza un símbolo no convencional "z", cuya definición debe quedar determinada por la divulgación (véase la Introducción del presente documento). Puesto que "z" está definido como cualquier aminoácido, el símbolo convencional utilizado para representar ese aminoácido es "X." Por lo tanto, la secuencia deberá representarse como una secuencia única:

GGGGX (SEQ ID NO: 4)

de preferencia anotada y debería anotarse con la clave de caracterización REGION, la localización de característica ">5" (corresponde a >5), con un calificador NOTE "note" con el valor "La totalidad de la secuencia de aminoácidos 1-5 puede repetirse una vez o más."

Con arreglo al párrafo 27, el símbolo "X" se interpretará como cualquiera de los símbolos "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" o "V", excepto cuando se utilice con una descripción detallada en el cuadro de características. Como en este ejemplo "X" representa "cualquier aminoácido", deberá anotarse con la clave de caracterización VARIANT y el calificador "note" con el valor "X puede ser cualquier aminoácido".

Por lo tanto, si se pretende que "X" represente "cualquier aminoácido", debería anotarse con la clave de caracterización "VARIANT" y un calificador NOTE "note" con el valor "X puede ser cualquier aminoácido".

Cuando sea posible, cada "X" deberá anotarse individualmente. Sin embargo, toda región que contiene residuos "X" contiguos, o numerosos residuos "X" dispersos en toda la secuencia, puede describirse en conjunto con la clave de caracterización "VARIANT" utilizando la sintaxis "x..y" como descriptor de localización, donde x e y son las posiciones del primer y último residuo "X", y un calificador NOTE "note" con el valor "X puede ser cualquier aminoácido".

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de PI podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 3.c), 7.b) y 27.

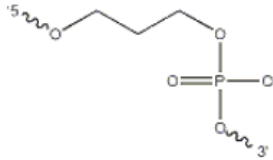
Párrafo 3.g) Definición de “nucleótido”

Ejemplo 3.g)-1: Secuencia de nucleótidos interrumpida por un espaciador C3

En una solicitud de patente se describe la secuencia siguiente:

atgcatgcatgcncggcatgcatgc

donde n = un espaciador C3 con la estructura siguiente:



Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

La secuencia enumerada contiene dos segmentos de nucleótidos específicamente definidos separados por un espaciador C3.

El espaciador C3 no es un nucleótido conforme al párrafo 3.g); el símbolo convencional “n” se utiliza de manera no convencional (véase la Introducción del presente documento). En consecuencia, cada segmento es una secuencia de nucleótidos distinta. Puesto que cada segmento contiene más de 10 nucleótidos específicamente definidos, ambos deben incluirse en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Cada segmento debe incluirse en una lista de secuencias como una secuencia distinta, cada una con su propio identificador de secuencia:

atgcatgcatgc (SEQ ID NO: 5)

cgcatgcatgc (SEQ ID NO: 6)

La citosina en cada segmento que está unida al espaciador C3 debería describirse en más detalle en un cuadro de características utilizando la clave de caracterización “misc_feature” y el calificador “note”. El valor calificador “note”, que es “texto libre”, debería indicar la presencia del espaciador, que está unido a otro ácido nucleico e identificar el espaciador, ya sea mediante su nombre químico completo, no abreviado, ya sea mediante su nombre común, es decir, espaciador C3.

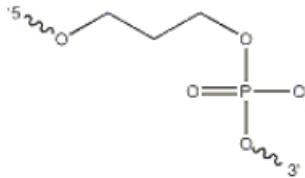
Párrafos pertinentes de la Norma ST.26: 3.g), 7.a) y 15.

Ejemplo 3.g)-2: Secuencia de nucleótidos con residuos alternativos, incluido un espaciador C3

En una solicitud de patente se describe la secuencia siguiente:

atgcatgcatgcncggcatgcatgc

donde n = c, a, g o un espaciador C3 con la estructura siguiente:



Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

Hay 24 residuos específicamente definidos en la secuencia enumerada interrumpida por la variable "n." Deberá consultarse la explicación de la secuencia en la divulgación para determinar si "n" se utiliza de manera convencional o no convencional (véase la Introducción del presente documento).

En la divulgación se indica que n = c, a, g o un espaciador C3. El símbolo "n" es un símbolo convencional utilizado de manera no convencional, puesto que se describe que puede incluir un espaciador C3, que no responde a la definición de nucleótido. También se describe que el símbolo "n" puede incluir "c", "a" o "g"; por lo tanto, la Norma ST.26 exige la inclusión de la secuencia de 25 nucleótidos en una lista de secuencias. Puesto que los dos segmentos separados por el espaciador C3 son secuencias distintas de la secuencia de 25 nucleótidos, también pueden incluirse las dos secuencias de 12 nucleótidos.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

En el ejemplo se indica que "n = c, a, g o un espaciador C3". Como se ha señalado más arriba, un espaciador C3 no es un nucleótido. Con arreglo al párrafo 15, el símbolo "n" no debe utilizarse para representar otra cosa que un nucleótido; por lo tanto, el símbolo "n" no puede representar un espaciador C3 en una lista de secuencias.

En el párrafo 15 también se establece que si correspondiera utilizar un símbolo de ambigüedad, deberá utilizarse el símbolo más restrictivo. El símbolo "v" representa "a or c o g" conforme al Cuadro 1 de la Sección 1 del Anexo I, y es más restrictivo que "n".

Cuando la variable "n" en el ejemplo es c, a, o g, la secuencia única enumerada mediante sus residuos que incluye las realizaciones más divulgadas y, por lo tanto, es la secuencia más abarcadora (véase la Introducción del presente documento) que debe incluirse en una lista de secuencias, es:

atgcatgcatgcvcggcatgcatgc (SEQ ID NO: 7)

Se alienta encarecidamente a incluir cualesquiera secuencias adicionales indispensables para la divulgación o las reivindicaciones de la invención, como se ha señalado en la Introducción del presente documento.

Cuando la variable "n" en el ejemplo es un espaciador C3, la secuencia puede considerarse como dos segmentos distintos de nucleótidos específicamente definidos a cada lado de la variable "n", es decir, atgcatgcatgc (SEQ ID NO: 8); y cggcatgcatgc (SEQ ID NO: 9). Si fuera indispensable para la divulgación o las reivindicaciones, esas dos secuencias deberían incluirse también en la lista de secuencias, cada una con su propio identificador de secuencia.

La citosina en cada segmento que está unida al espaciador C3 debería describirse en más detalle en un cuadro de características utilizando la clave de caracterización "misc_feature" y el calificador "note". El valor calificador "note", que es "texto libre", debería indicar la presencia del espaciador, que está unido a otro ácido nucleico e identificar el espaciador ya sea mediante su nombre químico completo, no abreviado, ya sea mediante su nombre común, es decir, espaciador C3.

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de PI podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 3.g), 7.a) y 15.

Ejemplo 3.g)-3: Sitio abásico

En una solicitud de patente se describe la secuencia siguiente:

gagcattgac-AP-taaggct

donde AP es un sitio abásico

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

Los residuos específicamente definidos de la secuencia enumerada están interrumpidos por un sitio abásico. El lado 5' del sitio abásico contiene 10 nucleótidos y el lado 3' del sitio abásico contiene 7 nucleótidos. En el párrafo 3.g)ii)2) se define un sitio abásico como un "nucleótido" cuando forma parte de una secuencia de nucleótidos. En consecuencia, el sitio abásico de este ejemplo se considera como un "nucleótido" a los fines de determinar si se exige incluir la secuencia en una lista de secuencias, y de qué forma. Por lo tanto, los residuos a cada lado del sitio abásico forman parte de una secuencia única enumerada que contiene un total de 18 nucleótidos, 17 de los cuales son específicamente definidos. Por consiguiente, la secuencia debe ser incluida como secuencia única en una lista de secuencias, según lo exige el párrafo 7.a) de la Norma ST.26.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia debe incluirse en una lista de secuencias como:

gagcattgacntaaggct (SEQ ID NO: 10)

El sitio abásico debe representarse mediante el símbolo "n" y debe describirse en más detalle en un cuadro de características. El medio preferido de anotación está dado por la clave de caracterización "modified_base" y el calificador obligatorio "mod_base" con el valor "OTHER". Debe incluirse un calificador "note" que describa la base modificada como sitio abásico.

Párrafos pertinentes de la Norma ST.26: 3.g), 7.a) y 17.

Ejemplo 3.g)-4: Análogos de ácido nucleico

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

PO₄-tagttcattgactaaggctccccattgact-OH

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

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ – Los distintos residuos que constituyen una secuencia de ANG se consideran nucleótidos con arreglo al párrafo 3.g)i)2) de la Norma ST.26. Es decir que la secuencia tiene más de diez nucleótidos enumerados y “específicamente definidos” y por ello se exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Las secuencias de ANG no tienen un extremo 5' y un extremo 3', sino un extremo 3' y un extremo 2'. El extremo 3', que suele ilustrarse con un grupo fosfato terminal, corresponde al extremo 5' del ADN o el ARN. (Cabe señalar que otros análogos de ácido nucleico pueden corresponder de manera diferente a los extremos 5' y 3' del ADN y el ARN). Conforme al párrafo 11, debe incluirse en una lista de secuencias “en el sentido de izquierda a derecha que imite el sentido 5' a 3'”. Por lo tanto, debe incluirse en una lista de secuencias como:

tagttcattgactaaggctccccattgact (SEQ ID NO: 11)

La secuencia debe describirse en un cuadro de características utilizando la clave de caracterización “modified_base” y el calificador obligatorio “mod_base” con la abreviatura “OTHER”. Debe incluirse un calificador “note” con el nombre completo, no abreviado, de los nucleótidos modificados, como “ácidos nucleicos glicólicos” o “2,3- dihidroxipropil nucleósidos”. Un único elemento INSDFeature puede utilizarse para describir toda la secuencia como un ANG en la que el rango de la INSDFeature_location es “1..30”.

Párrafos pertinentes de la Norma ST.26: 3.d), 3.g), 7.a), 11, 16, 18, 65 y 66.

Párrafo 3.k) Definición de “específicamente definido”

Ejemplo 3.k)-1: Símbolos de ambigüedad de nucleótidos

5' NNG KNG KNG K 3'

N y K son códigos de ambigüedad IUPAC-IUB

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

Los códigos de ambigüedad IUPAC-IUB corresponden a la lista de símbolos de nucleótidos que figuran en el Cuadro 1 de la Sección 1 del Anexo I. Con arreglo al párrafo 3.k), un nucleótido específicamente definido es todo nucleótido distinto a los representados por el símbolo “n” en el Anexo I. Por lo tanto, “K” y “G” son nucleótidos específicamente definidos y “N” no es un nucleótido específicamente definido.

La secuencia enumerada no tiene diez o más nucleótidos específicamente definidos y, por lo tanto, el párrafo 7.a) de la Norma ST.26 no exige que se incluya en una lista de secuencias.

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

Con arreglo al párrafo 8, “Una lista de secuencias no deberá incluir ninguna secuencia que tenga menos de diez nucleótidos específicamente definidos [...]”. La secuencia enumerada no tiene diez o más nucleótidos específicamente definidos; por lo tanto, no deberá incluirse en una lista de secuencias.

Párrafos pertinentes de la Norma ST.26: 3.k), 7.a), 8 y 13.

Ejemplo 3.k)-2: Símbolo de ambigüedad “n” utilizado de manera tanto convencional como no convencional

En una solicitud se divulga la secuencia artificial: 5'-AATGCCGGAN-3'. En la divulgación se establece además que:

- i) en una realización, N es cualquier nucleótido;
- ii) en una realización, N es opcional, pero, de preferencia, es G;
- iii) en una realización, N es K;
- iv) en una realización, N es C.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

La secuencia enumerada contiene 9 nucleótidos específicamente definidos y un símbolo “N.” Debe consultarse la explicación de la secuencia en la divulgación para determinar si el símbolo “N” se utiliza de manera convencional (véase la Introducción del presente documento).

Del examen de las realizaciones divulgadas i) a iv) de la secuencia enumerada se desprende que la realización más abarcadora del símbolo “N” es “cualquier nucleótido”. En la realización más abarcadora, “N” en la secuencia enumerada se utiliza de manera convencional.

En algunas realizaciones, se indica que “N” es un residuo específicamente definido (es decir, “N es C” en la parte iv)). Sin embargo, solo se considera la realización más abarcadora (es decir, “N es cualquier nucleótido”) para determinar si una secuencia debe incluirse en una lista de secuencias. Así pues, la secuencia enumerada que debe evaluarse es 5'-AATGCCGGAN-3'.

Sobre la base de ese análisis, la secuencia enumerada, es decir, AATGCCGGAN, no contiene diez nucleótidos específicamente definidos. Por lo tanto, el párrafo 7.a) de la Norma ST.26 no exige la inclusión de la secuencia en una lista de secuencias, pese al hecho de que “n” también se define como nucleótido específico en algunas realizaciones.

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

La secuencia “AATGCCGGAN” no debe incluirse en una lista de secuencias.

Sin embargo, una secuencia alternativa descrita puede incluirse en una lista de secuencias si el símbolo “N” se sustituye por un nucleótido específicamente definido.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Se alienta **encarecidamente** a que se incluyan secuencias que representan realizaciones que constituyen una parte fundamental de la invención. La inclusión de esas secuencias permite una búsqueda más minuciosa y ofrece información al público acerca de la materia respecto de la cual se solicita una patente.

En el caso del ejemplo que figura más arriba, se recomienda encarecidamente que las tres secuencias adicionales siguientes se incluyan en la lista de secuencias, cada una con su propio identificador:

aatgccggag (SEQ ID NO: 12)

aatgccggak (SEQ ID NO: 13)

aatgccggac (SEQ ID NO: 14)

Si no se incluyen las tres secuencias que figuran más arriba, ni una menos, el nucleótido que sustituye al símbolo “n” debería anotarse para describir las alternativas. Por ejemplo, si de ellas solo se incluye en la lista de secuencias SEQ ID NO: 12, debería utilizarse la clave de caracterización “misc_difference” con la localización de característica “10”, junto con dos calificadores “replace” en los que el valor de uno sería “g” y el del otro sería “c”.

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de PI podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 3.k), 7.a), 8 y 13.

Ejemplo 3.k)-3: Símbolo de ambigüedad “n” utilizado de manera no convencional

En una solicitud se divulga la secuencia: 5'-aatgttggan-3'

donde n es c

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

Con arreglo al párrafo 3.k), un nucleótido “específicamente definido” es todo nucleótido distinto a los representados por el símbolo “n” en el Cuadro I de la Sección I del Anexo I.

En este ejemplo, “n” se utiliza de manera no convencional para representar únicamente “c”. La divulgación no indica que “n” se utilice de manera convencional para representar “todo nucleótido”. Por lo tanto, la secuencia debe interpretarse como si se hubiera utilizado en ella el símbolo convencional equivalente, es decir, “c” (véase la Introducción del presente documento). Por consiguiente, la secuencia enumerada que debe considerarse es:

5'-aatgttggac-3'

Esta secuencia tiene diez nucleótidos específicamente definidos y el párrafo 7.a) de la Norma ST.26 exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia debe incluirse en una lista de secuencias como: aatgttggac (SEQ ID NO: 15)

Párrafos pertinentes de la Norma ST.26: 3.k) y 7.a).

Ejemplo 3.k)-4: Los símbolos de ambigüedad distintos de “n” son “específicamente definidos”

En una solicitud de patente se describe la secuencia siguiente:

5' NNG KNG KNG KAG VCR 3'

donde N, K, V y R son códigos de ambigüedad IUPAC-IUB

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

Los códigos de ambigüedad IUPAC-IUB corresponden a la lista de símbolos de nucleótidos definidos en el Cuadro 1 de la Sección 1 del Anexo I. Con arreglo al párrafo 3.k), un nucleótido “específicamente definido” es todo nucleótido distinto a los representados por el símbolo “n” en el Cuadro 1 de la Sección 1 del Anexo I. Por lo tanto, “K”, “V” y “R” son “nucleótidos específicamente definidos”.

La secuencia tiene 11 nucleótidos enumerados y “específicamente definidos” y el párrafo 7.a) de la Norma ST.26 exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia debe incluirse en una lista de secuencias como:

nngkngkngkagvcr (SEQ ID NO: 16)

Párrafos pertinentes de la Norma ST.26: 3.k), 7.a) y 15.

Ejemplo 3.k)-5: Abreviatura de ambigüedad "Xaa" utilizada de manera no convencional

En una solicitud de patente se describe la secuencia siguiente:

Xaa-Tyr-Glu-Xaa-Xaa-Xaa-Leu

donde Xaa en la posición 1 es cualquier aminoácido, Xaa en la posición 4 es Lys, Xaa en la posición 5 es Gly y Xaa en la posición 6 es leucina o isoleucina.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

El péptido enumerado en la fórmula tiene tres aminoácidos específicamente definidos en las posiciones 2, 3 y 7. El primer aminoácido está representado por una abreviatura convencional, es decir, Xaa, que representa cualquier aminoácido. Sin embargo, el 4º, el 5º y el 6º aminoácido están representados por una abreviatura convencional utilizada de manera no convencional (véase la Introducción del presente documento). Por lo tanto, se consulta la explicación de la secuencia en la divulgación para determinar la definición de "Xaa" en esas posiciones. Puesto que "Xaa" en las posiciones 4 a 6 está indicado como un aminoácido específico, debe interpretarse la secuencia como si en ella se hubieran utilizado las abreviaturas convencionales equivalentes, es decir, Lys, Gly y (Leu o Ile). Por consiguiente, la secuencia contiene cuatro o más aminoácidos específicamente definidos y debe incluirse en una lista de secuencias, según lo exige el párrafo 7.b) de la Norma ST.26.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

En la secuencia se utiliza la abreviatura convencional "Xaa" de manera no convencional. Por lo tanto, debe consultarse la explicación de la secuencia en la divulgación para determinar la definición de "Xaa" en las posiciones 4, 5 y 6. La explicación define "Xaa" como una lisina en la posición 4, una glicina en la posición 5 y una leucina o isoleucina en la posición 6. Los símbolos convencionales para esos aminoácidos son K, G y J respectivamente. Por lo tanto, la secuencia debería representarse en la lista de secuencias como:

XYEKGJL (SEQ ID NO: 17)

Con arreglo al párrafo 27, el símbolo "X" se interpretará como cualquiera de los símbolos "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" o "V", excepto cuando se utilice con una descripción detallada en el cuadro de características. Por lo tanto, si se pretende que "X" represente "en la posición 1 de SEQ ID NO: 17 representa cualquier aminoácido", debería anotarse con la clave de caracterización "VARIANT" y un calificador "NOTE" con el valor "X puede ser cualquier aminoácido".

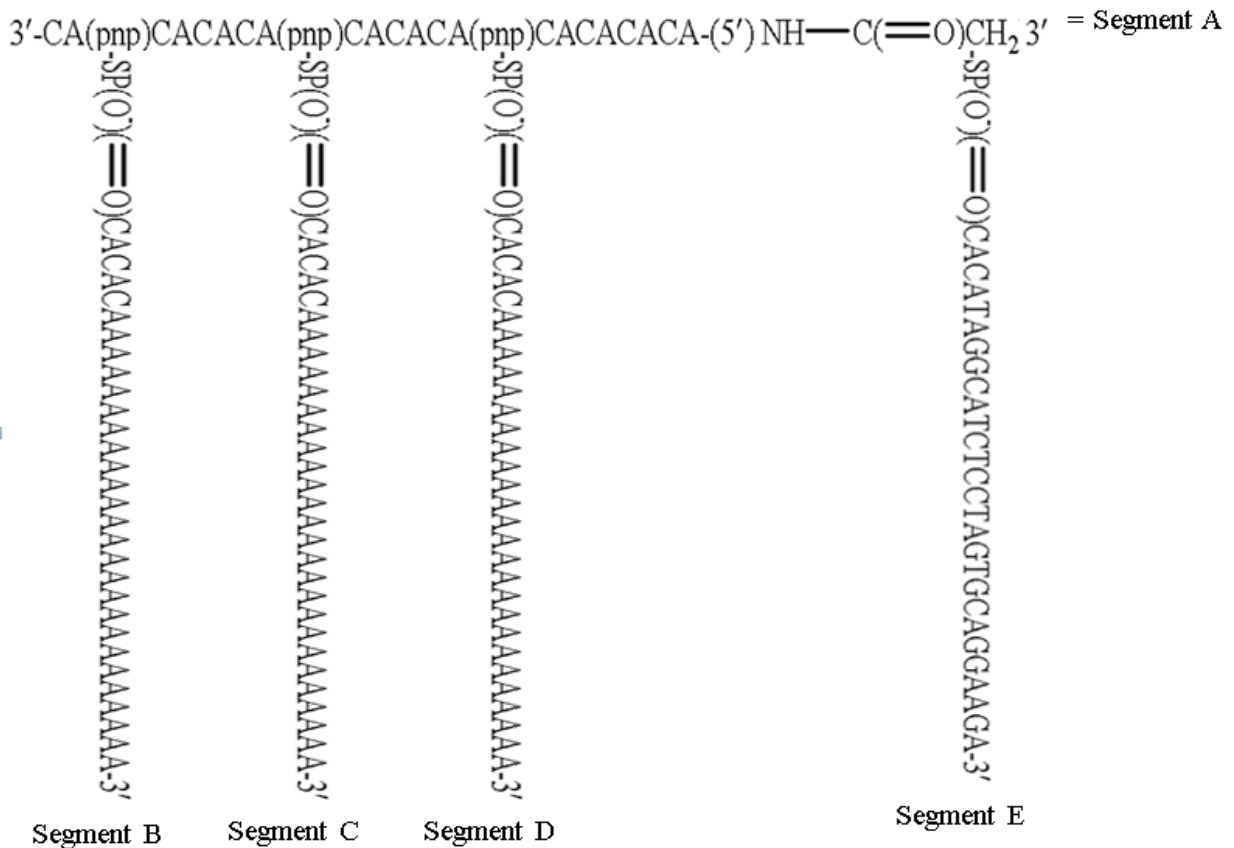
Cuando sea posible, cada "X" debería anotarse individualmente. Sin embargo, toda región que contiene residuos "X" contiguos, o numerosos residuos "X" dispersos en toda la secuencia, puede describirse en conjunto con la clave de caracterización "VARIANT" utilizando la sintaxis "x..y" como descriptor de localización, donde x e y son las posiciones del primer y último residuo "X", y un calificador "NOTE" con el valor "X puede ser cualquier aminoácido".

Párrafos pertinentes de la Norma ST.26: 3.k), 7.b), 26 y 27.

Párrafo 7.a) – Secuencia de nucleótidos exigida en una lista de secuencias

Ejemplo 7.a)-1: Secuencia de nucleótidos ramificada

En la descripción se divulga la siguiente secuencia de nucleótidos ramificada:



donde "pnp" es un enlace o monómero que contiene una funcionalidad del bromoacetilamino;

3'-CA(pnp)CACACA(pnp)CACACA(pnp)CACACACA-(5')NH—C(=O)CH₂ 3' es el segmento A;

SP(O)(=O)CACACAAA 3' es los segmentos B, C y D; y

SP(O)(=O)CACATAGGCATCTCCTAGTGCAGGAAGA 3' es el segmento E.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ – los cuatro segmentos verticales B a E deben incluirse en una lista de secuencias

NO – el segmento horizontal A no debe incluirse en una lista de secuencias

La figura que se reproduce más arriba es un ejemplo de una secuencia de ácido nucleico ramificada de "tipo peine" que contiene cinco segmentos lineales: el segmento horizontal A y los cuatro segmentos verticales B a E.

Con arreglo al párrafo 7.a), las regiones lineales de secuencias de nucleótidos ramificadas que contienen diez o más nucleótidos específicamente definidos, cuando los nucleótidos adyacentes están unidos en el sentido de 3' a 5', deben incluirse en una lista de secuencias.

Los cuatro segmentos verticales B a E contienen, cada uno, más de diez nucleótidos específicamente definidos, y en ellos los nucleótidos adyacentes están unidos en el sentido de 3' a 5'; por lo tanto se exige que cada uno de ellos se incluya en una lista de secuencias.

En el segmento horizontal A, las regiones lineales de la secuencia de nucleótidos están enlazadas por la fracción no nucleótido "pnp" y cada una de esas regiones lineales enlazadas contiene menos de diez nucleótidos específicamente definidos. Por lo tanto, puesto que ninguna región del segmento A contiene diez o más nucleótidos específicamente definidos y ningún nucleótido adyacente está unido en el sentido de 3' a 5', el párrafo 7.a) de la Norma ST.26 no exige que se incluyan en una lista de secuencias.

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

Con arreglo al párrafo 8, "una lista de secuencias no deberá incluir ninguna secuencia que tenga menos de diez nucleótidos específicamente definidos [...]".

Ninguna región del segmento A contiene diez o más nucleótidos específicamente definidos con nucleótidos adyacentes unidos en el sentido de 3' a 5'; por lo tanto, no debe incluirse en una lista de secuencias como secuencia distinta con su propio identificador de secuencia.

Sin embargo, los segmentos B, C, D y E pueden anotarse indicando que están enlazados al segmento A.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Los segmentos B, C y D son idénticos y deben incluirse en una lista de secuencias como secuencia única:

cacacaaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO: 18)

La primera "c" de la secuencia debería describirse en más detalle ~~como un nucleótido modificado~~ utilizando la clave de caracterización "misc_feature" y el calificador "note" con el valor, por ejemplo, "Esta secuencia es una de cuatro ramas de un polinucleótido ramificado."

El segmento E debe incluirse en una lista de secuencias como secuencia única:

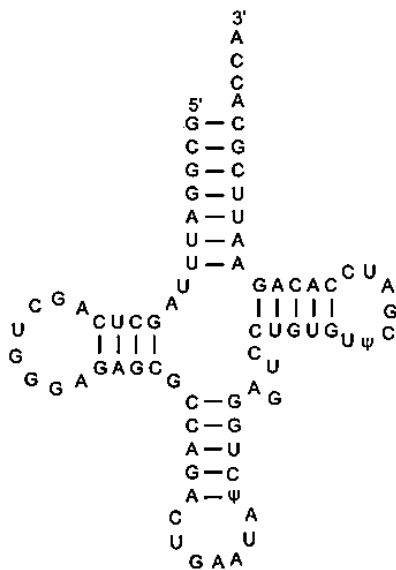
cacataggcatctcctagtcaggaaga (SEQ ID NO: 19)

La primera "c" de la secuencia debería describirse en más detalle ~~como un nucleótido modificado~~ utilizando la clave de caracterización "misc_feature" y el calificador "note" con el valor, por ejemplo, "Esta secuencia es una de cuatro ramas de un polinucleótido ramificado."

Párrafos pertinentes de la Norma ST.26: 7.a), 8, 11, 13 y 17.

Ejemplo 7.a)-2: Secuencia de nucleótidos lineal con una estructura secundaria

En una solicitud de patente se describe la secuencia siguiente:



donde Ψ es pseudouridina.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

La secuencia de nucleótidos contiene 73 nucleótidos enumerados y específicamente definidos. Así pues, en el ejemplo hay 10 o más nucleótidos "específicamente definidos" y, según exige el párrafo 7)a) de la Norma ST.26, debe incluirse en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Al consultar la divulgación se observa que "Ψ" es equivalente a pseudouridina. El único símbolo convencional que puede utilizarse para representar la pseudouridina es "n"; por lo tanto, "Ψ" es un símbolo no convencional utilizado para representar el símbolo convencional "n" (véase la Introducción del presente documento). Por consiguiente, debe interpretarse que la secuencia tiene dos símbolos "n", en lugar de los dos símbolos "Ψ".

El símbolo "u" no debe utilizarse para representar el uracilo en una molécula de ARN en la lista de secuencias. Con arreglo al párrafo 14, el símbolo "t" se interpretará como uracilo en ARN. La secuencia debe incluirse como:

gcggttagctcagctgggagagcgcagactgaatanctggagtcctgtgtncgatccacagaattcgcacca (SEQ ID NO: 20)

El valor del calificador obligatorio "mol_type" de la clave de caracterización obligatoria "source" es "tRNA". Puede proporcionarse información adicional con la clave de caracterización "tRNA" y cualquier calificador o calificadores apropiados.

Los residuos "n" deben describirse en más detalle en un cuadro de características utilizando la clave de caracterización "modified_base" y el calificador obligatorio "mod_base" con la abreviatura "p" para la pseudouridina como valor calificador (véase el Cuadro 2 del Anexo 1).

Párrafos pertinentes de la Norma ST.26: 7.a), 11, 13, 14, 17, 62, 84 y secciones 2 y 5 del Anexo I, clave de caracterización 5.43.

Ejemplo 7.a)-3: Símbolos de ambigüedad de nucleótidos utilizados de manera no convencional

En una solicitud de patente se describe la secuencia siguiente:

5' GATC-MDR-MDR-MDR-MDR-GTAC 3'

La explicación de la secuencia en la divulgación indica asimismo: "Un "DR Element" consiste en la secuencia 5' ATCAGCCAT 3'. Un DR Element mutante, o MDR, es un elemento DR en el que, en el medio, 5 nucleótidos, CAGCC, han mutado a TTTTT."

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

En la secuencia enumerada se utiliza el símbolo "MDR". Cuando no queda claro si al utilizar un símbolo en una secuencia se pretende que sea un símbolo convencional, es decir, un símbolo que figura en el Cuadro 3 de la Sección 3 del Anexo 1, o un símbolo no convencional, debe consultarse la explicación de la secuencia en la divulgación para tomar una decisión a ese respecto (véase la Introducción del presente documento). Con arreglo al Cuadro 3, "MDR" podría interpretarse como tres símbolos convencionales (m = a o c, d = a o g o t/u, r = g o a) o como abreviatura que equivale a la notación abreviada para alguna otra estructura.

Al consultar la divulgación se observa que un elemento MDR es equivalente a 5' ATTTTTTAT 3'. Se considera que las letras "MDR" son símbolos convencionales utilizados de manera no convencional; por lo tanto, debe interpretarse la secuencia como si hubiera sido divulgada utilizando los símbolos convencionales equivalentes. Por consiguiente, la secuencia enumerada que se considera para su inclusión en una lista de secuencias es:

5' GATC ATTTTTTAT ATTTTTTAT ATTTTTTAT ATTTTTTAT GTAC 3'

La secuencia enumerada tiene 44 nucleótidos específicamente definidos y el párrafo 7.a) de la Norma ST.26 exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia debe incluirse en una lista de secuencias como:

gatcatttttatatttttatatttttatatttttatgtac (SEQ ID NO: 21)

Párrafos pertinentes de la Norma ST.26: 7.a) y 13.

Ejemplo 7.a)-4: Símbolos de ambigüedad de nucleótidos utilizados de manera no convencional

En una solicitud de patente se describe la secuencia siguiente:

5' ATTC-N-N-N-N-GTAC 3'

La explicación de la secuencia en la divulgación indica asimismo que "N" consiste en la secuencia 5' ATACGCACT 3'.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

En la secuencia enumerada se utiliza el símbolo "N". Debe consultarse la explicación de la secuencia en la divulgación para determinar si "N" se utiliza de manera convencional o no convencional (véase la Introducción del presente documento).

Al consultar la divulgación se observa que "N" es equivalente a 5' ATACGCACT 3'. Así pues, "N" es un símbolo convencional utilizado de manera no convencional. Por consiguiente, la secuencia debe interpretarse como si se hubiera divulgado utilizando los símbolos convencionales equivalentes:

5' ATTC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3'

La secuencia enumerada tiene 44 nucleótidos específicamente definidos y el párrafo 7.a) de la Norma ST.26 exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia debe incluirse en una lista de secuencias como:

atcctacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO: 22)

Párrafos pertinentes de la Norma ST.26: 7.a) y 13.

Ejemplo 7.a)-5: Símbolos no convencionales de nucleótidos

En una solicitud de patente se describe la secuencia siguiente:

5' GATC-β-β-β-β-GTAC 3'

La explicación de la secuencia en la divulgación indica asimismo que "β" consiste en la secuencia 5' ATACGCACT 3'.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

En la secuencia enumerada se utiliza el símbolo no convencional "β". Debe consultarse la explicación de la secuencia en la divulgación para determinar el significado de "β" (véase la Introducción del presente documento).

Al consultar la divulgación se observa que "β" es equivalente a 5' ATACGCACT 3'. Así pues, "β" es un símbolo no convencional utilizado para representar una secuencia de nueve símbolos convencionales específicamente definidos. Por consiguiente, la secuencia debe interpretarse como si se hubiera divulgado utilizando los símbolos convencionales equivalentes:

5' GATC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3'

La secuencia enumerada tiene 44 nucleótidos específicamente definidos y el párrafo 7.a) de la Norma ST.26 exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia debe incluirse en una lista de secuencias como:

gatcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO: 23)

Párrafos pertinentes de la Norma ST.26: 7.a) y 13.

Ejemplo 7.a)-6: Símbolos no convencionales de nucleótidos

En una solicitud de patente se describe la secuencia siguiente:

5' GATC-β-β-β-β-GTAC 3'

La explicación de la secuencia en la divulgación indica asimismo que "β" es igual a adenina, inosina, o pseudouridina.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

En la secuencia enumerada se utiliza el símbolo no convencional "β". Debe consultarse la explicación de la secuencia en la divulgación para determinar el significado de "β" (véase la Introducción del presente documento).

Al consultar la divulgación se observa que "β" es equivalente a adenina, inosina, o pseudouridina. El único símbolo convencional que puede utilizarse para representar "adenina, inosina, o pseudouridina" es "n"; por lo tanto, "β" es un símbolo no convencional utilizado para representar el símbolo convencional "n". Por consiguiente, debe interpretarse que la secuencia tiene cuatro símbolos "n" (mostrados como "N" a continuación) en lugar de los cuatro símbolos "β":

5' GATC-N-N-N-N-GTAC 3'

La secuencia enumerada tiene solo ocho nucleótidos específicamente definidos y el párrafo 7.a) de la Norma ST.26 no exige su inclusión en una lista de secuencias.

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

La secuencia enumerada, 5' GATC-N-N-N-N-GTAC 3' no debe incluirse en una lista de secuencias.

Sin embargo, una secuencia alternativa divulgada puede incluirse en una lista de secuencias si al menos 2 de los símbolos "n" son sustituidos por adenina, dando como resultado una secuencia con al menos 10 o más nucleótidos específicamente definidos.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?


Una posible representación permitida es:

gatcaaaagtac (SEQ ID NO: 24)

En el ejemplo que figura más arriba, los cuatro nucleótidos de adenina que reemplazan los símbolos β deberían anotarse para señalar que esas posiciones podrían ser sustituidas por inosina o pseudouridina.

Debería utilizarse la clave de caracterización "misc_difference" con una localización de característica 5-8 y un calificador "note" con el valor, por ejemplo, "Un nucleótido en cualquiera de las posiciones entre 5 y 8 puede sustituirse por inosina o pseudouridina". Puesto que esas alternativas son nucleótidos modificados, será necesario utilizar la clave de caracterización "modified_base" junto con el calificador "mod_base". El valor para el calificador "mod_base" puede ser "OTHER" junto con un calificador "note" y el valor "i o p".

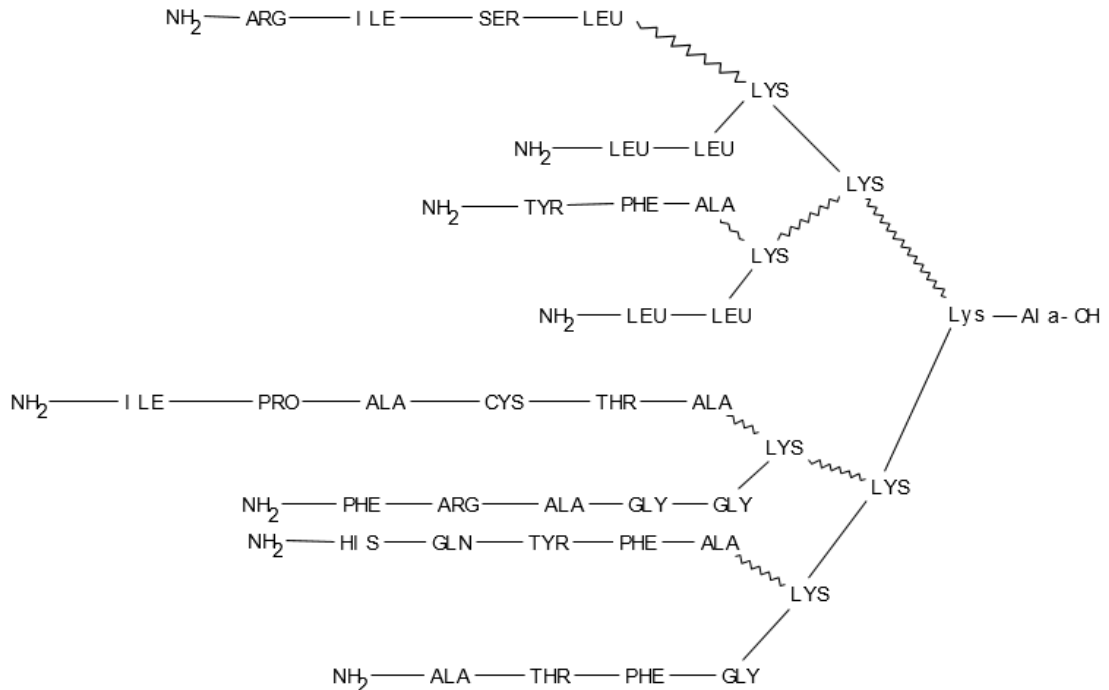
Es posible realizar otras sustituciones.

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de  podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 7.a), 8, 13 y 17.

Ejemplo 7.b)-2: Secuencia de aminoácidos ramificada

En la solicitud se describe una secuencia ramificada en la cual los residuos lisina se utilizan como andamiaje para formar ocho ramas a las que están unidas múltiples cadenas lineales de péptidos. La lisina es un aminoácido dibásico, que provee dos sitios para la unión de péptidos. El péptido se ilustra de la manera siguiente:

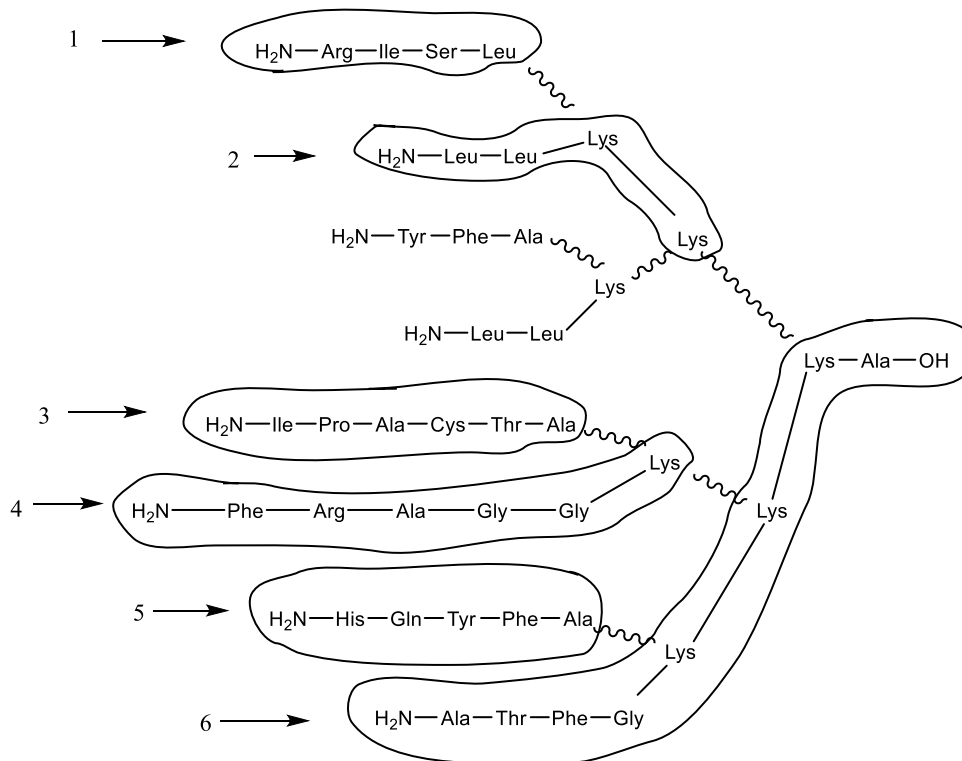


En el péptido ramificado que figura más arriba, las uniones entre la lisina y otro aminoácido indicadas mediante ~~—~~ representan un enlace amida entre la terminal amina de la lisina y el extremo carboxilo del aminoácido unido. Las uniones indicadas mediante ~~~~~ representan un enlace amida entre el amino en la cadena lateral de la lisina y el extremo carboxilo del aminoácido unido.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

En el ejemplo se divulga una secuencia ramificada en la que los residuos lisina se utilizan como andamiaje. El párrafo 7.b) exige que la región no ramificada o lineal de la secuencia, que contiene cuatro o más aminoácidos específicamente definidos, se incluya en una lista de secuencias. Las regiones lineales del péptido ramificado en el ejemplo que figura más arriba, que tienen cuatro o más aminoácidos específicamente definidos, figuran más abajo rodeadas por un círculo:



El párrafo 7.b) de la Norma ST.26 exige la inclusión de los péptidos 1 a 6, más arriba, en una lista de secuencias.

Los péptidos cuya inclusión en la lista de secuencias no se exige son los siguientes:

YFA

LLK

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

Con arreglo al apartado 8, una lista de secuencias no deberá incluir ninguna secuencia que tenga menos de cuatro aminoácidos específicamente definidos.

Los péptidos YFA y LLK contienen, cada uno, solo tres aminoácidos específicamente definidos y, por lo tanto, no deberán incluirse en una lista de secuencias como secuencias distintas, cada una con su propio número de identificación.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Los péptidos 1 a 6 deben representarse con identificadores de secuencia distintos:

RISL (SEQ ID NO: 26)

LLKK (SEQ ID NO: 27)

IPACTA (SEQ ID NO: 28)

FRAGGK (SEQ ID NO: 29)

HQYFA (SEQ ID NO: 30)

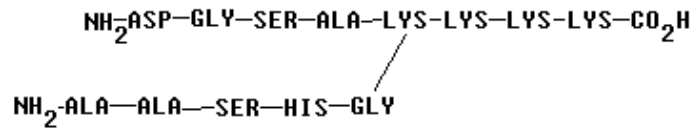
ATFGKKKA (SEQ ID NO: 31)

El enlace cruzado **se anota, de preferencia, puede anotarse** utilizando la clave de caracterización "SITE" y el calificador obligatorio "NOTE" **note** con el valor, por ejemplo, "Esta secuencia es una parte de una secuencia de aminoácidos ramificada". Con arreglo al párrafo 29.30 de la norma ST.26, las secuencias SEQ ID NO: 27, 29 y 31 deben incluir una anotación por cada lisina para indicar que es un aminoácido modificado, utilizando la clave de caracterización "SITE" junto con el calificador "NOTE" **calificador "note"** que describa que la cadena lateral de la lisina está unida por medio de un enlace amida a otra secuencia. **Preferiblemente, cada Cada** una de las secuencias SEQ ID NO: 26, 28 y 30 debería incluir una anotación para indicar que el aminoácido C-terminal está unido a otra secuencia, utilizando la clave de caracterización "SITE" junto con el calificador "NOTE" **calificador "note"**.

Párrafos pertinentes de la Norma ST.26: 7.b), 8, 26, 29, 30 y 31.

Ejemplo 7.b)-3: Secuencia de aminoácidos ramificada

Péptido de la secuencia siguiente:



El enlace entre el residuo glicina terminal en la secuencia inferior y la lisina en la secuencia superior se hace mediante una unión amida entre el carboxi-terminal de la glicina y el amino en la cadena lateral de la lisina.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

La región no ramificada o lineal de una secuencia que contiene cuatro o más aminoácidos específicamente definidos, debe incluirse en una lista de secuencias. En el ejemplo que figura más arriba, las regiones lineales del péptido ramificado que tienen más de cuatro aminoácidos son:



El párrafo 7.b) de la Norma ST.26 exige la inclusión de las secuencias 1 y 2 en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Las secuencias 1 y 2 deben representarse con identificadores de secuencia distintos:

DGSAKKK (SEQ ID NO: 32)

AASHG (SEQ ID NO: 33)

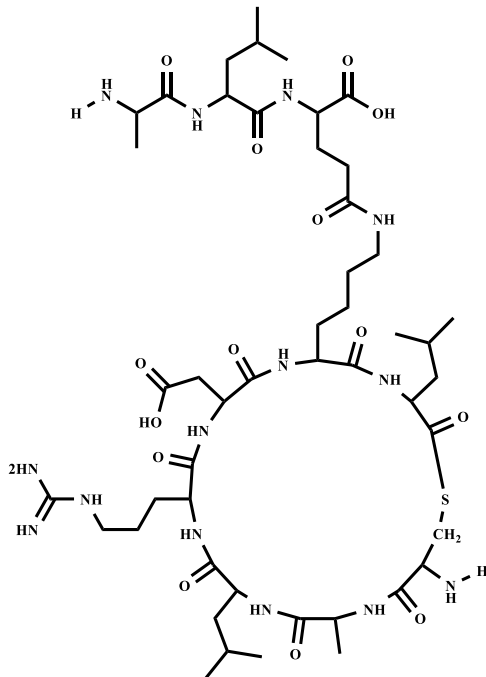
La secuencia DGSAKKK debe incluir una anotación para indicar que la lisina en la posición 5 es un aminoácido modificado, utilizando la clave de caracterización "SITE" junto con el calificador "NOTE note" que describa que la cadena lateral de la lisina está enlazada a otra secuencia por medio de una conexión amida. De preferencia, la secuencia AASHG debería incluir una anotación para indicar que la glicina en la posición 5 está enlazada a otra secuencia, utilizando la clave de caracterización "SITE" junto con el calificador "NOTE note".

La secuencia DSAKKK debería incluir una anotación para indicar que la 5ª lisina es un aminoácido modificado, utilizando la clave de caracterización "SITE" junto con el calificador "NOTE note" que describa que la lisina enlaza el péptido AASHG. De preferencia, la secuencia AASHG debería incluir una anotación para indicar que la 5ª glicina está enlazada a DGSAKKK, utilizando la clave de caracterización "SITE" junto con el calificador "NOTE note".

Párrafos pertinentes de la Norma ST.26: 7.b), 26, 29, 30 y 31.

Ejemplo 7.b)-4: Péptido cíclico que contiene una secuencia de aminoácidos ramificada

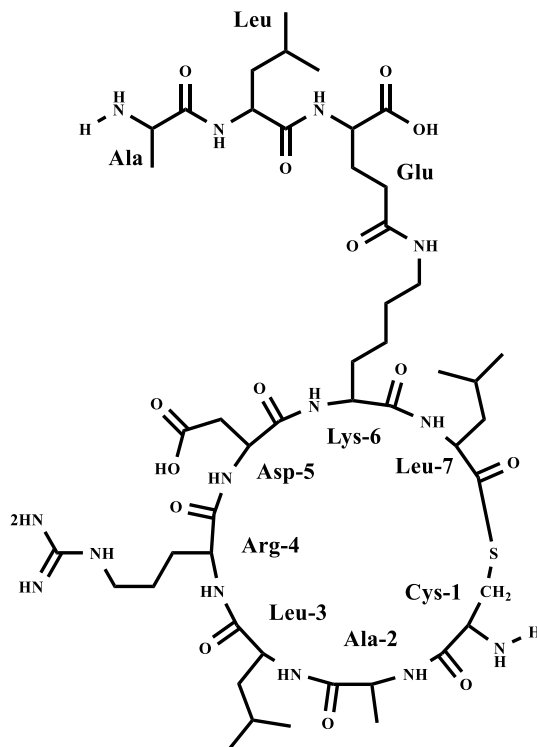
En una solicitud de patente se describe la secuencia siguiente:



La cisteína y la leucina de la estructura cíclica están enlazadas por medio de la cadena lateral de la cisteína y del terminal carboxilo de la leucina.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

La estructura que se muestra a continuación corresponde a una secuencia cíclica ramificada de aminoácidos que contiene los siguientes aminoácidos:



Puesto que la cadena lateral de la cisteína y el carboxilo terminal de la leucina intervienen en la ciclación, el extreme N-terminal del péptido cíclico se encuentra en la posición Cys-1.

SÍ - la región cíclica del péptido

El apartado b) del párrafo 7 de la Norma ST.26 exige que toda región lineal de una secuencia ramificada que contenga cuatro o más aminoácidos específicamente definidos, en la que los aminoácidos formen un único esqueleto peptídico, se incluya en una lista de secuencias. En el ejemplo anterior, la región cíclica del péptido ramificado tiene más de cuatro aminoácidos y, por lo tanto, debe incluirse en una lista de secuencias.

NO - la ramificación tripéptidica del péptido

No se exige la inclusión de la ramificación tripeptídica Ala-Leu-Glu en la lista de secuencias.

Pregunta 2: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

Con arreglo al párrafo 8, una lista de secuencias no deberá incluir ninguna secuencia que tenga menos de cuatro aminoácidos específicamente definidos.

La ramificación tripéptidica contiene solo tres aminoácidos específicamente definidos y, por lo tanto, no deberá incluirse en una lista de secuencias como una secuencia distinta, con su propio identificador de secuencia.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Si bien este ejemplo ilustra un péptido de configuración circular, el anillo no está compuesto exclusivamente de residuos de aminoácidos unidos por enlaces peptídicos, como se indica en el párrafo 25. Puesto que la ciclación de la secuencia de aminoácidos se produce por medio de la cadena lateral de la cisteína (Cys) y el carboxilo terminal de la leucina (Leu), a la cisteína se le debe asignar la posición 1 dentro de la región cíclica del péptido. Por consiguiente, la secuencia debe representarse como:

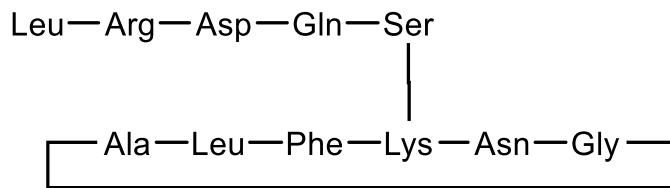
CALRDKL (SEQ ID NO: 89)

Como se indica en la figura anterior, la secuencia de aminoácidos se cicla mediante un enlace tioéster entre la cadena lateral de la cisteína y el carboxilo terminal de la leucina. Debe utilizarse la clave de caracterización "SITE" para describir la cisteína modificada que forma el enlace entre cadenas con la leucina. El elemento localización de características son los números de residuo de aminoácidos con enlaces cruzados en formato "x.y", es decir, "1..7". El calificativo obligatorio "NOTE" "note" debería indicar la naturaleza del enlace, por ejemplo, "cysteine leucine thioester (Cys-Leu)", para especificar que las posiciones Cys-1 y Leu-7 están enlazadas por medio de un enlace tioéster. Además, debe anotarse la lisina en la posición 6 para indicar que se ha modificado, utilizando la clave de caracterización "SITE" junto con el calificativo obligatorio "NOTE" "note" cuando el valor calificador describa que la cadena lateral de la lisina enlaza el tripéptido ALE.

Párrafos de la Norma ST.26: 7 b), 8, 25, 26, 29, 30, 31, 66.c) y 70

Ejemplo 7.b)-5: Péptido cíclico que contiene una secuencia de aminoácidos ramificada

En una solicitud de patente se divulga el siguiente péptido cíclico ramificado:

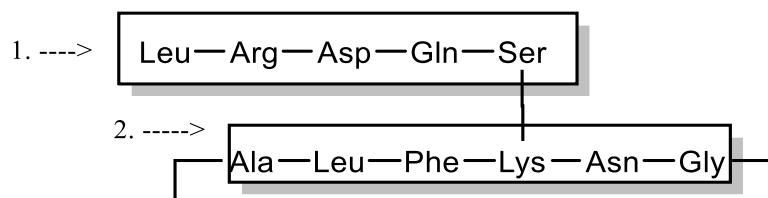


La serina y la lisina están enlazadas por medio de un enlace amídico entre el carboxi-terminal de la serina y la amina de la cadena lateral de la lisina.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

El párrafo 7.b) exige la inclusión de cualquier secuencia que contenga cuatro o más aminoácidos específicamente definidos y que pueda representarse como una región lineal de una secuencia ramificada en una lista de secuencias. En el ejemplo anterior, el péptido contiene una región cíclica en la que los aminoácidos están unidos por enlaces peptídicos, y una región ramificada que se une a una cadena lateral del lisina en la región cíclica. Las regiones de este péptido ramificado que pueden representarse como lineales y que contienen cuatro o más aminoácidos específicamente definidos son:



La Norma ST.26 exige la inclusión de las secuencias 1 y 2 de este péptido cíclico ramificado en una lista de secuencias, cada una con su propio identificador de secuencia.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia 1 debe representarse como:

LRDQS (SEQ. ID. NO: 90)

De preferencia, la secuencia se anota **1 puede anotarse** utilizando la clave de caracterización "SITE" junto con el **calificativo "NOTE" calificador "note"** para describir que la serina en la posición 5 está unida a otra secuencia por medio de un enlace amídico entre una serina y la cadena lateral de una lisina de la otra secuencia.

La secuencia 2 es un péptido cíclico. El párrafo 25 indica que, cuando una secuencia de aminoácidos tiene una configuración circular y carece de amino y carboxiterminales, el solicitante debe elegir el residuo de aminoácidos para la posición 1. Por consiguiente, la secuencia puede representarse como:

ALFKNG (SEQ. ID. NO: 91)

Alternativamente, cualquier otro aminoácido de la secuencia podría designarse como residuo de la posición 1. La secuencia ALFKNG debe describirse más detalladamente utilizando la clave de caracterización "SITE" junto con el **calificativo "NOTE" calificador "note"** para describir que la cadena lateral de la lisina del residuo de la posición 4 está unida mediante un enlace amídico a otra secuencia. Este enlace de la cadena lateral modifica la lisina y, conforme con el párrafo 30 de la Norma ST.26, un aminoácido modificado debe describirse más detalladamente en el cuadro de características. Además, debe proporcionarse una clave de caracterización "REGION" y un **calificativo "NOTE" calificador "note"** para indicar que el péptido ALFKNG es circular.

Párrafos de la Norma ST.26: 7.b), 25, 26, 30 y 31

Párrafo 11.a) – Secuencia de nucleótidos de doble cadena - totalmente complementarias

Ejemplo 11.a)-1: Secuencia de nucleótidos de doble cadena – longitudes iguales

En una solicitud de patente se describe la siguiente secuencia de ADN de doble cadena:

3' -CCGGTTAACGCTA-5'

5' -GGCCAATTGCGAT-3'

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

Cada una de las secuencias de nucleótidos enumeradas tiene más de 10 nucleótidos específicamente definidos. Al menos una cadena debe incluirse en la lista de secuencias, porque las dos cadenas de esta secuencia de nucleótidos de doble cadena son totalmente complementarias entre sí.

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

Si bien debe incluirse en la lista de secuencias la secuencia de una sola cadena, pueden incluirse las secuencias de ambas cadenas, cada una con su identificador de secuencia.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia de ADN de doble cadena debe representarse o bien como secuencia única o bien como dos secuencias distintas. Cada secuencia incluida en la lista de secuencias debe estar representada en sentido 5' a 3' y debe atribuírsele su propio identificador de secuencia.

atcgcaattggcc (cadena superior) (SEQ ID NO: 34)

y/o

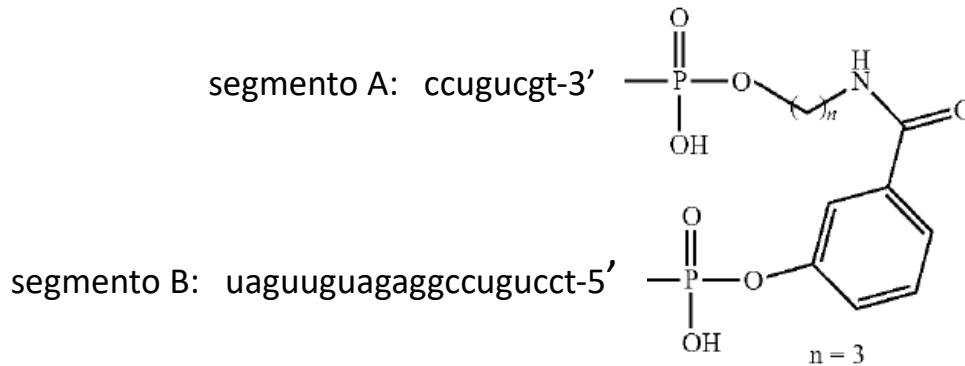
ggccaattgcat (cadena inferior) (SEQ ID NO: 35)

Párrafos de la Norma ST.26: 7.a), 11.a) y 13.

Párrafo 14 – Símbolo “t” interpretado como uracilo en ARN

Ejemplo 14-1: El símbolo “t” representa el uracilo en ARN

En una solicitud de patente se describe el compuesto siguiente:



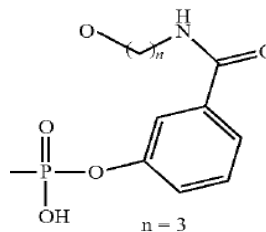
donde el segmento A y el segmento B son secuencias de ARN.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ – segmento B

NO – segmento A

La secuencia enumerada contiene dos segmentos de nucleótidos específicamente definidos separados por la siguiente estructura “de unión”:



La estructura de unión no es un nucleótido conforme al párrafo 3.g); por lo tanto, cada segmento debe considerarse como una secuencia distinta. El segmento B contiene más de 10 nucleótidos específicamente definidos y el párrafo 7.a) de la Norma ST.26 exige su inclusión en una lista de secuencias. El segmento A contiene tan solo ocho nucleótidos específicamente definidos y por lo tanto no se exige su inclusión en una lista de secuencias.

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

El segmento A contiene menos de 10 nucleótidos específicamente definidos y, conforme al párrafo 8 de la Norma ST.26, no debe incluirse en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

El segmento B es una molécula de ARN; por lo tanto, el elemento "INSDSeq_moltype" debe ser "ARN." El símbolo "u" no debe utilizarse para representar uracilo en una molécula de ARN en una lista de secuencias. Con arreglo al párrafo 14, el símbolo "t" se interpretará como uracilo en ARN. Por consiguiente, el segmento B debe incluirse en la lista de secuencias como:

tcctgtccggagatggtgat (SEQ ID NO: 40)

La timina en ARN se considera como un nucleótido modificado, es decir, uracilo modificado, y debe representarse en la secuencia como "t" y describirse en más detalle en un cuadro de características. En consecuencia, la timina en posición 1 debe describirse en más detalle utilizando la clave de caracterización "modified_base", el calificador "mod_base" con "OTHER" como valor calificador y un calificador "note" con "timina" como valor calificador.

La timina, es decir, el uracilo modificado, en posición 1 también debería describirse en más detalle en un cuadro de características utilizando la clave de caracterización "misc_feature" y un calificador "note". El oxígeno 5' de la timidina está unido por medio del enlazador ácido (4-(3-hidroxibenzamida) butil) fosfínico a otra secuencia de nucleótidos. Cuando sea posible, la otra secuencia podrá indicarse directamente como un valor del calificador "note".

Párrafos pertinentes de la Norma ST.26: 3.g), 7.a), 8, 13, 14, 19 y 54

Párrafo 27 – Debería utilizarse el símbolo de ambigüedad de aminoácidos más restrictivo

Ejemplo 27-1: Fórmula abreviada de un aminoácido

(GGGz)₂

donde z es cualquier aminoácido.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

La secuencia está divulgada como fórmula. (GGGz)₂ es simplemente una forma abreviada de representar la secuencia GGGzGGGz. Convencionalmente, una secuencia se expande, en primer lugar, y luego se determina la definición de cualquier variable, es decir, "z".

En la secuencia se utiliza el símbolo no convencional "z". La definición de "z" debe determinarse a partir de la explicación de la secuencia en la divulgación, que define ese símbolo como cualquier aminoácido (véase la Introducción del presente documento). En el ejemplo no se prevén limitaciones respecto de "z", por ejemplo, que sea igual cada vez que aparece. **El péptido del ejemplo tiene ocho aminoácidos enumerados, seis de los cuales son residuos de glicina específicamente definidos, y los dos restantes son la variable "z" que debería representarse en esta secuencia utilizando el símbolo convencional "X". El párrafo 7.b) de la Norma ST.26 exige la inclusión de la secuencia en una lista de secuencias como secuencia única con un único identificador.**

Cabe señalar que la secuencia sigue estando comprendida en el párrafo 7.b) a pesar del hecho de que los residuos enumerados y específicamente definidos no sean contiguos.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

En la secuencia se utiliza el símbolo no convencional "z" que, conforme a la divulgación, es un aminoácido. El símbolo convencional utilizado para representar "todo aminoácido" es "X". Por lo tanto, la secuencia debe representarse como la secuencia única expandida:

GGGXGGGX (SEQ ID NO: 41)

Con arreglo al párrafo 27, el símbolo "X" se interpretará como cualquiera de los símbolos "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" o "V", excepto cuando se utilice con una descripción detallada en el cuadro de características. **Per lo tanto, si se pretende que "Como en este ejemplo "X" represente a "cualquier aminoácido", debería anotarse con la clave de caracterización "VARIANT" y un calificador "NOTE note" con el valor "X puede ser cualquier aminoácido".**

Cuando sea posible, cada "X" debería anotarse individualmente. Sin embargo, toda región que contiene residuos "X" contiguos, o numerosos residuos "X" dispersos en toda la secuencia, puede describirse en conjunto con la clave de caracterización "VARIANT" utilizando la sintaxis "x..y" como descriptor de localización, donde x e y son las posiciones del primer y último residuo "X", y un calificador "NOTE note" con el valor "X puede ser cualquier aminoácido".

Además, en el ejemplo no se divulga que "z" sea el mismo aminoácido en ambas posiciones en la secuencia expandida. Sin embargo, si "z" se divulga como el mismo aminoácido en ambas posiciones, deberá proporcionarse la clave de caracterización "VARIANT" y un calificador "NOTE note" declarando que "X" en las posiciones 4 y 8 puede ser cualquier aminoácido, siempre y cuando se trate del mismo en ambas posiciones.

Párrafos pertinentes de la Norma ST.26: 3.c), 7.b) y 27.

Ejemplo 27-2: Fórmula abreviada – menos de cuatro aminoácidos específicamente definidos

Un péptido de la fórmula (Gly-Gly-Gly-z)_n

En la divulgación se declara asimismo que z es cualquier aminoácido y

- i) la variable n tiene cualquier longitud; o
- ii) la variable n es 2-100, de preferencia 3

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

Del examen de las dos realizaciones divulgadas, i) y ii), del péptido enumerado de la fórmula, se desprende que "n" puede ser de "cualquier longitud"; por lo tanto, la realización más abarcadora de "n" es indeterminada. Puesto que "n" es indeterminado, el péptido de la fórmula no puede expandirse hasta una longitud determinada y, por lo tanto, debe considerarse la fórmula no expandida.

El péptido enumerado en la fórmula no expandida ("n" = 1) tiene tres aminoácidos específicamente definidos, cada uno de los cuales es Gly, y el símbolo "z". Convencionalmente "Z" es el símbolo correspondiente a "glutamina o ácido glutámico"; sin embargo, en el ejemplo se define "z" como "cualquier aminoácido" (véase la Introducción del presente documento). En virtud de la Norma ST.26, un aminoácido que no esté específicamente definido está representado mediante "X". A la luz de este análisis, el péptido enumerado, es decir, GGGX, no contiene cuatro aminoácidos específicamente definidos. Por lo tanto, el párrafo 7.b) de la Norma ST.26 no exige su inclusión, a pesar del hecho de que "n" también está definido como valor numérico específico en algunas realizaciones.

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

En el ejemplo se proporciona un valor numérico específico para la variable "n," es decir, un límite inferior de 2, un límite superior de 100 y un valor exacto de 3. Toda secuencia que contenga como mínimo cuatro aminoácidos específicamente definidos puede incluirse en la lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Se prefiere una secuencia que contenga 100 copias de GGGX (SEQ ID NO: 42). En una anotación adicional debería indicarse que hasta 98 copias de GGGX podrían suprimirse. Se alienta encarecidamente a incluir realizaciones específicas adicionales que constituyan una parte fundamental de la invención.

Con arreglo al párrafo 27, el símbolo "X" se interpretará como cualquiera de los símbolos "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" o "V", excepto cuando se utilice con una descripción más detallada en el cuadro de características. **Por lo tanto, si se pretende que "X" represente " en SEQ ID NO: 42 representa "cualquier aminoácido", debería anotarse con la clave de caracterización VARIANT y un calificador NOTE "note" con el valor "X puede ser cualquier aminoácido".**

Cuando sea posible, cada "X" debería anotarse individualmente. Sin embargo, toda región que contiene residuos "X" contiguos, o numerosos residuos "X" dispersos en toda la secuencia, pueden describirse conjuntamente con la clave de característica VARIANT utilizando la sintaxis "x.y" como descriptor de localización, donde x e y son las posiciones del primer y último residuo "X", y un calificador **NOTE "note"** con el valor "X puede ser cualquier aminoácido".

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de **P.I.P.I** podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 3.c), 7.b), 26 y 27.

Ejemplo 27-3: Fórmula abreviada - cuatro o más aminoácidos específicamente definidos

Un péptido de la fórmula (Gly-Gly-Gly-z)_n

donde z es cualquier aminoácido y la variable n es 2-100, de preferencia, 3.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

El péptido enumerado de la fórmula tiene tres aminoácidos específicamente definidos, cada uno de los cuales es Gly, y el símbolo "z". Convencionalmente "Z" es el símbolo correspondiente a "glutamina o ácido glutámico"; sin embargo, en la descripción de este ejemplo, "z" se define como "cualquier aminoácido" (véase la Introducción del presente documento). En virtud de la Norma ST.26, un aminoácido que no esté específicamente definido está representado mediante "X". A la luz de este análisis, el péptido repetido enumerado no contiene cuatro aminoácidos específicamente definidos. Sin embargo, en la descripción se prevé un valor numérico específico para la variable "n," es decir, un límite inferior de 2 y un límite superior de 100. Por lo tanto, en el ejemplo se divulga un péptido que tiene al menos seis aminoácidos específicamente definidos en la secuencia GGGzGGGz, y la Norma ST.26 exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Puesto que "z" representa cualquier aminoácido, el símbolo convencional utilizado para representar el cuarto y el octavo aminoácidos es "X."

La norma ST.26 exige la inclusión en una lista de secuencias tan solo de la secuencia única que ha sido enumerada mediante sus residuos. Por lo tanto, al menos una secuencia que contenga 2, 3 o 100 copias de GGGX debe incluirse en la lista de secuencias; sin embargo, se prefiere la secuencia más abarcadora que contenga 100 copias de GGGX (SEQ ID NO: 42) (véase la Introducción del presente documento). En este último caso, podría indicarse en una anotación adicional que hasta 98 copias de GGGX podrían eliminarse. Se alienta encarecidamente a incluir dos secuencias adicionales que contengan 2 y 3 copias de GGGX, respectivamente (SEQ ID NO: 44-45).

Con arreglo al párrafo 27, el símbolo "X" se interpretará como cualquiera de los símbolos "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" o "V", excepto cuando se utilice con una descripción detallada en el cuadro de características. Por lo tanto, si se pretende que "Como en este ejemplo" "X" represente a "cualquier aminoácido", debería anotarse con la clave de caracterización "VARIANT" y un calificador "NOTE note" con el valor "X puede ser cualquier aminoácido".

Cuando sea posible, cada "X" debería anotarse individualmente. Sin embargo, toda región que contiene residuos "X" contiguos, o numerosos residuos "X" dispersos en toda la secuencia, puede describirse en conjunto con la clave de caracterización "VARIANT" utilizando la sintaxis "x..y" como descriptor de localización, donde x e y son las posiciones del primer y último residuo "X", y un calificador "NOTE note" con el valor "X puede ser cualquier aminoácido".

Por otra parte, en el ejemplo no se divulga que la variable "z" sea la misma las dos veces que aparece, en la secuencia expandida. Sin embargo, si "z" se divulga como el mismo aminoácido en todas las localizaciones, una clave de caracterización VARIANT y un calificador NOTE note deberían indicar que "X" en todas las posiciones puede ser cualquier aminoácido, en la medida en que se trate del mismo en todas las localizaciones.

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representativo no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de P.I.P.I. podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 3.c), 7.b), 26 y 27.

Párrafo 28 – Secuencias de aminoácidos separadas por símbolos internos de terminación

Ejemplo 28-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada

En una solicitud de patente se describen las secuencias siguientes:

caattcaggg tggatgaat atg gcg ccc aat acg caa acc gcc tct ccc cgc
Met Ala Pro Asn Thr Gln Thr Ala Ser Pro Arg

gcg ttg gca | gat tca tta atg cag ctg gca cga cag gtt tcc cga ctg
Ala Leu Ala Asp Ser Leu Met Gln Leu Ala Arg Gln Val Ser Arg Leu

Protein A

gaa agc ggg cag tga atg acc atg att acg gat tca ctg gcc gtc gtt
Glu Ser Gly Gln Met Thr Met Ile Thr Asp Ser Leu Ala Val Val

tta caa cgt cgt gac tgg gaa aac cct ggc gtt acc caa ctt aat cgc
Leu Gln Arg Arg Asp Trp Glu Asn Pro Gly Val Thr Gln Leu Asn Arg

Protein B

ctt gca gca cat tgg tgt caa aaa taa taataaccgg atgtactatt
Leu Ala Ala His Trp Cys Gln Lys

tatccctg atg ctg cgt cgt cag gtg aat gaa gtc gct taa gcaatcaatg
Met Leu Arg Arg Gln Val Asn Glu Val Ala

Protein C

tcggatgcgg cgcgacgctt atccgaccaa catatcataa

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

En la solicitud se describe una secuencia de nucleótidos, que contiene codones de terminación, que codifica tres secuencias de aminoácidos distintas.

La secuencia de nucleótidos enumerada contiene más de 10 nucleótidos específicamente definidos y debe incluirse en una lista de secuencias como secuencia única.

Por lo que respecta a las secuencias de aminoácidos codificadas, el párrafo 28 exige que las secuencias de aminoácidos separadas por un símbolo interno de terminación, por ejemplo, un espacio en blanco, deben incluirse como secuencias distintas. Puesto que la "Protein A", la "Protein B" y la "Protein C" contienen, cada una de ellas, cuatro o más aminoácidos específicamente definidos, el párrafo 7.b) de la Norma ST.26 exige que cada una de ellas se incluya en una lista de secuencias y se le asigne su propio identificador de secuencia.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia de nucleótidos debe incluirse en una lista de secuencias como:

```
caattcaggggtggaatgatggcgcccaatacgcgaaaccgcctctccccgcggtggccgattcattaatggaagcgggcagtgatgaccatgattac  
ggattcactggccgctgtttacaacgctgctgactgggaaaaccctggcgttacccttaactgccttgacgacacattgggtcaaaaataataaaccg  
gatgtactatttaccctgatgctgcgctgcgagtgatgaagtcgctaagcaatcaatgctggatgcggcgacgcttatccgaccaacatacataa.  
(SEQ ID NO: 46)
```

La secuencia de nucleótidos debería describirse en más detalles utilizando una clave de caracterización "CDS" para cada una de las tres proteínas, y el elemento INSDFeature_location debe identificar la localización de cada secuencia codificante, incluyendo el codón de terminación. Además, para cada clave de caracterización "CDS", debería incluirse el calificador "translation" con la secuencia de aminoácidos de la proteína como valor calificador. En la solicitud no se divulga la tabla del código genético que corresponde a la traducción (véase el Cuadro 7 de la Sección 9 del anexo 1). Si corresponde aplicar la tabla de códigos estándar, el calificador "transl_table" no será necesario; sin embargo, si corresponde aplicar una tabla del código genético diferente, deberá indicarse el valor calificador adecuado del Cuadro 7 para el calificador "transl_table". Por último, debe incluirse el calificador "protein_id" con el valor calificador que indique el identificador de secuencia de cada una de las secuencias de aminoácidos traducidas.

Las secuencias de aminoácidos deben incluirse como secuencias distintas, atribuyendo a cada una su propio identificador de secuencia:

MAPNTQTASPRALADSLMQLARQVSRLESGQ (SEQ ID NO: 47)

MTMITDSLAVVLQRRDWENPGVTQLNRLAAHWCQK (SEQ ID NO: 48)

MLRRQVNEVA (SEQ ID NO: 49)

NOTA: Véase el "Ejemplo 90-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones" en la que figura una ilustración de una secuencia de aminoácidos traducida representada como secuencia única.

Párrafos pertinentes de la Norma ST.26: 7, 26, 28, 57, 89 a 92.

Párrafo 29 – Representación de un aminoácido “other”

Ejemplo 29-1: Símbolo de ambigüedad más restrictivo para un aminoácido “other”

En una solicitud de patente se describe la secuencia siguiente:

Ala-Hse-X₁-X₂-X₃-X₄-Tyr-Leu-Gly-Ser

donde X₁= Ala o Gly,

X₂= Ala o Gly,

X₃= Ala o Gly,

X₄= Ala o Gly y

Hse = Homoserina

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

El péptido enumerado contiene cinco aminoácidos específicamente definidos. El símbolo “X” se utiliza convencionalmente para representar dos aminoácidos como alternativa (véase la Introducción del presente documento).

Puesto que hay cinco aminoácidos específicamente definidos, es decir, Ala, Tyr, Leu, Gly y Ser, el párrafo 7.b) de la Norma ST.26 exige que la secuencia se incluya en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

El párrafo 29 exige que cualquier aminoácido “other” esté representado por el símbolo “X”. En el ejemplo, la secuencia contiene el aminoácido Hse en la segunda posición, que no figura en el Cuadro 3 de la Sección 3 del Anexo I. En consecuencia, Hse es un aminoácido “other” y debe representarse mediante el símbolo “X”.

X₁-X₄ son posiciones variantes, y cada una de ellas puede ser A o G. El símbolo de ambigüedad más restrictivo para las alternativas A o G es “X”. Por lo tanto, la secuencia puede ser representada como:

AXXXXXYLGS (SEQ ID NO: 50)

Se alienta encarecidamente a incluir cualesquiera secuencias específicas fundamentales para la divulgación o las reivindicaciones de la invención, tal como se expone en la Introducción del presente documento.

Puesto que el aminoácido Hse no figura en el Cuadro 4 de la Sección 4 del Anexo I, deberán proporcionarse una clave de caracterización “SITE” y un calificador “NOTEnote” con el nombre completo, no abreviado, de la homoserina en la forma prevista en el párrafo 30 de la Norma ST.26.

Con arreglo al párrafo 27, puesto que X₁-X₄ representa una alternativa de tan solo dos aminoácidos, es necesaria una descripción adicional. El párrafo 96 indica que debería utilizarse la clave de caracterización “VARIANT” con el calificador “NOTEnote” y el valor calificador “A o G”. Con arreglo al párrafo 34 de la Norma ST.26, habida cuenta de que esas posiciones son adyacentes y tienen la misma descripción, pueden describirse conjuntamente utilizando la sintaxis “3..6” como descriptor de localización en el elemento INSDFeature_location.

Párrafos pertinentes de la Norma ST.26: 3.a), 7.b), 25-27, 29, 30, 34, 66, 70, 71 y 96 a 97.

Ejemplo 29-2: Utilización del correspondiente aminoácido no modificado

En una solicitud de patente se describe la secuencia siguiente:

Ala-Hyl-Tyr-Leu-Gly-Ser-Nle-Val-Ser-5ALA

Donde Hyl = hidroxilisina (modificación postraducciona de la lisina), Nle = norleucina y 5ALA = ácido δ -aminolevulínico

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

El péptido enumerado contiene más de cuatro aminoácidos específicamente definidos; por lo tanto, la secuencia debe incluirse en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La hidroxilisina en la posición 2, la norleucina en la posición 7 y el ácido δ -aminolevulínico en la posición 10 son todos "aminoácidos modificados". En primer lugar, debemos considerar cada aminoácido modificado y determinar si debe ser representado por el correspondiente aminoácido no modificado o por la variable "X" de la secuencia. El párrafo 29 establece que los aminoácidos modificados "deberían representarse en la secuencia como los correspondientes aminoácidos no modificados, cuando sea posible".

Queda a discreción del solicitante decidir si un aminoácido modificado será representado por el correspondiente residuo no modificado o por las variables "X". No obstante, debería tenerse en cuenta la siguiente indicación: Si un aminoácido se modifica mediante la adición de una fracción, como la metilación o la acetilación, y la estructura básica del aminoácido no modificado correspondiente no suele cambiar, se recomienda la representación por el aminoácido no modificado. Si el aminoácido modificado es estructuralmente muy diferente del correspondiente aminoácido no modificado, se recomienda su representación mediante una "X".

La estructura de la hidroxilisina es casi idéntica a la de la lisina, salvo que el tercer carbono del grupo R está modificado con un grupo hidroxilo. Dado que la estructura básica del correspondiente residuo de lisina no modificado está intacta, la hidroxilisina debería representarse en la secuencia por lisina ("K"), no por "X".

La norleucina es un isómero de la leucina. El grupo R de la leucina es una cadena de 4 carbonos, ramificada en el segundo carbono. La norleucina también tiene un grupo R de 4 carbonos, pero es lineal y no ramificada. Por lo tanto, la norleucina no es simplemente el resultado de una modificación añadida a la leucina, sino una estructura completamente diferente (aunque relacionada). Por lo tanto, se recomienda que la norleucina se represente con una "X" en la lista de secuencias.

El ácido δ -aminolevulínico no es estructuralmente similar a ninguno de los aminoácidos del Cuadro 3 del Anexo I, por lo que se recomienda representarlo con una "X" en una lista de secuencias.

Por lo tanto, la secuencia debería incluirse en una lista de secuencias como:

AKYLGXSVSX (SEQ ID NO: 51)

El párrafo 30 requiere la anotación adicional de cada aminoácido modificado.

La hidroxilisina es una modificación postraducciona de la lisina. Por lo tanto, debe describirse utilizando la clave de caracterización "MOD_RES" junto con un calificador "note" que describa la modificación. Cabe señalar que la "hidroxilisina" figura en el Cuadro 4 de la Sección 4 del Anexo 1, "Lista de aminoácidos modificados". Por lo tanto, el valor del calificador "note" puede contener la abreviatura "Hyl" en lugar del nombre completo no abreviado "hidroxilisina".

La norleucina no es un residuo modificado postraduccionalmente, por lo que debe describirse utilizando la clave de caracterización "SITE" junto con un calificador "note" que describa la modificación. Cabe señalar que la "norleucina" también figura en el Cuadro 4 de la Sección 4 del Anexo 1, por lo que el valor del calificador "note" puede contener la abreviatura "Nle" en lugar del nombre completo no abreviado "norleucina".

El ácido δ -aminolevulínico tampoco es un residuo modificado postraduccionalmente, por lo que debe describirse utilizando la clave de caracterización "SITE" junto con un calificador "note" que describa la modificación. El ácido δ -aminolevulínico no figura en el Cuadro 4 de la Sección 4 del Anexo 1, por lo que el valor del calificador "note" debe contener el nombre completo no abreviado del residuo modificado, "ácido δ -aminolevulínico".

Párrafos pertinentes de la Norma ST.26: 3.a), 3.e), 7.b), 29 y 30

Párrafo 30 – Anotación de un aminoácido modificado

Ejemplo 30-1: Clave de caracterización “CARBOHYD”

En una solicitud de patente se describe un polipéptido con un aminoácido específicamente modificado, que contiene una cadena lateral glicosilada, que se caracteriza por el hecho de que la Cys correspondiente a las posiciones 4 y 15 del polipéptido forma una unión disulfuro, según la secuencia siguiente:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Asn(asialyloligosaccharide)-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

El péptido enumerado prevé 17 aminoácidos específicamente definidos. Hay 16 aminoácidos naturales, siendo el noveno (asparagina) glicosilado. Por lo tanto, la secuencia debe incluirse en una lista de secuencias, según lo exige el párrafo 7.b) de la Norma ST.26.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Con arreglo al párrafo 29 de la Norma ST.26, un aminoácido modificado debería representarse en la secuencia como el correspondiente aminoácido no modificado, cuando sea posible.

Por lo tanto, la secuencia debe incluirse en una lista de secuencias como:

LEYCLKRWNETHCAW (SEQ ID NO: 5452)

Es necesaria una descripción más detallada del aminoácido modificado. Debería utilizarse la clave de caracterización “CARBOHYD” junto con el calificador (obligatorio) “NOTE note” para indicar que en la posición 9 se produce la unión de una cadena de azúcar (un sialiloligosacárido) con asparagina. El calificador “NOTE note” describe el tipo de enlace, por ejemplo enlace N. El descriptor de localización en el elemento localización de característica es el número de posición del residuo de la asparagina modificada.

Además, hay una unión disulfuro entre los dos residuos Cys. Por lo tanto, ~~se utiliza~~ debería utilizarse la clave de caracterización “DISULFID” para describir un enlace cruzado entre cadenas. El elemento localización de característica son los números de posición de los residuos Cys enlazados en formato “x..y”, es decir, “4..15”. El calificador ~~NOTE no es~~ obligatorio ~~“note” debería describir el enlace disulfuro entre cadenas.~~

Párrafos pertinentes de la Norma ST.26: 3.a), 7.b), 26, 29, 30, 66.c), 70 y Sección 7 del Anexo I, clave de caracterización 7.4.

Párrafo 36 – Secuencias que contienen regiones de un número exacto de residuos “n” o “X” contiguos

Ejemplo 36-1: Secuencia con una región de un número conocido de residuos “X” representada como secuencia única

LL-100-KYMR

donde “-100-” entre los aminoácidos leucina y lisina refleja una región de 100 aminoácidos en la secuencia.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

El párrafo 36 de la Norma ST.26 exige la inclusión de una secuencia que contiene al menos cuatro aminoácidos específicamente definidos separados por una o más regiones de un número definido de residuos “X”.

En la secuencia divulgada se utiliza un símbolo no convencional, es decir “-100-.” La definición de “-100-” debe determinarse a partir de la explicación de la secuencia en la divulgación, que define este símbolo como 100 aminoácidos entre la leucina y la lisina (véase la Introducción del presente documento). Por lo tanto, “-100-” es una región definida de residuos “X”. Puesto que seis de los 106 aminoácidos en la secuencia son específicamente definidos, el párrafo 7.b) de la Norma ST.26 exige que la secuencia se incluya en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

El símbolo no convencional “-100-” se representa como 100 residuos “X” (puesto que cualquier símbolo utilizado para representar un aminoácido es equivalente tan solo a un residuo). Por lo tanto, una secuencia única de 106 aminoácidos de longitud, que contiene 100 residuos “X” entre LL y KYMR, debe incluirse en una lista de secuencias (SEQ ID NO: 523).

Esta secuencia contiene 100 variables “X” entre LL y KYMR. El valor por defecto de la Norma ST.26 para “X”, sin ninguna otra anotación, es “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” o “V” (párrafo 27). Si estas 100 variables “X” se definen como algo distinto de este valor por defecto, se debe proporcionar una anotación adecuada para cada variable “X”.

Párrafos pertinentes de la Norma ST.26: 7.b), 26, 27 y 36.

Ejemplo 36-2: Secuencia con múltiples regiones de un número o una serie conocidos de residuos "X" representada como secuencia única

Lys-z₂-Lys-z_m-Lys-z₃-Lys-z_n-Lys-z₂-Lys

donde z es cualquier aminoácido, m=20, n=19-20, z₂ significa que los pares de lisina están separados por cualesquiera dos aminoácidos y z₃ significa que los pares de lisina están separados por cualesquiera tres aminoácidos.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

En la secuencia divulgada se utiliza un símbolo no convencional, es decir "z." Por lo tanto, debe consultarse la divulgación para determinar la definición; "z" se define como cualquier aminoácido (véase la Introducción del presente documento). El símbolo convencional utilizado para representar cualquier aminoácido es "X". Considerando la presencia de variables "X", el péptido contiene seis residuos lisina que están enumerados y específicamente definidos, y se exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

En la secuencia se utiliza un símbolo no convencional "z", cuya definición debe determinarse a partir de la divulgación. Puesto que "z" se define como cualquier aminoácido, el símbolo convencional es "X."

El medio de representación preferido y más abarcador es (véase la Introducción del presente documento):

KXXKXXXXXXXXXXXXXXXXXXXXXXXXKXXKXXXXXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 5354)

donde z_n es igual a 20 "X", con una descripción adicional de que la variable "X" correspondiente a la posición 30 puede suprimirse.

Como alternativa, o además de lo expuesto más arriba, la secuencia puede ser representada como:

KXXKXXXXXXXXXXXXXXXXXXXXXXXXKXXKXXXXXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 5455)

donde z_n es igual a 19 "X", con una descripción adicional de que puede insertarse una variable "X" entre las posiciones 29 y 30.

Con arreglo al párrafo 27, el símbolo "X" se interpretará como cualquiera de los símbolos "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" o "V", excepto cuando se utilice con una descripción detallada en el cuadro de características. Por lo tanto, si se pretende que "X" represente "cualquier aminoácido", debería anotarse con la clave de caracterización "VARIANT" y un calificador "NOTE" con el valor "X puede ser cualquier aminoácido".

Puesto que "X" en SEQ ID NO: 54 y 55 representa "cualquier aminoácido", deberá anotarse con la clave de caracterización VARIANT y el calificador "note" con el valor "X puede ser cualquier aminoácido".

Cuando sea posible, cada "X" debería anotarse individualmente. Sin embargo, toda región que contiene residuos "X" contiguos, o numerosos residuos "X" dispersos en toda la secuencia, puede describirse en conjunto con la clave de caracterización "VARIANT" utilizando la sintaxis "x..y" como descriptor de localización, donde x e y son las posiciones del primer y último residuo "X", y un calificador "NOTE" con el valor "X puede ser cualquier aminoácido".

Párrafos pertinentes de la Norma ST.26: 26, 27 y 36.

Párrafo 37 – Secuencias que contienen regiones de un número desconocido de residuos “n” o “X” contiguos

Ejemplo 37-1: Una secuencia con regiones de un número desconocido de residuos “X” no debe representarse como secuencia única

Gly-Gly----Gly-Gly-Xaa-Xaa

donde el símbolo ---- es una región indefinida en la secuencia, Xaa es cualquier aminoácido y los residuos glicina y Xaa están conectados entre sí mediante enlaces peptídicos.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

NO

El párrafo 37 de la Norma ST.26 prohíbe la inclusión de toda secuencia que contenga una región indefinida; por lo tanto, no se exige la inclusión de la totalidad de la secuencia.

El párrafo 37 de la Norma ST.26 exige la inclusión de toda región de una secuencia adyacente a una región indefinida que contenga cuatro o más aminoácidos específicamente definidos. En el ejemplo que figura más arriba, no se exige la inclusión de las regiones adyacentes a la región indefinida, puesto que cada región contiene solo dos aminoácidos específicamente definidos.

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

NO – la totalidad de la secuencia, no

NO – cualquier región de la secuencia, no

El párrafo 37 de la Norma ST.26 no permite la inclusión de la totalidad de la secuencia.

El párrafo 8 de la Norma ST.26 no permite la inclusión de las regiones adyacentes a la región indefinida, puesto que cada región contiene solo dos aminoácidos específicamente definidos.

Párrafos pertinentes de la Norma ST.26: 7.b), 8, 26 y 37.

Ejemplo 37-2: Una secuencia con regiones de un número desconocido de residuos "X" no debe representarse como secuencia única

Gly-Gly----Gly-Gly-Ala-Gly-Xaa-Xaa

donde el símbolo ---- es una región indefinida en la secuencia, Xaa es cualquier aminoácido y los residuos glicina y Xaa están conectados entre sí mediante enlaces peptídicos.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

NO – la totalidad de la secuencia, no

SÍ – una región de la secuencia

El párrafo 37 de la Norma ST.26 prohíbe la inclusión de toda secuencia que contenga una región indefinida, pero exige la inclusión de toda región de una secuencia adyacente a una región indefinida que contenga cuatro o más aminoácidos específicamente definidos.

En el ejemplo que figura más arriba, la Norma ST.26 no exige (y prohíbe) la inclusión tanto de la totalidad de la secuencia que contiene una región indefinida como de la región Gly-Gly adyacente a la región indefinida, que contiene solo dos aminoácidos específicamente definidos. Sin embargo, la Norma ST.26 exige la inclusión de la región Gly-Gly-Ala-Gly- Xaa-Xaa adyacente a la región indefinida, puesto que contiene al menos cuatro aminoácidos específicamente definidos.

Pregunta 2: ¿Permite la Norma ST.26 la inclusión de la secuencia o secuencias?

NO – la totalidad de la secuencia y la región Gly-Gly, no

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La región de la secuencia adyacente a la región indefinida que contiene cuatro aminoácidos específicamente definidos debe representarse como:

GGAGXX (SEQ ID NO: 5758)

De preferencia, debería anotarse la secuencia de manera que indique que la secuencia representada forma parte de una secuencia mayor que contiene una región indefinida, utilizando la clave de caracterización "SITE", la localización de característica "1" y el calificador "NOTE" con el valor, por ejemplo, "Este residuo presenta un enlace N-terminal a un péptido que tiene un N-terminal Gly-Gly y una región de longitud indefinida".

Con arreglo al párrafo 27, el símbolo "X" se interpretará como cualquiera de los símbolos "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" o "V", excepto cuando se utilice con una descripción detallada en el cuadro de características. Por lo tanto, si se pretende que "X" represente "en SEQ ID NO: 58 representa cualquier aminoácido", debería anotarse con la clave de caracterización "VARIANT" y un calificador "NOTE" con el valor "X puede ser cualquier aminoácido".

Cuando sea posible, cada "X" debería anotarse individualmente. Sin embargo, toda región que contiene residuos "X" contiguos, o numerosos residuos "X" dispersos en toda la secuencia, puede describirse en conjunto con la clave de caracterización "VARIANT" utilizando la sintaxis "x..y" como descriptor de localización, donde x e y son las posiciones del primer y último residuo "X", y un calificador "NOTE" con el valor "X puede ser cualquier aminoácido".

Párrafos pertinentes de la Norma ST.26: 7.b), 8, 26, 27 y 37.

Párrafo 55 – Una secuencia de nucleótidos que contiene segmentos de ADN y ARN

Ejemplo 55-1: Molécula combinada ADN/ARN

En una solicitud de patente se divulga la secuencia de oligonucleótidos siguiente:

AGACCTTcggagucuccuguugaacagauagucaaguagauC

En la que las letras mayúsculas representan residuos de ADN y las minúsculas, residuos de ARN.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

La secuencia divulgada tiene más de diez nucleótidos enumerados y específicamente definidos; por lo tanto, se exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia de nucleótidos debe incluirse en una lista de secuencias como:

agaccttcggagtctcctgttgaacagatagtagtcaagtagatc (SEQ ID NO: 9293)

Obsérvese que los nucleótidos de uracilo deben representarse mediante el símbolo "t" en la lista de secuencias.

El párrafo 55 de la Norma ST.26 establece que para toda secuencia de nucleótidos que contenga segmentos de ADN y ARN el tipo de molécula deberá indicarse como "DNA" y describirse con detalle utilizando la clave de caracterización "fuente", el calificador obligatorio "organism" con el valor "synthetic construct" y el calificador obligatorio "mol_type" con el valor "other DNA". Además, cada segmento de la secuencia debe describirse con detalle utilizando la clave de caracterización "misc_feature", que incluye la localización del segmento y el calificador "note" que indica si el segmento es ADN o ARN. La secuencia divulgada contiene dos segmentos de ADN (posiciones de nucleótidos 1 a 7 y 43) y un segmento de ARN (posiciones de nucleótidos 8 a 42).

Párrafos pertinentes de la Norma ST.26: 7, 14, 55-56, and 83

Párrafo 89 – Clave de caracterización “CDS”

Ejemplo 89-1: Secuencia de nucleótidos codificante y secuencia de aminoácidos codificada

En una solicitud de patente se describe la siguiente secuencia de nucleótidos y su traducción:

atg acc gga aat aaa cct gaa acc gat gtt tac gaa att tta tga

Met Thr Gly Asn Lys Pro Glu Thr Asp Val Tyr Glu Ile Leu STOP

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

La secuencia de nucleótidos enumerada tiene más de diez nucleótidos específicamente definidos.

La secuencia de aminoácidos enumerada tiene más de cuatro aminoácidos específicamente definidos.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia de nucleótidos debe representarse como:

atgaccggaataaacctgaaaccgatgtttacgaaatttatga (SEQ ID NO: 5859)

La secuencia de nucleótidos debería describirse además utilizando la clave de caracterización “CDS” y el elemento INSDFeature_location deberá identificar la totalidad de la secuencia, incluyendo el codón de terminación (es decir, de la posición 1 a la 45). Asimismo, debería incluirse el calificador “translation” con el valor calificador “MTGNKPETDVYEIL”. La solicitud no divulga la tabla del código genético que corresponde aplicar a la traducción (véase el Cuadro 7 de la Sección 9 del Anexo 1). Si corresponde aplicar la tabla del código genético estándar, el calificador “transl_table” no será necesario; sin embargo, si corresponde aplicar una tabla del código genético diferente, deberá indicarse el valor calificador adecuado del Cuadro 7 para el calificador “transl_table”. Por último, deberá incluirse el calificador “protein_id” con el valor calificador que indique el identificador de secuencia de la secuencia de aminoácidos traducida.

La secuencia de aminoácidos debe presentarse por separado con su propio identificador de secuencia utilizando códigos de una letra, según se indica a continuación:

MTGNKPETDVYEIL (SEQ ID NO: 5860)

El “STOP” después de la secuencia de aminoácidos enumerada no debe incluirse en la secuencia de aminoácidos en la lista de secuencias.

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de **P.I.P.I.** podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 7.a), 7.b), 26, 28, 89, 90 y 92.

Ejemplo 89-2: Una localización de característica se extiende más allá de la secuencia divulgada

Una solicitud de patente contiene la figura siguiente, en la que se divulga una secuencia codificadora parcial y su secuencia de aminoácidos traducida:

```
cat cac gca gca gaa tgt gga ttt tgt cct caa caa tgg caa gtt cta 48
His His Ala Ala Glu Cys Gly Phe Cys Pro Gln Gln Trp Gln Val Leu
1   5   10   15

cgt ggg agt ctg tgc att tgt gag ggt cca gct gaa gga tgg ttc ata 96
Arg Gly Ser Leu Cys Ile Cys Glu Gly Pro Ala Glu Gly Trp Phe Ile
  20   25   30

tca aga tgt tgg tta tgg tgt ggg cct caa gtc caa ggc ttt atc ttt 144
Ser Arg Cys Trp Leu Trp Cys Gly Pro Gln Val Gln Gly Phe Ile Phe
  35   40   45

gga gaa ggc aag gaa gga ggc ggt gac aga cgg gct gaa gcg agc cct 192
Gly Glu Gly Lys Glu Gly Gly Gly Asp Arg Arg Ala Glu Ala Ser Pro
  50   55   60

cag gag ttt tgg gaa tgc act tgg                216
Gln Glu Phe Trp Glu Cys Thr Trp
65   70
```

Figure 1 - partial coding sequence of the *Homo sapiens ITCH1* gene, which encodes amino acids 20 through 91 of the 442 amino acid long ITCH1 protein.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

En la solicitud se divulga una secuencia de nucleótidos y su secuencia de aminoácidos traducida. La secuencia de nucleótidos enumerada contiene más de 10 nucleótidos específicamente definidos y deberá incluirse en una lista de secuencias.

La secuencia de aminoácidos contiene más de 4 aminoácidos específicamente definidos y también deberá incluirse en una lista de secuencias como una secuencia distinta con su propio identificador de secuencia.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia de nucleótidos debe incluirse en una lista de secuencias como:

```
catcacgcagcagaatgtggattttgtcctcaacaatggcaagttctacgtgggagctgtgcatttgagggtccagctgaaggatggtcatatcaagatgt  
tggttatgggtggcctcaagtcgaaggctttatcttggagaaggcaaggaggcggtgacagacgggctgaagcgagcctcaggagtttggga  
atgcacttgg (SEQ ID NO: 9394)
```

La secuencia de nucleótidos debería describirse con detalle utilizando una clave de caracterización "CDS". El elemento INSDFeature_location debe identificar la ubicación la caracterización "CDS" en la secuencia e incluir el codón de terminación.

En la figura se describe una secuencia de codificación parcial que no incluye el codón de inicio ni el codón de terminación. Sin embargo, la descripción de la secuencia indica que el codón de inicio se sitúa antes del nucleótido en la posición 1 y el codón de terminación se sitúa después del último nucleótido en la posición 216.

La Norma ST.26 establece que el descriptor de localización no deberá incluir números de residuos fuera de la serie de la secuencia indicada en el elemento INSDSeq_sequence. Por consiguiente, en el ejemplo anterior el descriptor de localización para la clave de caracterización CDS no puede incluir números de posición fuera del intervalo entre las posiciones 1 y 216. La posición del codón de terminación en el elemento INSDFeature_location debe representarse utilizando el símbolo ">" para indicar que el codón de terminación se sitúa después de la posición 216. Del mismo modo, el símbolo "<" puede utilizarse para indicar que el codón de inicio se sitúa después de la posición 1. Por lo tanto, el descriptor de localización para la clave de caracterización CDS debería aparecer como se indica a continuación:

<1...>216

Debe tenerse presente que "<" y ">" son caracteres reservados y serán sustituidos por "<" y ">", respectivamente, en la instancia XML de la lista de secuencias.

Debería incluirse el calificador "translation" con la secuencia de aminoácidos de la proteína como valor calificador.

En la figura no se divulga la tabla del código genético que corresponde aplicar a la traducción (véase el Cuadro 7 de la Sección 9 del Anexo 1). Si corresponde aplicar la tabla del "código estándar", el calificador "transl_table" no será necesario; sin embargo, si corresponde aplicar una tabla del código genético diferente, deberá indicarse el valor calificador adecuado del Cuadro 7 que figura en el Anexo I de la Norma ST. 26 para el calificador "transl_table". Por último, deberá incluirse el calificador "protein_id" con el valor calificador que indique el identificador de secuencia de la secuencia de aminoácidos traducida.

La secuencia de aminoácidos traducida deberá incluirse como una secuencia distinta con su propio identificador de secuencia:

```
HHA A E C G F C P Q Q W Q V L R G S L C I C E G P A E G W F I S R C W L W C G P Q V Q G F I F G E G K E G G D R R A E A S P Q E F W E C  
T W (SEQ ID NO: 9495)
```

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de PI podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 7, 41, 65, 66, 70, 71, 89 y 92

Párrafo 92 – Secuencia de aminoácidos codificada por una secuencia codificadora

Ejemplo 92-1: Secuencia de aminoácidos codificada por una secuencia codificadora con intrones

Una solicitud de patente contiene la figura siguiente, en la que se divulga una secuencia codificadora y su traducción:

```
atg aag act ttc gca gcc ttg ctt tcc gct gtc act ctc gcg ctc tcg
Met Lys Thr Phe Ala Ala Leu Leu Ser Ala Val Thr Leu Ala Leu Ser

gtg cgc gcc cag gcg gct gtc tgg agt caa t gtaagtgccg ctgcttttca
Val Arg Ala Gln Ala Ala Val Trp Ser Gln

ttgatacgag actctacgcc gagctgacgt gctaccgatat ag gt ggc ggt aca
Cys Gly Gly Thr

ccg ggt tgg acg gcc gag acc act tgc gtt gct ggt tcg gtt tgt acc
Pro Gly Trp Thr Gly Glu Thr Thr Cys Val Ala Gly Ser Val Cys Thr

tcc ttg agc tca gtgagcgact ttcaatccgt cgtcattgct cctcatgtat
Ser Leu Ser Ser

tgacgattgg ccttcatag tca tac tct caa tgc gtt ccg gcc tcc gca acg
Ser Tyr Ser Gln Cys Val Pro Gly Ser Ala Thr

tcc agc gct ccg gcg gcc ccc tca gcg aca act tca gcc ccc gca cct
Ser Ser Ala Pro Ala Ala Pro Ser Ala Thr Thr Ser Gly Pro Ala Pro

acg gac gga acg tgc tcg gcc agc ggg gca tgg ccg cca ttg acc tga
Thr Asp Gly Thr Cys Ser Ala Ser Gly Ala Trp Pro Pro Leu Thr Ter
```

Figura 1 – los nucleótidos que figuran en letra negra son regiones de intrones.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

En la solicitud se divulga una secuencia de nucleótidos y su traducción a aminoácidos. La secuencia de nucleótidos enumerada contiene más de 10 nucleótidos específicamente definidos y debe incluirse en una lista de secuencias como secuencia única.

La secuencia de nucleótidos contiene secuencia codificante (exones) separada por secuencia no codificante (intrones). La figura ilustra la traducción de la secuencia de nucleótidos como tres secuencias de aminoácidos no contiguas. Conforme a la leyenda de la figura, las regiones de nucleótidos que figuran en letra negra son secuencias de intrones que serán eliminadas de un transcripto de ARN antes de la traducción a proteína. En consecuencia, de hecho, las tres secuencias de aminoácidos son una única secuencia enumerada, contigua, que contiene más de cuatro aminoácidos específicamente definidos y debe incluirse en una lista de secuencias como secuencia única.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia de nucleótidos debe incluirse en una lista de secuencias como:

```
atgaagactttcgcagcctgtcttcgctgtcactctcgcgctctcggtgvcgcccaggcggctgtctggagtaagtaagtcgctgttttcattgatagc  
agactctacgccgagctgacgtgctaccgtataggtggcggtacaccgggtggacggcgagaccacttgcgttgctggttcggtttgacctccttgagctc  
agtgcgactttcaatccgctgctcattgctcctcatgtattgacgattggcctcatagtcatactctcaatgcgttcgggctccgcaacgtccagcgtccgg  
cggcccctcagcgcacaacttcaggccccgcacctacggacggaacgtgctcggccagcgggcatggccgccattgactga (SEQ ID NO:  
7475)
```

La secuencia de nucleótidos debería describirse en más detalle utilizando una clave de caracterización "CDS" y el elemento INSDFeature_location deberá identificar la localización de la secuencia codificante, incluyendo el codón de terminación indicado mediante "Ter". El elemento INSDFeature_location de la CDS deberá utilizar el operador de localización "join" para indicar que los productos de traducción codificados por las localizaciones indicadas están unidos y forman un polipéptido único y contiguo que utiliza el formato "join(x₁..y₁,x₂..y₂,x₃..y₃)", es decir, "join(1..79,142..212,272..400)". Además, debería incluirse el calificador "translation", con la secuencia de aminoácidos de la proteína como valor calificador. (Cabe señalar que el símbolo de terminación "Ter" en la última posición de la secuencia no debe incluirse en la secuencia de aminoácidos.) En la solicitud no se divulga la tabla del código genético que corresponde aplicar a la traducción (véase el Cuadro 7 de la Sección 9 del Anexo 1). Si corresponde aplicar la tabla del "código estándar", el calificador "transl_table" no será necesario; sin embargo, si corresponde aplicar una tabla del código genético diferente, deberá indicarse el valor calificador adecuado del Cuadro 7 para el calificador "transl_table". Por último, debe incluirse el calificador "protein_id" con el valor calificador que indique el identificador de secuencia de la secuencia de aminoácidos traducida.

La secuencia de aminoácidos debe incluirse como secuencia única:

```
MKTFALLSAVTLALSvRAQAaVWSQCgGTPGWTGETTCVAGSVCTSLSSSYsQCVPGSATSSAPAAPSATT  
SGPAPTDGTCsASGAWPPLT (SEQ ID NO: 7576)
```

Párrafos pertinentes de la Norma ST.26: 7, 26, 28, 57, 67 y 89 a 92.

Párrafo 93 – Secuencia principal y una variante, cada una de ellas enumerada mediante sus residuos

Ejemplo 93-1: Representación de variantes enumeradas

La descripción incluye el siguiente alineamiento de secuencias.

```
D. melanogaster          ACATTGAATCTCATACCACTTT
D. virilis              ...-..G...C...-G.....
D. simulans            GT..G.CG..GT..SGT.G...
```

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

En esta ámbito de la técnica, es común incluir “puntos” en un alineamiento de secuencias para indicar: “esta posición es la misma que la posición que se encuentra por encima de ella”. Por lo tanto, los “puntos” en las secuencias de las especies *D. virilis* y *D. simulans* se consideran nucleótidos enumerados y específicamente definidos, pues son simplemente una forma abreviada de indicar que una posición determinada es el mismo nucleótido que en *D. melanogaster*. Además, con el fin de maximizar el alineamiento, en los alineamientos de secuencias figura con frecuencia el símbolo “-” para indicar la ausencia de un residuo.

Por consiguiente, la secuencia de nucleótidos de las especies *D. melanogaster* y *D. simulans* contiene 22 nucleótidos enumerados y específicamente definidos, y la secuencia de nucleótidos de la especie *D. virilis* contiene 19. Así pues, el párrafo 7.a) de la Norma ST.26 exige que cada una de las secuencias se incluya en una lista de secuencias con un identificador de secuencia distinto.

En esta ámbito de la técnica, es común incluir “puntos” en un alineamiento de secuencias para indicar: “esta posición es la misma que la posición que se encuentra por encima de ella”. Por lo tanto, los “puntos” en las especies 2 y 3 se consideran nucleótidos enumerados y específicamente definidos, pues son simplemente una forma abreviada de indicar que una posición determinada es el mismo nucleótido que en la especie 1. Además, con el fin de maximizar el alineamiento, en los alineamientos de secuencias figura con frecuencia el símbolo “-” para indicar la ausencia de un residuo.

Por consiguiente, la secuencia de nucleótidos de las especies 1 y 3 contiene 22 nucleótidos enumerados y específicamente definidos, y la secuencia de nucleótidos de la especie 2 contiene 19. Así pues, el párrafo 7.a) de la Norma ST.26 exige que cada una de las secuencias se incluya en una lista de secuencias con un identificador de secuencia distinto.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia *Drosophila melanogaster* debe incluirse en una lista de secuencias como:

acattgaatctcataccacttt (SEQ ID NO: 6061)

La secuencia *Drosophila virilis* debe incluirse en una lista de secuencias como:

acatggatcccacgacttt (SEQ ID NO: 6462)

La secuencia *Drosophila simulans* debe incluirse en una lista de secuencias como:

gtatggcgtcgtatsgtatgttt (SEQ ID NO: 6263)

Párrafos pertinentes de la Norma ST.26: 7.a), 13 y 93.

Ejemplo 93-2: Representación de variantes enumeradas

La descripción incluye la siguiente tabla de un péptido y sus variantes funcionales. 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".

| Posición | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|------------|---|---|---|---|---|---|---|---|---|
| Secuencia | A | V | L | T | Y | L | R | G | E |
| Variante 1 | | | | | | | | | A |
| Variante 2 | | | P | | | P | | | |
| Variante 3 | | | A | I | G | Y | | | |
| Variante 4 | | | | | | | - | | |

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

Tal como se ha indicado, un espacio en blanco en esta tabla indica que un aminoácido en la variante es el mismo que el aminoácido correspondiente en la "Secuencia". Por lo tanto, los aminoácidos de las variantes de secuencias están enumerados y específicamente definidos.

Puestos que las cuatro variantes de secuencias contienen, cada una, más de cuatro aminoácidos enumerados y específicamente definidos, el párrafo 7b) de la Norma ST.26 exige que cada una de las secuencias se incluya en una lista de secuencias con un identificador de secuencia distinto.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

AVLTYLRGE (SEQ ID NO: 7677)

AVLTYLRGA (SEQ ID NO: 7778)

AVPTYPRGE (SEQ ID NO: 7879)

AVAIGYRGE (SEQ ID NO: 7980)

AVLTYLGE (SEQ ID NO: 8081)

Párrafos pertinentes de la Norma ST.26: 7.b), 26 y 93.

Ejemplo 93-3: Representación de una secuencia de consenso

Una solicitud de patente incluye la Figura 1, con el siguiente alineamiento múltiple de secuencias:

| | |
|--------------------------------|----------------------------------|
| <i>Consensus</i> | LEGneQFINAakIIRHPkYnrkTlnNDIMLIK |
| <i>Homo sapiens</i> | LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK |
| <i>Pongo abelii</i> | LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK |
| <i>Papio Anubis</i> | LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK |
| <i>Rhinopithecus roxellana</i> | LEGTEQFINAAKIIRHPNYNRITLDNDILLIK |
| <i>Pan paniscus</i> | LEGNEQFINAAKIIRHPKYNRITLNDIMLIK |
| <i>Rhinopithecus bieti</i> | LEGNEQFINATKIIRHPKYNGNTLNDIMLIK |
| <i>Rhinopithecus roxellana</i> | LEGNEQFINATQIIRHPKYNGNTLNDIMLIK |

2. La secuencia de consenso incluye letras mayúsculas para representar residuos de aminoácidos conservados, así como las letras minúsculas “n”, “a”, “k”, “r”, “l” y “m” que representan los residuos de aminoácidos predominantes entre las secuencias alineadas.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

Las letras minúsculas en la secuencia de consenso representan, cada una, un único residuo de aminoácido. En consecuencia, la secuencia de consenso, así como cada una de las siete secuencias restantes de la Figura 1, incluye al menos cuatro aminoácidos específicamente definidos. El párrafo 7.b) de la Norma ST.26 exige la inclusión de las ocho secuencias en la lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

Las letras minúsculas en la secuencia de consenso se utilizan como símbolos de ambigüedad para representar el aminoácido predominante entre las posibles variantes para una posición específica. Por lo tanto, las letras minúsculas “n”, “a”, “k”, “r”, “l” y “m” son símbolos convencionales utilizados de manera no convencional, y la secuencia de consenso debe representarse utilizando un símbolo de ambigüedad en lugar de cada una de las letras minúsculas.

Debería utilizarse el símbolo de ambigüedad más restrictivo. Para la mayoría de las posiciones en la secuencia de consenso, “X” es el símbolo de ambigüedad más restrictivo; sin embargo, el símbolo de ambigüedad más restrictivo para “D” o “N” en las posiciones 20 y 25 es “B”. La secuencia de consenso debería incluirse en la lista de secuencias como:

LEGXEQFINAXXIIRHPXYBXXTXBNNDIXLIK (SEQ ID NO: 8482)

Con arreglo al párrafo 27, el símbolo “X” se interpretará como cualquiera de los símbolos “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, o “V”, excepto cuando se utilice con una descripción más detallada en el cuadro de características. Por lo tanto, cada “X” en la secuencia de consenso deberá describirse en mayor detalle en un cuadro de características utilizando la clave de caracterización “VARIANT” y el calificador “NOTE note” para indicar las posibles variantes para cada posición.

Las siete secuencias restantes deben incluirse en la lista de secuencias como:

LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK (SEQ ID NO: 8283)

LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK (SEQ ID NO: 8284)

LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK (SEQ ID NO: 8485)

LEGTEQFINAAKIIRHPNYNRITLDNDILLIK (SEQ ID NO: 8586)

LEGNEQFINAAKIIRHPKYNRITLNDIMLIK (SEQ ID NO: 8687)

LEGNEQFINATKIIRHPKYNGNTLNDIMLIK (SEQ ID NO: 8788)

LEGNEQFINATQIIRHPKYNGNTLNDIMLIK (SEQ ID NO: 8889)

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de PI podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 7.b), 26, 27, 93 y 97.

Párrafo 94 – Variante de secuencia divulgada como secuencia única con residuos alternativos enumerados

Ejemplo 94-1: Representación de una secuencia única con aminoácidos alternativos enumerados

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

(i) Gly-Gly-Gly-[Leu o Ile]-Ala-Thr-[Ser o Thr]

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

La secuencia prevé cuatro aminoácidos específicamente definidos, y el párrafo 7.b) de la Norma ST.26 exige la inclusión de la secuencia en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

En el Cuadro 3 de la Sección 3 del Anexo I se define el símbolo de ambigüedad "J" como isoleucina o leucina. Por lo tanto, la representación preferida de la secuencia es:

GGGJATX (SEQ ID NO: ~~63~~64)

lo que exige una descripción más detallada en un cuadro de características utilizando la clave de caracterización "VARIANT" y el calificador "NOTE~~note~~" para indicar que "X" es serina o treonina.

Como alternativa, la secuencia puede representarse, por ejemplo, como:

GGGLATS (SEQ ID NO: ~~64~~65)

lo que exige una descripción más detallada en un cuadro de características utilizando la clave de caracterización "VARIANT" y el calificador "NOTE~~note~~" para indicar que L puede ser sustituido por I, y S puede ser sustituido por T.

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de PI podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 7.b), 8, 26, 27, 94 y 97.

Párrafo 95.a) – Variante de secuencia divulgada únicamente mediante referencia a una secuencia principal con múltiples variaciones independientes

Ejemplo 95.a)-1: Representación de una variante de secuencia mediante anotación de la secuencia principal

Una solicitud contiene la divulgación siguiente:

“El fragmento de péptido 1 es Gly-Leu-Pro-Xaa-Arg-Ile-Cys, donde Xaa puede ser 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.”

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

Sí

El “fragmento de péptido 1” en cada una de las tres realizaciones divulgadas provee al menos seis aminoácidos específicamente definidos; por lo tanto, la secuencia debe incluirse en una lista de secuencias, según lo exige el párrafo 7.b) de la Norma ST.26.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

En este ejemplo, la secuencia enumerada de “fragmento de péptido 1” se divulga tres veces, como tres realizaciones diferentes, cada una con una descripción alternativa de Xaa. En este ejemplo, “X” es el símbolo de ambigüedad más restrictivo para la posición Xaa.

La Norma ST.26 exige que se incluya solo una vez la secuencia enumerada divulgada. En la más abarcadora de las tres realizaciones, Xaa es cualquier aminoácido (véase la Introducción del presente documento). Por lo tanto, la secuencia que debe incluirse en la lista de secuencias es:

GLPXRIC (SEQ ID NO: 6566)

Con arreglo al párrafo 27, el símbolo "X" se interpretará como cualquiera de los símbolos "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" o "V", excepto cuando se utilice con una descripción detallada en el cuadro de características. Puesto que "X" representa cualquier aminoácido, debería anotarse con la clave de caracterización "VARIANT" y un calificador "NOTE" con el valor "X puede ser cualquier aminoácido".

Cuando sea posible, cada "X" debería anotarse individualmente. Sin embargo, toda región que contiene residuos "X" contiguos, o numerosos residuos "X" dispersos en toda la secuencia, puede describirse en conjunto con la clave de caracterización "VARIANT" utilizando la sintaxis "x.y" como descriptor de localización, donde x e y son las posiciones del primer y último residuo "X", y un calificador "NOTE" con el valor "X puede ser cualquier aminoácido".

Se alienta encarecidamente a incluir cualesquiera secuencias adicionales fundamentales para la divulgación o las reivindicaciones de la invención, según se expone en la Introducción del presente documento.

Para el ejemplo que figura más arriba, se recomienda encarecidamente la inclusión en la lista de secuencias de las siguientes tres secuencias adicionales, cada una con su propio identificador de secuencia:

GLPVRIC (SEQ ID NO: 6667)

GLPTRIC (SEQ ID NO: 6768)

GLPDRIC (SEQ ID NO: 6869)

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de P.I.P. podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 7.b), 26, 27 y 95.a).

Párrafo 95.b) – Variante de secuencia divulgada únicamente mediante referencia a una secuencia principal con múltiples variaciones interdependientes

Ejemplo 95.b)-1: Representación de distintas variantes de secuencias con múltiples variaciones interdependientes

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

cgaatg_{n1}cccactacgaatg_{n2}cacgaatg_{n3}cccaca

donde n₁, n₂ y n₃ puede ser a, t, g, o c.

Distintas variantes de secuencia se divulgan de la manera siguiente:

si n₁ es a, n₂ y n₃ son t, g, o c;

si n₁ es t, n₂ y n₃ son a, g, o c;

si n₁ es g, n₂ y n₃ son t, a, o c;

si n₁ es c, n₂ y n₃ son t, g, o a.

Pregunta 1: ¿Exige la Norma ST.26 la inclusión de la secuencia o secuencias?

SÍ

La secuencia tiene más de diez nucleótidos enumerados y “específicamente definidos”, y el párrafo 7.a) de la Norma ST.26 exige su inclusión en una lista de secuencias.

Pregunta 3: ¿Cómo deberían representarse la secuencia o secuencias en la lista de secuencias?

La secuencia enumerada contiene más de diez nucleótidos específicamente definidos y tres residuos “n”. La norma ST.26 exige la inclusión de la secuencia enumerada divulgada y si corresponde utilizar un símbolo de ambigüedad debería utilizarse el símbolo más restrictivo. En este ejemplo, n₁, n₂ y n₃ pueden ser a, t, g, o c, por lo cual “n” es el símbolo de ambigüedad más restrictivo. Por consiguiente, la secuencia que debe incluirse en la lista de secuencias es:

cgaatg_ncccactacgaatg_ncacgaatg_ncccaca (SEQ ID NO: 6970)

El párrafo 15 de la Norma ST.26 establece que el símbolo “n” se interpretará como “a”, “c”, “g” o “t/u”, excepto cuando se utilice acompañado de una descripción detallada en el cuadro de características. Dado que el valor de cada residuo “n” en esta secuencia es equivalente al valor por defecto “a”, “c”, “g” o “t”, no se requiere ninguna otra anotación.

La secuencia enumerada contiene variaciones en tres localizaciones distintas y esas variaciones aparecen de forma interdependiente. Se alienta **encarecidamente** a incluir secuencias adicionales que representen realizaciones adicionales que son parte fundamental de la invención, según se expone en la Introducción del presente documento. Por lo tanto, con arreglo al párrafo 95.b) de la Norma ST.26, las realizaciones adicionales deberían incluirse en una lista de secuencias como cuatro secuencias distintas, cada una con su propio identificador de secuencia:

cgaatg_bcccactacgaatg_bcacgaatg_bcccaca (SEQ ID NO: 7071)

cgaatg_vcccactacgaatg_vcacgaatg_vcccaca (SEQ ID NO: 7172)

cgaatg_hcccactacgaatg_hcacgaatg_hcccaca (SEQ ID NO: 7273)

cgaatg_dcccactacgaatg_dcacgaatg_dcccaca (SEQ ID NO: 7374)

(Cabe señalar que b = t, g, o c; v = a, g, o c; h = t, a, o c; y d = t, g, o a; véase el Cuadro 1 de la Sección 1 del Anexo I)

Con arreglo al párrafo 15 de la Norma ST.26, debe utilizarse el símbolo más restrictivo para representar las posiciones variables. En consecuencia, n₂ y n₃ no deben estar representados por “n” en la secuencia.

ATENCIÓN: La representación preferida de la secuencia indicada más arriba se refiere al suministro de una lista de secuencias en la fecha de presentación de una solicitud de patente. La misma representación no puede aplicarse a una lista de secuencias suministrada con posterioridad a la fecha de presentación de una solicitud de patente, puesto que es necesario tener en cuenta si una oficina de PI podría considerar que la información suministrada añade materia a la divulgación original.

Párrafos pertinentes de la Norma ST.26: 7.a), 15 y 95.b).

[Sigue el Apéndice del Anexo VI de la Norma ST.26]

APÉNDICE

SECUENCIAS DEL DOCUMENTO DE ORIENTACIÓN EN XML

El apéndice está disponible en:

<https://www.wipo.int/standards/en/docs/st26-annex-vi-appendix-guidance-document-sequences.xml>

[Sigue el Anexo VII de la Norma ST.26]

ANEXO VII

RECOMENDACIÓN PARA LA TRANSFORMACIÓN DE LA LISTA DE SECUENCIAS DE LA NORMA ST.25 CON ARREGLO A LA NORMA ST.26: EVENTUAL ADICIÓN O SUPRESIÓN DE MATERIA

Versión 1.45

Revisión aprobada Propuesta presentada por el Comité Equipo Técnico sobre Listas de Normas Técnicas de Secuencias para que el CWS la OMPI (CWS) examine y apruebe en su octava novena sesión, celebrada el 4 de diciembre de 2020

INTRODUCCIÓN

Los requisitos para la presentación de secuencias de nucleótidos y aminoácidos difieren de la Norma ST.25 a la Norma ST.26 de la OMPI. Por consiguiente, se ha planteado la cuestión de si la Norma ST.26 exigiría la adición o supresión de materia respecto de una lista de secuencias presentada en virtud de la Norma ST.26 como parte de una solicitud internacional que no esté fundada en una solicitud cuya prioridad se reivindica.

ALCANCE DEL DOCUMENTO

En el presente documento se examinan los requisitos obligatorios de la Norma ST.26 y las eventuales consecuencias de dichos requisitos. El presente documento no aborda todas las situaciones hipotéticas posibles; si, en virtud de la Norma ST.26, el medio de representación de la información contenida en una lista de secuencias conforme con la Norma ST.25 no resulta claro, dicha información siempre podrá incluirse en la descripción de la solicitud para evitar la supresión de materia.

RECOMENDACIONES RELATIVAS A LA EVENTUAL ADICIÓN O SUPRESIÓN DE MATERIA

El examen de las cuestiones recogidas en el presente documento demuestra que la conversión de la Norma ST.25 a la Norma ST.26 no debería dar lugar necesariamente a la adición o supresión de materia, en particular respecto de las listas de secuencias presentadas en virtud de la Norma ST.25 que guardan plena conformidad con ella. No obstante, ciertas situaciones hipotéticas requerirán que el solicitante actúe con cautela. Se han formulado diversas recomendaciones para evitar la necesidad de añadir o suprimir materia.

Situación hipotética 1

La Norma ST.25 utiliza identificadores numéricos para etiquetar distintos tipos de datos, por ejemplo, <110> para el nombre del solicitante. La Norma ST.26 utiliza términos del idioma inglés, como nombres de elementos y atributos, para el etiquetado de datos.

Recomendación

Los términos de la Norma ST.26 describen el tipo de contenido de datos; por lo tanto, la utilización de los nombres de elementos y los atributos de la Norma ST.26 no constituye adición de materia.

Situación hipotética 2

La Norma ST.26 exige expresamente la inclusión de: a) secuencias ramificadas; b) secuencias con D-aminoácidos D; c) análogos de nucleótidos; y d) secuencias con sitios abásicos. Bajo la Norma ST.25, el requisito relativo a la inclusión o prohibición de dichas secuencias no estaba claro.

Recomendación:

La divulgación contenida en la solicitud debería bastar para representar dichas secuencias en una lista de secuencias conforme con la Norma ST.26 si estas no han sido incluidas en una lista de secuencias conforme con la Norma ST.25. Para determinados tipos de información requerida en virtud de la Norma ST.26, deberá tenerse cuidado de no añadir más materia de la divulgada; véanse, por ejemplo, las explicaciones que figuran más adelante (en la situación hipotética 4) sobre el calificador "mol_type" para las secuencias de nucleótidos.

Situación hipotética 3

La Norma ST.26 excluye las secuencias con menos de 10 nucleótidos específicamente definidos (con exclusión del símbolo "n") y menos de 4 aminoácidos específicamente definidos (con exclusión del símbolo "X").

Recomendación:

Las secuencias excluidas pueden incluirse en el cuerpo de la solicitud en caso de que no se hayan incluido.

Situación hipotética 4

La Norma ST.26 prevé las claves de caracterización obligatorias "source", para todas las secuencias de nucleótidos, y "SOURCEsource", para todas las secuencias de aminoácidos, cada una de las cuales incluye dos calificadores

obligatorios. La Norma ST.25 prevé una clave de caracterización correspondiente para las secuencias de nucleótidos (que apenas se utiliza) sin los calificadores correspondientes y no prevé ninguna clave de caracterización correspondiente para las secuencias de aminoácidos.

Secuencias de nucleótidos

Norma ST.26 – clave de caracterización 5.37 “source”; calificadores obligatorios 6.44 “organism” y 6.38 “mol_type” (véase el párrafo 75 de la Norma ST.26)

| Calificador | Valor |
|-------------|---|
| mol_type | genomic DNA |
| | genomic RNA |
| | mRNA |
| | tRNA |
| | rRNA |
| | other DNA (se aplica a las moléculas sintéticas) |
| | other RNA (se aplica a las moléculas sintéticas) |
| | transcribed RNA |
| | viral cRNA |
| | unassigned DNA (se aplica cuando la molécula <i>in vivo</i> es desconocida) |
| | unassigned RNA (se aplica cuando la molécula <i>in vivo</i> es desconocida) |

Secuencias de aminoácidos

Norma ST.26 – clave de caracterización 7.30 “SOURCEsource”; calificadores obligatorios 8.3 “ORGANISMorganism” y 8.1 “MOL_TYPEmol_type” (véase el párrafo 75 de la Norma ST.26)

| Calificador | Valor |
|------------------|---------|
| MOL_TYPEmol_type | protein |

Recomendación:

El único motivo de preocupación radica en los valores del vocabulario controlado asociados con el calificador “mol_type” para las secuencias de nucleótidos. Algunas de las alternativas de valor indicadas anteriormente pueden no estar suficientemente fundadas en la divulgación. Sin embargo, la adición de materia puede evitarse utilizando el valor más genérico para una secuencia específica, por ejemplo, “other DNA” y “other RNA” para una molécula sintética o “unassigned DNA” y “unassigned RNA” para una molécula *in vivo*.

Situación hipotética 5

Cuando una secuencia incluye la variable “Xaa”, la Norma ST.25 requiere que se incluya más información sobre dicho residuo en el campo <223>, que acompaña a los campos <221> (nombre de característica) y <222> (localización de característica). La Norma ST.25 no prevé un valor por defecto para “Xaa” (“X” en la Norma ST.26). Sin embargo, la Norma ST.26 sí prevé un valor por defecto y, por lo tanto, no siempre se necesita más información. Dos de las anotaciones más utilizadas en las secuencias peptídicas para la variable “Xaa” o “X” son “any amino acid” o “any naturally occurring amino acid”. Estos términos podrían interpretarse para incluir aminoácidos distintos de los enumerados en las tablas de aminoácidos que figuran en la Norma ST.25 o ST.26. En virtud de la Norma ST.26, el valor por defecto para “X”, sin más anotaciones, es cualquiera de los 22 aminoácidos individuales enumerados en el Anexo I (véase el Cuadro 3 de la Sección 3). Este valor por defecto conforme con la Norma ST.26 puede constituir en sí mismo materia añadida o suprimida y, por lo tanto, afectar negativamente al alcance de una solicitud de patente en su transición de la Norma ST.25 a la Norma ST.26.

Recomendaciones:

- a) Cuando una lista de secuencias conforme con la Norma ST.25 contiene un nombre de característica <221>, una localización de característica <222> correspondiente con Xaa e información adicional <223> sobre esta y el nombre de característica <221> es también una clave de caracterización conforme con la Norma ST.26, por ejemplo, “SITE”, “VARIANT” o “UNSURE”, debería utilizarse la clave de caracterización conforme con la Norma ST.26. Además, para evitar la posibilidad de que se suprima materia, la información contenida en el campo <223> deberá incluirse en un calificador “NOTE” “note” conexo.
- b) Cuando la lista de secuencias conforme con la Norma ST.25 contiene un nombre de característica <221> y una localización de característica <222> correspondiente a Xaa, y el nombre de característica <221> no es una clave de caracterización conforme con la Norma ST.26, deberían utilizarse las claves de la caracterización “SITE” o “REGION” con arreglo a la Norma ST.26, según corresponda. Además, para evitar la posibilidad de que se suprima materia, la información contenida en el campo <223>, así como el nombre de característica <221> inapropiado, deberán incluirse en el calificador “NOTE” “note” conexo. Por ejemplo, si en virtud de la Norma ST.25 un listado utiliza un nombre de característica que no es conforme con la Norma ST.25 o la Norma ST.26, como

"Variable" en el campo <221>, junto con la información adicional "Xaa is any amino acid" en el campo <223>, con arreglo a la Norma ST.26 el valor del calificador "NOTE" "note" sería "Variable – Xaa is any amino acid".

c) Cuando una lista de secuencias conforme con la Norma ST.25 no contiene ningún campo <221>, <222> o <223> correspondiente a Xaa o cuando se incluyen los campos <221> y <222> correspondientes a Xaa, pero no se incluye ninguna información en el campo <223> conexo (es decir que, si bien se han producido, ninguna de las dos situaciones hipotéticas es conforme con la Norma ST.25), cualquier información contenida en el cuerpo de la solicitud que describa Xaa debería incluirse en el calificador "NOTE" "note" con arreglo a la Norma ST.26 junto con la clave de caracterización apropiada, por ejemplo, "SITE", "REGION" o "UNSURE", y su localización.

Situación hipotética 6

En virtud de la Norma ST.25, el uracilo está representado en la secuencia mediante "u" y la timina está representada mediante "t". En virtud de la Norma ST.26, el uracilo y la timina están representados en la secuencia mediante "t", sin anotaciones adicionales; "t" se interpreta como uracilo en ARN y timina en ADN.

Recomendaciones:

a) Cuando una secuencia de ADN contiene uracilo, la Norma ST.26 la considera un nucleótido modificado y exige que el uracilo se represente mediante "t" y se describa con más detalle utilizando la clave de caracterización "modified_base", el calificador "mod_base" con "OTHER" como valor calificador y el calificador "note" con "uracil" como valor calificador. En virtud de la Norma ST.26, esta anotación no se considera materia añadida cuando la secuencia de ADN conforme con la Norma ST.25 contiene una nucleobase "u".

b) Cuando una secuencia de ARN contiene timina, la Norma ST.26 la considera un nucleótido modificado y exige que la timina se represente mediante "t" y se describa más detalladamente utilizando la clave de caracterización "modified_base", el calificador "mod_base" con "OTHER" como valor calificador y el calificador "note" con "thymine" como valor calificador. En virtud de la Norma ST.26, esta anotación no se considera materia añadida cuando la secuencia de ARN conforme con la Norma ST.25 contiene una nucleobase "t".

Situación hipotética 7

Tanto en la Norma ST.25 como en la Norma ST.26, los nucleótidos o aminoácidos modificados deben tener una descripción detallada. En virtud de la Norma ST.26, la identidad de un nucleótido modificado podrá indicarse utilizando una de las abreviaturas que figuran en el Cuadro 2 de la Sección 2 del Anexo I, cuando proceda. En caso contrario, deberá indicarse el nombre completo, no abreviado, del nucleótido modificado. Del mismo modo, la identidad de un aminoácido modificado podrá indicarse utilizando una de las abreviaturas que figuran en el Cuadro 4 de la Sección 4 del Anexo I, cuando proceda. En caso contrario, deberá indicarse el nombre completo, no abreviado, del aminoácido modificado. Por el contrario, si un residuo modificado no figura en ningún cuadro de la Norma ST.25 no será necesario utilizar el nombre completo, no abreviado; con frecuencia en estos casos se utiliza su abreviatura.

Recomendaciones:

a) Cuando en la solicitud y en la lista de secuencias conforme con la Norma ST.25 solo se ha utilizado una abreviatura que no figura en el Cuadro 2 de la Sección 2 ni en el Cuadro 4 de la Sección 4 del Anexo I para un nucleótido o aminoácido modificado, y se conoce en la técnica de que se trata que la abreviatura se refiere a un nucleótido o aminoácido modificado específico, el uso del nombre completo, no abreviado, no constituirá en sí mismo adición de materia.

b) Cuando en la solicitud y en la lista de secuencias conformes con la Norma ST.25 solo se ha utilizado una abreviatura que no figura en el Cuadro 2 de la Sección 2 ni en el Cuadro 4 de la Sección 4 del Anexo I para un nucleótido o aminoácido modificado (siempre que la solicitud no contenga ninguna estructura química), y no se conoce en la técnica de que se trata que la abreviatura se refiere a un nucleótido o aminoácido modificado específico, es decir, la abreviatura o bien no se conoce en absoluto en la técnica, o bien podría representar distintos nucleótidos o aminoácidos modificados, no sería posible cumplir lo dispuesto en la Norma ST.26 en esta situación sin la adición de materia. En este caso, por supuesto, la solicitud de prioridad y la lista de secuencias serían imprecisas. Para evitar la posible supresión de materia, en un calificador "nota" o "NOTA" "note" con arreglo a la Norma ST.26 debería figurar la abreviatura de la lista de secuencias conforme con la Norma ST.25, así como el nombre completo, no abreviado, del nucleótido o aminoácido modificado. No se dará prioridad al nombre completo, no abreviado, del nucleótido o aminoácido modificado requerido en la lista de secuencias conforme con la Norma ST. 26 sobre la solicitud anterior. Al redactar la lista de secuencias y la divulgación de la solicitud originales (con arreglo a la Norma ST.25) deberá tenerse cuidado de incluir el nombre no abreviado a fin de evitar problemas en el futuro.

Situación hipotética 8

La Norma ST.25 prevé varias claves de caracterización que no contempla la Norma ST.26. Por lo tanto, los solicitantes deben tener cuidado de capturar la información contenida en dichas claves de caracterización con arreglo a la Norma ST.25 a fin de que sea conforme con la Norma ST.26 sin necesidad de añadir o suprimir materia.

Recomendaciones:

El siguiente cuadro proporciona orientación sobre el modo en que la información contenida en una clave de caracterización con arreglo a la Norma ST.25 guarda conformidad con la Norma ST.26 sin necesidad de añadir o suprimir materia. Los números 1 a 23 corresponden a las claves de caracterización relacionadas con las secuencias de nucleótidos y los números 24 a 43 a las claves de caracterización relacionadas con las secuencias de aminoácidos.

| N.º | Clave de caracterización <221> conforme con la Norma ST. 25 | Equivalente en la Norma ST.26 | | |
|-----|---|-------------------------------|---|---|
| | | Clave de caracterización | Calificador | Valor calificador |
| 1 | allele | misc_feature | allele | valor <223> |
| 2 | attenuator | regulatory ² | regulatory_class ² | "attenuator" |
| | | | note (si <223> está presente) | valor <223> |
| 3 | CAAT_signal | regulatory ² | regulatory_class ² | "CAAT_signal" |
| | | | note (si <223> está presente) | <223> value |
| 4 | conflict | misc_feature | note | "conflict" y valor <223> |
| 5 | enhancer | regulatory ² | regulatory_class ² | "enhancer" |
| | | | note (si <223> está presente) | valor <223> |
| 6 | GC_signal | regulatory ² | regulatory_class ² | "GC_signal" |
| | | | note (si <223> está presente) | valor <223> |
| 7 | LTR | mobile_element ² | rpt_type ² | "long terminal repeat" |
| | | | note (si <223> está presente) | valor <223> |
| 8 | misc_signal | regulatory ² | regulatory_class ² | "other" |
| | | | note (si <223> está presente) | valor <223> |
| 9 | mutation | variation | note | "mutation" y valor <223> |
| 10 | old_sequence | misc_feature | note | "old_sequence" y valor <223> |
| 11 | polyA_signal | regulatory ² | regulatory_class ² | "polyA_signal_sequence" |
| | | | note (si <223> está presente) | valor <223> |
| 12 | promoter | regulatory ² | regulatory_class ² | "promoter" |
| | | | note si <223> está presente) | valor <223> |
| 13 | RBS | regulatory ² | regulatory_class ² | "ribosome_binding_site" |
| | | | note (si <223> está presente) | valor <223> |
| 14 | repeat_unit
a) cuando no se utiliza
"repeat_region" | misc_feature | note | "repeat_unit" y valor <223> |
| | repeat_unit
a) cuando se utiliza
"repeat_region" | repeat_region | rpt_unit_range

note (si <223> está presente) | primer residuo...último residuo

valor <223> |
| 15 | satellite | repeat_region | satellite | "satellite" (o "microsatellite" o "minisatellite" – si se admite) |
| | | | note (si <223> está presente) | valor <223> |
| 16 | scRNA | ncRNA ² | ncRNA_class ² | "scRNA" |
| | | | note (si <223> está presente) | valor <223> |
| 17 | snRNA | ncRNA ² | ncRNA_class ² | "snRNA" |
| | | | note (si <223> está presente) | valor <223> |
| 18 | TATA_signal | regulatory ² | regulatory_class ² | "TATA_box" |
| | | | note (si <223> está presente) | valor <223> |
| 19 | terminator | regulatory ² | regulatory_class ² | "terminator" |
| | | | note (si <223> está presente) | valor <223> |
| 20 | 3'clip | misc_feature | note | "3'clip" y valor <223> |
| 21 | 5'clip | misc_feature | note | "5'clip" y valor <223> |

² La Norma ST.26 puede exigir que una característica específica de la Norma ST.25, por ejemplo, "TATA_signal", sea sustituida por una clave de caracterización, un calificador o un valor más amplios, por ejemplo, "regulatory", "regulatory_class" o "TATA_box". En tal caso, la característica conforme con la Norma ST.25 más restrictiva tendrá prioridad sobre la solicitud anterior. No obstante, no se dará prioridad sobre la solicitud anterior a todo el alcance de la clave de caracterización o el calificador conforme con la Norma ST.26, por ejemplo, "regulatory" o "regulatory_class".

| N.º | Clave de caracterización <221> conforme con la Norma ST. 25 | Equivalente en la Norma ST.26 | | |
|-----|---|--|-------------------------------|--|
| | | Clave de caracterización | Calificador | Valor calificador |
| 22 | -10_signal | regulatory ² | regulatory_class ² | "minus_10_signal" |
| | | | note (si <223> está presente) | valor <223> |
| 23 | -35_signal | regulatory ² | regulatory_class ² | "minus_35_signal" |
| | | | note (si <223> está presente) | valor <223> |
| 24 | NON_CONS | Esta característica se refiere a toda región que contenga un número desconocido de residuos en una única secuencia, lo que está prohibido tanto en virtud de la Norma ST.25 (párrafo 22) como de la Norma ST.26 (párrafo 37). Por consiguiente, toda región de residuos específicamente definidos contemplada en el párrafo 7 de la Norma ST.26 deberá figurar en la lista de secuencias como una secuencia distinta y tener su propio identificador de secuencia. Para evitar la eventual adición o supresión de materia, cada una de estas secuencias debería anotarse de manera que indicara que forma parte de una secuencia mayor que contiene una región indefinida. | | |
| | | REGION | NOTEnote | Descripción |
| | | Descripción – que indica dónde y con qué está enlazada la secuencia, por ejemplo, este residuo presenta un enlace N-terminal a un péptido que tiene un N-terminal Gly-Gly y una región de longitud indefinida. | | |
| 25 | SIMILAR | REGION | NOTEnote | "SIMILAR" y valor <223> si está presente |
| 26 | THIOETH | CROSSLNK | NOTEnote | "THIOETH" y valor <223> si está presente |
| | | Para más información sobre la localización, véase el Anexo I de la Norma ST.26 - Comentario sobre la clave de caracterización «CROSSLNK» | | |
| 27 | THIOLEST | CROSSLNK | NOTEnote | "THIOLEST" y valor <223> si está presente |
| | | Para más información sobre la localización, véase el Anexo I de la Norma ST.26 - Comentario sobre la clave de caracterización «CROSSLNK» | | |
| 28 | VARSP LIC | Examinada en la Situación hipotética 13 más adelante | | |
| 29 | ACETYLATION | MOD_RES | NOTEnote | "ACETYLATION" y valor <223> si está presente |
| | | | NOTEnote | Información exigida por el comentario sobre la clave de caracterización "MOD_RES" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 30 | AMIDATION | MOD_RES | NOTEnote | "AMIDATION" y valor <223> si está presente |
| | | | NOTEnote | Información exigida por el comentario sobre la clave de caracterización "MOD_RES" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 31 | BLOCKED | MOD_RES | NOTEnote | "BLOCKED" y valor <223> si está presente |
| | | | NOTEnote | Información exigida por el comentario sobre la clave de caracterización "MOD_RES" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |

| N.º | Clave de caracterización <221> conforme con la Norma ST. 25 | Equivalente en la Norma ST.26 | | |
|-----|---|-------------------------------|------------------|--|
| | | Clave de caracterización | Calificador | Valor calificador |
| 32 | FORMYLATION | MOD_RES | NOTE note | "FORMYLATION" y valor <223> si está presente |
| | | | NOTE note | Información exigida por el comentario sobre la clave de caracterización "MOD_RES" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 33 | GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION | MOD_RES | NOTE note | "GAMMA-CARBOXYLGLUTAMIC ACID HYDROXYLATION" y valor <223> si está presente |
| | | | NOTE note | Información exigida por el comentario sobre la clave de caracterización "MOD_RES" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 34 | METHYLATION | MOD_RES | NOTE note | "METHYLATION" y valor <223> si está presente |
| | | | NOTE note | Información exigida por el comentario sobre la clave de caracterización "MOD_RES" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 35 | PHOSPHORYLATION | MOD_RES | NOTE note | "PHOSPHORYLATION" y valor <223> si está presente |
| | | | NOTE note | Información exigida por el comentario sobre la clave de caracterización "MOD_RES" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 36 | PYRROLIDONE CARBOXYLIC ACID | MOD_RES | NOTE note | "PYRROLIDONE CARBOXYLIC ACID" y valor <223> si está presente |
| | | | NOTE note | Información exigida por el comentario sobre la clave de caracterización "MOD_RES" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 37 | SULFATATION | MOD_RES | NOTE note | "SULFATATION" y valor <223> si está presente |
| | | | NOTE note | Información exigida por el comentario sobre la clave de caracterización "MOD_RES" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 38 | MYRISTATE | LIPID | NOTE note | "MYRISTATE" y valor <223> si está presente |
| | | | NOTE note | Información exigida por el comentario sobre la clave de caracterización "LIPID" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 39 | PALMITATE | LIPID | NOTE note | "PALMITATE" y valor <223> si está presente |
| | | | NOTE note | Información exigida por el comentario sobre la clave de caracterización "LIPID" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |

| N.º | Clave de caracterización <221> conforme con la Norma ST. 25 | Equivalente en la Norma ST.26 | | |
|-----|---|-------------------------------|----------------------|--|
| | | Clave de caracterización | Calificador | Valor calificador |
| 40 | FARNESYL | LIPID | NOTE ^{note} | "FARNESYL" y valor <223> si está presente |
| | | | NOTE ^{note} | Información exigida por el comentario sobre la clave de caracterización "LIPID" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 41 | GERANYL-GERANYL | LIPID | NOTE ^{note} | "GERANYL-GERANYL" y valor <223> si está presente |
| | | | NOTE ^{note} | Información exigida por el comentario sobre la clave de caracterización "LIPID" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 42 | GPI-ANCHOR | LIPID | NOTE ^{note} | "GPI-ANCHOR" y valor <223> si está presente |
| | | | NOTE ^{note} | Información exigida por el comentario sobre la clave de caracterización "LIPID" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |
| 43 | N-ACYL DIGLYCERIDE | LIPID | NOTE ^{note} | "N-ACYL DIGLYCERIDE" y valor <223> si está presente |
| | | | NOTE ^{note} | Información exigida por el comentario sobre la clave de caracterización "LIPID" del Anexo I de la Norma ST.26 (ni necesidad de añadir materia) |

Situación hipotética 9

Algunas claves de caracterización contempladas en la Norma ST.25 y la Norma ST.26, tanto para las secuencias de nucleótidos como de aminoácidos, tienen calificadores obligatorios conformes con la Norma ST.26, como se indica a continuación. La clave de caracterización de secuencias de nucleótidos "modified base" también está presente en la Norma ST.25 y la Norma ST.26; sin embargo, las recomendaciones apropiadas figuran en la situación hipotética 7. Si bien la Norma ST.25 no contempla ningún calificador, prevé un campo <223> para el texto libre. Cuando la información contenida en un campo <223> conforme a la Norma ST.25 es apropiada como valor para el calificador obligatorio con arreglo a la Norma ST.26, entonces la información debería incluirse tal cual figura. Cuando no se ha proporcionado un campo <223> conforme con la Norma ST.25 o este contiene información que no es apropiada como valor para el calificador obligatorio con arreglo a la Norma ST.26, el solicitante deberá tener cuidado de capturar la información contenida en el campo <223> correspondiente a la clave de caracterización de manera que se ajuste a la Norma ST.26 sin tener que añadir o suprimir materia.

Secuencias de nucleótidos³

| Clave de caracterización | Calificador obligatorio |
|--------------------------|-------------------------|
| 5.12 - misc_binding | 6.3 - bound_moiety |
| 5.30 - protein_bind | 6.3 - bound_moiety |

Recomendaciones:

- a) Si el campo <223> con arreglo a la Norma ST.25 falta o es inadecuado y en la descripción de la solicitud se divulga el nombre de una molécula o complejo que puede unirse a la localización de característica del ácido nucleico, dicho nombre debería insertarse en el calificador "bound_moiety".
 - i) Toda información contenida en el campo <223> con arreglo a la Norma ST.25 cuya inclusión en el calificador obligatorio "bound_moiety" no sea apropiada debería insertarse en otro calificador apropiado de la clave de caracterización, por ejemplo, "note".
- b) Si el campo <223> con arreglo a la Norma ST.25 falta o es inapropiado y en la descripción de la solicitud no se divulga el nombre de la molécula o complejo que puede unirse a la localización de característica del ácido nucleico, debería utilizarse en su lugar la clave de caracterización "misc_feature" de la Norma ST.26 con el calificador "note".

³ Las referencias numéricas de la tabla siguiente remiten a los números de las claves de caracterización y los calificadores que figuran en el Anexo I de la Norma ST.26 - Vocabulario controlado.

- i) Si el campo <223> con arreglo a la Norma ST.25 falta, el valor del calificador "note" debería ser el nombre de la clave de caracterización conforme con la Norma ST.25;
- ii) Si el campo <223> con arreglo a la Norma ST.25 contiene información inapropiada, el valor del calificador "note" debería ser el nombre de la clave de caracterización conforme con la Norma ST.25 y la información contenida en el campo <223>.

Secuencias de aminoácidos⁴

| Clave de caracterización | Calificador obligatorio |
|--------------------------|-------------------------|
| 7.2 – BINDING | 8.2 – NOTE note |
| 7.4 – CARBOHYD | 8.2 – NOTE note |
| 7.10 – DISULFID | 8.2 – NOTE note |
| 7.11 – DNA_BIND | 8.2 – NOTE note |
| 7.12 – DOMAIN | 8.2 – NOTE note |
| 7.16 – LIPID | 8.2 – NOTE note |
| 7.17 – METAL | 8.2 – NOTE note |
| 7.18 – MOD_RES | 8.2 – NOTE note |
| 7.23 – NP_BIND | 8.2 – NOTE note |
| 7.29 – SITE | 8.2 – NOTE note |
| 7.39 – ZN_FING | 8.2 – NOTE note |

Recomendaciones:

- a) Si el valor del campo <223> con arreglo a la Norma ST.25 falta o es inapropiado y la descripción de la solicitud divulga información específica exigida por el calificador obligatorio, dicha información debería incluirse en el calificador obligatorio "NOTE" "note".
- i) Toda información contenida en el campo <223> con arreglo a la Norma ST.25 cuya inclusión en el calificador obligatorio "NOTE" "note" no sea apropiada (véanse la definición y el comentario de la clave de caracterización) debería incluirse en un segundo calificador "NOTE" "note".
- b) Si el valor del campo <223> de la Norma ST.25 falta o es inapropiado y la descripción de la solicitud no divulga la información específica exigida en el calificador obligatorio, debería utilizarse en su lugar la clave de caracterización "SITE" (para un aminoácido) o "REGION" (para un grupo de aminoácidos) conforme con la Norma ST.26 con el calificador "NOTE" "note".
- i) Si falta el campo <223> con arreglo a la Norma ST.25, el valor del calificador "NOTE" "note" debería ser el nombre de la clave de caracterización conforme con la Norma ST.25;
- ii) Si el campo <223> con arreglo a la Norma ST.25 contiene información inapropiada, el valor del calificador "NOTE" "note" debería ser el nombre de la clave de caracterización conforme con la Norma ST.25 y la información contenida en el campo <223>.

Situación hipotética 10

En virtud de la Norma ST.25, cada clave de caracterización específica contiene un campo <222> para indicar una localización de característica; sin embargo, no se exige una indicación de la localización respecto de la mayoría de las características ni un formato de información de localización normalizado. Además, la Norma ST.25 no prevé operadores de localización, como "join". La Norma ST.26 contempla descriptores de localización y operadores normalizados y cada característica deberá contener al menos un descriptor de localización. (Las caracterizaciones CDS son un caso especial y se examinan más adelante, en la Situación hipotética 11).

Recomendaciones:

- a) Si la lista de secuencias con arreglo a la Norma ST.25 contiene el campo <222>, la importación directa o al formato de la Norma ST.26 no conlleva la necesidad de añadir materia.
- b) Si la lista de secuencias con arreglo a la Norma ST.25 no contiene el campo <222>, pero la información sobre la ubicación figura en la descripción de la solicitud, la importación directa o al formato ST.26 no conlleva la necesidad de añadir materia.
- c) Si ni la lista de secuencias con arreglo a la Norma ST.25 ni la descripción de la solicitud contienen información sobre la localización, es de suponer que la característica se aplica a toda la secuencia. (Indicar una

⁴ Las referencias numéricas de la tabla siguiente remiten a los números de las claves de caracterización y los calificadores que figuran en el Anexo I de la Norma ST.26 - Vocabulario controlado.

localización que sea menor que la secuencia completa sin apoyo en la descripción de la solicitud probablemente constituirá adición o supresión de materia. Al redactar la lista de secuencias y la divulgación de la solicitud originales (con arreglo a la Norma ST.25) deberá tenerse cuidado de incluir la información sobre la localización en la medida de lo posible a fin de evitar problemas en el futuro.

Situación hipotética 11

En la Norma ST.25, una secuencia codificante que codifica un único polipéptido contiguo pero que está interrumpida por una o más secuencias no codificantes, por ejemplo, intrones, se indica como múltiples caracterizaciones CDS individuales, como se ilustra a continuación:

```
<220>
<221> CDS
<222> (1)..(571)

<220>
<221> CDS
<222> (639)..(859)
```

Por el contrario, la Norma ST.26 prevé un operador de localización "join" que indica que los polipéptidos codificados por las localizaciones indicadas se unen para formar un único polipéptido contiguo. (Nota: tanto la Norma ST.25 como la Norma ST.26 exigen que el codón de terminación se incluya en la localización de la caracterización "CDS").

Recomendaciones:

- a) Si en la lista de secuencias conforme con la Norma ST.25 o la descripción de la solicitud se indica claramente que las secuencias de polipéptidos codificadas por las múltiples caracterizaciones "CDS" individuales forman un único polipéptido contiguo, deberá representarse una secuencia de codificación interrumpida por un intrón en una única caracterización "CDS" con el operador de localización "join", como se ilustra a continuación, de manera que no será necesario añadir materia:

```
<INSDFeature_key>CDS</INSDFeature_key>
<INSDFeature_location>join(1..571,639..859)</INSDFeature_location>
```

- b) Si la lista de secuencias conforme con la Norma ST.25 o la descripción de la solicitud no indican que las secuencias de polipéptidos codificadas por las dos caracterizaciones CDS distintas forman un único polipéptido contiguo, es probable que el uso del operador de localización "join" constituya adición de materia.

Situación hipotética 12

La Norma ST.25 establece que los nombres de característica deben figurar en los Cuadros 5 o 6. Sin embargo, la normativa de los Estados Unidos de América recomienda la utilización de dichos nombres, pero no obliga a ella. Por lo tanto, una secuencia que forme parte de una lista de secuencias conforme con la Norma ST.25 (y cumpla la normativa de los Estados Unidos) podría tener una clave de caracterización "personalizada" que carezca de la correspondiente clave de caracterización conforme con la Norma ST.26. También es posible que no se haya proporcionado ningún nombre de característica para el campo <221> o que el campo <221> falte. Estas situaciones hipotéticas pueden abordarse de manera similar.

Recomendación:

La clave de caracterización "custom" de la Norma ST.25 puede representarse en una lista de secuencias conforme con la Norma ST.26 sin que sea necesario añadir materia, como se indica a continuación:

| Tipo | Clave de caracterización <221> conforme con la Norma ST.25 | Posible equivalente en la Norma ST.26 | | |
|------|--|---------------------------------------|----------------------|--|
| | | Clave de caracterización | Calificador | Valor del calificador |
| NA | Clave de caracterización "custom" | misc_feature | note | Nombre de clave de caracterización "custom" y valor <223> si está presente |
| AA | Clave de caracterización "custom" | SITE o REGION | NOTE note | Nombre de clave de caracterización "custom" y valor <223> si está presente |

Situación hipotética 13

La Norma ST.25 prevé la clave de caracterización "VARSPPLIC", definida como "descripción de las variantes de la secuencia producidas por un empalme diferencial". En virtud de la Norma ST.26, "VARSPPLIC" se ha sustituido por la clave de caracterización más amplia "VAR_SEQ", definida como "descripción de las variantes producidas por empalme

diferencial, utilización de promotor alternativo, iniciación alternativa y desplazamiento del marco ribosómico". Por lo tanto, una lista de secuencias conforme con la Norma ST.26 no debería utilizar "VAR_SEQ" como sustituto de "VARSPPLIC" sin más explicaciones.

Recomendación:

En virtud de la Norma ST.26, la clave de caracterización "VAR_SEQ" debería utilizarse con el calificador "NOTE" "note", cuyo valor debería incluir una explicación de su ámbito de aplicación más limitado en virtud de la norma ST.25, por ejemplo, "variante de la secuencia producida por un empalme diferencial". Toda información adicional contenida en un campo <223> conexo conforme con la Norma ST.25 debería incluirse también en el calificador "NOTE" "note".

Situación hipotética 14

Si la fuente de una secuencia es artificial, el campo <213> (Organism) conforme con la Norma ST.25 exige el calificador "Artificial Sequence". En virtud de la Norma ST.26, la clave de caracterización "source" "SOURCE" exige el calificador "organism" "ORGANISM", cuyo valor deberá indicarse como "synthetic construct" en lugar de "Artificial Sequence".

Recomendación:

En virtud de la Norma ST.26, deberá indicarse "synthetic construct" como valor del calificador "organism" "ORGANISM". Para evitar la posible supresión de materia, toda información explicativa contenida en el campo requerido <223> con arreglo a la Norma ST.25 debería incluirse como un calificador "note" "NOTE" (de la clave de caracterización "source" "SOURCE").

Situación hipotética 15

Si se desconoce el nombre científico del organismo fuente de una secuencia, el campo <213> ("organism") conforme con la Norma ST.25 requerirá el calificador "Unknown". En virtud de la Norma ST.26, la clave de caracterización "source" "SOURCE" requerirá el calificador "organism" "ORGANISM", cuyo valor deberá indicarse como "unidentified" en lugar de "Unknown".

Recomendación:

En virtud de la Norma ST.26, deberá indicarse "unidentified" como valor del calificador "organism" "ORGANISM". A fin de evitar la posible supresión de indicaciones, toda información explicativa contenida en el campo obligatorio <223> conforme con la Norma ST.25 debería incluirse en un calificador "note" "NOTE" (de la clave de caracterización "source" "SOURCE").

Situación hipotética 16

La Norma ST.25 permite que la enumeración de los aminoácidos pueda tener números negativos contados en forma regresiva desde el aminoácido adyacente al número 1 para los aminoácidos que precedan a la proteína madura, por ejemplo, las presecuencias, las prosecuencias y las preprosecuencias, así como las secuencias señal. La Norma ST.26 no permite que la localización de característica contenga números negativos.

Recomendaciones:

- a) Si la lista de secuencias conforme con la Norma ST.25 contiene una o más características representadas en un campo <221> y un campo <222> conexo que contiene números negativos y/o positivos, por ejemplo, "PROPEP" y/o "CHAIN", en la lista de secuencias conforme con la Norma ST.26 debería utilizarse una clave de característica apropiada, por ejemplo, "PROPEP" y/o "CHAIN". Puede utilizarse un calificador "NOTE" "note" con la información que figura en un campo <223>, si existe, como valor del calificador;
- b) Si la lista de secuencias conforme con la Norma ST.25 no contiene una o más características representadas en un campo <221> y un campo <222> conexo, sino que la información relativa a los números negativos y/o positivos figura en la descripción de la solicitud, por ejemplo, "PROPEP" y/o "CHAIN", en la lista de secuencias conforme con la Norma ST.26 debería utilizarse la clave de caracterización apropiada, por ejemplo, "PROPEP" y/o "CHAIN". Puede utilizarse un calificador "NOTE" "note" con la información que figura en la descripción de la solicitud, si existe, como valor del calificador;
- c) Si ni la lista de secuencias con arreglo a la Norma ST.25 ni la descripción de la solicitud contienen información que explique la razón de la numeración negativa y/o positiva, para evitar la posible supresión de materia de la lista de secuencias conforme con la Norma ST.26 deberá utilizarse la clave de característica "REGION" cuando la localización de característica abarque la región de la secuencia conforme con la Norma ST.25 con numeración negativa. Además, debería utilizarse un calificador "NOTE" "note" para indicar que la secuencia de aminoácidos está numerada negativamente en la lista de secuencias de la solicitud conforme con la Norma ST.25 cuya prioridad se reivindica.

Situación hipotética 17

La Norma ST.25 prevé la inclusión de información sobre la publicación en los campos <300> a <313>. La Norma ST.26 no prevé la inclusión de dicha información.

Recomendación:

En virtud de la Norma ST.25, la información contenida en los campos <300> a <313> debería incluirse en el cuerpo de la solicitud conexas, si no figura ya en él.

Situación hipotética 18

La Norma ST.25 no prevé un método normalizado para indicar que una región CDS de una secuencia de nucleótidos debe traducirse utilizando una tabla del código genético distinta de la tabla del código genético normalizada. Por el contrario, la Norma ST.26 prevé el calificador "transl_table", que puede utilizarse con la clave de caracterización CDS para indicar que la región debe traducirse utilizando una tabla del código genético alternativa. Cuando no se utiliza el calificador "transl_table" se asume la utilización de la tabla del código genético normalizada.

Recomendaciones:

- a) Si la lista de secuencias con arreglo a la Norma ST.25 o la descripción de la solicitud indican claramente que una región CDS debe traducirse utilizando una tabla del código genético alternativa, deberá utilizarse el calificador "transl_table" con el número de la tabla del código genético apropiada como valor calificador. Si no se utiliza el calificador "transl_table", es probable que constituya adición de materia, ya que por defecto se asumiría la tabla de "Standard Code". La omisión en la lista de secuencias con arreglo a la Norma ST.26 de la información sobre la tabla del código genético alternativa referenciada en la lista de secuencias con arreglo a la Norma ST.25 o la descripción de la solicitud probablemente conllevaría la supresión de materia.
- b) Si la lista de secuencias con arreglo a la Norma ST.25 o la descripción de la solicitud no indican que una región CDS debe traducirse utilizando una tabla del código genético alternativa, no debería utilizarse el calificador "transl_table", o debería utilizarse únicamente con el valor calificador "1", es decir, con la tabla de códigos normalizados. La utilización del calificador "transl_table" con un valor calificador distinto de "1" probablemente conllevaría la supresión o adición de materia.

Situación hipotética 19

La Norma ST.25 no prevé un método normalizado para indicar la localización de una característica, en particular, una contenida en un sitio o región que se extiende más allá del residuo o de la serie de residuos que se ha especificado, por ejemplo, una región CDS de una secuencia de nucleótidos que se extiende más allá de uno o ambos extremos de una secuencia divulgada. Por el contrario, el descriptor de localización de la característica con arreglo a la Norma ST.26 prevé un método normalizado para indicar la localización de dicho sitio o región utilizando los símbolos "<" o ">". Por ejemplo, la localización de característica "CDS" deberá incluir el codón de terminación, incluso cuando este no esté incluido en la propia secuencia divulgada, por ejemplo, indicando la localización 1..>321.

Recomendaciones:

- a) Cuando la lista de secuencias conforme con la Norma ST.25 no indica explícitamente que la localización de una característica se extiende más allá de la secuencia, pero dicha localización está fundamentada en la divulgación o se desprende claramente de la propia secuencia, por ejemplo, el codón de terminación de una característica "CDS" que no esté contenido en la secuencia, pueden utilizarse los símbolos "<" o ">" en la lista de secuencias con arreglo a la ST.26 sin necesidad de añadir materia.
- b) Cuando la lista de secuencias conforme con la Norma ST.25 no indica explícitamente que la localización de una característica se extiende más allá de la secuencia y dicha localización no esté fundamentada en la divulgación ni se desprenda claramente de la propia secuencia, el cumplimiento de la Norma ST.26 puede no ser posible si no se añade materia. En esta situación, la solicitud de prioridad y la lista de secuencias podrían considerarse incompletas y la descripción de la localización de característica en la lista de secuencias conforme con la Norma ST.26 no tendrá prioridad sobre la solicitud anterior. Al redactar la lista de secuencias y la divulgación de la solicitud originales (con arreglo a la Norma ST.25) deberá tenerse cuidado de incluir la información completa sobre las características.

Situación hipotética 20

El Apéndice I de la Norma ST. 25 exige que, cuando una secuencia de nucleótidos contiene tanto fragmentos de ADN como de ARN, el valor en el campo <212> será "ADN" y la molécula combinada de ADN/ARN deberá describirse con más detalle en la sección de características, en los campos <220> a <223>; sin embargo, la naturaleza exacta de dicha descripción detallada no está clara y este requisito no se sigue de forma rutinaria. El párrafo 55 de la Norma ST.26 exige que cada segmento (la Norma ST.26 utiliza el término "segmento" en lugar de "fragmento" en aras de la coherencia interna) de ADN y ARN de la molécula combinada se describa con la clave de caracterización "misc_feature", que incluye la localización del segmento, y el calificador "note", que indica si el segmento es ADN o ARN.

Recomendaciones:

- a) Si en la lista de secuencias conforme con la Norma ST.25 se describen los segmentos de ADN y ARN mediante una o más características utilizando la variable "misc_feature" en el campo <221>, las localizaciones

apropiadas en el campo <222> y las indicaciones que identifican qué segmentos son ADN o ARN en el campo <223>, la adaptación de esa información al formato de la Norma ST.26 utilizando "misc_feature" para cada segmento de ADN y de ARN no debería conllevar la necesidad de añadir materia.

- b) Si en la lista de secuencias conforme con la Norma ST.25 se describen los segmentos de ADN y ARN mediante una o más características utilizando una clave de caracterización distinta de "misc_feature" en el campo <221>, las localizaciones apropiadas en el campo <222> y las indicaciones que identifican qué segmentos son ADN o ARN en el campo <223>, la adaptación de esa información al formato de la Norma ST.26 utilizando la función "misc_feature" para cada segmento de ADN y ARN y un calificador "note" adicional con la clave de caracterización original en el campo <221> no debería conllevar la necesidad de añadir materia.
- c) Si la lista de secuencias conforme con la Norma ST.25 prevé que la identidad (ADN o ARN) y la localización de cada segmento se incluya en un campo <223> que no esté asociado a los campos <221> y <222>, por ejemplo, para representar una secuencia artificial, la adaptación de esa información al formato de la Norma ST.26 utilizando la función "misc_feature" para cada segmento de ADN y ARN no debería conllevar la necesidad de añadir materia.
- d) Si la lista de secuencias conforme con la Norma ST.25 describe la molécula en el cuadro de características utilizando "misc_feature" en los campos <221> y <223> para indicar que la molécula es una molécula combinada de ADN/ARN, pero no proporciona información sobre la localización de cada segmento, y
- i) si en la descripción figuran las localizaciones de cada segmento de ADN y ARN, la adaptación de dicha información al formato de la Norma ST.26 utilizando "misc_feature" para cada segmento de ADN y ARN no debería conllevar la necesidad de añadir materia.
- ii) Si en la descripción no figura la información sobre la localización de cada segmento de ADN y ARN, quizás no sea posible cumplir con las disposiciones de la Norma ST.26 sin añadir materia. En este caso, la solicitud de prioridad y la lista de secuencias podrían considerarse incompletas y las descripciones de localización de características contenidas en la lista de secuencias conforme con la Norma ST.26 no tendrán prioridad sobre la solicitud anterior. Al redactar la lista de secuencias y la divulgación de la solicitud originales (con arreglo a la Norma ST.25) deberá tenerse cuidado de incluir la información completa sobre las características.
- e) Si la lista de secuencias conforme con la Norma ST.25 describe una molécula utilizando una clave de caracterización distinta de "misc_feature" en el campo <221> e indicando que dicha molécula es una molécula combinada de ADN/ARN en el campo <223>, pero no proporciona información sobre la localización de cada segmento, y
- i) si en la descripción se indica la localización de cada segmento de ADN y ARN, la adaptación de dicha información al formato de la Norma ST.26 utilizando una variable "misc_feature" para cada segmento de ADN y ARN y un calificador "note" adicional con la clave de caracterización original como valor en el campo <221>, ello no debería conllevar la necesidad de añadir materia;
- ii) si la descripción no contiene la información sobre la localización de cada segmento de ADN y ARN, quizás no sea posible cumplir los requisitos de la Norma ST.26 sin añadir materia. En esta situación, la solicitud de prioridad y la lista de secuencias podrían considerarse incompletas y las descripciones de localización de las características mencionadas en la lista de secuencias conformes con la Norma ST.26 no tendrán prioridad sobre la solicitud anterior. Al redactar la lista de secuencias y la divulgación de la solicitud originales (con arreglo a la Norma ST.25) deberá tenerse cuidado de incluir la información completa sobre las características.
- f) Si en la lista de secuencias conforme con la Norma ST.25 se indica que la molécula es una molécula combinada de ADN/ARN en un campo <223>, por ejemplo, para representar una secuencia artificial, pero no se proporciona ninguna clave de caracterización ni información sobre la localización de cada segmento, y
- i) si en la descripción se indican las localizaciones de cada segmento de ADN y ARN, la incorporación de dicha información al formato de la Norma ST.26 utilizando "misc_feature" para cada segmento de ADN y ARN no debería conllevar la necesidad de añadir materia;
- ii) Si la descripción no contiene información sobre la localización de cada segmento de ADN y ARN, quizás no sea posible cumplir los requisitos de la Norma ST.26 sin añadir materia. En este caso, la solicitud de prioridad y la lista de secuencias podrían considerarse incompletas y las descripciones de localización sobre las características de la lista de secuencias conforme con la Norma ST.26 no tendrán prioridad sobre la solicitud anterior. Al redactar la lista de secuencias y la divulgación de la solicitud originales (con arreglo a la Norma ST.25) deberá tenerse cuidado de incluir la información completa sobre las características.

[Fin del Anexo VII y de la Norma]

[Sigue el Anexo II (Anexo III de la Norma ST.26)]