

# ■ Norma ST.26 da OMPI: AVANÇADO

## Webinário de formação

*Nota: Este é um módulo avançado e assume-se que os participantes já completaram o módulo de introdução à norma ST.26.*

*O curso de introdução à norma ST.26 pode ser acessado aqui:  
[https://www.wipo.int/meetings/en/details.jsp?meeting\\_id=62848](https://www.wipo.int/meetings/en/details.jsp?meeting_id=62848)*

# O que vamos cobrir hoje

- Chaves de caracterização e qualificadores normalmente usados
- Formatos das localizações de características
- Formatos dos valores de qualificadores e dos valores de qualificadores não em inglês
- Situações especiais - uracilo em ADN e timina em ARN: moléculas híbridas de ADN/ARN
- Análogos de nucleotídeos, aminoácidos D e sequências ramificadas
- Variantes de sequências

# Chaves de caracterização e qualificadores normalmente usados

# Chaves de caracterização e qualificadores

- As chaves de caracterização podem ser usadas para descrever em maior detalhe um ou mais resíduos de uma sequência identificada por sua localização
  - As chaves de caracterização das sequências de nucleotídeos estão enumeradas no Anexo I, Seção 5
  - As chaves de caracterização das sequências de nucleotídeos são escritas em minúsculas; por exemplo, "misc\_binding"
  - As chaves de caracterização das sequências de aminoácidos estão enumeradas no Anexo I, Seção 7
  - As chaves de caracterização das sequências de aminoácidos são escritas em maiúsculas; por exemplo, "REGION"
  
- Qualificadores podem ser usados para descrever características em maior detalhe
  - Os qualificadores das sequências de nucleotídeos estão enumerados no Anexo I, Seção 6
  - Os qualificadores das sequências de nucleotídeos são escritos em minúsculas; por exemplo, "allele"
  - Os qualificadores das sequências de aminoácidos estão enumerados no Anexo I, Seção 8
  - Os qualificadores de sequências de aminoácidos são escritos em maiúsculas; por exemplo, "NOTE"

# Chaves de caracterização e qualificadores

```

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

```

# Chaves de caracterização e qualificadores

- As chaves de caracterização são opcionais, com exceção da característica "source"/"SOURCE", que é obrigatória para cada sequência.
- Cada chave de caracterização tem uma lista de qualificadores que podem ser usados para descrever a característica em maior detalhe. A maioria dos qualificadores são opcionais; no entanto, algumas chaves de caracterização têm qualificadores obrigatórios.

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

- Os qualificadores "mol\_type"/"MOL\_TYPE" e "organism"/"ORGANISM" são obrigatórios para a característica "source"/"SOURCE".

# Chaves de caracterização e qualificadores

- Algumas chaves de caracterização têm limitações adicionais
  - âmbito de organismos; por exemplo, "C\_region" se limita a eucariotas
  - âmbito de moléculas; por exemplo, "D-loop" se limita a sequências de ADN

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

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "modified\_base"

- A chave de caracterização "modified\_base" e seu qualificador obrigatório "mod\_base" devem ser usados para descrever um nucleotídeo modificado (ST.26, parágrafo 16)
- Um "nucleotídeo modificado" é qualquer nucleotídeo que não seja:
  - deoxi-[a, g, c, or t] 3'-monofosfato
  - [a, g, c, or u] 3'-monofosfato (ST.26, parágrafo 3(f))
- Um "nucleotídeo modificado" deve ser representado sempre que possível pelo nucleotídeo não-modificado correspondente (no Anexo I, Seção 1, Tabela 1). Caso contrário, pode ser representado por "n". Por exemplo, "2'-O-metilcitidina" deve ser representada por "c" na sequência. "Queuosina" deve ser representada por "n". O símbolo "n" equivale a apenas um resíduo.
- O valor do qualificador obrigatório "mod\_base" deve ser selecionado entre os valores no Anexo I, Seção 2, Tabela 2. Se o valor for "other" (outro), então um qualificador "note" adicional deve conter o nome completo e não abreviado do resíduo modificado.



# Chaves de caracterização e qualificadores

Sequências de nucleotídeos: chave de caracterização "modified\_base"

- Exemplo: sequência de nucleotídeos com inosina na posição 15

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

- "Inosine
- " figura no Anexo I, Seção 2, Tabela 2, com a abreviatura "i"

# Chaves de caracterização e qualificadores

Sequências de nucleotídeos: chave de caracterização "modified\_base"

- Exemplo: sequência de nucleotídeos com xantina na posição 22

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>22</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>
```

- A "xantina" NÃO figura no Anexo I, Seção 2, Tabela 2; portanto, o valor do qualificador obrigatório "mod\_base" deve ser "OTHER" e um qualificador "note" adicional deve ser incluído com o valor "xantina".

# Chaves de caracterização e qualificadores

Sequências de nucleotídeos: chave de caracterização "modified\_base"

- "modified\_base" também pode ser usado para descrever um sítio abásico:

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

■ A sequência é caracterizada e descrita em maior detalhe por uma chave de caracterização "modified\_base" junto com um qualificador "mod\_base" com o valor "OTHER" e um qualificador "note" adicional com o valor "abasic site".

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- A chave de caracterização "CDS" pode ser usada para identificar sequências de codificação. A localização da característica CDS deve incluir o códon de terminação. (ST.26, par. 89)
- NÃO há qualificadores obrigatórios para a chave de caracterização "CDS"
- Os qualificadores normalmente utilizados incluem:

"pseudo"

"pseudogene"

"translation"

"transl\_table"

"codon\_start"

"transl\_except"

"protein\_id"

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Uma característica "CDS" pode ter apenas UM dos qualificadores seguintes:

Qualificador	Descrição	Valor
pseudo	indica que a característica CDS é não funcional e não tem tradução mas não é um pseudogene	Nenhum
pseudogene	indica que a característica CDS é um pseudogene e não tem tradução	processado não processado unitário alélico desconhecido
tradução	indica a sequência de aminoácidos derivada da tradução da CDS	abreviaturas de aminoácidos de uma letra

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Uma sequência de aminoácidos contendo 4 ou mais aminoácidos especificamente definidos codificados por uma sequência de codificação e divulgados em um qualificador de "tradução" deve ser incluída na listagem de sequências como uma sequência separada à qual seu próprio número identificador de sequência é atribuído.
- O número identificador de sequência atribuído à sequência de aminoácidos deve ser o valor fornecido como qualificador "protein\_id" dentro da chave de caracterização "CDS".
- O qualificador "ORGANISM" da chave de caracterização "SOURCE" da sequência de aminoácidos deve ser idêntico ao da sua sequência de codificação. (ST.26, par. 92)

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Qualificadores que podem modificar a sequência traduzida:

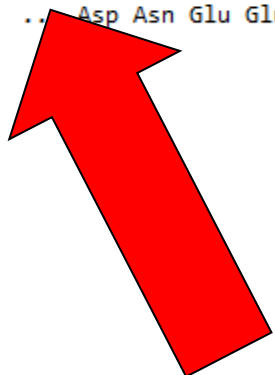
Qualificador	Descrição	Valor
transl_table	indica a tabela de códigos genéticos utilizada para traduzir a CDS; por predefinição é "1-Standard Code"	um número que corresponde a uma tabela de tradução no Anexo I, Seção 9
transl_except	indica a tradução de um códon que não está em conformidade com o código genético definido na "transl_table"	(pos:<location>, aa:<amino_acid>)
codon_start	indica a fase de leitura da CDS em relação à primeira base	1, 2 ou 3

# Chaves de caracterização e qualificadores

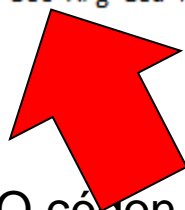
## Sequências de nucleotídeos: chave de caracterização "CDS"

- Exemplo - SEQ ID NO:1 é um fragmento de uma sequência de codificação proveniente de um gene mitocondrial de levedura (tabela de códigos genéticos "3-Yeast Mitochondrial Code"):

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



A sequência começa com um códon parcial



O códon nas posições 30-32  
codifica a selenocisteína  
(Sec)

- Que informações devem ser incluídas numa característica CDS para representar com exatidão esta sequência?



# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Exemplo - SEQ ID NO:1 é um fragmento de uma sequência de codificação proveniente de um gene mitocondrial de levedura (tabela de códigos genéticos "3-Yeast Mitochondrial Code"):

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

- Sequência representada na listagem de sequências como:

```
tggataatga agaagttaac gaagaatgta tgagattatt ttcaagaac gctcgtcatc taacatcaag gttgacataa
```

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Exemplo - SEQ ID NO:1 é um fragmento de uma sequência de codificação proveniente de um gene mitocondrial de levedura (tabela de códigos genéticos "3-Yeast Mitochondrial Code"):

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

- Chave de caracterização "CDS"
- Localização da característica: <1..80

O "<" indica que a região de codificação começa antes da posição 1

A localização inclui o códon de terminação nas posições 78-80

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Exemplo - SEQ ID NO:1 é um fragmento de uma sequência de codificação proveniente de um gene mitocondrial de levedura (tabela de códigos genéticos "3-Yeast Mitochondrial Code"):

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

- Chave de caracterização "CDS"
- Localização da característica:  
<1..80
- qualificador "codon\_start" com o valor "3"

O valor 3 do qualificador "codon\_start" indica que o primeiro códon completo começa na terceira posição dentro da localização

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Exemplo - SEQ ID NO:1 é um fragmento de uma sequência de codificação proveniente de um gene mitocondrial de levedura (tabela de códigos genéticos "3-Yeast Mitochondrial Code"):

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

- Chave de caracterização "CDS"
- Localização da característica: <1..80
- qualificador "codon\_start" com o valor "3"
- qualificador "transl\_table" com o valor "3"

As tabelas de códigos genéticos no Anexo 1, Seção 9, são utilizadas para determinar o valor do qualificador "transl\_table".

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Exemplo - SEQ ID NO:1 é um fragmento de uma sequência de codificação proveniente de um gene mitocondrial de levedura (tabela de códigos genéticos "3-Yeast Mitochondrial Code"):

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

- Chave de caracterização "CDS"
- Localização da característica: <1..80
- qualificador "codon\_start" com o valor "3"
- qualificador "transl\_table" com o valor "3"
- qualificador "transl\_except" com o valor "(pos:30..32,aa:Sec)"

O qualificador  
"transl\_except" coloca o  
aminoácido selenocisteína

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Exemplo - SEQ ID NO:1 é um fragmento de uma sequência de codificação proveniente de um gene mitocondrial de levedura (tabela de códigos genéticos "3-Yeast Mitochondrial Code"):

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

- Chave de caracterização "CDS"
- Localização da característica: <1..80
- qualificador "codon\_start" com o valor "3"
- qualificador "transl\_table" com o valor "3"
- qualificador "transl\_except" com o valor "(pos:30..32,aa:Sec)"
- qualificador "translation" com o valor "**DNEEVNEECURLFFKNARHTTSRLT**"

O códon de terminação não aparece no qualificador "translation"!

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Exemplo - SEQ ID NO:1 é um fragmento de uma sequência de codificação proveniente de um gene mitocondrial de levedura (tabela de códigos genéticos "3-Yeast Mitochondrial Code"):

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

- Chave de caracterização "CDS"
- Localização da característica: <1..80
- qualificador "codon\_start" com o valor "3"
- qualificador "transl\_table" com o valor "3"
- qualificador "transl\_except" com o valor "(pos:30..32,aa:Sec)"
- qualificador "translation" com o valor "DNEEVNEECURLFFKNARHTTSRLT"
- uma sequência individual de proteínas para a tradução

# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- Exemplo - SEQ ID NO:1 é um fragmento de uma sequência de codificação proveniente de um gene mitocondrial de levedura (tabela de códigos genéticos "3-Yeast Mitochondrial Code"):

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

- Chave de caracterização "CDS"
- Localização da característica: <1..80
- qualificador "codon\_start" com o valor "3"
- qualificador "transl\_table" com o valor "3"
- qualificador "transl\_except" com o valor "(pos:30..32,aa:Sec)"
- qualificador "translation" com o valor "DNEEVNEECURLFFKNARHTTSRLT"
- uma sequência individual de proteínas para a tradução
- o qualificador "protein\_id" com o número SEQ ID da proteína traduzida



# Chaves de caracterização e qualificadores

## Sequências de nucleotídeos: chave de caracterização "CDS"

- A localização da característica "CDS" pode utilizar o operador de localização "join" para conectar segmentos descontínuos de uma sequência e formar uma única região de codificação

`join(location1,location2)`

- A localização da característica "CDS" pode utilizar o operador "complement" para indicar que a característica está localizada na fita complementar da sequência especificada pelo descritor de localização

`complement(location)`

# Chaves de caracterização e qualificadores

Sequências de aminoácidos: chaves de caracterização normalmente usadas

Chave de caracterização	Descrição	Qualificadores obrigatórios?
SITE	indica um único sítio de interesse de um aminoácido	um qualificador NOTE obrigatório deve descrever o sítio
REGION	indica uma região de interesse	nenhum; NOTE é opcional
BINDING	indica o sítio de ligação para um grupo químico	um qualificador NOTE obrigatório deve conter o nome do grupo químico
UNSURE	descreve regiões de incerteza na sequência	nenhum; NOTE é opcional

# Chaves de caracterização e qualificadores

## Sequências de aminoácidos; aminoácidos modificados

- Um "aminoácido modificado" é qualquer aminoácido que não seja:

L-alanina	L-arginina	L-asparagina	ácido L-aspártico
L-cisteína	L-glutamina	ácido L-glutâmico	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-triptofano
	L-tirosina	L-valina	

(ST.26, par. 3(e))

- Um "aminoácido modificado" deve ser representado sempre que possível pelo aminoácido não modificado correspondente. Caso contrário, pode ser representado por "X". Por exemplo, "hidroxilisina" deve ser representada por "K" na sequência. "Ornitina" deve ser representada por "X". (ST.26, par. 29).
- O símbolo "X" equivale a apenas um resíduo.

# Chaves de caracterização e qualificadores

## Sequências de aminoácidos; aminoácidos modificados

- Várias chaves de caracterização podem ser utilizadas para indicar um "aminoácido modificado":

Chave de caracterização	Descrição
SITE	indica um aminoácido sem modificações pós-traducionais
MOD_RES	indica um aminoácido com modificações pós-traducionais
CARBOHYD	indica um aminoácido glicosilado
LIPID	indica a ligação covalente de uma fração lipídica a um aminoácido

- Um qualificador NOTE obrigatório deve ser incluído em cada uma das chaves de caracterização acima, com um valor que descreve a modificação.













# Formatos das localizações de características

# Formatos de localizações

## Descritores de localização para todos os tipos de moléculas

- Os descritores de localização são usados para identificar a localização de uma característica numa sequência
- A norma ST.26 tem requisitos obrigatórios para o formato dos descritores de localização
- Os seguintes formatos de descritores de localização podem ser usados para sequências de nucleotídeos e de aminoácidos:

Location descriptor type	Syntax	Description
Single residue number	x	Points to a single residue in the sequence.
Residue numbers delimitating a sequence span	x..y	Points to a continuous range of residues bounded by and including the starting and ending residues.
Residues before the first or beyond the last specified residue number	<x >x <x..y x..>y <x..>y	Points to a region including a specified residue or span of residues and extending beyond a specified residue. The '<' and '>' symbols may be used with a single residue or the starting and ending residue numbers of a span of residues to indicate that a feature extends beyond the specified residue number.

Norma ST.26 da OMPI, parágrafo 66(a)

# Formatos de localizações

## Descritores de localização para todos os tipos de moléculas

Location Example	Description
467	Points to residue 467 in the sequence.
340..565	Points to a continuous range of residues bounded by and including residues 340 and 565.
<1	Points to a feature location before the first residue.
<345..500	Indicates that the exact lower boundary point of a feature is unknown. The location begins at some residue previous to 345 and continues to and includes residue 500.
<1..888	Indicates that the feature starts before the first sequenced residue and continues to and includes residue 888.
1..>888	Indicates that the feature starts at the first sequenced residue and continues beyond residue 888.
<1..>888	Indicates that the feature starts before the first sequenced residue and continues beyond residue 888.

*Norma ST.26 da OMPI, parágrafo 70(a)*

# Formatos de localizações

## Descritores de localização exclusivamente para sequências de nucleotídeos

- O seguinte formato de descritores de localização pode ser usado APENAS para sequências de ADN e ARN:

Location descriptor type	Syntax	Description
A site between two adjoining nucleotides	$x^y$	Points to a site between two adjoining nucleotides, e.g., endonucleolytic cleavage site. The position numbers for the adjacent nucleotides are separated by a carat (^). The permitted formats for this descriptor are $x^{x+1}$ (for example 55 <sup>56</sup> ), or, for circular nucleotides, $x^1$ , where “x” is the full length of the molecule, i.e. 1000 <sup>1</sup> for circular molecule with length 1000.

*Norma ST.26 da OMPI, parágrafo 66(b)*

# Formatos de localizações

## Descritores de localização exclusivamente para sequências de nucleotídeos

- Quando o formato de localização  $x^y$  é usado,  $x$  e  $y$  devem ser resíduos adjacentes

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

*Norma ST.26 da OMPI, parágrafo 70(b)*

# Formatos de localizações

## Descritores de localização – caso especial das sequências de aminoácidos

- O descritor de localização x.y indica uma ligação cruzada intracadeia entre os dois resíduos indicados quando o descritor é usado com uma chave de caracterização "CROSSLINK" ou "DISULFID"

Location descriptor type	Syntax	Description
Residue numbers joined by an intrachain cross-link	x . y	Points to amino acids joined by an intrachain linkage when used with a feature that indicates an intrachain cross-link, such as "CROSSLNK" or "DISULFID".

*Norma ST.26 da OMPI, parágrafo 66(c)*

# Formatos de localizações

## Descritores de localização – caso especial das sequências de aminoácidos

```

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

```



# Formatos de localizações

## Operadores de localização para sequências de nucleotídeos

- Três operadores de localização estão disponíveis para uso em sequências de ADN e ARN: “join”, “order” e “complement”
- O uso dos operadores de localização "join" implica que os resíduos de nucleotídeos descritos pelos descritores de localização são colocados fisicamente em contato por processos biológicos (ST.26, parágrafo 68)
- Localizações usando "join" e "order" devem ter pelo menos dois descritores de localização separados por vírgulas
- "complement" pode ser usado em combinação com "join" ou "order"

Location syntax	Location description
<code>join(location, location, ..., location)</code>	The indicated locations are joined (placed end-to-end) to form one contiguous sequence.
<code>order(location, location, ..., location)</code>	The elements are found in the specified order but nothing is implied about whether joining those elements is reasonable.
<code>complement(location)</code>	Indicates that the feature is located on the strand complementary to the sequence span specified by the location descriptor, when read in the 5' to 3' direction or in the direction that mimics the 5' to 3' direction.

# Formatos de localizações

## Operadores de localização para sequências de nucleotídeos

Location example	Description
<code>join(12..78,134..202)</code>	Indicates that regions 12 to 78 and 134 to 202 should be joined to form one contiguous sequence.
<code>order(15..228,341..502)</code>	Indicates that regions 15 to 228 and 341 to 502 are present in the specified order
<code>complement(34..126)</code>	Starts at the nucleotide complementary to 126 and finishes at the nucleotide complementary to nucleotide 34 (the feature is on the strand complementary to the presented strand).
<code>complement(join(2691..4571,4918..5163))</code>	Joins nucleotides 2691 to 4571 and 4918 to 5163, then complements the joined segments (the feature is on the strand complementary to the presented strand).
<code>join(complement(4918..5163),complement(2691..4571))</code>	Complements regions 4918 to 5163 and 2691 to 4571, then joins the complemented segments (the feature is on the strand complementary to the presented strand).

*Norma ST.26 da OMPI, parágrafo 70(b)*

# Perguntas?

# Formatos dos valores de qualificadores

# Valores de qualificadores

- Os qualificadores definem características em maior detalhe

# Valores de qualificadores

- Os qualificadores definem características em maior detalhe
- Os qualificadores consistem em um nome de qualificador e muitas vezes um valor de qualificador

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

# Valores de qualificadores

- Os qualificadores definem características em maior detalhe
- Os qualificadores consistem em um nome de qualificador e muitas vezes um valor de qualificador

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

- Cada chave de caracterização tem uma lista de qualificadores permitidos para esta característica. Algumas chaves de caracterização têm qualificadores obrigatórios.

# Valores de qualificadores

- Os qualificadores definem características em maior detalhe
- Os qualificadores consistem em um nome de qualificador e muitas vezes um valor de qualificador

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

- Cada chave de caracterização tem uma lista de qualificadores permitidos para esta característica. Algumas chaves de caracterização têm qualificadores obrigatórios.
- As chaves de caracterização e seus qualificadores permitidos figuram na norma ST.26, Anexo I, Seção 5 (sequências de nucleotídeos) e Seção 7 (sequências de aminoácidos).



# Valores de qualificadores

- Exemplo - a chave de caracterização "misc\_binding" tem um qualificador obrigatório, "bound\_moiety", e 6 qualificadores opcionais:

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

(ST.26, Anexo I, Seção 5.12)

# Valores de qualificadores

- Os qualificadores definem características em maior detalhe
- Os qualificadores consistem em um nome de qualificador e muitas vezes um valor de qualificador

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

- Cada chave de caracterização tem uma lista de qualificadores permitidos para esta característica. Algumas chaves de caracterização têm qualificadores obrigatórios.
- As chaves de caracterização e seus qualificadores permitidos figuram na norma ST.26, Anexo I, Seção 5 (sequências de nucleotídeos) e Seção 7 (sequências de aminoácidos).
- Os qualificadores de sequências de nucleotídeos figuram com suas descrições na norma ST.26, Anexo I, Seção 6 (sequências de nucleotídeos) e Seção 8 (sequências de aminoácidos).

# Valores de qualificadores

---

6.3. Qualifier	<code>bound_moiety</code>
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	<code>&lt;INSDQualifier_value&gt;GAL4&lt;/INSDQualifier_value&gt;</code>
Comment	A single <code>bound_moiety</code> qualifier is permitted on the "misc_binding", "oriT" and "protein_bind" features.

---

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

(ST.26, Anexo I, Seção 6.3 e 6.22)

# Valores de qualificadores

## Tipos de formatos

- Os valores de qualificadores têm vários tipos de formato:
  1. Qualificadores com escolhas predefinidas de valores;
  2. Qualificadores com um formato definido de valores;
  3. Qualificadores nos quais o valor é uma sequência;
  4. Qualificadores SEM NENHUM valor;
  5. Qualificadores com valores de "free text" (texto livre)
    - um subconjunto dos valores de qualificadores de "texto livre" é categorizado como sendo "language-dependent" (dependente da língua)

# Valores de qualificadores

## Tipos de formatos – Escolhas predefinidas de valores

- Qualificadores com escolhas predefinidas de valores
- Exemplos:

"codon\_start" – os valores podem ser "1", "2" ou "3"

---

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

# Valores de qualificadores

## Tipos de formatos – Escolhas predefinidas de valores

- Exemplos: "rpt\_type" tem um conjunto limitado de escolhas de valores

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

# Valores de qualificadores

## Tipos de formatos – Escolhas predefinidas de valores

- No caso de qualificadores com escolhas predefinidas de valores, WIPO Sequence apresenta todos os valores permitidos numa lista suspensa pré-preenchida:

▼ QUALIFIERS

Add qualifier

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

▼ SEQUENCE

tgccgatatt	atccataagc	atgggctct	ggaagtgccg	ctgacggccc	agaagggcgt
cttactgtta	caagaagagt	atcatgacgc	gcatcatagc	cacttagacc	cgcacttatg
gttgccgcgc	caaaatatca	ttgcctctgt	gcaaggattg	gataaacata	ctgctgagtt

180

# Valores de qualificadores

## Tipos de formatos – Formato definido de valores

- Qualificadores com um formato definido de valores

- Exemplo: "anticodon" – o valor deve ter o formato

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

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



# Valores de qualificadores

## Tipos de formatos - sequências

- Qualificadores nos quais o valor é uma sequência
- Exemplo: "translation" – o valor deve ser uma sequência usando abreviaturas de aminoácidos de uma letra

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

# Valores de qualificadores

## Tipos de formatos - sequências

- Qualificador "replace" – o valor pode ser um único resíduo de nucleotídeos, uma sequência de resíduos, ou ser deixado vazio
- Se o valor de "replace" for deixado vazio, isto significa a supressão do resíduo indicado na característica correspondente

---

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

# Valores de qualificadores

## Tipos de formatos - sequências

- Qualificador "replace" – o valor pode ser um único resíduo de nucleotídeos, uma sequência de resíduos, ou ser deixado vazio
- Se o valor de "replace" for deixado vazio, isto significa a supressão do resíduo indicado na característica correspondente

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

**"Valor vazio"**

# Valores de qualificadores

## Tipos de formatos - NENHUM valor

- Qualificadores sem NENHUM valor

- Exemplos:

“environmental\_sample”

“germline”

“macronuclear”

“proviral”

---

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

- WIPO Sequence não permite adicionar um valor a qualificadores sem nenhum valor

- Estes qualificadores não devem ter um elemento "INSDQualifier\_value" vazio

# Valores de qualificadores

## Tipos de formatos – "free text" (texto livre)

- Muitos qualificadores tem um formato de valores "texto livre"
- A norma ST.26, parágrafo 3(n), define "texto livre" como "*...um tipo de formato de valores de certos qualificadores, que se apresenta sob a forma de uma frase de texto descritivo ou um outro formato especificado (como indicado no Anexo I).*"
- Os valores de qualificadores de texto livre são limitados a 1000 caracteres (ST.26, par. 86)
- Um subconjunto de qualificadores com formatos de valores de "texto livre" é "*dependente da língua*"
- Texto livre "dependente da língua" "pode necessitar de tradução para procedimentos nacionais, regionais ou internacionais". ST.26, parágrafo 3(o)

# Valores de qualificadores

## Tipos de formatos – "free text" (texto livre)

- Qualificadores que têm valores de "texto livre dependentes da língua" podem precisar de tradução
- Como determinar se um qualificador com um valor no formato de texto livre é dependente da língua?
  - A norma ST.26, Anexo I, Seção 6, Tabela 5, enumera todos os qualificadores de sequências de nucleotídeos com valores no formato de texto livre dependente da língua
  - A norma ST.26, Anexo I, Seção 8, Tabela 6, enumera todos os qualificadores de sequências de aminoácidos com valores no formato de texto livre dependente da língua
  - Consulte o "Formato obrigatório do valor" na descrição do qualificador

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>

# Valores de qualificadores

## Tipos de formatos – "texto livre"

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

# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- Os valores dos qualificadores de "texto livre dependentes da língua" podem ser fornecidos em duas línguas na listagem de sequências em XML: inglês e uma língua que não seja o inglês (ST.26, parágrafo 87)



# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- Os valores dos qualificadores "Language-Dependent Free Text" (texto livre dependente da língua) podem ser fornecidos em duas línguas na listagem de sequências em XML: inglês e uma língua que não seja o inglês (ST.26, parágrafo 87)
- Um valor de qualificador dependente da língua em inglês deve ser fornecido em um elemento INSDQualifier\_value

# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- Os valores dos qualificadores de "texto livre dependentes da língua" podem ser fornecidos em duas línguas na listagem de sequências em XML: inglês e uma língua que não seja o inglês (ST.26, parágrafo 87)
- Um valor de qualificador dependente da língua em inglês deve ser fornecido em um elemento INSDQualifier\_value
- Um valor de qualificador dependente da língua em qualquer língua que não seja o inglês deve ser fornecido em um elemento NonEnglishQualifier\_value

# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- Os valores dos qualificadores de "texto livre dependentes da língua" podem ser fornecidos em duas línguas na listagem de sequências em XML: inglês e uma língua que não seja o inglês (ST.26, parágrafo 87)
- Um valor de qualificador dependente da língua em inglês deve ser fornecido em um elemento `INSDQualifier_value`
- Um valor de qualificador dependente da língua em qualquer língua que não seja o inglês deve ser fornecido em um elemento `NonEnglishQualifier_value`
- Um elemento `NonEnglishQualifier_value` é permitido APENAS para qualificadores com valores no formato de texto livre dependente da língua (ST.26, parágrafo 87(b))

# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

```

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

```

# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- Quando uma listagem de sequências XML contém valores de qualificadores não em inglês:

1. o elemento raiz da listagem de sequências em XML deve conter um atributo "nonEnglishFreeTextLanguageCode" com uma abreviatura apropriada de 2 letras do código da língua (ST.26, parágrafos 43 e 87(b));

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

# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- Quando uma listagem de sequências em XML contém valores de qualificadores não em inglês:
  1. o elemento raiz da listagem de sequências em XML deve conter um atributo "nonEnglishFreeTextLanguageCode" com uma abreviatura apropriada de 2 letras do código da língua (ST.26, parágrafos 43 e 87(b));
  2. TODOS os qualificadores dependentes da língua na listagem de sequências devem ter valores na língua indicada no atributo "nonEnglishFreeTextLanguageCode";

# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- Quando uma listagem de sequências em XML contém valores de qualificadores não em inglês:
  1. o elemento raiz da listagem de sequências em XML deve conter um atributo "nonEnglishFreeTextLanguageCode" com uma abreviatura apropriada de 2 letras do código da língua (ST.26, parágrafos 43 e 87(b));
  2. TODOS os qualificadores dependentes da língua na listagem de sequências devem ter valores na língua indicada no atributo "nonEnglishFreeTextLanguageCode";
  3. Quando NonEnglishQualifier\_value e INSDQualifier\_value estiverem ambos presentes para um único qualificador, as informações contidas nos dois elementos devem ser equivalentes (ST.26, parágrafo 87(c)).

# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- Atributo "id" do INSDQualifier – o que é?
- A norma ST.26, parágrafo 87(d), estipula: *"Para qualificadores dependentes da língua, o elemento INSDQualifier pode incluir um atributo id opcional. O valor deste atributo deve estar no formato "q" seguido por um número inteiro positivo, por exemplo "q23", e deve pertencer exclusivamente a um elemento INSDQualifier, quer dizer o valor do atributo deve ser utilizado apenas uma vez em um arquivo de listagem de sequências."*

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



# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- Atributos "id" do INSDQualifier:
  - Identificam univocamente os valores de qualificadores que podem necessitar de tradução para serem exportados por WIPO Sequence para um arquivo XLIFF;
  - São opcionais;
  - Permitidos apenas para qualificadores dependentes da língua;
  - Devem ser exclusivos em uma listagem de sequências;
  - São automaticamente adicionados a um arquivo XML gerado por WIPO Sequence.

# Valores de qualificadores

## Tipos de formatos – "Texto livre dependente da língua"

- O atributo “originalFreeTextLanguageCode”

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

- A norma ST.26, parágrafo 43, define o atributo “originalFreeTextLanguageCode” como sendo "O código da língua...da única língua original na qual os qualificadores de texto livre dependentes da língua foram preparados."
- Este atributo é OPCIONAL

Situações especiais:  
Moléculas híbridas de ADN/ARN  
Uracilo em ADN e  
Timina em ARN

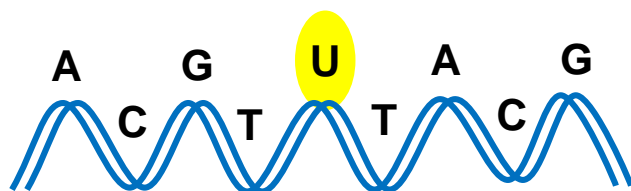
# Situações especiais

- Lembrete: o símbolo "u" para uracilo não é permitido na norma ST.26
- No ADN, "t" é timina
- No ARN, "t" é uracilo
- Dois cenários devem ser considerados:
  1. Molécula de ADN com nucleobase de uracilo ou molécula de ARN com nucleobase de timina;
  2. Molécula híbrida de ADN/ARN

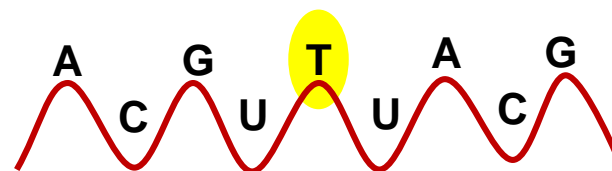
# Situações especiais

## Uracilo em ADN e timina em ARN

- Se a sequência tem uma espinha dorsal de ADN com uma nucleobase de uracilo, OU uma espinha dorsal de ARN com uma nucleobase de timina – descrever como "nucleotídeo modificado"



ADN



ARN

A norma ST.26, parágrafo 14, se aplica:

"14. O símbolo "t" é interpretado como timina no ADN e uracilo no ARN. O uracilo em ADN ou a timina em ARN é considerado como um **nucleotídeo modificado** e deve ser descrito em maior detalhe na tabela de características especificada no parágrafo 19."

# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a sequência de ARN seguinte:

**5' -cgucccacgugtccgaggua - 3'**

- Note a "timina" na posição 12. Este resíduo deve ser anotado como um nucleotídeo modificado.
- A norma ST.26, parágrafo 19, estipula: *"Uracilo em ADN ou timina em ARN são considerados nucleotídeos modificados e devem ser representados por "t" na sequência e ser descritos em maior detalhe na tabela de características usando a chave de caracterização "modified\_base", o valor "OTHER" para o qualificador "mod\_base", e o valor "uracilo" ou "timina", respectivamente, para o qualificador "note"."*

# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a sequência de ARN seguinte:

5' -cguccacgugtccgaggua - 3'

- Note a "timina" na posição 12. Este resíduo deve ser anotado como um nucleotídeo modificado.
- A norma ST.26, parágrafo 19, estipula: *Uracilo em ADN ou timina em ARN são considerados nucleotídeos modificados e devem ser representados por "t" na sequência e ser descritos em maior detalhe na tabela de características usando a chave de caracterização "modified\_base", o qualificador "mod\_base" com "OTHER" como valor do qualificador, e a "note" do qualificador com "uracilo" ou "timina", respectivamente, como valor do qualificador.*

# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a sequência de ARN seguinte:

**5' -cguccacgugtccgaggua - 3'**

- ✓ Todos os resíduos uracilo devem ser representados pelo símbolo "t".  
Portanto, a sequência deve ser representada na listagem de sequências como:

**cgtccccacgtgtccgaggta**



# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a sequência de ARN seguinte:

**5' -cguccacgugtccgaggua - 3'**

- ✓ Todos os resíduos uracilo devem ser representados pelo símbolo "t". Portanto, a sequência deve ser representada na listagem de sequências como:

**cgtccccacgtgtccgaggta**

- ✓ Chave de caracterização "modified\_base" com a localização "12"

# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a sequência de ARN seguinte:

**5' -cguccacgugtccgaggua - 3'**

- ✓ Todos os resíduos uracilo devem ser representados pelo símbolo "t".  
Portanto, a sequência deve ser representada na listagem de sequências como:
- ✓ **cgtccccacgtgtccgaggta**
- ✓ Chave de caracterização "modified\_base" com localização "12"
- ✓ Qualificador "mod\_base" com o valor "OTHER"

# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a sequência de ARN seguinte:

**5' -cguccacgugtccgaggua - 3'**

- ✓ Todos os resíduos uracilo devem ser representados pelo símbolo "t".  
Portanto, a sequência deve ser representada na listagem de sequências como:

**cgtccccacgtgtccgaggta**

- ✓ Chave de caracterização "modified\_base" com a localização "12"
- ✓ Qualificador "mod\_base" com o valor "OTHER"
- ✓ Qualificador "note" com o valor "timina"

# Situações especiais

## Sequências de ADN e ARN

```

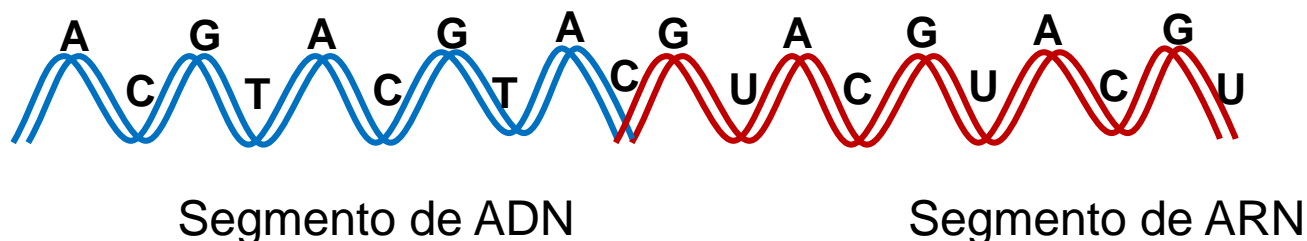
<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length>20</INSDSeq_length>
    <INSDSeq_moltype>RNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..20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

```

# Situações especiais

## Sequências híbridas de ADN/ARN

- Se uma sequência é uma molécula híbrida, quer dizer parte da espinha dorsal é de ADN e parte da espinha dorsal é de ARN:



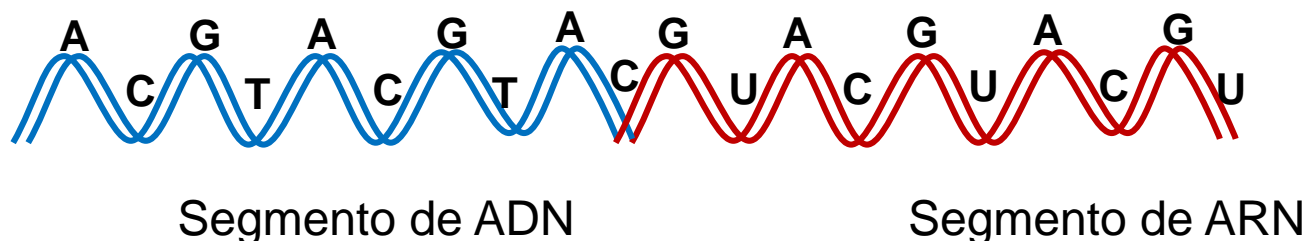
A norma ST.26, parágrafo 55, se aplica:

*"55. No caso de uma sequência de nucleotídeos contendo segmentos tanto de ADN quanto de ARN de um ou mais nucleotídeos, a molécula deve ser indicada como sendo do tipo ADN. A molécula combinada de ADN/ARN deve ser descrita em maior detalhe na tabela de características, usando a chave de caracterização "source" e o qualificador obrigatório "organism" com o valor "synthetic construct" (estrutura sintética) e o qualificador obrigatório "mol\_type" com o valor "other DNA". Cada segmento de ADN e ARN da molécula combinada de ADN/ARN deve ser descrito em maior detalhe por meio da chave de caracterização "misc\_feature" e o qualificador "note", que indica se o segmento é de ADN ou ARN."*

# Situações especiais

## Sequências híbridas de ADN/ARN

- Se uma sequência é uma molécula híbrida, quer dizer parte da espinha dorsal é de ADN e parte da espinha dorsal é de ARN:



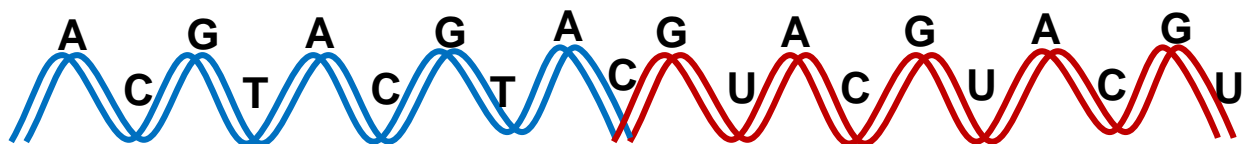
A norma ST.26, parágrafo 55, se aplica:

"55. No caso de uma sequência de nucleotídeos que contém segmentos tanto de ADN quanto de ARN de **um ou mais nucleotídeos, a molécula deve ser indicada como sendo do tipo ADN. A molécula combinada de ADN/ARN deve ser descrita em maior detalhe na tabela de características, usando a chave de caracterização "source" e o qualificador obrigatório "organism" com o valor "synthetic construct" (estrutura sintética) e o qualificador obrigatório "mol\_type" com o valor "other DNA". Cada segmento de ADN e ARN da molécula combinada de ADN/ARN deve ser descrito em maior detalhe por meio da chave de caracterização "misc\_feature" e do qualificador "note", que indica se o segmento é de ADN ou ARN.**"

# Situações especiais

## Sequências híbridas de ADN/ARN

- Se uma sequência é uma molécula híbrida, quer dizer parte da espinha dorsal é de ADN e parte da espinha dorsal é de ARN:



Segmento de ADN

Segmento de ARN

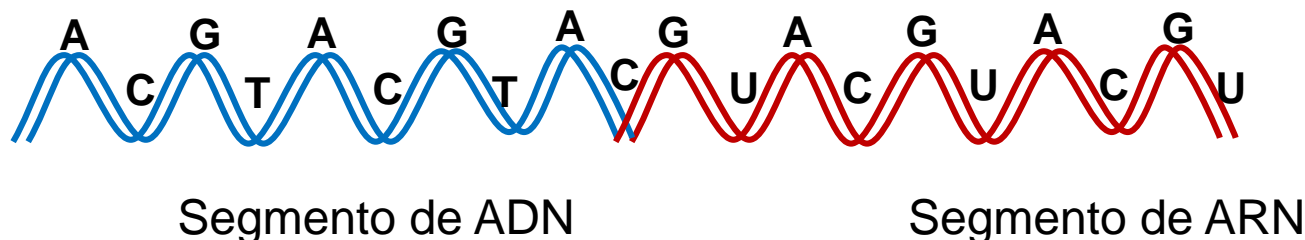
A norma ST.26, parágrafo 55, se aplica:

"55. No caso de uma sequência de nucleotídeos que contem segmentos tanto de ADN quanto de ARN de um ou mais nucleotídeos, a molécula deve ser indicada como sendo do tipo ADN. A molécula combinada de ADN/ARN deve ser descrita em maior detalhe na tabela de características, usando a chave de caracterização "source" e o qualificador obrigatório "organism" com o valor "synthetic construct" (estrutura sintética) e o qualificador obrigatório "mol\_type" com o valor "other DNA". Cada segmento de ADN e ARN da molécula combinada de ADN/ARN deve ser descrito em maior detalhe por meio da chave de caracterização "misc\_feature" e o qualificador "note", que indica se o segmento é de ADN ou ARN."

# Situações especiais

## Sequências híbridas de ADN/ARN

- Se uma sequência é uma molécula híbrida, quer dizer parte da espinha dorsal é de ADN e parte da espinha dorsal é de ARN:



A norma ST.26, parágrafo 55, se aplica:

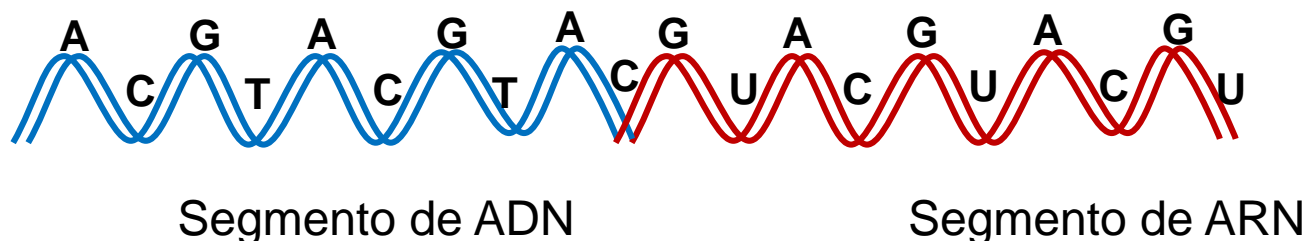
"55. No caso de uma sequência de nucleotídeos contendo segmentos tanto de ADN quanto de ARN de um ou mais nucleotídeos, a molécula deve ser indicada como sendo do tipo ADN. A molécula combinada de ADN/ARN deve ser descrita em maior detalhe na tabela de características, usando a chave de caracterização "source" e o qualificador obrigatório "organism" com o valor "synthetic construct" (estrutura sintética) e o qualificador obrigatório "mol\_type" com o valor "other DNA". Cada segmento de ADN e ARN da molécula combinada de ADN/ARN deve ser descrito em maior detalhe por meio da chave de caracterização "misc\_feature" e o qualificador "note", que indica se o segmento é de ADN ou ARN."



# Situações especiais

## Sequências híbridas de ADN/ARN

- Se uma sequência é uma molécula híbrida, quer dizer parte da espinha dorsal é de ADN e parte da espinha dorsal é de ARN:



A norma ST.26, parágrafo 55, se aplica:

"55. No caso de uma sequência de nucleotídeos contendo segmentos tanto de ADN quanto de ARN de um ou mais nucleotídeos, a molécula deve ser indicada como sendo do tipo ADN. A molécula combinada de ADN/ARN deve ser descrita em maior detalhe na tabela de características, usando a chave de caracterização "source" e o qualificador obrigatório "organism" com o valor "synthetic construct" (estrutura sintética) e o qualificador obrigatório "mol\_type" com o valor "other DNA". Cada segmento de ADN e ARN da molécula combinada de ADN/ARN deve ser descrito em maior detalhe por meio da chave de caracterização "misc\_feature" e o qualificador "note", que indica se o segmento é de ADN ou ARN."

# Situações especiais

## Sequências híbridas de ADN/ARN

Um pedido divulga a seguinte sequência híbrida de ADN/ARN:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

onde os símbolos em maiúsculas representam a parte ADN e os símbolos em minúsculas a parte ARN.

- Os resíduos 1-6 e 27-32 são de ADN, os resíduos 7-26 de ARN
- Devem ser considerados:
  1. a designação do organismo
  2. o tipo de molécula e mol\_type
  3. a identificação dos segmentos de ADN e ARN

# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a seguinte sequência híbrida de ADN/ARN:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

onde os símbolos em maiúsculas representam a parte ADN e os símbolos em minúsculas a parte ARN.

- A norma ST.26, parágrafo 55, estipula: *"...o tipo de molécula deve ser indicado como sendo ADN. A molécula combinada de ADN/ARN deve ser descrita em maior detalhe na tabela de características, usando a chave de caracterização "source" e o qualificador obrigatório "organism" com o valor "synthetic construct" (estrutura sintética) e o qualificador obrigatório "mol\_type" com o valor "other DNA"."*

- ✓ Tipo de molécula = "DNA"
- ✓ Qualificador nome de organismo = "synthetic construct"
- ✓ Qualificador mol\_type = "other DNA"

# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a seguinte sequência híbrida de ADN/ARN:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

onde os símbolos em maiúsculas representam a parte ADN e os símbolos em minúsculas a parte ARN.

- ✓ Todos os resíduos uracilo devem ser representados pelo símbolo "t".  
Portando, a sequência deve ser representada na listagem de sequências como:

**acctgccgtcccacgtgtccgaggtagcatta**

# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a sequência híbrida de ADN/ARN seguinte:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

onde os símbolos em maiúsculas representam a parte ADN e os símbolos em minúsculas a parte ARN.

- A norma ST.26, parágrafo 55, estipula: *"Cada segmento de ADN e ARN da molécula combinada de ADN/ARN deve ser descrito em maior detalhe por meio da chave de caracterização "misc\_feature" e do qualificador "note", que indica se o segmento é de ADN ou ARN."*
- ✓ Três segmentos = três chaves de caracterização "misc\_feature"

# Situações especiais

## Sequências de ADN e ARN

Um pedido divulga a sequência híbrida de ADN/ARN seguinte:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

onde os símbolos em maiúsculas representam a parte ADN e os símbolos em minúsculas a parte ARN.

- ✓ Segmento 1, resíduos 1-6:  
chave de caracterização "misc\_feature" com a localização "1..6"  
Qualificador "note" com o valor "DNA"
- ✓ Segmento 2, resíduos 7-26:  
chave de caracterização "misc\_feature" com localização "7..26"  
Qualificador "note" com valor "RNA"
- ✓ Segmento 3, resíduos 27-32  
chave de caracterização "misc\_feature" com a localização "27..32"  
Qualificador "note" com valor "DNA"

# Situações especiais

## Sequências de ADN e ARN

```

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

```

# Situações especiais

## Sequências de ADN e ARN

```

<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>1..6</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier id="q11">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>DNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>7..26</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier id="q12">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>RNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>27..32</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier id="q13">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>DNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>acctgccgtccccacgtgtccgaggtagcatta</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
</ST26SequenceListing>

```

Segmento de ADN 1

Segmento de ARN 2

Segmento de ADN 3



# Perguntas?

# Análogos de nucleotídeos, aminoácidos D e sequências ramificadas

# Novos tipos de moléculas

## Análogos de nucleotídeos

- As sequências de ácidos nucleicos contendo um ou mais análogos de nucleotídeos estão sujeitas às regras da norma ST.26
- Os análogos de nucleotídeos estão incluídos na definição de "nucleotídeo" na norma ST.26, parágrafo 3(g)(2):

*"um análogo de um 2'-deoxiribose-5'-monofosfato ou ribose-5'-monofosfato que, quando formando a espinha dorsal de um análogo de ácido nucleico, resulta em um arranjo de nucleobases que imita o arranjo de nucleobases em ácidos nucleicos contendo uma espinha dorsal de 2'-deoxiribose-5'-monofosfato ou ribose-5'-monofosfato, onde o análogo de ácido nucleico é capaz de emparelhar bases com um ácido nucleico complementar"*

- Análogos comuns de nucleotídeos incluem os ácidos peptonucleicos (APNs), os ácidos glicolnucleicos (AGN), os ácidos treonucleicos e os morfolinós
- Devem ser representados da esquerda para a direita, direção que imita a direção de 5' a 3'. (Norma ST.26, parágrafo 11)

# Novos tipos de moléculas

## Análogos de nucleotídeos

Um pedido de patente divulga a sequência seguinte de ácido glicolnucleico (AGN):

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

Onde a extremidade  $\text{PO}_4$  da sequência imita a extremidade 5' de uma sequência de ADN.

✓ Esta sequência deve ser incluída numa listagem de sequências

# Novos tipos de moléculas

## Análogos de nucleotídeos

Um pedido de patente divulga a sequência seguinte de ácido glicolnucleico (AGN):



Onde a extremidade  $\text{PO}_4$  da sequência imita a extremidade 5' de uma sequência de ADN.

- ✓ Esta sequência deve ser incluída numa listagem de sequências
- ✓ A extremidade  $\text{PO}_4$  imita a extremidade 5', a sequência deve portanto ser representada na direção mostrada

# Novos tipos de moléculas

## Análogos de nucleotídeos

Um pedido de patente divulga a sequência seguinte de ácido glicolnucleico (AGN):

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

Onde a extremidade  $\text{PO}_4$  da sequência imita a extremidade 5' de uma sequência de ADN.

- ✓ Esta sequência deve ser incluída numa listagem de sequências
- ✓ A extremidade  $\text{PO}_4$  imita a extremidade 5', a sequência deve portanto ser representada na direção mostrada
- ✓ A sequência inteira deve ser anotada com a chave de caracterização "modified\_base", um qualificador "mod\_base" com o valor "OTHER", e um qualificador "note" que inclui o nome completo e não abreviado dos nucleotídeos modificados, tais como "ácidos glicolnucleicos"

# Novos tipos de moléculas

## Análogos de nucleotídeos

```

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

```

*Nota: uma discussão extensa deste exemplo se encontra na Norma ST.26 da OMPI, Anexo VI, Exemplo 3(g)-4.*

# Novos tipos de moléculas

## D-aminoácidos

- As sequências de aminoácidos contendo um ou mais D-aminoácidos estão sujeitas às regras da norma ST.26
- Os D-aminoácidos estão incluídos na definição de "aminoácido" na norma ST.26, parágrafo 3(a):

*"aminoácido" significa qualquer aminoácido que possa ser representado usando quaisquer dos símbolos constantes do Anexo I (ver Seção 3, Tabela 3). Tais aminoácidos incluem, entre outros, D-aminoácidos e aminoácidos contendo cadeias laterais modificadas ou sintéticas."*

- Os D-aminoácidos devem sempre que possível ser representados na sequência com o símbolo do L-aminoácido não modificado correspondente
- devem ser descritos na tabela de características como aminoácido modificado



# Novos tipos de moléculas

## D-aminoácidos

Um pedido de patente descreve a sequência seguinte:

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

- ✓ Esta sequência deve ser incluída numa listagem de sequências

# Novos tipos de moléculas

## D-aminoácidos

Um pedido de patente descreve a sequência seguinte:

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

- ✓ Esta sequência deve ser incluída numa listagem de sequências
- ✓ Deve ser representada como: AGKLGGM
- ✓ A alanina na posição 1, o ácido glutâmico na posição 2 e a metionina na posição 6 devem todos ser anotados com uma chave de caracterização "SITE" e um qualificador "NOTE" contendo o nome completo e não abreviado do aminoácido correspondente

# Novos tipos de moléculas

## D-aminoácidos

```

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

```

*Nota: uma discussão extensa de um exemplo parecido se encontra na Norma ST.26 da OMPI, Anexo VI, Exemplo 3(a)-1.*

# Novos tipos de moléculas

## Sequências ramificadas

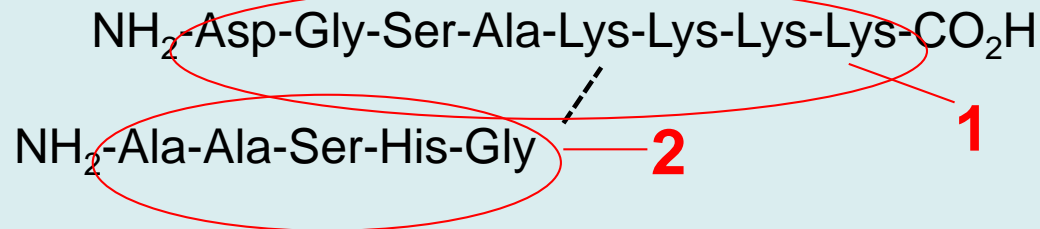
- As sequências ramificadas de ácidos nucleicos e as sequências ramificadas de aminoácidos estão sujeitas às regras da norma ST.26
- As regiões lineares de sequências ramificadas devem ser incluídas numa listagem de sequências quando contêm 10 ou mais nucleotídeos especificamente definidos ou 4 ou mais aminoácidos especificamente definidos (Norma ST.26 da OMPI, parágrafo 7)
- Cada região linear de uma sequência ramificada que atenda ao requisito de comprimento mínimo deve ser incluída como uma sequência individual com seu próprio número identificador (SEQ ID)
- Deve ser considerado o número de resíduos especificamente definidos em cada região linear individual, não o número total de resíduos especificamente definidos na estrutura



# Novos tipos de moléculas

## Sequências ramificadas

Um pedido de patente divulga um peptídeo com a sequência seguinte:



Onde ----- indica uma ligação amida entre o carboxi-terminal da glicina e a cadeia lateral da lisina

- ✓ Ambas as regiões lineares contêm  $\geq 4$  aminoácidos especificamente definidos, portanto ambas devem ser incluídas numa listagem de sequências
- ✓ Cada região linear deve ser incluída como uma sequência individual com seu próprio número identificador (SEQ ID)



# Novos tipos de moléculas

## Sequências ramificadas

1

```

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

```

2

```

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

```

Nota: uma discussão extensa de um exemplo parecido se encontra na norma ST.26 da OMPI, Anexo VI, Exemplo 7(b)-3.



# Variantes de sequências

# Variantes de sequências

- A norma ST.26, parágrafo 3(m), define uma "**sequência variante**" como

*"uma sequência de nucleotídeos ou de aminoácidos contendo **uma ou mais diferenças com relação a uma sequência primária**. Estas diferenças podem incluir resíduos alternativos (ver parágrafos 15 e 27), resíduos modificados (ver parágrafos 3(g), 3(h), 16 e 29), supressões, inserções e substituições. Ver parágrafos 93-95."*

- O modo como uma sequência variante é divulgada determina como deve ser representada numa listagem de sequências.
- A norma ST.26, parágrafos 93-95, determina como as variantes devem ser representadas.

# Variantes de sequências

## Parágrafo 93

*Parágrafo 93: Uma sequência primária e qualquer variante desta sequência, cada uma divulgada pela enumeração de seus resíduos e abrangidas pelo parágrafo 7, devem ser todas incluídas na listagem de sequências e seu próprio número identificador de sequência deve ser atribuído a cada uma.*



Se cada variante for enumerada individualmente, então cada variante deve ter seu próprio número identificador (SEQ ID)!

# Variantes de sequências

## Parágrafo 93

Um pedido de patente inclui uma figura com o alinhamento seguinte de sequências múltiplas:

Consensus	LEGNEQFINA <b>ak</b> IIRHP <b>ky</b> nrk <b>tl</b> n <b>nd</b> in <b>l</b> IK
Homo sapiens	LEGNEQFINAAAKIIRHPQYDRKTLNNDIMLIK
Pongo abelii	LEGNEQFINAAAKIIRHPQYDRKTVNNDIMLIK
Pan paniscus	LEGNEQFINAAAKIIRHPKYNRITLNNDIMLIK
Rhinopithecus bieti	LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK
Rhinopithecus roxellana	LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK

As minúsculas representam os resíduos predominantes de aminoácidos nas sequências alinhadas.

- ✓ Cada uma das seis sequências enumeradas deve ser incluída na listagem de sequências como uma sequência individual com seu próprio número identificador (SEQ ID).

*Nota: uma discussão extensa de um exemplo parecido se encontra na Norma ST.26 da OMPI, Anexo VI, Exemplo 93-3.*

# Variantes de sequências

## Anotação adequada

- A chave de caracterização adequada para anotar uma variante depende do tipo de molécula e da natureza da variação:

Type of sequence	Feature Key	Qualifier	Use
Nucleic acid	variation	replace or note	Naturally occurring mutations and polymorphisms, e.g., alleles, RFLPs.
Nucleic acid	misc_difference	replace or note	Variability introduced artificially, e.g., by genetic manipulation or by chemical synthesis.
Amino acid	VAR_SEQ	NOTE	Variant produced by alternative splicing, alternative promoter usage, alternative initiation and ribosomal frameshifting.
Amino acid	VARIANT	NOTE	Any type of variant for which VAR_SEQ is not applicable.

*Norma ST.26 da OMPI, parágrafo 96*

# Variantes de sequências

## Símbolo de ambiguidade mais restritivo

- A norma ST.26, parágrafos 15 e 27, estipula que quando um símbolo de ambiguidade é necessário, "o símbolo de ambiguidade mais restritivo deve ser usado..."

O que isso significa?

Por exemplo:

Em uma sequência de nucleotídeos, se uma posição pode ser "a ou c", use o símbolo de ambiguidade "m" em vez de "n".

Em uma sequência de aminoácidos, se uma posição pode ser "L ou I", use o símbolo de ambiguidade "J" em vez de "X".

Lembrem-se que "n" e "X" têm valores predefinidos, portanto sempre que "n" ou "X" são usados com um valor que não seja o valor predefinido, uma anotação é necessária.

# Variantes de sequências

## Parágrafo 94

*Parágrafo 94: Qualquer sequência variante divulgada como uma única sequência com resíduos alternativos enumerados em uma ou mais posições deve ser incluída na listagem de sequências e deve ser representada por uma única sequência, onde os resíduos alternativos enumerados são representados pelo símbolo de ambiguidade mais restritivo (ver parágrafos 15 e 27).*



Se as variantes NÃO FOREM enumeradas individualmente mas simplesmente mostradas como resíduos variáveis na sequência primária, NÃO necessitam de seu próprio número identificador (SEQ ID) individual!

# Variantes de sequências

## Parágrafo 94

Um pedido de patente divulga um peptídeo com a sequência:

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

- ✓ Pode ser incluído na listagem de sequências como uma única sequência
- ✓ A representação preferida da sequência é: GGGJATX
- ✓ [Leu ou Ile] deve ser representado pelo símbolo de ambiguidade mais restritivo "J"
- ✓ [Ser ou Thr] deve ser representado pelo símbolo "X" junto com a chave de caracterização "VARIANT" com um qualificador "note" que indica que X é serina ou treonina

*Nota: uma discussão extensa deste exemplo se encontra na Norma ST.26 da OMPI, Anexo VI, Exemplo 94-1.*



# Variantes de sequências

## Parágrafo 95

*Parágrafo 95: Qualquer sequência variante divulgada apenas por referência a uma ou mais supressões, inserções ou substituições numa sequência primária na listagem de sequências deve ser incluída na listagem de sequências. Quando incluída na listagem de sequências, tal sequência variante:*

*(a) pode ser representada por uma anotação da sequência primária, sempre que contenha uma ou mais variações em uma única localização ou em múltiplas localizações distintas e que estas variações ocorram de maneira independente;*

*(b) deve ser representada como uma sequência individual e seu próprio número identificador de sequência lhe deve ser atribuído sempre que contenha variações em múltiplas localizações distintas e que estas variações ocorram de maneira interdependente; e*

*(c) deve ser representada como uma sequência individual e seu próprio número identificador de sequência lhe deve ser atribuído sempre que contenha uma sequência substituída contendo mais de 1000 resíduos (ver parágrafo 86).*



"Por referência a uma ou mais supressões, inserções ou substituições" significa que as variantes são divulgadas em prosa.

# Variantes de sequências

## Parágrafo 95(a)

Um pedido contem a divulgação seguinte:

Fragmento de peptídeo 1: Gly-Leu-Pro-Xaa-Arg-Ile-Cys  
onde Xaa é qualquer aminoácido

\* \* \*

... em outro modo de realização, o fragmento de peptídeo 1 é Gly-Leu-Pro-Xaa-Arg-Ile-Cys, onde Xaa pode ser Val, Thr ou Asp.....

\* \* \*

... em outro modo de realização, o fragmento de peptídeo 1 é Gly-Leu-Pro-Xaa-Arg-Ile-Cys, onde Xaa pode ser Val.....

*A sequência primária, Gly-Leu-Pro-Xaa-Arg-Ile-Cys, contem uma ou mais variações em uma única localização e estas variações ocorrem de maneira independente*

# Variantes de sequências

## Parágrafo 95(a)

- ✓ A sequência primária deve ser incluída na listagem de sequências e as variantes devem ser representadas por anotações da sequência primária
- ✓ O modo de realização mais abrangente deve ser a versão incluída na listagem de sequências – neste exemplo, é a versão onde Xaa é "qualquer aminoácido"
- ✓ A sequência deve ser representada como GLPXRIC e requer a chave de caracterização "VARIANT" na posição 4 e um qualificador "note" indicando que Xaa é "qualquer aminoácido"
- ✓ Embora isso não seja necessário, recomenda-se que as três variantes seguintes sejam incluídas na listagem de sequências como sequências individuais:
  - GLPVRIC
  - GLPIRIC
  - GLPDRIC

*Nota: uma discussão extensa deste exemplo se encontra na norma ST.26 da OMPI, Anexo VI, Exemplo 95(a)-1*

# Variantes de sequências

## Parágrafo 95(b)

Um pedido de patente divulga a sequência consenso seguinte:

aatgn<sub>1</sub>cccacgaatgn<sub>2</sub>cac

onde n<sub>1</sub> e n<sub>2</sub> podem ser a, t, g, ou c.

Várias sequências variantes são divulgadas da seguinte forma:

se n<sub>1</sub> é a, então n<sub>2</sub> é t, g ou c

se n<sub>1</sub> é t, então n<sub>2</sub> é a, g ou c

se n<sub>1</sub> é g, então n<sub>2</sub> é t, a ou c

se n<sub>1</sub> é c, então n<sub>2</sub> é t, g ou a

*A sequência consenso contém variações em múltiplas localizações distintas e essas variações ocorrem de maneira interdependente*

# Variantes de sequências

## Parágrafo 95(b)

- ✓ A sequência consenso deve ser incluída na listagem de sequências e as variantes devem ser representadas como sequências individuais
- ✓ O modo de realização mais abrangente deve ser a versão incluída na listagem de sequências, onde  $n_1$  e  $n_2$  "podem ser a, t, g ou c"
- ✓ A sequência deve ser representada como: `aatgncccacgaaatgncac`
- ✓ Nenhuma anotação é necessária para n já que "n" é interpretado como sendo "a", "c", "g" ou "t", na ausência de uma chave de caracterização (ver norma ST.26, parágrafo 15)
- ✓ Embora isso não seja necessário, recomenda-se fortemente que as quatro variantes seguintes sejam incluídas na listagem de sequências como sequências individuais:

`aatgacccacgaaatgbcac`

(b = t, g ou c)

`aatgtcccacgaaatgvcac`

(v = a, g ou c)

`aatggcccacgaaatghcac`

(h = t, a ou c)

`aatgccccacgaaatgdcac`

(d = t, g ou a)

*Nota: uma discussão extensa de um exemplo parecido se encontra na Norma ST.26 da OMPI, Anexo VI, Exemplo 95(b).*

WIPO PUBLIC

# Variantes de sequências

## Parágrafo 95(c)

Um pedido contem a divulgação seguinte:

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

onde Xaa é Gly ou uma inserção da sequência

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

*Uma sequência variante contem uma sequência inserida ou substituída de mais de 1000 resíduos*

# Variantes de sequências

## Parágrafo 95(c)

- ✓ A variante onde Xaa = uma inserção de mais de 1000 resíduos deve ser incluída na listagem de sequências como uma sequência individual com seu próprio número identificador (SEQ ID).
- ✓ A variante onde Xaa = Gly também será incluída na listagem de sequências como uma sequência individual com seu próprio número identificador (SEQ ID).

...-MGLPRGRICK-...

# Variantes de sequências

**Pergunta:** Que parágrafo se aplica à divulgação seguinte?

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



# Variantes de sequências

**Pergunta:** Que parágrafo se aplica à divulgação seguinte?

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

**A:** Parágrafo 94:

*"Qualquer sequência variante divulgada como uma única sequência com resíduos alternativos enumerados em uma ou mais posições deve ser incluída na listagem de sequências e deve ser representada por uma única sequência, onde os resíduos alternativos enumerados são representados pelo símbolo de ambiguidade mais restritivo."*

# Variantes de sequências

**Pergunta:** Que parágrafo se aplica à divulgação seguinte?

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

Um espaço em branco na tabela indica que um aminoácido na variante é o mesmo que o aminoácido correspondente na "Sequência" e um "-" indica a supressão do aminoácido correspondente na "Sequência".

# Variantes de sequências

**Pergunta:** Que parágrafo se aplica à divulgação seguinte?

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

Um espaço em branco na tabela indica que um aminoácido na variante é o mesmo que o aminoácido correspondente na "Sequência" e um "-" indica a supressão do aminoácido correspondente na "Sequência".

**A:** Parágrafo 93

*"Uma sequência primária e qualquer variante desta sequência, cada uma divulgada pela enumeração de seus resíduos e abrangidas pelo parágrafo 7, devem ser todas incluídas na listagem de sequências e seu próprio número identificador de sequência deve ser atribuído a cada uma."*

# Variantes de sequências

**Pergunta:** Que parágrafo se aplica à divulgação seguinte?

A sequência seguinte é divulgada em uma figura:

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

A especificação contém as informações seguintes sobre a Sequência 1:

... em alguns modos de realização, a valina na posição 7 da sequência 1 é substituída por alanina...

# Variantes de sequências

**Pergunta:** Que parágrafo se aplica à divulgação seguinte?

A sequência seguinte é divulgada em uma figura:

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

A especificação contém as informações seguintes sobre a Sequência 1:

... em alguns modos de realização, a valina na posição 7 da sequência 1 é substituída por alanina...

**A:** Parágrafo 95(a): *"Qualquer sequência variante divulgada apenas por referência a uma ou mais supressões, inserções ou substituições numa sequência primária na listagem de sequências deve ser incluída na listagem de sequências. Quando incluída na listagem de sequências, tal sequência variante:*

*(a) pode ser representada por uma anotação da sequência primária, sempre que contenha uma ou mais variações em uma única localização ou em múltiplas localizações distintas e que estas variações ocorram de maneira independente;"*

# Variantes de sequências

**Pergunta:** Que parágrafo se aplica à divulgação seguinte?

A sequência seguinte é divulgada em uma figura:

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

A especificação contém as informações seguintes sobre a Sequência 1:

... se a valina na posição 7 da sequência 1 for substituída por alanina então a alanina na posição 10 será substituída por valina...

# Variantes de sequências

**Pergunta:** Que parágrafo se aplica à divulgação seguinte?

A sequência seguinte é divulgada em uma figura:

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

A especificação contém as informações seguintes sobre a Sequência 1:

... se a valina na posição 7 da sequência 1 for substituída por alanina então a alanina na posição 10 será substituída por valina...

**A:** Parágrafo 95(b): *Qualquer sequência variante divulgada apenas por referência a uma ou mais supressões, inserções ou substituições numa sequência primária na listagem de sequências deve ser incluída na listagem de sequências. Quando incluída na listagem de sequências, tal sequência variante:*

*(b) deve ser representada como uma sequência individual e seu próprio número identificador de sequência lhe deve ser atribuído sempre que contenha variações em múltiplas localizações distintas e que estas variações ocorram de maneira interdependente;*

# Perguntas?