

# ■ Стандарт ВОИС ST.26 УГЛУБЛЕННЫЙ УРОВЕНЬ

Учебный вебинар

*Примечание: этот модуль является продолжением обучения и рассчитан на то, что участники уже прошли модуль по основам стандарта ST.26.*

*Базовый учебный курс по стандарту ST.26 доступен по ссылке:  
[https://www.wipo.int/meetings/ru/details.jsp?meeting\\_id=62848](https://www.wipo.int/meetings/ru/details.jsp?meeting_id=62848)*

# Аспекты, которые будут сегодня рассмотрены

- Часто используемые ключи характеристик и квалификаторы
- Формат указания местоположения характеристик
- Формат значений квалификаторов и значения квалификаторов не на английском языке
- Особый случай: урацил в ДНК и тимин в РНК; комбинированные молекулы ДНК/РНК
- Аналоги нуклеотидов, D-аминокислоты и разветвленные последовательности
- Варианты последовательностей

# Часто используемые ключи характеристик и квалификаторы

# Ключи характеристик и квалификаторы

- Ключи характеристик могут использоваться для описания одного или нескольких остатков в последовательности, идентифицированных по местоположению
  - Ключи характеристик для нуклеотидных последовательностей приводятся в разделе 5 приложения I
  - Ключи характеристик для нуклеотидных последовательностей указываются в нижнем регистре, например «misc\_binding»
  - Ключи характеристик для аминокислотных последовательностей приводятся в разделе 7 приложения I
  - Ключи характеристик для аминокислотных последовательностей указываются в верхнем регистре, например «REGION»
  
- Квалификаторы могут использоваться для описания характеристик
  - Квалификаторы для нуклеотидных последовательностей приводятся в разделе 6 приложения I
  - Квалификаторы для нуклеотидных последовательностей указываются в нижнем регистре, например «allele»
  - Квалификаторы для аминокислотных последовательностей приводятся в разделе 8 приложения I
  - Квалификаторы для аминокислотных последовательностей указываются в верхнем регистре, например «NOTE»

# Ключи характеристик и квалификаторы

```

<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>

```

# Ключи характеристик и квалификаторы

- Ключи характеристик указывать необязательно, за исключением характеристики «source»/«SOURCE», которая является обязательной для каждой последовательности.
- Каждый ключ характеристики имеет список квалификаторов, которые можно использовать для описания характеристики. Большинство квалификаторов являются необязательными, но некоторые ключи характеристик имеют обязательные квалификаторы.

5.31. Ключ характеристики	regulatory
Определение	любой участок последовательности, который задействован в регуляции транскрипции, трансляции, репликации или структуры хроматина;
Обязательные квалификаторы	regulatory_class
Необязательные квалификаторы	allele bound_moiety function gene gene_synonym map note operon phenotype pseudo pseudogene standard_name

- Характеристика «source»/«SOURCE» имеет обязательные квалификаторы «mol\_type»/«MOL\_TYPE» и «organism»/«ORGANISM».

# Ключи характеристик и квалификаторы

- Некоторые ключи характеристик имеют дополнительные ограничения
  - organism score (виды организмов); например ключ «C\_region» ограничен эукариотами
  - molecule score (виды молекул); например ключ «D-loop» ограничен последовательностями ДНК

5.4.	Ключ характеристики	D-Loop
	Определение	петля смещения; область митохондриальной ДНК, в которой короткий участок РНК взаимодействует с одной из цепей ДНК, вытесняя исходную комплементарную цепь ДНК в этой области; также используется для описания замены участка одной цепи в дуплексной ДНК одноцепочечной ДНК, захваченной извне, в реакции, катализируемой белком RecA
	Необязательные квалификаторы	allele gene gene_synonym map note
	Тип молекул	DNA

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «modified\_base»

- Ключ характеристики «modified\_base» и его обязательный квалификатор «mod\_base» используются для описания модифицированного нуклеотида (стандарт ST.26, пункт 16)
- «Модифицированный нуклеотид» означает любой нуклеотид, кроме:
  - дезокси-[a, g, c или t] 3'-монофосфата
  - [a, g, c или u] 3'-монофосфата (стандарт ST.26, пункт 3(f))
- «Модифицированный нуклеотид», когда это возможно, должен быть представлен соответствующим немодифицированным нуклеотидом (таблица 1, раздел 1 приложения I). В противном случае он должен быть представлен как «n». Например, «2'-О-метилцитидин» должен быть представлен в последовательности как «c». «Квеуозин» должен быть представлен как «n». Символ «n» эквивалентен только одному остатку.
- Значение обязательного квалификатора «mod\_base» необходимо выбирать из значений, указанных в таблице 2 раздела 2 приложения I. Если выбрано значение «other», то дополнительный квалификатор «note» должен содержать полное, несокращенное название модифицированного остатка.



# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «modified\_base»

- Пример: нуклеотидная последовательность с инозином в позиции 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>
```

- В таблице 2 раздела 2 приложения I для инозина указано сокращение «i»

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «modified\_base»

- Пример: нуклеотидная последовательность с ксантином в позиции 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>
```

- Ксантин НЕ указан в таблице 2 раздела 2 приложения I; поэтому обязательный квалификатор «mod\_base» должен иметь значение «OTHER» и должен присутствовать дополнительный квалификатор «note» со значением «xanthine».

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «modified\_base»

- Ключ характеристики «modified\_base» может быть использован также для описания участка, лишённого азотистых оснований (abasic site):

```
<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>
```

- Участок, лишённый азотистых оснований, может быть представлен в последовательности символом «n» и далее описан при помощи ключа характеристики «modified\_base» с квалификатором «mod\_base» со значением «OTHER» и дополнительным квалификатором «note» со значением «abasic site».

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Ключ характеристики «CDS» может использоваться для идентификации кодирующих последовательностей. Местоположение характеристики «CDS» должно включать стоп-кодон (стандарт ST.26, пункт 89)
- Ключ характеристики «CDS» НЕ имеет обязательных квалификаторов
- Часто используемые квалификаторы:

pseudo

pseudogene

translation

transl\_table

codon\_start

transl\_except

protein\_id

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Характеристика «CDS» может иметь только ОДИН из следующих квалификаторов:

Квалификатор	Описание	Значение
pseudo	Указывает, что характеристика «CDS» нефункциональна и не имеет трансляции, но не является псевдогеном	Не имеет значения
pseudogene	Указывает, что характеристика «CDS» является псевдогеном и не имеет трансляции	processed unprocessed unitary allelic unknown
translation	Указывает, что аминокислотная последовательность получена в результате трансляции характеристики «CDS»	Однобуквенные обозначения аминокислот

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Аминокислотная последовательность, которая содержит 4 или более специально определенных аминокислот, кодируется кодирующей последовательностью и раскрыта в квалификаторе «translation», должна быть включена в перечень последовательностей как отдельная последовательность и ей должен быть присвоен отдельный идентификационный номер последовательности.
- Идентификационный номер последовательности, присвоенный аминокислотной последовательности, должен быть указан в качестве значения в классификаторе «protein\_id» ключа характеристики «CDS».
- Квалификатор «ORGANISM» ключа характеристики «SOURCE» для аминокислотной последовательности должен быть идентичен аналогичному квалификатору для ее кодирующей последовательности (стандарт ST.26, пункт 92).

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Квалификаторы, которые могут модифицировать транслированную последовательность:

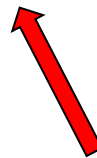
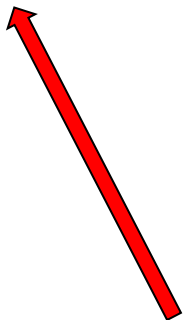
Квалификатор	Описание	Значение
transl_table	Указывает таблицу генетических кодов, использованную для трансляции CDS; значение по умолчанию «1-Standard Code» («1 – стандартный код»)	Число, соответствующее таблице трансляции в разделе 9 приложения I
transl_except	Указывает на трансляцию кодона, не соответствующего генетическому коду, определенному в квалификаторе «transl_table»	(pos:<местоположение>, aa:<аминокислота>)
codon_start	Указывает рамку считывания CDS относительно первого основания	1, 2 или 3

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Пример: последовательность № 1 представляет собой фрагмент кодирующей последовательности из митохондриального гена дрожжей (таблица генетических кодов «3-Yeast Mitochondrial Code» («3 – митохондриальный код дрожжей»))

```
...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 *
```



Кодон в позициях 30-32 кодирует селеноцистеин (Sec)

Последовательность начинается с частичного кодона

- Какую информацию следует включить в характеристику «CDS», чтобы точно представить эту последовательность?



# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Пример: последовательность № 1 представляет собой фрагмент кодирующей последовательности из митохондриального гена дрожжей (таблица генетических кодов «3-Yeast Mitochondrial Code» («3 – митохондриальный код дрожжей»))

```
...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 *
```

- В перечне последовательностей последовательность представлена следующим образом:

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

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Пример: последовательность № 1 представляет собой фрагмент кодирующей последовательности из митохондриального гена дрожжей (таблица генетических кодов «3-Yeast Mitochondrial Code» («3 – митохондриальный код дрожжей»))

```
...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 *
```

- Ключ характеристики «CDS»
- Местоположение характеристики: <1..80

Символ “<” указывает, что кодирующий участок начинается перед позицией 1

Заданный местоположением участок включает стоп-кодон в позициях 78-80

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Пример: последовательность № 1 представляет собой фрагмент кодирующей последовательности из митохондриального гена дрожжей (таблица генетических кодов «3-Yeast Mitochondrial Code» («3 – митохондриальный код дрожжей»))

```
...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 *
```

- Ключ характеристики «CDS»
- Местоположение характеристики: <1..80
- Квалификатор «codon\_start» со значением «3»

Значение «3» квалификатора «codon\_start» указывает, что первый полный кодон начинается в третьей позиции заданного местоположением участка

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Пример: последовательность № 1 представляет собой фрагмент кодирующей последовательности из митохондриального гена дрожжей (таблица генетических кодов «3-Yeast Mitochondrial Code» («3 – митохондриальный код дрожжей»))

```
...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 *
```

- Ключ характеристики «CDS»
- Местоположение характеристики:  
<1..80
- Квалификатор «codon\_start» со значением «3»
- Квалификатор «transl\_table» со значением «3»

Для определения значения квалификатора «transl\_table» используются таблицы генетических кодов в разделе 9 приложения

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Пример: последовательность № 1 представляет собой фрагмент кодирующей последовательности из митохондриального гена дрожжей (таблица генетических кодов «3-Yeast Mitochondrial Code» («3 – митохондриальный код дрожжей»))

```
...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 *
```

- Ключ характеристики «CDS»
- Местоположение характеристики: <1..80
- Квалификатор «codon\_start» со значением «3»
- Квалификатор «transl\_table» со значением «3»
- Квалификатор «transl\_except» со значением «(pos:30..32,aa:Sec)»

Квалификатор  
«transl\_except» вставит  
аминокислоту  
селеноцистеин

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Пример: последовательность № 1 представляет собой фрагмент кодирующей последовательности из митохондриального гена дрожжей (таблица генетических кодов «3-Yeast Mitochondrial Code» («3 – митохондриальный код дрожжей»))

```
...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 *
```

- Ключ характеристики «CDS»
- Местоположение характеристики: <1..80
- Квалификатор «codon\_start» со значением «3»
- Квалификатор «transl\_table» со значением «3»
- Квалификатор «transl\_except» со значением «(pos:30..32,aa:Sec)»
- Квалификатор «translation» со значением «DNEEVNEECURLFFKNARHTTSRLT»

Стоп-кодон не отображается в квалификаторе «translation»!

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Пример: последовательность № 1 представляет собой фрагмент кодирующей последовательности из митохондриального гена дрожжей (таблица генетических кодов «3-Yeast Mitochondrial Code» («3 – митохондриальный код дрожжей»))

```
...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 *
```

- Ключ характеристики «CDS»
- Местоположение характеристики: <1..80
- Квалификатор «codon\_start» со значением «3»
- Квалификатор «transl\_table» со значением «3»
- Квалификатор «transl\_except» со значением «(pos:30..32,aa:Sec)»
- Квалификатор «translation» со значением «DNEEVNEECURLFFKNARHTTSRLT»
- Отдельная белковая последовательность для трансляции

# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- Пример: последовательность № 1 представляет собой фрагмент кодирующей последовательности из митохондриального гена дрожжей (таблица генетических кодов «3-Yeast Mitochondrial Code» («3 – митохондриальный код дрожжей»))

```
...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 *
```

- Ключ характеристики «CDS»
- Местоположение характеристики: <1..80
- Квалификатор «codon\_start» со значением «3»
- Квалификатор «transl\_table» со значением «3»
- Квалификатор «transl\_except» со значением «(pos:30..32,aa:Sec)»
- Квалификатор «translation» со значением «DNEEVNEECURLFFKNARHTTSRLT»
- Отдельная белковая последовательность для трансляции
- Квалификатор «protein\_id» с идентификационным номером (SEQ ID) транслированного белка



# Ключи характеристик и квалификаторы

## Нуклеотидные последовательности: ключ характеристики «CDS»

- При указании местоположения характеристики «CDS» можно использовать оператор местоположения «join» для объединения прерывающихся сегментов последовательности в одну кодирующую область

join(местоположение1, местоположение2)

- При указании местоположения характеристики «CDS» можно использовать оператор «complement» для указания того, что характеристика расположена на нити ДНК, являющейся комплементарной относительно последовательности, определенной указателем местоположения

complement(местоположение)

# Ключи характеристик и квалификаторы

Аминокислотные последовательности: часто используемые ключи характеристик

Ключ характеристики	Описание	Обязательные квалификаторы?
SITE	Указывает интересующий единичный аминокислотный участок	Обязательный квалификатор «NOTE» должен указывать участок
REGION	Указывает интересующую область	Нет; квалификатор «NOTE» является необязательным
BINDING	Указывает участок связывания для химической группы	Обязательный квалификатор «NOTE» должен содержать название химической группы
UNSURE	Указывает на области неопределенности в последовательности	Нет; квалификатор «NOTE» является необязательным

# Ключи характеристик и квалификаторы

## Аминокислотные последовательности: модифицированные аминокислоты

■ «Модифицированная аминокислота» — любая аминокислота, кроме следующих:

L-аланина	L-аргинина	L-аспарагина	L-аспарагиновой кислоты
L-цистеина	L-глутамина	L-глутаминовой кислоты	L-глицина
L--гистидина	L-изолейцина	L-лейцина	L-лизина
L-метионина	L-фенилаланина	L-пролина	L-пирролизина
L-серина	L-селеноцистеина	L-треонина	L-триптофана
L-тирозина	L-валина		

(Стандарт ST.26, пункт 3(e))

■ Модифицированная аминокислота должна быть представлена соответствующей немодифицированной аминокислотой, когда это возможно. В противном случае она может быть представлена символом «X». Например, гидроксизин должен быть представлен символом «K» в последовательности. Орнитин должен быть представлен символом «X» (стандарт ST.26, пункт 29).

■ Символ «X» эквивалентен только одному остатку.

# Ключи характеристик и квалификаторы

## Аминокислотные последовательности: модифицированные аминокислоты

- Для указания «модифицированной аминокислоты» можно использовать несколько ключей характеристик:

Ключ характеристики	Описание
SITE	Указывает на не посттрансляционно модифицированную аминокислоту
MOD_RES	Указывает на посттрансляционно модифицированную аминокислоту
CARBOHYD	Указывает на гликозилированную аминокислоту
LIPID	Указывает на ковалентную связь липидного фрагмента с аминокислотой

- Каждый из указанных выше ключей характеристик должен иметь обязательный квалификатор «NOTE» со значением, описывающим модификацию.





# Ключи характеристик и квалификаторы

Аминокислотные последовательности:  
модифицированные аминокислоты

Gly-Ser-**N-acetylAla**-Ser-Asp-Val-Orn-Lys-Asn-Val-Leu  
1 5 10

Аланин в позиции 3 посттрансляционно модифицирован в клетке и превратился в n-ацетилаланин

асиалилолигосахарид

- Последовательность представлена в перечне последовательностей следующим образом:

**GSASDVXKNVL**

- Ключ характеристики «MOD\_RES» со значением местоположения «3» и квалификатор «NOTE» со значением «N-acetylalanine»

# Ключи характеристик и квалификаторы

## Аминокислотные последовательности: модифицированные аминокислоты

Gly-Ser-N-acetylAla-Ser-Asp-Val-Orn-Lys-Asn-Val-Leu

1

5

10

Аланин в позиции 3 посттрансляционно модифицирован в клетке и превратился в n-ацетилаланин

асиалилолигосахарид

- Последовательность представлена в перечне последовательностей следующим образом:

**GSASDVXKNVL**

- Ключ характеристики «MOD\_RES» со значением местоположения «3» и квалификатор «NOTE» со значением «N-acetylalanine»
- Ключ характеристики «SITE» со значением местоположения «7» и квалификатор «NOTE» со значением «ornithine»



# Ключи характеристик и квалификаторы

## Аминокислотные последовательности: модифицированные аминокислоты

Gly-Ser-N-acetylAla-Ser-Asp-Val-Orn-Lys-**Asn**-Val-Leu  
 1 5 10

Аланин в позиции 3 посттрансляционно модифицирован в клетке и превратился в n-ацетилаланин

асиалилолигосахарид

- Последовательность представлена в перечне последовательностей следующим образом:

**GSASDVXKNVL**

- Ключ характеристики «MOD\_RES» со значением местоположения «3» и квалификатор «NOTE» со значением «N-acetylalanine»
- Ключ характеристики «SITE» со значением местоположения «7» и квалификатор «NOTE» со значением «ornithine»
- Ключ характеристики «CARBOHYD» со значением местоположения «9» и квалификатор «NOTE» со значением «Asn side-chain linked to asialyloligosaccharide»

# Формат указания местоположения характеристик

# Формат указания местоположения

## Дескрипторы местоположения для молекул всех типов

- Дескрипторы местоположения используются для указания местоположения характеристики в последовательности
- Стандарт ST.26 предусматривает обязательные требования к формату дескрипторов местоположения
- Для нуклеотидных и аминокислотных последовательностей можно использовать следующий формат дескрипторов местоположения:

Дескриптор местоположения	Синтаксис	Описание
Номер единичного остатка	x	Указывает на единичный остаток в последовательности.
Номера остатков, ограничивающие некоторый участок последовательности	x..y	Указывает на ограниченный непрерывный диапазон остатков, включая стартовый и конечный остатки.
Остатки до первого или после последнего указанного номера остатка	<x >x <x..y x..>y <x..>y	Указывает на область, включающую определенный остаток или диапазон остатков и простирающуюся за пределы указанного остатка. Символы "<" и ">" могут быть использованы с номером единичного остатка или номерами стартового и конечного остатков диапазона для указания того, что характеристика выходит за пределы указанного номера остатка.

Стандарт ВОИС ST.26, пункт 66(a)

# Формат указания местоположения

## Дескрипторы местоположения для молекул всех типов

Пример местоположения	Описание
467	Указывает на остаток 467 в последовательности.
340..565	Указывает на непрерывный диапазон остатков, ограниченный остатками 340 и 565 (оба – включительно).
<1	Указывает на местоположение характеристики перед первым остатком.
<345..500	Указывает на то, что точная нижняя граница характеристики неизвестна. Местоположение начинается с некоторого остатка, предшествующего 345, и продолжается до остатка 500 включительно.
<1..888	Указывает на то, что характеристика начинается до первого остатка последовательности и продолжается до остатка 888 включительно.
1..>888	Указывает на то, что характеристика начинается с первого остатка последовательности и продолжается за пределами остатка 888.
<1..>888	Указывает на то, что характеристика начинается до первого остатка последовательности и продолжается за пределами остатка 888.

Стандарт ВОИС ST.26, пункт 70(a)

# Формат указания местоположения

## Дескрипторы местоположения, используемые только для нуклеотидных последовательностей

- Дескриптор местоположения в следующем формате можно использовать **ТОЛЬКО** для последовательностей ДНК и РНК:

Дескриптор местоположения	Синтаксис	Описание
Участок между двумя смежными нуклеотидами	$x^y$	Указывает на участок между двумя смежными нуклеотидами, например место разложения эндонуклеазы. Номера позиций смежных нуклеотидов отделяются каратом (^). Допустимыми форматами для этого дескриптора являются $x^{x+1}$ (например, $55^{56}$ ), или, для циклических нуклеотидов $x^1$ , где «x» это полная длина молекулы, то есть $1000^1$ для циклической молекулы длиной 1000.

Стандарт ВОИС ST.26, пункт 66(b)

# Формат указания местоположения

Дескрипторы местоположения, используемые только для нуклеотидных последовательностей

- При использовании дескриптора местоположения в формате x^y значения x и y должны представлять собой смежные остатки

Пример местоположения	Описание
123^124	Указывает на участок между остатками 123 и 124.
867^1	В циклической молекуле с 867 остатками указывает на участок между остатком, указанным как позиция 1, и остатком, указанным как позиция 867.

*Стандарт ВОИС ST.26, пункт 70(b)*

# Формат указания местоположения

## Дескрипторы местоположения: особенности для аминокислотных последовательностей

- Дескриптор местоположения x..y указывает на внутрицепочечную перекрестную связь между двумя указанными остатками при использовании с ключом характеристики «CROSSLNK» или «DISULFID»

Дескриптор местоположения	Синтаксис	Описание
Номера остатков, соединенных внутрицепочечной перекрестной связью	x..y	Указывает на аминокислоты, соединенные внутрицепочечной связью, при использовании с характеристикой, которая указывает на наличие внутрицепочечной перекрестной связи, например «CROSSLNK» или «DISULFID».

Стандарт ВОИС ST.26, пункт 66(c)

# Формат указания местоположения

## Дескрипторы местоположения: особенности для аминокислотных последовательностей

```

<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>QKKMIQFFKITHRYYYDIIIEHLCAKYDMNSVISNALFAKLNLMQYTDGLSTHEKIIINTSNPLTFSIVISLQRCVINLGGST</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```



# Формат указания местоположения

## Операторы местоположения для нуклеотидных последовательностей

- Имеется три оператора местоположения для использования с последовательностями ДНК и РНК: «join», «order» и «complement»
- Использование операторов «join» подразумевает, что остатки, описываемые дескрипторами местоположения, физически контактируют в биологических процессах (стандарт ST.26, пункт 68)
- При использовании операторов «join» и «order» должно быть указано минимум два дескриптора местоположения, разделенных запятой
- Оператор «complement» можно использовать в сочетании с операторами «join» или «order»

Синтаксис местоположения	Описание местоположения
join (местоположение, местоположение, ..., местоположение)	Указанные местоположения объединяются (друг за другом), образуя одну непрерывную последовательность.
order (местоположение, местоположение, ... местоположение)	Элементы следуют в указанном порядке, при этом не подразумевается необходимость их объединения.
complement (местоположение)	Указывает, что характеристика расположена на нити ДНК, являющейся комплементарной относительно диапазона последовательности, определенного указателем местоположения, при считывании в направлении от 5' к 3' или в направлении, имитирующем направление от 5' к 3'.

# Формат указания местоположения

## Операторы местоположения для нуклеотидных последовательностей

Пример местоположения	Описание
123^124	Указывает на участок между остатками 123 и 124.
join(12..78,134..202)	Указывает, что участки 12–78 и 134–202 должны быть объединены в одну непрерывную последовательность.
complement(34..126)	Начинается с нуклеотида, комплементарного нуклеотиду 126, и заканчивается нуклеотидом, комплементарным нуклеотиду 34 (характеристика находится на нити, комплементарной представленной нити).
complement(join(2691..4571, 4918..5163))	Объединяет нуклеотиды 2691–4571 и 4918–5163, а затем комплементирует соединенные сегменты (характеристика находится на нити, комплементарной представленной нити).
join(complement(4918..5163), complement(2691..4571))	Комплементирует участки 4918–5163 и 2691–4571, а затем соединяет комплементированные сегменты (характеристика находится на нити, комплементарной представленной нити).

Стандарт ВОИС ST.26, пункт 70(b)

# Вопросы?

# Формат значений квалификаторов

# Значения квалификаторов

- Квалификаторы помогают подробнее определить характеристики.

# Значения квалификаторов

- Квалификаторы помогают подробнее определить характеристики.
- Квалификатор состоит из названия квалификатора и часто значения квалификатора.

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

# Значения квалификаторов

- Квалификаторы помогают подробнее определить характеристики.
- Квалификатор состоит из названия квалификатора и часто значения квалификатора.

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

- Каждый ключ характеристики имеет список допустимых квалификаторов. Некоторые ключи характеристик имеют обязательные квалификаторы.

# Значения квалификаторов

- Квалификаторы помогают подробнее определить характеристики.
- Квалификатор состоит из названия квалификатора и часто значения квалификатора.

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

- Каждый ключ характеристики имеет список допустимых квалификаторов. Некоторые ключи характеристик имеют обязательные квалификаторы.
- Ключи характеристик и соответствующие допустимые квалификаторы приводятся в разделе 5 (нуклеотидные последовательности) и разделе 7 (аминокислотные последовательности) приложения I к стандарту ST.26.



# Значения квалификаторов

- Пример: ключ характеристики «misc\_binding» имеет один обязательный квалификатор («bound\_moiety») и 6 необязательных квалификаторов:

5.12.	Ключ характеристики	misc_binding
	Определение	участок в нуклеиновой кислоте, который ковалентно или нековалентно связывает другой фрагмент, который не может быть описан никаким другим ключом связывания (primer_bind или protein_bind)
	Обязательные квалификаторы	bound_moiety
	Необязательные квалификаторы	allele function gene gene_synonym map note
	Комментарий	обратите внимание, что для описания сайтов связывания рибосом необходимо использовать ключ характеристики «regulatory» и квалификатор «regulatory_class» со значением «ribosome_binding_site».

(Стандарт ST.26, приложение I, раздел 5.12)

# Значения квалификаторов

- Квалификаторы помогают подробнее определить характеристики.
- Квалификатор состоит из названия квалификатора и, часто, значения квалификатора.

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

- Каждый ключ характеристики имеет список допустимых квалификаторов. Некоторые ключи характеристик имеют обязательные квалификаторы.
- Ключи характеристик и соответствующие допустимые квалификаторы приводятся в разделе 5 (нуклеотидные последовательности) и разделе 7 (аминокислотные последовательности) приложения I к стандарту ST.26.
- Квалификаторы для нуклеотидных последовательностей вместе с их описанием приводятся в разделе 6 (нуклеотидные последовательности) и разделе 8 (аминокислотные последовательности) приложения I к стандарту ST.26.

# Значения квалификаторов

---

6.3.	Квалификатор	bound_moiety
	Определение	название молекулы/комплекса, который может связываться с данной характеристикой
	Формат обязательного значения	свободный текст
		Зависящий от языка: это значение может потребовать перевода для прохождения национальных/региональных процедур
	Пример	<INSDQualifier_value>GAL4</INSDQualifier_value>
	Комментарий	Для характеристик «misc_binding», «oriT» и «protein_bind» допускается использование одного квалификатора «bound_moiety».

---

6.22.	Квалификатор	gene
	Определение	символ гена, соответствующего области последовательности
	Формат обязательного значения	свободный текст
	Пример	<INSDQualifier_value>ilvE</INSDQualifier_value>
	Комментарий	Используйте квалификатор «gene» для указания символа гена; используйте квалификатор «standard_name» для указания полного названия гена.

(Стандарт ST.26, приложение I, разделы 6.3 и 6.22)

# Значения квалификаторов

## Форматы

- Имеется несколько форматов значений квалификаторов:
  1. Квалификаторы с заданным списком значений;
  2. Квалификаторы со значением в заданном формате;
  3. Квалификаторы, значения которых представляют собой последовательность;
  4. Квалификаторы БЕЗ значения;
  5. Квалификаторы со значениями в виде «свободного текста»  
- значения квалификаторов в виде «свободного текста» относятся к категории «зависящих от языка»

# Значения квалификаторов

Форматы: заданный список значений

- Квалификаторы с заданным списком значений

- Примеры:

«codon\_start» может иметь значение «1», «2» или «3»

---

6.9.	Квалификатор	codon_start
	Определение	указывает смещение, при котором может быть найден первый полный кодон кодирующей характеристики, относительно первого основания этой характеристики
	Формат обязательного значения	1 или 2 или 3
	Пример	<INSDQualifier_value>2</INSDQualifier_value>

# Значения квалификаторов

## Форматы: заданный список значений

### ■ Примеры: «rpt\_type» имеет ограниченное число возможных значений

6.60.	Квалификатор	rpt_type
	Определение	структура и распределение повторяющейся последовательности
	Формат обязательного значения	Одно из следующих слов (выражений) из контролируемой лексики: 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
	Пример	<INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value>
	Комментарий	Определение значений: tandem – повтор, который находится рядом с другим повтором в той же ориентации; direct – повтор, который находится не всегда рядом, но имеет ту же ориентацию; inverted – пара повторов в обратной ориентации друг относительно друга на одной и той же молекуле;

# Значения квалификаторов

Форматы: заданный список значений

- Для квалификаторов с заданным списком значений в приложении WIPO Sequence все допустимые значения указаны в выпадающем списке значений

▼ 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
gttggcgccg	caaaatatca	ttgcctctgt	gcaaggattg	gataaacata	ctgctgagtt

180

# Значения квалификаторов

Формат: значение в заданном формате

- Квалификаторы со значением в заданном формате
- Пример: «anticodon» должен иметь значение в следующем формате  
«(pos:<местоположение>,aa:<аминокислота>,seq:<текст>)»

6.2.	Квалификатор	anticodon
	Определение	расположение антикодона тРНК и аминокислоты, которую он кодирует
	Формат обязательного значения	(pos:<местоположение>,aa:<аминокислота>,seq:<текст>) где <местоположение> – позиция антикодона, <аминокислота> – трехбуквенная аббревиатура кодируемой аминокислоты, а <текст> – последовательность антикодона
	Пример	<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>



# Значения квалификаторов

## Форматы: последовательности

- Квалификаторы, значения которых представляют собой последовательность
- Пример: «translation» должен иметь значение, представляющее собой последовательность из однобуквенных обозначений **АМИНОКИСЛОТ**

6.79.	Квалификатор	translation
	Определение	однобуквенная аминокислотная последовательность, полученная либо из стандартного (или универсального) генетического кода, либо из таблицы, как указано в квалификаторе «transl_table» и как определено исключением в квалификаторе «transl_except»
	формат обязательного значения	непрерывная строка однобуквенных обозначений аминокислот из раздела 3 настоящего приложения, «X» используется для исключений AA.
	Пример	<INSDQualifier_value>MASTFPPWYRGCASTPSLKGLIMCTW</INSDQualifier_value>
	Комментарий	используется только с характеристикой «CDS»; должен сопровождаться квалификатором «protein_id», если результат трансляции содержит четыре или более специально определенных аминокислот; см. информацию о квалификаторе «transl_table», содержащую определение и сведения о том, где находятся таблицы генетических кодов; только один из квалификаторов «translation», «pseudo» и «pseudogene»  разрешено использовать для дальнейшего аннотирования характеристики CDS.

# Значения квалификаторов

## Форматы: последовательности

- Квалификатор «replace» может иметь единичный нуклеотидный остаток или последовательность остатков в качестве значения или может быть оставлен пустым
- Пустое значение квалификатора «replace» означает удаление остатка, указанного в соответствующей характеристике

---

6.57.	Квалификатор	replace
	Определение	указывает, что последовательность, определенная местоположением характеристики, заменяется последовательностью, указанной в значении квалификатора; если в квалификаторе нет последовательности (т.е. нет значения), это означает удаление последовательности
	формат обязательного значения	свободный текст
	Пример	<INSDQualifier_value>a</INSDQualifier_value> <INSDQualifier_value></INSDQualifier_value> - для удаления

# Значения квалификаторов

## Форматы: последовательности

- Квалификатор «replace» может иметь единичный нуклеотидный остаток или последовательность остатков в качестве значения или может быть оставлен пустым
- Пустое значение квалификатора «replace» означает удаление остатка, указанного в соответствующей характеристике

6.57.	Квалификатор	replace
	Определение	указывает, что последовательность, определенная местоположением характеристики, заменяется последовательностью, указанной в значении квалификатора; если в квалификаторе нет последовательности (т.е. нет значения), это означает удаление последовательности
	формат обязательного значения	свободный текст
	Пример	<pre>&lt;INSDQualifier_value&gt;a&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;&lt;/INSDQualifier_value&gt; - для удаления</pre>

**Пустое значение**



# Значения квалификаторов

## Формат: отсутствие значения

- Квалификаторы БЕЗ значения

- Примеры:

«environmental\_sample»    «germline»    «macronuclear»    «proviral»

6.51.	Квалификатор	proviral
	Определение	этот квалификатор используется для того, чтобы отметить последовательность, полученную из вируса или фага и интегрированную в геном другого организма
	Формат значения	не имеет значения

- Приложение WIPO Sequence не позволяет добавлять какое-либо значение квалификаторам без значения
- У этих квалификаторов элемент «INSDQualifier\_value» не должен быть пустым

# Значения квалификаторов

## Формат: свободный текст

- Многие квалификаторы имеют значения в формате свободного текста.
- В пункте 3(n) стандарта ST.26 приводится определение понятия «свободный текст»: «...тип формата значений для определенных квалификаторов, представляющий собой описательную текстовую фразу или другой определенный формат (как указано в приложении I)».
- Длина значения квалификаторов в виде свободного текста ограничена 1000 символов (стандарт ST.26, пункт 86).
- Подгруппа квалификаторов со значениями в виде свободного текста называется «зависящими от языка» квалификаторами.
- Зависящий от языка свободный текст «может потребовать перевода для прохождения национальных, региональных или международных процедур» (стандарт ST.26, пункт 3(o)).

# Значения квалификаторов

## Формат: свободный текст

- Квалификаторы со значениями в формате свободного текста могут потребовать перевода
- Как определить, является ли квалификатор со значением в формате свободного текста зависящим от языка?
  - В таблице 5 в разделе 6 приложения I к стандарту ST.26 перечислены все квалификаторы нуклеотидных последовательностей с зависящими от языка значениями в формате свободного текста
  - В таблице 6 в разделе 8 приложения I к стандарту ST.26 перечислены все квалификаторы аминокислотных последовательностей с зависящими от языка значениями в формате свободного текста
  - См. раздел «Mandatory value format» («Формат обязательного значения») в описании квалификатора

6.5.	Квалификатор	cell_type
	Определение	тип клетки, из которой была получена последовательность
	Формат обязательного значения	свободный текст
		Зависящий от языка: это значение может потребовать перевода для прохождения национальных/региональных процедур
	Пример	<INSDQualifier_value>leukocyte</INSDQualifier_value>

# Значения квалификаторов

## Формат: свободный текст

6.20.	Квалификатор	frequency
	Определение	частота, с которой встречается та или иная характеристика
	формат обязательного значения	<b>свободный текст</b> в виде дроби, указывающей долю популяции, имеющую данную характеристику
	Пример	<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.	Квалификатор	function
	Определение	функция, которую выполняет последовательность
	формат обязательного значения	<b>свободный текст</b> Зависящий от языка: это значение может потребовать перевода для прохождения национальных/региональных процедур
	Пример	<pre>&lt;INSDQualifier_value&gt;essential for recognition of cofactor &lt;/INSDQualifier_value&gt;</pre>
	Комментарий	Квалификатор «function» используется, когда название гена и/или название продукта не передают функцию, которую выполняет последовательность.

# Значения квалификаторов

Формат: зависящий от языка свободный текст

- В перечне последовательностей в формате XML значения квалификаторов в формате «зависящего от языка свободного текста» могут быть указаны на двух языках: на английском и на любом другом языке (стандарт ST.26, пункт 87)



# Значения квалификаторов

Формат: зависящий от языка свободный текст

- В перечне последовательностей в формате XML значения квалификаторов в формате «зависящего от языка свободного текста» могут быть указаны на двух языках: на английском и на любом другом языке (стандарт ST.26, пункт 87)
- Значение зависящего от языка квалификатора на английском языке должно быть указано в элементе «INSDQualifier\_value»

# Значения квалификаторов

Формат: зависящий от языка свободный текст

- В перечне последовательностей в формате XML значения квалификаторов в формате «зависящего от языка свободного текста» могут быть указаны на двух языках: на английском и на любом другом языке (стандарт ST.26, пункт 87)
- Значение зависящего от языка квалификатора на английском языке должно быть указано в элементе «INSDQualifier\_value»
- Значение зависящего от языка квалификатора на любом отличном от английского языке должно быть указано в элементе «NonEnglishQualifier\_value»

# Значения квалификаторов

Формат: зависящий от языка свободный текст

- В перечне последовательностей в формате XML значения квалификаторов в формате «зависящего от языка свободного текста» могут быть указаны на двух языках: на английском и на любом другом языке (стандарт ST.26, пункт 87)
- Значение зависящего от языка квалификатора на английском языке должно быть указано в элементе «INSDQualifier\_value»
- Значение зависящего от языка квалификатора на любом отличном от английского языке должно быть указано в элементе «NonEnglishQualifier\_value»
- Элемент «NonEnglishQualifier\_value» разрешен ТОЛЬКО для квалификаторов со значением в формате зависящего от языка свободного текста (стандарт ST.26, пункт 87(b))

# Значения квалификаторов

Формат: зависящий от языка свободный текст

```

<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">

```

# Значения квалификаторов

Формат: зависящий от языка свободный текст

- Если перечень последовательностей в формате XML содержит значения квалификаторов не на английском языке:

1. Корневой элемент перечня последовательностей в формате XML должен содержать атрибут «nonEnglishFreeTextLanguageCode» с двухбуквенным кодом соответствующего языка (стандарт ST.26, пункты 43 и 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">
```

# Значения квалификаторов

Формат: зависящий от языка свободный текст

- Если перечень последовательностей в формате XML содержит значения квалификаторов не на английском языке:

1. Корневой элемент перечня последовательностей в формате XML должен содержать атрибут «nonEnglishFreeTextLanguageCode» с двухбуквенным кодом соответствующего языка (стандарт ST.26, пункты 43 и 87(b));

2. ВСЕ зависящие от языка квалификаторы в перечне последовательностей должны иметь значения на языке, указанном в атрибуте «nonEnglishFreeTextLanguageCode»;

# Значения квалификаторов

## Формат: зависящий от языка свободный текст

- Если перечень последовательностей в формате XML содержит значения квалификаторов не на английском языке:

1. Корневой элемент перечня последовательностей в формате XML должен содержать атрибут «nonEnglishFreeTextLanguageCode» с двухбуквенным кодом соответствующего языка (стандарт ST.26, пункты 43 и 87(b));
2. ВСЕ зависящие от языка квалификаторы в перечне последовательностей должны иметь значения на языке, указанном в атрибуте «nonEnglishFreeTextLanguageCode»;
3. Если у одного квалификатора имеются значения «NonEnglishQualifier\_value» и «INSDQualifier\_value», то информация, содержащаяся в этих двух элементах, должна быть эквивалентной (стандарт ST.26, пункт 87(c)).

# Значения квалификаторов

Формат: зависящий от языка свободный текст

- Для чего нужен атрибут «id» элемента «INSDQualifier»?
- В пункте 87(d) стандарта ST.26 сказано следующее: «Для зависящих от языка квалификаторов элемент “INSDQualifier” может включать необязательный атрибут “id”. Значение этого атрибута должно иметь следующий формат: буква “q”, за которой следует положительное целое число, например “q23”, и должно быть уникальным для каждого элемента “INSDQualifier”, т.е. данное значение атрибута должно использоваться в файле перечня последовательностей только один раз».

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



# Значения квалификаторов

Формат: зависящий от языка свободный текст

- Атрибуты «id» элементов «INSDQualifier»:
  - Используются для уникальной идентификации значений квалификаторов, которые могут потребовать перевода для экспорта в файл в формате XLIFF в приложении WIPO Sequence;
  - Являются необязательными;
  - Разрешены только для зависящих от языка квалификаторов;
  - Должны быть уникальными в перечне последовательностей;
  - Добавляются автоматически в перечень последовательностей в формате XML, генерируемый приложением WIPO Sequence.

# Значения квалификаторов

Формат: зависящий от языка свободный текст

- Атрибут «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">
```

- В пункте 43 стандарта ST.26 атрибут «originalFreeTextLanguageCode» определен как «код языка... для одного оригинального языка, на котором представлены квалификаторы в формате зависящего от языка свободного текста»

- Этот атрибут является необязательным

Особый случай:  
комбинированные молекулы  
ДНК/РНК  
урацил в ДНК и тимин в РНК

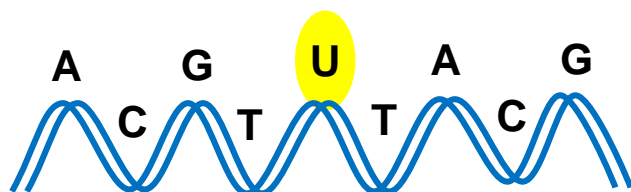
# Особый случай

- Напоминание: использование «u» для обозначения урацила не разрешено стандартом ST.26
- В ДНК «t» означает тимин
- В РНК «t» означает урацил
- Рассмотрим две ситуации:
  1. Молекула ДНК с урациловым нуклеотидным основанием или молекула РНК с тиминным нуклеотидным основанием
  2. Комбинированная молекула ДНК/РНК

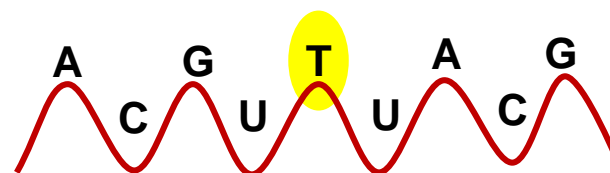
# Особый случай

## Урацил в ДНК и тимин в РНК

- Если последовательность имеет основную цепь ДНК с урациловым нуклеотидным основанием или основную цепь РНК с тиминным нуклеотидным основанием, такая последовательность является «модифицированным нуклеотидом»



ДНК



РНК

Применяется пункт 14 стандарта ST.26 14:

«14. Символ “t” означает тимин в ДНК и урацил в РНК. Урацил в ДНК или тимин в РНК считается **модифицированным нуклеотидом** и требует дальнейшего описания в таблице характеристик в соответствии с пунктом 19».

# Особый случай

## ДНК- и РНК-последовательности

В заявке раскрыта следующая РНК-последовательность:

5' -сгисссасгигтссгагга-3'

- Обратите внимание на «тимин» в позиции 12. Этот остаток требует аннотации в качестве модифицированного нуклеотида.
- В пункте 19 стандарта ST.26 сказано следующее: *«Урацил в ДНК или тимин в РНК считаются модифицированными нуклеотидами и должны быть представлены в последовательности как “t” и далее описаны в таблице характеристик с помощью специального ключа “modified\_base”, квалификатора “mod\_base” со значением “OTHER” и квалификатора “note” со значением “uracil” или “thymine”, соответственно».*

# Особый случай

## ДНК- и РНК-последовательности

В заявке раскрыта следующая РНК-последовательность:

5' -сгисссасгиггтссгаггуа-3'

- Обратите внимание на «тимин» в позиции 12. Этот остаток требует аннотации в качестве модифицированного нуклеотида.
- В пункте 19 стандарта ST.26 сказано следующее: «Урацил в ДНК или тимин в РНК считаются модифицированными нуклеотидами и должны быть представлены в последовательности как “t” и далее описаны в таблице характеристик с помощью специального ключа “modified\_base”, квалификатора “mod\_base” со значением “OTHER” и квалификатора “note” со значением “uracil” или “thymine”, соответственно».

# Особый случай

## ДНК- и РНК-последовательности

В заявке раскрыта следующая РНК-последовательность:

5' -сгисссасгиггссгaggua -3'

- ✓ Все урациловые остатки должны быть представлены символом «t». Таким образом, в перечне последовательностей эта последовательность должна выглядеть следующим образом:

сgtccсacgtgгссгaggta



# Особый случай

## ДНК- и РНК-последовательности

В заявке раскрыта следующая РНК-последовательность:

5' -сgисссасgигtссgaggua -3'

- ✓ Все урациловые остатки должны быть представлены символом «t». Таким образом, в перечне последовательностей эта последовательность должна выглядеть следующим образом:

сgtсссасgtgtссgaggta

- ✓ Ключ характеристики «modified\_base» со значением местоположения «12»

# Особый случай

## ДНК- и РНК-последовательности

В заявке раскрыта следующая РНК-последовательность:

5' -сgисссасgигтссgаgгuа -3'

- ✓ Все урациловые остатки должны быть представлены символом «t». Таким образом, в перечне последовательностей эта последовательность должна выглядеть следующим образом:

сgтсссасgтгтссgаgгtа

- ✓ Ключ характеристики «modified\_base» со значением местоположения «12»
- ✓ Квалификатор «mod\_base» со значением «OTHER»

# Особый случай

## ДНК- и РНК-последовательности

В заявке раскрыта следующая РНК-последовательность:

**5' -сgисссасgигтссgаgгuа -3'**

- ✓ Все урациловые остатки должны быть представлены символом «t». Таким образом, в перечне последовательностей эта последовательность должна выглядеть следующим образом:

**сgтсссасgтгтссgаgгtа**

- ✓ Ключ характеристики «modified\_base» со значением местоположения «12»
- ✓ Квалификатор «mod\_base» со значением «OTHER»
- ✓ Квалификатор «note» со значением «thymine»

# Особый случай ДНК- и РНК-последовательности

```

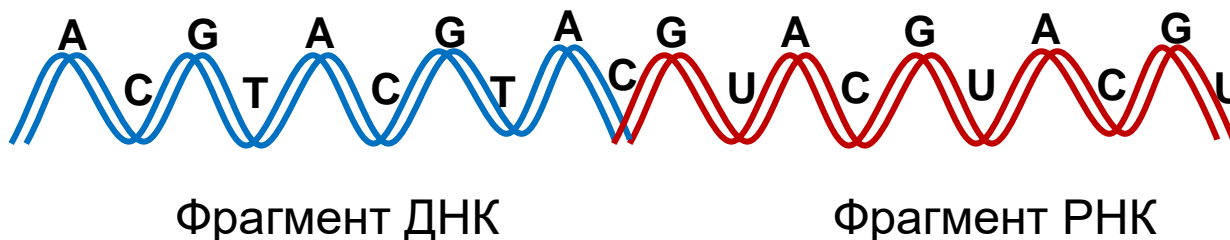
<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length>20</INSDSeq_length>
    <INSDSeq_moltype>RNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..20</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>cgTcccacgtgtcgaggta</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

# Особый случай

## Комбинированные последовательности ДНК/РНК

- Если последовательность является комбинированной молекулой, т.е. одна часть основной цепочки представляет собой ДНК, а другая — РНК:



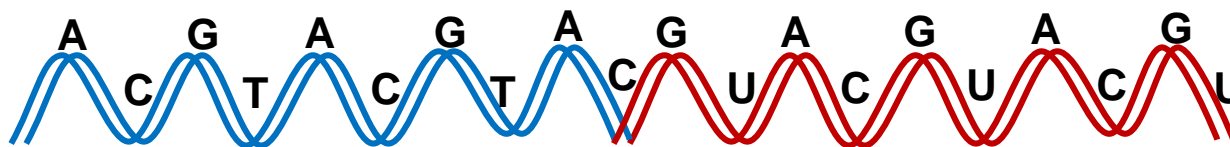
Применяется пункт 55 стандарта ST.26:

*«55. Для нуклеотидной последовательности, которая содержит как ДНК, так и РНК фрагменты одного или более нуклеотидов, тип молекулы должен быть указан как “ДНК”. Комбинированная молекула ДНК/РНК должна быть далее описана в таблице характеристик с помощью ключа характеристики “source” и обязательного квалификатора “organism” со значением “synthetic construct” и обязательного квалификатора “mol\_type” со значением “other DNA”. Каждый ДНК и РНК фрагмент комбинированной молекулы ДНК/РНК должен быть в дальнейшем описан при помощи ключа характеристики “misc\_feature” и квалификатора “note”, который указывает, является ли фрагмент ДНК или РНК».*

# Особый случай

## Комбинированные последовательности ДНК/РНК

- Если последовательность является комбинированной молекулой, т.е. одна часть основной цепочки представляет собой ДНК, а другая — РНК:



Фрагмент ДНК

Фрагмент РНК

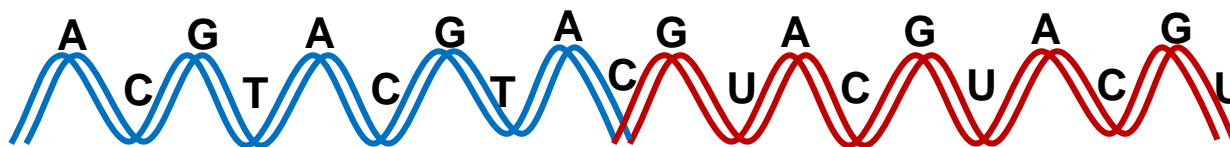
Применяется пункт 55 стандарта ST.26:

«55. Для нуклеотидной последовательности, которая содержит как ДНК, так и РНК фрагменты одного или более нуклеотидов, **тип молекулы должен быть указан как “ДНК”**. Комбинированная молекула ДНК/РНК должна быть далее описана в таблице характеристик с помощью ключа характеристики “source” и обязательного квалификатора “organism” со значением “synthetic construct” и обязательного квалификатора “mol\_type” со значением “other DNA”. Каждый ДНК и РНК фрагмент комбинированной молекулы ДНК/РНК должен быть в дальнейшем описан при помощи ключа характеристики “misc\_feature” и квалификатора “note”, который указывает, является ли фрагмент ДНК или РНК».

# Особый случай

## Комбинированные последовательности ДНК/РНК

- Если последовательность является комбинированной молекулой, т.е. одна часть основной цепочки представляет собой ДНК, а другая — РНК:



Фрагмент ДНК

Фрагмент РНК

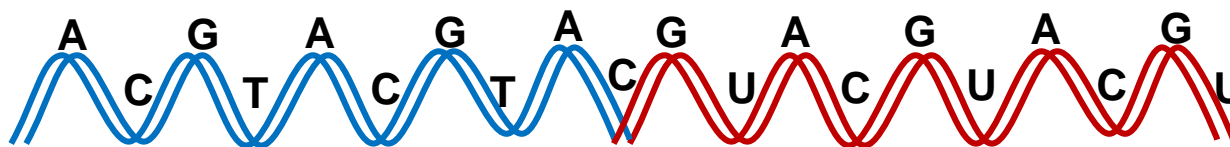
Применяется пункт 55 стандарта ST.26:

«55. Для нуклеотидной последовательности, которая содержит как ДНК, так и РНК фрагменты одного или более нуклеотидов, тип молекулы должен быть указан как «ДНК». Комбинированная молекула ДНК/РНК должна быть далее описана в таблице характеристик с помощью ключа характеристики «source» и обязательного квалификатора **«organism» со значением «synthetic construct»** и обязательного квалификатора «mol\_type» со значением «other DNA». Каждый ДНК и РНК фрагмент комбинированной молекулы ДНК/РНК должен быть в дальнейшем описан при помощи ключа характеристики «misc\_feature» и квалификатора «note», который указывает, является ли фрагмент ДНК или РНК».

# Особый случай

## Комбинированные последовательности ДНК/РНК

- Если последовательность является комбинированной молекулой, т.е. одна часть основной цепочки представляет собой ДНК, а другая — РНК:



Фрагмент ДНК

Фрагмент РНК

Применяется пункт 55 стандарта ST.26:

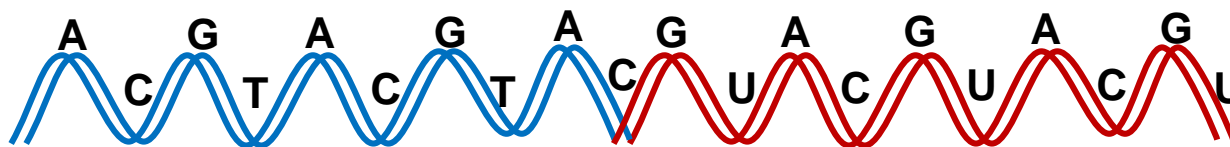
«55. Для нуклеотидной последовательности, которая содержит как ДНК, так и РНК фрагменты одного или более нуклеотидов, тип молекулы должен быть указан как «ДНК». Комбинированная молекула ДНК/РНК должна быть далее описана в таблице характеристик с помощью ключа характеристики «source» и обязательного квалификатора «organism» со значением «synthetic construct» и **обязательного квалификатора «mol\_type» со значением «other DNA»**. Каждый ДНК и РНК фрагмент комбинированной молекулы ДНК/РНК должен быть в дальнейшем описан при помощи ключа характеристики «misc\_feature» и квалификатора «note», который указывает, является ли фрагмент ДНК или РНК».



# Особый случай

## Комбинированные последовательности ДНК/РНК

- Если последовательность является комбинированной молекулой, т.е. одна часть основной цепочки представляет собой ДНК, а другая — РНК:



Фрагмент ДНК

Фрагмент РНК

Применяется пункт 55 стандарта ST.26:

«55. Для нуклеотидной последовательности, которая содержит как ДНК, так и РНК фрагменты одного или более нуклеотидов, тип молекулы должен быть указан как “ДНК”. Комбинированная молекула ДНК/РНК должна быть далее описана в таблице характеристик с помощью ключа характеристики “source” и обязательного квалификатора “organism” со значением “synthetic construct” и обязательного квалификатора “mol\_type” со значением “other DNA”. Каждый ДНК и РНК фрагмент комбинированной молекулы ДНК/РНК должен быть в дальнейшем описан при помощи **ключа характеристики “misc\_feature” и квалификатора “note”, который указывает, является ли фрагмент ДНК или РНК».**

# Особый случай

## Комбинированные последовательности ДНК/РНК

В заявке раскрыта следующая комбинированная последовательность ДНК/РНК:

**5' -АССТGСсgисссасgигиссgаgгуаGСАТТА-3'**

где символы в верхнем регистре означают фрагмент ДНК, а символы в нижнем регистре — фрагмент РНК.

- Остатки 1-6 и 27-32 — ДНК, остатки 7-26 — РНК
- Вопросы:
  1. Значение квалификатора «organism»
  2. Тип молекулы и значение квалификатора «mol\_type»
  3. Идентификация фрагментов ДНК и РНК

# Особый случай

## Комбинированные последовательности ДНК/РНК

В заявке раскрыта следующая комбинированная последовательность ДНК/РНК:

**5' -АССТGСсgисссасgигиссgаgгuаGСАТТА-3'**

где символы в верхнем регистре означают фрагмент ДНК, а символы в нижнем регистре — фрагмент РНК.

- В пункте 55 стандарта ST.26 сказано следующее: *«...тип молекулы должен быть указан как “ДНК”. Комбинированная молекула ДНК/РНК должна быть далее описана в таблице характеристик с помощью ключа характеристики “source” и обязательного квалификатора “organism” со значением “synthetic construct” и обязательного квалификатора “mol\_type” со значением “other DNA”».*
- ✓ Тип молекулы = «DNA»
- ✓ Значение квалификатора «organism» = «synthetic construct»
- ✓ Значение квалификатора «mol\_type» = «other DNA»

# Особый случай

## Комбинированные последовательности ДНК/РНК

В заявке раскрыта следующая комбинированная последовательность ДНК/РНК:

**5' -АССТGСсgисссасgигиссgаgguаGСАТТА-3'**

где символы в верхнем регистре означают фрагмент ДНК, а символы в нижнем регистре — фрагмент РНК.

- ✓ Все урациловые остатки должны быть представлены символом «t». Таким образом, в перечне последовательностей эта последовательность должна выглядеть следующим образом:

**асctgссgtcccасgtgtccgaggtagcattа**

# Особый случай

## Комбинированные последовательности ДНК/РНК

В заявке раскрыта следующая комбинированная последовательность ДНК/РНК:

**5' -ACCTGCcсgисссасgигисссgаgгуаGCАТТА-3'**

где символы в верхнем регистре означают фрагмент ДНК, а символы в нижнем регистре — фрагмент РНК.

- В пункте 55 стандарта ST.26 сказано следующее: *«Каждый ДНК и РНК фрагмент комбинированной молекулы ДНК/РНК должен быть в дальнейшем описан при помощи ключа характеристики “misc\_feature” и квалификатора “note”, который указывает, является ли фрагмент ДНК или РНК».*
- ✓ Три фрагмента = три ключа характеристики «misc\_feature»

# Особый случай

## Комбинированные последовательности ДНК/РНК

В заявке раскрыта следующая комбинированная последовательность ДНК/РНК:

**5' –ACCTGCcgcucssascgucgucssgagguagCATTA–3'**

где символы в верхнем регистре означают фрагмент ДНК, а символы в нижнем регистре — фрагмент РНК.

- ✓ Фрагмент 1, остатки 1-6:  
Ключ характеристики «misc\_feature» со значением местоположения «1..6»  
Квалификатор «note» со значением «DNA»
- ✓ Фрагмент 2, остатки 7-26:  
Ключ характеристики «misc\_feature» со значением местоположения «7..26»  
Квалификатор «note» со значением «RNA»
- ✓ Фрагмент 3, остатки 27-32  
Ключ характеристики «misc\_feature» со значением местоположения «27..32»  
Квалификатор «note» со значением «DNA»

# Особый случай

## Комбинированные последовательности ДНК/РНК

```

<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>

```

# Особый случай

## Комбинированные последовательности ДНК/РНК

```

<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>

```

Фрагмент ДНК 1

Фрагмент РНК 2

Фрагмент ДНК 3

WIPO  
WORLD  
INTELLECTUAL PROPERTY  
ORGANIZATION



# Вопросы?

# Аналоги нуклеотидов, D- аминокислоты и разветвленные последовательности

# Новые типы молекул

## Аналоги нуклеотидов

- В отношении последовательностей нуклеиновых кислот, содержащих один или несколько аналогов нуклеотидов, применяются правила стандарта ST.26
- О нуклеотидных аналогах речь идет в определении понятия «нуклеотид» в пункте 3(g)(2) стандарта ST.26:

*«аналог 2'-дезоксирибозо-5'-монофосфата или рибозо 5'-монофосфата, который при образовании основной цепи аналога нуклеиновой кислоты приводит к расположению нуклеотидных оснований, которые имитируют расположение нуклеотидов в нуклеиновых кислотах, содержащих 2'-дезоксирибозо-5'-монофосфата или рибозо-5-монофосфата, где аналог нуклеиновой кислоты способен к сдваиванию оснований с комплементарной нуклеиновой кислотой»*

- Примеры аналогов нуклеотидов: пептидные нуклеиновых кислотах (ПНК), гликолевые нуклеиновые кислоты (ГНК), треозные нуклеиновые кислоты и морфолины
- Должны быть представлены в направлении слева направо, которое имитирует направление от 5' к 3' (стандарт ST.26, пункт 11)

# Новые типы молекул

## Аналоги нуклеотидов

В патентной заявке раскрыта следующая последовательность гликолевой нуклеиновой кислоты (ГНК):

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

где  $\text{PO}_4$  — конец последовательности, имитирующий конец 5' последовательности ДНК.

- ✓ Эту последовательность необходимо включить в перечень последовательностей

# Новые типы молекул

## Аналоги нуклеотидов

В патентной заявке раскрыта следующая последовательность гликолевой нуклеиновой кислоты (ГНК):



где  $\text{PO}_4$  — конец последовательности, имитирующий конец 5' последовательности ДНК.

- ✓ Эту последовательность необходимо включить в перечень последовательностей
- ✓ Конец  $\text{PO}_4$  имитирует конец 5', поэтому последовательность должна быть представлена в показанном направлении

# Новые типы молекул

## Аналоги нуклеотидов

В патентной заявке раскрыта следующая последовательность гликолевой нуклеиновой кислоты (ГНК):

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

где  $\text{PO}_4$  — конец последовательности, имитирующий конец 5' последовательности ДНК.

- ✓ Эту последовательность необходимо включить в перечень последовательностей
- ✓ Конец  $\text{PO}_4$  имитирует конец 5', поэтому последовательность должна быть представлена в показанном направлении
- ✓ Вся последовательность должна быть аннотирована с использованием ключа характеристики «modified\_base», квалификатора «mod\_base» со значением «OTHER» и квалификатора «note», содержащего полное, несокращенное название модифицированных нуклеотидов, например, «гликолевые нуклеиновые кислоты».

# Новые типы молекул

## Аналоги нуклеотидов

```

- <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>

```

*Примечание: этот случай подробно рассматривается в примере 3(g)-4 в приложении VI к стандарту ВОИС ST.26.*

# Новые типы молекул

## D-аминокислоты

- В отношении аминокислотных последовательностей, содержащих одну или более D-аминокислот, применяются правила стандарта ST.26
- О D-аминокислотах речь идет в определении понятия «аминокислота» в пункте 3(a) стандарта ST.26:

*«аминокислота» означает любую аминокислоту, которая может быть представлена с помощью любого символа, определенного в приложении I (см. раздел 3, Таблица 3). Такие аминокислоты включают, в частности, D-аминокислоты и аминокислоты, содержащие модифицированные или синтетические боковые цепи».*

- D- аминокислоты должны быть представлены в последовательности в виде символа соответствующей немодифицированной L-аминокислоты, когда это возможно
- Должны быть описаны в таблице характеристик в качестве модифицированной аминокислоты



# Новые типы молекул

## D-аминокислоты

В патентной заявке описана следующая последовательность:

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

- ✓ Эту последовательность необходимо включить в перечень последовательностей

# Новые типы молекул

## D-аминокислоты

В патентной заявке описана следующая последовательность:

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

- ✓ Эту последовательность необходимо включить в перечень последовательностей
- ✓ Должна быть представлена как AGKLGМ
- ✓ Аланин в позиции 1, глутаминовая кислота в позиции 2 и метионин в позиции 6 должны быть аннотированы с использованием ключа характеристики «SITE» и квалификатора «NOTE» с указанием полного, несокращенного названия соответствующей аминокислоты

# Новые типы молекул

## D-аминокислоты

```

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

```

Примечание: похожий случай подробно рассматривается в примере 3(a)-1 в приложении VI к стандарту ВОИС ST.26.

# Новые типы молекул

## Разветвленные последовательности

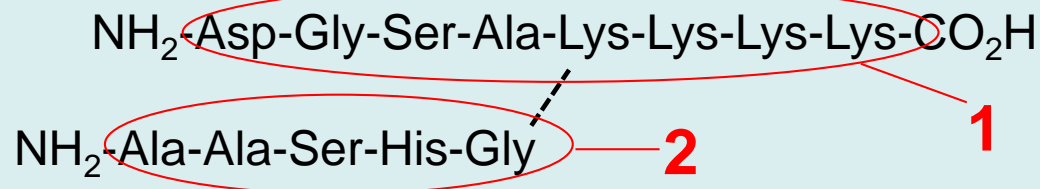
- В отношении разветвленных последовательностей нуклеиновых кислот и разветвленных последовательностей аминокислот применяются правила стандарта ST.26
- Линейные участки разветвленных последовательностей должны быть включены в перечень последовательностей, если они содержат 10 или более специально определенных нуклеотидов или 4 или более специально определенных аминокислот (стандарт ВОИС ST.26, пункт 7)
- Каждый линейный участок разветвленной последовательности, отвечающий требованию минимальной длины, должен быть включен в перечень в виде отдельной последовательности с отдельным идентификационным номером последовательности
- Необходимо учитывать количество специально определенных остатков в каждом отдельном линейном участке, а не общее количество специально определенных остатков в структуре



# Новые типы молекул

## Разветвленные последовательности

В патентной заявке описан пептид со следующей последовательностью:



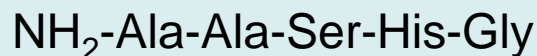
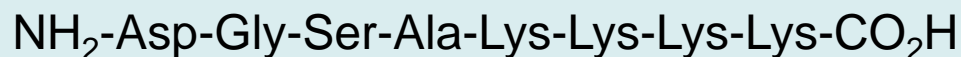
где----- указывает на амидную связь между карбоксильным концом глицина и боковой цепью лизина

- ✓ Оба линейных участка содержат  $\geq 4$  специально определенных аминокислот, поэтому оба они включены в перечень последовательностей
- ✓ Каждый линейный участок должен быть включен в перечень в виде отдельной последовательности с отдельным идентификационным номером последовательности

# Новые типы молекул

## Разветвленные последовательности

В патентной заявке описан пептид со следующей последовательностью:



где----- указывает на амидную связь между карбоксильным концом глицина и боковой цепью лизина

- ✓ Оба линейных участка содержат  $\geq 4$  специально определенных аминокислот, поэтому оба они включены в перечень последовательностей
- ✓ Каждый линейный участок должен быть включен в перечень в виде отдельной последовательности с отдельным идентификационным номером последовательности
- ✓ Обе последовательности должны быть аннотированы с указанием местоположения и характера амидной связи

# Новые типы молекул

## Разветвленные последовательности

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>

```

Примечание: похожий случай подробно рассматривается в примере 7(b)-3 в приложении VI к стандарту ВОИС ST.26.



# Варианты последовательностей

# Варианты последовательностей

- Понятие «**вариант последовательности**» определено в пункте 3(m) стандарта ST.26 следующим образом:

*«нуклеотидная или аминокислотная последовательность, которая содержит одно или несколько **отличий по сравнению с исходной последовательностью**. Эти отличия могут включать альтернативные остатки (см. пункты 15 и 27), модифицированные остатки (см. пункты 3(g), 3(h), 16 и 29), удаления, вставки и замены. См. пункты 93-95».*

- Способ раскрытия варианта последовательности определяет то, как он должен быть представлен в перечне последовательностей.
- Правила представления вариантов изложены в пунктах 93-95 стандарта ST.26.

# Варианты последовательностей

## Пункт 93

*Пункт 93: Исходная последовательность и любые варианты этой последовательности, которые раскрыты путем перечисления их остатков и на которые распространяется пункт 7, должны быть включены в перечень последовательностей под отдельным идентификационным номером последовательности.*



Если каждый вариант указывается отдельно, то каждый вариант должен иметь отдельный идентификационный номер последовательности!

# Варианты последовательностей

## Пункт 93

Патентная заявка содержит рисунок, на котором показано множественное выравнивание последовательностей:

Consensus	LEG <b>n</b> EQFINA <b>ak</b> IRHP <b>k</b> <b>nrk</b> TL <b>n</b> NDI <b>m</b> LIK
Homo sapiens	LEGNEQFINAAKIRHPQYDRKTLNNDIMLIK
Pongo abelii	LEGNEQFINAAKIRHPQYDRKTVNNDIMLIK
Pan paniscus	LEGNEQFINAAKIRHPKYNRI TLNNDIMLIK
Rhinopithecus bieti	LEGNEQFINATKIRHPKYNNGNTLNNDIMLIK
Rhinopithecus roxellana	LEGNEQFINATQIRHPKYNNGNTLNNDIMLIK

Буквы в нижнем регистре обозначают основные аминокислотные остатки в выровненных последовательностях.

- ✓ Каждая из шести приведенных последовательностей должна быть включена в перечень последовательностей в виде отдельной последовательности с отдельным идентификационным номером.

*Примечание: похожий случай подробно рассматривается в примере 93-3 в приложении VI к стандарту ВОИС ST.26.*

# Варианты последовательностей

## Правильная аннотация

- Правильный ключ характеристики для аннотирования варианта зависит от типа молекулы и характера изменения:

Тип последовательности	Ключ характеристики	Квалификатор	Использование
Нуклеиновая кислота	variation	replace или note	Естественно возникающие мутации и полиморфизмы, например аллели, ПДРФ.
Нуклеиновая кислота	misc_difference	replace или note	Вариабельность, вносимая искусственно, например путем генетических манипуляций или химического синтеза.
Аминокислота	VAR_SEQ	NOTE	Вариант, полученный путем проведения альтернативного сплайсинга, использования альтернативного промотора, альтернативного инициирования и рибосомного сдвига рамки.
Аминокислота	VARIANT	NOTE	Любой вариант, для которого VAR_SEQ не применим.

Стандарт ВОИС ST.26, пункт 96

# Варианты последовательностей

## Наиболее ограничивающий неоднозначный символ

- В пунктах 15 и 27 стандарта ST.26 сказано, что в случае необходимости использования неоднозначного символа «следует использовать наиболее ограничивающий неоднозначный символ...»

Что это значит?

### Пример:

Если в нуклеотидной последовательности какая-либо позиция имеет вид «a or c», следует использовать неоднозначный символ «m» вместо «n».

Если в нуклеотидной последовательности какая-либо позиция имеет вид «L or I», следует использовать неоднозначный символ «J» вместо «X».

Следует помнить, что «n» и «X» имеют значения по умолчанию, поэтому каждый раз, когда «n» или «X» используется для чего-то, отличного от значения по умолчанию, требуется аннотация.

# Варианты последовательностей

## Пункт 94

*Пункт 94: Любой вариант последовательности, раскрытой как одна последовательность с указанием альтернативных вариантов остатков на одной или нескольких позициях, должен быть включен в перечень последовательностей и должен быть представлен в виде одной последовательности, в которой указанные альтернативные варианты остатков представлены наиболее ограничивающим неоднозначным символом (см. пункты 15 и 27).*



Если варианты НЕ указываются отдельно, а просто показаны как изменяемые остатки в исходной последовательности, не требуется, чтобы варианты имели отдельный идентификационный номер последовательности!

# Варианты последовательностей

## Пункт 94

В патентной заявке раскрыта пептидная последовательность:  
Gly-Gly-Gly-[Leu or Ile]-Ala-Thr-[Ser or Thr]

- ✓ Может быть включена в перечень последовательностей в виде отдельной последовательности
- ✓ Предпочтительное представление последовательности: GGGJATX
- ✓ [Leu or Ile] должно быть представлено наиболее ограничивающим неоднозначным символом «J»
- ✓ [Ser or Thr] должно быть представлено символом «X» с ключом характеристики «VARIANT», имеющим квалификатор «note», в котором указано, что «X» означает серин или треонин

*Примечание: этот случай подробно рассматривается в примере 94-1 в приложении VI к стандарту ВОИС ST.26.*



# Варианты последовательностей

## Пункт 95

*Пункт 95: Любой вариант последовательности, раскрытой только посредством ссылки на удаление(я), добавление(я) или замещение(я) в исходной последовательности в перечне последовательностей, следует включать в перечень последовательностей. Такие варианты последовательностей, включенные в перечень последовательностей:*

*(a) могут быть представлены путем аннотирования исходной последовательности, если она содержит вариант(ы) на отдельной позиции или множестве разных позиций и появление этих вариантов происходит независимо;*

*(b) должны быть представлены как отдельная последовательность с отдельным идентификационным номером последовательности, если они содержат варианты на множестве разных позиций и возникновение данных вариантов является взаимосвязанным; и*

*(c) должны быть представлены как отдельная последовательность с отдельным идентификационным номером последовательности, если в последовательности содержатся замещения и добавления в более чем 1000 остатков (см. пункт 86).*



«Ссылки на удаление(я), добавление(я) или замещение(я)» означают, что варианты раскрываются в виде обычного текста.

# Варианты последовательностей

## Пункт 95(a)

В заявке раскрывается следующее:

Пептидный фрагмент 1: Gly-Leu-Pro-Хаа-Arg-Ile-Cys  
где Хаа — любая аминокислота

\* \* \*

... в еще одном воплощении пептидный фрагмент 1 — Gly-Leu-Pro-Хаа-Arg-Ile-Cys, где Хаа может быть Val, Thr или Asp.....

\* \* \*

... в еще одном воплощении пептидный фрагмент 1 — Gly-Leu-Pro-Хаа-Arg-Ile-Cys, где Хаа может быть Val...

*Исходная последовательность Gly-Leu-Pro-Хаа-Arg-Ile-Cys содержит вариант(ы) в одной позиции и появление этих вариантов происходит независимо*

# Варианты последовательностей

## Пункт 95(a)

- ✓ Исходная последовательность должна быть включена в перечень последовательностей, а варианты могут быть представлены путем аннотирования исходной последовательности
- ✓ Наилучшим воплощением должна быть версия, включенная в перечень последовательностей; в данном примере это версия, где «Хаа — любая аминокислота»
- ✓ Последовательность должна быть представлена как GLPXRIC и требует ключа характеристики «VARIANT» для позиции 4 и квалификатора «note» с указанием того, что Хаа означает любую аминокислоту
- ✓ Хотя это не обязательно, рекомендуется, чтобы следующие три варианта были включены в перечень последовательностей как отдельные последовательности:

GLPVRIC

GLPTRIC

GLPDRIC

*Примечание: этот случай подробно рассматривается в примере 95(a)-1 в приложении VI к стандарту ВОИС ST.26*

# Варианты последовательностей

## Пункт 95(b)

В патентной заявке описана следующая консенсусная последовательность:

aatgn<sub>1</sub>сссасgaaatgn<sub>2</sub>сас

где n<sub>1</sub> и n<sub>2</sub> могут представлять собой а, t, g или с.

Раскрыты следующие несколько вариантов последовательности:

если n<sub>1</sub> — а, то n<sub>2</sub> — t, g или с

если n<sub>1</sub> — t, то n<sub>2</sub> — а, g или с

если n<sub>1</sub> — g, то n<sub>2</sub> — t, а или с

если n<sub>1</sub> — с, то n<sub>2</sub> — t, g или а

*Консенсусная последовательность содержит варианты на множестве разных позиций и возникновение данных вариантов является взаимосвязанным*

# Варианты последовательностей

## Пункт 95(b)

- ✓ Консенсусная последовательность должна быть включена в перечень последовательностей, а варианты должны быть представлены в виде отдельных последовательностей
- ✓ Наилучшим воплощением должна быть версия, включенная в перечень последовательностей; в данном примере это версия, где  $n_1$  и  $n_2$  «могут представлять собой a, t, g или c»
- ✓ Последовательность должна быть представлена как aatgnсссacgaatgncac
- ✓ «n» не требует аннотации, поскольку в отсутствие ключа характеристики «n» может представлять собой «a», «c», «g» или «t» (см. пункт 15 стандарта ST.26)
- ✓ Хотя это не обязательно, но рекомендуется, чтобы следующие три варианта были включены в перечень последовательностей как отдельные последовательности:

aatgaсссacgaatgbcac (b = t, g или c)

aatgtсссacgaatgvcac (v = a, g или c)

aatggсссacgaatghcac (h = t, a или c)

aatgcсссacgaatgdcac (d = t, g или a)

*Примечание: похожий случай подробно рассматривается в примере 95(b) в приложении VI к стандарту ВОИС ST.26.*

# Варианты последовательностей

## Пункт 95(с)

В заявке раскрывается следующее:

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

где Хаа — Gly или добавление последовательности

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

*В одном варианте последовательности содержатся добавления или замещения в более чем 1000 остатков*

# Варианты последовательностей

## Пункт 95(с)

- ✓ Вариант, в котором Хаа = добавление более 1000 остатков, должен быть включен в перечень последовательностей как отдельная последовательность с отдельным идентификационным номером последовательности.
- ✓ Вариант, в котором Хаа = Gly, также должен быть включен в перечень последовательностей как отдельная последовательность с отдельным идентификационным номером последовательности.

...-MGLPRGRICK-...

# Варианты последовательностей

**Вопрос:** Какой пункт применяется в отношении следующего раскрытия?

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



# Варианты последовательностей

**Вопрос:** Какой пункт применяется в отношении следующего раскрытия?

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

**Ответ:** пункт 94

*«Любой вариант последовательности, раскрытой как одна последовательность с указанием альтернативных вариантов остатков на одной или нескольких позициях, должен быть включен в перечень последовательностей и должен быть представлен в виде одной последовательности, в которой указанные альтернативные варианты остатков представлены наиболее ограничивающим неоднозначным символом».*

# Варианты последовательностей

**Вопрос:** Какой пункт применяется в отношении следующего раскрытия?

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							-		

Пустое место в таблице означает, что аминокислота в варианте такая же, что и соответствующая аминокислота в последовательности, а “-” означает удаление соответствующей аминокислоты в последовательности.

# Варианты последовательностей

**Вопрос:** Какой пункт применяется в отношении следующего раскрытия?

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							-		

Пустое место в таблице означает, что аминокислота в варианте такая же, что и соответствующая аминокислота в последовательности, а “-” означает удаление соответствующей аминокислоты в последовательности.

**Ответ:** пункт 93

*«Исходная последовательность и любые варианты этой последовательности, которые раскрыты путем перечисления их остатков и на которые распространяется пункт 7, должны быть включены в перечень последовательностей под отдельным идентификационным номером последовательности».*

# Варианты последовательностей

**Вопрос:** Какой пункт применяется в отношении следующего раскрытия?

На рисунке раскрыта следующая последовательность:

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

Спецификация содержит следующую информацию о последовательности 1:

... в некоторых воплощениях валин в позиции 7 последовательности 1  
заменен на аланин ...

# Варианты последовательностей

**Вопрос:** Какой пункт применяется в отношении следующего раскрытия?

На рисунке раскрыта следующая последовательность:

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

Спецификация содержит следующую информацию о последовательности 1:

... в некоторых воплощениях валин в позиции 7 последовательности 1  
заменен на аланин ...

**Ответ.** Пункт 95(а): *Любой вариант последовательности, раскрытой только посредством ссылки на удаление(я), добавление(я) или замещение(я) в исходной последовательности в перечне последовательностей, следует включать в перечень последовательностей. Такие варианты последовательностей, включенные в перечень последовательностей:*

*(а) могут быть представлены путем аннотирования исходной последовательности, если она содержит вариант(ы) на отдельной позиции или множестве разных позиций и появление этих вариантов происходит независимо;*»

# Варианты последовательностей

**Вопрос:** Какой пункт применяется в отношении следующего раскрытия?

На рисунке раскрыта следующая последовательность:

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

Спецификация содержит следующую информацию о последовательности 1:

... если валин в позиции 7 последовательности 1 заменяется на аланин, то аланин в позиции 10 заменяется на валин...

# Варианты последовательностей

**Вопрос:** Какой пункт применяется в отношении следующего раскрытия?

На рисунке раскрыта следующая последовательность:

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

Спецификация содержит следующую информацию о последовательности 1:

... если валин в позиции 7 последовательности 1 заменяется на аланин, то аланин в позиции 10 заменяется на валин...

**Ответ.** Пункт 95(b): *Любой вариант последовательности, раскрытой только посредством ссылки на удаление(я), добавление(я) или замещение(я) в исходной последовательности в перечне последовательностей, следует включать в перечень последовательностей. Такие варианты последовательностей, включенные в перечень последовательностей:*

*(b) должны быть представлены как отдельная последовательность с отдельным идентификационным номером последовательности, если они содержат варианты на множестве разных позиций и возникновение данных вариантов является взаимосвязанным;*

# Вопросы?