**СТАНДАРТ ST. 26**

# Рекомендуемый Стандарт представления перечней нуклеотидных и аминокислотных последовательностей с использованием языка XML (Расширяемого языка разметки)

*Версия 2.0*

Пересмотр одобрен Комитетом по Стандартам ВОИС (КСВ)
на своей тринадцатой сессии 14 ноября 2025

*Редакционное примечание подготовлено Международным бюро*

Тринадцатая сессия Комитета по Стандартам ВОИС приняла решение о том, что версия 2.0 Стандарта ВОИС ST.26 вступит в силу 1 июля 2027 г. и будет применяться ко всем заявкам, поданным на эту дату или после неё

Только параграф 8 данного Стандарта будет применяться со дня вступления в силу версии 2.0, независимо от даты подачи заявки на патент

СОДЕРЖАНИЕ

**ВВЕДЕНИЕ** ---------- **3**
**ОПРЕДЕЛЕНИЯ** ---------- **3**
**ОБЛАСТЬ ПРИМЕНЕНИЯ** ---------- **6**
**ОТСЫЛКИ** ---------- **6**
**ПРЕДСТАВЛЕНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ** ---------- **7**
**Нуклеотидные последовательности** **7**
**Аминокислотные последовательности** **11**
**Представление особых ситуаций** **14**
**СТРУКТУРА ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В XML** ---------- **14**
**Корневой элемент** **15**
**Часть общей информации** **16**
**Часть с данными последовательности** **21**
**Таблица характеристик** **23**
**Функциональные ключи** **23**
**Обязательные функциональные ключи** **24**
**Местоположение признака** **24**
**Квалификаторы признака** **27**
**Обязательные квалификаторы признаков** **28**
**Элементы квалификатора** **28**
**Свободный текст** **31**
**Кодирующие последовательности.** **33**
**Варианты** **34**

**Приложения**

Приложение 1 – Контролируемая лексика
Приложение 2 – Определение типа документа (DTD) для перечня последовательностей
Приложение 3 – Образец перечня последовательностей (XML файл)
Приложение 4 – Подмножество символов из Таблицы Unicode Basic Latin Code
Приложение 5 – Дополнительные требования для обмена данными (только для ВИС)
Приложение 6 – Руководство пользователя с иллюстрированными примерами
Дополнение – Руководство по последовательностям в XML
Приложение 7 – Рекомендации по преобразованию перечня последовательностей из ST.25 в ST.26: потенциально добавленные или удалённые признаки изобретения.

## СТАНДАРТ ST. 26

# Рекомендуемый Стандарт представления перечней нуклеотидных и аминокислотных последовательностей с использованием языка XML (Расширяемого языка разметки)

*Версия 2.0*

*Пересмотр одобрен Комитетом по Стандартам ВОИС (КСВ)
на своей тринадцатой сессии*

ВВЕДЕНИЕ

1. Данный Стандарт определяет требования к раскрытию в заявках на патент последовательностей нуклеотидов и аминокислот для включения в перечень последовательностей, представляет методы раскрытия и определение типа документа (Document Type Definition, далее - DTD) для перечней последовательностей в XML (eXtensible Markup Language (Расширяемого Языка Разметки)). Рекомендуется, чтобы ведомства по интеллектуальной собственности принимали любые перечни последовательностей, соответствующие данному Стандарту, поданные как часть заявки на патент или касающиеся заявки на патент.

2. Цель данного Стандарта:

a) позволить заявителю единообразно составлять перечень последовательностей в заявке на патент, который был бы приемлем как для международных, так и для национальных или региональных процедур;

b) повысить достоверность и качество представления последовательностей для упрощения их распространения среди заявителей, широкой публики и экспертов,

c) облегчить поиск данных о последовательностях; и

d) обеспечить обмен данными о последовательностях в электронной форме, а также ввод соответствующих данных в компьютеризованные базы данных.

ОПРЕДЕЛЕНИЯ

3. Для целей данного Стандарта выражение:

a) «аминокислота» означает органическую молекулу, содержащую аминогруппу и карбоксильную группу. Аминокислоты, которые должны быть включены в это определение, для целей данного Стандарта, включают аминокислоты, перечисленные в Приложении I, Раздел 3, таблица 3 (как D, так и L стереоизомеры), и их модификации, включая модификации, приводящие к образованию бета-аминокислоты, гамма аминокислоты или N-замещенные глицины (см. параграф 29), аминокислоты, содержащие модифицированные боковые цепочки, а также аминокислоты, перечисленные в Приложении I, Раздел 4, таблица 4, но не ограничиваясь ими. Аминокислоты будут рассматриваться как немодифицированные L-альфа аминокислоты, если они далее не описаны в таблице характеристик как модифицированные в соответствии с параграфом 30. Для целей настоящего Стандарта остаток пептидной нуклеиновой кислоты (PNA) не считается аминокислотой, но считается нуклеотидом, как указано далее в параграфе 3(g)(i)(2).

b) «контролируемая лексика» - это терминология, содержащаяся в Стандарте, которая должна быть использована для описания свойств последовательностей, т.е. аннотирования области или интересующей позиции, определяемых в Приложении 1.

c) «перечисление остатков» означает раскрытие последовательности в патентной заявке путем перечисления по порядку каждого остатка последовательности, в которой:

(i) остаток представляет собой имя, аббревиатуру, символ или структуру (например, HHHHHHQ или HisHisHisHisHisHisGln); или

(ii) множество остатков представлено сокращенной формулой (например, $His_6Gln$).

d) "намеренно пропущенная последовательность", также известная как пустая последовательность, обозначает метку-заполнитель (placeholder) сохранения нумерации последовательностей в перечне последовательностей для согласованности с раскрытием заявки, например, если последовательность удаляется из описания после подготовки перечня последовательностей.

e) «модифицированная аминокислота» означает любую аминокислоту, как описано в параграфе 3 (a), другую, чем следующие L-альфа аминокислоты: L-аланин, L-аргинин, L-аспарагин, L-аспарагиновая кислота, L-цистеин, L-глутамин, L-глутаминовая кислота, L-глицин, L-гистидин, L-изолейцин, L-лейцин, L-лизин, L-метионин, L-фенилаланин, L-пролин, L-пирролизин, L-серин, L-селеноцистеин, L-треонин, L-триптофан, L-тирозин или L-валин.

f) «модифицированный нуклеотид» означает любой нуклеотид, как описано в параграфе 3 (g), отличный от 5'-монофосфата дезоксиаденозина, 5'-монофосфата дезоксигуазидина, 5'-монофосфата дезоксицитидина, 5'-монофосфата дезокситимидина, 5'-монофосфата аденозина, 5'-монофосфата гуанозина, 5'- монофосфата цитидина или 5'-монофосфата уридина.

g) «нуклеотид» означает любой нуклеотид или аналог нуклеотида, который может быть представлен с помощью любого символа, определённого в Приложении 1 (см. раздел 1, Таблица 1), где нуклеотид или аналог нуклеотида содержат:

(i) основную часть, выбранную из:

(1) 2' дезоксирибоза 5' монофосфата (основной цепи дезоксирибонуклеотида) или рибоза 5' монофосфата (основной цепи рибонуклеотида); или

(2) 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 основных частей включают аминокислоты, такие как в пептидных нуклеиновых кислотах, молекулы гликоля, как в гликоль нуклеиновых кислотах, молекулы треофуранозилового сахара, как в треозо-нуклеиновых кислотах, морфолиновые кольца и фосфородиамидатные группы, как в морфолинах, и молекулы циклогексенила, как в циклогексенильных нуклеиновых кислотах.

и

(ii) основную часть, которая либо:

(1) присоединена к нуклеотидному основанию, включая модифицированное или синтетическое пуриновое или пиримидиновое нуклеотидное основание; или

(2) лишена пуриновой или пиримидиновой нуклеотидной группы, когда нуклеотид является частью нуклеотидной последовательности, называемой «AP-site» или «abasic site» (участок ДНК, лишенный азотистых оснований).

h) “остаток” означает любой отдельный нуклеотид или аминокислоту, или их соответствующие аналоги в последовательности.

i) «идентификационный номер последовательности» представляет собой уникальный номер (целое число), присваиваемый каждой последовательности в перечне последовательностей.

j) «перечень последовательностей» означает как часть описания поданной заявки на патент, так и документ, поданный позднее заявки, который включает раскрытие нуклеотидной и/или аминокислотной последовательности(ей), а также другую имеющуюся информацию, как предусмотрено настоящим Стандартом.

k) «специально определённые» означает любые нуклеотиды, кроме тех, которые представлены с помощью "n", и любые аминокислоты, кроме тех, которые представлены с помощью "X", из перечисленных в Приложении 1 (см. раздел 1, Таблица 1 и раздел 3, Таблица 3 соответственно).

l) «неизвестный» нуклеотид или аминокислота означает, что отдельный нуклеотид или аминокислота представлены, но их идентификационные данные неизвестны или нераскрыты.

m) «вариант последовательности» означает нуклеотидную или аминокислотную последовательность, которая содержит одно или несколько отличий относительно первичной последовательности. Эти отличия могут включать альтернативные остатки (см. Параграфы 15 и 27), модифицированные остатки (см. параграфы 3 (e), 3 (f), 3 (g), 3 (h), 16 и 29), удаления, вставки и замены. Смотрите параграфы с 93 по 95.

n) «свободный текст» - это тип формата для некоторых квалификаторов, представленный в форме фразы описательного текста или другой указанный формат (как указано в Приложении I). Смотрите параграф 85.

o) «свободный текст, зависящий от языка» означает значение свободного текста некоторых квалификаторов, которое может потребовать перевода для международных, национальных или региональных процедур. Смотрите параграф 87.

4. Для целей настоящего Стандарта слово (слова):

(a) «может» относится к необязательному или допустимому методу, но не к требованию.

(b) «должен» относится к требованию данного Стандарта; игнорирование этого требования приведет к несоблюдению Стандарта.

(c) «не должно» относится к запрету в Стандарте.

(d) «следует» относится к настоятельно рекомендованному подходу, но не к требованию.

(e) «не следует» относится к настоятельно не рекомендованному методу, но не к запрету.

ОБЛАСТЬ ПРИМЕНЕНИЯ

5. Настоящий Стандарт устанавливает требования к представлению перечня последовательностей нуклеотидов и аминокислот в заявках на патент.

6. Перечни последовательностей, соответствующие данному Стандарту (далее - перечни последовательностей), содержат часть общей информации и часть с данными последовательностей. Перечень последовательностей должен быть представлен как единый файл в формате Расширяемого Языка Разметки (Extensible Markup Language, далее - XML), использующего определение типа документа (Document Type Definition, далее - DTD), представленные в Приложении 2. Часть общей информации с содержащейся в ней библиографической информацией служит исключительно для объединения её с перечнем последовательностей в патентной заявке, для которой этот перечень был представлен. Часть с данными последовательностей состоит из одного или более элементов данных последовательности, каждый из которых содержит информацию об одной последовательности. Элементы данных последовательности включают различные свойства (характеристики) ключей и квалификаторов последовательностей, на основе спецификаций Совместной Международной Базы Данных Нуклеотидных Последовательностей (International Nucleotide Sequence Database Collaboration, далее - INSDC) и Консорциума UniProt.

7. Для целей данного Стандарта последовательность, для которой необходим перечень последовательностей, раскрывается путём перечисления её остатков и может быть представлена:

a) неразветвлённой последовательностью или линейным участком разветвлённой последовательности, содержащей десять и более специально определённых нуклеотидов, где смежные нуклеотиды соединены либо:

(i) от 3’ к 5’ (или от 5' к 3') фосфодиэфирной связью; либо

(ii) любой химической связью, которая приводит к такому расположению соседних нуклеотидных оснований, которые имитируют расположение нуклеотидных оснований в нуклеиновых кислотах природного происхождения; или

b) неразветвлённой последовательностью или линейным участком разветвлённой последовательности, содержащей четыре и более специально определённых аминокислот, где аминокислоты образуют одну пептидную основу, т.е. смежные аминокислоты соединены амидными связями, сформированными между карбоксильной группой одной аминокислоты и аминогруппой другой аминокислоты.

8. Перечень последовательностей может включать последовательности с менее чем десятью специально определёнными нуклеотидами или с менее чем четырьмя специально определёнными аминокислотами в качестве последовательности, которой дан её собственный идентификационный номер последовательности.

ОТСЫЛКИ

9. Следующие Стандарты и документы являются релевантными по отношению к данному Стандарту:

Совместная Международная База Данных Нуклеотидных Последовательностей (INSDC) http://www.insdc.org/;

| | |
|---|---|
| Международный Стандарт ISO 639-1:2002 | Коды для обозначения названий языков Часть 1: Коды Alpha-2; |
| Консорциум UniProt: | http://www.uniprot.org/ |
| W3C XML 1.0 | http://www.w3.org/ |
| Стандарт ВОИС ST.2 | Стандартное представление календарных дат на опубликованных документах с использованием григорианского календаря |
| Стандарт ВОИС ST.3 | Рекомендуемый Стандарт на двубуквенные коды для представления стран, административных единиц и межправительственных организаций; |
| Стандарт ВОИС ST.16 | Рекомендуемые стандартные коды для идентификации различных видов патентных документов |
| Стандарт ВОИС ST.25 | Стандарт по представлению перечней последовательностей нуклеотидов и аминокислот в патентных заявках |

ПРЕДСТАВЛЕНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

10. Каждой последовательности, охваченной параграфом 7, должен быть присвоен отдельный идентификационный номер, включая последовательность, которая идентична участку более длинной последовательности. Идентификационный номер последовательности должен начинаться с 1 и последовательно увеличиваться на целое число. При отсутствии последовательности представляется идентификационный номер последовательности, т.е. на месте намеренно пропущенной последовательности должен использоваться “000” (см. параграф 58). Общее количество последовательностей должно быть указано в перечне последовательностей и должно равняться общему количеству идентификационных номеров последовательностей, независимо от того, будет ли это номер присутствующей последовательности или код «000» намеренно пропущенной последовательности.

### *Нуклеотидные последовательности*

11. Нуклеотидная последовательность должна быть представлена единой цепочкой слева направо в направлении от 5' к 3' или в направлении слева направо, которое имитирует направление от 5' к 3'. Обозначения 5' и 3' или любые другие подобные обозначения не должны быть включены в последовательность. Нуклеотидная последовательность, состоящая их двух цепочек, раскрывается путём перечисления остатков обеих цепочек, которые должны быть представлены как:

a) единая последовательность или как две отдельных последовательности, каждой из которых присвоен свой собственный идентификационный номер последовательности, в тех случаях, когда две отдельные цепочки полностью комплементарны друг другу; или

b) две отдельные последовательности, каждой из которых присвоен свой собственный идентификационный номер последовательности, в тех случаях, когда две отдельные цепочки не полностью комплементарны друг другу.

12. Для целей настоящего Стандарта первый нуклеотид, представленный в последовательности, является остатком на позиции номер 1. Когда нуклеотидная последовательность имеет циклическую (кольцевую) конфигурацию, заявитель может выбрать нуклеотид с положением остатка номер 1. Нумерация непрерывна по всей последовательности в направлении от 5' к 3' или в направлении, которое имитирует направление от 5' до 3'. Последний номер позиции остатка должен быть равен числу нуклеотидов в последовательности.

13. Все нуклеотиды в последовательности должны быть представлены с использованием символов, определённых в Приложении 1 (см. Глава 1, Таблица 1). При этом должны использоваться только строчные буквы. Любой символ, используемый для представления нуклеотида, является эквивалентом только одного остатка.

14. Символ “t” будет истолкован как тимин в ДНК и урацил в РНК. Урацил в ДНК или тимин в РНК считаются модифицированным нуклеотидом и должны быть в дальнейшем описаны в таблице характеристик, как представлено в параграфе 19.

15. В случае, когда присвоен неоднозначный символ (представляющий два и более альтернативных нуклеотида), должен быть использован наиболее ограничивающий символ, как перечислено в Приложении 1 (Глава 1, Таблица 1). Например, если нуклеотидами данной позиции могут быть “a” или “g”, следует использовать скорее “r”, чем “n”. Символ “n” будет рассматриваться как любой из немодифицированных “a” или “c” или “g” или “t/u”, за исключением случаев, где он используется с дальнейшим описанием в таблице характеристик. Символ “n” не должен использоваться для представления чего-либо, кроме нуклеотида. Символ “n” может использоваться для представления только одного модифицированного или “неизвестного” нуклеотида, вместе с последующим описанием в таблице характеристик, как предусмотрено в параграфах 16,17, 21 или 93-96. Для представления вариантов последовательности, то есть альтернативы, удаления, вставки или замены, смотри параграфы 94-100.

16. Модифицированные нуклеотиды должны быть представлены в последовательности как соответствующие немодифицированные нуклеотиды, т.е. как “a”, “c”, “g” или “t” всякий раз, когда это возможно. Символ “n” должен использоваться для представления модифицированного нуклеотида только когда ни один из других символов в Приложении I (см. Раздел 1, Таблица 1) не подходит. Модифицированный нуклеотид с нуклеиновым основанием, идентичным нуклеиновому основанию немодифицированного нуклеотида, указанный в Приложении I, Раздел 1, Таблица 1, должен быть представлен в последовательности как соответствующий немодифицированный нуклеотид. Например, остаток пептидной нуклеиновой кислоты (PNA) с гуаниновым нуклеиновым основанием должен быть представлен в последовательности как “g”, а не как “n”. Символ “n” является эквивалентом только одного остатка.

17. Модифицированный нуклеотид должен быть в дальнейшем описан в таблице характеристик (см. параграф 60 и последующие) с использованием специального ключа “modified_base” и обязательного квалификатора “mod_base” в сочетании с единственной аббревиатурой из Приложения 1 (см. Глава 2, Таблица 2) в качестве значения квалификатора; если аббревиатурой является “OTHER” (“ДРУГОЙ”), то полное несокращённое название модифицированного нуклеотида должно быть приведено как значение квалификатора “note”. Для перечисления альтернативных модифицированных нуклеотидов значение квалификатора «OTHER» может использоваться в сочетании с дополнительным квалификатором “note” («примечания») (см. параграфы 95 и 96). Аббревиатуры (или полные названия), приведенные в Приложении 1 (см. Глава 2, Таблица 2) не должны быть использованы в самой последовательности.

18. Нуклеотидная последовательность, включающая один или несколько участков следующих друг за другом модифицированных нуклеотидов, которые имеют одну и ту же основную часть (см. параграф 3 (g) (i) (2)), должна быть дополнительно описана в таблице характеристик, как это предусмотрено в параграфе 17. Модифицированные нуклеотиды каждого такого участка могут быть совместно описаны в одном элементе INSDFeature, как это предусмотрено в параграфе 22. В качестве значения в квалификаторе “note” («примечания») должно быть указано наиболее ограничивающее несокращённое химическое название, которое охватывает все модифицированные нуклеотиды в пределах участка или перечень химических названий всех нуклеотидов в диапазоне. Например, последовательность нуклеиновой кислоты гликоля, содержащая нуклеотиды нуклеинов «a», «c», «g» или «t», может быть описана в квалификаторе “note” как «2,3-дигидроксипропилнуклеозиды». Альтернативно, такая же последовательность может быть описана в квалификаторе “note” как «2,3-дигидроксипропиладенин, 2,3-дигидроксипропилтимин, 2,3-дигидроксипропилгуанин или 2,3-дигидроксипропилцитозин». Если отдельный модифицированный нуклеотид на участке включает дополнительную модификацию, то модифицированный нуклеотид также должен быть дополнительно описан в таблице характеристик, как это предусмотрено в параграфе 17.

19. Урацил в ДНК или тимин в РНК считаются модифицированными нуклеотидами и должны быть представлены в последовательности как “t” и далее описаны в таблице функций с помощью специального ключа “modified_base”, квалификатора "mod_base" с "OTHER" а качестве значения квалификатора, и квалификатора “note” с "урацилом" или "тимином" соответственно, как значение квалификатора.

20. Следующие примеры иллюстрируют модифицированные нуклеотиды, представленные в соответствии с параграфами с 16 по 18 выше:

Пример 1: Модифицированный нуклеотид, использующий аббревиатуру из Приложения 1 (см. Глава 2, Таблица 2)

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

Пример 2: Модифицированный нуклеотид, использующий “OTHER” из Приложения 1 (см. Глава 2, Таблица 2)

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

Пример 3. Нуклеотидная последовательность, состоящая из модифицированных нуклеотидов, охватываемых параграфом 3 (g) (i) (2), с двумя отдельными нуклеотидами, которые включают в себя дополнительную модификацию

```
<INSDFeature>
 <INSDFeature_key>modified_base</INSDFeature_key>
 <INSDFeature_location>1..954</INSDFeature_location>
 <INSDFeature_quals>
       <INSDQualifier>
             <INSDQualifier_name>mod_base</INSDQualifier_name>
             <INSDQualifier_value>OTHER</INSDQualifier_value>
       </INSDQualifier>
       <INSDQualifier>
             <INSDQualifier_name>note</INSDQualifier_name>
             <INSDQualifier_value>2,3-dihydroxypropyl
             nucleosides</INSDQualifier_value>
       </INSDQualifier>
 </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
      <INSDFeature_key>modified_base</INSDFeature_key>
      <INSDFeature_location>439</INSDFeature_location>
      <INSDFeature_quals>
            <INSDQualifier>
                  <INSDQualifier_name>mod_base</INSDQualifier_name>
                  <INSDQualifier_value>i</INSDQualifier_value>
            </INSDQualifier>
      </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
      <INSDFeature_key>modified_base</INSDFeature_key>
      <INSDFeature_location>684</INSDFeature_location>
      <INSDFeature_quals>
            <INSDQualifier>
                  <INSDQualifier_name>mod_base</INSDQualifier_name>
                  <INSDQualifier_value>OTHER</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier>
                  <INSDQualifier_name>note</INSDQualifier_name>
                  <INSDQualifier_value>xanthine</INSDQualifier_value>
            </INSDQualifier>
      </INSDFeature_quals>
</INSDFeature>
```

21. Любой “unknown” («неизвестный») нуклеотид должен быть представлен в последовательности как “n”. “Unknown” («неизвестный») нуклеотид должен быть в дальнейшем описан в таблице характеристик (см. параграф 60 и далее) с помощью функционального ключа “unsure” (“неопределённый”). Символ “n” является эквивалентом только одного остатка.

22. Участок, содержащий известное число непрерывных нуклеотидных остатков, для которых применяют одинаковые описания, может быть описан с использованием одного элемента `INSDFeature` с синтаксисом “x..y” в качестве дескриптора местонахождения в элементе `INSDFeature_location` (см. параграфы с 64 по 71). Для представления вариантов последовательности, т.е. удаления, альтернатив, добавления или замещения, см. параграфы 93 - 100.

23. Следующий пример иллюстрирует представление области модифицированных нуклеотидов, для которых используют одинаковые описания, в соответствии с разделом 22 выше:

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>358..485</INSDFeature_location>
  <INSDFeature_quals>
       <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
       </INSDQualifier>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>isoguanine</INSDQualifier_value>
       </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

*Аминокислотные последовательности*

24. Аминокислоты в аминокислотной последовательности должны быть представлены в амино – карбокси направлении слева направо. Амино и карбокси группы не должны быть представлены в последовательности.

25. Для целей настоящего Стандарта, первая аминокислота в последовательности является остатком на позиции номер 1, включая аминокислоты, предшествующие зрелому белку, например, пред-последовательности, про-последовательности, пред-про-последовательности и сигнальные последовательности. Когда аминокислотная последовательность имеет циклическую (кольцевую) конфигурацию, и кольцо состоит исключительно из аминокислотных остатков, связанных амидными связями, то есть последовательность не имеет амино- и карбокси- концов, заявитель должен выбрать аминокислоту с положением остатка номер 1. Нумерация непрерывна по всей последовательности в амино – карбокси направлении.

26. Все аминокислоты в последовательности должны быть представлены с использованием символов, определённых в Приложении 1 (см. Главу 3, Таблицу 3). Должны использоваться только заглавные буквы. Любой символ, используемый для представления аминокислоты, является эквивалентом только одного остатка.

27. В случае, если уместным является неоднозначный символ (представляющий две и более альтернативных аминокислоты), следует использовать наиболее ограничивающий символ как определено в Приложении 1 (см. Глава 3, Таблица 3). Например, если аминокислота на данной позиции может быть аспарагиновой кислотой или аспарагином, предпочтительнее использовать символ “B”, нежели символ “X”. Символ “X” будет истолкован как одна из немодифицированных L-альфа аминокислот “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, или “V”, за исключением случаев использования последующего описания в таблице характеристик. Символ “X” не должен использоваться для представления чего-либо иного, кроме аминокислоты. Одиночная модифицированная или «неизвестная» аминокислота может быть представлена символом “X”, когда не подходит ни один другой символ из Приложения I (см. Раздел 3, Таблица 3), вместе с последующим описанием в таблице характеристик, например, как предусмотрено в параграфах 29, 30, 32 или 93 – 98. Для представления вариантов последовательности, т.е. альтернатив, удаления, добавления или замещения, смотри параграфы 93 - 100.

28. Раскрытые аминокислотные последовательности, разделённые символами внутреннего разграничения, например, ”Ter” («конец») или звездочкой “*” или точкой “.“ или пробелом, должны быть включены как отдельные последовательности для каждой аминокислотной последовательности, которая содержит, по крайней мере, четыре специально определённые аминокислоты, и упомянуты в параграфе 7. Каждой такой отдельной последовательности должен быть присвоен собственный идентификационный номер последовательности. Разделительные символы и пробелы не должны быть включены в последовательности перечня последовательностей (см. параграф 57).

29. Когда это возможно, модифицированные аминокислоты должны быть представлены в последовательности символом соответствующих немодифицированных аминокислот. Символ “X” должен использоваться для представления модифицированной аминокислоты только когда не походит ни один из других символов в Приложении I (см. Глава 3, Таблица 3). Модифицированная аминокислота, имеющая одиночную содержащую карбон боковую цепочку, которая идентична боковой цепочке в L-альфа аминокислоте, указанной в Приложении I, Глава 3, Таблица 3, и которая расположена на:

(а) аминоазоте остатка вместо альфа карбона (например, N-замещенный глицин), **или**

(б) любом из основных атомов аминокислоты, имеющей расширенную карбоновую основу, такую как бета аминокислоты или гамма аминокислоты, **или**

(с) альфа карбоне, но как D-стереоизомере,

должна быть представлена в последовательности символом соответствующей немодифицированной аминокислоты, т.е. аминокислоты из Приложения I, Главы 3, Таблицы 3, имеющей идентичную боковую цепочку. Модифицированную аминокислоту, содержащую множество карбон-содержащих боковых цепочек, следует представлять с помощью соответствующей немодифицированной аминокислоты или как «другую» аминокислоту (например, “X”), что наиболее подходяще. Символ “X” является эквивалентом только одного остатка.

30.Модифицированная аминокислота должна быть далее описана в таблице характеристик (см. параграф 60 и последующие). Там, где это применимо, следует использовать функциональные ключи «CARBOHYD» или «LIPID» вместе с обязательным квалификатором “note” («примечание»). Функциональный ключ “MOD_RES” следует использовать для других посттрансляционно модифицированных аминокислот вместе с обязательным квалификатором “note”. Модифицированную аминокислоту, для которой “CARBOHYD”, “LIPID” или “MOD_RES” не применимы, т.е. не посттрансляционно модифицированную аминокислоту, следует описывать с помощью функционального ключа “SITE” вместе с обязательным квалификатором “note”. Значение квалификатора “note” должно быть или аббревиатурой, определённой в Приложении 1 (см. Глава 4, Таблица 4), или полным, несокращённым названием модифицированной аминокислоты. Аббревиатуры, определённые в Таблице 4, о которых речь идёт выше, или полные, несокращенные названия не должны использоваться в самой последовательности.

31. Следующие примеры иллюстрируют представление модифицированных аминокислот в соответствии с параграфом 30 выше:

Пример 1: Посттрансляционно модифицированная аминокислота

```
<INSDFeature>
  <INSDFeature_key>MOD_RES</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
```

```
      <INSDQualifier_value>3Hyp</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

Пример 2: Непосттрансляционно модифицированная аминокислота

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name> note </INSDQualifier_name>
      <INSDQualifier_value>Orn</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

Пример 3: D-аминокислота

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>9</INSDFeature_location>
    <INSDFeature_quals>
       <INSDQualifier>
             <INSDQualifier_name> note </INSDQualifier_name>
             <INSDQualifier_value>D-Arginine</INSDQualifier_value>
       </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

32. Любая "unknown" («неизвестная») аминокислота, должна быть представлена как символ “X” в последовательности. "Unknown" аминокислота, указанная как “X”, должна быть в дальнейшем описана в таблице характеристик (см. параграф 60 и последующие) с помощью функционального ключа “UNSURE” и возможно (по выбору) квалификатора “note”. Символ “X” является эквивалентом только одного остатка.

33. Следующие примеры иллюстрируют представление "unknown" или “other” аминокислоты в соответствии с параграфом 32 выше:

```
<INSDFeature>
  <INSDFeature_key>UNSURE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name> note </INSDQualifier_name>
      <INSDQualifier_value>A or V</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
```

34. Участки, содержащие известное число смежных аминокислотных остатков, для которых используют одинаковые описания, могут быть совместно описаны с помощью синтаксиса “x..y”, как описание местонахождения в элементе `INSDFeature_location` (см. параграфы 64-71). Для представления вариантов последовательности, т.е. удаления, альтернатив, добавления или замещения, см. параграфы 93 - 100.

*Представление особых ситуаций*

35. Последовательность, раскрытая путём перечисления её остатков, которые построены как единая непрерывная последовательность из одного или более несмежных сегментов более длинной последовательности или сегментов из различных последовательностей, должна быть включена в перечень последовательностей с присвоенным ей собственным идентификационным номером последовательности.

36. Последовательность, которая содержит участки специально определённых остатков, отделённых одной или более областями с конфигурациями остатков “n” или “X” (см. параграфы 15 и 27, соответственно), где раскрыто точное число “n” или “X” остатков на каждом участке, должна быть включена в перечень последовательностей как одна последовательность с присвоенным ей собственным идентификационным номером последовательности.

37. Последовательность, которая содержит участки специально определённых остатков, отделённых одним или более интервалами неизвестных или нераскрытых остатков, не должна быть представлена в перечне последовательностей как отдельная последовательность. Каждый участок специально определённых остатков, который охвачен параграфом 7, должен быть включен в перечень последовательностей в виде отдельной последовательности с присвоенным ей собственным идентификационным номером последовательности.

СТРУКТУРА ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В XML

38. На основании параграфа 6 выше, файл перечня последовательностей в XML в соответствии с этим Стандартом состоит из:

(a) части общей информации, которая содержит информацию о патентной заявке, к которой относится перечень последовательностей; и

(b) части с данными последовательности, которая содержит один или более элементов данных последовательности, каждый из которых, в свою очередь, содержит информацию об одной последовательности.

Пример перечня последовательностей представлен в Приложении 3.

39. Перечень последовательностей должен быть представлен в XML 1.0 с использованием DTD, представленных в Приложении 2 «Определение типа документа (DTD) для перечня последовательностей».

(a) Первая строка XML файла должна содержать XML указание:

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

(b) Вторая строка XML файла должна содержать указание типа документа (DOCTYPE):

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

40. Весь электронный перечень последовательностей должен содержаться в одном файле.

Файл должен быть закодирован с использованием юникода UTF-8 со следующими ограничениями:

a) информация, содержащаяся в элементах `ApplicantName`, `InventorName` и `InventionTitle` части общей информации, и `NonEnglishQualifier_value` части данных последовательности, может состоять из любых допустимых символов Unicode, указанных в спецификации XML 1.0, за исключением контрольных кодовых точек Unicode 0000-001F и 007F-009F. Зарезервированные символы «, &,», < и > (кодовые точки Unicode 0022, 0026, 0027, 003C и 003E соответственно), подлежат замене, как указано в параграфе 41; **и**

b) информация, содержащаяся во всех других элементах и атрибутах части общей информации и во всех других элементах и атрибутах части данных последовательности, должна состоять из печатных символов (включая пробел) из таблицы кодов Unicode Basic Latin (т. е. ограничено кодовыми точками Unicode 0020, через 007E - см. Приложение IV). Зарезервированные символы «, &,‘, < и > (кодовые точки Unicode 0022, 0026, 0027, 003C и 003E соответственно) должны быть заменены, как указано в параграфе 41.

41. В случае перечня последовательностей в XML не должны использоваться ссылки на числовые символы[1], а перечисленные зарезервированные символы должны быть заменены соответствующими, заранее определёнными элементами, когда они используются в значении признака или содержании элемента:

| Зарезервированные символы | Заранее определённые элементы |
|---|---|
| < | < |
| > | > |
| & | & |
| " | " |
| ' | ' |

Смотри параграф 71 в качестве примера. Единственно допустимыми ссылками на символы являются заранее определённые элементы, указанные в данном параграфе.

42. Все обязательные элементы должны быть заполнены (за исключением предусмотренного в разделе 58 для намеренно пропущенной последовательности). Необязательные элементы, для которых недоступно содержание, не должны появляться в случае XML (за исключением случаев, предусмотренных в параграфе 97, для представления удаления в последовательности в качестве значения квалификатора “replace” («замена»)).

### *Корневой элемент*

43. В соответствии с данным Стандартом корневым элементом в XML является элемент ST26SequenceListing, имеющий следующие признаки:

| Признак | Описание | Обязательно/Необязательно |
|---|---|---|
| dtdVersion | Версия DTD, используемая для создания этого файла в формате “V#_#”, например. “V1_3”. | Обязательно |

[1] Ссылка на числовой символ относится к символу его кодовой точки универсального набора символов / Unicode и использует формат:
«& # nnnn;» или «& # xhhhh;», где «nnnn» - это кодовая точка в десятичной форме, а «hhhh» - это кодовая точка в шестнадцатеричной форме.

| Признак | Описание | Обязательно/Необязательно |
|---|---|---|
| fileName | Файл с названием перечня последовательностей. | Необязательно |
| softwareName | Название программного обеспечения, создавшего этот файл. | Необязательно |
| softwareVersion | Версия программного обеспечения, создавшего этот файл. | Необязательно |
| productionDate | Дата создания файла перечня последовательностей (format “CCYY-MM-DD”). | Необязательно |
| originalFreeTextLanguageCode | Код языка (см. ссылку в параграфе 9 на ISO 639-1: 2002) оригинала, на котором были подготовлены квалификаторы свободного текста, зависящего от языка. | Необязательно |
| nonEnglishFreeTextLanguageCode | Код языка (см. ссылку в параграфе 9 на ISO 639-1: 2002) для элементов NonEnglishQualifier_value | Обязательно, когда элемент NonEnglishQualifier_value присутствует в перечне последовательностей |

44. Следующий пример иллюстрирует корневой элемент ST26SequenceListing и его признаки в XML, согласно параграфу 43 выше:

```
<ST26SequenceListing dtdVersion="V1_3" fileName="US11-405455-SEQL.xml"
softwareName=" WIPO Sequence" softwareVersion="1.0"
productionDate="2022-05-10" originalFreeTextLanguageCode="de"
nonEnglishFreeTextLanguageCode="fr" >
  {...}*
</ST26SequenceListing>

*{...} представляет часть общей информации и часть данных
последовательности, которые не включены в этот пример.
```

*Часть общей информации*

45. Элементы части общей информации относятся к информации о заявке на патент и содержат следующие понятия:

| Элемент | Описание | Обязательно/Необязательно |
|---|---|---|
| `ApplicationIdentification`<br><br>`ApplicationIdentification состоит из:` | Идентификация заявки, для которой представлен перечень последовательностей | Обязательно, если перечень последовательностей предоставлен в любое время после присвоения номера заявки |
| `IPOfficeCode` | Код ведомства, куда подают заявку в соответствии со Стандартом ВОИС ST.3 | Обязательно |
| `ApplicationNumberText` | Номер заявки, как установлено ведомством, в которое подают заявку (например, PCT/IB2013/99999) | Обязательно |
| `FilingDate` | Дата подачи заявки, для которой представлен перечень последовательностей (WIPO ST.2 формат "CCYY-MM- DD", используется 4-значное обозначение календарного года, 2-значное календарного месяца и 2-значное дня в течение календарного месяца, например, 2015-01-31) | Обязательно, если перечень последовательностей предоставлен в любое время после установления даты подачи |
| `ApplicantFileReference` | Единый уникальный идентификатор, присвоенный заявителю для идентификации конкретной заявки, введённый в символах, определённых в параграфе 40 (b) | Обязательно, если перечень последовательностей предоставлен в любое время до присвоения номера заявки; в противном случае необязательно |
| `EarliestPriorityApplicationIdentification` | Идентификация наиболее ранней приоритетной заявки (также содержащей `IPOfficeCode`, `ApplicationNumberText` и `FilingDate`, смотри `ApplicationIdentification` выше) | Обязательно, если испрашивается приоритет |
| `ApplicantName` | Имя первого указанного заявителя, введённое в символах, определённых в параграфе 40 (a). Данный элемент включает обязательный атрибут languageCode как определено в параграфе 47. | Обязательно |

| Признак | Описание | Обязательно/Необязательно |
|---|---|---|
| | | |
| ApplicantNameLatin | Если ApplicantName введено в символах иных, чем это определено в параграфе 40 b), перевод или транслитерация имени первого указанного заявителя должны быть также введены в символах, определённых в параграфе 40 b) | Обязательно, если ApplicantName содержит нелатинские символы |
| InventorName | Имя первого указанного изобретателя, введённое в символах, определённых в параграфе 40 (a) | Необязательно |
| InventorNameLatin | Если InventorName введено в символах иных, чем это определено в параграфе 40 b), перевод или транслитерация имени первого указанного изобретателя могут быть также введены в символах, определённых в параграфе 40 b) | Необязательно |
| InventionTitle | Название изобретения, введённое в символах, определённых в параграфе 40 (a) на языке подачи. Перевод названия изобретения на дополнительные языки может быть введен в символах, определённых в параграфе 40 (a) используя дополнительный элемент InventionTitle. Данный элемент включает обязательный признак languageCode, как указано в параграфе 48. В названии изобретения должно быть от двух до семи слов. | Обязательно на языке подачи. Необязательно для дополнительных языков. |
| SequenceTotalQuantity | Общее количество всех последовательностей в перечне, в том числе намеренно пропущенные последовательности (также известные как пустые последовательности) (см. параграф 10). | Обязательно |

46. Следующие примеры иллюстрируют представление части общей информации о перечне последовательностей, согласно разделу 45 выше:

Пример 1: перечень последовательностей, представленный до идентификации заявки и даты подачи

```
<?xml version=”1.0” encoding=“UTF-8”?>
<!DOCTYPE ST26SequenceListing PUBLIC “-//WIPO//DTD Sequence Listing
1.3//EN” “ST26SequenceListing_V1_3.dtd”>
<ST26SequenceListing dtdVersion=“V1_3” fileName=“Invention_SEQL.xml”
softwareName= "WIPO Sequence" softwareVersion=“1.0” productionDate=“2022-05-
10” originalFreeTextLanguageCode=“en” nonEnglishFreeTextLanguageCode=“ja” >
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2013/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode=“en”>GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode=“en”>Keiko Nakamura</InventorName>
  <InventionTitle languageCode=“en”>SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
  <SequenceTotalQuantity>9</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber=“1”> {...}* </SequenceData>
  <SequenceData sequenceIDNumber=“2”> {...} </SequenceData>
  <SequenceData sequenceIDNumber=“3”> {...} </SequenceData>
  <SequenceData sequenceIDNumber=“4”> {...} </SequenceData>
  <SequenceData sequenceIDNumber=“5”> {...} </SequenceData>
  <SequenceData sequenceIDNumber=“6”> {...} </SequenceData>
  <SequenceData sequenceIDNumber=“7”> {...} </SequenceData>
  <SequenceData sequenceIDNumber=“8”> {...} </SequenceData>
  <SequenceData sequenceIDNumber=“9”> {...} </SequenceData>
</ST26SequenceListing>

*{...} представляет релевантную информацию для каждой последовательности,
которая не включена в данный пример.
```

Пример 2: перечень последовательностей, представленный после идентификации заявки и даты подачи

```
<?xml version=“1.0” encoding=“UTF-8”?>
<!DOCTYPE ST26SequenceListing PUBLIC “-//WIPO//DTD Sequence Listing
1.3//EN” “ST26SequenceListing_V1_3.dtd”>
<ST26SequenceListing dtdVersion=“1_3” fileName=“Invention_SEQL.xml”
softwareName ”="WIPO Sequence" softwareVersion=“1.0” productionDate=“2022-
05-10” originalFreeTextLanguageCode=“en”onEnglishFreeTextLanguageCode=“ja”>
   <ApplicationIdentification>
    <IPOfficeCode>US</IPOfficeCode>
    <ApplicationNumberText>14/999,999</ApplicationNumberText>
      <FilingDate>2015-01-05</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
```

```
    <ApplicationNumberText>PCT/IB2014/099999</ApplicationNumberText>
      <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode="en">Keiko Nakamura</InventorName>
  <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
  <SequenceTotalQuantity>9</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1"> {...}* </SequenceData>
  <SequenceData sequenceIDNumber="2"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="3"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="4"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="5"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="6"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="7"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="8"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>*

*{...} представляет релевантную информацию для каждой последовательности,
которая не включена в данный пример.
```

47. Название заявителя и, необязательно, имя изобретателя должны быть указаны в элементах `ApplicantName` и `InventorName`, соответственно, так как они в основном приводятся на языке, на котором подана заявка. Для каждого элемента должен быть указан соответствующий языковой код (см. отсылку в параграфе 9 к ISO 639-1:2002) в параметре `languageCode`. Когда указание названия заявителя содержит символы иные, чем это определено в параграфе 40 (b), транслитерация или перевод названия заявителя должны быть также указаны в символах латинского алфавита в элементе `ApplicantNameLatin`. Когда указание имени изобретателя содержит символы, отличные от символов латинского алфавита, транслитерация или перевод имени изобретателя могут быть также указаны в символах латинского алфавита в элементе `InventorNameLatin`.

48. Название изобретения должно быть указано в элементе `InventionTitle` на языке подачи и может быть также указано на других языках, используя для этого элементы `InventionTitle` (см. таблицу в параграфе 45). Соответствующий языковой код (см. отсылку в параграфе 9 к ISO 639-1:2002) должен быть указан в параметре `languageCode` элемента.

49. Следующий пример иллюстрирует представление имени и названия изобретения согласно параграфам 47 и 48 выше:

Пример: Название заявителя и имя изобретателя каждое представлено в буквах японского и латинского алфавита, и название изобретения, представленное в буквах японского, английского и французского алфавита

```
<ApplicantName languageCode="ja">出願製薬株式会社</ApplicantName>
<ApplicantNameLatin>Shutsugan Pharmaceuticals Kabushiki
Kaisha</ApplicantNameLatin>
<InventorName languageCode ="ja">特許太郎</InventorName>
<InventorNameLatin>Taro Tokkyo</InventorNameLatin>
<InventionTitle languageCode="ja"> efg タンパク質をコードするマウス abcd-1 遺伝子
</InventionTitle>
<InventionTitle languageCode="en"> Mus musculus abcd-1 gene for efg protein
</InventionTitle>
```

```
<InventionTitle languageCode="fr"> Gène abcd-1 de Mus musculus pour
protéine efg </InventionTitle>
```

*Часть с данными последовательности*

50. Часть с данными последовательности должна состоять из одного или более элементов `SequenceData`, каждый элемент содержит информацию об одной последовательности.

51. Каждый элемент `SequenceData` должен иметь обязательный признак `sequenceIDNumber`, в котором содержится идентификационный номер (см. параграф 10) для каждой последовательности. Например:

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

52. Элемент SequenceData содержит зависимый элемент INSDSeq, который в свою очередь содержит следующие зависимые элементы:

| Элемент | Описание | Обязательно/Не включено | |
|---|---|---|---|
| | | Последовательности | Намеренно пропущенные последовательности |
| `INSDSeq_length` | Длина последовательности | Обязательно | Обязательно без значения |
| `INSDSeq_moltype` | Тип молекулы | Обязательно | Обязательно без значения |
| `INSDSeq_division` | Указание, что последовательность относится к заявке на патент | Обязательно со значением "PAT" | Обязательно без значения |
| `INSDSeq_feature-table` | Список аннотаций последовательностей | Обязательно | Не должны быть включены |
| `INSDSeq_sequence` | Последовательность | Обязательно | Обязательно со значением “000” |

53. Элемент INSDSeq_length должен показывать число нуклеотидов или аминокислот последовательности, содержащейся в элементе INSDSeq_sequence. Например:

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

54. Элемент `INSDSeq_moltype` должен показывать тип молекулы, которая была представлена. Для нуклеотидных последовательностей, включая последовательности аналогов нуклеотидов, тип молекулы должен быть указан как ДНК или РНК. Для аминокислотных последовательностей тип молекулы должен быть указан как AA. (Данный элемент отличен от квалификатора “mol_type”, рассмотренных в параграфах 55 и 84). Например:

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

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 функционального ключа “source” и обязательного квалификатора “organism” со значением “synthetic construct” и обязательного квалификатора “mol_type” со значением “other ДНК“. Каждый ДНК и РНК фрагмент комбинированной ДНК/РНК молекулы должен быть в дальнейшем описан с функциональным ключом “misc_feature” и квалификатором “note”, который указывает, является ли фрагмент ДНК или РНК.

56. Следующий пример иллюстрирует описание нуклеотидной последовательности, содержащей фрагменты и ДНК, и РНК, согласно параграфу 55 выше.

```
<INSDSeq>
  <INSDSeq_length>120</INSDSeq_length>
  <INSDSeq_moltype>DNA</INSDSeq_moltype>
  <INSDSeq_division>PAT</INSDSeq_division>
  <INSDSeq_feature-table>
    <INSDFeature>
      <INSDFeature_key>source</INSDFeature_key>
      <INSDFeature_location>1..120</INSDFeature_location>
      <INSDFeature_quals>
            <INSDQualifier>
                  <INSDQualifier_name>organism</INSDQualifier_name>
                  <INSDQualifier_value>synthetic construct</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier>
                  <INSDQualifier_name>mol_type</INSDQualifier_name>
                  <INSDQualifier_value>other DNA</INSDQualifier_value>
            </INSDQualifier>
      </INSDFeature_quals>
    </INSDFeature>
    <INSDFeature>
      <INSDFeature_key>misc_feature</INSDFeature_key>
      <INSDFeature_location>1..60</INSDFeature_location>
      <INSDFeature_quals>
        <INSDQualifier>
          <INSDQualifier_name>note</INSDQualifier_name>
          <INSDQualifier_value>DNA fragment</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_quals>
    </INSDFeature>
    <INSDFeature>
      <INSDFeature_key>misc_feature</INSDFeature_key>
      <INSDFeature_location>61..120</INSDFeature_location>
      <INSDFeature_quals>
        <INSDQualifier>
          <INSDQualifier_name>note</INSDQualifier_name>
          <INSDQualifier_value>RNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_quals>
    </INSDFeature>
  </INSDSeq_feature-table>
  <INSDSeq_sequence>
cgacccacgcgtccgaggaaccaaccatcacgtttgaggacttcgtgaaggaattggataatacccgtccctaccaaaatgg
cgagcgccgactcattgctcctcgtaccgtcgagcggc
  </INSDSeq_sequence>
</INSDSeq>
```

57. Элемент `INSDSeq_sequence` должен раскрывать последовательность. В последовательность должны быть включены только соответствующие символы, указанные в Приложении 1 (см. Глава 1, Таблица 1 и Глава 3, Таблица 3). Последовательность не должна включать цифры, знаки препинания или пробелы.

58. Намеренно пропущенные последовательности должны быть включены в перечень последовательностей и представлены следующим образом:

a) элемент `SequenceData` и его признак `sequenceIDNumber` с идентификационным номером последовательности для пропущенной последовательности в качестве его значения;

b) элементы `INSDSeq_length`, `INSDSeq_moltype`, `INSDSeq_division` представлены, но без значения;

c) элемент INSDSeq_feature-table не должен быть включен, **и**

d) элемент `INSDSeq_sequence` с цепочкой "000" в качестве значения.

59. Следующий пример иллюстрирует представление намеренно пропущенной последовательности, согласно параграфу 58 выше:

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

*Таблица характеристик*

60. Таблица характеристик содержит информацию о месте и роли различных областей определённой последовательности. Таблица характеристик необходима для каждой последовательности, за исключением любой намеренно пропущенной последовательности, в этом случае она не приводится. Таблица характеристик содержится в элементе `INSDSeq_feature-table`, который включает один или более элементов `INSDFeature`.

61. Каждый элемент `INSDFeature` описывает один признак и содержит следующие зависимые элементы:

| Элемент | Описание | Обязательно/Необязательно |
|---|---|---|
| `INSDFeature_key` | Слово или аббревиатура, указывающие признак | Обязательно |
| `INSDFeature_location` | Участок последовательности, соответствующий признаку | Обязательно |
| `INSDFeature_quals` | Квалификатор, содержащий вспомогательную информацию о признаке | Обязательно, где функциональный ключ требует одного или более квалификаторов, например источник; в противном случае - необязательно |

*Функциональные ключи*

62. Приложение 1 содержит единый перечень функциональных ключей, которые должны использоваться в соответствии с данным Стандартом, совместно с единым перечнем соответствующих квалификаторов и указанием, относится ли данный квалификатор к обязательным или необязательным. Единый перечень функциональных ключей для нуклеотидных последовательностей определён в Главе 5 Приложения 1, а для аминокислотных последовательностей – в Главе 7 Приложения 1.

*Обязательные функциональные ключи*

63. Функциональный ключ “source” является обязательным для всех нуклеотидных последовательностей и для всех аминокислотных последовательностей, за исключением любых намеренно пропущенных последовательностей. Каждая последовательность должна иметь единственный функциональный ключ “source”, охватывающий всю последовательность. Когда последовательность образуется из нескольких источников, данные источники могут быть в дальнейшем описаны в таблице характеристик с помощью функционального ключа “misc_feature” и квалификатора “note” для нуклеотидных последовательностей и функционального ключа “REGION” и квалификатора “note” для аминокислотных последовательностей.

*Местоположение признака*

64. Обязательный элемент `INSDFeature_location` должен содержать, по крайней мере, одно описание местоположения, которое определяет место или участок, соответствующий признаку последовательности в элементе `INSDSeq_sequence`. Аминокислотные последовательности должны содержать один и только один дескриптор местоположения в обязательном элементе INSDFeature_location. Нуклеотидные последовательности могут иметь более одного дескриптора местоположения в обязательном элементе INSDFeature_location, когда используются вместе с одним или более оператор(ом) местоположения (см. параграфы 67-70).

65. Описанием местоположения может быть номер единичного остатка, участок, ограниченный диапазоном смежных остатков, или место, или участок, который оказывается за пределами определённого остатка или диапазона остатков. Дескриптор местоположения не должен включать нумерацию остатков за пределами диапазона последовательности в элементе INSDSeq_sequence. Только для нуклеотидных последовательностей дескриптор местоположения может быть местом между двумя соседними номерами остатков. Множество дескрипторов местоположения должны использоваться совместно с оператором местоположения, когда признак соответствует прерывающимся местам или участкам нуклеотидной последовательности (смотри параграфы 67-70).

66. Синтаксис для каждого типа дескриптора местоположения указан в таблице ниже, где x и y являются номерами остатков, указанных как положительные целые числа, не больше, чем длина последовательности в элементе `INSDSeq_sequence`, а x меньше, чем y.

(a) Дескрипторы местоположения для нуклеотидных и аминокислотных последовательностей:

| Тип дескриптора местоположения | Синтаксис | Описание |
|---|---|---|
| Номер единичного остатка | `x` | Указывает на единичный остаток в последовательности |
| Номера остатков, ограничивающие интервал последовательности | `x..y` | Указывает на ограниченный непрерывный диапазон остатков, включая стартовый и конечный остатки. |
| Остатки перед первым или после последнего определённого номера остатка | `<x`<br>`>x`<br>`<x..y`<br>`x..>y`<br>`<x..>y` | Указывает на область, включающую определённый остаток или диапазон остатков и определённый остаток, выходящий за рамки диапазона. Символы '<' и '>' могут быть использованы с единичным остатком или для указания возможного выхода из диапазона номера определённого остатка. |

(b) Дескрипторы местоположения только для нуклеотидных последовательностей:

| Тип дескриптора местоположения | Синтаксис | Описание |
|---|---|---|
| Место между номерами двух смежных остатков | x^y | Указывает на место между двумя смежными остатками, например место эндонуклеолитического расщепления. Номера позиций для смежных остатков разделяются 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 с длиной 1000. |

(c) Дескрипторы местоположения только для аминокислотных последовательностей:

| Тип дескриптора местоположения | Синтаксис | Описание |
|---|---|---|
| Номера остатков соединены внутрицепочечной перекрестной связью | x..y | Указывает на аминокислоты, соединенные внутрицепочечной связью при использовании признака, который указывает на перекрестную связь внутри цепи, например “CROSSLNK” или “DISULFID”. |

67. Элемент INSDFeature_location нуклеотидных последовательностей может содержать один или несколько операторов местоположения. Оператор местоположения является префиксом либо для одного дескриптора местоположения, либо для комбинации дескрипторов местоположения, соответствуя одному, но прерывающемуся признаку, и указывает, где в указанной последовательности находится местоположение, соответствующее признаку, или как построен признак. Перечень операторов местоположения с их определениями приводится ниже. Операторы местоположения могут использоваться только для нуклеотидов.

| Синтаксическая структура местоположения | Описание местоположения |
|---|---|
| join(location,location, ..., location) | Указанные местоположения объединены (размещены вплотную друг к другу), чтобы сформировать одну непрерывную последовательность. |
| order(location,location, ..., location) | Элементы находятся в указанном порядке, но ничто не подразумевает обоснованности их соединения. |
| complement(location) | Указывает, что признак расположен в цепи, дополнительной к интервалу последовательности, указанному дескриптором местоположения, при чтении в направлении 5’ - 3’ или в направлении, которое имитирует направление 5 '- 3'. |

68. Соединение и порядок расположения операторов местоположения требуют, чтобы были представлены, по крайней мере, две запятые, разделяющие дескрипторы местоположения. Дескрипторы местоположения включаются между двумя соседними остатками, например, x^y, и не должны использоваться в соединении или указании порядка расположения.

Использование оператора определения местоположения соединения подразумевает, что остатки, описываемые дескрипторами местоположения, физически вступили в контакт в биологических процессах (например, экзоны, которые вносят свой вклад в признак участка кодирования)

69. Оператор местоположения “complement” может быть использован в сочетании либо с “join”, либо с “order” в одном и том же местоположении. Сочетания “join” и “order” в одном и том же местоположении не должны использоваться.

70. Следующие примеры иллюстрируют местоположение признаков согласно параграфам 64 – 69 выше:

а) Местоположение для нуклеотидных и аминокислотных последовательностей:

| **Пример местоположения** | **Описание** |
|---|---|
| `467` | Указывает на остаток 467 в последовательности. |
| `340..565` | Указывает на непрерывный диапазон остатков, охватывающий и включающий остатки 340 и 565. |
| `<1` | Указывает на признак местоположения перед первым остатком. |
| `<345..500` | Указывает, что точный нижний предел признака неизвестен. Местоположение начинается с некоторого остатка, предшествующего 345 и продолжается до остатка 500 включительно. |
| `<1..888` | Указывает, что признак начинается перед первым остатком последовательности и продолжается до остатка 888, включая его |
| `1..>888` | Указывает, что признак начинается на первом остатке последовательности и продолжается за пределами остатка 888. |
| `<1..>888` | Указывает, что признак начинается перед первым остатком последовательности и продолжается за пределами остатка 888. |

b) Местоположение только для нуклеотидных последовательностей:

| Пример местоположения | Описание |
|---|---|
| 23^124 | Указывает на местоположение между остатками 123 и 124. |
| | |
| join(12..78,134..202) | Указывает на то, что участки с 12 по 78 и со 134 по 202 следует объединить в одну непрерывную последовательность. |
| | |
| | |
| complement(34..126) | Начинается с нуклеотида, комплементарного к 126, и заканчивается на нуклеотиде, комплементарном к нуклеотиду 34 (признак находится на цепи, комплементарной к представленной цепи). |
| complement(join(2691..4571,4918..5163)) | Соединяет нуклеотиды с 2691 по 4571 и с 4918 по 5163, затем дополняет соединенные участки (признак находится на цепи, комплементарной представленной) |
| join(complement(4918..5163),complement(2691..4571)) | Дополняет участки с 4918 по 5163 и с 2691 по 4571, затем соединяет дополненные участки (признак находится на цепи, комплементарной представленной) |

(c) Местоположения только для аминокислотных последовательностей:

| Пример местоположения | Описание |
|---|---|
| 340..565 | Указывает, что аминокислоты в позициях 340 и 565 соединены посредством внутрицепочечной связи при использовании признака, который указывает внутрицепочечную перекрестную ссылку, такую как «CROSSLNK» или «DISULFID». |

71. В случае XML перечня последовательностей знаки “<” и “>” в дескрипторе местоположения должны быть заменены заранее определёнными элементами (см. параграф 41). Например:

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

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

*Квалификаторы признака*

72. Квалификаторы используются для предоставления информации о признаках в дополнение к предоставляемой функциональным ключом и местоположением признака. Есть три типа форматов для размещения различных типов информации, предоставляемой квалификаторами, а именно:

a) свободный текст (смотри параграфы 85 - 87);

b) контролируемая лексика или числовые значения (например, номер или дата); **и**

c) последовательности.

73. Раздел 6 Приложения 1 предоставляет эксклюзивный перечень квалификаторов и их форматы конкретных значений, если они есть, для каждого функционального ключа нуклеотидной последовательности, а Раздел 8 предоставляет эксклюзивный перечень квалификаторов и их форматы конкретных значений, если они есть, для каждого функционального ключа аминокислотной последовательности.

74. Любая последовательность, охваченная параграфом 7, которая представлена как значение квалификатора, должна быть включена в виде отдельной последовательности с присвоенным ей собственным идентификационным номером последовательности (см. параграф 10).

*Обязательные квалификаторы признаков*

75. Один обязательный функциональный ключ, например, “source” для нуклеотидных последовательностей и аминокислотных последовательностей, требует двух обязательных квалификаторов, “organism” и “mol_type”. Некоторые необязательные функциональные ключи также требуют обязательных квалификаторов.

*Элементы квалификатора*

76. Элемент `INSDFeature_quals` содержит один или более элементов `INSDQualifier`. Каждый элемент `INSDQualifier` представляет один квалификатор и состоит из трёх зависимых элементов и одного необязательного атрибута, указанных ниже:

| Элемент | Описание | Обязательно/Необязательно |
|---|---|---|
| `INSDQualifier_name` | Название квалификатора (смотри Приложение 1, Разделы 6 и 8) | Обязательно |
| `INSDQualifier_value` | Значение квалификатора, если имеется, в определённом формате (смотри Приложение 1, Разделы 6 и 8) и состоящее из тех символов, которые предусмотрены в параграфе 40(b). | Обязательно, когда точно определено (смотри параграф 87 и Приложение 1, Разделы 6 и 8) |
| `NonEnglishQualifier_value` | Значение квалификатора, если имеется, в определённом формате (смотри Приложение 1, Разделы 6 и 8) и состоящее из тех символов, которые предусмотрены в параграфе 40(a). | Обязательно, когда точно определено (смотри параграф 87 и Приложение 1, Разделы 6 и 8) |

| Элемент | Описание | Обязательно/Необязательно |
|---|---|---|
| id | Квалификатор со значением свободного текста, зависящим от языка, может быть однозначно идентифицирован с помощью необязательного XML-атрибута 'id' в элементе INSDQualifier (см. параграф 87(d)). Значение атрибута 'id' должно начинаться с буквы 'q' и продолжаться любым положительным числом. Значение атрибута 'id' должно быть уникальным для одного элемента INSDQualifier, т.е. значение атрибута должно использоваться только один раз в файле перечня последовательностей. | Необязательно |

77. Квалификатор организма, т.е. “organism” для нуклеотидных последовательностей (смотри Приложение 1, Раздел 6) и “organism” для аминокислотных последовательностей (смотри Приложение 1, Раздел 8) должен раскрывать источник, например, отдельный организм или происхождение последовательности. Обозначения организма следует выбирать из таксономической базы данных.

78. Если последовательность - природного происхождения и организм источник имеет род на латыни и обозначение вида, то обозначение должно быть использовано в качестве значения квалификатора. Предпочтительное английское общепринятое название может быть указано с использованием квалификатора “note” для нуклеотидных последовательностей и аминокислотных последовательностей, но не должно быть использовано в значении квалификатора организма.

79. Следующие примеры иллюстрируют организм источник последовательности, согласно параграфам 77 – 78 выше:

Пример 1: Источник для нуклеотидной последовательности:

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

```
  </INSDFeature>
</INSDSeq_feature-table>
```

Пример 2: Источник для аминокислотной последовательности:

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..174</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>
</INSDSeq_feature-table>
```

80. Если последовательность - природного происхождения и организм источник принадлежит к известному роду на латыни, но вид организма не указан или не опознан, то в качестве значения квалификатора организма должно указываться латинское обозначение рода, сопровождаемое “sp.”. Например:

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

81. Если последовательность - природного происхождения и исходный организм не имеет обозначения рода и вида на латыни, то значение квалификатора организма должно указываться как “unidentified” («неидентифицированный»). Любая известная таксономическая информация должна быть указана в квалификаторе “note”.
Например:

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

82. Если последовательность является природной и исходный организм не имеет обозначения рода и вида на латыни, такой как вирус, то должно быть использовано другое приемлемое научное название (например “Canine adenovirus type 2”) как значение квалификатора. Например:

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

83. Если последовательность - не природного происхождения, значение квалификатора организма должно быть указано как “synthetic construct”. Дальнейшая информация в отношении способа генерации последовательности может быть указана с помощью квалификатора “note”.

Например:

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..40</INSDFeature_location>
    <INSDFeature_quals>
      <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>
        <INSDQualifier_name> note </INSDQualifier_name>
        <INSDQualifier_value>synthetic peptide used as assay for
antibodies</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_quals>
  </INSDFeature>
</INSDSeq_feature-table>
```

84. Квалификатор “mol_type” для нуклеотидных последовательностей (см. Приложение I, Раздел 6) и квалификатор “mol_type” для аминокислотных последовательностей (см. Приложение I, Раздел 8) должны раскрывать тип молекулы, представленной в последовательности. Эти квалификаторы отличны от элемента INSDSeq_moltype, рассмотренного в параграфе 54:

(a) Для нуклеотидной последовательности значением квалификатора “mol_type” должно быть одно из перечисленных: “genomic DNA”, “genomic RNA”, “mRNA”, “tRNA”, “rRNA”, “other RNA”, “other DNA”, “transcribed RNA”, “viral cRNA”, “unassigned DNA”, или “unassigned RNA”. Если последовательность неприродного происхождения, например, значением квалификатора “organism” является “synthetic construct”, значением квалификатора “mol_type” должно быть или “other RNA” или “other DNA”;

Для аминокислотных последовательностей значением квалификатора “mol_type” является “protein”.

*Свободный текст*

85. Свободный текст, как указано в параграфе 3, - это тип значения формата для некоторых квалификаторов, представленный в форме фразы описательного текста или в другом указанном формате (как указано в Приложении 1).

86. Использование свободного текста должно быть ограничено несколькими короткими терминами, необходимыми для понимания характеристик последовательности. Для каждого квалификатора, другого, чем квалификатор "translation", свободный текст не должен превышать 1000 знаков (символов).

87. Свободный текст, зависящий от языка, как указано в параграфе 3(o), - это значение свободного текста некоторых квалификаторов, которое зависит от языка в том смысле, что может потребовать перевода для международных, национальных или региональных процедур. Квалификаторы для нуклеотидных последовательностей с форматом значений свободного текста, зависящим от языка, были определены в Приложении I, Раздел 6, Таблица 5. Квалификаторы для аминокислотных последовательностей с форматом значений свободного текста, зависящим от языка, были определены в Приложении I, Раздел 8, Таблица 6.

(a) Свободный текст, зависящий от языка, должен быть представлен в элементе INSDQualifier_value на английском языке или в элементе NonEnglishQualifier_value на языке, отличном от английского, или в обоих элементах. Обратите внимание, что если название организма представляет собой латинское название рода и вида, перевод не требуется. Технические термины и имена собственные, происходящие от неанглийских слов, которые используются во всем мире, рассматриваются как английские для цели значения элемента

INSDQualifier_value (например, «in vitro», «in vivo»).

(b) Если элемент NonEnglishQualifier_value присутствует в перечне последовательностей, в атрибуте nonEnglishFreeTextLanguageCode в корневом элементе должен быть указан (см. Параграф 43) соответствующий код языка (см. Ссылку в параграфе 9 на ISO 639-1: 2002). Все элементы NonEnglishQualifier_value в одном перечне последовательностей должны иметь значения на языке, указанном в атрибуте nonEnglishFreeTextLanguageCode. Элемент NonEnglishQualifier_value разрешен только для квалификаторов, которые имеют формат значения свободного текста, зависящий от языка.

(c) Когда оба элемента NonEnglishQualifier_value и INSDQualifier_value присутствуют для одного квалификатора, информация, содержащаяся в этих двух элементах, должна быть эквивалентной. Одно из следующих условий должно быть истинным: NonEnglishQualifier_value содержит перевод значения INSDQualifier_value; **или** `INSDQualifier_value` содержит перевод значения `NonEnglishQualifier_value`; либо оба элемента содержат перевод значения квалификатора с языка, указанного в атрибуте `originalFreeTextLanguageCode` (см. параграф 43).

(d) Для квалификаторов со значением свободного текста, зависящего от языка, элемент INSDQualifier может включать необязательный атрибут “id”. Значение этого атрибута должно быть в следующем формате: “q”, за которым следует положительное целое число, например «q23» и должно быть уникальным для одного элемента INSDQualifier, т.е. значение атрибута должно использоваться только один раз в файле перечня последовательностей.

88. Приведённые ниже примеры иллюстрируют представление свободного текста, зависящего от языка, как рассматривается в параграфе 87.

Пример 1: свободный текст, зависящий от языка, в элементе INSDQualifier_value:

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

Пример 2: свободный текст, зависящий от языка, в элементе INSDQualifier_value и элементе NonEnglishQualifier_value:

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

Пример 3: свободный текст, зависящий от языка, в элементе NonEnglishQualifier_value:

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

*Кодирующие последовательности.*

89. Для идентификации кодирующих последовательностей, т.е. последовательностей нуклеотидов, которые соответствуют последовательности аминокислот в протеине и стоп кодон, может быть использован функциональный ключ “CDS”. Элемент `INSDFeature_location` должен определять местоположение признака «CDS» и включать стоп кодон.

90. Квалификаторы "transl_table" и "translation" могут быть использованы с функциональным ключом “CDS” (смотри Приложение 1). Там, где не используется квалификатор "transl_table", предполагается использование Таблицы стандартных кодов (смотри Приложение 1, Раздел 9, таблица 7).

91. Квалификатор «transl_except» должен использоваться с функциональным ключом «CDS» и квалификатором “translation” для идентификации кодона, который кодирует пирролизин или селеноцистеин.

92. Аминокислотная последовательность, закодированная кодирующей последовательностью и раскрытая в квалификаторе “translation”, который охватывается параграфом 7, должна быть включена в перечень последовательностей с присвоенным ей собственным идентификационным номером. Идентификационный номер последовательности, присвоенный аминокислотной последовательности, должен быть представлен как значение в квалификаторе "protein_id" с функциональным ключом “CDS”. Квалификатор “organism” функционального ключа “source” для аминокислотной последовательности должен быть идентичен его кодирующей последовательности. Например:

```
      <INSDFeature>
        <INSDFeature_key>CDS</INSDFeature_key>
        <INSDFeature_location>1..507</INSDFeature_location>
        <INSDFeature_quals>
          <INSDQualifier>
            <INSDQualifier_name>transl_table</INSDQualifier_name>
            <INSDQualifier_value>11</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier>
            <INSDQualifier_name>translation</INSDQualifier_name>
            <INSDQualifier_value>MLVHLERTTIMFDFSSLINLPLIWGLLIAIAVLLYILMDGFDL
            GIGILLPFAPSDKCRDHMISSIAPFWDGNETWLVLGGGGLFAAFPLAYSILMPAFYIPIIIMLLGLI
            VRGVSFEFRFKAEGKYRRLWDYAFHFGSLGAAFCQGMILGAFIHGVEVNGRNFSGGQLM
            </INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier>
                <INSDQualifier_name>protein_id</INSDQualifier_name>
```

```
            <INSDQualifier_value>89</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_quals>
    </INSDFeature>
```

*Варианты*

93. Исходная последовательность и любые варианты этой последовательности, каждая раскрытая перечислением её остатков и охваченная параграфом 7, должна быть каждая включена в перечень последовательностей с назначенным ей своим собственным идентификационным номером последовательности.

94. Любой вариант последовательности, раскрытой как единая последовательность с перечислением альтернативных остатков на одной или более позициях, должен быть включен в перечень последовательностей и его следует представлять единой последовательностью, в которой перечисленные альтернативные остатки представлены наиболее ограничивающим неоднозначным символом (см. параграфы 15 и 27).

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

a) могут быть представлены путём аннотирования исходной последовательности, где она содержит вариант(-Ы) на одном местоположении или множестве разных местоположений, и возникновение этих вариантов происходит независимо;

b) следует представлять как отдельную последовательность с собственным присвоенным ей идентификационным номером последовательности, где они содержат варианты на множестве разных местоположений и возникновение этих вариантов является взаимозависимым; **и**

c) должны быть представлены как отдельная последовательность с собственным присвоенным ей идентификационным номером последовательности, там, где она содержит вставленную или замещенную последовательность, содержащую сверх 1000 остатков (см. параграф 86).

96. Таблица ниже показывает правильное использование функциональных ключей и квалификаторов для вариантов нуклеиновой и аминокислотной последовательности:

| **Тип последовательности** | **Функциональный ключ** | **Квалификатор** | **Использование** |
|---|---|---|---|
| Нуклеиновая кислота | variation | “replace” or “note” | Природные мутации и полиморфизмы, например, Alleles, RFLPs |
| Нуклеиновая кислота | misc_difference | “replace” or “note” | Вариабельность, вносимая искусственно, т.е. путём генетических манипуляций или путём химического синтеза |

| Тип последовательности | Функциональный ключ | Квалификатор | Использование |
|---|---|---|---|
| Аминокислота | VAR_SEQ | note | Вариант, полученный путем проведения альтернативного сплайсинга, использования альтернативного промотора, альтернативного инициирования и рибосомного сдвига рамки генетического кода |
| Аминокислота | VARIANT | note | Любой вариант, для которого VAR_SEQ не применим. |

97. Аннотирование последовательности для конкретного варианта должно включать функциональный ключ и квалификатор, как указано в таблице выше, и местоположение признака. Значением квалификатора “replace” должен быть только один альтернативный нуклеотид или нуклеотидная последовательность, использующая только символы, определённые в Разделе 1, Таблице 1 или пустое значение. Перечень альтернативных остатков может быть представлен как значение в квалификаторе “note”. В частности, перечень альтернативных аминокислот должен быть представлен как значение в квалификаторе “note”, где “X” используется в последовательности, и представляет значение другое, чем «любая из немодифицированных L-альфа аминокислот ‘A’, ‘R’, ‘N’, ‘D’, ‘C’, ‘Q’, ‘E’, ‘G’, ‘H’, ‘I’, ‘L’, ‘K’, ‘M’, ‘F’, ‘P’, ‘O’, ‘S’, ‘U’, ‘T’, ‘W’, ‘Y’ или 'V' " (см. параграф 27). Удаление должно быть представлено пустым значением квалификатора для квалификатора "replace" или указанием в “note”, что остаток может быть удалён. Замещение или добавление остатка(ов) должно быть представлено в квалификаторах "replace" или “note”. Для квалификаторов "replace" и “note” форматом значений является свободный текст, который не должен превышать 1000 знаков, как указано в параграфе 86. Смотри параграф 100 для последовательностей, охваченных параграфом 7, которые представлены как вставка или замещение значения квалификатора.

98. Символы, определённые в Приложении 1 (см. Разделы с 1 по 4, Таблицы с 1 по 4 соответственно), должны быть использованы для представления вариантов остатков, где необходимо. Для квалификатора “note”, где вариант остатка является модифицированным остатком, не определённым в таблицах 2 или 4 Приложения 1, полное несокращенное название модифицированного остатка должно быть представлено в качестве значения квалификатора. Модифицированные остатки должны быть в дальнейшем описаны в таблице характеристик, как предусмотрено в параграфах 17 или 30.

99. Следующие примеры иллюстрируют представление вариантов, согласно параграфам 95 – 98 выше.

Пример 1: Функциональный ключ «misc_difference» для перечисленных альтернативных нуклеотидов. Символ “n” на позиции 53 последовательности может быть одним из пяти альтернативных нуклеотидов.

```
<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>53</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>w, cmnm5s2u, mam5u, mcm5s2u, or
p</INSDQualifier_value>
    </INSDQualifier>
```

```
      </INSDFeature_quals>
    </INSDFeature>
    <INSDFeature>
     <INSDFeature_key>modified_base</INSDFeature_key>
     <INSDFeature_location>53</INSDFeature_location>
     <INSDFeature_quals>
           <INSDQualifier>
                 <INSDQualifier_name>mod_base</INSDQualifier_name>
                 <INSDQualifier_value>OTHER</INSDQualifier_value>
           </INSDQualifier>
           <INSDQualifier>
                 <INSDQualifier_name>note</INSDQualifier_name>
                 <INSDQualifier_value>cmnm5s2u, mam5u, mcm5s2u, or
    p</INSDQualifier_value>
           </INSDQualifier>
     </INSDFeature_quals>
</INSDFeature>
```

Пример 2: Функциональный ключ “misc_difference” для удаления в нуклеотидной последовательности. Нуклеотид в позиции 413 удалён.

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

Пример 3: Функциональный ключ “misc_difference” для вставки в нуклеотидную последовательность. Последовательность “atgccaaatat” вставлена между позициями 100 и 101 исходной последовательности.

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

Пример 4: Функциональный ключ “variation” для замещения в нуклеотидной последовательности. Цитозин замещает нуклеотид, указанный на позиции 413 последовательности.

```
     <INSDFeature>
           <INSDFeature_key>variation</INSDFeature_key>
           <INSDFeature_location>413</INSDFeature_location>
           <INSDFeature_quals>
                 <INSDQualifier>
                       <INSDQualifier_name>replace</INSDQualifier_name>
                       <INSDQualifier_value>c</INSDQualifier_value>
                 </INSDQualifier>
           </INSDFeature_quals>
```

Пример 5: Функциональный ключ “VARIANT” для замещения в аминокислотной последовательности. Аминокислота в положении 100 может быть заменена на I, A, F, Y, alle, Melle или Nle.

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

Пример 6: Функциональный ключ “VARIANT” для замещения в аминокислотной последовательности. Аминокислота в положении 100 может быть заменена на любую аминокислоту, кроме Lys, Arg или His.

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

100. Последовательность, охваченная параграфом 7, которая при аннотировании исходной последовательности представлена как вставка или замещение в значении квалификатора, должна также быть включена в перечень последовательностей с присвоенным ей собственным идентификационным номером последовательности.

[Приложение 1 к Стандарту ST.26 следует]

# Приложение 1

Контролируемая лексика

Версия 2.0

*Пересмотр одобрен Комитетом по Стандартам ВОИС (КСВ)*
*на своей тринадцатой сессии 14 ноября 2025*

Содержание

**Раздел 1: Перечень нуклеотидов....2**

**Раздел 2: Перечень модифицированных нуклеотидов....2**

**Раздел 3: Перечень аминокислот....5**

**Раздел 4: Перечень модифицированных аминокислот....6**

**Раздел 5: Функциональные ключи для нуклеотидных последовательностей....7**

**Раздел 6: Квалификаторы для нуклеотидных последовательностей....26**

**Раздел 7: Функциональные ключи для аминокислотных последовательностей....53**

**Раздел 8: Квалификаторы для аминокислотных последовательностей....61**

**Раздел 9: Таблицы генетических кодов....63**

## Раздел 1: Перечень нуклеотидов

Условные обозначения основных нуклеотидов, которые используются в перечне последовательностей, представлены в Таблице 1. Символ “t” будет истолкован как тимин в ДНК и урацил в РНК, если он используется без последующего описания. В случае, когда присвоен неоднозначный символ (представляющий два и более альтернативных основания), должен быть использован наиболее ограничивающий символ. Например, в случае если основаниями данной позиции могут быть “a или g”, следует использовать скорее “r”, чем “n”. Символ “n” будет рассматриваться как “a или c или g или t/u”, если он используется без дополнительного описания.

Таблица 1: Перечень обозначений нуклеотидов

| Символ | Определение |
|---|---|
| a | аденин /adenine |
| c | цитозин /cytosine |
| g | гуанин /guanine |
| t | тимин в ДНК/урацил в РНК (t/u)<br>thymine in DNA/uracil in RNA (t/u) |
| m | a или c |
| r | a или g |
| w | a или t/u |
| s | c или g |
| y | c или t/u |
| k | g или t/u |
| v | a или c или g; не t/u |
| h | a или c или t/u; не g |
| d | a или g или t/u; не c |
| b | c или g или t/u; не a |
| n | a или c или g или t/u; неизвестная или другая |

## Раздел 2: Перечень модифицированных нуклеотидов

Аббревиатуры, перечисленные в Таблице 2, являются единственными разрешёнными значениями для квалификатора mod_base. Если определённый модифицированный нуклеотид не представлен в нижеприведённой таблице, тогда в качестве его значения должна использоваться аббревиатура “OTHER”. Если аббревиатурой является “OTHER”, то полное несокращённое название модифицированного нуклеотида должно быть приведено в квалификаторе “note”. Аббревиатуры, приведенные в Таблице 2, не должны быть использованы в самой последовательности.

Таблица 2: Перечень модифицированных нуклеотидов

| Аббревиатура | Определение |
|---|---|
| ac4c | 4-ацетилцитидин<br>4-acetylcytidine |
| chm5u | 5-(карбоксигидроксиметил)уридин<br>5-(carboxyhydroxylmethyl)uridine |
| cm | 2′-O-метилцитидин<br>2′-O-methylcytidine |

| **Аббревиатура** | **Определение** |
|---|---|
| cmnm5s2u | 5-карбоксиметиламинометил-2-тиоуридин<br>(5-carboxymethylaminomethyl-2-thiouridine) |
| cmnm5u | 5-карбоксиметиламинометилуридин<br>(5-carboxymethylaminomethyluridine) |
| dhu | дигидроуридин<br>(dihydrouridine) |
| fm | 2′-O-метилпсевдоуридин<br>(2′-O-methylpseudouridine) |
| gal q | бета,D-галактозилквеуозин<br>(beta,D-galactosylqueuosine) |
| gm | 2′-O-метилгуанозин<br>(2′-O-methylguanosine) |
| i | инозин<br>(inosine) |
| i6a | N6-изопениладенозин<br>(N6-isopentenyladenosine) |
| m1a | 1-метиладенозин<br>(1-methyladenosine) |
| m1f | 1-метилпсевдоуридин<br>(1-methylpseudouridine) |
| m1g | 1-метилгуанозин<br>(1-methylguanosine) |
| m1i | 1-метилинозин<br>(1-methylinosine) |
| m22g | 2,2-диметилгуанозин<br>(2,2-dimethylguanosine) |
| m2a | 2-метиладенозин<br>(2-methyladenosine) |
| m2g | 2-метилгуанозин<br>(2-methylguanosine) |
| m3c | 3-метилцитидин<br>(3-methylcytidine) |
| m4c | N4-methylcytosine<br>(N4-methylcytosine) |
| m5c | 5-метилцитидин<br>(5-methylcytidine) |
| m6a | N6-метиладенозин<br>(N6-methyladenosine) |
| m7g | 7-метилгуанозин<br>(7-methylguanosine) |
| mam5u | 5-метиламинометилуридин<br>(5-methylaminomethyluridine) |
| mam5s2u | 5- метоксиаминометил-2-тиоуридин<br>(5-methoxyaminomethyl-2-thiouridine) |
| man q | бета,D-манозилквеуозин<br>(beta,D-mannosylqueuosine) |

| **Аббревиатура** | **Определение** |
|---|---|
| mcm5s2u | 5-метиламинометил-2-тиоуридин<br>(5-methylaminomethyl-2-thiouridine) |
| mcm5u | 5-метоксикарбонилметилуридин<br>(5-methoxycarbonylmethyluridine) |
| mo5u | 5-метоксиуридин<br>(5-methoxyuridine) |
| ms2i6a | 2-метилтио-N6-изопентиладенозин<br>(2-methylthio-N6-isopentenyladenosine) |
| ms2t6a | N-((9-бета-D- рибофуранозил-2- метилтиопурин-6-ил) карбомоил)треонин<br>(N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine) |
| mt6a | N-((9-бета-D- рибофураносилпурин-6-ил)N- метилкарбомоил)треонин<br>(N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine) |
| mv | уридин-5-оксоуксусная кислота-метилэфир<br>(uridine-5-oxoacetic acid-methylester) |
| o5u | уридин-5-оксоуксусная кислота (v)<br>(uridine-5-oxyacetic acid (v)) |
| osyw | вибутоксозин<br>(wybutoxosine) |
| p | псевдоуридин<br>(pseudouridine) |
| q | квеуозин<br>(queuosine) |
| s2c | 2-тиоцитидин<br>(2-thiocytidine) |
| s2t | 5-метил-2-тиоуридин<br>(5-methyl-2-thiouridine) |
| s2u | 2-тиоуридин<br>(2-thiouridine_ |
| s4u | 4-тиоуридин<br>(4-thiouridine) |
| m5u | 5-метилуридин<br>(5-methyluridine) |
| t6a | N-((9-бета-D-рибофураназилпурин-6-ил)карбомоил)треонин<br>(N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine) |
| tm | 2’-O-метил-5-метилуридин<br>(2’-O-methyl-5-methyluridine) |
| um | 2’-O-метилуридин<br>(2’-O-methyluridine) |
| yw | вибутозин<br>(wybutosine) |
| x | 3-(3-амино-3-карбоксипропил)уридин, (acp3)u<br>(3-(3-amino-3-carboxypropyl)uridine, (acp3)u) |
| OTHER | (требует квалификатора "note") |

## Раздел 3: Перечень аминокислот

Условные обозначения аминокислот, которые используются в последовательностях, представлены в Таблице 3. Когда присвоен неоднозначный символ (представляющий две или более альтернативных аминокислоты), должен быть использован наиболее ограничивающий символ. Например, если аминокислота в данной позиции может быть аспарагиновой кислотой или аспарагином, следует использовать скорее символ “B”, чем “X”. Символ “X” будет рассматриваться как одна из немодифицированных L-альфа-аминокислот “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, или “V”, когда это используется без дополнительного описания.

Таблица 3: Перечень обозначений аминокислот

| Символ | Определение |
|---|---|
| A | Аланин/Alanine |
| R | Аргинин/Arginine |
| N | Аспарагин/Asparagine |
| D | Аспарагиновая кислота (Аспартат)/<br>Aspartic acid (Aspartate) |
| C | Цистеин /Cysteine |
| Q | Глутамин /Glutamine |
| E | Глутаминовая кислота (Глутамат)/<br>Glutamic acid (Glutamate) |
| G | Глицин/Glycine |
| H | Гистидин/Histidine |
| I | Изолейцин/Isoleucine |
| L | Лейцин/Leucine |
| K | Лизин/Lysine |
| M | Метионин/Methionine |
| F | Фенилаланин/Phenylalanine |
| P | Пролин/Proline |
| O | Пирролизин/Pyrrolysine |
| S | Серин/Serine |
| U | Селеноцистеин/Selenocysteine |
| T | Треонин/Threonine |
| W | Триптофан/Tryptophan |
| Y | Тирозин/Tyrosine |
| V | Валин/Valine |
| B | Аспарагиновая кислота или Аспарагин/Aspartic acid or Asparagine |
| Z | Глутамин или Глутаминовая кислота/Glutamine or Glutamic acid |
| J | Лейцин или Изолейцин/Leucine or Isoleucine |
| X | A или R или N или D или C или Q или E или G или H или I или L или K или M или F или P или O или S или U или T или W или Y или V;<br>“unknown” или “other” |

:

## Раздел 4: Перечень модифицированных аминокислот

В Таблице 4 перечислены единственно разрешённые аббревиатуры для модифицированных аминокислот в обязательном квалификаторе “note” для функциональных ключей “MOD_RES” или “SITE”. Значение для квалификатора “note” должно быть или одной из аббревиатур из этой таблицы, где это возможно, или полным, несокращённым названием модифицированной аминокислоты. Аббревиатуры (или полные названия), приведённые в данной таблице, не должны быть использованы в самой последовательности.

Таблица 4: 4 Перечень модифицированных аминокислот

| **Аббревиатура** | **Модифицированная аминокислота** |
|---|---|
| Aad | 2-аминоадипиновая кислота/2-Aminoadipic acid |
| bAad | 3-аминоадипиновая кислота/3-Aminoadipic acid |
| bAla | бета-аланин, бета-аминопропионовая кислота/<br>beta-Alanine, beta-Aminoproprionic acid |
| Abu | 2-аминобутириновая кислота/2-Aminobutyric acid |
| 4Abu | 4-аминобутириновая кислота, пиперидиновая кислота/<br>4-Aminobutyric acid, piperidinic acid |
| Acp | 6-аминокапроиновая кислота/6-Aminocaproic acid |
| Ahe | 2-аминогептаноиновая кислота/2-Aminoheptanoic acid |
| Aib | 2-аминоизобутириновая кислота/2-Aminoisobutyric acid |
| bAib | 3-аминоизобутириновая кислота/3-Aminoisobutyric acid |
| Apm | 2-аминопимелиновая кислота/2-Aminopimelic acid |
| Dbu | 2,4-диаминобутириновая кислота /2,4-Diaminobutyric acid |
| Des | десмозин/Desmosine |
| Dpm | 2,2’-диаминопимелиновая кислота/2,2’-Diaminopimelic acid |
| Dpr | 2,3-диаминопропионовая кислота/2,3-Diaminoproprionic acid |
| EtGly | N-этилглицин/N-Ethylglycine |
| EtAsn | N-этиласпарагин/N-Ethylasparagine |
| Hyl | гидроксилизин/Hydroxylysine |
| aHyl | алло-гидроксилизин/allo-Hydroxylysine |
| 3Hyp | 3-гидроксипролин/3-Hydroxyproline |
| 4Hyp | 4-гидроксипролин/4-Hydroxyproline |
| Ide | Изодесмозин/Isodesmosine |
| aIle | алло-изолейцин/allo-Isoleucine |
| MeGly | N-метилглицин, саркозин/N-Methylglycine, sarcosine |
| MeIle | N- метилизолейцин/N-Methylisoleucine |
| MeLys | 6-N-метиллизин/6-N-Methyllysine |
| MeVal | N-метилвалин/N-Methylvaline |
| Nva | Норвалин/Norvaline |
| Nle | Норлейцин/Norleucine |
| Orn | Орнитин/Ornithine |

## Раздел 5: Функциональные ключи для нуклеотидных последовательностей

Данный раздел содержит перечень разрешённых функциональных ключей для использования в нуклеотидных последовательностях и перечень обязательных и необязательных квалификаторов. Функциональные ключи перечислены в алфавитном порядке. Функциональные ключи могут быть использованы для ДНК или для РНК, если другое не указано в поле “Molecule scope”. Некоторые функциональные ключи могут подходить для использования с искусственными последовательностями в дополнение к конкретным “organism scope”.

Названия функциональных ключей в случае XML последовательности должны использоваться в точности так, как они приводятся в описаниях ниже в поле “Feature key”, за исключением функциональных ключей 3’UTR и 5’UTR. Для функциональных ключей 3’UTR и 5’UTR см. поле “Comment” в описании.

| | |
|---|---|
| 5.1.Feature Key | C_region |
| Definition | constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; includes one or more exons depending on the particular chain |
| Optional qualifiers | allele |
| gene | |
| gene_synonym | |
| map | |
| note | |
| product | |
| pseudo | |
| pseudogene | |
| standard_name | |
| Organism scope | eukaryotes |

| | |
|---|---|
| 5.2.Feature Key | CDS |
| Definition | coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature may include amino acid conceptual translation |
| Optional qualifiers | allele |
| circular_RNA | |
| codon_start | |
| EC_number | |
| exception | |
| function | |
| gene | |
| gene_synonym | |
| map | |
| note | |
| number | |
| operon | |
| product | |
| protein_id | |
| pseudo | |
| pseudogene | |

| | |
|---|---|
| ribosomal_slippage<br>standard_name<br>translation<br>transl_except<br>transl_table<br>trans_splicing | |
| Comment | codon_start qualifier has valid value of 1 or 2 or 3, indicating the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature; transl_table defines the genetic code table used if other than the Standard or universal genetic code table; genetic code exceptions outside the range of the specified tables are reported in transl_except qualifier; only one of the qualifiers translation, pseugogene or pseudo are permitted with a CDS feature key; when the translation qualifier is used, the protein_id qualifier is mandatory if the translation product contains four or more specifically defined amino acids |

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

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

| | |
|---|---|
| 5.5. Feature Key | D_segment |
| Definition | Diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain |
| Optional qualifiers<br>Gene | allele |

| | |
|---|---|
| gene_synonym | |
| map | |
| note | |
| product | |
| pseudo | |
| pseudogene | |
| standard_name | |
| Organism scope | eukaryotes |

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

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

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

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

| | |
|---|---|
| 5.10. Feature Key | J_segment |
| Definition | joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains |
| Optional qualifiers | allele<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name |
| Organism scope | eukaryotes |

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

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

| | |
|---|---|
| 5.13 Feature Key | misc_difference |
| Definition | featured sequence  differs from the presented sequence at this location and cannot be described by any other Difference key (variation, or modified_base) |
| Optional qualifiers<br>clone<br>compare<br>gene<br>gene_synonym<br>map<br>note<br>phenotype<br>replace<br>standard_name | allele |

| | |
|---|---|
| Comment | the misc_difference feature key must be used to describe variability introduced artificially, e.g., by genetic manipulation or by chemical synthesis; use the replace qualifier to annotate a deletion, insertion, or substitution. The variation feature key must be used to describe naturally occurring genetic variability. |

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

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

| | |
|---|---|
| 5.16 Feature Key | misc_RNA |
| Definition | any transcript or RNA product that cannot be defined by other RNA keys (prim_transcript, precursor_RNA, mRNA, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, ncRNA, rRNA and tRNA) |
| Optional qualifiers | allele<br>function<br>gene<br>gene_synonym |

map
note
operon
product
pseudo
pseudogene
standard_name
trans_splicing

| | |
|---|---|
| 5.17 Feature Key | misc_structure |
| Definition | any secondary or tertiary nucleotide structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop) |
| Optional qualifiers | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>standard_name |

| | |
|---|---|
| 5.18 Feature Key | mobile_element |
| Definition | region of genome containing mobile elements |
| Mandatory qualifiers | mobile_element_type |
| Optional qualifiers | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>rpt_family<br>rpt_type<br>standard_name |

| | |
|---|---|
| 5.19 Feature Key | modified_base |
| Definition | the indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value) |
| Mandatory qualifiers | mod_base |
| Optional qualifiers | allele<br>Frequency<br>gene<br>gene_synonym<br>map<br>note |
| Comment | value for the mandatory mod_base qualifier is limited to the |

| | |
|---|---|
| | restricted vocabulary for modified base abbreviations in Section 2 of this Annex. |

| | |
|---|---|
| 5.20. Feature Key | mRNA |
| Definition | messenger RNA; includes 5' untranslated region (5'UTR), coding sequences (CDS, exon) and 3' untranslated region (3'UTR) |
| Optional qualifiers<br>circular_RNA<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>pseudo<br>pseudogene<br>standard_name<br>trans_splicing | allele |

| | |
|---|---|
| 5.21. Feature Key | ncRNA |
| Definition | a non-protein-coding gene, other than ribosomal RNA and transfer RNA, the functional molecule of which is the RNA transcript |
| Mandatory qualifiers | ncRNA_class |
| Optional qualifiers<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>pseudo<br>pseudogene<br>standard_name<br>trans_splicing | allele |
| Comment | the ncRNA feature must not be used for ribosomal and transfer RNA annotation, for which the rRNA and tRNA feature keys must be used, respectively |

| | |
|---|---|
| 5.22 Feature Key | N_region |
| Definition | extra nucleotides inserted between rearranged immunoglobulin segments |
| Optional qualifiers<br>gene<br>gene_synonym<br>map<br>note | allele |

| | |
|---|---|
| product | |
| pseudo | |
| pseudogene | |
| standard_name | |
| Organism scope | eukaryotes |

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

| | |
|---|---|
| 5.24. Feature Key | oriT |
| Definition | origin of transfer; region of a DNA molecule where transfer is initiated during the process of conjugation or mobilization |
| Optional qualifiers | allele |
| bound_moiety | |
| direction | |
| gene | |
| gene_synonym | |
| map | |
| note | |
| rpt_family | |
| rpt_type | |
| rpt_unit_range | |
| rpt_unit_seq | |
| standard_name | |
| Molecule Scope | DNA |
| Comment | rep_origin must be used to describe origins of replication; direction qualifier has legal values left, right and both, however only left and right are valid when used in conjunction with the oriT feature; origins of transfer can be present in the chromosome; plasmids can contain multiple origins of transfer |



| | |
|---|---|
| 5.25 Feature Key | polyA_site |
| Definition | site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation |

| | |
|---|---|
| Optional qualifiers<br>gene<br>gene_synonym<br>map<br>note | allele |
| Organism scope | eukaryotes and eukaryotic viruses |

| | |
|---|---|
| 5.26. Feature Key | precursor_RNA |
| Definition | any RNA species that is not yet the mature RNA product; may include ncRNA, rRNA, tRNA,5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR) |
| Optional qualifiers<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>standard_name<br>trans_splicing | allele |
| Comment | used for RNA which may be the result of post-transcriptional processing; if the RNA in question is known not to have been processed, use the prim_transcript key |

| | |
|---|---|
| 5.27. Feature Key | prim_transcript |
| Definition | primary (initial, unprocessed) transcript; may include ncRNA, rRNA, tRNA, 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR) |
| Optional qualifiers<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>standard_name | allele |



| | |
|---|---|
| 5.28. Feature Key | primer_bind |
| Definition | non-covalent primer binding site for initiation of replication, transcription, or reverse transcription; includes site(s) for synthetic e.g., PCR primer elements |

| | |
|---|---|
| Optional qualifiers | allele<br>gene<br>gene_synonym<br><br>map<br>note<br>standard_name |
| Comment | used to annotate the site on a given sequence to which a primer molecule binds - not intended to represent the sequence of the primer molecule itself; since PCR reactions most often involve pairs of primers, a single primer_bind key may use the order(location,location) operator with two locations, or a pair of primer_bind keys may be used |

| | |
|---|---|
| 5.29. Feature Key | propeptide |
| Definition | propeptide coding sequence; coding sequence for the domain of a proprotein that is cleaved to form the mature protein product. |
| Optional qualifiers | allele<br><br>function<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name |

| | |
|---|---|
| 5.30. Feature Key | protein_bind |
| Definition | non-covalent protein binding site on nucleic acid |
| Mandatory qualifiers | bound_moiety |
| Optional qualifiers | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>standard_name |
| Comment | note that the regulatory feature key and regulatory_class qualifier |



| | |
|---|---|
| | with the value ”ribosome_binding_site” must be used to describe ribosome binding sites |

| | |
|---|---|
| 5.31. Feature Key | regulatory |
| Definition | any region of a sequence that functions in the regulation of transcription, translation, replication or chromatin structure; |

| | |
|---|---|
| Mandatory qualifiers | regulatory_class |
| Optional qualifiers | allele |
| | bound_moiety |
| | function |
| | gene |
| | gene_synonym |
| | map |
| | note |
| | operon |
| | phenotype |
| | pseudo |
| | pseudogene |
| | standard_name |

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

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

| | |
|---|---|
| 5.34. Feature Key | rRNA |
| Definition | mature ribosomal RNA; RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins |

| | |
|---|---|
| Optional qualifiers | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>pseudo<br>pseudogene<br>standard_name |
| Comment | rRNA sizes should be annotated with the product qualifier |

| | |
|---|---|
| 5.35. Feature Key | S_region |
| Definition | switch region of immunoglobulin heavy chains; involved in the rearrangement of heavy chain DNA leading to the expression of a different immunoglobulin class from the same B-cell |
| Optional qualifiers | allele<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name |
| Organism scope | eukaryotes |

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



| | |
|---|---|
| 5.37. Feature Key | source |
| Definition | identifies the source of the sequence; this key is mandatory; every sequence will have a single source key spanning the entire sequence |

| | |
|---|---|
| Mandatory qualifiers | organism<br>mol_type |
| Optional qualifiers | cell_line<br>cell_type<br>chromosome<br>clone<br>clone_lib<br>collected_by<br>collection_date<br>cultivar<br>dev_stage<br>ecotype<br>environmental_sample<br>germline<br>haplogroup<br>haplotype<br>host<br>identified_by<br>isolate<br>isolation_source<br>lab_host<br>lat_lon<br>macronuclear<br>map<br>mating_type<br>note<br>organelle<br>PCR_primers<br>plasmid<br>pop_variant<br>proviral<br>rearranged<br>segment<br>serotype<br>serovar<br>sex<br>strain<br>sub_clone<br>sub_species<br>sub_strain<br>tissue_lib<br>tissue_type<br>variety |
| Molecule scope | any |



| | |
|---|---|
| 5.38. Feature Key | stem_loop |
| Definition | hairpin; a double-helical region formed by base-pairing between adjacent (inverted) complementary sequences in a single strand of RNA or DNA |
| Optional qualifiers | allele<br>function<br>gene |

| | |
|---|---|
| | gene_synonym<br>map<br>note<br>operon<br>standard_name |

| | |
|---|---|
| 5.39. Feature Key | STS |
| Definition | sequence tagged site; short, single-copy DNA sequence that characterizes a mapping landmark on the genome and can be detected by PCR; a region of the genome can be mapped by determining the order of a series of STSs |
| Optional qualifiers | allele<br>gene<br>gene_synonym<br>map<br>note<br>standard_name |
| Molecule scope | DNA |
| Comment | STS location to include primer(s) in primer_bind key or primers |

| | |
|---|---|
| 5.40. Feature Key | telomere |
| Definition | region of biological interest identified as a telomere and which has been experimentally characterized |
| Optional qualifiers | note<br>rpt_type<br>rpt_unit_range<br>rpt_unit_seq<br>standard_name |
| Comment | the telomere feature describes the interval of DNA that corresponds to a specific structure at the end of the linear eukaryotic chromosome which is required for the integrity and maintenance of the end; this region is unique compared to the rest of the chromosome and represents the physical end of the chromosome |

| | |
|---|---|
| 5.41. Feature Key | tmRNA |
| Definition | transfer messenger RNA; tmRNA acts as a tRNA first, and then as an mRNA that encodes a peptide tag; the ribosome translates this mRNA region of tmRNA and attaches the encoded peptide tag to the C-terminus of the unfinished protein; this attached tag targets the protein for destruction or proteolysis |
| Optional qualifiers | allele<br>function<br>gene<br>gene_synonym<br>map |

note
product
pseudo
pseudogene
standard_name
tag_peptide

| | |
|---|---|
| 5.42. Feature Key | transit_peptide |
| Definition | transit peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organellar protein; this domain is involved in post-translational import of the protein into the organelle |
| Optional qualifiers | allele |
| function | |
| gene | |
| gene_synonym | |
| map | |
| note | |
| product | |
| pseudo | |
| pseudogene | |
| standard_name | |

| | |
|---|---|
| 5.43. Feature Key | tRNA |
| Definition | mature transfer RNA, a small RNA molecule (75-85 bases long) that mediates the translation of a nucleic acid sequence into an amino acid sequence |
| Optional qualifiers | allele |
| circular_RNA | |
| anticodon | |
| function | |
| gene | |
| gene_synonym | |
| map | |
| note | |
| operon | |
| product | |
| pseudo | |
| pseudogene | |
| standard_name | |
| trans_splicing | |



| | |
|---|---|
| 5.44. Feature Key | unsure |
| Definition | a small region of sequenced bases, generally 10 or fewer in its length, whichcould not be confidently identified. Such a region might contain called bases (a, t, g, or c), or a mixture of called-bases and uncalled-bases ('n'). |
| Optional qualifiers | allele |
| compare | |
| gene | |

| | |
|---|---|
| | gene_synonym<br>map<br>note<br>replace |
| Comment | use the replace qualifier to annotate a deletion, insertion, or substitution. |

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

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



| | |
|---|---|
| | standard_name |
| Organism scope | eukaryotes |

| | |
|---|---|
| 5.47. Feature Key | variation |
| Definition | a related strain contains stable mutations from the same gene (e.g., RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others) |
| Optional qualifiers | allele |

compare
frequency
gene
gene_synonym
map
note
phenotype
product
replace
standard_name

Comment — used to describe alleles, RFLP's,and other naturally occurring mutations and polymorphisms; use the replace qualifier to annotate a deletion, insertion, or substitution; variability arising as a result of genetic manipulation (e.g. site directed mutagenesis) must be described with the misc_difference feature;

5.48. Feature Key — 3'UTR

Definition — 1) region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein;
2) region at the 3' end of an RNA virus (following the last stop codon) that is not translated into a protein;

Optional qualifiers — allele
function
gene
gene_synonym
map
note
standard_name
trans_splicing

Comment — The apostrophe character has special meaning in XML, and must be substituted with "'" in the value of an element. Thus "3'UTR" must be represented as "3'UTR" in the XML file, i.e., <INSDFeature_key>3'UTR</INSDFeature_key>.

5.49. Feature Key — 5'UTR

Definition — 1) region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein;



2) region at the 5' end of an RNA virus (preceding the first initiation codon) that is not translated into a protein;

Optional qualifiers — allele
function
gene
gene_synonym
map
note
standard_name

trans_splicing

Comment — The apostrophe character has special meaning in XML, and must be substituted with “'” in the value of an element. Thus “5’UTR” must be represented as “5'UTR” in the XML file, i.e., <INSDFeature_key>5'UTR</INSDFeature_key>.

## Раздел 6: Квалификаторы для нуклеотидных последовательностей

Данный раздел содержит перечень квалификаторов, которые должны быть использованы для признаков в нуклеотидных последовательностях. Квалификаторы перечислены в алфавитном порядке.

Если формат значения “none”, элементы INSDQualifier_value и NonEnglishQualifier_value не должны использоваться.

Если формат значения представляет собой свободный текст, который определяется как зависящий от языка, должно использоваться одно из следующих значений:

1) элемент INSDQualifier_value; **или**

2) элемент NonEnglishQualifier_value; **или**

3) как элемент INSDQualifier_value, так и элемент NonEnglishQualifier_value.

Там, где формат значений отличается от “none”, но не идентифицируется как свободный текст, зависящий от языка, должен использоваться элемент INSDQualifier_value, а элемент NonEnglishQualifier_value не должен использоваться.

ПОЖАЛУЙСТА, ОБРАТИТЕ ВНИМАНИЕ: Любое значение квалификатора, предусмотренное для квалификатора со значением «свободный текст» (“free text”), зависящий от языка, формат может потребовать перевода для международных, национальных или региональных процедур. Квалификаторы, перечисленные в следующей таблице, считаются имеющими значения свободного текста, зависящего от языка:

Таблица 5: Список значений квалификаторов для нуклеотидных последовательностей с зависимыми от языка значениями свободного текста

| **Часть** | **Квалификатор свободного текста, зависящего от языка** |
|---|---|
| 6.3 | bound_moiety |
| 6.5 | cell_type |
| 6.8 | clone |
| 6.9 | clone_lib |
| 6.11 | collected_by |
| 6.14 | cultivar |
| 6.15 | dev_stage |
| 6.18 | ecotype |
| 6.21 | frequency |
| 6.22 | function |
| 6.24 | gene_synonym |
| 6.26 | haplogroup |
| 6.28 | host |
| 6.29 | identified_by |
| 6.30 | isolate |
| 6.31 | isolation_source |
| 6.32 | lab_host |
| 6.36 | mating_type |
| 6.41 | note |
| 6.45 | organism |
| 6.47 | phenotype |
| 6.49 | pop_variant |
| 6.50 | product |
| 6.66 | serotype |
| 6.67 | serovar |
| 6.68 | sex |
| 6.69 | standard_name |
| 6.70 | strain |
| 6.71 | sub_clone |

| Часть | Квалификатор свободного текста, зависящего от языка |
|---|---|
| 6.72 | sub_species |
| 6.73 | sub_strain |
| 6.75 | tissue_lib |
| 6.76 | tissue_type |
| 6.81 | variety |

| | | |
|---|---|---|
| 6.1. | Qualifier | allele |
| | Definition | name of the allele for the given gene |
| | Mandatory Value format | free text |
| | Example | <INSDQualifier_value>adh1-1</INSDQualifier_value> |
| | Comment | all gene-related features (exon, CDS etc) for a given gene should share the same allele qualifier value; the allele qualifier value must, by definition, be different from the gene qualifier value; when used with the variation feature key, the allele qualifier value should be that of the variant. |

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

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

| | | |
|---|---|---|
| 6.4. | Qualifier | cell_line |
| | Definition | cell line from which the sequence was obtained |

| | |
|---|---|
| Mandatory Value format | free text |
| Example | <INSDQualifier_value>MCF7</INSDQualifier_value> |

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

| | |
|---|---|
| 6.6. Qualifier | chromosome |
| Definition | chromosome (e.g., Chromosome number) from which the sequence was obtained |
| Mandatory Value format | free text |
| Example | <INSDQualifier_value>1</INSDQualifier_value><br><INSDQualifier_value>X</INSDQualifier_value> |

| | |
|---|---|
| 6.7. Qualifier | circular_RNA |
| Definition | indicates that exons are out-of-order or overlapping because this spliced RNA product is a circular RNA (circRNA) created by backsplicing, for example when a downstream exon in the gene is located 5' of an upstream exon in the RNA product |
| Value format | none |
| Comment | should be used on features such as CDS, mRNA, tRNA and other features that are produced as a result of a backsplicing event. This qualifier should be used only when the splice event is indicated in the "join" operator, eg join(complement(69611..69724),139856..140087) |

| | |
|---|---|
| 6.8. Qualifier | clone |
| Definition | clone from which the sequence was obtained |
| Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/International/National/Regional procedures |
| Example | <INSDQualifier_value>lambda-hIL7.3</INSDQualifier_value> |
| Comment | a source feature must not contain more than one clone should qualifier; where the sequence was obtained from multiple clones it may be further described in the feature table using the feature key misc_feature and a note qualifier to specify the multiple clones. |

| | |
|---|---|
| 6.9. Qualifier | clone_lib |
| Definition | clone library from which the sequence was obtained |
| Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/International/National/Regional procedures |
| Example | <INSDQualifier_value>lambda-hIL7</INSDQualifier_value> |

| | |
|---|---|
| 6.10. Qualifier | codon_start |
| Definition | indicates the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature. |
| Mandatory Value format | 1 or 2 or 3 |
| Example | <INSDQualifier_value>2</INSDQualifier_value> |

| | |
|---|---|
| 6.11. Qualifier | collected_by |
| Definition | name of persons or institute who collected the specimen |
| Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/International/National/Regional procedures |
| Example | <INSDQualifier_value>Dan Janzen</INSDQualifier_value> |

| | |
|---|---|
| 6.12. Qualifier | collection_date |
| Definition | date that the specimen was collected |
| Mandatory Value format | YYYY-MM-DD, YYYY-MM or YYYY |
| Example | <INSDQualifier_value>1952-10-21</INSDQualifier_value><br><INSDQualifier_value>1952-10</INSDQualifier_value><br><INSDQualifier_value>1952</INSDQualifier_value> |

| | |
|---|---|
| 6.13. Qualifier | compare |
| Definition | Reference details of an existing public INSD entry to which a comparison is made |
| Mandatory Value format | [accession-number.sequence-version] |
| Example | <INSDQualifier_value>AJ634337.1</INSDQualifier_value> |
| Comment | This qualifier may be used on the following features: misc_difference, unsure, and variation. Multiple compare qualifiers with different contents are allowed within a single feature. This qualifier is not intended for large-scale annotation of variations, such as SNPs. |

| | | |
|---|---|---|
| 6.14. | Qualifier | cultivar |
| | Definition | cultivar (cultivated variety) of plant from which sequence was obtained |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/International/National/Regional procedures |
| | Example | <INSDQualifier_value>Nipponbare</INSDQualifier_value><br><INSDQualifier_value>Tenuifolius</INSDQualifier_value><br><INSDQualifier_value>Candy Cane</INSDQualifier_value><br><INSDQualifier_value>IR36</INSDQualifier_value> |
| | Comment | 'cultivar' is applied solely to products of artificial selection; use the variety qualifier for natural, named plant and fungal varieties. |

| | | |
|---|---|---|
| 6.15. | Qualifier | dev_stage |
| | Definition | if the sequence was obtained from an organism in a specific developmental stage, it is specified with this qualifier |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/International/National/Regional procedures) |
| | Example | <INSDQualifier_value>fourth instar larva</INSDQualifier_value> |

| | | |
|---|---|---|
| 6.16. | Qualifier | direction |
| | Definition | direction of DNA replication |
| | Mandatory Value format | left, right, or both<br>where left indicates toward the 5' end of the sequence (as presented) and right indicates toward the 3' end |
| | Example | <INSDQualifier_value>left</INSDQualifier_value> |
| | Comment | The values left, right, and both are permitted when the direction qualifier is used to annotate a rep_origin feature key. However, only left and right values are permitted when the direction qualifier is used to annotate an oriT feature key. |

| | | |
|---|---|---|
| 6.17. | Qualifier | EC_number |
| | Definition | Enzyme Commission number for enzyme product of sequence |
| | Mandatory Value format | free text |
| | Example | <INSDQualifier_value>1.1.2.4</INSDQualifier_value><br><INSDQualifier_value>1.1.2.-</INSDQualifier_value><br><INSDQualifier_value>1.1.2.n</INSDQualifier_value><br><INSDQualifier_value>1.1.2.n1</INSDQualifier_value> |

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

| | | |
|---|---|---|
| 6.20. | Qualifier | exception |
| | Definition | indicates that the coding region cannot be translated using standard biological rules |
| | Mandatory Value format | One of the following controlled vocabulary phrases:<br>RNA editing<br>rearrangement required for product<br>annotated by transcript or proteomic data |
| | Example | <INSDQualifier_value>RNA editing</INSDQualifier_value><br><INSDQualifier_value>rearrangement required for product</INSDQualifier_value> |
| | Comment | only to be used to describe biological mechanisms such as RNA editing; protein translation of a CDS with an exception qualifier will be different from the corresponding conceptual translation; must not be used where transl_except qualifier would be adequate, e.g., in case of stop codon completion use. |

| | | |
|---|---|---|
| 6.21. | Qualifier | frequency |
| | Definition | frequency of the occurrence of a feature |
| | Mandatory Value format | free text representing the proportion of a population carrying the feature expressed as a fraction<br>Language-dependent: this value may require translation for International/International/National/Regional procedures |
| | Example | <INSDQualifier_value>23/108</INSDQualifier_value><br><INSDQualifier_value>1 in 12</INSDQualifier_value><br><INSDQualifier_value>0.85</INSDQualifier_value> |

| | | |
|---|---|---|
| 6.22. | Qualifier | function |
| | Definition | function attributed to a sequence |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/International/National/Regional procedures |
| | Example | <INSDQualifier_value>essential for recognition of cofactor</INSDQualifier_value> |
| | Comment | The function qualifier is used when the gene name and/or product name do not convey the function attributable to a sequence. |

| | | |
|---|---|---|
| 6.23. | Qualifier | gene |
| | Definition | symbol of the gene corresponding to a sequence region |
| | Mandatory Value format | free text |
| | Example | <INSDQualifier_value>ilvE</INSDQualifier_value> |

| | | |
|---|---|---|
| | Comment | Use gene qualifier to provide the gene symbol; use standard_name qualifier to provide the full gene name. |

| | | |
|---|---|---|
| 6.24. | Qualifier | gene_synonym |
| | Definition | synonymous, replaced, obsolete or former gene symbol |
| | Mandatory Value format | free text |
| | Language-dependent: this value may require translation for | International/International/National/Regional procedures |
| | Example | <INSDQualifier_value>Hox-3.3</INSDQualifier_value> in a feature where the gene qualifier value is Hoxc6 |
| | Comment | used where it is helpful to indicate a gene symbol synonym; when the gene_synonym qualifier is used, a primary gene symbol must always be indicated in a gene qualifier |

| | | |
|---|---|---|
| 6.25. | Qualifier | germline |
| | Definition | the sequence presented has not undergone somatic rearrangement as part of an adaptive immune response; it is the unrearranged sequence that was inherited from the parental germline |
| | Value format | none |
| | Comment | germline qualifier must not be used to indicate that the source of the sequence is a gamete or germ cell; germline and rearranged qualifiers must not be used in the same source feature; germline and rearranged qualifiers must only be used for molecules that can undergo somatic rearrangements as part of an adaptive immune response; these are the T-cell receptor (TCR) and immunoglobulin loci in the jawed vertebrates, and the unrelated variable lymphocyte receptor (VLR) locus in the jawless fish (lampreys and hagfish); germline and rearranged qualifiers should not be used outside of the Craniata (taxid=89593) |

| | | |
|---|---|---|
| 6.26. | Qualifier | haplogroup |
| | Definition | name for a group of similar haplotypes that share some sequence variation. Haplogroups are often used to track migration of population groups. |
| | Mandatory Value format | free text |
| | Language-dependent: this value may require translation for | International/International/National/Regional procedures |
| | Example | <INSDQualifier_value>H*</INSDQualifier_value> |

| | | |
|---|---|---|
| 6.27. | Qualifier | haplotype |
| | Definition | name for a specific set of alleles that are linked together on the same physical chromosome. In the absence of recombination, each haplotype is inherited as a unit, and may be used to track gene |

flow in populations.

Mandatory Value format free text

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

6.28. Qualifier host

Definition natural (as opposed to laboratory) host to the organism from which sequenced molecule was obtained

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

Example <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
<INSDQualifier_value>Homo sapiens 12 year old girl</INSDQualifier_value>
<INSDQualifier_value>Rhizobium NGR234</INSDQualifier_value>

6.29. Qualifier identified_by

Definition name of the expert who identified the specimen taxonomically

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

Example <INSDQualifier_value>John Burns</INSDQualifier_value>

6.30. Qualifier isolate

Definition individual isolate from which the sequence was obtained

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

Example <INSDQualifier_value>Patient #152</INSDQualifier_value>
<INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>

6.31. Qualifier isolation_source

Definition describes the physical, environmental and/or local geographical

source of the biological sample from which the sequence was derived

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

Examples <INSDQualifier_value>rumen isolates from standard Pelleted ration-fed steer #67</INSDQualifier_value>
<INSDQualifier_value>permanent Antarctic sea

ice</INSDQualifier_value>
<INSDQualifier_value>denitrifying activated sludge from carbon_limited continuous reactor</INSDQualifier_value>

Comment used only with the source feature key; source feature keys containing an environmental_sample qualifier should also contain an isolation_source qualifier

6.32. Qualifier lab_host

Definition scientific name of the laboratory host used to propagate the source organism from which the sequenced molecule was obtained

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

Example <INSDQualifier_value>Gallus gallus</INSDQualifier_value>
<INSDQualifier_value>Gallus gallus embryo</INSDQualifier_value>
<INSDQualifier_value>Escherichia coli strain DH5 alpha</INSDQualifier_value>
<INSDQualifier_value>Homo sapiens HeLa cells</INSDQualifier_value>

Comment the full binomial scientific name of the host organism should be used when known; extra conditional information relating to the host may also be included

6.33. Qualifier lat_lon

Definition geographical coordinates of the location where the specimen was collected

Mandatory Value format free text - degrees latitude and longitude in format "d[d.dddd] N|S d[dd.dddd] W|E"

Example <INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value>
<INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value>

6.34. Qualifier macronuclear

Definition if the sequence shown is DNA and from an organism which undergoes chromosomal differentiation between macronuclear and micronuclear stages, this qualifier is used to denote that the sequence is from macronuclear DNA

Value format none

6.35. Qualifier map

Definition genomic map position of feature

Mandatory Value format free text

Example <INSDQualifier_value>8q12- q13</INSDQualifier_value>

6.36. Qualifier mating_type

Definition mating type of the organism from which the sequence was obtained; mating type is used for prokaryotes, and for eukaryotes that undergo meiosis without sexually dimorphic gametes

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

Examples <INSDQualifier_value>MAT-1</INSDQualifier_value>
<INSDQualifier_value>plus</INSDQualifier_value>
<INSDQualifier_value>-</INSDQualifier_value>
<INSDQualifier_value>odd</INSDQualifier_value>
<INSDQualifier_value>even</INSDQualifier_value>"

Comment mating_type qualifier values male and female are valid in the prokaryotes, but not in the eukaryotes;
for more information, see the entry for the sex qualifier.

6.37. Qualifier mobile_element_type

Definition type and name or identifier of the mobile element which is described by the parent feature

Mandatory Value format <mobile_element_type>[:<mobile_element_name>]
where <mobile_element_type> is one of the following:
transposon
retrotransposon
integron
insertion sequence
non-LTR retrotransposon
SINE
MITE
LINE
other

Example <INSDQualifier_value>transposon:Tnp9</INSDQualifier_value>

Comment mobile_element_type is legal on mobile_element feature key only. Mobile element should be used to represent both elements which are currently mobile, and those which were mobile in the past. Value "other" for <mobile_element_type> requires a <mobile_element_name>

6.38. Qualifier mod_base

Definition abbreviation for a modified nucleotide base

Mandatory Value format modified base abbreviation chosen from this Annex, Section 2

Example <INSDQualifier_value>m5c</INSDQualifier_value>
<INSDQualifier_value>OTHER</INSDQualifier_value>

| | | |
|---|---|---|
| | Comment | specific modified nucleotides not found in paragraph 2 of this Annex are annotated by entering OTHER as the value for the mod_base<br><br>qualifier and including a note qualifier with the full name of the modified base as its value |
| 6.39. | Qualifier | mol_type |
| | Definition | molecule type of sequence |
| | Mandatory Value format | One chosen from the following:<br>genomic DNA<br>genomic RNA<br>mRNA<br>tRNA<br>rRNA<br>other RNA<br>other DNA<br>transcribed RNA<br>viral cRNA<br>unassigned DNA<br>unassigned RNA |
| | Example | <INSDQualifier_value>genomic DNA</INSDQualifier_value><br><INSDQualifier_value>other RNA</INSDQualifier_value> |
| | Comment | mol_type qualifier is mandatory on the source feature key; the value "genomic DNA" does not imply that the molecule is nuclear (e.g., organelle and plasmid DNA must be described using "genomic DNA"); ribosomal RNA genes must be described using "genomic DNA";<br><br>"rRNA" must only be used if the ribosomal RNA molecule itself has been sequenced; values "other RNA" and "other DNA" must be applied to synthetic molecules, values "unassigned DNA", "unassigned RNA" must be applied where in vivo molecule is unknown. |
| 6.40. | Qualifier | ncRNA_class |
| | Definition | a structured description of the classification of the non-coding RNA described by the ncRNA parent key |
| | Mandatory Value format | TYPE<br>where TYPE is one of the following controlled vocabulary terms or phrases:<br>antisense_RNA<br>autocatalytically_spliced_intron<br>circRNA<br>ribozyme<br>hammerhead_ribozyme<br>lncRNA<br>RNase_P_RNA<br>RNase_MRP_RNA<br>telomerase_RNA<br>guide_RNA<br>sgRNA<br>rasiRNA |

scRNA
scaRNA
siRNA
pre_miRNA
miRNA
piRNA
snoRNA
snRNA
SRP_RNA
vault_RNA
Y_RNA
Other

Example <INSDQualifier_value>autocatalytically_spliced_intron</INSDQualifier_value>
<INSDQualifier_value>siRNA</INSDQualifier_value>
<INSDQualifier_value>scRNA</INSDQualifier_value>
<INSDQualifier_value>other</INSDQualifier_value>

Comment specific ncRNA types not yet in the ncRNA_class controlled vocabulary must be annotated by entering other as the ncRNA_class qualifier value, and providing a brief explanation of novel ncRNA_class in a note qualifier

6.41. Qualifier note

Definition any comment or additional information

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

Example <INSDQualifier_value>A comment about the feature</INSDQualifier_value>

6.42. Qualifier number

Definition a number to indicate the order of genetic elements (e.g., exons or introns) in the 5’ to 3’ direction

Mandatory Value format free text (with no whitespace characters)

Example <INSDQualifier_value>4</INSDQualifier_value>
<INSDQualifier_value>6B</INSDQualifier_value>

Comment text limited to integers, letters or combination of integers and/or letters represented as a data value that contains no whitespace characters; any additional terms should be included in a standard_name qualifier. Example: a number qualifier with a value of 2A and a standard_name qualifier with a value of long

6.43. Qualifier operon

Definition name of the group of contiguous genes transcribed into a single transcript to which that feature belongs

Mandatory Value format free text

Example <INSDQualifier_value>lac</INSDQualifier_value>

6.44. Qualifier organelle

Definition type of membrane-bound intracellular structure from which the sequence was obtained

Mandatory Value format One of the following controlled vocabulary terms and phrases:
chromatophore
hydrogenosome
mitochondrion
nucleomorph
plastid
mitochondrion:kinetoplast
plastid:chloroplast
plastid:apicoplast
plastid:chromoplast
plastid:cyanelle
plastid:leucoplast
plastid:proplastid,

Examples <INSDQualifier_value>chromatophore</INSDQualifier_value>
<INSDQualifier_value>hydrogenosome</INSDQualifier_value>
<INSDQualifier_value>mitochondrion</INSDQualifier_value>
<INSDQualifier_value>nucleomorph</INSDQualifier_value>
<INSDQualifier_value>plastid</INSDQualifier_value>
<INSDQualifier_value>mitochondrion:kinetoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:chloroplast</INSDQualifier_value>
<INSDQualifier_value>plastid:apicoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:chromoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:cyanelle</INSDQualifier_value>
<INSDQualifier_value>plastid:leucoplast</INSDQualifier_value>
<INSDQualifier_value>plastid:proplastid</INSDQualifier_value>

6.45. Qualifier organism

Definition scientific name of the organism that provided the sequenced genetic material, if known, or the available taxonomic information if the organism is unclassified; or an indication that the sequence is a synthetic construct

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

Example <INSDQualifier_value>Homo sapiens</INSDQualifier_value>

6.46. Qualifier PCR_primers

Definition PCR primers that were used to amplify the sequence. A single /PCR_primers qualifier should contain all the primers used for a single PCR reaction. If multiple forward or reverse primers are present in a single PCR reaction, multiple sets of fwd_name/fwd_seq

| | | |
|---|---|---|
| | | or rev_name/rev_seq values will be present |
| | Mandatory Value format | [fwd_name: XXX1, ]fwd_seq: xxxxx1,[fwd_name: XXX2, ]fwd_seq: xxxxx2, [rev_name: YYY1, ]rev_seq: yyyyy1,[rev_name: YYY2, ]rev_seq: yyyyy2 |
| | Example | <INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt,rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value><INSDQualifier_value>fwd_name: hoge1, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<i>gtgtatcttact</INSDQualifier_value>INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt, fwd_name: CO1P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value> |
| | Comment | fwd_seq and rev_seq are both mandatory; fwd_name and rev_name are both optional.Both sequences must be presented in 5’ to 3’ order. The sequences must be given in the symbols from Annex 1, paragraph 0,1, except for the modified bases, which must be enclosed within angle brackets < >. In XML, the angle brackets < and > must be substituted with < and > since they are reserved characters in XML. |
| 6.47. | Qualifier | phenotype |
| | Definition | phenotype conferred by the feature, where phenotype is defined as a physical, biochemical or behavioural characteristic or set of characteristics |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>erythromycin resistance</INSDQualifier_value> |
| 6.48. | Qualifier | plasmid |
| | Definition | name of naturally occurring plasmid from which the sequence was obtained, where plasmid is defined as an independently replicating genetic unit that cannot be described by chromosome or segment qualifiers |
| | Mandatory Value format | free text |
| | Example | <INSDQualifier_value>pC589</INSDQualifier_value> |
| 6.49. | Qualifier | pop_variant |
| | Definition | name of subpopulation or phenotype of the sample from which the sequence was derived |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>pop1</INSDQualifier_value> |

```
<INSDQualifier_value>Bear Paw</INSDQualifier_value>
```

| | | |
|---|---|---|
| 6.50. | Qualifier | product |
| | Definition | name of the product associated with the feature, e.g., the mRNA of an mRNA feature, the polypeptide of a CDS, the mature peptide of a mat_peptide, etc. |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>trypsinogen</INSDQualifier_value> (when qualifier appears in CDS feature)<br><INSDQualifier_value>trypsin</INSDQualifier_value> (when qualifier appears in mat_peptide feature)<br><INSDQualifier_value>XYZ neural-specific transcript</INSDQualifier_value> (when qualifier appears in mRNA feature) |
| 6.51. | Qualifier | protein_id |
| | Definition | protein sequence identification number, an integer used in a sequence listing to designate the protein sequence encoded by the coding sequence identified in the corresponding CDS feature key and translation qualifier |
| | Mandatory Value format | an integer greater than zero |
| | Example | <INSDQualifier_value>89</INSDQualifier_value> |
| 6.52. | Qualifier | proviral |
| | Definition | this qualifier is used to flag sequence obtained from a virus or phage that is integrated into the genome of another organism |
| | Value format | none |
| 6.53. | Qualifier | pseudo |
| | Definition | indicates that this feature is a non-functional version of the element named by the feature key |
| | Value format | none |
| | Comment | The qualifier pseudo should be used to describe non-functional genes that are not formally described as pseudogenes, e.g., CDS has Other reasons may include sequencing or assembly errors. In order to annotate pseudogenes the qualifier pseudogene must be used, indicating the TYPE of pseudogene |

6.54. Qualifier pseudogene

Definition indicates that this feature is a pseudogene of the element named by the feature key

Mandatory Value format TYPE
where TYPE is one of the following controlled vocabulary terms or phrases:
processed
unprocessed
unitary
allelic
unknown

Example <INSDQualifier_value>processed</INSDQualifier_value>
<INSDQualifier_value>unprocessed</INSDQualifier_value>
<INSDQualifier_value>unitary</INSDQualifier_value>
<INSDQualifier_value>allelic</INSDQualifier_value>
<INSDQualifier_value>unknown</INSDQualifier_value>

Comment Definitions of TYPE values:
processed - the pseudogene has arisen by reverse transcription of a mRNA into cDNA, followed by reintegration into the genome. Therefore, it has lost any intron/exon structure, and it might have a pseudo-polyA-tail.
unprocessed - the pseudogene has arisen from a copy of the parent gene by duplication followed by accumulation of random mutations. The changes, compared to their functional homolog, include insertions, deletions, premature stop codons, frameshifts and a higher proportion of non-synonymous versus synonymous substitutions.
unitary - the pseudogene has no parent. It is the original gene, which is functional is some species but disrupted in some way (indels, mutation, recombination) in another species or strain.
allelic - a (unitary) pseudogene that is stable in the population but importantly it has a functional alternative allele also in the population. i.e., one strain may have the gene, another strain may have the pseudogene. MHC haplotypes have allelic pseudogenes.
unknown - the submitter does not know the method of pseudogenization.

6.55. Qualifier rearranged

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

Value format none

Comment The rearranged qualifier must not be used to annotate chromosome rearrangements that are not involved in an adaptive immune response; germline and rearranged qualifiers must not be used in the same source feature; germline and rearranged qualifiers must only be used for molecules that can undergo somatic rearrangements as part of an adaptive immune response; these are the T-cell receptor (TCR) and immunoglobulin loci in the jawed vertebrates, and the unrelated

variable lymphocyte receptor (VLR) locus in the jawless fish (lampreys and hagfish); germline and rearranged qualifiers should not be used outside of the Craniata (taxid=89593)

6.56. Qualifier: recombination_class

Definition: a structured description of the classification of recombination hotspot region within a sequence

Mandatory Value format: TYPE
where TYPE is one of the following controlled vocabulary terms or phrases:
meiotic
mitotic
non_allelic_homologous
chromosome_breakpoint
other

Example: <INSDQualifier_value>meiotic recombination</INSDQualifier_value>

<INSDQualifier_value>chromosome_breakpoint</INSDQualifier_value>

Comment: specific recombination classes not yet in the recombination_class controlled vocabulary must be annotated by entering “other” as the recombination_class qualifier value and providing a brief explanation of the novel recombination_class in a note qualifier

6.57. Qualifier: regulatory_class

Definition: a structured description of the classification of transcriptional, translational, replicational and chromatin structure related regulatory elements in a sequence

Mandatory Value format: TYPE
where TYPE is one of the following controlled vocabulary terms or phrases:
attenuator
CAAT_signal
DNase_I_hypersensitive_site
enhancer
enhancer_blocking_element
GC_signal
imprinting_control_region
insulator
locus_control_region
matrix_attachment_region
minus_35_signal
minus_10_signal
polyA_signal_sequence
promoter
recoding_stimulatory_region
recombination_enhancer
replication_regulatory_region
response_element
ribosome_binding_site

| | | |
|---|---|---|
| | | riboswitch<br>silencer<br>TATA_box<br>terminator<br>transcriptional_cis_regulatory_region<br>uORF<br>other |
| | Example | <INSDQualifier_value>promoter</INSDQualifier_value><br><INSDQualifier_value>enhancer</INSDQualifier_value><br><INSDQualifier_value>ribosome_binding_site</INSDQualifier_value> |
| | Comment | specific regulatory classes not yet in the regulatory_class controlled vocabularymust be annotated by entering “other” as the regulatory_class qualifier value and providing a brief explanation of the novel regulatory_class in a note qualifier |
| 6.58. | Qualifier | replace |
| | Definition | indicates that the sequence identified in a feature’s location is replaced by the sequence shown in the qualifier’s value; if no sequence (i.e., no value) is contained within the qualifier, this indicates a deletion |
| | Mandatory Value format | free text |
| | Example | <INSDQualifier_value>a</INSDQualifier_value><br><INSDQualifier_value></INSDQualifier_value> - for a deletion |
| 6.59. | Qualifier | ribosomal_slippage |
| | Definition | during protein translation, certain sequences can program ribosomes to change to an alternative reading frame by a mechanism known as ribosomal slippage |
| | Value format | none |
| | Comment | a join operator, e.g.: [join(486..1784,1787..4810)] must be used in the CDS feature location to indicate the location of ribosomal_slippage |
| 6.60. | Qualifier | rpt_family |
| | Definition | type of repeated sequence; "Alu" or "Kpn", for example |
| | Mandatory Value format | free text |
| | Example | <INSDQualifier_value>Alu</INSDQualifier_value> |

| | | |
|---|---|---|
| 6.61. | Qualifier | rpt_type |
| | Definition | structure and distribution of repeated sequence |
| | Mandatory Value format | One of the following controlled vocabulary terms or phrases:<br>tandem<br>direct<br>inverted<br>flanking<br>nested<br>terminal<br>dispersed<br>long_terminal_repeat<br>non_ltr_retrotransposon_polymeric_tract<br>centromeric_repeat<br>telomeric_repeat<br>x_element_combinatorial_repeat<br>y_prime_element<br>other |
| | Example | <INSDQualifier_value>inverted</INSDQualifier_value><br><INSDQualifier_value>long_terminal_repeat</INSDQualifier_value> |
| | Comment | Definitions of the values:<br>tandem - a repeat that exists adjacent to another in the same orientation;<br>direct - a repeat that exists not always adjacent but is in the same orientation;<br>inverted - a repeat pair occurring in reverse orientation to one another on the same molecule;<br>flanking - a repeat lying outside the sequence for which it has functional significance (eg. transposon insertion target sites);<br>nested - a repeat that is disrupted by the insertion of another element;<br>dispersed - a repeat that is found dispersed throughout the genome;<br>terminal - a repeat at the ends of and within the sequence for which it has functional significance (eg. transposon LTRs);<br>long_terminal_repeat - a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses;<br>non_ltr_retrotransposon_polymeric_tract - a polymeric tract, such as poly(dA), within a non LTR retrotransposon;<br>centromeric_repeat - a repeat region found within the modular centromere;<br>telomeric_repeat - a repeat region found within the telomere;<br>x_element_combinatorial_repeat - a repeat region located between the X element and the telomere or adjacent Y' element;<br>y_prime_element - a repeat region located adjacent to telomeric repeats or X element combinatorial repeats, either as a single copy or tandem repeat of two to four copies;<br>other - a repeat exhibiting important attributes that cannot be described by other values. |

| | | |
|---|---|---|
| 6.62. | Qualifier | rpt_unit_range |
| | Definition | location of a repeating unit expressed as a range |
| | Mandatory Value format | <base_range> - where <base_range> is the first and last base (separated by two dots) of a repeating unit |
| | Example | <INSDQualifier_value>202..245</INSDQualifier_value> |
| | Comment | used to indicate the base range of the sequence that constitutes a repeating unit within the region specified by the feature keys oriT and repeat_region. |

| | | |
|---|---|---|
| 6.63. | Qualifier | rpt_unit_seq |
| | Definition | identity of a repeat sequence |
| | Mandatory Value format | free text |
| | Example | <INSDQualifier_value>aagggc</INSDQualifier_value><br><INSDQualifier_value>ag(5)tg(8)</INSDQualifier_value><br><INSDQualifier_value>(AAAGA)6(AAAA)1(AAAGA)12</INSDQualifier_value> |
| | Comment | used to indicate the literal sequence that constitutes a repeating unit within the region specified by the feature keys oriT and repeat_region |

| | | |
|---|---|---|
| 6.64. | Qualifier | satellite |
| | Definition | identifier for a satellite DNA marker, composed of many tandem repeats (identical or related) of a short basic repeated unit |
| | Mandatory Value format | <satellite_type>[:<class>][ <identifier>] - where <satellite_type> is one of the following:<br>satellite<br>microsatellite<br>minisatellite |
| | Example | <INSDQualifier_value>satellite: S1a</INSDQualifier_value><br><INSDQualifier_value>satellite: alpha</INSDQualifier_value><br><INSDQualifier_value>satellite: gamma III</INSDQualifier_value><br><INSDQualifier_value>microsatellite: DC130</INSDQualifier_value> |
| | Comment | many satellites have base composition or other properties that differ from those of the rest of the genome that allows them to be identified. |

| | | |
|---|---|---|
| 6.65. | Qualifier | segment |
| | Definition | name of viral or phage segment sequenced |
| | Mandatory Value format | free text |
| | Example | <INSDQualifier_value>6</INSDQualifier_value> |

| | | |
|---|---|---|
| 6.66. | Qualifier | serotype |
| | Definition | serological variety of a species characterized by its antigenic properties |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>B1</INSDQualifier_value> |
| | Comment | used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for the prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms". |

| | | |
|---|---|---|
| 6.67. | Qualifier | serovar |
| | Definition | serological variety of a species (usually a prokaryote) characterized by its antigenic properties |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>O157:H7</INSDQualifier_value> |
| | Comment | used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms". |

| | | |
|---|---|---|
| 6.68. | Qualifier | sex |
| | Definition | sex of the organism from which the sequence was obtained; sex is used for eukaryotic organisms that undergo meiosis and have sexually dimorphic gametes |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Examples | <INSDQualifier_value>female</INSDQualifier_value><br><INSDQualifier_value>male</INSDQualifier_value><br><INSDQualifier_value>hermaphrodite</INSDQualifier_value><br><INSDQualifier_value>unisexual</INSDQualifier_value><br><INSDQualifier_value>bisexual</INSDQualifier_value><br><INSDQualifier_value>asexual</INSDQualifier_value><br><INSDQualifier_value>monoecious</INSDQualifier_value> [or monecious]<br><INSDQualifier_value>dioecious</INSDQualifier_value> [or diecious] |
| | Comment | The sex qualifier should be used (instead of mating_type qualifier) in the Metazoa, Embryophyta, Rhodophyta & Phaeophyceae; mating_type |

| | | |
|---|---|---|
| | | qualifier should be used (instead of sex qualifier) in the Bacteria, Archaea & Fungi; neither sex nor mating_type qualifiers should be used in the viruses; outside of the taxa listed above, mating_type qualifier should be used unless the value of the qualifier is taken from the vocabulary given in the examples above |
| 6.69. | Qualifier | standard_name |
| | Definition | accepted standard name for this feature |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>dotted</INSDQualifier_value> |
| | Comment | use standard_name qualifier to give full gene name, but use gene qualifier to give gene symbol (in the above example gene qualifier value is Dt). |
| 6.70. | Qualifier | strain |
| | Definition | strain from which sequence was obtained |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>BALB/c</INSDQualifier_value> |
| | Comment | feature entries including a strain qualifier must not include the environmental_sample qualifier |
| 6.71. | Qualifier | sub_clone |
| | Definition | sub-clone from which sequence was obtained |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>lambda-hIL7.20g</INSDQualifier_value> |
| | Comment | a source feature must not contain more than one sub_clone qualifier; to indicate that the sequence was obtained from multiple sub_clones, multiple sources may be further described using the feature key “misc_feature” and the qualifier “note” |
| 6.72. | Qualifier | sub_species |
| | Definition | name of sub-species of organism from which sequence was obtained |
| | Mandatory Value format | free text |

| | | |
|---|---|---|
| | | Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>lactis</INSDQualifier_value> |
| 6.73. | Qualifier | sub_strain |
| | Definition | name or identifier of a genetically or otherwise modified strain from which sequence was obtained, derived from a parental strain (which should be annotated in the strain qualifier). sub_strain from which sequence was obtained |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>abis</INSDQualifier_value> |
| | Comment | must be accompanied by a strain qualifier in a source feature;if the parental strain is not given, the modified strain should be annotated in the strain qualifier instead of sub_strain. For example, either a strain qualifier with the value K-12 and a substrain qualifier with the value MG1655 or a strain qualifier with the value MG1655 |
| 6.74. | Qualifier | tag_peptide |
| | Definition | base location encoding the polypeptide for proteolysis tag of tmRNA and its termination codon |
| | Mandatory Value format | <base_range> - where <base_range> provides the first and last base (separated by two dots) of the location for the proteolysis tag |
| | Example | <INSDQualifier_value>90..122</INSDQualifier_value> |
| | Comment | it is recommended that the amino acid sequence corresponding to the tag_peptide be annotated by describing a 5’ partial CDS feature; e.g., CDS with a location of <90..122 |
| 6.75. | Qualifier | tissue_lib |
| | Definition | tissue library from which sequence was obtained |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>tissue library 772</INSDQualifier_value> |
| 6.76. | Qualifier | tissue_type |
| | Definition | tissue type from which the sequence was obtained |
| | Mandatory Value format | free text |

| | | |
|---|---|---|
| | | Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>liver</INSDQualifier_value> |

| | | |
|---|---|---|
| 6.77. | Qualifier | transl_except |
| | Definition | translational exception: single codon the translation of which does not conform to genetic code defined by organism or transl_table. |
| | Mandatory Value format | (pos:<location>,aa:<amino_acid>) where <amino_acid> is the three letter abbreviation for the amino acid coded by the codon at the base_range position |
| | Example | <INSDQualifier_value>(pos:213..215,aa:Trp) </INSDQualifier_value><br><INSDQualifier_value>(pos:462..464,aa:OTHER) </INSDQualifier_value><br><INSDQualifier_value>(pos:1017,aa:TERM) </INSDQualifier_value><br><INSDQualifier_value>(pos:2000..2001,aa:TERM) </INSDQualifier_value> |
| | Comment | if the amino acid is not one of the specific amino acids listed in Annex 1, paragraph 3, use OTHER as <amino_acid> and provide the name of the unusual amino acid in a note qualifier; for modified amino-acid selenocysteine use three letter abbreviation 'Sec' (one letter symbol 'U' in amino-acid sequence) for <amino _acid>; for modified amino-acid pyrrolysine use three letter abbreviation 'Pyl' (one letter symbol 'O'in amino-acid sequence) for <amino _acid>; for partial termination codons where TAA stop codon is completed by the addition of 3' A residues to the mRNA either a single base_position or a base_range is used for the location, see the third and fourth examples above, in conjunction with a note qualifier indicating 'stop codon completed by the addition of 3' A residues to the mRNA'. |

| | | |
|---|---|---|
| 6.78. | Qualifier | transl_table |
| | Definition | definition of genetic code table used if other than universal or standard genetic code table. Tables used are described in this Annex |
| | Mandatory Value format | <integer><br>where <integer> is the number assigned to the genetic code table |
| | Example | <INSDQualifier_value>3</INSDQualifier_value> - example where the yeast mitochondrial code is to be used |
| | Comment | if the transl_table qualifier is not used to further annotate a CDS feature key, then the CDS is translated using the Standard Code (i.e. Universal Genetic Code). Genetic code exceptions outside the range of specified tables are reported in transl_except qualifiers. |

| | | |
|---|---|---|
| 6.79. | Qualifier | trans_splicing |
| | Definition | indicates that exons from two RNA molecules are ligated in intermolecular reaction to form mature RNA |
| | Value format | none |

| | | |
|---|---|---|
| | Comment | should be used on features such as CDS, mRNA and other features that are produced as a result of a trans-splicing event. This qualifier must be used only when the splice event is indicated in the "join" operator, e.g., join(complement(69611..69724),139856..140087) in the feature location |

| | | |
|---|---|---|
| 6.80. | Qualifier | translation |
| | Definition | one-letter abbreviated amino acid sequence derived from either the standard (or universal) genetic code or the table as specified in a transl_table qualifier and as determined by an exception in the transl_except qualifier |
| | Mandatory Value format | contiguous string of one-letter amino acid abbreviations from this Annex paragraph 3, "X" is to be used for AA exceptions. |
| | Example | <INSDQualifier_value>MASTFPPWYRGCASTPSLKGLIMCTW</INSDQualifier_value> |
| | Comment | to be used with CDS feature only; must be accompanied by protein_id qualifier when the translation product contains four or more specifically defined amino acids; see transl_table for definition and location of genetic code Tables; only one of the qualifiers translation, pseudo and pseudogene are permitted to further annotate a CDS feature. |

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

## Раздел 7: Функциональные ключи для аминокислотных последовательностей

Данный раздел содержит перечень разрешенных функциональных ключей, которые будут использоваться в аминокислотных последовательностях. Функциональные ключи перечислены в алфавитном порядке.

| | | |
|---|---|---|
| 7..1 | Feature Key | ACT_SITE |
| | Definition | Amino acid(s) involved in the activity of an enzyme |
| | Optional qualifiers | note |
| | Comment | Each amino acid residue of the active site must be annotated separately with the ACT_SITE feature key. The corresponding amino acid residue number must be provided as the location descriptor in the feature location element. |

| | | |
|---|---|---|
| 7.2. | Feature Key | BINDING |
| | Definition | Binding site for any chemical group (co-enzyme, prosthetic group, etc.). The chemical nature of the group is indicated in the note qualifier |
| | Mandatory qualifiers | note |
| | Comment | Examples of values for the “note” qualifier: “Heme (covalent)” and “Chloride.” Where appropriate, the features keys CA_BIND, DNA_BIND, METAL,and NP_BIND should be used rather than BINDING. |

| | | |
|---|---|---|
| 7.3. | Feature Key | CA_BIND |
| | Definition | Extent of a calcium-binding region |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.4. | Feature Key | CARBOHYD |
| | Definition | Glycosylation site |
| | Mandatory qualifiers | note |
| | Comment | This key describes the occurrence of the attachment of a glycan (mono- or polysaccharide) to a residue of the protein. The type of linkage (C-, N- or O-linked) to the protein is indicated in the “note” qualifier. If the nature of the reducing terminal sugar is known, its abbreviation is shown between parentheses. If three dots ’...’ follow the abbreviation this indicates an extension of the carbohydrate chain. Conversely no dots means that a monosaccharide is linked. Examples of values used in the “note” qualifier:N-linked (GlcNAc...); O-linked (GlcNAc); O-linked (Glc...);C-linked (Man)partial;O-linked (Ara...). |

| | | |
|---|---|---|
| 7.5. | Feature Key | CHAIN |
| | Definition | Extent of a polypeptide chain in the mature protein |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.6. | Feature Key | COILED |
| | Definition | Extent of a coiled-coil region |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.7. | Feature Key | COMPBIAS |
| | Definition | Extent of a compositionally biased region |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.8. | Feature Key | CONFLICT |
| | Definition | Different sources report differing sequences. |
| | Optional qualifiers | note |
| | Comment | Examples of values for the “note” qualifier: Missing; K -> Q; GSDSE -> RIRLR; V -> A. |

| | | |
|---|---|---|
| 7.9. | Feature Key | CROSSLNK |
| | Definition | Post translationally formed amino acid bonds. |
| | Mandatory qualifiers | note |
| | Comment | Covalent linkages of various types formed between two proteins (interchain cross-links) or between two parts of the same protein (intrachain cross-links); except for cross-links formed by disulfide bonds, for which the “DISULFID” feature key is to be used. For an interchain cross-link, the location descriptor in the feature location element is the residue number of the amino acid cross-linked to the other protein. For an intrachain cross-link, the location descriptor in the feature location element is the residue numbers of the cross-linked amino acids in “x..y” format, e.g. ”42..50”. The note qualifier indicates the nature of the cross-link; at least specifying the name of the conjugate and the identity of the two amino acids involved. Examples of values for the “note” qualifier: “Isoglutamyl cysteine thioester (Cys-Gln);” “Beta-methyllanthionine (Cys-Thr);” and “Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin)” |

| | | |
|---|---|---|
| 7.10. | Feature Key | DISULFID |
| | Definition | Disulfide bond |

| | | |
|---|---|---|
| | Optional qualifiers | note |
| | Comment | For an interchain disulfide bond, the location descriptor in the feature location element is the residue number of the cysteine linked to the other protein. For an intrachain cross-link, the location descriptor in the feature location element is the residue numbers of the linked cysteines in “x..y” format, e.g. ”42..50”. For interchain disulfide bonds, the note qualifier indicates the nature of the cross-link, by identifying the other protein, for example, “Interchain (between A and B chains)” |
| 7.11. | Feature Key | DNA_BIND |
| | Definition | Extent of a DNA-binding region |
| | Mandatory qualifiers | note |
| | Comment | The nature of the DNA-binding region is given in the note qualifier. Examples of values for the “note” qualifier: “Homeobox” and “Myb 2” |
| 7.12. | Feature Key | DOMAIN |
| | Definition | Extent of a domain, which is defined as a specific combination of secondary structures organized into a characteristic three-dimensional structure or fold |
| | Mandatory qualifiers | note |
| | Comment | The domain type is given in the NOTE qualifier. Where several copies of a domain are present, the domains are numbered. Examples of values for the “note” qualifier: “Ras-GAP” and “Cadherin 1” |
| 7.13. | Feature Key | HELIX |
| | Definition | Secondary structure: Helices, for example, Alpha-helix; 3(10) helix; or Pi-helix |
| | Optional qualifiers | note |
| | Comment | This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a ’loop’ or ’random-coil’ structure. |
| 7.14. | Feature Key | INIT_MET |
| | Definition | Initiator methionine |
| | Optional qualifiers | note |
| | Comment | The location descriptor in the feature location element is “1”. This |

| | | |
|---|---|---|
| | | feature key indicates the N-terminal methionine is cleaved off. This feature is not used when the initiator methionine is not cleaved off. |
| 7.15. | Feature Key | INTRAMEM |
| | Definition | Extent of a region located in a membrane without crossing it |
| | Optional qualifiers | note |
| 7.16. | Feature Key | LIPID |
| | Definition | Covalent binding of a lipid moiety |
| | Mandatory qualifiers | note |
| | Comment | The chemical nature of the bound lipid moiety is given in the NOTE qualifier, indicating at least the name of the lipidated amino acid. Examples of values for the “note” qualifier: “N-myristoyl glycine”; “GPI-anchor amidated serine” and “S-diacylglycerol cysteine.” |
| 7.17. | Feature Key | METAL |
| | Definition | Binding site for a metal ion. |
| | Mandatory qualifiers | note |
| | Comment | The note qualifier indicates the nature of the metal. Examples of values for the “note” qualifier: “Iron (heme axial ligand)”” and “Copper”. |
| 7.18. | Feature Key | MOD_RES |
| | Definition | Posttranslational modification of a residue |
| | Mandatory qualifiers | note |
| | Comment | The chemical nature of the modified residue is given in the NOTE qualifier, indicating at least the name of the post-translationally modified amino acid. If the modified amino acid is listed in Section 4 of this Annex, the abbreviation may be used in place of the the full name. Examples of values for the “NOTE” qualifier: “N-acetylalanine”; “3-Hyp”; and “MeLys” or “N-6-methyllysine" |
| 7.19. | Feature Key | MOTIF |
| | Definition | Short (up to 20 amino acids) sequence motif of biological interest |
| | Optional qualifiers | note |
| 7.20. | Feature Key | MUTAGEN |
| | Definition | Site which has been experimentally altered by mutagenesis |

| | | |
|---|---|---|
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.21. | Feature Key | NON_STD |
| | Definition | Non-standard amino acid |
| | Optional qualifiers | note |
| | Comment | This key describes the occurrence of non-standard amino acids selenocysteine (U) and pyrrolysine (O) in the amino acid sequence. |

| | | |
|---|---|---|
| 7.22. | Feature Key | NON_TER |
| | Definition | The residue at an extremity of the sequence is not the terminal residue |
| | Optional qualifiers | note |
| | Comment | If applied to position 1, this means that the first position is not the N-terminus of the complete molecule. If applied to the last position, it means that this position is not the C-terminus of the complete molecule. |

| | | |
|---|---|---|
| 7.23. | Feature Key | NP_BIND |
| | Definition | Extent of a nucleotide phosphate-binding region |
| | Mandatory qualifiers | note |
| | Comment | The nature of the nucleotide phosphate is indicated in the note qualifier. Examples of values for the “note” qualifier: “ATP” and “FAD”. |

| | | |
|---|---|---|
| 7.24. | Feature Key | PEPTIDE |
| | Definition | Extent of a released active peptide |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.25. | Feature Key | PROPEP |
| | Definition | Extent of a propeptide |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.26. | Feature Key | REGION |
| | Definition | Extent of a region of interest in the sequence |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.27. | Feature Key | REPEAT |
| | Definition | Extent of an internal sequence repetition |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.28. | Feature Key | SIGNAL |
| | Definition | Extent of a signal sequence (prepeptide) |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.29. | Feature Key | SITE |
| | Definition | Any interesting single amino-acid site on the sequence that is not defined by another feature key. It can also apply to an amino acid bond which is represented by the positions of the two flanking amino acids |
| | Mandatory qualifier | note |
| | Comment | When SITE is used to annotate a modified amino acid the value for the qualifier “note” must either be an abbreviation set forth in in Section 4 of this Annex or the complete, unabbreviated name of the modified amino acid. |

| | | |
|---|---|---|
| 7.30. | Feature Key | source |
| | Definition | Identifies the source of the sequence; this key is mandatory; every sequence will have a single source feature spanning the entire sequence |
| | Mandatory qualifiers<br>organism | mol_type |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.31. | Feature Key | STRAND |
| | Definition | Secondary structure: Beta-strand; for example Hydrogen bonded beta-strand or residue in an isolated beta-bridge |
| | Optional qualifiers | note |
| | Comment | This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a ’loop’ or ’random-coil’ structure. |

| | | |
|---|---|---|
| 7.32. | Feature Key | TOPO_DOM |
| | Definition | Topological domain |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.33. | Feature Key | TRANSMEM |
| | Definition | Extent of a transmembrane region |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.34. | Feature Key | TRANSIT |
| | Definition | Extent of a transit peptide (mitochondrion, chloroplast, thylakoid, cyanelle, peroxisome etc.) |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.35. | Feature Key | TURN |
| | Definition | Secondary structure Turns, for example, H-bonded turn (3-turn, 4-turn or 5-turn) |
| | Optional qualifiers | note |
| | Comment | This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a 'loop' or 'random-coil' structure. |

| | | |
|---|---|---|
| 7.36. | Feature Key | UNSURE |
| | Definition | Uncertainties in the sequence |
| | Optional qualifiers | note |
| | Comment | Used to describe region(s) of an amino acid sequence for which the authors are unsure about the sequence presentation. |

| | | |
|---|---|---|
| 7.37 | Feature Key | VARIANT |
| | Definition | Authors report that sequence variants exist. |
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.38. | Feature Key | VAR_SEQ |
| | Definition | Description of sequence variants produced by alternative splicing, alternative promoter usage, alternative initiation and ribosomal frameshifting |

| | | |
|---|---|---|
| | Optional qualifiers | note |

| | | |
|---|---|---|
| 7.39. | Feature Key | ZN_FING |
| | Definition | Extent of a zinc finger region |
| | Mandatory qualifiers | note |
| | Comment | The type of zinc finger is indicated in the note qualifier. For example: “GATA-type” and “NR C4-type” |

## Раздел 8: Квалификаторы для аминокислотных последовательностей

Данный раздел содержит перечень разрешенных квалификаторов, которые будут использоваться в аминокислотных последовательностях.

Если формат значения представляет собой свободный текст, зависящий от языка, должно использоваться одно из следующих значений:

1) элемент `INSDQualifier_value`; **или**

2) элемент `NonEnglishQualifier_value`; **или**

3) как элемент `INSDQualifier_value`, так и элемент `NonEnglishQualifier_value`.

Если формат значения не определяется как свободный текст, зависящий от языка, должен использоваться элемент `INSDQualifier_value`, а элемент `NonEnglishQualifier_value` не должен использоваться.

ПОЖАЛУЙСТА, ОБРАТИТЕ ВНИМАНИЕ: Любое значение квалификатора, предусмотренное для квалификатора со значением формата “free text” («свободный текст»), может потребовать перевода для международных, национальных или региональных процедур. Считается, что квалификаторы, перечисленные в следующей таблице, имеют значения свободного текста, зависящие от языка:

Таблица 6: Список квалификаторов со значениями свободного текста, зависящими от языка, для аминокислотных последовательностей

| Раздел | Квалификатор свободного текста, зависящего от языка |
|---|---|
| 8.2 | note |
| 8.3 | organism |

| | | |
|---|---|---|
| 8.1. | Qualifier | mol_type |
| | Definition | In vivo molecule type of sequence |
| | Mandatory Value format | protein |
| | Example | <INSDQualifier_value>protein</INSDQualifier_value> |
| | Comment | The "mol_type" qualifier is mandatory on the source feature key. |

| | | |
|---|---|---|
| 8.2. | Qualifier | note |
| | Definition | Any comment or additional information |
| | Mandatory Value format | free text<br>Language-dependent: this value may require translation for International/National/Regional procedures |
| | Example | <INSDQualifier_value>Heme (covalent)</INSDQualifier_value> |
| | Comment | The “note” qualifier is mandatory for the feature keys: BINDING; CARBOHYD; CROSSLNK; DISULFID; DNA_BIND; DOMAIN; LIPID; METAL; MOD_RES; NP_BIND; SITE and ZN_FING |

| | | |
|---|---|---|
| 8.3. | Qualifier | organism |
| | Definition | Scientific name of the organism that provided the peptide |
| | Mandatory Value format | free text |
| | Language-dependent: this value may require translation for International/National/Regional procedures | |
| | Example | <INSDQualifier_value>Homo sapiens</INSDQualifier_value> |
| | Comment | The “organism” qualifier is mandatory for the source feature key. |

## Раздел 9: Таблицы генетических кодов

Таблица 7 представляет Таблицы Генетических Кодов, которые должны использоваться для преобразования кодирующих последовательностей. Значением для квалификатора trans_table является номер, присвоенный соответствующей таблице генетического кода. Там, где признак CDS описывается квалификатором преобразования, но не квалификатором transl_table, для преобразования по умолчанию используется Standard Code (стандартный код) – 1. (Примечание: Таблицы генетических кодов 7, 8, 15 и с 17 по 20 и 32 не существуют, поэтому эти числа отсутствуют в Таблице 7).

Table 7: Genetic Code Tables

| 1 - Standard Code |
|---|
| AAs = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG<br>Starts = ---M---------------M---------------M----------------------------<br>Base1 = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg<br>Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg<br>Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag |
| 2 - Vertebrate Mitochondrial Code |
| AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSS**VVVVAAAADDEEGGGG<br>Starts = --------------------------------MMMM---------------M------------<br>Base1 = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg<br>Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg<br>Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag |
| 3 - Yeast Mitochondrial Code |
| AAs = FFLLSSSSYY**CCWWTTTTPPPPHHQQRRRRIIMMTTTTNNKKSSRRVVVVAAAADDEEGGGG<br>Starts = --------------------------------MM--------------- M------------<br>Base1 = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg<br>Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg<br>Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag |
| 4 - Mold, Protozoan, Coelenterate Mitochondrial Code & Mycoplasma/Spiroplasma Code |
| AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG<br>Starts = --MM---------------M------------MMMM---------------M------------<br>Base1 = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg<br>Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg<br>Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag |
| 5 - Invertebrate Mitochondrial Code |
| AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSSSSVVVVAAAADDEEGGGG<br>Starts = ---M----------------------------MMMM---------------M------------<br>Base1 = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg<br>Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg<br>Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag |
| 6 - Ciliate, Dasycladacean and Hexamita Nuclear Code |
| AAs = FFLLSSSSYYQQCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG<br>Starts = --------------------------------M----------------------------<br>Base1 = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg<br>Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg<br>Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag |
| 9 - Echinoderm and Flatworm Mitochondrial Code |
| AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNNKSSSSVVVVAAAADDEEGGGG<br>Starts = --------------------------------M---------------M------------<br>Base1 = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg<br>Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg<br>Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag |

10 - Euplotid Nuclear Code

```
  AAs  =  FFLLSSSSYY**CCCWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------------------------------M----------------------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

11 - Bacterial l, Archaeal and Plant Plastid Code

```
  AAs  =  FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ---M---------------M------------MMMM---------------M------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

12 - Alternative Yeast Nuclear Code

```
  AAs  =  FFLLSSSSYY**CC*WLLLSPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ------------------M---------------M----------------------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

13 - Ascidian Mitochondrial Code

```
  AAs  =  FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSSGGVVVVAAAADDEEGGGG
Starts =  ---M------------------------------MM---------------M------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

14 - Alternative Flatworm Mitochondrial Code

```
  AAs  =  FFLLSSSSYYY*CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts =  ----------------------------------M----------------------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
Starts =  ----------------------------------M----------------------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

16 - Chlorophycean Mitochondrial Code

```
  AAs  =  FFLLSSSSYY*LCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------------------------------M----------------------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

21 - Trematode Mitochondrial Code

```
  AAs  =  FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts =  ----------------------------------M---------------M------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

22 - Scenedesmus obliquus Mitochondrial Code

```
  AAs  =  FFLLSS*SYY*LCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------------------------------M----------------------------
Base1  =  ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

23 - Thraustochytrium Mitochondrial Code

```
  AAs  =  FF*LSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  --------------------------------M--M---------------M------------
Base1  =  ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

24 - Pterobranchia Mitochondrial Code

```
  AAs  =  FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSSKVVVVAAAADDEEGGGG
Starts =  ---M---------------M---------------M---------------M------------
Base1  =  ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

25 - Candidate Division SR1 and Gracilibacteria Code

```
  AAs  = FFLLSSSSYY**CCGWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M ------------------------------- M---------------M ------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

26 - Pachysolen tannophilus Nuclear Code

```
  AAs  = FFLLSSSSYY**CC*WLLLAPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = --------------------M---------------M----------------------------
Base1  = ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

27 - Karyorelict Nuclear Code

```
  AAs  =  FFLLSSSSYYQQCCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ---------------*--------------------M----------------------------
Base1  =  ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

28 - Condylostoma Nuclearr Code

```
  AAs  =  FFLLSSSSYYQQCCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------**--*--------------------M----------------------------
Base1  =  ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

29 - Mesodinium Nuclear Code

```
  AAs  =  FFLLSSSSYYYYCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------------------------------M----------------------------
Base1  =  ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

30 - Peritrich Nuclear Code

```
  AAs  =  FFLLSSSSYYEECC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------------------------------M----------------------------
Base1  =  ttttttttttttttttccccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

31 - Blastocrithidia Nuclear Code

```
  AAs  =  FFLLSSSSYYEECCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts =  ----------**----------------------M----------------------------
Base1  =  tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2  =  ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3  =  tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

33 - Cephalodiscidae Mitochondrial UAA-Tyr Code

```
AAs = FFLLSSSSYYY*CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSSKVVVVAAAADDEEGGGG
Starts = ---M--------------M---------------M---------------M------------
Base1 = tttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaggggggggggggggg
Base2 = ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3 = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
```

[Приложение 2 следует]

# Приложение 2

ОПРЕДЕЛЕНИЕ ТИПА ДОКУМЕНТА ДЛЯ ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ (DTD)

Версия 1.3

*Пересмотр одобрен Комитетом по Стандартам ВОИС (КСВ) на своей одиннадцатой сессии 8 декабря 2023*

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Annex II of ST.26, Document Type Definition (DTD) for Sequence Listing

This entity may be identified by the PUBLIC identifier:
********************************************************************************
****
PUBLIC "-//WIPO//DTD SEQUENCE LISTING 1.3//EN" "ST26SequenceListing_V1_3.dtd"
********************************************************************************
****
* PUBLIC DTD URL

* http://www.wipo.int/standards/DTD/ST26SequenceListing_V1_3.dtd
********************************************************************************

Revision of Annex II to WIPO Standard ST.26 is submitted for approval by the Committee
on WIPO Standards (CWS) at its tenth session.

********************************************************************************
* CONTACTS
********************************************************************************
xml.standards@wipo.int


********************************************************************************
* NOTES
********************************************************************************
The sequence data part is a subset of the complete INSDC DTD V.1.5 that only covers
the requirements of WIPO Standard ST.26.

********************************************************************************
* REVISION HISTORY
********************************************************************************
2022-11-25: Comment related to filename approved at CWS/10 (no update to version
number)
2021-11-05: Revised Version 1.3 approved at CWS/9 (small edits to the comments)
2020-05-20: Version 1.3 approved at CWS/8.
Changes:
- Optional originalFreeTextLanguageCode attribute added to to allow applicants to
indicate the language of the free text in the original sequence listing.
- Optional nonEnglishFreeTextLanguageCode attribute added to to allow applicants to
indicate the language of the free text provided in the element .
- Optional id attribute added to INSDQualifier to facilitate comparison of language -
dependent qualifier values between sequence listings.
- Optional element added to element to allow applicants to type language-dependent
qualifiers in a non-English Language with the characters set forth in paragraph 40(a)
of the ST.26 main body document.

2018-10-19: Version 1.2 approved at CWS/6.
Changes:
<INDSQualifier*> changed to <INDSQualifier+> for alignment with business eneds and
advice from NCBI (an INSDFeature quals element (if present) should have one or more
INDSQualifier elements).

2017-06-02: Version 1.1 approved at the CWS/5
```

```
Changes:
Comments added to <INSDSeq_length>, <INSDSeq_division> and <INSDSeq_sequence> to
clarify the reason of the differences between the INSDC DTD v.1.5 and ST26 Sequence
Listing DTD V1_1.
******************************************************************************
2016-03-24: Version 1.0 adopted by the CWS/4Bis
2014-03-11: Final draft for adoption.
******************************************************************************

ST26SequenceListing
******************************************************************************
* ROOT ELEMENT
******************************************************************************
-->
<!ELEMENT ST26SequenceListing  ((ApplicantFileReference |
(ApplicationIdentification,ApplicantFileReference?)),
EarliestPriorityApplicationIdentification?,(ApplicantName,
ApplicantNameLatin?)?,(InventorName,InventorNameLatin?)?,InventionTitle+,SequenceTotal
Quantity,SequenceData+) >
<!--The elements ApplicantName and InventorName are optional in this DTD to facilitate
the conversion between various encoding schemes-->
<!--originalFreeTextLanguageCode:
The language code (see reference in paragraph 9 to ISO 639-1:2002) for the single
original language in which the language-dependent free text qualifiers
(NonEnglishQualifier_value) were prepared.
-->
<!--nonEnglishFreeTextLanguageCode:
The language code (see reference in paragraph 9 to ISO 639-1:2002) for the language in
which the language-dependent free text qualifiers (NonEnglishQualifier_value)
currently correspond.
-->
<!--fileName:
By default the file name will be set to the value provided for the project name in
WIPO Sequence. If the value is identical to the actual ST.26 XML filename, it should
be noted that Offices may enforce their requirements for the filename used which may
restrict which characters are allowable for submitted electronic files. It is also
acceptable for the value of the filename attribute and the actual file name to be
different. Please refer to the WIPO Sequence and ST.26 Knowledge Base for further
details on Offices' naming conventions for electronic files
--->
<!ATTLIST ST26SequenceListing
               dtdVersion  CDATA #REQUIRED
               fileName  CDATA #IMPLIED
               softwareName  CDATA #IMPLIED
               softwareVersion  CDATA #IMPLIED
               productionDate  CDATA #IMPLIED  >
               originalFreeTextLanguageCode CDATA #IMPLIED
               nonEnglishFreeTextLanguageCode CDATA #IMPLIED
>
<!--ApplicantFileReference
Applicant's or agent's file reference, mandatory if application identification not
provided.
-->
<!ELEMENT ApplicantFileReference  (#PCDATA) >

<!--ApplicationIdentification
Application identification for which the sequence listing is submitted, when
available.
-->
<!ELEMENT ApplicationIdentification  (IPOfficeCode?,ApplicationNumberText,
               FilingDate?) >

<!--EarliestPriorityApplicationIdentification
Identification of the earliest priority application, which Contains IPOfficeCode,
ApplicationNumberText and FilingDate elements.
-->
<!ELEMENT EarliestPriorityApplicationIdentification  (IPOfficeCode,
ApplicationNumberText,FilingDate?) >

<!—ApplicantName
```

```
The name of the first mentioned applicant in characters set forth in paragraph 40 a)
of the ST.26 main body document.
-->
<!--languageCode: Appropriate language code from ISO 639-1 - Codes for the
representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT ApplicantName  (#PCDATA) >
<!ATTLIST ApplicantName
               languageCode  CDATA #REQUIRED
>
<!—ApplicantNameLatin
Where ApplicantName is typed in characters other than those as set forth in paragraph
40 b), a translation or transliteration of the name of the first mentioned applicant
must also be typed in characters as set forth in paragraph 40 b).
-->
<!ELEMENT ApplicantNameLatin  (#PCDATA) >

<!--InventorName
Name of the first mentioned inventor typed in the characters as set forth in paragraph
40 a).-->
<!--languageCode: Appropriate language code from ISO 639-1 - Codes for the
representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventorName  (#PCDATA) >
<!ATTLIST InventorName
               languageCode  CDATA #REQUIRED
>
<!--InventorNameLatin
Where InventorName is typed in characters other than those as set forth in paragraph
40 b), a translation or transliteration of the first mentioned inventor may also be
typed in characters as set forth in paragraph 40 b).
-->
<!ELEMENT InventorNameLatin  (#PCDATA) >

<!--InventionTitle
Title of the invention typed in the characters as set forth in paragraph 40 a) in the
language of filing. A translation of the title of the invention into additional
languages may be typed in the characters as set forth in paragraph 40 a) using
additional InventionTitle elements. The title of invention should be between two to
seven words.
-->
<!--languageCode: Appropriate language code from ISO 639-1 - Codes
for the representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventionTitle  (#PCDATA) >
<!ATTLIST InventionTitle
               languageCode  CDATA    #REQUIRED
>
<!--SequenceTotalQuantity
Indicates the total number of sequences in the document.
Its purpose is to be quickly accessible for automatic processing.
-->
<!ELEMENT SequenceTotalQuantity  (#PCDATA) >

<!--SequenceData
Data for individual Sequence.
For intentionally skipped sequences see the ST.26 main body document.
-->
<!ELEMENT SequenceData  (INSDSeq) >
<!ATTLIST SequenceData
               sequenceIDNumber  CDATA    #REQUIRED  >

<!--IPOfficeCode
ST.3 code. For example, if the application identification is PCT/IB2013/099999, then
IPOfficeCode value will be IB.
-->
<!ELEMENT IPOfficeCode  (#PCDATA) >

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

```
-->
<!ELEMENT ApplicationNumberText  (#PCDATA)
>
<!--FilingDate
The date of filing of the patent application for which the sequence listing is
submitted ST.2 format (paragraphs 7 (a) and 11) "CCYY-MM-DD", using a 4-digit calendar
year, a 2-digit calendar month and a 2-digit day within the calendar month, e.g.,
2015-01-31
-->
<!ELEMENT FilingDate  (#PCDATA) >
<!--*********************************************************************************
* INSD Part
**********************************************************************************

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

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

This part is based on:

The International Nucleotide Sequence Database (INSD) collaboration.

INSDSeq provides the elements of a sequence as presented in the GenBank/EMBL/DDBJ-
style flatfile formats. Not all elements are used here.
-->
<!--INSDSeq
Sequence data. Changed INSD V1.5 DTD elements, INSDSeq_division and INSDSeq_sequence
from optional to mandatory per business requirements.
-->
<!ELEMENT INSDSeq  (INSDSeq_length,INSDSeq_moltype,INSDSeq_division,
               INSDSeq_other-seqids?,INSDSeq_feature-table?,INSDSeq_sequence) >

<!--INSDSeq_length
The length of the sequence.INSDSeq_length allows only integer.
-->
<!ELEMENT INSDSeq_length  (#PCDATA) >

<!--INSDSeq_moltype
Admissible values: DNA, RNA, AA
-->
<!ELEMENT INSDSeq_moltype  (#PCDATA) >

<!--INSDSeq_division
Indication that a sequence is related to a patent application. Must be populated with
the value PAT.
-->
<!ELEMENT INSDSeq_division  (#PCDATA) >

<!--INSDSeq_other-seqids
In the context of data exchange with database providers, the IPOs should populate for
each sequence the element INSDSeq_other-seqids with one INSDSeqid containing a
reference to the corresponding published patent and the sequence identification.
-->
<!ELEMENT INSDSeq_other-seqids  (INSDSeqid?) >

<!--INSDSeq_feature-table
Information on the location and roles of various regions within a particular sequence.
Whenever the element INSDSeq_feature-table is used, it must contain at least one
feature.
-->
<!ELEMENT INSDSeq_feature-table  (INSDFeature+) >

<!--INSDSeq_sequence
The residues of the sequence. The sequence must not contain numbers, punctuation or
whitespace characters.
-->
<!ELEMENT INSDSeq_sequence  (#PCDATA) >
```

```
<!--INSDSeqid
Intended for the use of IPOs in data exchange only.

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

Example:
pat|WO|2013999999|A1|123456

This represents the 123456th sequence from WO patent publication No. 2013999999 (A1)
-->
<!ELEMENT INSDSeqid  (#PCDATA) >

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

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

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

<!--INSDFeature_quals
List of qualifiers containing auxiliary information about a feature.
-->
<!ELEMENT INSDFeature_quals  (INSDQualifier+) >
<!--INSDQualifier
Additional information about a feature.
For coding sequences and variants see the ST.26 main body document.
-->
<!--id
Unique identifier for the INSDQualifier to facilitate comparison of versions of a
sequence listing specifically having language-dependent qualifier values in different
languages.
-->
<!ELEMENT INSDQualifier  (INSDQualifier_name,INSDQualifier_value?,
NonEnglishQualifier_value?)>
<!ATTLIST INSDQualifier
         id ID #IMPLIED
>
<!--INSDQualifier_name
Name of the qualifier.
-->
<!ELEMENT INSDQualifier_name  (#PCDATA) >
<!--INSDQualifier_value
Value of the qualifier. Where the qualifier is language-dependent its value must be in
the English language and typed with the characters set forth in paragraph 40 (b).
-->
<!ELEMENT INSDQualifier_value  (#PCDATA) >
<!--NonEnglishQualifier_value
Value of a language-dependent qualifier in a language that is not English and typed
with the characters set forth in paragraph 40 (a). The language is indicated with the
attribute nonEnglishFreeTextLanguageCode.
-->
<!ELEMENT NonEnglishQualifier_value (#PCDATA)>
```

[Приложение 3 следует]

**Приложение 3**

ОБРАЗЕЦ ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ (XML файл)

Версия 2.0

*Одобрено Комитетом по Стандартам ВОИС (КСВ)*
*на своей тринадцатой сессии 14 ноября 2025*

Приложение 3 доступно по адресу:
https://www.wipo.int/standards/en/xml_material/st26/st26-annex-iii-sequence-listing-specimen.xml

[Приложение 4 следует]

# Приложение 4

**Подмножество символов из Таблицы Unicode Basic Latin Code для использования в перечнях последовательностей с XML**

*Версия 2.0*

*Одобрено Комитетом по Стандартам ВОИС (КСВ)*
*на своей тринадцатой сессии 14 ноября 2025*

Символ амперсанда (0026) разрешен только как часть заранее определенного значения. Кавычка (0022), апостроф (0027), знак меньше чем (003C) и знак больше чем (003E) должны быть представлены их заранее определенными значениями. Кроме того, символ амперсанда (0026) должен быть представлен его заранее определенным значением при использовании его в качестве амперсанда в значении атрибута или содержания элемента.

| Unicode code point | Character | name |
|---|---|---|
| 0020 | | SPACE |
| 0021 | ! | EXCLAMATION MARK |
| 0022 | “ | QUOTATION MARK |
| 0023 | # | NUMBER SIGN |
| 0024 | $ | DOLLAR SIGN |
| 0025 | % | PERCENT SIGN |
| 0026 | & | AMPERSAND |
| 0027 | ‘ | APOSTROPHE |
| 0028 | ( | LEFT PARENTHESIS |
| 0029 | ) | RIGHT PARENTHESIS |
| 002A | * | ASTERISK |
| 002B | + | PLUS SIGN |
| 002C | , | COMMA |
| 002D | - | HYPHEN-MINUS |
| 002E | . | FULL STOP |
| 002F | / | SOLIDUS |
| 0030 | 0 | DIGIT ZERO |
| 0031 | 1 | DIGIT ONE |
| 0032 | 2 | DIGIT TWO |
| 0033 | 3 | DIGIT THREE |
| 0034 | 4 | DIGIT FOUR |
| 0035 | 5 | DIGIT FIVE |
| 0036 | 6 | DIGIT SIX |
| 0037 | 7 | DIGIT SEVEN |
| 0038 | 8 | DIGIT EIGHT |
| 0039 | 9 | DIGIT NINE |
| 003A | : | COLON |

| | | |
|---|---|---|
| 003B | ; | SEMICOLON |
| 003C | < | LESS - THAN -SIGN |
| 003D | = | EQUALS SIGN |
| 003E | > | GREATER - THAN -SIGN |
| 003F | ? | QUESTION MARK |
| 0040 | @ | COMMERCIAL AT |
| 0041 | A | LATIN CAPITAL LETTER A |
| 0042 | B | LATIN CAPITAL LETTER B |
| 0043 | C | LATIN CAPITAL LETTER C |
| 0044 | D | LATIN CAPITAL LETTER D |
| 0045 | E | LATIN CAPITAL LETTER E |
| 0046 | F | LATIN CAPITAL LETTER F |
| 0047 | G | LATIN CAPITAL LETTER G |
| 0048 | H | LATIN CAPITAL LETTER H |
| 0049 | I | LATIN CAPITAL LETTER I |
| 004A | J | LATIN CAPITAL LETTER J |
| 004B | K | LATIN CAPITAL LETTER K |
| 004C | L | LATIN CAPITAL LETTER L |
| 004D | M | LATIN CAPITAL LETTER M |
| 004E | N | LATIN CAPITAL LETTER N |
| 004F | O | LATIN CAPITAL LETTER O |
| 0050 | P | LATIN CAPITAL LETTER P |
| 0051 | Q | LATIN CAPITAL LETTER Q |
| 0052 | R | LATIN CAPITAL LETTER R |
| 0053 | S | LATIN CAPITAL LETTER S |
| 0054 | T | LATIN CAPITAL LETTER T |
| 0055 | U | LATIN CAPITAL LETTER U |
| 0056 | V | LATIN CAPITAL LETTER V |
| 0057 | W | LATIN CAPITAL LETTER W |
| 0058 | X | LATIN CAPITAL LETTER X |
| 0059 | Y | LATIN CAPITAL LETTER Y |
| 005A | Z | LATIN CAPITAL LETTER Z |
| 005B | [ | LEFT SQUARE BRACKET |
| 005C | \ | REVERSE SOLIDUS |
| 005D | ] | RIGHT SQUARE BRACKET |
| 005E | ^ | CIRCUMFLEX ACCENT |
| 005F | _ | LOW LINE |
| 0060 | ` | GRAVE ACCENT |
| 0061 | a | LATIN SMALL LETTER A |

| | | |
|---|---|---|
| 0062 | b | LATIN SMALL LETTER B |
| 0063 | c | LATIN SMALL LETTER C |
| 0064 | d | LATIN SMALL LETTER D |
| 0065 | e | LATIN SMALL LETTER E |
| 0066 | f | LATIN SMALL LETTER F |
| 0067 | g | LATIN SMALL LETTER G |
| 0068 | h | LATIN SMALL LETTER H |
| 0069 | i | LATIN SMALL LETTER I |
| 006A | j | LATIN SMALL LETTER J |
| 006B | k | LATIN SMALL LETTER K |
| 006C | l | LATIN SMALL LETTER L |
| 006D | m | LATIN SMALL LETTER M |
| 006E | n | LATIN SMALL LETTER N |
| 006F | o | LATIN SMALL LETTER O |
| 0070 | p | LATIN SMALL LETTER P |
| 0071 | q | LATIN SMALL LETTER Q |
| 0072 | r | LATIN SMALL LETTER R |
| 0073 | s | LATIN SMALL LETTER S |
| 0074 | t | LATIN SMALL LETTER T |
| 0075 | u | LATIN SMALL LETTER U |
| 0076 | v | LATIN SMALL LETTER V |
| 0077 | w | LATIN SMALL LETTER W |
| 0078 | x | LATIN SMALL LETTER X |
| 0079 | y | LATIN SMALL LETTER Y |
| 007A | z | LATIN SMALL LETTER Z |
| 007B | { | LEFT CURLY BRACKET |
| 007C | \| | VERTICAL LINE |
| 007D | } | RIGHT CURLY BRACKET |
| 007E | ~ | TILDE |

[Приложение 5 следует]

## Приложение 5

## Дополнительные требования для обмена данными (только для патентных ведомств)

*Версия 2.0*
*Одобрено Комитетом по Стандартам ВОИС (КСВ)*
*на своей тринадцатой сессии 14 ноября 2025*

В контексте обмена данными с базами данных провайдеров, ВИС для каждой последовательности должны заполнить элемент INSD_other-seqids с одним INSDSeqid, содержащим отсылку к соответствующему опубликованному патенту и идентификационный номер последовательности в следующем формате:

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

где {office code} - это код ведомства промышленной собственности, публикующего патентный документ в соответствии со стандартом ST.3; {document kind code} - это код для идентификации различных видов патентных документов, как указано в стандарте ST.16; {publication number} - это номер публикации заявки или патента; и {Sequence identification number} - это номер последовательности в данной заявке или патенте.

Пример:

pat|WO|2013999999|A1|123456.

Который будет преобразован в настоящий XML пример как:

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

Где “123456” - это 123456-я последовательность из публикации WO номер 2013999999 (A1).

[Приложение 6 следует]

# Приложение 6

## РУКОВОДСТВО С ПОЯСНИТЕЛЬНЫМИ ПРИМЕРАМИ

*Версия 2.0*

*Одобрено Комитетом по Стандартам ВОИС (КСВ)*
*на своей тринадцатой сессии 14 ноября 2025*

СОДЕРЖАНИЕ

ВВЕДЕНИЕ ……..…… ……..3.26.6.1
УКАЗАТЕЛЬ ПРИМЕРОВ …..…… ……..3.26.6.9
ПРИМЕРЫ. . ……..……..3.26.6.13
ДОПОЛНЕНИЕ… ……....……3.26.6.103

## ВВЕДЕНИЕ

Данный Стандарт указывает в качестве одной из своих целей “позволить заявителю единообразно составлять перечень последовательностей в заявке на патент, который был бы приемлем как для международных, так и для национальных или региональных процедур.” Цель настоящего Руководства позаботиться о том, чтобы все заявители и Ведомства интеллектуальной собственности (ВИС) понимали и принимали необходимые условия для включения и представления раскрытий последовательностей, так чтобы данная цель была достигнута.

Руководство состоит из данного Введения, указателя примеров, примеров раскрытий последовательностей и Дополнения, содержащего перечень последовательностей в XML с последовательностями, приведёнными в примерах. Данное Введение поясняет некоторые понятия и терминологию, используемые в последующем документе. Примеры иллюстрируют требования конкретных параграфов Стандарта, и каждый пример был обозначен номером наиболее релевантного параграфа. Некоторые примеры дополнительно иллюстрируют и другие параграфы, поэтому соответствующие перекрестные ссылки указаны в конце каждого примера. Указатель приводит номера страниц для примеров. Каждой последовательности в примере, которая или должна, или может быть включена в перечень последовательностей, был присвоен идентификационный номер последовательности (SEQ ID NO) и она представлена в XML формате в Дополнении к настоящему документу.

Для каждого примера любая пояснительная информация, представленная вместе с последовательностью, должна рассматриваться как полнота раскрытия, относящаяся к данной последовательности. Приведенные ответы учитывают только информацию, явно представленную в примере.

Рекомендации, представленные в этом документе, направлены на подготовку перечня последовательностей для представления на дату подачи патентной заявки. Подготовка перечня последовательностей для представления после даты подачи патентной заявки должна учитывать, может ли предоставленная информация рассматриваться ВИС, как дополнительный объект (признак) к первоначальному раскрытию. Следовательно, возможно, что руководство, представленное в этом документе, может быть неприменимо к перечню последовательностей, представленному после даты подачи заявки на патент.

*Подготовка перечня последовательностей*

Подготовка перечня последовательностей для заявки на патент предусматривает рассмотрение следующих вопросов:

1. Требует ли параграф 7 ST.26 включения конкретной раскрытой последовательности?

2. Если включения конкретной раскрытой последовательности не требуется, допускается ли включение такой последовательности согласно ST.26?

3. Если ST.26 требует или допускает включение конкретной раскрытой последовательности, то как следует эту последовательность представлять в перечне последовательностей?

Относительно первого вопроса, параграф 7 ST.26 (с некоторыми ограничениями) требует включения раскрытой последовательности в заявке на патент путем перечисления его остатков, где последовательность содержит десять или более конкретно определенных нуклеотидов или четыре или более конкретно определенных аминокислоты.

Что касается второго вопроса, параграф 8 ST.26 разрешает включение последовательностей, имеющих менее десяти конкретно определенных нуклеотидов или четырех конкретно определенных аминокислот.

Ясное понимание «перечисления его остатков» и «конкретно определенного» необходимо, чтобы ответить на эти два вопроса.

Что касается третьего вопроса, этот документ представляет раскрытие последовательностей, которое иллюстрирует множество сценариев в сочетании с подробным рассмотрением предпочтительных способов представления каждой последовательности или, если последовательность содержит несколько вариантов - «последовательность, имеющую наибольший охват», в соответствии с данным Стандартом. Поскольку невозможно рассмотреть каждый возможный сценарий нетипичной последовательности, в этом Руководстве делается попытка изложить рассуждения, лежащие в основе подхода к каждому примеру и порядок применения положений ST.26, так чтобы такие же рассуждения могли быть применены к другим сценариям последовательности, не приведенным в качестве примера.

Перечисление остатков

Параграф 3(c) ST.26 определяет «перечисление остатков» как раскрытие последовательности в патентной заявке путем перечисления по порядку каждого остатка последовательности, в которой (i) остаток представляет собой имя, аббревиатуру, символ или структуру; (ii) множество остатков представлено сокращенной формулой. Последовательность должна раскрываться в патентной заявке путём «перечисления остатков» с использованием условно принятых символов, которые представляют собой символы нуклеотидов, представленные в Таблице 1 Раздела 1 Приложения 1 к ST.26 (т.е. символы нижнего регистра или их эквиваленты верхнего регистра [1]) и символы аминокислот, представленные в Таблице 3 Раздела 3 Приложения 1 к ST.26 (то есть символы верхнего регистра или их эквиваленты в нижнем регистре[1]) или трехбуквенные сокращения, установленные далее в Приложении 6, Таблица B. В дальнейшем эти нуклеотидные и аминокислотные символы называются условно принятыми (конвенциональными). Символы, отличные от указанных в этих таблицах, далее называются «нерекомендованными».

[1] ПРИМЕЧАНИЕ: Несмотря на то, что в раскрытии заявки нуклеотиды или аминокислоты могут быть представлены символами как нижнего, так и верхнего регистра, для последовательности, включённой в перечень последовательностей для представления нуклеотидной последовательности должны использоваться только строчные буквы (см. п. 13 ST.26), а для представления аминокислотной последовательности должны использоваться только заглавные буквы (см. п. 26 ST.26).

Когда представление остатка описывается как эквивалент условно принятому символу или сокращению (например, “$Z_1$” означает “A”), или определённой последовательности условно принятых символов (например, “$Z_1$” означает “agga”), тогда последовательность интерпретируется так, как если бы она была раскрыта с использованием эквивалентов условно принятых символов или сокращений, чтобы определить, требует ли параграф 7 Стандарта ST.26 включения его в перечень последовательностей. Когда в качестве “нестандартного” символа нуклеотида используется неоднозначный символ (например, X1 = инозин или псевдоуридин), но не эквивалентный одному из условно принятых неоднозначных символов в Разделе 1, Таблица 1 (т. е. “m”, “r”, “w”, “s”, “y”, “k”, “v”, “h”, “d”, “b”, или “n”), то остаток интерпретируется как «n» остаток для определения, подразумевает ли параграф 7 ST.26 включение этой последовательности в перечень последовательностей. Аналогично, если в качестве “нестандартного” символа аминокислоты используется неоднозначный символ (например, «Z1» означает «A», «G», «S» или «T»), но не эквивалентный одному из условно принятых неоднозначных символов в Разделе 3, Таблица 3 (т. е, B, Z, J или X), то остаток интерпретируется как “X” остаток для определения, подразумевает ли параграф 7 ST.26 включение этой последовательности в перечень последовательностей.

Будьте внимательны при раскрытии последовательности рекомендуемым образом (условно принятыми символами), однако в противном случае, если последовательность раскрыта нерекомендуемым образом (“нестандартными” символами), может быть необходима консультация для объяснения раскрытия последовательности.

Даже если использовались условно принятые символы, по-прежнему должны даваться объяснения раскрытия последовательности для подтверждения того, что символ используется рекомендуемым образом. Если символ используется нерекомендуемым образом, такое пояснение необходимо, чтобы определить, подразумевается ли включение последовательности в перечень последовательностей по параграфу 7 Стандарта ST.26.

Конкретно определённые

В параграфе 3(k) ST.26 термин «конкретно определённые» означает любые нуклеотиды, кроме тех, которые представлены с помощью символа "n", и любые аминокислоты, кроме тех, которые представлены с помощью символа "X", из перечисленных в Приложении 1, где «n» и «X» используются условно принятым образом, как описано в Разделе 1, Таблица 1 (т. е. «a или c или g или t/u; «неизвестный» или «другой») и в Разделе 3, Таблица 3 (т.е. A или R или N или D или C или Q или E или G или H или I или L или K или M или F или P или O или S или U или T или W или Y или V, «неизвестный» или «другой») соответственно. Обсуждение выше, касающееся условно принятых символов или сокращений и их использования условно принятым или нерекомендованным образом, будет принято во внимание, чтобы определить, являются ли нуклеотид или аминокислота «конкретно определенными».

Наиболее представительная последовательность

Когда последовательность, которая удовлетворяет требованиям параграфа 7, раскрывается путем перечисления ее остатков только один раз в заявке, но описывается по-разному в нескольких вариантах осуществления, например в одном варианте осуществления, «X» в одном или нескольких местах может представлять собой любую аминокислоту, а в других вариантах осуществления «X» может представлять собой только ограниченное количество аминокислот, ST.26 требует включения в перечень последовательностей только одной последовательности, которая была раскрыта перечислением её остатков. В соответствии с параграфами 15 и 27, где такая последовательность содержит неоднозначные символы «n» или «X», «n» будет истолковываться как любой из немодифицированных нуклеотидов “a”, “c”, “g”, или “t/u”, а “X” будет истолковываться как любая из немодифицированных L-альфа-аминокислот “A, “R”, N”, “D”, “C”, “Q”,

“E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” или “V, соответственно, в отсутствие дальнейших пояснений.

Следовательно, единственной последовательностью, которую необходимо включить, является раскрытая последовательность, имеющая наибольший охват (наиболее представительная). Последовательностью, имеющей наибольший охват, является единственная последовательность, имеющая варианты остатков, которые представлены наиболее ограничивающими неоднозначными символами, включающими в себя наиболее раскрытые варианты осуществления. Аналогично, когда последовательность раскрывается путем перечисления ее остатков только один раз, но длина последовательности может варьироваться из-за изменения количества копий, самый длинный вариант последовательности считается последовательностью, имеющей наибольший охват. Например, рассмотрим последовательность, которая может содержать от 2 до 5 повторяющихся областей. В этом случае вариант с 5 копиями повтора является последовательностью, имеющей наибольший охват, и должен быть включен в перечень последовательностей. Однако включение дополнительных конкретных последовательностей настоятельно рекомендуется, где это целесообразно, например, тех, которые представляют дополнительные варианты осуществления, и являются ключевой частью изобретения. Включение дополнительных последовательностей позволяет проводить более детальный поиск и обеспечивает публикацию объекта, по которому испрашивается патент.

*Использование неоднозначного символа*

Надлежащее использование неоднозначного символа “n” в перечне последовательностей

Символ “n”

a. не должен использоваться для представления чего бы то ни было, кроме одного нуклеотида;

b. будет истолкован как один из следующих немодифицированных нуклеотидов “a”, “c”, “g”, или “t/u”, за исключением случаев, когда он используется с последующим описанием;

c. должен быть использован вместе с последующим описанием для представления любого из следующих нуклеотидов:

    i. природный или неприродный модифицированный нуклеотид, содержащий модифицированное нуклеиновое основание, который не может быть представлен любым другим символом из Приложения 1 (см. Глава 1; Таблица 1),

    ii. “неизвестный” нуклеотид, т.е., неустановленный, не раскрытый или не определённый;

    iii. «abasic site» (АП-сайт, апуриновый/апиримидиновый сайт).

d. должен быть использован для представления модифицированного нуклеотида, содержащего нуклеиновое основание, идентичное нуклеиновым основаниям нуклеотидов, перечисленных в перечне в Приложении 1, Раздел 1, Таблица 1; **и**

e. может быть использован для представления вариантов последовательности, то есть альтернатив, удалений, вставок или замен, где «n» является наиболее ограничивающим неоднозначным символом

Надлежащее использование неоднозначного символа “X” в перечне последовательностей

Символ “X”

a. не должен использоваться для представления чего бы то ни было, кроме одной аминокислоты;

b. будет истолкован как одна из немодифицированных L-альфа-аминокислот “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, or “V”, за исключением случаев, когда он используется с последующим описанием;

c. должен быть использован вместе с последующим описанием для представления любой из следующих аминокислот:

i. натуральная или ненатуральная модифицированная аминокислота, содержащая модифицированную боковую цепь, которая не может быть представлена иначе, как любым другим символом в Приложении 1 (см. Глава 3; Таблица 3),

ii. “неизвестная” аминокислота, т.е., неустановленная, не раскрытая, или не определённая; или

d. не должен быть использован для представления модифицированной аминокислоты с одиночной карбон содержащей боковой цепью, идентичной боковой цепи L-альфа-аминокислот, перечисленных в Приложении I, Раздел 1, Таблица 3; **и**

e. может быть использован для представления вариантов последовательности, то есть альтернатив, удалений, вставок или замен, где “X” является наиболее ограничивающим неоднозначным символом.

Аннотирование модифицированных остатков

Настоящий Стандарт требует, чтобы «модифицированные» остатки были аннотированы в соответствии с параграфом 17 для нуклеотидов и в соответствии с параграфом 30 для аминокислот.

Параграф 3(e) ST.26 определяет «модифицированную аминокислоту» как любую аминокислоту, описанную в параграфе 3(a), другую чем следующие L-альфа-аминокислоты: L-аланин, L-аргинин, L-аспарагин, L-аспарагиновая кислота, L-цистеин, L-глутамин, L-глутаминовая кислота, L-глицин, L-гистидин, L-изолейцин, L-лейцин, L-лизин, L-метионин, L- фенилаланин, L-пролин, L-пирролизин, L-серин, L-селеноцистеин, L-треонин, L-триптофан, L-тирозин или L-валин. Точно так же Стандарт определяет «модифицированный нуклеотид» как любой нуклеотид, как описано в параграфе 3(g), кроме дезоксиаденозина 5’-монофосфата, дезоксигуанозина 5’-монофосфата, дезоксицитидина 5’-монофосфата, дезокситимидина-5’-монофосфата, аденозина-5’-монофосфата, гуанозина-5’-монофосфата, цитидина 5’-монофосфата, или уридина 5’-монофосфата (ST.26, п. 3(f)).

Основываясь на приведенных выше определениях, модификации нуклеиновых оснований или сахаро-фосфатного остова нуклеиновой кислоты и модификации аминокислотных боковых цепей, или амидного остова пептида приводят к одному или нескольким «модифицированным нуклеотидам» или «модифицированным аминокислотам» соответственно. Следовательно, такие нуклеотиды и аминокислоты должны быть аннотированы. Примеры модификации остова включают фосфоротиоатные связи в нуклеиновых кислотах, аналоги нуклеотидов, таких как пептидно -

нуклеиновые кислоты (PNAs), гликоль - нуклеиновые кислоты (GNAs), N- замещенные глицины и D-аминокислоты.

Обратите внимание, что модификация концевой аминокислоты пептида или концевого нуклеотида нуклеиновой кислоты не обязательно приводит к «модифицированной аминокислоте» или «модифицированному нуклеотиду». Нужно посмотреть на концевую модификацию и определить, изменяет ли модификация химическую структуру остатка таким образом, что остаток попадает вне рамок исключений, изложенных в параграф 3(e) и 3(f). Например, С-концевой остаток, который связан со структурой (такой как часть разветвленной последовательности – см. пептид №2 в примере 7(b)-3) через условно принятую амидную связь, не считается «модифицированным остатком». Точно так же N-концевой остаток, который является амином, связанным с биотином, не считается «модифицированным остатком». В обоих сценариях структура остатка, участвующего в С-концевой или N-концевой связи, не изменена по сравнению с условно принятыми аминокислотами, указанными в параграфе 3(e) Стандарта.

Напротив, концевые модификации, которые изменяют химическую структуру остатка, считаются «модифицированными остатками» и должны быть аннотированы. Например, метилирование С-конца в примере 3(c)-1 действительно изменяет химическую структуру концевого остатка, так как метильная группа заменяет гидроксил, обычно присутствующий в альфа-карбоксильной группе. Следовательно, этот метилированный лизин должен быть помечен как «модифицированный остаток».

Обратите внимание, что заявитель должен будет оценить каждую модификацию концевого остатка в пронумерованной последовательности и определить, изменилась ли структура концевого остатка. Если структура модифицированного остатка отличается от условно принятых аминокислот или нуклеотидов, указанных в параграфах 3(e) и 3(f) Стандарта, то модификация должна быть аннотирована.

Наконец, всегда рекомендуется, чтобы заявители включали в свои перечни последовательностей как можно больше информации, чтобы представлять свои раскрытия как можно точнее. Таким образом, даже если модификация не требует аннотирования, её желательно включать.

Следует отметить, что аннотирование вариантов перечисляемой исходной последовательности должно соответствовать требованиям параграфов 93-100 ST.26. Модификации, которые раскрываются как варианты перечисляемой последовательности, могут не включаться в перечень последовательностей. Для определения термина аннотирование вариантов см. параграфы 93-95 ST.26.

Представление модифицированных остатков

ST.26 требует, чтобы модифицированный нуклеотид с нуклеиновым основанием, идентичным нуклеиновым основаниям, перечисленным в Приложении I, Раздел 1, Таблица 1, был представлен в последовательности с помощью символа для соответствующего немодифицированного нуклеотида. Аналогично, модифицированная аминокислота с единичной карбон-содержащей боковой цепью, которая идентична боковой цепи в L-альфа-аминокислоте, перечисленной в Приложении I, Раздел 3, Таблица 3, должна быть представлена в последовательности с помощью символа соответствующей немодифицированной аминокислоты. Все другие модифицированные нуклеотиды и аминокислоты должны быть представлены в перечне последовательностей с помощью символа для соответствующего немодифицированного остатка, когда это возможно (см. параграфы 16 и 29). Для модифицированных остатков, которые не имеют нуклеиновых оснований, идентичных нуклеиновым основаниям нуклеотидов, перечисленных в Приложении I, Раздел 1, таблица 1, или боковой цепи, идентичной боковой цепи в L-альфа-аминокислоте, перечисленной в Приложении I, Раздел 3, Таблица 3, представление с помощью соответствующих немодифицированных остатков является (см. параграф 4(d) “строго рекомендованным подходом, но не требованием” (т.е. “should” («следует»), а не “must” («должен»)). В таких случаях остаётся на усмотрение заявителя решить, будет ли модифицированный остаток представлен символом для соответствующего немодифицированного остатка или параметрами “n” или “X”.

Как общее эмпирическое правило - если нуклеиновое основание или аминокислотная боковая цепь модифицированы путем, который не сильно меняет структуру, тогда рекомендуется представление символом немодифицированного остатка. В общем, добавление фрагмента к нуклеиновому основанию или боковой цепи, такое, как путём метилирования, гликозилирования или ацетилирования, несущественно изменяет структуру нуклеинового основания или аминокислотной боковой цепи. Исходя из этого, метилированный аденин в перечне последовательностей следует предпочтительно представлять с помощью “a”, а не ”n”. Гликозилированный серин в перечне последовательностей следует предпочтительно представлять с помощью “S”, а не “X”. Однако, если модифицированное неклеиновое основание или аминокислотная боковая цепь очень отличны по структуре от любого немодифицированного остатка, тогда рекомендуется представление с помощью” n” или “X”. Например, норлейцин явялется изомером лейцина, и его боковая цепь представляет собой линейную структуру из 4 атомов углерода. Лейцин тоже содержит 4-х углеродную боковую цепь, но это – ответвление у второго углерода. Поэтому боковая цепь норлейцина имеет совершенно другую структуру, чем боковая цепь лицина. Соответственно, рекомендуется, чтобы норлейцин был представлен как “X” в перечне последовательностей, а не “L”.

То, как заявитель выберет способ обозначения модифицированного нуклеотида или аминокислоты, которые не обязательно должны быть представлены символом соответствующего немодифицированного остатка, будет определять, является ли этот остаток “конкретно определенным”. Нуклеотид является «конкретно определенным», когда он представлен чем-либо, кроме “n”, а аминокислота «конкретно определена», когда она представлена чем-либо, кроме «X» (см. ST.26, параграф 3(k)). Следовательно, 2'O-метиладенозин, представленный как «a» в последовательности, является конкретно определенным, тогда как норлейцин, представленный как «X» в последовательности, не является конкретно определённым.

*Таблица A – Условно принятые символы нуклеотидов и определения*

| Символ | Определения |
|---|---|
| a | аденин / Adenine |
| c | цитозин / Cytosine |
| g | гуанин / Guanine |
| t | тимин в ДНК/урацил в РНК<br>Thymine in DNA /Uracil in RNA (t/u) |
| m | a или c |
| r | a или g |
| w | a или t/u |
| s | c или g |
| y | c или t/u |
| k | g или t/u |
| v | a или c или g; не t/u |
| h | a или c или t/u; не g |
| d | a или g или t/u; не c |
| b | c или g или t/u; не a |
| n | a или c или g или t/u; “unknown” или “other” |

*Таблица B – Условно принятые символы аминокислот, трёхбуквенные коды и определения*

| Символ | 3-буквенный код | Определение |
|---|---|---|
| A | Ala | Аланин/Alanine |
| R | Arg | Аргинин/Arginine |
| N | Asn | Аспарагин/Asparagine |
| D | Asp | Аспарагиновая кислота (Аспартат)/ Aspartic acid (Aspartate) |
| C | Cys | Цистеин /Cysteine |
| Q | Gln | Глутаминовая кислота (Глутамат)/ Glutamic acid (Glutamate) |
| E | Glu | Глутамин /Glutamine |
| G | Gly | Глицин/Glycine |
| H | His | Гистидин/Histidine |
| I | Ile | Изолейцин/Isoleucine |
| L | Leu | Лейцин/Leucine |
| K | Lys | Лизин/Lysine |
| M | Met | Метионин/Methionine |
| F | Phe | Фенилаланин/Phenylalanine |
| P | Pro | Пролин/Proline |
| O | Pyl | Пирролизин/Pyrrolysine |
| S | Ser | Серин/Serine |
| U | Sec | Селеноцистеин/Selenocysteine |
| T | Thr | Треонин/Threonine |
| W | Trp | Триптофан/Tryptophan |
| Y | Tyr | Тирозин/Tyrosine |
| V | Val | Валин/Valine |
| B | Asx | Аспарагиновая кислота или Аспарагин/Aspartic Acid or Asparagine |
| Z | Glx | Глутамин или Глутаминовая кислота/Glutamine or Glutamic acid |
| J | Xle | Лейцин или Изолейцин/Leucine or Isoleucine |
| X | Xaa | A или R или N или D или C или Q или E или G или H или I или L или K или M или F или P или O или S или U или T или W или Y или V; "unknown" или "other" |

УКАЗАТЕЛЬ ПРИМЕРОВ

**Страница**

*Параграф 3(a) – определение “аминокислоты”* ……**13**

**Пример 3(a)-1: D-аминокислоты** ……**13**

**Пример 3(a)-2: Органическая молекула с аминогруппой и карбоксильной группой** **14**

*Параграф 3(c) – определение “перечисления остатков”* …… **15**

**Пример 3(c)-1: Перечисление аминокислот в химической структуре** …… **15**

**Пример 3(c)-2: Сокращенная формула для аминокислотной последовательности** …. **17**

*Параграф 3(g) – Определение “нуклеотид”* ……**18**

**Пример 3(g)-1: Нуклеотидная последовательность прервана C3 спейсером** …… **18**

**Пример 3(g)-2: Нуклеотидная последовательность с альтернативными остатками, включая C3 спейсер** …… **19**

**Пример 3(g)-3: Abasic site (AP-сайт)** …… **21**

**Пример 3(g)-4: Аналоги нуклеиновой кислоты** …… **22**

*Параграф 3(k) – Определение “конкретно определённый”*…… ..**23**

**Пример 3(k)-1: Неоднозначные символы нуклеотидов** …… **23**

**Пример 3(k)-2: Неоднозначный символ “n”, используемый как условно принятым, так и нерекомендованным образом** …… **24**

**Пример 3(k)-3: Неоднозначный символ “n” используемый нерекомендованным образом** …… **26**

**Пример 3(k)-4: Неоднозначные символы, отличные от “n” являются “конкретно определёнными”** …… **27**

**Пример 3(k)-5: Неоднозначное сокращение “Xaa” используется нерекомендованным образом** …… **28**

*Параграф 7(a) – Нуклеотидные последовательности обязательные в перечне последовательностей*……..….**29**

**Пример 7(a)-1: Разветвленная нуклеотидная последовательность** …… **29**

**Пример 7(a)-2: Линейная нуклеотидная последовательность, имеющая вторичную структуру** …… **32**

**Пример 7(a)-3: Неоднозначные символы нуклеотидов, используемые нерекомендуемым образом** …… **34**

**Пример 7(a)-4: Неоднозначные символы нуклеотидов, используемые нерекомендуемым образом** …… **35**

**Пример 7(a)-5: Нерекомендованные символы нуклеотидов** …… **36**

**Пример 7(a)-6: Нерекомендованные символы нуклеотидов** …… **37**

**Пример 7(a)-7: Инвертированные нуклеотиды I**…… ….**38**

**Пример 7(a)-8: Инвертированные нуклеотиды II**……**39**

*Параграф 7(b) – Аминокислотные последовательности, требуемые (обязательные) в перечне последовательностей*……..**42**

**Пример 7(b)-1: Четыре и более конкретно определённых аминокислоты ..... 42**

**Пример 7(b)-2: Разветвленная аминокислотная последовательность ..... 43**

**Пример 7(b)-3 Разветвленная аминокислотная последовательность..... 46**

**Пример 7(b)-4: Циклический пептид, содержащий разветвленную аминокислотную последовательность..... 48**

**Пример 7(b)-5: Циклический пептид, содержащий разветвленную аминокислотную последовательность …… ……...50**

*Параграф 11(a) – Нуклеотидная последовательность из двух цепочек – полностью комплементарная* ……..**52**

**Пример 11(a)-1: Нуклеотидная последовательность из двух цепочек – одинаковой длины..... 52**

*Параграф 11(b) – Двухцепочечная нуклеотидная последовательность – не полностью комплементарная* ……..**53**

**Пример 11(b)-1: Двухцепочечная нуклеотидная последовательность разной длины 53**

**Пример 11(b)-2: Двухцепочечная нуклеотидная последовательность – без сегмента спаривания оснований..... 55**

*Параграф 12 – Циклическая последовательность нуклеотидов*……**56**

**Пример 12-1: Циклическая последовательность нуклеотидов ……..56**

*Параграф 14 – Символ “t” рассматривается как урацил в RNA*…… ……**55**

**Пример 14-1: Символ “t” представляет урацил в RNA ..... 57**

*Параграф 27* – Должен быть использован наиболее ограничивающий неоднозначный символ…..**59**

**Пример 27-1: Сокращённая формула для аминокислотной последовательности……59**

**Пример 27-2: Сокращенная формула - менее четырех конкретно определенных аминокислот ..... 60**

**Пример 27-3: Сокращённая формула – четыре или более конкретно определённых аминокислот ..... 62**

*Параграф 28 – Аминокислотные последовательности, разделённые символами внутреннего разграничения*……**63**

**Пример 28-1: Кодирующая нуклеотидная последовательность и кодируемая аминокислотная последовательность..... 63**

*Параграф 29 – Представление «другой» аминокислоты*……..**65**

**Пример 29-1: Наиболее ограничивающий неоднозначный символ для “other” аминокислоты ..... 65**

**Пример 29-2: Использование соответствующей немодифицированной аминокислоты . 67**

**Пример 29-3: Пептоидный полимер I 69**

**Пример 29-4: Пептоидный полимер II 71**

**Пример 29-5: Бета-аминокислоты 75**

**Пример 29-6: Пептидно-пептоидный гибрид 77**

*Параграф 30 – Аннотирование модифицированной аминокислоты*……..……...**78**

**Пример 30-1 – Функциональный ключ “CARBOHYD”**..... **78**

**Пример 30-2: Посттрансляционно модифицированные аминокислоты**.....**79**

*Параграф 36 - Последовательности, содержащие участки с точным количеством смежных «n» или «X» остатков*……..**80**

**Пример 36-1: Последовательность с участком известного количества «X» остатков, представленная как единая последовательность** ..... **80**

**Пример 36-2: Последовательность с несколькими участками с известным количеством или диапазоном «X» остатков, представленных в виде одной последовательности 81**

**Пример 36-3: Последовательность с несколькими участками с известным количеством или диапазоном «X» остатков, представленных в виде одной последовательности 82**

*Параграф 37 - Последовательности, содержащие участки с неизвестным количеством «n» или «X» смежных остатков* ……**84**

**Пример 37-1: Последовательность с участками с неизвестным количеством «X» остатков не должна быть представлена как единая последовательность……..84**

**Пример 37-2: Последовательность с участками с неизвестным количеством «X» остатков не должна быть представлена как единая последовательность**..... **85**

*Параграф 55 – Нуклеотидная последовательность, которая содержит как ДНК, так и РНК фрагменты*…… .**87**

**Пример 55-1: Комбинированная молекула ДНК / РНК……...87**

*Параграф 89 – Функциональный ключ “CDS”*…… **88**

**Пример 89-1: Кодирующая нуклеотидная последовательность и кодируемая аминокислотная последовательность**..... **88**

**Пример 89-2: Местоположение признака, не предусмотренного (выходящего за пределы) в раскрытии последовательности**..... **89**

*Параграф 92 - Аминокислотная последовательность, кодируемая кодирующей последовательностью*……**91**

**Пример 92-1: Аминокислотная последовательность, кодируемая кодирующей последовательностью с интронами**..... **91**

*Параграф 93 - Исходная последовательность и вариант, перечисленные каждый своими остатками*…… ……...**93**

**Пример 93-1: Представление перечисляемых вариантов I** ..... **93**

**Пример 93-2: Представление перечисляемых вариантов II** ..... **94**

**Пример 93-3: Представление согласованной последовательности** ..... **95**

*Параграф 94 – Вариант последовательности, раскрытой как единая последовательность с перечисленными альтернативными остатками*……..**97**

**Пример 94-1: Представление одной последовательности с перечисленными альтернативными аминокислотами** ..... **97**

**Пример 94-2: Представление одиночной последовательности с перечисленными альтернативными аминокислотами, которые могут быть модифицированными аминокислотами**.....**98**

*Параграф 95 (a) - Вариант последовательности, раскрытой только посредством ссылки на исходную последовательность с несколькими независимыми вариациями*……..……**99**

**Пример 95 (a) -1: Представление варианта последовательности посредством аннотирования исходной последовательности ..... 99**

*Параграф 95 (b) - вариант последовательности, раскрытой только посредством ссылки на основную последовательность с множеством взаимозависимых вариаций*……...**101**

**Пример 95 (b) -1: Представление отдельных вариантов последовательностей с несколькими взаимозависимыми вариациями ..... 101**

ПРИМЕРЫ

*Параграф 3(a) - Определение “аминокислоты”*

**Пример 3(a)-1: D-аминокислоты**

Заявка на патент описывает следующую последовательность:

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

**Вопрос 1: Требует ли ST.26 включения последовательности(-тей)?**

**ДА.**

Параграф 3(a) Стандарта определяет «аминокислоту» как включающую «D-аминокислоты» и аминокислоты, содержащие модифицированные или синтетические боковые цепи. На основании этого определения, приведённый пептид содержит семь аминокислот. ”Lys” в позиции 3 и “**Gly**” в позиции 5 являются конкретно определенными. Параграф 29 данного Стандарта требует представления D-аминокислоты, имеющей одиночную карбон-содержащую боковую цепь, которая идентична L-альфа аминокислоте, содержащейся в Приложении I, Раздел 3, Таблица 3, символом соответствующей немодифицированной аминокислоты. Соответственно, D-Ala в позиции 1, **D-Glu** в позиции 2 и D-Met в позиции 6 должны быть представлены как A,” “E,” and “M”, соответственно, и также являются конкретно определенными. Следовательно, минимально, последовательность содержит 5 конкретно определенных аминокислот и должна быть включена в перечень последовательностей, как того требует параграф 7(b) ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

“Nle” в позиции 4 и “D –Nle” в позиции 7 не имеют боковых цепей, идентичных боковой цепи в L-альфа аминокислоте, содержащейся в Приложении I, Раздел 3, Таблица 3. Следовательно, норлейцеин и D–норлейцеин могут быть представлены в последовательности символом “X” или наиболее подходящим аминокислотным символом.

Параграф 25 указывает, что, когда аминокислотные последовательности имеют циклическую (кольцевую) конфигурацию и кольцо состоит исключительно из аминокислотных остатков, связанных амидными связями, заявитель должен выбрать аминокислоту с положением остатка номер 1. Соответственно, последовательность может быть представлена как:

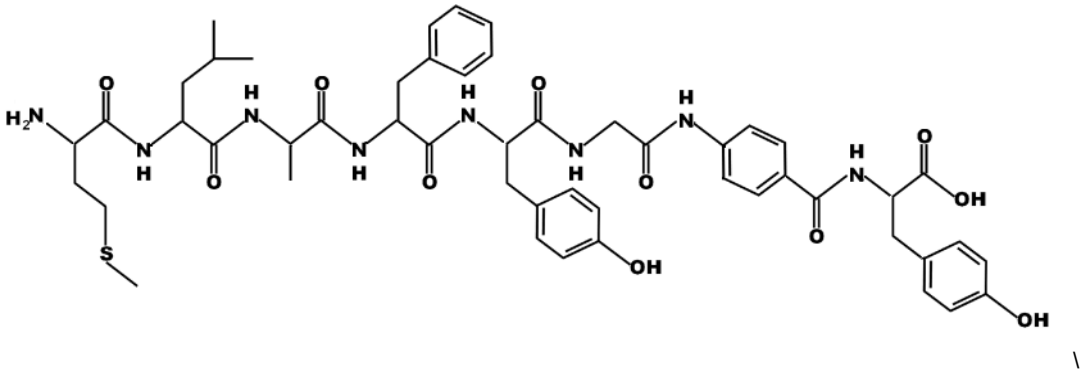
AEKXGMX (SEQ ID NO: 1)

или, иначе, с любой другой аминокислотой в последовательности с положением остатка номер 1. Для каждой D-аминокислоты должен быть указан функциональный ключ «SITE» и квалификатор «note» с полным, несокращённым названием D-аминокислоты в качестве значения квалификатора, например, D-аланин и D-норлейцин. Кроме того, функциональный ключ «SITE» и квалификатор «note» должны быть обеспечены аббревиатурой для L-норлейцина в качестве значения квалификатора, то есть «Nle», как указано в Таблице 4 Раздела 4 Приложения I. Наконец, должны быть приведены функциональный ключ “REGION” и квалификатор “note”, указывающие, что пептид является циклическим (кольцевым).

**Соответствующие параграфы ST.26: 3(a),** 7(b), 25, 26, 29, 30 и 31

**Пример 3(a)-2: Органическая молекула с аминогруппой и карбоксильной группой**

Заявка на патент раскрывает пептидную последовательность Met-Leu-Ala-Phe-Tyr-Gly-(4-aminobenzoic acid)-Tyr, имеющую следующую структуру:



**Вопрос 1: Требует ли ST.26 включения последовательности (-ей)?**

**ДА.**

Параграф 7(b) требует включения любой неразветвленной последовательности, которая содержит четыре или более конкретно определенных аминокислот, где аминокислоты формируют единый пептидный остов. Поэтому в этой последовательности семь конкретно определенных аминокислот. Параграф 7(b) ST.26 требует, чтобы последовательность была включена в список последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

4-аминобензойная кислота, представленная в 7-й позиции, рассматривается как аминокислота в смысле параграфа 3(a). Параграф 3(a) указывает, что аминокислота является “органической молекулой, содержащей аминогруппу и карбоксильную группу”. 4-аминобензойная кислота, представленная в 7-й позиции, является органической молекулой, содержащей аминогруппу и карбоксильную группу, следовательно, она соответствует определению аминокислоты. Так как 4-аминобензойная кислота соединена с соседними аминокислотами амидными связями как часть единого пептидного остова, раскрытый пептид из восьми остатков должен быть представлен в перечне последовательностей как единая последовательность.

4-аминобензойная кислота является “модифицированной аминокислотой”, которая не имеет боковой цепи, которая была бы идентична боковой цепи какой-либо из L-альфа аминокислот, перечисленных в перечне Приложения I, Раздел 3, Таблица 3. В соответствии с параграфом 29 СТ.26 модифицированную аминокислоту следует представлять в последовательности как соответствующую немодифицированную аминокислоту всякий раз, когда это возможно. Символ “X” следует использовать только для представления модифицированной кислоты, когда ни один из других символов в Приложении I, Раздел 3, Таблица 3 не подходит. Соответственно, 4-аминобензойная кислота должна быть представлена в последовательности с помощью символа “X”, и последовательность должна быть представлена как

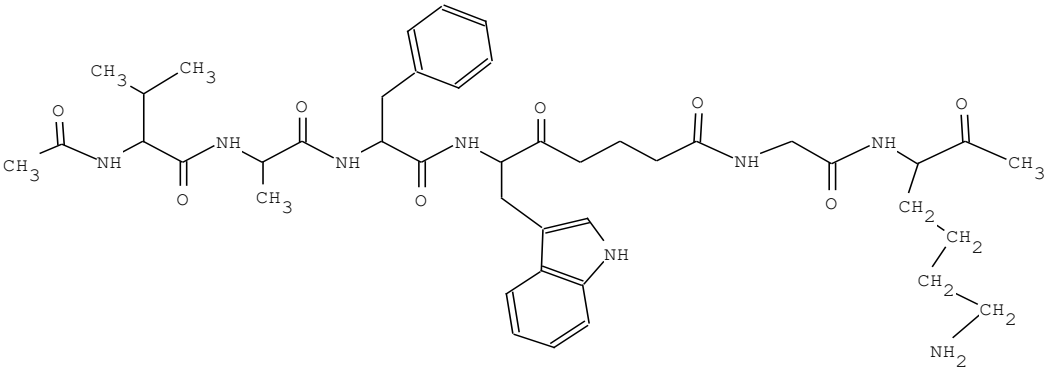
MLAFYGXY (SEQ ID NO: 103).

4-аминобензойная кислота в позиции7 должна быть описана. в соответствии с параграфом 29, используя функциональный ключ “SITE“ и обязательный квалификатор “note“ со значением, которое является полным несокращенным наименованием модифицированной аминокислоты, в данном случае “4-аминобензойная кислота“.

**Соответствующие параграфы ST.26: 3(a),** 3(e), 7(b), 29, 30 и 31

*Параграф 3(c) – определение “перечисления остатков”*

**Пример 3(c)-1: Перечисление аминокислот в химической структуре**



**Вопрос 1: Требует ли ST.26 включения последовательности (-тей)?**

**ДА**

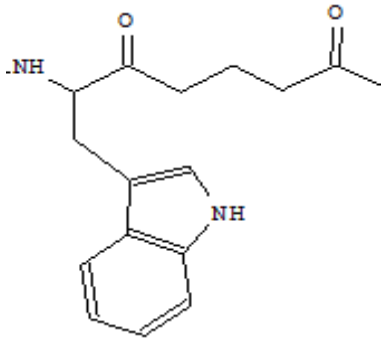
Представленный пептид, проиллюстрированный как структура, содержит, по меньшей мере, четыре конкретно определенные аминокислоты. Следовательно, последовательность должна быть включена в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность следует представить как:

VAFWGK (SEQ ID NO: 2)

Боковая цепь четвертого остатка идентична боковой цепи в L-альфа-аминокислоте, указанной в перечне Приложения I, Раздел 3, Таблица 3.



В соответствии с параграфом 29 она должна быть представлена в последовательности, используя символ соответствующей немодифицированной аминокислоты, которая является триптофаном (символ W), Как предписывается параграфом 30, этот модифицированный остаток должен быть далее описан с использованием функционального ключа «SITE» совместно с квалификатором «note». Квалификатор «note» предоставляет полное, несокращенное название модифицированного триптофана на позиции 4 перечисляемого пептида, например “6-amino-7-(1H-indol-3-yl)-5-oxoheptanoic acid”.

Метилирование С-конца данного пептида изменяет химическую структуру концевого лизина, поскольку -OH на конце заменяется на -CH3. Из-за этого структурного изменения лизин в последовательности считается «модифицированной амниокислотой».

Соответственно, функциональный ключ SITE» и квалификатор “note” требуются для указания на метилирование C-конца. Однако валин на N конце не считается «модифицированной аминокислотой», поскольку добавленная ацетильная группа валина включает обычную пептидную связь. Ацетилирование не изменяет структуру валина. Соответственно, добавленные функциональный ключ «SITE» и квалификатор «note» должны быть включены для указания ацетилирования N конца.

**Соответствующие параграфы ST.26: 3(a),** 3(c), 7(b), 29, 30 и 31

**Пример 3(c)-2: Сокращенная формула для аминокислотной последовательности**

$(G_4z)_n$

Где G= Глицин, z = любая аминокислота, а переменная n могут быть любым целым числом.

**Вопрос 1: Требует ли ST.26 включения последовательности (-тей)?**

**ДА**

Описание указывает, что «n» может быть «любым целым числом»; следовательно, наиболее охватывающий вариант $(G_4z)_n$ является неопределенным. Поскольку «n» является неопределенным, пептид из формулы не может быть расширен до определенной длины, и, следовательно, должна быть рассмотрена нерасширенная формула.

Представленный пептид с нерасширенной формулой («n» = 1) содержит четыре конкретно определенные аминокислоты, каждая из которых представляет собой Gly, и символ «z». Условно «Z» является символом «глютамина или глутаминовой кислоты»; однако, пример определяет «z» как «любая аминокислота». В соответствии с ST.26 аминокислота, которая конкретно не определена, обозначена как «X». На основании этого анализа представленный пептид, то есть GGGGX, содержит четыре глициновых остатка, которые перечислены и конкретно определены. Таким образом, параграф 7(b) ST.26 требует включения последовательности в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

В последовательности используется “нестандартный” символ «z», определение которого должно быть установлено из описания (см. Введение к этому документу). Поскольку «z» определяется как любая аминокислота, рекомендуемым символом, используемым для обозначения этой аминокислоты, является «X». Следовательно, последовательность должна быть представлена в виде одной последовательности:

GGGGX (SEQ ID NO: 4)

и снабжена комментарием с функциональным ключом REGION, местоположением признака “>5” (соответствует >5), с квалификатором note со значением «Вся последовательность аминокислот 1-5 может повторяться один или несколько раз».

Согласно параграфу 27, «X» будет истолковываться как любая из немодифицированных L-альфа-аминокислот «A», «R», «N», «D», «C», «Q», «E», «G», «H», «I», «L», «K», «M», «F», «P», «O», «S», «U», «T», «W», «Y» или «V», за исключением случаев, когда он используется с последующим описанием в таблице характеристик. В данном примере «X» представляет «любую аминокислоту», что шире в охвате, чем подразумеваемое значение «X», установленное далее в параграфе 27, поскольку «любая аминокислота» охватывает модифицированные или другие аминокислоты. Поэтому «X» в данной последовательности должен быть аннотирован с помощью функционального ключа VARIANT и квалификатора note со значением: «X может быть любой аминокислотой».

**ПРЕДУПРЕЖДЕНИЕ:** Выбранное представление последовательности, указанной выше, направлено на предоставление перечня последовательностей на дату подачи патентной заявки. К перечню последовательностей, представленному после даты подачи патентной заявки, такое же представление может не применяться, поскольку необходимо рассмотреть вопрос о том, может ли предоставленная информация быть рассмотрена IPO, на предмет добавления объекта (признака) к первоначальному раскрытию.

**Соответствующие параграфы ST.26: 3(c),** 7(b) и27

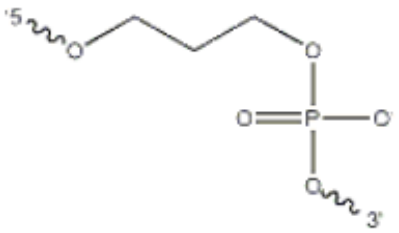
*Параграф 3(g) - Определение “нуклеотид”*

**Пример 3(g)-1: Нуклеотидная последовательность прервана C3 спейсером**

В патентной заявке раскрыта следующая последовательность:

atgcatgcatgcncggcatgcatgc

где n = C3 спейсер со следующей структурой:




**Вопрос 1: Требует ли ST.26 включения последовательности (-ей)?**

**ДА**

Представленная последовательность содержит два сегмента конкретно определенных нуклеотидов, разделенных спейсером C3.

Спейсер C3 не является нуклеотидом в соответствии с параграфом 3 (g); условно принятый символ «n» используется нерекомендуемым образом (см. Введение к этому документу). Следовательно, каждый сегмент представляет собой отдельную нуклеотидную последовательность. Поскольку каждый сегмент содержит более 10 конкретно определенных нуклеотидов, оба должны быть включены в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Каждый сегмент должен быть включён в перечень последовательностей, как отдельная последовательность, каждая со своим собственным идентификационным номером последовательности:

atgcatgcatgc (SEQ ID NO: 5)

cggcatgcatgc (SEQ ID NO: 6)

В каждом сегменте цитозин, который присоединен к спейсеру C3, должен быть дополнительно описан в таблице характеристик с использованием функционального ключа «misc_feature» и квалификатора «note». Значение квалификатора «note», которое является «свободным текстом» (“free text”), должно указывать на наличие спейсера, который присоединен к другой нуклеиновой кислоте, и идентифицировать спейсер либо по его полному не сокращенному химическому названию, либо по его общепринятому названию, например, C3 спейсер.

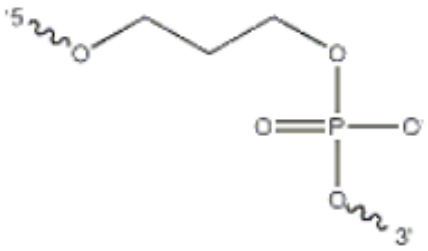
**Соответствующие параграфы ST.26: 3(g),** 7(a), и 15

**Пример 3(g)-2: Нуклеотидная последовательность с альтернативными остатками, включая C3 спейсер**

В патентной заявке раскрыта следующая последовательность:

atgcatgcatgcncggcatgcatgc

Где n = c, a, g, или C3 спейсер со следующей структурой:




**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

В представленной последовательности имеется 24 конкретно определенных остатка, разделённых переменной «n». Пояснение в раскрытии последовательности необходимо принять во внимание, чтобы определить, используется ли «n» рекомендуемым или нерекомендуемым образом (см. Введение к этому документу).

В раскрытии указывается, что n = c, a, g или спейсер C3. «n» является условно принятым символом, используемым нерекомендуемым способом, поскольку он описан как включающий спейсер C3, который не соответствует определению нуклеотида. Символ «n» также описывается как включающий «c», «a» или «g»; следовательно, ST.26 требует включения 25 нуклеотидной последовательности в перечень последовательностей. Поскольку два сегмента, разделенных спейсером C3, являются различными последовательностями из 25 нуклеотидной последовательности, то могут быть включены две 12 нуклеотидные последовательности.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

В примере указывается, что «n = c, a, g или спейсер C3». Как описано выше, спейсер C3 не является нуклеотидом. Согласно параграфу 15, символ «n» не должен использоваться для обозначения чего-либо, кроме нуклеотида; поэтому символ «n» не может представлять спейсер C3 в перечне последовательностей.

В параграфе 15 также указывается, что в тех случаях, когда применим неоднозначный символ, следует использовать наиболее ограничивающий символ. Символ «v» обозначает «a или c или g» в соответствии с Приложением I, Разделом 1, Таблицей 1, и является более ограничительным, чем «n».

Если переменная «n» в этом примере представляет собой c, a или g, то единственная последовательность, перечисленная её остатками, включает в себя наиболее раскрытые варианты осуществления и, следовательно, является самой охватывающей последовательностью (см. Введение к этому документу), которая должна быть включена в перечень последовательностей:

atgcatgcatgcvcggcatgcatgc (SEQ ID NO: 7)

Настоятельно рекомендуется включать любые дополнительные последовательности, существенные для раскрытия или формулы изобретения, как описано во введении к этому документу.

Если в данном примере переменная «n» представляет собой спейсер C3, последовательность можно рассматривать как два отдельных сегмента конкретно определенных нуклеотидов с каждой стороны от переменной «n», то есть atgcatgcatgc (SEQ ID NO: 8); и cggcatgcatgc (SEQ ID NO: 9). Если это существенно для раскрытия или формулы изобретения, эти две последовательности также должны быть включены в перечень последовательностей, каждая со своим собственным идентификационным номером последовательности.

Цитозин в каждом сегменте, который присоединен к спейсеру C3, должен быть дополнительно описан в таблице характеристик с использованием функционального ключа «misc_feature» и квалификатора «note». Значение квалификатора «note», которое представляет собой “free text” («свободный текст»), должно указывать на наличие спейсера, который присоединен к другой нуклеиновой кислоте, и идентифицировать спейсер либо по его полному не сокращенному химическому названию, либо по общепринятому названию, например, C3 спейсер.

**ПРЕДУПРЕЖДЕНИЕ:** Выбранное представление последовательности, указанной выше, направлено на предоставление перечня последовательностей на дату подачи патентной заявки. К перечню последовательностей, представленному после даты подачи патентной заявки, такое же представление может не применяться, поскольку необходимо рассмотреть вопрос о том, может ли предоставленная информация быть рассмотрена IPO, на предмет добавления объекта (признака) к первоначальному раскрытию.

**Соответствующие параграфы ST.26: 3(g)**, 7(a), и 15

**Пример 3(g)-3: AP-сайт (Abasic site)**

В патентной заявке описана следующая последовательность:

gagcattgac-AP-taaggct

Где AP это AP-сайт (abasic site - участок с удаленными азотистыми основаниями)

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Конкретно определенные остатки представленной последовательности разделяются посредством abasic site. 5 ' позиция abasic site содержит 10 нуклеотидов, а 3' позиция abasic site содержит 7 нуклеотидов. Параграф 3 (g) (ii) (2) определяет abasic site как «нуклеотид», когда он является частью нуклеотидной последовательности. Следовательно, abasic site в этом примере считается «нуклеотидом» для определения того, должна ли последовательность быть включена в перечень последовательностей и каким образом. Соответственно, остатки на каждой позиции abasic site являются частью единой пронумерованной последовательности, содержащей всего 18 нуклеотидов, 17 из которых конкретно определены. Следовательно, последовательность должна быть включена как одна последовательность в перечень последовательностей, как того требует параграф 7(a) ST.26.

**Вопрос 3 Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность должна быть включена в перечень последовательностей как:

gagcattgacntaaggct (SEQ ID NO: 10)

Abasic site должен быть представлен как «n» и должен быть дополнительно описан в таблице характеристик. Предпочтительным средством описания является функциональный ключ «updated_base» и обязательный квалификатор «mod_base» со значением «OTHER». Должен быть включен квалификатор “note”, который описывает модифицированное основание как abasic site.

**Соответствующие параграфы ST.26: 3(g),** 7(a), и 17

**Пример 3(g)-4: Аналоги нуклеиновой кислоты**

В заявке на патент раскрыта следующая последовательность гликоль нуклеиновой кислоты (GNA):

$PO_4$-tagttcattgactaaggctccccattgact-OH

Где левый конец последовательности имитирует 5’-конец ДНК последовательности.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Отдельные остатки, которые входят в состав последовательности GNA, считаются модифицированными нуклеотидами согласно параграфам 3(f) и 3 (g) (i) (2) ST.26. Параграф 16 ST.26 требует, чтобы модифицированный нуклеотид, который имеет нуклеиновую основу, идентичную нуклеиновой основе в немодифицированном нуклеотиде, указанном в Приложении I, Раздел 1, Таблица 1, представляется в последовательности символом соответствующего немодифицированного нуклеотида, делая каждый остаток конкретно определенным.

Соответственно, последовательность имеет более десяти перечисленных и «конкретно определенных» нуклеотидов и должна быть включена в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательности GNA не имеют 5'-конца и 3'-конца, а вернее 3'-конца и 2'-конца. 3'-конец, который обычно обозначается как имеющий концевую фосфатную группу, соответствует 5'-концу ДНК или РНК. (Обратите внимание, что другие аналоги нуклеиновых кислот могут по-разному соответствовать 5'-концу и 3'-концу ДНК и РНК.) Согласно параграфу 11, она должна быть включена в перечень последовательностей «в направлении слева направо, что имитирует направление от 5'-конца к 3'-концу ". Следовательно, она должна быть включена в перечень последовательностей как:

tagttcattgactaaggctccccattgact (SEQ ID NO: 11)

Каждый нуклеотид из последовательности должен быть описан в таблице характеристик с использованием функционального ключа «modified _base» и обязательного квалификатора «mod_base» с аббревиатурой «OTHER». Квалификатор “note” должен быть включен с полным без сокращения названием модифицированных нуклеотидов. В соответствии с параграфом 18 ST.26 так как все модифицированные нуклеотиды из последовательности разделяют одну и ту же основную часть, участок может быть совместно описан в таблице характеристик с “1..30“ как указателем местоположения, функциональный ключ «modified_base», обязательный квалификатор «mod_base» с аббревиатурой «OTHER» и квалификатором “note“ со значением наиболее строгого названия химического вещества без сокращений, которое охватывает все из модифицированных нуклеотидов в данном участке, такое как «гликолевые нуклеиновые кислоты» или «2,3-дигидроксипропильные нуклеозиды».

**Соответствующие параграфы ST.26:** 3(f), **3(g)**, 7(a), 11, 16, 17, 18, 22, 65, и 66

*Параграф 3(k) - Определение “конкретно определённый”*

**Пример 3(k)-1: Неоднозначные символы нуклеотидов**

5’ NNG KNG KNG K 3’

N и K являются IUPAC-IUB неоднозначными кодами

**Вопрос 1: Требует ли ST.26 включения последовательности(-ей)?**

**НЕТ**

Неоднозначные коды IUPAC-IUB соответствуют списку символов нуклеотидов, определенному в Приложении I, Раздел 1, Таблица 1. В соответствии с параграфом 3 (k) условно определенным нуклеотидом является любой нуклеотид, перечисленный в Приложении I, отличный от тех, которые представлены символом «n». Следовательно, «K» и «G» представляют собой конкретно определенные нуклеотиды, а «N» не является конкретно определенным нуклеотидом.

Представленная последовательность не имеет десяти или более конкретно определенных нуклеотидов и, следовательно, согласно параграфу 7(a) ST.26 не должна включаться в перечень последовательностей.

**Вопрос 2: Разрешает ли ST.26 включение последовательности(ей)?**

**ДА**

Согласно параграфу 8 «Перечень последовательностей может включать последовательности, имеющие менее десяти конкретно определенных нуклеотидов…».

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

При включении последовательность следует представлять в перечне последовательностей как:

nngkngkngk (SEQ ID NO: 104)

Обратите внимание, что значение по умолчанию “n”, указанное в параграфе 15, соответствует коду неоднозначности IUPAC-IUB для “n”.

**Соответствующие параграфы ST.26: 3(k)**, 7(a), 8, и 13

**Пример 3(k)-2: Неоднозначный символ “n” используемый как условно принятым, так и нерекомендованным образом**

В заявке раскрыта искусственная последовательность: 5’-AATGCCGGAN-3’. Раскрытие также указывает:

(i) в одном варианте, N – любой нуклеотид;

(ii) в одном варианте, N необязательно, но предпочтительно G;

(iii) в одном варианте, N равно K;

(iv) в одном варианте, N равно C.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**НЕТ**

Представленная последовательность содержит 9 конкретно определённых нуклеотидов и “N.” Необходимо принять во внимание пояснения к последовательности в раскрытии, чтобы установить используется ли символ “N” рекомендованным образом (см. Введение к этому документу).

Рассмотрение раскрытых вариантов осуществления (i) - (iv) представленной последовательности показывает, что наиболее охватывающим вариантом «N» является «любой нуклеотид». В наиболее охватывающем варианте осуществления «N» в представленной последовательности используется обычным способом.

В некоторых вариантах «N» описывается как конкретно определенный остаток (то есть «N представляет собой C» в части (iv)). Однако только наиболее охватывающий вариант (то есть «N представляет собой любой нуклеотид») учитывается при определении, должна ли последовательность быть включена в перечень последовательностей. Таким образом, представленная последовательность, которая должна быть оценена, это

5’-AATGCCGGAN-3 ’.

На основании этого анализа представленная последовательность, то есть AATGCCGGAN, не содержит десяти конкретно определенных нуклеотидов. Следовательно, параграф 7 (a) ST.26 не требует включения последовательности в перечень последовательностей, несмотря на то, что в некоторых вариантах осуществления «n» также определяется как конкретные нуклеотиды.

**Вопрос 2: Разрешает ли ST.26 включение последовательности(ей)?**

**ДА**

Последовательность “AATGCCGGAN” может быть включена в перечень последовательностей.

Альтернативно или в дополнение, альтернативная последовательность может быть включена в перечень последовательностей, если «N» заменить на конкретно определённый нуклеотид.

**Вопрос 3 Как следует включать эту последовательность(-ти) в перечень последовательностей?**

При включении последовательность, имеющую меньше чем десять конкретно определенных нуклеотидов, следует представлять в перечне последовательностей как:

aatgccggan (SEQ ID NO: 105).

В соответствии с параграфом 15 “n“ будет трактоваться как что-либо из немодифицированных “a”, “c”, “g” or “t/u”, за исключением случаев, когда далее используется описание в таблице характеристик.

В данном примере “n“ представляет «любой нуклеотид», охват которого шире, чем по умолчанию определяется для “n“ в параграфе 15, т.к. «любой нуклеотид» охватывает модифицированные или другие нуклеотиды. Поэтому “n“ должно быть аннотировано с использованием функционального ключа “misc_difference” и квалификатора note“, который описывает “n“ как «любой нуклеотид». Квалификатор “note“ может также относиться к предпочтительным вариантам выбора: «“n“ может быть любым нуклеотидом, предпочтительно g, k или c».

Включение дополнительных последовательностей, представляющих разные варианты, которые являются основной частью изобретения, настоятельно рекомендуется. Включение этих последовательностей позволяет проводить более качественный поиск и обеспечивает официальное информирование об объекте, по которому испрашивается патент.

Для приведенного выше примера настоятельно рекомендуется включить в перечень последовательностей следующие три дополнительные последовательности, каждая из которых имеет собственный идентификационный номер последовательности:

aatgccggag (SEQ ID NO: 12)

aatgccggak (SEQ ID NO: 13)

aatgccggac (SEQ ID NO: 14)

Если какая-либо из трёх вышеуказанных последовательностей включена в перечень последовательностей, нуклеотид, который заменяет «n», должен быть аннотирован с помощью функционального ключа “misc_difference” с обязательным квалификатором “note“ или «replace» для описания возможных альтернатив.

**ПРЕДУПРЕЖДЕНИЕ:** Предпочтительное представление последовательностей, указанное выше как SEQ ID NOos: 12-14, направлено на предоставление перечня последовательностей на дату подачи патентной заявки. К перечню последовательностей, представленному после даты подачи патентной заявки, такое же представление может не применяться, поскольку необходимо рассмотреть вопрос о том, может ли предоставленная информация быть рассмотрена IPO на предмет добавления признака изобретения к первоначальному раскрытию.

**Соответствующие параграфы ST.26: 3(k)**, 7(a), 8, 13 и 15

**Пример 3(k)-3: Неоднозначный символ “n” используемый нерекомендованным образом**

Заявка раскрывает последовательность: 5’-aatgttggan-3’

Где n есть c

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

В соответствии с параграфом 3 (k) «конкретно определенный» нуклеотид - это любой нуклеотид, отличный от нуклеотида, который представлен символом «n», представленным в Приложении 1, Раздел 1, Таблица 1.

В этом примере «n» используется нерекомендованным образом для представления только «c». В раскрытии не указывается, что «n» используется обычным образом для обозначения «любого нуклеотида». Следовательно, последовательность должна интерпретироваться так, как если бы в последовательности использовался эквивалентный условный символ, то есть «c» (см. Введение к этому документу). Соответственно, представленная последовательность, которая должна учитываться это:

5’-aatgttggac-3’

Данная последовательность имеет десять конкретно определённых нуклеотидов и по требованию параграфа 7(a) ST.26 должна быть включена в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность должна быть включена в перечень последовательностей как:
aatgttggac (SEQ ID NO: 15)

**Соответствующие параграфы ST.26: 3(k)** и 7(a)

**Пример 3(k)-4: Неоднозначные символы, отличные от “n” являются “конкретно определёнными”**

В патентной заявке раскрыта следующая последовательность:

5’ NNG KNG KNG KAG VCR 3’

где N, K, V, и R это IUPAC-IUB неоднозначные коды

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Неоднозначные коды IUPAC-IUB соответствуют списку символов нуклеотидов, определенных в Приложении I, Раздел 1, Таблица 1. В соответствии с параграфом 3 (k) «конкретно определённым» нуклеотидом является любой нуклеотид, отличный от тех, которые представлены символом «n», перечисленные в Приложении I, Раздел 1, Таблица 1. Следовательно, «K», «V» и «R» являются «конкретно определенными» нуклеотидами.

Последовательность имеет одиннадцать перечисленных и «конкретно определённых» нуклеотидов и по требованию параграфа 7(a) ST.26 должна быть включена в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность должна быть включена в перечень последовательностей как: nngkngkngkagvcr (SEQ ID NO: 16)

**Соответствующие параграфы ST.26: 3(k)**, 7(a) и 15

**Пример 3(k)-5: Неоднозначное сокращение “Xaa” используется нерекомендованным образом**

В патентной заявке раскрыта следующая последовательность:

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

Где Xaa на позиции 1 любая аминокислота, Xaa на позиции 4 это Lys, Xaa на позиции 5 это Gly, а Xaa на позиции 6 - Лейцин или изолейцин.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Представленный в формуле пептид содержит три конкретно определенных аминокислоты на позициях 2, 3 и 7. Первая аминокислота представлена условно принятым сокращением, то есть Xaa, обозначая любую аминокислоту. Однако 4-я, 5-я и 6-я аминокислоты представлены условно принятой аббревиатурой, используемой нерекомендуемым способом (см. Введение в этот документ). Следовательно, пояснение последовательности в описании приведено для определения «Xaa» на указанных позициях. Поскольку «Xaa» на позициях 4-6 указаны как конкретная аминокислота, последовательность должна интерпретироваться так, как если бы в последовательности использовались эквивалентные условно принятые сокращения, то есть Lys, Gly и (Leu или Ile). Следовательно, последовательность содержит четыре или более конкретно определенных аминокислот и должна быть включена в перечень последовательностей, как того требует ST.26 параграф 7 (b).

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

В последовательности используется условно принятое сокращение «Xaa» нерекомендуемым способом. Следовательно, пояснение последовательности в описании должно быть использовано для определения «Xaa» на позициях 4, 5 и 6. В пояснении «Xaa» определяется как лизин на позиции 4, глицин на позиции 5 и лейцин или изолейцин на позиции 6. Условно принятыми символами для этих аминокислот являются K, G и J соответственно. Следовательно, последовательность должна быть представлена в перечне последовательностей как:

XYEKGJL (SEQ ID NO: 17)

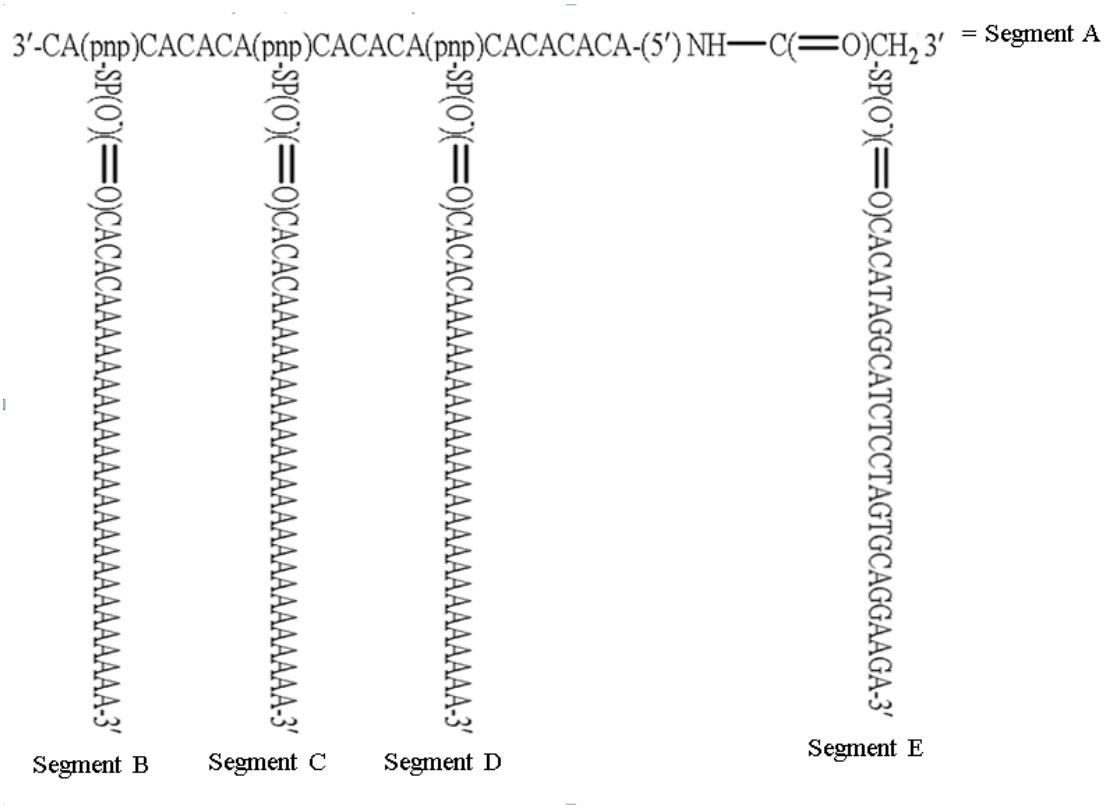
Согласно параграфу 27, «X» будет рассматриваться как любая из следующих немодифицированных L-альфа-аминокислот: «A», «R», «N», «D», «C», «Q», «E», «G», «H» , «I», «L», «K», «M», «F», «P», «O», «S», «U», «T», «W», «Y» или «V», за исключением случаев, когда они используются с последующим описанием в таблице характеристик. «X» в положении 1 SEQ ID NO: 17 представляет собой «любую аминокислоту», которая шире по охвату, чем значение «X» по умолчанию, данное в параграфе 27, поскольку «любая аминокислота» охватывает модифицированные или другие аминокислоты. Следовательно, «X» в данной последовательности должна быть аннотирована с помощью функционального ключа VARIANT и квалификатора note со значением: «X может быть любой аминокислотой».

**Соответствующие параграфы ST.26: 3(k)**, 7(b), 26 и 27

*Параграф 7(a) – Нуклеотидные последовательности обязательные в перечне последовательностей*

**Пример 7(a)-1: Разветвлённая нуклеотидная последовательность**

Описание раскрывает следующую разветвленную нуклеотидную последовательность:




где "pnp" представляет собой сцепление или мономер, содержащий бромацетиламинную функциональную группу;

3’-CA(pnp)CACACA(pnp)CACACA(pnp)CACACACA-(5’)NH—C(=O)$CH_2$ 3’ - это сегмент A;

SP($O^-$)(=O)CACACAAAAAAAAAAAAAAAAAAAAA 3’ - это сегменты B, C и D; и

SP($O^-$)(=O)CACATAGGCATCTCCTAGTGCAGGAAGA 3’ - это сегмент E.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(ей)?**

**ДА** – четыре вертикальных сегмента B-E должны быть включены в перечень последовательностей.

**НЕТ** – горизонтальный сегмент A не должен быть включён в перечень последовательностей как одиночная последовательность.

Вышеприведенная фигура является примером “comb-type” разветвленной последовательности нуклеиновой кислоты, содержащей пять линейных сегментов: горизонтальный сегмент A и четыре вертикальных сегмента B-E.

В соответствии с параграфом 7 (a) линейные участки разветвленных нуклеотидных последовательностей, содержащих десять или более конкретно определенных нуклеотидов, в которых соседние нуклеотиды соединены от 3 'до 5', должны быть включены в перечень последовательностей.

Каждый из четырех вертикальных сегментов B-E содержит более десяти конкретно определенных нуклеотидов, где соседние нуклеотиды соединены от 3 'до 5', и поэтому каждый из них должен быть включен в перечень последовательностей.

В горизонтальном сегменте A линейные участки нуклеотидной последовательности связаны не-нуклеотидным фрагментом «pnp», и каждый из этих связанных линейных участков содержит менее десяти конкретно определенных нуклеотидов. Следовательно, поскольку ни одна область сегмента A не содержит десять или более конкретно определенных нуклеотидов, в которых соседние нуклеотиды соединены от 3 'до 5', согласно параграфу 7(a) ST.26 они не должны включаться в перечень последовательностей.

**Вопрос 2: Разрешает ли ST.26 включение данной последовательности(ей)?**

**ДА**

Согласно параграфу 8 ST.26 «Перечень последовательностей может включать последовательности, имеющие менее десяти конкретно определенных нуклеотидов…».

Сегмент A содержит четыре линейных участка нуклеотидной последовательности, разделенных не нуклеотидными фрагментами “pnp“. Следовательно, сегмент A не может быть включен в перечень последовательностей как одиночная последовательность и должен рассматриваться как четыре отдельных последовательности. Каждая из этих четырех отдельных последовательностей может быть включен в перечень последовательностей со своими собственными SEQ ID номерами.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Сегменты B, C и D идентичны и должны быть включены в перечень последовательностей как одна последовательность:

cacacaaaaaaaaaaaaaaaaaaaaaaaa. (SEQ ID NO: 18)

Первое «c» в последовательности должно быть далее описано с использованием функционального ключа «misc_feature» и квалификатора «note» со значением, например, «Эта последовательность является одной из четырех ветвей разветвленного полинуклеотида».

Сегмент E должен быть включен в перечень последовательностей как одна последовательность:

cacataggcatctcctagtgcaggaaga. (SEQ ID NO: 19)

Первое «c» в последовательности должно быть далее описано с использованием функционального ключа «misc_feature» и квалификатора «note» со значением, например, «Эта последовательность является одной из четырех ветвей разветвленного полинуклеотида».

Четыре линейных последовательности, представленные в сегменте A, могут быть представлены в перечне последовательностей с 5’ до 3” как представлено далее:

acacacac (SEQ ID NO: 106)

acacac (SEQ ID NO: 107)

ac (SEQ ID NO: 108)

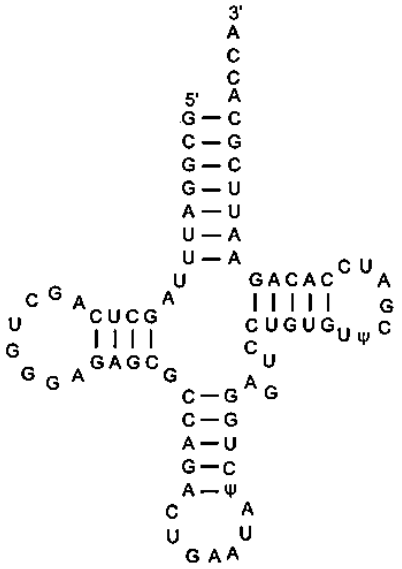
Обратите внимание, что два внутренних сегмента последовательностей в Сегменте A идентичны, поэтому один SEQ ID номер достаточен для обоих.

Каждый сегмент должен быть аннотирован с использованием функционального ключа “misc_feature” и квалификатора “note“ для указания того, что они модифицированы на 5’- ом или 3’-ем концах с помощью “pnp“ и соединены вместе как часть большей разветвленной структуры.

**Соответствующие параграфы ST.26: 7(a)**, 8, 11, 13 и 17

**Пример 7(a)-2: Линейная нуклеотидная последовательность, имеющая вторичную структуру**

В патентной заявке описана следующая последовательность:




Где Ψ это псевдоуридин.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Нуклеотидная последовательность содержит семьдесят три перечисленных и конкретно определенных нуклеотида. Таким образом, в примере имеется десять или более «конкретно определенных» нуклеотидов, и, как того требует параграф 7(a) ST.26, они должны быть включены в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Экспертиза раскрытия информации указывает, что «Ψ» эквивалентно псевдоуридину. Единственный условно принятый символ, который можно использовать для обозначения псевдоуридина, это «n»; следовательно, «Ψ» является нерекомендованным символом, используемым для обозначения условно принятого символа «n» (см. Введение к этому документу). Соответственно, последовательность должна толковаться, как имеющая два «n» символа вместо двух «Ψ» символов.

Символ «u» не должен использоваться для обозначения урацила в молекуле РНК в перечне последовательностей. Согласно параграфу 14, символ «t» будет толковаться как урацил в РНК. Последовательность должна быть включена как:

gcggatttagctcagctgggagagcgccagactgaatanctggagtcctgtgtncgatccacagaattcgcacca (SEQ ID NO: 20)

Значением обязательного квалификатора «mol_type» обязательного функционального ключа «source» является «tRNA». Дополнительная информация может быть предоставлена с помощью функционального ключа «tRNA» и любого соответствующего квалификатора (ов).

Остатки «n» должны быть дополнительно описаны в таблице характеристик с использованием функционального ключа «updated_base» и обязательного квалификатора «mod_base» с аббревиатурой «p» для псевдоуридина в качестве значения квалификатора (см. Приложение 1, таблица 2).

**Соответствующие параграфы ST.26: 7(a)**, 11, 13, 14, 17, 62, 84 и Приложение 1, Главы 2 и 5, функциональный ключ 5.43.

**Пример 7(a)-3: Неоднозначные символы нуклеотидов, используемые нерекомендуемым образом**

В патентной заявке описана следующая последовательность:

5’ GATC-MDR-MDR-MDR-MDR-GTAC 3’

Дальнейшее объяснение последовательности в описании указывает: «DR элемент» состоит из последовательности 5 ’ATCAGCCAT 3’. Мутантный DR-элемент, или MDR, представляет собой DR-элемент, в котором 5 средних нуклеотидов, CAGCC, мутировали в TTTTT».

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

В представленной последовательности используется символ «MDR». В случае, если неясно, используется ли символ в последовательности, как условно принятый символ, то есть символ, представленный в Приложении 1, Раздел 3, Таблица 3, или как неоднозначный символ, нужно определится (см. Введение к этому документу), приняв во внимание объяснение последовательности в описании. Согласно Таблице 3, «MDR» может интерпретироваться как три условно принятых символа (m = a или c, d = a или g или t / u, r = g или a) или как аббревиатура, которое является условным сокращением для некоторых других структур.

Экспертиза раскрытия информации показывает, что элемент MDR эквивалентен 5’ATTTTTTAT 3’. Буквы «MDR» считаются условно принятыми символами, используемыми нерекомендованным способом; следовательно, последовательность должна интерпретироваться так, как если бы она была раскрыта с использованием эквивалентных условно принятых символов. Соответственно, представленная последовательность, которая рассматривается для включения в перечень последовательностей следующая:

5’ GATC ATTTTTTAT ATTTTTTAT ATTTTTTAT ATTTTTTAT GTAC 3’

Представленная последовательность имеет 44 конкретно определенных нуклеотида и согласно параграфу 7(a) ST.26 требует включения в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность должна быть включена в перечень последовательностей как:

gatcatttttatatttttatatttttatatttttatgtac (SEQ ID NO: 21)

**Соответствующие параграфы ST.26: 7(a) и** 13

**Пример 7(a)-4: Неоднозначные символы нуклеотидов, используемые нерекомендуемым образом**

В патентной заявке описана следующая последовательность:

5’ ATTC-N-N-N-N-GTAC 3’

Дальнейшее объяснение последовательности в описании указывает, что «N» состоит из последовательности 5 ’ATACGCACT 3’.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

В представленной последовательности используется символ «N». Необходимо принять во внимание объяснение последовательности в описании, чтобы определить, используется ли «N» условно принятым или нерекомендуемым образом (см. Введение к этому документу).

Экспертиза раскрытия информации показывает, что «N» эквивалентно 5 ’ATACGCACT 3’. Таким образом, «N» - это условно принятый символ, используемый нерекомендуемым образом. Соответственно, последовательность должна интерпретироваться так, как если бы она была раскрыта с использованием эквивалентных условно принятых символов:

5’ ATTC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3’

Представленная последовательность имеет 44 конкретно определенных нуклеотида и согласно параграфу 7 (a) ST.26 требует включения в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность должна быть включена в перечень последовательностей как:

attcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO: 22)

**Соответствующие параграфы ST.26: 7(a) и** 13

**Пример 7(a)-5: Нерекомендованные символы нуклеотидов**

В патентной заявке описана следующая последовательность:

5’ GATC-β-β-β-β-GTAC 3’

Дальнейшее объяснение последовательности в описании указывает, что “β” состоит из последовательности 5’ ATACGCACT 3’.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

В представленной последовательности используется нерекомендованный символ “β”. Необходимо принять во внимание объяснение последовательности в описании, чтобы определить, что означает “β” (см. Введение к этому документу).

Экспертиза раскрытия информации показывает, что “β” эквивалентно 5’ ATACGCACT 3’. Таким образом, «β» является нерекомендованным символом, используемым для представления последовательности из девяти конкретно определенных условно принятых символов. Соответственно, последовательность должна интерпретироваться так, как если бы она была раскрыта с использованием эквивалентных условно принятых символов:

5’ GATC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3’

Представленная последовательность имеет 44 конкретно определенных нуклеотида и согласно параграфу 7(a) ST.26 требует включения в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность должна быть включена в перечень последовательностей как:

gatcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO: 23)

**Соответствующие параграфы ST.26: 7(a) и** 13

**Пример 7(a)-6: Нерекомендованные символы нуклеотидов**

В патентной заявке описана следующая последовательность:

5’ GATC-β-β-β-β-GTAC 3’

Дальнейшее объяснение последовательности в описании указывает, что “β” равно аденину, инозину или псевдоуридину.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

В представленной последовательности используется нерекомендованный символ “β”. Необходимо принять во внимание объяснение последовательности в описании, чтобы определить, что означает “β” (см. Введение к этому документу).

Экспертиза раскрытия информации показывает, что “β” эквивалентно аденину, инозину или псевдоуридину. Единственным условно принятым символом, который можно использовать для обозначения «аденина, инозина или псевдоуридина», является «n»; следовательно, «β» является нерекомендованным символом, используемым для представления условно принятого символа «n». Соответственно, последовательность должна интерпретироваться как имеющая четыре символа «n» (показано как «N» ниже) вместо четырех символов «β»:

5’ GATC-N-N-N-N-GTAC 3’

Представленная последовательность имеет только восемь конкретно определенных нуклеотида и согласно параграфу 7(a) ST.26 не требует включения в перечень последовательностей.

**Вопрос 2: Разрешает ли ST.26 включение последовательности(ей)?**

**ДА**

Параграф 8 ST.26 разрешает включение последовательностей с менее чем десятью конкретно определенными последовательностями. Поэтому представленная последовательность 5’ GATC-N-N-N-N-GTAC 3’ может быть включена в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность следует представить в перечне последовательностей как:

gatcnnnngtac (SEQ ID NO: 109)

“n” остатков на позициях 5-8 следует затем аннотировать в таблице характеристик, используя функциональный ключ “misc_difference” и обязательный квалификатор “note“ со значением, указывающим, что “n” – это аденин, инозин или псевдоуридин. Поскольку две альтернативы являются модифицированными нуклеотидами, также должны быть использованы функциональный ключ “modified_base” вместе с квалификатором “mod_base”. У квалификатора “mod_base” должно быть значение “OTHER”. Требуются два квалификатора примечаний (“note“), один – со значением “i”, а другой – со значением “p”.

**Соответствующие параграфы ST.26: 7(a)**, 8, 13, и 17

**Пример 7(a)-7: Инвертированные нуклеотиды I**

В патентной заявке раскрыта следующая одноцепочечная последовательность ДНК:

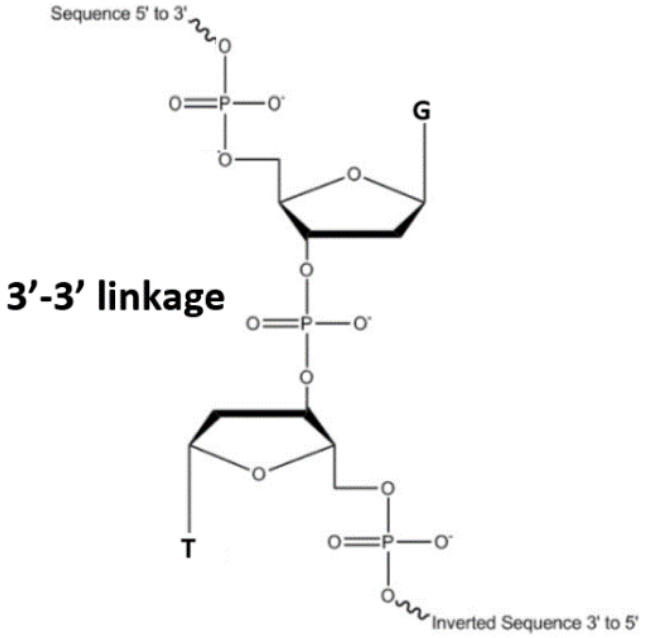
5'attgactaagtgttccccattgact5'

При этом направленность последовательности изменяется внутри цепи из-за обратной связи 3'-3' между остатками 12 и 13. Подчеркнутая часть последовательности ориентирована от 3' до 5' слева направо.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Первые 12 остатков изображены в стандартной ориентации от 5’ к 3’. Буква «g» в позиции 12 связана с буквой «t» в позиции 13 через обратную связь от 3’ к 3’:




Остальная часть молекулы, изображенная в позициях с 13 по 25, находится в противоположной ориентации – от 3’ к 5’.

Параграф 11 ST.26 требует, чтобы нуклеотидная последовательность была представлена в направлении от 5’ к 3’ слева направо.

Следовательно, чтобы правильно представить эту молекулу в перечне последовательностей, она должна быть представлена двумя последовательностями: одна последовательность для позиций с 1 по 12, а вторая для позиций с 13 по 25. Каждая часть последовательности содержит десять или более «конкретно определенных» последовательностей нуклеотидов и, как того требует параграф 7(a) ST.26, должны быть включены в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Позиции 1–12 должны быть включены в перечень последовательностей как:

attgactaagtg (SEQ ID NO.:99)

Позиция 12 должна быть описана в таблице характеристик с использованием функционального ключа «misc_feature» и квалификатора «note» со значением, указывающим, что остаток связан с инвертированной нуклеотидной последовательностью через 3'-3' фосфодиэфирную связь с тимидин-3'-монофосфатом на позиции 13 последовательности tcagttacccctt.

Позиции 13-25 должны быть включены в перечень последовательностей как:

tcagttacccctt (SEQ ID NO.: 100)

Обратите внимание, что эта последовательность имеет обратную ориентацию по отношению к тому, как она была изображена в исходном описании, так что теперь она ориентирована от 5’ к 3’, слева направо. Позиция 13 должна быть описана в таблице характеристик с использованием функционального ключа «misc_feature» и квалификатора «note» со значением, указывающим, что остаток связан с инвертированной нуклеотидной последовательностью через 3'-3'-фосфодиэфирную связь с гуанозином 3'- монофосфат ан позиции 12 последовательности attgactaagtg.

**Соответствующие параграфы ST.26: 7(a)**, 11

**Пример 7(a)-8: Инвертированные нуклеотиды II**

В патентной заявке раскрыта следующая последовательность ДНК:

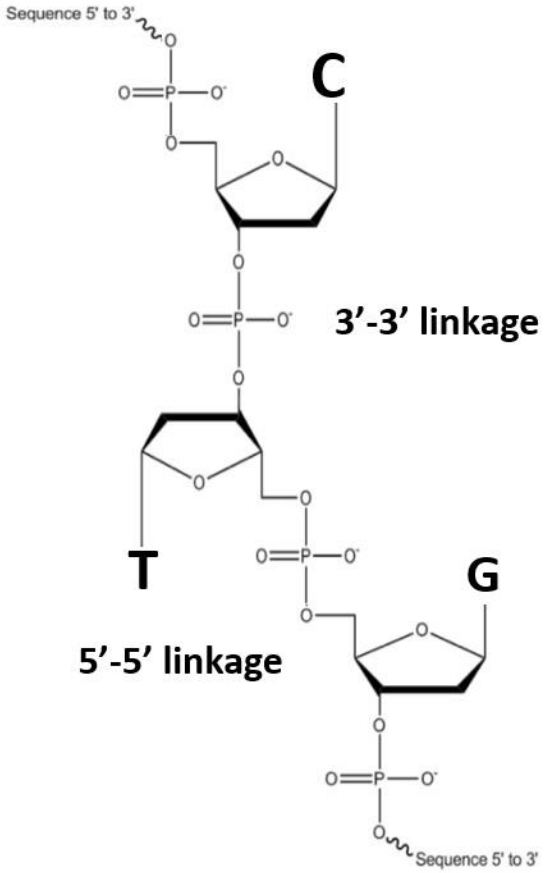
5'-attgactaagtgctgcccattgact-3'

При этом подчеркнутый остаток тимина (позиция 14) представляет собой инвертированный нуклеотид, который связан с цитозином через 3'-3'-фосфодиэфирную связь и с гуанином через 5'-5'-фосфодиэфирную связь.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(ей)?**

**ДА**

Инвертированный тимидин в позиции 14 прерывает направленность последовательности от 5' к 3' путем введения связи 3'-3' между остатками 13 и 14 и связи 5'-5' между остатками 14 и 15:




Инвертированный тимидин связывает первую часть последовательности, остатки 1–13, и вторую часть последовательности, остатки 15–25. Каждая часть последовательности

содержит, по меньшей мере, 10 пронумерованных и конкретно определенных нуклеотидов. Таким образом, каждая часть последовательности должна быть включена в перечень последовательностей, как того требует параграф 7(a) ST.26. Одиночный тимидин, связывающий две части последовательности, не может быть включен в перечень последовательностей как отдельная последовательность, поскольку он не представляет собой последовательность, состоящую как минимум из 10 конкретно определенных нуклеотидов.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Инвертированный тимидин на позиции 14 прерывает направленность последовательности от 5' к 3' путем введения связи 3'-3' между остатками 13 и 14 и связи 5'-5' между остатками 14 и 15. ST.26, параграф 11 требует, чтобы нуклеотидная последовательность была представлена в направлении от 5' к 3' слева направо. Следовательно, чтобы правильно представить эту молекулу в перечне последовательностей, она должна быть представлена двумя последовательностями: первая последовательность для остатков с 1 по 13, и вторая последовательность для остатков с 15 по 25.

Позиции 1–13 должны быть включены в перечень последовательностей как:

attgactaagtgc (SEQ ID NO: 101)

Позиция 13 должна быть описана в таблице характеристик с использованием функционального ключа «misc_feature» и квалификатора «note» со значением, указывающим, что остаток связан посредством 3'-3'-фосфодиэфирной связи с тимидин-3'-монофосфатом, который в свою очередь связан через 5'-5' фосфодиэфирную связь с g остатком на позиции 1 последовательности gcccattgact.

Позиции 15-25 должны быть включены в список последовательностей как:

gcccattgact (SEQ ID NO: 102)

Позиция 1 должна быть описана в таблице характеристик с использованием функционального ключа «misc_feature» и квалификатора «note» со значением, указывающим, что остаток связан через 5'-5'-фосфодиэфирную связь с тимидин-3'-монофосфатом, который в свою очередь связан через 3'-3' фосфодиэфирную связь с остатком на позиции 13 последовательности attgactaagtgc.

**Соответствующие параграфы ST.26: 7(a)**, 11

*Параграф 7(b) – Аминокислотные последовательности, требуемые (обязательные) в перечне последовательностей*

**Пример 7(b)-1: Четыре и более конкретно определённых аминокислоты**

XXXXXXXXDXXXXXXXXXXFXXXXXXXXXXXXXXXXXXXXXXXXXXXXAXXXXXXXXXXXXXXXX
XXXXGXXXXX

Где X = любая аминокислота

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Представленный пептид содержит четыре конкретно определенные аминокислоты. Символ «X» используется для обозначения остальных аминокислот как любой аминокислоты (см. Введение к этому документу).

Поскольку существуют четыре конкретно определенные аминокислоты, то есть Asp, Phe, Ala и Gly, параграф 7(b) ST.26 требует, чтобы последовательность была включена в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность должна быть представлена как:

XXXXXXXXDXXXXXXXXXXFXXXXXXXXXXXXXXXXXXXXXXXXXXXXAXXXXXXXXXXXXXXXX
XXXXGXXXXX (SEQ ID NO: 25)

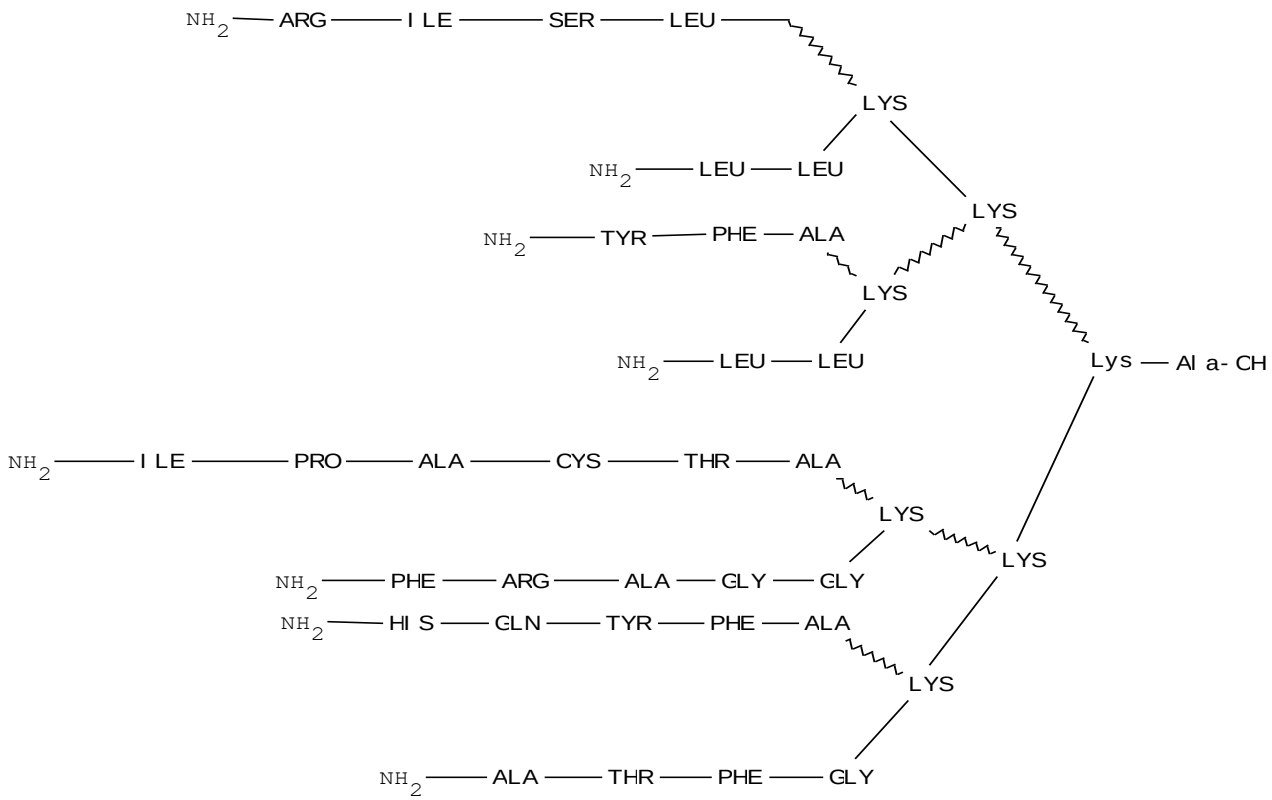
Согласно параграфу 27, «X» будет толковаться как одна из немодифицированных L-альфа-аминокислот: «A», «R», «N», «D», «C», «Q», «E», «G», «H», «I», «L», «K», «M», «F», «P», «O», «S», «U», «T», «W», «Y» или «V», за исключением случаев, когда он используется с последующим описанием в таблице характеристик. «X» в SEQ ID NO: 25 представляет собой «любую аминокислоту», которая шире по охвату, чем определение «X» по умолчанию, данное в параграфе 27, поскольку «любая аминокислота» охватывает модифицированные или другие аминокислоты. Поэтому “X” остатки в данной последовательности должны быть аннотированы с помощью функционального ключа VARIANT и квалификатора note со значением «“X” может быть любой аминокислотой».

По возможности, каждый “X” должен быть аннотирован индивидуально. Однако участок смежных остатков “X” может быть совместно описан с помощью функционального ключа VARIANT с использованием синтаксиса «x..y» в качестве дескриптора местоположения, где x и y - позиции первого и последнего остатков “X”, а также квалификатор note со значением «X может быть любой аминокислотой».

**Соответствующие параграфы ST.26: 7(b)** и 27

**Пример 7(b)-2: Разветвленная аминокислотная последовательность**

В заявке описана разветвленная последовательность, где остатки лизина используются в качестве каркасного ядра для образования восьми ветвей, к которым присоединены многочисленные линейные пептидные цепи. Лизин является двухосновной аминокислотой, обеспечивающей его двумя участками для амидной связи. Разветвленная последовательность иллюстрируется следующим образом:



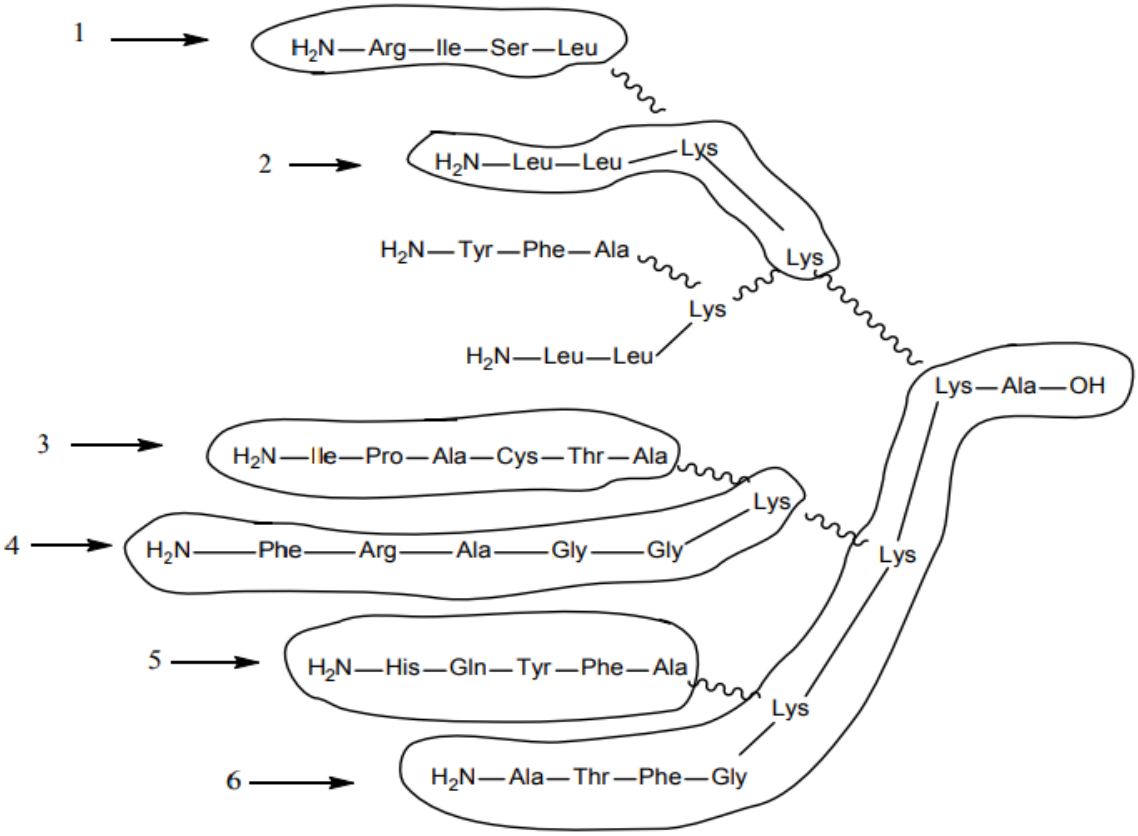

В указанном выше разветвленном пептиде связи между лизином и прилегающей аминокислотой, обозначенные как ———, представляют собой амидную связь между концевым амином лизина и карбоксильным концом связанной аминокислоты. Связи, обозначенные как ∿∿∿, представляют собой амидную связь между амином боковой цепи лизина и карбоксильным концом связанной аминокислоты.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(ей)?**

**ДА**

В примере раскрыта разветвленная последовательность, где остатки лизина используются в качестве каркаса. Поскольку иллюстрация делает различия между основой амидных связей ( ——— ) и связями боковых цепей ( ∿∿∿ ), линейные участки разветвленной последовательности, которые формируют каждую основу, очевидны. Параграф 7(b) требует, чтобы неразветвленный или линейный участок последовательности, содержащий четыре или более конкретно определенных аминокислоты, в которых аминокислоты формируют единую пептидную основу, были включены в перечень последовательностей.

Эти разветвленные последовательности состоят из 8-ми линейных пептидов; линейные пептиды, имеющие четыре или более аминокислот, обведены на рисунке замкнутыми кривыми:




Пептиды 1 и 5-8 содержат по четыре или более конкретно определенных аминокислот. Поэтому параграф 7(b) ST.26 требует включения пептидов 1 и 5-8 в перечень последовательностей.

Пептид 2 содержит 2 конкретно определенных аминокислоты (Leu-Leu). Два остатка Lys на позициях 3 и 4 модифицированы посредством амидной связи между их аминами боковой цепи и карбоксильным концом связанной аминокислоты. В соответствии с параграфом 29 ST.26 модифицированные аминокислоты следует представлять в последовательности, когда возможно, как соответствующие немодифицированные аминокислоты. Следовательно, пептид 2 следует включить в перечень последовательностей как “LLKK”.

Пептиды 3 и 4 каждый содержит только 3 аминокислоты; поэтому не требуется, чтобы они были включены в перечень последовательностей.

**Вопрос 2: Разрешает ли ST.26 включение этой последовательности(ей)?**

**ДА**

Согласно параграфу 8, перечень последовательностей может включать последовательности, имеющие менее четырех конкретно определенных аминокислот.

Пептиды 3 и 4 каждый содержит менее чем четыре конкретно определенные аминокислоты. Поэтому их включение в перечень последовательностей со своими собственными идентификационными номерами последовательностей является опциональным (оставляется на усмотрение заявителя).

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Все из модифицированных лизиновых остатков из вышеприведенных пептидов следует представлять в последовательности с символом “K”, а не “X”, в соответствии с рекомендациями параграфа 29 в ST.26.

Пептиды 1 и 5-8 должны быть включены в перечень последовательностей отдельными идентификаторами последовательности, и их следует представлять следующим образом:

RISL (SEQ ID NO: 26)

IPACTA (SEQ ID NO: 28)

FRAGGK (SEQ ID NO: 29)

HQYFA (SEQ ID NO: 30)

ATFGKKKA (SEQ ID NO: 31)

Дополнительно, настоятельно рекомендуется, чтобы пептид 2 был включен в перечень последовательностей со своим собственным идентификатором последовательности, и представлен следующим образом:

LLKK (SEQ ID NO: 27).

В соответствии с параграфом 30 ST.26 модифицированные лизиновые остатки в каждой последовательности должны быть аннотированы с использованием функционального ключа “SITE совместно с квалификатором “note“. В значении квалификатора “note следует описать, что боковые цепи лизина присоединены к другой последовательности через амидные связи. Дополнительно, каждая последовательность в разветвленной структуре может быть аннотирована с помощью функционального ключа “SITE“ и обязательного квалификатора “note“ со значением, например, «Данная последовательность является одной частью разветвленной аминокислотной последовательности».

Пептиды 3 и 4 могут быть включены в перечень последовательностей и представлены следующим образом:

YFA (SEQ ID NO: 110)

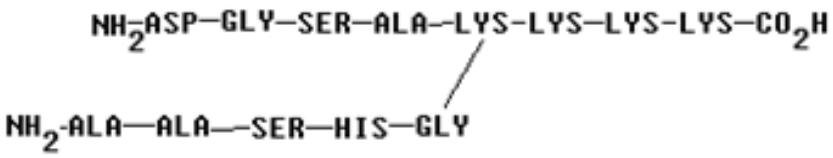
LLK (SEQ ID NO: 111)

Эти последовательности следует аннотировать подобно тому, как объяснено выше для SEQ ID 26-31.

**Соответствующие параграфы ST.26: 7(b)**, 8, 26, 29, 30, и 31

**Пример 7(b)-3: Разветвленная аминокислотная последовательность II**

Пептид следующей последовательности:




Связь между концевым остатком глицина в нижней последовательности и лизином в верхней последовательности осуществляется через амидную связь между карбокси-концом глицина и аминоконцевой боковой цепью лизина.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-ей)?**

**ДА**

Неразветвленный или линейный участок последовательности, содержащий четыре или более конкретно определенных аминокислот, должен быть включен в перечень последовательностей. В приведенном выше примере линейные участки разветвленного пептида, которые имеют более четырех аминокислот, представляют собой:




Параграф 7 (b) ST.26 требует включения последовательностей 1 и 2 в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Лизиновый остаток в позиции 5 последовательности 1 модифицирован за счет амидной связи между амином боковой цепи лизина и карбоксильным концом С-концевого глицина. В соответствии с параграфом 7(b) ST.26 модифицированную аминокислоту следует представлять в последовательности как соответствующие немодифицированные аминокислоты, когда это возможно. Поэтому модифицированный лизиновый остаток следует представлять в последовательности как “K“.

Соответственно, последовательности 1 и 2 следует представить следующим образом:

DGSAKKKK (SEQ ID NO: 32)

AASHG (SEQ ID NO: 33)

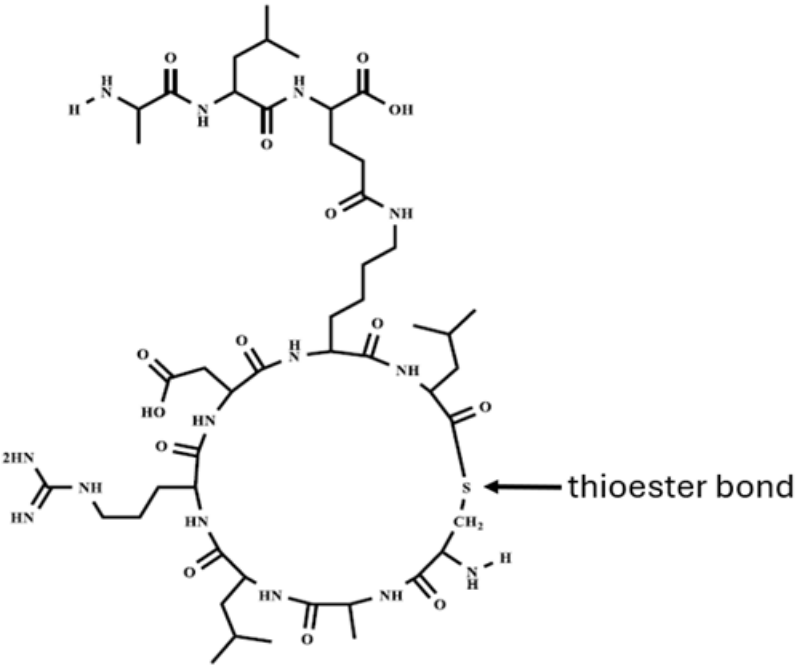
Последовательность DGSAKKKK должна включать аннотацию, указывающую, что лизин на позиции № 5 является модифицированной аминокислотой, используя функциональный ключ «SITE» вместе с квалификатором «note», описывающим, что боковая цепь лизина

связана через амидную связь с другой последовательностью. Последовательность AASHG должна включать аннотацию, указывающую, что глицин в позиции номер 5 связан с другой последовательностью с помощью функционального ключа «SITE» вместе с квалификатором «note».

**Соответствующие параграфы ST.26: 7(b)**, 26, 29, 30, и 31

**Пример 7(b)-4: Циклический пептид, содержащий разветвленную аминокислотную последовательность**

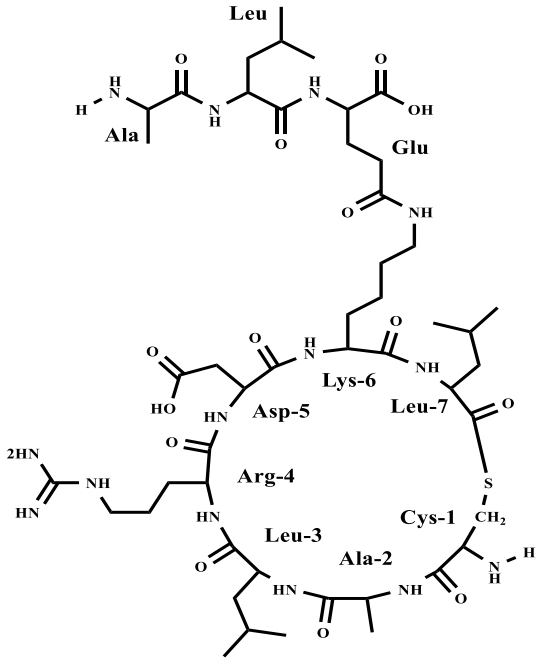
Заявка на патент раскрывает следующую структуру:




Цистеин и лейцин в циклической структуре связаны тиоэфирной связью между боковой цепью Cys и карбоксильным концом Leu.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(ей)?**

На структуре показана разветвлённая циклическая аминокислотная последовательность, которая содержит следующие аминокислоты:



Поскольку боковая цепь Cys и карбоксильный конец Leu участвуют в циклизации, N-конец циклического пептида находится в Cys-1. Боковая цепь Lys в позиции 6 циклической структуры связана через амидную цепь с боковой цепью Glu на C-конце пептида из трех аминокислот Ala-Leu-Glu для формирования разветвленной последовательности.

**ДА – циклическая область пептида**

Параграф 7 (b) ST.26 требует, чтобы линейный участок разветвленной последовательности, содержащий четыре или более конкретно определенных аминокислот, где аминокислоты образуют одну пептидную основу, должен быть включен в перечень последовательностей. В приведенном выше примере циклический участок разветвленного пептида имеет более четырех конкретно определенных аминокислот, и, следовательно, должен быть включен в перечень последовательностей.

**НЕТ – трипептидная ветвь пептида**

Трипептидная ветвь Ala-Leu-Glu не обязательно должна присутствовать в перечне последовательностей.

**Вопрос 2: Разрешает ли ST.26 включение последовательности(ей)?**

**ДА**

Согласно параграфу 8, перечень последовательностей может включать какие-либо последовательности, имеющие менее четырех конкретно определенных аминокислот.

Трипептидная ветвь содержит только три конкретно определенные аминокислоты и, следовательно, её включение в перечень последовательностей в виде отдельной последовательности с собственным идентификационным номером оставляется на усмотрение заявителя.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Хотя этот пример иллюстрирует пептид, который имеет круговую конфигурацию, кольцо не состоит только из аминокислотных остатков в пептидных связях, как указано в параграфе 25. Поскольку циклизация аминокислотной последовательности происходит через боковую цепь цистеина (Cys) и карбоксильный конец лейцина (Leu) цистеину должен быть присвоен номер позиции 1 на циклическом участке пептида. Соответственно, последовательность должна быть представлена как:

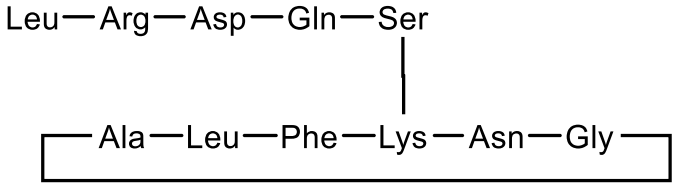
CALRDKL (SEQ ID NO: 90).

Как показано на фигуре выше, аминокислотная последовательность циклизуется посредством конъюгации тиоэфира между боковой цепью цистеина и карбокси-концом лейцина. Функциональный ключ “SITE” должен использоваться для описания модифицированного цистеина, который образует внутрицепочечную связь с лейцином. Элемент местоположения признака - это номера остатков перекрёстно связанных аминокислот в формате «x..y», то есть «1..7)». Обязательный квалификатор “note” должен указывать характер связи, например «цистеин-лейцин-тиоэфир (Cys-Leu)», чтобы указать, что Cys-1 и Leu-7 связаны через тиоэфирную связь. Кроме того, чтобы указать, что лизин на позиции № 6 модифицирован, он должен быть снабжён аннотацией, с использованием функционального ключа «SITE» вместе с обязательным квалификатором «note», где значение квалификатора описывает, что боковая цепь лизина связывает трипептид ALE.

**Соответствующие параграфы ST.26: 7(b)**, 8, 25, 26, 29 30, 31, 66(c) и 70

**Пример 7(b)-5: <u>Циклический пептид, содержащий разветвленную аминокислотную последовательность II</u>**

Заявка на патент раскрывает следующий разветвленный циклический пептид:

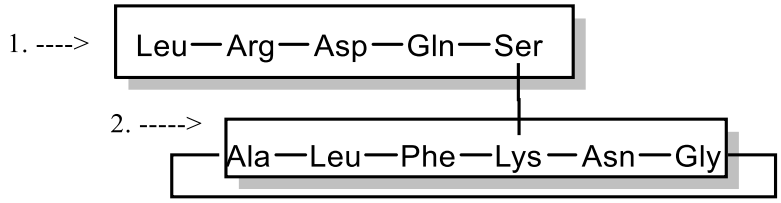



Ser и Lys связаны через амидную связь между карбокси-концом серина и амином в боковой цепи Lys.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(ей)?**

**ДА**

Параграф 7 (b) требует включения любой последовательности, которая содержит четыре или более конкретно определенных аминокислот и которая может быть представлена в виде линейного участка разветвленной последовательности в перечне последовательностей. В приведенном выше примере пептид содержит циклический участок, на котором аминокислоты связаны амидными связями, и разветвленный участок, который присоединен к боковой цепи Lys на циклическом участке. Участки данного разветвленного пептида, которые могут быть представлены как линейные и которые содержат четыре или более конкретно определенных аминокислот, представляют собой:




ST.26 требует включения последовательностей 1 и 2 данного циклического разветвленного пептида в перечень последовательностей, со своим собственным идентификационным номером последовательности у каждой.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность 1 должна быть представлена как:

LRDQS (SEQ. ID. NO: 91)

Последовательность 1 может быть аннотирована с использованием функционального ключа «SITE» вместе с квалификатором «note» для описания того, что серин в позиции 5 связан с другой последовательностью через амидную связь между основой карбоксильной группы Ser и боковой цепью Lys в другой последовательности.

Последовательность 2 представляет собой циклический пептид. В параграфе 25 указано, что, когда аминокислотная последовательность имеет кольцевую форму и не имеет амино- и

карбокси- концов, заявитель должен выбрать аминокислотный остаток на позиции номер 1. Хотя любая аминокислота в последовательности может быть обозначена как позиция остатка на позиции номер 1, с точки зрения описания разветвленного циклического пептида Ala представляется подходящей для выбора в качестве первого остатка. Lys в позиции 4 является модифицированной благодаря амидной связи между её боковой цепью и карбоксильным концом Ser в последовательности 1. В соответствии с параграфом 29 ST.26 модифицированные аминокислоты следует представлять в последовательности как соответствующие немодифицированные аминокислоты, когда это возможно. Поэтому модифицированный остаток Lys следует в последовательности представлять как “K“, а не “X”. Следовательно, последовательность следует представить как:

ALFKNG (SEQ. ID. NO: 92)

В качестве альтернативы, любая другая аминокислота в последовательности может быть обозначена как позиция остатка номер 1. Lys на позиции 4 должен быть далее описан с использованием функционального ключа «SITE» вместе с квалификатором «note», чтобы пояснить, что боковая цепь связана через амидную связь с другой последовательностью. Эта связь боковой цепи модифицирует Lys, и, в соответствии с параграфом 30 ST.26, модифицированная аминокислота должна быть дополнительно описана в таблице характеристик. Более того, следует представить функциональный ключ “REGION” и квалификатор “note” для указания, что пептид ALFKNG является кольцевым.

**Соответствующие параграфы ST.26: 7(b)**, 25, 26, 30 и 31

*Параграф 11(a) – Нуклеотидная последовательность из двух цепочек – полностью комплементарная*

**Пример 11(a)-1: Нуклеотидная последовательность из двух цепочек – одинаковой длины**

Заявка на патент описывает следующую последовательность ДНК из двух цепочек:

3′-CCGGTTAACGCTA-5′

5′-GGCCAATTGCGAT-3′

**Вопрос 1: Требует ли ST.26 включения этой последовательности(ей)?**

**ДА**

Каждая представленная нуклеотидная последовательность имеет более 10 конкретно определенных нуклеотидов. По крайней мере, одна цепь должна быть включена в перечень последовательностей, поскольку две цепи этой двухцепочечной нуклеотидной последовательности полностью комплементарны друг другу.

**Вопрос 2: Разрешает ли ST.26 включение последовательности(ей)?**

**ДА**

Хотя последовательность только одной цепи должна быть включена в перечень последовательностей, последовательности обеих цепей могут быть включены, каждая со своим собственным идентификационным номером последовательности.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность ДНК из двух цепочек должна быть представлена в виде одной последовательности или двух отдельных последовательностей. Каждая последовательность, включенная в перечень последовательностей должна быть представлена в направлении от 5 ’к 3’ и должна иметь свой собственный идентификационный номер последовательности.

atcgcaattggcc (верхняя цепь) (SEQ ID NO: 34)

and/or

ggccaattgcgat (нижняя цепь) (SEQ ID NO: 35)

**Соответствующие параграфы ST.26:** 7(a), **11(a)**, и 13

*Параграф 11(b) – Двухцепочечная нуклеотидная последовательность – не полностью комплементарная*

**Пример 11(b)-1: Двухцепочечная нуклеотидная последовательность – разной длины**

Заявка на патент содержит следующий чертеж и подпись к чертежу:

```
5′-tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc-3′
                          ||||||||||||||||
                          gggtaactgantccgc
```

Промоторный участок человеческого гена ABC1 (верхняя цепь) связан с зондом пептидной нуклеиновой кислоты (PNA) (нижняя цепь), где «n» в зонде PNA представляет собой универсальное основание PNA, выбранное из группы, состоящей из 5-нитроиндола и 3-нитроиндола.

**Вопрос 1: Требует ли ST.26 включения последовательности(-тей)?**

**ДА** – промоторный участок человеческого гена ABC1 (верхняя цепь)

Верхняя цепь имеет более десяти перечисленных и «конкретно определенных» нуклеотидов и должна быть включена в перечень последовательностей.

**ДА** – зонд PNA (нижняя цепь)

Нижняя цепь также должна быть включена в перечень последовательностей со своим собственным идентификационным номером последовательности, потому что две цепи не полностью комплементарны друг другу.

Отдельные остатки, которые содержат PNA или «пептидную нуклеиновую кислоту», считаются модифицированными нуклеотидами в соответствии с параграфами 3(f) и 3(g) ST.26. Более того, параграф 16 ST.26 устанавливает, что модифицированные нуклеотиды с нуклеиновой основой идентичной нуклеиновой основе в немодифицированном нуклеотиде, указанном в перечисленных в перечне Приложения I, Раздел 1, Таблица 1, должны быть представлены в последовательности как соответствующие немодифицированные нуклеотиды. За исключением «n» в позиции 6, зонд PNA состоит из (2-аминоэтил)-глициновых единиц, соединённых с нуклеиновыми основаниями, идентичными к нуклеиновым основаниям немодифицированных нуклеотидов, перечисленных в перечне Приложения I, Раздел 1, Таблица 1. Следовательно, каждая позиция, за исключением позиции 6, должна быть представлена символом для соответствующего немодифицированного нуклеотида. Следовательно, зонд PNA имеет более 10 перечисленных и «конкретно определенных» нуклеотидов и должен быть включен в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Верхняя цепь должна быть представлена в перечне последовательностей как:

tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc (SEQ ID NO: 36)

Нижняя цепь представляет собой пептидную нуклеиновую кислоту и поэтому не имеет 3 'и 5' окончания. Согласно параграфу 11, она должна быть включена в перечень последовательностей «в направлении слева направо, которое имитирует направление от 5' к 3'». Следовательно, она должна быть включена в перечень последовательностей как:

cgcctnagtcaatggg (SEQ ID NO: 37)

Каждый нуклеотид в нижней цепи должен быть описан в таблице характеристик с использованием функционального ключа “modified_base” и обязательного квалификатора «mod_base» с аббревиатурой «OTHER». Должен быть включен квалификатор “note” с полным несокращенным названием модифицированного нуклеотида. В соответствии с параграфом 18 ST.26, так как все модифицированные нуклеотиды в последовательности, используют один и тот же фрагмент основы, участок может быть совместно описан в таблице характеристик с дескриптором локации «1..16», функциональным ключом “modified_base”, обязательным квалификатором «mod_base» с аббревиатурой «OTHER» и квалификатором “note” со значением, которое является наиболее

строгим не сокращенным химическим наименованием, охватывающим все из модифицированных нуклеотидов в области, таких как «N- (2-аминоэтил) глициновые нуклеозиды».

Поскольку остаток «n» в позиции 6 включает две возможные вариации нуклеиновой основы - 5-нитроиндол и 3-нитроиндол, он должен быть далее описан как вариант в таблице характеристик с использованием функционального ключа “misc_difference” и обязательного квалификатора “note“ со значением, которое включает полное несокращенное имя двух возможных модифицированных нуклеотидов: «N- (2-аминоэтил) глицин 5-нитроиндол или N- (2-аминоэтил) глицин 3-нитроиндол».

Последовательности пептидных нуклеиновых кислот не встречаются в природе. Поэтому в соответствии с параграфом 83 ST.26 квалификатор “organism” функционального ключа “source” должен иметь значение «синтетическая конструкция», а обязательный квалификатор “mol_type” должен иметь значение “other DNA”.

**Соответствующие параграфы ST.26:** 3(g), 7(a), **11(b)**, 17, 18 и 83

**Пример 11(b)-2: Двухцепочечная нуклеотидная последовательность – без сегмента спаривания оснований**

Заявка на патент описывает следующую последовательность ДНК из двух цепочек:

```
3′-CCGGTTAGCTTATACGCTAGGGCTA-5′
   |||||||       ||||||||||||
5′-GGCCAATATGGCTTGCGATCCCGAT-3′
```

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Каждая цепь представленной двухцепочечной нуклеотидной последовательности имеет более 10 конкретно определенных нуклеотидов. Обе цепи должны быть включены в перечень последовательностей со своим собственным идентификационным номером последовательности, поскольку две цепи не являются полностью комплементарными друг другу.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность каждой цепи должна быть представлена в направлении от 5 ’к 3’, и ей должен быть присвоен собственный идентификационный номер последовательности:

atcgggatcgcatattcgattggcc (top strand) (SEQ ID NO: 38)

и

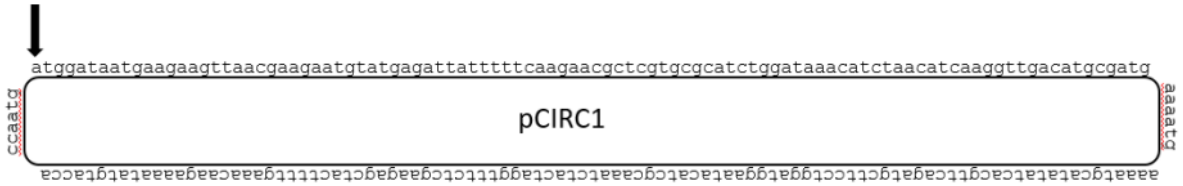
ggccaatatggcttgcgatcccgat (bottom strand) (SEQ ID NO: 39)

**Соответствующие параграфы ST.26:** 7(a), **11(b)**, и 13

*Параграф 12 – Циклическая последовательность нуклеотидов*

**Пример 12-1: Циклическая последовательность нуклеотидов**

Заявка на патент содержит следующий рисунок, раскрывающий последовательность ДНК плазмиды pCIRC1:




**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Перечисленная последовательность нуклеотидов содержит более 10 конкретно определенных нуклеотидов. Следовательно, последовательность должна быть включена в перечень последовательностей в соответствии с требованиями параграфа 7(a) ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Согласно параграфу 12 ST.26, когда последовательности нуклеотидов имеют кольцевую конфигурацию, заявитель должен выбрать нуклеотид в позиции остатка номер 1. Для целей данного примера будет использоваться остаток «a» в положении 1, обозначенный стрелкой на рисунке. Однако любой остаток может быть выбран в качестве положения 1. Если остаток указан стрелкой как положение 1, последовательность должна быть включена в перечень последовательностей как:

atggataatgaagaagttaacgaagaatgtatgagattatttttcaagaacgctcgtgcgcatctggataaa
catctaacatcaaggttgacatgcgatgaaaatgaaaatgcatatatcacgttcagatgcttcctggatgga
atacatcgcaaatctactaggtttctcgaagagctacttttgaaacaagaaaatatgtaccaccaatg
(SEQ ID NO: 98)

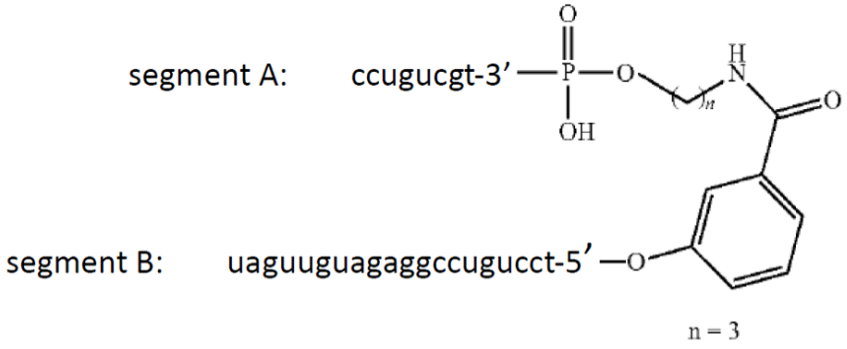
Последовательность должна быть дополнительно описана с использованием функционального ключа «misc_feature» с местоположением «212^1», что указывает на то, что последний остаток в последовательности, позиция 212, связан с остатком 1. Квалификатор «note» должен быть включен со значением, указывающим, что молекула является круговой.

**Соответствующие параграфы ST.26:** 7(a), **12** и Приложение 1, Часть 5, функциональный ключ 5.15

*Параграф 14 – Символ “t” рассматривается как урацил в RNA*

**Пример 14-1: Символ “t” представляет урацил в RNA**

В патентной заявке описана следующая композиция:



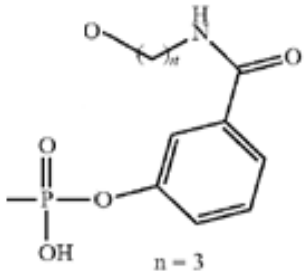

Где фрагмент A и фрагмент B являются RNA последовательностями.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(тей)?**

**ДА – фрагмент B**

**НЕТ – фрагмент A**

Представленная последовательность содержит два фрагмента конкретно определённых нуклеотида, разделённых следующей “линкерной” структурой:




Линкерная структура не является нуклеотидом согласно параграфу 3 (g); поэтому каждый сегмент должен рассматриваться как отдельная последовательность. Сегмент B содержит более 10 конкретно определенных нуклеотидов, и параграф 7(a) ST.26 требует включения в перечень последовательностей. Фрагмент A содержит только восемь конкретно определенных нуклеотидов и, следовательно, не требуется включать его в перечень последовательностей.

**Вопрос 2: Разрешает ли ST.26 включение этой последовательности (ей)?**

**ДА**

Поскольку фрагмент A содержит менее 10 конкретно определенных нуклеотидов, его включение в перечень последовательностей является опциональным (оставляется на усмотрение заявителя) согласно параграфу 8 Стандарта ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Сегмент B представляет собой молекулу RNA; следовательно, элемент «INSDSeq_moltype» должен быть «RNA». Символ «u» не должен использоваться для представления урацила в молекуле RNA в перечне последовательностей. Согласно параграфу 14, символ «t» будет толковаться как урацил в RNA, если далее не описан в таблице характеристик. Соответственно, сегмент B должен быть включен в перечень последовательностей как:

tcctgtccggagatgttgat (SEQ ID NO: 40)

Тимин в RNA считается модифицированным нуклеотидом, то есть модифицированным урацилом, и должен быть представлен в последовательности как «t» и дополнительно описан в таблице характеристик, как указано в параграфе 19 данного Стандарта. Соответственно, тимин в позиции 1 должен быть дополнительно описан с использованием функционального ключа “modified_base”, квалификатора «mod_base» с «OTHER» в качестве значения квалификатора и квалификатора «note» с «тимином» в качестве значения квалификатора.

Тимин, то есть модифицированный урацил, на позиции 1 также должен быть дополнительно описан в таблице характеристик с использованием функционального ключа «misc_feature» и квалификатора «note» со значением, например, “The 5 ' фосфат тимидина присоединяется через линкер 3-гидроксибензамидо-N-пропил-3-фосфат к другой нуклеотидной последовательности». Там, где это практически возможно, другая последовательность может быть непосредственно указана как значение в квалификаторе “note”.

Если сегмент A будет включен в перечень последовательностей, его следует представить как:

cctgtcgt (SEQ ID NO: 113)

Элемент “INSDSeq_moltype” должен быть "RNA". Символ “u” не должен использоваться для представления урацила в молекуле RNA в перечне последовательностей. В соответствии с параграфом 14 символ "t" будет интерпретироваться как урацил в РНК при отсутствии дальнейших пояснений.

В соответствии с параграфом 19 ST.26 тимин в позиции 8 должен быть далее описан в таблице характеристик использованием функционального ключа “modified_base”, квалификатора “mod_base” со значением квалификатора “OTHER” и квалификатора “note” с “thymine” в качестве значения квалификатора.

Тимин, т.е. модифицированный урацил, в позиции 8 также должен быть далее описан в таблице характеристик, используя функциональный ключ “misc_feature” и квалификатор “note” со значением, например:

«3' конец последовательности присоединяется через линкер 3-гидроксибензамидо-N-пропил-3- фосфат к 5' конец последовательности tccuguccggagauguugau»

**Соответствующие параграфы ST.26:** 3(g), 7(a), 8, 13, **14**, 19, и 54

*Параграф 27 – Следует использовать наиболее ограничивающий неоднозначный символ*

**Пример 27-1: Сокращённая формула для аминокислотной последовательности**

(GGGz), где z - любая аминокислота.

**Вопрос 1: Требует ли ST.26 включения последовательности(тей)?**

**ДА**

Последовательность раскрыта в виде формулы. $(GGGz)_2$ - это просто сокращенный способ представления последовательности GGGzGGGz. Обычно последовательность сначала расширяется, и после чего даётся определение любой переменной, то есть «z».

В последовательности используется “нестандартный” символ «z». Определение «z» должно быть установлено из поясняющей информации к последовательности в описании, которая определяет этот символ как любую аминокислоту (см. Введение к этому документу). Пример не предусматривает какого-либо ограничения для «z», например, что оно является одинаковым в каждом случае.

Пептид в примере имеет восемь перечисленных аминокислот, шесть из которых являются конкретно определенными остатками глицина, а оставшиеся две представляют собой переменную «z», которая должна быть представлена в этой последовательности с помощью общепринятого символа «X». Параграф 7(b) ST.26 требует включения последовательности в перечень последовательностей в виде единой последовательности с одним идентификационным номером последовательности.

Обратите внимание, что последовательность все еще охватывается параграфом 7(b), несмотря на тот факт, что перечисленные и конкретно определенные остатки не являются смежными.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

В последовательности используется “нестандартный” символ «z», который согласно описанию представляет собой любую аминокислоту. Традиционный символ, используемый для обозначения «любая аминокислота», представляет собой «X». Следовательно, последовательность должна быть представлена как одна расширенная последовательность:

GGGXGGGX (SEQ ID NO: 41)

В соответствии с параграфом 27, “X” будет рассматриваться как одна из немодифицированных L-альфа-аминокислот “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, или “V”, кроме случаев, когда он используется с последующим описанием в таблице характеристик. В данном примере «X» представляет собой «любую аминокислоту», которая шире по охвату чем значение «X» по умолчанию, данное в параграфе 27, поскольку «любая аминокислота» охватывает модифицированные или другие аминокислоты. Поэтому каждая «X» должна быть аннотирована с помощью функционального ключа VARIANT и квалификатора note со значением «X может быть любой аминокислотой».

Хотя пример не раскрывает, что «z» является одной и той же аминокислотой на обоих позициях в расширенной последовательности, если «z» была раскрыта как одна и та же аминокислота на обеих позициях, то в значении квалификатора «note» для «X» в позиции 4 следует указать, что «X» может быть любой аминокислотой, пока она является той же самой аминокислотой, как в позиции 8. В значении квалификатора «note» для «X» в позиции 8 следует указывать, что она может быть любой аминокислотой, пока она является той же самой аминокислотой, как в позиции 4.

**Соответствующие параграфы ST.26:** 3(c), 7(b) and **27**

**Пример 27-2: Сокращенная формула - менее четырех конкретно определенных аминокислот**

Формула пептида $(Gly\text{-}Gly\text{-}Gly\text{-}z)_n$

Описание далее указывает, что z является любой аминокислотой и

(i) переменная n – любой длины; или

(ii) переменная n - 2-100, предпочтительно 3

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**НЕТ**

Рассмотрение обоих раскрытых вариантов осуществления (i) и (ii) представленной формулы пептида показывает, что «n» может быть «любой длины»; следовательно, наиболее охватывающий вариант «n» является неопределенным. Поскольку «n» является неопределенным, формула пептида не может быть расширена до определенной длины, и, следовательно, должна быть рассмотрена нерасширенная формула.

Представленный в нерасширенной формуле пептид («n» = 1) содержит три конкретно определенные аминокислоты, каждая из которых представляет собой Gly, и символ «z». Условно принятый символ «Z» является символом «глютамина или глутаминовой кислоты»; тем не менее, пример определяет «z» как «любую аминокислоту» (см. Введение к этому документу). В соответствии со стандартом ST.26 аминокислота, которая конкретно не определена, представляется как «X». На основании этого анализа представленный пептид, то есть GGGX, не содержит четырех конкретно определенных аминокислот. Поэтому параграф 7(b) ST.26 не требует включения данной последовательности, несмотря на то, что в некоторых вариантах осуществления «n» также определено как конкретное числовое значение.

**Вопрос 2: Разрешает ли ST.26 включение последовательности(ей)?**

**ДА**

В соответствии с параграфом 8, последовательности, содержащие менее, чем четыре конкретно определенных аминокислоты, может быть включена в перечень последовательностей. Поэтому перечисленный пептид в нерасширенной формуле (“n” = 1), который даёт три конкретно определенных аминокислоты, может быть включен в перечень последовательностей.

В дополнение или альтернативно, варианты со конкретными числовыми значениями для изменяемого “n”, т.е. меньше ограничения из 2, больше ограничения из 100 и точное значение 3, могут быть включены в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

В соответствии с введением к Приложению VI, когда длина последовательности может варьироваться из-за изменения количества копий, самый длинный вариант последовательности, считается наиболее охватывающим. В данном примере самым длинным вариантом, который определен (т.е. имеет точное значение “n”), является вариант, в котором n=100. Поэтому последовательность, содержащая 100 копий GGGX является предпочтительной для включения в перечень последовательностей (SEQ ID NO: 42). Последовательность следует аннотировать функциональным ключом “VARIANT” и

квалификатором “note” со значением, указывающим, что до 98 копий GGGX могут быть удалены. Настоятельно рекомендуется включение дополнительных конкретных вариантов осуществления, которые являются ключевой частью изобретения.

Согласно параграфу 27, «X» будет истолковываться как любая из немодифицированных L-альфа-аминокислот «A», «R», «N», «D», «C», «Q», «E», «G», «H», «I», «L», «K», «M», «F», «P», «O», «S», «U», «T», «W», «Y» или «V», за исключением случаев, когда она используется с дальнейшим описанием в таблице характеристик. «X» в SEQ ID NO: 42 представляет собой «любую аминокислоту», что шире по охвату, чем значение «X» по умолчанию, данное в параграфе 27, которое охватывает и модифицированные и другие кислоты. Следовательно, каждая «X» должна быть индивидуально аннотирована с помощью функционального ключа VARIANT и квалификатора “note” со значением: «X может быть любой аминокислотой».

Кроме того, или менее предпочтительное представление, следующая последовательность может быть включена в перечень последовательностей:

GGGX (SEQ ID NO: 118)

«X» следует аннотировать тем же путем, как указано выше для SEQ ID NO: 42.

**ПРЕДУПРЕЖДЕНИЕ:** Предпочтительное представление последовательности, указанной выше, направлено на предоставление перечня последовательностей на дату подачи заявки на патент. К перечню последовательностей, представленному после даты подачи патентной заявки, такое же представление может не применяться, поскольку необходимо рассмотреть вопрос о том, может ли предоставленная информация быть рассмотрена Ведомством на предмет добавления объекта (признака) к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 3(c), 7(b), 26 и **27**

**Пример 27-3: Сокращённая формула – четыре или более конкретно определённых аминокислот**

Формула пептида $(Gly\text{-}Gly\text{-}Gly\text{-}z)_n$

Где z – любая аминокислота, а переменная n составляет от 2 до100, предпочтительно 3.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Пептид, представленный формулой, содержит три конкретно определенные аминокислоты, каждая из которых представляет собой Gly, и символ «z». Обычно «Z» является символом «глутамина или глутаминовой кислоты»; однако описание в этом примере определяет «z» как «любую аминокислоту» (см. Введение к этому документу). В соответствии с ST.26 аминокислота, которая конкретно не определена, обозначается как «X». На основании этого анализа представленный повторный пептид не содержит четырех специфически определенных аминокислот. Однако в описании приводится конкретное числовое значение для переменной «n», т.е. нижний предел 2 и верхний предел 100. Следовательно, в примере раскрыт пептид, имеющий по меньшей мере шесть конкретно определенных аминокислот в последовательности GGGzGGGz, для которого по ST.26 требуется включение в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность в перечень последовательностей?**

Поскольку «z» представляет любую аминокислоту, условно принятым символом, используемым для обозначения четвертой и восьмой аминокислот, является «X».

ST.26 требует включения в перечень последовательностей только одной последовательности, которая была раскрыта перечислением её остатков. Следовательно, по крайней мере одна последовательность, содержащая любую из 2, 3 или 100 копий GGGX, должна быть включена в перечень последовательностей; однако наиболее охватывающая последовательность, содержащая 100 копий GGGX, является предпочтительной (SEQ ID NO: 42) (см. Введение к этому документу). В последнем случае в последующей аннотации можно указать, что до 98 копий GGGX могут быть удалены. Настоятельно рекомендуется включать две дополнительные последовательности, содержащие 2 и 3 копии GGGX соответственно (SEQ ID NO: 44-45).

Согласно параграфу 27 «X» будет истолкован как любая из немодифицированных L-альфа-аминокислот «A», «R», «N», «D», «C», «Q», «E», «G», «H», «I», «L», «K», «M», «F», «P», «O», «S», «U», «T», «W», «Y» или «V», за исключением случаев, когда она используется с последующим описанием в таблице характеристик. «X» в этом примере представляет собой «любую аминокислоту», что шире по охвату, чем значение «X» по умолчанию, данное в параграфе 27, поскольку «любая аминокислота» охватывает модифицированные или другие аминокислоты. Следовательно, каждая «X» должна быть индивидуально аннотирована с помощью функционального ключа VARIANT и квалификатора note” со значением: «X может быть любой аминокислотой».

Кроме того, в данном примере не показано, что переменная «z» одинакова в каждом из двух случаях в расширенной последовательности. Однако, если «z» раскрывается как одна и та же аминокислота во всех местоположениях, то функциональный ключ VARIANT и квалификатор note должны указывать, что «X» на всех позициях может быть любой аминокислотой, если они одинаковы во всех местоположениях.

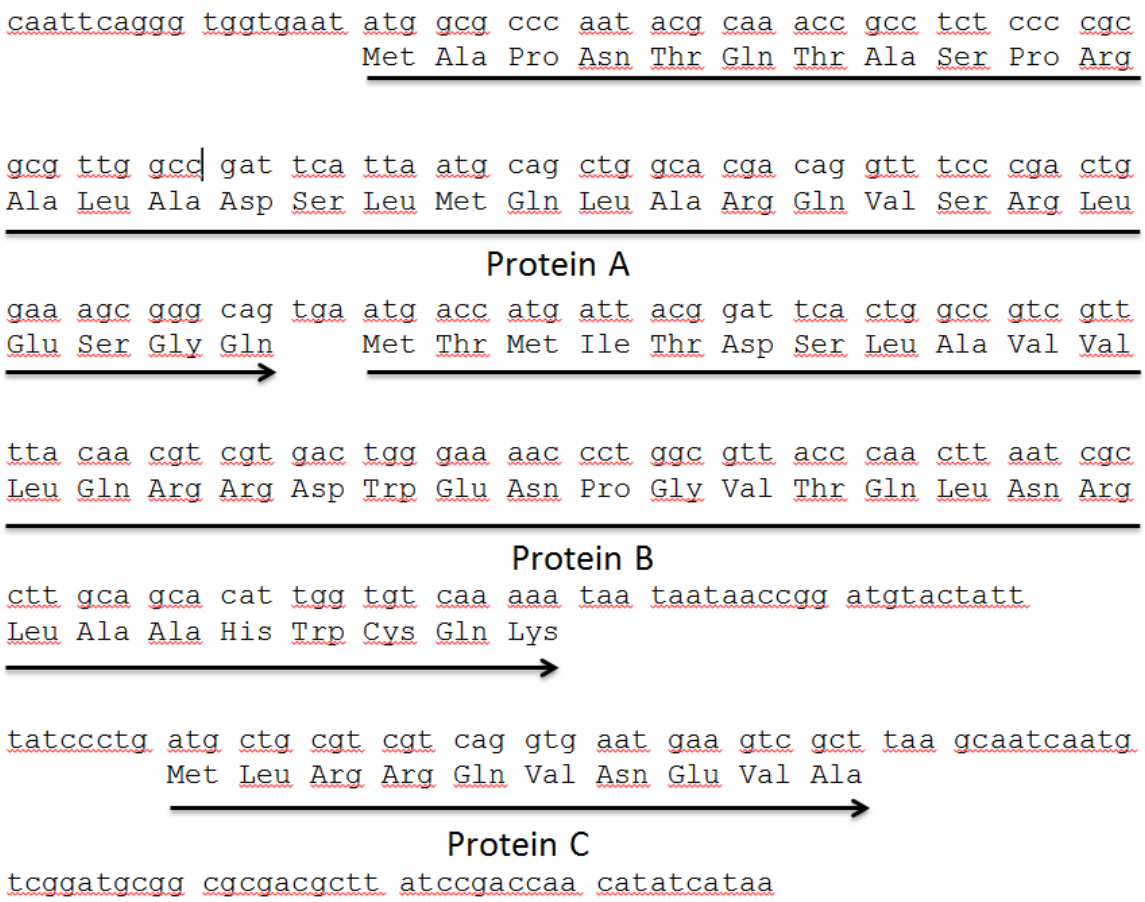
**ПРЕДУПРЕЖДЕНИЕ:** Предпочтительное представление последовательности, указанной выше, направлено на предоставление перечня последовательностей на дату подачи патентной заявки. К перечню последовательностей, представленному после даты подачи патентной заявки, такое же представление может не применяться, поскольку необходимо учесть, может ли предоставленная информация быть рассмотрена Ведомством на предмет добавления объекта (признака) к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 3(c), 7(b), 26 и **27**

*Параграф 28 – Аминокислотные последовательности, разделённые символами внутреннего разграничения*

***Пример 28-1: Кодирующая нуклеотидная последовательность и кодируемая аминокислотная последовательность***

В патентной заявке описана следующая последовательность:




**Вопрос 1: Требует ли ST.26 включения данной последовательности(-ей)?**

**ДА**

В заявке описана нуклеотидная последовательность, содержащая стоп - кодоны, которая кодирует три отдельные аминокислотные последовательности.

Представленная нуклеотидная последовательность содержит более 10 условно определенных нуклеотидов и должна быть включена в перечень последовательностей в виде одной последовательности.

Что касается кодируемых аминокислотных последовательностей, в параграфе 28 требуется, чтобы аминокислотные последовательности, разделенные внутренним символом-терминатором, таким как пробел, были включены в виде отдельных

последовательностей. Поскольку каждый из «Протеина A», «Протеина B» и «Протеина C» содержит четыре или более конкретно определенных аминокислот, параграф 7(b) ST.26 требует, чтобы каждая из них была включена в перечень последовательностей, и ей должен быть присвоен её собственный идентификационный номер последовательности.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Нуклеотидная последовательность должна быть включена в перечень последовательностей как:

caattcagggtggtgaatatggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggccaggcaggt gagcaggctggaaagcgggcagtgaatgaccatgattacggattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggc gttacccaacttaatcgccttgcagcacattggtgtcaaaaataataataaccggatgtactatttatccctgatgctgcgtcgtcaggtgaa tgaagtcgcttaagcaatcaatgtcggatgcggcgcgacgcttatccgaccaacatatcataa (SEQ ID NO: 46).

Нуклеотидная последовательность должна быть в дальнейшем описана с использованием функционального ключа «CDS» для каждого из трех белков, а элемент INSDFeature_location должен идентифицировать местоположение каждой кодирующей последовательности, включая стоп-кодон. Кроме того, для каждого функционального ключа «CDS» должен быть включен квалификатор «translation» с аминокислотной последовательностью белка в качестве значения квалификатора. В заявке не раскрывается таблица генетического кода, которая применяется к переводу (см. Приложение 1, Раздел 9, Таблица 7). Если используется Таблица стандартных кодов, то квалификатор «transl_table» не нужен; однако, если применяется другая таблица генетических кодов, тогда для квалификатора «transl_table» должно быть указано соответствующее значение квалификатора из таблицы 7. Наконец, квалификатор «protein _id» должен быть включен со значением квалификатора, указывающим идентификационный номер последовательности каждой из преобразованных аминокислотных последовательностей.

Аминокислотные последовательности должны быть включены в виде отдельных последовательностей, каждой из которых присвоен свой собственный идентификационный номер последовательности:

MAPNTQTASPRALADSLMQLARQVSRLESGQ (SEQ ID NO: 47)

MTMITDSLAVVLQRRDWENPGVTQLNRLAAHWCQK (SEQ ID NO: 48)

MLRRQVNEVA (SEQ ID NO: 49)

ПРИМЕЧАНИЕ. См. «Пример 90-1. Аминокислотная последовательность, кодируемая кодирующей последовательностью с интронами», для иллюстрации преобразованной аминокислотной последовательности, представленной в виде одной последовательности.

**Соответствующие параграфы ST.26:** 7, 26, **28**, 57, 89-92

*Параграф 29 – Представление модифицированной аминокислоты*

**Пример 29-1: Наиболее ограничивающий неоднозначный символ для “other” («другой») аминокислоты**

Заявка на патент описывает следующую последовательность:

Ala-Hse-$X_1$-$X_2$-$X_3$-$X_4$-Tyr-Leu-Gly-Ser

Где, $X_1$= Ala or Gly,

$X_2$= Ala or Gly,

$X_3$= Ala or Gly,

$X_4$= Ala or Gly, и

Hse = Homoserine

**Вопрос 1: Требует ли ST.26 включения последовательности(-тей)?**

**ДА**

Приведённый пептид содержит пять конкретно определенных аминокислот. Символ «X» обычно используется для обозначения двух аминокислот в качестве альтернативы (см. Введение к данному документу).

Поскольку существует по меньшей мере пять специфически определенных аминокислот, а именно, Ala, Tyr, Leu, Gly и Ser, параграф 7 (b) ST.26 требует, чтобы последовательность была включена в перечень последовательностей. Символ «X»

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Гомосерин на позиции 2 данной последовательности является альфа-аминокислотой с боковой цепью -CH2-CH2-OH. Это – модифицированная аминокислота, имеющая боковую цепь, которая не идентична боковой цепи любой из L-альфа-аминокислот, перечисленных в Приложении I, Раздел 3, Таблица 3, В соответствии с параграфом 29 ST.26 модифицированная аминокислота, когда это возможно, должна быть представлена в последовательности как немодифицированная аминокислота. Символ «X» следует использовать для представления модифицированной аминокислоты, когда не подходит ни один из других символов в Приложении I, Раздел 3, Таблица 3. Следовательно, гомосерин может быть представлен с помощью «X» или наиболее подходящего символа аминокислоты.

$X_1$-$X_4$ являются вариантами позиций на остатках 3-6, каждая из которых может быть A или G. Наиболее ограничивающим неоднозначным символом для альтернатив A или G является «X». Следовательно, последовательность может быть представлена как:

AXXXXXYLGS (SEQ ID NO: 50)

Настоятельно рекомендуется включать любые конкретные последовательности, имеющие существенное значение для описания или формулы изобретения, как обсуждалось во введении к настоящему документу.

В соответствии с параграфом 30, Hse должен быть описан с использованием функционального ключа “SITE” и квалификатора “note”. Поскольку аминокислота Hse не найдена в Приложении I, Раздел 4, Таблица 4, значение квалификатора “note” должно быть полным несокращенным наименованием гомосерина.

Согласно параграфу 27, поскольку $X_1$-$X_4$ представляют собой альтернативу только 2 аминокислотам, требуется последующее описание. Параграф 96 указывает, что функциональный ключ «VARIANT» должен использоваться с квалификатором «note» и значением квалификатора «A или G». Согласно параграфу 34 ST.26, т.к. эти позиции являются смежными и имеют одинаковое описание, они могут быть совместно описаны с использованием синтаксиса «3..6» в качестве дескриптора местоположения в элементе INSDFeature_location.

**Соответствующие параграфы ST.26:** 3(a), 7(b), 25-27, **29**, 30, 34, 66, 70, 71 и 96-97

**Пример 29-2: Использование соответствующей немодифицированной аминокислоты**

Заявка на патент описывает следующую последовательность:

Ala-Hyl-Tyr-Leu-Gly-Ser-Nle-Val-Ser-Orn

Где Hyl = гидроксилизин (посттрансляционная модификация лизина), Nle = норлейцин и Orn = орнитин.

**Вопрос 1: Требует ли ST.26 включения данной последовательности(-тей)?**

**ДА**

Приведённый пептид содержит более четырех конкретно определенных аминокислот; следовательно, последовательность должна быть включена в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Гидроксилизин в положении 2, норлейцин в положении 7 и орнитин в положении 10 все являются «модифицированными аминокислотами». Во-первых, мы должны рассмотреть каждую модифицированную аминокислоту с точки зрения параграфа 29, чтобы определить, следует ли её представлять в последовательности соответствующей немодифицированной аминокислотой или переменной «X». Ни гидроксилизин, ни норлейцин, ни орнитин не содержат боковой цепи, идентичной таковой в какой-либо аминокислоте, перечисленной в Приложении I, Раздел 3, Таблица 3. Следовательно, каждую модифицированную аминокислоту «следует представлять в последовательности как соответствующую немодифицированную аминокислоту, когда это возможно."

На выбор заявителя оставляется решить, будет ли модифицированная аминокислота представлена соответствующим немодифицированным остатком или переменной «X». Тем не менее, следует принять во внимание следующее руководство: Если боковая цепь аминокислоты модифицируется путем добавления небольшого фрагмента, и основная структура боковой цепи остаётся в основном сохраненной, тогда рекомендуется представление немодифицированной аминокислотой. Если модифицированная боковая цепь структурно сильно отличается от таковой в соответствующей немодифицированной аминокислоте, то рекомендуется обозначение «X».

Структура гидроксилизина почти идентична лизину, за исключением того, что третий углерод в боковой цепи модифицирован гидроксильной группой. Поскольку основная структура соответствующего немодифицированного остатка лизина не изменена, рекомендуется, чтобы гидроксилизин представлялся в последовательности как «K», а не «X».

Норлейцин представляет собой изомер лейцина. Боковая цепь лейцина представляет собой цепь из 4 атомов углерода, разветвленную у второго атома углерода. Хотя норлейцин также имеет 4-углеродную боковую цепь, она является линейной и неразветвленной. Следовательно, норлейцин — это не просто результат фрагмента, добавленного к лейцину, а совершенно другая (хотя и родственная) структура. Поэтому рекомендуется, чтобы норлейцин был представлен в последовательности символом «X», а не “L”.

Структура орнитина в наибольшей степени подобна лизину; лизин имеет 4-х углеродную боковую цепь, а орнитин имеет 3-х углеродную боковую цепь. Орнитин не может быть создан путем добавления маленького фрагмента к лизину или другой аминокислоте,

перечисленной в Приложении I, Таблица 3, поэтому рекомендуется, чтобы орнитин был представлен в последовательности символом «X», а не “K”.

Следовательно, данную последовательность в перечне последовательностей следует представлять как:

AKYLGSXVSX (SEQ ID NO: 51)

Параграф 30 требует дополнительного аннотирования каждой модифицированной аминокислоты.

Гидроксилизин представляет собой посттрансляционную модификацию лизина. Следовательно, он должен быть описан с использованием функционального ключа «MOD_RES» вместе с квалификатором «note», описывающим модификацию. Обратите внимание, что «гидроксилизин» указан в Приложении 1, Разделе 4, Таблице 4 «Перечень модифицированных аминокислот». Следовательно, в значении квалификатора «note» нужно исвпользовать аббревиатуру «Hyl» вместо полного несокращенного названия «гидроксилизин».

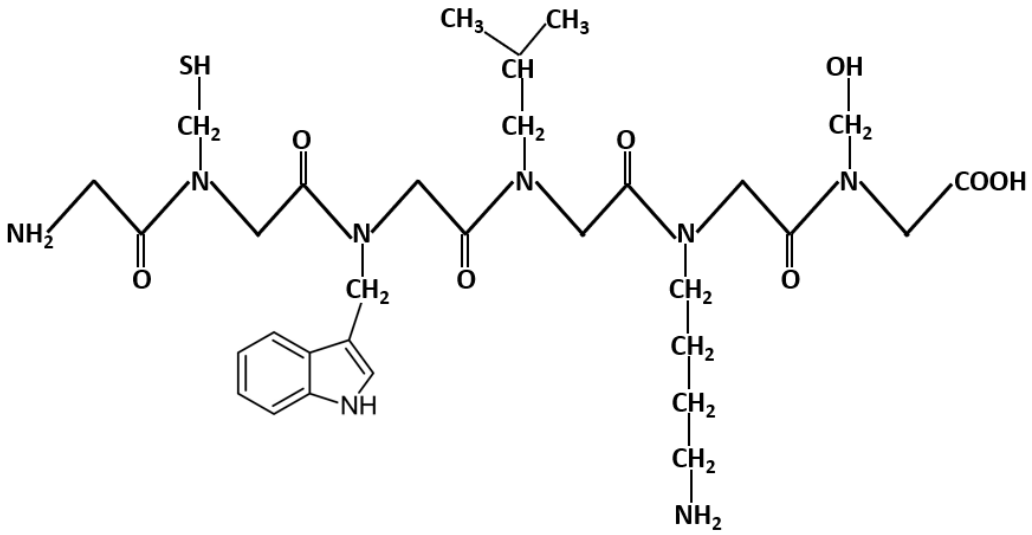
Норлейцин не является посттрансляционно модифицированным остатком, поэтому он должен быть описан с использованием функционального ключа “SITE” вместе с квалификатором «note», описывающим модификацию. Обратите внимание, что «норлейцин» также указан в Приложении 1, Раздел 4, Таблица 4. Поэтому в значении квалификатора «note» нужно использовать аббревиатуру «Nle» вместо полного, несокращенного названия «норлейцин».

Орнитин также не является посттрансляционно модифицированным остатком, поэтому его следует описывать с помощью функционального ключа “SITE” вместе с квалификатором «note», описывающим модификацию. Орнитин указан в Приложении 1, Разделе 4, Таблице 4 с сокращением “Orn”. Следовательно, в значении квалификатора «note» нужно использовать сокращение “Orn”.

**Соответствующие параграфы ST.26:** 3(a), 3(e), 7(b), 29 и 30

**Пример 29-3 Пептоидный полимер I**

Заявка на патент раскрывает следующий пептидомиметик “пептоидный” полимер из N-замещенного глицина:




**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**
**ДА**

N-замещенный глицин охватывается определением «аминокислоты», данным в параграфе 3(a) данного Стандарта. Параграф 29 данного Стандарта требует представления модифицированной аминокислоты, которая имеет единственную углеродсодержащую боковую цепь, идентичную боковой цепи L-альфа-аминокислоты, указанной в Приложении I, раздел 3, таблица 3, которая расположена на азотистой части аминокислотного остатка, с помощью символа соответствующей немодифицированной аминокислоты. В раскрытой пептоидной структуре:

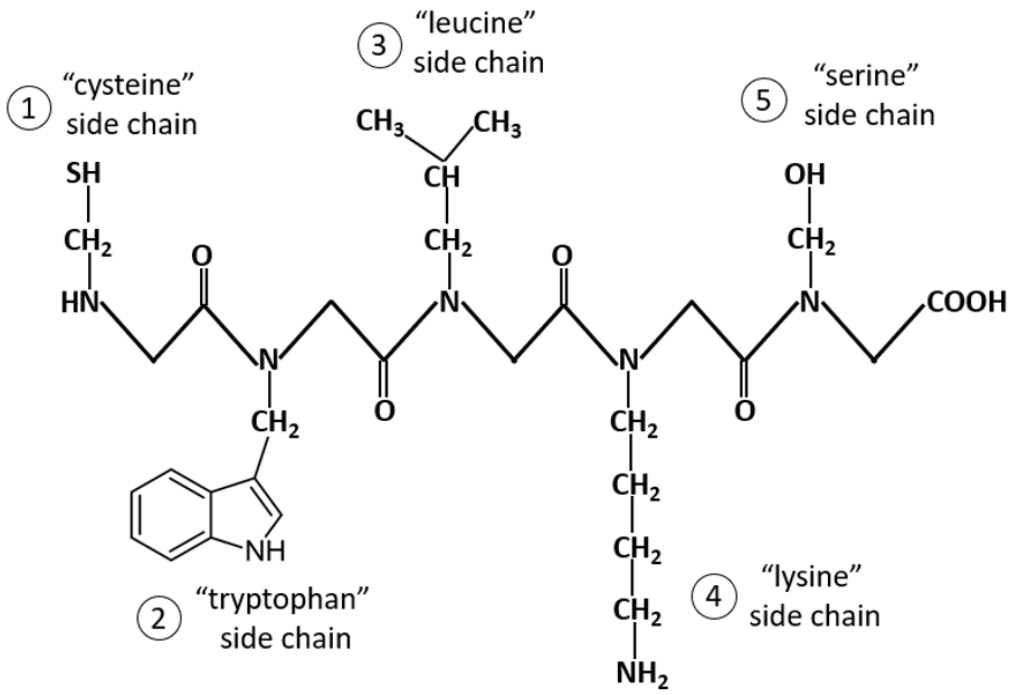
- боковая цепь, прикрепленная к азоту первого остатка, идентична боковой цепи цистеина;

- боковая цепь, прикрепленная к азоту второго остатка, идентична боковой цепи триптофана;

- боковая цепь, прикрепленная к азоту третьего остатка, идентична боковой цепи лейцина;

- боковая цепь, прикрепленная к азоту четвертого остатка, идентична боковой цепи лизина;

- боковая цепь, прикрепленная к азоту пятого остатка, идентична боковой цепи серина.




Следовательно, раскрытый пептид содержит пять модифицированных аминокислот, которые являются конкретно определенными, и эта последовательность должна быть включена в перечень последовательностей, как требуется параграфом 7(b) Стандарта ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Данная последовательность должна быть включена в перечень последовательностей как:

CWLKS (SEQ ID NO: 114)

Каждый остаток должен быть описан в таблице характеристик с использованием функционального ключа “SITE“ и обязательного квалификатора “note“ со значением, которое описывает остаток с помощью его химического наименования.

- “**C**” в позиции 1 может быть описан как “N-(sulfanylmethyl)glycine”.
- “**W**” в позиции 2 может быть описан как “N-(1H-indol-3-ylmethyl)glycine”.
- **“L”** в позиции 3 может быть описан как “N-(isobutyl)glycine”.
- **“K”** в позиции 4 может быть описан как “N-(aminobutyl)glycine”.
- **“S”** в позиции 5 может быть описан как “N-(hydroxymethyl)glycine”.

**ПРЕДУПРЕЖДЕНИЕ:** предпочтительное представление последовательностей, указанных выше, и их аннотаций предназначено для представления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 3(a), 3(e), 7(b), 29, 30 и 34

**Пример 29-4 Пептоидный полимер II**

Заявка на патент раскрывает следующее:

Аналоги пептидов, охватывающие остатки N-замещенного глицина, такие как последовательность NLys-Npa-N4hb-Nnle-NLys-NTrp, являются значительно более стабильными *in vivo* чем натуральные пептиды.

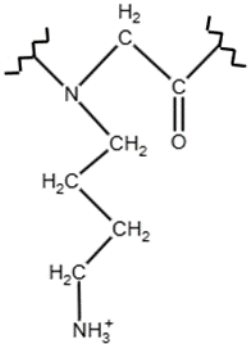
Последующее раскрытие указывает, что:

- “NLys” является N-(4-aminobutyl)glycine
- “Npa” является N-(benzyl)glycine
- “N4hb” является N-(4-hydroxybutyl)glycine
- “Nnle” является N-(butyl)glycine
- “NTrp” является N-[1-(indol-3-yl)methyl]glycine

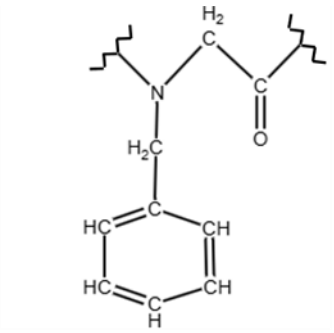
**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**
**ДА**

Указанная последовательность охватывает шесть остатков, имеющих нерекомендованные аминокислотные сокращения “NLys”, “Npa”, “N4hb”, “Nnle” и “NTrp”.

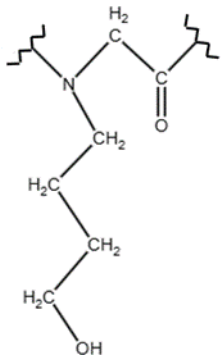
“NLys” является N-(4-aminobutyl)glycine со структурой:



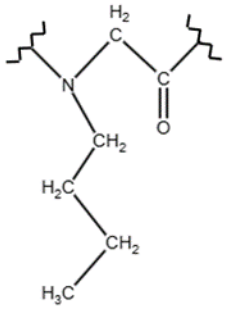

“Npa” является N-(benzyl)glycine со структурой:



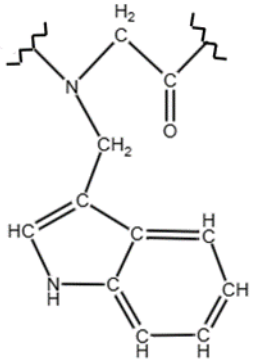
“N4hb” является N-(4-hydroxybutyl)glycine со структурой:



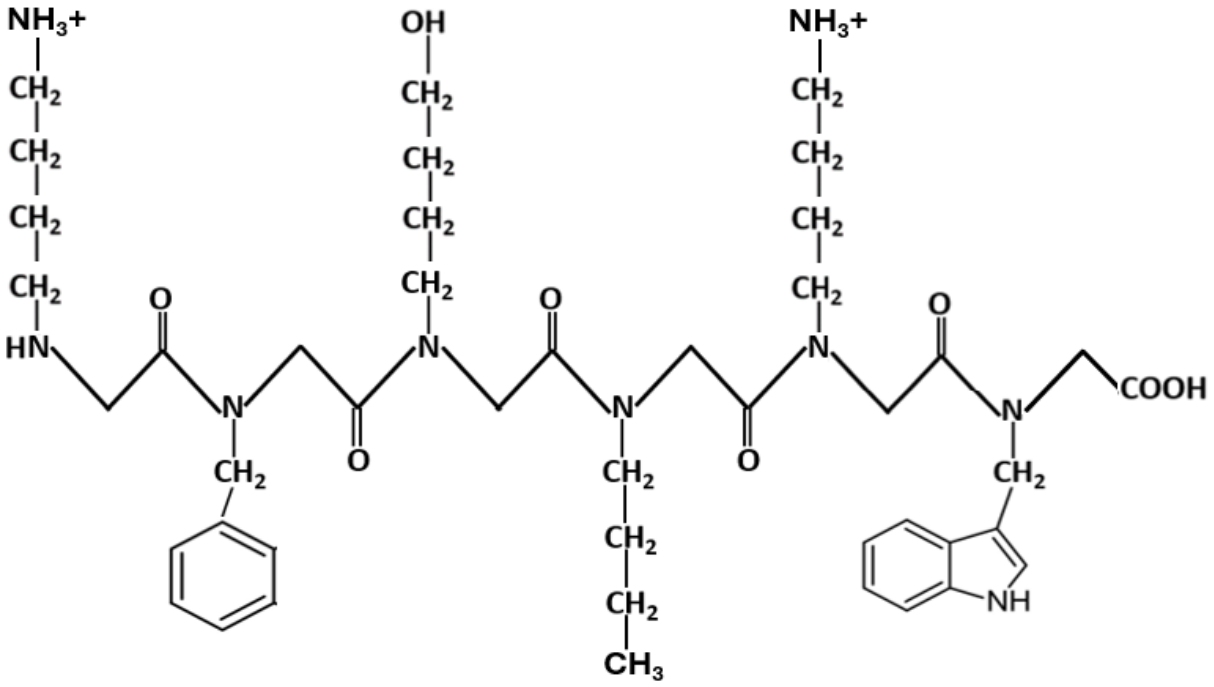
“Nnle” является N-(butyl)glycine со структурой:



“NTrp” является N-[1-(indol-3-yl)methyl]glycine со структурой:



Соответственно, структура данной раскрытой последовательности выглядит так:




Каждый остаток является N-замещенным глицином, который охватывается определением «аминокислоты», данным в параграфе 3(a) данного Стандарта. Параграф 29 данного Стандарта требует представления модифицированной аминокислоты, которая имеет единственную углеродсодержащую боковую цепь, идентичную боковой цепи L-альфа-аминокислоты, указанной в Приложении I, раздел 3, таблица 3, которая расположена на аминоазоте остатка, с помощью символа соответствующей немодифицированной аминокислоты. В раскрытой последовательности:

- боковая цепь, прикрепленная к азоту первого остатка, идентична боковой цепи лизина, следовательно, этот остаток должен быть представлен в последовательности как “K”;

- боковая цепь, прикрепленная к азоту второго остатка, идентична боковой цепи фенилаланина, следовательно, этот остаток должен быть представлен в последовательности как “F”;

- боковая цепь, прикрепленная к азоту третьего остатка, НЕ идентичен боковым цепям какой-либо из L-альфа-аминокислот, перечисленных в перечне Приложения I, Раздел 3, Таблица 3, следовательно, этот остаток может быть представлен в последовательности как “X” или наиболее подходящий аминокислотный символ;

- боковая цепь, прикрепленная к азоту четвертого остатка, НЕ идентична боковым цепям какой-либо из L-альфа-аминокислот, перечисленных в перечне Приложения I, Раздел 3, Таблица 3, следовательно, этот остаток может быть представлен в последовательности как “X” или наиболее подходящий аминокислотный символ;

- боковая цепь, прикрепленная к азоту пятого остатка, идентична боковой цепи лизина, следовательно, этот остаток должен быть представлен в последовательности как “K”;

- боковая цепь, прикрепленная к азоту шестого остатка, идентична боковой цепи триптофана, следовательно, этот остаток должен быть представлен в последовательности как “W”.

С точки зрения вышеприведенного анализа данная раскрытая последовательность содержит минимум четыре конкретно определенных аминокислоты и должна быть включена в перечень последовательностей, как этого требует параграф 7(b) ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Если “X” представляется наиболее подходящим символом для третьего и четвертого остатков, то последовательность должна быть включена в перечень последовательностей как:

KFXXKW (SEQ ID NO:115)

Каждый остаток должен быть описан в таблице характеристик с использованием функционального ключа “SITE“ и обязательного квалификатора “note“ со значением, которое описывает остаток как его химическое наименование.

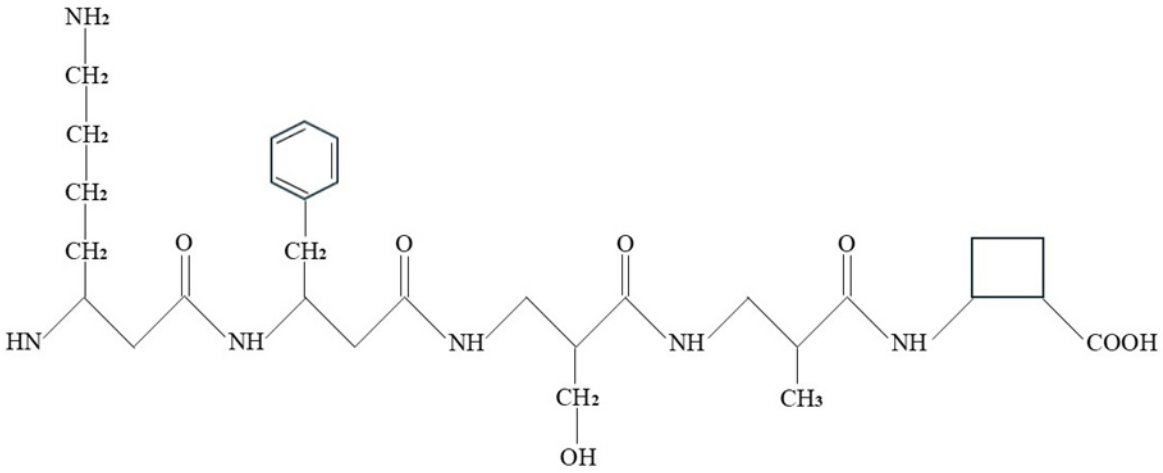
- The “K” в позиции 1 может быть описан как “N-(4-aminobutyl)glycine.”
- The “F” в позиции 2 может быть описан как “N-(benzyl)glycine.”
- The “X” в позиции 3 может быть описан как “N-(4-hydroxybutyl) glycine.”
- The “X” в позиции 4 может быть описан как “N-(butyl) glycine.”
- The “K” в позиции 5 может быть описан как “N-(4-aminobutyl)glycine.”
- The “W” в позиции 6 может быть описан как “N-[1-(indol-3-yl)methyl]glycine.”

**ПРЕДУПРЕЖДЕНИЕ**: предпочтительное представление данных последовательностей, указанное выше, и их аннотаций предназначено для предоставления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 3(a), 3(e), 7(b), 29, 30

Пример 29-5 **Бета-аминокислоты**

Заявка на патент раскрывает следующую последовательность:




**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Последовательность содержит бета-аминокислоты, которые охватываются определением аминокислот, данным в параграфе 3(a) данного Стандарта. Бета-аминокислота имеет аминогруппу, прикрепленную к бета-углероду, второму углероду, удаленному от карбоксильной группы. Параграф 29 этого Стандарта требует представления модифицированной аминокислоты, которая имеет одиночную углеродсодержащую боковую цепь, идентичную боковой цепи в L-альфа-аминокислоте, содержащейся в перечне Приложения I, Раздел 3, Таблица 3, размещенной на любом основном атоме аминокислоты, имеющей расширенную основу, символом соответствующей немодифицированной аминокислоты.

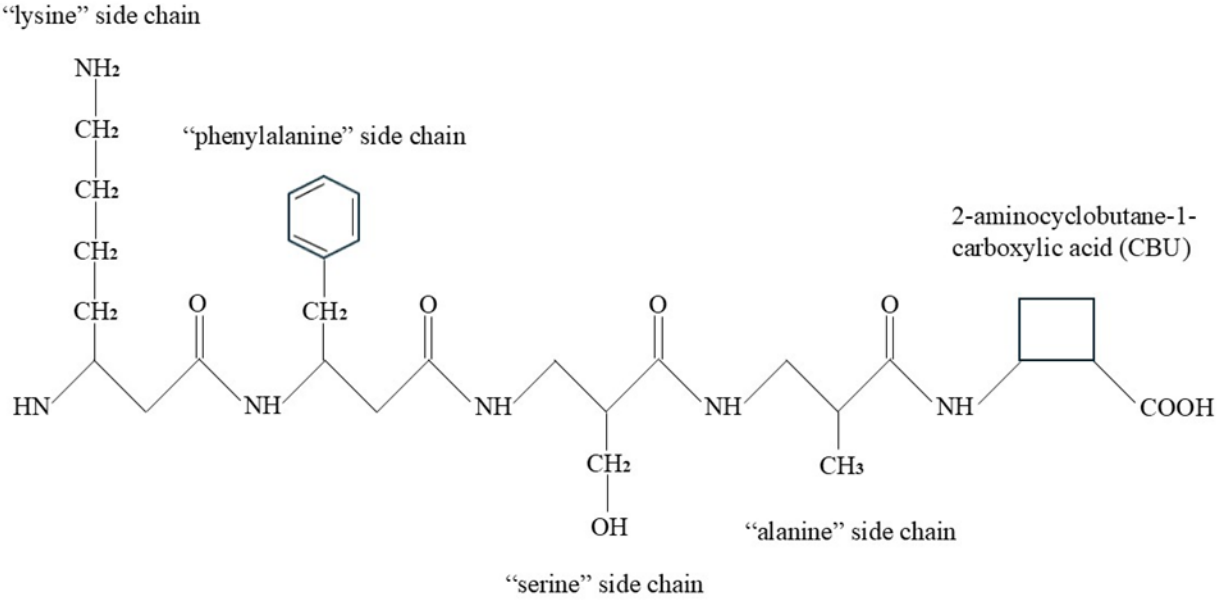
Первый остаток содержит одиночную боковую цепь, прикрепленную к расширенной основе, которая идентична боковой цепи лизина.

Второй остаток содержит одиночную боковую цепь, прикрепленную к расширенной основе, которая идентична боковой цепи фенилаланина.

Третий остаток содержит одиночную боковую цепь, прикрепленную к расширенной основе, которая идентична боковой цепи серина.

Четвертый остаток содержит одиночную боковую цепь, прикрепленную к расширенной основе, которая идентична боковой цепи аланина.

Пятый остаток содержит цикличную боковую цепь, которая НЕ идентична какой-либо боковой цепи L-альфа-аминокислоты, содержащейся в перечне Приложения I, Раздел 3, Таблица 3. Следовательно, она может быть представлена символом “X” или символом наиболее подходящей аминокислоты.




Следовательно, раскрытый пептид содержит по меньшей мере четыре модифицированных аминокислоты, которые конкретно определены, и последовательность должна быть включена в перечень последовательностей, как это требуется параграфом 7(b) ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Если "X" представляется наиболее подходящим символом для пятого остатка, последовательность должна быть включена в перечень последовательностей как:

KFSAX (SEQ ID NO: 119)

Каждый остаток должен быть описан в таблице характеристик с использованием функционального ключа "SITE" и обязательного квалификатора "note" со значением, которое описывает этот остаток его химическим наименованием.

– The "K" в позиции 1 может быть описан как "3,7-diaminoheptanoic acid."
– The "F" в позиции 2 может быть описан как "3-amino-4-phenylbutanoic acid."
– The "S" в позиции 3 может быть описан как "3-amino-2-(hydroxymethyl)propanoic acid."
– The "A" в позиции 4 может быть описан как "3-amino-2-methylpropanoic acid."
– The "X" в позиции 5 может быть описан как "2-aminocyclobutane-1-carboxylic acid."

**ПРЕДУПРЕЖДЕНИЕ**: предпочтительное представление данных последовательностей, указанное выше, и их аннотаций предназначено для предоставления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 3(a), 3(e), 7(b), 27, 29, 30

**Пример 29-6 Пептид-пептоидный гибрид**

Заявка на патент раскрывает следующее:

«Антибактериальная активность недекапептида Lys-Lys-Leu-Leu-Lys-Leu-Leu-Lys-Leu-Leu-Leu была усилена, когда пять позиций были замещены соответствующим пептоидной (т.е. N-замещенным глицином) единицей, что привело к пептид-пептоидной гибридной последовательности Lys-**Lys**-Leu-Leu-Lys-**Leu**-Leu-**Lys**-Leu-**Leu**-Leu. Выделенные жирным шрифтом буквы представляют соответствующие пептоидные единицы.»

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**
**ДА**

Обе оригинальных пептидных последовательности Lys-Lys-Leu-Leu-Lys-Leu-Leu-Lys-Leu-Leu-Leu и пептид-пептоидная гибридная последовательность Lys-**Lys**-Leu-Leu-Lys-**Leu**-Leu-**Lys**-Leu-**Leu**-Leu содержат по меньшей мере четыре конкретно определенных аминокислоты и должны быть включены в перечень последовательностей, как это требуется параграфом 7(b) ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Оригинальная пептидная последовательность должна быть включена в перечень последовательностей как:

KKLLKLLKLLL (SEQ ID NO: 116)

Все пять «пептоидных единиц» в пептид-пептоидной гибридной последовательности являются N-замещенными глициновыми остатками, которые охватываются определением «аминокислоты», данным в параграфе 3(a) данного Стандарта. Параграф 29 данного Стандарта требует представления модифицированной аминокислоты, которая имеет единственную углеродсодержащую боковую цепь, идентичную боковой цепи L-альфа-аминокислоты, указанной в Приложении I, раздел 3, таблица 3, которая расположена на аминоазоте остатка, с помощью символа соответствующей немодифицированной аминокислоты. Таким образом, каждый пептоид «Lys» должен быть представлен в последовательности как «K» и каждый пептоид «Leu» как «L». Следовательно, пептид-пептоидная гибридная последовательность должна быть включена в перечень последовательностей как:

KKLLKLLKLLL (SEQ ID NO: 117).

Каждый пептоидный остаток должен каждый быть описан отдельно в таблице характеристик с использованием функционального ключа SITE“ и обязательного квалификатора “note“ со значением, которое описывает остаток его химическим наименованием:

–пептоидный остаток “Lys” в позициях 2 и 8 могут быть описаны как “N-(4-aminobutyl)glycine.”
–пептоидный остаток “Leu” в позициях 4, 6, и 10 могут быть описаны как “N-(isobutyl)glycine.”

**Соответствующие параграфы ST.26:** 3(a), 3(e), 7(b), 27, 29, 30

*Параграф 30 – Аннотирование модифицированной аминокислоты*

**Пример 30-1 – Функциональный ключ “CARBOHYD”**

В заявке на патент описан полипептид со специфически модифицированной аминокислотой, содержащей гликолизированную боковую цепь, далее характеризующийся тем, что Cys, соответствующий позициям 4 и 15 полипептида, образует дисульфидную связь, согласно следующей последовательности:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Asn(asialyloligosaccharide)-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

**Вопрос 1: Требует ли ST.26 включения данной последовательности(-тей)?**

**ДА**

Представленный пептид содержит 17 конкретно определенных аминокислот. Так как это – более, чем четыре конкретно определенных аминокислоты, последовательность должна быть включена в перечень последовательностей, как того требует параграф 7(b) ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Asn в позиции 9 является модифицированным в силу того, что к боковой цепи прикреплён асиалилолигосахарид. В соответствии с параграфом 29 ST.26, модифицированную аминокислоту следует представлять в последовательности, когда это возможно, как соответствующую немодифицированную аминокислоту.

Соответственно, последовательность следует включать в перечень последовательностей как:

LEYCLKRWNETISHCAW (SEQ ID NO: 52)

Модифицированная аминокислота должна быть далее описана в таблице характеристик. Функциональный ключ «CARBOHYD» совместно с (обязательным) квалификатором «note» должен быть использован для указания на наличие присоединения сахарной цепи (асиалилолигосахарида) к Asn в позиции 9. Квалификатор «note» должен также описать тип связи, например, N-связанный. Дескриптором местоположения в признаке местоположения элемента является номер позиции остатка модифицированного Asn.

Кроме того, между двумя остатками Cys существует дисульфидная связь. Поэтому функциональную ключ «DISULFID» следует использовать для описания внутрицепочечной перекрёстной связи. Значение признака местоположения элемента - это номера позиций остатков связанных остатков Cys в формате “x..y”, т.е. «4…15». Обязательный квалификатор «note» должен описывать внутрицепочечную дисульфидную связь.

**Соответствующие параграфы ST.26:** 3(a), 7(b), 26, 29, 30, 66(c), 70 и Приложение I, раздел 7, функциональный ключ 7.4

**Пример 30-2 – Посттрансляционно модифицированные аминокислоты**

В заявке на патент описан следующий полипептид:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Glu-Thr-Ile-Ser-His

где Arg на позиции 7 является посттрансляционно деиминированым до цитруллина.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Перечисленный пептид содержит 13 конкретно определенных аминокислот. Поэтому последовательность должна быть включена в перечень последовательностей, как того требует параграф 7(b) ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Согласно параграфу 29 ST.26, модифицированную аминокислоту следует представлять в последовательности как соответствующую немодифицированную аминокислоту, когда это возможно.

Следовательно, последовательность должна быть включена в перечень последовательностей как:

LEYCLKRWETISH (SEQ ID NO: 97)

где символ «R» используется для обозначения аргинина на позиции 7.

При этом требуется последующее описание, указывающее, что аргинин на позиции 7 является модифицированным до цитруллина. Преобразование аргинина в цитруллин является посттрансляционной модификацией. Следовательно, функциональный ключ «MOD_RES» следует использовать вместе с обязательным квалификатором «note», чтобы указать, что аргинин деиминирован до формы цитруллина. Дескриптор местоположения в элементе местоположения признака представляет собой номер позиции остатка модифицированного аргинина.

**Соответствующие параграфы ST.26:** 3(a), 7(b), 30 и Приложение I, Раздел 7, функциональный ключ 7.18

*Параграф 36 - Последовательности, содержащие участки с точным количеством смежных «n» или «X» остатков*

**Пример 36-1: Последовательность с участком известного количества «X» остатков, представленная как единая последовательность**

LL-100-KYMR

Где «-100-» между аминокислотами лейцина и лизина показывает 100-аминокислотный участок в последовательности.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Параграф 36 ST.26 требует включения последовательности, которая содержит, по меньшей мере, четыре конкретно определенные аминокислоты, разделенные одним или несколькими участками определенного количества «X» остатков.

В раскрытой последовательности используется нерекомендованный символ, т. е. «-100-». Определение «-100-» должно быть приведено из пояснения к последовательности в описании, которое определяет этот символ как 100 аминокислот между лейцином и лизином (см. Введение в этот документ). Следовательно, «-100-» представляет собой определенный участок остатков «X». Поскольку шесть из 106 аминокислот в последовательности конкретно определены, параграф 7(b) ST.26 требует, чтобы последовательность была включена в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Нерекомендованный символ «-100-» представлен в виде 100 “X“ остатков (поскольку любой символ, используемый для обозначения аминокислоты, эквивалентен только одному остатку). Следовательно, в перечень последовательностей должна быть включена одна последовательность длиной 106 аминокислот, содержащая 100 остатков «X» между LL и KYMR (SEQ ID NO: 53).

Данная последовательность содержит 100 “X“ остатков между LL и KYMR. В ST.26 по умолчанию значением для «X» без дополнительного аннотирования является любая из немодифицированных L-альфа-аминокислот «A», «R», «N», «D», «C», «Q», «E», «G», «H», “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, «W», «Y» или «V» (параграф 27). Если эти 100 переменных «X» определены как что-либо иное, чем это значение по умолчанию, то для каждой переменной «X» должна быть предоставлена соответствующая аннотация.

**Соответствующие параграфы ST.26:** 7(b), 26, 27 и **36**

**Пример 36-2. Последовательность с несколькими участками с известным количеством или диапазоном «X» остатков, представленных в виде одной последовательности**

Lys-$z_2$-Lys-$z_m$-Lys-$z_3$-Lys-$z_n$-Lys-$z_2$-Lys

Где z - любая аминокислота, m = 20, n = 19-20, $z_2$ означает, что между лизинами находятся любые две аминокислоты, а $z_3$ означает, что между лизинами находятся любые три аминокислоты.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

В раскрытой последовательности используется нерекомендованный символ, а именно - «z». Следовательно, необходимо ознакомиться с информацией об определении его формулировки, чтобы найти, что “z” определяется как любая аминокислота (см. Введение к данному документу). Рекомендованным символом для представления любой аминокислоты является “X”. Учитывая наличие “X” остатков, пептид содержит шесть остатков лизина, которые перечислены и конкретно определены, что требует их включения в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

В последовательности используется нерекомендованный символ «z», определение которого должно быть указано в раскрытии. Раз «z» определяется как любая аминокислота, рекомендованным символом является “X”.

Предпочтительным и наиболее охватывающим значением для представления является (см. Введение к данному документу):

KXXKXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 54)

Где $z_n$ равно 20 “X“, с последующим описанием, что остаток “X”, соответствующий позиции 30, может быть удален.

Альтернативно или в дополнение к вышеизложенному, последовательность может быть представлена как:
KXXKXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 55)

Где $z_n$ равно 19 “X“ с последующим описанием, что остаток “X” между позициями 29 и 30 может быть вставлен.

Согласно параграфу 27, “X” будет истолковываться как любая из немодифицированных L-альфа-аминокислот A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” или “V”, за исключением случаев, когда он используется с дальнейшим описанием в таблице характеристик. “X” в SEQ ID No. 54 и 55 представляет собой «любую аминокислоту», что шире по охвату, чем значение по умолчанию для “X”, данное в параграфе 27, поскольку «любая аминокислота» охватывает и модифицированные, и другие аминокислоты. Следовательно, каждый “X” должен быть аннотирован с помощью функционального ключа VARIANT и квалификатора note со значением: «X может быть любой аминокислотой».

По возможности, каждый “X” должен быть аннотирован индивидуально. Однако участок из непрерывных остатков “X” может быть совместно описан с помощью функционального ключа VARIANT, используя синтаксис «x..y» в качестве дескриптора местоположения, где x и y - позиции первого и последнего остатков “X”, а также квалификатора “note“ со значением: «X может быть любой аминокислотой».

**Соответствующие параграфы ST.26:** 26, 27 и **36**

**Пример 36-3. Последовательность со множеством участков известного количества или диапазоном** “X” **остатков, представленных в виде одной последовательности**

K-$z_2$-K-$z_m$-K-$z_3$-K-$z_n$-K-$z_2$-K

Где z - любая аминокислота, где m = 15-25, предпочтительно 20-22, где n = 15-25, предпочтительно 19-20, где $z_2$ означает, что лизины по его краям разделены любыми двумя аминокислотами, а $z_3$ означает, что лизины по его краям разделены любыми тремя аминокислотами.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**ДА**

Последовательность в данном примере использует нерекомендованный символ, а именно - «z». Поэтому рекомендуется всестороннее раскрытие для формулировки определения «z» (см. Введение к этому документу). Описание определяет этот символ как любую аминокислоту. Условно принятым символом, используемым для обозначения остатка, определяемого как «любая аминокислота», является “X”. Учитывая присутствие остатков “X”, пептид содержит шесть остатков лизина, которые перечислены и конкретно определены, т.е. должны быть включены в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

В последовательности используется нерекомендованный символ «z», определение которого должно быть изложено в раскрытии. Поскольку «z» определено как любая аминокислота, условно принятым символом для неё является “X”. Предпочтительным и наиболее обобщающим способом представления является:

KXXKXXXXXXXXXXXXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 56)

(где m=25 и n=25). Участок “$z_m$” и участок “$z_n$” должны каждый быть аннотирован с использованием функционального ключа «VARIANT» наряду с обязательным квалификатором “note” со значением, указывающим, что до 10 остатков “X” в участке может быть удалено.

Настоятельно рекомендуется включение любых конкретных последовательностей, существенных для описания или формулы изобретения, как обсуждалось во введении к данному документу.

Альтернативно, последовательность может быть представлена как:

KXXKXXXXXXXXXXXXXXKXXXKXXXXXXXXXXXXXXXXKXXK (SEQ ID NO: 57)

(где m=15 и n=15). Участок ““$z_m$” и участок ““$z_n$” должны каждый быть аннотирована с использованием функционального ключа «VARIANT» наряду с обязательным квалификатором “note“ со значением, указывающим, что может быть вставлено до 10 остатков “X.

В качестве дальнейших альтернатив могут быть включены любые или все возможные варианты.

В соответствии с параграфом 27, «X» будет истолковываться как любая из немодифицированных L-альфа-аминокислот “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” “V”, за исключением

случаев, когда он используется с дальнейшим описанием в таблице характеристик. «X» в SEQ ID No. 56 и 57 представляет собой «любую аминокислоту», что шире по охвату, чем значение «X» по умолчанию, данное в параграфе 27, поскольку «любая аминокислота» охватывает модифицированные и другие аминокислоты. Поэтому каждый «X» должен быть аннотирован с помощью функционального ключа VARIANT и квалификатора “note” со значением: «X может быть любой аминокислотой».

Где практически осуществимо, каждый «X» должен быть аннотирован индивидуально. Однако участок смежных остатков «X» может быть совместно описан с помощью функционального ключа VARIANT, используя синтаксис «x..y» в качестве дескриптора местоположения, где x и y - позиции первого и последнего остатков «X», а также квалификатора “note” со значением: «X может быть любой аминокислотой».

**ПРЕДУПРЕЖДЕНИЕ:** предпочтительное представление данной последовательности, указанное выше, предназначено для предоставления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 27 и **36**

*Параграф 37 - Последовательности, содержащие участки с неизвестным количеством смежных «n» или «X» остатков*

***Пример 37-1: Последовательность с участками с неизвестным количеством «X» остатков не должна быть представлена как единая последовательность I***

Gly-Gly----Gly-Gly-Xaa-Xaa

где символ ---- представляет собой неопределенный разрыв (пробел) в последовательности, где Xaa представляет собой любую аминокислоту, и где остатки глицина и Xaa связаны друг с другом посредством пептидных связей.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**НЕТ**

Параграф 37 ST.26 запрещает включение любой последовательности, которая содержит неопределенный пробел; следовательно, включение всей последовательности не требуется.

Параграф 37 ST.26 действительно требует включения любого участка последовательности, смежного с неопределенным пробелом, который содержит четыре или более конкретно определенных аминокислот. В приведенном выше примере включение какого-либо участка, смежного с неопределенным пробелом, не требуется, поскольку каждый участок содержит только две конкретно определенные аминокислоты.

**Вопрос 2: Разрешает ли ST.26 включение данной последовательности(-ей)?**

**НЕТ –** для последовательности целиком

Параграф 37 ST.26 НЕ разрешает включение ВСЕЙ последовательности.

**ДА –** каждый участок, смежный с неопределенным пробелом, может быть включен в перечень последовательностей.

Параграф 8 ST.26 разрешает включение какого-либо участка, смежного с неопределенным пробелом, содержащего каждый только две конкретно определенные аминокислоты.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

При условии включения, каждый участок, смежный с неопределенным пробелом, следует представлять в перечне последовательностей как:

GG (SEQ ID NO: 120)

GGXX (SEQ ID NO: 121)

Две последовательности следует аннотировать для указания, что они являются частью б**о**льшей последовательности, которая содержит неопределенный пробел, с помощью использования функционального ключа “SITE” и обязательного квалификатора “note”, где значение квалификатора “note” указывает, что C-концевой остаток последовательности GG связан с N-концевым остатков последовательности GGXX с помощью пробела неопределенной длины.

**Соответствующие параграфы ST.26:** 7(b), 8, 26 и **37**

***Пример 37-2: Последовательность с участками с неизвестным количеством «X» остатков не должна быть представлена как единая последовательность II***

Gly-Gly----Gly-Gly-Ala-Gly-Xaa-Xaa

в которой символ ---- представляет собой неопределенный пробел в последовательности, где Xaa представляет собой любую аминокислоту, а остатки Gly и Xaa связаны друг с другом посредством пептидных связей.

**Вопрос 1: Требует ли ST.26 включения последовательности(ей)?**

**НЕТ –** для последовательности целиком.

**ДА –** участок последовательности

Параграф 37 ST.26 запрещает включение любой последовательности, которая содержит неопределенный пробел, но требует включения какого-либо участка последовательности, смежного с неопределенным пробелом, который содержит четыре или более конкретно определенных аминокислоты.

В приведенном выше примере ST.26 не требует (и запрещает) включение всей последовательности, которая содержит неопределенный пробел. Однако ST.26 требует включения участка Gly-Gly-Ala-Gly-Xaa-Xaa, смежного с неопределенным пробелом, поскольку он содержит по меньшей мере четыре конкретно определенных аминокислоты.

**Вопрос 2: Разрешает ли ST.26 включение последовательности(ей)?**

**НЕТ –** всей последовательности

**ДА** – участок Gly-Gly может быть включен

Параграф 37 ST.26 не разрешает включение всей последовательности.

Параграф 8 ST.26 разрешает включение последовательностей, имеющих менее четырех конкретно определенных аминокислот.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Участок последовательности, примыкающий к неопределенному пробелу, который содержит четыре конкретно определенные аминокислоты, должен быть представлен как:

GGAGXX (SEQ ID NO: 58)

Эту последовательность следует аннотировать, чтобы указать, что представленная последовательность является частью большей последовательности, которая содержит неопределенный пробел, с помощью использования функционального ключа “SITE”, характеристики местоположения «1» и квалификатора “note” со значением, например, «Данный остаток прикреплён N-концом к пептиду неопределенной длины, который включает N-концевую Gly-Gly последовательность».

Согласно параграфу 27, “X” будет истолковываться как любая из немодифицированных L-альфа-аминокислот “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” или “V”, за исключением случаев, когда он используется с дальнейшим описанием в таблице характеристик. “X” в SEQ ID No. 58 представляет собой «любую аминокислоту», что шире по охвату, чем значение “X” по умолчанию, данное в параграфе 27. Поскольку «любая аминокислота» включает

модифицированные и другие аминокислоты. Следовательно, каждый “X” должен быть аннотирован с помощью функционального ключа VARIANT и квалификатора “note” со значением: «“X” может быть любой аминокислотой».

По возможности, каждый “X” должен быть аннотирован индивидуально. Однако участок смежных остатков “X” может быть совместно описан с помощью функционального ключа VARIANT с использованием синтаксиса “x..y” в качестве дескриптора местоположения, где x и y - позиции первого и последнего остатков “X”, а также квалификатора “note” со значением: «“X” может быть любой аминокислотой».

Участок последовательности, смежный с пробелом неопределенно длины, если он включен, который содержит две конкретно определенных аминокислоты, следует представлять, как:

GG (SEQ ID NO: 122).

Эту последовательность следует аннотировать для указания, что представляемая последовательность является частью более длинной последовательности, которая содержит неопределенный пробел, с помощью использования функционального ключа “SITE”, характеристики местоположения «2» и квалификатора “note” со значением: «Пептидная последовательность, С-концевая по отношению к этому остатку, является неопределенной, за исключением конечных остатков GGAGXX».

**Соответствующие параграфы ST.26:** 7(b), 8, 26, 27 и **37**

*Параграф 55 – Нуклеотидная последовательность, которая содержит сегменты ДНК и РНК*

**Пример 55-1: Комбинированная молекула ДНК / РНК**

Заявка на патент описывает следующую олигонуклеотидную последовательность:

AGACCTTcggagucuccuguugaacagauagucaaaguagauC

В которой буквы верхнего регистра представляют ДНК остатки, а буквы нижнего регистра представляют РНК остатки.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Раскрытая последовательность имеет более десяти перечисленных и конкретно определенных нуклеотидов; следовательно, её необходимо включить в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Нуклеотидная последовательность должна быть включена в перечень последовательностей как:

agaccttcggagtctcctgttgaacagatagtcaaagtagatc (SEQ ID NO: 93)

Обратите внимание, что нуклеотиды урацила должны быть представлены в перечне последовательностей символом «t».

Параграф 55 ST.26 предписывает, что нуклеотидная последовательность, содержащая сегменты как ДНК, так и РНК, должна указываться как молекула типа «ДНК» и должна быть далее описана с использованием функционального ключа “source” и как обязательного квалификатора “organism” со значением “синтетическая конструкция”, так и обязательного квалификатора “mol_type” со значением “другая ДНК”. Кроме того, каждый сегмент последовательности должен быть далее описан с помощью функционального ключа «misc_feature» и элемента характеристики местоположения со значением, которое указывает местоположение сегмента, и квалификатора “note”, который указывает, является ли сегмент ДНК или РНК. Раскрытая последовательность содержит два сегмента ДНК (позиции нуклеотидов 1-7 и 43) и один сегмент РНК (нуклеотидные позиции 8-42).

**Соответствующие параграфы ST.26:** 7, 14, **55**-56 и 83

*Параграф 89 – Функциональный ключ “CDS”*

**Пример 89-1: Кодирующая нуклеотидная последовательность и кодируемая аминокислотная последовательность**

Заявка на патент описывает следующую нуклеотидную последовательность и ее трансляцию**:**

```
atg acc gga aat aaa cct gaa acc gat gtt tac gaa att tta tga

Met Thr Gly Asn Lys Pro Glu Thr Asp Val Tyr Glu Ile Leu STOP
```

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Представленная нуклеотидная последовательность имеет более десяти конкретно определенных нуклеотидов.

Представленная аминокислотная последовательность имеет более четырех конкретно определенных аминокислот.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Нуклеотидная последовательность должна быть представлена как:

atgaccggaaataaacctgaaaccgatgtttacgaaattttatga (SEQ ID NO: 59)

Нуклеотидная последовательность должна быть дополнительно описана с использованием функционального ключа «CDS», а элемент INSDFeature_location должен идентифицировать всю последовательность, включая стоп-кодон (то есть, позиции с 1 по 45). Кроме того, квалификатор “translation” следует включать со значением квалификатора «MTGNKPETDVYEIL». В заявке не раскрывается таблица генетического кода, которая применяется к трансляции (см. Приложение 1, Раздел 9, Таблица 7). Если применяется таблица стандартных кодов, то квалификатор «trans_table» не нужен; однако, если применяется другая таблица генетических кодов, тогда для квалификатора «trans_table» должно быть указано соответствующее значение квалификатора из Таблицы 7. Наконец, квалификатор «protein_id» должен быть включен со значением квалификатора, указывающим идентификационный номер последовательности транслированной аминокислотной последовательности.

Аминокислотная последовательность должна быть отдельно представлена своим собственным идентификационным номером последовательности с использованием однобуквенных кодов следующим образом:

MTGNKPETDVYEIL (SEQ ID NO: 60)

STOP, следующий за представленной аминокислотной последовательностью, не должен быть включен в аминокислотную последовательность в перечне последовательностей.

**ПРЕДУПРЕЖДЕНИЕ:** предпочтительное представление данной последовательности, указанное выше, предназначено для предоставления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 7(a), 7(b), 26, 28, **89**, 90 и 92

**Пример 89-2: Местоположение признака выходит за пределы раскрытой последовательности**

Заявка на патент содержит следующую иллюстрацию, раскрывающую частично кодирующую последовательность и ее транслированную аминокислотную последовательность:

```
cat cac gca gca gaa tgt gga ttt tgt cct caa caa tgg caa gtt cta     48
His His Ala Ala Glu Cys Gly Phe Cys Pro Gln Gln Trp Gln Val Leu
1               5                  10                  15

cgt ggg agt ctg tgc att tgt gag ggt cca gct gaa gga tgg ttc ata     96
Arg Gly Ser Leu Cys Ile Cys Glu Gly Pro Ala Glu Gly Trp Phe Ile
            20                  25                  30

tca aga tgt tgg tta tgg tgt ggg cct caa gtc caa ggc ttt atc ttt    144
Ser Arg Cys Trp Leu Trp Cys Gly Pro Gln Val Gln Gly Phe Ile Phe
        35                  40                  45

gga gaa ggc aag gaa gga ggc ggt gac aga cgg gct gaa gcg agc cct    192
Gly Glu Gly Lys Glu Gly Gly Gly Asp Arg Arg Ala Glu Ala Ser Pro
    50                  55                  60

cag gag ttt tgg gaa tgc act tgg                                    216
Gln Glu Phe Trp Glu Cys Thr Trp
65                  70
```

Фигура 1 – частичная кодирующая последовательность гена Homo sapiens ITCH1, которая кодирует аминокислоты с 20 по 91 белка ITCH1 длиной 442 аминокислоты.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

В заявке раскрыта нуклеотидная последовательность и ее транслированная аминокислотная последовательность. Представленная нуклеотидная последовательность содержит более 10 конкретно определенных нуклеотидов и должна быть включена в перечень последовательностей.

Аминокислотная последовательность содержит более 4 конкретно определенных аминокислот и также должна быть включена в перечень последовательностей в виде отдельной последовательности с собственным идентификационным номером последовательности.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Нуклеотидная последовательность должна быть включена в перечень последовательностей как:

catcacgcagcagaatgtggattttgtcctcaacaatggcaagttctacgtgggagtctgtgcatttgtgagggtccagctgaaggatggtt
catatcaagatgttggttatggtgtgggcctcaagtccaaggctttatctttggagaaggcaaggaaggaggcggtgacagacgggctg
aagcgagccctcaggagttttgggaatgcacttgg (SEQ ID NO: 94)

Эту нуклеотидную последовательность далее следует описать с использованием функционального ключа «CDS». Элемент INSDFeature_location должен идентифицировать местоположение «CDS» в последовательности и должен включать в себя стоп-кодон.

Фигура 1 описывает частично кодирующую последовательность, которая не включает стартовый кодон или стоп-кодон. Однако описание последовательности указывает на то, что стартовый кодон расположен до нуклеотида на позиции 1, а стоп-кодон расположен после последнего нуклеотида на позиции 216.

ST.26 предписывает, что дескриптор местоположения не должен включать нумерацию остатков за пределами диапазона последовательности в элементе INSDSeq_sequence. Следовательно, в приведенном выше примере дескриптор местоположения для функционального ключа CDS не может включать номера позиций за пределами диапазона от 1 до 216. Местоположение стоп-кодона в элементе INSDFeature_location должно быть представлено с использованием символа «>», чтобы указать, что стоп-кодон расположен после позиции 216. Аналогично, символ «<» может использоваться для указания того, что местоположение стартового кодона находится выше позиции 1. Таким образом, дескриптор местоположения для функционального ключа CDS должен выглядеть следующим образом:

<1..>216

Обратите внимание, что «<» и «>» являются запасными символами и будут заменены на «<» и «>» соответственно в XML представлении перечня последовательностей.

Квалификатор “translation” следует включить с аминокислотной последовательностью белка в качестве значения квалификатора. Фигура 1 не раскрывает таблицу генетического кода, который применяет трансляцию (см. Приложение 1, Раздел 9, Таблица 7). Если применяется таблица стандартного кода, то квалификатор «trans_table» не нужен; однако, если применяется таблица другого генетического кода, то для квалификатора «trans_table» должно быть указано подходящее значение квалификатора из таблицы 7 Приложения I к ST.26. Наконец, в характеристику CDS должен быть включен квалификатор «protein _id» со значением квалификатора, указывающим на идентификационный номер транслированной аминокислотной последовательности.

Транслированная аминокислотная последовательность должна быть включена как отдельная последовательность с собственным идентификационным номером последовательности:

HHAAECGFCPQQWQVLRGSLCICEGPAEGWFISRCWLWCGPQVQGFIFGEGKEGGGDRRAEA
SPQEFWECTW (SEQ ID NO: 95)

**ПРЕДУПРЕЖДЕНИЕ:** предпочтительное представление данной последовательности, указанное выше, предназначено для предоставления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 7, 41, 65, 66, 70, 71, 89 и 92

*Параграф 92 – Аминокислотная последовательность, кодированная кодирующей последовательностью*

**Пример 92-1: Аминокислотная последовательность, кодированная кодирующей последовательностью с интронами**

Заявка на патент содержит следующую иллюстрацию, раскрывающую кодирующую последовательность и ее трансляцию:

```
atg aag act ttc gca gcc ttg ctt tcc gct gtc act ctc gcg ctc tcg
Met Lys Thr Phe Ala Ala Leu Leu Ser Ala Val Thr Leu Ala Leu Ser

gtg cgc gcc cag gcg gct gtc tgg agt caa t gtaagtgccg ctgcttttca
Val Arg Ala Gln Ala Ala Val Trp Ser Gln

ttgatacgag actctacgcc gagctgacgt gctaccgtat ag gt  ggc ggt aca
                                                Cys Gly Gly Thr

ccg ggt tgg acg ggc gag acc act tgc gtt gct ggt tcg gtt tgt acc
Pro Gly Trp Thr Gly Glu Thr Thr Cys Val Ala Gly Ser Val Cys Thr

tcc ttg agc tca gtgagcgact ttcaatccgt cgtcattgct cctcatgtat
Ser Leu Ser Ser

tgacgattgg ccttcatag tca tac tct caa tgc gtt ccg ggc tcc gca acg
                    Ser Tyr Ser Gln Cys Val Pro Gly Ser Ala Thr

tcc agc gct ccg gcg gcc ccc tca gcg aca act tca ggc ccc gca cct
Ser Ser Ala Pro Ala Ala Pro Ser Ala Thr Thr Ser Gly Pro Ala Pro

acg gac gga acg tgc tcg gcc agc ggg gca tgg ccg cca ttg acc tga
Thr Asp Gly Thr Cys Ser Ala Ser Gly Ala Trp Pro Pro Leu Thr Ter
```

Фигура 1 – нуклеотиды, выделенные жирным шрифтом, являются интронными участками.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Заявка раскрывает нуклеотидную последовательность и ее аминокислотную трансляцию. Представленная нуклеотидная последовательность содержит более 10 конкретно определенных нуклеотидов и должна быть включена в перечень последовательностей в виде одной последовательности.

Нуклеотидная последовательность содержит кодирующую последовательность (экзоны), отделенную некодирующей последовательностью (интроны). На фигуре 1 изображена трансляция нуклеотидной последовательности в виде трех несмежных аминокислотных последовательностей. В соответствии с подписью к фигуре 1, выделенные жирным шрифтом участки нуклеотидов являются интронными последовательностями, которые будут сплайсированы из РНК-транскрипта перед трансляцией в белок. Соответственно, три аминокислотные последовательности на самом деле представляют собой единую непрерывную нумерованную последовательность, которая содержит более четырех

конкретно определенных аминокислот и должна быть включена в перечень последовательностей в виде единой последовательности.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Нуклеотидная последовательность должна быть включена в перечень последовательностей как:

atgaagactttcgcagccttgctttccgctgtcactctcgcgctctcggtgcgcgcccaggcggctgtctggagtcaatgtaagtgccgctg
cttttcattgatacgagactctacgccgagctgacgtgctaccgtataggtggcggtacaccgggttggacgggcgagaccacttgcgttg
ctggttcggtttgtacctccttgagctcagtgagcgactttcaatccgtcgtcattgctcctcatgtattgacgattggccttcatagtcatactctc
aatgcgttccgggctccgcaacgtccagcgctccggcggccccctcagcgacaacttcaggccccgcacctacggacggaacgtgct
cggccagcggggcatggccgccattgacctga (SEQ ID NO: 75)

Нуклеотидная последовательность должна быть далее описана с использованием функционального ключа «CDS», а элемент INSDFeature_location должен идентифицировать местоположение кодирующей последовательности, включая стоп-кодон, обозначенный «Ter». В CDS INSDFeature_location должен использоваться оператор местоположения «join» с использованием формата «join(x1..y1,x2..y2,x3..y3)», чтобы указать, что продукты трансляции, закодированные указанными местоположениями, объединяются и образуют единый непрерывный полипептид, например,«join(1..79,142..212,272..400)». Кроме того, квалификатор “translation” должен быть включен с аминокислотной последовательностью белка в качестве значения квалификатора. (Обратите внимание, что символ окончания «Ter» на последней позиции последовательности не должен быть включен в аминокислотную последовательность). В заявке не раскрывается таблица генетического кода, которая применяется к трансляции (см. Приложение 1, раздел 9, таблица 7). Если применяется таблица «Стандартные коды», то квалификатор «trans_table» не нужен; однако, если применяется другая таблица генетического кода, тогда для квалификатора «trans_table» должно быть указано соответствующее значение квалификатора из таблицы 7. Наконец, квалификатор «protein_id» должен быть включен со значением квалификатора, указывающим идентификационный номер последовательности транслированной аминокислотной последовательности. Аминокислотная последовательность должна быть включена как единая последовательность:

MKTFAALLSAVTLALSVRAQAAVWSQCGGTPGWTGETTCVAGSVCTSLSSSYSQCVPGSATSS
APAAPSATTSGPAPTDGTCSASGAWPPLT (SEQ ID NO: 76)

**Соответствующие параграфы ST.26:** 7, 26, 28, 57, 67и 89-92

*Параграф 93 - Исходная последовательность и вариант, представленные каждая своими остатками*

**Пример 93-1: Представление описанных вариантов I**

Описание включает в себя следующее выравнивание последовательности.

```
D. melanogaster   ACATTGAATCTCATACCACTTT

D. virilis        ...-..G...C..--.G.....

D. simulans       GT..G.CG..GT..SGT.G...
```

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

В данной области техники принято включать «точки» в выравнивание последовательности, чтобы указать: «эта позиция совпадает с позицией над ней». Поэтому «точки» в последовательностях *D. virilis* и *D. simulans* считаются представленными и конкретно определенными нуклеотидами, так как они являются просто кратким способом обозначения того, что данная позиция является тем же нуклеотидом, что и у *D. melanogaster*. Кроме того, выравнивания последовательности часто отображаются символом «-», указывающим на отсутствие остатка, чтобы максимизировать выравнивание.

Соответственно, нуклеотидные последовательности *D. melanogaster* и *D. simulans* содержат двадцать два представленных и конкретно определенных нуклеотида, тогда как нуклеотидная последовательность *D. virilis* содержит девятнадцать. Таким образом, согласно параграфу 7(a) Стандарта ST.26 каждая последовательность должна быть включена в перечень последовательностей с отдельными идентификационными номерами последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Последовательность *Drosophila melanogaster* должна быть включена в перечень последовательностей как:

acattgaatctcataccacttt (SEQ ID NO: 61)

Последовательность *Drosophila virilis* должна быть включена в перечень последовательностей как:

acatggatcccacgacttt (SEQ ID NO: 62)

Последовательность *Drosophila simulans* должна быть включена в перечень последовательностей как:

gtatggcgtcgtatsgtagttt (SEQ ID NO: 63)

**Соответствующие параграфы ST.26:** 7(a), 13 и **93**

**Пример 93-2: Представление перечисленных вариантов II**

Описание включает следующую таблицу пептида и его функциональных вариантов. Пробел в таблице ниже указывает, что аминокислота в варианте такая же, как соответствующая аминокислота в «Последовательности», а «-» указывает на удаление соответствующей аминокислоты в «Последовательности».

| Позиция | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Последовательность | A | V | L | T | Y | L | R | G | E |
| Вариант 1 | | | | | | | | | A |
| Вариант 2 | | | P | | | P | | | |
| Вариант 3 | | | A | I | G | Y | | | |
| Вариант 4 | | | | | | | - | | |

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Как указано, пробел в этой таблице показывает, что аминокислота в варианте такая же, как соответствующая аминокислота в «последовательности». Следовательно, аминокислоты вариантов последовательностей представлены и конкретно определены.

Поскольку каждая из четырех вариантов последовательностей содержит более четырех представленных и конкретно определенных аминокислот, каждая из них, согласно параграфу 7(b) ST.26 должна быть включена в перечень последовательностей с отдельными идентификационными номерами последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

AVLTYLRGE (SEQ ID NO: 77)

AVLTYLRGA (SEQ ID NO: 78)

AVPTYPRGE (SEQ ID NO: 79)

AVAIGYRGE (SEQ ID NO: 80)

AVLTYLGE (SEQ ID NO: 81)

**Соответствующие параграфы ST.26:** 7(b), 26 и **93**

**Пример 93-3: Представление согласованной последовательности**

Заявка на патент включает Фигуру 1 со следующим множественным выравниванием последовательностей.

*Consensus* LEG**n**EQFINA**ak**IIRHP**k**Y**nrk**T**ln**NDI**m**LIK

*Homo sapiens* LEG**N**EQFINA**AK**IIRHP**Q**Y**DRK**T**LN**NDI**M**LIK

*Pongo abelii* LEG**N**EQFINA**AK**IIRHP**Q**Y**DRK**T**VN**NDI**M**LIK

*Papio Anubis* LEG**T**EQFINA**AK**IIRHP**D**Y**DRK**T**LN**NDI**L**LIK

*Rhinopithecus roxellana* LEG**T**EQFINA**AK**IIRHP**N**Y**NRI**T**LD**NDI**L**LIK

*Pan paniscus* LEG**N**EQFINA**AK**IIRHP**K**Y**NRI**T**LN**NDI**M**LIK

*Rhinopithecus bieti* LEG**N**EQFINA**TK**IIRHP**K**Y**NGN**T**LN**NDI**M**LIK

*Rhinopithecus roxellana* LEG**N**EQFINA**TQ**IIRHP**K**Y**NGN**T**LN**NDI**M**LIK

Согласованная последовательность включает заглавные буквы для представления консервативных аминокислотных остатков, тогда как строчные буквы «n», «a», «k», «r», «l» и «m» представляют предоминантные аминокислотные остатки среди выровненных последовательностей.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Каждая строчная буква в согласованной последовательности представляет собой аминокислотный остаток. Следовательно, согласованная последовательность, а также каждая из оставшихся семи последовательностей на фиг. 1 включает, по меньшей мере, четыре конкретно определенные аминокислоты. Параграф 7(b) ST.26 требует включения всех восьми последовательностей в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Строчные буквы в согласованной последовательности используются в качестве неоднозначных символов для представления предоминантной аминокислоты среди возможных вариантов для конкретной позиции. Поэтому строчные буквы «n», «a», «k», «r», «l» и «m» являются условно принятыми символами, используемыми нерекомендованным образом, и согласованная последовательность должна быть представлена с использованием неоднозначного символа на месте каждой из строчных букв.

Следует использовать наиболее ограничивающий неоднозначный символ. Для большинства позиций в согласованной последовательности “X” является наиболее ограничивающим неоднозначным символом; однако наиболее ограничивающим неоднозначным символом для «D» или «N» в позициях 20 и 25 является «B». Согласованная последовательность должна быть включена в перечень последовательностей как:

LEGXEQFINAXXIIRHPXYBXXTXBNDIXLIK (SEQ ID NO: 82)

В соответствии с параграфом 27 символ “X” будет истолковываться как какая-либо из немодифицированных L-альфа-аминокислот “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” или “V”, за исключением случаев, когда он используется с дальнейшим описанием в таблице характеристик. Следовательно, каждый “X” в согласованной последовательности должен быть далее описан в таблице характеристик с использованием функционального ключа «VARIANT» и

квалификатора «note» для указания возможных вариантов для каждой позиции.

Оставшиеся семь последовательностей должны быть включены в перечень последовательностей как:

LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK (SEQ ID NO: 83)

LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK (SEQ ID NO: 84)

LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK (SEQ ID NO: 85)

LEGTEQFINAAKIIRHPNYNRITLDNDILLIK (SEQ ID NO: 86)

LEGNEQFINAAKIIRHPKYNRITLNNDIMLIK (SEQ ID NO: 87)

LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 88)

LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 89)

**ПРЕДУПРЕЖДЕНИЕ**: предпочтительное представление данной последовательности, указанное выше, предназначено для предоставления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 7(b), 26, 27, 93 и 97

*Параграф 94 – Вариант последовательности, раскрытой как единая последовательность с перечисленными альтернативными остатками*

**Пример 94-1: Представление одной последовательности с перечисленными альтернативными аминокислотами**

В патентной заявке заявлена пептидная последовательность:

(i) Gly-Gly-Gly-[Leu or Ile]-Ala-Thr-[Ser or Thr]

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Последовательность представляет пять конкретно определенных аминокислоты, и параграф 7(b) ST.26 требует включения последовательности в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Таблица 3 Приложения I, Раздела 3 определяет неоднозначный символ «J» как изолейцин или лейцин. Поэтому предпочтительным представлением последовательности является:

GGGJATX (SEQ ID NO: 64)

что требует дальнейшего описания в таблице характеристик с использованием функционального ключа «VARIANT» и квалификатора «note», чтобы указать, что “X” представляет собой серин или треонин.

Альтернативно, последовательность может быть представлена, например, как:

GGGLATS (SEQ ID NO: 65)

что требует дальнейшего описания в таблице характеристик с использованием функционального ключа «VARIANT» и квалификатора «note», чтобы указать, что L может быть заменена на I, а S может быть заменена на T.

**ПРЕДУПРЕЖДЕНИЕ**: предпочтительное представление данной последовательности, указанное выше, предназначено для предоставления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 7(b), 8, 26, 27, 94 и 97

**Пример 94-2: Представление единой последовательности с перечисленными альтернативными аминокислотами, которые могут быть модифицированными аминокислотами**

В патентной заявке описан следующий полипептид:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Xaa-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

где Xaa может быть Ile, Ala, Phe, Tyr, aIle, MeIle или Nle.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Перечисленный пептид содержит 16 конкретно определенных аминокислот. Следовательно, последовательность должна быть включена в перечень последовательностей в соответствии с требованиями параграфа 7(b) ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Наиболее ограничивающим неоднозначным символом, который может охватывать “Ile, Ala, Phe, Tyr, aIle, MeIle или Nle” является “X”. Следовательно, последовательность должна быть включена в перечень последовательностей как:

LEYCLKRWXETISHCAW (SEQ ID NO: 96)

Альтернативные аминокислотные остатки, охватываемые “X” на позиции 9, включают модифицированные аминокислоты (т.е. aIle, MeIle и Nle). Параграф 30 ST.26 требует, чтобы «[a] модифицированная аминокислота была далее описана в таблице характеристик». Однако параграф 30 не требует использования какого-либо конкретного функционального ключа для описания модифицированных аминокислот. Хотя в параграфе 30 описывается использование функциональных ключей «CARBOHYD», «LIPID», «MOD_RES» и «SITE», эти функциональные ключи больше подходят для сценариев, когда модифицированная аминокислота не входит в список альтернатив для определенного местоположения. В этом примере функциональный ключ «VARIANT» удовлетворяет требованию параграфа 30, поскольку он позволяет включить все альтернативы для варианта локации. Таким образом, для описания варианта локации на позиции 9 следует использовать функциональный ключ «VARIANT» с квалификатором “note” и “Ile, Ala, Phe, Tyr, aIle, MeIle или Nle” в качестве значения квалификатора. Использование второго функционального ключа, такого как «SITE», с квалификатором “note” может быть использовано для дальнейшей идентификации каждой альтернативной модифицированной аминокислоты, найденной на позиции 9.

**Соответствующие параграфы ST.26:** 3(a), 7(b), 27, 30, **94**, 96 и Приложение I, Раздел 4, Таблица 4.

*Параграф 95(a) - Вариант последовательности, раскрытой только посредством ссылки на исходную последовательность со множеством независимых вариаций*

**Пример 95(a)-1: Представление варианта последовательности посредством аннотирования исходной последовательности**

Заявка содержит следующее описание:

«Пептидный фрагмент 1 представляет собой Gly-Leu-Pro-Xaa-Arg-Ile-Cys, где Xaa может быть любой аминокислотой….

В другом варианте осуществления пептидный фрагмент 1 представляет собой Gly-Leu-Pro-Xaa-Arg-Ile-Cys, где Xaa может представлять собой Val, Thr или Asp….

В другом варианте осуществления пептидный фрагмент 1 представляет собой Gly-Leu-Pro-Xaa-Arg-Ile-Cys, где Xaa может представлять собой Val».

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

«Пептидный фрагмент 1» в каждом из трех раскрытых вариантов осуществления содержит по меньшей мере шесть конкретно определенных аминокислот; следовательно, последовательность должна быть включена в перечень последовательностей, как того требует параграф 7(b) ST.26.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

В данном примере представленная последовательность «Пептидного фрагмента 1» раскрыта в виде трех разных вариантов осуществления, каждое с альтернативным описанием Xaa. В этом примере “X” является наиболее ограничивающим неоднозначным символом для позиции Xaa.

ST.26 требует включения раскрытой представленной последовательности только один раз. В наиболее охватывающем из трех вариантов осуществления Xaa является любой аминокислотой (см. Введение к данному документу). Следовательно, последовательность, которая должна быть включена в перечень последовательностей, представляет собой:

GLPXRIC (SEQ ID NO: 66)

В соответствии с параграфом 27, “X” будет истолковываться как любая из немодифицированных L-альфа-аминокислот “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y” или “V”, за исключением случаев, когда он используется с дальнейшим описанием в таблице характеристик. “X” в SEQ ID NO: 66 представляет собой «любую аминокислоту», что шире по охвату, чем значение “X” по умолчанию, данное в параграфе 27, поскольку «любая аминокислота» охватывает модифицированные или другие аминокислоты. Следовательно, она должна быть аннотирована с помощью функционального ключа VARIANT и квалификатора “note” со значением: «X может быть любой аминокислотой».

Настоятельно рекомендуется включать любые дополнительные последовательности, существенные для описания или формулы изобретения, как говорится во введении к данному документу.

Для приведенного выше примера настоятельно рекомендуется, чтобы следующие дополнительные три последовательности были включены в перечень последовательностей, каждая из которых имеет свой собственный идентификационный номер последовательности:

GLPVRIC (SEQ ID NO: 67)

GLPTRIC (SEQ ID NO: 68)

GLPDRIC (SEQ ID NO: 69)

**ПРЕДУПРЕЖДЕНИЕ**: предпочтительное представление данной последовательности, указанное выше, предназначено для предоставления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 7(b), 26, 27 и **95(a)**

*Параграф 95(b) - Вариант последовательности, раскрытой только посредством ссылки на основную последовательность со множеством взаимозависимых вариаций*

**Пример 95(b)-1: Представление отдельных вариантов последовательностей со множеством взаимозависимых вариаций**

В заявке на патент описывается следующая согласованная последовательность:

cgaatg**n**$_1$cccactacgaatg**n**$_2$cacgaatg**n**$_3$cccaca

где $n_1$, $n_2$, и $n_3$ могут быть a, t, g или c.

Несколько вариантов последовательностей раскрыты следующим образом:

если $n_1$ это a, то $n_2$ и $n_3$ это t, g или c;

если $n_1$ это t, то $n_2$ и $n_3$ это a, g или c;

если $n_1$ это g, то $n_2$ и $n_3$ это t, a или c;

если $n_1$ это c, то $n_2$ и $n_3$ это t, g или a.

**Вопрос 1: Требует ли ST.26 включения этой последовательности(-тей)?**

**ДА**

Последовательность имеет более десяти перечисленных и «конкретно определенных» нуклеотидов и, как требуется параграфом 7(a) ST.26, должна быть включена в перечень последовательностей.

**Вопрос 3: Как следует включать эту последовательность(-ти) в перечень последовательностей?**

Представленная последовательность содержит более десяти конкретно определенных нуклеотидов и три “n” остатка. ST.26 требует включения раскрытой перечисляемой последовательности, и там, где присвоен неоднозначный символ, должен быть использован наиболее ограничивающий символ. В данном примере $n_1$, $n_2$ и $n_3$ могут быть a, t, g или c, поэтому “n” является наиболее ограничивающим неоднозначным символом. Следовательно, последовательность, которая должна быть включена в перечень последовательностей, это:

cgaatgncccactacgaatgncacgaatgncccaca (SEQ ID NO: 70)

Параграф 15 ST.26 гласит, что «символ “n” будет толковаться как любой из немодифицированных “a”, “c”, “g” или “t/u”, за исключением случаев, когда он используется с дальнейшим описанием в таблице характеристик. Поскольку значение каждого остатка “n” в этой последовательности равно значениям по умолчанию “a”, “c”, “g” или “t”, дальнейшая аннотация не требуется.

Представленная последовательность содержит вариации в трех разных местоположениях, и возникновение вариаций является взаимозависимым. Включение дополнительных последовательностей, которые представляют дополнительные варианты осуществления, что является ключевой частью изобретения, **настоятельно** рекомендуется, как говорилось во Введении к этому документу. Следовательно, в соответствии с параграфом 95(b) ST.26, следующие дополнительные варианты осуществления должны быть включены в перечень последовательностей в виде четырех отдельных последовательностей, каждая из которых имеет свой собственный идентификационный номер последовательности.

cgaatgacccactacgaatgbcacgaatgbcccaca (SEQ ID NO: 71)

cgaatgtcccactacgaatgvcacgaatgvcccaca (SEQ ID NO: 72)

cgaatggcccactacgaatghcacgaatghcccaca (SEQ ID NO: 73)

cgaatgccccactacgaatgdcacgaatgdcccaca (SEQ ID NO: 73)

(При этом b = t, g или c; v = a, g или c; h = t, a или c; и d = t, g или a; см. Приложение 1, Раздел 1, Таблица 1).

В соответствии с параграфом 15 ST.26, наиболее ограничивающий символ должен использоваться для представления изменяемых позиций. Следовательно, $n_2$ и $n_3$ не должны быть представлены в последовательности как “n”.

**ПРЕДУПРЕЖДЕНИЕ**: предпочтительное представление данной последовательности, указанное выше, предназначено для предоставления перечня последовательностей на дату подачи заявки на патент. Такое же представление может не быть применимо к перечню последовательностей, предоставляемому после даты подачи заявки на патент, т.к. необходимо обдумать вопрос, может ли предоставляемая позже информация быть добавлена патентным Ведомством к первоначальному раскрытию.

**Соответствующие параграфы ST.26:** 7(a), 15 и 95(b)

[Дополнение к Приложению 6 ST.26 следует]

## ДОПОЛНЕНИЕ

### Руководство по последовательностям в XML

Дополнение доступно по адресу:

https://www.wipo.int/standards/en/xml_material/st26/st26-annex-vi-appendix-guidance-document-sequences_v1_7.xml

[Приложение VII следует]

## Приложение 7

РЕКОМЕНДАЦИИ ПО ПРЕОБРАЗОВАНИЮ ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ИЗ ST.25 В ST.26:

ПОТЕНЦИАЛЬНОЕ ДОБАВЛЕНИЕ ИЛИ УДАЛЕНИЕ ПРИЗНАКОВ ИЗОБРЕТЕНИЯ.

*Версия 2.0*

*Одобрено Комитетом по Стандартам ВОИС (КСВ)
на своей тринадцатой сессии 14 ноября 2025*

*Введение*

В Стандартах ВОИС ST.25 и ST.26 требования к представлению нуклеотидных и аминокислотных последовательностей различаются. Вследствие чего был поднят вопрос о том, потребует ли представление перечня последовательностей, как части международной заявки, в соответствии со Стандартом ST.26, добавления или удаления какой-либо части признаков изобретения, которая может не поддерживаться заявкой, по которой испрашивается приоритет.

*Область применения*

В этом документе рассматриваются обязательные требования Стандарта ST.26 и любые потенциальные последствия этих требований. Данный документ не рассматривает все возможные сценарии; если способы представления в ST.26 информации, содержащейся в перечне последовательностей ST.25, не ясны, то эта информация всегда может быть включена в описание заявки, чтобы избежать исключения части признаков изобретения.

*Рекомендации для возможного добавления или исключения части признаков изобретения*

Рассмотрение вопросов, содержащихся в этом документе, показывает, что сам по себе переход от ST.25 к ST.26 по существу не ведёт к добавлению или удалению части признаков изобретения, в частности, когда перечень последовательностей в ST.25 полностью соответствует Стандарту ST.25. Однако существуют определенные сценарии, которые потребуют от заявителя осторожности. Рекомендации были предоставлены, чтобы избежать добавления или исключения части признаков изобретения.

*Сценарий 1*

ST.25 использует числовые идентификаторы для теговой разметки различных типов данных, например, <110> для имени заявителя. ST.26 использует термины на английском языке в качестве названий и атрибутов элементов для тегирования данных.

Рекомендация:

Термины ST.26 просто описывают тип содержания данных; следовательно, использование названий и атрибутов элементов ST.26 не является дополнением к признакам изобретения.

*Сценарий 2*

ST.26 явным образом требует включения: (a) разветвленных последовательностей; (b) последовательностей с D-аминокислотами; (c) нуклеотидных аналогов и (d) последовательностей, содержащих «AP-site» или «abasic site» (участки с удаленными азотистыми основаниями). В соответствии с ST.25 требование включения или запрета таких последовательностей неясно.

Рекомендация:

Раскрытие изобретения, содержащегося в заявке, должно быть достаточно для представления этих последовательностей в перечне последовательностей по ST.26, несмотря на то, что они, возможно, не были включены в перечень последовательностей по ST.25.

Для определенных типов информации, по требованию ST.26, необходимо соблюдать осторожность, чтобы не добавлять признаков изобретения помимо тех, что приведены в описании, например, см. обсуждение ниже (в Сценарии 4) по квалификатору mol_type для нуклеотидных последовательностей.

*Сценарий 3*

ST.26 разрешает включение последовательностей с менее чем 10-ю конкретно определенными нуклеотидами (не включая “n”) и менее чем 4-мя специально определенными аминокислотами (не включая “X”).

Рекомендация:

Последовательности менее чем с 10-ю конкретно определенными нуклеотидами и менее чем 4-мя специально определенными аминокислотами могут оставаться в перечне последовательностей в ST.26. Альтернативно, эти последовательности могут быть удалены из перечня последовательностей. Если выбрана вторая опция, следует быть осторожным и убедиться, что удаленные последовательности были включены в тело заявки, если эти последовательности еще не были определены там.

*Сценарий 4*

Как для нуклеотидных последовательностей, так и для аминокислотных последовательностей ST.26 имеет обязательный функциональный ключ "source" с двумя обязательными квалификаторами, одним из которых является “mol_type”. ST.25 имеет соответствующий функциональный ключ для нуклеотидных последовательностей (который используется редко) без соответствующих квалификаторов, и нет соответствующего функционального ключа для аминокислотных последовательностей.

**Нуклеотидные последовательности**

ST.26 – функциональный ключ 5.37 “source”; обязательный квалификатор 6.39 ‘mol_type” (см. параграф 75 ST.26).

| Квалификатор | Значение |
|---|---|
| mol_type | genomic DNA |
| | genomic RNA |
| | mRNA |
| | tRNA |
| | rRNA |
| | other DNA (applies to synthetic molecules) |
| | other RNA (applies to synthetic molecules) |
| | transcribed RNA |
| | viral cRNA |
| | unassigned DNA (applies where *in vivo* molecule is unknown) |
| | unassigned RNA (applies where *in vivo* molecule is unknown) |

**Аминокислотные последовательности**

ST.26 – функциональный ключ 7.30 source; обязательный квалификатор 8.1 mol_type (см. параграф 75 ST.26)

| Квалификатор | Значение |
|---|---|
| mol_type | protein |

Рекомендация:

Единственный вопрос, вызывающий озабоченность - это значения контролируемой лексики, связанные с квалификатором “mol_type” для нуклеотидных последовательностей. Некоторые из значений, перечисленных выше, могут быть недостаточно раскрыты в описании. Однако можно

избежать добавления дополнительных признаков изобретения путем использования наиболее общего значения для конкретной последовательности, например, “другая ДНК” и “другая РНК” для синтетической молекулы и “нераспределённая ДНК” и “нераспределённая РНК” для *in vivo* молекулы.

*Сценарий 5*

Когда последовательность включает “Xaa”, ST.25 требует, чтобы дополнительная информация, касающаяся этого остатка, была включена в поле <223>, которое сопровождает поля <221> (название признака) и <222> (местоположение признака). ST.25 не предоставляет значение для “Xaa” по умолчанию (в ST.26 - “X”). В то же время ST.26 даёт такое значение по умолчанию, и, следовательно, дополнительная информация не всегда требуется. Две из наиболее часто используемых аннотаций в пептидных последовательностях - «любая аминокислота» или «любая аминокислота природного происхождения» для вариантов “Xaa” или “X”. Этот язык может быть интерпретирован для включения аминокислот других, чем перечислены в таблицах аминокислот, содержащихся или в ST.25, или в ST.26. Значением по умолчанию в ST.26 для «X» без дальнейшей аннотации является любая из 22 отдельных аминокислот, перечисленных в Приложении 1 (см. Раздел 3, Таблица 3). Это значение по умолчанию в ST.26 может представлять собой добавление или удаление признаков изобретения, и, следовательно, негативно влиять на объем патентной заявки при переходе от ST.25 к ST.26.

Рекомендации:

(a) Когда перечень последовательностей в ST.25 включает в себя название признака <221>, местоположение признака <222>, соответствующее Xaa, и дополнительную информацию <223> о Xaa, а название признака <221> также является соответствующим функциональным ключом в ST.26, например, SITE, VARIANT или UNSURE, то следует использовать функциональный ключ ST.26. Кроме того, чтобы избежать потенциального удаления части признаков изобретения, информация в поле <223> должна быть включена в сопровождающий квалификатор “note”.

(b) Когда перечень последовательностей в ST.25 включает в себя название признака <221>, местоположение признака <222>, соответствующее Xaa, и дополнительную информацию <223> о Xaa, а название признака <221> не является функциональным ключом в ST.26, то следует использовать функциональные ключи ST.26 SITE или REGION, в зависимости от ситуации. Кроме того, чтобы избежать потенциального удаления части признаков изобретения, информация в поле <223>, а также неподходящее название признака <221> должны быть включены в сопровождающий квалификатор «note». Например, в перечне ST.25 использовано имя признака «<221> Variable», которого нет в ST.25 или ST.26, вместе с последующей информацией <223>: «Xaa является любой аминокислотой». В этом примере значение квалификатора “note” в ST.26 было бы: «Variable - Xaa является любой аминокислотой».

(c) Когда в перечне последовательностей по ST.25 не приводятся поля <221>, <222> или <223>, соответствующие Xaa, или когда поля <221> и <222>, соответствующие Xaa, включены, но в соответствующем поле <223> нет информации (ни один из сценариев не соответствует ST.25, но тем не менее это произошло), любая информация, содержащаяся в заявке для описания «Xaa», должна быть включена в квалификатор «note» ST.26 вместе с соответствующим функциональным ключом, например SITE, REGION, или UNSURE, и местоположением.

*Сценарий 6*

В ST.25 урацил представлен в последовательности как «u», а тимин представлен как «t». В ST.26 урацил и тимин представлены в последовательности как «t» и без дальнейших пояснений; «t» представляет урацил в РНК и тимин в ДНК.

Рекомендации:

(a) Когда последовательность ДНК содержит урацил, ST.26 рассматривает его как модифицированный нуклеотид, и требует, чтобы урацил был представлен как «t» и далее был описан с использованием функционального ключа «modified_base», квалификатора «mod_base» с «OTHER» в качестве значения квалификатора, и квалификатором «note» с «uracil» в качестве

значения квалификатора. Эта аннотация ST.26 не считается добавлением признаков изобретения, в котором последовательность ДНК по ST.25 содержала «u».

(b) Когда последовательность РНК содержит тимин, ST.26 рассматривает его как модифицированный нуклеотид, и требует, чтобы тимин был представлен как «t» и дополнительно описан с использованием функционального ключа «modified_base», квалификатора «mod_base» с «OTHER» в качестве значения квалификатора и квалификатором «note» с “thymine” в качестве значения квалификатора. Эта аннотация в ST.26 не считается добавлением признаков изобретения, в котором последовательность РНК ST.25 содержит «t».

*Сценарий 7*

Как в ST.25, так и в ST.26, модифицированные нуклеотиды или аминокислоты должны иметь последующее описание. В ST.26 идентификация модифицированного нуклеотида может быть указана с использованием аббревиатуры из Приложения 1, Раздела 2, Таблицы 2, где это применимо. В противном случае должно быть указано полное не сокращенное название модифицированного нуклеотида. Аналогичным образом, обозначение модифицированной аминокислоты может быть указано с использованием аббревиатуры из Приложения 1, Раздела 4, Таблицы 4, где это применимо, В противном случае должно быть указано полное несокращенное название модифицированной аминокислоты. В отличие от этого, если модифицированный остаток не содержится в таблице ST.25, использование полного, не сокращенного названия не требуется, и нередко вместо него используется сокращение.

Рекомендации:

(a) Когда использовалось только сокращенное название, либо для модифицированного нуклеотида, либо для модифицированной аминокислоты, как в заявке, так и в перечне последовательностей по ST.25, и оно не указано в Приложении 1, Разделе 2, Таблице 2 или Разделе 4, Таблице 4, а сокращенное название известно в данной области техники для ссылки только на один конкретный модифицированный нуклеотид или модифицированную аминокислоту, тогда использование полного, не сокращенного названия само по себе не будет являться дополнением признаков изобретения.

(b) Когда только сокращенное название, которое отсутствует в Приложении 1, Разделе 2, Таблице 2 или Разделе 4, Таблице 4 использовалось как в заявке, так и в перечне последовательностей по ST.25 либо для модифицированного нуклеотида, либо для модифицированной аминокислоты (а заявка не содержит химической структуры), и сокращенное название в данной области техники не известно как ссылка на один конкретный модифицированный нуклеотид или модифицированную аминокислоту, т.е. сокращение либо вообще не известно в данной области, либо может представлять множество различных модифицированных нуклеотидов или модифицированных аминокислот, тогда соблюдение ST.26 в этой ситуации невозможно без введения дополнительных признаков изобретения. Конечно, в этом случае приоритетная заявка и перечень последовательностей сами по себе неопределённы. Во избежание возможного удаления части признаков изобретения сокращенное название из перечня последовательностей по ST.25 следует разместить в квалификаторе «note» ST.26 в дополнение к значению полного несокращенного названия модифицированного нуклеотида или модифицированной аминокислоты. Полное несокращенное название модифицированного нуклеотида или модифицированной аминокислоты, требуемое в перечне последовательностей ST.26, не будет приоритетным по более ранней заявке. Следует позаботиться о том, чтобы при составлении исходных (в ST.25) перечня последовательностей и описания заявки, в них было включено несокращенное название для предотвращения проблем в будущем.

*Сценарий 8*

ST.25 содержит ряд функциональных ключей, которые не содержатся в ST.26. Поэтому заявители должны позаботиться о том, чтобы ввести информацию, содержащуюся в этих функциональных ключах ST.25, способом, совместимым с ST.26, без введения или удаления части признаков изобретения.

Рекомендации:

В следующей таблице приведены рекомендации относительно того, каким образом информация, содержащаяся в функциональном ключе бывшего ST.25, может быть включена в соответствии со стандартом ST.26 без введения или удаления части признаков изобретения. Номера 1-23 являются функциональными ключами, связанными с нуклеотидными последовательностями, а номера 24-43 являются функциональными ключами, связанными с аминокислотными последовательностями.

| No. | ST.25 Функциональный ключ <221> | эквивалент ST.26 | | |
|---|---|---|---|---|
| | | **Функциональный ключ** | **Квалификатор** | **Значение квалификатора** |
| 1 | allele | misc_feature | allele | значение <223> |
| 2 | attenuator | regulatory1 | regulatory_class[1] | “attenuator” |
| | | | note (если <223> есть) | значение <223> |
| 3 | CAAT_signal | regulatory[1] | regulatory_class[1] | “CAAT_signal” |
| | | | note (если <223> есть) | значение <223> |
| 4 | conflict | misc_feature | note | “conflict” и значение <223> |
| 5 | enhancer | regulatory[1] | regulatory_class[1] | “enhancer” |
| | | | note (если <223> есть) | значение <223> |
| 6 | GC_signal | regulatory[1] | regulatory_class[1] | “GC_signal” |
| | | | note (если <223> есть) | <223> value |
| 7 | LTR | mobile_element[1] | rpt_type[1]1 | “long_terminal_repeat” |
| | | | note (если <223> есть) | <223> value |
| 8 | misc_signal | regulatory[1] | regulatory_class[1] | “other” |
| | | | note (если <223> есть) | значение <223> |
| 9 | mutation | variation | note | “mutation” и значение <223> |
| 10 | old_sequence | misc_feature | note | “old_sequence” и значение <223> |
| 11 | polyA_signal | regulatory[1] | regulatory_class[1] | “polyA_signal_sequence” |
| | | | note (если есть <223>) | значение <223> |
| 12 | promoter | regulatory[1] | regulatory_class[1] | “promoter” |
| | | | note (если есть <223>) | значение <223> |
| 13 | RBS | regulatory[1] | regulatory_class[1] | “ribosome_binding_site” |
| | | | note (если есть <223>) | значение <223> |
| 14 | repeat_unit (a) when repeat_region not used | misc_feature | note | “repeat_unit” и значение <223> |
| | repeat_unit (b) when repeat_region used | repeat_region | rpt_unit_range | 1st residue..last residue |
| | | | note (если есть <223>) | значение <223> |

[1] ST.26 может потребовать, чтобы конкретная функция ST.25, например, TATA_signal, была заменена более общим функциональным ключом/квалификатором/значением, например, regulatory/regulatory_class/ TATA_box.

| No. | ST.25 Функциональный ключ <221> | эквивалент ST.26 | | |
|---|---|---|---|---|
| | | **Функциональный ключ** | **Квалификатор** | **Значение квалификатора** |
| 15 | satellite | repeat_region | satellite | “satellite” (или “microsatellite”или “minisatellite” – если поддерживается) |
| | | | note (если <223> есть) | значение <223> |
| 16 | scRNA | ncRNA[1] | ncRNA_class[1] | “scRNA” |
| | | | note (если есть <223>) | значение <223> |
| 17 | snRNA | ncRNA[1] | ncRNA_class[1] | “snRNA” |
| | | | note (если есть <223>) | значение <223> |
| 18 | TATA_signal | regulatory[1] | regulatory_class[1] | “TATA_box”2 |
| | | | note | TATA_signal and (if present): значение <223> |
| 19 | terminator | regulatory[1] | regulatory_class[1] | “terminator” |
| | | | note (если есть <223>) | значение <223> |
| 20 | 3’clip | misc_feature | note | “3’clip” и значение <223> |
| 21 | 5’clip | misc_feature | note | “5’clip” и значение <223> |
| 22 | -10_signal | regulatory[1] | regulatory_class[1] | “minus_10_signal” |
| | | | note (если есть <223>) | значение <223> |
| 23 | -35_signal | regulatory[1] | regulatory_class1 | “minus_35_signal” |
| | | | note (если есть <223>) | значение <223> |
| 24 | NON_CONS | Данный признак относится к пропуску неопределённого числа остатков в единичной последовательности, что запрещено как в ST.25 (параграф 22), так и в ST.26 (параграф 37). Следовательно, каждый участок специально определённых остатков, охваченных в ST.26 параграфом 7, должен быть включен в перечень последовательностей в виде отдельной последовательности и должен иметь свой собственный идентификационный номер последовательности. Чтобы избежать добавления / удаления признаков изобретения, каждая такая последовательность должна быть снабжена примечанием, чтобы указать, что она является частью большей последовательности, которая содержит неопределенный пропуск. | | |
| | | REGION | note | Description |
| | | Description (описание) - где и с чем связана последовательность, например, этот остаток связан с N-концом пептида, имеющим N-концевой Gly-Gly и разрыв неопределенной длины. | | |
| 25 | SIMILAR | REGION | note | “SIMILAR” и, если есть. значение <223> |
| 26 | THIOETH | CROSSLNK | note | “THIOETH” и, если есть, значение <223> |
| | | Дополнительную информацию см. в комментариях к функциональному ключу CROSSLNK в Приложении 1 Стандарта ST.26 | | |
| 27 | THIOLEST | CROSSLNK | note | “THIOLEST” и, если есть, значение <223> |
| | | Дополнительную информацию см. в CROSSLNK в Приложении 1 Стандарта ST.26 Комментарии к функциональному ключу | | |
| 28 | VARSPLIC | Обсуждается в сценарии 13 ниже | | |

[2] Во избежание добавления признаков изобретения, что может привести к частичной потере приоритета, рекомендуется включать более ограничивающий термин "TATA_signal" в квалификаторе " note ", как показано в таблице выше (пункт № 18). Если в редких случаях Заявитель считает, что использование значения "TATA_box" для квалификатора "regulatory_class" не подходит, вместо него можно использовать значение: "other". В этом случае термин «TATA_signal» должен быть включен в квалификатор «note», связанный с функциональным ключом «regulatory».

| No. | ST.25 Функциональный ключ <221> | эквивалент ST.26 | | |
|---|---|---|---|---|
| | | **Функциональный ключ** | **Квалификатор** | **Значение квалификатора** |
| 29 | ACETYLATION | MOD_RES | note | “ACETYLATION” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа MOD_RES в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 30 | AMIDATION | MOD_RES | note | “AMIDATION” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа MOD_RES в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 31 | BLOCKED | MOD_RES | note | “BLOCKED” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа MOD_RES в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 32 | FORMYLATION | MOD_RES | note | “FORMYLATION” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа MOD_RES в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 33 | GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION | MOD_RES | note | “GAMMA-CARBOXYLGLUTAMIC ACID HYDROXYLATION” и, если есть,. значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа MOD_RES в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 34 | METHYLATION | MOD_RES | note | “METHYLATION” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа MOD_RES в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 35 | PHOSPHORYLATION | MOD_RES | note | “PHOSPHORYLATION” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа MOD_RES в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |

| No. | ST.25 Функциональный ключ <221> | эквивалент ST.26 | | |
|---|---|---|---|---|
| | | **Функциональный ключ** | **Квалификатор** | **Значение квалификатора** |
| 36 | PYRROLIDONE CARBOXYLIC ACID | MOD_RES | note | “PYRROLIDONE CARBOXYLIC ACID” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа MOD_RES в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 37 | SULFATATION | MOD_RES | note | “SULFATATION” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа MOD_RES в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 38 | MYRISTATE | LIPID | note | “MYRISTATE” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа LIPID в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 39 | PALMITATE | LIPID | note | “PALMITATE” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа LIPID в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 40 | FARNESYL | LIPID | note | “FARNESYL” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа LIPID в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 41 | GERANYL-GERANYL | LIPID | note | “GERANYL-GERANYL” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа LIPID в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 42 | GPI-ANCHOR | LIPID | note | “GPI-ANCHOR” и, если есть, значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа LIPID в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |
| 43 | N-ACYL DIGLYCERIDE | LIPID | note | “N-ACYL DIGLYCERIDE” и, если есть. значение <223> |
| | | | note | Информация, требуемая Комментарием функционального ключа LIPID в Приложении 1 к Стандарту ST.26, если возможно (без добавления признаков изобретения) |

*Сценарий 9*

Отдельные функциональные ключи, представленные как в ST.25, так и в ST.26, как для нуклеотидных, так и для аминокислотных последовательностей, имеют обязательные квалификаторы в ST.26, как указано ниже. Функциональный ключ “modified_base” нуклеотидной последовательности также присутствует как в ST.25, так и в ST.26, однако Сценарий 7 содержит соответствующие рекомендации.

В ST.25 не было никаких квалификаторов, но было поле <223> free text (произвольный текст). Когда информация, содержащаяся в поле <223> ST.25, подходит в качестве значения для обязательного квалификатора в ST.26, то эта информация должна быть включена. Когда поле <223> ST.25 либо не было указано, либо содержит информацию, которая не подходит в качестве значения для обязательного квалификатора в ST.26, тогда заявители должны позаботиться о том, чтобы ввести информацию, содержащуюся в функциональном ключе ST.25 /поле <223> в соответствии со стандартом ST.26 без введения или удаления части признаков изобретения.

**Нуклеотидные последовательности** [3]

| Функциональный ключ | Обязательный квалификаторв |
|---|---|
| 5.12 - misc_binding | 6.3 - bound_moiety |
| 5.30 - protein_bind | 6.3 - bound_moiety |

Рекомендации:

(a) Если поле <223> в ST.25 отсутствует или не подходит, а в описании заявки раскрыто название молекулы/комплексного соединения, которое может связываться с местоположением признака нуклеиновой кислоты, то это название должно быть включено в квалификатор «bound_moiety».

(i) Любая информация, содержащаяся в поле <223> в ST.25, которая не подходит для включения в квалификатор «bound_moiety», должна быть включена в подходящий необязательный квалификатор функционального ключа, например, “note”.

(b) Если поле <223> в ST.25 отсутствует или не подходит, а описание заявки не раскрывает названия молекулы/комплексного соединения, которое может связываться с местоположением признака нуклеиновой кислоты, тогда в ST.26 должен быть использован функциональный ключ “misc_feature” вместо misc_binding или protein_bind с квалификатором “note”.

(i) Если поле <223> ST.25 отсутствовало, значением квалификатора “note” должно быть название функционального ключа ST.25;

(ii) Если поле <223> в ST.25 содержало неподходящую информацию, значением квалификатора “note” должно быть название функционального ключа ST.25 и информация из поля <223>.

## Аминокислотные последовательности [4]

| Функциональный ключ | Обязательный квалификатор |
|---|---|
| 7.2 – BINDING | 8.2 – note |
| 7.4 – CARBOHYD | 8.2 – note |
| 7.10 – DISULFID | 8.2 – note |
| 7.11 – DNA_BIND | 8.2 – note |
| 7.12 – DOMAIN | 8.2 – note |
| 7.16 – LIPID | 8.2 – note |
| 7.17 – METAL | 8.2 – note |
| 7.18 – MOD_RES | 8.2 – note |
| 7.23 – NP_BIND | 8.2 – note |
| 7.29 – SITE | 8.2 – note |
| 7.39 – ZN_FING | 8.2 – note |

[3] Номера, приведённые в таблице относятся к номерам функционального ключа и квалификатора ST.26 Приложения 1 Контролируемая лексика.
[4] Числовые ссылки в таблице ниже относятся к номерам функциональных ключей и квалификаторов Стандарта ST.26, Приложение I, Контролируемая лексика.

Рекомендации:

(a) Если поле <223> в ST.25 отсутствует или не подходит, а описание заявки раскрывало конкретную информацию, требуемую в обязательном квалификаторе, то эту информацию следует включить в обязательный квалификатор “note”.

(i) Любую информацию, содержащуюся в поле <223> в ST.25, которая не подходит для включения в обязательный квалификатор ”note” (см. определение функционального ключа и комментарий) следует включить во второй квалификатор “note”.

(b) Если поле <223> ST.25 отсутствует или не целесообразно, а описание заявки не раскрывало конкретную информацию, необходимую в обязательном квалификаторе, то в ST.26 вместо этого следует использовать функциональный ключ “SITE” (для одной аминокислоты) или «REGION» (для ряда аминокислот) с квалификатором “note”.

(i) Если поле <223> в ST.25 отсутствует, то в качестве значения квалификатора ”note” следует использовать название функционального ключа в ST.25;

(ii) Если поле <223> ST.25 содержало неподходящую информацию, то в качестве значения квалификатора “note” следует использовать название функционального ключа ST.25 и информацию из поля <223>.

*Сценарий 10*

Каждый конкретный функциональный ключ в ST.25 имеет поле <222> для указания местоположения признака; однако ST.25 не требует указания местоположения для большинства признаков, и формат информации о местоположении не стандартизирован. Более того, ST.25 не имеет операторов местоположения, например “join”. ST.26 имеет стандартизированные дескрипторы и операторы местоположения, и каждый признак должен содержать, по крайней мере, один дескриптор местоположения. (Признаки CDS являются особым случаем и обсуждаются ниже в сценарии 11).

Рекомендации:

(a) Если в перечне последовательностей в ST.25 имелось поле <222>, прямой импорт или импорт в формат ST.26 не должен вызывать какого-либо дополнительного рассмотрения по изменению признаков изобретения;

(b) Если в перечне последовательностей в ST.25 не было поля <222>, но информация о местоположении содержалась в описании заявки, то прямое импортирование или импорт в формат ST.26 не должен вызывать какого-либо дополнительного рассмотрения вопроса изменения признаков изобретения;

(c) Если ни перечень последовательностей ST.25, ни описание заявки не содержали информацию о местоположении, то, вероятно, признак относится ко всей последовательности в целом. (Указание местоположения, которое меньше всей последовательности целиком, без подтверждения в описании заявки, вероятно, будет представлять собой добавление/удаление признаков изобретения). Следует позаботиться о том, чтобы исходный перечень последовательностей (в ST.25) и описание заявки включали информацию о местоположении, чтобы расширить возможность избежать проблем в будущем.

*Сценарий 11*

В ST.25 кодирующая последовательность, которая кодировала единичный непрерывный полипептид, но которая была прервана одной или несколькими некодирующими последовательностями, например, интронами, была указана как множество отдельных CDS, как проиллюстрировано ниже:

<220>
<221> CDS
<222> (1) (571)

<220>
<221> CDS
<222> (639) (859)

В отличие от ST.25, ST.26 имеет оператор местоположения соединения, который указывает, что полипептиды, кодируемые указанными местоположениями, объединяются и образуют один непрерывный полипептид. (Примечание: и ST.25, и ST.26 требуют, чтобы стоп кодон был включен в признак местоположения CDS.)

Рекомендации:

(a) Если в перечне последовательностей по ST.25 или в описании заявки ясно указано, что полипептидные последовательности, кодируемые множественными отдельными CDS признаками, образуют один непрерывный полипептид, то кодирующая последовательность, прерванная интроном в одиночном CDS признаке должна быть представлена с оператором я местоположения соединения, как проиллюстрировано ниже, таким образом, что не вводится никаких дополнительных признаков изобретения:

<INSDFeature_key>CDS</INSDFeature_key>
<INSDFeature_location>join(1..571,639..859)</INSDFeature_location>

(b) Если в перечне последовательностей по ST.25 или в описании заявки не указано, что полипептидные последовательности, кодируемые двумя отдельными CDS признаками, образуют единый непрерывный полипептид, то использование оператора местоположения соединения, вероятно, представляло бы собой дополнительные признаки изобретения.

*Сценарий 12*

Стандарт ST.25 устанавливает, что названия признаков должны совпадать с одним из названий в таблицах 5 или 6. Однако в нормативных актах США указано, что эти названия признаков были рекомендованы, но не обязательны. Следовательно, последовательность в перечне последовательностей ST.25 (соответствующая нормативам США) может иметь «пользовательское» название функционального ключа без соответствующего функционального ключа в ST.26. Возможно также, что название признака не было приведено для поля <221>, или поле <221> отсутствует. Такие сценарии могут быть обработаны сходным образом.

Рекомендации:

«Пользовательское» название функционального ключа из ST.25 может быть в перечне последовательностей по ST.26 представлено следующим образом, не изменяя признаков изобретения:

| Тип | ST.25 Функциональный ключ <221> | Потенциальный эквивалент ST.26 | | |
|---|---|---|---|---|
| | | Функциональный ключ | Квалификатор | Значение квалификатора |
| NA | “Custom” feature key | misc_feature | note | «пользовательское» название функционального ключа и, если имеется, значение поля <223> |
| AA | “Custom” feature key | SITE or REGION | note | «пользовательское» название функционального ключа и, если имеется, значение поля <223> |

*Сценарий 13*

ST.25 содержит функциональный ключ «VARSPLIC», определенный как «описание вариантов последовательности, полученных альтернативным сплайсингом». В ST.26 «VARSPLIC» был заменен более общим функциональным ключом VAR_SEQ, определенным как «описание вариантов последовательности, полученных с помощью альтернативного сплайсинга, альтернативного использования промотора, альтернативного инициирования и изменение структуры рибосом». Поэтому в перечне последовательностей по ST.26 не следует использовать «VAR_SEQ» вместо «VARSPLIC» без дальнейшего пояснения.

Рекомендации:

В ST.26 признак «VAR_SEQ» следует использовать с квалификатором «note», в значение которого следует включать пояснение более узкой области применения ST.25, например: «вариант последовательности, полученный альтернативным сплайсингом». Любая дополнительная информация, содержащаяся в сопровождающем поле <223> ST.25, также должна быть включена в квалификатор «note».

*Сценарий 14*

Если источник последовательности был искусственным, поле <213> «Организм» в ST.25 требует фразы “Artificial Sequence” («Искусственная последовательность»). В ST.26 функциональный ключ “source” требует квалификатора “organism”, значение которого должно быть указано как “synthetic construct” («синтетическая конструкция»), а не “Artificial Sequence” («искусственная последовательность»).

Рекомендация:

Значение для квалификатора “organism” в ST.26 должно быть указано как “synthetic construct” («синтетическая конструкция»). Чтобы избежать потенциального удаления части признаков изобретения, любая пояснительная информация, содержащаяся в требуемом ST.25 поле <223>, следует включать в квалификатор “note” (функционального ключа “source”).

*Сценарий 15*

Если научное название организма - источника последовательности неизвестно, поле <213> Organism в ST.25 требует включения термина “Unknown” («Неизвестный»). В ST.26 функциональный ключ “source” требует наличия квалификатора “organism”, значение которого должно быть указано как “unidentified” («неустановленный»), а не “Unknown” («неизвестный»).

Рекомендация:

Значение для квалификатора “organism” в ST.26 должно быть указано как “unidentified” («неустановленный»)ё. Чтобы избежать потенциального удаления части признаков изобретения, любую пояснительную информацию, содержащуюся в требуемом по ST.25 поле <223>, следует включать в квалификатор “note” (функционального ключа “source”).

*Сценарий 16*

ST.25 допускает для представленных аминокислот опциональное включение отрицательных чисел, считая в обратном порядке, начиная с аминокислоты под номером 1, для аминокислот, предшествующих зрелому белку, например, пред-последовательности, про-последовательности, пред-про-последовательности и сигнальные последовательности. ST.26 не допускает использования отрицательных чисел в местоположении признака.

Рекомендации:

(a) Если перечень последовательностей в ST.25 имел признак или признаки, представленные в поле <221> и сопутствующем поле <222>, которые содержали отрицательную и/или положительную нумерацию, например, «PROPEP» и/или «CHAIN», то в

перечне последовательностей по ST.26 следует использовать соответствующий функциональный ключ, например, «PROPEP» и/или «CHAIN». Квалификатор «note» может быть использован с информацией из поля <223>, если таковая имелась, в качестве значения квалификатора;

(b) Если в перечне последовательностей по ST.25 не было признака или признаков, представленных в поле <221> и сопутствующем поле <222>, но информация относительно отрицательной и/или положительной нумерации содержалась в описании заявки, то в перечне последовательностей в ST.26 следует использовать соответствующий функциональный ключ, например, «PROPEP» и/или «CHAIN». В других случаях можно использовать функциональный ключ “REGION”. Квалификатор “note” может использоваться с информацией из описания заявки, если таковая имеется, в качестве значения квалификатора;

(c) Если ни перечень последовательностей в ST.25, ни описание заявки не содержат информации, поясняющей отрицательный и/или положительный порядок номеров, то во избежание потенциального удаления признаков изобретения в перечне последовательностей по ST.26, следует использовать функциональный ключ “REGION”, где местоположение признака охватывает отрицательно пронумерованный участок в последовательности в ST.25. Кроме того, для указания того, что аминокислотная последовательность была отрицательно пронумерована в перечне последовательностей по ST.25 к заявке, для которой испрашивается приоритет, следует использовать квалификатор “note”.

*Сценарий 17*

ST.25 предусматривает публикацию информации в полях с <300> по <313>. ST.26 не предусматривает включение такой информации.

Рекомендация:

Информацию, содержащуюся в полях с <300> по <313> ST.25, следует вводить в сопровождающие перечень последовательностей материалы заявки, если она еще не содержится в них.

*Сценарий 18*

ST.25 не даёт стандартизированного способа указать, что CDS участок нуклеотидной последовательности был преобразован с использованием таблицы генетических кодов, отличной от стандартной таблицы генетических кодов. ST.26, напротив, имеет квалификатор “transl_table”, который может быть использован с функциональным ключом «CDS», чтобы указать, что участок должен быть преобразован с использованием альтернативной таблицы генетических кодов. Если квалификатор “transl_table” не используется, предполагается использование стандартной таблицы генетических кодов.

Рекомендации:

(a) -> Если в перечне последовательностей по ST.25 или в описании заявки четко указано, что CDS участок должен быть преобразован с использованием альтернативной таблицы генетических кодов, то должен быть использован квалификатор «trans_table» с соответствующим номером таблицы генетических кодов в качестве значения квалификатора. Отказ от использования квалификатора «trans_table», скорее всего, станет добавлением признаков изобретения, так как по умолчанию будет предполагаться таблица «Стандартный код». Не включение в перечень последовательностей по ST.26 информации о таблице альтернативных генетических кодов из перечня последовательностей по ST.25 или из описания заявки, вероятно, будет представлять собой удаление части признаков изобретения.

(b) -> Если в перечне последовательностей по ST.25 или в описании заявки не указано, что CDS участок должен быть преобразован с использованием таблицы альтернативных генетических кодов, то не следует использовать квалификатор “transl_table”, или использовать его только со значением квалификатора “1”, т. е. с таблицей Стандартных кодов. Использование квалификатора «trans_table» с любым значением, отличным от «1», скорее всего, будет означать добавление или удаление части признаков изобретения.

*Сценарий 19*

ST.25 не даёт стандартизированного способа указать местоположение признака, в частности, имеющего месторасположение или участок, который выходит за пределы указанного остатка или диапазона остатков, например, участок CDS нуклеотидной последовательности, который выходит за пределы одного или обоих концов раскрытой последовательности. В отличие от этого, в ST.26 дескриптор местоположения признака предусматривает стандартизированный способ указания такого месторасположения или участка с помощью символов «<» или «>». Например, местоположение признака «CDS» должно включать в себя стоп кодон, даже когда стоп кодон не включен в саму раскрытую последовательность, путём указания местоположения, например, 1..>321.

Рекомендации:

(a) Когда в перечне последовательностей по ST.25 однозначно не указывалось, что местоположение признака выходит за пределы последовательности, и такое местоположение либо поддерживалось в описании заявки, либо ясно из самой последовательности, например, стоп-кодон CDS, который не содержится в последовательности, то символы «<» или «>» в перечне последовательностей ST.26 могут быть использованы, не добавляя признаков к изобретению.

b) -> Когда в перечне последовательностей по ST.25 однозначно не указывалось, что местоположение признака выходит за пределы последовательности, и такое местоположение не подтверждалось ни в раскрытии заявки, и не было ясно из самой последовательности, то соблюдение ST.26, без введения дополнительных признаков изобретения, может быть невозможным в этой ситуации. В этом случае приоритетная заявка и перечень последовательностей сами по себе, вероятно, не полны. В этой ситуации дескриптор местоположения признака в перечне последовательностей по ST.26 не будет приоритетным по более ранней заявке. Следует заранее позаботиться о том, чтобы разработать исходные перечень последовательностей (в ST.25) и описание заявки так, чтобы они включали полную информацию о признаках.

*Сценарий 20*

Приложение 1 ST.25 требует, чтобы в случаях, когда нуклеотидная последовательность содержит фрагменты как ДНК, так и РНК, в поле <212> следует ставить значение «ДНК», а комбинированную молекулу ДНК/РНК следует далее описать в разделе признаков с <220> по <223>; однако точная природа дальнейшего описания неясна, и это требование обычно не соблюдается. Параграф 55 ST.26 требует, чтобы каждый сегмент ДНК и РНК (ST.26 для внутренней согласованности использует термин «сегмент», а не «фрагмент») комбинированной молекулы ДНК/РНК должен быть далее описан с помощью функционального ключа «misc_feature», который включает местоположение сегмента и квалификатор “note”, который указывает, является ли сегмент ДНК или РНК.

Рекомендации:

(a) Если перечень последовательностей по ST.25 описывал сегменты ДНК и РНК в одном или нескольких признаках, используя <221> misc_feature, соответствующие местоположения в <222> и указания в <223> о том, какие сегменты были ДНК или РНК, то включение этой информации в формат ST.26, используя misc_feature для каждого сегмента ДНК и РНК не должно вызывать каких-либо вопросов о добавлении части признаков изобретения.

(b) Если перечень последовательностей по ST.25 описывал сегменты ДНК и РНК в одном или более признаках, используя функциональный ключ в <221> другой, чем misc_feature, соответствующие местоположения в <222> и указания в <223>, идентифицирующие, какие сегменты являются ДНК или РНК, то включение этой информации в формат ST.26 с использованием misc_feature для каждого сегмента ДНК и РНК и дополнительного квалификатора «note» с исходным функциональным ключом <221> в качестве значения, не должно поднимать вопрос о добавлении или удалении части признаков изобретения;

(c) Если перечень последовательностей в ST.25 обеспечивает идентификацию (ДНК или РНК) и местоположение каждого сегмента в поле <223>, которое не связано с полями <221> и <222>, например, определение для искусственной последовательности, то включение этой информации в формат ST.26, используя misc_feature для каждого сегмента ДНК и РНК, не должно вызывать вопрос о добавлении части признаков изобретения;

(d) Если перечень последовательностей в ST.25 описывал молекулу в признаке, используя <221> misc_feature и <223>, отмечая, что молекула является комбинированной молекулой ДНК/РНК, но не предоставляя информации о местоположении для каждого сегмента, и

(i) если в описании указаны местоположения каждого сегмента ДНК и РНК, то включение этой информации в формат ST.26, используя misc_feature для каждого сегмента ДНК и РНК, не должно вызывать каких-либо дополнительных вопросов по изменению объема изобретения;

(ii) если описание не содержит информацию о местонахождении каждого сегмента ДНК и РНК, то в этой ситуации компиляция с ST.26 без введения дополнительного объема изобретения может оказаться невозможной. В этом случае приоритетная заявка и перечень последовательностей сами по себе, вероятно, не полны. В этой ситуации любые описания местоположения признаков в перечне последовательностей по ST.26 не будут приоритетными по более ранней заявке. Следует позаботиться о том, чтобы при составлении исходных (в ST.25), перечня последовательностей и описание заявки включить в них полную информацию о признаках.

(e) Если перечень последовательностей в ST.25 описывал молекулу в признаке, используя функциональный ключ в <221>, другой, чем misc_feature, и <223>, отмечая, что молекула является комбинированной молекулой ДНК/РНК, но информация о местоположении для каждого сегмента не предоставлена, и

(i) если в описании приведены местоположения каждого сегмента ДНК и РНК, то включение этой информации в формат ST.26, используя misc_feature для каждого сегмента ДНК и РНК и дополнительный квалификатор «note» с исходным функциональным ключом <221> в качестве его значения, не должно вызывать вопрос о добавлении или удалении части признаков изобретения;

(ii) если описание не содержит информацию о местонахождении каждого сегмента ДНК и РНК, то компиляция со стандартом ST.26 без введения дополнительного объема изобретения может оказаться невозможной. В этом случае приоритетная заявка и перечень последовательностей сами по себе, вероятно, не полны. В этой ситуации любые описания местоположения признаков в перечне последовательностей по ST.26 не будут приоритетными по более ранней заявке. Следует позаботиться о том, чтобы составлять исходный (в ST.25) перечень последовательностей и описание заявки, включая в него полную информацию о признаках.

(f) Если в перечне последовательностей по ST.25 отмечено, что молекула является комбинированной молекулой ДНК/РНК в поле <223>, например, дано пояснение к искусственной последовательности, но не предоставлено никакого функционального ключа или информации о местоположении каждого сегмента, и

(i) если в описании указаны местоположения каждого сегмента ДНК и РНК, то включение этой информации в формат ST.26 с использованием misc_feature для каждого сегмента ДНК и РНК не должно вызывать каких-либо дополнительных вопросов по изменению объёма изобретения;

(ii) если описание не содержит информацию о местонахождении каждого сегмента ДНК и РНК, то в этой ситуации соответствие со стандартом ST.26 без введения дополнительного объёма изобретения может оказаться невозможным. В этом случае приоритетная заявка и перечень последовательностей сами по себе, вероятно, не полны. В такой ситуации любые описания местоположения признаков в перечне последовательностей в ST.26 не будут приоритетными по более ранней заявке. Следует позаботиться о том, чтобы составлять исходный (в ST.25) перечень последовательностей и описание заявки, включая в него полную информацию о признаках.

[Конец Приложения 7 и Стандарта]

[Приложение II (пересмотренное добавление к Приложению IV ST.26) следует]

[Конец документа]