

СТАНДАРТ ST. 26

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

Версия 1.7

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

Редакционное примечание подготовлено Международным бюро

Десятая сессия Комитета по стандартам ВОИС приняла решение о том, что версия 1.6 стандарта ВОИС ST.26 вступит в силу 1 июля 2024 г. До того момента следует продолжать использовать версию 1.6 стандарта ST.26.

Стандарт публикуется с целью информирования Ведомств интеллектуальной собственности и других заинтересованных сторон.

Содержание

ВВЕДЕНИЕ	3
ОПРЕДЕЛЕНИЯ	3
ОБЛАСТЬ ПРИМЕНЕНИЯ	6
ОТСЫЛКИ.....	7
ПРЕДСТАВЛЕНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ.....	7
Нуклеотидные последовательности	7
Аминокислотные последовательности	11
Представление особых ситуаций	13
СТРУКТУРА ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В XML	14
Корневой (основной) элемент	15
Часть общей информации	16
Часть с данными последовательности	21
Таблица характеристик	23
Функциональные ключи	24
Обязательные функциональные ключи	24
Местоположение признака	24
Квалификаторы признака	28
Обязательные квалификаторы признаков	28
Элементы квалификатора	29
Свободный текст	331
Кодирующие последовательности.	34
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 добавленный или удалённый объект (признак).

СТАНДАРТ ST. 26

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

Версия 1.7

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

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 любую аминокислоту, которая может быть представлена с помощью любого символа, определённого в Приложении 1 (см. раздел 3, Таблица 3). Такие аминокислоты включают, в частности, D-аминокислоты и аминокислоты, содержащие модифицированные или синтетические боковые цепи. Аминокислоты будут истолкованы как немодифицированные L-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 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.

с) «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, His₆Gln).

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.

е) «модифицированная аминокислота» означает любую аминокислоту, как описано в параграфе 3 (а), отличную от L-аланина, L-аргинина, L-аспарагина, L-аспарагиновой кислоты, L-цистеина, L-глутамина, L-глутаминовой кислоты, L-глицина, L-гистидина, L-изолейцина, L-лейцина, L-лизина, L-метионина, L-фенилаланина, L-пролина, L-пирролизина, L-серина, L-селеноцистеина, L-треонина, L-триптофана, L-тирозина или L-валина.

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

г) «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 сдвигу оснований с комплементарной нуклеиновой кислотой; примеры основных частей включают аминокислоты, такие как в пептидных нуклеиновых кислотах, молекулы гликоля, как в гликоль нуклеиновых кислотах, молекулы тетрофуранозидового сахара, как в тетроза-нуклеиновых кислотах, морфолиновые кольца и фосфородиамидатные группы, как в морфолинах, и молекулы циклогексанила, как в циклогексанильных нуклеиновых кислотах.

и

(ii) основной частью является либо:

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

(2) отсутствие 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.

к) «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 последовательность» 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 (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; игнорирование этого требования приведет к несоблюдению Стандарта.

- (с) «не должно» относится к запрету в Стандарте.
- (d) «следует» относится к настоятельно рекомендованному подходу, но не к требованию.
- (е) «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 патентных заявках.

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:

а) 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

б) неразветвлённой последовательностью или линейным участком разветвлённой последовательности, содержащей четыре и более специально определённых аминокислот, где аминокислоты образуют одну пептидную основу и где смежные аминокислоты соединены пептидными связями.

8. Перечень последовательностей не должен включать никаких последовательностей с менее чем десятью специально определёнными нуклеотидами или с менее чем четырьмя специально определёнными аминокислотами, как последовательность со своим собственным присвоенным идентификационным номером 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](#)

Стандарт ВОИС [ST.3](#)

Стандарт ВОИС [ST.16](#)

Стандарт ВОИС [ST.25](#)

Стандартное представление календарных дат
на опубликованных документах с
использованием григорианского календаря
Рекомендуемый Стандарт на двубуквенные
коды для представления стран,
административных единиц и
межправительственных организаций;
Рекомендуемые стандартные коды для
идентификации различных видов патентных
документов
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. раздел 58). Общее количество последовательностей должно быть указано в перечне последовательностей и должно равняться общему количеству идентификационных номеров последовательности, независимо от того, будет ли это некоторая последовательность или код «000».

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 из двух цепочек, раскрывается путём перечисления остатков обеих цепочек, которые должны быть представлены как:

а) 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

б) 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 используются 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 считается модифицированным нуклеотидом и должен быть в дальнейшем описан 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, в случае если нуклеотидами 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”, 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. Любой модифицированный нуклеотид в последовательности, который не может быть представлен иначе, как любой другой символ в Приложении 1 (см. Глава 1, Таблица 1), т.е., «другой» нуклеотид, такой как неприродные нуклеотиды, должен быть представлен символом “n.” Символ “n” является эквивалентом только одного основания.

17. Модифицированный нуклеотид должен быть в дальнейшем описан в таблице свойств (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”), 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” как значением квалификатора и квалификатора “note” с “урацилом” или “timiном”, 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_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

```
<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>xanthine</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</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_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>2,3-dihydroxypropyl
        nucleosides</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>439</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

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

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

22. Участок, содержащий известное число смежных остатков “a”, “c”, “g”, “t” или “n”, 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 - 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_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>isoguanine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Аминокислотные последовательности

24. Аминокислоты в аминокислотной последовательности должны быть представлены в аминокислотном 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 пептидными связями, то есть последовательность не имеет аминокислотных концов, заявитель должен выбрать аминокислоту с положением остатка номер 1. Нумерация непрерывна по всей последовательности в аминокислотном 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, как альтернативу, две и более 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 один из символов “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 случая использования последующего описания в таблице характеристик. Символ “X” не должен использоваться для представления чего-либо иного, кроме аминокислоты. Символ “X” может быть использован для представления только одной аминокислоты вместе с последующим описанием в таблице характеристик, 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, включая D-аминокислоты, должны быть представлены в последовательности, как соответствующие немодифицированные аминокислоты. Любая модифицированная аминокислота в последовательности, которая не может быть представлена любым символом из Приложения 1 (см. Глава 3, Таблица 3), т.е. «другая» аминокислота, должна быть представлена как "X". 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 квалификатором "note" («ПРИМЕЧАНИЕ»). Функциональный ключ "MOD_RES" следует использовать для других посттрансляционно модифицированных аминокислот совместно с квалификатором "note"; в других случаях функциональный ключ "SITE" следует использовать совместно с квалификатором "note". Значение квалификатора "note" должно быть или аббревиатурой (сокращением), 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, несокращённые 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_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>3Hyp</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Пример 2: Непосттрансляционно модифицированная аминокислота

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name> note </INSDQualifier_name>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

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

Пример 3: D-аминокислота

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

32. Любая "unknown" («неизвестная») аминокислота, должна быть представлена как символ "X" в последовательности. "Unknown" («неизвестная») 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" («неизвестной») или "other" («другой») аминокислоты в соответствии с параграфом 32 выше:

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

34. Участки, содержащие известное число смежных остатков "X", для которых используют одинаковые описания, могут быть совместно описаны с помощью синтаксиса "x..y", как описание местонахождения в элементе INSDFeature_location (см. параграфы 64-70). 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 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 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 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 объявление:

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

- (b) Вторая строка XML файла должна содержать объявление 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 уникода UTF-8 со следующими ограничениями:

- а) 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. Резервированные символы «, &,», < и > (кодированные точки Unicode 0022, 0026, 0027, 003C и 003E соответственно), подлежат замене, как указано в параграфе 41; а также
- б) информация, содержащаяся во всех других элементах и атрибутах части общей информации и во всех других элементах и атрибутах части данных последовательности должны 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 символы¹ и 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. Единственными 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 (основной) 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
fileName	Файл с названием перечня последовательностей.	Необязательно
softwareName	Название программного обеспечения, создавшего этот файл.	Необязательно
softwareVersion	Версия программного обеспечения, создавшего этот файл.	Необязательно

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
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 патентной заявке и содержат следующие понятия:

Элемент	Описание	Обязательно/ Необязательно
ApplicationIdentification	Идентификация заявки, для которой представлен перечень последовательностей	Обязательно, если перечень последовательностей предоставлен в любое время после присвоения номера заявки
ApplicationIdentification состоит из:		
IPOfficeCode	Код ведомства подачи в соответствии со Стандартом ВОИС ST.3	Обязательно
ApplicationNumberText	Номер заявки, как установлено ведомством подачи (например, номер заявки PCT/IB2013/99999)	Обязательно
FilingDate	Дата подачи патентной 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	Отдельный уникальный идентификатор, присвоенный заявителю для идентификации отдельной заявки, введенный в символах, определенных 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

Признак	Описание	Обязательно/Необязательно
ApplicantName	Имя первого указанного заявителя, введённое в символах, определённых в параграфе 40 (а). Данный элемент включает обязательный атрибут <code>languageCode</code> как определено в параграфе 47.	Обязательно
ApplicantNameLatin	Если <code>ApplicantName</code> 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 <code>ApplicantName</code> содержит нелатинские символы
InventorName	Имя первого указанного изобретателя, введённое в символах, определённых в параграфе 40 (а)	Необязательно
InventorNameLatin	Если <code>InventorName</code> 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 (а) на языке подачи. Перевод названия изобретения на дополнительные языки может быть введен в символах, определённых в параграфе 40 (а) используя дополнительный элемент <code>InventionTitle</code> . Данный элемент включает обязательный признак <code>languageCode</code> , как указано в параграфе 48. В названии изобретения должно быть от двух до семи слов.	Обязательно на языке подачи. Необязательно для дополнительных языков.

Признак	Описание	Обязательно/Необязательно
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_SEQ1.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_SEQ1.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

разделам 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 патентной заявке	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 АА. (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_qual>
        <INSDQualifier>
          <INSDQualifier_name>organism</INSDQualifier_name>
          <INSDQualifier_value>synthetic construct</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
          <INSDQualifier_name>mol_type</INSDQualifier_name>
          <INSDQualifier_value>other DNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_qual>
    </INSDFeature>
    <INSDFeature>
      <INSDFeature_key>misc_feature</INSDFeature_key>
      <INSDFeature_location>1..60</INSDFeature_location>
      <INSDFeature_qual>
        <INSDQualifier>
          <INSDQualifier_name>note</INSDQualifier_name>
          <INSDQualifier_value>DNA fragment</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_qual>
    </INSDFeature>
    <INSDFeature>
      <INSDFeature_key>misc_feature</INSDFeature_key>
      <INSDFeature_location>61..120</INSDFeature_location>
```

```
<INSDFeature_qual>  
  <INSDQualifier>  
    <INSDQualifier_name>note</INSDQualifier_name>  
    <INSDQualifier_value>RNA</INSDQualifier_value>  
  </INSDQualifier>  
</INSDFeature_qual>  
</INSDFeature>  
</INSDSeq_feature-table>  
<INSDSeq_sequence>  
cgaccacgcgtccgaggaaccaaccatcacgtttgaggacttcgtgaaggaattggataatacccgctccctaccaaattgg  
cgagcgccgactcattgctcctcgtaccgctcgagcggc  
</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
<code>INSDFeature_key</code>	Слово или аббревиатура, указывающие признак	Обязательно
<code>INSDFeature_location</code>	Участок последовательности, соответствующий признаку	Обязательно
<code>INSDFeature_qual</code>	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 элемент `INSDSeq_sequence`. Только для нуклеотидных последовательностей дескриптор местоположения может быть местом между двумя соседними номерами остатков. Повторяющиеся описания местоположений могут быть использованы в соединении с оператором местоположения, когда признаку соответствуют дискретные места или области нуклеотидной последовательности (смотри параграфы 67-70).

66. Синтаксис для каждого типа описания 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, ограниченных диапазоном последовательности	$x..y$	Указывает на ограниченный непрерывный диапазон остатков, включая стартовый и конечный остаток.
Остатки перед первым или после последнего определённого номера остатка	$<x$ $>x$ $<x..y$ $x..>y$ $<x..>y$	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 разложения эндонуклеазы, номера позиций для смежных остатков, отделённых 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.

(с) Deskriptory mestopolozheniya tol'ko dlya aminokislotnykh posledovatel'nostey:

Тип дескриптора местоположения	Синтаксис	Описание
Количество остатков присоединено с помощью перекрестной связи внутри цепи	<code>x..y</code>	Указывает на аминокислоты, соединенные внутрицепочечной связью при использовании признака, который указывает на перекрестную ссылку внутри цепи, например "CROSSLNK" или "DISULFID".

67. Элемент `INSDFeature_location` нуклеотидных последовательностей может содержать один или несколько операторов местоположения. Оператор местоположения является префиксом для одного описания местоположения или комбинации дескрипторов местоположения, соответствующих одному, но характерному признаку, и определяет местоположение соответствующего признака в указанной последовательности или как построен признак. Перечень операторов местоположения с их определениями приводится ниже. Операторы местоположения могут использоваться только для нуклеотидов.

Синтаксическая структура местоположения	Описание местоположения
<code>join(location, location, ..., location)</code>	Указатели местоположений объединены (сквозное размещение), чтобы сформировать непрерывную последовательность.
<code>order(location, location, ..., location)</code>	Элементы находятся в определённом порядке, но ничто не предполагает целесообразности их соединения.
<code>complement(location)</code>	Указывает, что признак расположен в дополнительной цепи диапазона последовательности, определённого указателем местоположения, когда считывается в 5' - 3' направлении или в направлении, которое имитирует 5' - 3' направление.

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^u`, и не должны использоваться в соединении или указании порядка расположения. Использование операторов соединения предполагает, что остатки, описываемые дескрипторами местоположения, физически контактируют в биологических процессах (например, экзоны, способствующие кодирующим свойствам участка)

69. Оператор местоположения "complement" может быть использован в сочетании либо с "join", либо с "order" в одном и том же месте. Сочетания "join" и "order" в одном и том же месте не допускается.

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 ограниченный непрерывный диапазон остатков, остатки 340 и 565.
<1	Указывает на признак местоположения перед первым остатком.
<345..500	Указывает на то, что точный нижний предел признака неизвестен. Размещение начинается с некоторого остатка, предшествующего 345 и продолжается до остатка 500 включительно.
<1..888	Указывает на то, что признак начинается перед первым остатком последовательности и продолжается до остатка 888 включительно.
1..>888	Указывает на то, что признак начинается с первого остатка последовательности и продолжается за пределами остатка 888.
<1..>888	Указывает на то, 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 позицию между остатками 123 и 124.
join(12..78,134..202)	Указывает на то, что области с12 по 78 и с 134 по 202 должны быть объединены в одну непрерывную последовательность.
complement(34..126)	Начинается с нуклеотида, комплементарного 126 основанию, и заканчивается на нуклеотиде, комплементарном нуклеотиду 34 (признак находится на цепи, комплементарной представленной).
complement(join(2691..4571,4918..5163))	Последовательность нуклеотидов с 2691 по 4571 и с 4918 по 5163, которая комплементарна объединенным сегментам (признак находится на цепи, комплементарной представленной)

Пример местоположения	Описание
<code>join (complement (4918..5163) , complement (2691..4571))</code>	Объединение с 4918 по 5163 и с 2691 по 4571 оснований комплементарных цепей (признак находится на цепи, комплементарной представленной)

(с) Местоположение только для аминокислотных последовательностей:

Пример местоположения	Описание
<code>340..565</code>	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 перечне последовательностей XML символы “<” 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 получения информации о признаках в дополнение к той, которая даётся через функциональный ключ и местоположение признака. Есть три типа форматов для представления различных типов информации, предоставляемых квалификаторами, а именно:

- а) свободный текст (смотри параграфы 85 - 87);
- б) контролируемая лексика или числовые значения (например, номер или дата); и
- с) последовательности.

73. В Главе 6 Приложения 1 представлен единый перечень квалификаторов с определёнными значениями форматов, и если имеется, функциональным ключом для каждой последовательности нуклеотидов и в главе 8 содержится единый перечень квалификаторов и указанные форматы значений, если таковые имеются, 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 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_qual` 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
<code>INSDQualifier_name</code>	Название квалификатора (смотри Приложение 1, главы 6 и 8)	Обязательно
<code>INSDQualifier_value</code>	Значение квалификатора, если имеется, в определённом формате (смотри Приложение 1, главы 6 и 8) и состоящее из символов, как предусмотрено в пункте 40 (b).	Обязательно, когда точно определено (смотри параграф 87 и Приложение 1, главы 6 и 8)
<code>NonEnglishQualifier_value</code>	Значение квалификатора, если имеется, в определённом формате (смотри Приложение 1, главы 6 и 8) и состоящее из символов, как предусмотрено в пункте 40 (a).	Обязательно, когда точно определено (смотри параграф 87 и Приложение 1, главы 6 и 8)
<code>id</code>	Квалификатор со значением свободного текста, зависящим от языка, может быть однозначно идентифицирован с помощью необязательного XML-атрибута 'id' в элементе <code>INSDQualifier</code> (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 <code>INSDQualifier</code> , 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, глава 6) и “organism” для аминокислотных последовательностей (смотри Приложение 1, глава 8) должен раскрывать источник, например, отдельный организм или происхождение последовательности. Указанный организм должен быть выбран из систематической базы данных.

78. Если последовательность является природной и исходный организм имеет латинский род и обозначение вида, то обозначение должно быть использовано в качестве значения квалификатора. Предпочтительное английское общепринятое название может быть приведено с использованием квалификатора “note” для нуклеотидных последовательностей и аминокислотных последовательностей, но не должно быть использовано в значении квалификатора организма.

79. Следующие примеры иллюстрируют источник организм 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_qual>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>Solanum lycopersicum</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>note</INSDQualifier_name>
        <INSDQualifier_value>common name: tomato</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>genomic DNA</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>
</INSDSeq_feature-table>
```

Пример 2: Источник для аминокислотной последовательности:

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..174</INSDFeature_location>
    <INSDFeature_qual>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>protein</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>
</INSDSeq_feature-table>
```

80. Если последовательность является природной и исходный организм принадлежит к известному роду на латыни, но вид организма не указан или не опознан, то в качестве значения квалификатора должно указываться латинское обозначение рода, сопровождаемое “sp.”.

Например:

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

81. Если последовательность является природной и исходный организм не имеет обозначения рода и вида на латыни, то значение квалификатора организма должно указываться как “unidentified” («неидентифицированный»). Любая известная таксонометрическая информация должна быть указана в квалификаторе “note” для нуклеотидных последовательностей и квалификаторе “note” для аминокислотных последовательностей. 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 природная, значение квалификатора организма должно быть указано как “synthetic construct”. Дальнейшая информация в отношении способа генерации последовательности может быть указана с помощью квалификатора “note” для нуклеотидной последовательности и квалификатора “note” для аминокислотной последовательности. Например:

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

84. Квалификатор “mol_type” для нуклеотидных последовательностей (см. Приложение I, Глава 6) и “mol_type” для аминокислотных последовательностей (см. Приложение I, Глава 8) должен раскрывать тип молекулы, представленной в последовательности. Данный квалификатор отличен от элемента INSDSeq_moltype, описанного 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”;

(b) Для аминокислотных последовательностей значением квалификатора “mol_type” является “protein”.

Свободный текст

85. Свободный текст, как указано в параграфе 3 это тип 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, кроме квалификатора “translation”, свободный текст не должен превышать 1000 знаков (символов).

87. Свободный текст, зависящий от языка, как указано в параграфе 3, представляет собой 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, указаны в Приложении I, Раздел 6, Таблица 5. Квалификаторы для аминокислотных последовательностей с форматом значений свободного текста, зависящим от языка, указаны 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, считаются английскими в смысле 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) Если элементы `NonEnglishQualifier_value` и `INSDQualifier_value` оба присутствуют для одного квалификатора, информация, содержащаяся в 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, зависящим от языка, элемент INSDQualifier может включать в себя необязательный идентификатор атрибута. 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_qual>
    <INSDQualifier id="q1">
      <INSDQualifier_name>function</INSDQualifier_name>
      <INSDQualifier_value>binds to regulatory protein
Est3</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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_qual>
    <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_qual>
</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_qual>
    <INSDQualifier id="q1034">
      <INSDQualifier_name>note</INSDQualifier_name>
```

```
<NonEnglishQualifier_value>clive la chaîne glucidique
</NonEnglishQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

Кодирующие последовательности.

89. Для идентификации кодирующих последовательностей может быть использован функциональный ключ “CDS”, т.е. последовательности нуклеотидов, которые соответствуют последовательности аминокислот в белке и стоп кодоне. 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). Если квалификатор “`transl_table`” не применяется, допускается использование Таблицы стандартных кодов (смотри Приложение 1, глава 9, таблица 7).

91. Квалификатор «`transl_except`» должен использоваться с функциональным ключом «CDS» и квалификатором “`translation`” («перевод») 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 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 своей кодирующей последовательности.

Например:

```
<INSDFeature>
  <INSDFeature_key>CDS</INSDFeature_key>
  <INSDFeature_location>1..507</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>transl_table</INSDQualifier_name>
      <INSDQualifier_value>11</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>translation</INSDQualifier_name>
      <INSDQualifier_value>
MLVHLERTTIMFDFSSLINLPLIWGLLIAIAVLLYILMDGFDLGIGILLPFAPSDKCRDHMISSIAPFWDGNETWLV
LGGGGLFAAFPLAYSILMPAFYIPIIIMLLGLIVRGVSFEFRFKAEGKYRRLWDYAFHFGSLGA AFCQGMILGAFIH
GVEVNGRNFSGGQLM
      </INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>protein_id</INSDQualifier_name>
      <INSDQualifier_value>89</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Варианты

93. Исходная последовательность и любые варианты этой последовательности, раскрытые перечислением её остатков и охваченные параграфом 7, должны быть включены в перечень последовательностей со своим собственным присвоенным идентификационным номером последовательности.

94. Любой вариант последовательности, раскрытой как одна последовательность с перечислением альтернативных остатков на одной или нескольких позициях, должен быть включен в перечень последовательностей и должен быть представлен одной последовательностью, в которой перечислены альтернативные остатки, представленные 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:

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

б) должны быть представлены как отдельная последовательность с собственным присвоенным идентификационным номером последовательности, если они содержат варианты на множестве разных позиций и возникновение данных вариантов является взаимосвязанным; и

с) должны быть представлены как отдельная последовательность с собственным присвоенным идентификационным номером последовательности, в случае, если в последовательности содержатся замещения и добавления в более чем 1000 остатков. (см. параграф 86).

96. Таблица ниже показывает правильный способ использования 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 определённого 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 квалификаторах "note". В частности, перечень альтернативных аминокислот должен быть представлен как значение в квалификаторе "note", где «X» используется в последовательности, и представляет значение за исключением следующих «любой из '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 представлено в параграфе 86. Смотри параграф 100 для последовательностей, охваченных параграфом 7, которые представлены как замещение или добавление значения квалификатора.

98. Символы, определённые в Приложении 1 (см. главы с 1 по 4, таблицы с 1 по 4 соответственно), если необходимо, должны быть использованы для представления вариантов остатков. 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. "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_qual>
```

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

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

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

Пример 3: Функциональный ключ “misc_difference” для добавления в нуклеотидной последовательности. Последовательность “atgccaaatat” добавлена между положениями 100 и 101 исходной последовательности.

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

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

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

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

```
<INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>I, A, F, Y, aIle, MeIle, or Nle
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>MOD_RES</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name> note </INSDQualifier_name>
      <INSDQualifier_value>aIle, MeIle, or
Nle</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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_qual>
    <INSDQualifier>
      <INSDQualifier_name> note </INSDQualifier_name>
      <INSDQualifier_value>not K, R, or H</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

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

идентификационным номером последовательности.

[Приложение 1 к Стандарту ST.26 следует]