



СТАНДАРТ ST.25

СТАНДАРТ ПО ПРЕДСТАВЛЕНИЮ ПЕРЕЧНЕЙ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НУКЛЕОТИДОВ И АМИНОКИСЛОТ В ПАТЕНТНЫХ ЗАЯВКАХ

*Стандарт принят 28 мая 1998 г. на двадцать второй сессии
Исполнительно-координационного Комитета ПКИПС*

Рекомендуется, чтобы ведомства по промышленной собственности применяли положения "Стандарта по представлению перечней последовательностей нуклеотидов и аминокислот в международных заявках, поданных в соответствии с договором РСТ", как это изложено в Приложении С к Административным Инструкциям РСТ и *mutatis mutandis* в отношении всех патентных заявок, отличных от международных заявок РСТ, имея в виду, что отдельные положения, специфические для процедур и требований РСТ, могут применяться только к международным заявкам РСТ^(*). Текст стандарта РСТ воспроизводится на последующих страницах.

^(*) Если на 1-ое июля 1998 г. национальное законодательство и практика, применяемая любым из ведомств, не соответствует положениям первых двух предложений параграфа 3 "Стандарта по представлению перечней последовательностей нуклеотидов и аминокислот в международных заявках РСТ", то данное ведомство может решить не следовать указанным положениям до тех пор, пока сохраняется такое несоответствие.



ПРИЛОЖЕНИЕ С

СТАНДАРТ ПО ПРЕДСТАВЛЕНИЮ ПЕРЕЧНЕЙ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НУКЛЕОДИТОВ И АМИНОКИСЛОТ В МЕЖДУНАРОДНЫХ ЗАЯВКАХ, ПОДАВАЕМЫХ В СООТВЕТСТВИИ С ДОГОВОРом РСТ

ВВЕДЕНИЕ

1. Данный стандарт был разработан с тем, чтобы обеспечить стандартизацию представления перечней последовательностей нуклеотидов и аминокислот в международных патентных заявках. Данный стандарт должен обеспечить заявителю возможность составления единого перечня последовательностей, который бы приемлем для всех получающих ведомств, Международных поисковых органов и органов международной предварительной экспертизы для целей международной фазы, а также для всех указанных и выбранных ведомств для целей национальной фазы. Стандарт предназначен для повышения точности и качества представления перечней последовательностей нуклеотидов и аминокислот, приводимых в международных заявках, облегчения представления и удобного доведения последовательностей до сведения заявителей, широкой публики и экспертов, для содействия поиску данных о последовательностях, обеспечения обмена данными о последовательностях в электронной форме, а также вводу соответствующей информации в компьютеризованные базы данных.

ОПРЕДЕЛЕНИЯ

2. Для целей данного стандарта :

(i) выражение “Перечень последовательностей” означает часть описания заявки в том виде, как она подана, или документ, представленный в дополнение к заявке, которые содержат детальное раскрытие последовательностей нуклеотидов и/или аминокислот, а также другой приводимой информации;

(ii) последовательности, которые включаются, являются неразветвленными последовательностями из четырех или более аминокислот или неразветвленными последовательностями из десяти или более нуклеотидов. Разветвленные последовательности, последовательности с меньшим числом нуклеотидов или аминокислот, чем это определено, а также последовательности, включающие нуклеотиды или аминокислоты, отличные от тех, что приведены в таблицах 1, 2, 3 и 4 Приложения 2, специально исключаются из данного определения;

(iii) “нуклеотиды” охватывают только такие нуклеотиды, которые могут быть представлены с использованием символов, определенных в Таблице 1 Приложения 2. Модификации, например, метилированные основания, могут быть описаны, как это определено в таблице 2 Приложения 2, но они не должны быть показаны в нуклеотидной последовательности в явном (эксплицитном) виде;

(iv) “аминокислотами” являются такие L-аминокислоты, которые обычно встречаются в натуральных белках и перечислены в таблице 3 Приложения 2. Последовательности аминокислот, которые содержат, по крайней мере, одну D-аминокислоту, не предназначены для включения в данное определение. Любая последовательность аминокислот, которая содержит аминокислоты, модифицированные посттрансляционно, могут быть описаны как последовательность аминокислот, которая начально транслирована с использованием символов, приведенных в таблице 3 приложения 2, с модифицированными позициями, например, гидроксигликозилирования или гликоксилирования, как это описано в таблице 4 Приложения 2, но эти модификации не должны быть показаны эксплицитно в последовательности аминокислот. Данным определением охватывается любая пептид или белок, которые могут быть выражены как последовательность символов, приведенных в таблице 3 приложения 2, в сочетании с каким-либо другим описанием, приведенным в другом месте, отражающим, например, ненормальные связи, перекрестные связи (например, десульфидная связь), концевые кэпы, непептидные связи и т.д.;

(v) “идентификатор последовательности” представляет собой уникальное целое число, которое соответствует номеру SEQ ID NO, приписываемому каждой последовательности в перечне;

(vi) “цифровой идентификатор” представляет собой трехзначное число, которое представляет специфический элемент данных;

(vii) “словарь языка нейтральных выражений” представляет собой контролируемый словарь, используемый в перечне последовательностей, который представляет собой научные термины, предписываемые провайдерами баз данных последовательностей (включая научные наименования, определители и их словарно контролируемые значения, символы, появляющиеся в таблицах 1, 2, 3 и 4 Приложения 2, а также ключи характеристик, появляющиеся в таблицах 5 и 6 Приложения 2);



(viii) “компетентный орган” представляет собой Международный поисковый орган, в обязанности которого входит проведение международного поиска по данной международной заявке, или Международный орган предварительной экспертизы, в обязанности которого входит проведение предварительной экспертизы по данной международной заявке, или указанное/выбранное ведомство, в отношении которого была начата обработка данной международной заявки.

ПЕРЕЧЕНЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

3. Перечень последовательностей в том виде, как он определен в параграфе 2(i), в тех случаях, когда он представляется вместе с заявкой, должен помещаться в конце этой заявки. Эта часть должна быть озаглавлена как: “Перечень последовательностей”, начинаться на новой странице и предпочтительно иметь независимую нумерацию страниц. Перечень последовательностей образует самостоятельную часть описания, поэтому согласно параграфу 36 нет необходимости описывать данные последовательности еще где-либо в описании изобретения.

4. В тех случаях, когда перечень последовательностей, определенный в параграфе 2(i), не содержится в заявке при ее подаче, но представляется дополнительно как отдельный документ, после подачи заявки (см. параграф 37), он должен быть озаглавлен как: “Перечень последовательностей” и иметь независимую нумерацию страниц. Первоначальная нумерация последовательностей (см. параграф 5), использованная в заявке при ее подаче, должна сохраняться в перечне последовательностей, предоставляемом позже.

5. Каждой последовательности должен быть присвоен отдельный идентификатор последовательности. Идентификаторы последовательностей должны начинаться с 1 и увеличиваться последовательно на целое число. Если для некоторого идентификатора отсутствует последовательность, код 000 должен появляться под цифровым идентификатором <400>, начиная с новой строки после SEQ ID NO. Цифровому идентификатору <160 > должно соответствовать общее число номеров SEQ ID NO, независимо от того, следуют ли за ним некоторая последовательность или код 000.

6. В описании изобретения, формуле или рисунках заявки последовательности, представленные в перечне, должны иметь отсылку с помощью соответствующего идентификатора последовательности и иметь перед собой “SEQ ID NO:”.

7. Последовательности нуклеотидов и аминокислот должны представляться, по крайней мере, с помощью одной из следующих трех возможностей:

- (i) только последовательностью нуклеотидов;
- (ii) только последовательностью аминокислот;
- (iii) последовательностью нуклеотидов совместно с соответствующей последовательностью аминокислот.

Для последовательностей, представленных в формате, определенном выше согласно альтернативе (iii), последовательность аминокислот должна раскрываться отдельно в перечне последовательностей как отдельная последовательность аминокислот, имеющая отдельный числовой идентификатор последовательности.

ПОСЛЕДОВАТЕЛЬНОСТИ НУКЛЕОТИДОВ

Используемые символы

8. Последовательность нуклеотидов должна быть представлена единой цепочкой слева направо в направлении от 5' -конца к 3' -концу. Термы 3' и 5' не должны быть представлены в последовательности.

9. Основания нуклеотидной последовательности должны представляться с помощью однобуквенного кода для знаков нуклеотидных последовательностей. При этом используются только строчные буквы в соответствии с перечнем, приводимым в таблице 1 Приложения 2.

10. Модифицированные основания должны быть представлены как соответствующие немодифицированные основания или как “n” в самой последовательности, если модифицированное основание является одним из тех, которые перечислены в таблице 2 Приложения 2, а модификация далее должна быть описана в разделе характеристик перечня последовательностей с использованием кодов, приведенных в таблице 2 Приложения 2. Эти коды могут быть использованы в описании или разделе характеристик перечня последовательностей, но не в самой последовательности (см. также параграф 32). Символ “n” является эквивалентом только одного неизвестного или модифицированного нуклеотида.



Используемый формат

11. Последовательность нуклеотидов должна содержать максимум 60 оснований на одну строку с пробелом между каждой группой, состоящей из 10 оснований.
12. Основания последовательности нуклеотидов (включая интроны) должны быть перечислены группами из 10 оснований, за исключением тех, что в кодирующей части последовательности. Оставшиеся основания числом менее 10 в конце некодирующих частей последовательности должны быть сгруппированы вместе и отделены от смежных групп пробелом.
13. Основания кодирующих частей последовательности нуклеотидов должны быть перечислены тройками (кодонами)
14. Нумерация нуклеотидов должна начинаться с первого основания последовательности с номера 1. Она должно быть непрерывной по всей последовательности в направлении от 5' к 3'. Она должна иметь маркировку в правом поле рядом со строкой, содержащей однобуквенные коды для оснований с указанием номера последнего основания в данной строке. Способ нумерации нуклеотидной последовательности, изложенный выше, остается применимым к нуклеотидным последовательностям, имеющим круговую конфигурацию, с тем исключением, что обозначение первого основания нуклеотидной последовательности может быть сделано по выбору заявителя.
15. Последовательность нуклеотидов, состоящая из одного или более несмежных сегментов более длинной последовательности или из сегментов из различных последовательностей, должна нумероваться как отдельная последовательность с отдельным идентификатором последовательности. Последовательность с разрывом или несколькими разрывами должна нумероваться как совокупность отдельных последовательностей, имеющих самостоятельные идентификаторы последовательностей, причем число последовательностей должно равняться числу непрерывных цепочек из элементов последовательностей.

ПОСЛЕДОВАТЕЛЬНОСТИ АМИНОКИСЛОТ

Используемые символы

16. Аминокислоты в белковой или пептидной последовательности должны быть перечислены в направлении от amino к карбоксилу слева направо. Аминогруппы и карбоксильные группы не должны отображаться в последовательности.
17. Аминокислоты должны представляться с помощью трехбуквенных кодов, в которых первая буква является прописной (заглавной) и которые соответствуют списку, приведенному в таблице 3 Приложения 2. Последовательность аминокислот, содержащая символ пробела или внутренней терминации (например, "Ter" или "*" или ".") могут представляться не как единая последовательность аминокислот, а как отдельные последовательности аминокислот (см. также параграф 22).
18. Модифицированные или необычные аминокислоты должны представляться как соответствующие немодифицированные аминокислоты или как "Хаа" в самой последовательности в случае, если модифицированная аминокислота является одной из тех, что перечислены в таблице 4 Приложения 2 и если модификация далее описывается в разделе характеристик перечня последовательностей с использованием кодов, приводимых в таблице 4 Приложения 2. Эти коды могут быть использованы в описании или разделе характеристик перечня последовательностей, но не в самой последовательности (см. также параграф 32). Символ "Хаа" является эквивалентом только одной неизвестной или модифицированной аминокислоты.

Используемый формат

19. Белковая или пептидная последовательность должна иметь максимум 16 аминокислот на одну строку с постановкой символа пробела между каждой аминокислотой
20. Аминокислоты, соответствующие кодонам в кодирующих частях последовательности нуклеотидов, должны помещаться непосредственно под соответствующими кодонами. В тех случаях, когда кодон разделен интроном, символ аминокислоты должен помещаться внизу под частью кодона, содержащего два нуклеотида



21. Нумерация аминокислот должна начинаться с первой аминокислоты последовательности с номера 1. Факультативно аминокислоты, предшествующие зрелому белку, например, в случае наличия пред-последовательностей, про-последовательностей, пред-про-последовательностей и сигнальных последовательностей, могут иметь отрицательные номера с обратным отсчетом, начиная с аминокислоты, ближайшей к номеру 1. Ноль(0) не используется, если нумерация аминокислот использует отрицательные числа с тем , чтобы отличить зрелый белок. Нумерация имеет маркировку под последовательностью для каждых пяти аминокислот. Способ нумерации для последовательностей аминокислот, описанный выше, остается применимым для последовательностей аминокислот, имеющих круговую конфигурацию, за исключением того, что обозначение первой аминокислоты в последовательности может быть выбрано по усмотрению заявителя.

22. Последовательность аминокислот, состоящая из одного или более несмежных сегментов более длинной последовательности или из сегментов разных последовательностей, должна нумероваться как отдельная последовательность, имеющая отдельный идентификатор последовательности. Последовательность, имеющая один или несколько разрывов, должна нумероваться как совокупность отдельных последовательностей, имеющих отдельные идентификаторы последовательности, причем число отдельных последовательностей должно быть равным числу непрерывных цепочек из элементов последовательностей

ДРУГАЯ НАЛИЧНАЯ ИНФОРМАЦИЯ В ПЕРЕЧНЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

23. Порядок следования элементов информации в перечне последовательностей должен соответствовать порядку, в котором эти элементы перечислены в списке цифровых идентификаторов элементов данных, определенных в Приложении 1.

24. Для представления элементов информации в перечне последовательностей должны использоваться только цифровые идентификаторы элементов данных, определенные в Приложении 1. Описательная часть соответствующих цифровых идентификаторов не должна использоваться. Идентифицируемая информация должна следовать непосредственно после цифрового идентификатора, в то время как в перечне последовательностей должны присутствовать только такие цифровые идентификаторы, для которых присутствует соответствующая информация. Исключением этого требования являются два цифровых идентификатора :<220> и <300>, которые соответственно служат заголовками для “Характеристика” и “Информация о публикации” и ассоциируются с информацией соответственно в цифровых идентификаторах с <221> по <223> и с <301> по <313>. Если информация о характеристиках и публикации приводится в перечне последовательностей под соответствующими цифровыми идентификаторами, то соответственно должны включаться цифровые идентификаторы <220> и <300>, но оставаться незаполненными. В целом, незаполненная (пустая) строка должна включаться между цифровыми идентификаторами, если цифра в первой или второй позиции данного цифрового идентификатора меняется. Исключением из этого общего правила является требование, чтобы незаполненная строка не появлялась перед цифровым идентификатором <310>. Дополнительно к этому пустая строка должна предшествовать любому повторяющемуся цифровому идентификатору.

Обязательные элементы данных

25. Перечень последовательностей должен включать в дополнение и непосредственно перед действительными последовательностями нуклеотидов и /или аминокислот следующие элементы информации, определенные в Приложении 1 (обязательные элементы данных):

<110>	Имя заявителя
<120>	Название изобретения
<160>	Номер SEQ ID NO
<210>	SEQ ID NO : x
<211>	Длина
<212>	Тип
<213>	Организм
<400>	Последовательность

В тех случаях, когда имя заявителя (цифровой идентификатор <110>) написано буквами, отличающимися от латинского алфавита, оно должно быть указано также буквами латинского алфавита либо посредством транслитерации, либо посредством перевода на английский язык.



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – ST.25

Страница: 3.25.6

Элементы данных, за исключением тех, которые появляются под цифровыми идентификаторами <110>, <120> и <160>, должны быть повторены для каждой последовательности, включаемой в перечень последовательностей. Если для некоторого идентификатора отсутствует последовательность (см. параграф 5 выше и SEQ ID NO: 4 в примере, описанном в Приложении 3 данного стандарта) , то обязательными являются только элементы данных под цифровыми идентификаторами <210> и <400>.

26. В дополнение к элементам данных, указанным выше в параграфе 25, в том случае, когда перечень последовательностей подается одновременно с заявкой, к которой он относится или в любое другое время до того, как заявке присвоен номер, в перечень последовательностей должен быть включен следующий элемент данных:

<130>	Номер(отсылка к) досье заявки
-------	-------------------------------

27. В дополнение к элементам данных, указанным выше в параграфе 25, в том случае, когда перечень последовательностей представляется в ответ на запрос из компетентного органа или в любое другое время после присвоения заявке номера, в перечень последовательностей включаются следующие элементы:

<140>	текущая патентная заявка
<141>	дата подачи заявки

28. В дополнение к элементам данных, указанным выше в параграфе 25, в том случае, когда перечень последовательностей подается в отношении заявки, испрашивающей приоритет более ранней заявки, в последовательность перечней включаются следующие элементы:

<150>	более ранняя патентная заявка
<151>	дата подачи более ранней заявки

29. Если в последовательности используются “п” или “Хаа», или модифицированное основание, или модифицированная/необычная L-аминокислота, то следующие элементы данных являются обязательными:

<220>	Характеристика
<221>	Наименование/ключ
<222>	Расположение
<223>	Прочая информация

30. Если организм (цифровой идентификатор <213>) является “искусственной последовательностью” или “неизвестной”, то следующие элементы данных являются обязательными:

<220>	Характеристика
<223>	Прочая информация

Факультативные элементы данных

31. Все элементы данных, определенные в Приложении 1 и не упомянутые выше в параграфах с 25 по 30, являются факультативными (факультативные элементы данных).

Представление характеристик

32. Если приводятся характеристики последовательностей (то-есть, цифровой идентификатор 220), то они должны быть описаны с помощью “ключей характеристик”, определенных в таблицах 5 и 6 Приложения 2.⁽¹⁾

⁽¹⁾ Данные таблицы содержат извлечения из Таблиц характеристик (нуклеотидные последовательности)/Общий Банк/DDBJ/EMBL и таблиц характеристик(Последовательности аминокислот) SWISS PROT/



Свободный текст

33. “Свободным текстом” являются формулировки, описывающие характеристики последовательностей под цифровым идентификатором <223> (Прочая информация), в которых не используется нейтральная языковая лексика, как это указано в параграфе 2(vii).

34. Использование свободного текста должно быть ограничено до нескольких коротких терминов, необходимых для понимания данной последовательности. Оно не должно превосходить четырех строк, в каждой строке максимум 65 знаков для каждого присутствующего элемента данных при их представлении на английском языке. Любая прочая информация включается в основную часть описания с использованием тех же формулировок.

35. Любой свободный текст должен быть предпочтительно на английском языке.

36. В тех случаях, когда перечень последовательностей в своей описательной части содержит свободный текст, то этот свободный текст должен быть повторен в основной части описания в том же самом виде. Рекомендуется, чтобы свободный текст в изложении основной части описания изобретения был бы помещен в специальный раздел описания, озаглавленный “свободный текст перечня последовательностей”.

ПЕРЕЧЕНЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ, ПРЕДСТАВЛЯЕМЫЙ ПОЗЖЕ

37. Любой перечень последовательностей, который не содержался в заявке на момент ее подачи, но который представляется позже, не должен выходить за рамки раскрытия содержания заявки на момент ее подачи и должен сопровождаться заявлением на этот счет. Это означает, что перечень последовательностей, представляемый после подачи заявки, должен содержать только такие последовательности, которые были описаны на момент подачи заявки.

38. Любой перечень последовательностей, который не содержался в заявке на момент ее подачи, не входит в состав заявки. Однако согласно положениям Правил РСТ 13ter, 26.3 и 91 и статьи РСТ 91 имеется возможность внесения изменения для перечня последовательностей, включаемого в заявку, которое подается в качестве исправления в соответствии с Правилами РСТ 13ter или 26.3, внесения поправок согласно правила РСТ 91(в случае очевидных ошибок) или изменений согласно статье РСТ 34 , или представления самого перечня в соответствии со статьей РСТ 34 в качестве дополнения к заявке.

МАШИНОЧИТАЕМАЯ ФОРМА ПРЕДСТАВЛЕНИЯ ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

39. Копия перечня последовательностей должна быть также представлена в машиночитаемой (компьютерной) форме в дополнение к перечню последовательностей , включаемому в заявку, если этого требует компетентный орган.

40. Любой перечень последовательностей, представляемый в машиночитаемой форме в дополнение к машинописному перечню последовательностей, должен быть идентичным машинописному перечню и должен сопровождаться заявлением относительно того, что “информация, записанная в компьютерной форме, является идентичной данному машинописному перечню последовательностей”.

41. Полная копия перечня последовательностей, позволяющая ее распечатку, должна размещаться в одном электронном файле предпочтительно на одной дискете или на одном электронном носителе, который принимается данным компетентным органом. Файл, записанный на дискете или ином электронном носителе, приемлемом для данного компетентного органа, должен быть закодирован с использованием кодировочной страницы IBM 437⁽²⁾, кодировочной страницы IBM 932⁽³⁾ или иной совместимой кодировочной страницы. Совместимые кодировочные страницы, согласно требованиям, например, для знаков японского, китайского, арабского языков, кириллического, греческого или еврейского алфавитов, должны быть таковыми, чтобы знакам латинского алфавита и цифрам в них были присвоены те же самые шестнадцатиричные позиции, что и в выше названных кодировочных страницах.

42. Машиночитаемая форма должна предпочтительно создаваться с помощью специализированного программного обеспечения, такого как, PatentIn или другого лицензионного программного продукта; она может создаваться с помощью иных средств при условии, что перечень последовательностей, представляемый на дискете или ином электронном носителе, приемлемом для данного компетентного органа, может быть прочитан операционной системой для ПЭВМ, которая приемлема для компетентного органа.

⁽²⁾ IBM является зарегистрированным товарным знаком International Business Machine Corporation, США.

⁽³⁾ Указанные кодировочные листы являются де факто стандартами для персональных компьютеров.



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – ST.25

Страница: 3.25.8

43. Сжатие (компрессия) файла допустима в случае использования дискеты в качестве носителя при условии, что сжатый файл представлен в саморазвертывающемся формате, обеспечивающем декомпрессию в операционной системе ПЭВМ, используемой компетентным органом.
44. Дискета или иной электронный носитель, приемлемый для компетентного органа, должны иметь ярлык, прикрепленный к нему постоянным образом, на котором напечатано или написано от руки печатными буквами: имя заявителя, название изобретения, ссылочный номер, дата, на которую произведена запись, название операционной системы и наименование компетентного органа.
45. Если дискета или иной электронный носитель, приемлемый для компетентного органа, представляется после даты подачи заявки, то ярлык должен также включать дату подачи и номер заявки.
46. Любое исправление машинописного перечня последовательностей, подаваемое в соответствии с Правилами РСТ 13ter.1(a) (i) или 26.3, любые поправки очевидных ошибок в машинописном перечне последовательностей, представляемые в соответствии с правилом РСТ 91 или любые изменения, которые включает машинописный перечень и который представлен в соответствии со статьей 34, должны сопровождаться машиночитаемой формой перечня последовательностей, содержащего соответствующие исправления поправки или изменения.

[Приложения с 1 по 3 следуют]



ПРИЛОЖЕНИЯ

- Приложение 1: Цифровые идентификаторы
- Приложение 2: Символы нуклеотидов и аминокислот и Таблицы характеристик
- Таблица 1: Перечень нуклеотидов
- Таблица 2: Перечень модифицированных нуклеотидов
- Таблица 3: Перечень аминокислот
- Таблица 4: Перечень модифицированных и необычных аминокислот
- Таблица 5: Перечень ключей характеристик, относящихся к последовательностям нуклеотидов
- Таблица 6: Перечень ключей характеристик, относящихся к последовательностям белков
- Приложение 3: Пример перечня последовательностей



ПРИЛОЖЕНИЕ 1

ЦИФРОВЫЕ ИДЕНТИФИКАТОРЫ

Только цифровые идентификаторы, определенные ниже, могут использоваться в перечнях последовательностей, представляемых в заявках. Формулировки заголовков элементов данных, приведенных ниже, не должны включаться в перечни последовательностей.

Цифровые идентификаторы обязательных элементов данных, т.е., элементов данных, которые должны обязательно включаться во все перечни последовательностей (см. параграф 25 данного стандарта: элементы 110, 120, 160, 210, 211, 212, 213 и 400), и цифровые идентификаторы элементов данных, которые должны включаться в специальных случаях, определенных в данном стандарте (см. параграфы 26, 27, 28, 29 и 30 данного стандарта: элементы 130, 140, 141, 150 и 151, а также 220 по 223) помечены символом "М" I

Цифровые идентификаторы факультативных элементов данных (см. параграф 31 данного стандарта) помечены символом "О".

Цифровой идентификатор	Описание цифрового идентификатора	Обязательный (М) Необязательный (О)	Комментарии
<110>	имя заявителя	М	В тех случаях, когда имя заявителя написано буквами иного алфавита, чем латинского, оно должно быть также указано буквами латинского алфавита либо посредством транслитерации, либо посредством перевода на английский язык
<120>	название изобретения	М	
<130>	ссылка на досье (дело заявки)	М, в случаях, указанных в пар. 26 данного Стандарта	См. параграф 26 данного стандарта
<140>	текущая патентная заявка	М, в случаях, указанных в пар. 27 данного стандарта	См. пар. 27 данного стандарта; текущая патентная заявка должна быть идентифицирована в следующем порядке: двубуквенным кодом страны согласно Стандарту ВОИС ST.3 и номером заявки (в формате, используемом данным ведомством по промышленной собственности, в которое подается данная текущая заявка) или для международной заявки с помощью номера международной заявки
<141>	дата подачи текущей заявки	М, в случаях, указанных в пар.27 данного стандарта	См. пар.27 данного стандарта; дата должна быть указана согласно стандарта ВОИС ST.2 (CCYY MM DD)



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – ST.25

Страница: 3.25.11

Приложение 1, страница 2

Цифровой идентификатор	Описание цифрового идентификатора	Обязательный (М) Необязательный (О)	Комментарии
<150>	более ранняя патентная заявка	М, в случаях, указанных в пар. 28 данного стандарта	См. пар. 28 данного стандарта; более ранняя заявка должна быть идентифицирована в следующем порядке: с помощью двухбуквенного кода, указанного согласно стандарту ВОИС ST.3 , и номера заявки(в формате, используемом тем ведомством по промышленной собственности, в которое была подана более ранняя заявка) или в случае международной заявки с помощью номера международной заявки
<151>	дата подачи более ранней заявки	М, в случаях, указанных в пар. 28 данного стандарта	См. пар. 28 данного стандарта; дата должна быть указана в соответствии со стандартом ВОИС ST.2 (CCYY MM DD)
<160>	номер SEQ ID NO	М	
<170>	программное обеспечение	О	
<210>	информация в отношении SEQ ID NO: x	М	соответствует целому числу, представляющему данный номер SEQ ID NO
<211>	длина	М	длина последовательности выражается числом пар оснований или аминокислот
<212>	тип	М	тип молекулы, находящейся в последовательности SEQ ID NO: x, либо ДНК, либо РНК, либо ПРТ(PRT); если последовательность нуклеотидов содержит фрагменты и ДНК и РНК, принимаемое значение должно быть ДНК, и дополнительно комбинированная молекула ДНК/РНК должна быть описана в разделе характеристик с <220> по< 223>
<213>	организм	М	родо-видовое (то-есть, научное наименование) или "Искусственная последовательность" или "неизвестное"
<220>	характеристика	М, в случаях, указанных в пар. 29 и 30 данного стандарта	оставляется незаполненным; см. Пар. 29 и 30 данного стандарта; описание точек биологической значимости в последовательности SEQ ID NO:x; (может быть повторено в зависимости от числа указанных характеристик)
<221>	имя/ключ	М, в случаях, указанных в пар. 29 данного стандарта	см. пар. 29 данного стандарта; используются только те ключи, которые описаны в таблице 5 или 6 Приложения 2



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – ST.25

Страница: 3.25.12

Приложение 1, страница 3

Цифровой идентификатор	Описание цифрового идентификатора	Обязательный (М) Необязательный (О)	Комментарии
<222>	местоположение	М, в случаях, указанных в пар. 29 данного стандарта	см. пар. 29 данного стандарта; - с (номера первого основания/аминокислоты в характеристике) - по(номер последнего основания/аминокислоты в характеристике) - пары оснований (номера относятся к позициям пар оснований в последовательности нуклеотидов) -аминокислоты(номера соотносятся с позициями остаточных аминокислот в последовательности аминокислот) - расположение характеристики в дополнительной цепочке относительно той, которая приведена в перечне последовательностей
<223>	прочая информация	М, в случаях, указанных в пар. 29 и 30 данного стандарта	см. пар. 29 и 30 данного стандарта; любая другая информация с использованием терминов словаря нейтрального языка или свободный текст(предпочтительно на английском языке); всякий свободный текст должен быть повторен в основной части описания с теми же самыми формулировками(см. пар. 36 данного стандарта); в тех случаях, когда имеется какое-либо модифицированное основание или модифицированная/необычная L-аминокислота, содержащаяся в таблицах 2 и 4 Приложения 2, должен использоваться символ, ассоциированный с данным основанием или аминокислотой из таблиц 2 и 4 Приложения 2
<300>	информация о публикации	О	оставить незаполненным; повторить раздел в отношении каждой публикации
<301>	авторы	О	
<302>	название	О	название публикации
<303>	Журнал	О	название журнала, в котором опубликованы данные
<304>	том	О	том журнала, в котором опубликованы данные
<305>	выпуск	О	выпуск журнала, в котором опубликованы данные
<306>	страницы	О	номера страниц, на которых приведены данные
<307>	дата	О	дата издания журнала, в котором опубликованы данные; если возможно, дата указывается в соответствии со стандартом ВОИС ST.2 (CCYY MM DD)



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – ST.25

Страница: 3.25.13

Приложение 1, страница 4

Цифровой идентификатор	Описание цифрового идентификатора	Обязательный (М) Необязательный (О)	Комментарии
<308>	номер доступа в базе данных	О	номер доступа, присвоенный базой данных, включая название базы данных
<309>	дата ввода в базу данных	О	дата ввода в базу данных; дата должна быть указана в соответствии со стандартом ВОИС ST.2 (CCYY MM DD)
<310>	номер документа	О	номер документа, только для ссылок на патенты; полная идентификация документа делается в следующем порядке: двубуквенный код, указываемый в соответствии со стандартом ВОИС ST.3 , номер публикации, указываемый в соответствии со стандартом ST.6 , и код вида документа, указываемый в соответствии со стандартом ST.16
<311>	дата подачи	О	дата подачи документа, только для ссылок на патент; дата должна быть указана в соответствии со стандартом ST.2 (CCYY MM DD)
<312 >	дата публикации	О	дата подачи документа, только для ссылок на патент; дата должна быть указана в соответствии со стандартом ST.2 (CCYY MM DD)
<313>	релевантные остатки в SEQ ID NO:х: с по	О	
<400>	последовательность	М	SEQ ID NO: х должен следовать за цифровым идентификатором и должен помещаться на строке, предшествующей данной последовательности (см. Приложение 3)

[Приложение 2 следует]



ПРИЛОЖЕНИЕ 2

ТАБЛИЦЫ СИМВОЛОВ НУКЛЕОТИДОВ И АМИНОКИСЛОТ, А ТАКЖЕ ХАРАКТЕРИСТИК

Таблица 1: Перечень нуклеотидов

Символ	Значение	Происхождение обозначения символа
a	a	<u>a</u> denine (аденин)
g	g	<u>g</u> uanine (гуанин)
c	c	<u>c</u> ytosine (цитозин)
t	t	<u>t</u> hymine (тимин)
u	u	<u>u</u> racil (урацил)
г	g или a	pu <u>r</u> ine (пурин)
у	t/u или c	py <u>r</u> imidine (пиримидин)
m	a или c	a <u>m</u> ino (амин)
k	g или t/u	<u>k</u> eto (кето)
s	g или c	<u>s</u> trong (сильные взаимодействия, связи 3 Н)
w	a или t/u	<u>w</u> eak (слабые взаимодействия, связи 2Н)
b	g или c или t/c	not a (не a)
d	a или g или t/u	not c (не c)
h	a или c или t/u	not g (не g)
v	a или g или c	not t, not u (не t , не u)
n	a или g или c или t/u, неизвестное или прочее	<u>a</u> ny (любое)



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – СТ.25

Страница: 3.25.15

Приложение 2, страница 2

Таблица 2: Перечень модифицированных нуклеотидов

Символ	Значение
ac4c	4- ацетилцитидин (4-acetylcytidine)
chm5u	5-(карбоксихидроксиметил)уридин 5-(carboxyhydroxymethyl)uridine
cm	2'- O- метилцитидин (2'-O-methylcytidine)
cmnm5s2u	5-карбоксиметиламинометил-2-тиоуридин (5-carboxymethylaminomethyl-2-thiouridine)
cmnm5u	5-карбоксиметиламинометилуридин (5-carboxymethylaminomethyluridine)
d	дигидроуридин (dihydrouridine)
fm	2'-О-метилпсевдоуридин (2'-O-methylpseudouridine)
gal q	бета, D-галактозилквеозин (beta, D-galactosylqueosine)
gm	2-О- метилгуанозин (2'-O-methylguanosine)
i	инозин (inosine)
i6a	N6-изопениладенозин (N6-isopenyladenosine)
m1a	1- метиладенозин (1-methyladenosine)
m1f	1-метилпсевдоуридин (1-methylpseudouridine)
m1g	1-метилгуанозин (1-methylguanosine)
m1i	1-метиинозин (1-methylinosine)
m22g	2,2-диметилгуанозин (2,2-dimethylguanosine)
m2a	2-метиладенозин (2-methyladenosine)
m2g	2-метилгуанозин (2-methylguanosine)
m3c	3-метилцитидин (3-methylcytidine)
m5c	5-метилцитидин (5-methylcytidine)
m6a	N6- метиладенозин (N6-methyladenosine)
m7g	7-метилгуанозин (7-methylguanosine)
mam5u	5-метиламинометилуридин (5-methylaminomethyluridine)
mam5s2u	5-метоксиаминометил-2-тиоуридин (5-methoxyaminomethyl-2-thiouridine)
man q	бета, D-манозилквеозин (beta, D-mannosylqueosine)
mcm5s2u	5-метоксикарбонилметил-2-тиоуридин (5-methoxycarbonylmethyl-2-thiouridine)
mcm5u	5-метоксикарбонилметилуридин (5-methoxycarbonylmethyluridine)
mo5u	5-метоксиуридин (5-methoxyuridine)
ms2i6a	2-метилтио-N6-изопентиладенозин (2-methylthio-N6-isopentyladenosine)
ms2t6a	N-((9-бета-D-рибофуранозил-2-метилтиопурин-6-)карбомойл)треонин N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoil)threonine
mt6a	N-((9-бета-D-рибофуранозилпурин-6)метилкарбомойл)треонин N-((9-beta-D-ribofuranosylpurine-6-yl)-methylcarbamoil)threonine
mv	уридин-5-оксоуксусная кислота-метилэфир (uridine-5-oxoacetic acid-methylester)
o5u	уридин-5-оксиуксусная кислота (uridine-5-oxoacetic acid)
osyw	вибутоксозин (wybutoxosine)
p	псевдоуридин (pseudouridine)
q	квеозин (queosine)
s2c	2-тиоцитидин (2-thiocytidine)
s2t	5- метил-2-тиоуридин (5-methyl-2-thiouridine)
s2u	2-тиоуридин (2-thiouridine)
s4u	4- тиоуридин (4-thiouridine)
t	5- метилуридин (5-methyluridine)
t6a	N-((9-бета-D-рибофуранозилпурин-6)карбомойл)треонин N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoil)threonine
tm	2'-О-метил-5-метилуридин (2'-O-methyl-5-methyluridine)
um	2'-О-метилуридин (2'-O-methyluridine)
ym	вибутозин (wybutosine)
x	3-(3-амино-3-карбокси-пропил)уридин, (acp3)u / 3-(3-amino-3-carboxy-propyl)uridine,(acp3)u



Приложение 2, страница 3

Таблица 3: Перечень аминокислот

Символ	Значение
Ala	Аланин (Alanine)
Cys	Цистеин (Cysteine)
Asp	Аспарагиновая кислота (Aspartic Acid)
Glu	Глутаминовая кислота (Glutamic Acid)
Phe	Фенилаланин (Phenylalanine)
Gly	Глицин (Glycine)
His	Гистидин (Histidine)
Ile	Изолейцин (Isoleucine)
Lys	Лизин (Lysine)
Leu	Лейцин (Leucine)
Met	Метионин (Methionine)
Asn	Аспарагин (Asparagine)
Pro	Пролин (Proline)
Gln	Глутамин (Glutamine)
Arg	Аргенин (Arginine)
Ser	Серин (Serine)
Thr	Треонин (Threonine)
Val	Валин (Valine)
Trp	Триптофан (Tryptophan)
Tyr	Тирозин (Tyrosine)
Asx	Asp или Asn
Glx	Glu или Gln
Xaa	неизвестная или прочая



Приложение 2, страница 4

Таблица 4: Перечень модифицированных и необычных аминокислот

Символ	Значение
Aad	2-аминоадипиновая кислота (2-Aminoadipic acid)
bAad	3-аминоадипиновая кислота (3-Aminoadipic acid)
bAla	бета-аланин, бета-аминопропионовая кислота (beta-Alanine, beta-Aminopropionic acid)
Abu	2-аминобутириновая кислота (2-Aminobutyric acid)
4Abu	4-аминобутириновая кислота, пиперидиновая кислота (4-Aminobutyric acid, piperidinic acid)
Acр	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-Diaminopropionic 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)
alle	Алло-изолейцин (allo-Isoleucine)
MeGly	N-метилглицин, саркозин (N-Methylglycine, sarcosine)
Melle	N-метилизолойцин (N-Methylisoleucine)
MeLys	6-N-метиллизин (6-N-Methyllysine)
MeVal	N-метилвалин (N-Methylvaline)
Nva	Норвалин (Norvaline)
Nle	Норлейцин (Norleucine)
Orn	Орнитин (Ornithine)



Приложение 2, страница 5

Таблица 5: Перечень ключей характеристик, относящихся к последовательностям нуклеотидов

Ключ	Описание
allele (аллель)	Родственный индивид или штамм содержит устойчивые, чередующиеся формы того же самого гена, который отличается от представленной последовательности в данном местоположении (а возможно и в других)
attenuator (ослабитель)	(1) область ДНК, в которой происходит регуляция окончания транскрипции, которая управляет экспрессией некоторых бактериальных оперонов; (2) сегмент последовательности, расположенный между промотором и первым структурным геном, что вызывает частичную терминацию транскрипции
C-region (C-область, константная)	константная область легких и тяжелых цепей иммуноглобулина, а также цепей типа альфа рецепторов Т-клеток, бета и гамма; включает один или более экзонов в зависимости от вида цепи
CAAT –signal (сигнал типа CAAT)	CAAT- блок; часть консервативной последовательности, расположенной на расстоянии примерно в 75 bp(пар оснований) выше от начальной точки эукариотных единиц транскрипции, которые могут быть вовлечены в полимеразное связывание РНК; консенсус = GG (C или T) CAATCT
CDS	кодирующая последовательность; последовательность нуклеотидов, которая соответствует последовательности аминокислот в белке (местоположение включает терминирующий кодон); характеристика включает концептуальную трансляцию аминокислот
conflict (конфликт)	независимые детерминации “одной и той же” цепи различаются в данном участке(сайте) или области
D-loop (D-петля)	петля замещения; область в пределах митохондриального ДНК, в которой короткий фрагмент (участок) РНК спаривается с одной цепью ДНК, замещая цепь исходного партнера в этой области; используется также для описания замещения участка одной цепи дуплексной ДНК на одноцепочечный фрагмент в реакции, вызываемой RecA белком
D-segment (D-сегмент)	D-сегмент тяжелой иммуноглобулиновой цепи, а также бета цепи рецепторов Т-клеток
enhancer (усилитель)	цис-активная последовательность, которая увеличивает использование (некоторых) эукариотных промоторов и может действовать в двух направлениях и в любом местоположении (влево, вправо) относительно промотора
exon (экзон)	участок генома, который кодирует часть сплайсированного мРНК; может содержать 5'UTR, все CDS и 3'UTR
GC-signal (GC-сигнал)	GC-блок; консервативная область, богатая-GC, расположенная выше(слева) от начальной точки эукариотических единиц транскрипции, которая может происходить множественным способом или в обоих направлениях; консенсус = GGGCGG
gene (ген)	область биологического интереса, идентифицируемая как ген и для которой присваивается имя
iDNA (iДНК)	промежуточная ДНК; ДНК, которая удаляется посредством любого из нескольких видов рекомбинации
Intron (интрон)	Сегмент ДНК, который считан (транскриптирован), но удален из транскрипта путем сплайта вместе с последовательностями (экзонами) по обе стороны от него
J-segment (J-сегмент)	J-сегмент(соединения) легких и тяжелых иммуноглобулиновых цепей, а также цепей типа бета, гамма и альфа рецепторов Т-клеток
LTR	Длинный концевой повтор (LTR-последовательность), последовательность, непосредственно повторяемая на обоих концах определенной последовательности, обычно обнаруживаемой в ретровирусах
mat_peptide	Кодирующая последовательность зрелого пептида или белка; кодирующая последовательность для зрелого или конечного пептида или белкового продукта в результате пост-трансляционной модификации; местоположение не включает кодона терминации (в отличие от соответствующего CDS)



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – ST.25

Страница: 3.25.19

Приложение 2, страница 6

Ключ	Описание
misc_binding	Участок(сайт) в нуклеиновой кислоте, который связывает ковалентно или нековалентно с другим участком, что не может быть описано ни каким другим ключом связывания (primer_bind или protein_bind)
misc_difference	Характерная последовательность отличается от той, которая представлена в данной словарной статье и не может быть описана ни каким другим ключом различия (conflict, unsure, old_sequence,mutation,variation, allele или modified_base)
misc_feature	Область биологического интереса, которая не может быть описана никаким другим ключом; новая или редкая характеристика
misc_recomb	Место любого обобщенного, характерного для данного участка события или события репликативной рекомбинации, в котором происходит разрыв или воссоединение дуплексных ДНК, которые не могут быть описаны другими ключами рекомбинации(idNA и virion) или определителями исходных ключей (/insertion_seq, /transposon, /proviral)
misc_RNA	любой транскрипт или продукт РНК, который не может быть определен другими ключами РНК (prim_transcript, precursor_RNA, mRNA, 5'clip, 3'clip,5'UTR,3'UTR, exon, CDS, sig_peptide,transit_peptide, mat_peptide, intron, polyA_site,rRNA, tRNA, scRNA, и snRNA)
misc_signal	любая область, содержащая сигнал управления или изменения генной функции или экспрессии, которая не может быть описана другими ключами сигналов (promoter, CAAT_signal, TATA_signal, --35_signal, -10_signal,GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator и rep_origin)
misc_structure	любая вторичная или третичная структура или конформация, которая не может быть описана другими ключами Структуры(stem_loop и D-loop)
modified_base	Указанный нуклеотид является модифицированным нуклеотидом и должен быть заменен на указанную молекулу (представленную в значении определителя mod_base)
mRNA	Мессенджер(посредник) РНК; включает 5' нетранслированную область(5'UTR), кодирующие последовательности (CDS,exon), и 3' нетранслированную область (3' UTR)
mutation	Родственный штамм имеет внезапное, ненаследованное изменение в последовательности в данном месте
N_region	Дополнительные нуклеотиды, вставленные между переупорядоченными сегментами иммуноглобулина
old_sequence	Представляемая последовательность является возвратом к предыдущей версии последовательности в данном месте
polyA_signal	Область распознавания, необходимая для эндонуклеазного расщепления транскрипта РНК, за которым следует полиаденилирование; консенсус = AATAAA
polyA_site	Участок (сайт) в транскрипте РНК, к которому добавляются адениновые остатки путем посттранскрипционного полиаденилирования
precursor_RNA	Любой вид РНК, который не является зрелым продуктом РНК; может включать 5' усеченную область (5' clip), 5' нетранслированную область(5'UTR), кодирующие последовательности(CDS, exon), 3' нетранслированную область(3'UTR) и 3' усеченную область (3'clip)
prim_transcript	Первичный (начальный, необработанный) транскрипт; включает 5' усеченную область(5'clip), 5' нетранслированную область(5'UTR), кодирующие последовательности(CDS,exon), вставочные последовательности (intron), 3' нетранслированную область(3'UTR) и 3' усеченную область(3'clip)
primer_bind	Сайт нековалентного связывания праймера для инициации репликации, транскрипции или обратной транскрипции; включает сайт(ы) для синтетических элементов PCR -праймера
promoter	Область молекулы ДНК, вовлеченная в полимеразное связывание РНК для инициации транскрипции
protein_bind	Участок(сайт) нековалентного связывания белка в нуклеиновой кислоте
RBS	Сайт рибосомного связывания
repeat_region	Область генома, содержащая повторяющиеся единицы
repeat_unit	Одиночный элемент повтора



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – ST.25

Страница: 3.25.20

Приложение 2, страница 7

Ключ	Описание
rep_origin	Начало репликации; сайт начала дублирования нуклеиновой кислоты для получения двух идентичных копий
rRNA	Зрелая рибосомная РНК; компонент РНК рибонуклеопротеиновой частицы (рибосомы), которая собирает аминокислоты в белки
S_region	Область переключения тяжелых цепей иммуноглобулина; вовлеченная в перестройку тяжелых цепей ДНК, приводящей к экспрессии иммуноглобулина отличающегося класса из той же самой В-клетки
satellite	Несколько tandemных повторов (идентичных или родственных) из короткой базовой повторяющейся единицы; многие имеют состав основания или другие свойства, отличающиеся от средней генома, что позволяет им отделяться от основной массы (основного диска) геномной ДНК
scRNA	Низкомолекулярная цитоплазмическая РНК; любая из нескольких низкомолекулярных цитоплазмических молекул РНК, присутствующих в цитоплазме, а (иногда) и в ядре эукариота
sig_peptide	Кодирующая последовательность сигнального пептида; кодирующая последовательность для N-концевого домена выделенного белка; этот домен вовлекается в присоединение рождающегося полипептида к мембране; лидерная последовательность
snPNA	Низкомолекулярная ядерная РНК; любая из нескольких видов низкомолекулярных РНК, ограниченных ядром; несколько snRNA, вовлеченных в сплайсинг или другую реакцию процессинга РНК
source	Идентифицирует биологический источник специфицированного участка последовательности; данный ключ является обязательным; каждая запись должна иметь, как минимум, единый ключ источника, охватывающий всю последовательность; на одну последовательность допускается более, чем один ключ
stem_loop	Петля типа "шпилька"; область двойной спирали, образованная спариванием оснований между смежными (инвертированными) комплементарными последовательностями в единую цепь РНК или ДНК
STS	Размеченный сайт(участок) последовательности; короткая единственная копия последовательности ДНК, которая характеризует картирование генома и может быть обнаружена с помощью PCR; область генома может быть картирована путем определения порядка серий STS
TATA_signal	ТАТА блок; блок Гольдберга-Хогнесса; консервативный септамер, богатый АТ, находящийся на расстоянии 25 bp до начальной точки каждой единицы транскрипта эукариотной РНК полимеразы II, которая может быть вовлечена в положенин энзима для правильного инициирования; консенсус= TATA (А илиТ)А(А илиТ)
terminator	последовательность ДНК, расположенная в конце транскрипта или рядом с областью промотера, которая заставляет РНК полимеразу закончить транскрипцию; может быть также сайтом связывания репрессорного белка
transit_peptide	Кодирующая последовательность транзитного пептида; кодирующая последовательность для N-концевого домена ядерно-кодированного органоидного белка; этот домен вовлекается в посттрансляционный перенос белка в органеллу
tRNA	Зрелая транспортная РНК, небольшая молекула РНК (длиной 75-85 оснований), которая содействует трансляции последовательности нуклеиновой кислоты в последовательность аминокислот
unsure	Автор не уверен в точной последовательности в данной области
V_region	Вариабельная область легких и тяжелых цепей иммуноглобулина, а также цепей типа бета, гамма и альфа рецепторов Т-клеток; коды для вариабельной аминокислотной терминальной части; может состоять из V_ сегментов, D_ сегментов, N_ областей и J_ сегментов
V_segment	Вариабельный сегмент легких и тяжелых цепей иммуноглобулина, а также цепей типа бета, гамма и альфа рецепторов Т-клеток; коды для многих вариабельных областей и последних нескольких аминокислот лидерных пептидов



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – СТ.25

Страница: 3.25.21

Приложение 2, страница 8

Ключ	Описание
variation	Родственный штамм содержит стабильные мутации из одного и того же гена (например, RFLP (полиморфизм длины рестрикционных фрагментов), другие полиморфизмы и т.д.), которые отличаются от представленной последовательности в данном месте (и возможно в других)
3'clip	3'- область транскрипта- предшественника, которая отсекается во время процессинга
3'UTR	Область рядом с 3' концом зрелого транскрипта (следующего за кодоном терминации), который не транслируется в белок
5'clip	5'-область транскрипта - предшественника, которая отсекается во время процессинга
5'UTR	область рядом с 5' концом зрелого транскрипта (предшествующего кодону инициации), который не транслируется в белок
-10_signal	блок Прибнова; консервативная область на расстоянии 10 bp выше начальной точки единиц бактериальной транскрипции, которые могут быть вовлечены в полимеразу связывания РНК; консенсус = TAtAaT
-35_signal	консервативный гексамер на расстоянии 35 bp выше начальной точки единиц бактериальной транскрипции ; консенсус = TTGACa [] или TGTTGACA[]



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – ST.25

Страница: 3.25.22

Приложение 2, страница 9

Таблица 6: Перечень ключей характеристик, относящихся к последовательностям белков

Ключ	Описание
CONFLICT	различные статьи сообщают о различных последовательностях
VARIANT	авторы сообщают о том, что существуют варианты последовательности
VARSPPLIC	описание вариантов последовательности, полученных альтернативным сплайсингом
MUTAGEN	сайт, который был изменен экспериментально
MOD_RES	пост-трансляционная модификация остатка
ACETYLYATION	N-терминального (концевого) или иного
AMIDATION	в целом на С-конце зрелого активного пептида
BLOCKED	неопределенной группы блокирования на N- или С-конце
FORMYLATION	N-концевого метионина
GAMMACARBOXYGLUTAMIC ACID HYDROXYLATION	аспарагина, аспарагиновой кислоты, пролина или лизина
METHYLATION	в целом лизина или аргинина
PHOSPORYLATION	серина, треонина, тирозина, аспарагиновой кислоты или гистидина
PYROLIDONE CARBOXYLIC ACID	N-концевого глутамата, который образует внутренний циклический лактам
SULFATATION	в целом тирозина
LIPID	ковалентное связывание липоидной части
MYRISTATE	миристинной группы, присоединенной через амидную связь к N-концевому остатку глицина зрелой формы белка или к внутреннему остатку лизина
PALMITATE	пальмитатной группы, присоединенной через тиоэфирную связь к остатку цистеина или через эфирную связь к остатку серина или треонина
FARNESYL	фарнесильной группы, присоединенной через тиоэфирную связь к остатку цистеина
GERANYL-GERANYL	геранил-гераниловой группы, присоединенной через тиоэфирную связь к остатку цистеина
GPI-ANCHOR	гликосил-фосфатидилинозитоловой (GPI) группы, соединенной с альфа-карбоксылльной группой С-концевого остатка зрелой формы белка
N-ACYL DIGLYCERIDE	N-концевого цистеина зрелой формы прокариотического липопротеина с амидосвязанной жирной кислотой и глицериновой группой, к которой присоединены две жирные кислоты с помощью эфирной связи
DISULFID	дисульфидная связь; концевые точки "ОТ" и "ДО" представляют собой два остатка, которые связаны дисульфидной связью между цепями; если концевые точки "ОТ" и "ДО" являются идентичными и существует, дисульфидная связь между цепями, и данное поле описания указывает на природу поперечной сшивки
THIOLEST	тиоэфирная связь; концевые точки "ОТ" и "ДО" представляют два остатка, которые связаны тиоэфирной связью
THIOETH	тиоэфирная связь; концевые точки "ОТ" и "ДО" представляют два остатка, которые связаны тиоэфирной связью
CARBOHYD	сайт(участок) гликозилирования; в поле описания указывается природа углевода(если известна)



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – ST.25

Страница: 3.25.23

Приложение 2, страница 10

Ключ	Описание
METAL	сайт(участок) связывания иона металла; в поле описания указывается природа металла
BINDING	сайт связывания любой химической группы (ко-энзимов, простетической группы и т.д.); в поле описания указывается химическая природа группы
SIGNAL	протяженность сигнальной последовательности (препептида)
TRANSIT	протяженность транзитного пептида (митохондриального, хлорпластического или для микротельца)
PROPEP	протяженность пропептида
CHAIN	протяженность полипептидной цепи в зрелом белке
PEPTIDE	протяженность выделенного активного пептида
DOMAIN	протяженность интересующего домена в цепи; в поле описания указывается природа данного домена
CA_BIND	протяженность области связывания кальция
DNA_BIND	протяженность области связывания ДНК
NP_BIND	протяженность области связывания нуклеотидфосфата; в поле описания указывается природа нуклеотидфосфата
TRANSMEM	протяженность трансмембранной области
ZN_FING	протяженность области цинк-фингера
SIMILAR	величина сходства с другой последовательностью белка; в поле описания дается точная информация относительно данной последовательности
REPEAT	величина повтора внутренней последовательности
HELIX	вторичная структура: спирали, например, альфа-спираль, 3(10) спираль или Pi- спираль
STRAND	вторичная структура: бета-цепь, например, бета-цепь с водородной связью или остаток в изолированном бета-мосте
TURN	витки вторичной структуры, например, виток с Н-связью (3-виток, 4-виток или 5-виток)
ACT_SITE	аминокислота(ы), вовлеченная(ые) в ферментативную активность
SITE	любой другой интересующий сайт (участок) в последовательности
INIT_MET	известно, что данная последовательность начинается с метионина в качестве инициатора
NON_TER	остаток в конце последовательности не является концевым (терминальным) остатком; если это относится к позиции 1, то означает, что первая позиция не является N -концом всей молекулы; если это относится к последней позиции, то означает, что данная позиция не является С-концом всей молекулы; для данного ключа отсутствует поле описания
NON_CONS	непоследовательные остатки; указывает на то, что два остатка в последовательности являются не последовательными и что между ними находится ряд остатков вне последовательности
UNSURE	неопределенности в последовательности; используется для описания области(ей) последовательности, в отношении которой(ых) авторы не уверены в точности определения последовательности

[Приложение 3 следует]



ПРИЛОЖЕНИЕ 3

ПРИМЕР ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

<110> Smith, John; Smithgene Inc.

<120> Example of a Sequence Listing

<130> 01-00001

<140> PCT/EP98/00001
<141> 1998-12-31

<150> US 08/999,999
<151> 1997-10-15

<160> 4

<170> PatentIn version 2.0

<210> 1
<211> 389
<212> DNA
<213> Paramecium sp.

<220>
<221> CDS
<222> (279)...(389)

<300>
<301> Doe, Richard
<302> Isolation and Characterization of a Gene Encoding a Protease
from Paramecium sp.
<303> Journal of Genes
<304> 1
<305> 4
<306> 1-7
<307> 1988-06-31
<308> 123456
<309> 1988-06-31

<400> 1
agctgtagtc attcctgtgt cctctttctct ctgggcttct caccctgcta atcagatctc 60
agggagagtg tcttgaccct cctctgcctt tgcagcttca caggcaggca ggcaggcagc 120
tgatgtggca attgctggca gtgccacagg cttttcagcc aggcttaggg tgggttccgc 180
cgcggcgcgg cggccctct cgcgctctc tcgcgctct ctctcgtct cctctcgtc 240



РУКОВОДСТВО ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЛАСТИ ПРОМЫШЛЕННОЙ СОБСТВЕННОСТИ

Стандарты – СТ.25

Страница: 3.25.25

Приложение 3, страница 3

```
ggacctgatt  aggtgagcag  gaggaggggg  cagttagc      atg  gtt  tca  atg  ttc  agc  296
                Met  Val  Ser  Met  Phe  Ser
                1          5

ttg  tct  ttc  aaa  tgg  cct  gga  ttt  tgt  ttg  ttt  gtt  tgt  ttg  ttc  caa  344
Leu  Ser  Phe  Lys  Trp  Pro  Gly  Phe  Cys  Leu  Phe  Val  Cys  Leu  Phe  Gln
                10          15          20

tgt  ccc  aaa  gtc  ctc  ccc  tgt  cac  tca  tca  ctg  cag  ccg  aat  ctt  389
Cys  Pro  Lys  Val  Leu  Pro  Cys  His  Ser  Ser  Leu  Gln  Pro  Asn  Leu
                25          30          35

<210>          2
<211>          37
<212>          PRT
<213>          Paramecium sp.

<400>          2
Met  Val  Ser  Met  Phe  Ser  Leu  Ser  Phe  Lys  Trp  Pro  Gly  Phe  Cys  Leu
  1          5          10          15

Phe  Val  Cys  Leu  Phe  Gln  Cys  Pro  Lys  Val  Leu  Pro  Cys  His  Ser  Ser
          20          25          30

Leu  Gln  Pro  Asn  Leu
          35

<210>          3
<211>          11
<212>          PRT
<213>          Artificial Sequence

<220>
<223>          Designed peptide based on size and polarity to act as a
                linker between the alpha and beta chains of Protein XYZ.

<400>          3
Met  Val  Asn  Leu  Glu  Pro  Met  His  Thr  Glu  Ile
  1          5          10

<210>          4
<400>          4
000
```

[Конец Приложения 3 и Стандарта]