СТАНДАРТ ST.26

РЕКОМЕНДОВАНИЙ СТАНДАРТ З ПРЕДСТАВЛЕННЯ ПЕРЕЛІКІВ НУКЛЕОТИДНИХ ТА АМІНОКИСЛОТНИХ ПОСЛІДОВНОСТЕЙ З ВИКОРИСТАННЯМ XML (РОЗШИРЮВАНОЇ МОВИ РОЗМІТКИ)

Версія 1.6

Редакція затверджена Комітетом зі стандартів ВОІВ(КСВ) на його десятій сесії 25 листопада 2022 року

Редакційні зауваження підготовлені Міжнародним бюро

Десята сесія Комітету зі стандартів ВОІВ вирішила, що версія 1.6 стандарту ВОІВ ST.26 набуде чинності 1 липня 2023 року. Тим часом має використовуватися версія 1.5 стандарту ВОІВ ST.26.

Стандарт опубліковано з метою інформування відомств інтелектуальної власності та інших заінтересованих осіб.

ЗМІСТ

ВСТУП 3
СФЕРА ЗАСТОСУВАННЯ 5
ПОСИЛАННЯ 6
ПРЕДСТАВЛЕННЯ ПОСЛІДОВНОСТЕЙ 7
Нуклеотидні послідовності 7
Амінокислотні послідовності 11
Представлення окремих випадків 13
СТРУКТУРА ПЕРЕЛІКУ ПОСЛІДОВНОСТЕЙ В XML 13
Кореневий елемент 15
Частина із загальною інформацією 16
Частина з даними щодо послідовності 20
Таблиця функцій 22
Ключі функцій 23
Обов'язкові ключі функцій 23
Локалізація функцій 23
Кваліфікатори функцій 27
Обов'язкові кваліфікатори функцій 27
Елементи кваліфікаторів 27
Довільний текст 31
Кодуючі послідовності 32
Варіанти 33

ДОДАТКИ

Додаток I – Стандартизована лексика

Додаток II – Визначення типу документа (DTD) для переліку послідовностей

Додаток III – Зразок переліку послідовностей (файл XML)

Додаток IV – Набір символів з Таблиці кодів основних латинських символів Unicode для використання в екземплярі XML переліку послідовностей

Додаток V – Додаткові вимоги до обміну даними (лише для BIB)

Додаток VI – Керівний документ з ілюстрованими прикладами

Доповнення – Керівний документ щодо представлення послідовностей у форматі XML

Додаток VII - Рекомендації щодо перетворення переліку послідовностей зі Стандарту BOIB ST.25 на ST.26: потенційно доданий або видалений об’єкт

СТАНДАРТ ST.26

РЕКОМЕНДОВАНИЙ СТАНДАРТ З ПРЕДСТАВЛЕННЯ ПЕРЕЛІКІВ НУКЛЕОТИДНИХ ТА АМІНОКИСЛОТНИХ ПОСЛІДОВНОСТЕЙ З ВИКОРИСТАННЯМ XML

(РОЗШИРЮВАНОЇ МОВИ РОЗМІТКИ)

Версія 1.6

Редакція затверджена Комітетом зі стандартів ВОІВ(КСВ)
на його десятій сесії 25 листопада 2022 року

ВСТУП

1. У цьому стандарті визначено особливості представлення нуклеотидних та амінокислотних послідовностей у патентних заявках для включення до переліків послідовностей, способи розкриття характерних ознак послідовностей, якими вони мають бути представлені, а також визначення типу документа (DTD) для переліків послідовностей за схемою XML (розширювана мова розмітки). Відомствам інтелектуальної власності рекомендовано приймати будь-який перелік послідовностей, підготовлений відповідно до цього стандарту, як частину патентної заявки або як такий, що стосується патентної заявки.

2. Цей стандарт має на меті:

(a) надати заявникам можливість включати до патентних заявок лише один перелік послідовностей, придатний для цілей як міжнародної, так і національної або регіональної процедур;

(b) підвищити точність і якість представлення послідовностей для спрощення їх розповсюдження, що матиме користь як для заявників, так і для громадськості та експертів;

(c) спростити пошук даних щодо послідовностей; і

(d) забезпечити можливість обміну даними щодо послідовностей в електронній формі і внесення їх до комп'ютеризованих баз даних.

ВИЗНАЧЕННЯ

3. У цьому стандарті:

(a) термін "амінокислота" означає амінокислоту, що може бути представлена з використанням символів, визначених у Додатку I (див. Розділ 3, Таблиця 3). До таких амінокислот належать, наприклад, D-амінокислоти та амінокислоти, що містять модифікованіабо синтетичні бокові ланцюги. Амінокислоти розглядаються в таблиці функцій як немодифіковані L-амінокислоти, якщо не вказано інше відповідно до параграфа 30. Для цього стандарту залишок пептидно-нуклеїнової кислоти (ПНК) вважається не амінокислотою, а нуклеотидом, як зазначено в параграфі 3(g)(i)(2).

(b) вислів "стандартизована лексика" означає термінологію, що використовується у цьому стандарті і повинна використовуватися для опису функцій послідовності, тобто для анотацій до областей або сайтів, як визначено у Додатку I.

(c) "перелік залишків" означає розкриття послідовності в патентній заявці шляхом внесення до переліку по порядку кожного залишку послідовності, де:

(i) залишок представлений назвою, скороченням, символом або структурою (наприклад: HHHHHHQ або HisHisHisHisHisHisGln); або

(ii) численні залишки представлені скороченою формулою (наприклад: His6Gln)

(d) "навмисно пропущена послідовність", також відома як порожня послідовність, стосується вказівника місця заповнення, призначеного для збереження нумерації послідовностей у переліку послідовностей з метою узгодження з розкриттям у заявці, наприклад, якщо послідовність було видалено з розкриття, для уникнення порушення нумерації послідовностей як у розкритті, так і в переліку послідовностей.

(e) "модифікована амінокислота" означає амінокислоту, як описано в параграфі 3(a), окрім L-аланіну, L-аргініну, L-аспарагіну, L-аспарагінової кислоти, L-цистеїну, L-глутаміну, L-глутамінової кислоти, L-гліцину, L-гістидину, L-ізолейцину, L-лейцину, L-лізину, L-метіоніну, L-фенілаланіну, L-проліну, L-пірролізину, L-серину, L-селеноцистеїну, L-треоніну, L-триптофану, L-тирозину, або L-валіну.

(f) "модифікований нуклеотид" означає будь-який нуклеотид, як описано у параграфі 3(g), окрім дезоксиаденозин 3’-монофосфату, дезоксигуанозин 3’-монофосфату, дезоксицитидин 3’-монофосфату, дезокситимідин 3’-монофосфату, аденозин 3’-монофосфату, гуанозин 3’-монофосфату, цитидин 3’-монофосфату, або уридин 3’-монофосфату.

(g) термін "нуклеотид" означає будь-який нуклеотид або аналог нуклеотиду, що може бути представлений з використанням символів, визначених у Додатку I (див. Розділ 1, Таблиця 1), де нуклеотид або аналог нуклеотиду містить:

(i) основний компонент, обраний з:

(1) 2’ дезоксирибози 5’ монофосфату (основний компонент дезоксирибонуклеотиду) або рибози 5’ монофосфату (основний компонент рибонуклеотиду); або

(2) аналогу 2’ дезоксирибози 5’ монофосфату або рибози 5’ монофосфату, який при формуванні каркасу аналогу нуклеїнової кислоти призводить до впорядкування нуклеїнових основ, що імітує розташування нуклеїнових основ в нуклеїнових кислотах, що містять каркас 2' дезоксирибози 5' монофосфату або рибози 5' монофосфату, де аналог нуклеїнової кислоти здатний до спарювання основ разом з комплементарною нуклеїновою кислотою; приклади фрагментів каркасу, що включають амінокислоти, як у пептидних нуклеїнових кислотах, молекулах гліколю, так і в глюколевих нуклеїнових кислотах, молекулах треофуранозильного цукру, та в треосних нуклеїнових кислотах, морфолінових кільцях та фосфородіамідатних групах як у морфолінах, так і в циклогексенільних молекулах, так і у циклогексенільних нуклеїнових кислотах.

та

(ii) один з двох основних компонентів:

(1) таку, що приєднана до нуклеїнової основи, включаючи модифіковану або синтетичну пуринову чи піримідинову нуклеїнову основу; або

(2) таку, що не має пуринової чи піримідинової нуклеїнової основи, де нуклеотид є частиною нуклеотидної послідовності, що називається "AP сайт" або "апуриновий сайт".

(h) "залишок" означає будь-який окремий нуклеотид або будь-яку окрему амінокислоту,або їхні відповідні аналоги в послідовності.

(i) термін "ідентифікаційний номер послідовності" означає унікальний номер (ціле число), присвоєний послідовності в переліку послідовностей.

(j) вислів "перелік послідовностей" означає частину опису поданої патентної заявки або документ, що стосується заявки, поданий пізніше, в якому розкривається (-ються) нуклеотидна (-і) та/або амінокислотна (-і) послідовність (послідовності) з будь-яким додатковим описом, як передбачено цим Стандартом.

(k) вислів "специфічно визначений (-а)" означає будь-який нуклеотид, окрім нуклеотидів, представлених символом "n", або будь-яку амінокислоту, окрім амінокислот, представлених символом "X", зазначений (-у) в Додатку I (див. Розділ 1, Таблиця 1 та Розділ 3,Таблиця 3, відповідно).

(l) термін "невідомий (-а)" нуклеотид або амінокислота означає, що в послідовності присутній (-я) нуклеотид або амінокислота, природа якого (-ї) невідома або не розкривається.

(m) "варіантна послідовність" визначає будь-яку нуклеотидну або амінокислотну послідовність, що містить одну або більше відмінностей по відношенню до первинної послідовності. Ці відмінності можуть включати альтернативні залишки (див. параграфи 15 та 27), модифіковані залишки (див. параграфи 3(g), 3(h), 16 та 29), видалення, вставки та заміни. Див . параграфи 93-95.

(n) "довільний текст" являє собою тип формату значень для певних кваліфікаторів, представлений у формі описової текстової фрази або в іншому визначеному форматі (як вказано у Додатку I). Див. параграф 85.

(o) "довільний текст, що залежить від мови" означає параметр довільного тексту певних кваліфікаторів, що можуть потребувати перекладу для міжнародних, національних або регіональних процедур. Див. параграф 87.

4. У цьому стандарті слово (а):

(a) "може" стосується обов’язкового або допустимого підходу, але не вимоги.

(b) "повинен" стосується вимоги стандарту; ігнорування цієї вимоги призведе до невідповідності.

(c) "не повинен" стосується заборони стандарту.

(d) "має" стосується настійної рекомендації, але не вимоги.

(e) "не має" стосується неприйнятного підходу, але не заборони.

СФЕРА ЗАСТОСУВАННЯ

5. Цей стандарт визначає вимоги до представлення переліків нуклеотидних та амінокислотних послідовностей, що розкриваються у патентних заявках.

6. Перелік послідовностей за цим стандартом (далі – перелік послідовностей) містить частину із загальною інформацією і частину з даними щодо послідовності. Перелік послідовностей повинен представлятися в одному файлі за схемою XML з використанням DTD (визначення типу документа), наведеного у Додатку II. Бібліографічна інформація, що міститься у частині із загальною інформацією, призначена виключно для асоціювання переліку послідовностей з патентною заявкою, стосовно якої цей перелік послідовностей був

поданий. Частина з даними щодо послідовності містить один або більше елементів даних, кожний з яких містить інформацію про одну послідовність. Елементи даних послідовності включають різні ключі функцій і відповідні кваліфікатори на основі специфікацій Проєкту міжнародного співробітництва баз даних нуклеотидних послідовностей (INSDC) і UniProt.

7. У цьому стандарті послідовність, включена до переліку послідовностей, є послідовністю, що розкривається у заявці шляхом нумерування її залишків і може бути представлена:

(a) нерозгалуженою послідовністю або лінійною ділянкою розгалуженої послідовності з десятьма чи більше специфічно визначеними нуклеотидами, серед яких суміжні нуклеотиди з’єднуються:

(i) 3' до 5' (або 5' до 3') фосфодіефірним зв’язком; або

(ii) будь-яким хімічним зв’язком, який призводить до розташування суміжних нуклеїнових основ, що імітує розташування нуклеїнових основ в природних нуклеїнових кислотах; або

(b) нерозгалуженою послідовністю або лінійною ділянкою розгалуженої послідовності з чотирма чи більше специфічно визначеними амінокислотами, серед яких амінокислоти формують один пептидний каркас, тобто суміжні амінокислоти з’єднані пептидними зв'язками.

8. Перелік послідовностей не повинен містити, як послідовність з присвоєним власним ідентифікаційним номером послідовності, послідовності з менш ніж десятьма специфічно визначеними нуклеотидами або з менш, ніж чотирма специфічно визначеними амінокислотами.

ПОСИЛАННЯ

9. Для цього стандарту мають значення наступні ресурси і стандарти:

Проєкт міжнародного співробітництва баз даних нуклеотидних послідовностей (INSDC)	http://www.insdc.org/;
Міжнародний стандарт ISO 639-1:2002	"Коди для представлення назв мов. Частина 1: Код Альфа-2";
Консорціум UniProt	http://www.uniprot.org/;
Схема W3C XML 1.0	http://www.w3.org/;
Стандарт ВОІВ ST.2	"Стандартний спосіб представлення календарних дат з використанням григоріанського календаря";
Стандарт ВОІВ ST.3	"Рекомендований стандарт стосовно двобуквених кодів для представлення держав, інших адміністративних одиниць та міжурядових організацій";

СТАНДАРТ ВОІВ ST.16	"Рекомендовані стандартні коди для ідентифікації різних видів патентних документів";
СТАНДАРТ ВОІВ ST.25	"Стандарт з представлення переліків нуклеотидних та амінокислотних послідовностей у патентних заявках".

ПРЕДСТАВЛЕННЯ ПОСЛІДОВНОСТЕЙ

10. Кожній послідовності, що міститься в параграфі 7, повинен бути присвоєний окремий ідентифікаційний номер, включаючи послідовність, яка є ідентичною області довшої послідовності. Ідентифікаційні номери послідовностей повинні починатися з номера 1 і збільшуватися послідовно на одиницю. Якщо під ідентифікаційним номером послідовності не наведено жодної послідовності, тобто у випадку навмисно пропущеної послідовності, замість неї повинно бути зазначено "000" (див. параграф 58). У переліку послідовностей повинна бути вказана загальна кількість послідовностей, що повинна дорівнювати загальній кількості ідентифікаційних номерів послідовностей і після якої має наводитися послідовність або "000".

Нуклеотидні послідовності

11. Нуклеотидна послідовність повинна представлятися одним ланцюгом зліва направо у напрямку від 5'-кінця до 3'-кінця або у напрямку зліва направо, що імітує напрямок від 5'-кінця до 3'-кінця. Позначення 5' і 3' або інші подібні позначення не повинні бути включені до послідовності. Дволанцюгова нуклеотидна послідовність, що розкривається шляхом нумерування залишків обох спіралей, повинна бути представлена як:

(a) одна послідовність або дві окремі послідовності, кожна із власним ідентифікаційним номером, в яких дві окремі спіралі є повністю комплементарними одна до іншої, або

(b) дві окремі послідовності, кожна із власним ідентифікаційним номером, в яких дві спіралі не є повністю комплементарними одна до іншої.

12. Для цього стандарту перший нуклеотид, представлений в послідовності є положенням залишку номер 1. Якщо нуклеотидні послідовності кільцеві, тоді заявник повинен вибрати нуклеотид в положенні залишку номер 1. Нумерація наскрізна у всій послідовності у напрямку від 5'-кінця до 3'-кінця або у напрямку, що імітує напрямок від 5'- кінця до 3'-кінця. Останній номер положення залишку повинен дорівнювати кількості нуклеотидів у послідовності.

13. Усі нуклеотиди в послідовності повинні бути представлені з використанням символів, визначених у Додатку I (див. Розділ 1, Таблиця 1). Повинні використовуватись лише маленькі літери. Будь-який символ, що представляє нуклеотид, є еквівалентом лише одного залишку.

14. Символ "t" має розглядатися як тимін в ДНК і як урацил в РНК. Урацил у ДНК або тимін у РНК вважається модифікованим нуклеотидом і повинен докладно описуватись у таблиці функцій, як зазначено у параграфі 19.

15. Якщо може бути використаний багатозначний символ (для представлення двох або

більше альтернативних нуклеотидів), слід використовувати найбільш обмежувальний символ, як зазначено в Додатку I (див. Розділ 1, Таблиця 1). Наприклад, якщо нуклеотид у певному положенні може бути як "a", так і "g", слід використовувати "r", а не "n". Символ "n" має розглядатися як будь-який із символів "a", "c", "g" або "t/u", окрім випадків, коли він використовується для додаткового опису у таблиці функцій. Символ "n" повинен використовуватись для представлення виключно нуклеотидів. Окремий модифікований або "невідомий" нуклеотид може бути представлений символом "n" з додатковим описом у таблиці функцій, як зазначено у параграфах 16, 17,21 або 93-96. Для представлення варіантів послідовностей, тобто альтернатив, видалень, вставок та заміщень, див. параграфи 94 - 100.

16. Якщо можливо, модифіковані нуклеотиди мають бути представлені у послідовності як відповідні немодифіковані нуклеотиди, тобто "a", "c", "g" або "t". Будь-який модифікований нуклеотид у послідовності, що не може бути представлений жодним символом із Додатка I (див. Розділ 1, Таблиця 1), тобто нуклеотид "інший", наприклад нуклеотид, що не зустрічається у природі, повинен бути представлений символом "n". Символ "n" є еквівалентом лише одного залишку.

17. Модифікований нуклеотид повинен бути детально описаний в таблиці функцій (див. параграф 60 і далі) з використанням ключа функції "modified_base" та обов'язкового кваліфікатора "mod_base" разом із скороченням з Додатка I (див. Розділ 2, Таблиця 2) як значення кваліфікатора; якщо скороченням є "OTHER", у значенні кваліфікатора "note" повинна бути наведена повна нескорочена назва модифікованого нуклеотиду. Для переліку альтернативних модифікованих нуклеотидів значення кваліфікатора "OTHER" може використовуватись у поєднанні з кваліфікатором "note" (див. параграфи 97 та 98). Зазначені вище скорочення (або повні назви), наведені у Додатку I (див. Розділ 2, Таблиця 2), не повинні зазначатися безпосередньо у послідовності.

18. Нуклеотидна послідовність, включаючи одну або більше областей послідовних модифікованих нуклеотидів, які мають один основний компонент (див. параграф 3(g)(i)(2)), повинні надалі описуватися в таблиці функцій, як наведено у параграфі 17. Модифіковані нуклеотиди кожної такої області можуть бути спільно описані в єдиному елементі "INSDFeature", як наведено у параграфі 22. Більш обмежуюча нескорочена хімічна назва, що охоплює всі модифіковані нуклеотиди в діапазоні або перелік хімічних назв всіх нуклеотидів в діапазоні повинні наводитися як значення в кваліфікаторі "note". Наприклад, гліколева нуклеїнова кислотна послідовність, що містить нуклеотиди "a", "c", "g", або "t" може бути описана в кваліфікаторі "note" як "2,3-dihydroxypropyl nucleosides". В іншому випадку, така сама послідовність може бути описана в кваліфікаторі "note" як "2,3-dihydroxypropyladenine, 2,3-dihydroxypropylthymine, 2,3-dihydroxypropylguanine, or 2,3-dihydroxypropylcytosin". Якщо окремий модифікований нуклеотид в області включає додаткову модифікацію, тоді модифікований нуклеотид повинен детально описуватися в таблиці функцій, як наведено у параграфі 17.

19. Урацил в ДНК або тимін в РНК вважаються модифікованими нуклеотидами і повинні представлятися у послідовності як "t" з додатковим описом у таблиці функцій з використанням ключа функції "modified_base", кваліфікатора "mod_base" зі значенням "OTHER" та кваліфікатора "note" зі значенням "урацил" або "тимін" відповідно.

20. Нижче наведено приклади представлення модифікованих нуклеотидів відповідно до

параграфів 16-18:

Приклад 1: Модифікований нуклеотид з використанням скорочення з Додатка I (див. Розділ 2, Таблиця 2)

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

Приклад 2: Модифікований нуклеотид з використанням скорочення "OTHER" з Додатка I (див. Розділ 2, Таблиця 2)

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

Приклад 3: Нуклеотидна послідовність, що складається з модифікованих нуклеотидів, охоплена у параграфі 3(g)(i)(2) з двома окремими нуклеотидами, що включають в себе подальшу модифікацію

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

```
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>439</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>i</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
    <INSDFeature_key>modified_base</INSDFeature_key>
    <INSDFeature_location>684</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>xanthine</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

21. Будь-який "невідомий" нуклеотид повинен бути представлений у послідовності символом "n". "Невідомий" нуклеотид має бути детально описаний в таблиці функцій (див. параграф 60 і далі), з використанням ключа функції "unsure". Символ "n" є еквівалентом лише одного залишку.

22. Область, що містить відому кількість безперервних залишків "a", "c", "g", "t" або "n", яких стосується один і той самий опис, може бути описана з використанням елемента "INSDFeature" та синтаксису "x..y" як дескриптора локалізації в елементі `INSDFeature_location` (див. параграфи 64-71). Інформацію про представлення варіантів послідовностей, тобто видалення, альтернативи, вставки або заміщення, наведено у параграфах 94-100.

23. Нижче наведено приклад представлення області модифікованих нуклеотидів, для яких застосовується той самий опис, відповідно до параграфа 22:

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

Амінокислотні послідовності

24. Амінокислоти в амінокислотній послідовності повинні бути представлені зліва направо у напрямку від амінів до карбоксилів. Аміногрупи і карбоксильні групи не повинні наводитися саме у послідовності.

25. У цьому стандарті перша амінокислота в послідовності є положенням залишку номер 1, включаючи амінокислоти, що передують зрілому білку, наприклад, у пре-послідовностях, про-послідовностях, пре-про-послідовностях і сигнальних послідовностях. Якщо амінокислотна послідовність кільцева та кільце складається виключно з амінокислотних залишків, з'єднаних пептидними зв'язками, тобто послідовність не має амінокінця та карбоксильного кінця, тоді заявник повинен обрати амінокислоту в положенні залишку номер 1. Нумерація наскрізна по всій послідовності у напрямку від амінів до карбоксилів.

26. Усі амінокислоти в послідовності повинні бути представлені з використанням символів, визначених у Додатку I (Розділ 3, Таблиця 3). Повинні використовуватись лише великі літери. Будь-який символ, що представляє амінокислоту, є еквівалентом лише одного залишку.

27. Якщо може бути використаний багатозначний символ (для представлення двох або більше амінокислот у варіанті), слід використовувати найбільш обмежувальний символ, як наведено у Додатку I (див. Розділ 3, Таблиця 3). Наприклад, якщо амінокислота у певному положенні може бути як аспарагіновою кислотою, так і аспарагіном, слід використовувати символ "B", а не "X". Символ "X" має розглядатися як будь-який із символів "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" або "V", окрім випадків, коли він використовується для додаткового опису у таблиці функцій. Символ "X" не повинен використовуватись для представлення будь-чого іншого, крім амінокислот. Окрема модифікована або "невідома" амінокислота може бути представлена символом "X" з додатковим описом у таблиці функцій, наприклад, як зазначено у параграфах 29, 30, 32 або 94-98. Інформацію про представлення варіантів послідовностей, тобто альтернативи, видалення, вставки або заміщення, наведено у параграфах 94-100.

28. Розкриті амінокислотні послідовності, відокремлені в описі внутрішнім термінатором, представлені, наприклад "Ter", або зірочкою "*", або крапкою ".", або пробілом, повинні бути включені як окремі послідовності для кожної амінокислотної послідовності, що містить принаймні чотири специфічно визначені амінокислоти й охоплюється параграфом 7. Кожній такій окремій послідовності повинен присвоюватися власний ідентифікаційний номер послідовності. Термінатори і пробіли не повинні бути включені до послідовності у переліку послідовностей (див. параграф 57).

29. За будь-якої можливості, модифіковані амінокислоти, і D-амінокислоти включно, мають представлятися у послідовності як відповідні немодифіковані амінокислоти. Будь-яка модифікована кислота у послідовності, що не може бути представлена жодним символом із Додатка I (див. Розділ 3, Таблиця 3), тобто "інша" амінокислота, повинна бути представлена символом "X". Символ "X" є еквівалентом лише одного залишку.

30. Модифікована амінокислота повинна бути детально описана в таблиці функцій (див. параграф 60 і далі). Де це можливо, ключі функції "CARBOHYD" або "LIPID", слід використовувати разом з кваліфікатором "note". Ключ функції "MOD_RES" слід використовувати для інших пост-трансляційно модифікованих амінокислот разом з кваліфікатором "note"; в інших випадках слід використовувати ключ функції "SITE" разом з

кваліфікатором "note". Значенням кваліфікатора "note" повинне бути скорочення із Додатка I (див. Розділ 4, Таблиця 4) або повна, нескорочена назва модифікованої амінокислоти. Зазначені вище скорочення або повні, нескорочені назви, наведені у Таблиці 4, не повинні зазначатися безпосередньо у послідовності.

31. Нижче наведено приклади представлення модифікованих амінокислот відповідно до параграфа 30:

Приклад 1: Пост-трансляційно модифікована амінокислота

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

Приклад 2: Не пост-трансляційно модифікована амінокислота

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

Приклад 3: D-амінокислота

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

32. Будь-яка "невідома" амінокислота повинна бути представлена у послідовності символом "X". "Невідома" амінокислота, позначена "X", повинна бути детально описана у таблиці функцій (див. параграф 60 і далі), використовуючи ключ функції "UNSURE" і

необов'язковий кваліфікатор "note". Символ "X" є еквівалентом лише одного залишку.

33. Нижче наведено приклад представлення "невідомої" амінокислоти відповідно до параграфа 32:

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

34. Область, що містить відому кількість безперервних залишків "X", яких стосується один і той самий опис, може бути описана з використанням синтаксису "x..y" як дескриптора локалізації в елементі `INSDFeature_location` (див. параграфи 64-70). Інформацію про представлення варіантів послідовностей, тобто видалення, альтернативи, вставки або заміщення, наведено у параграфах 94-100.

Представлення окремих випадків

35. Послідовність, що розкривається шляхом нумерування її залишків і утворює єдину неперервну послідовність з одного або більше небезперервних фрагментів більшої послідовності або з фрагментів різних послідовностей, повинна включатися до переліку послідовностей, і їй повинен бути присвоєний власний ідентифікаційний номер.

36. Послідовність, яка містить області специфічно визначених залишків, відокремлених однією або більше областями безперервних залишків "n" або "X" (див. параграфи 15 і 27 відповідно), і в якій зазначено точну кількість залишків "n" або "X" у кожній області, повинна включатися до переліку послідовностей як одна послідовність, і їй повинен бути присвоєний власний ідентифікаційний номер.

37. Послідовність, яка містить області специфічно визначених залишків, відокремлених один від одного одним чи більше проміжками невідомої або нерозкритої кількості залишків, не повинна представлятися в переліку послідовностей як окрема послідовність. Кожна область специфічно визначених залишків, що охоплюється параграфом 7, повинна бути включена до переліку послідовностей як окрема послідовність із присвоєним власним ідентифікаційним номером.

СТРУКТУРА ПЕРЕЛІКУ ПОСЛІДОВНОСТЕЙ В XML

38. Відповідно до параграфа 6 цього стандарту, екземпляр XML переліку послідовностей складається з наступного:

(a) частина із загальною інформацією про патентну заявку, до якої відноситься перелікпослідовностей; і

(b) частина з даними щодо послідовності, яка містить один або більше елементів даних послідовності, кожний з яких, у свою чергу, містить інформацію про одну

послідовність.

Приклад переліку послідовностей наведено у Додатку III.

39. Перелік послідовностей повинен бути представлений за схемою XML 1.0 з використанням DTD, наведеного у Додатку II "Визначення типу документа для переліку послідовностей".

(a) Перший рядок екземпляра XML повинен містити декларацію XML:

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

(b) Другий рядок екземпляра XML повинен містити декларацію типу документа (DOCTYPE):

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

40. Весь перелік послідовностей в електронному вигляді повинен знаходитися в одному файлі. У файлі повинне використовуватися кодування Unicode UTF-8, але з такими обмеженнями:

(a) інформація, що міститься в елементах `ApplicantName`, `InventorName` та `InventionTitle` у частині із загальною інформацією, та значення частини даних послідовності `NonEnglishQualifier`, може наводитися будь-якими дійсними символами Unicode, що зазначені у специфікації XML 1.0, крім точок коду Unicode Control 0000-001F та 007F-009F. Зарезервовані символи “, &,‘, <, > (точки коду Unicode 0022, 0026, 0027, 003C та 003E відповідно), повинні бути замінені відповідно до параграфа 41; та

(b) інформація, що міститься у всіх інших елементах та атрибутах загальної інформаційної частини та у всіх інших елементах та атрибутах частини даних послідовності, повинна складатися з друкованих символів (включаючи пробіл) із таблиці кодів основних латинських символів Unicode (тобто точки коду Unicode обмежені з 0020 по 007E - див. Додаток IV). Зарезервовані символи ", &,‘, <i> (точки коду Unicode 0022, 0026, 0027, 003C та 003E відповідно) повинні бути замінені відповідно до параграфа 41.

41. В екземплярі XML переліку послідовностей, цифрові посилання[1] на символи не повинні використовуватися та наступні зарезервовані символи, у разі їх використання як значення атрибута або як змісту елемента, повинні бути замінені відповідними попередньо визначеними сутностями:

Зарезервований символ	Визначені елементи
<	<
>	>
&	&
"	"
'	'

Див. приклад у параграфі 71. Єдиними дозволеними посиланнями на сутності символів є

[1] Цифрове посилання на символ посилається на символ за його точкою коду Universal Character Set / Unicode і використовує формат: "& # nnnn;" або “& # xhhhh;”, де “nnnn” – точка коду в десятковій формі, а “hhhh” - в шістнадцятковій формі.

попередньо визначені сутності, викладені в цьому абзаці.

42. Усі обов'язкові елементи підлягають заповненню (окрім навмисно пропущених послідовностей відповідно до параграфа 58). Необов'язкові елементи без змісту не повинні наводитися в екземплярі XML (окрім представлення видалення в послідовності у значенні кваліфікатора "replace" відповідно до параграфа 97).

Кореневий елемент

43. Відповідно до цього стандарту, кореневим елементом екземпляра XML є елемент `ST26SequenceListing` з такими атрибутами:

Атрибут	Опис	Обов'язковий/ Необов'язковий
`dtdVersion`	Версія DTD, використана для створення цього файлу, у форматі "V#_#", наприклад "V1_2".	Обов'язковий
`fileName`	Ім'я файлу, що містить перелік послідовностей.	Необов'язковий
`softwareName`	Назва програмного забезпечення, в якому було створено цей файл.	Необов'язковий
`softwareVersion`	Версія програмного забезпечення, в якому було створено цей файл.	Необов'язковий
`productionDate`	Дата створення файлу, що містить перелік послідовностей (у форматі "РРРР-ММ-ДД").	Необов'язковий
`originalFreeTextLanguageCode`	Код мови (див. посилання у параграфі 9 на ISO 639-1:2020) для єдиної мови оригіналу, якою були підготовлені кваліфікатори довільного тексту, залежного від мови.	Необов’язковий
`nonEnglishFreeTextLanguageCode`	Код мови (див. посилання у параграфі 9 на ISO 639-1:2020) для елементів `nonEnglishFreeTextLanguageCode`	Обов’язковий, коли в списку послідовностей присутній елемент `nonEnglishFreeTextLanguageCode`

44. Приклад кореневого елемента `ST26SequenceListing` екземпляра XML та його атрибутів відповідно до параграфа 43:

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

*{...} представляє частину із загальною інформацією та частину з даними щодо послідовності, не відображені у цьому прикладі.

Частина із загальною інформацією

45. Елементи частини із загальною інформацією відносяться до інформації, що стосується патентної заявки, а саме:

Елемент	Опис	Обов'язковий/ Необов'язковий
`ApplicationIdentification`	Ідентифікаційні дані заявки, до якої відноситься перелік послідовностей	Обов'язковий, якщо перелік послідовностей подається після присвоєння номера заявки
Елемент `ApplicationIdentification` складається з:		
`IPOfficeCode`	Код відомства подання за стандартом ВОІВ ST.3	Обов'язковий
`ApplicationNumberText`	Номер заявки, наданий відомством, що подало заявку (наприклад, PCT/IB2013/099999).	Обов'язковий
`FilingDate`	Дата подання патентної заявки, стосовно якої подається перелік послідовностей (у форматі "РРРР-ММ-ДД" за стандартом ВОІВ ST.2, де "РРРР" – календарний рік, ММ – календарний місяць, ДД – день у межах календарного місяця, наприклад 2015-01-31)	Обов'язковий, якщо перелік послідовностей подається після встановлення дати подання
`ApplicantFileReference`	Унікальний ідентифікаційний номер, присвоєний заявником для ідентифікації певної заявки, наведений символами відповідно до параграфа 40(b)	Обов'язковий, якщо перелік послідовностей подається до присвоєння номера заявки; в інших випадках – необов'язковий
`EarliestPriorityApplicationIdentification`	Ідентифікація заявки з більш раннім пріоритетом (також містить елементи `IPOfficeCode`, `ApplicationNumberText` і `FilingDate`, див. елементи `ApplicationIdentification`)	Обов'язковий, якщо заявлено пріоритет

ApplicantName	Ім'я першого вказаного заявника, представлене символами, визначеними у параграфі 40 (a). Цей елемент включає обов'язковий атрибут languageCode, як зазначено у параграфі 47.	Обов'язковий
ApplicantNameLatin	Якщо ApplicantName наведено символами, не зазначеними у параграфі 40(b), необхідно навести переклад або транслітерацію імені першого зазначеного заявника символами, зазначеними у параграфі 40(b)	Обов'язковий, якщо елемент ApplicantName містить інші символи ніж латинські.
InventorName	Ім'я першого вказаного винахідника, представлене символами, визначеними у параграфі 40(a). Цей елемент включає обов'язковий атрибут languageCode, як зазначено у параграфі 47.	Необов'язковий
InventorNameLatin	Якщо InventorName наведено символами, не зазначеними у параграфі 40(b), необхідно навести переклад або транслітерацію імені першого зазначеного винахідника символами, зазначеними у параграфі 40(b)	Необов'язковий
InventionTitle	Назва винаходу, вказана символами як зазначено у параграфі 40(a), мовою оригіналу. Переклад назви винаходу додатковими мовами може наводитися символами як зазначено у параграфі 40(a) з використанням додаткових елементів InventionTitle. Цей елемент включає обов'язковий атрибут languageCode як зазначено у параграфі 48. Назва винаходу має бути від двох до семи слів.	Обов'язковий для мови подання. Необов'язковий для додаткових мов.

SequenceTotalQuantity	Загальна кількість усіх послідовностей в переліку послідовностей, у тому числі навмисно пропущених послідовностей (відомих як порожні послідовності) (див. параграф 10).	Обов'язковий

46. Нижче наведено приклад представлення частини переліку послідовностей із загальною інформацією відповідно до параграфа 45.

Приклад 1: Перелік послідовностей, поданий до присвоєння заявці ідентифікаційних даних ідати подання

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_0.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="Invention_SEQL.xml"
softwareName="WIPO Sequence" softwareVersion="1.0"
productionDate="2022-05-10" originalFreeTextLanguageCode="en"
nonEnglishFreeTextLanguageCode="ja">
    <ApplicantFileReference>AB123</ApplicantFileReference>
    <EarliestPriorityApplicationIdentification>
        <IPOfficeCode>IB</IPOfficeCode>
        <ApplicationNumberText>PCT/IB2013/099999</ApplicationNumberText>
        <FilingDate>2014-07-10</FilingDate>
    </EarliestPriorityApplicationIdentification>
    <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
    <InventorName languageCode="en">Keiko Nakamura</InventorName>
    <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
    PROTEINS</InventionTitle>
    <SequenceTotalQuantity>9</SequenceTotalQuantity>
    <SequenceData sequenceIDNumber="1"> {...}* </SequenceData>
    <SequenceData  sequenceIDNumber="2">  {...}  </SequenceData>
    <SequenceData  sequenceIDNumber="3">  {...}  </SequenceData>
    <SequenceData  sequenceIDNumber="4">  {...}  </SequenceData>
    <SequenceData  sequenceIDNumber="5">  {...}  </SequenceData>
    <SequenceData  sequenceIDNumber="6">  {...}  </SequenceData>
    <SequenceData  sequenceIDNumber="7">  {...}  </SequenceData>
    <SequenceData sequenceIDNumber="8"> {...} </SequenceData>
    <SequenceData  sequenceIDNumber="9">  {...}  </SequenceData>
</ST26SequenceListing>
* {...} представляє відповідну інформацію для кожної послідовності, не
представлену в цьому прикладі.
```

Приклад 2: Перелік послідовностей, поданий після присвоєння заявці ідентифікаційних даних і дати подання

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.
3//EN" "ST26SequenceListing_V1_3.dtd">
```

```
<ST26SequenceListing dtdVersion="1_3" fileName="Invention_SEQL.xml"
softwareName="WIPO Sequence" softwareVersion="1.0"
productionDate="2022-05-10" originalFreeTextLanguageCode="en"
nonEnglishFreeTextLanguageCode="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>

* {...} представляє відповідну інформацію для кожної послідовності, не
представлену в цьому прикладі.
```

47. Ім'я заявника і, за бажанням, ім'я винахідника повинні бути зазначені в елементах `ApplicantName` та `InventorName` відповідно, оскільки їх зазвичай наводять мовою подання заявки. Для кожного елемента повинен бути зазначений відповідний код мови (див. посилання в параграфі 9 на ISO 639-1:2002) в атрибуті `languageCode`. Якщо вказане ім'я заявника містить інші символи, окрім символів латинської абетки, як зазначено у параграфі 40(b), необхідно також навести транслітерацію або переклад імені заявника символами латинської абетки в елементі `ApplicantNameLatin`. Якщо вказане ім'я винахідника містить інші символи, окрім символів латинської абетки, необхідно також навести транслітерацію або переклад імені винахідника символами латинської абетки в елементі `InventorNameLatin`.

48. Назва винаходу повинна бути зазначена в елементі `InventionTitle` мовою подання, а також може бути зазначена іншими мовами в декількох елементах `InventionTitle` (див. таблицю у параграфі 45). Для кожного елемента повинен бути зазначений відповідний код мови (див. посилання в параграфі 9 на ISO 639-1:2002) в атрибуті `languageCode`.

49. Нижче наведено приклад представлення імен і назви винаходу відповідно до параграфів 47 і 48.

Приклад: Імена заявника і винахідника представлені японськими і латинськими символами, а назва винаходу – японською, англійською та французькою мовами

```
<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>
```

Частина з даними щодо послідовності

50. Частина з даними щодо послідовності повинна містити один або більше елементів `SequenceData`, кожний з яких містить інформацію про одну послідовність.

51. Кожний елемент `SequenceData` повинен мати обов'язковий атрибут `sequenceIDNumber`, в якому міститься ідентифікаційний номер для кожної послідовності (див. параграф 10). Наприклад:

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

52. Елемент `SequenceData` повинен містити залежний елемент `INSDSeq`, що складається зтаких додаткових залежних елементів:

Елемент	Опис	Обов'язковий/Не включається	
		Послідовності	Навмисно пропущені послідовності
INSDSeq_length	Довжина послідовності	Обов'язковий	Обов'язковий без значення
INSDSeq_moltype	Тип молекули	Обов'язковий	Обов'язковий без значення
INSDSeq_division	Зазначення того, що послідовність пов'язана з патентною заявкою	Обов'язковий із значенням "PAT"	Обов'язковий без значення
INSDSeq_feature-table	Перелік тлумачень послідовності	Обов'язковий	НЕ повинен включатися
INSDSeq_sequence	Послідовність	Обов'язковий	Обов'язковий із значенням "000"

53. Елемент `INSDSeq_length` повинен розкривати кількість нуклеотидів або амінокислот послідовності, що міститься в елементі `INSDSeq_sequence`. Наприклад:

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

54. Елемент `INSDSeq_moltype` повинен розкривати тип представленої молекули. Для нуклеотидних послідовностей, включаючи послідовності аналогів нуклеотидів, тип молекули повинен бути вказаний як "ДНК" або "РНК". Для амінокислотних послідовностей, тип молекули повинен бути вказаний як "AA". (Цей елемент відрізняється від кваліфікатора "mol_type" , що розглядається у параграфах 55 і 84). Наприклад:

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

55. Для нуклеотидної послідовності, яка містить сегменти, що складаються з одного або більше нуклеотидів як ДНК, так і РНК, тип молекули повинно бути позначено як ДНК. Комбінована молекула ДНК/РНК повинна бути додатково описана у таблиці функцій з використанням ключа функції "source" та обов'язкового кваліфікатора "organism" із значенням "synthetic construct" *("синтетичний конструкт")*, а також обов'язкового кваліфікатора "mol_type" із значенням "other DNA" *("інша ДНК")*. Кожний сегмент ДНК і РНК комбінованої молекули ДНК/РНК повинен бути додатково описаний з використанням ключа функції "misc_feature" та кваліфікатора "note", в якому зазначається, чи це сегмент ДНК, чи РНК.

56. Нижче наведено приклад опису нуклеотидної послідовності, що містить сегменти РНК і ДНК, відповідно до параграфа 55.

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

```
            </INSDFeature_quals>
        </INSDFeature>
        <INSDFeature>
            <INSDFeature_key>misc_feature</INSDFeature_key>
            <INSDFeature_location>61..120</INSDFeature_location>
            <INSDFeature_quals>
                <INSDQualifier>
                    <INSDQualifier_name>note</INSDQualifier_name>
                    <INSDQualifier_value>RNA</INSDQualifier_value>
                </INSDQualifier>
            </INSDFeature_quals>
        </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>cgacccacgcgtccgaggaaccaaccatcacgtttgaggacttcgtgaaggaatt
    ggataatacccgtccctaccaaaatggcgagcgccgactcattgctcctcgtaccgtcgagcggc</INSDSe
    q_sequence>
</INSDSeq>
```

57. Елемент `INSDSeq_sequence` повинен розкривати послідовність. Лише відповідні символи, наведені у Додатку I (див. Розділ 1, Таблиця 1 і Розділ 3, Таблиця 3) повинні бути включені до послідовності. Послідовність не повинна включати цифри, розділові знаки або пробіли.

58. Навмисно пропущена послідовність повинна бути включена до переліку послідовностей та представлена таким чином:

(a) елемент `SequenceData` і його атрибут `sequenceIDNumber` з ідентифікаційним номером пропущеної послідовності як його значення;

(b) елементи `INSDSeq_length`, `INSDSeq_moltype` і `INSDSeq_division` наводяться, але без значень;

(c) елемент `INSDSeq_feature-table` не повинен включатися; і

(d) елемент `INSDSeq_sequence` із значенням "000" .

59. Нижче наведено приклад представлення навмисно пропущеної послідовності відповіднодо параграфа 58:

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

Таблиця функцій

60. Таблиця функцій містить інформацію про локалізацію та призначення різних ділянок у межах певної послідовності. Таблиця функцій повинна наводитися для кожної послідовності, окрім навмисно пропущених послідовностей. Таблиця функцій міститься в елементі `INSDSeq_feature-table`, який складається з одного або більше елементів `INSDFeature`.

61. Кожний елемент `INSDFeature` описує одну функцію і складається з таких залежних елементів:

Елемент	Опис	Обов'язковий/ Необов'язковий
`INSDFeature_key`	Слово чи абревіатура, що ідентифікує функцію	Обов'язковий
`INSDFeature_location`	Ділянка послідовності, що відповідає функції	Обов'язковий
`INSDFeature_quals`	Кваліфікатор з додатковою інформацією про функцію	Обов'язковий, якщо ключ функції потребує зазначення одного або більше кваліфікаторів, наприклад "source"; в інших випадках – необов'язковий

Ключі функцій

62. Додаток I містить виключний список ключів функцій, які повинні використовуватись за цим стандартом, а також виключний список відповідних кваліфікаторів із зазначенням того, чи є ці кваліфікатори обов'язковими або необов'язковими. Розділ 5 Додатка I містить виключний список ключів функцій для нуклеотидних послідовностей, а Розділ 7 Додатка I містить виключний список ключів функцій для амінокислотних послідовностей.

Обов'язкові ключі функцій

63. Ключ функції "source" є обов'язковим для всіх нуклеотидних послідовностей та для всіх амінокислотних послідовностей, окрім навмисно пропущених послідовностей. Кожна послідовність повинна мати один ключ функції "source", що охоплює всю послідовність. Якщо послідовність походить з кількох джерел, ці джерела можуть бути додатково описані в таблиці функцій за допомогою ключа функції "misc_feature" та кваліфікатора "note" для нуклеотидних послідовностей або за допомогою ключа функції "REGION" та кваліфікатора "note" для амінокислотних послідовностей.

Локалізація функцій

64. Обов'язковий елемент `INSDFeature_location` повинен містити принаймні один дескриптор локалізації, що визначає сайт або ділянку, яка відповідає функції послідовності в елементі `INSDSeq_sequence`. Послідовності амінокислот повинні містити один і тільки один дескриптор в обов’язковому елементі `INSDFeature` більше операторів локалізації (див. параграфи 67-70). Нуклеотидні послідовності можуть мати більше одного дескриптора в обов’язковому елементі `INSDFeature_location`, коли він використовується разом з одним або декількома операторами локалізації (див. параграфи 67–70).

65. Дескриптором локалізації може бути окремий залишок, ділянка, яка обмежує безперервний діапазон залишків, або сайт чи ділянка, що виходить за межі визначеного залишку чи діапазону залишків. Дескриптор локалізації не повинен включати нумерацію залишків поза діапазоном послідовності в елементі INSDSeq_sequence. Тільки для нуклеотидних послідовностей дескриптором локалізації може бути сайт між двома сусідніми

номерами залишків. Якщо функція відповідає перервним сайтам або ділянкам нуклеотидної послідовності, повинні використовуватись декілька дескрипторів локалізації у поєднанні з оператором локалізації (див. параграфи 67-70).

66. Синтаксис для кожного типу дескриптора локалізації зазначено у таблиці нижче, де x та y є номерами залишків, позначені як цілі додатні цілі числа менше довжини послідовності в елементі `INSDSeq_sequence`, а x менше y.

(a) Дескриптори локалізації для нуклеотидних та амінокислотних послідовностей:

Тип дескриптора локалізації	Синтаксис	Опис
Номер окремого залишку	`x`	Вказує на один залишок у послідовності.
Номери залишків, що обмежують діапазон послідовності	`x..y`	Вказує на обмежений безперервний діапазон залишків, включаючи початковий і кінцевий залишки

Залишки перед першим або після останнього вказаного номера залишку	`<x`
	`>x`
	`<x..y`
	`x..>y`
	`<x..>y`
		Вказує на ділянку, що включає визначений залишок або діапазон залишків і виходить за межі визначеного залишку. Символи '<' і '>' можуть використовуватися з одиничним залишком або з початковим та кінцевим номером залишку діапазону залишків для того, щоб вказати, що функція виходить за межі вказаного номера залишку.

(b) Дескриптори локалізації лише для нуклеотидних послідовностей:

Тип дескриптора локалізації	Синтаксис	Опис
Сайт між двома суміжними номерами залишків.	`x^y`	Вказує на сайт між двома суміжними залишками, наприклад сайт розщеплювання ендонуклеози. Номери положень суміжних залишків розділяються циркумфлексом (^). Допустимими форматами для цього дескриптора є x^x+1 (наприклад, 55^56), або, для кільцевих нуклеотидів, x^1, де "x" є довжиною молекули, тобто 1000^1 для кільцевої молекули з довжиною 1000.

(c) Дескриптори локалізації лише для амінокислотних послідовностей:

Тип дескриптора локалізації	Синтаксис	Опис
Номери залишків, об’єднаних внутрішньоланцюговою зшивкою	`x..y`	Вказує на амінокислоти, з’єднані внутрішньоланцюговим зв’язком, коли використовується функція, яка вказую на внутрішньоланцюгову зшивку, наприклад "CROSSLINK" або "DISULFID".

67. Елемент INSDFeature_location послідовностей нуклеотидів може містити один або декілька операторів локалізації. Оператор локалізації є префіксом до одного дескриптора локалізації або до комбінації дескрипторів локалізації, що відповідають єдиній, але перервній функції, і визначає локалізацію відповідної функції у вказаній послідовності або природу функції. Перелік операторів локалізації та їх визначень наведено нижче. Оператори локалізації можуть бути використані лише для нуклеотидів.

Синтаксис локалізації	Опис локалізації
`join(location,location, ... location)`	Вказані локалізації поєднуються (кінець-в-кінець) для утворення однієї безперервної послідовності.
`order(location,location, ... location)`	Елементи знаходяться в зазначеному порядку, але ніщо не передбачає доцільності з’єднання цих елементів.
`complement(location)`	Вказує, що функція розташована в ланцюзі, який є додатковим до діапазону послідовностей, визначеного дескриптором локалізації, коли зчитується у напрямку від 5'-кінця до 3'-кінця або у напрямку, що імітує напрямок від 5'-кінця до 3'-кінця.

68. Оператори локалізації "join" і "order" потребують зазначення принаймні двох дескрипторів локалізації, розділених комами. Дескриптори локалізації, що включають сайти між двома суміжними залишками, тобто x^y, не повинні використовуватись в локалізаціях "join" і "order". Використання оператора локалізації "join" передбачає, що залишки, які описуються дескрипторами локалізації, фізично контактують в біологічних процесах (наприклад, екзони, що сприяють функції кодуючої області).

69. Оператор локалізації "complement" може використовуватись разом з "join" або з "order" у межах однієї локалізації. У межах однієї локалізації не повинна використовуватись комбінація операторів локалізації "join" і "order".

70. Нижче наведено приклади локалізації функцій відповідно до параграфів 64-69:

(a) локалізації для нуклеотидних та амінокислотних послідовностей:

Приклад локалізації	Опис
`467`	Вказує на залишок 467 у послідовності.
`340..565`	Вказує на неперервний діапазон залишків, що поєднані залишками 340 і 565 і включають їх.
`<1`	Вказує на локалізацію функції перед першим залишком.
`<345..500`	Вказує на те, що точна нижня точка межі функції невідома. Локалізація починається із залишку, що передує залишку 345 і
	продовжується включно до залишку 500.

`<1..888`	Вказує на те, що функція починається до першого залишку послідовності і продовжується до залишку 888 включно.
`1..>888`	Вказує на те, що функція починається до першого залишку послідовності і продовжується після залишку 888.
`<1..>888`	Вказує на те, що функція починається перед першим залишком послідовності і продовжується після залишку 888.
`join(12..78,134..202)`	Вказує на те, що області 12-78 і 134-202 слід об'єднати для утворення однієї безперервної послідовності.

(b) локалізації виключно для нуклеотидних послідовностей:

Приклад локалізації	Опис
`123^124`	Вказує на сайт між залишками 123 і 124.
`join(12..78,134..202)`	Вказує на те, що області 12-78 і 134-202 слід об'єднати для утворення однієї безперервної послідовності.
`complement(34..126)`	Починається в нуклеотиді, що є комплементарним нуклеотиду 126, і закінчується на нуклеотиді, що є комплементарним нуклеотиду 34 (функція знаходиться у ланцюзі, що є комплементарним представленому ланцюгу).
`complement(join(2691..4571,4918..5163))`	Приєднує нуклеотиди з 2691 по 4571 та нуклеотиди з 4918 по 5163, потім стає комплементарним приєднаним сегментам (функція знаходиться у ланцюзі, що є комплементарним представленому ланцюгу).
`join(complement(4918..5163,complement(2691..4571))`	Комплементарні області з 4918 до 5163 та з 2691 до 4571, потім приєднуються до комплементарних сегментів (функція знаходиться у ланцюзі, що є комплементарним представленому ланцюгу).

(c) локалізації виключно для амінокислотних послідовностей:

Приклад локалізації	Опис
`340..565`	Вказує на те, що амінокислоти в положеннях 340 і 565 з’єднані за допомогою внутрішньоланцюгового зв’язку, коли використовується функція, яка вказує на внутрішньоланцюгову зшивку, наприклад "CROSSLINK" або "DISULFID".

71. Символи "<" і ">" у дескрипторі локалізації в екземплярі XML переліку послідовності повинні замінятися відповідними попередньо визначеними сутностями (див. параграф 41). Наприклад:

Локалізація функції "<1":
```
<INSDFeature_location>&lt;1</INSDFeature_location>
```

Локалізація функції "1..>888":
```
<INSDFeature_location>1..&gt;888</INSDFeature_location>
```

Кваліфікатори функцій

72. Кваліфікатори використовуються для надання інформації про функції додатково до інформації, що передається ключем функції та локалізацією функції. Існують три типи формату значення, призначені для різних видів інформації, яку несуть кваліфікатори, а саме:

(a) довільний текст (див. параграфи 85-87);

(b) стандартизована лексика або обчислювані значення (наприклад номер або дата); і

(c) послідовності.

73. У Розділі 6 Додатка I наведено виключний список кваліфікаторів та їхніх визначених форматів значень, за наявності, для кожного ключа функції нуклеотидної послідовності, а в Розділі 8 наведено виключний список кваліфікаторів та, за наявності, їхніх визначених форматів значень для кожного ключа функції амінокислотної послідовності.

74. Будь-яка послідовність, що охоплюється параграфом 7 і представляється як значення кваліфікатора, повинна включатися до переліку послідовностей окремо і з присвоєним власним ідентифікаційним номером послідовності (див. параграф 10).

Обов'язкові кваліфікатори функцій

75. Один обов'язковий ключ функції, тобто "source" для нуклеотидних послідовностей та амінокислотних послідовностей, потребує зазначення двох обов'язкових кваліфікаторів: "organism" і "mol_type". Деякі необов'язкові ключі функцій також вимагають обов'язкових кваліфікаторів.

Елементи кваліфікаторів

76. Елемент `INSDFeature_quals` містить один або більше елементів `INSDQualifier`. Кожний елемент `INSDQualifier` представляє один кваліфікатор і складається з таких трьох залежних елементів та одного необов’язкового атрибута:

Елемент/Атрибут	Опис	Обов'язковий/ Необов'язковий
`INSDQualifier_name`	Ім'я кваліфікатора (див. Додаток I, Розділи 6 і 8).	Обов'язковий

INSDQualifier_value	Значення кваліфікатора, за наявності, у визначеному форматі (див. Додаток I, Розділи 6 і 8) та складається символами, як зазначено у параграфі 40(b).	Обов'язковий, якщо визначений (див. Додаток I, Розділи 6 і 8)
NonEnglishQualifier_value	Значення кваліфікатора, за наявності, у визначеному форматі (див. Додаток I, Розділи 6 і 8) та складається символами, як зазначено у параграфі 40(a).	Обов’язковий, якщо визначений (див. параграф 87 та Додаток I, Розділи 6 і 8)
id	Кваліфікатор зі значенням залежного від мови довільного тексту можна однозначно ідентифікувати за допомогою необов'язкового XML-атрибута 'id' в елементі INSDQualifier (див. параграф 87(d)). Значення атрибута 'id' повинно починатися з літери 'q' і продовжуватися будь-яким натуральним числом. Значення атрибута 'id' повинне бути унікальним для одного елемента INSDQualifier, тобто значення атрибуту повинно бути використано тільки один раз у файлі переліку послідовності.	Необов'язковий

77. Кваліфікатор організмів, тобто "organism" для нуклеотидних послідовностей (див. Розділ 6 Додатка I) та "organism" для амінокислотних послідовностей (див. Розділ 8 Додатка I), повинен розкривати джерело, тобто один організм або походження послідовності. Назви організмів слід вибирати з таксономічної бази даних.

78. Якщо послідовність є природною, а рід і вид вихідного організму мають латинське позначення, це позначення повинне використовуватись як значення кваліфікатора. Переважну загальну назву англійською мовою можна зазначити, використовуючи кваліфікатор "note" для нуклеотидних послідовностей та амінокислотних послідовностей, проте вона не повинна зазначатися у значенні кваліфікатора "organism".

79. Нижче наведено приклади вихідного організму послідовності відповідно до параграфів 77 та 78:

Приклад 1: Джерело нуклеотидної послідовності

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

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

Приклад 2: Джерело амінокислотної послідовності

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

80. Якщо послідовність є природною, а вихідний організм має відомий рід латинською мовою, але його вид є невизначеним чи не вказаний, у значенні кваліфікатора "organism" повинен наводитися рід латинською мовою, за яким іде слідом "sp.". Наприклад:

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

81. Якщо послідовність є природною, проте рід і вид організму латинською мовою невідомі, у значенні кваліфікатора "organism" повинно бути зазначено "unidentified". Будь-яка відома таксономічна інформація має бути позначена в кваліфікаторі "note" для нуклеотидних послідовностей і в кваліфікаторі "note" для амінокислотних організмів. Наприклад:

```
<INSDQualifier_name>organism</INSDQualifier_name>
```

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

82. Якщо послідовність є природною, а рід і вид вихідного організму не мають позначення латинською мовою, наприклад вірус, у значенні кваліфікатора "organism" повинна використовуватись інша допустима наукова назва (наприклад, "Canine adenovirus type 2"). Наприклад:

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

83. Якщо послідовність не є природною, у значенні кваліфікатора "organism" повинно бути зазначено "synthetic construct". Додаткову інформацію стосовно того, як послідовність була отримана, можна зазначати з використанням кваліфікатора "note" для нуклеотидних послідовностей та кваліфікаторі "note" для амінокислотних послідовностей. Наприклад:

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

84. Кваліфікатор "mol_type" для нуклеотидних послідовностей (див. Додаток I, Розділ 6) та кваліфікатор "mol_type" для амінокислотних послідовностей (див. Додаток I, Розділ 8) повинні розкривати тип молекули, представленої у послідовності. Ці кваліфікатори відрізняються від елемента `INSDSeq_moltype`, зазначеного у параграфі 54:

(a) Для нуклеотидної послідовності, значення кваліфікатора "mol_type" повинне бути одним із таких: "genomic DNA", "genomic RNA", "Mrna", "tRNA", "rRNA", "other RNA", "other DNA", "transcribed RNA", "viral cRNA", "unassigned DNA" або "unassigned RNA". Якщо послідовність не є природною, тобто значенням кваліфікатора "organism" є "synthetic construct", у значенні кваліфікатора "mol_type" повинно бути зазначено "other RNA" або "other DNA";

(b) Для амінокислотних послідовностей значенням кваліфікатора "mol_type" є "protein".

Довільний текст

85. Довільний текст, як зазначено у параграфі 3 - це один із типів формату значення для певних кваліфікаторів, що представляється у формі описової текстової фрази або іншого зазначеного формату (як зазначено у Додатку I).

86. Використання довільного тексту повинно обмежуватися кількома короткими термінами, необхідними для розуміння характеристики послідовності. Довільний текст для кожного кваліфікатора крім кваліфікатора "translation" не повинен перевищувати 1000 символів.

87. Залежний від мови довільний текст, як зазначено у параграфі 3, є визначенням довільного тексту певних кваліфікаторів, що залежать від мови, оскільки це може вимагати перекладу для міжнародних, національних або регіональних процедур. Кваліфікатори для нуклеотидних послідовностей із залежним від мови довільним текстом визначені у Додатку I, Розділ 6, таблиця 5. Кваліфікатори для амінокислотних послідовностей із залежним від мови довільним текстом визначені у Додатку I, Розділ 8, таблиця 6.

(a) Залежний від мови довільний текст повинен бути представлений в елементі `INSDQualifier_value` англійською мовою або в елементі `NonEnglishQualifier_value` іншою мовою, або в обох елементах. Варто звернути увагу, що якщо назва організму – це латинська назва роду або виду, переклад не потрібен. Технічні терміни та власні назви, що походять не від англійських слів, що вживаються в міжнародному масштабі, вважаються англійськими з метою визначення елемента `INSDQualifier_value` (наприклад, "in vitro", "in vivo").

(b) Якщо елемент `NonEnglishQualifier_value` присутній у списку послідовностей, відповідний код мови (див. посилання в параграфі 9 на ISO 539-1:2002), повинен бути вказаний в кореневому елементі атрибуту `nonEnglishFreeTextLanguageCode` (див параграф 43). Всі елементи `NonEnglishQualifier_value` в одному переліку послідовностей повинні мати значення у мові, зазначеній у атрибуті `nonEnglishFreeTextLanguageCode`. Елемент `NonEnglishQualifier_value` дозволяється лише для кваліфікаторів, які мають формат залежного від мови довільного тексту.

(c) Коли елементи `NonEnglishQualifier_value` та `INSDQualifier_value` присутні в одному кваліфікаторі, інформація, що міститься у двох елементах, повинна бути еквівалентною. Одна з наступних умов повинна виконуватись: елемент `NonEnglishQualifier_value` містить переклад значення `INSDQualifier_value`; або, елемент `INSDQualifier_value` містить переклад значення `NonEnglishQualifier_value`; або, обидва елементи містять переклад значення кваліфікатора з мови, вказаної в атрибуті `originalFreeTextLanguageCode` (див. параграф 43).

(d) Для кваліфікаторів із залежним від мови значенням довільного тексту, елемент `INSDQualifier` може містити необов’язковий атрибут `id`. Значення цього атрибута повинно бути у форматі "q", за яким слід додати ціле число, наприклад "q23", і має бути унікальним для одного елемента `INSDQualifier`, тобто значення атрибута має використовуватись лише один раз у файлі переліку послідовностей.

88. Наступні приклади ілюструють подання залежного від мови довільного тексту, як

зазначалось в параграфі 87.

Приклад 1: залежний від мови довільний текст в елементі `INSDQualifier_value`:

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

Приклад 2: залежний від мови довільний текст в елементах INSDQualifier_value та NonEnglishQualifier_value:

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

Приклад 3: залежний від мови довільний текст у елементі NonEnglishQualifier_value:

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

Кодуючі послідовності

89. Ключ функції "CDS" може використовуватись для ідентифікації кодуючих послідовностей, тобто послідовностей нуклеотидів, що відповідають послідовностям амінокислот у білку або у стоп-кодоні. Локалізація функції "CDS" в обов’язковому елементі `INSDFeature_location` повинна включати стоп-кодон.

90. Кваліфікатори "transl_table" і "translation" можуть використовуватися з ключем функції "CDS" (див. Додаток I). Якщо кваліфікатор "transl_table" не використовується, передбачається використання Таблиці стандартних кодів (див. Додаток I, Розділ 9, Таблиця 7).

91. Кваліфікатор "transl_except" повинен використовуватись з ключем функції "CDS" та кваліфікатором "translation" для ідентифікації кодону, який кодує пірролізин або селеноцистеїн.

92. Амінокислотна послідовність, закодована кодуючою послідовністю та розкрита в кваліфікаторі "translation", що охоплюється параграфом 7, повинна включатися до переліку послідовностей, а також їй повинен бути присвоєний власний ідентифікаційний номер послідовності. Ідентифікаційний номер послідовності, присвоєний амінокислотній послідовності, повинен наводитися як значення в кваліфікаторі "protein_id" з ключем функції "CDS". Кваліфікатор "organism" ключа функції "source" для амінокислотної послідовності повинен бути ідентичним такому ж кваліфікатору його кодуючої послідовності. Наприклад:

```
<INSDFeature>
    <INSDFeature_key>CDS</INSDFeature_key>
    <INSDFeature_location>1..507</INSDFeature_location>
    <INSDFeature_quals>
        <INSDQualifier>
            <INSDQualifier_name>transl_table</INSDQualifier_name>
           <INSDQualifier_value>11</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
            <INSDQualifier_name>translation</INSDQualifier_name>
            <INSDQualifier_value>MLVHLERTTIMFDFSSLINLPLIWGLLIAIAVLLYILMDGF
            DLGIGILLPFAPSDKCRDHMISSIAPFWDGNETWLVLGGGGLFAAFPLAYSILMPAFYIPII
            IMLLGLIVRGVSFEFRFKAEGKYRRLWDYAFHFGSLGAAFCQGMILGAFIHGVEVNGRNFSG
            GQLM</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
            <INSDQualifier_name>protein_id</INSDQualifier_name>
            <INSDQualifier_value>89</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_quals>
</INSDFeature>
```

Варіанти

93. Первинна послідовність і будь-який варіант цієї послідовності, що розкриваються шляхом перерахування їх залишків і охоплюються параграфом 7, повинні бути включені до переліку послідовностей та їм має бути присвоєно власний ідентифікаційний номер послідовності.

94. Будь-який варіант послідовності, що розкривається однією послідовністю з переліком альтернативних залишків в одному або більше положеннях, повинен бути включений до переліку послідовностей та представлений однією послідовністю, в якій перелік альтернативних залишків представлений найбільш обмежуючим багатозначним символом (див. параграфи 15 та 27).

95. Будь-який варіант послідовності, що розкривається лише шляхом посилання на видалення, вставку (-и), або заміщення у первинній послідовності в переліку послідовностей, слід включити до переліку послідовностей. Якщо передбачено в переліку послідовностей, такий варіант послідовності:

(a) може бути представлений шляхом анотування первинної послідовності, яка містить

варіант (варіанти) в одній локалізації або декількох окремих локалізаціях і частота цих варіантів є незалежною;

(b) має бути представлений як окрема послідовність з присвоєним власним ідентифікаційним номером послідовності, яка містить варіанти в декількох окремих локалізаціях і частота цих варіантів є незалежною; і

(c) повинен бути представлений як окрема послідовність з присвоєним власним ідентифікаційним номером послідовності, яка містить включену або заміщену послідовність, що містить більше 1000 залишків (див. параграф 86).

96. У таблиці нижче наведено способи належного використання кваліфікаторів і ключів функцій для варіантів послідовностей нуклеїнової кислоти та амінокислоти:

Тип Послідовності	Ключ функції	Кваліфікатор	Застосування
Нуклеїнова кислота	variation	replace або note	Природні мутації та поліморфізми, наприклад алелі, ПДРФ.
Нуклеїнова кислота	misc_difference	replace або note	Мінливість, закладена штучно, наприклад генетичною маніпуляцією або хімічним синтезом.
Амінокислота	VAR_SEQ	note	Варіант, отриманий шляхом альтернативного сплайсингу, використання альтернативного промотора, альтернативної ініціації та рибосомного фреймшифтингу.
Амінокислота	VARIANT	note	Будь-який тип варіанту, для якого ключ функції "VAR_SEQ" не застосовується.

97. Тлумачення послідовності для певного варіанта повинна містити ключ функції та кваліфікатор, як зазначено у таблиці вище, а також локалізацію функції. Значення кваліфікатора "replace" має бути лише окремий альтернативний нуклеотид або нуклеотидна послідовність з використанням символів, визначених у Розділі 1, Таблиці 1, або порожнє. Список альтернативних залишків може наводитись як значення в кваліфікаторі "note".Зокрема, список альтернативних амінокислот повинен наводитись як значення в кваліфікаторі "note", де "X" використовується в послідовності та представляє значення відмінне від "будь-якого із символів ‘A’, ‘R’, ‘N’, ‘D’, ‘C’, ‘Q’, ‘E’, ‘G’, ‘H’, ‘I’, ‘L’, ‘K’, ‘M’, ‘F’, ‘P’, ‘O’, ‘S’, ‘U’, ‘T’, ‘W’, ‘Y’, або ‘V’"(див. параграф 27). Видалення повинне представлятися порожнім значенням кваліфікатора для кваліфікатора "replace" чи зазначенням в кваліфікаторі "note", що залишок може бути видалений. Вставлений (-і) або заміщений (-і) залишок (-ки) повинен (-ні) бути передбачений (-ні) в кваліфікаторі "replace" або "note". Форматом значення для кваліфікаторів "replace" та "note" є довільний текст, що не повинен перевищувати 1000 символів, як зазначено у параграфі 86. Див. параграф 100 стосовно послідовностей, охоплених параграфом 7, які представляються як вставка або як заміщення у значенні кваліфікатора.

98. Якщо доцільно, символи, визначені у Додатку I (див. Таблиці 1-4 у Розділах 1-4 відповідно), слід використовувати для представлення варіантів залишків. Для кваліфікатора "note", якщо варіант залишку є модифікованим залишком, не визначеним у Таблиці 2 або Таблиці 4 Додатка I, у значенні кваліфікатора повинна зазначатися повна, нескорочена назва модифікованого залишку. Модифіковані залишки повинні бути описані нижче у таблиці функцій, як наведено у параграфі 17 або 30.

99. Нижче наведено приклади представлення варіантів відповідно до параграфів 95-98:

Приклад 1: Ключ функції "misc_difference" для переліку альтернативних нуклеотидів. "n" в положенні 53 послідовності може бути одним з п’яти альтернативних нуклеотидів.

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

Приклад 2: Ключ функції "misc_difference" для видалення з нуклеотидної послідовності. Видалення нуклеотиду в положенні 413 послідовності.

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

Приклад 3: Ключ функції "misc_difference" для вставки до нуклеотидної послідовності. Включення послідовності "atgccaaatat" між положеннями 100 і 101 первинної послідовності.

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

Приклад 4: Ключ функції "variation" для заміни в нуклеотидній послідовності. Заміна нуклеотида в положенні 413 послідовності цитозином.

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

Приклад 5: Ключ функції "VARIANT" для заміни в амінокислотній послідовності.

Амінокислоту в положенні 100 послідовності можна замінити на `I, A, F, Y, aIle, MeIle`, або `Nle`.

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

Приклад 6: Ключ функції "VARIANT" для заміни в амінокислотній послідовності. Амінокислоту в положенні 100 послідовності можна замінити будь-якою амінокислотою окрім Lys, Arg або His.

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

100. Послідовність, що охоплюється параграфом 7, яка наводиться як вставка або заміщення у значенні кваліфікатора для тлумачення первинної послідовності, також повинна бути включена до переліку послідовностей, та їй повинен бути присвоєний власний ідентифікаційний номер.

[Додаток I наведено далі]