

## СТАНДАРТ ST.26

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

*Версія 1.7*

*Редакція, схвалена Комітетом зі стандартів VOIB (КСВ)  
на його одинадцятій сесії 8 грудня 2023 року*

*Редакційні зауваження підготовлені Міжнародним бюро*

*На одинадцятій сесії Комітету зі стандартів VOIB було прийнято рішення, що  
версія 1.7 стандарту VOIB ST.26 набуде чинності 1 липня 2024 року. Тим часом  
має використовуватися версія 1.6 стандарту VOIB ST.26.*

## ЗМІСТ

|  |    |
|--|----|
| ВСТУП.....                                       | 3  |
| ВИЗНАЧЕННЯ.....                                  | 3  |
| СФЕРА ЗАСТОСУВАННЯ.....                          | 5  |
| ПОСИЛАННЯ.....                                   | 6  |
| ПРЕДСТАВЛЕННЯ ПОСЛІДОВНОСТЕЙ .....               | 7  |
| <i>Нуклеотидні послідовності</i> .....           | 7  |
| <i>Амінокислотні послідовності</i> .....         | 11 |
| <i>Представлення окремих випадків</i> .....      | 13 |
| СТРУКТУРА ПЕРЕЛІКУ ПОСЛІДОВНОСТЕЙ В XML.....     | 13 |
| <i>Кореневий елемент</i> .....                   | 15 |
| <i>Частина із загальною інформацією</i> .....    | 16 |
| <i>Частина з даними щодо послідовності</i> ..... | 20 |
| <i>Таблиця функцій</i> .....                     | 23 |
| <i>Ключі функцій</i> .....                       | 23 |
| <i>Обов'язкові ключі функцій</i> .....           | 23 |
| <i>Локалізація функцій</i> .....                 | 23 |
| <i>Кваліфікатори функцій</i> .....               | 27 |
| <i>Обов'язкові кваліфікатори функцій</i> .....   | 27 |
| <i>Елементи кваліфікаторів</i> .....             | 27 |
| <i>Довільний текст</i> .....                     | 31 |
| <i>Кодуючі послідовності</i> .....               | 32 |
| <i>Варіанти</i> .....                            | 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 1.7*

*Редакція, схвалена Комітетом зі стандартів ВОІВ (КСВ)  
на його одинадцятій сесії 8 грудня 2023 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 модифіковані або 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: ННННННQ 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 5'-монофосфату, дезоксигуанозин 5'-монофосфату, дезоксицитидин 5'-монофосфату, дезокситимідин 5'-монофосфату, аденозин 5'-монофосфату, гуанозин 5'-монофосфату, цитидин 5'-монофосфату, або уридин 5'-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 нуклеїною 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 окремих 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 (-і) 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') фосфодієфірним 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:

Проект міжнародного співробітництва баз даних нуклеотидних послідовностей (INSDC)

<http://www.insdc.org/>;

Міжнародний стандарт [ISO 639-1:2002](https://www.iso.org/standard/55861.html)

"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 VOIB [ST.2](#)

"Стандартний спосіб представлення календарних дат з використанням григоріанського календаря";

Стандарт VOIB [ST.3](#)

"Рекомендований стандарт стосовно двобуквених кодів для представлення держав, інших адміністративних одиниць та міжурядових організацій";

СТАНДАРТ VOIB [ST.16](#)

"Рекомендовані стандартні коди для ідентифікації різних видів патентних документів";

СТАНДАРТ VOIB [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 окремих 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_qual>
    <INSDFqualifier>
```



```
        <INSDQualifier_name>mod_base</INSDQualifier_name>
        <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
</INSDFeature_qual>
</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_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>xanthine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

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

```

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

```

21. 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_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>isoguanine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

*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 аспарагіною 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 І (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_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>3Hyp</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

#### Приклад 2: Не пост-трансляційно модифікована амінокислота

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

#### Приклад 3: D-амінокислота

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

32. Будь-яка "невідомо" амінокислота повинна бути представлена у послідовності символом "X". "Невідомо" 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_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>A or V</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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 безперервних 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:

- (а) частина із загальною інформацією про патентну заявку, до якої відноситься перелік послідовностей; 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<sup>1</sup> 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 |
|-----------------------|--------------------|
| <                     | &lt;               |
| >                     | &gt;               |
| &                     | &amp;              |
| "                     | &quot;             |
| '                     | &apos;             |

<sup>1</sup> 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.

Див. приклад у параграфі 71. Єдиними дозволеними посиланнями на сутності символів є 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/<br>Необов'язковий  |
|---|--|--|
| <code>dtdVersion</code>                     | Версія DTD, використана для створення цього файлу, у форматі "V#_#", наприклад "V1_2".   | Обов'язковий   |
| <code>fileName</code>                       | Ім'я файлу, що містить перелік послідовностей.   | Необов'язковий   |
| <code>softwareName</code>                   | Назва програмного забезпечення, в якому було створено цей файл.  | Необов'язковий   |
| <code>softwareVersion</code>                | Версія програмного забезпечення, в якому було створено цей файл.   | Необов'язковий   |
| <code>productionDate</code>                 | Дата створення файлу, що містить перелік послідовностей (у форматі "RRRR-ММ-ДД").  | Необов'язковий   |
| <code>originalFreeTextLanguageCode</code>   | 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   |
| <code>nonEnglishFreeTextLanguageCode</code> | Код мови (див. посилання у параграфі 9 на ISO 639-1:2020) для елементів <code>nonEnglishFreeTextLanguageCode</code>  | Обов'язковий, коли в списку послідовностей присутній елемент <code>nonEnglishFreeTextLanguageCode</code> |

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

```
<ST26SequenceListing dtdVersion="V1_3" fileName="US11-405455-
SEQ1.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/<br>Необов'язковий   |
|---|--|---|
| <p>ApplicationIdentification</p> <p>Елемент ApplicationIdentification складається з:</p> <p>IPOfficeCode</p> <p>ApplicationNumberText</p> | <p>Ідентифікаційні дані заявки, до якої відноситься перелік послідовностей</p> <p>Код відомства подання за стандартом ВОІВ <a href="#">ST.3</a></p> <p>Номер заявки, наданий відомством, що подало заявку (наприклад, РСТ/ІВ2013/099999).</p>                            | <p>Обов'язковий, якщо перелік послідовностей подається після присвоєння номера заявки</p> <p>Обов'язковий</p> <p>Обов'язковий</p> |
| FilingDate  | <p>Дата подання патентної заявки, стосовно якої подається перелік послідовностей (у форматі "РРРР-ММ-ДД" за стандартом ВОІВ <a href="#">ST.2</a>, де "РРРР" – календарний рік, ММ – календарний місяць, ДД – день у межах календарного місяця, наприклад 2015-01-31)</p> | Обов'язковий, якщо перелік послідовностей подається після встановлення дати подання   |
| ApplicantFileReference  | <p>Унікальний ідентифікаційний номер, присвоєний заявником для ідентифікації певної заявки, наведений символами відповідно до параграфу 40(b)</p>  | Обов'язковий, якщо перелік послідовностей подається до присвоєння номера заявки; в інших випадках – необов'язковий                |
| EarliestPriorityApplication Identification  | <p>Ідентифікація заявки з більш раннім пріоритетом (також містить елементи IPOfficeCode, ApplicationNumberText і FilingDate, див. елементи ApplicationIdentification)</p>  | Обов'язковий, якщо заявлено пріоритет   |



| Елемент            | Опис  | Обов'язковий/<br>Необов'язковий   |
|--------------------|---|---|
| ApplicantName      | Ім'я першого вказаного заявника, представлене символами, визначеними у параграфі 40 (а). Цей елемент включає обов'язковий атрибут <code>languageCode</code> , як зазначено у параграфі 47.  | Обов'язковий  |
| ApplicantNameLatin | Якщо <code>ApplicantName</code> 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 <code>ApplicantName</code> містить інші символи ніж латинські. |
| InventorName       | Ім'я першого вказаного винахідника, представлене символами, визначеними у параграфі 40(а). Цей елемент включає обов'язковий атрибут <code>languageCode</code> , як зазначено у параграфі 47.  | Необов'язковий  |
| InventorNameLatin  | Якщо <code>InventorName</code> 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(а), мовою оригіналу. Переклад назви винаходу додатковими мовами може наводитися символами як зазначено у параграфі 40(а) з використанням додаткових елементів <code>InventionTitle</code> . Цей елемент включає обов'язковий атрибут <code>languageCode</code> як зазначено у параграфі 48. Назва | Обов'язковий для мови подання.<br>Необов'язковий для додаткових мов.                      |

| Елемент               | Опис   | Обов'язковий/<br>Необов'язковий |
|-----------------------|--|---------------------------------|
|                       | винаходу має бути від двох до семи слів.   |                                 |
| 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 і дати 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_SEQ1.xml" softwareName="WIPO Sequence"
softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="en"
nonEnglishFreeTextLanguageCode="ja">
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2013/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode="en">Keiko Nakamura</InventorName>
  <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
  <SequenceTotalQuantity>9</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="2"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="3"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="4"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="5"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="6"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="7"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="8"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>
```

\* {...} представляє відповідну інформацію для кожної послідовності, не представлена 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_SEQ1.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 представлена 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 з таких додаткових залежних елементів:

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

| Елемент          | Опис          | Обов'язковий/Не включається |                                  |
|------------------|---------------|-----------------------------|----------------------------------|
|                  |               | Послідовності               | Навмисно пропущені послідовності |
| 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 "АА". (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 конструктор"*), 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_qual>
        <INSDQualifier>
          <INSDQualifier_name>organism</INSDQualifier_name>
          <INSDQualifier_value>syntheti
c
construct</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
          <INSDQualifier_name>mol_type</INSDQualifier_name>
          <INSDQualifier_value>other DNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_qual>
    </INSDFeature>
  </INSDSeq_feature-table>
</INSDSeq>
```

```

</INSDFeature>
<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>1..60</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>DNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>61..120</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>RNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>cgaccacgcgctccgaggaaccaaccatcacgtttgaggacttcgtgaaggaatt
ggataatacccgctccctacaaaatggcgagcgccgactcattgtcctcgtaccgctcgagcggc</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 відповідно до параграфу 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/<br>Необов'язковий   |
|-----------------------------------|---|---|
| <code>INSDFeature_key</code>      | Слово чи аббревіатура, що ідентифікує функцію     | Обов'язковий  |
| <code>INSDFeature_location</code> | Ділянка послідовності, що відповідає функції      | Обов'язковий  |
| <code>INSDFeature_qual</code>     | 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 | $<x$<br>$>x$<br>$<x..y$<br>$x..>y$<br>$<x..>y$ | 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 ( $\wedge$ ). 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 вказує 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  |
|--|---|
| <code>join(location, location, ... location)</code>  | Вказані локалізації поєднуються (кінець-в-кінець) для утворення однієї безперервної послідовності.  |
| <code>order(location, location, ... location)</code> | Елементи знаходяться в зазначеному порядку, але ніщо не передбачає доцільності з'єднання цих елементів.   |
| <code>complement(location)</code>                    | 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  $x^u$ , 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:

(а) 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.                                     |

| Приклад локалізації   | Опис   |
|-----------------------|--|
| <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_qual` 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/<br>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 атрибута 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_qual>
    <INSDQualifier>
      <INSDQualifier_name>organism</INSDQualifier_name>
      <INSDQualifier_value>Solanum
        lycopersicum</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>common name: tomato</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>mol_type</INSDQualifier_name>
      <INSDQualifier_value>genomic DNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>

```

## Приклад 2: Джерело амінокислотної послідовності

```

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

```

80. 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 зазначити 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_qual>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>synthetic construct</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>protein</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>note</INSDQualifier_name>
        <INSDQualifier_value>synthetic peptide used as assay for
        antibodies</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>
</INSDSeq_feature-table>
```

84. Кваліфікатор "mol\_type" для нуклеотидних послідовностей (див. Додаток І, Розділ 6) та кваліфікатор "mol\_type" для амінокислотних послідовностей (див. Додаток І, 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 кореновому 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_qual>
  <INSDQualifier id="q1">
    <INSDQualifier_name>function</INSDQualifier_name>
    <INSDQualifier_value>binds      to      regulatory
                                protein
    Est3</INSDQualifier_value>
  </INSDQualifier>
</INSDFeature_qual>
</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_qual>
    <INSDQualifier id="q45">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>cleaves
        carbohydratechain</INSDQualifier_value>
      <NonEnglishQualifier_value>clive la chaîne glucidique
      </NonEnglishQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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_qual>
    <INSDQualifier id="q1034">
      <INSDQualifier_name>note</INSDQualifier_name>
      <NonEnglishQualifier_value>clive la chaîne glucidique
      </NonEnglishQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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_qual>
    <INSDQualifier>
      <INSDQualifier_name>transl_table</INSDQualifier_name>
      <INSDQualifier_value>11</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>translation</INSDQualifier_name>
      <INSDQualifier_value>MLVHLERTTIMFDFSSLINLPLIWGLLIAIAVLLYILMDGF
        DLGIGILLPFAPSDKCRDHMISSIAFPWDGNETWLVLGGGGLFAAFPLAYSILMPAFYIPII
        IMLLGLIVRGVSFEFRFKAEGKYRRLWDYAFHFGLGAAFCQGMILGAFIHGVEVNGRNFSG
        GQLM</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>protein_id</INSDQualifier_name>
      <INSDQualifier_value>89</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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 (-), або заміщення у первинній послідовності в переліку послідовностей, слід включити до переліку послідовностей. Якщо передбачено в переліку послідовностей, такий варіант послідовності:

- а) 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<br>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 (-i) або заміщений (-i) 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 зазначитися 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_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value> w, cmnm5s2u, mam5u, mcm5s2u, or p
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>53</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>cmnm5s2u, mam5u, mcm5s2u, or
      p</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

**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_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value></INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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_qual>
```

```

    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value>atgccaaatat</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value>c</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</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_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>I, A, F, Y, aIle, MeIle, or Nle</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>MOD_RES</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>aIle, MeIle, or Nle</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

**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_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>

```

---

```
<INSDQualifier_value>not K, R, or H</INSDQualifier_value>  
</INSDQualifier>  
</INSDFeature_qual>  
</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]

**ДОДАТОК І****СТАНДАРТИЗОВАНА ЛЕКСИКА**

Версія 1.7

*Редакція, схвалена Комітетом зі стандартів ВОІВ (КСВ)  
на його одинадцятій сесії 8 грудня 2023 року*

**ЗМІСТ**

|   |    |
|---|----|
| РОЗДІЛ 1: ПЕРЕЛІК НУКЛЕОТИДІВ .....                             | 39 |
| РОЗДІЛ 2: ПЕРЕЛІК МОДИФІКОВАНИХ НУКЛЕОТИДІВ .....               | 40 |
| РОЗДІЛ 3: ПЕРЕЛІК АМІНОКИСЛОТ .....                             | 42 |
| РОЗДІЛ 4: ПЕРЕЛІК МОДИФІКОВАНИХ АМІНОКИСЛОТ .....               | 43 |
| РОЗДІЛ 5: КЛЮЧІ ФУНКЦІЙ ДЛЯ НУКЛЕОТИДНИХ ПОСЛІДОВНОСТЕЙ .....   | 44 |
| РОЗДІЛ 6: КВАЛІФІКАТОРИ ДЛЯ НУКЛЕОТИДНИХ ПОСЛІДОВНОСТЕЙ .....   | 60 |
| РОЗДІЛ 7: КЛЮЧІ ФУНКЦІЙ ДЛЯ АМІНОКИСЛОТНИХ ПОСЛІДОВНОСТЕЙ ..... | 84 |
| РОЗДІЛ 8: КВАЛІФІКАТОРИ ДЛЯ АМІНОКИСЛОТНИХ ПОСЛІДОВНОСТЕЙ ..... | 91 |
| РОЗДІЛ 9: ТАБЛИЦІ ГЕНЕТИЧНОГО КОДУ .....                        | 93 |

## РОЗДІЛ 1: ПЕРЕЛІК НУКЛЕОТИДІВ

Коди для нуклеотидних основ, які слід використовувати в переліках послідовностей, представлені у Таблиці 1. Символ "t" тлумачиться як тимін у ДНК і як урацил у РНК, якщо використовується без додаткового опису. Якщо доцільно використати неоднозначний символ (що представляє собою дві або більше основ у варіанті), слід скористатися найбільш обмежувальним символом. Наприклад, якщо основа у певному положенні може бути "a або g", слід використовувати "r", а не "n". Символ "n" тлумачиться як "a або c, або g, або t/u", якщо використовується без додаткового опису.

Таблиця 1: Перелік нуклеотидів

| Символ | Нуклеотид                                    |
|--------|--|
| a      | аденін (adenine)                             |
| c      | цитозин (cytosine)                           |
| g      | гуанін (guanine)                             |
| t      | тимін (thymine) в ДНК/урацил в РНК (t/u)     |
| m      | a або c                                      |
| r      | a або g                                      |
| w      | a або t/u                                    |
| s      | c або g                                      |
| y      | c або t/u                                    |
| k      | g або t/u                                    |
| v      | a або c, або g; не t/u                       |
| h      | a або c, або t/u; не g                       |
| d      | a або g, або t/u; не c                       |
| b      | c або g, або t/u; не a                       |
| n      | a або c, або g, або t/u; невідомий або інший |

## РОЗДІЛ 2: ПЕРЕЛІК МОДИФІКОВАНИХ НУКЛЕОТИДІВ

Скорочення, наведені у Таблиці 2, є виключними допустимими значеннями для кваліфікатора "mod\_base". Якщо певний модифікований нуклеотид відсутній у таблиці нижче, його значенням повинне бути "OTHER". Якщо значенням є "OTHER", в кваліфікаторі "note" повинна наводитися повна нескорочена назва модифікованої основи. Скорочення, наведені у Таблиці 2, не повинні зазначатися безпосередньо у послідовності.

Таблиця 2: Перелік модифікованих нуклеотидів

| Скорочення | Модифікований нуклеотид               |
|------------|---------------------------------------|
| ac4c       | 4-ацетилцитидин                       |
| chm5u      | 5-(карбоксигідроксиметил)уридин       |
| cm         | 2'-О-метилцитидин                     |
| cmnm5s2u   | 5-карбоксиметиламінометил-2-тіоуридин |
| cmnm5u     | 5-карбоксиметиламінометилуридин       |
| dhu        | дигідроуридин                         |
| fm         | 2'-О-метилпсевдоуридин                |
| gal q      | бета,D-галактозилквеуозин             |
| gm         | 2'-О-метилгуанозин                    |
| i          | інозин                                |
| i6a        | N6-ізопентениладенозин                |
| m1a        | 1-метиладенозин                       |
| m1f        | 1-метилпсевдоуридин                   |
| m1g        | 1-метилгуанозин                       |
| m1i        | 1-метилінозин                         |
| m22g       | 2,2-диметилгуанозин                   |
| m2a        | 2-метиладенозин                       |
| m2g        | 2-метилгуанозин                       |
| m3c        | 3-метилцитидин                        |
| m4c        | N4-метилцистозин                      |
| m5c        | 5-метилцитидин                        |
| m6a        | N6-метиладенозин                      |
| m7g        | 7-метилгуанозин                       |
| mam5u      | 5-метиламінометилуридин               |
| mam5s2u    | 5- метиламінометил-2-тіоуридин        |
| man q      | бета-D-манозилквеуозин                |
| mcm5s2u    | 5-метоксикарбонілметил-2-тіоуридин    |
| mcm5u      | 5-метоксикарбонілметилуридин          |



| Скорочення | Модифікований нуклеотид   |
|------------|---|
| mo5u       | 5-метоксиуридин   |
| ms2i6a     | 2-метилтіо-N6-ізопентениладенозин                                 |
| ms2t6a     | N-((9-бета-D-рибофуранозил-2-метилтіопурин-6-їл)карбамоїл)треонін |
| mt6a       | N-((9-бета-D-рибофуранозилпурин-6-їл)N-метилкарбамоїл)треонін     |
| mv         | уридин-5-оксицтова кислота-метилефір                              |
| o5u        | уридин-5-оксицтова кислота (v)                                    |
| osyw       | вибутоксозин  |
| p          | псевдоуридин  |
| q          | квеуозин  |
| s2c        | 2-тіоцитидин  |
| s2t        | 5-метил-2-тіоуридин   |
| s2u        | 2-тіоуридин   |
| s4u        | 4-тіоуридин   |
| m5u        | 5- метилуридин  |
| t6a        | N-((9-бета-D-рибофуранозилпурин-6-їл)-карбамоїл)треонін           |
| tm         | 2'-O-метил-5-метилуридин  |
| um         | 2'-O-метилуридин  |
| yw         | вибутозин   |
| x          | 3-(3-аміно-3-карбокси-пропіл)уродин, (аср3)u                      |
| OTHER      | (вимагає зазначення кваліфікатора "note").                        |

## РОЗДІЛ 3: ПЕРЕЛІК АМІНОКИСЛОТ

Коди амінокислот для використання у послідовностях наведено у Таблиці 3. Якщо може бути використаний двозначний символ (для представлення двох або більше амінокислот у варіанті), слід використовувати найбільш обмежувальний символ. Наприклад, якщо амінокислота у певному положенні може бути як аспарагіноюю кислотою, так і аспарагіном, слід використовувати символ "B", а не "X". Символ "X" тлумачиться як "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", або "V", коли він використовується без додаткового опису.

Таблиця 3: Перелік амінокислот

| Символ | Амінокислота   |
|--------|--|
| A      | аланін   |
| R      | аргінін  |
| N      | аспарагін  |
| D      | аспарагінова кислота (аспартат)  |
| C      | цистеїн  |
| Q      | глутамін   |
| E      | глутамінова кислота (глутамат)   |
| G      | гліцин   |
| H      | гістидин   |
| I      | ізолейцин  |
| L      | лейцин   |
| K      | лізин  |
| M      | метіонін   |
| F      | фенілаланін  |
| P      | пролін   |
| O      | пірролізин   |
| S      | серин  |
| U      | селеноцистеїн  |
| T      | треонін  |
| W      | триптофан  |
| Y      | тирозин  |
| V      | валін  |
| B      | аспарагінова кислота або аспарагін   |
| Z      | глутамін або глутамінова кислота   |
| J      | лейцин або ізолейцин   |
| X      | A або R або N або D або C або Q або E або G або H або I або L або K або M або F або P або O або S або U або T або W або Y або V; "невідомо" або "інша" |

## РОЗДІЛ 4: ПЕРЕЛІК МОДИФІКОВАНИХ АМІНОКИСЛОТ

У Таблиці 4 наведено виключні допустимі скорочення для модифікованої або нестандартної амінокислоти в обов'язковому кваліфікаторі "NOTE" для ключів функцій "MOD\_RES" або "SITE". Значенням для кваліфікатора "NOTE" повинне бути, за можливості, скорочення з цієї таблиці або повна, нескорочена назва модифікованої амінокислоти. Скорочення (або повні назви), наведені у цій таблиці, не повинні зазначатися безпосередньо у послідовності.

Таблиця 4: Перелік модифікованих амінокислот

| Скорочення | Модифікована або нестандартна амінокислота   |
|------------|--|
| Aad        | 2-аміноадипінова кислота                     |
| bAad       | 3-аміноадипінова кислота                     |
| bAla       | бета-аланін, бета-амінопропіонова кислота    |
| Abu        | 2-аміномасляна кислота                       |
| 4Abu       | 4-аміномасляна кислота, піперидинова кислота |
| Asp        | 6-амінокапронова кислота                     |
| Ahe        | 2-аміногептанова кислота                     |
| Aib        | 2-аміноізомасляна кислота                    |
| bAib       | 3-аміноізомасляна кислота                    |
| Apm        | 2-амінопімелінова кислота                    |
| Dbu        | 2,4-діаміномасляна кислота                   |
| Des        | десмозин                                     |
| Dpm        | 2,2'-діамінопімелінова кислота               |
| Dpr        | 2,3-діамінопропіонова кислота                |
| EtGly      | N-етилгліцин                                 |
| EtAsn      | N-етиласпарагін                              |
| Hyl        | гідроксилізін                                |
| aHyl       | алло-гідроксилізін                           |
| 3Hyp       | 3-гідроксипролін                             |
| 4Hyp       | 4-гідроксипролін                             |
| Ide        | ізодесмозин                                  |
| alle       | алло-ізолейцин                               |
| MeGly      | N-метилгліцин, саркозин                      |
| Melle      | N-метилізолейцин                             |
| MeLys      | 6-N-метиллізін                               |
| MeVal      | N-метилвалін                                 |
| Nva        | норвалін                                     |
| Nle        | норлейцин                                    |
| Orn        | орнітин                                      |

## РОЗДІЛ 5: КЛЮЧІ ФУНКЦІЙ ДЛЯ НУКЛЕОТИДНИХ ПОСЛІДОВНОСТЕЙ

Цей розділ містить перелік допустимих ключів функцій для нуклеотидних послідовностей, а також обов'язкові та необов'язкові кваліфікатори. Ключі функцій наведені в алфавітному порядку. Ключі функцій можуть використовуватись як для ДНК, так і для РНК, якщо в графі "Тип молекули" не зазначено інше. Деякі ключі функцій можуть підходити для використання із штучними послідовностями на додачу до ключа функції "вид організму".

Ключі функцій в екземплярі XML переліку послідовностей повинні називатися так само як і в описі нижче під заголовком "Ключ функції", окрім ключів функцій "3'UTR" і "5'UTR". Див. "Примітки" в описі до ключів функцій "3'UTR" і "5'UTR".

|      |                             |   |
|------|-----------------------------|---|
| 5.1. | Ключ функції                | C_region  |
|      | Визначення                  | Константна область легких і важких ланцюгів імуноглобуліну, альфа-, бета- і гама-ланцюгів Т-клітинного рецептора; включає один або більше екзонів залежно від ланцюга   |
|      | Необов'язкові кваліфікатори | allele<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard name   |
|      | Вид організму               | еукаріоти   |
| 5.2. | Ключ функції                | CDS   |
|      | Визначення                  | Кодуюча послідовність; послідовність нуклеотидів, що відповідає послідовності амінокислот у білку (локалізація включає термінальний кодон); функція може включати концептуальну трансляцію амінокислоти   |
|      | Необов'язкові кваліфікатори | allele<br>codon_start<br>EC_number<br>exception<br>function<br>gene<br>gene_synonym<br>map<br>note<br>number<br>operon<br>product<br>protein_id<br>pseudo<br>pseudogene<br>ribosomal_slippage<br>standard_name<br>translation transl_<br>except transl_<br>table trans_<br>splicing |
|      | Примітки                    | Кваліфікатор "codon_start" має дійсне значення "1", "2" або "3", що вказує на зміщення, де розташований перший  |

|      |  |   |
|------|--|---|
|      |  | повний кодон кодуючої функції відносно першої основи цієї функції; кваліфікатор "transl_table" визначає використовувану таблицю генетичного коду, якщо вона відрізняється від стандартної або універсальної таблиці генетичного коду; винятки генетичного коду поза межами визначеної таблиці наводяться в кваліфікаторі "transl_except"; лише один з кваліфікаторів "translation", "pseudogene" або "pseudo" допустимий для використання з ключем функції "CDS"; коли використовується кваліфікатор "translation", кваліфікатор "protein_id" є обов'язковим, якщо продукт трансляції містить чотири і більше специфічно визначені амінокислоти |
| 5.3. | Ключ функції<br>Визначення<br>Необов'язкові кваліфікатори<br>Примітки      | centromere<br>Біологічно досліджувана область, ідентифікована як центромера і досліджена експериментальним шляхом<br>note<br>standard_name<br>Функція "centromere" описує інтервал ДНК, що відповідає області, де утримуються хроматиди і формується кінетохор  |
| 5.4. | Ключ функції<br>Визначення<br>Необов'язкові кваліфікатори<br>Тип молекули  | D-loop<br>Петля зміщення; область в межах мітохондріальної ДНК, на якій короткий проміжок РНК парується з одним ланцюгом ДНК, витісняючи оригінальний парний ланцюг ДНК з цієї області; також використовується для опису заміщення у реакції області одного ланцюга двоспіральної ДНК іншим ланцюгом, прискореного RecA-протеїном<br>allele<br>gene<br>gene_synonym<br>map<br>note<br>ДНК   |
| 5.5. | Ключ функції<br>Визначення<br>Необов'язкові кваліфікатори<br>Вид організму | D_segment<br>D-ділянка важкого ланцюга імуноглобуліну і бета-ланцюга Т-клітинного рецептора<br>allele<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name<br>еукаріоти   |
| 5.6. | Ключ функції<br>Визначення   | exon<br>Область генома, в якій закодована частина сплайсованої мРНК, рРНК або тРНК; може містити нетрансльовані ділянки на 5'-кінці (5'UTR), усі CDS і нетрансльовані ділянки на 3'-кінці (3'UTR)   |

|      |                             |   |
|------|-----------------------------|---|
|      | Необов'язкові кваліфікатори | allele<br>EC_number<br>function<br>gene<br>gene_synonym<br>map<br>note<br>number<br>product<br>pseudo<br>pseudogene<br>standard_name<br>trans_splicing  |
| 5.7. | Ключ функції                | gene  |
|      | Визначення                  | Область біологічної важливості, що визначається як ген і якій була присвоєна назва  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>pseudo<br>pseudogene<br>phenotype<br>standard_name<br>trans_splicing  |
|      | Примітки                    | Функція "gene" описує інтервал ДНК, що відповідає генетичній ознаці або генетичному фенотипу; функція по суті не зв'язана зі своїм положенням на кінцях; її мета – представляти область, де розташований ген. |
| 5.8. | Ключ функції                | iDNA  |
|      | Визначення                  | Проміжна ДНК; ДНК, що знищується у процесі будь-якої рекомбінації   |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>number<br>standard_name  |
|      | Тип молекули                | ДНК   |
|      | Примітки                    | Наприклад, у соматичній обробці імуноглобулінових генів.  |
| 5.9. | Ключ функції                | intron  |
|      | Визначення                  | Ділянка ДНК, що транскрибується, але видаляється з транскрипта шляхом сплайсингу послідовностей (екзонів) з усіх її сторін  |

|      |                             |   |
|------|-----------------------------|---|
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>number<br>pseudo<br>pseudogene<br>standard_name<br>trans_splicing  |
| 5.10 | Ключ функції                | J_segment   |
| .    | Визначення                  | З'єднувальна ділянка легких і важких ланцюгів імуноглобуліну, альфа-, бета- і гама-ланцюгів Т-клітинного рецептора  |
|      | Необов'язкові кваліфікатори | allele<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name   |
|      | Вид організму               | еукаріоти   |
| 5.11 | Ключ функції                | mat_peptide   |
| .    | Визначення                  | Кодуюча послідовність зрілого пептиду або білка; кодуюча послідовність для зрілого або кінцевого пептиду або білка, що є продуктом пост-трансляційної модифікації; локалізація не охоплює термінуючий кодон (на відміну від відповідної кодуючої послідовності) |
|      | Необов'язкові кваліфікатори | allele<br>EC_number<br>function<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name  |
| 5.12 | Ключ функції                | misc_binding  |
| .    | Визначення                  | Сайт нуклеїнової кислоти, що ковалентно або нековалентно зв'язує другий компонент, який не можна описати будь-яким іншим ключем зв'язування ("primer_bind" або "protein_bind")  |
|      | Обов'язкові кваліфікатори   | bound_moiety  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note   |
|      | Примітки                    | Слід зазначити, що регуляторний ключ функції та кваліфікатор "regulatory_class" зі значенням "ribosome_binding_site" повинні використовуватися для  |

## опису сайтів зв'язування рибосом

|      |                             |  |
|------|-----------------------------|--|
| 5.13 | Ключ функції                | misc_difference  |
| .    | Визначення                  | У цьому місці вибрана послідовність відрізняється від представленої послідовності і не може бути описана жодним іншим диференціальним ключем ("variation" або "modified_base")   |
|      | Необов'язкові кваліфікатори | allele<br>clone<br>compare<br>gene<br>gene_synonym<br>map<br>note<br>phenotype<br>replace<br>standard_name   |
|      | Примітки                    | Ключ функції "misc_difference" повинен використовуватись для опису мінливості, закладеної штучно, наприклад, шляхом генетичного маніпулювання або хімічного синтезування; для опису видалення, вставки або заміщення повинен використовуватись кваліфікатор "replace". Варіативний ключ функції повинен використовуватись для опису природної генетичної мінливості. |
| 5.14 | Ключ функції                | misc_feature   |
| .    | Визначення                  | Біологічно досліджувана область, яку не можна описати жодним іншим ключем функції; нова або рідкісна функція   |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note number<br>phenotype<br>product<br>pseudo<br>pseudogene<br>standard_name  |
|      | Примітки                    | Цей ключ не повинен використовуватись лише для простого позначення ділянки з метою надання зауважень до неї або використання її в локалізації іншої функції  |
| 5.15 | Ключ функції                | misc_recomb  |
| .    | Визначення                  | Сайт будь-якої узагальненої, сайт-специфічної події, або події реплікативної рекомбінації, на якій присутній розрив і відновлення двоспиральної ДНК, що не можуть бути описані іншими ключами рекомбінації або кваліфікаторами вихідного ключа (proviral)  |
|      | Необов'язкові кваліфікатори | allele<br>gene<br>gene_synonym<br>map<br>note<br>recombination_class<br>standard_name  |
|      | Тип молекули                | ДНК  |
| 5.16 | Ключ функції                | misc_RNA   |
| .    |                             |  |



|      |                             |  |
|------|-----------------------------|--|
|      | Визначення                  | Будь-який транскрипт або продукт РНК, який не можна визначити іншими ключами РНК (prim_transcript, precursor_RNA, mRNA, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA site, ncRNA, rRNA, tRNA) |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>pseudo<br>pseudogene<br>standard_name<br>trans_splicing  |
| 5.17 | Ключ функції                | misc_structure   |
| .    | Визначення                  | Будь-яка вторинна або третинна структура чи конформація, яку не можна описати іншими ключами структури (stem_loop або D-loop)  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>standard_name   |
| 5.18 | Ключ функції                | mobile_element   |
| .    | Визначення                  | Область генома, що містить мобільні елементи   |
|      | Обов'язкові кваліфікатори   | mobile_element_type  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>rpt_family<br>rpt_type<br>standard_name   |
| 5.19 | Ключ функції                | modified_base  |
| .    | Визначення                  | Вказаний нуклеотид є модифікованим нуклеотидом і має бути замінений на вказану молекулу (зазначену під кваліфікатором mod_base)  |
|      | Обов'язкові кваліфікатори   | mod_base   |
|      | Необов'язкові кваліфікатори | allele<br>frequency<br>gene<br>gene_synonym<br>map<br>note   |
|      | Примітки                    | Значення обов'язкового кваліфікатора "mod_base" обмежено визначеним словником скорочень для модифікованих основ, наведеним у Розділі 2 цього додатку.  |

|      |                             |  |
|------|-----------------------------|--|
| 5.20 | Ключ функції                | mRNA   |
| .    | Визначення                  | Матрична РНК; включає нетрансльовану область 5'-кінця ("5'UTR"), кодує послідовності ("CDS", "exon") і нетрансльовану область 3'-кінця ("3'UTR")   |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>pseudo<br>pseudogene<br>standard_name<br>trans_splicing  |
| 5.21 | Ключ функції                | ncRNA  |
| .    | Визначення                  | Білок-некодуєчий ген, окрім рибосомної РНК і транспортної РНК, функціональною молекулою якого є транскрипт РНК   |
|      | Обов'язкові кваліфікатори   | ncRNA_class  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>pseudo<br>pseudogene<br>standard_name<br>trans_splicing  |
|      | Примітки                    | Функція "ncRNA" не повинна використовуватися для опису рибосомної і транспортної РНК, для яких повинні використовуватись ключі функцій "rRNA" і "tRNA" відповідно                        |
| 5.22 | Ключ функції                | N_region   |
| .    | Визначення                  | Додаткові нуклеотиди, що вставляються між перегрупованими фрагментами імуноглобуліну   |
|      | Необов'язкові кваліфікатори | allele<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name  |
|      | Вид організму               | еукаріоти  |
| 5.23 | Ключ функції                | operon   |
| .    | Визначення                  | Область, що містить поліцистронний транскрипт, який включає кластер генів, що контролюються одними і тими ж регуляторними послідовностями / одним промотором в одному біологічному шляху |
|      | Обов'язкові                 | operon   |

|      |                             |   |
|------|-----------------------------|---|
|      | кваліфікатори               |   |
|      | Необов'язкові кваліфікатори | allele<br>function<br>map<br>note<br>phenotype<br>pseudo<br>pseudogene<br>standard_name   |
| 5.24 | Ключ функції                | oriT  |
| .    | Визначення                  | Точка початку перенесення; область молекули ДНК, в якій починається перенесення під час процесу кон'югації або мобілізації  |
|      | Необов'язкові кваліфікатори | allele<br>bound_moiety<br>direction<br>gene<br>gene_synonym<br>map<br>note<br>rpt_family<br>rpt_type<br>rpt_unit_range<br>rpt_unit_seq<br>standard_name   |
|      | Тип молекули                | ДНК   |
|      | Примітки                    | Ключ функції "rep_origin" повинен використовуватись для опису точок початку реплікації; кваліфікатор "direction" має дійсні значення "left", "right" і "" "both", проте лише "left" і "right" є допустимими для використання у поєднанні з функцією "oriT"; точки початку переносу можуть знаходитися в хромосомі; плазміди можуть містити множину точок початку переносу |
| 5.25 | Ключ функції                | polyA_site  |
| .    | Визначення                  | Сайт транскрипта РНК, до якої будуть додані аденінові залишки шляхом пост-транскрипційного поліаденілування   |
|      | Необов'язкові кваліфікатори | allele<br>gene<br>gene_synonym<br>map<br>note   |
|      | Вид організму               | еукаріоти та еукаріотичні віруси  |
| 5.26 | Ключ функції                | precursor_RNA   |
| .    | Визначення                  | Будь-який тип РНК, який не є зрілим продуктом РНК; може охоплювати нкРНК, рРНК, тРНК, нетрансльовану область 5'-кінця ("5'UTR"), кодуєчі послідовності ("CDS", "exon"), вставні послідовності ("intron") і нетрансльовану область 3'-кінця ("3'UTR")  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product  |

|      |                             |   |
|------|-----------------------------|---|
|      |                             | standard_name<br>trans_splicing   |
|      | Примітки                    | Використовується для РНК, яка може бути результатом пост-транскрипційної обробки; якщо відомо, що вказана РНК не є обробленою, повинен використовуватись ключ "prim_transcript"   |
| 5.27 | Ключ функції                | prim_transcript   |
| .    | Визначення                  | Первинний (початковий, необроблений) транскрипт; може охоплювати нкРНК, рРНК, тРНК, нетрансльовану область 5'-кінця ("5'UTR"), кодуєчі послідовності ("CDS", "exon"), вставні послідовності ("intron") і нетрансльовану область 3'-кінця ("3'UTR")  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>standard_name  |
| 5.28 | Ключ функції                | primer_bind   |
| .    | Визначення                  | Сайт нековалентного зв'язування праймера для ініціації реплікації, транскрипції або зворотної транскрипції; охоплює сайт (сайти) для синтетичних елементів, наприклад праймерів для ПЛР   |
|      | Необов'язкові кваліфікатори | allele<br>gene<br>gene_synonym<br>map<br>note<br>standard_name  |
|      | Примітки                    | Використовується для опису сайту заданої послідовності, з яким зв'язана молекула праймера – не призначений для представлення послідовності власне молекули праймера; оскільки в реакціях ПЛР зазвичай беруть участь пари праймерів, один ключ "primer_bind" може використовувати оператор "order(location,location)" з двома локалізаціями, або ж можна використовувати пару ключів "primer_bind" |
| 5.29 | Ключ функції                | propeptide  |
| .    | Визначення                  | Кодуюча послідовність пропептиду; кодуєча послідовність для домену протеїну, тобто розщеплюється, щоб сформувати зрілий білковий продукт  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name   |
| 5.30 | Ключ функції                | protein_bind  |
| .    | Визначення                  | Сайт нековалентного білкового зв'язку в нуклеїновій кислоті   |

|      |                             |   |
|------|-----------------------------|---|
|      | Обов'язкові кваліфікатори   | bound_moiety  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>standard_name  |
|      | Примітки                    | Слід зазначити, що ключ характеристики та кваліфікатор "regulatory_class" зі значенням "ribosome_binding_site" повинен використовуватися для опису сайтів зв'язування рибосом |
| 5.31 | Ключ функції                | regulatory  |
| .    | Визначення                  | Будь-яка область послідовності, що функціонує при регуляції транскрипції, трансляції, реплікації або структура хроматину;   |
|      | Обов'язкові кваліфікатори   | regulatory_class  |
|      | Необов'язкові кваліфікатори | allele<br>bound_moiety<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>phenotype<br>pseudo<br>pseudogene<br>standard_name                                     |
| 5.32 | Ключ функції                | repeat_region   |
| .    | Визначення                  | Область генома, що містить повторювані одиниці  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>rpt_family<br>rpt_type<br>rpt_unit_range<br>rpt_unit_seq<br>satellite<br>standard_name                           |
| 5.33 | Ключ функції                | rep_origin  |
| .    | Визначення                  | Джерело реплікації; сайт початку подвоювання нуклеїнової кислоти для отримання двох ідентичних копій  |
|      | Необов'язкові кваліфікатори | allele<br>direction<br>function<br>gene<br>gene_synonym   |

|      |                             |   |
|------|-----------------------------|---|
|      |                             | map<br>note<br>standard_name  |
|      | Примітки                    | Кваліфікатор напрямку має такі допустимі значення: left, right або both   |
| 5.34 | Ключ функції                | rRNA  |
| .    | Визначення                  | Зріла рибосомна РНК; компонент РНК рибонуклеопротеїнової частки (рибосоми), яка збирає амінокислоти в білки   |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>pseudo<br>pseudogene<br>standard_name   |
|      | Примітки                    | Розміри рРНК мають бути зазначені в кваліфікаторі "product"   |
| 5.35 | Ключ функції                | s_region  |
| .    | Визначення                  | Область перемикавання важких ланцюгів імуноглобуліну; бере участь в перебудові важких ланцюгів ДНК, що призводить до експресії імуноглобуліну іншого класу з тієї ж самої В-клітини                             |
|      | Необов'язкові кваліфікатори | allele<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name   |
|      | Вид організму               | еукаріоти   |
| 5.36 | Ключ функції                | sig_peptide   |
| .    | Визначення                  | Кодуюча послідовність сигнального пептиду; кодуюча послідовність для N-кінцевого домену виділеного білка; цей домен бере участь у приєднуванні поліпептиду, що збільшується, до лідерної послідовності мембрани |
|      | Необов'язкові кваліфікатори | allele<br>function gene gene_synonym<br>map<br>note<br>product pseudo pseudogene<br>standard_name   |
| 5.37 | Ключ функції                | source  |
| .    | Визначення                  | Ідентифікує джерело послідовності; цей ключ є обов'язковим; кожна послідовність повинна містити один ключ "source", що охоплює всю послідовність  |
|      | Обов'язкові кваліфікатори   | organism mol_type   |

|                             |              |  |
|-----------------------------|--------------|--|
| Необов'язкові кваліфікатори |              | cell_line<br>cell_type<br>chromosome<br>clone<br>clone_lib<br>collected_by<br>collection_date<br>cultivar<br>dev_stage<br>ecotype<br>environmental_sample<br>germline<br>haplogroup<br>haplotype<br>host<br>identified_by<br>isolate<br>isolation_source<br>lab_host<br>lat_lon<br>macronuclear<br>map<br>mating_type<br>note<br>organelle<br>PCR_primers<br>plasmid<br>pop_variant<br>proviral<br>rearranged<br>segment<br>serotype<br>serovar<br>sex<br>strain<br>sub_clone<br>sub_species<br>sub_strain<br>tissue_lib<br>tissue_<br>type<br>variety |
| Тип молекули                |              | будь-який  |
| 5.38                        | Ключ функції | stem_loop  |
| .                           | Визначення   | "Шпилька"; двоспіральна область, утворена шляхом парування основ між суміжними (інвертованими) комплементарними послідовностями в одному ланцюгу РНК або ДНК   |
| Необов'язкові кваліфікатори |              | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>standard_name   |
| 5.39                        | Ключ функції | STS  |
| .                           | Визначення   | ДНК-маркувальний сайт; коротка унікальна послідовності ДНК, що описує картування генома і може бути виявлена за допомогою ПЛР; область генома може бути картована шляхом визначення порядку серій ДНК-маркувальних сайтів  |
| Необов'язкові кваліфікатори |              | allele   |

|      |                             |   |
|------|-----------------------------|---|
|      | Тип молекули                | gene<br>gene_synonym<br>map<br>note<br>standard_name<br>ДНК   |
|      | Примітки                    | Локалізація ДНК-маркувального сайту для включення праймера (- ів) до ключа "primer_bind" або до праймерів   |
| 5.40 | Ключ функції                | telomere  |
| .    | Визначення                  | Біологічно досліджувана область, ідентифікована як теломера і досліджена експериментальним шляхом   |
|      | Необов'язкові кваліфікатори | note<br>rpt_type<br>rpt_unit_range<br>rpt_unit_seq<br>standard_name   |
|      | Примітки                    | Функція "telomere" описує інтервал ДНК, що відповідає певній структурі наприкінці лінійної еукаріотичної хромосоми, необхідній для забезпечення цілісності й підтримки кінця; ця область є унікальною у порівнянні з іншими ділянками хромосоми і представляє собою фізичний кінець хромосоми |
| 5.41 | Ключ функції                | tmRNA   |
| .    | Визначення                  | Транспортна матрична РНК; tmRNA спершу діє як тРНК, а потім як мРНК, що кодує пептидну мітку; рибосома транлює цю область мРНК до tmRNA і приєднує закодовану пептидну мітку до С-кінця незавершеного білка; ця приєднана мітка визначає білок для деструкції або протеолізу                  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name tag_peptide   |
| 5.42 | Ключ функції                | transit_peptide   |
| .    | Визначення                  | Кодуюча послідовність транзитного пептиду; кодуюча послідовність для N-кінцевого домену ядерно-кодованого органоїдного білка; цей домен бере участь у пост-трансляційному перенесенні білка до органели   |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name   |
| 5.43 | Ключ функції                | tRNA  |
| .    | Визначення                  | Зріла транспортна РНК; мала молекула РНК (довжиною 75-85  |



|      |                             |   |
|------|-----------------------------|---|
|      | Необов'язкові кваліфікатори | основ), яка допомагає в процесі трансляції послідовності нуклеїнових кислот в амінокислотну послідовність<br>allele<br>anticodon<br>function<br>gene<br>gene_synonym<br>map<br>note<br>operon<br>product<br>pseudo<br>pseudogene<br>standard_name<br>trans_splicing |
| 5.44 | Ключ функції                | unsure  |
| .    | Визначення                  | Невелика область секвенованих основ, як правило, довжиною 10, або менше, яку не можна впевнено ідентифікувати. Така область може містити названі основи (a, t, g, або c), або суміш названих основ та неназваних основ ('n').                                       |
|      | Необов'язкові кваліфікатори | allele<br>compare<br>gene<br>gene_synonym<br>map<br>note replace  |
|      | Примітки                    | для видалення, вставки або заміни слід використовувати кваліфікатор "replace".  |
| 5.45 | Ключ функції                | V_region  |
| .    | Визначення                  | Варіабельна область легких і важких ланцюгів імуноглобуліну, а також альфа-, бета- і гама-ланцюгів Т-клітинного рецептора; коди для варіабельної термінальної аміно-частини; може складатися з таких ключів: "V segment", "D segment", "N region" і "J segment"     |
|      | Необов'язкові кваліфікатори | allele<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name   |
|      | Вид організму               | еукаріоти   |
| 5.46 | Ключ функції                | V_segment   |
| .    | Визначення                  | Варіабельна ділянка легких і важких ланцюгів імуноглобуліну, а також альфа-, бета- і гама-ланцюгів Т-клітинного рецептора; коди для більшої частини варіабельної області ("V_region") та декількох останніх амінокислот лідерного пептиду                           |
|      | Необов'язкові кваліфікатори | allele<br>gene<br>gene_synonym<br>map<br>note<br>product<br>pseudo<br>pseudogene<br>standard_name   |

|      | Вид організму               | еукаріоти   |
|------|-----------------------------|---|
| 5.47 | Ключ функції                | variation   |
| .    | Визначення                  | Споріднений штаб містить стабільні мутації з того ж гена (наприклад ПДРФ, поліморфізми тощо), які відрізняються від наведеної послідовності в цій локалізації (а, можливо, і в інших)   |
|      | Необов'язкові кваліфікатори | allele<br>compare<br>frequency<br>gene<br>gene_synonym<br>map<br>note<br>phenotype<br>product<br>replace<br>standard_name   |
|      | Примітки                    | Використовується для опису алелів, ПДРФ та інших природних мутацій і поліморфізмів; для опису видалення, вставки або заміщення використовується кваліфікатор "replace"; мінливість, що виникає внаслідок генетичного маніпулювання (наприклад, спрямованого мутагенезу), повинна описуватися за допомогою функції "misc_difference" |
| 5.48 | Ключ функції                | 3'UTR   |
| .    | Визначення                  | 1) Область на 3'-кінці зрілого транскрипта (після стоп-кодона), що не транслюється у білок;<br>2) Область на 3'-кінці РНК-вірусу (після останнього стоп-кодона), що не транслюється у білок;  |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>standard_name<br>trans_splicing  |
|      | Примітки                    | Знак апострофа має особливе значення в XML, тож у значенні елемента він повинен замінятися на "&apos;". Таким чином, "3'UTR" у файлі XML повинне бути представлено як "3&apos;UTR", тобто<br><INSDFeature_key>3&apos;UTR</INSDFeature_key>.   |
| 5.49 | Ключ функції                | 5'UTR   |
| .    | Визначення                  | 1) Область на 5'-кінці зрілого транскрипта (перед ініціювальним кодоном), що не транслюється у білок;<br>Область на 5'-кінці РНК-вірусу (перед першим ініціювальним кодоном), що не транслюється у білок;   |
|      | Необов'язкові кваліфікатори | allele<br>function<br>gene<br>gene_synonym<br>map<br>note<br>standard_name<br>trans_splicing  |
|      | Примітки                    | Знак апострофа має особливе значення в XML, тож у значенні елемента він повинен замінятися на "&apos;". Таким чином, "5'UTR" у файлі XML повинне бути представлено як "5&apos;UTR", тобто   |

<INSDFeature\_key>5&apos;UTR</INSDFeature\_key>.

## РОЗДІЛ 6: КВАЛІФІКАТОРИ ДЛЯ НУКЛЕОТИДНИХ ПОСЛІДОВНОСТЕЙ

Цей розділ містить перелік кваліфікаторів для використання з функціями у нуклеотидних послідовностях. Кваліфікатори наведені в алфавітному порядку.

Якщо формат значення – “none”, елементи `INSDQualifier_value` та `NonEnglishQualifier_value` не повинні використовуватись.

Якщо формат значення – це довільний текст, який визначається як залежний від мови, слід використовувати один з наступних елементів

- 1) елемент `INSDQualifier_value`; або
- 2) елемент `NonEnglishQualifier_value`; або
- 3) обидва елементи `INSDQualifier_value` та `NonEnglishQualifier_value`.

Якщо формат значення – це щось інше, ніж “none”, але визначений як залежний від мови довільний текст, то повинен використовуватись елемент `INSDQualifier_value`, а елемент `NonEnglishQualifier_value` не повинен використовуватись.

**СЛІД ЗАУВАЖИТИ, ЩО:** Будь-яке значення кваліфікатора, передбачене для кваліфікатора з форматом значення залежного від мови довільного тексту може вимагати перекладу для міжнародних, національних або регіональних процедур. Кваліфікатори, наведені в наступній таблиці, вважаються такими, що мають значення залежного від мови довільного тексту:

Таблиця 5: Список кваліфікаторів із залежними від мови значеннями довільного тексту для нуклеотидних послідовностей

| Секція | Кваліфікатор залежного від мови довільного тексту |
|--------|---|
| 6.3    | <code>bound_moiety</code>                         |
| 6.5    | <code>cell_type</code>                            |
| 6.8    | <code>clone</code>                                |
| 6.9    | <code>clone_lib</code>                            |
| 6.11   | <code>collected_by</code>                         |
| 6.14   | <code>cultivar</code>                             |
| 6.15   | <code>dev_stage</code>                            |
| 6.18   | <code>ecotype</code>                              |
| 6.21   | <code>frequency</code>                            |
| 6.22   | <code>function</code>                             |
| 6.24   | <code>gene_synonym</code>                         |
| 6.26   | <code>haplogroup</code>                           |
| 6.28   | <code>host</code>                                 |
| 6.29   | <code>identified_by</code>                        |

---

|      |                  |
|------|------------------|
| 6.30 | isolate          |
| 6.31 | isolation_source |
| 6.32 | lab_host         |
| 6.36 | mating_type      |
| 6.41 | note             |
| 6.45 | organism         |
| 6.47 | phenotype        |
| 6.49 | pop_variant      |
| 6.50 | product          |
| 6.66 | serotype         |
| 6.67 | serovar          |
| 6.68 | sex              |
| 6.69 | standard_name    |
| 6.70 | strain           |
| 6.71 | sub_clone        |
| 6.72 | sub_species      |
| 6.73 | sub-strain       |
| 6.75 | tissue_lib       |
| 6.76 | tissue_type      |
| 6.81 | variety          |

|      |                               |  |
|------|-------------------------------|--|
| 6.1. | Кваліфікатор                  | allele   |
|      | Визначення                    | Назва алеля певного гена   |
|      | Формат обов'язкового значення | довільний текст  |
|      | Приклади                      | <INSDQualifier_value>adh1-1</INSDQualifier_value>  |
|      | Примітки                      | Всі функції, пов'язані з певним геном ("exon", "CDS" тощо), мають поділяти те саме значення кваліфікатора "allele"; як усталено, значення кваліфікатора "allele" повинне відрізнятися від значення кваліфікатора "gene"; значення кваліфікатора "allele" має відповідати такому ж значенню у варіанті. |
| 6.2. | Кваліфікатор                  | anticodon  |
|      | Визначення                    | Локалізація антикодона тРНК і амінокислоти, яку вона кодує   |
|      | Формат обов'язкового значення | (pos:<location>,aa:<amino_acid>,seq:<text>), де "<location>" є положенням антикодона, "<amino_acid>" є трибуквеною скороченою назвою закодованої амінокислоти, а "<text>" є послідовністю антикодона   |
|      | Приклади                      | <INSDQualifier_value>(pos:34..36,aa:Phe,seq:aaa)</INSDQualifier_value><br><INSDQualifier_value>(pos:join(5,495..496,aa:Leu,seq:taa)</INSDQualifier_value><br><INSDQualifier_value>(pos:complement(4156..4158),aa:Glu,seq:tg)</INSDQualifier_value>   |
| 6.3. | Кваліфікатор                  | bound_moiety   |
|      | Визначення                    | Назва молекули/комплекса, що можуть бути пов'язані з відповідною функцією  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|      | Приклади                      | <INSDQualifier_value>GAL4</INSDQualifier_value>  |
|      | Примітки                      | Для функцій "misc_binding", "oriT" і "protein_bind" допустимий один кваліфікатор "bound_moiety".   |
| 6.4. | Кваліфікатор                  | cell_line  |
|      | Визначення                    | Клітинна лінія, з якої було отримано послідовність   |
|      | Формат обов'язкового значення | довільний текст  |
|      | Приклади                      | <INSDQualifier_value>MCF7</INSDQualifier_value>  |
| 6.5. | Кваліфікатор                  | cell_type  |
|      | Визначення                    | Тип клітини, з якої було отримано послідовність  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|      | Приклади                      | <INSDQualifier_value>leukocyte</INSDQualifier_value>   |
| 6.6. | Кваліфікатор                  | chromosome   |
|      | Визначення                    | Хромосома (наприклад, номер хромосоми), з якої було отримано послідовність   |

|                               |   |
|-------------------------------|---|
| Формат обов'язкового значення | довільний текст   |
| Приклади                      | <INSDQualifier_value>1</INSDQualifier_value><br><INSDQualifier_value>X</INSDQualifier_value>  |
| 6.7. Кваліфікатор             | circular_RNA  |
| Визначення                    | вказує на те, що екзони не впорядковані або перекриваються, оскільки ця сплайсифікована РНК є циркулярною РНК (цирРНК), створеною шляхом зворотного сплайсингу, наприклад, коли наступний за ним екзон у гені розташований за 5' від попереднього екзону в РНК-продукті                         |
| Формат обов'язкового значення | відсутній   |
| Примітки                      | слід використовувати для таких елементів, як CDS, мРНК, тРНК та інших елементів, які утворюються в результаті події зворотного сплайсингу. Цей кваліфікатор слід використовувати лише коли подія сплайсингу вказана в операторі "join", наприклад join(complement(69611..69724),139856..140087) |
| 6.8. Кваліфікатор             | clone   |
| Визначення                    | Клон, з якого було отримано послідовність   |
| Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)   |
| Приклади                      | <INSDQualifier_value>lambda-hIL7.3</INSDQualifier_value>  |
| Примітки                      | Функція "source" не повинна містити більше одного кваліфікатора клону; якщо послідовність було отримано з множини клонів, це можна додатково описати в таблиці функцій, використовуючи ключ функції "misc_feature" та кваліфікатор "note" для зазначення множини клонів.                        |
| 6.9. Кваліфікатор             | clone_lib   |
| Визначення                    | Бібліотека клонів, з якої було отримано послідовність   |
| Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)   |
| Приклади                      | <INSDQualifier_value>lambda-hIL7</INSDQualifier_value>  |
| 6.10                          | Кваліфікатор  |
| .                             | codon_start   |
| Визначення                    | Вказує на зміщення, де розташований перший повний кодон, відносно першої основи відповідної функції.  |
| Формат обов'язкового значення | 1, або 2, або 3   |
| Приклади                      | <INSDQualifier_value>2</INSDQualifier_value>  |
| 6.11                          | Кваліфікатор  |
| .                             | collected_by  |
| Визначення                    | Ім'я особи або назва установи, яка зберігає зразок  |
| Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для національних/регіональних процедур)   |

|          |   |  |
|----------|---|--|
| Приклади | <INSDQualifier_value>Dan Janzen</INSDQualifier_value> |  |
| 6.12     | Кваліфікатор  | collection_date  |
| .        | Визначення  | Дата, коли зразок був прийнятий на збереження  |
|          | Формат обов'язкового значення                         | PPPP-ММ-ДД, PPPP-ММ або PPPP   |
|          | Приклади  | <INSDQualifier_value>1952-10-21</INSDQualifier_value><br><INSDQualifier_value>1952-10</INSDQualifier_value><br><INSDQualifier_value>1952</INSDQualifier_value>   |
|          | Примітки  | "PPPP" - представлення року у форматі чотиризначного числа. "ММ" - представлення місяця у форматі двозначного числа. "ДД" - представлення дня місяця у форматі двозначного числа   |
| 6.13     | Кваліфікатор  | compare  |
| .        | Визначення  | Відомості про існуючий публічний запис в INSD, з яким здійснюється порівняння  |
|          | Формат обов'язкового значення                         | [accession-number.sequence-version]  |
|          | Приклади  | <INSDQualifier_value>AJ634337.1</INSDQualifier_value>  |
|          | Примітки  | Цей кваліфікатор може використовуватися з такими функціями: misc_difference, unsure і variation. В межах однієї функції допустимі декілька кваліфікаторів "compare" з різним вмістом.<br><br>Цей кваліфікатор не призначений для масштабного описування варіантів, таких як одонуклеотидні поліморфізми. |
| 6.14     | Кваліфікатор  | cultivar   |
| .        | Визначення  | сорт (культурний сорт) рослин, з якого було отримано послідовність   |
|          | Формат обов'язкового значення                         | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|          | Приклади  | <INSDQualifier_value>Nipponbare</INSDQualifier_value><br><INSDQualifier_value>Tenuifolius</INSDQualifier_value><br><INSDQualifier_value>Candy Cane</INSDQualifier_value><br><INSDQualifier_value>IR36</INSDQualifier_value>  |
|          | Примітки  | Кваліфікатор "cultivar" застосовується виключно для продуктів штучної селекції; для відомих природних сортів рослин і грибів слід використовувати кваліфікатор "variety".  |
| 6.15     | Кваліфікатор  | dev_stage  |
| .        | Визначення  | Якщо послідовність було отримано з організму на певній стадії розвитку, це зазначається в цьому кваліфікаторі  |
|          | Формат обов'язкового значення                         | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|          | Приклади  | <INSDQualifier_value>fourth instar larva</INSDQualifier_value>   |



|      |                               |   |
|------|-------------------------------|---|
| 6.16 | Кваліфікатор                  | direction   |
| .    |                               |   |
|      | Визначення                    | Напрямок реплікації ДНК   |
|      | Формат обов'язкового значення | "left", "right" або "both", де "left" означає напрямок до 5'-кінця послідовності (як показано), а "right" означає напрямок до 3'-кінця  |
|      | Приклади                      | <INSDQualifier_value>left</INSDQualifier_value>   |
|      | Примітки                      | Значення "left", "right" і "both" дозволяється використовувати, якщо кваліфікатор "direction" використовується для опису ключа функції "rep_origin". Проте, якщо кваліфікатор "direction" використовується для опису ключа функції "oriT", дозволяється використовувати лише значення "left" і "right".   |
| 6.17 | Кваліфікатор                  | EC_number   |
| .    |                               |   |
|      | Визначення                    | Шифр КФ для ферментного продукту послідовності  |
|      | Формат обов'язкового значення | довільний текст   |
|      | Приклади                      | <INSDQualifier_value>1.1.2.4</INSDQualifier_value><br><INSDQualifier_value>1.1.2.-</INSDQualifier_value><br><INSDQualifier_value>1.1.2.n</INSDQualifier_value><br><INSDQualifier_value>1.1.2.n1</INSDQualifier_value>   |
|      | Примітки                      | Чинні значення шифрів КФ визначено у списку, підготовленому Номенклатурним комітетом Міжнародного союзу біохімії та молекулярної біології (NC-IUBMB) (опублікований в Номенклатурі ферментів 1992 року (видавництво Academic Press, Сан Дієго) або у новішій її редакції). Формат представляє собою рядок із чотирьох цифр, розділених крапками; щонайбільше три цифри з кінця рядка можуть бути замінені на тире "-" для позначення непевного присвоєння. Символи, що включають символ "n", наприклад, "n", "n1" і т.п., можуть використовуватись в останньому положенні замість цифри, якщо шифр КФ очікує присвоєння. Слід зауважити, що подібні неповні шифри КФ не схвалюються NC-IUBMB. |
| 6.18 | Кваліфікатор                  | ecotype   |
| .    |                               |   |
|      | Визначення                    | Популяція у межах певного виду, що демонструє генетичні фенотипові ознаки, які відображають адаптацію до локального навколишнього середовища  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)   |
|      | Приклади                      | <INSDQualifier_value>Columbia</INSDQualifier_value>   |
|      | Примітки                      | Прикладом такої популяції є популяція, що адаптувала свій листяний опушений покрив трихом у відповідь на особливо сонячне навколишнє середовище. Кваліфікатор "ecotype" зазвичай застосовується до стандартного генетичного фонду <i>Arabidopsis thaliana</i> , але може застосовуватись до будь-якого безчерешкового організму.  |
| 6.19 | Кваліфікатор                  | environmental_sample  |
| .    |                               |   |
|      | Визначення                    | Визначає послідовності, отримані прямою молекулярною ізоляцією із значного середовищного зразка ДНК (шляхом ПЛР з або без подальшого клонування продукту, DGGE або іншими невідомими методами) без надійної ідентифікації   |

|                 |                               |  |
|-----------------|-------------------------------|--|
|                 |                               | первинного організму. Середовищні зразки включають клінічні зразки, вміст кишківника та інші послідовності з невідомих організмів, які можуть бути пов'язані з певним хазяїном. Вони не включають ендосимбіонти, які можуть бути напевно виділені з конкретного хазяїна, організми з робочих зразків, які легко ідентифікуються, проте не є культивованими (наприклад чимало ціанобактерій), або фітоплазми, які можуть бути легко виділені з хворих рослин (навіть незважаючи на те, що їх не можна виростити в чистих культурах) |
| Формат значення |                               | немає  |
| Примітки        |                               | Використовується лише з ключем функції "source"; ключі функції "source", які містять кваліфікатор "environmental_sample", також мають включати кваліфікатор "isolation_source"; функція "source", що включає кваліфікатор "environmental_sample" не повинна включати кваліфікатор "strain".  |
| 6.20            | Кваліфікатор                  | exception  |
|                 | Визначення                    | Вказує на те, що кодуючу ділянку не можливо транслювати за стандартними біологічними правилами   |
|                 | Формат обов'язкового значення | Один із таких стандартизованих висловів:<br>RNA editing<br>rearrangement required for product annotated by transcript or proteomic data  |
|                 | Приклади                      | <INSDQualifier_value>RNA editing</INSDQualifier_value><br><INSDQualifier_value>rearrangement required for product</INSDQualifier_value>  |
|                 | Примітки                      | Повинен використовуватись лише для опису біологічних механізмів на кшталт "редагування" РНК; білкова трансляція кодуючої послідовності з кваліфікатором "exception" відрізнятиметься від відповідної концептуальної трансляції; не повинен використовуватись, якщо кваліфікатор "transl_except" буде повноцінним, наприклад, у випадку використання стоп-кодона.   |
| 6.21            | Кваліфікатор                  | frequency  |
|                 | Визначення                    | Частота повторюваності функції   |
|                 | Формат обов'язкового значення | довільний текст, що показує співвідношення популяції з функцією, у вигляді дробу<br>Залежить від мови: це значення може потребувати перекладу для міжнародних/національних/регіональних процедур   |
|                 | Приклади                      | <INSDQualifier_value>23/108</INSDQualifier_value><br><INSDQualifier_value>1 з 12</INSDQualifier_value><br><INSDQualifier_value>0.85</INSDQualifier_value>  |
| 6.22            | Кваліфікатор                  | function   |
|                 | Визначення                    | Функція, що відноситься до послідовності   |
|                 | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для національних/регіональних процедур)<br>Залежить від мови: це значення може потребувати перекладу для міжнародних/національних/регіональних процедур  |
|                 | Приклади                      | <INSDQualifier_value>essential for recognition of cofactor<br></INSDQualifier_value>   |

|          |  |  |
|----------|--|--|
| Примітки | Кваліфікатор "function" використовується, якщо назва гена та/або назва продукту не передають функцію, що відноситься до послідовності. |  |
| 6.23     | Кваліфікатор   | gene   |
| .        | Визначення   | Символ гена, що відповідає ділянці послідовності   |
|          | Формат обов'язкового значення  | довільний текст  |
|          | Приклади   | <INSDQualifier_value>ilvE</INSDQualifier_value>  |
|          | Примітки   | Для зазначення символу гена використовується кваліфікатор "gene"; для зазначення повної назви гена використовується кваліфікатор "standard_name".  |
| 6.24     | Кваліфікатор   | gene_synonym   |
| .        | Визначення   | Синонімічний, заміщуваний, застарілий або колишній символ гена   |
|          | Формат обов'язкового значення  | довільний текст<br><br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
| 6.25     | Кваліфікатор   | germline   |
| .        | Визначення   | Представлена послідовність не піддавалася соматичній перебудові як частині адаптивної імунної реакції; це неперебудована послідовність, успадкована від батьківської генеративної лінії  |
|          | Формат значення  | немає  |
|          | Примітки   | Кваліфікатор "germline" не повинен використовуватись для позначення того, що джерелом послідовності є гамета або статеві клітини; кваліфікатори "germline" і "rearranged" не повинні використовуватись в одній функції "source"; кваліфікатори "germline" і "rearranged" повинні використовуватись лише для молекул, що здатні до соматичної перебудови як частини адаптивної імунної реакції; до них належать локуси Т-клітинних рецепторів (ТКР) та імуноглобулінові локуси у щелепних хребетних, а також непов'язані локуси варіабельних лімфоцитних рецепторів (ВЛР) у безщелепних риб (міног і міксин); кваліфікатори "germline" і "rearranged" не повинні використовуватись за межами класу Craniata (taxid=89593) |
| 6.26     | Кваліфікатор   | haplogroup   |
| .        | Визначення   | Назва групи однакових гаплотипів, що поділяють певну зміну послідовності. Гаплогрупи часто використовуються для відстежування міграції популяційних груп.  |
|          | Формат обов'язкового значення  | довільний текст<br><br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|          | Приклади   | <INSDQualifier_value>H*</INSDQualifier_value>  |
| 6.27     | Кваліфікатор   | haplotype  |
| .        | Визначення   | Назва певної групи алелів, зв'язаних разом на одній фізичній хромосомі. За відсутності рекомбінації, кожен гаплотип успадковується як окрема одиниця і може  |

|      |                               |  |
|------|-------------------------------|--|
|      |                               | використовуватись для відстежування потоку генів у популяціях.   |
|      | Формат обов'язкового значення | довільний текст  |
|      | Приклади                      | <INSDQualifier_value>Dw3 B5 Cw1 A1</INSDQualifier_value>   |
| 6.28 | Кваліфікатор                  | host   |
|      | Визначення                    | природний (на відміну від лабораторного) організм-хазяїн, з якого було отримано секвеновану молекулу   |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для національних/регіональних процедур)  |
|      | Приклади                      | <INSDQualifier_value>Homo sapiens</INSDQualifier_value><br><INSDQualifier_value>Homo sapiens 12 year old girl</INSDQualifier_value><br><INSDQualifier_value>Rhizobium NGR234</INSDQualifier_value>   |
| 6.29 | Кваліфікатор                  | identified_by  |
|      | Визначення                    | Ім'я фахівця, який таксономічно ідентифікував зразок.  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|      | Приклади                      | <INSDQualifier_value>John Burns</INSDQualifier_value>  |
| 6.30 | Кваліфікатор                  | isolate  |
|      | Визначення                    | Окремий ізолят, з якого було отримано послідовність.   |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|      | Приклади                      | <INSDQualifier_value>Patient #152</INSDQualifier_value><br><INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>   |
| 6.31 | Кваліфікатор                  | isolation_source   |
|      | Визначення                    | Описує фізичне, середовищне та/або місцеве географічне джерело біологічного зразка, з якого було отримано послідовність.   |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|      | Приклади                      | <INSDQualifier_value>rumen isolates from standard Pelleted ration-fed steer #67</INSDQualifier_value><br><INSDQualifier_value>permanent Antarctic sea ice</INSDQualifier_value><br><INSDQualifier_value>denitrifying activated sludge from carbon_limited continuous reactor</INSDQualifier_value> |
|      | Примітки                      | Використовується лише з ключем функції "source"; ключі функцій "source", що містять кваліфікатор "environmental_sample", мають також містити кваліфікатор "isolation_source".  |
| 6.32 | Кваліфікатор                  | lab_host   |

|      |                               |   |
|------|-------------------------------|---|
|      | Визначення                    | Наукова назва лабораторного хазяїна, що використовується для поширювання вихідного організму, з якого було отримано секвеновану молекулу  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)   |
|      | Приклади                      | <INSDQualifier_value>Gallus gallus</INSDQualifier_value><br><INSDQualifier_value>Gallus gallus embryo</INSDQualifier_value><br><INSDQualifier_value>Escherichia coli strain DH5 alpha</INSDQualifier_value><br><INSDQualifier_value>Homo sapiens HeLa cells</INSDQualifier_value> |
|      | Примітки                      | Якщо відомо, має використовуватись повна біноміальна назва організму-хазяїна; також може включатись додаткова умовна інформація стосовно хазяїна  |
| 6.33 | Кваліфікатор                  | lat_lon   |
|      | Визначення                    | Географічні координати місцезнаходження, де було прийнято зразок на збереження  |
|      | Формат обов'язкового значення | довільний текст - градуси широти і довготи у форматі "Г[гг.гггг] пн.ш. пд.ш. Г[гг.гггг] зх.д. сх.д."  |
|      | Приклади                      | <INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value><br><INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value>   |
| 6.34 | Кваліфікатор                  | macronuclear  |
|      | Визначення                    | Якщо представлена послідовність є ДНК і отримана з організму, що піддається хромосомній диференціації між макронуклеарною і мікронуклеарною стадіями, цей кваліфікатор використовується для позначення того, що послідовність отримана з макронуклеарної ДНК                      |
|      | Формат значення               | немає   |
| 6.35 | Кваліфікатор                  | map   |
|      | Визначення                    | Положення функції на геномній карті   |
|      | Формат обов'язкового значення | довільний текст   |
|      | Приклади                      | <INSDQualifier_value>8q12-q13</INSDQualifier_value>   |
| 6.36 | Кваліфікатор                  | mating_type   |
|      | Визначення                    | Тип парування організму, з якого було отримано послідовність; тип парування застосовується до прокаріотів, а також до еукаріотів, здатних до мейозу без статеву диморфних гамет   |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для національних/регіональних процедур)   |
|      | Приклади                      | <INSDQualifier_value>MAT-1</INSDQualifier_value><br><INSDQualifier_value>plus</INSDQualifier_value><br><INSDQualifier_value>-</INSDQualifier_value><br><INSDQualifier_value>odd</INSDQualifier_value><br><INSDQualifier_value>even</INSDQualifier_value>                          |

|                               |  |                     |
|-------------------------------|--|---------------------|
| Примітки                      | Значення "male" і "female" кваліфікатора "mating_type" дійсні для прокариотів, а не для еукаріотів; детальна інформація міститься у записі стосовно кваліфікатора "sex".   |                     |
| 6.37                          | Кваліфікатор   | mobile_element_type |
| Визначення                    | Тип і назва або кваліфікатор мобільного елемента, що описується первинною функцією   |                     |
| Формат обов'язкового значення | <mobile_element_type>[:<mobile_element_name>], де <mobile_element_type> є одним із наступних термінів:<br>transposon<br>retrotransposon<br>integron<br>insertion sequence<br>non-LTR retrotransposon<br>SINE<br>MITE<br>LINE<br>other  |                     |
| Приклади                      | <INSDQualifier_value>transposon:Tnp9</INSDQualifier_value>   |                     |
| Примітки                      | Кваліфікатор "mobile_element_type" дійсний лише для ключа функції "mobile_element". Мобільний елемент має використовуватись для представлення обох елементів, що наразі є мобільними, а також елементів, які були мобільними раніше.<br>Значення "other" для <mobile_element_type> потребує зазначення <mobile_element_name> |                     |
| 6.38                          | Кваліфікатор   | mod_base            |
| Визначення                    | скорочення для модифікованої нуклеотидної основи   |                     |
| Формат обов'язкового значення | Одне із скорочень для модифікованих основ, наведених у Розділі 2 цього додатку   |                     |
| Приклади                      | <INSDQualifier_value>m5c</INSDQualifier_value><br><INSDQualifier_value>OTHER</INSDQualifier_value>   |                     |
| Примітки                      | Певні модифіковані нуклеотиди, не зазначені у Розділі 2 цього Додатку, описуються словом "OTHER", що вказується як значення кваліфікатора "mod_base", і шляхом включення кваліфікатора "note" з повною назвою модифікованої основи як його значенням   |                     |
| 6.39                          | Кваліфікатор   | mol_type            |
| Визначення                    | Молекулярний тип послідовності   |                     |
| Формат обов'язкового значення | Одне із таких значень:<br>genomic DNA<br>genomic RNA<br>mRNA<br>tRNA<br>rRNA<br>other RNA<br>other DNA<br>transcribed RNA<br>viral cRNA<br>unassigned DNA<br>unassigned RNA  |                     |
| Приклади                      | <INSDQualifier_value>genomic DNA</INSDQualifier_value><br><INSDQualifier_value>other RNA</INSDQualifier_value>   |                     |
| Примітки                      | Кваліфікатор "mol_type" є обов'язковим для ключа функції "source"; значення "genomic DNA" не означає, що молекула  |                     |

є ядерною (наприклад, органельна і плазмідна ДНК повинні описуватися, використовуючи значення "genomic DNA"); гени рибосомної ДНК повинні описуватися, використовуючи значення "genomic DNA"; значення "rRNA" повинна використовуватись, лише якщо було секвеновано власне молекулу рибосомної ДНК; значення "other RNA" і "other DNA" повинні застосовуватись до штучних молекул, значення "unassigned DNA" і "unassigned RNA" повинні застосовуватись, якщо молекула *in vivo* невідома.

|      |                               |  |
|------|-------------------------------|--|
| 6.40 | Кваліфікатор                  | ncRNA_class  |
| .    | Визначення                    | Структурний опис класифікації некодуючої РНК, описаний первинним ключем "ncRNA"  |
|      | Формат обов'язкового значення | TYPE<br>де "TYPE" є одним із таких стандартизованих термінів або висловів:<br>antisense_RNA<br>autocatalytically_spliced_intron<br>circRNA<br>ribozyme<br>hammerhead_ribozyme<br>lncRNA<br>RNase_P_RNA<br>RNase_MRP_RNA<br>telomerase_RNA<br>guide_RNA<br>sgRNA<br>rasiRNA<br>scRNA<br>scaRNA<br>siRNA<br>pre_miRNA<br>miRNA<br>piRNA<br>snoRNA<br>snRNA<br>SRP_RNA<br>vault_RNA<br>Y_RNA<br>other |
|      | Приклади                      | <INSDQualifier_value>autocatalytically_spliced_intron<br></INSDQualifier_value><br><INSDQualifier_value>siRNA</INSDQualifier_value><br><INSDQualifier_value>scRNA</INSDQualifier_value><br><INSDQualifier_value>other</INSDQualifier_value>  |
|      | Примітки                      | Певні типи некодованих РНК, позначення яких відсутні у стандартизованій лексиці для кваліфікатора "ncRNA_class", повинні бути описані шляхом наведення "other" як значення кваліфікатора "ncRNA_class" і надання відповідного стислого пояснення в кваліфікаторі "note"  |
| 6.41 | Кваліфікатор                  | note   |
| .    | Визначення                    | Будь-які примітки або додаткова інформація   |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|      | Приклади                      | <INSDQualifier_value>A comment about the feature</INSDQualifier_value>   |
| 6.42 | Кваліфікатор                  | number   |

|      |                               |  |
|------|-------------------------------|--|
|      | Визначення                    | Число для позначення порядку генетичних елементів (наприклад, екзонів або інтронів) у напрямку від 5'-кінця до 3'-кінця  |
|      | Формат обов'язкового значення | довільний текст (без пробілів)   |
|      | Приклади                      | <INSDQualifier_value>4</INSDQualifier_value><br><INSDQualifier_value>6B</INSDQualifier_value>  |
|      | Примітки                      | Текст, обмежений цілими числами, буквами або комбінаціями цілих чисел і/або букв, що представляються як значення даних, які не містять пробілів; будь-які додаткові терміни слід включати до кваліфікатора "standard_name".<br>Приклад:<br>кваліфікатор "number" із значенням "2A" та кваліфікатор "standard_name" із значенням "long"   |
| 6.43 | Кваліфікатор                  | operon   |
|      | Визначення                    | Назва групи безперервних генів, транскрибованих в один транскрипт, до якого належить функція   |
|      | Формат обов'язкового значення | довільний текст  |
|      | Приклади                      | <INSDQualifier_value>lac</INSDQualifier_value>   |
| 6.44 | Кваліфікатор                  | organelle  |
|      | Визначення                    | Тип внутрішньоклітинної структури, зв'язаної з мембраною, з якої було отримано послідовність   |
|      | Формат обов'язкового значення | Один із таких стандартизованих термінів або висловів:<br>chromatophore<br>hydrogenosome<br>mitochondrion<br>nucleomorph<br>plastid<br>mitochondrion:kinetoplast<br>plastid:chloroplast<br>plastid:apicoplast<br>plastid:chromoplast<br>plastid:cyanelle<br>plastid:leucoplast<br>plastid:proplastid  |
|      | Приклади                      | <INSDQualifier_value>chromatophore</INSDQualifier_value><br><INSDQualifier_value>hydrogenosome</INSDQualifier_value><br><INSDQualifier_value>mitochondrion</INSDQualifier_value><br><INSDQualifier_value>nucleomorph</INSDQualifier_value><br><INSDQualifier_value>plastid</INSDQualifier_value><br><INSDQualifier_value>mitochondrion:kinetoplast</INSDQualifier_value><br><br><INSDQualifier_value>plastid:chloroplast</INSDQualifier_value><br><br><INSDQualifier_value>plastid:apicoplast</INSDQualifier_value><br><INSDQualifier_value>plastid:chromoplast</INSDQualifier_value><br><br><INSDQualifier_value>plastid:cyanelle</INSDQualifier_value> |



|      |                               |   |
|------|-------------------------------|---|
|      |                               | <pre>&lt;INSDQualifier_value&gt;plastid:leucoplast&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;plastid:proplastid&lt;/INSDQualifier_value&gt;</pre>  |
| 6.45 | Кваліфікатор                  | organism  |
| .    | Визначення                    | Наукова назва організму, що надав генетичний матеріал послідовності, якщо відомо, або доступна таксономічна інформація, якщо організм не є класифікованим; або зазначення того, що послідовність є синтетичним конструктом  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)   |
|      | Приклади                      | <INSDQualifier_value>Homo sapiens</INSDQualifier_value>   |
| 6.46 | Кваліфікатор                  | PCR_primers   |
| .    | Визначення                    | Праймери ПЛР, використані для ампліфікації послідовності. Один кваліфікатор "/PCR_primers" має містити всі праймери, що використовуються для ПЛР. Якщо в одній ПЛР присутні декілька прямих або зворотних праймерів, має бути присутня множина наборів значень "fwd_name/fwd_seq or rev_name/rev_seq"   |
|      | Формат обов'язкового значення | [fwd_name: XXX1, ]fwd_seq: xxxxx1,[fwd_name: XXX2, ]fwd_seq: xxxxx2, [rev_name: YYY1, ]rev_seq: yyyyy1,[rev_name: YYY2, ]rev_seq: yyyyy2  |
|      | Приклади                      | <pre>&lt;INSDQualifier_value&gt;fwd_name:      C01P1,      fwd_seq: ttgatttttttggtcaycwgaagt,rev_name:      C01R4,      rev_seq: ccwvytardcctarraartggtg&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;fwd_name: hogel, fwd_seq: cgkgtgtatcttact,      rev_name:      hoge2,      rev_seq: cg&lt;lt;i&gt;gt;gtgtatcttact&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;fwd_name:      C01P1,      fwd_seq: ttgatttttttggtcaycwgaagt,      fwd_name:      C01P2,      fwd_seq: gatacacaggtcaycwgaagt,      rev_name:      C01R4,      rev_seq: ccwvytardcctarraartggtg&lt;/INSDQualifier_value&gt;</pre> |
|      | Примітки                      | "fwd_seq" і "rev_seq" є обов'язковими; "fwd_name" і "rev_name" є необов'язковими.<br>Обидві послідовності повинні представлятися у порядку 5'>3'. Послідовності повинні представлятися символами, наведеними у Розділі 1 цього Додатка, окрім модифікованих основ, які повинні поміщатися в кутові дужки (< >). У коді XML, кутові дужки "<" і ">" повинні замінюватися на "&lt;" і "&gt;", оскільки в XML вони є зарезервованими символами.  |
| 6.47 | Кваліфікатор                  | phenotype   |
| .    | Визначення                    | Фенотип, що підтверджується функцією, причому фенотип визначається як фізична, біохімічна або поведінкова характеристика чи сукупність характеристик  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)   |
|      | Приклади                      | <INSDQualifier_value>erythromycin resistance</INSDQualifier_value>  |
| 6.48 | Кваліфікатор                  | plasmid   |
| .    |                               |   |

|      |                               |   |
|------|-------------------------------|---|
|      | Визначення                    | Назва природної плазмиди, з якої було отримано послідовність; плазіда визначається як незалежно реплікаційна генетична одиниця, що не може бути описана кваліфікаторами "chromosome" або "segment"  |
|      | Формат обов'язкового значення | довільний текст   |
|      | Приклади                      | <INSDQualifier_value>pC589</INSDQualifier_value>  |
| 6.49 | Кваліфікатор                  | pop_variant   |
|      | Визначення                    | Назва субпопуляції або фенотипу зразка, з якого походить послідовність  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)   |
|      | Приклади                      | <INSDQualifier_value>pop1</INSDQualifier_value><br><INSDQualifier_value>Bear Paw</INSDQualifier_value>  |
| 6.50 | Кваліфікатор                  | product   |
|      | Визначення                    | Назва продукту, пов'язаного з функцією, наприклад, мРНК функції "mRNA", поліпептид функції "CDS", зрілий пептид функції "mat_peptide" тощо.   |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)   |
|      | Приклади                      | <INSDQualifier_value>trypsinogen</INSDQualifier_value><br>(якщо кваліфікатор присутній у функції "CDS")<br><INSDQualifier_value>trypsin</INSDQualifier_value> (якщо кваліфікатор присутній у функції "mat_peptide")<br><INSDQualifier_value>XYZ neural-specific transcript</INSDQualifier_value> (якщо кваліфікатор присутній у функції "mRNA") |
| 6.51 | Кваліфікатор                  | protein_id  |
|      | Визначення                    | Ідентифікаційний номер білкової послідовності; ціле число, що використовується в переліку послідовностей для позначення білкової послідовності, закодованої кодуючою послідовністю, визначеною відповідним ключем функції "CDS" та кваліфікатором "translation"   |
|      | Формат обов'язкового значення | ціле число більше нуля  |
|      | Приклади                      | <INSDQualifier_value>89</INSDQualifier_value>   |
| 6.52 | Кваліфікатор                  | proviral  |
|      | Визначення                    | Цей кваліфікатор використовується для позначення послідовності, отриманої з вірусу або бактеріофагу, інтегрованого в геном іншого організму   |
|      | Формат значення               | немає   |
| 6.53 | Кваліфікатор                  | pseudo  |
|      | Визначення                    | Вказує на те, що ця функція є нефункціональною версією елемента, названого ключем функції   |
|      | Формат значення               | немає   |
|      | Примітки                      | Кваліфікатор "pseudo" має використовуватись для опису   |

нефункціональних генів, які формально не описуються як псевдогени, наприклад, кодуєча послідовність не має трансляції з інших причин ніж псевдогенетичні події. Інші причини можуть включати помилки секвенування або складання

Для опису псевдогенів повинен використовуватись кваліфікатор "pseudogene" із зазначенням типу (TYPE) псевдогена.

|      |                               |  |
|------|-------------------------------|--|
| 6.54 | Кваліфікатор                  | pseudogene   |
|      | Визначення                    | Вказує на те, що ця функція є псевдогеном елемента, названого ключем функції   |
|      | Формат обов'язкового значення | TYPE<br>де "TYPE" є одним із таких стандартизованих термінів або висловів:<br>processed<br>unprocessed<br>unitary<br>allelic<br>unknown  |
|      | Приклади                      | <INSDQualifier_value>processed</INSDQualifier_value><br><INSDQualifier_value>unprocessed</INSDQualifier_value><br><INSDQualifier_value>unitary</INSDQualifier_value><br><INSDQualifier_value>allelic</INSDQualifier_value><br><INSDQualifier_value>unknown</INSDQualifier_value>   |
|      | Примітки                      | Визначення значень "TYPE":<br>processed - псевдоген з'явився унаслідок зворотної транскрипції мРНК у кДНК з подальшою реінтеграцією в геном. Таким чином, він утратив будь-яку інтронну/екзонну структуру і можливо набув псевдо-полі (A) хвіст.<br>unprocessed - псевдоген з'явився з копії батьківського гена внаслідок дублювання з подальшим накопиченням випадкових мутацій. Зміни, порівняно з їх функціональним гомологом, включають вставки, видалення, передчасні стоп-кодони, зсуви рамки і більшу частку несинонімічних (проти синонімічних) замінів.<br>unitary - псевдоген не має першоджерела. Він є оригінальним геном, функціональним в одному виді, проте порушеним якимось чином в іншому виді або штамі.<br>allelic - (унітарний) псевдоген, стабільний в популяції та, щонайважливіше, має функціональний альтернативний алель, що також знаходиться у популяції. Тобто, один штам може мати ген, а інший штам може мати псевдоген. Гаплотипи МНС мають алельні псевдогени.<br>unknown - подавець не знає способу псевдогенізації. |
| 6.55 | Кваліфікатор                  | rearranged   |
|      | Визначення                    | Послідовність, представлена у записі, піддавалася соматичній перебудові як частині адаптивної імунної реакції; вона не є неперебудованою послідовністю, успадкованою від батьківської генеративної лінії   |
|      | Формат значення               | немає  |
|      | Примітки                      | Кваліфікатор "rearranged" не повинен використовуватись для анотування хромосомних перебудов, які не беруть участь в адаптивній імунній реакції; кваліфікатори "germline" і "rearranged" не повинні використовуватись в одній функції "source"; кваліфікатори "germline" і  |

|      |                               |   |
|------|-------------------------------|---|
|      |                               | "rearranged" повинні використовуватись лише для молекул, здатних до соматичної перебудови як частини адаптивної імунної реакції; до них належать локуси Т-клітинних рецепторів (ТКР) та імуноглобулінові локуси у щелепних хребетних, а також непов'язані локуси варіабельних лімфоцитних рецепторів (ВЛР) у безщелепних риб (міног і міксин); кваліфікатори "germline" і "rearranged" не повинні використовуватись за межами класу Craniata (taxid=89593)  |
| 6.56 | Кваліфікатор                  | recombination_class   |
|      | Визначення                    | структурований опис класифікації області гарячих точок рекомбінації в межах послідовності   |
|      | Формат обов'язкового значення | TYPE<br>де "TYPE" є одним із таких стандартизованих термінів або висловів:<br>meiotic<br>mitotic_<br>non_allelic_homologous<br>chromosome_breakpoint<br>other   |
|      | Приклади                      | <INSDQualifier_value>meiotic</INSDQualifier_valu e><br><INSDQualifier_value>chromosome_breakpoint</INSDQualifier_val ue>  |
|      | Примітки                      | Певні класи рекомбінації, позначення яких відсутні у стандартизованій лексиці для кваліфікатора "recombination_class", повинні бути описані шляхом наведення "other" як значення кваліфікатора "recombination_class" і надання відповідного стислого пояснення в кваліфікаторі "note"   |
| 6.57 | Кваліфікатор                  | regulatory_class  |
|      | Визначення                    | структурований опис класифікації транскрипційних, трансляційних, реплікаційних та регуляторних елементів, що стосуються структури хроматину, в послідовності  |
|      | Формат обов'язкового значення | TYPE<br>де "TYPE" є одним із таких стандартизованих термінів або висловів:<br>attenuator<br>CAAT_signal<br>DNase_I_hypersensitive_site<br>enhancer<br>enhancer_blocking_element<br>GC_signal<br>imprinting_control_region<br>insulator<br>locus_control_region<br>matrix_attachment_region<br>minus_35_signal<br>minus_10_signal<br>polyA_signal_sequence<br>promoter<br>recoding_stimulatory_region<br>replication_regulatory_region<br>response_element<br>ribosome_binding_site<br>riboswitch<br>silencer<br>TATA_box<br>terminator<br>transcriptional_cis_regulatory_region |

|          |                               |  |
|----------|-------------------------------|--|
|          |                               | other  |
| Приклади |                               | <INSDQualifier_value>promoter</INSDQualifier_value><br><INSDQualifier_value>enhancer</INSDQualifier_value><br><INSDQualifier_value>ribosome_binding_site</INSDQualifier_value>   |
| Примітки |                               | Певні регуляторні класи, позначення яких відсутні у стандартизованій лексиці для кваліфікатора "regulatory_class", повинні бути описані шляхом наведення "other" як значення кваліфікатора "regulatory_class" і надання відповідного стислого пояснення в кваліфікаторі "note"                   |
| 6.58     | Кваліфікатор                  | replace  |
|          | Визначення                    | Вказує на те, що послідовність, ідентифікована в локалізації функції, заміщується послідовністю, наведеною у значенні кваліфікатора; якщо кваліфікатор не містить жодної послідовності (тобто, не має значення), це означає видалення  |
|          | Формат обов'язкового значення | довільний текст  |
|          | Приклади                      | <INSDQualifier_value>a</INSDQualifier_value><br><INSDQualifier_value></INSDQualifier_value> - для видалення  |
| 6.59     | Кваліфікатор                  | ribosomal_slippage   |
|          | Визначення                    | Під час білкової трансляції певні послідовності можуть програмувати рибосоми на перехід до альтернативної рамки зчитування за допомогою механізму, відомого як рибосомне прослизання   |
|          | Формат значення               | немає  |
|          | Примітки                      | Оператор "join", наприклад: [join(486..1784,1787..4810)] повинен використовуватись в локалізації функції кодуючої послідовності для позначення локалізації "ribosomal_slippage"  |
| 6.60     | Кваліфікатор                  | rpt_family   |
|          | Визначення                    | Тип повторюваної послідовності; наприклад, "Alu" або "Kpn"   |
|          | Формат обов'язкового значення | довільний текст  |
|          | Приклади                      | <INSDQualifier_value>Alu</INSDQualifier_value>   |
| 6.61     | Кваліфікатор                  | rpt_type   |
|          | Визначення                    | Структура та розподіл повторюваної послідовності   |
|          | Формат обов'язкового значення | Один із таких стандартизованих термінів або висловів:<br>tandem<br>direct<br>inverted<br>flanking<br>nested<br>terminal dispersed<br>long terminal repeat<br>non_ltr_retrotransposon_polymeric_tract<br>centromeric_repeat<br>telomeric_repeat x_element_combinatorial_repeat<br>y_prime_element |

|                               |  |                |
|-------------------------------|--|----------------|
| Приклади                      | <pre>other &lt;INSDQualifier_value&gt;inverted&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;long_terminal_repeat&lt;/INSDQualifier_value&gt;</pre>   |                |
| Примітки                      | <p>Визначення значень:</p> <p>tandem - повтор, суміжний з іншим повтором у тій самій орієнтації;</p> <p>direct - повтор, не завжди суміжний, проте у тій самій орієнтації;</p> <p>inverted - пара повторів, що зустрічається у зворотній орієнтації один до одного у тій самій молекулі;</p> <p>flanking - повтор поза межами послідовності, для якої він є функціонально важливим (наприклад, цільовий сайт вставки транспозонів);</p> <p>nested - повтор, який порушується шляхом вставки іншого елемента;</p> <p>dispersed - повтор, розосереджений по всьому геному;</p> <p>terminal - повтор на кінцях і в межах послідовності, для якої він є функціонально важливим (наприклад, довгий кінцевий повтор транспозонів);</p> <p>long_terminal_repeat - послідовність, яка безпосередньо повторюється на обох кінцях певної послідовності, зокрема, що зустрічається в ретровірусах;</p> <p>non_ltr_retrotransposon_polymeric_tract - полімерний тракт, такий як полі(дА), в межах недовгого кінцевого повтору (LTR) ретротранспозону;</p> <p>centromeric_repeat - область повтору, що знаходиться в межах модульного центромера;</p> <p>telomeric_repeat - область повтору, що знаходиться в межах теломера;</p> <p>x_element_combinatorial_repeat - область повтору, розташована між елементом "X" та теломером або суміжним елементом "Y"; y_prime_element - область повтору, розташована поряд з теломерними повторами або комбінаторними повторами елемента "X" як одна копія або як тандемний повтор від двох до чотирьох копій;</p> <p>other - повтор, що виявляє важливі властивості, які не можуть бути описані іншими значеннями.</p> |                |
| 6.62                          | Кваліфікатор   | rpt_unit_range |
| Визначення                    | Локалізація повторюваної одиниці, виражена як діапазон   |                |
| Формат обов'язкового значення | <base_range>, де <base_range> є першою та останньою основами (розділеними двома крапками) повторюваної одиниці   |                |
| Приклади                      | <INSDQualifier_value>202..245</INSDQualifier_value>  |                |
| Примітки                      | Використовується для позначення меж основи послідовності, що є повторюваною одиницею, в межах ділянки, визначеної ключами функцій "oriT" і "repeat_region".  |                |
| 6.63                          | Кваліфікатор   | rpt_unit_seq   |
| Визначення                    | Ідентифікація повторюваної послідовності   |                |
| Формат обов'язкового значення | довільний текст  |                |
| Приклади                      | <pre>&lt;INSDQualifier_value&gt;aagggc&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;ag(5)tg(8)&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;(AAAGA)6(AAAA)1(AAAGA)12&lt;/INSDQualifier_value&gt;</pre>   |                |
| Примітки                      | Використовується для позначення буквенної послідовності, що становить повторювану одиницю в межах області,   |                |

|      |                               |  |
|------|-------------------------------|--|
|      |                               | визначеної ключами функцій "orit" і "repeat_region"  |
| 6.64 | Кваліфікатор                  | satellite  |
|      | Визначення                    | Кваліфікатор ДНК-маркера сателіта, що складається з множини тандемних повторів (ідентичних або пов'язаних) короткої базової повторюваної одиниці   |
|      | Формат обов'язкового значення | <satellite_type>[:<class>][ <identifier>], де <satellite_type> є одним із наступних термінів: satellite; microsatellite; minisatellite   |
|      | Приклади                      | <INSDQualifier_value>satellite: Sla</INSDQualifier_value><br><INSDQualifier_value>satellite: alpha</INSDQualifier_value><br><INSDQualifier_value>satellite: III</INSDQualifier_value> gamma<br><INSDQualifier_value>microsatellite: DC130</INSDQualifier_value>                  |
|      | Примітки                      | Багато сателітів мають нуклеотидний склад або інші властивості, які відрізняються від складу і властивостей решти генома, що дозволяє ідентифікувати їх.   |
| 6.65 | Кваліфікатор                  | segment  |
|      | Визначення                    | Назва секвенованого вірусного або бактеріофагового фрагменту   |
|      | Формат обов'язкового значення | довільний текст  |
|      | Приклади                      | <INSDQualifier_value>6</INSDQualifier_value>   |
| 6.66 | Кваліфікатор                  | serotype   |
|      | Визначення                    | Серологічне різноманіття видів, що відрізняється своїми антигенними властивостями  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|      | Приклади                      | <INSDQualifier_value>B1</INSDQualifier_value>  |
|      | Примітки                      | Використовується лише з ключем функції "source"; Бактеріологічний кодекс рекомендує використовувати термін 'серовар' замість терміну 'серотип' для прокаріотів; див. Доповнення 10.В "Внутрішньовидові терміни" Міжнародного кодексу номенклатури бактерій (редакція 1990 року). |
| 6.67 | Кваліфікатор                  | serovar  |
|      | Визначення                    | Серологічне різноманіття видів (зазвичай прокаріотична), що характеризується своїми антигенними властивостями  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)  |
|      | Приклади                      | <INSDQualifier_value>O157:H7</INSDQualifier_value>   |
|      | Примітки                      | Використовується лише з ключем функції "source"; Бактеріологічний кодекс рекомендує використовувати термін 'серовар' замість терміну 'серотип' для прокаріотів; див. Доповнення 10.В "Внутрішньовидові терміни" Міжнародного кодексу номенклатури бактерій (редакція 1990 року). |
| 6.68 | Кваліфікатор                  | sex  |

|      |                               |   |
|------|-------------------------------|---|
|      | Визначення                    | Стать організму, з якого було отримано послідовність; стать застосовується до еукаріотичних організмів, які здатні до мейозу і мають статеві диморфні гамети  |
|      | Формат обов'язкового значення | довільний текст<br>Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур   |
|      | Приклади                      | <INSDQualifier_value>female</INSDQualifier_value><br><INSDQualifier_value>male</INSDQualifier_value><br><INSDQualifier_value>hermaphrodite</INSDQualifier_value><br><INSDQualifier_value>unisexual</INSDQualifier_value><br><INSDQualifier_value>bisexual</INSDQualifier_value><br><INSDQualifier_value>asexual</INSDQualifier_value><br><INSDQualifier_value>monoecious</INSDQualifier_value><br><INSDQualifier_value>dioecious</INSDQualifier_value>  |
|      | Примітки                      | Кваліфікатор "sex" має використовуватись (замість кваліфікатора "mating_type") для багатоклітинних тварин, ембріофітів, червоних і бурих водоростей; кваліфікатор "mating_type" має використовуватись (замість кваліфікатора "sex") для бактерій, архей і грибів; ні кваліфікатор "sex", ні кваліфікатор "mating_type" не мають використовуватися для вірусів; за межами таксонів, перелічених вище, має використовуватись кваліфікатор "mating_type", за винятком, якщо значення кваліфікатора взяте із словника у прикладах, наведених вище |
| 6.69 | Кваліфікатор                  | standard_name   |
|      | Визначення                    | Загальноприйнята назва цієї функції   |
|      | Формат обов'язкового значення | довільний текст<br>це значення може вимагати перекладу для міжнародних/національних/регіональних процедур   |
|      | Приклади                      | <INSDQualifier_value>dotted</INSDQualifier_value>   |
|      | Примітки                      | Кваліфікатор "standard_name" використовується для зазначення повної назви гена, а кваліфікатор "gene" – для зазначення символу гена (у прикладі, наведеному вище, значенням кваліфікатора "gene" є "Dt").   |
| 6.70 | Кваліфікатор                  | strain  |
|      | Визначення                    | Штам, з якого було отримано послідовність   |
|      | Формат обов'язкового значення | довільний текст<br>Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур   |
|      | Приклади                      | <INSDQualifier_value>BALB/c</INSDQualifier_value>   |
|      | Примітки                      | Записи функцій, що включають кваліфікатор "strain", не повинні включати кваліфікатор "environmental_sample"   |
| 6.71 | Кваліфікатор                  | sub_clone   |
|      | Визначення                    | Субклон, з якого було отримано послідовність  |
|      | Формат обов'язкового значення | довільний текст<br>Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур   |
|      | Приклади                      | <INSDQualifier_value>lambda-hIL7.20g</INSDQualifier_value>  |
|      | Примітки                      | Функція "source" не повинна містити більше одного кваліфікатора "sub_clone"; для позначення того, що  |



послідовність було отримано з множини субклонів, множина джерел може бути додатково описана, використовуючи ключ функцій "misc\_feature" та кваліфікатор "note"

|      |                               |   |
|------|-------------------------------|---|
| 6.72 | Кваліфікатор                  | sub_species   |
| .    | Визначення                    | Назва підвиду організму, з якого було отримано послідовність  |
|      | Формат обов'язкового значення | довільний текст<br>Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур   |
|      | Приклади                      | <INSDQualifier_value>lactis</INSDQualifier_value>   |
| 6.73 | Кваліфікатор                  | sub_strain  |
| .    | Визначення                    | Назва або ідентифікатор штаму, модифікованого генетично або іншим чином, з якого було отримано послідовність; походить з батьківського штаму (має бути зазначено в кваліфікаторі "strain"). Субштам, з якого було отримано послідовність  |
|      | Формат обов'язкового значення | довільний текст<br>Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур   |
|      | Приклади                      | <INSDQualifier_value>abis</INSDQualifier_value>   |
|      | Примітки                      | Повинен супроводжуватися кваліфікатором "strain" у функції "source"; Якщо батьківський штам не зазначено, модифікований штам має бути описано в кваліфікаторі "strain", а не в кваліфікаторі "substrain". Наприклад, може бути вказано кваліфікатор "strain" із значенням "K-12" та кваліфікатор "substrain" із значенням "MG1655" або ж лише кваліфікатор "strain" із значенням "MG1655" |
| 6.74 | Кваліфікатор                  | tag_peptide   |
| .    | Визначення                    | Локалізація основи, яка кодує поліпептид для маркера протеолізу тмРНК і його термінуючий кодон  |
|      | Формат обов'язкового значення | <base_range>, де у <base_range> вказується перша та остання основи (розділені двома крапками) локалізації для маркера протеолізу  |
|      | Приклади                      | <INSDQualifier_value>90..122</INSDQualifier_value>  |
|      | Примітки                      | Рекомендовано, щоб амінокислотна послідовність, яка відповідає кваліфікатору "tag_peptide", анутовалася частковою функцією "CDS" 5'-кінця; наприклад, кодуєча послідовність з локалізацією <90..122   |
| 6.75 | Кваліфікатор                  | tissue_lib  |
| .    | Визначення                    | Бібліотека тканин, з якої було отримано послідовність   |
|      | Формат обов'язкового значення | довільний текст<br>Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур   |
|      | Приклади                      | <INSDQualifier_value>tissue library<br>772</INSDQualifier_value>  |
| 6.76 | Кваліфікатор                  | tissue_type   |
| .    | Визначення                    | Тип тканини, з якої було отримано послідовність   |
|      | Формат обов'язкового значення | довільний текст<br>Залежно від мови: це значення може вимагати перекладу для  |

|          |                               |   |
|----------|-------------------------------|---|
|          |                               | міжнародних/національних/регіональних процедур  |
| Приклади |                               | <INSDQualifier_value>liver</INSDQualifier_value>  |
| 6.77     | Кваліфікатор                  | transl_except   |
|          | Визначення                    | Трансляційний виняток: один кодон, трансляція якого не відповідає генетичному коду, визначеному в кваліфікаторі "organism" або "transl_table"   |
|          | Формат обов'язкового значення | (pos:location,aa:<amino_acid>), де "<amino_acid>" є трибуквеним скороченням для амінокислоти, закодованої кодоном у положенні "base_range"  |
|          | Приклади                      | <INSDQualifier_value>(pos:213..215,aa:Trp)<br></INSDQualifier_value><br><INSDQualifier_value>(pos:462..464,aa:OTHER)<br></INSDQualifier_value><br><INSDQualifier_value>(pos:1017,aa:TERM)<br></INSDQualifier_value><br><INSDQualifier_value>(pos:2000..2001,aa:TERM)<br></INSDQualifier_value>  |
|          | Примітки                      | Якщо амінокислота не є однією з певних амінокислот, наведених у переліку в Розділі 3 цього Додатка, слід використовувати значення "OTHER" як "<amino_acid>" і вказувати назву нестандартної амінокислоти в кваліфікаторі "note"; для модифікованого амінокислотного селеноцистеїну слід використовувати трибуквене скорочення 'Sec' (однобуквений символ 'U' в амінокислотній послідовності) як "<amino_acid>"; для модифікованого амінокислотного піролізину слід використовувати трибуквене скорочення 'Pyl' (однобуквений символ 'O' в амінокислотній послідовності) як "<amino_acid>"; для часткових кінцевих кодонів, якщо стоп-кодон ТАА доповнений шляхом додавання залишків 3'A до мРНК, одне значення "<base_position>" або "<base_range>" використовується для локалізації (див. третій і четвертий приклади вище) у поєднанні з кваліфікатором "note", де зазначається наступне: 'стоп-кодон, доповнений шляхом додавання залишків 3'A до мРНК'. |
| 6.78     | Кваліфікатор                  | transl_table  |
|          | Визначення                    | Визначення використовуваної таблиці генетичного коду, якщо вона не є універсальною або стандартною таблицею генетичного коду. Таблиці для використання наведено в цьому Додатку.  |
|          | Формат обов'язкового значення | <integer>, де "<integer>" є номером присвоєної таблиці генетичного коду   |
|          | Приклади                      | <INSDQualifier_value>3</INSDQualifier_value>  |
|          | Примітки                      | Якщо кваліфікатор "transl_table" не використовується для додаткового опису ключа функції "CDS", то кодуєча послідовність транслюється за допомогою стандартного коду (тобто Універсального генетичного коду). Винятки з генетичного коду поза межами визначених таблиць наводяться в кваліфікаторі "transl_except".   |
| 6.79     | Кваліфікатор                  | transl_splicing   |
|          | Визначення                    | Вказує на те, що екзони з двох молекул РНК лігуються під час міжмолекулярної реакції для утворення зрілої РНК   |
|          | Формат значення               | немає   |

|                               |   |             |
|-------------------------------|---|-------------|
| Примітки                      | Має використовуватись для таких функцій як CDS, mRNA та інших функцій, отриманих унаслідок транс-сплайсинга. Цей кваліфікатор повинен використовуватись, лише якщо сплайсинг указаний в операторі "join", наприклад,<br><code>join(complement(69611..69724),139856..140087)</code> в локалізації функції  |             |
| 6.80                          | Кваліфікатор  | translation |
| Визначення                    | Однобуквене скорочення назви амінокислотної послідовності, узятє із стандартного (або універсального) генетичного коду або таблиці, як зазначено в кваліфікаторі "transl_table" і як визначено винятком в кваліфікаторі "transl_except"   |             |
| Формат обов'язкового значення | безперервний рядок однобуквених скорочень назв амінокислот з Розділу 3 цього Додатка, "X" повинен використовуватись для AA-винятків   |             |
| Приклади                      | <code>&lt;INSDQualifier_value&gt;MASTFPPWYRGCASTPSLKGLIMCTW&lt;/INSDQualifier_value&gt;</code>  |             |
| Примітки                      | Має використовуватись лише з функцією "CDS"; повинен супроводжуватись кваліфікатором "protein_id", якщо продукт трансляції містить чотири або більше специфічно визначені амінокислоти; визначення і місце розташування таблиць генетичного коду містяться у записі стосовно кваліфікатора "transl_table"; для додаткового опису функції "CDS" допускається використання лише одного з кваліфікаторів "translation", "pseudo" і "pseudogene". |             |
| 6.81                          | Кваліфікатор  | variety     |
| Визначення                    | Різновид (= varietas, офіційний ранг системи Ліннея) організму, з якого було отримано послідовність   |             |
| Формат обов'язкового значення | довільний текст<br>Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур   |             |
| Приклади                      | <code>&lt;INSDQualifier_value&gt;insularis&lt;/INSDQualifier_value&gt;</code>   |             |
| Примітки                      | Кваліфікатор "cultivar" використовується для культивованих різновидів рослин, тобто продуктів штучної селекції; різновиди, крім різновидів рослин і грибів, мають описуватися за допомогою кваліфікатора "note", наприклад, із значенням<br><code>&lt;INSDQualifier_value&gt;breed:Cukorova&lt;/INSDQualifier_value&gt;</code>  |             |

## РОЗДІЛ 7: КЛЮЧІ ФУНКЦІЙ ДЛЯ АМІНОКИСЛОТНИХ ПОСЛІДОВНОСТЕЙ

Цей розділ містить перелік допустимих ключів функцій для використання для амінокислотних послідовностей. Ключі функцій наведені в алфавітному порядку.

|      |                             |  |
|------|-----------------------------|--|
| 7.1. | Ключ функції                | ACT_SITE   |
|      | Визначення                  | Амінокислота (-и), залучена (-і) до активності ферменту  |
|      | Необов'язкові кваліфікатори | NOTE   |
|      | Примітки                    | Кожний амінокислотний залишок активного сайту повинен бути анотований окремо з використанням ключа функції "ACT_SITE". Відповідний номер амінокислотного залишку повинен бути вказаний як дескриптор локалізації в елементі локалізації функції.   |
| 7.2. | Ключ функції                | BINDING  |
|      | Визначення                  | Зв'язувальний сайт для будь-якої хімічної групи (коферментної, простетичної тощо). Хімічні властивості групи зазначаються в кваліфікаторі "note".  |
|      | Обов'язкові кваліфікатори   | NOTE   |
|      | Примітки                    | Приклади значень для кваліфікатора "note": "Neme (covalent)" і "Chloride". Якщо можливо, слід використовувати ключі функцій "CA_BIND", "DNA_BIND", "METAL" та "NP_BIND", а не "BINDING".   |
| 7.3. | Ключ функції                | CA_BIND  |
|      | Визначення                  | Протяжність кальцій-зв'язувальної ділянки  |
|      | Необов'язкові кваліфікатори | NOTE   |
| 7.4. | Ключ функції                | CARBOHYD   |
|      | Визначення                  | Сайт глікозилювання  |
|      | Обов'язкові кваліфікатори   | note   |
|      | Примітки                    | Цей ключ описує кількість випадків приєднання глікану (моно- або полісахариду) до залишку білка. Тип зв'язку (C-, N- або O-linked) з білком вказується в кваліфікаторі "note". Якщо відомо природу відновлюваного термінального цукру, його скорочення зазначається у дужках. Якщо після скорочення стоять три крапки, це вказує на продовження вуглеводного ланцюга. І навпаки, відсутність крапок означає, що моносахарид є зв'язаним. Тип зв'язку (C-, N- або O-linked) з білком вказується в кваліфікаторі "note". Приклади значень, що використовуються в кваліфікаторі "note": N-linked (GlcNAc...); O-linked (GlcNAc); O-linked (Glc...); частково C-linked (Man); O-linked (Ara...). |
| 7.5. | Ключ функції                | CHAIN  |
|      | Визначення                  | Протяжність поліпептидного ланцюга у зрілому білку.  |
|      | Необов'язкові кваліфікатори | note   |

|       |   |  |
|-------|---|--|
| 7.6.  | Ключ функції                                  | COILED   |
|       | Визначення                                    | Протяжність області суперспіралі   |
|       | Необов'язкові кваліфікатори                   | note   |
| 7.7.  | Ключ функції                                  | COMPBIAS   |
|       | Визначення                                    | Протяжність композиційно зміщеної області  |
|       | Необов'язкові кваліфікатори                   | note   |
| 7.8.  | Ключ функції                                  | CONFLICT   |
|       | Визначення                                    | В різних джерелах повідомляється про різні послідовності.  |
|       | Необов'язкові кваліфікатори                   | note   |
|       | Примітки                                      | Приклади значень для кваліфікатора "NOTE": Missing; K -> Q; GSDSE -> RIRLR; V ->A.   |
| 7.9.  | Ключ функції                                  | CROSSLNK   |
|       | Визначення                                    | Пост-трансляційно сформовані амінокислотні зв'язки.  |
|       | Обов'язкові кваліфікатори                     | note   |
|       | Примітки                                      | Ковалентні зв'язки різних типів, утворені між двома білками (міжланцюгові поперечні зв'язки) або між двома частинами одного білка (міжланцюгові поперечні зв'язки); окрім поперечних зв'язків, утворених дисульфідними зв'язками, для яких повинен використовуватись ключ функції "DISULFID". Для міжланцюгового поперечного зв'язку, дескриптором локалізації в елементі локалізації функції є номер залишку амінокислоти, поперечно зв'язаної з іншим білком. Для міжланцюгового поперечного зв'язку, дескриптором локалізації в елементі локалізації функції є номери залишків поперечно зв'язаних амінокислот у форматі "x.y", наприклад "42..50".<br>Кваліфікатор "note" вказує на природу поперечного зв'язку; зазначається принаймні назва кон'югата і назви двох залучених амінокислот. Приклади значень кваліфікатора "note": "Isoglutamyl cysteine thioester (Cys-Gln);" "Beta-methyllanthionine (Cys-Thr);" and "Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin)". |
| 7.10. | Ключ функції                                  | DISULFID   |
|       | Визначення                                    | Дисульфідний зв'язок   |
|       | Необов'язкові<br>Обов'язкові<br>кваліфікатори | note   |
|       | Примітки                                      | Для міжланцюгового дисульфідного зв'язку, дескриптором локалізації в елементі локалізації функції є номер залишку цистеїну, зв'язаного з іншим білком. Для міжланцюгового  |

поперечного зв'язку, дескрипторами локалізації в елементі локалізації функції є номери залишків зв'язаних цистеїнів у форматі "х.у", наприклад "42.50". Для міжланцюгових дисульфідних зв'язків, кваліфікатор "NOTE" вказує на природу поперечного зв'язку, ідентифікуючи інший білок, наприклад "Міжланцюговий зв'язок (між ланцюгами А і В).

|       |                             |  |
|-------|-----------------------------|--|
| 7.11  | Ключ функції                | DNA_BIND   |
|       | Визначення                  | Протяжність ДНК-зв'язувальної ділянки  |
|       | Обов'язкові кваліфікатори   | note   |
|       | Примітки                    | Природа ДНК-зв'язувальної ділянки зазначається в кваліфікаторі "note". Приклади значень для кваліфікатора "note": "Homebox" і "Myb 2".   |
| 7.12. | Ключ функції                | DOMAIN   |
|       | Визначення                  | Протяжність домену, що визначається як певна комбінація вторинних структур, упорядкованих у характерній тривимірній структурі або укладці.   |
|       | Обов'язкові кваліфікатори   | note   |
|       | Примітки                    | Тип домену зазначається в кваліфікаторі "NOTE". За наявності декількох копій домену, домени нумеруються. Приклади значень для кваліфікатора "note": "Ras-GAP" і "Cadherin 1".  |
| 7.13. | Ключ функції                | HELIX  |
|       | Визначення                  | Вторинна структура: Helices, наприклад Alpha-helix; 3(10) helix; або Pi-helix  |
|       | Необов'язкові кваліфікатори | note   |
|       | Примітки                    | Ця функція використовується лише для білків з відомою третинною структурою. Зазначаються лише три типи вторинної структури: спіралі (ключ "HELIX"), бета-ланцюги (ключ "STRAND") і повороти (ключ "TURN"). Залишки, не зазначені в одному із цих класів, знаходяться у структурі "випадкової спіралі" або "петлі". |
| 7.14. | Ключ функції                | INIT_MET   |
|       | Визначення                  | Ініціювальний метіонін   |
|       | Необов'язкові кваліфікатори | note   |
|       | Примітки                    | Дескриптором локалізації в елементі локалізації функції є "1". Цей ключ функції вказує на відщеплення N-кінцевого метіоніну. Ця функція не використовується, якщо ініціювальний метіонін не відщеплюється.   |
| 7.15. | Ключ функції                | INTRAMEM   |
|       | Визначення                  | Протяжність області, яка розташована у мембрані і не виходить за її межі   |

|       |                             |  |
|-------|-----------------------------|--|
|       | Необов'язкові кваліфікатори | note   |
| 7.16. | Ключ функції                | LIPID  |
|       | Визначення                  | Ковалентне зв'язування ліпідного компонента  |
|       | Обов'язкові кваліфікатори   | note   |
|       | Примітки                    | Хімічна природа зв'язаного ліпідного компонента наводяться в кваліфікаторі "note", де принаймні зазначається назва ліпідованої амінокислоти. Приклади значень кваліфікатора "note": "N-myristoyl glycine"; "GPI-anchor amidated serine" та "S-diacylglycerol cysteine".  |
| 7.17. | Ключ функції                | METAL  |
|       | Визначення                  | Сайт зв'язування для іона металу   |
|       | Обов'язкові кваліфікатори   | note   |
|       | Примітки                    | В кваліфікаторі "note" зазначається природа металу. Приклади значень кваліфікатора "note": "Iron(heme axial ligand)"; " i "Copper".  |
| 7.18. | Ключ функції                | MOD_RES  |
|       | Визначення                  | Пост-трансляційна модифікація залишку  |
|       | Обов'язкові кваліфікатори   | note   |
|       | Примітки                    | Хімічна природа модифікованого залишку наводиться в кваліфікаторі "NOTE" із зазначенням принаймні назви пост-трансляційно модифікованої амінокислоти. Якщо модифікована кислота зазначена у Розділі 4 цього Додатка, замість повної назви може використовуватись скорочена. Приклади значень для кваліфікатора "note": "N-acetylalanine"; "3-Nup"; i "MeLys" або "N-6-methyllysine". |
| 7.19. | Ключ функції                | MOTIF  |
|       | Визначення                  | Короткий (до 20-и амінокислот) біологічно досліджуваний мотив послідовності.   |
|       | Необов'язкові кваліфікатори | note   |
| 7.20. | Ключ функції                | MUTAGEN  |
|       | Визначення                  | Сайт, змінений експериментальним шляхом мутагенезу   |
|       | Необов'язкові кваліфікатори | note   |
| 7.21. | Ключ функції                | NON_STD  |

|       |                             |   |
|-------|-----------------------------|---|
|       | Визначення                  | Нестандартна амінокислота   |
|       | Необов'язкові кваліфікатори | note  |
|       | Примітки                    | Цей ключ описує кількість нестандартних амінокислот селеноцистеїну (U) та пірролізину (O) в амінокислотній послідовності.   |
| 7.22. | Ключ функції                | NON_TER   |
|       | Визначення                  | Залишок наприкінці послідовності не є кінцевим (термінальним)   |
|       | Необов'язкові кваліфікатори | note  |
|       | Примітки                    | Якщо це стосується положення 1, то перше положення не є N-кінцем усієї молекули. Якщо це стосується останнього положення, то це положення не є C-кінцем усієї молекули. |
| 7.23. | Ключ функції                | NP_BIND   |
|       | Визначення                  | Протяжність ділянки зв'язування нуклеотидфосфату  |
|       | Обов'язкові кваліфікатори   | note  |
|       | Примітки                    | Властивості нуклеотидфосфату зазначаються в кваліфікаторі "note". Приклади значень для кваліфікатора "note": "ATP" і "FAD".   |
| 7.24. | Ключ функції                | PEPTIDE   |
|       | Визначення                  | Протяжність вивільненого активного пептиду  |
|       | Необов'язкові кваліфікатори | note  |
| 7.25. | Ключ функції                | PROPEP  |
|       | Визначення                  | Протяжність пропептиду  |
|       | Необов'язкові кваліфікатори | note  |
| 7.26. | Ключ функції                | REGION  |
|       | Визначення                  | Протяжність дослідної ділянки у послідовності   |
|       | Необов'язкові кваліфікатори | note  |
| 7.27. | Ключ функції                | REPEAT  |
|       | Визначення                  | Протяжність внутрішнього повторення послідовності   |
|       | Необов'язкові кваліфікатори | note  |



|       |                             |   |
|-------|-----------------------------|---|
| 7.28. | Ключ функції                | SIGNAL  |
|       | Визначення                  | Протяжність сигнальної послідовності (препептиду)   |
|       | Необов'язкові кваліфікатори | note  |
| 7.29. | Ключ функції                | SITE  |
|       | Визначення                  | Будь-який дослідний амінокислотний сайт у послідовності, не визначений іншим ключем функції. Він також може застосовуватися до амінокислотного зв'язку, представленого положеннями двох амінокислот, що фланкують   |
|       | Обов'язкові кваліфікатори   | note  |
|       | Примітки                    | Якщо ключ функції "SITE" використовується для анотування модифікованої амінокислоти, значенням кваліфікатора "note" повинне бути одне із скорочень, наведених у Розділі 4 цього Додатка, або повна, нескорочена назва модифікованої амінокислоти.   |
| 7.30. | Ключ функції                | SOURCE  |
|       | Визначення                  | Ідентифікує джерело послідовності; цей ключ є обов'язковим; кожна послідовність повинна мати одну функцію "SOURCE", що охоплює всю послідовність  |
|       | Обов'язкові кваліфікатори   | MOL_TYPE ORGANISM   |
|       | Необов'язкові кваліфікатори | note  |
| 7.31. | Ключ функції                | STRAND  |
|       | Визначення                  | Вторинна структура: бета-ланцюг, наприклад бета-ланцюг з водневим зв'язком або залишок в ізольованому бета-містку   |
|       | Необов'язкові кваліфікатори | note  |
|       | Примітки                    | Ця функція застосовується лише до білків з відомою третинною структурою. Зазначаються лише три типи вторинної структури: спіралі (ключ "HELIX"), бета-ланцюги (ключ "STRAND") і повороти (ключ "TURN"). Залишки, не зазначені в одному із цих класів, знаходяться у структурі "випадкової спіралі" або "петлі". |
| 7.32. | Ключ функції                | TOPO_DOM  |
|       | Визначення                  | Топологічний домен  |
|       | Необов'язкові кваліфікатори | note  |
| 7.33. | Ключ функції                | TRANSMEM  |
|       | Визначення                  | Протяжність трансмембранної області   |
|       | Необов'язкові кваліфікатори | note  |
| 7.34. | Ключ функції                | TRANSIT   |
|       | Визначення                  | Протяжність транзитного пептиду (мітохондрії, хлоропласту, тилакоїду, ціанели, пероксисоми тощо)  |

|       |                             |   |
|-------|-----------------------------|---|
|       | Необов'язкові кваліфікатори | note  |
| 7.35. | Ключ функції                | TURN  |
|       | Визначення                  | Петлі вторинної структури, наприклад петля з Н-зв'язком (3- петля, 4-петля або 5-петля)   |
|       | Необов'язкові кваліфікатори | note  |
|       | Примітки                    | Ця функція застосовується лише до білків з відомою третинною структурою. Зазначаються лише три типи вторинної структури: спіралі (ключ "HELIX"), бета-ланцюги (ключ "STRAND") і повороти (ключ "TURN"). Залишки, не зазначені в одному із цих класів, знаходяться у структурі "випадкової спіралі" або "петлі". |
| 7.36. | Ключ функції                | UNSURE  |
|       | Визначення                  | Невизначеності в послідовності  |
|       | Необов'язкові кваліфікатори | note  |
|       | Примітки                    | Використовується для опису ділянки (-ок) амінокислотної послідовності, щодо представлення якої (-их) автори не впевнені.  |
| 7.37. | Ключ функції                | VARIANT   |
|       | Визначення                  | Автори повідомляють про існування варіантів послідовності.  |
|       | Необов'язкові кваліфікатори | note  |
| 7.38. | Ключ функції                | VAR_SEQ   |
|       | Визначення                  | Опис варіантів послідовності, отриманих шляхом альтернативного сплайсингу, використанням альтернативного промоутера, альтернативної ініціації та рибосомного фреймшифтингу.   |
|       | Необов'язкові кваліфікатори | note  |
| 7.39. | Ключ функції                | ZN_FING   |
|       | Визначення                  | Протяжність ділянки "цинкові пальці"  |
|       | Обов'язкові кваліфікатори   | note  |
|       | Примітки                    | Тип "цинкових пальців" вказується в кваліфікаторі "note". Наприклад: "GATA-type" і "NR C4-type".  |

## РОЗДІЛ 8: КВАЛІФІКАТОРИ ДЛЯ АМІНОКИСЛОТНИХ ПОСЛІДОВНОСТЕЙ

Цей розділ містить перелік допустимих кваліфікаторів для використання для амінокислотних послідовностей.

Якщо формат значення – це довільний текст, який визначається як залежний від мови, слід використовувати один з наступних елементів:

- 1) елемент `INSDQualifier_value`; або
- 2) елемент `NonEnglishQualifier_value`; або
- 3) обидва елементи `INSDQualifier_value` та `NonEnglishQualifier_value`.

Якщо формат значення – це щось інше, ніж "none", але визначений як залежний від мови довільний текст, то повинен використовуватись елемент `INSDQualifier_value`, а елемент `NonEnglishQualifier_value` не повинен використовуватись.

СЛІД ЗАУВАЖИТИ, ЩО: Будь-яке значення кваліфікатора, надане для кваліфікатора з мовно- залежним форматом значення довільного тексту, може вимагати перекладу для міжнародних, національних або регіональних процедур. Кваліфікатори, наведені в наступній таблиці, вважаються такими, що мають значення залежного від мови довільного тексту:

Таблиця 6: Список значень кваліфікаторів із залежними від мови значеннями довільного тексту для амінокислотних послідовностей

| Секція | Кваліфікатор залежного від мови довільного тексту |
|--------|---|
| 8.2    | note  |
| 8.3    | organism  |

|      |                               |   |
|------|-------------------------------|---|
| 8.1. | Кваліфікатор                  | <code>mol_type</code>   |
|      | Визначення                    | Тип молекул послідовності <i>in vivo</i>  |
|      | Формат обов'язкового значення | білок   |
|      | Приклад                       | <code>&lt;INSDQualifier_value&gt;protein&lt;/INSDQualifier_value&gt;</code>   |
|      | Примітки                      | Кваліфікатор "mol_type" є обов'язковим для ключа функції "source".  |
| 8.2. | Кваліфікатор                  | <code>note</code>   |
|      | Визначення                    | Будь-які примітки або додаткова інформація  |
|      | Формат обов'язкового значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для міжнародних/національних/регіональних процедур)         |
|      | Приклад                       | <code>&lt;INSDQualifier_value&gt;Heme (covalent)&lt;/INSDQualifier_value&gt;</code>   |
|      | Примітки                      | Кваліфікатор "note" є обов'язковим для таких ключів функцій:<br><code>BINDING; CARBOHYD; CROSSLNK; DISULFID; DNA BIND; DOMAIN;</code> |

---

LIPID; METAL; MOD\_RES; NP\_BIND, SITE і ZN\_FING.

---

|      |                                     |  |
|------|-------------------------------------|--|
| 8.3. | Кваліфікатор                        | organism   |
|      | Визначення                          | Наукова назва організму, з якого отримано пептид   |
|      | Формат<br>обов'язкового<br>значення | довільний текст<br>(Залежно від мови: це значення може вимагати перекладу для<br>міжнародних/національних/регіональних процедур) |
|      | Приклад                             | <INSDQualifier_value>Homo sapiens</INSDQualifier_value>  |
|      | Примітки                            | Кваліфікатор "organism" є обов'язковим для ключа функції<br>"source".  |

## РОЗДІЛ 9: ТАБЛИЦІ ГЕНЕТИЧНОГО КОДУ

У Таблиці 7 наведено генетичні коди для трансляції кодуючих послідовностей. Значенням кваліфікатора `trans_table` є номер, присвоєний відповідній таблиці генетичних кодів. Якщо функція "CDS" описується кваліфікатором трансляції, а не кваліфікатором `transl_table`, для трансляції слід використовувати таблицю "1 – Standard Code". (Примітка: Таблиць генетичного коду 7, 8, 15 і 17-20 не існує, тому ці номери відсутні у Таблиці 7).

Таблиця 7: Таблиці генетичного коду

| <b>1 – Стандартний код</b>   |   |
|--|---|
| AAs  | = FLLSSSSYY**CC*WLLLLPPPHHQRRRIIMTTTTNNKSSRRVVVAAAADDEEGGGG     |
| Початок  | = --- M.....M.....M.....  |
| Основа1  | = tttttttttttttcccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggggg |
| Основа2  | = ttttcccaaaaggggttttcccaaaaggggttttcccaaaaggggttttcccaaaagggg  |
| Основа3  | = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag  |
| <b>2 – Мітохондріальний код хребетних</b>  |   |
| AAs  | = FLLSSSSYY**CCWLLLLPPPHHQRRRIIMTTTTNNKSS**VVVAAAADDEEGGGG      |
| Початок  | = .....MMMM.....M.....  |
| Основа1  | = tttttttttttttcccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggggg |
| Основа2  | = ttttcccaaaaggggttttcccaaaaggggttttcccaaaaggggttttcccaaaagggg  |
| Основа3  | = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag  |
| <b>3 – Мітохондріальний код дріжджів</b>   |   |
| AAs  | = FLLSSSSYY**CCWTTTTPPPHHQRRRIIMTTTTNNKSSRRVVVAAAADDEEGGGG      |
| Початок  | = .....MM.....M.....  |
| Основа1  | = tttttttttttttcccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggggg |
| Основа2  | = ttttcccaaaaggggttttcccaaaaggggttttcccaaaaggggttttcccaaaagggg  |
| Основа3  | = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag  |
| <b>4 – Мітохондріальний код плісняви, найпростіших, кишковопорожнинних та код мікоплазми/спіроплазми</b> |   |
| AAs  | = FLLSSSSYY**CCWLLLLPPPHHQRRRIIMTTTTNNKSSRRVVVAAAADDEEGGGG      |
| Початок  | = .. MM.....M.....MMMM.....M.....                               |
| Основа1  | = tttttttttttttcccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggggg |
| Основа2  | = ttttcccaaaaggggttttcccaaaaggggttttcccaaaaggggttttcccaaaagggg  |
| Основа3  | = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag  |
| <b>5 – Мітохондріальний код безхребетних</b>   |   |
| AAs  | = FLLSSSSYY**CCWLLLLPPPHHQRRRIIMTTTTNNKSSVVVAAAADDEEGGGG        |
| Початок  | = ... M.....MMMM.....M.....                                     |
| Основа1  | = tttttttttttttcccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggggg |
| Основа2  | = ttttcccaaaaggggttttcccaaaaggggttttcccaaaaggggttttcccaaaagggg  |
| Основа3  | = tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag  |
| <b>6 – Ядерний код інфузорій, дазикладильних та гексамітів</b>   |   |
| AAs  | = FLLSSSSYYQCC*WLLLLPPPHHQRRRIIMTTTTNNKSSRRVVVAAAADDEEGGGG      |
| Початок  | = .....M.....   |
| Основа1  | = tttttttttttttcccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggggg |
| Основа2  | = ttttcccaaaaggggttttcccaaaaggggttttcccaaaaggggttttcccaaaagggg  |





|  |   |  |
|--|---|--|
| Основа1  | = | ttttttttttttcccccccccccccaaaaaaaaaaaaaaaggggggggggggggg  |
| Основа2  | = | ttttcccaaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg  |
| Основа3  | = | tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag   |
| <b>29 - Ядерний код Mesodinium</b>                       |   |  |
| AAs  | = | FFLLSSSSYYYYCC*WLLLLPPPHHQRRRRRIIMTTTTNNKSSRRVVVAAAADDEEGGGG   |
| Початок  | = | .....M.....  |
| Основа1  | = | ttttttttttttcccccccccccccaaaaaaaaaaaaaaaggggggggggggggg  |
| Основа2  | = | ttttcccaaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg  |
| Основа3  | = | tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag   |
| <b>30 - Ядерний код перитрихів</b>                       |   |  |
| AAs  | = | FFLLSSSSYYEECC*WLLAPPPPHHQRRRRRIIMTTTTNNKSSRRVVVAAAADDEEGGGWLLLLPP<br>PPHHQQR RRRIIMTTTTNNKSSRRVVVAAAADDEEGGGG |
| Початок  | = | .....M.....  |
| Основа1  | = | ttttttttttttcccccccccccccaaaaaaaaaaaaaaaggggggggggggggg  |
| Основа2  | = | ttttcccaaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg  |
| Основа3  | = | tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag   |
| <b>31 - Ядерний код Blastocrithidia</b>                  |   |  |
| AAs  | = | FFLLSSSSYYEECCWLLLLPPPHHQRRRRRIIMTTTTNNKSSRRVVVAAAADDEEGGGG  |
| Початок  | = | .....**.....M.....   |
| Основа1  | = | ttttttttttttcccccccccccccaaaaaaaaaaaaaaaggggggggggggggg  |
| Основа2  | = | ttttcccaaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg  |
| Основа3  | = | tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag   |
| <b>33 - Мітохондріальний код UAA-Tyr Cephalodiscidae</b> |   |  |
| AAs  | = | FFLLSSSSYYY*CCWLLLLPPPHHQRRRRRIIMTTTTNNKSSKVVVAAAADDEEGGGG   |
| Початок  | = | --M.....*.....M.....M.....M.....   |
| Основа1  | = | ttttttttttttcccccccccccccaaaaaaaaaaaaaaaggggggggggggggg  |
| Основа2  | = | ttttcccaaaaaggggtttccccaaaaggggtttccccaaaaggggtttccccaaaagggg  |
| Основа3  | = | tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag   |

[Додаток II наведено далі]



## ДОДАТОК II

### ВИЗНАЧЕННЯ ТИПУ ДОКУМЕНТА (DTD) ДЛЯ ПЕРЕЛІКУ ПОСЛІДОВНОСТЕЙ

#### Версія 1.3

*Редакція, схвалена Комітетом зі стандартів ВОІВ (КСВ)  
на його одинадцятій сесії 8 грудня 2023 року*

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Додаток II до стандарту ВОІВ ST.26, визначення типу документа (DTD) для
переліку послідовностей

Цю сутність можна ідентифікувати за допомогою кваліфікатора "PUBLIC":
*****PUBLIC
"--//WIPO//DTD SEQUENCE LISTING 1.3//EN" "ST26SequenceListing_V1_3.dtd"
*****
* URL-адреса публічного DTD
*https://www.wipo.int/standards/dtd/ST26SequenceListing_V1_3.dtd
*****

*Перегляд Додатка II до стандарту ВОІВ ST.26 був затверджений Комітетом зі
*стандартів ВОІВ (КСВ) на його десятій сесії.

*****
* КОНТАКТНА ІНФОРМАЦІЯ
*****
*
*xml.standards@wipo.int
*
*****
* ПРИМІТКИ
*****
*
*Частина з даними щодо послідовності є підрозділом INSDC DTD V.1.5, що містить лише
*вимоги стандарту ВОІВ ST.26.
*
*****
* ІСТОРІЯ ПЕРЕГЛЯДУ
* *****
2022-11-25: Коментар щодо імені файлу затверджено на десятій сесії КСВ (без
оновлення номера версії)
2021-11-05: Перегляд версії 1.3 затверджено на дев'ятій сесії КСВ (незначні
зміни до коментарів)
2020-05-20: Версія 1.3 затверджена на восьмій сесії КСВ.
Зміни:
- Необов'язковий атрибут originalFreeTextLanguageCode доданий до
<ST26SequenceListing>, щоб дозволити заявникам вказати мову довільного тексту
оригінальному переліку послідовностей.
- Необов'язковий атрибут nonEnglishFreeTextLanguageCode доданий до
<ST26SequenceListing>, щоб дозволити заявникам вказувати мову довільного тексту в
елементі <NonEnglishQualifier_value>.
- Необов'язковий атрибут id, доданий до INSDQualifier для полегшення порівняння
значень, що залежать від мови, між переліками послідовностей.
- Необов'язковий елемент <NonEnglishQualifier_value> доданий до елемента
<INSDQualifier>, щоб дозволити заявникам набирати неанглійською мовою кваліфікатори, що
залежать від мови, за допомогою символів, зазначених у параграфі40(a) основного тексту
стандарту ВОІВ ST.26.
2018-10-19: Версія 1.2 затверджена на шостій сесії КСВ.
```

Зміни:

<INSDQualifier\*> замінено на <INSDQualifier+> для приведення у відповідність з бізнес-потребами та рекомендаціями Національного центру біотехнологічної інформації (NCBI) (елемент INSDFeature\_qual (за наявності) повинен мати один або більше елементів INSDQualifier).

2017-06-02: Версія 1.1 затверджена на п'ятій сесії КСВ

Зміни:

Додано примітки до <INSDSeq\_length>, <INSDSeq\_division> та <INSDSeq\_sequence> для пояснення причини відмінностей між INSDC DTD v.1.5 та Переліком послідовностей за стандартом VOIB ST.26 DTD V1\_1.

2016-03-24: Версія 1.0 затверджена на відновленій четвертій сесії КСВ

2014-03-11: Остаточний проєкт для затвердження.

\*\*\*\*\*

ST26SequenceListing

\*\*\*\*\*

\* КОРЕНЕВИЙ ЕЛЕМЕНТ

\*\*\*\*\*

-->

```
<!ELEMENT ST26SequenceListing ((ApplicantFileReference | (ApplicationIdentification,
ApplicantFileReference?)), EarliestPriorityApplicationIdentification?,
(ApplicantName, ApplicantNameLatin?)?, (InventorName, InventorNameLatin?)?,
InventionTitle+, SequenceTotalQuantity, SequenceData+)>
```

<!--Елементи "ApplicantName" та "InventorName" є необов'язковими у цьому DTD для спрощення конвертації між різними схемами кодування-->

<!--originalFreeTextLanguageCode:

Код мови (див. посилання в параграфі 9 на ISO 639-1:2002) для єдиної мови оригіналу, якою були підготовлені кваліфікатори залежного від мови довільного тексту (NonEnglishQualifier\_value).

-->

<! - nonEnglishFreeTextLanguageCode:

Код мови (див. посилання в параграфі 9 на ISO 639-1: 2002) для мови, якій наразі відповідають кваліфікатори залежного від мови довільного тексту (NonEnglishQualifier\_value).

-->

<!--fileName:

За замовчуванням ім'я файлу буде встановлено для значення, наведеного для назви проєкту у WIPO Sequence. Якщо значення ідентичне фактичному імені XML-файлу відповідно до стандарту VOIB ST.26, слід зазначити, що відомства можуть встановлювати до імені файлу, що застосовується, свої вимоги, які можуть обмежувати символи, допустимі у поданих електронних файлах. Також допускаються відмінності між значенням атрибута імені файлу та фактичним іменем файлу. Більш детальна інформація про правила відомств щодо іменування електронних файлів міститься у WIPO Sequence та Базі знань ST.26.

---->

<!ATTLIST ST26SequenceListing

dtdVersion CDATA #REQUIRED fileName CDATA

#IMPLIED

softwareName CDATA #IMPLIED softwareVersion

CDATA

#IMPLIED

productionDate CDATA

#IMPLIED

originalFreeTextLanguageCode CDATA #IMPLIED nonEnglishFreeTextLanguageCode

CDATA #IMPLIED

>

<!--ApplicantFileReference

Посилання на файл заявника або представника обов'язкове, якщо не вказано ідентифікаційні дані заявки.

-->

<!ELEMENT ApplicantFileReference (#PCDATA)>

<!--ApplicationIdentification

За наявності - ідентифікаційні дані заявки, стосовно якої подано перелік послідовностей.

-->

<!ELEMENT ApplicationIdentification (IPOfficeCode, ApplicationNumberText, FilingDate?)>

<!--EarliestPriorityApplicationIdentification

Ідентифікація заявки з більш раннім пріоритетом, що містить елементи "IPOfficeCode", "ApplicationNumberText" і "FilingDate".

-->

```
<!ELEMENT EarliestPriorityApplicationIdentification (IPOfficeCode,
ApplicationNumberText, FilingDate?)>

<!--ApplicantName
Ім'я першого вказаного заявника, представлене символами, визначеними у параграфі 40(a)
основної частини стандарту BOIB ST.26.
-->
<!--languageCode: Відповідний код мови за стандартом ISO 639-1 "Коди для представлення
назв мов. Частина 1: Код Альфа-2"
-->
<!ELEMENT ApplicantName (#PCDATA)>
<ATTLIST ApplicantName
languageCode CDATA #REQUIRED
>
<!--ApplicantNameLatin
Якщо елемент "ApplicantName" наведено символами, не зазначеними у параграфі 40(b),
необхідно навести переклад або транслітерацію імені першого зазначеного заявника
символами, зазначеними у параграфі 40(b) основної частини стандарту BOIB ST.26.
-->
<!ELEMENT ApplicantNameLatin (#PCDATA)>
<!--InventorName
Ім'я першого вказаного винахідника, представлене символами, визначеними у параграфі
40(a). -->
<!--languageCode: Відповідний код мови за стандартом ISO 639-1 "Коди для представлення
назв мов. Частина 1: Код Альфа-2"
-->
<!ELEMENT InventorName (#PCDATA)>
<ATTLIST InventorName
languageCode CDATA #REQUIRED
>
<!--InventorNameLatin
Якщо елемент "InventorName" наведено символами, не зазначеними у параграфі 40(b),
необхідно навести переклад або транслітерацію імені першого зазначеного винахідника
символами, зазначеними у параграфі 40(b).
-->
<!ELEMENT InventorNameLatin (#PCDATA)>
<!--InventionTitle
Назва винаходу вказана символами, як зазначено у параграфі 40(a) основної частини
стандарту BOIB ST.26, мовою оригіналу. Переклад назви винаходу додатковими мовами може
наводитися символами, як зазначено у параграфі 40(a) з використанням додаткових
елементів "InventionTitle". Бажано не більше 2-7 слів.
-->
<!--languageCode: Відповідний код мови за стандартом ISO 639-1 "Коди для представлення
назв мов. Частина 1: Код Альфа-2"
-->
<!ELEMENT InventionTitle (#PCDATA) >
<ATTLIST InventionTitle
languageCode CDATA #REQUIRED
>
<!--SequenceTotalQuantity
Вказує на загальну кількість послідовностей у документі.
Має бути швидкодоступним для автоматичної обробки.
-->
<!ELEMENT SequenceTotalQuantity (#PCDATA)>
<!--SequenceData
Дані стосовно певної послідовності.
Інформація стосовно навмисно пропущених послідовностей міститься в основній частині
стандарту BOIB ST.26.
-->
<!ELEMENT SequenceData (INSDSeq)>
<ATTLIST SequenceData
sequenceIDNumber CDATA #REQUIRED
>
<!--IPOfficeCode
Код за стандартом BOIB ST.3. Наприклад, якщо ідентифікаційними даними заявки є
"РСТ/ІВ2013/099999", то значенням "IPOfficeCode" є "МБ" для позначення Міжнародного бюро
ВОІВ.
-->
```

```

<!ELEMENT IOfficeCode (#PCDATA)>
<!--ApplicationNumberText
Ідентифікаційні дані заявки, надані відомством подання (наприклад, РСТ/ІВ2013/999999)
-->
<!ELEMENT ApplicationNumberText (#PCDATA)>

<!--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. Детальну інформацію
викладено у параграфах 7(a) і 11 стандарту ВОІВ ST.2.
-->
<!ELEMENT FilingDate (#PCDATA)>
<!--*****
* Частина INSD
*****

Метою частини INSD цього DTD є визначення специфічних DTD для переліків послідовностей
для підтримки роботи відомств інтелектуальної власності та спрощення обміну даними з
публічними депозитаріями.

Частина INSD є підмножиною INSD DTD v1.5 і може використовуватись лише для генерування
екземпляра XML, оскільки вона не підтримує всю структуру INSD.

Ця частина заснована на:

Проекті міжнародного співробітництва баз даних нуклеотидних послідовностей (INSDC).

"INSDSeq" містить елементи послідовності у форматах, як їх зазначено в GenBank/EMBL/DDBJ.
Тут використовуються не всі елементи.
-->
<!--INSDSeq
Дані щодо послідовності. Змінено елементи "INSD V1.5 DTD", "INSDSeq_division" та
"INSDSeq_sequence" з необов'язкових на обов'язкові для бізнес-потреб.
-->
<!ELEMENT INSDSeq (INSDSeq_length,INSDSeq_moltype,INSDSeq_division, INSDSeq_other-
seqids?,INSDSeq_feature-table?,INSDSeq_sequence)>
<!--INSDSeq_length
Довжина послідовності. "INSDSeq_length" дозволяються тільки цілі числа.
-->
<!ELEMENT INSDSeq_length (#PCDATA)>
<!--INSDSeq_moltype
Допустимі значення: DNA, RNA, AA
-->
<!ELEMENT INSDSeq_moltype (#PCDATA)>
<!--INSDSeq_division
Зазначення того, що послідовність відноситься до патентної заявки. Повинен бути
заповнений значенням "PAT".
-->
<!ELEMENT INSDSeq_division (#PCDATA)>
<!--INSDSeq_other-seqids
В рамках обміну даними з провайдерами баз даних, патентні відомства мають заповнювати
елемент "INSDSeq_other-seqids" для кожної послідовності одним елементом "INSDSeqid" з
посиланням на відповідний опублікований патент та ідентифікаційними даними послідовності.
-->
<!ELEMENT INSDSeq_other-seqids (INSDSeqid?)>
<!--INSDSeq_feature-table
Інформація про локалізацію та призначення різних ділянок у межах певної послідовності.
Елемент "INSDSeq_feature-table" повинен завжди містити принаймні одну функцію.
-->
<!ELEMENT INSDSeq_feature-table (INSDFeature+)>
<!--INSDSeq_sequence
Залишки послідовності. Послідовність не повинна містити цифри, розділові знаки або
пробіли.
-->
<!ELEMENT INSDSeq_sequence (#PCDATA)>
<!--INSDSeqid
Призначений для використання відомствами інтелектуальної власності лише під час обміну
даними.

```

## Формат:

pat|{код відомства}|{номер публікації}|{код виду документа}|{ідентифікаційний номер послідовності}

де кодом відомства є код відомства інтелектуальної власності, яке опублікувало патентний документ; номером публікації є номер опублікованої заявки або патенту; кодом виду документа є буквенний код для ідентифікації патентного документа відповідно до стандарту ВОІВ ST.16; а ідентифікаційним номером послідовності є номер послідовності в заявці або у патенті

## Приклад:

pat|WO|2013999999|A1|123456

Це 123456-а послідовність у патентній публікації WO2013999999 (A1)

```
-->
<!ELEMENT INSDSeqid (#PCDATA)>
<!--INSDFeature Опис
однієї функції.
-->
<!ELEMENT INSDFeature (INSDFeature_key,INSDFeature_location,INSDFeature_qual?)>
<!--INSDFeature_key
Слово чи аббревіатура, що ідентифікує функцію.
-->
<!ELEMENT INSDFeature_key (#PCDATA)>
<!--INSDFeature_location
Ділянка представленої послідовності, що відповідає функції.
-->
<!ELEMENT INSDFeature_location (#PCDATA)>
<!--INSDFeature_qual
Перелік кваліфікаторів з додатковою інформацією про функцію.
-->
<!ELEMENT INSDFeature_qual (INSDQualifier+)>
<!--INSDQualifier
Додаткова інформація про функцію.
Інформація про кодуєчі послідовності та варіанти міститься в основній частині стандарту
ВОІВ ST.26.
-->
<!--id
Унікальний ідентифікатор INSDQualifier для полегшення порівняння версій переліків
послідовностей, особливо тих, що містять залежні від мови значення кваліфікаторів різними
мовами.
-->
<!ELEMENT INSDQualifier (INSDQualifier name, INSDQualifier value?,
NonEnglishQualifier value?)>
<!ATTLIST INSDQualifier
id ID #IMPLIED
>
<!--INSDQualifier name
Ім'я кваліфікатора.
-->
<!ELEMENT INSDQualifier_name (#PCDATA)>
<!--INSDQualifier value
Значення кваліфікатора. Якщо кваліфікатор залежить від мови, то його значення повинно
бути англійською мовою та наводитися символами, як зазначено у параграфі 40 (b).
-->
<!ELEMENT INSDQualifier_value (#PCDATA)>
<!--NonEnglishQualifier value
Значення залежного від мови кваліфікатора мовою, яка не є англійською, наводиться
символами, як зазначено у параграфі 40 (a). Мова позначається атрибутом
nonEnglishFreeTextLanguageCode.
-->
<!ELEMENT NonEnglishQualifier_value (#PCDATA)>
```

[Додаток III наведено далі]

### **ДОДАТОК III**

#### **ЗРАЗОК ПЕРЕЛІКУ ПОСЛІДОВНОСТЕЙ (файл XML)**

*Версія 1.4*

*Редакція, схвалена Комітетом зі стандартів ВОІВ (КСВ)  
на його одинадцятій сесії 8 грудня 2023 року*

Додаток III доступний за адресою:  
[https://www.wipo.int/standards/en/xml\\_material/st26/st26-specimen.xml](https://www.wipo.int/standards/en/xml_material/st26/st26-specimen.xml) .  
[annex-iii-sequence-listing-](#)

[Додаток IV наведено далі]

## ДОДАТОК IV

## НАБІР СИМВОЛІВ З ТАБЛИЦІ КОДІВ ОСНОВНИХ ЛАТИНСЬКИХ СИМВОЛІВ UNICODE ДЛЯ ВИКОРИСТАННЯ В ЕКЗЕМПЛЯРІ XML ПЕРЕЛІКУ ПОСЛІДОВНОСТІ

Версія 1.3

*Редакція, схвалена Комітетом зі стандартів VOIB (КСВ)  
на його одинадцятій сесії 8 грудня 2023 року*

Амперсанд (0026) допускається лише як частина попередньо визначеної сутності. Лапки (0022), апостроф (0027), знак "менше" (003C) і знак "більше" (003E) повинні представлятися за допомогою їх попередньо визначених сутностей. Крім того, символ амперсанд (0026) повинен бути представлений попередньо визначеною сутністю, якщо він використовується як амперсанд в значенні атрибута або змісті елемента.

| Код<br>Unicode | Символ | Назва              |
|----------------|--------|--------------------|
| 0020           |        | ПРОБІЛ             |
| 0021           | !      | ЗНАК ОКЛИКУ        |
| 0022           | "      | ЛАПКИ              |
| 0023           | #      | ОКТОТОРП           |
| 0024           | \$     | ЗНАК ДОЛАРА        |
| 0025           | %      | ЗНАК ВІДСОТКА      |
| 0026           | &      | АМПЕРСАНД          |
| 0027           | '      | АПОСТРОФ           |
| 0028           | (      | ЛІВА КРУГЛА ДУЖКА  |
| 0029           | )      | ПРАВА КРУГЛА ДУЖКА |
| 002A           | *      | ЗІРОЧКА            |
| 002B           | +      | ЗНАК ПЛЮС          |
| 002C           | ,      | КОМА               |
| 002D           | -      | ДЕФІС / ЗНАК МІНУС |
| 002E           | .      | КРАПКА             |
| 002F           | /      | СКІСНА РИСКА       |
| 0030           | 0      | ЦИФРА НУЛЬ         |
| 0031           | 1      | ЦИФРА ОДИН         |
| 0032           | 2      | ЦИФРА ДВА          |
| 0033           | 3      | ЦИФРА ТРИ          |

| <b>Код<br/>Unicode</b> | <b>Символ</b> | <b>Назва</b>              |
|------------------------|---------------|---------------------------|
| 0034                   | 4             | ЦИФРА ЧОТИРИ              |
| 0035                   | 5             | ЦИФРА П'ЯТЬ               |
| 0036                   | 6             | ЦИФРА ШІСТЬ               |
| 0037                   | 7             | ЦИФРА СІМ                 |
| 0038                   | 8             | ЦИФРА ВІСІМ               |
| 0039                   | 9             | ЦИФРА ДЕВ'ЯТЬ             |
| 003A                   | :             | ДВОКРАПКА                 |
| 003B                   | ;             | КРАПКА З КОМОЮ            |
| 003C                   | <             | ЗНАК "МЕНШЕ"              |
| 003D                   | =             | ЗНАК РІВНОСТІ             |
| 003E                   | >             | ЗНАК "БІЛЬШЕ"             |
| 003F                   | ?             | ЗНАК ПИТАННЯ              |
| 0040                   | @             | КОМЕРЦІЙНЕ ЕТ             |
| 0041                   | A             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА A |
| 0042                   | B             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА B |
| 0043                   | C             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА C |
| 0044                   | D             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА D |
| 0045                   | E             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА E |
| 0046                   | F             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА F |
| 0047                   | G             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА G |
| 0048                   | H             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА H |
| 0049                   | I             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА I |
| 004A                   | J             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА J |
| 004B                   | K             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА K |
| 004C                   | L             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА L |
| 004D                   | M             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА M |
| 004E                   | N             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА N |
| 004F                   | O             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА O |
| 0050                   | P             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА P |
| 0051                   | Q             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА Q |
| 0052                   | R             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА R |
| 0053                   | S             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА S |
| 0054                   | T             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА T |
| 0055                   | U             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА U |
| 0056                   | V             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА V |



| <b>Код<br/>Unicode</b> | <b>Символ</b> | <b>Назва</b>                |
|------------------------|---------------|-----------------------------|
| 0057                   | W             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА W   |
| 0058                   | X             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА X   |
| 0059                   | Y             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА Y   |
| 005A                   | Z             | ВЕЛИКА ЛАТИНСЬКА ЛІТЕРА Z   |
| 005B                   | [             | ЛІВА КВАДРАТНА ДУЖКА        |
| 005C                   | \             | ЗВОРОТНА СКІСНА РИСКА       |
| 005D                   | ]             | ПРАВА КВАДРАТНА ДУЖКА       |
| 005E                   | ^             | ЦИРКУМФЛЕКС                 |
| 005F                   | _             | ЗНАК ПІДКРЕСЛЮВАННЯ         |
| 0060                   | `             | ГРАВІС                      |
| 0061                   | a             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА A |
| 0062                   | b             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА B |
| 0063                   | c             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА C |
| 0064                   | d             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА D |
| 0065                   | e             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА E |
| 0066                   | f             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА F |
| 0067                   | g             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА G |
| 0068                   | h             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА H |
| 0069                   | i             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА I |
| 006A                   | j             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА J |
| 006B                   | k             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА K |
| 006C                   | l             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА L |
| 006D                   | m             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА M |
| 006E                   | n             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА N |
| 006F                   | o             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА O |
| 0070                   | p             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА P |
| 0071                   | q             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА Q |
| 0072                   | r             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА R |
| 0073                   | s             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА S |
| 0074                   | t             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА T |
| 0075                   | u             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА U |
| 0076                   | v             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА V |
| 0077                   | w             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА W |
| 0078                   | x             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА X |
| 0079                   | y             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА Y |

| <b>Код<br/>Unicode</b> | <b>Символ</b> | <b>Назва</b>                |
|------------------------|---------------|-----------------------------|
| 007A                   | z             | МАЛЕНЬКА ЛАТИНСЬКА ЛІТЕРА Z |
| 007B                   | {             | ЛІВА ФІГУРНА ДУЖКА          |
| 007C                   |               | ВЕРТИКАЛЬНА РИСКА           |
| 007D                   | }             | ПРАВА ФІГУРНА ДУЖКА         |
| 007E                   | ~             | ТИЛЬДА                      |

[Додаток V наведено далі]

---

**ДОДАТОК V****ДОДАТКОВІ ВИМОГИ ДО ОБМІНУ ДАНИМИ (ЛИШЕ ДЛЯ ПАТЕНТНИХ ВІДОМСТВ)***Версія 1.5*

*Редакція, схвалена Комітетом зі стандартів BOIB (КСВ)  
на його одинадцятій сесії 8 грудня 2023 року*

В рамках обміну даними з провайдерами баз даних (членами INSD), ВІВ мають заповнювати елемент `INSDSeq_other-seqids` для кожної послідовності одним елементом `INSDSeqid` з посиланням на відповідний опублікований патент та ідентифікаційним номером послідовності у такому форматі:

pat|{код відомства}|{номер публікації}|{код виду документа}|{ідентифікаційний номер  
послідовності}

де кодом відомства є код відомства інтелектуальної власності, яке опублікувало патентний документ, відповідно до стандарту BOIB ST.3; кодом виду документа є код для ідентифікації різних видів патентних документів відповідно до стандарту BOIB ST.16; номером публікації є номер опублікованої заявки або патенту; а ідентифікаційним номером послідовності є номер послідовності в заявці або у патенті.

Приклад:

pat|WO|2013999999|A1|123456

Відповідний зразок за схемою XML:

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

де "123456" є 123456-ю послідовністю в публікації WO2013999999 (A1).

[Додаток VI наведено далі]

**ДОДАТОК VI****КЕРІВНИЙ ДОКУМЕНТ З ІЛЮСТРОВАНИМИ ПРИКЛАДАМИ***Версія 1.7*

*Редакція, схвалена Комітетом зі стандартів ВОІВ (КСВ)  
на його одинадцятій сесії 8 грудня 2023 року*

**ЗМІСТ**

|                       |     |
|-----------------------|-----|
| ВСТУП.....            | 108 |
| ІНДЕКС ПРИКЛАДІВ..... | 116 |
| ПРИКЛАДИ.....         | 119 |
| ДОПОВНЕННЯ.....       | 202 |

**ВСТУП**

У цьому Стандарті одним із завдань визначено: "надати заявникам можливість скласти єдиний перелік послідовностей у патентній заявці, прийнятний для цілей як міжнародних, так і національних або регіональних процедур". Мета цього Керівного документа - забезпечення можливості взаєморозуміння та домовленості для всіх заявників та Відомств інтелектуальної власності (ВІВ) стосовно вимог щодо включення та представлення опису послідовностей, таким чином, щоб досягти мети Стандарту.

Цей керівний документ складається зі вступу, індексу прикладів, прикладів опису послідовності, а також доповнення, що містить перелік послідовностей у форматі XML та послідовності з прикладів. Цей вступ пояснює певні поняття та термінологію, використані в решті частини цього документа. Приклади ілюструють вимоги конкретних параграфів стандарту і кожен приклад позначено найбільш відповідним номером абзацу. Деякі приклади детально ілюструють інші параграфи, а відповідні перехресні посилання вказуються в кінці кожного прикладу. В індексі вказано номери сторінок для прикладів та будь-яких зазначених перехресних посилань. Кожній послідовності в прикладі, яка повинна або може бути включена до переліку послідовностей, присвоєно ідентифікаційний номер послідовності (SEQ ID NO); вона відображається у форматі XML у [Доповненні](#) до цього документа.

Для кожного прикладу будь-яка пояснювальна інформація, представлена з послідовністю, повинна розглядатися як повний опис інформації щодо цієї послідовності. Надані відповіді стосуються лише інформації, безпосередньо представленої в прикладі.

Рекомендації, наведені в цьому документі, спрямовані на підготовку переліку послідовностей для надання на дату подання патентної заявки. Підготовка переліку послідовностей для надання після дати подання патентної заявки повинна враховувати, чи може надана інформація бути розглянута ВІВ, для додання цього об'єкту до первинно поданого опису. Тому цілком можливо, що рекомендації, наведені в цьому документі, можуть не застосовуватися до переліку послідовностей, наданих після дати подання

патентної заявки.

#### *Підготовка переліку послідовностей*

Підготовка переліку послідовностей до патентної заявки вимагає розгляду наступних питань:

1. Чи вимагає параграф 7 стандарту VOIB ST.26 включення певної розкритої послідовності?
2. Чи дозволено включення послідовності стандартом VOIB ST.26, якщо включення певної розкритої послідовності не вимагається?
3. Яким чином ця послідовність буде представлена в переліку послідовностей, якщо включення певної розкритої послідовності вимагається або дозволяється стандартом VOIB ST.26?

Стосовно першого запитання, параграф 7 стандарту VOIB ST.26 (з певними обмеженнями) вимагає включення послідовності, описаної в патентній заявці, шляхом перерахування її залишків, де послідовність містить десять або більше специфічно визначених нуклеотидів або чотири чи більше специфічно визначених амінокислоти.

Стосовно другого запитання, параграф 8 стандарту VOIB ST.26 забороняє включення будь-яких послідовностей, що містять менше десяти специфічно визначених нуклеотидів або чотирьох специфічно визначених амінокислот.

Для відповіді на ці два запитання необхідне чітке розуміння понять "представлення залишків" і "специфічно визначених".

Стосовно третього запитання, цей документ містить опис послідовностей, які демонструють різноманітні сценарії разом з повним обговоренням переважних засобів представлення кожної послідовності або, якщо послідовність містить декілька варіантів - "найбільш охоплюючої послідовності" відповідно до цього стандарту. Оскільки неможливо відреагувати на всі ймовірні сценарії послідовності, цей керівний документ намагається викласти пояснення підходу до кожного прикладу та способу, до якого застосовуються положення стандарту VOIB ST.26, таким чином, щоб однакові пояснення могли застосовуватися до інших сценаріїв послідовностей, які не наведені в прикладі.

#### Перерахування залишків

У параграфі 3(с) стандарту VOIB ST.26 "перелік залишків" визначено як опис послідовності в патентній заявці шляхом переліку по порядку кожного залишку послідовності, де (i) залишок представлений назвою, скороченням, символом або структурою; або (ii) кілька залишків представлені скороченою формулою. Послідовність повинна бути розкрита в патентній заявці шляхом "переліку залишків" за допомогою стандартних символів, які є символами нуклеотидів, викладеними в Таблиці 1 Розділу 1 Додатку I стандарту VOIB ST.26 (наприклад, символи нижнього регістру або їх еквіваленти верхнього регістру<sup>2</sup>) та символи амінокислот, викладені в Таблиці 3 Розділу 3 Додатку I стандарту VOIB ST.26 (тобто символи верхнього регістру або їх еквіваленти нижнього регістру<sup>1</sup>). Символи, крім тих, що

<sup>2</sup> ПРИМІТКА. Хоча опис заявки може представляти нуклеотиди або амінокислоти з символами нижнього регістру або верхнього регістру, для послідовності, включеної до переліку послідовностей, повинні використовуватись лише символи нижнього регістру для представлення нуклеотидної послідовності (див. параграф 13 стандарту VOIB ST.26), а для представлення амінокислотної послідовності повинні використовуватись лише символи верхнього регістру (див. параграф 26 стандарту VOIB ST.26)

викладені в цих таблицях, є "нестандартними".

Якщо представлення залишку розкривається як еквівалент стандартного символу або скорочення (наприклад, "Z<sub>1</sub>" означає "A"), або певної послідовності стандартних символів (наприклад, "Z<sub>1</sub>" означає "agga"), то послідовність тлумачиться так, ніби вона була розкрита за допомогою еквівалентного стандартного (-их) символу(-ів) або скорочення (-ь), щоб визначити, чи параграф 7 стандарту BOIB ST.26 вимагає включення до переліку послідовностей, чи параграф 8 забороняє включення. Якщо нестандартний символ нуклеотиду використовується як багатозначний символ (наприклад, X1 = інозин або псевдоуридин), але не є еквівалентом одного зі стандартних багатозначних символів в Розділі 1, Таблиці 1 (тобто "m", "r", "w", "s", "y", "k", "v", "h", "d", "b", або "n"), то залишок інтерпретується як залишок "n", щоб визначити, чи вимагає параграф 7 стандарту BOIB ST.26 включення послідовності до переліку послідовностей, чи параграф 8 стандарту ST.26 забороняє включення. Аналогічно, якщо нестандартний символ амінокислоти використовується як багатозначний символ (наприклад, ". "A", "G", "S" або "T"), але не є еквівалентом одного зі стандартних багатозначних символів в Розділі 3, Таблиці 3 (тобто, B, Z, J, або X), тоді залишок інтерпретується як залишок "X" для визначення, чи вимагає параграф 7 стандарту BOIB ST.26 включення послідовності до переліку послідовностей, чи параграф 8 стандарту BOIB ST.26 забороняє таке включення.

Варто звернути увагу на опис послідовностей з використанням стандартних символів; проте, якщо послідовності описуються інакше, може знадобитися ознайомлення з описом для визначення значення нестандартного представлення.

Якщо використовується стандартний символ, повинно бути прийнято до уваги пояснення послідовності в розкритті для підтвердження, що символ використовується стандартним способом. Якщо символ використовується нестандартним способом, необхідно визначити роз'яснення, чи вимагає параграф 7 стандарту BOIB ST.26 включення до переліку послідовностей, чи параграф 8 забороняє таке включення.

#### Специфічно визначені

Параграф 3(k) стандарту BOIB ST.26 визначає: "специфічно визначений" - будь-який нуклеотид, відмінний від тих, що представлені символом "n" та будь-яка амінокислота, відмінна від тих, що представлені символом "X", перелічені у Додатку I, де "n" та "X" використовуються стандартним способом, як описано в Розділі 1, Таблиці 1 (тобто "a або c або g або t/u; "невідомий" або "інший") та в Розділі 3, Таблиці 3 (тобто A або R або N або D або C або Q або E або G або H або I або L або K або M або F або P або O або S або U або T або W або Y або V "невідомий" або "інший"), відповідно. Зазначення вище стосовно стандартних символів або нестандартних символів або скорочення та їх застосування стандартним або нестандартним способом буде враховуватися для визначення того, чи є нуклеотид або амінокислота "специфічно визначеним (-ою)".

#### Найбільш охоплююча послідовність

Якщо послідовність, яка відповідає вимогам параграфа 7, розкривається шляхом перерахування її залишків в заявці лише один раз, але по-різному описується в декількох варіантах здійснення, наприклад, в одному варіанті "X" в одній або декількох локалізаціях може бути будь-якою амінокислотою, але в інших варіантах здійснення "X" може бути лише обмеженою кількістю амінокислот, стандарт BOIB ST.26 вимагає включення до переліку послідовностей лише однієї послідовності, яка була перерахована її залишками. Відповідно до параграфів 15 та 27, якщо така послідовність містить декілька багатозначних символів "n" або "X", то "n" або "X" тлумачиться як будь-який нуклеотид або

амінокислота, відповідно, за відсутності додаткового пояснення. Отже, єдина послідовність, яку потрібно включити, - це описана більш охоплююча послідовність. Найбільш охоплююча послідовність - це одинична послідовність, з варіантами залишків, представленими більш обмежувальними багатозначними символами, які містять найбільш описані варіанти здійснення. Подібним чином, якщо послідовність описана шляхом перерахування її залишків лише один раз, але довжина послідовності може змінюватися через варіацію кількості копій, найдовший варіант здійснення послідовності вважається найбільш охоплюючою послідовністю. Наприклад, послідовність, що містить повторювану область, яка може варіюватися від 2 до 5 копій, як перераховано. Варіант здійснення з 5 копіями повтору є найбільш охоплюючою послідовністю і має бути включений до переліку послідовностей. Проте включення додаткових специфічних послідовностей настійно рекомендується, де це доцільно, наприклад, таких, які представляють додаткові варіанти здійснення, що є ключовою частиною винаходу. Включення додаткових послідовностей дозволяє здійснити більш ретельний пошук та публічно повідомляти про об'єкт, щодо якого витребується патент.

### *Використання багатозначного символу*

#### Правильне використання багатозначного символу "n" в переліку послідовностей

##### Символ "n"

- a. не повинен використовуватися для позначення нічого, крім одного нуклеотиду;
- b. має тлумачитись як будь-який із "a", "c", "g" або "t/u", за винятком випадків, коли він використовується з подальшим описом;
- c. слід використовувати для представлення будь-якого з наступних нуклеотидів разом з додатковим описом:
  - i. модифікований нуклеотид, наприклад, природний, синтетичний або неприродний, який не може бути представлений будь-яким іншим символом у Додатку I (див. Розділ 1, Таблиця 1)
  - ii. "невідомий" нуклеотид, тобто невизначений, нерозкритий або не встановлений;
  - iii. апуриновий сайт; або
- d. може бути використаний для представлення варіанту послідовності, тобто альтернативи, видалення, вставки або заміни, де "n" є найбільш обмежувальним багатозначним символом.

#### Правильне використання багатозначного символу "X" у переліку послідовностей

##### Символ "X"

- a. не повинен використовуватись для позначення нічого іншого, крім однієї амінокислоти;
- b. має тлумачитись як будь-який з "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" або "V", за винятком випадків, коли він використовується з детальним описом;
- c. слід використовувати для представлення будь-якої з наступних амінокислот разом

з наступним описом:

- i. модифікована амінокислота, наприклад, природна, синтетична або штучна, що не може бути інакше представлена будь-яким іншим символом у Додатку I (див. Розділ 3, Таблиця 3);
  - ii. "невідома" амінокислота, тобто невизначена, нерозкрита або невстановлена; або
- d. може використовуватися для представлення варіанту послідовності, тобто альтернатив, видалень, вставок або заміщень, де "X" є найбільш обмежувальним багатозначним символом.

Тлумачення модифікованих залишків

Цей стандарт вимагає, тлумачення "модифікованих" залишків згідно з параграфом 17 для нуклеотидів та згідно з параграфом 30 для амінокислот.

У параграфі 3(e) Стандарту VOIB ST.26 "модифікована амінокислота" визначена як будь-яка амінокислота, описана в параграфі 3(a), крім 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-піролізину, L-серину, L-селеноцистеїну, L-треоніну, L-триптофану, L-тирозину або L-валіну. Подібним чином у стандарті визначено "модифікований нуклеотид" як будь-який нуклеотид, як описано в параграфі 3(g), крім дезоксиаденозин 5'-монофосфату, дезоксигуанозин 3'-монофосфату, дезоксцитидин 5'-монофосфату, дезокситимідин 5'-монофосфату, аденозин 5'-монофосфат, гуанозин 5'-монофосфату, цитидин 5'-монофосфату або уридин 5'-монофосфату (Стандарт VOIB ST.26, параграф 3(f)).

Виходячи з наведених вище визначень, модифікації нуклеїнових основ або каркасу цукрофосфату нуклеїнової кислоти та модифікації груп R амінокислот або пептидного каркасу пептиду призводять до одного або більше "модифікованих нуклеотидів" або "модифікованих амінокислот" відповідно. Тому такі нуклеотиди та амінокислоти вимагають тлумачення. Приклади модифікацій каркасу включають аналоги нуклеотидів, такі як пептидні нуклеїнові кислоти (ПНК) і гліколеві нуклеїнові кислоти (GNA) і D-амінокислоти.

Варто зазначити, що модифікація кінцевої амінокислоти пептиду або кінцевого нуклеотиду нуклеїнової кислоти не обов'язково призводить до "модифікованої амінокислоти" або "модифікованого нуклеотиду". Необхідно переглянути кінцеву модифікацію та визначити, чи модифікація змінює хімічну структуру залишку таким чином, що залишок виходить за межі винятків, наведених у параграфах 3(e) та 3(f). Наприклад, пептид, у якому С-кінцевий залишок зв'язаний зі структурою (наприклад, частиною розгалуженої послідовності – див. пептид №2 у прикладі 7(b)-3) через стандартний амідний зв'язок, не вважається "модифікованим залишком" і тому не вимагає тлумачення. Аналогічно, пептид, в якому N-кінцевий залишок є амідним зв'язком з біотином, не вважається "модифікованим залишком" і тому не вимагає тлумачення. В обох сценаріях структура залишку, залученого до С-кінцевого або N-кінцевого зв'язку, не змінюється порівняно зі стандартними амінокислотами, зазначеними в параграфі 3(e) Стандарту.

Навпаки, кінцеві модифікації, які змінюють хімічну структуру залишку, вважаються "модифікованими залишками" і повинні бути анотовані. Наприклад, метилювання С-кінця в прикладі 3(c)-1 справді змінює хімічну структуру кінцевого залишку, оскільки метильна група замінює гідроксил, який зазвичай міститься в альфа-карбоксылній групі.



Таким чином, цей метильований лізін повинен бути позначений як "модифікований залишок".

Варто зауважити, що заявник повинен оцінювати кожну модифікацію кінцевого залишку в межах переліченої послідовності та приймати рішення щодо того, чи змінюється структура кінцевого залишку. Якщо модифікована структура залишків відрізняється від звичайних амінокислот або нуклеотидів, зазначених у параграфах 3(e) та 3(f) Стандарту, тоді модифікація повинна бути анотована.

Нарешті, завжди рекомендується, щоб заявники включали якомога більше інформації у свої переліки послідовностей, щоб представити свої описи якомога точніше. Таким чином, навіть якщо модифікація не вимагає анотації, бажано її включити.

Втім, варто зауважити, що тлумачення варіантів перерахованої первинної послідовності має відповідати вимогам параграфів 93-100 Стандарту VOIB ST.26. Модифікації, які описуються як варіанти пронумерованої послідовності, можуть не вимагатися для включення до переліку послідовностей. Для визначення тлумачення варіантів див. Стандарт VOIB ST.26, параграфи 93-95.

**Представлення модифікованих залишків**

У Стандарті VOIB ST.26 вказано, що модифіковані нуклеотиди та амінокислоти повинні бути представлені в переліку послідовностей як відповідний немодифікований залишок, коли це можливо (див. параграфи 16 і 29). Варто зауважити, що ця рекомендація - "потрібно" – "настійно рекомендований підхід, але не вимога" (див. пункт 4(d)). Заявник вирішує, чи буде модифікований залишок представлений відповідним немодифікованим залишком або змінними "n" або "X".

Як правило, якщо залишок модифіковано додаванням компонента, наприклад метилуванням або ацетилюванням, і структура немодифікованого залишку загалом не змінюється, тоді рекомендовано представлення за допомогою немодифікованого залишку. Наприклад, метильований аденозин бажано позначати "a" в переліку послідовностей. Однак, якщо модифікований залишок структурно відрізняється від будь-якого немодифікованого залишку, тоді рекомендується "n" або "X". Наприклад, норлейцин є ізомером лейцину, а його бічний ланцюг є лінійною структурою з 4 атомів вуглецю. Лейцин також має бічний ланцюг з 4 вуглеців, але він розгалужений на другому вуглеці. Тому норлейцин є не просто результатом модифікації, доданої до лейцину, а зовсім іншою (хоча й спорідненою) структурою. Тому рекомендується позначати норлейцин у переліку послідовностей символом "X".

Нуклеотид є "специфічно визначеним", коли він представлений будь-яким символом, крім "n", а амінокислота є "специфічно визначеною", коли представлена будь-яким символом, крім "X" (див. Стандарт VOIB ST.26, параграф 3(k)). Таким чином, 2'-О-метиладенозин, представлений символом "a" у послідовності, є специфічно визначеним, тоді як норлейцин, представлений символом "X" у послідовності, не є специфічно визначеним.

*Таблиця А - Стандартні символи нуклеотидів, скорочення та назви*

| Символ | Скорочення | Назва нуклеотиду |
|--------|------------|------------------|
| a      |            | аденін           |
| c      |            | цитозин          |

| Символ | Скорочення   | Назва нуклеотиду                  |
|--------|--|-----------------------------------|
| g      |  | гуанін                            |
| t      |  | тимін у ДНК<br>урацил у РНК (t/u) |
| m      | а або с  |                                   |
| r      | а або g  |                                   |
| w      | а або t/u  |                                   |
| s      | с або g  |                                   |
| y      | с або t/u  |                                   |
| k      | g або t/u  |                                   |
| v      | а або с або g; не t/u                                |                                   |
| h      | а або с або t/u; не g                                |                                   |
| d      | а або g або t/u; не с                                |                                   |
| b      | с або g або t/u; не а                                |                                   |
| n      | а або с або g або t/u;<br>"невідомий"<br>або "інший" |                                   |

Таблиця В – Стандартні символи амінокислот, скорочення та назви

| Символ | Трибуквене скорочення | Назва амінокислоти                 |
|--------|-----------------------|------------------------------------|
| A      | Ala                   | аланін                             |
| R      | Arg                   | аргінін                            |
| N      | Asn                   | аспарагін                          |
| D      | Asp                   | аспарагінова кислота<br>(аспартат) |
| C      | Cys                   | цистеїн                            |
| E      | Glu                   | глутамінова кислота<br>(глутамат)  |
| Q      | Gln                   | глутамін                           |
| G      | Gly                   | гліцин                             |
| H      | His                   | гістидин                           |
| I      | Ile                   | ізолейцин                          |
| L      | Leu                   | лейцин                             |
| K      | Lys                   | лізин                              |
| M      | Met                   | метіонін                           |
| F      | Phe                   | фенілаланін                        |
| P      | Pro                   | пролін                             |
| O      | Pyl                   | пірролізин                         |
| S      | Ser                   | серин                              |
| U      | Sec                   | селеноцистеїн                      |
| T      | Thr                   | треонін                            |
| W      | Trp                   | триптофан                          |
| Y      | Tyr                   | тирозин                            |
| V      | Val                   | валін                              |
| B      | Asx                   | аспарагінова кислота або           |

|   |     |   |
|---|-----|---|
|   |     | аспарагін   |
| Z | Glx | глутамін або глутамінова кислота  |
| J | Xle | лейцин або ізолейцин  |
| X | Xaa | A або R або N або D<br>або C або Q або E або<br>G або H або I або L<br>або K або M або F або<br>P або O або S або U<br>або T або W або Y або<br>V, "невідома"<br>або "інша" |

## ІНДЕКС ПРИКЛАДІВ

|   |     |
|---|-----|
| <i>Параграф 3(а) Визначення "амінокислота"</i> .....  | 119 |
| <b>Приклад 3(а)-1: D-амінокислоти</b> .....   | 119 |
| <i>Параграф 3(с) – Визначення "перелік залишків"</i> .....  | 120 |
| <b>Приклад 3(с)-1: Перелік амінокислот за хімічною структурою</b> .....   | 120 |
| <b>Приклад 3(с)-2: Скорочена формула для амінокислотної послідовності</b> .....   | 122 |
| <i>Параграф 3(г) - Визначення "нуклеотид"</i> .....   | 124 |
| <b>Приклад 3(г)-1: Нуклеотидна послідовність, перервана спейсером С3</b> .....  | 124 |
| <b>Приклад 3(г)-2: Нуклеотидна послідовність з альтернативними залишками, включаючи спейсер С3</b> .....                  | 125 |
| <b>Приклад 3(г)-3: Апуриновий сайт</b> .....  | 127 |
| <b>Приклад 3(г)-4: Аналоги нуклеїнових кислот</b> .....   | 128 |
| <i>Параграф 3 (к) Визначення "специфічно визначений"</i> .....  | 129 |
| <b>Приклад 3(к)-1: Багатозначні символи нуклеотидів</b> .....   | 129 |
| <b>Приклад 3(к)-2: Багатозначний символ "n", який використовується як стандартним, так і нестандартним способом</b> ..... | 130 |
| <b>Приклад 3(к)-3: Багатозначний символ "n", який використовується нестандартним способом</b> .....                       | 132 |
| <b>Приклад 3(к)-4: Багатозначні символи, окрім "n", є "специфічно визначеними"</b> .....                                  | 133 |
| <b>Приклад 3(к)-5: Багатозначне скорочення "Хаа" використовується нестандартним способом</b> .....                        | 134 |
| <i>Параграф 7(а) - Нуклеотидні послідовності, які необхідно зазначати в переліку послідовностей ...</i>                   | 136 |
| <b>Приклад 7(а)-1: Розгалужена нуклеотидна послідовність</b> .....  | 136 |
| <b>Приклад 7(а)-2: Лінійна нуклеотидна послідовність, що має вторинну структуру</b> .....                                 | 138 |
| <b>Приклад 7(а)-3: Багатозначні символи нуклеотидів, що використовуються нестандартним способом</b> .....                 | 140 |
| <b>Приклад 7(а)-4: Багатозначні символи нуклеотидів, що використовуються нестандартним способом</b> .....                 | 141 |
| <b>Приклад 7(а)-5: Нестандартні символи нуклеотидів</b> .....   | 142 |
| <b>Приклад 7(а)-6: Нестандартні символи нуклеотидів</b> .....   | 143 |
| <b>Приклад 7(а)-7: Інвертовані нуклеотиди I</b> .....   | 145 |
| <b>Приклад 7(а)-8: Інвертовані нуклеотиди II</b> .....  | 147 |
| <i>Параграф 7(б) - Амінокислотні послідовності, які необхідно зазначати в переліку послідовностей</i>                     | 149 |
| <b>Приклад 7(б)-1: Чотири або більше специфічно визначені амінокислоти</b> .....  | 149 |
| <b>Приклад 7(б)-2: Розгалужена амінокислотна послідовність</b> .....  | 150 |
| <b>Приклад 7(б)-3: Розгалужена амінокислотна послідовність</b> .....  | 153 |
| <b>Приклад 7(б)-4: Циклічний пептид, який містить розгалужену амінокислотну послідовність</b> .                           | 154 |
| <b>Приклад 7(б)-5: Циклічний пептид, який містить розгалужену амінокислотну послідовність</b> .                           | 157 |
| <i>Параграф 11(а) - Дволанцюгова нуклеотидна послідовність - повністю комплементарна</i> .....                            | 159 |
| <b>Приклад 11(а)-1: Дволанцюгова нуклеотидна послідовність - однакові довжини</b> .....                                   | 159 |
| <i>Параграф 11(б) - Дволанцюгова нуклеотидна послідовність - не повністю комплементарна</i> .....                         | 160 |
| <b>Приклад 11(б)-1: Дволанцюгова нуклеотидна послідовність - різні довжини</b> .....                                      | 160 |
| <b>Приклад 11(б)-2: Дволанцюгова нуклеотидна послідовність – відсутній сегмент спарювання основ</b> .....                 | 161 |
| <i>Параграф 12 – Кільцева послідовність нуклеотидів</i> .....   | 162 |
| <b>Приклад 12-1: Кільцева послідовність нуклеотидів</b> .....   | 162 |
| <i>Параграф 14 – Символ "t" позначає урацил в РНК</i> .....   | 163 |
| <b>Приклад 14-1: Символ "t" представляє урацил у РНК</b> .....  | 163 |

|   |     |
|---|-----|
| <i>Параграф 27 – Має використовуватись найбільш обмежувальний багатозначний символ амінокислоти</i> .....   | 165 |
| <b>Приклад 27-1: Скорочена формула для нуклеотидної послідовності</b> .....   | 165 |
| <b>Приклад 27-2: Скорочена формула - менше чотирьох специфічно визначених амінокислот</b> ..  | 167 |
| <b>Приклад 27-3: Скорочена формула - чотири або більше специфічно визначені амінокислоти</b>  | 169 |
| <i>Параграф 28 - Амінокислотні послідовності, розділені внутрішніми термінаторами</i> .....   | 171 |
| <b>Приклад 28-1: Кодуюча нуклеотидна послідовність і кодована амінокислотна послідовність</b>   | 171 |
| <i>Параграф 29 - Представлення "іншої" амінокислоти</i> .....   | 173 |
| <b>Приклад 29-1: Найбільш обмежувальний багатозначний символ для "іншої" амінокислоти</b> ..  | 173 |
| <b>Приклад 29-2: Використання відповідної немодифікованої амінокислоти</b> .....  | 174 |
| <i>Параграф 30 - Тлумачення модифікованих амінокислот</i> .....   | 176 |
| <b>Приклад 30-1: Ключ функції "CARBODHYD"</b> .....   | 176 |
| <b>Приклад 30-2 – Посттрансляційно модифіковані амінокислоти</b> .....  | 177 |
| <i>Параграф 36 - Послідовності, що містять області точної кількості суміжних залишків "n" або "X"</i><br>.....  | 178 |
| <b>Приклад 36-1: Послідовність з областю відомої кількості залишків "X", представлена у вигляді однієї послідовності</b> .....                                | 178 |
| <b>Приклад 36-2: Послідовність з кількома областями відомої кількості або діапазоном залишків "X", представлена як одна послідовність</b> .....               | 179 |
| <b>Приклад 36-3: Послідовність з кількома областями відомої кількості або діапазоном залишків "X", представлена як одна послідовність</b> .....               | 181 |
| <i>Параграф 37 - Послідовності, що містять області невідомої кількості суміжних залишків "n" або "X"</i><br>.....   | 183 |
| <b>Приклад 37-1: Послідовність з областями невідомої кількості залишків "X" не повинна бути представлена як одна послідовність</b> .....                      | 183 |
| <b>Приклад 37-2: Послідовність з областями невідомої кількості залишків "X" не повинна бути представлена як одна послідовність</b> .....                      | 184 |
| <i>Параграф 55 – Нуклеотидна послідовність, яка містить сегменти ДНК та РНК</i> .....   | 186 |
| <b>Приклад 55-1: Комбінована молекула ДНК/РНК</b> .....   | 186 |
| <i>Параграф 89 - Ключ функції "CDS"</i> .....   | 187 |
| <b>Приклад 89-1: Кодуюча нуклеотидна послідовність і кодована амінокислотна послідовність</b>   | 187 |
| <b>Приклад 89-2: Локалізація функції виходить за межі розкритої послідовності</b> .....   | 188 |
| <i>Параграф 92 - Амінокислотна послідовність, кодована кодуючою послідовністю</i> .....   | 190 |
| <b>Приклад 92-1: Амінокислотна послідовність, кодована кодуючою послідовністю з інтронами</b><br>.....  | 190 |
| <i>Параграф 93 - Первинна послідовність і її варіанти, кожний з яких перерахований за своїми залишками</i> .....  | 192 |
| <b>Приклад 93-1: Представлення перерахованих варіантів</b> .....  | 192 |
| <b>Приклад 93-2: Представлення перерахованих варіантів</b> .....  | 193 |
| <b>Приклад 93-3: Представлення консенсусної послідовності</b> .....   | 194 |
| <i>Параграф 94 - Варіантна послідовність, розкрита як одна послідовність з перерахованими альтернативними залишками</i> .....                                 | 196 |
| <b>Приклад 94-1: Представлення однієї послідовності з перерахованими альтернативними амінокислотами</b> .....   | 196 |
| <b>Приклад 94-2 – Представлення однієї послідовності з перерахованими альтернативними амінокислотами, які можуть бути модифікованими амінокислотами</b> ..... | 197 |
| <i>Параграф 95(a) – Варіантна послідовність, яка розкривається лише за посиланням на первинну послідовність з кількома незалежними варіантами</i> .....       | 198 |

---

|   |            |
|---|------------|
| <b>Приклад 95(а)-1: Представлення варіантної послідовності шляхом тлумачення первинної послідовності.....</b>   | <b>198</b> |
| <i>Параграф 95(в) - Варіантна послідовність, що розкривається лише за допомогою посилання на первинну послідовність з кількома взаємозалежними варіантами .....</i> | <i>200</i> |
| <b>Приклад 95(в)-1: Представлення окремих варіантних послідовностей з кількома взаємозалежними варіантами .....</b>   | <b>200</b> |

## ПРИКЛАДИ

### *Параграф 3(a) Визначення "амінокислота"*

#### **Приклад 3(a)-1: D-амінокислоти**

Патентна заявка описує наступну послідовність:

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

#### **Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності(ей)?**

##### **ТАК**

Параграф 3(a) стандарту VOIB ST.26 визначає "амінокислоту" як таку, що містить "D-амінокислоти" та амінокислоти, що містять модифіковані або синтетичні бічні ланцюги. Виходячи з цього визначення, перерахований пептид містить п'ять амінокислот, які є специфічно визначеними (D-Ala, D-Glu, Lys, Gly, D-Met). Тому, відповідно до параграфа 7(b) стандарту VOIB ST.26, послідовність має бути включена до переліку послідовностей.

#### **Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Параграф 29 вимагає, щоб D-амінокислоти були представлені у послідовності як відповідна немодифікована L-амінокислота. Крім того, будь-яка модифікована амінокислота, яка не може бути представлена будь-яким іншим символом у Таблиці 3 Розділу 3 Додатку I, повинна бути представлена символом "X".

У цьому прикладі послідовність містить три D-амінокислоти, які можуть бути представлені немодифікованою L-амінокислотою в Таблиці 3 Розділу 3 Додатка I, однією L-амінокислотою (Nle) і однією D-амінокислотою (D-Nle), що повинна бути представлена символом "X".

Параграф 25 вказує на те, що коли амінокислотні послідовності є кільцевим і кільце складається виключно з амінокислотних залишків, з'єднаних пептидними зв'язками, заявник повинен обрати амінокислоту в положенні залишку №1. Відповідно, послідовність може бути представлена як:

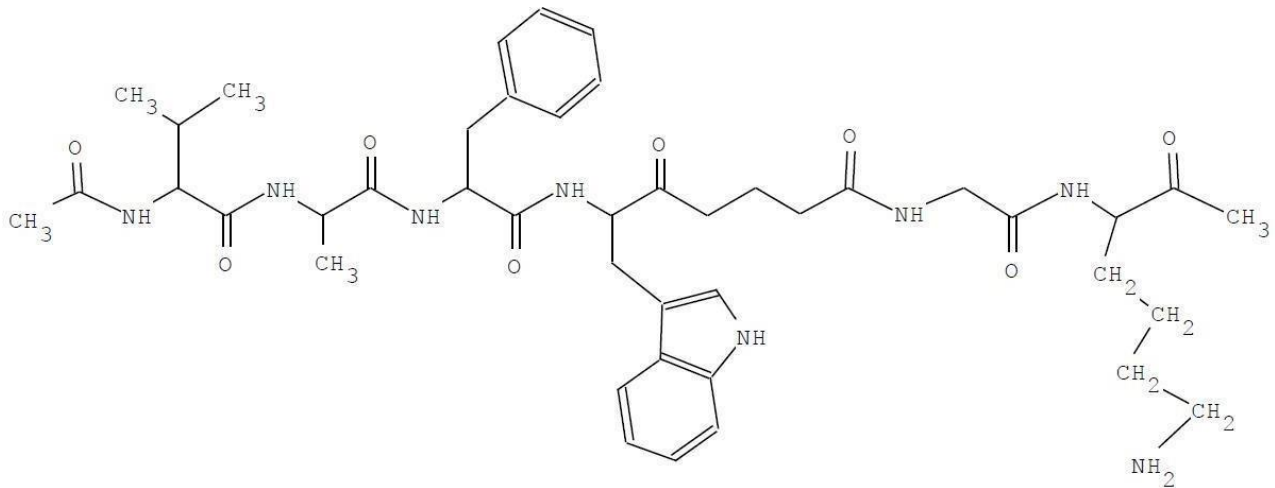
AEKXGMX (SEQ ID NO: 1)

або іншим способом, з будь-якою іншою амінокислотою у послідовності в 1 положенні залишку. Ключ функції "SITE" та кваліфікатор "note" повинні бути вказані для кожної D-амінокислоти повною, нескороченою назвою D-амінокислоти в якості значення кваліфікатора, наприклад, D-аланін і D-норлейцин. Крім того, ключ функції "SITE" та кваліфікатор "note" повинні мати скорочення L-норлейцину в якості значення кваліфікатора, тобто "Nle", як зазначено в Таблиці 4 Розділу 4 Додатка I. Нарешті, ключ функції "REGION" та кваліфікатор "note" мають бути надані для позначення кільцевого пептиду.

#### **Відповідні параграфи стандарту VOIB ST.26: 3(a), 7(b), 25, 26, 29, 30, та 31**

Параграф 3(с) – Визначення "перелік залишків"

Приклад 3(с)-1: Перелік амінокислот за хімічною структурою



Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

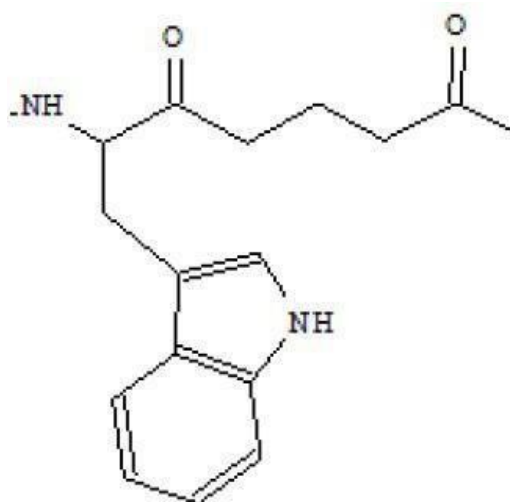
**ТАК**

Перерахований пептид, зображений як структура, містить принаймні чотири специфічно визначені амінокислоти. Тому послідовність повинна бути включена до переліку послідовностей.

Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Послідовність може бути представлена як:

VAFXGK (SEQ ID NO: 2)



де "X" представляє "іншу" модифіковану амінокислоту: , яка вимагає ключ функції "SITE" разом з кваліфікатором "note". Кваліфікатор "note" надає повну, нескорочену назву модифікованого триптофану в положенні 4 перерахованого пептиду, наприклад,



"6-аміно-7-(1H-індол-3-їл)-5-оксогектанова кислота". Далі додатковий ключ функції "SITE" та кваліфікатор "note" необхідні для позначення ацетилювання N-кінця та метилювання C-кінця.

В іншому випадку, послідовність може бути представлена як:

VAFW (SEQ ID NO: 3)

Ключ функції "SITE" та кваліфікатор "note" необхідні для позначення модифікації триптофану в положенні 4 переліченого пептиду із значенням: "С-кінець з'єднаний глютаральдегідним мостом з дипептидом GK". Далі додатковий ключ функції "SITE" в положенні 1 та кваліфікатор "note" необхідні для позначення ацетилювання N-кінця.

**Відповідні параграфи стандарту VOIB ST.26): 3(с), 7(b), 29, 30, та 31**

**Приклад 3(с)-2: Скорочена формула для амінокислотної послідовності**(G<sub>4</sub>z)<sub>n</sub>

Де G= Гліцин, z = будь-яка амінокислота та змінна "n" можуть бути будь-яким цілим числом.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?****ТАК**

Опис показує, що "n" може бути "будь-яким цілим числом"; отже, найбільш охоплюючий варіант здійснення "n" невизначений. Оскільки "n" невизначений, пептид формули не може бути розширений до певної довжини і тому повинна бути розглянута нерозширена формула.

Перелічений пептид у нерозширеній формулі ("n" = 1) забезпечує чотири специфічно визначені амінокислоти, кожна з яких є Gly і символ "z". Зазвичай "Z" є символом "глутаміну або глутамінової кислоти"; однак, приклад визначає "z" як "будь-яку амінокислоту". Відповідно до стандарту VOIB ST.26 амінокислота, яка не є специфічно визначеною, представляється "X". На підставі цього аналізу перерахований пептид, тобто GGGGX, містить чотири залишки гліцину, які є перерахованими та специфічно визначеними. Таким чином, параграф 7(b) стандарту VOIB ST.26 вимагає включення послідовності до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (ності) у переліку послідовностей?**

Послідовність використовує нестандартний символ "z", визначення якого повинно бути надано в описі (див. Вступ до цього документа). Оскільки "z" визначається як будь-яка амінокислота, стандартний символ, що використовується для представлення цієї амінокислоти є "X". Отже, послідовність повинна бути представлена як одна послідовність:

GGGGX (SEQ ID NO: 4)

переважно анотована ключем функції "REGION", локалізація функції "&gt;5" (відповідає >5), з кваліфікатором "note" зі значенням "Вся послідовність амінокислот 1-5 може повторюватися один або більше разів".

Відповідно до параграфа 27, "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 із додатковим описом у таблиці функцій. Тому, якщо символ "X" повинен використовуватись для представлення "будь-якої амінокислоти", тоді він має бути описаний з використанням ключа функції "VARIANT" та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".

Де доцільно, кожен "X" має бути описаний окремо. Проте, область безперервних залишків "X" або множина залишків "X", розподілена в послідовності, може бути спільно описана за допомогою ключа функції "VARIANT" з використанням синтаксису "x..y" в якості дескриптора локалізації, де "x" та "y" – положення першого та останнього залишків "X"; та кваліфікатора "note" зі значенням "X може бути будь-якою

амінокислотою".

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту VOIB ST.26: 3(c), 7(b) та 27**

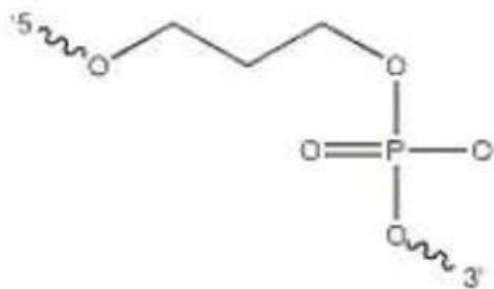
Параграф 3(g) - Визначення "нуклеотид"

**Приклад 3(g)-1: Нуклеотидна послідовність, перервана спейсером C3**

Патентна заявка описує таку послідовність:

atgcatgcatgcnscggcatgcatgc

де n = спейсер C3 з наступною структурою:



**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Перерахована послідовність містить два сегменти специфічно визначених нуклеотидів, відокремлених спейсером C3.

Спейсер C3 не є нуклеотидом відповідно до параграфу 3(g); стандартний символ "n" використовується нестандартним способом (див. Вступ до цього документа). Отже, кожен сегмент є окремою нуклеотидною послідовністю. Оскільки кожен сегмент містить більше 10 специфічно визначених нуклеотидів, вони повинні бути включені до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Кожен сегмент повинен бути включений до переліку послідовностей як окрема послідовність, кожна з яких має власний ідентифікаційний номер послідовності:

atgcatgcatgc (SEQ ID NO: 5)

cggcatgcatgc (SEQ ID NO: 6)

Цитозин у кожному сегменті, який приєднується до спейсера C3, слід додатково описати в таблиці функцій використовуючи ключ функції "misc\_feature" і кваліфікатор "note". Значення кваліфікатора "note", яке є "довільним текстом", має вказувати на наявність спейсера, що з'єднаний з іншою нуклеїновою кислотою та ідентифікувати спейсер або за допомогою повної нескорченої хімічної назви, або за допомогою загальної назви, наприклад спейсер C3.

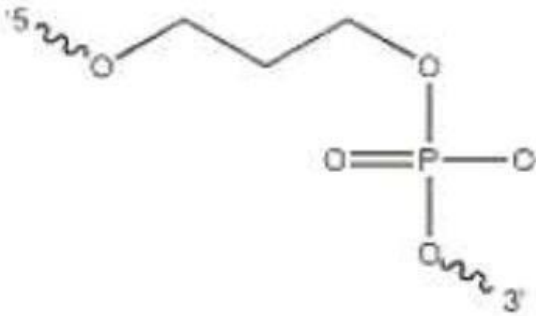
**Відповідні параграфи стандарту VOIB ST.26: 3(g), 7(a) та 15**

### Приклад 3(g)-2: Нуклеотидна послідовність з альтернативними залишками, включаючи спейсер С3

Патентна заявка описує таку послідовність:

atgcatgcatgcncgcatgcatgc

де n = c, a, g або спейсер С3 з наступною структурою:



#### Питання 1: Чи вимагає стандарт ВОІВ ST.26 включення послідовності (-ей)?

**ТАК**

У переліку послідовностей, що переривається змінною "n" міститься 24 специфічно визначені залишки. Потрібно враховувати пояснення послідовності в описі для того, щоб визначити, чи "n" використовується стандартним або нестандартним способом (див. Вступ до цього документа).

Опис показує, що n = c, a, g або спейсер С3. "n" - це стандартний символ, який використовується нестандартним способом, оскільки він описується як символ, що містить спейсер С3, який не відповідає визначенню нуклеотиду. Символ "n" також описується як такий, що містить "c", "a" або "g"; тому стандарт ВОІВ ST.26 вимагає включення 25 нуклеотидної послідовності до переліку послідовностей. Оскільки два сегменти, розділені спейсером С3, є послідовностями, відмінними від 25 нуклеотидної послідовності, дві 12 нуклеотидні послідовності також можуть бути включені.

#### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Приклад вказує на те, що "n = c, a, g, або спейсер С3". Як зазначалося вище, спейсер С3 не є нуклеотидом. Відповідно до параграфу 15, символ "n" не повинен використовуватися для позначення нічого іншого, окрім нуклеотиду; отже, символ "n" не може представляти спейсер С3 в переліку послідовностей.

У параграфі 15 також зазначається, що, якщо багатозначний символ є доречним, тоді слід використовувати більш обмежувальний символ. Символ "v" представляє "a або c або g" відповідно до Таблиці 1 Розділу 1 Додатку I, який є більш обмежувальним, ніж "n".

Якщо змінна "n" в прикладі є c, a або g, тоді одна послідовність, наведена її залишками, що містить найбільш розкриті варіанти здійснення, і тому є найбільш охоплюючою послідовністю (див. Вступ до цього документу), яка повинна бути включена до переліку послідовностей являє собою:

atgcatgcatgcvcgcatgcatgc (SEQ ID NO: 7)

Настійно рекомендується включення будь-яких додаткових послідовностей, необхідних для опису інформації або пунктів формули винаходу, як зазначалося у вступі до цього документа.

Якщо змінна "n" у прикладі є спейсером C3, то послідовність може розглядатися як два окремі сегменти специфічно визначених нуклеотидів по обидві сторони від змінної "n", тобто, atgcatgcatgc (SEQ ID NO: 8); і cggcatgcatgc (SEQ ID NO: 9). Якщо це важливо для опису або формули винаходу, то ці дві послідовності також мають бути включені до переліку послідовностей, кожна з власним ідентифікаційним номером послідовності.

Цитозин у кожному сегменті, який приєднується до спейсера C3, слід додатково описати в таблиці функцій використовуючи ключ функції "misc\_feature" і кваліфікатор "note". Значення кваліфікатора "note", яке є "довільним текстом", має вказувати на наявність спейсера, який з'єднаний з іншою нуклеїновою кислотою та ідентифікувати спейсер за допомогою його повної нескороченої хімічної назви або загальної назви, наприклад, спейсер C3.

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту VOIB ST.26: 3(g), 7(a) та 15**

**Приклад 3(g)-3: Апуриновий сайт**

Патентна заявка описує наступну послідовність:

`gagcattgac-AP-taaggct`

де AP - це апуриновий сайт

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Специфічно визначені залишки перерахованої послідовності перериваються апуриновим сайтом. 5' кінець апуринового сайту складається з 10 нуклеотидів, а 3' кінець апуринового сайту складається з 7 нуклеотидів. Параграф 3(g)(ii)(2) визначає апуриновий сайт як "нуклеотид", якщо він є частиною нуклеотидної послідовності. Отже, апуриновий сайт в цьому прикладі розглядається як "нуклеотид" для визначення того, чи послідовність повинна бути включена та яким чином послідовність має бути включена до переліку послідовностей. Відповідно, залишки на кожному кінці апуринового сайту є частиною єдиного переліку послідовностей, що містить 18 нуклеотидів загалом, 17 з яких специфічно визначені. Тому, відповідно до параграфа 7(a) стандарту VOIB ST.26, послідовність повинна бути включена до переліку послідовностей як одна послідовність.

**Питання 3: Як повинна бути представлена послідовність у переліку послідовностей?**

Послідовність повинна бути включена до переліку послідовностей як:

`gagcattgacntaaggct (SEQ ID NO: 10)`

Апуриновий сайт повинен бути представлений "n" і повинен бути додатково описаний в таблиці функцій. Бажаний спосіб тлумачення – ключ функції "modified\_base" та обов'язковий кваліфікатор "mod\_base" зі значенням "OTHER". Необхідно включити кваліфікатор "note", який описує модифіковану основу як апуриновий сайт.

**Відповідні параграфи стандарту VOIB ST.26: 3(g), 7(a) та 17**

**Приклад 3(g)-4: Аналоги нуклеїнових кислот**

Патентна заявка описує наступну послідовність гліколевої нуклеїнової кислоти (GNA):

PO<sub>4</sub> - tagttcattgactaaggctcccctgact-OH

У якій лівий кінець послідовності імітує 5'-кінець послідовності ДНК.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК** - Окремі залишки, які містять послідовність GNA, вважаються нуклеотидами відповідно до параграфу 3(g)(i)(2) стандарту VOIB ST.26. Відповідно, послідовність має більше десятих перелічених і "специфічно визначених" нуклеотидів і повинна бути включена до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовності GNA не мають 5'-кінця та 3'-кінця, але мають 3'-кінець і 2'-кінець. 3'-кінець, який зазвичай зображується як такий, що має термінальну фосфатну групу, відповідає 5'-кінцю ДНК або РНК. (Варто зауважити, що інші аналоги нуклеїнових кислот можуть відповідати 5'-кінцю та 3'-кінцю ДНК і РНК, відповідно). Відповідно до параграфу 11, вона повинна бути включена до переліку послідовностей "у напрямку зліва направо, що імітує напрямок від 5'-кінця до 3'-кінця". Тому вона повинна бути включена до переліку послідовностей як:

tagttcattgactaaggctcccctgact (SEQ ID NO: 11)

Послідовність повинна бути описана в таблиці функцій з застосуванням ключа функції, "modified\_base" та обов'язкового кваліфікатора "mod\_base" зі аббревіатурою "OTHER". Кваліфікатор "note" повинен бути включений до повної нескороченої назви модифікованих нуклеотидів, таких як "гліколеві нуклеїнові кислоти" або "2,3-дигідроксипропілнуклеозиди". Один елемент "INSDFeature" може використовуватись для опису всієї послідовності як GNA, якщо "INSDFeature\_location" має діапазон "1..30".

**Відповідні параграфи стандарту VOIB ST.26:** 3(d), 3(g), 7(a), 11, 16, 18, 65 та 66



*Параграф 3 (к) - Визначення "специфічно визначений"*

**Приклад 3(к)-1: Багатозначні символи нуклеотидів**

5' NNG KNG KNG K 3'

N та K являють собою багатозначні коди IUPAC-IUB

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ**

Багатозначні коди IUPAC-IUB відповідають переліку нуклеотидних символів, визначених у Таблиці 1 Розділу 1 Додатку I. Відповідно до параграфа 3(к), специфічно визначеним нуклеотидом є будь-який нуклеотид, відмінний від того, що представлений символом

"n", зазначений у Додатку I. Тому "K" та "G" є специфічно визначеними нуклеотидами, а "N" не є специфічно визначеним нуклеотидом.

Перерахована послідовність не має десяти або більше специфічно визначених нуклеотидів, і тому параграф 7(а) стандарту VOIB ST.26 не вимагає включення до переліку послідовностей.

**Питання 2: Чи дозволяє стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ**

Відповідно до параграфа 8, "Перелік послідовностей не повинен включати жодних послідовностей, що мають менше десяти специфічно визначених нуклеотидів...". Перерахована послідовність не містить десяти або більше специфічно визначених нуклеотидів; отже, вона не повинна бути включена до переліку послідовностей.

**Відповідні параграфи стандарту VOIB ST.26: 3(к), 7(а), 8 та 13**

**Приклад 3(к)-2: Багатозначний символ "n", який використовується як стандартним, так і нестандартним способом**

Заявка описує штучну послідовність: 5'-AATGCCGGAN-3'. Опис також встановлює:

- (i) в одному варіанті здійснення, N являє собою будь-який нуклеотид;
- (ii) в одному варіанті здійснення, N є обов'язковим, але переважно являє собою G;
- (iii) в одному варіанті здійснення, N являє собою K;
- (iv) в одному варіанті здійснення, N являє собою C.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**Ні**

Перелік послідовностей містить 9 специфічно визначених нуклеотидів та "N". Необхідно взяти до уваги пояснення послідовності в описі для визначення, чи символ "N" використовується стандартним способом (див. Вступ до цього документа).

Розгляд описаних варіантів здійснення (i) - (iv) переліку послідовностей показує, що найбільш охоплюючим варіантом здійснення "N" є "будь-який нуклеотид". У найбільш охоплюючому варіанті здійснення "N" в переліку послідовностей використовується стандартним способом.

У деяких варіантах здійснення "N" описано як специфічно визначені залишки (тобто "N являє собою C" у частині (iv)). Однак тільки найбільш охоплюючий варіант здійснення (тобто "N - будь-який нуклеотид") розглядається при визначенні, чи повинна послідовність бути включена до переліку послідовностей. Таким чином, перерахована послідовність, яка повинна бути оцінена, є 5'-AATGCCGGAN-3'.

На основі цього аналізу перелік послідовностей, тобто AATGCCGGAN, не містить десять специфічно визначених нуклеотидів. Отже, параграф 7(a) стандарту VOIB ST.26 не вимагає включення послідовності до переліку послідовностей, незважаючи на те, що в деяких варіантах здійснення "n" визначається також як специфічні нуклеотиди.

**Питання 2: Чи дозволяє стандарт VOIB ST.26 включення послідовності (-ей)?**

**Ні**

Послідовність "AATGCCGGAN" не повинна бути включена до переліку послідовностей.

Проте описана альтернативна послідовність може бути включена до переліку послідовностей, якщо "N" замінено на специфічно визначений нуклеотид.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

**Настійно** рекомендується включення послідовностей, що представляють варіанти здійснення, які є ключовою частиною винаходу. Включення цих послідовностей дає змогу здійснювати більш ретельний пошук і публічно повідомляти про об'єкт, щодо якого витребується патент.

Для наведеного вище прикладу, настійно рекомендується включити до переліку

---

послідовностей наступні три додаткові послідовності, кожна з яких має власний ідентифікаційний номер послідовності:

aatgccggag (SEQ ID NO: 12)

aatgccggak (SEQ ID NO: 13)

aatgccggac (SEQ ID NO: 14)

Якщо включено менше, ніж усі три вищезазначені послідовності, нуклеотид, який замінює "n", повинен бути розтлумачений для описання альтернатив. Наприклад, якщо тільки вищезазначена послідовність SEQ ID NO: 12 включена до переліку послідовностей, то ключ функції "misc\_difference" з локалізацією функції "10" повинен використовуватися разом з двома кваліфікаторами "replace", де значення для одного буде "k", а для іншого - "c".

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту ВОІВ ST.26: 3(k), 7(a), 8 та 13**

**Приклад 3(к)-3: Багатозначний символ "n", який використовується нестандартним способом**

В заявці описана послідовність: 5'-aatgttggan-3'

Де n являє собою с

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Відповідно до параграфа 3(к), "специфічно визначений" нуклеотид є будь-яким нуклеотидом, відмінним від тих, що представлені символом "n", який наведено в Таблиці 1 Розділу 1 Додатка I.

У цьому прикладі "n" використовується нестандартним способом для представлення лише "с". Описом не зазначено, що "n" використовується стандартним способом для позначення "будь-якого нуклеотиду". Тому послідовність повинна тлумачитися так, ніби еквівалентний стандартний символ, тобто "с", був використаний у послідовності (*див.* Вступ до цього документа). Відповідно, переліченою послідовністю, яку необхідно розглянути є:

5'-aatgttggac-3'

Ця послідовність має десять специфічно визначених нуклеотидів і відповідно до параграфа 7(a) стандарту VOIB ST.26 має бути включена до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність повинна бути включена до переліку послідовностей як: aatgttggac (SEQ ID NO: 15)

**Відповідні параграфи стандарту VOIB ST.26: 3(к) та 7(a)**

**Приклад 3(к)-4: Багатозначні символи, окрім "n", є "специфічно визначеними"**

Патентна заявка описує таку послідовність:

5' NNG KNG KNG KAG VCR 3'

де N, K, V та R – багатозначні коди IUPAC-IUB

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?****ТАК**

Багатозначні коди IUPAC-IUB відповідають переліку нуклеотидних символів, визначених у Таблиці 1 Розділу 1 Додатка I. Відповідно до параграфа 3(к), "специфічно визначений" нуклеотид є будь-яким нуклеотидом, відмінним від тих, що представлені символом "n", і наведені у Таблиці 1 Розділу 1 Додатку I. Отже, "K", "V" та "R" є "специфічно визначеними" нуклеотидами.

Послідовність має одинадцять перелічених та "специфічно визначених" нуклеотидів, і відповідно до параграфа 7(a) стандарту VOIB ST.26 повинна бути включена до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність повинна бути включена до переліку послідовностей як:

nngkngkngkagvcr (SEQ ID NO: 16)

**Відповідні параграфи стандарту VOIB ST.26: 3(к), 7(a) та 15**

**Приклад 3(к)-5: Багатозначне скорочення "Хаа" використовується нестандартним способом**

Патентна заявка описує таку послідовність:

Хаа-Tyr-Glu-Хаа-Хаа-Хаа-Leu

Де Хаа у положенні 1 є будь-якою амінокислотою, Хаа у положенні 4 являє собою Lys, Хаа у положенні 5 являє собою Gly та Хаа у положенні 6 являє собою лейцин або ізолейцин.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Перелічений у формулі пептид містить три специфічно визначені амінокислоти у положеннях 2, 3 і 7. Перша амінокислота представлена стандартним скороченням, тобто Хаа, що представляє будь-яку амінокислоту. Однак четверта, п'ята та шоста амінокислоти представлені стандартним скороченням, яке використовується нестандартним способом (див. Вступ до цього документа). Тому враховується пояснення послідовності в описі для надання визначення "Хаа" в цих положеннях. Оскільки "Хаа" в положеннях 4-6 вказується як специфічна амінокислота, послідовність повинна тлумачитися так, ніби в послідовності були використані еквівалентні стандартні скорочення, тобто Lys, Gly і (Leu або Ile). Отже, послідовність містить чотири або більше специфічно визначених амінокислот і повинна бути включена до переліку послідовностей, як того вимагає параграф 7(b) стандарту VOIB ST.26.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність використовує стандартне скорочення "Хаа" нестандартним способом. Тому, для того, щоб дати визначення "Хаа" у положеннях 4, 5 та 6, необхідно враховувати пояснення послідовності в описі. Пояснення визначає "Хаа" як лізин у положенні 4, гліцин у положенні 5 та лейцин або ізолейцин у положенні 6. Стандартними символами для цих амінокислот є, відповідно, K, G і J. Тому послідовність повинна бути представлена в переліку послідовностей як:

XUEKGLJ (SEQ ID NO: 17)

Відповідно до параграфа 27, "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 із додатковим описом у таблиці функцій. Тому, якщо символ "X" повинен використовуватись для представлення "будь-якої амінокислоти", тоді він має бути описаний з використанням ключа функції "VARIANT" та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".

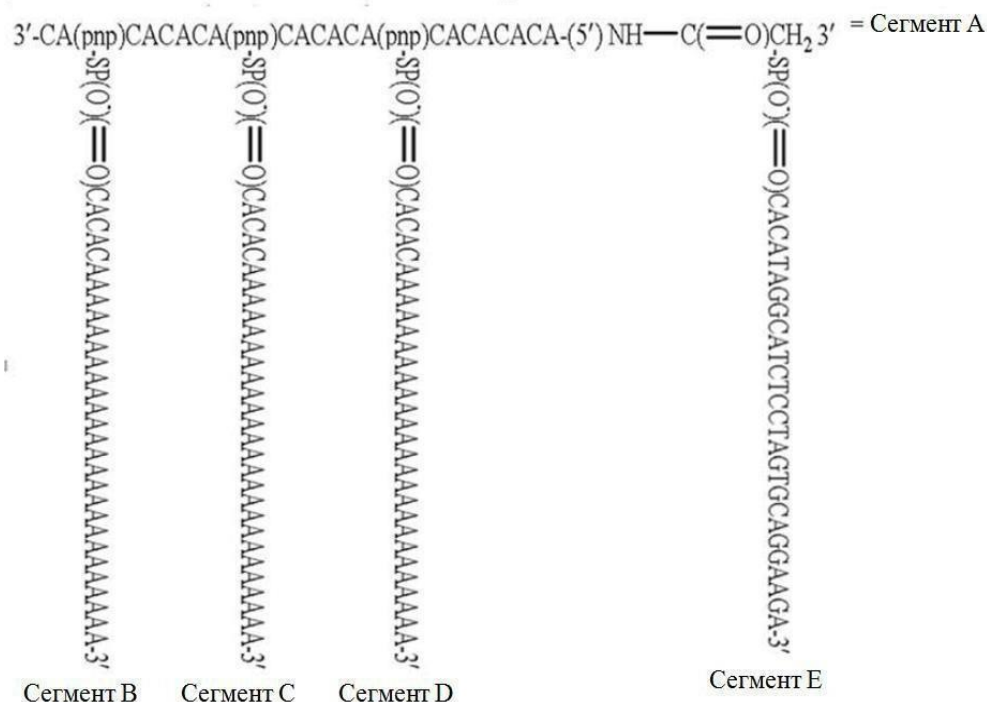
Де доцільно, кожен "X" має бути описаний окремо. Проте, область безперервних залишків "X" або множина залишків "X", розподілена в послідовності, може бути спільно описана за допомогою ключа функції "VARIANT" з використанням синтаксису "x.y" в якості дескриптора локалізації, де "x" та "y" – положення першого та останнього залишків "X"; та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".

**Відповідні параграфи стандарту ВОІВ СТ.26: 3(к), 7(b), 26 та 27**

Параграф 7(a) - Нуклеотидні послідовності, які необхідно зазначити в переліку послідовностей

### Приклад 7(a)-1: Розгалужена нуклеотидна послідовність

В описі розкривається наступна розгалужена нуклеотидна послідовність:



де "pnp" є зв'язком або мономером, що має функціональність бромацетиламіну;  
 $3'-\text{CA}(\text{pnp})\text{CACACA}(\text{pnp})\text{CACACA}(\text{pnp})\text{CACACACA}-(5')\text{NH}-\text{C}(=\text{O})\text{CH}_2 3'$  є сегментом А;  
 $\text{SP}(\text{O}^-)(=\text{O})\text{CACACAAAAA} 3'$  є сегментами В, С і D; і  
 $\text{SP}(\text{O}^-)(=\text{O})\text{CACATAGGCATCTCTAGTGCAGGAAGA} 3'$  є сегментом Е.

#### Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

**ТАК** - чотири вертикальні сегменти В-Е повинні бути включені до переліку послідовностей

**НІ** - горизонтальний сегмент А не повинен бути включений до переліку послідовностей

Наведена фігура є прикладом розгалуженої послідовності нуклеїнової кислоти "комбінованого типу", що містить п'ять лінійних сегментів: горизонтальний сегмент А і чотири вертикальні сегменти В-Е.

Відповідно до параграфа 7(a), лінійні області розгалужених нуклеотидних послідовностей, що містять десять або більше специфічно визначених нуклеотидів, в яких суміжні нуклеотиди поєднуються у напрямку від 3'- до 5'-кінця, повинні бути включені до переліку послідовностей.

Чотири вертикальні сегменти В-Е містять понад десять специфічно визначених нуклеотидів, в яких суміжні нуклеотиди поєднуються у напрямку від 3'- до 5'-кінця, і тому кожна послідовність повинна бути включена до переліку послідовностей.

У горизонтальному сегменті А лінійні області нуклеотидної послідовності зв'язані ненуклеотидним компонентом "pnp" і кожна з цих зв'язаних лінійних областей містить менше десяти специфічно визначених нуклеотидів. Тому, оскільки жодна область сегмента А не містить десять або більше специфічно визначених нуклеотидів, де



суміжні нуклеотиди поєднуються у напрямку від 3'- до 5'-кінця, відповідно до параграфа 7(a) стандарту VOIB ST.26, вони не є обов'язковими для включення до переліку послідовностей.

**Питання 2: Чи дозволяє стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ**

Відповідно до параграфа 8, "Перелік послідовностей не повинен включати жодних послідовностей з менш ніж десятьма специфічно визначеними нуклеотидами..." "

Жодна область сегмента А не містить десять або більше специфічно визначених нуклеотидів, в яких суміжні нуклеотиди поєднуються у напрямку від 3'- до 5'-кінця; отже, вона не повинна бути включена до переліку послідовностей як окрема послідовність з її власним ідентифікаційним номером послідовності.

Однак сегменти В, С, D та Е можуть бути анотовані, щоб вказати, що вони зв'язані з сегментом А.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Сегменти В, С та D ідентичні і повинні бути включені до переліку послідовностей як одна послідовність:

casasaaaaaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO: 18)

Перший "с" в послідовності слід додатково описати як модифікований нуклеотид, використовуючи ключ функції "misc\_feature" і кваліфікатор "note" зі значенням, наприклад: "Ця послідовність є однією з чотирьох гілок розгалуженого полінуклеотиду".

Сегмент Е повинен бути включений до переліку послідовностей як одна послідовність:

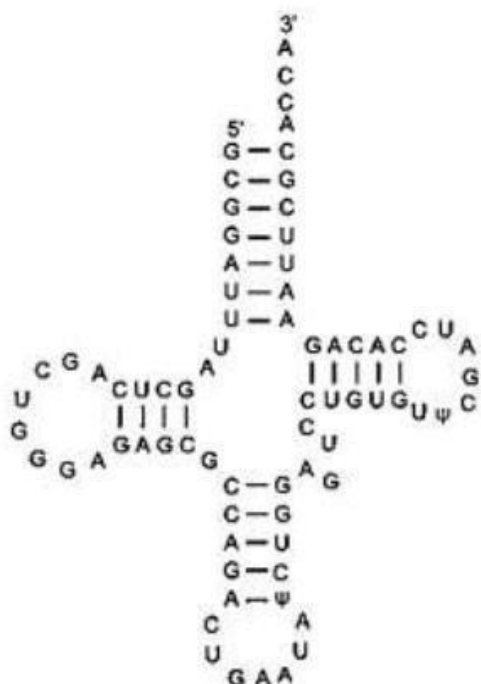
casataggcatctcctagtgacgaaga. (SEQ ID NO: 19)

Перший "с" в послідовності слід додатково описати як модифікований нуклеотид, використовуючи ключ функції "misc\_feature" і кваліфікатор "note" зі значенням, наприклад: "Ця послідовність є однією з чотирьох гілок розгалуженого полінуклеотиду".

**Відповідні параграфи стандарту VOIB ST.26: 7(a), 8, 11, 13 та 17**

## Приклад 7(а)-2: Лінійна нуклеотидна послідовність, що має вторинну структуру

Патентна заявка описує наступну послідовність:



Де Ψ являє собою псевдоуридин.

### Питання 1: Чи вимагає стандарт ВОІВ ST.26 включення послідовності (-ей)?

**ТАК**

Нуклеотидна послідовність містить 73 перерахованих і специфічно визначених нуклеотиди. Таким чином, приклад має десять або більше "специфічно визначених" нуклеотидів, і, як того вимагає параграф 7(а) стандарту ВОІВ ST.26, повинна бути включена до переліку послідовностей.

### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

З опису випливає, що "Ψ" є еквівалентом псевдоуридину. Єдиний стандартний символ, що може бути використаний для позначення псевдоуридину являє собою "n"; отже, "Ψ" - це стандартний символ, який використовується для позначення стандартного символу "n" (див. Вступ до цього документа). Відповідно, послідовність повинна тлумачитися як така, що має два символи "n" замість двох символів "Ψ".

Символ "u" не повинен використовуватися для позначення урацилу в молекулі РНК у переліку послідовностей. Відповідно до параграфа 14, символ "t" буде тлумачитися як урацил в РНК. Послідовність повинна бути включена як:

gcggttagctcagctgggagagcgccagactgaatanctggagtcctgtgtncgatccacagaattcgacca (SEQ ID NO: 20)

Значенням обов'язкового кваліфікатора "mol\_type" обов'язкового ключа функції "source" є "tRNA". Додаткова інформація може надаватися за допомогою ключа функції "tRNA" та будь-яким (-и) відповідним (-и) кваліфікатором (-ами).

---

Залишки "n" повинні бути додатково описані в таблиці функцій використовуючи ключ функції "modified\_base" і обов'язковий кваліфікатор "mod\_base" зі скороченням "p" для псевдоурдину як значення кваліфікатора (див. Таблиця 2 Додатку I).

**Відповідні параграфи стандарту VOIB ST.26: 7(a), 11, 13, 14, 17, 62, 84 та ключ функції 5.43 Розділів 2 та 5 Додатка I**

**Приклад 7(а)-3: Багатозначні символи нуклеотидів, що використовуються нестандартним способом**

Патентна заявка описує таку послідовність:

5' GATC-MDR-MDR-MDR-MDR-GTAC 3'

Пояснення послідовності в описі далі вказує на те, що "DR Element" складається з послідовності 5' ATCAGCCAT 3'. Мутантний DR Element, або MDR, є елементом DR, в якому середні 5 нуклеотидів, CAGCC, мутовані до TTTTT".

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?****ТАК**

Перерахована послідовність використовує символ "MDR". Якщо незрозуміло, чи символ, що використовується в послідовності, призначений як стандартний символ, тобто символ, зазначений у Таблиці 3 Розділу 3 Додатку I, чи як нестандартний символ, для надання визначення необхідно враховувати пояснення послідовності в описі (*див.* Вступ до цього документа). Згідно з Таблицею 3, "MDR" можна тлумачити як три стандартні символи (m = a або c, d = a або g або t / u, r = g або a) або як аббревіатуру, яка є скороченим позначенням іншої структури.

З опису випливає, що елемент MDR еквівалентний 5' ATTTTTTAT 3'. Літери "MDR" вважаються стандартними символами, що використовуються нестандартним способом; отже, послідовність повинна тлумачитись так, ніби вона була описана за допомогою еквівалентних стандартних символів. Відповідно, перерахована послідовність, яка розглядається для включення до переліку послідовностей, являє собою:

5' GATC ATTTTTTAT ATTTTTTAT ATTTTTTAT ATTTTTTAT GTAC 3'

Перерахована послідовність має 44 специфічно визначених нуклеотиди і, відповідно до параграфа 7(а) стандарту VOIB ST.26, повинна бути включена до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність повинна бути включена до переліку послідовностей як:

gatcatttttatattttatattttatatttttatgtac (SEQ ID NO: 21)

**Відповідні параграфи стандарту VOIB 26: 7(а) та 13**

**Приклад 7(а)-4: Багатозначні символи нуклеотидів, що використовуються нестандартним способом**

Патентна заявка описує таку послідовність:

5' АТТС-N-N-N-N-GTAC 3'

Пояснення послідовності в описі вказує на те, що "N" складається з послідовності 5' АТАСГСАСТ 3'.

**Питання 1: Чи вимагає стандарт ВОІВ ST.26 включення послідовності (-ей)?****ТАК**

Перерахована послідовність використовує символ "N". Потрібно врахувати пояснення послідовності в описі для того, щоб визначити, чи "N" використовується стандартним або нестандартним способом (*див.* Вступ до цього документа).

З опису випливає, що "N" еквівалентна 5' АТАСГСАСТ 3'. Таким чином, "N" є стандартним символом, що використовується нестандартним способом. Відповідно, послідовність повинна тлумачитися як ніби вона була описана за допомогою еквівалентних стандартних символів:

5' АТТС-АТАСГСАСТ-АТАСГСАСТ-АТАСГСАСТ-АТАСГСАСТ-GTAC 3'

Перерахована послідовність має 44 специфічно визначених нуклеотиди і, відповідно до параграфа 7(а) стандарту ВОІВ ST.26, повинна бути включена до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність повинна бути включена до переліку послідовностей як:

atccatagcactatagcactatagcactatagcactgtac (SEQ ID NO: 22)

**Відповідні параграфи стандарту ВОІВ ST.26: 7(а) та 13**

**Приклад 7(а)-5: Нестандартні символи нуклеотидів**

Патентна заявка описує наступну послідовність:

5' GATC-β-β-β-β-GTAC 3'

Пояснення послідовності в описі далі вказує на те, що "β" складається з послідовності 5' ATACGCACT 3'.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Перерахована послідовність використовує нестандартний символ "β". Потрібно врахувати пояснення послідовності в описі для того, щоб дати визначення "β" (див. Вступ до цього документа).

З опису випливає, що "β" є еквівалентною 5' ATACGCACT 3'. Таким чином, "β" є нестандартним символом, який використовується для позначення послідовності, що складається з дев'яти специфічно визначених стандартних символів. Відповідно, послідовність повинна тлумачитися так, ніби вона була описана за допомогою еквівалентних стандартних символів:

5' GATC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3'

Перерахована послідовність має 44 специфічно визначених нуклеотиди і, відповідно до параграфа 7(а) стандарту VOIB ST.26, повинна бути включена до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність повинна бути включена до переліку послідовностей як:

gatcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO: 23)

**Відповідні параграфи стандарту VOIB ST.26: 7(а) та 13**

**Приклад 7(а)-6: Нестандартні символи нуклеотидів**

Патентна заявка описує таку послідовність:

5' GATC-β-β-β-β-GTAC 3'

Пояснення послідовності в описі далі вказує на те, що "β" дорівнює аденіну, інозину або псевдоуридину.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ**

Перерахована послідовність використовує нестандартний символ "β". Потрібно врахувати пояснення послідовності в описі для того, щоб дати визначення "β" (див. Вступ до цього документа).

З опису випливає, що "β" еквівалентно аденіну, інозину або псевдоуридину. Єдиний стандартний символ, який може використовуватись для представлення "аденіну, інозину або псевдоуридину" – це "n"; отже, "β" є нестандартним символом, що використовується для позначення стандартного символу "n". Відповідно, послідовність повинна тлумачитись як така, що має чотири символи "n" замість чотирьох символів "β":

5' GATC-N-N-N-N-GTAC 3'

Перерахована послідовність містить лише вісім специфічно визначених нуклеотидів і, відповідно до параграфу 7(а) стандарту VOIB ST.26, не повинна бути включена до переліку послідовностей.

**Питання 2: Чи дозволяє стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ**

Перерахована послідовність, 5' GATC-N-N-N-N-GTAC 3' не повинна бути включена до переліку послідовностей.

Проте, описана альтернативна послідовність може бути включена до переліку послідовностей, якщо принаймні 2 з "n" символів замінюються аденіном, в результаті чого утворюється послідовність принаймні з 10 або більше специфічно визначених нуклеотидів.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Одне з можливих дозволених представлень:

gatcaaaagtac (SEQ ID NO: 24)

У наведеному вище прикладі чотири аденінові нуклеотиди, які замінюють β-символи, мають бути описані для зазначення, що ці положення можуть бути замінені інозином або псевдоуридином.

Ключ функції "misc\_difference" має використовуватися з локалізацією функції 5-8 та кваліфікатором "note" зі значенням, наприклад: "Нуклеотид у будь-якому положенні 5- 8 може бути заміщений інозином або псевдоуридином". Оскільки ці альтернативи є модифікованими нуклеотидами, то вимагається ключ функції "modified\_base" разом з кваліфікатором "mod\_base". Значення для кваліфікатора "mod\_base" може бути "OTHER" з кваліфікатором "note" та значенням "i або p".

Можливі інші перестановки.

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище,

---

спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту ВОІВ ST.26: 7(a), 8, 13 та 17**



## Приклад 7(а)-7: Інвертовані нуклеотиди I

У патентній заявці розкривається наступна послідовність одноланцюгової ДНК:

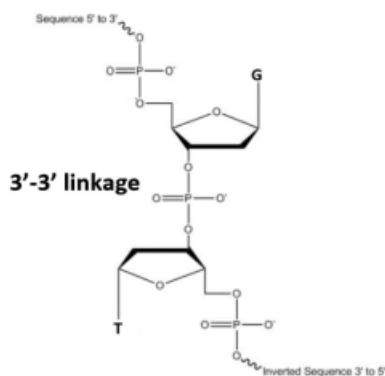
5'attgactaagtgtgttccccattgact5'

При цьому напрямок/направленість послідовності змінюється в межах ланцюга за рахунок зворотного зв'язку 3' на 3' між залишками 12 і 13. Підкреслена частина послідовності орієнтована на 3' - 5' зліва направо.

### Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

**ТАК**

Перші 12 залишків зображені в стандартній орієнтації 5' на 3'. "g" в положенні 12 з'єднаний з "t" в положенні 13 за допомогою зворотного зв'язку 3' на 3':



Решта молекули, зображена в положеннях з 13 по 25, має протилежну орієнтацію - 3' до 5'.

Параграф 11 стандарту VOIB ST.26 вимагає, щоб нуклеотидна послідовність була представлена в напрямку 5' - 3' зліва направо. Тому, щоб правильно представити цю молекулу в переліку послідовностей, вона повинна бути представлена двома послідовностями - одна послідовність для положень з 1 по 12, а друга - для положень з 13 по 25. Кожна частина послідовності містить десять або більше "специфічно визначених" нуклеотидів і, відповідно до вимог параграфу 7(а) стандарту VOIB ST.26, повинна бути включена до переліку послідовностей.

### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Положення 1-12 повинні бути включені до переліку послідовностей як:

attgactaagtg (SEQ ID №: 99)

Положення 12 повинно бути описане в таблиці функцій з використанням ключа функції "misc\_feature" і кваліфікатора "note" зі значенням, що вказує на те, що залишок з'єднаний з інвертованою нуклеотидною послідовністю через 3'-3' фосфодієфірний зв'язок з тимідин-3'-монофосфатом.

Положення 13-25 повинні бути включені до переліку послідовностей як:

tcagttaccctt (SEQ ID №: 100)

Варто зауважити, що ця послідовність має зворотну орієнтацію по відношенню до того, як вона була зображена в первинно поданому описі, таким чином, що вона тепер орієнтована 5' на 3', зліва направо. Положення 13 повинно бути описане в таблиці функцій

---

з використанням функції "misc\_feature" і кваліфікатора "note" зі значенням, що вказує на те, що залишок з'єднаний з інвертованою нуклеотидною послідовністю через 3'-3' фосфодієфірний зв'язок з тимідин-3'-монофосфатом.

**Відповідні параграфи стандарту VOIB ST.26:** 7(a), 11

## Приклад 7(а)-8: Інвертовані нуклеотиди II

У патентній заявці розкривається наступна послідовність ДНК:

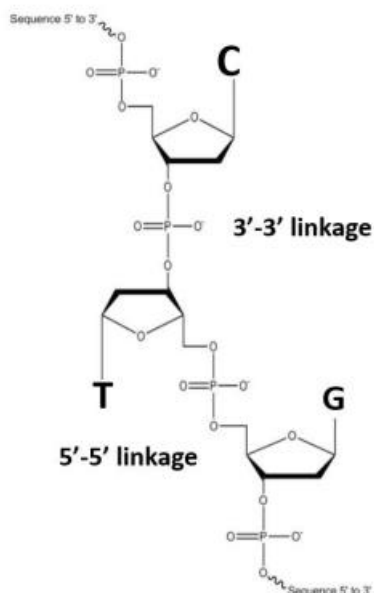
5' - attgactaagactgactgactgact - 3'

Де підкреслений залишок тиміну (положення 14) є інвертованим нуклеотидом, який з'єднаний з цитозином через фосфодієфірний зв'язок 3'-3', а з гуаніном - фосфодієфірним зв'язком 5'-5'.

### Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

**ТАК**

Інвертований тимідин в положенні 14 перериває спрямованість послідовності 5' до 3', вводячи зв'язок 3'-3' між залишками 13 і 14 та 5'-5' зв'язку між залишками 14 і 15:



Інвертований тимідин з'єднує першу частину послідовності, залишки 1-13, і другу частину послідовності, залишки 15-25. Кожна частина послідовності містить щонайменше 10 перелічених і специфічно визначених нуклеотидів. Таким чином, кожна частина послідовності повинна бути включена до переліку послідовностей відповідно до вимог параграфу 7(а) стандарту VOIB ST.26. Одиночний тимідин, що з'єднує дві частини послідовності, не може бути включений до переліку послідовностей як окрема послідовність, оскільки він не є послідовністю з щонайменше 10 специфічно визначених нуклеотидів.

### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Інвертований тимідин у положенні 14 перериває напрямок послідовності від 5' до 3', вводячи зв'язок 3'-3' між залишками 13 і 14 та зв'язок 5'-5' між залишками 14 і 15. Параграф 11 стандарту VOIB ST.26 вимагає, щоб нуклеотидна послідовність була представлена в напрямку 5' - 3' зліва направо. Тому, щоб правильно представити цю молекулу в переліку послідовностей, вона повинна бути представлена двома послідовностями - першою послідовністю для залишків з 1 по 13 і другою послідовністю для залишків з 15 по 25.

Положення 1-13 повинні бути включені до переліку послідовностей у вигляді:

attgactaagtgc (SEQ ID NO: 101)

Положення 13 повинно бути описане в таблиці функцій з використанням ключа функції

---

"misc\_feature" та кваліфікатора "note" з значенням, що вказує на те, що залишок з'єднаний з іншою послідовністю за допомогою 3'-3' фосфодієфірного зв'язку з тимідин-3'-монофосфатом, який, у свою чергу, з'єднаний з іншою послідовністю за допомогою 5'-5' фосфодієфірного зв'язку.

Положення 15-25 повинні бути включені до переліку послідовностей як:

gcccattgact (SEQ ID NO: 102)

Положення 1 повинне бути описане в таблиці функцій з використанням ключа функції "misc\_feature" та кваліфікатора "note" зі значенням, що вказує на те, що залишок з'єднаний з іншою послідовністю через 5'-5' фосфодієфірний зв'язок з тимідин 3'-монофосфатом, який, в свою чергу, з'єднаний з іншою послідовністю через 3'-3' фосфодієфірний зв'язок.

**Відповідні параграфи стандарту VOIB ST.26:** 7(a), 11

Параграф 7(b) - Амінокислотні послідовності, які необхідно зазначати в переліку послідовностей

**Приклад 7(b)-1: Чотири або більше специфічно визначені амінокислоти**

XXXXXXXXDXXXXXXXXXXFXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAXXXXXXXXXXXXXXXXXXXXXXXXXGXXX  
XX

Де X = будь-яка амінокислота

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Перелічений пептид містить чотири специфічно визначені амінокислоти. Символ "X" звичайно використовується для позначення амінокислот, які лишилися, як будь-якої амінокислоти (див. Вступ до цього документа).

Оскільки існує чотири специфічно визначені амінокислоти, тобто Asp, Phe, Ala та Gly, параграф 7(b) стандарту VOIB ST.26 вимагає, щоб ця послідовність була включена до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність повинна бути представлена як:

XXXXXXXXDXXXXXXXXXXFXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAXXXXXXXXXXXXXXXXXXXXXXXXXGXXX  
XX (SEQ ID NO: 25)

Відповідно до параграфа 27, "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",

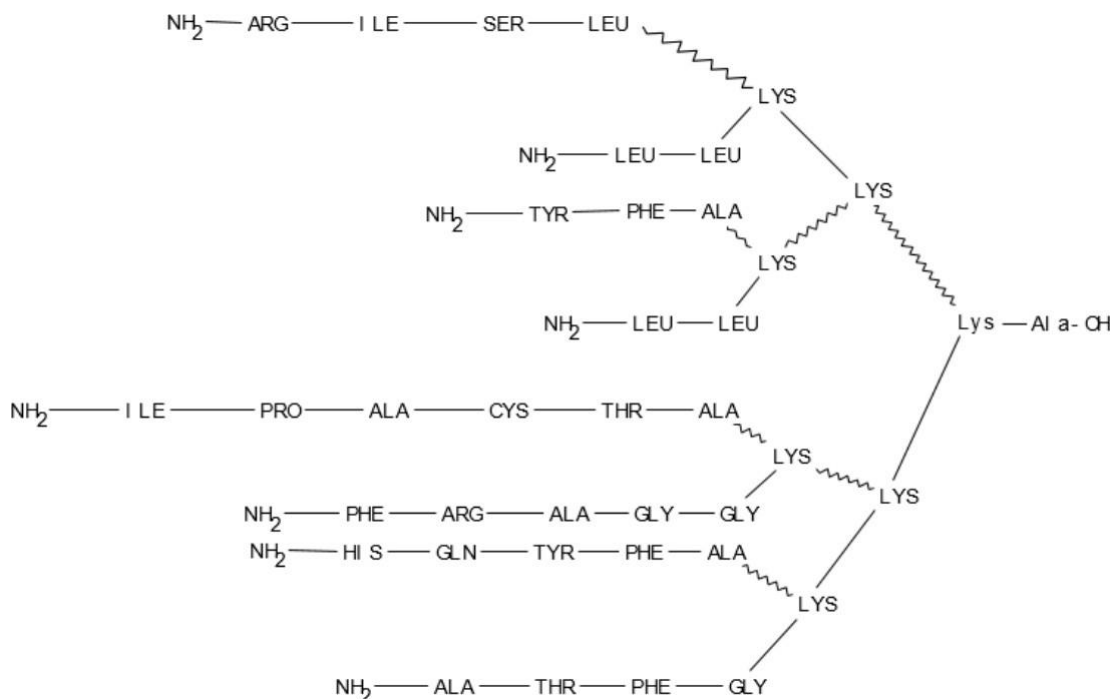
окрім випадків, коли він використовується із додатковим описом у таблиці функцій. Тому, якщо символ "X" повинен використовуватись для представлення "будь-якої амінокислоти", тоді він має бути описаний з використанням ключа функції "VARIANT" та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".


Де доцільно, кожен "X" має бути описаний окремо. Проте, область безперервних залишків "X" або множина залишків "X", розподілена в послідовності, може бути спільно описана за допомогою ключа функції "VARIANT" з використанням синтаксису "x.y" в якості дескриптора локалізації, де "x" та "y" – положення першого та останнього залишків "X"; та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".

**Відповідні параграфи стандарту VOIB ST.26: 7(b), 8 та 27**

### Приклад 7(b)-2: Розгалужена амінокислотна послідовність

Заявка описує розгалужену послідовність, в якій залишки лізину використовуються як основа каркасу для утворення восьми гілок, до яких прикріплено кілька лінійних пептидних ланцюгів. Лізин є двоосновною амінокислотою, що забезпечує її двома сайтами для пептидних зв'язків. Пептид ілюструється таким чином:

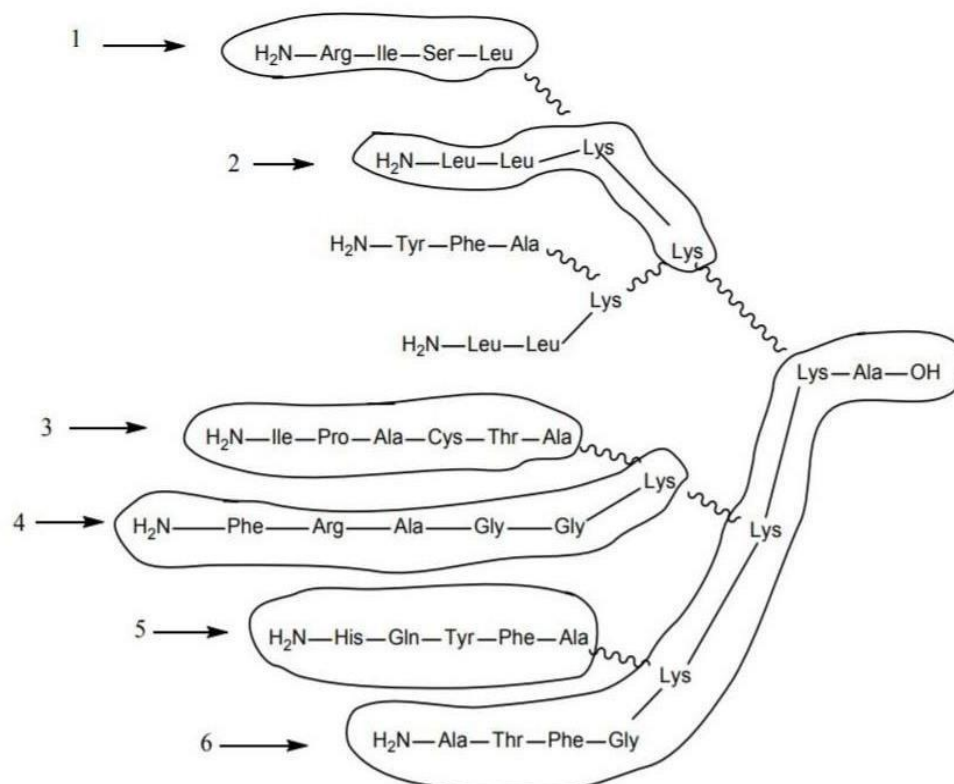


У вищезазначеному розгалуженому пептиді зв'язки між лізином та іншою амінокислотою, зображені —, представляють амідний зв'язок між кінцевим аміном лізину та карбоксильним кінцем зв'язаної амінокислоти. Зв'язки, позначені , представляють амідний зв'язок між бічним ланцюгом аміну лізину та карбоксильним кінцем зв'язаної амінокислоти.

#### Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

**ТАК**

Приклад описує розгалужену послідовність, в якій залишки лізину використовуються в якості каркасу. Параграф 7(b) вимагає, щоб нерозгалужена або лінійна область послідовності, що містить чотири або більше специфічно визначених амінокислоти, була включена до переліку послідовностей. У наведеному вище прикладі лінійні області розгалуженого пептиду, що мають чотири або більше специфічно визначені амінокислоти обведені лініями:



Параграф 7(b) стандарту VOIB ST.26 вимагає включення пептидів 1-6 до переліку послідовностей.

Пептиди, які не потрібно включати до переліку послідовностей:

YFA

LLK

### Питання 2: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

**НІ**

Відповідно до параграфа 8 перелік послідовностей не повинен включати будь-які послідовності з менш ніж чотирма специфічно визначеними амінокислотами.

Кожен з пептидів YFA та LLK містить лише три специфічно визначені амінокислоти, і тому вони не повинні бути включені до переліку послідовностей як окремі послідовності з власними ідентифікаційними номерами послідовностей.

### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Пептиди 1-6 повинні бути представлені окремими ідентифікаторами послідовності:

RISL (SEQ ID NO: 26)

LLKK (SEQ ID NO: 27)

IPACTA (SEQ ID NO: 28)

---

FRAGGK (SEQ ID NO: 29)

HQYFA (SEQ ID NO: 30)

ATFGKKKA (SEQ ID NO: 31)

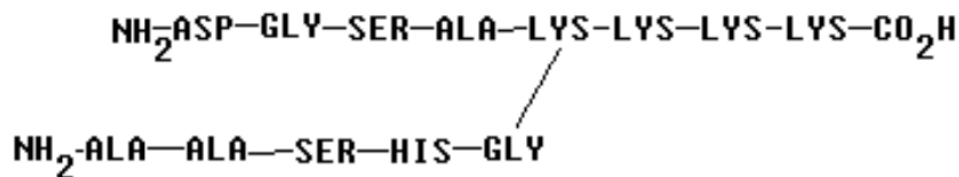
Розгалужена структура може бути розтлумачена, використовуючи ключ функції "SITE" та обов'язковий кваліфікатор "note" зі значенням, наприклад: "Ця послідовність є однією частиною розгалуженої амінокислотної послідовності". Відповідно до параграфа 29 Стандарту VOIB ST.26, послідовності SEQ ID NO 27, 29 та 31 повинні включати тлумачення для кожного лізину для зазначення, що він є модифікованою амінокислотою з використанням ключа функції "SITE" разом з кваліфікатором "note", який описує бічний ланцюг лізину, з'єднаний амідним зв'язком з іншою послідовністю. Рекомендовано, щоб кожна з послідовностей SEQ ID NO 26, 28 та 30 включала тлумачення для зазначення, що С-кінцева амінокислота, зв'язана з іншою послідовністю за допомогою ключа функції "SITE" разом з кваліфікатором "note".

**Відповідні параграфи стандарту VOIB ST.26: 7(b), 8, 26, 29, 30 та 31**



**Приклад 7(b)-3: Розгалужена амінокислотна послідовність**

Пептид наступної послідовності:

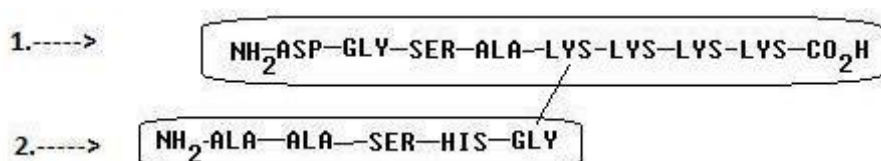


Зв'язок між кінцевим залишком гліцину в нижній послідовності та лізином у верхній послідовності проходить через амідний зв'язок між карбоксильним кінцем гліцину та амінокінцевим бічним ланцюгом лізину.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Нерозгалужена чи лінійна область послідовності, яка містить чотири або більше специфічно визначених амінокислоти, повинна бути включена до переліку послідовностей. У наведеному вище прикладі лінійні області розгалуженого пептиду мають більше чотирьох амінокислот:



Параграф 7(b) стандарту VOIB ST.26 вимагає включення послідовностей 1 та 2 до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовності 1 і 2 повинні бути представлені окремими ідентифікаторами послідовності:

DGSAKKKK (SEQ ID NO: 32)

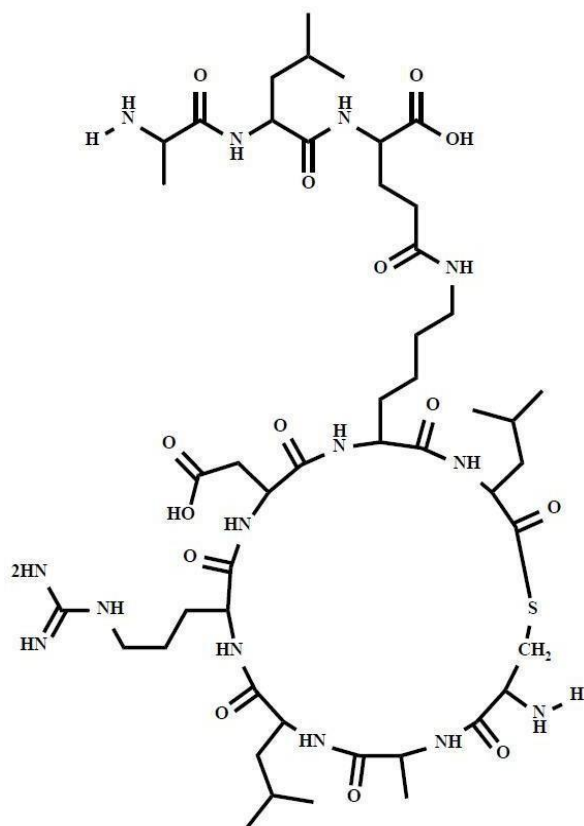
AASHG (SEQ ID NO: 33)

Послідовність DGSAKKKK повинна містити тлумачення, яке вказує на те, що лізин в положенні 5 є модифікованою амінокислотою, використовуючи ключ функції "SITE" разом із кваліфікатором "note", який описує, що бічний ланцюг лізину з'єднаний амідним зв'язком з іншою послідовністю. Бажано, щоб послідовність AASHG мала тлумачення, яке вказує на те, що гліцин в положенні 5 зв'язаний з іншою послідовністю, за допомогою ключа функції "SITE" разом з кваліфікатором "note".

**Відповідні параграфи стандарту VOIB ST.26: 7(b), 26, 29, 30 та 31**

**Приклад 7(b)-4: Циклічний пептид, який містить розгалужену амінокислотну послідовність**

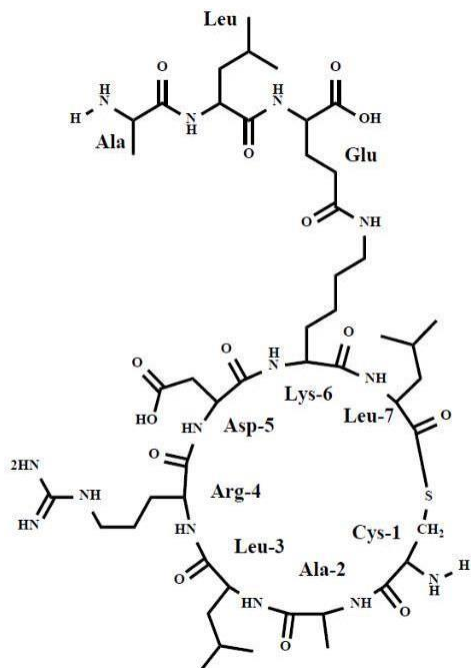
Патентна заявка описує наступну структуру:



Цистеїн та лейцин в циклічній структурі зв'язані бічним ланцюгом Cys та карбонільним кінцем Leu.

**Питання 1: Чи вимагає стандарт ВОІВ ST.26 включення послідовності (-ей)?**

Показана структура має розгалужену циклічну амінокислотну послідовність, яка містить наступні амінокислоти:



Оскільки бічний ланцюг Cys та карбоксиновий кінець Leu беруть участь у циклізації, N-кінець циклічного пептиду розташований на Cys-1.

ТАК – циклічна область пептиду

Відповідно до параграфу 7(b) Стандарту ВОІВ лінійна ділянка розгалуженої послідовності, яка містить чотири або більше специфічно визначені амінокислоти, де амінокислоти утворюють один пептидний каркас, повинна бути включена до переліку послідовностей. В наведеному вище прикладі циклічна область розгалуженого пептиду має більше, ніж чотири амінокислоти, і тому має бути включена до переліку послідовностей.

НІ – трипептидна гілка пептиду

Трипептидна гілка Ala-Leu-Glu не обов'язково має бути в переліку послідовностей.

## Питання 2: Чи вимагає стандарт ВОІВ ST.26 включення послідовності (-ей)?

НІ

Відповідно до параграфу 8, перелік послідовностей не повинен включати будь-яку послідовність з менше, ніж чотирма специфічно визначеними амінокислотами.

Трипептидна гілка містить лише три специфічно визначені амінокислоти і тому не повинна бути включена до переліку послідовностей як окрема послідовність з власним ідентифікаційним номером.

## Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Оскільки цей приклад ілюструє кільцевий пептид, то кільце не складається виключно з амінокислотних залишків в пептидних зв'язках, як зазначено в параграфі 25. Так як циклізування амінокислотної послідовності відбувається за допомогою бічного ланцюга цистеїну (Cys) та карбоксинового кінця лейцину (Leu), цистеїну має бути присвоєно номер положення 1 в межах циклічної області пептиду. Відповідно, послідовність має бути представлена як:

CALRDKL (SEQ ID NO: 90).

Як показано на фігурі вище, амінокислотна послідовність циклізується шляхом

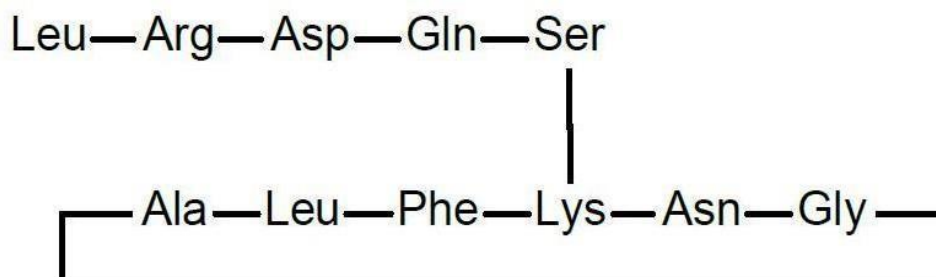
---

тіоестерної кон'югації між бічним ланцюгом цистеїну та карбоксильним кінцем лейцину. Ключ функції "SITE" повинен використовуватись для опису модифікованого цистеїну, який утворює міжланцюговий зв'язок з лейцином. Оператор локалізації елементу це залишкові числа перехрестних амінокислот у форматі "x.y", наприклад, "1..7". Обов'язковий кваліфікатор "note" повинен позначати природу зв'язку, наприклад, "Тіоестер цистеїн лейцину (Cys-Leu)", щоб встановити, що Cys-1 та Leu-7 зв'язані тіоестерним зв'язком. Крім того, лізин в положенні 6 повинен мати тлумачення для зазначення, що він модифікований шляхом використання ключа функції "SITE" разом з обов'язковим кваліфікатором "note", де значення кваліфікатора описує, що бічний ланцюг лізину з'єднується з трипептидом ALE.

**Відповідні параграфи стандарту VOIB ST.26: 7(b), 8, 25, 26, 29, 30, 31, 66(c) та 70**

**Приклад 7(b)-5: Циклічний пептид, який містить розгалужену амінокислотну послідовність**

Патентна заявка описує наступний розгалужений циклічний пептид:

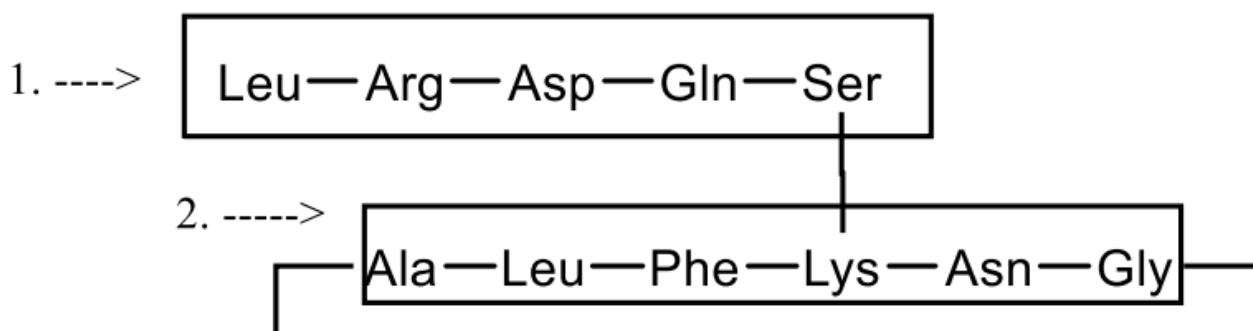


Ser та Lys зв'язані амідним зв'язком між карбоксильним кінцем серину та аміном у бічному ланцюзі Lys.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Параграф 7(b) вимагає включення послідовності, що містить принаймні чотири специфічно визначені амінокислоти, та які можуть бути представлені лінійною областю розгалуженої послідовності в переліку послідовностей. В наведеному вище прикладі пептид містить циклічну область, в якій амінокислоти з'єднуються пептидними зв'язками, та розгалужена область, з'єднана з бічним ланцюгом лізину (Lys) в циклічній області. Області цього розгалуженого пептиду, які можуть бути представлені лінійно та які містять чотири або більше специфічно визначені амінокислоти:



Стандарт VOIB ST.26 вимагає включення послідовностей 1 та 2 даного циклічного розгалуженого пептиду в перелік послідовностей, кожна з яких має власний ідентифікаційний номер.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність 1 повинна бути представлена як:

LRDQS (SEQ ID NO: 91)

Переважно, послідовність тлумачиться за допомогою ключа функції "SITE" разом з кваліфікатором "note", щоб описати, що серин у положенні 5 зв'язаний з іншою послідовністю амідним зв'язком між Ser і бічним ланцюгом Lys в іншій послідовності.

---

Послідовність 2 - циклічний пептид. В параграфі 25 вказано, що коли послідовність амінокислот є кільцевою конфігурацією і не має аміно- та карбокси-кінців, заявник повинен обрати амінокислотний залишок у положенні № 1. Відповідно, послідовність може бути представлена як:

ALFKNG (SEQ ID NO: 92)

В іншому випадку будь-яка інша амінокислота в послідовності може бути позначена як положення залишку № 1. Послідовність ALFKNG повинна бути додатково описана за допомогою ключа функції "SITE" разом з кваліфікатором "note", щоб описати, що бічний ланцюг Lys в положенні залишку №4 зв'язаний амідним зв'язком з іншою послідовністю. Цей бічний ланцюг модифікує Lys, і згідно з параграфом 30 Стандарту VOIB ST.26, модифіковану амінокислоту необхідно додатково описати в таблиці функцій. Крім того, слід надати ключ функції "REGION" та кваліфікатор "note", для зазначення, що пептид ALFKNG є кільцевим.

**Відповідні параграфи стандарту VOIB ST.26: 7(b), 25, 26, 30 та 31**

*Параграф 11(a) - Дволанцюгова нуклеотидна послідовність - повністю комплементарна*

**Приклад 11(a)-1: Дволанцюгова нуклеотидна послідовність - однакові довжини**

Патентна заявка описує таку дволанцюгову ДНК-послідовність:

3' -CCGGTTAACGCTA-5'

5' -GGCCAATTGCGAT-3'

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Кожна перелічена нуклеотидна послідовність містить більше 10 специфічно визначених нуклеотидів. Принаймні один ланцюг повинен бути включений до переліку послідовностей, тому що два ланцюги цієї дволанцюгової нуклеотидної послідовності є повністю комплементарними один одному.

**Питання 2: Чи дозволяє стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Хоча послідовність з лише одного ланцюга повинна бути включена до переліку послідовностей, послідовності з двох ланцюгів можуть бути включені, кожна із власним ідентифікаційним номером послідовності.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність дволанцюгової ДНК повинна бути представлена або як одна послідовність, або як дві окремі послідовності. Кожна послідовність, включена до переліку послідовностей, повинна бути представлена у напрямку від 5'-кінця до 3'-кінця та їй повинен бути присвоєний власний ідентифікаційний номер послідовності.

atcgcaattggcc (верхній ланцюг) (SEQ ID NO: 34)

і/або

ggccaattgcgat (нижній ланцюг) (SEQ ID NO: 35)

**Відповідні параграфи стандарту VOIB ST.26: 7(a), 11(a) та 13**

*Параграф 11(b) - Дволанцюгова нуклеотидна послідовність - не повністю комплементарна*

### Приклад 11(b)-1: Дволанцюгова нуклеотидна послідовність - різні довжини

Патентна заявка містить таку Фігуру і таке пояснення:

```
5' -tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc-3'
      |||
      gggtaactgantccgc
```

Область промотора людського гена ABC1 (верхній ланцюг), зв'язана ПНК-зондом (нижній ланцюг). Де "n" в ПНК-зонді є універсальною основою ПНК, вибраною з групи, що складається з 5-нітроіндолу та 3-нітроіндолу.

### Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

**ТАК** - промоторна область ABC1 (верхній ланцюг)

Верхній ланцюг містить більше десяти перерахованих та "специфічно визначених" нуклеотидів, та повинен бути включений до переліку послідовностей.

**ТАК** – ПНК-зонд (нижній ланцюг)

Нижній ланцюг також повинен бути включений до переліку послідовностей з власним ідентифікаційним номером послідовності, тому що два ланцюги не повністю комплементарні один одному. Окремі залишки, які містять ПНК або "пептидо-нуклеїнову кислоту" вважаються нуклеотидами відповідно до параграфу 3(g) стандарту VOIB ST.26. Тому нижній ланцюг містить більше 10 перерахованих та "специфічно визначених" нуклеотидів і повинен бути включений до переліку послідовностей.

### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Верхній ланцюг повинен бути включений до переліку послідовностей як:  
tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc (SEQ ID NO: 36)

Нижній ланцюг є пептидо-нуклеїною кислотою, і тому не має 3'-кінця та 5'-кінця. Відповідно до абзацу 11, він повинен бути включений до переліку послідовностей "у напрямку зліва направо, який імітує напрямок від 5'-кінця до 3'-кінця". Тому він повинен бути включений до переліку послідовностей як:

cgcctnagtcaatggg (SEQ ID NO: 37)

Кваліфікатор "organism" ключа функції "source" повинен мати значення "synthetic construct" та обов'язковий кваліфікатор "mol\_type" зі значенням "other DNA". Нижній ланцюг повинен бути описаний в таблиці функцій, використовуючи ключ функції "modified\_base" та обов'язковий кваліфікатор "mod\_base" з аббревіатурою "OTHER".

Кваліфікатор "note" повинен бути включений разом із повною нескороченою назвою модифікованих нуклеотидів, такою як "N-(2-аміноетил) гліцин нуклеозиди".

Залишок "n" повинен бути далі описаний в таблиці функцій з використанням ключа функції "modified\_base" та обов'язкового кваліфікатора "mod\_base" з аббревіатурою "OTHER". Кваліфікатор "note" повинен бути включений з повною нескороченою назвою модифікованого нуклеотиду: "N-(2-аміноетил) гліцин 5-нітроіндол або N-(2-аміноетил) гліцин 3-нітроіндол".

**Відповідні параграфи стандарту VOIB ST.26:** 3(g), 7(a), 11(b), 17 та 18



**Приклад 11(b)-2: Дволанцюгова нуклеотидна послідовність – відсутній сегмент спарювання основ**

Патентна заявка описує таку дволанцюгову послідовність ДНК:

```
3' -CCGGTTAGCTTATACGCTAGGGCTA-5'  
      |||||      |||||  
5' -GGCCAATATGGCTTGCGATCCCGAT-3'
```

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Кожний ланцюг перерахованої дволанцюгової нуклеотидної послідовності має більше 10 специфічно визначених нуклеотидів. Обидва ланцюги повинні бути включені до переліку послідовностей, кожен з яких має власний ідентифікаційний номер послідовності, тому що два ланцюги не повністю комплементарні один одному.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність кожного ланцюга повинна бути представлена у напрямку від 5'-кінця до 3'-кінця і мати власний ідентифікаційний номер послідовності:

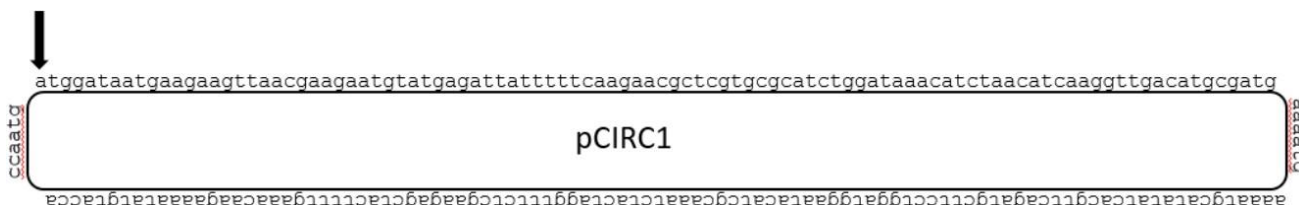
atcgggatcgcatattcgattggcc (верхній ланцюг) (SEQ ID NO: 38) і  
ggccaatatggcttgcatcccgat (нижній ланцюг) (SEQ ID NO: 39)

**Відповідні параграфи стандарту VOIB ST.26: 7(a), 11(b) та 13**

## Параграф 12 – Кільцева послідовність нуклеотидів

### Приклад 12-1: Кільцева послідовність нуклеотидів

Заявка на патент містить наступний малюнок, який розкриває послідовність ДНК плазміди pCIRC1:



#### Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

##### ТАК

Перерахована нуклеотидна послідовність містить більше 10 специфічно визначених нуклеотидів. Таким чином, послідовність має бути включена до переліку послідовностей, як того вимагає параграф (7)(a) стандарту VOIB ST.26.

#### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Згідно з параграфом 12 стандарту VOIB ST.26, коли нуклеотидні послідовності мають кільцеву конфігурацію, заявник повинен обрати нуклеотид у положенні залишку під номером 1. Для цілей цього прикладу залишок "а", позначений стрілкою на малюнку, буде використано як положення 1. Однак будь-який залишок може бути обраний як положення 1. Із зазначеним залишком за стрілкою як положення 1, послідовність має бути включена до переліку послідовностей як:

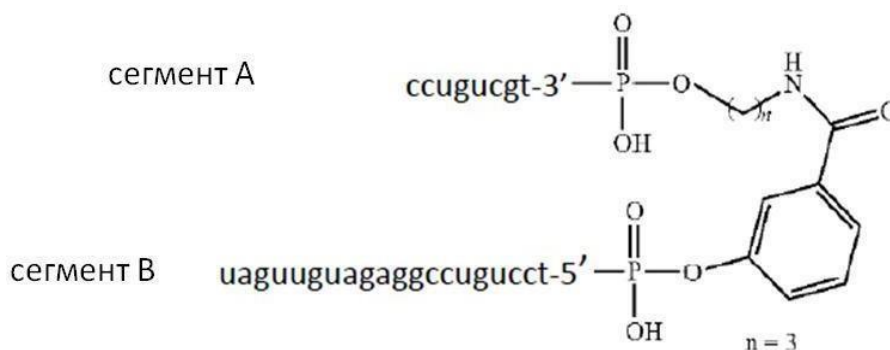
```
atggataatgaagaagttaacgaagaatgtatgagattatTTTTTcaagaacgctcgtgcgcatct
ggataaacatctaacaatcaagggttgacatgcatgaaatgaaatgcatatatcacgttcagat
gcttcctggatggaatacatcgcaaatctactagggtttctcgaagagctacttttgaacaagaa
aatatgtaccaccaatg (SEQ ID NO: 98)
```

Послідовність має бути додатково описана за допомогою ключа функції "misc\_feature" із локалізацією "212^1", яка вказує, що останній залишок у послідовності, положення 212, пов'язаний із залишком 1. Кваліфікатор "note". повинен бути включений разом із значенням, яке вказує на те, що молекула є кільцевою.

**Відповідні параграфи стандарту VOIB ST.26:** 7(a), 12 і ключ функції 5.15 Розділу 5 Додатка I

Параграф 14 – Символ "t" позначає урацил в РНК

**Приклад 14-1: Символ "t" представляє урацил у РНК**



У патентній заявці описується така сполука:

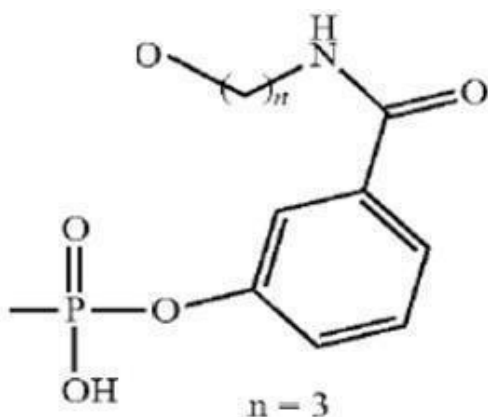
Де сегмент А і сегмент В є послідовностями РНК.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК - сегмент В**

**НІ - сегмент А**

Перелік послідовностей містить два сегменти специфічно визначених нуклеотидів, розділених наступною структурою "лінкер":



Структура лінкера не є нуклеотидом відповідно до параграфу 3(g); отже, кожний сегмент повинен розглядатися як окрема послідовність. Сегмент В містить більше 10 специфічно визначених нуклеотидів і, відповідно до параграфу 7(a) стандарту VOIB ST.26, має бути включений до переліку послідовностей. Сегмент А містить лише 8 специфічно визначених нуклеотидів і тому не має бути включений до переліку послідовностей.

**Питання 2: Чи дозволяє стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ**

Сегмент А містить менше 10 специфічно визначених нуклеотидів, і відповідно до параграфу 8 Стандарту VOIB ST.26, не повинен бути включеним до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Сегмент В - це молекула РНК; отже, елемент "INSDSeq\_moltype" повинен бути "РНК".

---

Символ "u" не повинен використовуватись для позначення урацилу в молекулі РНК у переліку послідовностей. Відповідно до параграфа 14, символ "t" слід тлумачити як урацил у РНК. Відповідно, сегмент В повинен бути включений до переліку послідовностей як:

tcctgtccggagatggtgat (SEQ ID NO: 40)

Тимін у РНК вважається модифікованим нуклеотидом, тобто модифікованим урацилом, і повинен бути представлений в послідовності як "t" і додатково описаний в таблиці функцій. Відповідно, тимін у положенні 1 повинен бути додатково описаний за допомогою ключа функції "modified\_base", кваліфікатора "mod\_base" зі значенням кваліфікатора "OTHER" і кваліфікатора "note" зі значенням кваліфікатора "thymine".

Тимін, тобто модифікований урацил, в положенні 1 також повинен бути додатково описаний у таблиці функцій з використанням ключа функції "misc\_feature" та кваліфікатора "note" зі значенням, наприклад, "5'-фосфат тимідину приєднується за допомогою лінкера 3-гідроксибензамідо-N-пропіл-3-фосфат до іншої нуклеотидної послідовності". Де доцільно, інша послідовність може бути безпосередньо вказана як значення в кваліфікаторі "note".

**Відповідні параграфи стандарту VOIB ST.26:** 3(g), 7(a), 8, 13, **14**, 19 та 54

Параграф 27 – Має використовуватись найбільш обмежувальний багатозначний символ амінокислоти

### Приклад 27-1: Скорочена формула для нуклеотидної послідовності

(GGGz)<sub>2</sub>

Де z являє собою будь-яку амінокислоту.

### Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

#### ТАК

Послідовність розкривається як формула. (GGGz)<sub>2</sub> - це просто скорочений спосіб представлення послідовності GGGzGGGz. Зазвичай, спочатку послідовність розкривається, а визначення будь-якої змінної, наприклад, "z" дається після цього.

Послідовність використовує нестандартний символ "z". Необхідно дати визначення "z" ґрунтуючись на поясненні послідовності в описі, яке визначає цей символ як будь-яку амінокислоту (*див.* Вступ до цього документа). Приклад не передбачає жодних обмежень для "z", наприклад, що вони є однаковими в кожному випадку.

Тому "z" є еквівалентом стандартному символу "X", і пептид у прикладі має вісім перерахованих амінокислот, шість з яких є специфічно визначеними залишками гліцину. Відповідно до параграфу 7(b) стандарту VOIB ST.26, послідовність має бути включена до переліку послідовностей як одна послідовність з одним ідентифікаційним номером послідовності.

Слід звернути увагу, що ця послідовність охоплюється параграфом 7(b), незважаючи на те, що перераховані та специфічно визначені залишки не є суміжними.

### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

У послідовності використовується нестандартний символ "z", який відповідно до опису є будь-якою амінокислотою. Стандартний символ, який використовується для позначення "будь-якої амінокислоти", є "X". Тому послідовність повинна бути представлена як одна розширена послідовність:

GGGXGGGX (SEQ ID NO: 41)

Відповідно до параграфу 27, "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 із додатковим описом у таблиці функцій. Тому, якщо символ "X" повинен використовуватись для представлення "будь-якої амінокислоти", тоді він має бути описаний з використанням ключа функції "VARIANT" та кваліфікатора "NOTE" зі значенням "X може бути будь-якою амінокислотою".

Де доцільно, кожен "X" має бути описаний окремо. Проте, область безперервних залишків "X" або множина залишків "X", розподілена в послідовності, може бути спільно описана за допомогою ключа функції "VARIANT" з використанням синтаксису "x.y" в якості дескриптора локалізації, де "x" та "y" – положення першого та останнього залишків "X"; та кваліфікатора "NOTE" зі значенням "X може бути будь-якою амінокислотою".

Крім того, приклад не розкриває, що "z" - однакова амінокислота в обох положеннях в розширеній послідовності. Однак, якщо "z" розкривається як одна і та ж амінокислота в обох положеннях, то ключ функції "VARIANT" та кваліфікатор "NOTE" має

супроводжуватися коментарем, в якому зазначається, що "X" у положеннях 4 та 8 може бути будь-якою амінокислотою, якщо вони однакові в обох положеннях.

**Відповідні параграфи стандарту VOIB ST.26: 3(c), 7(b) та 27**

**Приклад 27-2: Скорочена формула - менше чотирьох специфічно визначених амінокислот**

Пептид формули (Gly-Gly-Gly-z)<sub>n</sub>

Розкриття далі вказує, що z є будь-якою амінокислотою та

- (i) змінна n являє собою будь-яку довжину; або
- (ii) змінна n становить 2-100, переважно 3

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ**

Розгляд обох розкритих варіантів здійснення (i) та (ii) переліченого пептиду формули показує, що "n" може бути "будь-якої довжини"; отже, найбільш охоплюючий варіант здійснення "n" є невизначеним. Оскільки "n" є невизначеним, пептид формули не може бути розширений до певної довжини, а отже, має розглядатися нерозширена формула.

Перелічений пептид у нерозширеній формулі ("n" = 1) забезпечує три специфічно визначені амінокислоти, кожна з яких є Gly і символом "z". Зазвичай "Z" є символом "глутаміну або глутамінової кислоти"; однак, приклад визначає "z" як "будь-яку амінокислоту" (див. Вступ до цього документа). Відповідно до стандарту VOIB ST.26 - амінокислота, яка не є специфічно визначеною, позначається "X". На підставі цього аналізу перерахований пептид, тобто GGGX, не містить чотири специфічно визначені амінокислоти. Тому параграф 7(b) стандарту VOIB ST.26 не вимагає включення, незважаючи на те, що в деяких варіантах здійснення "n" також визначається як конкретні числові значення.

**Питання 2: Чи дозволяє стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Приклад надає конкретне чисельне значення для змінної "n", тобто нижня межа - 2, верхня межа - 100, та точне значення 3. Будь-яка послідовність, яка містить принаймні чотири специфічно визначені амінокислоти, може бути включена до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Перевага надається послідовності, що містить 100 копій GGGX (SEQ ID NO: 42). Наступне тлумачення повинне вказувати, що щонайбільше 98 копій GGGX можуть бути видалені. Настійно рекомендується включити додаткові конкретні варіанти здійснення, які є ключовою частиною винаходу.

Відповідно до параграфа 27, "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 із додатковим описом у таблиці функцій.

Тому, якщо символ "X" повинен використовуватись для представлення "будь-якої амінокислоти", тоді він має бути описаний з використанням ключа функції "VARIANT" та кваліфікатора "NOTE" зі значенням "X може бути будь-якою амінокислотою".

Де доцільно, кожен "X" має бути описаний окремо. Проте, область безперервних залишків "X" або множина залишків "X", розподілена в послідовності, може бути спільно описана за допомогою ключа функції "VARIANT" з використанням синтаксису "x..y" в якості дескриптора локалізації, де "x" та "y" – положення першого та останнього залишків "X"; та кваліфікатора "note" зі значенням "X може бути будь-якою

амінокислотою".

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту ВОІВ ST.26: 3 (c), 7(b), 26 та 27**



**Приклад 27-3: Скорочена формула - чотири або більше специфічно визначені амінокислоти**

Пептид формули (Gly-Gly-Gly-z)<sub>n</sub>

Де z - будь-яка амінокислота і змінна n становить 2-100, переважно 3.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?****ТАК**

Перелічений пептид формули містить три специфічно визначені амінокислоти, кожна з яких являє собою Gly, і символ "z". Зазвичай, "Z" є символом "глутаміну або глутамінової кислоти"; однак опис в цьому прикладі визначає "z" як "будь-яку амінокислоту" (див. Вступ до цього документа). Відповідно до стандарту VOIB ST.26 амінокислота, яка не є специфічно визначеною, позначена "X". На підставі цього аналізу перерахований повторюваний пептид не містить чотири специфічно визначені амінокислоти. Проте опис містить конкретне числове значення для змінної "n", тобто нижня межа - 2 і верхня межа - 100. Таким чином, у прикладі розкривається пептид, що має принаймні шість специфічно визначених амінокислот у послідовності GGGzGGGz, і відповідно до стандарту VOIB ST.26, повинен бути включений до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Оскільки "z" позначає будь-яку амінокислоту, то стандартний символ, що використовується для позначення четвертої та восьмої амінокислот - "X".

Стандарт VOIB ST.26 вимагає включення до переліку послідовностей лише однієї послідовності, яка була перерахована її залишками. Тому, принаймні одна послідовність, що містить будь-які 2, 3 або 100 копій GGGX, повинна бути включена до переліку послідовностей; однак перевага надається найбільш охоплюючій послідовності, що містить 100 копій GGGX (SEQID №: 42) (див. Вступ до цього документа). В останньому випадку додаткове тлумачення, що до 98 копій GGGX можуть бути видалені. Настійно рекомендується включити дві додаткові послідовності, що містять 2 та 3 копії GGGX, відповідно (SEQ ID NO: 44-45).

Відповідно до параграфу 27, "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 із додатковим описом у таблиці функцій. Тому, якщо символ "X" повинен використовуватись для представлення "будь-якої амінокислоти", тоді він має бути описаний з використанням ключа функції "VARIANT" та кваліфікатора "NOTE" зі значенням "X може бути будь-якою амінокислотою".

Де доцільно, кожен "X" має бути описаний окремо. Проте, область безперервних залишків "X" або множина залишків "X", розподілена в послідовності, може бути спільно описана за допомогою ключа функції "VARIANT" з використанням синтаксису "x..y" в якості дескриптора локалізації, де "x" та "y" – положення першого та останнього залишків "X"; та кваліфікатора "NOTE" зі значенням "X може бути будь-якою амінокислотою".

Крім того, приклад не описує, що змінна "z" однакова в кожному з двох випадків у розширеній послідовності. Проте, якщо "z" розкривається як однакова амінокислота у всіх локалізаціях, то ключ функції VARIANT і кваліфікатор NOTE має вказувати, що "X" у всіх положеннях може бути будь-якою амінокислотою, якщо вона є однаковою в усіх локалізаціях.

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище,

---

спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту VOIB ST.26: 3(c), 7(b), 26 та 27**

Параграф 28 - Амінокислотні послідовності, розділені внутрішніми термінаторами

**Приклад 28-1: Кодуюча нуклеотидна послідовність і кодована амінокислотна послідовність**

Патентна заявка описує наступні послідовності:

caattcaggg tggtgaat atg gcg ccc aat acg caa acc gcc tct ccc cgc  
Met Ala Pro Asn Thr Gln Thr Ala Ser Pro Arg

gcg ttg gca gat tca tta atg cag ctg gca cga cag gtt tcc cga ctg  
Ala Leu Ala Asp Ser Leu Met Gln Leu Ala Arg Gln Val Ser Arg Leu

**Білок А**

gaa agc ggg cag tga atg acc atg att acg gat tca ctg gcc gtc gtt  
Glu Ser Gly Gln Met Thr Met Ile Thr Asp Ser Leu Ala Val Val

tta caa cgt cgt gac tgg gaa aac cct ggc gtt acc caa ctt aat cgc  
Leu Gln Arg Arg Asp Trp Glu Asn Pro Gly Val Thr Gln Leu Asn Arg

**Білок В**

ctt gca gca cat tgg tgt caa aaa taa taataaccgg atgtactatt  
Leu Ala Ala His Trp Cys Gln Lys

tatccctg atg ctg cgt cgt cag gtg aat gaa gtc gct taa gcaatcaatg  
Met Leu Arg Arg Gln Val Asn Glu Val Ala

**Білок С**

tcggatgcgcgacgctt atccgaccaa catatcataa

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Заявка описує нуклеотидну послідовність, що містить термінаційні кодони, які кодують три різні амінокислотні послідовності.

Перелік нуклеотидних послідовностей містить більше 10 специфічно визначених нуклеотидів і повинен бути включений до переліку послідовностей як одна послідовність.

Що стосується кодованих амінокислотних послідовностей, то параграф 28 вимагає, щоб амінокислотні послідовності, розділені символом внутрішнього термінатора, таким як геп, повинні бути включені як окремі послідовності. Оскільки кожна послідовність: "Білок А", "Білок В" та "Білок С" містить чотири або більше специфічно визначених амінокислот, параграф 7(b) стандарту VOIB ST.26 вимагає включення кожної до переліку послідовностей і призначення власного ідентифікаційного номеру послідовності.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Нуклеотидна послідовність повинна бути включена до переліку послідовностей як:

caattcaggggtggtgaatatggcgccaataacgcaaacgcctctccccgcgcttgccgattcattaatggaagcgggca

---

gtgaatgaccatgattacggattcactggccgctgttttacaacgctgctgactgggaaaaccctggcgttacccaacttaatcgc  
cttgacgacattggtgtcaaaaataataataaccggatgtactatttatccctgatgctgctgctcaggtgaatgaagtcgct  
taagcaatcaatgctggatgcggcgacgcttatccgaccaacatatcataa (SEQ ID NO:46)

Нуклеотидна послідовність має бути додатково описана, використовуючи ключ функції "CDS" для кожного з трьох білків та елемент INSDFeature\_location слід визначити локалізацію кожної кодуєчої послідовності, включаючи стоп-кодон. Крім того, для кожного ключа функції "CDS" повинен бути включений кваліфікатор "translation" як значення кваліфікатора "amino acid sequence of the protein". Заявка не розкриває таблицю генетичних кодів, яка застосовується для трансляції (див. Таблицю 5 Розділу 9 Додатку 1). Якщо застосовується Таблиця стандартних кодів, то кваліфікатор "transl\_table" не потрібен; однак, якщо застосовується інша таблиця генетичних кодів, тоді для кваліфікатора "transl\_table" має бути зазначено відповідне значення кваліфікатора з Таблиці 5. Нарешті, кваліфікатор "protein\_id" повинен бути включений зі значенням кваліфікатора, що вказує ідентифікаційний номер послідовності кожної трансльованої амінокислотної послідовності.

Амінокислотні послідовності повинні бути включені як окремі послідовності, кожна з яких має власний ідентифікаційний номер послідовності:

MAPNTQTASPRALADSLMQLARQVSRLESGQ (SEQ ID NO: 47)

MTMITDSLAVVLQRRDWENPGVTQLNRLAAHWCCQK (SEQ ID NO: 48)

MLRRQVNEVA (SEQ ID NO: 49)

**ПРИМІТКА:** Див. "Приклад 90-1 Амінокислотна послідовність, кодована кодуєчою послідовністю з інтронами" для ілюстрації трансльованої амінокислотної послідовності, представленої як одна послідовність.

**Відповідні параграфи стандарту ВОІВ ST.26:** 7, 26, 28, 57, 89-92

*Параграф 29 - Представлення "іншої" амінокислоти*

**Приклад 29-1: Найбільш обмежувальний багатозначний символ для "іншої" амінокислоти**

Патентна заявка описує таку послідовність:

Ala-Hse-X1-X2-X3-X4-Tyr-Leu-Gly-Ser

Де X1 = Ala або Gly,

X2 = Ala або Gly,

X3 = Ala або Gly,

X4 = Ala або Gly, і

Hse = Гомосерин

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Перелічений пептид містить п'ять специфічно визначених амінокислот. Символ "X" зазвичай використовується для позначення двох амінокислот як альтернатив (*див.* Вступ до цього документа).

Оскільки існує п'ять специфічно визначених амінокислот, наприклад Ala, Tyr, Leu, Gly та Ser, параграф 7(b) стандарту VOIB ST.26, вимагає включення послідовності до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Параграф 29 вимагає, щоб будь-яка "інша" амінокислота повинна бути представлена символом "X". Послідовність у прикладі містить амінокислоту Hse в положенні 2, якої нема в Таблиці 3 Розділу 3 Додатку I. Відповідно, Hse є "іншою" амінокислотою і повинна бути представлена символом "X".

X1-X4 є варіантами положень, кожен з яких може бути A або G. Найбільш обмежувальним багатозначним символом для альтернатив A або G є "X". Тому послідовність може бути представлена як:

AXXXXXYLGS (SEQ ID NO: 50)

Включення будь-якої конкретної послідовності, необхідної для опису або формули винаходу, настійно рекомендується, як це було зазначено у вступі до цього документа.

Оскільки амінокислота Hse відсутня в Таблиці 4 Розділу 4 Додатку I, то ключ функції "SITE" та кваліфікатор "note" повинен надавати повну, нескорочену назву гомосерин, відповідно до параграфа 30 Стандарту VOIB ST.26.

Відповідно до параграфа 27, оскільки X1-X4 представляють альтернативу лише 2 амінокислот, тоді необхідний додатковий опис. Параграф 96 вказує на те, що ключ функції "VARIANT" має використовуватися з кваліфікатором "note" та значенням кваліфікатора "A або G". Відповідно до параграфа 34 стандарту VOIB ST.26, оскільки ці положення є суміжними та мають однаковий опис, вони можуть бути описані спільно, використовуючи синтаксис "3..6" як дескриптор локалізації в елементі INSDFeature\_location.

**Відповідні параграфи стандарту VOIB ST.26:** 3(a), 7(b), 25-27, **29**, 30, 34, 66, 70, 71 і 96-97

**Приклад 29-2: Використання відповідної немодифікованої амінокислоти**

Патентна заявка описує послідовність:

Ala-Hyl-Tyr-Leu-Gly-Ser-Nle-Val-Ser-5ALA

Де Hyl = гідроксилізин (посттрансляційна модифікація лізину), Nle = норлейцин і 5ALA = δ-амінолевулінова кислота

**Запитання 1: чи вимагає ST.26 включення послідовності(й)?**

**ТАК**

Перерахований пептид містить більше чотирьох конкретно визначених амінокислот; тому, послідовність повинна бути включена до переліку послідовностей.

**Запитання 3: Як послідовність (послідовності) має бути представлена в переліку послідовностей?**

Гідроксилізин у положенні 2, норлейцин у положенні 7 і δ-амінолевулінова кислота в положенні 10 — усі є «модифікованими амінокислотами». По-перше, необхідно розглянути кожну модифіковану амінокислоту та визначити, чи має вона бути представлена відповідною немодифікованою амінокислотою чи змінною "X" у послідовності. У параграфі 29 зазначено, що модифікована амінокислота «має бути представлена в послідовності як відповідна немодифікована амінокислота, коли це можливо».

Заявник вирішує, чи буде модифікована амінокислота представлена відповідним немодифікованим залишком або змінним символом "X". Однак необхідно враховувати наступне: якщо амінокислота модифікована шляхом додавання фрагмента, наприклад метилювання або ацетилювання, і основна структура відповідної немодифікованої амінокислоти загалом не змінюється, тоді рекомендується подавати немодифіковану амінокислоту. Якщо модифікована амінокислота структурно сильно відрізняється від відповідної немодифікованої амінокислоти, тоді рекомендується використовувати символ "X".

Структура гідроксилізину майже ідентична лізину, за винятком того, що третій вуглець у R-групі модифікований гідроксильною групою. Оскільки основна структура відповідного немодифікованого залишку лізину є інтактною, гідроксилізин повинен бути представлений у послідовності лізином («K»), а не «X».

Норлейцин є ізомером лейцину. R-група лейцину являє собою 4-вуглецевий ланцюг, розгалужений на другому вуглеці.

Норлейцин також має R-групу з 4 вуглецьями, але вона лінійна і не розгалужена. Таким чином, норлейцин є не просто результатом модифікації, доданої до лейцину, а зовсім іншою (хоча й спорідненою) структурою. Тому рекомендується, щоб норлейцин був представлений символом "X" у переліку послідовностей.

δ-амінолевулінова кислота не є структурно подібною до жодної з амінокислот, перелічених у Додатку 1, Таблиця 3. Тому рекомендується, щоб δ-амінолевулінова кислота була представлена символом "X" у переліку послідовностей.

Відповідно, послідовність має бути включена до переліку послідовностей як:  
AKYLG SXVSX (SEQ ID NO: 51)

Параграф 30 вимагає подальшого тлумачення кожної модифікованої амінокислоти.

Гідроксилізин є посттрансляційною модифікацією лізину. Тому його потрібно описати за допомогою ключа функції "MOD\_RES" разом із кваліфікатором "note", який описує модифікацію. Зверніть увагу, що "гідроксилізин" наведено в Додатку 1, Розділ 4, Таблиця

---

4, "Список модифікованих амінокислот". Отже, значення кваліфікатора "note" може містити аббревіатуру "Nyl" замість повної нескорченої назви «гідроксилізін».

Норлейцин не є посттрансляційно модифікованим залишком, тому його необхідно описати за допомогою ключа функції "SITE" разом із кваліфікатором "note", який описує модифікацію. Варто звернути увагу, що "норлейцин" також вказано в таблиці 4 розділу 4 Додатку 1. Тому значення кваліфікатора "note" може містити аббревіатуру "Nle" замість повної, нескорченої назви "норлейцин".

$\delta$ -амінолевулінова кислота також не є посттрансляційно модифікованим залишком, тому вона повинна бути описана за допомогою ключа функції «SITE» разом із кваліфікатором "note", який описує модифікацію.  $\delta$ -амінолевулінова кислота не вказана в таблиці 4 розділу 4 Додатку 1, тому значення кваліфікатора "note" повинно містити повну нескорочену назву модифікованого залишку, "δ-амінолевулінову кислоту".

**Відповідні параграфи стандарту VOIB ST.26:** 3(a), 3(e), 7(b), 29 і 30

## Параграф 30 - Тлумачення модифікованих амінокислот

### Приклад 30-1: Ключ функції "CARBODHYD"

Патентна заявка описує поліпептид зі специфічно модифікованою амінокислотою, що містить глікозильований бічний ланцюг, який характеризується тим, що Cys, який відповідає положенням 4 і 15 поліпептиду, утворює дисульфідний зв'язок, відповідно до наступної послідовності:

Leu-Glu-Tyr-Cys-Ley-Lys-Arg-Trp-Asn(азіалілолігосахарид))-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

### Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

#### ТАК

Перелічений пептид містить 17 специфічно визначених амінокислот. Існує 16 природних амінокислот, серед яких дев'ять (аспарагін) глікозильована. Тому, послідовність повинна бути включена до переліку послідовностей, як вимагається параграфом 7(b) стандарту VOIB ST.26.

### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Відповідно до параграфу 29 стандарту VOIB ST.26, модифікована амінокислота має бути представлена у послідовності як відповідна немодифікована амінокислота, якщо це можливо.

Тому послідовність повинна бути включена до переліку послідовностей як: LEYCLKRWNETHCAW (SEQ ID NO: 52)

Необхідний додатковий опис модифікованої амінокислоти. Ключ функції "CARBODHYD" разом з (обов'язковим) кваліфікатором "note" повинен використовуватися для позначення випадків приєднання цукрового ланцюга (asialyloligosaccharide) до аспарагіну в положенні 9. Кваліфікатор "note" описує тип зв'язування, наприклад, N-зв'язаний. Дескриптор локалізації в елементі локалізація функції являє собою номер положення залишку модифікованого аспарагіну.

Крім того, існує дисульфідний зв'язок між двома залишками Cys. Тому для опису міжланцюгового поперечного зв'язку використовується ключ функції "DISULFID". Елемент локалізації функції відображає номери положення залишків зв'язаних із залишками Cys у форматі "x..y", наприклад, "4..15". Кваліфікатор "note" не є обов'язковим.

**Відповідні параграфи стандарту VOIB ST.26:** 3(a), 7(b), 26, 29, **30** та ключ функції 7.4 Розділу 7 Додатка I



**Приклад 30-2 – Посттрансляційно модифіковані амінокислоти**

Патентна заявка на описує наступний поліпептид:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Glu-Thr-Ile-Ser-His

де Arg у положенні 7 є посттрансляційно деімінований до цитруліну.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Перерахований пептид містить 13 специфічно визначених амінокислот. Тому послідовність повинна бути включена до переліку послідовностей, як того вимагає параграф (7)(b) стандарту VOIB ST.26

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Відповідно до параграфа 29 стандарту VOIB ST.26, модифікована амінокислота повинна бути представлена в послідовності як відповідна немодифікована амінокислота, коли це можливо.

Таким чином, послідовність повинна бути включена до переліку послідовностей як:  
LEYCLKRWETISH (SEQ ID NO: 97)

де символ "R" використовується для позначення аргініну в положенні 7.

Необхідний додатковий опис, який вказує на те, що аргінін у положенні 7 може бути модифікований до цитруліну. Модифікація аргініну в цитрулін є посттрансляційною модифікацією. Тому ключ функції "MOD\_RES" слід використовувати разом з обов'язковим кваліфікатором "note", щоб вказати, що аргінін може бути деімінований до форми цитруліну. Дескриптор локалізації в елементі локалізації функції є номером позиції залишку модифікованого аргініну.

**Відповідні параграфи стандарту VOIB ST.26:** 3(a), 7(b), 30 і ключ функції 7.18 Розділу 7 Додатка I.

Параграф 36 - Послідовності, що містять області точної кількості суміжних залишків "n" або "X"

**Приклад 36-1: Послідовність з областю відомої кількості залишків "X", представлена у вигляді однієї послідовності**

LL-100-KYMR

Де "-100-" між амінокислотами лейцином та лізином позначає амінокислотну область послідовності, що містить 100 амінокислот.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Параграф 36 стандарту VOIB ST.26 вимагає включення послідовності, що містить принаймні чотири специфічно визначені амінокислоти, розділені однією або кількома областями з визначеною кількістю залишків "X".

Розкрита послідовність використовує нестандартний символ, тобто "-100-". Визначення "-100-" має впливати з пояснення послідовності в описі, яка визначає цей символ як 100 амінокислот між лейцином та лізином (*див.* Вступ до цього документа). Тому "-100" це визначена область "X" залишків. Оскільки шість зі 106 амінокислот у послідовності є специфічно визначеними, параграф 7(b) стандарту VOIB ST.26 вимагає включення послідовності до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Нестандартний символ "-100-" позначає 100 залишків "X" (оскільки будь-який символ, який використовується для позначення амінокислоти є еквівалентом лише одному залишку). Тому одна послідовність, довжиною 106 амінокислот, що містить 100 залишків "X" між LL і KYMR, повинна бути включена до переліку послідовностей (SEQ ID NO: 53).

Ця послідовність містить 100 змінних "X" між LL і KYMR. Значення ST.26 за замовчуванням для "X" без додаткових приміток є будь-яким із "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" або "V" (параграф 27). Якщо ці 100 змінних "X" визначені як щось інше, ніж це значення за замовчуванням, то для кожної змінної "X" необхідно надати правильне тлумачення.

**Відповідні параграфи стандарту VOIB ST.26: 7(b), 26, 27 та 36**

**Приклад 36-2: Послідовність з кількома областями відомої кількості або діапазоном залишків "X", представлена як одна послідовність**

Lys-z2-Lys-zm-Lys-z3-Lys-zn-Lys-z2-Lys

Де z являє собою будь-яку амінокислоту,  $m = 20$ ,  $n = 19-20$ , z2 означає, що пари лізінів розділені будь-якими двома амінокислотами, а z3 означає, що пари лізінів розділені будь-якими трьома амінокислотами.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Розкрита послідовність використовує нестандартний символ, тобто "z". Тому, слід враховувати, що опис має містити визначення "z" ; "z" в описі визначається як будь-яка амінокислота (див. Вступ до цього документа). Стандартним символом, який використовується для позначення будь-якої амінокислоти, є "X". Враховуючи наявність змінних "X", пептид містить шість перерахованих та специфічно визначених залишків лізину, які мають бути включені до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Послідовність використовує нестандартний символ "z", визначення якого повинно впливати з опису. Оскільки "z" визначається як будь-яка амінокислота, стандартним символом є "X."

Найкращим і найпоширенішим способом представлення є (див. Вступ до цього документа):

KXXKXXK (SEQ ID NO: 54)

Де  $z_n$  дорівнює 20 "X-ам", з подальшим описом, що змінна "X", яка відповідає положенню 30, може бути видалена.

В іншому випадку, або на додаток до вищезазначеного, послідовність може бути представлена як:

KXXKXXK (SEQ ID NO: 55)

Де  $z_n$  дорівнює 19 "X-ам", з подальшим описом, що змінна "X" може бути вставлена між положеннями з номерами 29 і 30.

Відповідно до параграфу 27, "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 із додатковим описом у таблиці функцій. Тому, якщо символ "X" повинен використовуватись для представлення "будь-якої амінокислоти", тоді він має бути описаний з використанням ключа функції "VARIANT" та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".

Де доцільно, кожен "X" має бути описаний окремо. Проте, область безперервних залишків "X" або множина залишків "X", розподілена в послідовності, може бути спільно описана за допомогою ключа функції "VARIANT" з використанням синтаксису "x..y" в якості дескриптора локалізації, де "x" та "y" – положення першого та останнього залишків "X"; та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".

**Відповідні параграфи стандарту VOIB ST.26: 26, 27 та 36**



---

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки.

Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту VOIB ST.26: 27 і 36**

Параграф 37 - Послідовності, що містять області невідомої кількості суміжних залишків "n" або "X"

**Приклад 37-1: Послідовність з областями невідомої кількості залишків "X" не повинна бути представлена як одна послідовність**

Gly-Gly ---- Gly-Gly-Хаа-Хаа

де символ----є невизначеним гепом у послідовності, де Хаа являє собою будь-яку амінокислоту, а гліцин і залишки Хаа з'єднуються один з одним за допомогою пептидних зв'язків.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ**

Параграф 37 стандарту VOIB ST.26 забороняє включення будь-якої послідовності, що містить невизначений геп; отже, включення всієї послідовності не вимагається.

Параграф 37 стандарту VOIB ST.26 вимагає включення будь-якої частини послідовності, суміжної з невизначеним гепом, що містить чотири або більше специфічно визначених амінокислоти. У наведеному вище прикладі включення будь-якої частини, що є суміжною з невизначеним гепом не вимагається, оскільки кожна частина містить лише дві специфічно визначені амінокислоти.

**Питання 2: Чи дозволяє стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ** - не всієї послідовності

**НІ** - не будь-якої частини послідовності

Параграф 37 стандарту VOIB ST.26 не дозволяє включення всієї послідовності.

Параграф 8 стандарту VOIB ST.26 не дозволяє включення будь-якої частини, суміжної з невизначеним гепом, оскільки кожна частина містить лише дві специфічно визначені амінокислоти.

**Відповідні параграфи стандарту VOIB ST.26: 7(b), 8, 26 та 37**

**Приклад 37-2: Послідовність з областями невідомої кількості залишків "X" не повинна бути представлена як одна послідовність**

Gly-Gly ---- Gly-Gly-Ala-Gly-Xaa-Xaa

де символ ---- є невизначеним гепом у послідовності, де Xaa є будь-якою амінокислотою, а залишки гліцину та Xaa з'єднані один з одним пептидними зв'язками.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ** - не всієї послідовності

**ТАК** - область послідовності

Параграф 37 стандарту VOIB ST.26 забороняє включення будь-якої послідовності, яка містить невизначений геп, але вимагає включення будь-якої області послідовності, суміжної з невизначеним гепом, який містить чотири або більше специфічно визначені амінокислоти.

У наведеному вище прикладі стандарт VOIB ST.26 не вимагає (та забороняє) включення як всієї послідовності, що містить невизначений геп, так і області Gly-Gly суміжної з невизначеним гепом, що містить лише дві специфічно визначені амінокислоти. Проте, стандарт VOIB ST.26 вимагає включення області Gly-Gly-Ala-Gly-Xaa-Xaa, суміжної з невизначеним гепом, оскільки він містить принаймні чотири специфічно визначені амінокислоти.

**Питання 2: Чи дозволяє стандарт VOIB ST.26 включення послідовності (-ей)?**

**НІ** - не всієї послідовності і не області Gly-Gly

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Область послідовності, суміжної з невизначеним гепом, яка містить чотири специфічно визначені амінокислоти, повинна бути представлена як:

GGAGXX (SEQ ID NO: 58)

Переважно, послідовність повинна бути розтлумачена, щоб вказати, що представлена послідовність є частиною більшої послідовності, яка містить невизначений геп, використовуючи ключ функції "SITE", локалізацію функції "1" та кваліфікатор "NOTE" зі значенням, наприклад: "Цей залишок зв'язаний N-кінцем з пептидом, що має Gly-Gly на N-кінці та геп невизначеної довжини".

Відповідно до параграфа 27, "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 із додатковим описом у таблиці функцій. Оскільки "X" представляє в SEQ ID NO: 58 "будь-яку амінокислоту", то він має бути описаний з використанням ключа функції "VARIANT" та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".

Де доцільно, кожен "X" має бути описаний окремо. Проте, область безперервних залишків "X" або множина залишків "X", розподілена в послідовності, може бути спільно описана за допомогою ключа функції "VARIANT" з використанням синтаксису "x..y" в якості дескриптора локалізації, де "x" та "y" – положення першого та останнього залишків "X"; та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".



**Відповідні параграфи стандарту ВОІВ СТ.26: 7(b), 8, 26, 27 та 37**

*Параграф 55 – Нуклеотидна послідовність, яка містить сегменти ДНК та РНК*

### **Приклад 55-1: Комбінована молекула ДНК/РНК**

Патентна заявка розкриває наступну олігонуклеотидну послідовність:

AGACCTTcggagucissiguugaacagauagusaagauguC

Де великі літери представляють залишки ДНК, а малі літери представляють залишки РНК.

### **Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

#### **ТАК**

Розкрита послідовність має більше, ніж десять перерахованих та специфічно визначених нуклеотидів; тому її потрібно включити до переліку послідовностей.

### **Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Нуклеотидна послідовність повинна бути включена до переліку послідовностей як:

agaccttcggagtctcctgttgaacagatagtcaaagtagatc (SEQ ID NO: 93)

Варто зазначити, що нуклеотиди урацилу повинні бути представлені символом "t" у переліку послідовностей.

Параграф 55 Стандарту VOIB ST.26 передбачає, що нуклеотидна послідовність, яка містить як сегменти ДНК, так і РНК, повинна бути зазначена як молекула типу "ДНК" і додатково описана за допомогою ключа функції "source" та обов'язкового кваліфікатора "organism" зі значенням "синтетичний конструкт" та обов'язкового кваліфікатора "mol\_type" зі значенням "інша ДНК". Крім того, кожен сегмент послідовності повинен бути додатково описаний за допомогою ключа функції "misc\_feature", який включає локалізацію сегмента, та кваліфікатора "note", який вказує, чи сегмент є ДНК або РНК. Розкрита послідовність містить два сегменти ДНК (положення нуклеотидів 1-7 та 43) та один сегмент РНК (положення нуклеотидів 8-42).

**Відповідні параграфи стандарту VOIB ST.26: 7, 14, 55-56 та 83**

---

*Параграф 89 - Ключ функції "CDS"*

**Приклад 89-1: Кодуюча нуклеотидна послідовність і кодована амінокислотна послідовність**

Патентна заявка описує таку нуклеотидну послідовність та її трансляцію:

```
atg acc gga aat aaa cct gaa acc gat gtt tac gaa att tta tga  
Met Thr Gly Asn Lys Pro Glu Thr Asp Val Tyr Glu Ile Leu STOP
```

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Перерахована нуклеотидна послідовність має більше десяти специфічно визначених нуклеотидів.

Перерахована амінокислотна послідовність має більше чотирьох специфічно визначених амінокислот.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Нуклеотидна послідовність повинна бути представлена як:  
atgaccggaataaacctgaaaccgatgtttacgaaattttatga (SEQ ID NO: 59)

Нуклеотидна послідовність має бути додатково описана, використовуючи ключ функції "CDS", а елемент INSDFeature\_location повинен ідентифікувати всю послідовність, включаючи стоп-кодон (наприклад, положення від 1 до 45). Крім того, кваліфікатор "translation" має бути включений зі значенням кваліфікатора "MTGNKRPETDVYEIL". Заявка не розкриває таблицю генетичних кодів, яка застосовується для трансляції (див. Таблиця 7 Розділу 9 Додатку I). Якщо застосовується Таблиця стандартних кодів, то кваліфікатор "transl\_table" є необов'язковим; однак, якщо застосовується інша таблиця генетичних кодів, то для кваліфікатора "transl\_table" потрібно вказати відповідне значення кваліфікатора з Таблиці 7. Нарешті, кваліфікатор "protein\_id" повинен бути включений зі значенням кваліфікатора, що вказує на ідентифікаційний номер трансльованої амінокислотної послідовності.

Амінокислотна послідовність повинна бути окремо представлена власним ідентифікаційним номером послідовності, використовуючи однобуквені коди наступним чином:

MTGNKRPETDVYEIL (SEQ ID NO: 60)

STOP після представленої амінокислотної послідовності не повинен включатися до амінокислотної послідовності в переліку послідовностей.

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту VOIB ST.26:** 7(a), 7(b), 26, 28, 89, 90 та 92

**Приклад 89-2: Локалізація функції виходить за межі розкритої послідовності**

Патентна заявка містить наступну фігуру, що описує часткову кодуєчу послідовність та її трансльовану амінокислотну послідовність:

```

cat cac gca gca gaa tgt gga ttt tgt cct caa caa tgg caa gtt cta      48
His His Ala Ala Glu Cys Gly Phe Cys Pro Gln Gln Trp Gln Val Leu
1           5           10           15

cgt ggg agt ctg tgc att tgt gag ggt cca gct gaa gga tgg ttc ata      96
Arg Gly Ser Leu Cys Ile Cys Glu Gly Pro Ala Glu Gly Trp Phe Ile
           20           25           30

tca aga tgt tgg tta tgg tgt ggg cct caa gtc caa ggc ttt atc ttt      144
Ser Arg Cys Trp Leu Trp Cys Gly Pro Gln Val Gln Gly Phe Ile Phe
           35           40           45

gga gaa ggc aag gaa gga ggc ggt gac aga cgg gct gaa gcg agc cct      192
Gly Glu Gly Lys Glu Gly Gly Gly Asp Arg Arg Ala Glu Ala Ser Pro
           50           55           60

cag gag ttt tgg gaa tgc act tgg                                     216
Gln Glu Phe Trp Glu Cys Thr Trp
65           70

```

Фігура 1 – часткова кодуєча послідовність гену *Homo sapiens ITCH1*, яка кодує амінокислоти 20 - 91 з білка ITCH1 довжиною 442 амінокислоти.

**Питання 1: Чи вимагає стандарт ВОІВ ST.26 включення послідовності (-ей)?****ТАК**

Заявка описує нуклеотидну послідовність та її трансльовану амінокислотну послідовність. Перелічена нуклеотидна послідовність містить більше, ніж 10 специфічно визначених нуклеотидів та повинна бути включена до переліку послідовностей.

Амінокислотна послідовність містить більше, ніж 4 специфічно визначені амінокислоти а також повинна бути включена до переліку послідовностей як окрема послідовність з власним ідентифікаційним номером.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Нуклеотидна послідовність повинна бути включена до переліку послідовностей як:

```

catcacgcagcagaatgtggattttgtcctcaacaatggcaagttctacgtgggagtctgtgcatttgtgagggtccagctgaaggatggt
tcatatcaagatgttggtatggtgtgggcctcaagtccaaggctttatctttggagaaggcaaggaaggaggcggtgacagacgggctg
aagcagaccctcaggagtttgggaatgcacttgg (SEQ ID NO: 94)

```

Нуклеотидна послідовність далі має бути описана за допомогою ключа функції "CDS". Елемент INSDFeature\_location повинен ідентифікувати локалізацію функції "CDS" у послідовності та повинен включати стоп-кодон.

На фігурі описана часткова кодуєча послідовність, яка не включає ініціувальний кодон або стоп-кодон. Однак опис послідовності вказує на те, що ініціувальний кодон знаходиться вище за нуклеотидом у положенні 1, а стоп-кодон знаходиться нижче від

останнього нуклеотиду в положенні 216.

Стандарт BOIB ST.26 вказує, що дескриптор локалізації не повинен включати нумерацію залишків за межами послідовності, в елементі INSDSeq\_sequence. Отже, у наведеному вище прикладі дескриптор локалізації ключа функції CDS не може включати номери положень поза діапазоном від 1 до 216. Локалізація стоп-кодона в елементі INSDFeature\_location повинна бути представлена символом ">", щоб вказати, що стоп-кодон розташований нижче положення 216. Аналогічно, символ "<" може бути використаний для зазначення, що розташування ініціувального кодона знаходиться вище положення 1. Таким чином, дескриптор локалізації ключа функції CDS повинен виглядати як:

<1..>216

Варто зазначити, що символи "<" та ">" є зарезервованими символами, які будуть замінені відповідно на "&lt;" та "&gt;" в екземплярі XML у переліку послідовностей.

Кваліфікатор "translation" повинен бути включений до амінокислотної послідовності білка як значення кваліфікатора. На фігурі не описано таблицю генетичного коду, яка застосовується для транслювання (див. Додаток 1, Розділ 9, Таблицю 7). Якщо застосовується таблиця Стандартного коду, то кваліфікатор "transl\_table" не потрібен; однак, якщо застосовується інша таблиця генетичного коду, то для кваліфікатора "transl\_table" має бути вказано відповідне значення кваліфікатора з таблиці 7 додатка I до Стандарту BOIB ST.26. Нарешті, кваліфікатор "protein\_id" повинен бути включений у функцію CDS зі значенням кваліфікатора, що зазначає ідентифікаційний номер трансльованої амінокислотної послідовності.

Трансльована амінокислотна послідовність повинна бути включена як окрема послідовність з власним ідентифікаційним номером:

HNAAECGFPCQQWQVLRGSLCICEGPAEGWFISRCWLWCGPQVQGFIFGEGKEGGDRRAEASPQEF  
WESTW (SEQ ID NO: 95)

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту BOIB ST.26:** 7, 41, 65, 66, 70, 71, 89 та 92

Параграф 92 - Амінокислотна послідовність, кодована кодуючою послідовністю

**Приклад 92-1: Амінокислотна послідовність, кодована кодуючою послідовністю з інтронами**

Патентна заявка містить таку Фігуру, що розкриває кодуючу послідовність та її трансляцію:

```

atg aag act ttc gca gcc ttg ctt tcc gct gtc act ctc gcg ctc tcg
Met Lys Thr Phe Ala Ala Leu Leu Ser Ala Val Thr Leu Ala Leu Ser

gtg cgc gcc cag gcg gct gtc tgg agt caa t gtaagtgccg ctgctttca
Val Arg Ala Gln Ala Ala Val Trp Ser Gln

ttgatacgag actctacgcc gagctgacgt gctaccgatat ag gt ggc ggt aca
Cys Gly Gly Thr

ccg ggt tgg acg ggc gag acc act tgc gtt gct ggt tcg gtt tgt acc
Pro Gly Trp Thr Gly Glu Thr Thr Cys Val Ala Gly Ser Val Cys Thr

tcc ttg agc tca gtgagcgact ttcaatccgt cgtcattgct cctcatgtat
Ser Leu Ser Ser

tgacgattgg ccttcatag tca tac tct caa tgc gtt ccg ggc tcc gca acg
Ser Tyr Ser Gln Cys Val Pro Gly Ser Ala Thr

tcc agc gct ccg gcg gcc ccc tca gcg aca act tca ggc ccc gca cct
Ser Ser Ala Pro Ala Ala Pro Ser Ala Thr Thr Ser Gly Pro Ala Pro

acg gac gga acg tgc tcg gcc agc ggg gca tgg ccg cca ttg acc tga
Thr Asp Gly Thr Cys Ser Ala Ser Gly Ala Trp Pro Pro Leu Thr Ter

```

Фігура 1 - нуклеотиди, виділені напівжирним шрифтом, є інтронними областями.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Заявка розкриває нуклеотидну послідовність та її амінокислотну трансляцію. Перелік нуклеотидних послідовностей містить більше 10 специфічно визначених нуклеотидів і повинен бути включений до переліку послідовностей як одна послідовність.

Нуклеотидна послідовність містить кодуючу послідовність (екзони), розділену некодуючою послідовністю (інтрони). Фігура відображає трансляцію нуклеотидної послідовності як трьох несуміжних амінокислотних послідовностей. Згідно з підписом до Фігури, виділені напівжирним шрифтом області нуклеотидів є інтронними послідовностями, які будуть з'єднані з транскриптом РНК перед трансляцією у білок. Відповідно, три амінокислотні послідовності насправді є однією, суміжною, перерахованою послідовністю, яка містить більше чотирьох специфічно визначених амінокислот, та повинна бути включена до переліку послідовностей як одна послідовність.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Нуклеотидна послідовність повинна бути включена до переліку послідовностей, як:

---

```
atgaagactttcgcagccttgctttccgctgtcactctcgcgctctcgggtgcgcgcccaggcggctgtctggagtcaatgtaagtgccg
ctgcttttcattgatacagagactctacgccgagctgacgtgctaccgtataggtggcggtacaccgggttggacggggcgagaccactt
gcgttgctggttcggtttgacctccttgagctcagtgagcgactttcaatccgtcgtcattgctcctcatgtattgacgattggccttcat
agtcatactctcaatgcgttccgggctccgcaacgtccagcgctccggcggccccctcagcgacaacttcaggccccgcacctacgg
acggaacgtgctcggccagcggggcatggccgacctga (SEQ ID NO: 75)
```

Нуклеотидна послідовність має бути додатково описана, використовуючи ключ функції "CDS", а елемент INSDFeature\_location повинен визначити локалізацію кодуєчої послідовності, у тому числі стоп-кодону, позначеного символом "Ter". Крім того, слід включити кваліфікатор "translation", з амінокислотною послідовністю білка як значенням кваліфікатора. (Зверніть увагу, що символ термінатора "Ter" в останньому положенні послідовності не повинен бути включений до амінокислотної послідовності). Заявка не розкриває таблицю генетичних кодів, яка застосовується для трансляції (див. Таблицю 5 Розділу 9 Додатку 1). Якщо застосовується таблиця "Стандартний код", то кваліфікатор "transl\_table" не є обов'язковим; проте, якщо застосовується інша таблиця генетичних кодів, то відповідне значення кваліфікатора з Таблиці 5 має бути вказане для кваліфікатора "transl\_table". Нарешті, кваліфікатор "protein\_id" повинен бути включений зі значенням кваліфікатора, що вказує ідентифікаційний номер послідовності трансльованої амінокислотної послідовності. Амінокислотна послідовність повинна бути включена як одна послідовність:

```
MKTFAALLSAVTLALSVRAQAAVWSQCGGTPGWTGETTCVAGSVCTSLSSSYSQCVPGSATSSAPAAPS
ATTSGPAPTDGTCSASGAWPPLT (SEQ ID NO: 76)
```

**Відповідні параграфи стандарту VOIB ST.26:** 7, 26, 28, 57, 67, 89-92

Параграф 93 - Первинна послідовність і її варіанти, кожний з яких перерахований за своїми залишками

### Приклад 93-1: Представлення перерахованих варіантів

Опис включає наступне вирівнювання послідовності.

|                        |                         |
|------------------------|-------------------------|
| <i>D. melanogaster</i> | ACATTGAATCTCATACCACTTT  |
| <i>D. virilis</i>      | ...-...G...C...-.G..... |
| <i>D. simulans</i>     | GT..G.CG..GT..SGT.G...  |

### Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?

#### ТАК

Загальноприйнято включення "крапок" до вирівнювання послідовності для позначення "це положення таке ж, як і те, що "розташоване над ним". Тому "крапки" у послідовностях *D. virilis* і *D. simulans* розглядаються як перераховані та специфічно визначені нуклеотиди, оскільки вони є просто коротким способом зазначення, що дане положення - такий же нуклеотид, як і *D. melanogaster*. Крім того, вирівнювання послідовностей часто відображає символ "-" для позначення відсутності залишку, щоб максимізувати вирівнювання.

Відповідно, нуклеотидні послідовності *D. melanogaster* та *D. simulans* містять двадцять два перераховані та специфічно визначені нуклеотиди, тоді як нуклеотидна послідовність *D. virilis* містять дев'ятнадцять. Таким чином, відповідно до параграфа 7(a) стандарту VOIB ST.26 кожна послідовність має бути включена до переліку послідовностей з окремими ідентифікаційними номерами послідовності.

### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Послідовність *Drosophila melanogaster* повинна бути включена до переліку послідовностей як:

acattgaatctcataccattt (SEQ ID NO: 61)

Послідовність *Drosophila virilis* повинна бути включена до переліку послідовностей як:

acatggatcccacgacttt (SEQ ID NO: 62)

Послідовність *Drosophila simulans* повинна бути включена до переліку послідовностей як:

gtatggcgtcgtatsgtagttt (SEQ ID NO: 63)

**Відповідні параграфи стандарту VOIB ST.26: 7(a), 13 та 93**



**Приклад 93-2: Представлення перерахованих варіантів**

Опис включає наступну таблицю пептиду та його функціональні варіанти. Пробіл в таблиці нижче вказує, що амінокислота у варіанті така ж, як і відповідна амінокислота в рядку "Послідовність", а "-" означає видалення відповідної амінокислоти в "Послідовності".

| Положення     | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---------------|---|---|---|---|---|---|---|---|---|
| Послідовність | A | V | L | T | Y | L | R | G | E |
| Варіант 1     |   |   |   |   |   |   |   |   | A |
| Варіант 2     |   |   | P |   |   | P |   |   |   |
| Варіант 3     |   |   | A | I | G | Y |   |   |   |
| Варіант 4     |   |   |   |   |   |   | - |   |   |

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?****ТАК**

Як зазначено, пробіл у цій таблиці вказує на те, що амінокислота в варіанті така ж, як і відповідна амінокислота в рядку "Послідовність". Тому амінокислоти варіантних послідовностей є перерахованими та специфічно визначеними.

Оскільки чотири варіантні послідовності містять більше чотирьох перелічених та специфічно визначених амінокислот, відповідно до параграфу 7(b) стандарту VOIB ST.26, кожна послідовність має бути включена до переліку послідовностей з окремим ідентифікаційним номером послідовності.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

AVLTYLRGE (SEQ ID NO: 77)

AVLTYLRGA (SEQ ID NO: 78)

AVPTYPRGE (SEQ ID NO: 79)

AVAIGYRGE (SEQ ID NO: 80)

AVLTYLGE (SEQ ID NO: 81)

**Відповідні параграфи стандарту VOIB ST.26: 7(b), 26 та 93**

### Приклад 93-3: Представлення консенсусної послідовності

Патентна заявка включає Фігуру 1 з наступним вирівнюванням декількох послідовностей.

|                                |  |
|--------------------------------|--|
| <i>Consensus</i>               | LEGnEQFINA <b>ak</b> IIRHP <b>kYnrk</b> TlnNDIM <b>LIK</b>           |
| <i>Homo sapiens</i>            | LEGnEQFINA <b>AKI</b> IIRHP <b>QYDRK</b> TLNNDIM <b>LIK</b>          |
| <i>Pongo abelii</i>            | LEGnEQFINA <b>AKI</b> IIRHP <b>QYDRK</b> TVNNDIM <b>LIK</b>          |
| <i>Papio Anubis</i>            | LEG <b>TE</b> QFINA <b>AKI</b> IIRHP <b>DYDRK</b> TLNNDI <b>LLIK</b> |
| <i>Rhinopithecus roxellana</i> | LEG <b>TE</b> QFINA <b>AKI</b> IIRHP <b>NYNRI</b> TLDNDI <b>LLIK</b> |
| <i>Pan paniscus</i>            | LEGnEQFINA <b>AKI</b> IIRHP <b>KYNRI</b> TlnNDIM <b>LIK</b>          |
| <i>Rhinopithecus bieti</i>     | LEGnEQFINA <b>TKI</b> IIRHP <b>KYNGN</b> TlnNDIM <b>LIK</b>          |
| <i>Rhinopithecus roxellana</i> | LEGnEQFINA <b>TQI</b> IIRHP <b>KYNGN</b> TlnNDIM <b>LIK</b>          |

Консенсусна послідовність включає великі літери для позначення консервативних амінокислотних залишків, а малі літери "n", "a", "k", "r", "l" і "m" являють собою переважаючі амінокислотні залишки серед вирівняної послідовності.

#### Питання 1: Чи вимагає стандарт BOIB ST.26 включення послідовності (-ей)?

##### ТАК

Кожна мала літера в консенсусній послідовності представляє собою один амінокислотний залишок. Отже, консенсусна послідовність, а також кожна інша з семи послідовностей, наведених у Фігурі 1, включає принаймні чотири специфічно визначені амінокислоти. Параграф 7(b) стандарту BOIB ST.26 вимагає включення всіх восьми послідовностей до переліку послідовностей.

#### Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?

Малі літери в консенсусній послідовності використовуються як багатозначні символи для позначення передомінантної амінокислоти серед можливих варіантів конкретного положення. Тому малі літери "n", "a", "k", "r", "l" та "m" являють собою стандартні символи, що використовуються нестандартним способом, а консенсусна послідовність повинна бути представлена за допомогою багатозначного символу замість кожної малої літери.

Має бути застосований найбільш обмежувальний багатозначний символ. Для більшості положень у консенсусній послідовності "X" є найбільш обмежувальним багатозначним символом; однак, найбільш обмежувальним багатозначним символом для "D" або "N" у положеннях 20 та 25 є "B". Консенсусна послідовність повинна бути включена до переліку послідовностей як:

LEGXEQFINAXXIIRHPXYVBXHTXBNDIXLIK (SEQ ID NO: 81)

Відповідно до параграфа 27, символ "X" буде тлумачитися як будь-який з "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" або "V",

за винятком випадків, коли він використовується з подальшим описом у таблиці функцій. Тому кожен "X" у консенсусній послідовності має бути додатково описаний у таблиці функцій, використовуючи ключ функції "VARIANT" та кваліфікатор "note", щоб вказати можливі варіанти для кожного положення.

Інші сім послідовностей повинні бути включені до переліку послідовностей як:

---

|  |      |    |     |     |
|--|------|----|-----|-----|
| LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK                 | (SEQ | ID | NO: | 83) |
| LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK                 | (SEQ | ID | NO: | 84) |
| LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK                 | (SEQ | ID | NO: | 85) |
| LEGTEQFINAAKIIRHPNYNRITLDNDILLIK                 | (SEQ | ID | NO: | 86) |
| LEGNEQFINAAKIIRHPKYNRITLNNDIMLIK (SEQ ID NO: 87) |      |    |     |     |
| LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 88) |      |    |     |     |
| LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 89) |      |    |     |     |

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту VOIB ST.26:** 7(b), 26, 27, 93 та 97

*Параграф 94 - Варіантна послідовність, розкрита як одна послідовність з перерахованими альтернативними залишками*

**Приклад 94-1: Представлення однієї послідовності з перерахованими альтернативними амінокислотами**

В патентній заявці заявлений пептид з послідовністю:

(i) Gly-Gly-Gly-[Leu або Ile]-Ala-Thr-[Ser або Thr]

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

У послідовності передбачено чотири специфічно визначені амінокислоти, а параграф 7(b) стандарту VOIB ST.26 вимагає включення послідовності до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Таблиця 3 Розділу 3 Додатку I визначає багатозначний символ "J" як ізолейцин або лейцин. Тому, бажано представляти послідовність як:

GGGJATX (SEQ ID NO: 64),

що вимагає подальшого опису в таблиці функцій, використовуючи ключ функції "VARIANT" та кваліфікатор "note" для зазначення, що "X" являє собою серин або треонін.

В іншому випадку, послідовність може бути представлена, наприклад, як: GGGLATS (SEQ ID NO: 65), що вимагає подальшого опису в таблиці функцій, використовуючи ключ функції "VARIANT" та кваліфікатор "note" для зазначення, що L можна замінити на I, а S можна замінити на T.

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту VOIB ST.26: 7(b), 8, 26, 27, 94**

**Приклад 94-2 – Представлення однієї послідовності з перерахованими альтернативними амінокислотами, які можуть бути модифікованими амінокислотами**

Патентна заявка описує поліпептид:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Xaa-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

де Xaa може бути Ile, Ala, Phe, Tyr, alle, Melle або Nle.

**Питання 1: Чи вимагає стандарт BOIB ST.26 включення послідовності (-ей)?****ТАК**

Перерахований пептид містить 16 спеціально визначених амінокислот. Таким чином, послідовність має бути включена до переліку послідовностей, відповідно до вимог параграфа (7)(b) стандарту BOIB ST.26.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Найбільш обмежувальний багатозначний символ, який може охоплювати "Ile, Ala, Phe, Tyr, alle, Melle або Nle", — це "X". Таким чином, послідовність має бути включена до переліку послідовностей як:

LEYCLKRWXETISHCAW (SEQ ID NO: 96)

Параграф 30 стандарту BOIB ST.26 вимагає, щоб «модифікована амінокислота була додатково описана в таблиці функцій».

Однак параграф 30 не вимагає використання будь-якого конкретного ключа функції для опису модифікованих амінокислот. Хоча параграф 30 описує використання ключів функцій "CARBOHYD", "LIPID", "MOD\_RES" і "SITE", ці ключі функцій більш підходять для сценаріїв, коли модифікована амінокислота не входить до списку альтернатив для певної локалізації. У цьому прикладі ключ функції "VARIANT" задовольняє вимозі параграфа 30, оскільки він дозволяє включати всі альтернативи для варіантного сайту. Отже, ключ функції "VARIANT" із кваліфікатором "note" та значенням кваліфікатора "Ile, Ala, Phe, Tyr, alle, Melle або Nle" як слід використовувати для опису сайту варіанту в положенні 9. Використання другий ключ функції, такий як "SITE" із кваліфікатором "note", може бути використаний для подальшої ідентифікації модифікованих амінокислот, знайдених у положенні 9.

**Відповідні параграфи стандарту BOIB ST.26:** 3(a), 7(b), 27, 30, 94, 96 і Додаток I, Розділ 4, Таблиця 4

*Параграф 95(a) – Варіантна послідовність, яка розкривається лише за посиланням на первинну послідовність з кількома незалежними варіантами*

### **Приклад 95(a)-1: Представлення варіантної послідовності шляхом тлумачення первинної послідовності**

Заявка містить такий опис:

"Пептидний фрагмент 1 являє собою Gly-Leu-Pro-Хаа-Arg-Ile-Cys, де Хаа може бути будь-якою амінокислотою ....

В іншому варіанті здійснення пептидний фрагмент 1 являє собою Gly-Leu-Pro-Хаа-Arg-Ile-Cys, де Хаа може бути Val, Thr або Asp ...

В іншому варіанті здійснення пептидний фрагмент 1 являє собою Gly-Leu-Pro-Хаа-Arg-Ile-Cys, де Хаа може бути Val ".

### **Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

"Пептидний фрагмент 1" у кожному з трьох описаних варіантів здійснення передбачає принаймні шість специфічно визначених амінокислот; отже, послідовність повинна бути включена до переліку послідовностей, як того вимагає параграф 7(b) стандарту VOIB ST.26.

### **Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

У цьому прикладі перерахована послідовність "Пептидний фрагмент 1" описується тричі, у трьох різних варіантах здійснення, кожен з яких має альтернативний опис Хаа. У цьому прикладі "X" є найбільш обмежувальним багатозначним символом положення Хаа.

Стандарт VOIB ST.26 вимагає включення описаної перерахованої послідовності лише один раз. У найбільш охоплюючому з трьох варіантів здійснення, Хаа є будь-якою амінокислотою (див. Вступ до цього документа). Тому послідовність, яка повинна бути включена до переліку послідовностей, являє собою:

GLPXRIC (SEQ ID NO: 66)

Відповідно до параграфа 27, "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 із додатковим описом у таблиці функцій. Тому, якщо символ "X" повинен використовуватись для представлення "будь-якої амінокислоти", тоді він має бути описаний з використанням ключа функції "VARIANT" та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".

Де доцільно, кожен "X" має бути описаний окремо. Проте, область безперервних залишків "X" або множина залишків "X", розподілена в послідовності, може бути спільно описана за допомогою ключа функції "VARIANT" з використанням синтаксису "x..y" в якості дескриптора локалізації, де "x" та "y" – положення першого та останнього залишків "X"; та кваліфікатора "note" зі значенням "X може бути будь-якою амінокислотою".

Включення будь-яких додаткових послідовностей, необхідних для опису інформації або формули винаходу, настійно рекомендується, як було зазначено у вступі до цього документа.

---

Для наведеного вище прикладу настійно рекомендується включити наступні три додаткові послідовності до переліку послідовностей, кожна з яких має власний ідентифікаційний номер послідовності:

GLPVVIC (SEQ ID NO: 67)

GLPTRIC (SEQ ID NO: 68)

GLPDRIC (SEQ ID NO: 69)

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту ВОІВ ST.26: 7(b), 26, 27 та 95(a)**

*Параграф 95(b) - Варіантна послідовність, що розкривається лише за допомогою посилання на первинну послідовність з кількома взаємозалежними варіантами*

**Приклад 95(b)-1: Представлення окремих варіантних послідовностей з кількома взаємозалежними варіантами**

Патентна заявка описує наступну консенсусну послідовність:

cgaaatgn<sub>1</sub>cccactacgaatgn<sub>2</sub>cacgaatgn<sub>3</sub>cccaca

де n<sub>1</sub>, n<sub>2</sub> та n<sub>3</sub> можуть бути a, t, g або c.

Кілька варіантних послідовностей розкриваються наступним чином:

якщо n<sub>1</sub> є a, то n<sub>2</sub> та n<sub>3</sub> є t, g або c;

якщо n<sub>1</sub> є t, то n<sub>2</sub> та n<sub>3</sub> є a, g або c;

якщо n<sub>1</sub> є g, то n<sub>2</sub> та n<sub>3</sub> є t, a або c;

якщо n<sub>1</sub> є c, то n<sub>2</sub> та n<sub>3</sub> є t, g або a.

**Питання 1: Чи вимагає стандарт VOIB ST.26 включення послідовності (-ей)?**

**ТАК**

Послідовність має більше десяти перерахованих та "специфічно визначених" нуклеотидів і відповідно до параграфа 7(a) стандарту VOIB ST.26 має бути включена до переліку послідовностей.

**Питання 3: Як слід представляти послідовність (-ності) у переліку послідовностей?**

Перерахована послідовність містить більше десяти специфічно визначених нуклеотидів та трьох "n" залишків. Стандарт VOIB ST.26 вимагає включення розкритої перерахованої послідовності і, якщо багатозначний символ є доречним, слід використовувати найбільш обмежувальний символ. У цьому прикладі n<sub>1</sub>, n<sub>2</sub> та n<sub>3</sub> можуть бути a, t, g або c, тому "n" є найбільш обмежувальним багатозначним символом. Отже, послідовність, яка повинна бути включена до переліку послідовностей:

cgaaatgncccactacgaatgncacgaatgncccaca (SEQ ID NO: 70)

В параграфі 15 стандарту VOIB ST.26 зазначено, що "символ "n" буде тлумачитися як будь-який із "a", "c", "g" або "t/u", за винятком випадків, коли він використовується з додатковим описом у таблиці функцій. Оскільки значення кожного залишку "n" у цій послідовності є еквівалентом значенню за замовчанням "a", "c", "g" або "t", додаткове тлумачення не потрібно.

Перерахована послідовність містить варіанти в трьох різних локалізаціях і виникнення варіацій є взаємозалежним. Як зазначено у вступі до цього документа, **настійно** рекомендується включення додаткових послідовностей, що представляють додаткові варіанти здійснення, які є ключовою частиною винаходу. Тому, згідно з параграфом 95(b) стандарту VOIB ST.26, додаткові варіанти здійснення повинні бути включені до переліку послідовностей як чотири окремі послідовності, кожна з яких має власний ідентифікаційний номер послідовності:



cgaatgacccactacgaatgbcacgaatgbcccaca (SEQ ID NO: 71)

cgaatgtcccactacgaatgvcacgaatgvcccaca (SEQ ID NO: 72)

cgaatggcccactacgaatghcacgaatghcccaca (SEQ ID NO: 73)

cgaatgccccactacgaatgdcacgaatgdcccaca (SEQ ID NO: 74)

(Варто зазначити, що b = t, g, або c; v = a, g або c; h = t, a або c; i d = t, g або a; див. Таблиця 1 Розділу 1 Додатку I)

Відповідно до параграфа 15 стандарту VOIB ST.26, найбільш обмежувальний символ повинен використовуватись для позначення змінних положень. Отже, n2 і n3 не повинні бути позначені як "n" у послідовності.

**ПОПЕРЕДЖЕННЯ.** Переважне представлення послідовності, зазначеної вище, спрямоване на надання переліку послідовностей на дату подання патентної заявки. Таке саме подання може бути неприйнятним щодо переліку послідовностей, який надається після дати подання патентної заявки, оскільки необхідно вирішити, чи може надана інформація бути розглянута ВІВ для додання об'єкту до первинно поданого опису.

**Відповідні параграфи стандарту VOIB ST.26:** 7(a), 15 та 95(b)

[Доповнення до Додатку VI стандарту ST.26 наведено далі]

## ДОПОВНЕННЯ

КЕРІВНИЙ ДОКУМЕНТ ЩОДО ПРЕДСТАВЛЕННЯ ПОСЛІДОВНОСТЕЙ У ФОРМАТІ XML

Доповнення доступне за адресою:

[https://www.wipo.int/standards/en/xml\\_material/st26/st26-annex-vi-appendix-guidance-document-sequences\\_v1\\_7.xml](https://www.wipo.int/standards/en/xml_material/st26/st26-annex-vi-appendix-guidance-document-sequences_v1_7.xml)

[Додаток VII наведено далі]

## ДОДАТОК VII

### РЕКОМЕНДАЦІЇ ЩОДО ПЕРЕТВОРЕННЯ ПЕРЕЛІКУ ПОСЛІДОВНОСТЕЙ ЗІ СТАНДАРТУ VOIB ST.25 НА ST.26: ПОТЕНЦІЙНО ДОДАНИЙ АБО ВИДАЛЕНИЙ ОБ'ЄКТ

*Версія 1.7*

*Редакція, схвалена Комітетом зі стандартів VOIB (КСВ)  
на його одинадцятій сесії 8 грудня 2023 року*

#### *Вступ*

Вимоги для представлення нуклеотидних та амінокислотних послідовностей відрізняються в стандартах VOIB ST.25 та ST.26. Тому, постало питання про те, чи вимагатиме Стандарт VOIB ST.26 додавання або видалення будь-якого об'єкта у переліку послідовностей, поданого як частина міжнародної заявки згідно зі стандартом VOIB ST.26, і який може не підтримуватися заявкою, відносно якої заявляється пріоритет.

#### *Сфера застосування*

Цей документ стосується обов'язкових вимог стандарту VOIB ST.26 та будь-яких потенційних наслідків цих вимог. Цей документ не розглядає кожного потенційного сценарію; якщо засоби представлення в Стандарті VOIB ST.26 інформації, що міститься в переліку послідовностей Стандарту VOIB ST.25 не є чіткими, то інформація завжди може включатися в опис заявки для уникнення видаленого об'єкта.

#### *Рекомендації стосовно потенційно доданих або видалених об'єктів*

Огляд питань, що містяться в цьому документі, демонструє, що перехід від стандарту VOIB ST.25 на стандарт VOIB ST.26 сам по собі не може призвести до доданого чи видаленого об'єкта, зокрема, якщо перелік послідовностей стандарту VOIB ST.25 повністю відповідає стандарту VOIB ST.25. Однак, існують певні сценарії, які вимагатимуть уваги заявника. Рекомендації було надано для уникнення додавання або видалення об'єктів.

#### *Сценарій 1*

Стандарт VOIB ST.25 використовує для маркування різних типів даних числові ідентифікатори, наприклад, <110> для Імені заявника. Стандарт VOIB ST.26 використовує терміни англійською мовою як назви елементів та атрибути для маркування даних.

#### Рекомендація:

Терміни, застосовані в стандарті VOIB ST.26 тільки описують тип змісту даних; тому, використання імен елементів та атрибутів відповідно до стандарту VOIB ST.26 не становить доданий об'єкт.

#### *Сценарій 2*

Стандарт VOIB ST.26 безпосередньо вимагає включення: (a) розгалужених послідовностей; (b) послідовностей з D-амінокислотами; (c) аналогів нуклеотидів; та (d) послідовностей з апуриновими сайтами. Відповідно до Стандарту VOIB ST.25, вимоги стосовно включення або

заборони таких послідовностей не є чіткими.

#### Рекомендація:

Опис, що міститься в заявці, має бути достатнім для представлення цих послідовностей в переліку послідовностей Стандарту BOIB ST.26, в той час як вони можуть бути не включені до переліку послідовностей Стандарту BOIB ST.25. Для певних типів інформації, як вимагає Стандарт BOIB ST.26, необхідно слідкувати, щоб не додати об'єкт поза описом, наприклад, див. опис нижче (в Сценарії 4) стосовно кваліфікатора "mol\_type" для нуклеотидних послідовностей.

#### *Сценарій 3*

Стандарт BOIB ST.26 виключає послідовності з менше, ніж 10 специфічно визначеними нуклеотидами (крім "n") та менше, ніж 4 специфічно визначеними амінокислотами (крім "X").

#### Рекомендація:

Виключені послідовності можуть бути включені до тіла заявки, якщо ці послідовності ще не були включені до нього раніше.

#### *Сценарій 4*

Як для нуклеотидних, так і для амінокислотних послідовностей стандарт BOIB ST.26 має обов'язковий ключ характеристики "source" з двома обов'язковими кваліфікаторами, один з яких "mol type". Стандарт BOIB ST.25 має відповідний ключ функції для нуклеотидних послідовностей (який рідко використовується) без відповідних кваліфікаторів, а для амінокислотних послідовностей не існує відповідного ключа функції.

#### Нуклеотидні послідовності

Стандарт BOIB ST.26 ключ характеристики 5.37 source; обов'язковий кваліфікатор 6.39 mol\_type (див. параграф 75 Стандарту BOIB ST.26)

| Кваліфікатор | Значення   |
|--------------|--|
| mol_type     | genomic DNA  |
|              | genomic RNA  |
|              | mRNA   |
|              | tRNA   |
|              | rRNA   |
|              | other DNA (застосовується до штучних молекул)                          |
|              | other RNA (застосовується до штучних молекул)                          |
|              | transcribed RNA  |
|              | viral cRNA   |
|              | unassigned DNA (застосовується, якщо молекула <i>in vivo</i> невідома) |
|              | unassigned RNA (застосовується, якщо молекула <i>in vivo</i> невідома) |

#### Амінокислотні послідовності

Стандарт BOIB ST.26 – ключ функції 7.30 source; обов'язковий кваліфікатор 8.1 mol\_type (див. параграф 75 Стандарту BOIB ST.26)

| Кваліфікатор | Значення |
|--------------|----------|
| mol_type     | protein  |

Рекомендація:

Єдиним проблемним питанням є значення в стандартизованій лексиці, пов'язані з кваліфікатором "mol\_type" для нуклеотидних послідовностей. Деякі значення, перелічені вище, можуть не підтримуватися в описі в достатньому обсязі. Доданого об'єкта, втім, можна уникнути шляхом використання найбільш загального значення для певної послідовності, тобто, "other DNA" та "other RNA" для синтетичної молекули та "unassigned DNA" та "unassigned RNA" для молекули в умовах *in vivo*.

Сценарій 5

Якщо послідовність включає "Хаа", тоді Стандарт BOIB ST.25 вимагає, щоб додаткова інформація, що стосується цього залишку, була включена в поле <223>, яке супроводжує поля <221> (назва функції) та <222> (локалізація функцій). Стандарт BOIB ST.25 не надає значення за замовчуванням для "Хаа" ("X" в Стандарті BOIB ST.26). Проте, Стандарт BOIB ST.26 надає таке значення за замовчуванням, і тому, додаткова інформація не завжди вимагається. Два з найчастіше вживаних тлумачень в пептидних послідовностях - це "будь-яка амінокислота" або "будь-яка природна амінокислота" для змінної "Хаа" або "X". Це пояснення можна тлумачити для включення амінокислот крім тих, що перелічені в таблицях амінокислот зі стандарту BOIB ST.25 або ST.26. Значенням за замовчуванням стандарту BOIB ST.26 для "X" без додаткового тлумачення є будь-яка з 22 окремих амінокислот, перерахованих в Додатку I (див. Розділ 3, Таблиця 3). Це значення за замовчуванням стандарту BOIB ST.26 може саме містити доданий або видалений об'єкт, і тому негативно впливає на обсяг прав патентної заявки при переході від стандарту BOIB ST.25 до стандарту BOIB ST.26.

Рекомендації:

(a) Якщо перелік послідовностей Стандарту BOIB ST.25 включає <221> назву функції, <222> локалізацію функції відповідно до "Хаа", та <223> додаткову інформацію щодо "Хаа", а <221> назва функції є також відповідним ключем функції Стандарту BOIB ST.26, наприклад, "SITE", "VARIANT", або "UNSURE", тоді має використовуватися ключ функції Стандарту BOIB ST.26. Крім того, для уникнення потенційно видаленого об'єкта, інформація в полі <223> повинна бути включена до супровідного кваліфікатора "note".

(b) Якщо перелік послідовностей Стандарту BOIB ST.25 включає <221> назву функції, <222> локалізацію функції відповідно до "Хаа", та <223> додаткову інформацію щодо "Хаа", а <221> назва функції не є ключем функції Стандарту BOIB ST.26, тоді мають використовуватися відповідно ключі функцій "SITE" або "REGION" Стандарту BOIB ST.26. Крім того, щоб уникнути потенційно видаленого об'єкта, інформація в полі <223>, так само, як і невідповідна <221> назва функції повинні бути включені до супровідного кваліфікатора "note". Наприклад, в переліку Стандарту BOIB ST.25 використовується назва функції, відсутня в Стандарті BOIB ST.25 або ST.26, <221> змінна разом із додатковою інформацією <223> "Хаа" є будь-якою амінокислотою. В цьому прикладі значення кваліфікатора "note" Стандарту BOIB ST.26 буде "Змінна - Хаа є будь-якою амінокислотою".

(c) Якщо перелік послідовностей Стандарту BOIB ST.25 не включає полів <221>, <222> або <223>, що відповідають "Хаа", або де поля <221> і <222>, що відповідають "Хаа", включені, але інформація не включена до відповідного поля

<223> (жоден зі сценаріїв не відповідає Стандарту VOIB ST.25, але, незважаючи на це, стався), будь-яка інформація, що міститься в тілі заявки для опису "Хаа", повинна бути включена до кваліфікатора "note" Стандарту VOIB ST.26 разом з відповідним ключем функції, наприклад "SITE", "REGION", або "UNSURE" і локалізацією.

### Сценарій 6

В Стандарті VOIB ST.25 урацил представлений в послідовності символом "u", а тимін представлений символом "t" без додаткового тлумачення. В Стандарті VOIB ST.26 і урацил, і тимін в послідовності представлені "t" без додаткового тлумачення; "t" представляє урацил в РНК та тимін в ДНК.

### Рекомендації:

(a) Якщо послідовність ДНК містить урацил, Стандарт VOIB ST.26 розглядає його як модифікований нуклеотид, і вимагає, щоб урацил був представлений як "t" і додатково описувався за допомогою ключа функції "modified\_base", кваліфікатора "mod\_base" з "OTHER" в якості значення кваліфікатора і кваліфікатора "note" з "uracil" в якості значення кваліфікатора. Це тлумачення Стандарту VOIB ST.26 не вважається доданим об'єктом, якщо послідовність ДНК Стандарту VOIB ST.25 містила "u".

(b) Якщо послідовність РНК містить тимін, Стандарт VOIB ST.26 розглядає його як модифікований нуклеотид, і вимагає, щоб тимін був представлений як "t" і додатково описаний з використанням ключа функції "modified\_base", кваліфікатора "mod\_base" з "OTHER" в якості значення кваліфікатора і кваліфікатора "note" з "thymine" в якості значення кваліфікатора. Це тлумачення Стандарту VOIB ST.26 не вважається доданим об'єктом, якщо послідовність РНК Стандарту VOIB ST.25 містила "t".

### Сценарій 7

В обох Стандартах VOIB ST.25 і ST.26 модифіковані нуклеотиди або амінокислоти повинні мати додатковий опис. У Стандарті VOIB ST.26 назва модифікованого нуклеотиду може бути вказана з використанням аббревіатури з Додатку I, Розділ 2, Таблиця 2, де це можливо. В іншому випадку повинна бути вказана повна нескорочена назва модифікованого нуклеотиду. Аналогічно, назва модифікованої амінокислоти може бути вказана з використанням аббревіатури з Додатку I, Розділ 4, Таблиця 4, де це можливо. В іншому випадку повинна бути вказана повна нескорочена назва модифікованої амінокислоти. В іншому випадку, якщо модифікований залишок не міститься в таблиці Стандарту VOIB ST.25, використання повної, нескороченої назви не потрібно, і не рідко, натомість використовується аббревіатура.

### Рекомендації:

(a) Якщо тільки скорочена назва, відсутня в Додатку I, Розділ 2, Таблиця 2 або Розділ 4, Таблиця 4, використовувалася як у заявці, так і в переліку послідовностей Стандарту VOIB ST.25 або для модифікованого нуклеотиду або модифікованої амінокислоти, і скорочена назва відома в даній області техніки для посилання тільки на один специфічно модифікований нуклеотид або модифіковану амінокислоту, тоді використання повної, нескороченої назви не становить доданий об'єкт.

(b) Якщо і у заявці, і в переліку послідовностей Стандарту VOIB ST.25 використовувалася тільки скорочена назва, відсутня в Додатку I, Розділ 2, Таблиця 2 або Розділ 4, Таблиця 4, для модифікованого нуклеотиду або модифікованої амінокислоти (і заявка не містить хімічної структури), а скорочена назва невідома в даній області техніки для посилання на один певний модифікований нуклеотид або модифіковану

амінокислоту, тобто аббревіатура або взагалі не відома в даній області техніки, або може представляти кілька різних модифікованих нуклеотидів або модифіковані амінокислоти, тоді дотримання Стандарту VOIB ST.26, без введення доданого об'єкта, не є можливим у цій ситуації. Звичайно, в цьому випадку пріоритетна заявка і перелік послідовностей самі по собі є нечіткими. Щоб уникнути потенційно видаленого об'єкта, скорочена назва з переліку послідовностей Стандарту VOIB ST.25 повинна бути поміщена в кваліфікатор "note" Стандарту VOIB ST.26 на додаток до значення повної нескороченої назви модифікованого нуклеотиду або модифікованої амінокислоти. Повна нескорочена назва модифікованого нуклеотиду або модифікованої амінокислоти, відповідно до вимоги в переліку послідовностей Стандарту VOIB ST.26, не матиме пріоритету перед попередньою заявкою. Щоб уникнути майбутніх протиріч, необхідно ретельно перевірити проект первинно поданого (відповідно до Стандарту VOIB ST.25) переліку послідовностей і опису заявки для включення нескороченої назви.

#### *Сценарій 8*

Стандарт VOIB ST.25 містить низку ключів функцій, які не містяться в Стандарті VOIB ST.26. Тому, заявники повинні подбати про збір інформації, що міститься в цих ключах функцій Стандарту VOIB ST.25, у спосіб, сумісний зі Стандартом VOIB ST.26, без введення доданого або видаленого об'єкта.

#### Рекомендації:

У наведеній нижче таблиці подано вказівки щодо способу, яким інформація, що міститься в попередньому ключі функції Стандарту VOIB ST.25, може бути включена відповідно до Стандарту VOIB ST.26 без введення доданого або видаленого об'єкта. Числа 1-23 є ключами функцій, що стосуються нуклеотидних послідовностей, а числа 24 - 43 є ключами функцій, що стосуються амінокислотних послідовностей.

| №. | Ключ функції <221> Стандарту VOIB ST.25                | Еквівалент в Стандарті VOIB ST.26 |   |   |
|----|--|-----------------------------------|---|---|
|    |  | Ключ функції                      | Кваліфікатор                                | Значення кваліфікатора  |
| 1  | allele   | misc_feature                      | allele                                      | значення <223>  |
| 2  | attenuator   | regulatory <sup>3</sup>           | regulatory_class <sup>1</sup>               | "attenuator"  |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 3  | CAAT_signal  | regulatory <sup>1</sup>           | regulatory_class <sup>1</sup>               | "CAAT_signal"   |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 4  | conflict   | misc_feature                      | note  | "conflict" та значення <223>  |
| 5  | enhancer   | regulatory <sup>1</sup>           | regulatory_class <sup>1</sup>               | "enhancer"  |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 6  | GC_signal  | regulatory <sup>1</sup>           | regulatory_class <sup>1</sup>               | "GC_signal"   |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 7  | LTR  | mobile_element <sup>1</sup>       | rpt_type <sup>1</sup>                       | "long_terminal_repeat"  |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 8  | misc_signal  | regulatory <sup>1</sup>           | regulatory_class <sup>1</sup>               | "other"   |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 9  | mutation   | variation                         | note  | "mutation " та значення <223>   |
| 10 | old_sequence   | misc_feature                      | note  | "old_sequence" та значення <223>  |
| 11 | polyA_signal   | regulatory <sup>1</sup>           | regulatory_class <sup>1</sup>               | "polyA_signal_sequence"   |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 12 | promoter   | regulatory <sup>1</sup>           | regulatory_class <sup>1</sup>               | "promoter"  |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 13 | RBS  | regulatory <sup>1</sup>           | regulatory_class <sup>1</sup>               | "ribosome_binding_site"   |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 14 | repeat_unit (a) якщо не використовується repeat_region | misc_feature                      | note  | "repeat_unit" та значення <223>   |
|    | repeat_unit (b) якщо використовується repeat_region    | repeat_region                     | rpt_unit_range<br>note (за наявності <223>) | 1ий залишок..останній залишок<br>значення <223>                             |
| 15 | satellite  | repeat_region                     | satellite                                   | "satellite" (або "microsatellite" або "minisatellite" – якщо підтримується) |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 16 | scRNA  | ncRNA <sup>1</sup>                | ncRNA_class <sup>1</sup>                    | "scRNA"   |
|    |  |                                   | note (за наявності <223>)                   | значення <223>  |
| 17 | snRNA  | ncRNA <sup>1</sup>                | ncRNA_class                                 | "snRNA"   |

<sup>3</sup> Стандарт VOIB ST.26 може вимагати, щоб конкретна функція Стандарту VOIB ST.25, наприклад, TATA\_signal, була замінена на ширший ключ функції/кваліфікатор/значення, наприклад, regulatory/regulatory\_class/TATA\_box



| No. | Ключ функції <221> Стандарту BOIB ST.25 | Еквівалент в Стандарті BOIB ST.26  |                               |  |
|-----|---|--|-------------------------------|--|
|     |   | Ключ функції   | Кваліфікатор                  | Значення кваліфікатора                                 |
|     |   |  | note (за наявності <223>)     | значення <223>   |
| 18  | TATA_signal                             | regulatory <sup>1</sup>  | regulatory_class <sup>1</sup> | "TATA_box" <sup>24</sup>                               |
|     |   |  | note                          | TATA_signal та (за наявності <223>):<br>значення <223> |
| 19  | terminator                              | regulatory <sup>1</sup>  | regulatory_class <sup>1</sup> | "terminator"   |
|     |   |  | note (за наявності <223>)     | значення <223>   |
| 20  | 3'clip                                  | misc_feature   | note                          | "3'clip" та значення <223>                             |
| 21  | 5'clip                                  | misc_feature   | note                          | "5'clip" та значення <223>                             |
| 22  | -10_signal                              | regulatory <sup>1</sup>  | regulatory_class <sup>1</sup> | "minus_10_signal"                                      |
|     |   |  | note (за наявності <223>)     | значення <223>   |
| 23  | -35_signal                              | regulatory <sup>1</sup>  | regulatory_class <sup>1</sup> | "minus_35_signal"                                      |
|     |   |  | note (за наявності <223>)     | <223>  |
| 24  | NON_CONS                                | Ця функція стосується гепу невідомої кількості залишків в одиничній послідовності, що заборонено як в Стандарті BOIB ST.25 (параграф 22), так і в Стандарті BOIB ST.26 (параграф 37). Отже, кожна область специфічно визначених залишків, яка охоплюється параграфом 7 Стандарту BOIB ST.26, повинна бути включена в перелік послідовностей як окрема послідовність і їй має бути присвоєно власний ідентифікаційний номер послідовності. Щоб уникнути доданого/видаленого об'єкта, кожна така послідовність повинна бути анотована, для зазначення, що вона є частиною більшої послідовності, яка містить невизначений геп. |                               |  |
|     |   | REGION   | note                          | Опис   |
|     |   | Опис - де і з чим зв'язана послідовність, наприклад, цей залишок зв'язаний N-кінцем з пептидом, що має N-кінець Gly-Gly і геп невизначеної довжини.  |                               |  |
| 25  | SIMILAR                                 | REGION   | note                          | "SIMILAR" та значення <223> за наявності               |
| 26  | THIOETH                                 | CROSSLNK   | note                          | "THIOETH" та значення <223> за наявності               |
|     |   | Для отримання додаткових відомостей про локалізацію функції, див. Стандарт BOIB ST.26 Додаток I, Перехресні посилання, Коментар до ключа функції   |                               |  |
| 27  | THIOLEST                                | CROSSLNK   | note                          | "THIOLEST" та значення <223> за наявності              |
|     |   | Для отримання додаткових відомостей про локалізацію функції, див. Стандарт BOIB ST.26 Додаток I, Перехресні посилання, Коментар до ключа функції   |                               |  |
| 28  | VARSPIC                                 | Описано нижче, в Сценарії 13   |                               |  |
| 29  | ACETYLATION                             | MOD_RES  | note                          | "ACETYLATION" та значення <223> за наявності           |

<sup>4</sup> Щоб уникнути додавання об'єкта, який може призвести до часткової втрати пріоритету, рекомендується включити більш обмежений термін "TATA\_signal" у кваліфікатор "note", як показано в таблиці вище (пункт № 18). Якщо в рідкісних випадках Заявник вважає, що використання значення "TATA\_box" для кваліфікатора "regulatory\_class" недоцільне, тоді значення: "other" може використовуватися замість "TATA\_box". У цьому випадку термін "TATA\_signal" має бути включений до кваліфікатора "note", пов'язаного з ключем функції "regulatory".

| No. | Ключ функції <221><br>Стандарту VOIB<br>ST.25 | Еквівалент в Стандарті VOIB ST.26 |              |   |
|-----|---|-----------------------------------|--------------|---|
|     |   | Ключ функції                      | Кваліфікатор | Значення кваліфікатора  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції MOD_RES Додатка I Стандарту VOIB ST.26, якщо це можливо (без доданого об'єкта) |
| 30  | AMIDATION                                     | MOD_RES                           | note         | "AMIDATION" та значення<223> за наявності   |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції MOD_RES Додатка I Стандарту VOIB ST.26, якщо це можливо (без доданого об'єкта) |
| 31  | BLOCKED                                       | MOD_RES                           | note         | "BLOCKED" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції MOD_RES Додатка I Стандарту VOIB ST.26, якщо це можливо (без доданого об'єкта) |
| 32  | FORMYLATION                                   | MOD_RES                           | note         | "FORMYLATION" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції MOD_RES Додатка I Стандарту VOIB ST.26, якщо це можливо (без доданого об'єкта) |
| 33  | GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION      | MOD_RES                           | note         | "GAMMA-CARBOXYLGLUTAMIC ACID HYDROXYLATION" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції MOD_RES Додатка I Стандарту VOIB ST.26, якщо це можливо (без доданого об'єкта) |
| 34  | METHYLATION                                   | MOD_RES                           | note         | "METHYLATION" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції MOD_RES Додатка I Стандарту VOIB ST.26, якщо це можливо (без доданого об'єкта) |
| 35  | PHOSPHORYLATION                               | MOD_RES                           | note         | "PHOSPHORYLATION" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції MOD_RES Додатка I Стандарту VOIB ST.26, якщо це можливо (без доданого об'єкта) |
| 36  | PYRROLIDONE CARBOXYLIC ACID                   | MOD_RES                           | note         | "PYRROLIDONE CARBOXYLIC ACID" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції MOD_RES Додатку I Стандарту VOIB ST.26, якщо це можливо (без доданого об'єкта) |

| No. | Ключ функції <221> Стандарту BOIB ST.25 | Еквівалент в Стандарті BOIB ST.26 |              |   |
|-----|---|-----------------------------------|--------------|---|
|     |   | Ключ функції                      | Кваліфікатор | Значення кваліфікатора  |
| 37  | SULFATATION                             | MOD_RES                           | note         | "SULFATATION" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції MOD_RES Додатка I Стандарту BOIB ST.26, якщо це можливо (без доданого об'єкта) |
| 38  | MYRISTATE                               | LIPID                             | note         | "MYRISTATE" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції LIPID Додатка I Стандарту BOIB ST.26, якщо це можливо (без доданого об'єкта)   |
| 39  | PALMITATE                               | LIPID                             | note         | "PALMITATE" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції LIPID Додатка I Стандарту BOIB ST.26, якщо це можливо (без доданого об'єкта)   |
| 40  | FARNESYL                                | LIPID                             | note         | "FARNESYL" та значення <223> за наявності   |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції LIPID Додатка I Стандарту BOIB ST.26, якщо це можливо (без доданого об'єкта)   |
| 41  | GERANYL-GERANYL                         | LIPID                             | note         | "GERANYL-GERANYL" та значення <223> за наявності  |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції LIPID Додатка I Стандарту BOIB ST.26, якщо це можливо (без доданого об'єкта)   |
| 42  | GPI-ANCHOR                              | LIPID                             | note         | "GPI-ANCHOR" та значення <223> за наявності   |
|     |   |                                   | note         | Інформація, яку вимагає коментар ключа функції LIPID Додатка I Стандарту BOIB ST.26, якщо це можливо (без доданого об'єкта)   |
| 43  | N-ACYLDIGLYCERIDE                       | LIPID                             | note         | "N-ACYL DIGLYCERIDE" та значення <223> за наявності   |
|     |   |                                   | note         | Інформація, яку вимагає Додаток I Стандарту BOIB ST.26, якщо це можливо (без доданого об'єкта)                                |

### Сценарій 9

Деякі ключі функцій, присутні як в Стандарті BOIB ST.25, так і в Стандарті BOIB ST.26, як для нуклеотидних послідовностей, так і для амінокислотних послідовностей, мають обов'язкові кваліфікатори в Стандарті BOIB ST.26, як зазначено нижче. Ключ функції нуклеотидної послідовності "modified\_base" також присутній як в Стандарті BOIB ST.25, так і в Стандарті BOIB ST.26; однак, Сценарій 7 містить відповідні рекомендації.

Стандарт BOIB ST.25 не мав жодних кваліфікаторів, але мав поле довільного тексту <223>. Коли інформація, що міститься в полі <223> Стандарту BOIB ST.25, підходить в якості значення для обов'язкового кваліфікатора Стандарту BOIB ST.26, то інформація повинна бути включена як така. Якщо поле <223> Стандарту BOIB ST.25 не було надано або містить інформацію, яка не є прийнятною в якості значення для обов'язкового кваліфікатора Стандарту BOIB ST.26, то заявники повинні подбати про збір інформації, що міститься в ключі функції Стандарту BOIB ST.25/полі <223> способом, сумісним зі Стандартом BOIB ST.26 без введення доданого або видаленого об'єкта.

Нуклеотидні послідовності<sup>5</sup>

| Ключ функції        | Обов'язковий кваліфікатор |
|---------------------|---------------------------|
| 5.12 - misc_binding | 6.3 - bound_moiety        |
| 5.30 - protein_bind | 6.3 - bound_moiety        |

#### Рекомендації:

(a) Якщо поле <223> Стандарту BOIB ST.25 відсутнє або неприйнятне, а опис заявки розкриває назву молекули/комплексу, який може зв'язуватися з локалізацією функції нуклеїнової кислоти, то ця назва повинна бути включена до кваліфікатора "bound\_moiety".

(i) Будь-яка інформація, що міститься в полі <223> Стандарту BOIB ST.25, яка є неприйнятною для включення до кваліфікатора "bound\_moiety", повинна бути вставлена у відповідний необов'язковий кваліфікатор ключа функції, наприклад, "note".

(b) Якщо поле <223> Стандарту BOIB ST.25 відсутнє або неприйнятне, і опис заявки не розкриває назви молекули/комплексу, які можуть зв'язуватися з локалізацією функції нуклеїнової кислоти, тоді ключ функції "misc\_feature" Стандарту BOIB ST.26 слід використовувати замість misc\_binding або protein\_bind, з кваліфікатором "note".

(i) Якщо поле <223> Стандарту BOIB ST.25 було відсутнє, то значення кваліфікатора "note" має бути назвою ключа функції Стандарту BOIB ST.25;

(ii) Якщо поле <223> Стандарту BOIB ST.25 містило невідповідну інформацію, то значення кваліфікатора "note" повинно бути назвою ключа функції Стандарту BOIB ST.25 і інформацією з поля <223>.

Амінокислотні послідовності<sup>6</sup>

| Ключ функції    | Обов'язковий кваліфікатор |
|-----------------|---------------------------|
| 7.2 – BINDING   | 8.2 – note                |
| 7.4 – CARBOHYD  | 8.2 – note                |
| 7.10 – DISULFID | 8.2 – note                |
| 7.11 – DNA_BIND | 8.2 – note                |
| 7.12 – DOMAIN   | 8.2 – note                |
| 7.16 – LIPID    | 8.2 – note                |
| 7.17 – METAL    | 8.2 – note                |

<sup>5</sup> Числові посилання в таблиці нижче стосуються ключа функції та номерів кваліфікаторів у Додатку I "Стандартизована лексика" Стандарту BOIB ST.26.

<sup>6</sup> Числові посилання в таблиці нижче стосуються ключа функції та номерів кваліфікаторів у Додатку I "Стандартизована лексика" Стандарту BOIB ST.26.

|                |            |
|----------------|------------|
| 7.18 – MOD_RES | 8.2 – note |
| 7.23 – NP_BIND | 8.2 – note |
| 7.29 – SITE    | 8.2 – note |
| 7.39 – ZN_FING | 8.2 – note |

#### Рекомендації:

- (a) Якщо поле <223> Стандарту VOIB ST.25 відсутнє або неприйнятне, а опис заявки розкриває конкретну інформацію, необхідну в обов'язковому кваліфікаторі, то ця інформація повинна бути включена до обов'язкового кваліфікатора "note".
- (i) Будь-яка інформація, що міститься в полі <223> Стандарту VOIB ST.25, яка не підходить для включення до обов'язкового кваліфікатора "note" (див. визначення ключа функції та коментар), повинна бути вставлена у другий кваліфікатор "note".
- (b) Якщо поле <223> Стандарту VOIB ST.25 відсутнє або неприйнятне, і опис заявки не розкриває конкретну інформацію, необхідну в обов'язковому кваліфікаторі, тоді слід використовувати натомість ключ функції "SITE" Стандарту VOIB ST.26 (для однієї амінокислоти) або "REGION" (для низки амінокислот) з кваліфікатором "note".
- (i) Якщо поле <223> Стандарту VOIB ST.25 відсутнє, то значення кваліфікатора "note" має бути ім'ям ключа функції Стандарту VOIB ST.25;
- (ii) Якщо поле <223> Стандарту VOIB ST.25 містило невідповідну інформацію, то значення кваліфікатора "note" має бути ім'ям ключа функції Стандарту VOIB ST.25 і інформацією з поля <223>.

#### *Сценарій 10*

Кожен конкретний ключ функції у Стандарті VOIB ST.25 має поле <222> для позначення локалізації функції; однак, Стандарт VOIB ST.25 не вимагає зазначення локалізації для більшості функцій, а формат інформації про локалізацію не стандартизований. Крім того, Стандарт VOIB ST.25 не має операторів локалізації, наприклад, "join". Стандарт VOIB ST.26 має стандартизовані дескриптори і оператори локалізації, і кожна функція повинна містити принаймні один дескриптор локалізації. (Функції CDS є окремим випадком і розглядаються нижче у Сценарії 11).

#### Рекомендації:

- (a) Якщо перелік послідовностей Стандарту VOIB ST.25 мав поле <222>, прямий імпорт або імпорт у формат Стандарту VOIB ST.26 не повинен викликати жодних питань щодо доданого об'єкта;
- (b) Якщо перелік послідовностей Стандарту VOIB ST.25 не мав поля <222>, але інформація про локалізацію містилася в описі заявки, то прямий імпорт або імпорт у формат Стандарту VOIB ST.26 не повинен викликати жодних питань щодо доданого об'єкта;
- (c) Якщо ні перелік послідовностей Стандарту VOIB ST.25, ні опис заявки не містили інформацію про локалізацію, то, можливо, ця функція застосовується до всієї послідовності. (Вказування локалізації, яка є меншою за всю послідовність без підтримки в описі заявки, ймовірно, становитиме доданий/видалений об'єкт). Щоб уникнути майбутніх протиріч, необхідно ретельно перевірити проект первинно поданого (відповідно до Стандарту VOIB ST.25) переліку послідовностей і опису заявки для включення інформації про локалізацію в максимально повному обсязі.

### Сценарій 11

У Стандарті BOIB ST.25 кодує послідовність, яка кодує одиничний суміжний поліпептид, але перервана однією або декількома некодуєчими послідовностями, наприклад інтронами, була позначена як безліч окремих функцій CDS, як показано нижче:

```
<220>
<221> CDS
<222> (1)..(571)
```

```
<220>
<221> CDS
<222> (639)..(859)
```

На противагу, Стандарт BOIB ST.26 має оператор локалізації "join", який вказує, що поліпептиди, закодовані зазначеними локалізаціями, з'єднані і утворюють єдиний суміжний поліпептид. (Примітка: як Стандарт BOIB ST.25, так і Стандарт BOIB ST.26 вимагають, щоб стоп-кодон був включений до локалізації функції CDS.)

#### Рекомендації:

(a) Якщо перелік послідовностей Стандарту BOIB ST.25 або опис заявки чітко вказують на те, що поліпептидні послідовності, кодовані множинними окремими функціями CDS, утворюють одиничний, суміжний поліпептид, то послідовність кодування, перервана інтроном в одиничній функції CDS, повинна бути представлена оператором локалізації "join", як показано нижче, таким чином, щоб не було введено жодного доданого об'єкта:

```
<INSDFeature_key>CDS</INSDFeature_key>
<INSDFeature_location>join(1..571,639..859)</INSDFeature_location>
```

(b) Якщо перелік послідовностей Стандарту BOIB ST.25 або опис заявки не вказують на те, що поліпептидні послідовності, кодовані двома окремими функціями CDS, утворюють одиничний суміжний поліпептид, тоді використання оператора локалізації "join", швидше за все, становитиме доданий об'єкт.

### Сценарій 12

Стандарт BOIB ST.25 вказує, що назви функцій повинні бути з таблиці 5 або 6. Проте, нормативні акти США зазначають, що ці назви функцій рекомендовані, але не обов'язкові. Отже, послідовність у переліку послідовностей Стандарту BOIB ST.25 (сумісна з нормативними актами США) може мати назву ключа функції "custom" без відповідного ключа функції у Стандарті BOIB ST.26. Можливо також, що для поля <221> відсутня назва функції або поле <221> відсутнє. Ці сценарії можуть здійснюватися схожим чином.

#### Рекомендація:

Назва ключа функції "custom" зі Стандарту BOIB ST.25 може бути представлена в переліку послідовностей Стандарту BOIB ST.26 без доданого об'єкта наступним чином:

| Тип | Ключ функції <221> Стандарту BOIB ST.25 | Потенційний еквівалент в Стандарті BOIB ST.26 |              |  |
|-----|---|---|--------------|--|
|     |   | Ключ функції                                  | Кваліфікатор | Значення кваліфікатора                                       |
| NA  | Ключ функції "custom"                   | misc_feature                                  | note         | назва ключа функції "custom" та значення <223>, за наявності |

| AA | Ключ функції "custom" | SITE або REGION | note | назва ключа функції "custom" та значення <223>, за наявності |
|----|-----------------------|-----------------|------|--|
|----|-----------------------|-----------------|------|--|

*Сценарій 13*

Стандарт BOIB ST.25 містить ключ функції "VARSP LIC", який визначається як "опис варіантів послідовностей, отриманих шляхом альтернативного сплайсингу". У Стандарті BOIB ST.26 "VARSP LIC" було замінено більш широким ключем функції VAR\_SEQ, визначеним як "опис варіантів послідовностей, отриманих шляхом альтернативного сплайсингу, використання альтернативного промотора, альтернативної ініціації і рибосомного фреймшифтингу". Отже, перелік послідовностей Стандарту BOIB ST.26 не повинен використовувати "VAR\_SEQ" як заміну "VARSP LIC" без додаткового пояснення.

Рекомендація:

У Стандарті BOIB ST.26 функція "VAR\_SEQ" повинна використовуватися з кваліфікатором "note", значення якого повинно включати пояснення більш вузької області Стандарту BOIB ST.25, наприклад, "варіант послідовностей, отриманий шляхом альтернативного сплайсингу". Будь-яка додаткова інформація, що міститься в супровідному полі <223> Стандарту BOIB ST.25, також повинна бути включена до кваліфікатора "note".

*Сценарій 14*

Якщо джерело послідовності було штучним, поле "Organism" <213> Стандарту BOIB ST.25 вимагає фрази "Artificial Sequence" ("Штучна послідовність"). У Стандарті BOIB ST.26 ключ функції "source" вимагає кваліфікатора "organism", значення якого повинно бути вказано як "synthetic construct", а не "Artificial Sequence".

Рекомендація:

Значення для кваліфікатора "organism" Стандарту BOIB ST.26 необхідно вказати як "synthetic construct". Щоб уникнути потенційно видаленого об'єкта, будь-яка пояснювальна інформація, що міститься в обов'язковому полі <223> Стандарту BOIB ST.25, повинна бути включена до кваліфікатора "note" (ключа функції "source").

*Сценарій 15*

Якщо наукова назва вихідного організму послідовності невідома, поле <213> "Organism" Стандарту BOIB ST.25 вимагає терміну "Unknown". У Стандарті BOIB ST.26 ключ функції "source" вимагає кваліфікатора "organism", значення якого повинно бути вказано як "unidentified", а не "Unknown".

Рекомендація:

Значення для кваліфікатора "organism" Стандарту BOIB ST.26 необхідно вказати як "unidentified". Щоб уникнути потенційно видаленого об'єкта, будь-яка пояснювальна інформація, що міститься в обов'язковому полі <223> Стандарту BOIB ST.25, повинна бути включена до кваліфікатора "note" (ключа функції "source").

*Сценарій 16*

Стандарт BOIB ST.25 дозволяє для нумерації амінокислот необов'язкове включення від'ємних чисел, починаючи зворотний відлік з амінокислоти, що знаходиться поруч з числом 1, для амінокислот, що передують зрілому білку, наприклад, пре-послідовності, про-послідовності, пре-про-послідовності і сигнальні послідовності. Стандарт BOIB ST.26 не дозволяє від'ємні числа у локалізації функції.

Рекомендації:

- (a) Якщо перелік послідовностей Стандарту BOIB ST.25 мав функцію або функції,

представлені в <221> і супроводжуючому полі <222>, яке містило від'ємну та/або додатну нумерацію, наприклад, "PROPER" та/або "CHAIN", тоді в переліку послідовностей Стандарту VOIB ST.26 слід використовувати відповідний ключ функції, наприклад, "PROPER" та/або "CHAIN". Кваліфікатор "note" може використовуватися з інформацією в полі <223>, якщо така є, в якості значення кваліфікатора;

(b) Якщо перелік послідовностей Стандарту VOIB ST.25 не має функції або функцій, представлених в полі <221> і супроводжуючому полі <222>, але інформація містилася в описі заявки щодо від'ємної та/або додатної нумерації, тоді в переліку послідовностей Стандарту VOIB ST.26 має використовуватися відповідний ключ функції, наприклад, "PROPER" та/або "CHAIN". В іншому випадку може використовуватися ключ функції "REGION". За наявності, кваліфікатор "note" може використовуватися з інформацією в описі заявки, як значення кваліфікатора;

(c) Якщо ні перелік послідовностей Стандарту VOIB ST.25, ні опис заявки, не містять інформації, що пояснює від'ємну та/або додатну нумерацію, то, щоб уникнути потенційно видаленого об'єкта в переліку послідовностей Стандарту VOIB ST.26, слід використовувати ключ функції "REGION", де локалізація функції охоплює ділянку з від'ємним номером у послідовності Стандарту VOIB ST.25. Крім того, відповідно до Стандарту VOIB ST.25, для зазначення, що амінокислотна послідовність була від'ємно пронумерована в переліку послідовностей заявки, відносно якої заявляється пріоритет, слід використовувати кваліфікатор "note".

#### *Сценарій 17*

Стандарт VOIB ST.25 передбачає інформацію про публікацію в полях <300> - <313>. Стандарт VOIB ST.26 не передбачає включення такої інформації.

#### Рекомендація:

Інформація, що міститься в полях <300> - <313> Стандарту VOIB ST.25 повинна бути вставлена в тіло супутньої заявки, якщо вона ще не міститься в ній.

#### *Сценарій 18*

Стандарт VOIB ST.25 не передбачає стандартизований спосіб для зазначення, що ділянка CDS нуклеотидної послідовності повинна бути трансльована з використанням таблиці генетичного коду, відмінної від стандартної таблиці генетичного коду. В порівнянні, Стандарт VOIB ST.26 має кваліфікатор "transl\_table", який може використовуватися з ключем функції "CDS" для зазначення, що ділянка має бути трансльована за допомогою альтернативної таблиці генетичного коду. Якщо кваліфікатор "transl\_table" не використовується, тоді передбачається використання стандартної таблиці генетичного коду.

#### Рекомендації:

(a) Якщо перелік послідовностей Стандарту VOIB ST.25 або опис заявки чітко вказує на те, що ділянка CDS повинна бути трансльована з використанням альтернативної таблиці генетичного коду, тоді кваліфікатор "transl\_table" повинен використовуватися з відповідним номером таблиці генетичного коду як значення кваліфікатора. Невикористання кваліфікатора "transl\_table", швидше за все, становитиме доданий об'єкт, оскільки передбачається таблиця "Standard Code" стандартного коду за замовчуванням. Нездатність включити до переліку послідовностей Стандарту VOIB ST.26 альтернативну інформацію таблиці генетичного коду з переліку послідовностей Стандарту VOIB ST.25 або з опису заявки, ймовірно,



становитиме видалений об'єкт.

(b) Якщо перелік послідовностей Стандарту VOIB ST.25 або опис заявки не вказують на те, що ділянка CDS повинна бути трансльована за допомогою альтернативної таблиці генетичного коду, то кваліфікатор "transl\_table" не повинен використовуватися або повинен використовуватися тільки зі значенням кваліфікатора "1, ", тобто таблиця стандартного коду. Використання кваліфікатора "transl\_table" з будь-яким значенням кваліфікатора, відмінним від "1", ймовірно, становитиме додані та видалені об'єкти.

### *Сценарій 19*

Стандарт VOIB ST.25 не забезпечує стандартизованого способу для позначення локалізації функції, зокрема, тієї, що міститься в сайті або ділянці, яка виходить за межі зазначеного залишку або проміжку залишків, наприклад, область CDS нуклеотидної послідовності, яка виходить за межі одного або обох кінців розкритої послідовності. В порівнянні, дескриптор локалізації об'єкта Стандарту VOIB ST.26 забезпечує стандартизований спосіб для зазначення локалізації такого сайту або ділянки за допомогою символів "<" або ">". Наприклад, локалізація функції "CDS" повинна включати в себе стоп-кодон, навіть якщо стоп-кодон не включений до самої розкритої послідовності, на кшталт шляхом зазначення локалізації 1..> 321.

#### Рекомендації:

(a) Якщо перелік послідовностей Стандарту VOIB ST.25 явно не вказує, що локалізація функції виходить за межі послідовності, але така локалізація або підтримується описом, або зрозуміла з самої послідовності, наприклад, стоп-кодон функції CDS, не міститься в послідовності, тоді символи "<" або ">" можуть бути використані в переліку послідовностей Стандарту VOIB ST.26 без додавання об'єкта.

(b) Якщо перелік послідовностей Стандарту VOIB ST.25 явно не вказує на те, що локалізація функції виходить за межі послідовності, і така локалізація не підтримується описом, і не є зрозумілою з самої послідовності, тоді відповідність Стандарту VOIB ST.26, без введення доданого об'єкта, ймовірно, неможлива в цій ситуації. У цьому випадку пріоритетна заявка і перелік послідовностей самі по собі є, можливо, неповними. У цій ситуації опису локалізації функції в переліку послідовностей Стандарту VOIB ST.26 не буде надано пріоритету перед попередньою заявкою. Щоб включити повну інформацію про функцію, необхідно ретельно перевірити проєкт первинно поданого (відповідно до Стандарту VOIB ST.25) переліку послідовностей і опису заявки.

### *Сценарій 20*

Додаток I Стандарту VOIB ST.25 вимагає, що там, де нуклеотидна послідовність містить як фрагменти ДНК, так і РНК, значення в <212> повинно бути "ДНК", а комбінована молекула ДНК/РНК далі описується в розділі функцій <220> - <223>; однак, точний характер додаткового опису не є чітким, і ця вимога не виконується. Параграф 55 Стандарту VOIB ST.26 вимагає, щоб кожен сегмент ДНК і РНК (в Стандарті VOIB ST.26 використовується "сегмент", а не "фрагмент" для внутрішньої узгодженості) комбінованої молекули ДНК/РНК був додатково описаний ключем функції "misc\_feature", який включає локалізацію сегмента, і кваліфікатор "note", який вказує, чи сегмент є ДНК або РНК.

#### Рекомендації:

(a) Якщо перелік послідовностей Стандарту VOIB ST.25 описує сегменти ДНК і

РНК в одній або більше функціях використовуючи <221> "misc\_feature", відповідні локалізації в <222> і зазначення в <223> щодо того, які сегменти були ДНК або РНК, то включення цієї інформації у форматі Стандарту VOIB ST.26, використовуючи функцію "misc\_feature" для кожного сегмента ДНК і РНК, не повинно викликати жодних питань щодо доданого об'єкта;

(b) Якщо перелік послідовностей Стандарту VOIB ST.25 описує сегменти ДНК і РНК в одній або більше функціях, використовуючи ключ функції в <221>, відмінний від "misc\_feature", відповідні локалізації в <222> і функції в <223>, що визначають, які сегменти є ДНК або РНК, тоді включення цієї інформації до формату Стандарту VOIB ST.26, з використанням "misc\_feature" для кожного сегмента ДНК і РНК і додаткового кваліфікатора "note" з вихідним ключем функції <221> в якості значення, не повинно викликати жодних питань щодо доданого або видаленого об'єкта;

(c) Якщо перелік послідовностей Стандарту VOIB ST.25 забезпечує ідентичність (ДНК або РНК) і локалізацію кожного сегмента в полі <223>, яке не пов'язане з полями <221> і <222>, наприклад, пояснення для штучної послідовності, а потім включення цієї інформації до формату Стандарту VOIB ST.26 з використанням "misc\_feature" для кожного сегмента ДНК і РНК, не повинно викликати жодних питань щодо доданого об'єкта;

(d) Якщо перелік послідовностей Стандарту VOIB ST.25 описував молекулу в функції з використанням <221> "misc\_feature" і <223>, зазначивши, що молекула є комбінованою молекулою ДНК/РНК, але не надає інформацію про локалізацію для кожного сегмента, і

(i) Якщо опис містив локалізацію кожного сегмента ДНК і РНК, то включення цієї інформації в формат Стандарту VOIB ST.26 з використанням функції "misc\_feature" для кожного сегмента ДНК і РНК не повинно викликати жодних питань щодо доданого об'єкта;

(ii) Якщо опис не містить інформації про локалізацію кожного сегмента ДНК і РНК, то сумісність зі Стандартом VOIB ST.26, без введення доданого об'єкта, може бути неможливою у цьому випадку. У цьому разі пріоритетна заявка і перелік послідовностей самі по собі є, можливо, неповними. У цій ситуації - будь-яким дескрипторам локалізації функцій в переліку послідовностей Стандарту VOIB ST.26 не буде надано пріоритету перед попередньою заявкою. Щоб включити повну інформацію про функцію, необхідно ретельно перевірити проект первинно поданого (відповідно до Стандарту VOIB ST.25) переліку послідовностей і опису заявки.

(e) Якщо в переліку послідовностей Стандарту VOIB ST.25 описана молекула в функції з використанням ключа функції в <221>, відмінної від "misc\_feature", і зазначенням <223>, що молекула є комбінованою молекулою ДНК/РНК, але не надає інформацію про локалізацію для кожного сегмента, і

(i) Якщо опис містив локалізації кожного сегмента ДНК і РНК, то включення цієї інформації до формату Стандарту VOIB ST.26 з використанням "misc\_feature" для кожного сегмента ДНК і РНК і додаткового кваліфікатора "note" з первинно поданим ключем функції <221> в якості значення, не

повинно викликати жодних питань щодо доданого чи видаленого об'єкта;

(ii) Якщо опис не містить інформації про локалізацію кожного сегмента ДНК і РНК, то сумісність зі Стандартом BOIB ST.26, без введення доданого об'єкта, може бути неможливою у цій ситуації. У цьому випадку пріоритетна заявка і перелік послідовностей самі по собі є, можливо, неповними. У цій ситуації будь-який опис локалізації функцій в переліку послідовностей Стандарту BOIB ST.26 не надаватиме пріоритету попередній заявці. Щоб включити повну інформацію про функцію, необхідно ретельно перевірити проєкт первинно поданого (відповідно до Стандарту BOIB ST.25) переліку послідовностей і опису заявки.

(f) Якщо в переліку послідовностей Стандарту BOIB ST.25 зазначено, що молекула є комбінованою молекулою ДНК/РНК в полі <223>, наприклад, пояснення для штучної послідовності "Artificial Sequence", але не надано інформацію про ключ функції або про локалізацію кожного сегмента, і

(i) Якщо в опис надано локалізації кожного сегмента ДНК і РНК, то включення цієї інформації до формату Стандарту BOIB ST.26 з використанням функції "misc\_feature" для кожного сегмента ДНК і РНК не повинно викликати жодних питань щодо доданого об'єкта;

(ii) Якщо опис не містить інформації про локацію кожного сегмента ДНК і РНК, то сумісність зі Стандартом BOIB ST.26, без введення доданого об'єкта, може бути неможливою у цій ситуації. У цьому випадку пріоритетна заявка і перелік послідовностей самі по собі є, можливо, неповними. У цій ситуації будь-який опис локалізацій функцій в переліку послідовностей Стандарту BOIB ST.26 не матиме пріоритету перед попередньою заявкою. Щоб включити повну інформацію про функцію, необхідно ретельно перевірити проєкт первинно поданого (відповідно до Стандарту BOIB ST.25) переліку послідовностей і опису заявки.

[Кінець Додатка VII і стандарту]