

СТАНДАРТ ST.25**СТАНДАРТ З ПРЕДСТАВЛЕННЯ ПЕРЕЛІКІВ ПОСЛІДОВНОСТЕЙ
НУКЛЕОТИДІВ ТА АМІНОКИСЛОТ У ПАТЕНТНИХ ЗАЯВКАХ**

*Редакція прийнята Робочою групою зі стандартів та документації ПКІТ
на її одинадцятій сесії 30 жовтня 2009 року*

Відомствам рекомендується застосовувати положення «Стандарту з представлення переліків послідовностей нуклеотидів та амінокислот у міжнародних патентних заявках, поданих відповідно до Договору про патентну кооперацію (РСТ)», як викладено в Додатку С до Адміністративних інструкцій РСТ, із внесенням необхідних змін, стосовно всіх патентних заявок, інших ніж міжнародні заявки РСТ, враховуючи те, що певні положення, характерні для процедур РСТ, і вимоги можуть застосовуватись лише до міжнародних заявок РСТ⁽¹⁾. Текст зазначеного вище стандарту, що стосується міжнародних заявок РСТ, представлено нижче.

⁽¹⁾ Якщо станом на 1 липня 2009 року національне законодавство та практика, що застосовуються відомством, не відповідають положенням параграфу 3(i) «Стандарту з представлення переліків послідовностей нуклеотидів та амінокислот у міжнародних патентних заявках, поданих відповідно до Договору про патентну кооперацію (РСТ)», згідно з яким перелік послідовностей, що міститься в поданій міжнародній заявці, «має бути представлений як окрема частина опису, розташовуватися наприкінці заявки, бажано, під назвою «Перелік послідовностей», починатися з нової сторінки, а також мати незалежну нумерацію сторінок», то Відомство може не дотримуватися зазначених положень, доки таку невідповідність не буде усунуто.

ДОДАТОК С

СТАНДАРТ З ПРЕДСТАВЛЕННЯ ПЕРЕЛІКІВ ПОСЛІДОВНОСТЕЙ НУКЛЕОТИДІВ ТА АМІНОКИСЛОТ У МІЖНАРОДНИХ ПАТЕНТНИХ ЗАЯВКАХ, ПОДАНИХ ВІДПОВІДНО ДО ДОГОВОРУ ПРО ПАТЕНТНУ КООПЕРАЦІЮ (РСТ)

ВСТУП

1. Даний стандарт розроблено з метою стандартизації представлення переліків послідовностей нуклеотидів та амінокислот у міжнародних патентних заявках. Метою цього стандарту є надання заявникові можливості скласти єдиний перелік послідовностей, прийнятний для всіх одержуючих відомств, міжнародних пошукових органів та органів міжнародної попередньої експертизи для цілей міжнародної фази, а також для всіх зазначених і вибраних відомств для цілей національної фази. Стандарт має на меті підвищити точність та якість представлення послідовностей нуклеотидів та амінокислот у міжнародних заявках, спростити представлення і розповсюдження послідовностей на користь заявників, громадськості та експертів, полегшити пошук даних стосовно послідовностей, сприяти обміну цими даними в електронній формі, а також введенню таких даних до комп'ютерних баз даних.

ВИЗНАЧЕННЯ

2. Для цього стандарту:

(i) вираз «перелік послідовностей» означає перелік послідовностей нуклеотидів та/або амінокислот, що надає детальне розкриття послідовностей нуклеотидів та/або амінокислот, а також іншу корисну інформацію;

(i-bis) вираз «перелік послідовностей, що складає частину міжнародної заявки» означає перелік послідовностей у поданій міжнародній заявці (як зазначено в параграфі 3) або перелік послідовностей, включений до міжнародної заявки шляхом коригування згідно з правилом 26, виправлення згідно з правилом 91 або внесення поправки згідно зі статтею 34(2)(b) до опису, в частині, що стосується послідовностей, які містяться у поданій міжнародній заявці (як зазначено в параграфі 3*bis*);

(i-ter) вираз «перелік послідовностей, що не складає частину міжнародної заявки» означає перелік послідовностей, що не входить до міжнародної заявки, але надається для цілей міжнародного пошуку або міжнародної попередньої експертизи (як зазначено в параграфах 4 та 4*bis*);

(ii) послідовностями, що включаються, є будь-які нерозгалужені послідовності, що складаються з чотирьох або більше амінокислот, або нерозгалужені послідовності, що складаються з десяти або більше нуклеотидів. Розгалужені послідовності, тобто послідовності з менш ніж чотирма визначеними нуклеотидами або амінокислотами, а також послідовності, що містять нуклеотиди або амінокислоти, відмінні від тих, що наведені в Таблицях 1, 2, 3 і 4 Доповнення 2, виключено з цього визначення навмисне;

(iii) термін «нуклеотиди» охоплює лише такі нуклеотиди, які можуть бути представлені з використанням символів, наведених у Таблиці 1 Доповнення 2. Модифікації, наприклад метильовані основи, можуть бути описані у спосіб, зазначений в Таблиці 2 Доповнення 2, але не повинні наводитися у послідовності нуклеотидів повністю;

(iv) термін «амінокислоти» означає L-амінокислоти, які зазвичай утворюються в білках природного походження, а також ті, що наведено в Таблиці 3 Доповнення 2. Послідовності амінокислот, які містять щонайменше одну D-амінокислоту, не охоплюються цим визначенням. Будь-яку послідовність амінокислот, що містить пост-трансляційно модифіковані амінокислоти, можна описати у вигляді послідовності амінокислот, яка спочатку була трансльована за допомогою символів, наведених у Таблиці 3 Доповнення 2, з модифікованими позиціями, наприклад гідроксилування та глікозилування, як описано в Таблиці 4 Доповнення 2, але такі модифікації не повинні представлятися у послідовності амінокислот повністю. Це визначення охоплює будь-який пептид або білок, який можна виразити за допомогою послідовності з використанням символів, наведених у Таблиці 3 Доповнення 2, у поєднанні з описом, розміщеним в іншій частині, для представлення, наприклад нестандартних зв'язків, перехресних зв'язків (наприклад дисульфідного зв'язку), кінцевих кепів, непептидних зв'язків тощо;

(v) термін «ідентифікатор послідовності» означає унікальне ціле число, що відповідає ідентифікаційному номеру (SEQ ID NO), який призначається кожній послідовності в переліку;

(vi) термін «цифровий ідентифікатор» означає тризначне число, що відтворює певний елемент даних;

(vii) термін «мовно-незалежна лексика» означає стандартизовану лексику, що використовується в переліку послідовностей і включає наукові терміни, визначені постачальниками баз даних послідовностей (у тому числі наукові назви, класифікатори та їх стандартизовані лексичні значення, символи, наведені в Таблицях 1, 2, 3 і 4 Доповнення 2, а також ключі характеристик, наведені в Таблицях 5 і 6 Доповнення 2);

(viii) термін «компетентний орган» означає міжнародний пошуковий орган, який здійснює міжнародний пошук і видає письмове рішення стосовно міжнародної заявки, або орган міжнародної попередньої експертизи, який здійснює міжнародну попередню експертизу міжнародної заявки.

ПЕРЕЛІКИ ПОСЛІДОВНОСТЕЙ

Перелік послідовностей, що складає частину міжнародної заявки

3. Перелік послідовностей, що міститься в поданій міжнародній заявці:

(i) повинен бути представлений як окрема частина опису наприкінці заявки бажано під назвою «Перелік послідовностей», починатися з нової сторінки та мати незалежну нумерацію сторінок³⁴; бажано, щоб перелік послідовностей не відтворювався у будь-якій іншій частині заявки; відповідно до параграфу 36, необов'язково описувати послідовності в іншій частині опису;

(ii) має представляти послідовності та іншу доступну інформацію в переліку послідовностей відповідно до параграфів 5-35;

(iii) у випадку, якщо міжнародна заявка була подана в електронній формі, повинен бути представлений у форматі електронного документа і подаватися засобами передачі даних відповідно до параграфа 37.

Zbis. Будь-яке коригування згідно з правилом 26, виправлення згідно з правилом 91 або поправка згідно зі статтею 34(2)(b), що вноситься до опису і стосується послідовностей у міжнародній заявці, поданій в електронній формі, має подаватися у вигляді переліку послідовностей в електронній формі та містити повний перелік з відповідними коригуваннями, виправленнями або поправками. Будь-який подібний перелік послідовностей:

(i) бажано наводити під назвою «Перелік послідовностей – Коригування», «Перелік послідовностей – Виправлення» або «Перелік послідовностей – Поправка» відповідно, а також повинен мати незалежну нумерацію сторінок⁽²⁾;

(ii) має представляти послідовності та іншу інформацію в переліку послідовностей відповідно до параграфів 5-35; якщо можливо, необхідно використовувати початкову нумерацію послідовностей в поданій міжнародній заявці (як зазначено в параграфі 5); в іншому випадку, послідовності мають бути пронумеровані відповідно до параграфа 5;

(iii) має бути представлений у форматі електронного документа і подаватися засобами передачі даних відповідно до параграфа 38.

Перелік послідовностей, що не складає частину міжнародної заявки

4. Перелік послідовностей, що подається відповідно до правила 13*ter* для цілей міжнародного пошуку або міжнародної попередньої експертизи:

(i) бажано наводити під назвою «Перелік послідовностей – Правило 13*ter*»;

(ii) має представляти послідовності та іншу доступну інформацію в переліку послідовностей відповідно до параграфів 5-35; якщо можливо, необхідно використовувати початкову нумерацію послідовностей в поданій міжнародній заявці (як зазначено в параграфі 5); в іншому випадку, послідовності мають бути пронумеровані відповідно до параграфа 5;

(iii) якщо надається на папері відповідно до правила 13*ter*.1(b), повинен мати незалежну нумерацію сторінок;

(iv) якщо надається в електронній формі, має бути представлений у форматі електронного документа і подаватися засобами передачі даних відповідно до параграфа 39;

(v) якщо надається в електронній формі разом з міжнародною заявкою, має бути ідентичним переліку послідовностей, що міститься в заявці, і супроводжуватися повідомленням наступного змісту: «інформація, записана в електронній формі і надана відповідно до правила 13*ter*, є ідентичною переліку послідовностей, що міститься в міжнародній заявці»;

⁽²⁾ **Редакційна примітка:** Незалежна нумерація сторінок не вимагається у випадку, якщо перелік послідовностей наводиться в міжнародній заявці, яка подається в електронній формі, і надається у форматі електронного документа відповідно до параграфа 40.

(vi) якщо надається після подачі міжнародної заявці, не повинен виходити за межі інформації, наведеної в міжнародній заявці, і має супроводжуватися відповідним повідомленням; будь-який подібний перелік послідовностей повинен містити лише ті послідовності, які було представлено в поданій міжнародній заявці.

4bis. Будь-яке коригування згідно з правилом 26, виправлення згідно з правилом 91 або поправка згідно зі статтею 34(2)(b), що вноситься до опису і стосується послідовностей у поданій міжнародній заявці, має супроводжуватися (для цілей міжнародного пошуку та міжнародної попередньої експертизи) переліком послідовностей в електронній формі у форматі електронного документа відповідно до параграфа 39, який повинен містити повний перелік, включаючи будь-яке коригування, виправлення або поправку, якщо цього вимагає компетентний орган, окрім випадків, якщо такий перелік в електронній формі вже доступний цьому органу в прийнятній для нього формі. Будь-який подібний перелік послідовностей в електронній формі:

(i) бажано наводити під назвою «Перелік послідовностей – Коригування – Правило 13ter», «Перелік послідовностей – Виправлення – Правило 13ter» або «Перелік послідовностей – Поправка – Правило 13ter» відповідно;

(ii) має представляти послідовності та іншу інформацію в переліку послідовностей відповідно до параграфів 5-35; якщо можливо, необхідно використовувати початкову нумерацію послідовностей в поданій міжнародній заявці (як зазначено в параграфі 5); в іншому випадку, послідовності мають бути пронумеровані відповідно до параграфа 5;

(iii) має подаватися засобами передачі даних відповідно до параграфа 39;

(iv) має бути ідентичним переліку послідовностей, поданому відповідно до параграфа *3bis* як коригування згідно з правилом 26, як виправлення згідно з правилом 91 або як поправка згідно зі статтею 34(2)(b) до опису, і повинен супроводжуватися повідомленням наступного змісту: «інформація, записана в електронній формі і надана відповідно до правила 13ter, є ідентичною переліку послідовностей, поданому як коригування згідно з правилом 26 (або як виправлення згідно з правилом 91, або як поправка згідно зі статтею 34(2)(b)) до опису у відношенні до послідовностей, наведених у поданій міжнародній заявці».

Якщо такий перелік послідовностей в електронній формі і таке повідомлення недоступні для компетентного органу, то будь-яке коригування, виправлення або поправка має братися цим органом до уваги для цілей міжнародного пошуку або попередньої експертизи в тому обсязі, який дозволить провести ефективний пошук або попередню експертизу без такого переліку послідовностей в електронній формі.

ПРЕДСТАВЛЕННЯ ПОСЛІДОВНОСТЕЙ

5. Кожній послідовності має бути призначений окремий ідентифікатор послідовності. Ідентифікатори послідовності мають починатися з цифри 1 і послідовно зростати на ціле число. Якщо для ідентифікатора відсутня послідовність, на наступному після SEQ ID NO рядку має зазначатися код «000» під цифровим ідентифікатором <400>. Цифровому ідентифікатору <160> має відповідати загальна кількість SEQ ID NO, незалежно від наявності послідовності або коду «000» після нього.

6. Послідовності, що наводяться в переліку послідовностей в описі, формулі або кресленнях заявки, мають позначатися відповідним ідентифікатором після "SEQ ID NO:".

7. Послідовності нуклеотидів та амінокислот мають представлятися принаймні за одним з наступних варіантів:

- (i) лише послідовністю нуклеотидів,
- (ii) лише послідовністю амінокислот,
- (iii) послідовністю нуклеотидів разом з відповідною послідовністю амінокислот.

Для послідовностей, що представляються за варіантом, зазначеним у пункті (iii), послідовність амінокислот у переліку послідовностей має представлятися окремо як послідовність з окремим цифровим ідентифікатором.

Послідовності нуклеотидів

Символи для використання

8. Послідовності нуклеотидів мають представлятися одним ланцюгом зліва направо у напрямку від 5'-кінця до 3'-кінця. Межі «3'» і «5'» не повинні зазначатися в послідовності.

9. Основи послідовностей нуклеотидів мають представлятися за допомогою однобуквенного коду для характеристик послідовності нуклеотидів. Необхідно використовувати лише літери нижнього регістру відповідно до переліку, наведеного в Таблиці 1 Доповнення 2.

10. Безпосередньо в послідовності, модифіковані основи, якщо вони включені до переліку, наведеного в Таблиці 2 Доповнення 2, мають представлятися як відповідні немодифіковані основи або ж за допомогою символу "n", а модифікація повинна бути описана далі у розділі характеристик переліку послідовностей за допомогою кодів, наведених у Таблиці 2 Доповнення 2. Ці коди можуть використовуватись в описі або у розділі характеристик переліку послідовностей, але не у власне послідовності (див. також параграф 32). Символ "n" є еквівалентом лише одного невідомого або модифікованого нуклеотида.

Формат для використання

11. При представленні послідовності нуклеотидів, кількість основ в одному рядку не має перевищувати 60; після кожних десяти основ має проставлятися пробіл.

12. Основи послідовності нуклеотидів (включаючи інтрони) мають представлятися групами по 10 основ у кожній, за виключенням тих, що представляються в кодуючих ділянках послідовності. Основи, що залишилися, кількість яких не перевищує 10 і які знаходяться наприкінці некодуючих ділянок послідовності, мають бути згруповані та відокремлені від сусідніх груп пробілом.

13. Основи кодуючих ділянок послідовності нуклеотидів мають наводитися триплетами (кодонами).

14. Нумерація нуклеотидів має починатися з першої основи послідовності з номера 1. Вона має бути неперервною у всій послідовності в напрямку від 5'-кінця до 3'-кінця. Нумерація проставляється у правому полі сторінки, поряд з рядком, що містить однобуквені коди основ, разом із зазначенням номеру останньої основи в цьому рядку. Зазначений вище спосіб

нумерування послідовностей нуклеотидів застосовується до послідовностей нуклеотидів, які мають кільцеву конфігурацію, при цьому перший нуклеотид у послідовності може бути позначений на розсуд заявника.

15. Послідовність нуклеотидів, що складається з однієї чи більше несуміжних ділянок довшої послідовності або з ділянок різних послідовностей, має нумеруватися як окрема послідовність з окремим ідентифікатором. Послідовність з одним або декількома проміжками має нумеруватись як сукупність окремих послідовностей з окремими ідентифікаторами, при цьому кількість окремих послідовностей має дорівнювати кількості неперервних ланцюгів даних цієї послідовності.

Послідовності амінокислот

Символи для використання

16. Амінокислоти в послідовності білків або пептидів мають наводитися зліва направо у напрямку від амінів до карбоксилів. Аміногрупи і карбоксильні групи не повинні наводитися безпосередньо у послідовності.

17. Амінокислоти мають наводитися за допомогою трибуквеного коду з великою першою літерою і повинні відповідати переліку, наведеному в Таблиці 3 Доповнення 2. Послідовність амінокислот, що містить пробіл або внутрішні термінатори (наприклад "Ter", "*" або "."), може наводитися не як одна послідовність, а як окремі послідовності амінокислот (див. параграф 22).

18. Безпосередньо в послідовності, модифіковані та нестандартні амінокислоти, якщо вони включені до переліку, наведеного у Таблиці 4 Доповнення 2, мають представлятися як відповідні немодифіковані амінокислоти або ж за допомогою символу "Хаа", а модифікація повинна бути описана далі у розділі характеристик переліку послідовностей за допомогою кодів, наведених у Таблиці 4 Доповнення 2. Ці коди можуть використовуватися в описі або у розділі характеристик переліку послідовностей, але не у власне послідовності (див. також параграф 32). Символ "Хаа" є еквівалентом лише однієї невідомої або модифікованої амінокислоти.

Формат для використання

19. При представленні послідовності білків або пептидів, кількість амінокислот в одному рядку не має перевищувати 16; після кожної амінокислоти має проставлятися пробіл.

20. Амінокислоти, що відповідають кодонам у кодуючих ділянках послідовності нуклеотидів, мають наводитися безпосередньо під відповідними кодонами. Якщо кодон розділений інтроном, то символ позначення амінокислоти має наводитися під частиною кодона, яка включає два нуклеотиди.

21. Нумерація амінокислот має починатися з першої амінокислоти в послідовності з номера 1. Крім того, амінокислоти, що передують зрілому білку, наприклад у випадку наявності пре-послідовностей, про-послідовностей, пре-про-послідовностей або сигнальних послідовностей, можуть позначатися від'ємними числами, починаючи з амінокислоти поряд з номером 1 у зворотному напрямку. Якщо в нумерації амінокислот використовуються від'ємні числа для позначення зрілого білка, то нуль (0) не використовується. Він має

наводиться під кожною п'ятою амінокислотою послідовності. Зазначений вище спосіб нумерування переліку послідовностей амінокислот застосовується до циклічних послідовностей амінокислот, при цьому перша амінокислота в послідовності позначається на розсуд заявника.

22. Послідовність амінокислот, що складається з однієї чи більше несуміжних ділянок довшої послідовності або з ділянок різних послідовностей, має нумеруватись як окрема послідовність з окремим ідентифікатором. Послідовність з одним або декількома пропусками має нумеруватись як сукупність окремих послідовностей з окремими ідентифікаторами, при цьому кількість окремих послідовностей має дорівнювати кількості послідовних ланцюгів даних цієї послідовності.

ІНША ДОСТУПНА ІНФОРМАЦІЯ В ПЕРЕЛІКУ ПОСЛІДОВНОСТЕЙ

23. Порядок елементів даних у переліках послідовностей має відповідати порядку, в якому ці елементи наведено у списку цифрових ідентифікаторів елементів даних, як зазначено в Доповненні 1.

24. Для представлення елементів даних у переліку послідовностей необхідно використовувати лише цифрові ідентифікатори елементів даних, наведені в Доповненні 1. Відповідні описи цифрових ідентифікаторів не повинні використовуватись. Інформація має наводитись безпосередньо після цифрового ідентифікатора, при цьому в переліку послідовностей мають зазначатись лише ті цифрові ідентифікатори, стосовно яких надається ця інформація. Винятками до цієї вимоги є цифрові ідентифікатори <220> і <300>, які слугують заголовками розділів «Характеристики» та «Інформація про публікацію» і відповідають даним під цифровими ідентифікаторами <221>-<223> і <301>-<313> відповідно. Якщо в переліку послідовностей під цими цифровими ідентифікаторами наводиться інформація про характеристику і публікацію, цифрові ідентифікатори <220> і <300> необхідно включати, проте залишати незаповненими. Зазвичай, порожній рядок має вставлятися між цифровими ідентифікаторами, якщо змінюється цифра у першій або другій позиції цифрового ідентифікатора. Винятком до цього загальноприйнятого правила є те, що порожній рядок не може передувати цифровому ідентифікатору <310>. Крім того, порожній рядок має передувати будь-якому цифровому ідентифікатору, що повторюється.

Обов'язкові елементи даних

25. Перелік послідовностей має містити, окрім власне послідовності нуклеотидів та/або амінокислот і безпосередньо передуючи їй, наступні елементи даних, визначені в Доповненні 1 (обов'язкові елементи даних):

<110>	Ім'я заявника
<120>	Назва винаходу
<160>	Кількість SEQ ID NO
<210>	SEQ ID NO: x
<211>	Довжина

<212>	Тип
<213>	Організм
<400>	Послідовність

Якщо ім'я заявника (цифровий ідентифікатор <110>) наводиться не латинськими літерами, його також необхідно точно транслітерувати або перекласти англійською мовою.

Елементи даних, окрім тих, що зазначаються під цифровими ідентифікаторами <110>, <120> і <160>, мають повторюватись для кожної послідовності в переліку. У випадку відсутності послідовності для ідентифікатора, обов'язковими є цифрові ідентифікатори <210> і <400> (див. параграф 5 і рядок "SEQ ID NO: 4" у зразку, наведеному в Доповненні 3).

26. Якщо перелік послідовностей надається раніше за присвоєння заявці номера, окрім елементів даних, зазначених у параграфі 25, до переліку послідовностей має бути включений такий елемент даних:

<130>	Посилання на дос'є заявки
-------	---------------------------

27. Якщо перелік послідовностей надається пізніше присвоєння заявці номера, окрім елементів даних, зазначених у параграфі 25, до переліку послідовностей мають бути включені такі елементи даних:

<140>	Діюча патентна заявка
<141>	Діюча дата подання

28. Якщо перелік послідовностей подається у відношенні заявки, що заявляє пріоритет більш ранньої заявки, окрім елементів даних, зазначених у параграфі 25, до переліку послідовностей мають бути включені такі елементи даних:

<150>	Більш рання патентна заявка
<151>	Більш рання дата подання

29. Якщо в послідовності використовується символ "n", символ "Хаа", модифікована основа або модифікована/нестандартна L-амінокислота, то наступні елементи даних є обов'язковими для включення до переліку:

<220>	Характеристика
<221>	Назва/ключ
<222>	Локалізація
<223>	Інша інформація

30. Якщо організмом (цифровий ідентифікатор <213>) є штучна послідовність (Artificial sequence) або він невідомий (Unknown), то наступні елементи даних є обов'язковими для включення до переліку:

<220>	Характеристика
<223>	Інша інформація

Необов'язкові елементи даних

31. Всі елементи даних, зазначені в Доповненні 1 і не зазначені у параграфах 25-30, є необов'язковими (необов'язкові елементи даних).

Представлення характеристик

32. При представленні характеристик послідовностей (цифровий ідентифікатор <220>), вони мають бути описані «ключами характеристик», зазначеними в Таблицях 5 і 6 Доповнення 2⁽³⁾.

Довільний текст

33. Довільний текст це формулювання з описом характеристик послідовності під цифровим ідентифікатором <223> (інша інформація), в якому не використовується мовно-незалежна лексика, визначена у параграфі 2(vii).

34. Довільний текст має обмежуватись декількома короткими термінами, важливими для розуміння послідовності. Якщо довільний текст наведено англійською мовою, він не повинен перевищувати 4 рядки по 65 символів максимум у кожному рядку для кожного елементу даних. Будь-яка інша додаткова інформація повинна наводитися в описі тією ж мовою, що і текст опису.

35. Будь-який довільний текст бажано наводити англійською мовою.

ПОВТОРЕННЯ ДОВІЛЬНОГО ТЕКСТУ В ОСНОВНІЙ ЧАСТИНІ ОПИСУ

36. Якщо перелік послідовностей, що складає частину міжнародної заявки, містить довільний текст, то цей текст має бути повторно наведений в описі тією ж мовою. Довільний текст, що наводиться мовою, якою викладено текст опису, рекомендовано наводити у певній частині опису під назвою «Довільний текст переліку послідовностей».

ПЕРЕЛІКИ ПОСЛІДОВНОСТЕЙ В ЕЛЕКТРОННІЙ ФОРМІ

37. Будь-який перелік послідовностей, зазначений у параграфі 3, що міститься в міжнародній заявці, яка подається в електронному вигляді, має надаватися у форматі електронного документа засобами передачі даних, визначеними одержуючим відомством з метою подання міжнародних заявок в електронній формі, за умови, якщо цей перелік надається у форматі електронного документа відповідно до параграфа 40 і, у разі

⁽³⁾ **Редакційна примітка:** Ці таблиці містять витяги з Таблиці характеристик у базі даних DDBJ/EMBL/GenBank (послідовності нуклеотидів) і з Таблиці характеристик у базі даних SWISS PROT (послідовності амінокислот).

необхідності, за допомогою засобів передачі даних, визначеними одержуючим відомством і компетентним органом.^{(4), (5)}

38. Будь-яка послідовність в електронній формі, як зазначено в параграфі 3*bis*, має надаватися у форматі електронного документа, визначеному одержуючим відомством (у випадку коригування) або компетентним органом (у випадку виправлення або внесення поправок), за умови, якщо цей перелік надається у форматі електронного документа відповідно до параграфа 40. У межах цього параграфа, будь-який перелік має подаватися засобами передачі даних, визначеними одержуючим відомством або компетентним органом; за можливості, перелік бажано подавати засобами передачі даних, визначеними як одержуючим відомством, так і компетентним органом.⁽⁶⁾

39. У межах цього параграфа, будь-який перелік послідовностей в електронній формі, зазначений у параграфі 4 і 4*bis*, що подається для цілей міжнародного пошуку або міжнародної попередньої експертизи, має надаватися у форматі електронного документа відповідно до параграфа 40 за допомогою засобів передачі даних, визначених компетентним органом.

40. Для цілей міжнародного пошуку і міжнародної попередньої експертизи, будь-який перелік послідовностей в електронній формі має міститися в одному електронному файлі, закодованому за допомогою кодової сторінки IBM⁽⁷⁾ Code Page 437, IBM Code Page 932⁽⁸⁾ або іншої сумісної кодової сторінки для представлення переліку послідовностей відповідно до параграфів 5-36 без включення жодних інших кодів. Сумісна кодова сторінка, необхідна, наприклад, для японських, китайських, кириличних, арабських, грецьких або єврейських символів, це така сторінка, в якій літери та цифри латинського алфавіту призначаються тим же шістнадцятковим позиціям, що й у зазначених вище кодових сторінках.

41. Будь-який перелік послідовностей у форматі електронного документа відповідно до параграфа 40 бажано створювати за допомогою спеціального програмного забезпечення, такого як PatentIn.

ПРОЦЕДУРИ В ЗАЗНАЧЕНИХ І ВИБРАНИХ ВІДОМСТВАХ

⁽⁴⁾ **Редакційна примітка:** Якщо перелік послідовностей в електронному вигляді, оформлений відповідно до цього стандарту, недоступний для компетентного органу у формі та спосіб, прийнятний для нього (тобто недоступний у форматі електронного документа відповідно до параграфа 40), компетентний орган може попросити заявника надати такий перелік у зазначеній вище формі (див. правило 13*ter*).

⁽⁵⁾ **Редакційна примітка:** Незалежно від формату електронного документа переліку послідовностей, необхідно дотримуватись просторового співвідношення (наприклад стовпчиків і рядків) елементів даних у переліку і формату власне послідовностей нуклеотидів та/або амінокислот, визначених у цьому Додатку.

⁽⁶⁾ **Редакційна примітка:** Якщо перелік послідовностей на заміну в електронній формі, включаючи будь-яке коригування, виправлення або поправку, недоступний для компетентного органу у формі та спосіб, прийнятний для нього (тобто недоступний у форматі електронного документа відповідно до параграфа 40), будь-яке подібне коригування, виправлення або будь-яка подібна поправка має братися цим органом до уваги для цілей міжнародного пошуку або попередньої експертизи в тому обсязі, який дозволить провести ефективний пошук або попередню експертизу без такого переліку послідовностей на заміну (див. параграф 4*bis*). Див. Редакційну примітку 38, яка також застосовується до будь-якого переліку послідовностей на заміну в електронній формі, зазначеного в параграфі 3*bis*.

⁽⁷⁾ **Редакційна примітка:** IBM є зареєстрованою торговельною маркою американської корпорації International Business Machines Corporation.

⁽⁸⁾ **Редакційна примітка:** Зазначені кодові сторінки є де-факто стандартами для персональних комп'ютерів.

42. З метою дотримання всіх вимог, визначених зазначеним або вибраним відомством, у якому вже почався розгляд міжнародної заявки, що містить одну або більше послідовностей нуклеотидів та/або амінокислот (див. правило 13ter.3):

(i) будь-яке посилання на одержуюче відомство або компетентний орган має тлумачитися як посилання на зазначене або вибране відомство;

(ii) будь-яке посилання на перелік послідовностей, що включається до міжнародної заявки в результаті коригування відповідно до правила 26, виправлення відповідно до правила 91 або поправки відповідно до статті 34(2)(b) до опису, в частині, що стосується послідовностей у поданій заявці, повинне також тлумачитися як таке, що включає будь-який перелік послідовностей, що міститься у заявці, відповідно до національного законодавства країни, в якій розташоване зазначене або вибране відомство, шляхом коригування (формального недоліку), виправлення (очевидної помилки) або внесення поправки до опису, в частині, що стосується послідовностей у поданій заявці;

(iii) будь-яке посилання на перелік послідовностей, що надається для цілей міжнародного пошуку або міжнародної попередньої експертизи, має тлумачитися як таке, що включає будь-який перелік, що надається зазначеному або вибраному відомству для проведення національного пошуку або експертизи;

(iv) зазначене або вибране відомство може запропонувати заявникові надати (у певний термін, прийнятний, враховуючи обставини) перелік послідовностей в електронній формі відповідно до цього стандарту для цілей національного пошуку та/або експертизи, якщо такий перелік не доступний відомству у прийнятній для нього формі та спосіб.

Доповнення

Доповнення 1: Цифрові ідентифікатори

Доповнення 2: Таблиця символів та характеристик амінокислот і нуклеотидів

Таблиця 1: Перелік нуклеотидів

Таблиця 2: Перелік модифікованих нуклеотидів

Таблиця 3: Перелік амінокислот

Таблиця 4: Перелік модифікованих і нестандартних амінокислот

Таблиця 5: Перелік ключів характеристик, що стосуються послідовностей нуклеотидів

Таблиця 6: Перелік ключів характеристик, що стосуються послідовностей білків

Доповнення 3: Зразок переліку послідовностей

[Доповнення 1-3 до Додатка С наведено далі]

Додаток С, Доповнення 1 Цифрові ідентифікатори

Лише цифрові ідентифікатори, зазначені нижче, можуть використовуватись в переліках послідовностей, що подаються в заявках. Назви елементів даних, наведені нижче (стовпчик «Опис цифрового ідентифікатора»), не повинні зазначатися в переліках послідовностей.

Цифрові ідентифікатори обов'язкових елементів даних, тобто елементів даних, які повинні зазначатися в усіх переліках послідовностей (див. цифрові ідентифікатори <110>, <120>, <160>, <210>, <211>, <212>, <213> і <400> у параграфі 25), і цифрові ідентифікатори елементів даних, які повинні зазначатися за умов, визначених цим стандартом (див. цифрові ідентифікатори <130>, <140>, <141>, <150>, <151> і <220>-<223> у параграфах 26, 27, 28, 29 і 30), позначені символом "О".

Цифрові ідентифікатори необов'язкових елементів даних (див. параграф 31) позначені символом "Н".

Цифровий ідентифікатор	Опис цифрового ідентифікатора	Обов'язковий (О) або необов'язковий (Н)	Примітки
<110>	Ім'я заявника	О	якщо ім'я заявника наводиться нелатинськими літерами, його також необхідно точно транслітерувати або перекласти англійською мовою
<120>	Назва винаходу	О	
<130>	Посилання на досьє заявки	О (за умов, визначених у параграфі 26)	див. параграф 26
<140>	Діюча патентна заявка	О (за умов, визначених у параграфі 27)	див. параграф 27; діюча патентна заявка ідентифікується за допомогою двобуквенного коду відповідно до стандарту VOIB ST.3 і номера національної заявки (у форматі, що використовується відомством промислової власності, до якого вона подається) або, у випадку міжнародної заявки, за допомогою номера міжнародної заявки

Цифровий ідентифікатор	Опис цифрового ідентифікатора	Обов'язковий (О) або необов'язковий (Н)	Примітки
<141>	Дата подання діючої патентної заявки	О (за умов, визначених у параграфі 27)	див. параграф 27; дата наводиться у форматі відповідно до стандарту VOIB ST.2 (PPPP MM ДД)
<150>	Більш рання патентна заявка	О (за умов, визначених у параграфі 28)	див. параграф 28; більш рання патентна заявка ідентифікується за допомогою двобуквенного коду відповідно до стандарту VOIB ST.3 і номера національної заявки (у форматі, що використовується відомством промислової власності, до якого вона подається) або, у випадку міжнародної заявки, за допомогою номера міжнародної заявки
<151>	Дата подання більш ранньої патентної заявки	О (за умов, визначених у параграфі 28)	див. параграф 27; дата наводиться у форматі відповідно до стандарту VOIB ST.2 (PPPP MM ДД)
<160>	Кількість SEQ ID NO	О	
<170>	Програмне забезпечення	Н	
<210>	SEQ ID NO: x	О	x – ціле число, яке є ідентифікаційним номером послідовності
<211>	Довжина	О	довжина послідовності, яка виражається в кількості пар основ або амінокислот

Цифровий ідентифікатор	Опис цифрового ідентифікатора	Обов'язковий (О) або необов'язковий (Н)	Примітки
<212>	Тип	О	тип молекули, секвенованої в послідовності з ідентифікаційним номером "SEQ ID NO: x": ДНК, РНК, або ПРТ (протеїн); якщо послідовність нуклеотидів містить фрагменти як ДНК, так і РНК, значенням буде «ДНК»; крім того, комбінована молекула ДНК/РНК має бути описана під ідентифікаторами <220>-<223>
<213>	Організм	О	видова, тобто наукова назва, або "Artificial sequence" (штучна послідовність), або "Unknown" (невідома)
<220>	Характеристика	О (за умов, визначених у параграфах 29 і 30)	не заповнюється; див. параграфи 29 і 30; опис особливостей біологічного значення в послідовності з ідентифікаційним номером "SEQ ID NO: x" (може повторюватись залежно від кількості зазначених характеристик)
<221>	Ключ	О (за умов, визначених у параграфі 29)	див. параграф 29; необхідно використовувати лише ключі, наведені у Таблицях 5 і 6 Доповнення 2

Цифровий ідентифікатор	Опис цифрового ідентифікатора	Обов'язковий (О) або необов'язковий (Н)	Примітки
<222>	Локалізація	О (за умов, визначених у параграфі 29)	<p>див. параграф 29;</p> <ul style="list-style-type: none"> - від (номер першої основи/амінокислоти в характеристиці); - до (номер останньої основи/амінокислоти в характеристиці); - пари основ (номери, що відповідають позиціям пар основ у послідовності нуклеотидів); - амінокислоти (номери, що відповідають позиціям амінокислотних залишків у послідовності амінокислот); - розміщення характеристики на ланцюгу, комплементарному до ланцюга, наведеного в переліку послідовностей
<223>	Інша інформація	О (за умов, визначених у параграфах 29 і 30)	<p>див. параграфи 29 і 30; будь-яка інша релевантна інформація, викладена з використанням мовно-незалежної лексики або довільним текстом (бажано англійською мовою); будь-який довільний текст має повторюватись в описі тією ж мовою (див. параграф 36); якщо послідовність містить будь-яку модифіковану основу або модифіковану/нестандартну L-амінокислоту, наведену в Таблицях 2 і 4 Доповнення 2, необхідно використовувати один із символів, наведених у Таблицях 2 і 4 Доповнення 2, що відповідає такій основі або амінокислоті</p>

Цифровий ідентифікатор	Опис цифрового ідентифікатора	Обов'язковий (О) або необов'язковий (Н)	Примітки
<300>	Інформація про публікацію	Н	не заповнюється; повторюється для кожної релевантної публікації
<301>	Автори	Н	
<302>	Назва	Н	назва публікації
<303>	Періодичне видання	Н	назва періодичного видання, в якому було опубліковано інформацію
<304>	Том	Н	том періодичного видання, в якому було опубліковано інформацію
<305>	Номер	Н	номер періодичного видання, в якому було опубліковано інформацію
<306>	Сторінки	Н	сторінки періодичного видання, на яких було опубліковано інформацію
<307>	Дата	Н	дата публікації періодичного видання; за можливості, дату необхідно наводити у форматі відповідно до стандарту VOIB ST.2 (PPPP MM ДД)
<308>	Номер у базі даних	Н	номер у базі даних, включаючи назву бази даних
<309>	Дата внесення до бази даних	Н	дата внесення до бази даних; дата наводиться у форматі відповідно до стандарту VOIB ST.2 (PPPP MM ДД)

Цифровий ідентифікатор	Опис цифрового ідентифікатора	Обов'язковий (О) або необов'язковий (Н)	Примітки
<310>	Номер документа	Н	номер документа (лише для патентних посилань); повнотекстовий документ позначається, в наступному порядку, за допомогою двобуквенного коду відповідно до стандарту VOIB ST.3, номера публікації відповідно до стандарту VOIB ST.6 і коду виду документа відповідно до стандарту VOIB ST.16
<311>	Дата подання	Н	дата подання документа (лише для патентних посилань); дата наводиться у форматі відповідно до стандарту VOIB ST.2 (PPPP MM ДД)
<312>	Дата публікації	Н	дата публікації документа (лише для патентних посилань); дата наводиться у форматі відповідно до стандарту VOIB ST.2 (PPPP MM ДД)
<313>	Релевантні залишки у послідовності з ідентифікаційним номером "SEQ ID NO: x": від ... до ...	Н	
<400>	Послідовність	О	Ідентифікаційний номер "SEQ ID NO: x" повинен наводитися після цифрового ідентифікатора в рядку, що передує послідовності (див. Доповнення 3)

[Доповнення 2 до Додатку С наведено далі]

Додаток С, Доповнення 2
Таблиця символів і характеристик амінокислот і нуклеотидів

Таблиця 1: Перелік нуклеотидів

Символ	Значення	Походження позначення
a	a	аденін (<u>a</u> denine)
g	g	гуанін (<u>g</u> uanine)
c	c	цитозин (<u>c</u> ytosine)
t	t	тимін (<u>t</u> hymine)
u	u	урацил (<u>u</u> racil)
r	g або a	пурин (pu <u>r</u> ine)
y	t/u або c	піримідин (py <u>r</u> imidine)
m	a або c	аміно (a <u>m</u> ino)
k	g або t/u	кето (<u>k</u> eto)
s	g або c	сильна (<u>s</u> trong) взаємодія третинний водневий зв'язок
w	a або t/u	слабка (<u>w</u> weak) взаємодія вторинний водневий зв'язок
b	g або c або t/u	не a
d	a або g або t/u	не c
h	a або c або t/u	не g
v	a або g або c	не t, не u
n	a або g або c або t/u, невідомий або інше	будь-який (<u>a</u> ny)

Таблиця 2: Перелік модифікованих нуклеотидів

Символ	Значення
ac4c	4-ацетилцитидин
chm5u	5-(карбоксигідроксиметил)уридин
cm	2'-О-метилцитидин
cmnm5s2u	5-карбоксиметиламінометил-2-тіоуридин
cmnm5u	5-карбоксиметиламінометилуридин
d	дигідроуридин
fm	2'-О-метилпсевдоуридин
gal q	бета, D-галактозинквеуозин
gm	2'-О-метилгуанозин
l	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
m5c	5-метилцитидин
m6a	N6-метиладенозин
m7g	7-метилгуанозин
mam5u	5-метиламінометилуридин
mam5s2u	5-метоксиамінометил-2-тіоуридин
man q	бета, D-маннозилквеуозин
mcm5s2u	5-метоксикарбонілметил-2-тіоуридин
mcm5u	5-метоксикарбонілметилуридин
mo5u	5-метоксиуридин
ms2i6a	2-метилтіо-N6-ізопентениладенозин

Символ	Значення
ms2t6a	N-((9-бета-D-рибофуранозил-2-метилгіопурин-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
osyw	вибутоксозин
p	псевдоуридин
q	квеуозин
s2c	2-тіоцитидин
s2t	5-метил-2-тіоуридин
s2u	2-тіоуридин
s4u	4-тіоуридин
t	5-метилуридин
t6a	N-((9-бета-D-рибофуранозилпурин-6-їл)-карбамоїл)треонін
tm	2'-О-метил-5-метилуридин
um	2'-О-метилуридин
yw	вибутозин
x	3-(3-аміно-3-карбокси-пропіл)уродин, (аср3)u

Таблиця 3: Перелік амінокислот

Символ	Значення
Ala	аланін
Cys	цистеїн
Asp	аспарагінова кислота
Glu	глутамінова кислота
Phe	фенілаланін
Gly	гліцин
His	гістидин
Ile	ізолейцин
Lys	лізин
Leu	лейцин
Met	метіонін
Asn	аспарагін
Pro	пролін
Gln	глутамін
Arg	аргінін
Ser	серин
Thr	треонін
Val	валін
Trp	триптофан
Tyr	тирозин
Asx	аспарагінова кислота або аспарагін
Glx	глутамінова кислота або глутамін
Xaa	невідомо або інше

Таблиця 4: Перелік модифікованих і нестандартних амінокислот

Символ	Значення
Aad	2-аміноадипінова кислота
bAad	3-аміноадипінова кислота
bAla	бета-аланін, бета-амінопропіонова кислота
Abu	2-аміномасляна кислота
4Abu	4-аміномасляна кислота, піперидинова кислота
Acpr	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-етилглїцин
EtAsn	N-етиласпарагін
Hyl	гїдроксилїзин
aHyl	алло-гїдроксилїзин
3Hyp	3-гїдроксипролін
4Hyp	4-гїдроксипролін
Ide	їзодесмозин
alle	алло-їзолейцин
MeGly	N-метилглїцин, саркозин
Melle	N-метилїзолейцин
MeLys	6-N-метиллїзин
MeVal	N-метилвалїн
Nva	норвалїн
Nle	норлейцин
Orn	орнїтин

Таблиця 5: Перелік ключів характеристик, що стосуються послідовностей нуклеотидів

Ключ	Опис
allele	споріднений індивід або штам містить стійкі, альтернативні форми одного й того ж гена, який відрізняється від представленої послідовності в даній локалізації (а, можливо, і в інших)
attenuator	(1) область ДНК, на якій відбувається регулювання термінації транскрипції та яка контролює експресію деяких бактеріальних оперонів; (2) ділянка послідовності, розташована між промотором і першим структурним геном, який викликає часткову термінацію транскрипції
C_region	константна область легких і важких ланцюгів імуноглобуліну, альфа-, бета- і гама-ланцюгів Т-клітинного рецептора; включає один або більше екзонів залежно від ланцюга
CAAT_signal	CAAT-бокс; частина консервативної послідовності, розташована на відстані приблизно 75 bp (пар основ) від початкової точки еукаріотичних одиниць транскрипції, які можуть брати участь у полімеразному зв'язуванні РНК; консенсус=GG (С або Т) CAATCT
CDS	кодуюча послідовність; послідовність нуклеотидів, що відповідає послідовності амінокислот у білку (локалізація включає термінувальний кодон); характеристика включає концептуальну трансляцію амінокислот
conflict	незалежні визначення «тієї ж самої» послідовності відрізняються на цьому сайті або в даній області
D-loop	петля зміщення; область в межах мітохондріальної ДНК, на якій короткий проміжок РНК паруються 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
D-segment	D-ділянка важкого ланцюга імуноглобуліну і бета-ланцюга Т-клітинного рецептора
enhancer	послідовність, що діє в цис-положенні, яка збільшує використання (деяких) еукаріотичних промоторів і може функціонувати в будь-якому напрямку та місці (вгору або вниз) відносно промотора
exon	область генома, в якій закодована частина сплайсованої матричної (інформаційної) РНК; може містити нетрансльовані ділянки на 5'-кінці (5'UTR), усі кодуючі послідовності (CDS) і нетрансльовані ділянки на 3'-кінці (3'UTR)

Ключ	Опис
GC_signal	GC-бокс; консервативна GC-збагачена область, розташована вище початкової точки еукаріотичних транскрипційних одиниць, яка може повторюватись декілька разів або у будь-якій орієнтації; консенсус=GGGCGG
gene	область біологічної важливості, що визначається як ген і якій була присвоєна назва
iDNA	проміжна ДНК; ДНК, що знищується у процесі будь-якої рекомбінації
intron	ділянка ДНК, що транскрибується, але видаляється з транскрипта шляхом сплайсингу послідовностей (екзонів) з усіх її сторін
J_segment	з'єднувальна ділянка легких і важких ланцюгів імуноглобуліну, альфа-, бета- і гама-ланцюгів Т-клітинного рецептора
LTR	довге термінальне повторення (LTR-послідовність); послідовність, яка безпосередньо повторюється на обох кінцях певної послідовності, і яка зазвичай зустрічається в ретровірусах
mat_peptide	кодуюча послідовність зрілого пептиду або білка; кодуюча послідовність для зрілого або кінцевого пептиду або білка, що є продуктом пост-трансляційної модифікації; локалізація не охоплює термінаторний кодон, на відміну від відповідної кодуючої послідовності
misc_binding	сайт нуклеїнової кислоти, що ковалентно або нековалентно зв'яже другий сайт, який не можна описати будь-яким іншим ключем зв'язування (primer_bind або protein_bind)
misc_difference	характерна послідовність відрізняється від послідовності, наведеної у записі, і не може бути описана будь-яким іншим ключем відмінності (conflict, unsure, old_sequence, mutation, variation, allele або modified_base)
misc_feature	область біологічної важливості, яку не можна описати будь-яким іншим ключем характеристики; нова або рідкісна характеристика
misc_recomb	сайт будь-якої узагальненої, сайт-специфічної події, або події реплікативної рекомбінації, на якій присутній розрив і відновлення двоспіральної ДНК, що не можуть бути описані іншими ключами рекомбінації (iDNA і virion) або ідентифікаторами вихідних ключів (/insertion_seq, /transposon, /proviral)
misc_RNA	будь-який транскрипт або продукт РНК, який не можна визначити іншими ключами РНК (prim_transcript, precursor_RNA, mRNA, 5'clip, 3'clip, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, rRNA, tRNA, scRNA або snRNA)

Ключ	Опис
misc_signal	будь-яка область, що містить сигнал управління або змінювання функції чи експресії генів, яку не можна описати іншими ключами сигналу (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator або rep_origin)
misc_structure	будь-яка вторинна або третинна структура чи конформація, яку не можна описати іншими ключами структури (stem_loop або D-loop)
modified_base	вказаний нуклеотид є модифікованим нуклеотидом і має бути замінений на вказану молекулу, зазначену під ідентифікатором mod_base
mRNA	матрична (інформаційна) РНК; включає нетрансльовану область 5'-кінця (5'UTR), кодуючі послідовності (CDS, exon) та нетрансльовану область 3'-кінця (3'UTR)
mutation	споріднений штам з раптовою спадковою зміною в послідовності у цій локалізації
N_region	додаткові нуклеотиди, що вставляються між перегрупованими ділянками імуноглобуліну
old_sequence	представлена послідовність змінює попередній варіант послідовності в даній локалізації
polyA_signal	область розпізнавання, необхідна для ендонуклеазного розщеплення транскрипта РНК, за яким слідує поліаденілування; консенсус=AATAAA
polyA_site	сайт транскрипта РНК, до якої будуть додані аденінові залишки шляхом пост-транскрипційного поліаденілування
precursor_RNA	будь-який тип РНК, який не є зрілим продуктом РНК; може містити урізану область 5'-кінця (5'clip), нетрансльовану область 5'-кінця (5'UTR), кодуючі послідовності (CDS, exon), вставні послідовності (intron), нетрансльовану область 3'-кінця (3'UTR) та урізану область 3'-кінця (3'clip)
prim_transcript	первинний (початковий, необроблений) транскрипт; містить урізану область 5'-кінця (5'clip), нетрансльовану область 5'-кінця (5'UTR), кодуючі послідовності (CDS, exon), вставні послідовності (intron), нетрансльовані області 3'-кінця (3'UTR) та урізані області 3'-кінця (3'clip)
primer_bind	сайт нековалентного зв'язування праймера для ініціації реплікації, транскрипції або зворотної транскрипції; включає сайт (сайти) для синтетичних елементів, наприклад праймерів для ланцюгової реакції полімерази (PCR)
promoter	область молекули ДНК, яка бере участь у зв'язуванні РНК-полімерази для ініціювання транскрипції
protein_bind	сайт нековалентного білкового зв'язку в нуклеїновій кислоті
RBS	сайт зв'язування рибосом

Ключ	Опис
repeat_region	область генома, що містить повторювані одиниці
repeat_unit	одиничний повторюваний елемент
rep_origin	джерело реплікації; сайт початку подвоювання нуклеїнової кислоти для отримання двох ідентичних копій
rRNA	зріла рибосомна РНК; компонент РНК рибонуклеопроїєїнової частки (рибосоми), яка збирає амінокислоти в білки
S_region	область перемикування важких ланцюгів імуноглобуліну; бере участь в перебудові важких ланцюгів ДНК, що призводить до експресії імуноглобуліну іншого класу з тієї ж самої В-клітини
satellite	велика кількість тандемних повторів (ідентичних або споріднених) короткої базової повторюваної одиниці; має базовий склад або іншу властивість, яка відрізняється від середнього геномного значення, що дозволяє їм виділятися з основної маси (основного банку) геномної ДНК
scRNA	мала цитоплазматична РНК; будь-яка з декількох малих цитоплазматичних молекул РНК, присутніх у цитоплазмі, а також (іноді) в ядрі еукаріота
sig_peptide	кодуюча послідовність сигнального пептида; кодуюча послідовність для N-кінцевого домену виділеного білка; цей домен бере участь у приєднанні поліпептиду, що збільшується, до мембрани; лідерна послідовність
snRNA	мала ядерна РНК; будь-яка з декількох видів малих РНК, обмежених ядром; декілька малих ядерних РНК, що беруть участь у сплайсингу або іншій реакції оброблення РНК
source	ідентифікує біологічне джерело визначеної довжини послідовності; цей ключ є обов'язковим; кожний запис повинен містити принаймні один ключ джерела, що охоплює всю послідовність; дозволяється зазначити декілька ключів для однієї послідовності
stem_loop	«шпилька»; двоспіральна область, утворена шляхом парування основ між суміжними (інвертованими) комплементарними послідовностями в одному ланцюгу РНК або ДНК
STS	ДНК-маркувальний сайт; коротка унікальна послідовності ДНК, що описує картування генома і може бути виявлена за допомогою полімеразної ланцюгової реакції (ПЛР); область генома може бути картована шляхом визначення порядку серій ДНК-маркувальних сайтів

Ключ	Опис
TATA_signal	ТАТА-бокс; бокс Гольдберга-Гогнесса; консервативний септамір, збагачений АТ, який розташований на відстані 25 bp (пар основ) до початкової точки кожної одиниці транскрипції еукаріотичної РНК-полімерази II, яка може брати участь в позиціюванні ензима для коректного ініціювання; консенсус=TATA(A або T)A(A або T)
terminator	послідовність ДНК, розташована наприкінці транскрипта або поряд з областю промотера, яка змушує РНК-полімеразу завершити транскрипцію; також може бути сайтом зв'язування репресорного білка
transit_peptide	кодуюча послідовність транзитного пептида; кодуюча послідовність для N-кінцевого домену ядерно-кодованого органоїдного білка; цей домен бере участь у пост-трансляційному перенесенні білка до органели
tRNA	зріла транспортна РНК; мала молекула РНК (довжиною 75-85 основ), яка допомагає в процесі трансляції послідовності нуклеїнових кислот у послідовність амінокислот
unsure	автор невпевнений у точній послідовності в даній області
V_region	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
V_segment	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
variation	споріднений штам містить стабільні мутації з того ж гена (наприклад, позиції довжин рестрикційних фрагментів (RFLP), поліморфізми тощо), які відрізняються від наведеної послідовності в цій локалізації (а, можливо, і в інших)
3'clip	кінцева область 3'-кінця попереднього транскрипта, яка урізається під час процесингу
3'UTR	область 3'-кінця зрілого транскрипта (після стоп-кодона), що не транслюється до білка
5'clip	кінцева область 5'-кінця попереднього транскрипта, яка урізається під час процесингу
5'UTR	область 5'-кінця зрілого транскрипта (перед кодоном ініціувальним кодоном), що не транслюється до білка

Ключ	Опис
-10_signal	бокс Прибнова; консервативна область на відстані 10 bp (пар основ) вище від початкової точки одиниць бактеріальної транскрипції, які можуть брати участь у зв'язуванні РНК-полімерази; консенсус=ТAtAaТ
-35_signal	консервативний гексамер на відстані 35 bp (пар основ) вище від початкової точки одиниць бактеріальної транскрипції; консенсус= ТТGACa [] або TGTTGACA []

Таблиця 6: Перелік ключів характеристик, що стосуються послідовностей білків

Ключ	Опис
CONFLICT	неспівпадіння послідовностей у різних публікаціях
VARIANT	автори повідомляють про існування різних варіантів послідовності
VARSP LIC	опис варіантів послідовності, отриманих шляхом альтернативного сплайсингу
MUTAGEN	ділянка, змінена експериментальним шляхом
MOD_RES	пост-трансляційна модифікація залишку
ACETYLATION	N-кінцевого або іншого
AMIDATION	по суті на C-кінці зрілого активного пептиду
BLOCKED	невизначеної групи блокування на N-кінці або C-кінці
FORMYLATION	N-термінального метіоніну
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	аспарагіну, аспарагінової кислоти, проліну або лізину
METHYLATION	по суті лізину або аргініну
PHOSPHORYLATION	серину, треоніну, тирозину, аспарагінової кислоти або гістидину
PYRROLIDONE CARBOXYLIC ACID	N-кінцевого глутамату, що утворив внутрішній циклічний лактам
SULFATATION	по суті тирозину
LIPID	ковалентне зв'язування ліпідної частини
MYRISTATE	міристатної групи, приєднаної до N-кінцевого залишку гліцину зрілої форми білка або до внутрішнього залишку лізину за допомогою амідного зв'язку
PALMITATE	пальмітатної групи, приєднаної до залишку цистеїну за допомогою тіоефірного зв'язку або до залишку серину чи треоніну за допомогою ефірного зв'язку
FARNESYL	фарнезильної групи, приєднаної до залишку цистеїну за допомогою тіоефірного зв'язку
GERANYL-GERANYL	гераніл-геранілової групи, приєднаної до залишку цистеїну за допомогою тіоефірного зв'язку
GPI-ANCHOR	глікозил-фосфатидилінозитолової (ГФІ) групи, приєднаної до альфа-карбоксильної групи C-кінцевого залишку зрілої форми білка

Ключ	Опис
N-ACYL DIGLYCERIDE	N-кінцевого цистеїну зрілої форми прокаріотичного ліпопротеїну з амідозв'язаною жирною кислотою і гліцериновою групою, до якої приєднані дві жирні кислоти за допомогою ефірного зв'язку
DISULFID	дисульфідний зв'язок; кінцеві точки «ВІД» і «ДО» представляють собою два залишки, поєднані за допомогою міжланцюгового дисульфідного зв'язку; якщо кінцеві точки «ВІД» і «ДО» співпадають, тоді дисульфідний зв'язок є міжланцюговим, а в рядку опису зазначаються властивості поперечного зв'язку
THIOLEST	складний тіолефірний зв'язок; кінцеві точки «ВІД» і «ДО» представляють собою два залишки, поєднані за допомогою складного тіолефірного зв'язку
THIOETH	тіоефірний зв'язок; кінцеві точки «ВІД» і «ДО» представляють собою два залишки, поєднані за допомогою тіоефірного зв'язку
CARBOHYD	ділянка глікозилування; в рядку опису зазначаються властивості вуглеводу (якщо відомо)
METAL	ділянка зв'язування іона металу; в рядку опису зазначаються властивості металу
BINDING	ділянка зв'язування будь-якої хімічної групи (коензимів, простетичної групи тощо); в рядку опису зазначаються хімічні властивості групи
SIGNAL	протяжність сигнальної послідовності (препептида)
TRANSIT	протяжність транзитного пептида (мітохондріального, хлоропластичного або для мікротільця)
PROPEP	протяжність препептида
CHAIN	протяжність поліпептидного ланцюга у зрілому білку
PEPTIDE	протяжність вивільненого активного пептида
DOMAIN	протяжність відповідного домену в послідовності; в рядку опису зазначаються властивості домену
CA_BIND	протяжність області кальцій-зв'язування
DNA_BIND	протяжність області зв'язування ДНК
NP_BIND	протяжність області зв'язування нуклеотидфосфату; в рядку опису зазначаються властивості нуклеотидфосфату
TRANSMEM	протяжність трансмембранної області

Ключ	Опис
ZN_FING	протяжність області білка «цинкові пальці»
SIMILAR	протяжність подібності з іншою послідовністю білків; в рядку опису наводиться точна інформація стосовно даної послідовності
REPEAT	протяжність повторення внутрішньої послідовності
HELIX	вторинна структура: спіралі, наприклад альфа-спіраль, 3(10)-спіраль або Pi-спіраль
STRAND	вторинна структура: бета-ланцюг, наприклад бета-ланцюг з водневим зв'язком або залишок в ізольованому бета-містку
TURN	петлі вторинної структури, наприклад петля з Н-зв'язком (3-петля, 4-петля або 5-петля)
ACT-SITE	Амінокислота (-и), що бере (-уть) участь в діяльності ензима
SITE	будь-яка інша ділянка в послідовності
INIT_MET	послідовність починається з ініціувального метіоніну
NON_TER	залишок наприкінці послідовності не є кінцевим (термінальним); якщо це стосується позиції 1, значить перша позиція не є N-кінцем усієї молекули; якщо це стосується останньої позиції, значить ця позиція не є С-кінцем усієї молекули; для цього ключа рядок опису не передбачено
NON_CONS	непослідовні залишки; вказує на те, що два залишки в послідовності не є послідовними і між ними знаходиться ще низка непослідовних залишків
UNSURE	невизначеності в послідовності; використовується для позначення області (-ей) послідовності, стосовно якої (яких) автори не впевнені щодо точності призначення послідовності

[Доповнення 3 до Додатка С наведено далі]

Додаток С, Доповнення 3
Зразок переліку послідовностей

```

<110>      Smith, John; Smithgene Inc.
<120>      Example of a Sequence Listing (зразок переліку послідовностей)
<130>      01-00001
<140>      PCT/EP98/00001
<141>      1998-12-31
<150>      US 08/999,999
<151>      1997-10-15
<160>      4
<170>      PatentIn version 2.0
<210>      1
<211>      389
<212>      DNA (ДНК)
<213>      Paramecium sp. (парамеції)
<220>
<221>      CDS (кодує послідовність)
<222>      (279)...(389)
<300>
<301>      Doe, Richard
<302>      Isolation and Characterization of a Gene Encoding a Protease from
      Paramecium sp. (Ізолювання і характеризування гена, що кодує протеази з
      парамеції)
<303>      Journal of Genes
<304>      1
<305>      4
<306>      1-7
<307>      1988-06-31
<308>      123456
<309>      1988-06-31
<400>      1
agctgtagtc  attcctgtgt  cctcttctct  ctgggcttct  Caccctgcta  atcagatctc          60
agggagagtg  tcttgaccct  cctctgcctt  tgcagcttca  Caggcaggca  ggcaggcagc          120
tgatgtggca  attgctggca  gtgccacagg  cttttcagcc  Aggcttaggg  tgggttccgc          180
cgcggcgcgg  cggccctct  cgcgctcctc  tcgcgcttct  Ctctcgctct  cctctcgctc          240
ggacctgatt  aggtgagcag  gaggaggggg  cagttagc    atg  Gtt  tca  atg  ttc  agc  296
              Met Val Ser Met Phe Ser
              1          5
ttg  tct  ttc  aaa  tgg  cct  gga  ttt  tgt  ttg  ttt  Gtt  tgt  ttg  ttc  caa  344
Leu  Ser  Phe  Lys  Trp  Pro  Gly  Phe  Cys  Leu  Phe  Val  Cys  Leu  Phe  Gln
              10          15          20
tgt  ccc  aaa  gtc  ctc  ccc  tgt  cac  tca  tca  ctg  Cag  ccg  aat  ctt          389
Cys  Pro  Lys  Val  Leu  Pro  Cys  His  Ser  Ser  Leu  Gln  Pro  Asn  Leu
              25          30          35

```

<210> 2
 <211> 37
 <212> PRT (ревертаза)
 <213> Paramecium sp. (парамеції)
 <400> 2
 Met Val Ser Met Phe Ser Leu Ser Phe Lys Trp Pro Gly Phe Cys Ley
 1 5 10 15
 Phe Val Cys Leu Phe Gln Cys Pro Lys Val Leu Pro Cys His Ser Ser
 20 25 30
 Leu Gln Pro Asn Leu
 35

<210> 3
 <211> 11
 <212> PRT (ревертаза)
 <213> Artificial sequence (штучна послідовність)

<220>
 <223> Designed peptide based on size and polarity to act as a linker between the alpha and beta chains of Protein XYZ. (Синтезований пептид має розмір і полярність, що дозволяють йому виконувати роль зв'язувального елемента (лінкера) між альфа- і бета-ланцюгами протеїна XYZ.)

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

<210> 4

<400> 4

000

[Кінець Доповнення 3 і Стандарту]