

ST. 26 標準

建議使用XML(可擴展標記語言)表示核苷酸及胺基酸序列表的標準

版本 1.5

WIPO標準委員會(CWS)於2021年11月5日第九屆會議批准修訂

國際局編輯備註

WIPO大會第54屆會議(第25屆例行會議)批准了WIPO ST. 26標準在國家、地區及國際層面的新大爆炸實施日期，即2022年7月1日。同時，建議繼續使用ST. 25標準

本標準發布供專利主管機關及其他相關單位參考

目錄

簡介	2
定義	2
範圍	5
參考	5
序列表示	6
核苷酸序列	6
胺基酸序列	10
特殊情況表示	12
XML序列表結構	12
根要素(<i>Root element</i>)	14
基本資訊部分	15
序列資料部分	18
特徵表(<i>Feature table</i>)	21
特徵關鍵(<i>Feature keys</i>)	21
必要特徵關鍵 (<i>Mandatory feature keys</i>)	21
特徵位置(<i>Feature location</i>)	22
特徵限定詞(<i>Feature qualifiers</i>)	25
必要特徵限定詞(<i>Mandatory feature qualifiers</i>)	25
限定詞要素(<i>Qualifier elements</i>)	25
自由文本(<i>free text</i>)	28
編碼序列	30
變體(<i>Variants</i>)	31

附錄

附錄I-受控詞彙(*Controlled vocabulary*)

附錄II-序列表文檔類型定義(*DTD*)

附錄III-序列表例示(*XML文件*)

附錄IV-用於XML序列表的Unicode基本拉丁文代碼子集

附錄V-附加資料交換要件(僅供IPOs)

附錄VI-例示指南文件

 附件- XML格式的指導文件序列(*Guidance document sequence in XML*)

附錄VII-序列表ST. 25轉換為ST. 26的建議：可能添加或刪除的主題

ST.26 標準

建議使用 XML(可擴展標記語言)表示核苷酸及胺基酸序列 表的標準

版本 1.5

WIPO標準委員會(CWS)於2021年11月5日第9屆會議批准修訂

簡介

1. 本標準定義專利申請案核苷酸及胺基酸序列揭露的表示方式以及XML(eXtensible Markup Language, 可擴展標記語言)序列表的文檔類型定義(Document Type Definition, DTD)。建議專利主管機關接受任何符合本標準的序列表作為專利申請的一部分或專利申請相關文件。
2. 本標準目的：
 - (a) 允許申請人在專利申請中制定1份符合國際及國家或地區程序要求的序列表；
 - (b) 提高序列呈現的準確及品質，以利於序列資料傳播，使申請人、公眾及審查人員受益；
 - (c) 促進序列資料的檢索；以及
 - (d) 允許以電子形式交換序列資料並引入電腦數據庫。

定義

3. 就本標準而言，用語定義：
 - (a) 「胺基酸(amino acid)」係指任何可以使用附錄I中列出的任何符號表示的任何胺基酸(參見第3節表3)。所述胺基酸包含，特別是D-胺基酸及含有修飾或合成側鏈的胺基酸。若未根據第30段進一步於特徵表中描述修飾，則胺基酸視為未修飾的L-胺基酸。基於本標準之目的，肽

核酸(PNA)殘基不被視為胺基酸，而被視為第3(g)(i)(2)段所列的核苷酸。

- (b) 「受控詞彙(controlled vocabulary)」是本標準中在描述序列特徵時必須使用的術語，即附錄I中規定之感興趣區域或位點相關註釋。
- (c) 「殘基列舉(enumeration of its residues)」係指按順序列出各殘基以揭露專利申請的序列，其中：
 - (i) 殘基由名稱、縮寫、符號或結構表示(例如：HHHHHHHQ或HisHisHisHisHisHisGln)。
 - (ii) 多個殘基由簡寫式(shorthand formula)表示(例如：His₆Gln)。
- (d) 「刻意跳過序列(intentionally skipped sequence)」，也稱為空序列，指為了與專利申請揭露一致而保留序列表中序列編號的佔位，例如，從揭露中刪除序列以避免對序列表中的序列重新編號。
- (e) 「修飾胺基酸(modified amino acid)」係指除了L-丙胺酸、L-精胺酸、L-天冬醯胺酸、L-天冬胺酸、L-半胱胺酸、L-麩醯胺酸、L-麩胺酸、L-甘胺酸、L-組胺酸、L-白胺酸異白胺酸、L-白胺酸、L-離胺酸、L-甲硫胺酸、L-苯丙胺酸、L-脯胺酸、L-吡咯離胺酸、L-絲胺酸、L-硒代半胱胺酸、L-蘇胺酸、L-色胺酸、L-酪胺酸或L-纈胺酸外，於第3(a)段中描述的任何胺基酸。
- (f) 「修飾核苷酸(modified nucleotide)」係指除了脫氧腺苷3'-單磷酸、脫氧鳥苷3'-單磷酸、脫氧胞苷3'-單磷酸、脫氧胸苷3'-單磷酸、腺苷3'-單磷酸、鳥苷3'-單磷酸、胞苷3'-單磷酸或尿苷3'-單磷酸外，於第3(g)段中描述的任何核苷酸。
- (g) 「核苷酸(nucleotide)」係指任何核苷酸或核苷酸類似物，可以使用附錄I所述任何符號表示(參見第1節表1)，其中核苷酸或核苷酸類似物包含：
 - i. 骨架部分(moiety)選自：
 - (1) 2' 去氧核糖5' 單磷酸(去氧核糖核苷酸的骨架部分)或核糖5' 單磷酸(核糖核苷酸的骨架部分)；或者
 - (2) 2' 去氧核糖5' 單磷酸或核糖5' 單磷酸的類似物，當形成核酸類似物的主鏈時，導致模擬含有2' 去氧核糖5' 單磷酸或核糖5' 單磷酸骨架的核酸中的核鹼基排列，其中核酸類似物能夠與互補核酸鹼基配對。骨架部分的實例包含，如肽核酸中的胺基酸，如乙二醇核酸中的乙二醇分子，如蘇糖核酸中的蘇糖呋喃糖分子，如嗎啉中的嗎啉環及二胺基磷酸酯基團，以及如環己烯基核酸中的環

己烯基分子。

以及

ii. 骨架部分是：

- (1) 連接核鹼基，包含修飾或合成的嘌呤或嘧啶核鹼基；或者
 - (2) 當核苷酸是序列的一部分，缺少嘌呤或嘧啶核鹼基時，稱為「AP位點」或「無鹼基位點」。
- (h) 「殘基(residue)」係指序列中的任何單個核苷酸或胺基酸或其類似物。
- (i) 「序列識別號(sequence identification number)」係指序列表中分配給各序列的唯一編號(整數)。
- (j) 「序列表(sequence listing)」係指專利申請時提交說明書的一部分或申請後提交的文件，其中包含揭露的核苷酸及/或胺基酸序列，以及如本標準規定之任何進一步描述。
- (k) 「具體定義(specifically defined)」係指除了以符號「n」所表示者外的任何核苷酸，及除了以符號「X」所表示者以外的任何胺基酸，列於附錄I(分別參見第1節表1以及第3節表3)。
- (l) 「未知(unknown)核苷酸或胺基酸」係指所呈現的單一核苷酸或胺基酸其身份未知或未揭露。
- (m) 「變體序列(variant sequence)」係指包含1個或多個差異的核苷酸或胺基酸序列。相對主要序列，差異可能包含替代殘基(alternative residues)(參見第15及27段)、修飾殘基(modified residues)(參見第3(g)、3(h)、16及29段)、刪除(deletions)、插入(insertions)及置換(substitutions)。參見第93至95段。
- (n) 「自由文本(free text)」係指某些限定詞的值格式，是以描述性文字短語或其他指定格式(如附錄I所示)的形式呈現。參見第85段。
- (o) 「語言依賴自由文本(language-dependent free text)」係指某些限定詞的自由文本內容，可能因國家、區域或國際程序要求須進行翻譯。參見第87段。

4. 就本標準而言，詞語：

- (a) 「可以(may)」係指可選或允許的方法，但不要求。
- (b) 「必須(must)」係指本標準的要求；無視要求將導致不合規定。
- (c) 「不得(must not)」係指本標準的禁止。
- (d) 「建議(should)」係指強烈鼓勵，但不要求。

(e) 「不建議(should not)」係指強烈反對，但不禁止。

範圍

5. 本標準規定專利申請中揭露序列之核苷酸及胺基酸序列表的記載要求。
6. 符合本標準之序列表(以下簡稱序列表)包含基本資訊部分及序列資料部分。序列表必須使用附錄II所提供的文檔類型定義作為單個XML文件。基本資訊部分包含書目資訊，僅用於將序列表與提交序列表的專利申請關聯。序列資料部分由1個或多個序列資料要素所組成，各序列資料要素包含一序列的資訊。序列資料要素包含根據國際核苷酸序列資料庫協作組織(International Nucleotide Sequence Database Collaboration, INSDC)及UniProt規範之各種特徵關鍵(feature keys)及其接續的限定詞(subsequent qualifiers)。
7. 就本標準而言，要求包含於序列表的序列，係指在專利申請以殘基列舉揭露的序列，其可表示為：
 - (a) 包含10個或更多具體定義核苷酸的非分支序列或分支序列的線性區域，其中相鄰核苷酸透過以下方式連接：
 - i. 3'到5'(或5'到3')磷酸二酯鍵。
 - ii. 任何化學鍵，其使相鄰核鹼基以模擬天然存在核酸中核鹼基的排列方式排列。
 - (b) 包含4個或更多具體定義胺基酸的非分支序列或分支序列的線性區域，其中胺基酸會形成單一胜肽骨架，即相鄰胺基酸透過肽鍵連接一起。
8. 序列表中分配有其專屬序列識別號的序列，不得包含任何少於10個具體定義核苷酸或少於4個具體定義胺基酸的序列。

參考

9. 以下標準及資源與本標準有相關性：

國際核苷酸序列資料庫協作組織	http://www.insdc.org/
國際標準 ISO 639-1:2002	語言名稱表示代碼-第1部分：Alpha-2代碼

UniProt聯盟	http://www.uniprot.org/
W3C XML 1.0	http://www.w3.org/
WIPO ST.2標準	以公曆(Gregorian calendar)指定曆法日期的標準方式
WIPO ST.3標準	關於代表國家、其他實體及政府間組織之雙字母代碼的建議標準
WIPO ST.16標準	識別各類專利文獻之建議標準代碼
WIPO ST.25標準	專利申請的核苷酸及胺基酸序列表記載標準

序列表示

10. 根據第7段所涵蓋的各序列(包含及1個長序列之某個區域完全相同的序列)都必須分配有其獨立的序列識別號。序列識別號必須以數字1開頭，並按整數連續遞增編碼。若該序列識別號下沒有序列，即指刻意跳過序列，則必須使用「000」代替序列(參見第58段)。序列表中必須指明總序列數，並且總序列數必須等於序列識別號的總數，無論序列識別號下是有1序列或是「000」。

核苷酸序列

11. 核苷酸序列必須僅以單一鏈從左到右、以5'到3'方向，或者從左到右、模擬5'到3'的方向表示，並且序列中不得包含5'及3'的標示或任何其他類似標示。透過列舉2條鏈的殘基來揭露雙鏈核苷酸序列，並且必須依下述方式表示：
- 彼此完全互補的2條鏈，可作為單一序列表示，或者是作為2個獨立序列，各序列皆分配其專屬序列識別號，或
 - 彼此不完全互補2條鏈，必須視為2個獨立的序列，各序列皆分配其專屬序列識別號。
12. 就本標準而言，序列中出現的第1個核苷酸是第1個殘基，若核苷酸序列呈環狀構形時，申請人必須選擇作為第1個殘基的核苷酸。整個序列中是從5'到3'或模擬5'到3'的方向連續編號，最後1個殘基位點編號必須等於序列中的核苷酸總數。
13. 序列中的所有核苷酸必須使用附錄I中規定的符號來表示(參見第1節表1)，且只能使用小寫字母。任何用以表示核苷酸的符號都只相當於1個殘基。

14. 符號「t」在DNA中被解釋為胸腺嘧啶，在RNA中被解釋為尿嘧啶。DNA中的尿嘧啶或RNA中的胸腺嘧啶則被視為修飾核苷酸，必須在第19段提供的特徵表中進一步描述。
15. 當需要使用歧義符號(表示2個或多個替代核苷酸)，建議使用最嚴格的符號，如附錄I(第1節表1)中所列舉者。例如，若指定位置的核苷酸可為「a」或「g」，則建議使用「r」而非「n」。除非在特徵表中進一步描述，否則符號「n」被解釋為「a」、「c」、「g」或「t/u」中之任一者。符號「n」不得用於表示核苷酸以外的其他物。如第16、17、21或93-96段中所涵蓋者，單一修飾或「未知」核苷酸可以符號「n」表示，並且在特徵表中進一步描述。有關序列變體的表示，包含替代(alternatives)、刪除(deletions)、插入(insertions)或置換(substitutions)，請參見第93至100段。
16. 修飾核苷酸建議在序列中儘可能以對應之未修飾核苷酸表示，即「a」、「c」、「g」或「t」。序列中任何不能用附錄I中之任何其他符號表示的修飾核苷酸(參見第1節表1)，例如，「其他(other)」核苷酸，指非天然存在的核苷酸，必須以符號「n」表示，且符號「n」只相當於1個殘基。
17. 修飾核苷酸必須在特徵表中進一步描述(參見第60段及以下)，使用特徵關鍵「modified_base」及必要限定詞「mod_base」，以及單一個附錄I中的縮寫(參見第2節表2)；若縮寫為「OTHER」，則必須提供修飾核苷酸的完整非縮寫名稱作為限定詞「note」的值(value)。若要列出替代殘基的修飾核苷酸，可以將限定詞值「OTHER」進一步與限定詞「note」結合使用(參見第97及98段)。上述附錄I(參見第2節表2)提供的縮寫詞(或全名)不得用於序列本身。
18. 核苷酸序列包含1個或多個共用相同骨架的連續修飾核苷酸區域(參見第3(g)(i)(2)段)，必須依第17段所述方式在特徵表中進一步描述。此區域的修飾核苷酸可以根據第22段的規定，在一個INSDFeature要素中聯合描述，而限定詞「note」下必須使用一個涵蓋該範圍內所有修飾核苷酸的最嚴格、非縮寫的化學名稱，或列出該範圍內所有核苷酸之最嚴格、非縮寫的化學名稱作為該限定詞值。例如，含有「a」、「c」、「g」或「t」核鹼基的乙二醇核酸序列，可以在限定詞「note」中描述為「2,3-二羥丙基核苷(2,3-dihydroxypropyl nucleosides)」；又或者，可以在限定詞「note」中描述為「2,3-二羥

丙基腺嘌呤(2, 3-dihydroxypropyladenine)、2, 3-二羥基丙基胸腺嘧啶(2, 3-dihydroxypropylthymine)、2, 3-二羥丙基鳥嘌呤(2, 3-dihydroxypropylguanine)或2, 3-二羥丙基胞嘧啶(2, 3-dihydroxypropylcytosine)；若該區域中的單個修飾核苷酸包含有額外修飾，還必須依第17段所述方式在特徵表中進一步描述該額外修飾之核苷酸。

19. DNA中的尿嘧啶或RNA中的胸腺嘧啶被視為修飾核苷酸，必須在序列中表示為「t」，並在特徵表中使用特徵關鍵「modified_base」、限定詞「mod_base」及限定詞「note」進一步描述，其中限定詞「mod_base」的值使用OTHER，限定詞「note」的值則分別使用「尿嘧啶(uracil)」或「胸腺嘧啶(thymine)」。

20. 根據上述第16至18段，以下實例說明修飾核苷酸的表示方式：

實例1：使用附錄I中的縮寫來表示修飾核苷酸(參見第2節表2)

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

實例2：使用附錄I中之「OTHER」來表示修飾核苷酸(參見第2節表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>
```

實例3：一核苷酸序列，由第3(g)(i)(2)段中所涵蓋修飾核苷酸所組成，並且包含2個具進一步修飾的單獨核苷酸

```

<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. 任何「unknown」核苷酸必須在序列中用符號「n」表示。一個「unknown」核苷酸建議在特徵表(參見第60段及以下)中使用特徵關鍵「unsure」進一步描述。符號「n」只相當於1個殘基。
22. 一包含已知數量的連續「a」、「c」、「g」、「t」或「n」殘基且適用相同描述的區域，可以使用單一INSDFeature要素進行聯合描述，並使用語法「x.y」作為INSDFeature_location要素的位置描述(參見第64至71段)。另有關序列變體(即替代、刪除、插入或置換)的表示，請參見第93至100段。
23. 根據上述第22段，以下實例說明修飾核苷酸區域的表示方式：

```

<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>

```

胺基酸序列

24. 胺基酸序列中的胺基酸殘基必須從左到右，以胺基到羧基的方向表示。序列中不得出現胺基及羧基。
25. 就本標準而言，序列中的第1個胺基酸是第1個殘基，包含成熟蛋白(mature protein)之前的胺基酸，例如前序列(pre-sequences)、原序列(pro-sequences)、前原序列(pre-pro-sequences)及訊息序列(signal sequences)。當胺基酸序列呈環狀構形，且僅由透過肽鍵連接胺基酸殘基以組成環狀時，即該序列沒有胺基及羧基末端，申請人必須選擇作為第1個殘基位置的胺基酸，且將整個序列從胺基到羧基方向連續編號。
26. 序列中所有胺基酸必須使用附錄I中規定的符號表示(參見第3節表3)，只能使用大寫字母，且用於表示胺基酸的任何符號只相當於1個殘基。
27. 當需要使用歧義符號(ambiguity symbol)(代表2種或多種替代胺基酸)時，建議使用最嚴格的符號，如附錄I(第3節表3)中所列。例如，若指定位置的胺基酸可為天冬胺酸或天冬醯胺酸，建議使用符號「B」，而非「X」，除非在特徵表中進一步描述，因為符號「X」被解釋為「A」、「R」、「N」、「D」、「C」、「Q」、「E」、「G」、「H」、「I」、「L」、「K」、「M」、「F」、「P」、「O」、「S」、「U」、「T」、「W」、「Y」或「V」中之任一者。符號「X」不得用於表示胺基酸以外的其他物。單一修飾或「未知」胺基酸可以符號「X」表示，並在特徵表中進一步描述，如第29、30、32或93-98段所涵蓋。有關序列變體(如替代、刪除、插入或置換)的表示，請參見第93至100段。
28. 以內部終止符號(例如用「Ter」、星號「*」、句點「.»或空格)作分隔所揭露的胺基酸序列，若各分段序列包含至少4個具體定義胺基酸且為第7段所涵蓋，必須在序列表中表示個別分段序列，並須分配其專屬序列識別號。終止符號及空格不得包含在序列表的序列中(參見第57段)。

29. 修飾胺基酸，包含D-胺基酸，在序列中建議儘可能以對應的未修飾胺基酸表示。序列中任何不能用附錄I修飾胺基酸(參見第3節表3)表示者，即「其他(other)」胺基酸，必須用「X」表示。符號「X」只相當於1個殘基。
30. 修飾胺基酸必須於特徵表中進一步描述(參見第60段及以下)。在可能情況下，建議特徵關鍵「CARBOHYD」或「LIPID」與限定詞「note」併用。對於轉譯後修飾的胺基酸，建議特徵關鍵「MOD_RES」與限定詞「note」併用；除此以外的修飾胺基酸，則建議特徵關鍵「SITE」及限定詞「note」併用。限定詞「note」的值必須為附錄I中規定的縮寫(參見第4節表4)，或為修飾胺基酸之完整非縮寫的名稱。上述表4中所列出的縮寫或完整非縮寫的名稱不得在序列本身中使用。
31. 根據上述第30段，以下實例說明修飾的胺基酸的表示方式：

實例1：轉譯後修飾的胺基酸

```
<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>
```

實例2：非轉譯後修飾的胺基酸

```
<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>
```

實例3：D-胺基酸

```
<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. 任何「unknown」胺基酸必須在序列中以符號「X」表示。指定為「X」的「unknown」胺基酸必須在特徵表中使用特徵關鍵「UNSURE」及非必要限定詞「note」進一步描述(參見第60段及以下)。符號「X」只相當於1個殘基。
33. 根據上述第32段，以下實例說明「unknown」胺基酸的表示方式：

```
<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. 一包含已知數量的連續「X」殘基，且適用相同描述的區域，可以使用語法「x.y」作為INSDFeature_location要素中的位置描述(參見第64至70段)。有關序列變體(如替代、刪除、插入或替換)的表示，請參見第93至100段。

特殊情況表示

35. 一個以列舉其殘基所揭露的序列，該序列是由1個或多個不相鄰片段構成的單一連續序列，並且前述片段是較長的序列或來自於不同序列的片段，則該序列必須包含在序列表中並分配其專屬序列識別號。
36. 一個包含數個具體定義殘基區域的序列，區域間由1個或多個連續「n」或「X」殘基分隔開來(分別參見第15及27段)，且每個區域中之「n」或「X」殘基的確切數目被揭露，則該序列必須作為單獨序列包含在序列表中，並分配其專屬序列識別號。
37. 一個包含數個具體定義殘基區域的序列，且區域間由1個或多個未知或未揭露數量的殘基分隔開來，則該序列不得作為單獨序列表示在序列表中。但每個具體定義殘基區域只要符合第7段要求，都必須作為單獨序列包含在序列表中，並分配其專屬序列識別號。

XML序列表結構

38. 根據上述第6段，符合本標準的XML序列表文件包含以下內容：

- (a) 基本資訊部分，包含序列表所對應的專利申請資訊；及
 - (b) 序列資料部分，包含1或多個序列資料要素，各要素包含1個序列資訊。
- 序列表實例如附錄III所示。

39. 序列表必須以XML 1.0格式呈現，並使用附錄II「序列表的文檔類型定義」。

- (a) XML實例的第1行必須包含XML聲明：

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

- (b) XML實例的第2行必須包含文檔類型(document type, DOCTYPE)聲明：

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

40. 整個電子序列表必須包含在1個文件中，該文件必須使用Unicode UTF-8編碼，並有以下限制：

- (a) 基本資訊部分的ApplicantName要素、InventorName要素及InventionTitle要素中包含的資訊，以及序列資料部分的 NonEnglishQualifier_value要素中包含的資訊可以由XML 1.0規範中指定的任何有效Unicode字元組成，但Unicode控制代碼點(control code points) 0000-001F及007F-009F除外。保留字元(reserved characters) “&”、<及> (分別為Unicode代碼點(code points) 0022、0026、0027、003C及003E)，必須按第41段的規定進行替換；及
- (b) 基本資訊部分及序列資料部分的所有其他要素及屬性所包含的資訊，必須由可列印字元(包含空格字元)組成，且這些字元必須來自Unicode基本拉丁字元表(Basic Latin code table)(即僅限於Unicode代碼點0020到007E—參見附錄IV)。保留字元 “&”、<及>(分別為Unicode代碼點0022、0026、0027、003C及003E)必須按第41段中的規定進行替換。

41. 在XML序列表的實例中，不得使用數值字元參考(numeric character references)¹，並且在屬性值或要素值中使用以下保留字元時，必須用對應的預定義實體(predefined entities)替換：

保留字元	預定義實體
<	<
>	>

¹ 數值字元參考指的是以其Universal Character Set/Unicode code point之字元，使用格式：“&#nnnn;”或“&#xhhhh;”，其中“nnnn”為10進位制之代碼點，而“hhhh”為16進位制之代碼點。

&	&
“	"
’	'

參見第71段例示。唯一允許字元是本段中規定的預定義實體。

42. 在XML實例中，所有必要要素(mandatory elements)皆必須填寫(除第58段中規定的刻意跳過序列外)。對於無法得知內容的非必要要素則建議不要包含在XML實例中(除非是第97段規定在限定詞「replace」的值表示序列中的刪除)。

根要素(Root element)

43. 根據本標準，XML實例的根要素是ST26SequenceListing要素，具有以下屬性：

屬性	描述	必要(Mandatory)/ 非必要(Optional)
dtdVersion	此文件所使用的DTD版本格式為「V#_#」，例如「V1_3」	必要
fileName	序列表文件名稱	非必要
softwareName	生成此文件的軟體名稱	非必要
softwareVersion	生成此文件的軟體版本	非必要
productionDate	序列表文件的製作日期(格式「CCYY-MM-DD」)	非必要
originalFreeTextLanguageCode	語言依賴性自由文本限定詞所涉及的單一原始語言的語言代碼(參見第9段中的參考資料ISO 639-1:2002)	非必要
nonEnglishFreeTextLanguageCode	NonEnglishQualifier_value要素中的語言代碼(參見第9段中的參考資料ISO 639-1:2002)	當序列表中存在NonEnglishQualifier_value要素時，則為必須

44. 根據上述第43段的XML實例，以下實例說明ST26SequenceListing根要素及其屬性：

```
<ST26SequenceListing dtdVersion="V1_3" fileName="US11_405455_SEQ1.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="de" nonEnglishFreeTextLanguageCode="fr">
  {...}*
</ST26SequenceListing>
```

*{...} 表示例示中沒有包含基本資訊部分及序列資料部分。

基本資訊部分

45. 基本資訊部分涉及專利申請資訊，如下：

要素	描述	必要/非必要
ApplicationIdentification	檢送序列表的申請識別號(application identification)	在取得申請號(application number)之後提交序列表時，則為必要
ApplicationIdentification由以下部分組成：		
IPOfficeCode	申請專利局的ST. 3代碼	必要
ApplicationNumberText	申請專利局提供的專利申請號(例如，PCT/IB2013/099999)	必要
FilingDate	檢送序列表的專利申請案的申請日(ST. 2格式「CCY Y-MM-DD」，使用4位日曆年、2位日曆月及2位數字日，例如2015-01-31)	FilingDate
ApplicantFileReference	由申請人指定的單一特定識別號，按第40(b)段所規定的字元表示，用以識別其特定的申請	當序列表是在取得專利申請號之前的任何時間提交時，則為必要；否則，非必要
EarliestPriorityApplicationIdentification	最早優先申請號(同時包含IPOfficeCode、ApplicationNumberText及FilingDate，參見上述ApplicationIdentification)	主張優先權時，則為必要
ApplicantName	按第40(a)段規定的字元表示的第一位提及申請人姓名。該要素包含第47段所規定的必要屬性languageCode	必要
ApplicantNameLatin	若ApplicantName以第40(b)段規定以外的字元表示，則必須以第40(b)段規定的字元表示第一位申請人姓名的翻譯或音譯	當ApplicantName包含非拉丁文字元時，則為必要

要素	描述	必要/非必要
InventorName	按第40(a)段規定的字元表示的第一位提及發明人姓名。該要素包含第47段規定的必要屬性language Code	非必要
InventorNameLatin	若InventorName以第40(b)段規定以外的字元表示，亦可以第40(b)段中規定的字元表示第一位發明人姓名翻譯或音譯	非必要
InventionTitle	以申請文件所用語言，按第40(a)段規定的字元表示的發明名稱。將發明名稱翻譯成額外語言，可以使用額外的InventionTitle要素，按第40(a)段規定的字元表示。該要素包含第48段所規定的必要屬性languageCode。發明名稱建議介於2到7個字之間	在申請文件所用語言中，是必要。 在額外使用語言中，則為非必要
SequenceTotalQuantity	序列表中所有序列的總數，包含刻意跳過序列(也稱為空序列)(參見第10段)	必要

46. 根據上文第45段，以下實例說明序列表的基本資訊部分：

實例1：在取得申請號及申請日之前提交的序列表

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="Invention_SEQ1.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="ja">
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2013/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode="en">Keiko Nakamura</InventorName>
  <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
```

```

<SequenceTotalQuantity>9</SequenceTotalQuantity>
<SequenceData sequenceIDNumber="1"> {...}* </SequenceData>
<SequenceData sequenceIDNumber="2"> {...} </SequenceData>
<SequenceData sequenceIDNumber="3"> {...} </SequenceData>
<SequenceData sequenceIDNumber="4"> {...} </SequenceData>
<SequenceData sequenceIDNumber="5"> {...} </SequenceData>
<SequenceData sequenceIDNumber="6"> {...} </SequenceData>
<SequenceData sequenceIDNumber="7"> {...} </SequenceData>
<SequenceData sequenceIDNumber="8"> {...} </SequenceData>
<SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>

```

*{...} 表示在例示中沒有包含各序列的相關資訊。

實例2：在取得申請號及申請日之後提交的序列表

```

<?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>*

```

*{...} 表示在例示中沒有包含各序列的相關資訊。

47. 申請人姓名及(非必要的)發明人姓名必須分別在ApplicantName要素及InventorName要素中表示，且通常以申請時所用語言表示。各要素的languageCode屬性中必須表示適當的語言代碼(參見第9段中的參考資料ISO 639-1:2002)。若申請人名稱包含第40(b)段所規定的拉丁文字母以外的字元，也必須在ApplicantNameLatin要素中，以拉丁文字母的字元表示申請人名稱的音譯或翻譯。若發明人姓名包含拉丁文字母以外的字元，也可在InventorNameLa

tin要素中，以拉丁文字母的字元表示發明人姓名的音譯或翻譯。

48. 發明名稱必須在InventionTitle要素中以申請時所用的語言表示，也可使用多個InventionTitle要素，並以其他語言表示(參見第45段表格)，必須在languageCode要素的屬性中表示適當的語言代碼(參見第9段中的參考資料ISO 639-1:2002)。

49. 根據上述第47及48段，以下例示說明發明名稱及姓名的表示方式：

實例：申請人姓名及發明人姓名分別以日文及拉丁文字母表示，發明名稱以日文、英文及法文表示

```
<ApplicantName languageCode="ja">出願製藥株式会社</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>
```

序列資料部分

50. 序列資料部分必須由1個或多個SequenceData要素組成，各要素包含1個序列的資訊。

51. 各個SequenceData要素必須有1個必要屬性sequenceIDNumber，表示序列識別號(參見第10段)。例如：

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

52. SequenceData要素必須包含一INSDSeq從屬要素(dependent element)，其是由以下更多的從屬要素組成：

要素	描述	必要/不包含	
		序列	刻意跳過序列
INSDSeq_length	序列長度	必要	必要 無值
INSDSeq_moltype	分子類型	必要	必要 無值
INSDSeq_division	與專利申請相關的序列	必要 值「PAT」	必要 無值
INSDSeq_feature-table	序列註釋列表	必要	不得有
INSDSeq_sequence	序列	必要	必要 有值，記載「000」

53. INSDSeq_length要素必須揭露INSDSeq_sequence要素中序列的核苷酸或胺基酸數量。例如：

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

54. INSDSeq_moltype要素必須揭露分子類型。對於核苷酸序列(包含核苷酸類似物)，分子類型必須表示為DNA或RNA；對於胺基酸序列，分子類型必須表示為AA(該要素不同於第55及84段中的限定詞「mol_type」)。例如：

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

55. 對於1個同時包含1或多個核苷酸的DNA及RNA片段的核苷酸序列，其分子類型必須被表示為DNA。該結合DNA/RNA分子必須以特徵關鍵「source」、必要限定詞「organism」(其值為「synthetic construct」)以及必要限定詞「mol_type」(其值為「other DNA」)，在特徵表中進一步描述。該結合DNA/RNA分子中的各個DNA及RNA片段，必須以特徵關鍵「misc_feature」及限定詞「note」在特徵表中進一步描述，以表示該片段是DNA或是RNA。

56. 根據上述第55段，以下實例說明一個同時包含DNA及RNA片段之核苷酸序列的表示方式：

```

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

```

57. INSDSeq_sequence要素必須揭露序列，序列中必須只能包含附錄I中規定的適當符號(參見第1節表1及第3節表3)，不得包含數字、標點符號或空白字元。
58. 刻意跳過序列必須包含在序列表中，並且表示如下：
- (a) SequenceData要素及其sequenceIDNumber屬性以刻意跳過序列的序列識別號為其值；
 - (b) INSDSeq_length要素、INSDSeq_moltype要素及INSDSeq_division要素必須存在，但無值；
 - (c) 不得包含INSDSeq_feature-table要素；以及
 - (d) INSDSeq_sequence要素以字元串「000」為其值。

59. 根據上述第58段，以下實例說明刻意跳過序列的表示方式：

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

特徵表 (Feature table)

60. 特徵表包含一特定序列中各個區域的位置及作用的相關資訊。除刻意跳過序列不得包含有特徵表外，各個序列都必須有特徵表。特徵表包含在INSDSeq_feature-table要素中，該要素由1個或多個INSDFeature要素組成。

61. 各INSDFeature要素描述1個特徵，並且由下列相關要素組成：

要素	描述	必要/非必要
INSDFeature_key	特徵的單詞或縮寫	必要
INSDFeature_location	特徵對應的序列區域	必要
INSDFeature_qual	包含有關特徵的輔助資訊的限定詞	當特徵關鍵需要1個或多個限定詞(例如,來源)時,則為必要;否則,則非必要

特徵關鍵 (Feature keys)

62. 附錄I包含本標準下必須使用的特徵關鍵專用列表(exclusive listing)以及連同相關限定詞專用列表，並且指明這些限定詞為必要或是非必要。附錄I第5節提供核苷酸序列的特徵關鍵專用列表，而第7節提供胺基酸序列的特徵關鍵專用列表。

必要特徵關鍵 (Mandatory feature keys)

63. 對於所有核苷酸序列及所有胺基酸序列，除刻意跳過序列外，都必須記載特徵關鍵「source」。各個序列必須有1個跨越整個序列的「source」特徵關鍵。當1個序列源自於多個來源時，這些來源可以在特徵表中進一步描述，對於核苷酸序列使用特徵關鍵「misc_feature」及限定詞「note」，對於胺基酸序列則是使用特徵關鍵「REGION」及限定詞「note」。

特徵位置 (Feature location)

64. INSDFeature_location必要要素必須包含至少1個位置描述(location descriptor)，其定義對應INSDDseq_sequence要素中之序列的1個位點或1個區域。對於胺基酸序列，INSDFeature_location必要要素必須包含1個(且只能1個)位置描述。對於核苷酸序列，當1個或多個位置運算符(<及>符號)結合使用時，INSDFeature_location必要要素可以包含多於1個位置描述(參見第67至70段)。
65. 位置描述可以是單一殘基編號、連續殘基編號範圍的區域、或是特定殘基或特定殘基區間並且延伸到特定殘基之外的位點或區域。位置描述不得超出INSDDseq_sequence要素中序列的殘基編號範圍。僅限在核苷酸序列的位置描述，可以是2個相鄰殘基編號之間的位點。當特徵是對應到核苷酸序列的不連續位點或區域時，則多個位置描述必須結合使用位置運算符(<及>符號)結合使用(參見第67至70段)。
66. 位置描述的語法如下表所示，其中x及y是殘基編號，以正整數表示，且不大於INSDDseq_sequence要素中序列長度，且x小於y。

(a) 核苷酸及胺基酸序列的位置描述：

位置描述類型	句法	描述
單一殘基編號	x	序列中的單一殘基
殘基編號界定序列區間(span)	x..y	連續殘基範圍並且包含起始及終止殘基
特定殘基編號之前或之後、特定殘基編號區間的起始殘基之前、末端殘基之後	<x >x <x..y x..>y <x..>y	特定殘基或特定殘基區間並且延伸到指定殘基之外的區域。「<及>」符號可以與單一殘基或是殘基區間的起始及終止殘基編號併用，以表示超出特定殘基編號的特徵

(b) 僅適用於核苷酸序列的位置描述：

位置描述類型	句法	描述
2個相鄰核苷酸之間的位點	x^y	2個相鄰核苷酸之間的位點，例如核酸內切裂解位點。相鄰核苷酸位點編號以插入符號(^)作分隔。表示為x^x+1(例如55^56)，或者，對於環狀核苷酸則表示為

		x^1 ，其中「x」是分子全長，例如，對於全長為1000鹼基的環狀分子，表示為 1000^1
--	--	--

(c) 僅適用於胺基酸序列的位置描述：

位置描述類型	句法	描述
透過鏈內交聯(cross-link)連接的殘基編號	x..y	當使用一鏈內交聯的特徵(例如「CROSSLINK」或「DISULFID」)時，表示是透過鏈內交聯以連接胺基酸

67. 核苷酸序列的INSDFeature_location要素可包含1個或多個位置運算符(join, order, complement)。位置運算符是前綴(prefix)，可用於單一位置描述(descriptor)，或是對應單一但不連續特徵的位置描述的組合(a combination of location descriptors)，並且指定所示序列上與特徵對應的位置，或者是如何構建該特徵。以下提供位置運算符列表及其定義，該等位置運算符僅適用於核苷酸。

位置語法	位置描述
join(location, location, ..., location)	所指示位置被連接(首尾相連)以形成一連續序列
order(location, location, ..., location)	以指示順序排列，但並不意謂是否合理
complement(location)	當以5'到3'方向或是模擬5'到3'方向讀取時，表示該特徵是與位置描述指定的序列區間互補

68. 連接(join)及排序(order)位置運算符要求至少提供2個以逗號分隔的位置描述。連接或排序位置不得使用2個相鄰殘基的間位點的位置描述(即 x^y)。使用連接位置運算符意味著所描述位置的殘基透過生物學過程(例如，編碼區域的外顯子)進行物理連接。

69. 互補(complement)可與連接或排序結合在同一位置上使用。但不得在同一位置上使用連接及排序的組合。

70. 根據上述第64至69段，以下例示說明特徵位置：

(a) 核苷酸及胺基酸序列的位置：

位置例示	描述
467	序列中的第467個殘基位點
340..565	由第340個及第565個殘基界定之一段包含該兩個殘基的連續殘基範圍
<1	特徵位置在第1個殘基之前
<345..500	特徵的正確下界點未知。該位置開始在第345個殘基之前的某個殘基，並且延伸到第500個殘基
<1..888	由第1個殘基之前開始並且延伸到第888個殘基
1..>888	由第1個殘基開始並且延伸到第888個殘基之後
<1..>888	由第1個殘基之前開始並且延伸到第888個殘基之後

(b) 僅適用於核苷酸序列的位置表示：

位置例示	描述
123^124	第123及124個殘基之間的位點
join(12..78, 134..202)	表示第12到78個區域以及第134到202個區域連接起來形成一連續序列
complement(34..126)	起始位置互補於第126個核苷酸，並且終止位置互補於第34個核苷酸（該特徵位在與所示鏈互補的鏈上）
complement(join(2691..4571, 4918..5163))	連接第2691到4571個及第4918到5163個核苷酸，然後與該連接的片段互補（該特徵位在與所示鏈互補的鏈上）
join(complement(4918..5163), complement(2691..4571))	與第4918到5163個核苷酸的區域互補及與第2691到4571個核苷酸的區域互補，並且將互補片段連接起來（該特徵位在與所示鏈互補的鏈上）

(c) 僅適用於胺基酸序列的位置表示：

位置例示	描述
340..565	當使用鏈內交聯特徵(例如「CROSSLINK」或「DISULFID」)時，表示第34

	0個殘基及第565個殘基透過鏈內交聯連接一起
--	------------------------

71. 在XML序列表實例中，位置描述中的字元「<」及「>」必須替換為對應的預定義實體(參見第41段)。例如：

特徵位置「<1」:

```
<INSDFeature_location>&lt;1</INSDFeature_location>
```

特徵位置「1..>888」:

```
<INSDFeature_location>1..&gt;888</INSDFeature_location>
```

特徵限定詞(Feature qualifiers)

72. 限定詞是除了特徵關鍵及特徵位置所表示的資訊之外，用於補充有關特徵的資訊。限定詞所表示的不同類型資訊，適用以下3種類型的值格式：

- (a) 自由文本(參見第85至87段)；
- (b) 受控詞彙或列舉數值(例如，數字或日期)；以及
- (c) 序列。

73. 附錄I第6節提供用於核苷酸序列特徵關鍵的限定詞以及其指定值格式的專用列表，第8節提供用於胺基酸序列特徵關鍵的限定詞的專用列表。

74. 以限定詞值提供且被第7段涵蓋之任何序列，必須單獨包含在序列表中，並分配有其專屬序列識別號(參見第10段)。

必要特徵限定詞(Mandatory feature qualifiers)

75. 對於核苷酸序列及胺基酸序列，一個必要特徵關鍵，如「來源」，需要2個必要限定詞「生物體(organism)」及「分子類型(mol_type)」。有些非必要特徵關鍵也是需要必要限定詞。

限定詞要素(Qualifier elements)

76. INSDFeature_qual要素包含1個或多個INSDQualifier要素，各個INSDQualifier要素表示一限定詞，並且由以下3個從屬要素組成：

要素	描述	必要/非必要
INSDQualifier_name	限定詞的名稱(參見附錄 I 第6及8節)	必要
INSDQualifier_value	若有的話, 限定詞的值以特定格式(參見附錄 I 第6節及第8節)表示, 並且按第40(b)段規定的字元組成	當有特定時, 則為必要(參見第87段及附錄 I 第6節及第8節)
NonEnglishQualifier_value	若有的話, 限定詞的值以特定格式(參見附錄 I 第6節及第8節)表示, 並且按第40(a)段規定的字元組成	當有特定時, 則為必要(參見第87段及附錄 I 第6節及第8節)

77. 生物體限定詞, 如核苷酸序列的「organism」(參見附錄I第6節)及胺基酸序列的「organism」(參見附錄I第8節)必須揭露序列的來源, 例如, 單一生物體或起源(origin)。生物體名稱建議從分類學數據庫(taxonomy database)中選擇。

78. 若序列是天然存在且來源生物體有其指定的拉丁屬名及種名, 則該指定名稱必須用作限定詞的值。對於核苷酸序列及胺基酸序列, 可以使用限定詞「note」指定優選的英文通用名稱, 但該通用名稱不得作為生物體限定詞的值。

79. 根據上述第77及78段, 以下實例說明序列來源的表示方式:

實例1: 核苷酸序列來源

```
<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>
```

實例2：胺基酸序列來源

```
<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. 若序列為天然存在，且來源生物有一已知的拉丁屬名，但未指定或未識別其物種，則生物體限定詞值必須表示為拉丁屬名且其後接「sp.」。例如：

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

81. 若序列是天然存在的，但拉丁生物屬名及種名未知，則生物體限定詞值必須表示為「unidentified」。任何已知的分類學資訊建議在核苷酸序列的限定詞「note」及胺基酸序列的限定詞「note」中表示。例如：

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

82. 若序列是天然存在的，並且來源生物體沒有拉丁屬名及種名，例如病毒，則必須使用另一可接受的學名(例如，「Canine adenovirus type 2」)作為生物體限定詞的值。例如：

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

83. 若序列不是天然存在的，則生物體限定詞的值必須表示為「synthetic construct」。對於該等核苷酸序列及胺基酸序列，可以使用限定詞「note」來表示有關序列生成方法的進一步資訊。例如：

```

<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. 核苷酸序列的限定詞「mol_type」(參見附錄I第6節)及胺基酸的限定詞「mol_type」(參見附錄I第8節)必須表示序列分子類型。該等限定詞不同於第54段中討論的INSDSeq_moltype要素：

(a) 對於核苷酸序列，限定詞「mol_type」的值必須為以下之一：

「genomic DNA」，「genomic RNA」、「mRNA」、「tRNA」、「rRNA」、「other RNA」、「other DNA」、「transcribed RNA」、「viral cRNA」、「unassigned DNA」或「unassigned RNA」。若序列不是天然存在的，即限定詞「organism」的值為「synthetic construct」，限定詞「mol_type」的值必須是「other RNA」或「other DNA」；

(b) 對於胺基酸序列，限定詞「mol_type」的值為「protein」。

自由文本(*free text*)

85. 如第3(n)段所述，自由文本是一種用於某些限定詞的值格式，以描述性文字短語(descriptive text phrase)形式或其他指定格式形式表示(如附錄I所示)。

86. 自由文本的使用必須限制在少數個對於理解序列特徵不可或缺的短語(short terms)。對於每個限定詞，自由文本不得超過1000個字元。

87. 如第3(o)段所述，語言依賴自由文本是某些語言依賴限定詞的自由文本值，該值在國家、區域或國際程序上可能要求翻譯。附錄I第6節表5表示，核苷酸序列具有語言依賴的自由文本值格式的限定詞。附錄I第8節表6則表示，胺基酸序列具有語言依賴的自由文本值格式的限定詞。

- (a) 語言依賴自由文本必須在INSDQualifier_value要素中以英語表示，或在NonEnglishQualifier_value要素中以非英語語言表示，或是兩者皆呈現。對於生物名稱的拉丁屬名及種名，則不需要翻譯。國際上使用源自非英語單詞的技術術語及專有名稱，被視為英語，並且可以在INSDQualifier_value要素的值中使用(例如「in vitro」、「in vivo」)。
- (b) 若在序列表中存在NonEnglishQualifier_value要素，則必須在根要素的nonEnglishFreeTextLanguageCode屬性中表示適當的語言代碼(參見第9段中的參考資料ISO 639-1:2002)(參見第43段)。一序列表中所有NonEnglishQualifier_value要素的值，必須使用nonEnglishFreeTextLanguageCode 屬性中所表示語言。NonEnglishQualifier_value要素只允許用於具有語言依賴自由文本值格式的限定詞。
- (c) 若NonEnglishQualifier_value及INSDQualifier_value都存在於一限定詞中，這2個要素中包含的資訊必須相符。符合以下條件之一：NonEnglishQualifier_value包含INSDQualifier_value值的翻譯；或者，INSDQualifier_value包含NonEnglishQualifier_value值的翻譯；或者，2個要素都包含由originalFreeTextLanguageCode屬性指定的語言所翻譯的限定詞的值(請參閱第43段)。
- (d) 對於語言依賴限定詞，INSDQualifier要素可以包含1個非必要id 屬性。此屬性的值必須採用「q」後跟正整數(例如「q23」)的格式，並且只能專屬於一個INSDQualifier要素，即屬性的值在序列表文件中只能使用1次。

88. 根據上述第87段，以下實例說明語言依賴自由文本的表示方式

實例1：INSDQualifier_value要素中的語言依賴自由文本：

```
<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>
```

實例2：INSDQualifier_value要素及NonEnglishQualifier_value要素中的語言依賴自由文本：

```

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

```

實例3：NonEnglishQualifier_value要素中的語言依賴自由文本：

```

<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>

```

編碼序列

89. 特徵關鍵「CDS」可用於表示編碼序列，即與蛋白質中胺基酸序列對應的核苷酸序列及終止密碼子。INSDFeature_location必要元素對於特徵關鍵「CDS」的位置表示必須包括終止密碼子。
90. 限定詞「transl_table」及「translation」可與特徵關鍵「CDS」一起使用（參見附錄I）。在不使用限定詞「transl_table」的情況下，推定使用標準密碼表(Standard Code Table)(參見附錄I，第9節，表7)。
91. 限定詞「transl_except」必須與特徵關鍵「CDS」及限定詞「translation」一起使用，以表示1個編碼吡咯離胺酸(pyrrolysine)或硒半胱胺酸(selenocysteine)的密碼子。
92. 在序列表中必須包含由編碼序列編碼並以第7段涵蓋的限定詞「translation」表示的胺基酸序列，並指定其專屬序列識別號。該胺基酸序列的序列識別號必須作為特徵關鍵「CDS」中限定詞「protein_id」的值。對應的胺基酸序列的特徵關鍵「source」中限定詞「organism」的值必須相符。例如：

```

<INSDFeature>
  <INSDFeature_key>CDS</INSDFeature_key>
  <INSDFeature_location>1..507</INSDFeature_location>
  <INSDFeature_qualifiers>
    <INSDQualifier>
      <INSDQualifier_name>transl_table</INSDQualifier_name>
      <INSDQualifier_value>1</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>translation</INSDQualifier_name>
      <INSDQualifier_value>MLVHLERTTIMDFSSLINLPLIWGLLIAIAVLLYLMDGFDLIGIGILL
      PFAPSDKCRDHMISSIAFFWDGNETWLVGGGGLFAAFPLAYSILMPAFYIPIIIMLLGLIVRGVSFEFR
      FKAEGKYRRLWDYAFHFGSLGA AFCQGMILGAFIHGVEVNGRNFSGGQLM
    </INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>protein_id</INSDQualifier_name>
      <INSDQualifier_value>89</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qualifiers>
</INSDFeature>

```

變體(Variants)

93. 一級序列及該序列的任何變體(每一個都藉由列舉殘基表示並包含在第7段中)，每個都必須包含在序列表中並分配其專屬序列識別號。
94. 單個序列在1個或多個位置上具有列舉替代殘基表示任何變體序列，必須包含在序列表中，並且應由單獨序列表示，其中列舉替代殘基由最嚴格的歧義符號表示(參見第15及27段)。
95. 序列表中一級序列的缺失、插入或取代表示的任何變體序列建議包括在序列表中。若包含在序列表中，此類變體序列：
 - (a) 若其在單個位置或多個不同位置包含變異，並且這些變異是獨立發生的，可以由一級序列的註釋表示；
 - (b) 若其包含多個不同位置的變異，並且這些變異的發生是相互依存的，應以單獨序列表示並指定其專屬序列識別號；及
 - (c) 若其含有超過1000個殘基的插入或取代序列，應以單獨序列表示並指定其專屬序列識別號(參見第86段)。
96. 下表說明核酸及胺基酸序列變體的特徵關鍵及限定詞的正確使用：

序列類型	特徵關鍵	限定詞	使用
核酸	variation	replace 或 note	自然發生的突變及多態性，例如等位基因(alleles)、RFLP

核酸	misc_difference	replace或note	人為引入的可變性，例如，透過基因操作或化學合成
胺基酸	VAR_SEQ	note	可變剪接(alternative splicing)產生的變體，替代啟動子使用，替代起始及核糖體框移
胺基酸	VARIANT	note	VAR_SEQ不適用的任何類型變體

97. 特定變體的序列註釋必須包括特徵關鍵及限定詞，如上表所示，以及特徵位置。限定詞「replace」的值必須是僅使用第1節表1中列符號的單個替代核苷酸或核苷酸序列，或者為空。可以提供替代殘基列表作為限定詞「note」中的值。特別是，當「X」用於序列中，必須提供替代胺基酸列表作為限定詞「note」中的值，其中並且表示不同於「A」、「R」、「N」、「D」、「C」、「Q」、「E」、「G」、「H」、「I」、「L」、「K」、「M」、「F」、「P」、「O」、「S」、「U」、「T」、「W」、「Y」或「V」中之任一者的值(參見第27段)。刪除必須表示為限定詞「replace」的值為空，或在限定詞「note」中指示殘基可能被刪除。插入或替換必須在限定詞「replace」或「note」中提供插入或替換的殘基。限定詞「replace」及「note」的值格式是自由文本，不得超過1000個字元，如第86段中所規定。請參見第100段中涵蓋於第7段的序列，這些序列具有插入或替換作為限定詞的值。

98. 附錄I中規定的符號(分別參見第1至4節表1至4)應在適當情況下用於表示變異殘基。對於限定詞「note」，如果變體殘基是附錄I表2或表4中未列出的修飾殘基，則必須提供修飾殘基的完整非縮寫名稱作為限定詞的值。必須在第17或30段提供的特徵表中進一步描述修飾殘基。

99. 根據上述第95至98段，以下實例說明變體的表示方式：

實例1：列舉替代核苷酸的特徵關鍵「misc_difference」。

序列第53位的「n」可以是5個替代核苷酸之一。

```

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

```

實例2：核苷酸序列缺失的特徵關鍵「misc_difference」。

該序列第413位的核苷酸被刪除。

```

<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>

```

實例3：用於插入核苷酸序列的特徵關鍵「misc_difference」。

序列「atgccaaatat」被插入到一級序列的第100及101位之間。

```

<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>

```

實例4：核苷酸序列中具有核苷酸替換的特徵關鍵「variation」。
胞嘧啶取代序列第413位的核苷酸。

```

<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>

```

實例5：胺基酸序列中具有胺基酸替換的特徵關鍵「VARIANT」。
序列第100位的胺基酸可以用I、A、F、Y、aIle、MeIle或Nle替換。

```

<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>

```

實例6：胺基酸序列中具有胺基酸替換的特徵關鍵「VARIANT」。

序列第100位的胺基酸可以用除Lys、Arg或His之外的任何胺基酸替換。

```
<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. 在第7段中提供的限定詞值註解為插入或替換的一級序列，該序列也必須包含在序列表中並分配其專屬序列識別號。

附錄 I

受控詞彙

版本 1.5

WIPO標準委員會(CWS)於2021年11月5日第九屆會議上批准修訂

目錄

第 1 節：核苷酸列表	2
第 2 節：修飾核苷酸列表	3
第 3 節：胺基酸列表	5
第 4 節：修飾胺基酸列表	6
第 5 節：核苷酸序列的特徵關鍵	7
第 6 節：核苷酸序列的限定詞	29
第 7 節：胺基酸序列的特徵關鍵	56
第 8 節：胺基酸序列限定詞	63
第 9 節：遺傳密碼表	65

第 1 節：核苷酸列表

序列表中使用的核苷酸鹼基符號見表1。符號「t」在沒有進一步描述的情況下被解釋為DNA中的胸腺嘧啶及RNA中的尿嘧啶。如果歧義符號(代表2個或多個鹼基)是合適的，則應使用最嚴格的符號。例如，如果給定位置的鹼基可能是「a」或「g」，則應使用「r」，而不是「n」。符號「n」在沒有進一步說明的情況下被解釋為「a或c或g或t/u」。

表1：核苷酸符號列表

符號	定義
a	腺嘌呤(adenine)
c	胞嘧啶(cytosine)
g	鳥嘌呤(guanine)
t	DNA中的胸腺嘧(thymine)啶/RNA中的尿嘧啶(uracil)(t/u)
m	a或c
r	a或g
w	a或t/u
s	c或g
y	c或t/u
k	G或t/u
v	a或c或g；非t/u
h	a或c或t/u；不是g
d	a或g或t/u；不是c
b	c或g或t/u；不是a
n	a或c或g或t/u；「未知」或「其他」

第 2 節：修飾核苷酸列表

表2所列縮寫是限定詞「mod_base」唯一允許的值。如果下表中沒有特定的修飾核苷酸，則必須使用縮寫「OTHER」作為其值。如果以縮寫「OTHER」為值，則必須使用限定詞「note」提供修飾鹼基的完整非縮寫名稱。表2中提供的縮寫詞不得用於序列本身。

表2：修飾核苷酸列表

縮寫	定義
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
dhu	dihydrouridine
fm	2'-O-methylpseudouridine
gal q	beta-D-galactosylqueuosine
gm	2'-O-methylguanosine
i	inosine
i6a	N6-isopentenyladenosine
m1a	1-methyladenosine
m1f	1-methylpseudouridine
m1g	1-methylguanosine
m1i	1-methylinosine
m22g	2, 2-dimethylguanosine
m2a	2-methyladenosine
m2g	2-methylguanosine
m3c	3-methylcytidine
m4c	N4-methylcytosine
m5c	5-methylcytidine
m6a	N6-methyladenosine
m7g	7-methylguanosine
mam5u	5-methylaminomethyluridine
mam5s2u	5-methylaminomethyl-2-thiouridine
man q	beta-D-mannosylqueuosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine

縮寫	定義
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine
mv	uridine-5-oxoacetic acid-methylester
o5u	uridine-5-oxyacetic acid (v)
osyw	wybutoxosine
p	pseudouridine
q	queuosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
m5u	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	wybutosine
x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u
OTHER	(requires note qualifier)

第 3 節：胺基酸列表

序列中使用的胺基酸符號如表3所示。如果歧義符號(代表2個或多個替代胺基酸)是合適的，則應該使用最嚴格的符號。例如，如果給定位置的胺基酸可能是天冬胺酸或天冬醯胺，則應使用符號「B」，而不是「X」。符號「X」在沒有進一步說明的情況下將被解釋為「A」、「R」、「N」、「D」、「C」、「Q」、「E」、「G」、「H」、「I」、「L」、「K」、「M」、「F」、「P」、「O」、「S」、「U」、「T」、「W」、「Y」或「V」。

表3：胺基酸符號列表

符號	定義
A	Alanine
R	Arginine
N	Asparagine
D	Aspartic acid (Aspartate)
C	Cysteine
Q	Glutamine
E	Glutamic acid (Glutamate)
G	Glycine
H	Histidine
I	Isoleucine
L	Leucine
K	Lysine
M	Methionine
F	Phenylalanine
P	Proline
O	Pyrrolysine
S	Serine
U	Selenocysteine
T	Threonine
W	Tryptophan
Y	Tyrosine
V	Valine
B	Aspartic acid或Asparagine
Z	Glutamine或Glutamic acid
J	Leucine或Isoleucine
X	A或R或N或D或C或Q或E或G或H或I或L或K或M或F或P或O或S或U或T或W或Y或V；「未知」或「其他」

第 4 節：修飾胺基酸列表

表4列出了在特徵關鍵「MOD_RES」或「SITE」的必要限定詞「note」中修飾胺基酸唯一允許的縮寫。限定詞「note」的值必須是此表中的縮寫(在適當的情況下)，或修飾胺基酸的完整、非縮寫的名稱。本表中提供的縮寫(或全稱)不得在序列本身中使用。

表4：修飾胺基酸列表

縮寫	修飾胺基酸
Aad	2-Aminoadipic acid
bAad	3-Aminoadipic acid
bAla	beta-Alanine, beta-Aminopropionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4-Diaminobutyric acid
Des	Desmosine
Dpm	2,2'-Diaminopimelic acid
Dpr	2,3-Diaminopropionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
Ide	Isodesmosine
alle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
MeIle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine
Orn	Ornithine

第 5 節：核苷酸序列的特徵關鍵

本節包含允許用於核苷酸序列的特徵關鍵列表，並列出了必要及非必要的限定詞。特徵關鍵按字母順序排列。除非在「分子的範圍(Molecule scope)」下另有說明，否則特徵關鍵可用於DNA或RNA。除了指定的「生物體的範圍(organism scope)」之外，某些特徵關鍵可能還適用於人工序列。

XML序列列表的實例中使用的特徵關鍵名稱必須與其在下面的描述中出現在「特徵關鍵」的完全一樣，特徵關鍵3' UTR及5' UTR除外。請參閱3' UTR及5' UTR特徵關鍵說明中的「註釋」。

5.1	特徵關鍵	C_region
	定義	免疫球蛋白輕鏈及重鏈的恆定區，以及T細胞受體 α 、 β 及 γ 鏈；基於特定鏈上的1個或多個外顯子
	非必要限定詞	allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物體的分類	真核生物

5.2	特徵關鍵	CDS
	定義	編碼序列；與蛋白質中胺基酸序列相對應的核苷酸序列(位置包括終止密碼子)；特徵可能包括胺基酸概念性轉譯
	非必要限定詞	allele circular_RNA codon_start EC_number exception function gene gene_synonym

		map note number operon product protein_id pseudo pseudogene ribosomal_slippage standard_name translation transl_except transl_table trans_splicing
	註釋	限定詞codon_start的有效值為1或2或3，表示可以找到編碼特徵的第1個完整密碼子的偏移量，相對於該特徵的第1個鹼基；transl_table定義除標準或通用遺傳密碼表外使用的遺傳密碼表；在限定詞transl_except中記載指定表範圍的外的例外遺傳密碼；特徵關鍵CDS僅允許用於限定詞translation、偽基因(pseudogene)或偽基因(pseudo)中之一；當使用限定詞translation時，如果轉譯產物包含4個或更多個具體定義的胺基酸，則限定詞protein_id是必要的

5.3	特徵關鍵	centromere
	定義	確定為著絲粒的生物學興趣區域，並已通過實驗表徵
	非必要限定詞	note standard_name
	註釋	著絲粒特徵描述對應於染色單體所在區域並形成著絲粒的DNA間隔

5.4	特徵關鍵	D-loop
	定義	位移環；線粒體DNA中的一個區域，其中一小段RNA與一條DNA鏈配對，取代該區域中原來的伴侶DNA(partner DNA)鏈；也用於描述在RecA蛋白催化的反應中單鏈入侵者對雙鏈DNA的一個區域的

		置換
	非必要限定詞	allele gene gene_synonym map note
	分子的分類	DNA

5.5	特徵關鍵	D_segment
	定義	免疫球蛋白重鏈及T細胞受體 β 鏈的多樣性片段
	非必要限定詞	allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物體的分類	真核生物

5.6	特徵關鍵	exon
	定義	基因組編碼spliced mRNA、rRNA及tRNA部分的區域；可能含有5' UTR、所有CDS 及3' UTR
	非必要限定詞	allele EC_number function gene gene_synonym map note number product pseudo pseudogene standard_name trans_splicing

5.7	特徵關鍵	gene
	定義	確定為基因並為其指定名稱的生物學感興趣區域
	非必要限定詞	allele function gene gene_synonym map note operon product pseudo pseudogene phenotype standard_name trans_splicing
	註釋	基因特徵描述對應於遺傳性狀或表型的DNA區間；根據定義，該特徵不受其末端位置的嚴格限制；其代表基因所在的區域

5.8	特徵關鍵	iDNA
	定義	干預DNA；通過幾種重組中之任何一種而被消除的DNA
	非必要限定詞	allele function gene gene_synonym map note number standard_name
	分子的分類	DNA
	註釋	例如，在體細胞加工中的免疫球蛋白基因

5.9	特徵關鍵	intron
	定義	被轉錄的DNA片段，但通過將其兩側的序列(外顯子)拼接在一起而從轉錄本中移除
	非必要限定詞	allele

		function gene gene_synonym map note number pseudo pseudogene standard_name trans_splicing
--	--	--

5.10	特徵關鍵	J_segment
	定義	連接免疫球蛋白輕鏈及重鏈以及T細胞受體 α 、 β 及 γ 鏈的片段
	非必要限定詞	allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物體的分類	真核生物

5.11	特徵關鍵	mat_peptide
	定義	成熟肽或蛋白質編碼序列；轉譯後修飾後成熟或最終肽或蛋白質產物的編碼序列；該位置不包括終止密碼子(與相應的CDS不同)
	非必要限定詞	allele EC_number function gene gene_synonym map note product pseudo

		pseudogene standard_name
--	--	-----------------------------

5.12	特徵關鍵	misc_binding
	定義	核酸中不能用任何其他結合關鍵(primer_bind或protein_bind)描述的共價或非共價結合另一部分(moiety)的位點
	必要限定詞	bound_moiety
	非必要限定詞	allele function gene gene_synonym map note
	註釋	請注意，描述核糖體結合位點時，必須使用特徵關鍵regulatory及限定詞regulatory_class，並以「ribosome_binding_site」為限定詞的值

5.13	特徵關鍵	misc_difference
	定義	特色序列(featured sequence)與此位置的呈現序列(presented sequence)不同，且不能用任何其他差異關鍵(variation或modified_base)來描述
	非必要限定詞	allele clone compare gene gene_synonym map note phenotype replace standard_name
	註釋	特徵關鍵misc_difference必須用於描述人為引入的可變性，例如，通過基因操作或化學合成；使用限定詞replace來註釋刪除、插入或替換。必須使用特徵關鍵variation來描述自然發生的遺傳變異性。

5.14	特徵關鍵	misc_feature
	定義	無法用任何其他特徵關鍵描述的生物學感興趣區域；一個新的或罕見的特徵
	非必要限定詞	allele function gene gene_synonym map note number phenotype product pseudo pseudogene standard_name
	註釋	當需要只是標記一個區域以便對其進行評論或在另一個功能的位置使用它時，不應使用此關鍵

5.15	特徵關鍵	misc_recomb
	定義	任何普遍的、特定於位點的或複製重組事件的位點，其中雙鏈DNA的斷裂及重聚無法用其他重組關鍵或source關鍵的限定詞(proviral)描述
	非必要限定詞	allele gene gene_synonym map note recombination_class standard_name
	分子的分類	DNA

5.16	特徵關鍵	misc_RNA
	定義	任何不能由其他RNA關鍵定義的轉錄本或RNA產物(prim_transcript、precursor_RNA、mRNA、5' UTR、3' UTR、exon、CDS、sig_peptide、transit_peptide、mat_peptide、intron、polyA_site、ncRNA、rRNA及tRNA)

	非必要限定詞	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
--	--------	---

5.17	特徵關鍵	misc_structure
	定義	其他結構關鍵(stem_loop及D-loop)無法描述的任何二級或三級核苷酸結構或構形(conformation)
	非必要限定詞	allele function gene gene_synonym map note standard_name

5.18	特徵關鍵	mobile_element
	定義	含有移動元件的基因組區域
	必要限定詞	mobile_element_type
	非必要限定詞	allele function gene gene_synonym map note rpt_family rpt_type standard_name

5.19	特徵關鍵	modified_base
	定義	指示的核苷酸是經過修飾的核苷酸，應由指示的分子替換(在限定詞mod_base的值中指明)
	必要限定詞	mod_base
	非必要限定詞	allele frequency gene gene_synonym map note
	註釋	必要限定詞mod_base的值限於附錄第2節受限詞彙表中的修飾核苷酸縮寫

5.20	特徵關鍵	mRNA
	定義	信使核糖核酸；包括5'非翻譯區(5' UTR)、編碼序列(CDS、外顯子)及3'非翻譯區(3' UTR)
	非必要限定詞	allele circular_RNA function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing

5.21	特徵關鍵	ncRNA
	定義	一種非蛋白質編碼基因，不同於ribosomal RNA及transfer RNA，其功能分子是RNA轉錄本
	必要限定詞	ncRNA_class
	非必要限定詞	allele function gene gene_synonym

		map note operon product pseudo pseudogene standard_name trans_splicing
	註釋	ncRNA特徵不得用於ribosomal及transfer RNA 註釋，必須分別使用特徵關鍵rRNA及tRNA

5.22	特徵關鍵	N_region
	定義	在重排的免疫球蛋白片段之間插入的額外核苷酸
	非必要限定詞	allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物體的分類	真核生物

5.23	特徵關鍵	operon
	定義	包含多順反子(polycistronic)轉錄物的區域， 包括受相同調節序列/啟動子控制並處於相同生 物途徑中的一組基因
	必要限定詞	operon
	非必要限定詞	allele function map note phenotype pseudo pseudogene standard_name

5.24	特徵關鍵	oriT
	定義	轉移起點；結合或轉移過程中開始轉移的DNA分子區域
	非必要限定詞	allele bound_moiety direction gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq standard_name
	分子的分類	DNA
	註釋	rep_origin 必須用於描述複製起點；限定詞 direction 具有 left、right 及 both 的允許值，但是當與 oriT 特徵結合使用時，只有 left 及 right 有效；轉移的起點可以存在於染色體中；質體可以包含多個轉移起點

5.25	特徵關鍵	polyA_site
	定義	RNA轉錄本上的位點，通過轉錄後聚腺苷酸化將腺嘌呤殘基添加到該位點
	非必要限定詞	allele gene gene_synonym map note
	生物體的分類	真核生物及真核病毒

5.26	特徵關鍵	precursor_RNA
	定義	任何尚未成為成熟RNA產物的RNA種類；可能包括 ncRNA、rRNA、tRNA、5' 非翻譯區(5' UTR)、編碼序列(CDS、外顯子)、間插序列(內含子)及3' 非翻譯區(3' UTR)

	非必要限定詞	allele function gene gene_synonym map note operon product standard_name trans_splicing
	註釋	可能用於轉錄後加工的RNA；如果問題中的RNA已知未經加工處理，請使用prim_transcript作為特徵關鍵

5.27	特徵關鍵	prim_transcript
	定義	初級(起始、未經處理的)轉錄物；可能包括ncRNA、rRNA、tRNA、5'非翻譯區(5' UTR)、編碼序列(CDS、外顯子)、間插序列(內含子)及3'非翻譯區(3' UTR)
	非必要限定詞	allele function gene gene_synonym map note operon standard_name

5.28	特徵關鍵	primer_bind
	定義	於啟動複製、轉錄或逆轉錄時的非共價引子結合位點；包括用於合成的位點，如PCR引子元件
	非必要限定詞	allele gene gene_synonym map note standard_name
	註釋	用於註解一給定序列上的引子分子結合位點，並

		非代表引子分子本身的序列；由於PCR反應通常涉及成對的引子，因此單個primer_bind特徵關鍵可在2個位置上使用操縱子(operator)次序(location, location)，或使用一對primer_bind特徵關鍵來表示
--	--	---

5. 29	特徵關鍵	propeptide
	定義	前肽編碼序列；蛋白原(proprotein)結構域的編碼序列，其可被切割以形成成熟蛋白質產物
	非必要限定詞	allele function gene gene_synonym map note product pseudo pseudogene standard_name

5. 30	特徵關鍵	protein_bind
	定義	核酸上的非共價蛋白結合位點
	必要限定詞	bound_moiety
	非必要限定詞	allele function gene gene_synonym map note operon standard_name
	註釋	注意，必須使用特徵關鍵regulatory及值為「ribosome_binding_site」的限定詞regulatory_class來描述核糖體結合位點

5. 31	特徵關鍵	regulatory
	定義	在調控轉錄、轉譯、複製或染色質結構中起作用之一序列中的任何區域

	必要限定詞	regulatory_class
	非必要限定詞	allele bound_moiety function gene gene_synonym map note operon phenotype pseudo pseudogene standard_name

5.32	特徵關鍵	repeat_region
	定義	包含重複單元的基因組區域
	非必要限定詞	allele function gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq satellite standard_name

5.33	特徵關鍵	rep_origin
	定義	複製起點；核酸複製的起始位點以產生2個相同的拷貝
	非必要限定詞	allele direction function gene gene_synonym map note

		standard_name
	註釋	direction限定詞具有效值：left、right或both

5.34	特徵關鍵	rRNA
	定義	成熟的核糖體 RNA；將胺基酸組入蛋白質的核糖核蛋白顆粒(核糖體)的RNA部分
	非必要限定詞	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name
	註釋	rRNA大小應以product限定詞註解

5.35	特徵關鍵	S_region
	定義	免疫球蛋白重鏈的轉換區；參與重鏈DNA的重排，造成同一B細胞表達不同類型的免疫球蛋白
	非必要限定詞	allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物體的分類	真核生物

5.36	特徵關鍵	sig_peptide
	定義	信號肽編碼序列；分泌蛋白N端結構域的編碼序列；該結構域參與將新生多肽連接到膜前導序列
	非必要限定詞	allele function

		gene gene_synonym map note product pseudo pseudogene standard_name
--	--	---

5.37	特徵關鍵	source
	定義	用於識別序列的來源；此特徵關鍵為必要的；每個序列都有單一個含括整個序列的特徵關鍵source
	必要限定詞	Organism mol_type
	非必要限定詞	cell_line cell_type chromosome clone clone_lib collected_by collection_date cultivar dev_stage ecotype environmental_sample germline haplogroup haplotype host identified_by isolate isolation_source lab_host lat_lon macronuclear map mating_type

		<p>note organelle PCR_primers plasmid pop_variant proviral rearranged segment serotype serovar sex strain sub_clone sub_species sub_strain tissue_lib tissue_type variety</p>
	分子的分類	任何分子

5.38	特徵關鍵	stem_loop
	定義	髮夾(hairpin)；在單股RNA或DNA中的相鄰(倒置)互補序列間，由鹼基配對所形成的雙螺旋區域
	非必要限定詞	<p>allele function gene gene_synonym map note operon standard_name</p>

5.39	特徵關鍵	STS
	定義	序列標記位點；短的、單拷貝的DNA序列，其特徵是基因組上的定位標誌，可以被PCR檢測；可以透過確定一系列STS的順序來定位基因組
	非必要限定詞	allele

		gene gene_synonym map note standard_name
	分子的分類	DNA
	註釋	在關鍵primer_bind或引子中包含引子的STS位置

5.40	特徵關鍵	telomere
	定義	生物學感興趣的區域，被確定為端粒，並被實驗歸納特徵
	非必要限定詞	note rpt_type rpt_unit_range rpt_unit_seq standard_name
	註釋	telomere特徵描述一DNA區間，係與線性真核染色體末端的特定結構相對應，為末端完整性及維持所必需；該區域與染色體的其餘部分相比是獨一無二的，代表染色體的物理性末端

5.41	特徵關鍵	tmRNA
	定義	轉送傳訊RNA；tmRNA首先作為tRNA，然後再作為編碼胜肽標籤的mRNA；核醣體轉譯tmRNA之此mRNA區域，並將編碼的胜肽標籤附加到未完成的蛋白質C末端；該附加標籤針對蛋白質進行破壞或蛋白水解
	非必要限定詞	allele function gene gene_synonym map note product pseudo pseudogene standard_name

		tag_peptide
--	--	-------------

5.42	特徵關鍵	transit_peptide
	定義	轉運肽(transit peptide)編碼序列；核編碼胞器蛋白的N端結構域的編碼序列；該結構域參與蛋白質轉譯後輸入至胞器
	非必要限定詞	allele function gene gene_synonym map note product pseudo pseudogene standard_name

5.43	特徵關鍵	tRNA
	定義	成熟的轉送RNA，一種小RNA分子(長度為75-85個鹼基)，介導核酸序列轉譯成胺基酸序列
	非必要限定詞	allele circular_RNA anticodon function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing

5.44	特徵關鍵	unsure
	定義	一個已定序鹼基的小區域，長度通常為10個或更少，無法可靠地識別。這樣的區域可能包含所謂

		的鹼基(a、t、g或c)或所謂的鹼基及非所謂鹼基('n')的混合物
	非必要限定詞	allele compare gene gene_synonym map note replace
	註釋	使用限定詞replace來註釋刪除、插入或取代

5.45	特徵關鍵	V_region
	定義	免疫球蛋白輕鏈及重鏈的可變區段，以及T細胞受體 α 、 β 及 γ 鏈；編碼可變胺基末端部分；可以由V_segments、D_segments、N_regions及J_segments組成
	非必要限定詞	allele gene gene_synonym map note product pseudo pseudogene standard_name
	生物體的分類	真核生物

5.46	特徵關鍵	V_segment
	定義	免疫球蛋白輕鏈及重鏈的可變區段，以及T細胞受體 α 、 β 及 γ 鏈；編碼大部分可變區(V_region)及前導肽的最後幾個胺基酸
	非必要限定詞	allele gene gene_synonym map note product pseudo

		pseudogene standard_name
	生物體的分類	真核生物

5.47	特徵關鍵	variation
	定義	相關菌株包含來自同一基因的穩定突變（如限制片段長度多型性(RFLP)、多型性等），該位置(及其他可能的位置)呈現的序列不同
	非必要限定詞	allele compare frequency gene gene_synonym map note phenotype product replace standard_name
	註釋	用於描述對偶基因、RFLP及其他自然發生的突變及多型性；使用限定詞replace來註釋刪除、插入或取代；由於基因操作(例如：定點誘變)而產生的變異性必須用misc_difference 特徵來描述

5.48	特徵關鍵	3' UTR
	定義	1)成熟轉錄物3' 端區域(在終止密碼子之後)，不會轉譯成蛋白質； 2)RNA 病毒3' 端區域(最後一個起始密碼子之後)，不會被轉譯成蛋白質
	非必要限定詞	allele function gene gene_synonym map note standard_name trans_splicing

	註釋	撇號字元在 XML 中有特殊含義，在要素的值中必須用「'」代替。因此，「3' UTR」在XML文件中必須表示為「3'UTR」，即<INSDFeature_key>3'UTR </INSDFeature_key>
--	----	--

5. 49	特徵關鍵	5' UTR
	定義	1)成熟轉錄物5'端區域（在起始密碼子之前），不會轉譯成蛋白質； 2)RNA病毒5'端區域（第一個起始密碼子之前），不會被轉譯成蛋白質
	非必要限定詞	allele function gene gene_synonym map note standard_name trans_splicing
	註釋	撇號字元在XML中有特殊含義，在要素的值中必須用「'」代替。因此，「5' UTR」在XML文件中必須表示為「5'UTR」，即 <INSDFeature_key>5'UTR </INSDFeature_key>

第 6 節：核苷酸序列的限定詞

本節包含用於核苷酸序列特徵的限定詞列表，各限定詞按字母順序排列。

若值格式為「無」，則不得使用INSDQualifier_value要素，且不得使用NonEnglishQualifier_value要素。

若值格式標識為語言依賴性的自由文本，則須使用以下格式之一：

- 1、 INSDQualifier_value要素。
- 2、 NonEnglishQualifier_value要素。
- 3、 INSDQualifier_value要素及NonEnglishQualifier_value要素。

若值格式不是「無」但未標識為語言依賴性自由文本，則須使用INSDQualifier_value要素且不得使用NonEnglishQualifier_value要素。

請注意：限定詞具有值格式為語言依賴性「自由文本」的限定詞值，可能需因國家或區域程序要求進行翻譯。下表所列限定詞的值為語言依賴性自由文本：

表5：值為語言依賴性自由文本的核苷酸序列限定詞列表

項次	限定詞
6.3	bound_moiety
6.5	cell_type
6.8	clone
6.9	clone_lib
6.11	collected_by
6.14	cultivar
6.15	dev_stage
6.18	ecotype
6.22	function
6.24	gene_synonym
6.26	haplogroup
6.28	host
6.29	identified_by
6.30	isolate
6.31	isolation_source
6.32	lab_host
6.36	mating_type
6.41	note

項次	限定詞
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	限定詞	allele
	定義	給定基因的等位基因名稱
	必要值格式	自由文本
	例示	<INSDQualifier_value>adh1-1</INSDQualifier_value>
	註釋	給定基因的所有基因相關特徵(外顯子、CDS等)應共有相同的等位基因限定詞值；根據定義，限定詞allele的值必須不同於限定詞gene的值；當與特徵關鍵variation一起使用時，該限定詞allele的值應為該變異體

6.2	限定詞	anticodon
	定義	tRNA反密碼子的位置及其所編碼的胺基酸
	必要值格式	(pos:<location>, aa:<amino_acid>, seq:<text>), 其中<location>是反密碼子的位置,<amino_acid>是編碼胺基酸的三字母縮寫,<text>是反密碼子的序列
	例示	<INSDQualifier_value>(pos:34..36, aa:Phe, seq:aaa </INSDQualifier_value> <INSDQualifier_value>(pos:join(5,495..496), aa:Leu, seq:taa </INSDQualifier_value> <INSDQualifier_value>(pos:complement(4156..4158), aa:Glu, seq:ttg </INSDQualifier_value>

6.3	限定詞	bound_moiety
	定義	可能與給定特徵結合的分子/複合物的名稱
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>GAL4</INSDQualifier_value>
	註釋	在「misc_binding」、「oriT」及「protein_bind」等特徵中，允許使用單個限定詞bound_moiety

6.4	限定詞	cell_line
	定義	所獲得序列源自的細胞系
	必要值格式	自由文本
	例示	<INSDQualifier_value>MCF7</INSDQualifier_value>

6.5	限定詞	cell_type
	定義	所獲得序列源自的細胞類型 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	必要值格式	自由文本
	例示	<INSDQualifier_value>leukocyte</INSDQualifier_value>

6.6	限定詞	chromosome
	定義	所獲得序列源自的染色體(例如染色體編號)
	必要值格式	自由文本
	例示	<INSDQualifier_value>1</INSDQualifier_value> <INSDQualifier_value>X</INSDQualifier_value>

6.7	限定詞	circular_RNA
	定義	表示外顯子亂序或重疊，因為這種剪切RNA係由反向剪切所產生的環狀RNA(circRNA)，例如，當基因中的下游外顯子位於RNA產物的上游外顯子的5'時
	必要值格式	無
	註釋	應用於如CDS、mRNA、tRNA等特徵及因反向剪切事件而產生的其他特徵。只有當剪切事件係指「加入」操縱子時，此限定詞應被使用，例如 join(complement(69611..69724), 139856..140087)

6.8	限定詞	clone
	定義	所獲得序列源自的殖株
	必要值格式	自由文本

		語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>lambda-hIL7.3</INSDQualifier_value>
	註釋	一個source特徵不得包含多個限定詞clone；如果該序列係從多個殖株中獲得者，可使用特徵關鍵misc_feature及限定詞note在特徵表中進一步描述以指定多個殖株。

6.9	限定詞	clone_lib
	定義	所獲得序列源自的殖株文庫
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>lambda-hIL7</INSDQualifier_value>

6.10	限定詞	codon_start
	定義	表示在1個編碼特徵中可以被找到的第1個完整密碼子的offset，相對於該特徵的第1個鹼基
	必要值格式	1或2或3
	例示	<INSDQualifier_value>2</INSDQualifier_value>

6.11	限定詞	collected_by
	定義	採集該檢體的人名或機構名稱
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Dan Janzen</INSDQualifier_value>

6.12	限定詞	collection_date
	定義	檢體被採集的日期
	必要值格式	YYYY-MM-DD、YYYY-MM或YYYY
	例示	<INSDQualifier_value>1952-10-21</INSDQualifier_value> <INSDQualifier_value>1952-10</INSDQualifier_value> <INSDQualifier_value>1952</INSDQualifier_value>
	註釋	「YYYY」是代表年份的4位數值，「MM」是代表月份的2位數值，「DD」是代表1個月中的第幾天的2位數值

6.13	限定詞	compare
	定義	與現有公知的INSD條目相比較的參考資訊詳情
	必要值格式	登入編號.序列版本[accession-number.sequence-version]
	例示	<INSDQualifier_value>AJ634337.1</INSDQualifier_value>

	註釋	此限定詞可用於以下特徵：misc_difference、unsure及variation，單個特徵中允許多個具不同內容的限定詞compare，此限定詞不適用於大規模註解變異，如SNP
--	----	---

6.14	限定詞	cultivar
	定義	所獲得序列源自的栽培種(栽培變種)
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Nipponbare</INSDQualifier_value> <INSDQualifier_value>Tenuifolius</INSDQualifier_value> <INSDQualifier_value>Candy Cane</INSDQualifier_value> <INSDQualifier_value>IR36</INSDQualifier_value>
	註釋	「cultivar」僅適用於人工篩選的產物，對天然的、已命名的植物及真菌變種係使用variety限定詞

6.15	限定詞	dev_stage
	定義	若序列獲取自處於特定發育階段的生物體，則以此限定詞具體說明
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>fourth instar larva</INSDQualifier_value>

6.16	限定詞	direction
	定義	DNA複製方向
	必要值格式	左、右或兩者 其中「左」指朝向序列模板股的5'端，「右」指朝向3'端
	例示	<INSDQualifier_value>left</INSDQualifier_value>
	註釋	方向限定詞用於註釋特徵關鍵「rep_origin」時，其值可為左、右或兩者；但若方向限定詞係用於註釋特徵關鍵「oriT」，則其值僅可為左或右。

6.17	限定詞	EC_number
	定義	序列的酶的Enzyme Commission number (EC編號)
	必要值格式	自由文本
	例示	<INSDQualifier_value>1.1.2.4</INSDQualifier_value> <INSDQualifier_value>1.1.2.-</INSDQualifier_value> <INSDQualifier_value>1.1.2.n</INSDQualifier_value>

		<INSDQualifier_value>1.1.2.n1</INSDQualifier_value>
	註釋	可接受的EC編號值係依國際生物化學及分子生物學聯盟命名委員會(NCIUBMB)編制列表(發表於Enzyme Nomenclature 1992, Academic Press, San Diego或其更新版)定義。其表示格式為以句點分隔的一串4個數字；從數字串末端開始最多3個數字可用破折號「-」代替，以表示不確定的分配。當EC編號正在等待分配，可以在最後位置使用包括「n」的符號，例如「n」、「n1」等。請注意，此類不完整的EC編號未經NC-IUBMB認可

6.18	限定詞	ecotype
	定義	顯示基於基因給定物種的一群體，其表型特徵反映適應棲息地當地
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Columbia</INSDQualifier_value>
	註釋	例如：為了適應特別陽光充足的棲息地，一群體適應比正常葉子更毛茸茸的葉子。「ecotype」通常適用於擬南芥(<i>Arabidopsis thaliana</i>)的標準基因庫，但亦適用於任何固著生物(<i>sessile organism</i>)。

6.19	限定詞	environmental_sample
	定義	識別由分離自大量環境DNA樣本的直接分子獲取的序列(經由PCR，其包含或不包含隨後產物選殖、DGGE或其他不知名方法)，其未可靠的辨識來源生物。環境樣本包括臨床樣本、腸道內容物及其他來自可能與特定宿主有關的不知名生物體的序列，不包括可由特定宿主可靠地回收的內共生體、來自易於辨識但未培養的田間樣本生物體(例如：許多藍細菌)或從患病植物可靠地回收的植物菌質體(<i>phytoplasmas</i>)(即使其未於在無菌條件下培養)
	值格式	無
	註釋	僅與特徵關鍵「source」併用，特徵關鍵「source」包含限定詞「environmental_sample」時，建議亦包含限定詞「isolation_source」；來源特徵包含限定詞「environmental_sample」時，不得包含限定詞「strain」。

6.20	限定詞	exception
	定義	表示編碼區域不能以標準生物學規則轉譯

	必要值格式	以下受控詞彙短語之一： RNA editing rearrangement required for product annotated by transcript or proteomic data
	例示	<INSDQualifier_value>RNA editing</INSDQualifier_value> <INSDQualifier_value>rearrangement required for product</INSDQualifier_value>
	註釋	僅用於描述如RNA編輯等的生物機制，帶有限定詞「exception」的CDS蛋白質轉譯不同於相應概念上的轉譯；不得用於使用限定詞「transl_excep」已足夠時，例如，在終止密碼子完成使用的情形

6.21	限定詞	frequency
	定義	特徵出現的頻率
	必要值格式	自由文本，表示攜帶該特徵的族群的比例，以分數表示
	例示	<INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value>

6.22	限定詞	function
	定義	歸屬於序列的功能
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>essential for recognition of cofactor </INSDQualifier_value>
	註釋	當基因名稱及及/或產物名稱無法表達可歸屬於序列的功能時，使用限定詞「function」

6.23	限定詞	gene
	定義	對應於序列區域的基因的符號
	必要值格式	自由文本
	例示	<INSDQualifier_value>ilvE</INSDQualifier_value>
	註釋	使用限定詞「gene」提供基因符號；使用限定詞「standard_name」提供完整基因名稱

6.24	限定詞	gene_synonym
	定義	同義、替換、過時或以前的基因符號
	必要值格式	自由文本

		語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Hox-3.3</INSDQualifier_value> 在限定詞「gene」的值为 Hoxc6 的特徵中
	註釋	用於有助於指示基因符號的同義詞，使用限定詞「gene_synonym」時，在限定詞「gene」中須指明主要基因符號

6.25	限定詞	germline
	定義	出現的序列未作為適應性免疫反應的一部分經體細胞重排，其係遺傳自親代種系的未重排序列
	必要值格式	無
	註釋	限定詞「germline」不得用於指示序列來源為配子或生殖細胞；在相同來源特徵中不得使用限定詞「germline」及「rearranged」；限定詞「germline」及「rearranged」只能用於可作為適應性免疫反應一部分而進行體細胞重排的分分子；有領脊椎動物的T細胞受體(TCR)及免疫球蛋白基因座，以及無領魚（七鰓鰻及盲鰻）中不相關的可變淋巴細胞受體(VLR)基因座；除了有頭動物(Craniata; taxid=89593)外，建議不使用限定詞「germline」及「rearranged」

6.26	限定詞	haplogroup
	定義	一群共享某些序列變異的相似單倍型的名稱。單倍群通常用於追蹤群體的遷移。
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>H*</INSDQualifier_value>

6.27	限定詞	haplotype
	定義	在同一染色體上連在一起的特定等位基因的名稱。在無重組情況下，每個單倍型遺傳自一個單元，因此可用於追蹤族群的基因流動。
	必要值格式	自由文本
	例示	<INSDQualifier_value>Dw3 B5 Cw1 A1</INSDQualifier_value>

6.28	限定詞	host
	定義	獲取序列分子的生物體的天然(相對於實驗室)宿主
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>

		<INSDQualifier_value>Homo sapiens 12 year old girl</INSDQualifier_value> <INSDQualifier_value>Rhizobium NGR234</INSDQualifier_value>
--	--	---

6. 29	限定詞	identified_by
	定義	確定標本分類的專家姓名
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>John Burns</INSDQualifier_value>

6. 30	限定詞	isolate
	定義	獲取序列的隔離個體
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Patient #152</INSDQualifier_value> <INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>

6. 31	限定詞	isolation_source
	定義	描述獲取序列的生物樣本的物理、環境及/或當地地理來源
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>rumen isolates from standard Pelleted ration-fed steer #67</INSDQualifier_value> <INSDQualifier_value>permanent Antarctic sea ice</INSDQualifier_value> <INSDQualifier_value>denitrifying activated sludge from carbon_limited continuous reactor</INSDQualifier_value>
	註釋	僅與特徵關鍵「source」併用，特徵關鍵「source」包含限定詞「environmental_sample」時，建議亦包含限定詞「isolation_source」

6. 32	限定詞	lab_host
	定義	實驗室用於繁殖的宿主的學名，其來源生物體的序列已知
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Gallus gallus</INSDQualifier_value>

		<p><INSDQualifier_value>Gallus gallus embryo</INSDQualifier_value></p> <p><INSDQualifier_value>Escherichia coli strain DH5 alpha</INSDQualifier_value></p> <p><INSDQualifier_value>Homo sapiens HeLa cells</INSDQualifier_value></p>
	註釋	已知時，建議使用宿主生物體的完整二名法學名，亦可包括與宿主相關的額外條件資訊

6.33	限定詞	lat_lon
	定義	樣本採集地點的地理座標
	必要值格式	自由文本，經緯度格式為「d[d. dddd] N S d[dd. dddd] W E」
	例示	<p><INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value></p> <p><INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value></p> <p>></p>

6.34	限定詞	macronuclear
	定義	若所示序列為DNA，且係來自染色體分化進行至大核階段及小核階段間的生物體，則此限定詞用於表示該序列來自大核DNA
	值格式	無

6.35	限定詞	map
	定義	特徵的基因組圖譜位置
	必要值格式	自由文本
	例示	<INSDQualifier_value>8q12-q13</INSDQualifier_value>

6.36	限定詞	mating_type
	定義	獲取序列的生物體的交配型，交配型用於原核生物，而真核生物係進行減數分裂而不具性別二型性配子
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<p><INSDQualifier_value>MAT-1</INSDQualifier_value></p> <p><INSDQualifier_value>plus</INSDQualifier_value></p> <p><INSDQualifier_value>-</INSDQualifier_value></p> <p><INSDQualifier_value>odd</INSDQualifier_value></p> <p><INSDQualifier_value>even</INSDQualifier_value></p>

	註釋	限定詞「mating_type」的值male及female在原核生物可被接受，但在真核生物中不可接受；詳細資訊請參閱限定詞「sex」的條目。
--	----	--

6.37	限定詞	mobile_element_type
	定義	以親代特徵描述的移動式元件的類型及名稱或辨識號
	必要值格式	<mobile_element_type>[:<mobile_element_name>] 其中<mobile_element_type>為以下之一： transposon retrotransposon integron insertion sequence non-LTR retrotransposon SINE MITE LINE other
	例示	<INSDQualifier_value>transposon:Tnp9</INSDQualifier_value>
	註釋	「mobile_element_type」僅允許用於特徵關鍵「mobile_element」。移動式元件建議用於表示目前地移動及過去地移動。<mobile_element_type> 的值為「other」時，要求使用 <mobile_element_name>

6.38	限定詞	mod_base
	定義	修飾核苷酸鹼基的縮寫
	必要值格式	修飾核苷酸鹼基的縮寫選自本附錄第2節
	例示	<INSDQualifier_value>m5c</INSDQualifier_value> <INSDQualifier_value>OTHER</INSDQualifier_value>
	註釋	未見於本附錄第2節的特定修飾核苷酸，以「OTHER」為限定詞「mod_base」的值進行註解，並使用限定詞「note」，其值為修飾鹼基的全名

6.39	限定詞	mol_type
	定義	序列的分子類型
	必要值格式	選自下列選項之一： genomic DNA genomic RNA mRNA

		tRNA rRNA other RNA other DNA transcribed RNA viral cRNA unassigned DNA unassigned RNA
	例示	<INSDQualifier_value>genomic DNA</INSDQualifier_value> <INSDQualifier_value>other RNA</INSDQualifier_value>
	註釋	限定詞「mol_type」在特徵關鍵「source」為必要的；值為「genomic DNA」並非意指其為核分子(例如：胞器及質體DNA必須以「genomic DNA」描述)；核糖體RNA基因必須以「genomic DNA」描述；「rRNA」僅用於核糖體RNA本身已被定序時；值為「other RNA」及「other DNA」須被用於合成分子，值「unassigned DNA」及「unassigned RNA」須用於活體內未知分子。

6. 40	限定詞	ncRNA_class
	定義	結構性描述非編碼RNA的分類
	必要值格式	TYPE 其中TYPE是以下受控詞彙術語或短語之一： antisense_RNA autocatalytically_spliced_intron circRNA ribozyme hammerhead_ribozyme lncRNA RNase_P_RNA RNase_MRP_RNA telomerase_RNA guide_RNA sgRNA rasiRNA scrRNA scaRNA siRNA pre_miRNA

		miRNA piRNA snoRNA snRNA SRP_RNA vault_RNA Y_RNA Other
	例示	INSDQualifier_value>autocatalytically_spliced_intron </INSDQualifier_value> <INSDQualifier_value>siRNA</INSDQualifier_value> <INSDQualifier_value>scRNA</INSDQualifier_value> <INSDQualifier_value>other</INSDQualifier_value>
	註釋	特定ncRNA類別未列於受控詞彙ncRNA_class，其註釋必須以「other」作為限定詞「ncRNA_class」的值，並使用限定詞「note」對新ncRNA_class提供簡要解釋

6.41	限定詞	note
	定義	任何註釋或附加資訊
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value> A comment about the feature </INSDQualifier_value>

6.42	限定詞	number
	定義	用於指明遺傳因子(例如：外顯子或內含子)上由5'到3'方向順序的數字
	必要值格式	自由文本(不含空白字元)
	例示	<INSDQualifier_value>4</INSDQualifier_value> <INSDQualifier_value>6B</INSDQualifier_value>
	註釋	文字僅限於整數、字母或整數及字母的組合，表示數值不包含空白字元；建議任何附加術語均包含於限定詞「standard_name」，例示：限定詞「number」的值為2A，並使用限定詞「standard_name」值為「long」

6.43	限定詞	operon
	定義	一組相鄰基因的名稱，其轉錄為該特徵所屬的單個轉錄本
	必要值格式	自由文本

	例示	<INSDQualifier_value>lac</INSDQualifier_value>
--	----	--

6. 44	限定詞	organelle
	定義	一種以膜界定範圍的細胞內構造，可從中獲得序列
	必要值格式	使用以下受控詞彙術語或短語之一： chromatophore hydrogenosome mitochondrion nucleomorph plastid mitochondrion:kinetoplast plastid:chloroplast plastid:apicoplast plastid:chromoplast plastid:cyanelle plastid:leucoplast plastid:proplastid
	例示	<INSDQualifier_value>chromatophore</INSDQualifier_value> <INSDQualifier_value>hydrogenosome</INSDQualifier_value> <INSDQualifier_value>mitochondrion</INSDQualifier_value> <INSDQualifier_value>nucleomorph</INSDQualifier_value> <INSDQualifier_value>plastid</INSDQualifier_value> <INSDQualifier_value>mitochondrion:kinetoplast</INSDQualifier_value> <INSDQualifier_value>plastid:chloroplast</INSDQualifier_value> <INSDQualifier_value>plastid:apicoplast</INSDQualifier_value> <INSDQualifier_value>plastid:chromoplast</INSDQualifier_value> <INSDQualifier_value>plastid:cyanelle</INSDQualifier_value> <INSDQualifier_value>plastid:leucoplast</INSDQualifier_value> <INSDQualifier_value>plastid:proplastid</INSDQualifier_value>

6. 45	限定詞	organism
	定義	若為已知，則提供定序遺傳物質的生物體學名；若生物體未分類，則提供適當的分類學資訊；表明該序列為合成構建體
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>

6. 46	限定詞	PCR_primers
	定義	PCR引子係用於擴增序列，單個限定詞「PCR_primers」建議包含用於單個PCR反應的所有引子，若單個PCR反應存在多個正向或反向引子，將使用多組「fwd_name/fwd_seq」或「rev_name/rev_seq」值
	必要值格式	[fwd_name: XXX1,]fwd_seq: xxxxx1 [fwd_name: XXX2,]fwd_seq: xxxxx2 [rev_name:YYY1,]rev_seq: yyyyy1 [rev_name: YYY2,]rev_seq: yyyyy2
	例示	<INSDQualifier_value>fwd_name: C01P1, fwd_seq: ttgatttttggtcayccwgaagt, rev_name:C01R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value> <INSDQualifier_value>fwd_name: hogel, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<i><i>gtgtatcttact</INSDQualifier_value> <INSDQualifier_value>fwd_name: C01P1, fwd_seq: ttgatttttggtcayccwgaagt, fwd_name:C01P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: C01R4, rev_seq:ccwvytardcctarraartgttg</INSDQualifier_value>
	註釋	fwd_seq及rev_seq均為必要的；而fwd_name及rev_name均為非必要的，2個序列皆須以5'到3'的順序呈現。除修飾鹼基外，序列須使用本附錄第1節所示符號，並須以尖括號<>括起來。在XML中，由於尖括號<及>為XML的保留字元，須以<及>取代

6. 47	限定詞	phenotype
	定義	特徵賦予的表現型，其中表現型被解釋為物理、生化或行為特色或一組特色
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>erythromycin resistance</INSDQualifier_value>

6. 48	限定詞	plasmid
	定義	獲取序列的天然質體的名稱，其中質體被定義為獨立複製基因體單元，且不能以限定詞chromosome或segment描述
	必要值格式	自由文本

	例示	<INSDQualifier_value>pC589</INSDQualifier_value>
6. 49	限定詞	pop_variant
	定義	序列來源樣本的亞群或表現型名稱
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>pop1</INSDQualifier_value> <INSDQualifier_value>Bear Paw</INSDQualifier_value>
6. 50	限定詞	product
	定義	與特徵相關產物的名稱，例如：mRNA特徵的mRNA、CDS的多肽、mat_peptide的成熟肽等。
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>trypsinogen</INSDQualifier_value> (when qualifier appears inCDS feature) <INSDQualifier_value>trypsin</INSDQualifier_value> (when qualifier appears inmat_peptide feature) <INSDQualifier_value>XYZ neural-specific transcript</INSDQualifier_value> (whenqualifier appears in mRNA feature)
6. 51	限定詞	protein_id
	定義	蛋白質序列識別號，其係用於序列表的整數，以指定由相應特徵關鍵CDS及限定詞translation識別的編碼序列所編碼的蛋白質序列
	必要值格式	大於0的整數
	例示	<INSDQualifier_value>89</INSDQualifier_value>
6. 52	限定詞	proviral
	定義	該限定詞用於標記從病毒或噬菌體獲得的序列，該序列整合到另一個生物體的基因組中
	必要值格式	無
6. 53	限定詞	pseudo
	定義	指示此特徵為以特徵關鍵命名的非功能版元件
	值格式	無
	註釋	限定詞「pseudo」建議用於描述未正式描述為偽基因的非

		功能基因，例如：由於偽基因化情狀以外的其他原因使CDS沒有轉譯，而其他原因可能包括序列錯誤或組裝錯誤。為註釋偽基因，須使用限定詞「pseudogene」指明偽基因的類型。
--	--	---

6. 54	限定詞	pseudogene
	定義	以特徵關鍵命名的元件，表示此特徵為偽基因
	必要值格式	TYPE 其中 TYPE 是以下受控詞彙術語或短語之一： processed unprocessed unitary allelic unknown
	例示	<INSDQualifier_value>processed</INSDQualifier_value> <INSDQualifier_value>unprocessed</INSDQualifier_value> <INSDQualifier_value>unitary</INSDQualifier_value> <INSDQualifier_value>allelic</INSDQualifier_value> <INSDQualifier_value>unknown</INSDQualifier_value>
	註釋	TYPE值的定義： processed：偽基因的發生係經由mRNA逆轉錄為cDNA，而後再融入基因體中，因此，其已失去內含子/外顯子結構，且可能具有偽多腺核苷酸尾(pseudo-polyA-tail) unprocessed：偽基因的發生係經由染色體複製複製親本基因，而後積累隨機突變。相較於其具功能的同源物，其改變包括插入、缺失、過早終止密碼子、移框，以及相對同義取代具有更高比例的變異取代 unitary：該偽基因不具親源，其係原始基因，在部分物種中具有功能，但在其他種物種或品系中被以某種方式(插入或缺失、突變、重組)破壞 allelic：一種(unitary)偽基因，其於群體中穩定，但重要的是其於群體中具有一個功能性替代等位基因(即一種菌株具有該基因，另一種菌株具有偽基因)。MHC單倍型具有allelic偽基因 unknown：提交者不知偽基因化的方法

6. 55	限定詞	rearranged
	定義	此條目中出現的序列係作為適應性免疫反應的一部分經體

		細胞重排，其非遺傳自親代種系的未重排序列
	值格式	無
	註釋	限定詞「rearranged」不得用於註釋非涉及適應性免疫反應的染色體重排；在相同來源特徵中不得使用限定詞「germline」及「rearranged」；限定詞「germline」及「rearranged」只能用於可作為適應性免疫反應一部分而進行體細胞重排的分分子；有領脊椎動物的T細胞受體(TCR)及免疫球蛋白基因座，以及無領魚(七鰓鰻及盲鰻)中不相關的可變淋巴細胞受體(VLR)基因座；除了有頭動物(Craniata; taxid=89593)外，建議不使用限定詞「germline」及「rearranged」

6.56	限定詞	recombination_class
	定義	結構性描述序列中重組熱點區域的分類
	必要值格式	TYPE 其中TYPE是以下受控詞彙術語或短語之一： meiotic mitotic non_allelic_homologous chromosome_breakpoint other
	例示	<INSDQualifier_value>meiotic</INSDQualifier_value> <INSDQualifier_value>chromosome_breakpoint</INSDQualifier_value>
	註釋	特定重組類別未列於受控詞彙recombination_class，其註釋必須以「other」作為限定詞「recombination_class」的值，並使用限定詞「note」對新recombination_class提供簡要解釋

6.57	限定詞	regulatory_class
	定義	結構性描述序列中轉錄、轉譯、複製及染色質結構相關調控元件的分類
	必要值格式	TYPE 其中 TYPE 是以下受控詞彙術語或短語之一： attenuator CAAT_signal DNase_I_hypersensitive_site enhancer

		<p>enhancer_blocking_element GC_signal imprinting_control_region insulator locus_control_region matrix_attachment_region minus_35_signal minus_10_signal polyA_signal_sequence promoter recoding_stimulatory_region recombination_enhancer replication_regulatory_region response_element ribosome_binding_site riboswitch silencer TATA_box terminator transcriptional_cis_regulatory_region uORF other</p>
	例示	<p><INSDQualifier_value>promoter</INSDQualifier_value> <INSDQualifier_value>enhancer</INSDQualifier_value> <INSDQualifier_value>ribosome_binding_site</INSDQualifier_value></p>
	註釋	<p>特定調控類別未列於受控詞彙regulatory_class，其註釋必須以「other」作為限定詞「regulatory_class」的值，並使用限定詞「note」對新regulatory_class提供簡要解釋</p>

6. 58	限定詞	replace
	定義	表示以特徵位置標識的序列被限定詞值所示的序列取代；若限定詞中不包含序列（即沒有值），則表示刪除
	必要值格式	自由文本
	例示	<p><INSDQualifier_value>a</INSDQualifier_value> <INSDQualifier_value></INSDQualifier_value> - for a deletion</p>

6.59	限定詞	ribosomal_slippage
	定義	在蛋白質轉譯過程中，某些序列透過核糖體滑移機制，規劃核糖體改變為替代閱讀框(alternative reading frame)
	值格式	無
	註釋	特徵位置CDS中須使用連接運算符(例如：[join(486..1784,1787..4810)])，以指示 ribosomal_slippage的位置

6.60	限定詞	rpt_family
	定義	重複序列的類型，例如：Alu或Kpn
	必要值格式	自由文本
	例示	<INSDQualifier_value>Alu</INSDQualifier_value>

6.61	限定詞	rpt_type
	定義	重複序列的結構及分佈
	必要值格式	使用以下受控詞彙術語或短語之一： tandem direct inverted flanking nested terminal dispersed long_terminal_repeat non_ltr_retrotransposon_polymeric_tract centromeric_repeat telomeric_repeat x_element_combinatorial_repeat y_prime_element other
	例示	<INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value>
	註釋	值的定義 tandem：以相同方向鄰近於另一個的重複 direct：不總是鄰近但方向相同的重複 inverted：在同一分子上以相反方向出現的成對重複 flanking：位於具功能意義的序列的外的重複（例如：目

		<p>標位點的轉位子插入)</p> <p>nested：經由插入另一元件而中斷的重複</p> <p>terminal：分散於整個基因組的重複</p> <p>dispersed：在具功能意義的序列末端及內部重複(例如：轉位子LTR)</p> <p>long_terminal_repeat：在定義序列的兩端直接重複的序列，通常見於逆轉錄病毒</p> <p>non_ltr_retrotransposon_polymeric_tract：非LTR逆轉錄轉位子內聚合束，例如：poly(dA)</p> <p>centromeric_repeat：在模組化著絲粒內發現的重複區域</p> <p>telomeric_repeat：在端粒內的重複區域</p> <p>x_element_combinatorial_repeat：位於X元件及端粒或鄰近的Y'元件間的重複區域</p> <p>y_prime_element：鄰近於端粒重複序列或X元件組合重複序列的重複區域，其為單次複製或為2到4次複製的串聯重複</p> <p>other：展現無法以其他值描述的重要屬性的重複</p>
--	--	--

6. 62	限定詞	rpt_unit_range
	定義	以範圍表示重複單元的位置
	必要值格式	<base_range> - 其中<base_range>是重複單元的第1個及最後1個鹼基(由2個點分隔)
	例示	<INSDQualifier_value>202..245</INSDQualifier_value>
	註釋	用於表示序列的鹼基範圍，其係構成特徵關鍵「oriT」及「repeat_region」指定區域的重複單元

6. 63	限定詞	rpt_unit_seq
	定義	重複序列本身
	必要值格式	自由文本
	例示	<INSDQualifier_value>aagggc</INSDQualifier_value> <INSDQualifier_value>ag(5)tg(8)</INSDQualifier_value> <INSDQualifier_value>(AAAGA)6(AAAA)1(AAAGA)12</INSDQualifier_value>
	註釋	用於表示文字序列，其係構成特徵關鍵「oriT」及「repeat_region」指定區域的重複單元。

6. 64	限定詞	satellite
	定義	隨體DNA(satellite DNA)標記的標識符，由短、基本的重複

		單元的許多串聯重複(相同或相關)組成
必要值格式		<satellite_type>[:<class>][<identifier>] - 其中 <satellite_type>為下列之一： satellite microsatellite minisatellite
例示		<INSDQualifier_value>satellite: Sla</INSDQualifier_value> <INSDQualifier_value>satellite: alpha</INSDQualifier_value> <INSDQualifier_value>satellite: gamma III</INSDQualifier_value> <INSDQualifier_value>microsatellite: DC130</INSDQualifier_value>
註釋		許多隨體的鹼基組成或其他特性不同於基因組的其餘部分，因此可以被識別

6.65	限定詞	segment
	定義	定序的病毒或噬菌體片段名稱
	必要值格式	自由文本
	例示	<INSDQualifier_value>6</INSDQualifier_value>

6.66	限定詞	serotype
	定義	以其抗原特性為特徵的物種的血清學變種
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>B1</INSDQualifier_value>
	註釋	僅與特徵關鍵source一起使用；細菌學法典(Bacteriological Code)建議原核生物使用術語「血清型(serovar)」替代「血清型(serotype)」；參見國際細菌命名法典(International Code of Nomenclature of Bacteria)(1990年修訂版)附錄 10.B「種下名詞(Infraspecific Terms)」。

6.67	限定詞	serovar
	定義	以其抗原特性為特徵的物種(通常是原核生物)的血清學變種
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>O157:H7</INSDQualifier_value>
	註釋	僅與特徵關鍵source一起使用；細菌學法典(Bacteriological

		cal Code)建議原核生物使用術語「血清型(serovar)」而不是「血清型(serotype)」；參見國際細菌命名法典(International Code of Nomenclature of Bacteria)(1990年修訂版)附錄 10.B「種下名詞(Infraspecific Terms)」。
--	--	---

6. 68	限定詞	sex
	定義	獲得該序列的生物體性別；性別用於經減數分裂並具有兩性異型配子的真核生物
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>female</INSDQualifier_value> <INSDQualifier_value>male</INSDQualifier_value> <INSDQualifier_value>hermaphrodite</INSDQualifier_value> <INSDQualifier_value>unisexual</INSDQualifier_value> <INSDQualifier_value>bisexual</INSDQualifier_value> <INSDQualifier_value>asexual</INSDQualifier_value> <INSDQualifier_value>monoecious</INSDQualifier_value> [or mon ecious] <INSDQualifier_value>dioecious</INSDQualifier_value> [or diec ious]
	註釋	後生動物(Metazoa)、有胚植物亞界(Embryophyta)、紅藻門(Rhodophyta)及褐藻門(Phaeophyceae)應使用限定詞sex(而不是限定詞mating_type)；細菌、古細菌及真菌應使用限定詞mating_type(而不是限定詞sex)；病毒不應使用限定詞sex及mating_type；在前述所列分類外，建議使用限定詞mating_type，除非限定詞的值取自上面例示提供的詞彙表

6. 69	限定詞	standard_name
	定義	此關鍵接受的標準名稱
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>dotted</INSDQualifier_value>
	註釋	使用限定詞standard_name 提供完整的基因名稱，但使用限定詞gene提供基因符號(在上述例示中限定詞gene值為Dt)。

6. 70	限定詞	strain
	定義	獲得該序列的品系
	必要值格式	自由文本

		語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>BALB/c</INSDQualifier_value>
	註釋	含有限定詞strain的特徵項目不得包含限定詞environmental_sample

6.71	限定詞	sub_clone
	定義	獲得該序列的亞殖株
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>lambda-hIL7.20g</INSDQualifier_value>
	註釋	特徵source不得包含多個限定詞sub_clone；為了表示該序列是從多個亞殖株中獲得的，可以使用特徵關鍵「misc_feature」及限定詞「note」進一步描述多重來源。

6.72	限定詞	sub_species
	定義	獲得該序列的生物亞種名稱
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>lactis</INSDQualifier_value>

6.73	限定詞	sub_strain
	定義	獲得該序列的基因或其他修飾菌株的名稱或標識符，源自於親代菌株(應在限定詞strain中註釋)。sub_strain係指獲得該序列者
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>abis</INSDQualifier_value>
	註釋	必須在特徵source中附有限定詞strain；如果未提供親代菌株，則修飾菌株應在限定詞strain而不是sub_strain中註釋。例如，限定詞strain的值為K-12且限定詞substrain的值為MG1655，或限定詞strain的值為MG1655

6.74	限定詞	tag_peptide
	定義	編碼tmRNA蛋白水解標籤及其終止密碼子的多肽的鹼基位置
	必要值格式	<base_range> - 其中<base_range>提供蛋白水解標記第1個及最後1個鹼基(以2個點分隔)的位置
	例示	<INSDQualifier_value>90..122</INSDQualifier_value>
	註釋	建議使用5' 部分CDS特徵對tag_peptide對應的胺基酸序列

		進行註釋，例如CDS的位置<90..122
--	--	-----------------------

6.75	限定詞	tissue_lib
	定義	取得序列的組織庫
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>tissue library 772</INSDQualifier_value>

6.76	限定詞	tissue_type
	定義	取得該序列的組織類型
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>liver</INSDQualifier_value>

6.77	限定詞	transl_except
	定義	轉譯例外：單一密碼子的轉譯不符合生物體或transl_table定義的遺傳密碼
	必要值格式	(pos:<location>, aa:<amino_acid>) 其中<amino_acid>是鹼基範圍(base_range)位置的密碼子所編碼的胺基酸的3字母縮寫
	例示	<INSDQualifier_value>(pos:213..215, aa:Trp) </INSDQualifier_value> <INSDQualifier_value>(pos:462..464, aa:OTHER) </INSDQualifier_value> <INSDQualifier_value>(pos:1017, aa:TERM) </INSDQualifier_value> > <INSDQualifier_value>(pos:2000..2001, aa:TERM) </INSDQualifier_value>
	註釋	如果胺基酸不是附錄第3節中列出的特定胺基酸之一，則<amino_acid>使用OTHER，並在限定詞Note中提供該不常見的胺基酸名稱；對於修飾的胺基酸硒半胱胺酸，<amino_acid>使用3個字母的縮寫「Sec」（胺基酸序列中的單字母符號「U」）；對於修飾的胺基酸吡咯離胺酸，<amino_acid>使用3個字母的縮寫「Pyl」（胺基酸序列中的單字母符號「O」）；對於部分終止密碼子，如TAA終止密碼子透過在mRNA添加3' A殘基來完成，使用base_position或base_range說明其位置，參見上述第3、4項例示，並同時使用限定詞Note指明「透過在mRNA添加 3' A殘基完成終止密碼子」

6.78	限定詞	transl_table
	定義	定義所用為通用或標準遺傳密碼表以外的遺傳密碼表。使用的表格如附錄
	必要值格式	<integer> 其中<integer>是分配給遺傳密碼表的數字
	例示	<INSDQualifier_value>3</INSDQualifier_value> 例示使用酵母菌線粒體代碼
	註釋	若特徵關鍵CDS未使用限定詞transl_table進一步註釋，則表示CDS使用標準代碼(即通用遺傳代碼)進行轉譯 若使用指定表範圍之外的例外遺傳密碼，需用限定詞 transl_except說明

6.79	限定詞	trans_splicing
	定義	表明來自2個RNA的外顯子在分子間反應，連接形成成熟RNA
	值格式	無
	註釋	應該用於特徵CDS、mRNA及其他因反式剪接而產生的特徵。只有當運算符「join」指示剪接時，才必須使用此限定詞，如：特徵location為 join(complement(69611..69724), 139856..140087)

6.80	限定詞	translation
	定義	來自標準(或通用)遺傳密碼，或限定詞transl_table指定與限定詞transl_except例外確定的表格的單字母縮寫胺基酸序列
	必要值格式	來自附錄第3節的單字母胺基酸縮寫連續字元串，「X」用於A例外。
	例示	<INSDQualifier_value>MASTFPPWYRGCASTPSLKGLIMCTW</INSDQualifier_value>
	註釋	僅與特徵CDS一起使用；當轉譯產物包含4個或更多個明確定義胺基酸時，必須附有限定詞protein_id；見transl_table遺傳密碼表的定義及位置；特徵CDS僅允許使用限定詞translation、pseudo及pseudogene之一作進一步註釋。

6.81	限定詞	variety
	定義	序列來源的生物體種類(= varietas，正式的林奈等級)。
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯

	例示	<INSDQualifier_value>insularis</INSDQualifier_value>
	註釋	<p>對栽培植物品種(即人工選擇的產品)使用限定詞cultivar。植物及真菌變異以外的品種應使用限定詞Note進行註釋，例如：值為</p> <p><INSDQualifier_value>breed:Cukorova</INSDQualifier_value></p>

第 7 節：胺基酸序列的特徵關鍵

本節包含允許用於胺基酸序列的特徵鍵列表。特徵關鍵按字母順序排列。

7.1	特徵關鍵	ACT_SITE
	定義	參與酶活性的胺基酸
	非必要限定詞	note
	註釋	必須用特徵關鍵 ACT_SITE 分別註釋活性位點的每個胺基酸殘基，特徵 location 要素中的位置描述符必須提供相對應的胺基酸殘基編號

7.2	特徵關鍵	BINDING
	定義	任何化學基團(輔酶、輔基等)的結合位點。在限定詞 Note 指示該基團的化學性質
	必要限定詞	note
	註釋	限定詞「Note」的值例示：「血基質(共價)」及「氯化物」。在適當的情況下，應使用特徵關鍵 CA_BIND、DNA_BIND、METAL 及 NP_BIND，而不是 BINDING

7.3	特徵關鍵	CA_BIND
	定義	鈣結合區的範圍
	非必要限定詞	note

7.4	特徵關鍵	CARBOHYD
	定義	醣基化位點
	必要限定詞	note
	註釋	該關鍵描述聚醣(單醣或多醣)與蛋白質殘基的連接情況。在限定詞 Note 指示與蛋白質的連接類型(C-、N-或O-連接)，若已知還原末端糖的性質，其縮寫顯示於括號內，縮寫後面有三個點「...」，表示醣鏈的延伸；反之，沒有點表示連接單醣。限定詞「Note」的值例示：N-liked (GlcNAc...)、O-liked (GlcNAc)；O-liked (Glc...)、C-liked (Man) 部分、O-liked (Ara...)

7.5	特徵關鍵	CHAIN
	定義	成熟蛋白質中多肽鏈的範圍
	非必要限定詞	note

7.6	特徵關鍵	COILED
	定義	雙螺旋(coiled-coil)的範圍
	非必要限定詞	note

7.7	特徵關鍵	COMPBIAS
	定義	組成偏差區域的範圍
	非必要限定詞	note

7.8	特徵關鍵	CONFLICT
	定義	不同的來源報告不同的序列
	非必要限定詞	note
	註釋	限定詞「Note」的值例示：缺失、K->Q、GSDSE->RIRLR、V->A。

7.9	特徵關鍵	CROSSLNK
	定義	轉譯後形成的胺基酸鍵
	必要限定詞	note
	註釋	2種蛋白質之間(鏈間交聯)或同一蛋白質的2個部分之間(鏈內交聯)形成的各種類型共價鍵,除由雙硫鍵形成的交聯外,使用特徵關鍵「DISULFID」。對於鏈間交聯,特徵location要素中的位置描述符是與其他蛋白質交聯的胺基酸殘基數。對於鏈內交聯,特徵location要素中的位置描述符是「x..y」格式交聯胺基酸殘基編號,例如「42..50」。限定詞Note表示交聯的性質,至少指定共軛的名稱及所涉及的2個胺基酸的特性。限定詞「Note」值例示:「異麩胺醯半胱胺酸硫酯(Cys-Gln)」、「β-甲基羊毛硫胺酸(Cys-Thr)」及「甘胺醯基離胺酸異肽(Lys-Gly)(與泛素的G-Cter互鏈)」

7.10	特徵關鍵	DISULFID
	定義	雙硫鍵
	必要限定詞	note
	註釋	對於鏈間雙硫鍵,特徵location要素中的位置描述符是與其他蛋白質連接的半胱胺酸的殘基數。對於鏈內交聯,特徵location要素中的位置描述符是以「x..y」格式表示鏈接的半胱胺酸的殘基編號,例如「42..50」。對於鏈間雙硫鍵,限定詞Note通過識別其他蛋白質來指示交聯的性質,例如「鏈間(A及B鏈之間)」

7.11	特徵關鍵	DNA_BIND
	定義	DNA結合區域的範圍
	必要限定詞	note
	註釋	限定詞Note提供DNA結合區域的性質。限定詞「Note」值的例示：「同源匣(Homeobox)」及「Myb 2」。

7.12	特徵關鍵	DOMAIN
	定義	域(domain)的範圍,定義為組成特徵三維結構或折疊的二級結構的特定組合
	必要限定詞	note
	註釋	限定詞Note提供域的類型。如果1個域存在多種版本,則對域進行編號。限定詞「Note」值的例示：「Ras-GTP酶活化蛋白」及「鈣黏蛋白1」

7.13	特徵關鍵	HELIX
	定義	二級結構螺旋(helix),例如 α 螺旋、3(10)螺旋或pi螺旋
	非必要限定詞	note
	註釋	這種特徵只能用於已知三級結構的蛋白質。只有3種類型的二級結構被指定:螺旋(關鍵HELIX)、 β 鏈(關鍵STRAND)及轉折(關鍵TURN)。非特定於前述任一類別的殘基為位於「環(loop)」或「隨機線圈(random-coil)」結構中

7.14	特徵關鍵	INIT_MET
	定義	起始甲硫胺酸
	非必要限定詞	note
	註釋	特徵location要素中的位置描述符為「1」。該特徵關鍵表示N末端甲硫胺酸被切除。當起始甲硫胺酸未被切除時,不使用此特徵

7.15	特徵關鍵	INTRAMEM
	定義	位於膜中但不穿過膜的區域的範圍
	非必要限定詞	note

7.16	特徵關鍵	LIPID
	定義	脂質部分的共價結合

	必要限定詞	note
	註釋	限定詞Note提供結合脂質部分的化學性質，至少表明脂化胺基酸的名稱。限定詞「Note」值的例示：「N-肉豆蔻醯甘胺酸」、「多醣磷脂肌醇錨定醯胺化絲胺酸」及「S-二醯基甘油半胱胺酸」

7.17	特徵關鍵	METAL
	定義	金屬離子的結合位點
	必要限定詞	note
	註釋	限定詞Note表示金屬的性質。限定詞「Note」值的例示：「鐵(血基質軸向配體)」及「銅」

7.18	特徵關鍵	MOD_RES
	定義	殘基的轉譯後修飾
	必要限定詞	note
	註釋	在限定詞Note中提供修飾殘基的化學性質，至少表明轉譯後修飾胺基酸的名稱。如果修飾胺基酸列於附錄第4節，則可以使用縮寫代替全稱。限定詞「Note」值的例示：「N-乙醯丙胺酸」、「3-Hyp」及「MeLys」或「N-6-甲基離胺酸」

7.19	特徵關鍵	MOTIF
	定義	具有生物學意義之短(最多20個胺基酸)序列模體
	非必要限定詞	note

7.20	特徵關鍵	MUTAGEN
	定義	經突變誘發實驗改變的位點
	非必要限定詞	note

7.21	特徵關鍵	NON_STD
	定義	非標準胺基酸
	非必要限定詞	note
	註釋	該關鍵僅描述胺基酸序列中出現非標準胺基酸硒半胱胺酸(U)及吡咯離胺酸(O)

7.22	特徵關鍵	NON_TER
	定義	序列盡頭的殘基不是末端殘基
	非必要限定詞	note

	註釋	如果應用於位置1，表示第一個位置不是完整分子的N末端；如果應用於最後一個位置，則表示該位置不是完整分子的C末端。
--	----	--

7.23	特徵關鍵	NP_BIND
	定義	核苷酸磷酸鹽結合區的範圍
	必要限定詞	note
	註釋	在限定詞Note中指示核苷酸的性質。限定詞「note」的值的例示：「ATP」及「FAD」

7.24	特徵關鍵	PEPTIDE
	定義	釋放的活性胜肽的範圍
	非必要限定詞	note

7.25	特徵關鍵	PROPEP
	定義	原肽的範圍
	非必要限定詞	note

7.26	特徵關鍵	REGION
	定義	序列中感興趣區域的範圍
	非必要限定詞	note

7.27	特徵關鍵	REPEAT
	定義	內部序列重複的範圍
	非必要限定詞	note

7.28	特徵關鍵	SIGNAL
	定義	訊息序列(前肽)的範圍
	非必要限定詞	note

7.29	特徵關鍵	SITE
	定義	序列上未被另一個特徵關鍵定義的任何胺基酸位點，其也可以應用於胺基酸鍵，表示兩側胺基酸的位置
	必要限定詞	note
	註釋	當SITE用於註釋修飾胺基酸時，限定詞「Note」的值必須是附錄第4節中規定的縮寫，或者是該修飾胺基酸完整非縮寫的名稱

7.30	特徵關鍵	source
	定義	識別序列的來源；這個關鍵是必要的；每個序列都有一個橫跨整個序列的特徵source
	必要限定詞	mol_type organism
	非必要限定詞	note

7.31	特徵關鍵	STRAND
	定義	二級結構： β 鏈(Beta-strand)，例如氫鍵 β 鏈或獨立 β 鏈的殘基
	非必要限定詞	note
	註釋	這種特徵只能用於已知三級結構的蛋白質。只有3種類型的二級結構被指定：螺旋(關鍵HELIX)、 β 鏈(關鍵STRAND)及轉折(關鍵TURN)。非特定於前述任一類別的殘基為位於「環(loop)」或「隨機線圈(random-coil)」結構中

7.32	特徵關鍵	TOPO_DOM
	定義	拓撲區域
	非必要限定詞	note

7.33	特徵關鍵	TRANSMEM
	定義	穿膜區域的範圍
	非必要限定詞	note

7.34	特徵關鍵	TRANSIT
	定義	轉運胜肽(線粒體、葉綠體、類囊體、藍藻、過氧化物酶體等)的範圍
	非必要限定詞	note

7.35	特徵關鍵	TURN
	定義	二級結構轉折，例如氫鍵轉折(3-轉折、4-轉折或5-轉折)
	非必要限定詞	note
	註釋	這種特徵只能用於已知三級結構的蛋白質。只有3種類型的二級結構被指定：螺旋(關鍵HELIX)、 β 鏈(關鍵STRAND)及轉折(關鍵TURN)。非特定於前述任一類別的殘基為位於「環(loop)」或「隨機線圈(random-coil)」

		結構中
7.36	特徵關鍵	UNSURE
	定義	序列中的不確定者
	非必要限定詞	note
	註釋	用於描述作者所呈現序列中不確定的胺基酸序列區域
7.37	特徵關鍵	VARIANT
	定義	作者報告存在序列變異
	非必要限定詞	note
7.38	特徵關鍵	VAR_SEQ
	定義	描述由選擇性剪接、選擇性啟動子使用、選擇性起始及核糖體框移產生的序列變異
	非必要限定詞	note
7.39	特徵關鍵	ZN_FING
	定義	鋅指(zinc finger)區域的範圍
	必要限定詞	note
	註釋	鋅指的類型在限定詞note中指明，例如：「GATA型」及「NR C4型」

第 8 節：胺基酸序列限定詞

本節包含允許用於胺基酸序列的限定詞列表。

若值格式標識為語言依賴性的自由文本，則必須使用下列格式之一：

- 1、 INSDQualifier_value要素。
- 2、 NonEnglishQualifier_value要素。
- 3、 INSDQualifier_value要素及NonEnglishQualifier_value要素。

若值格式未標識為語言依賴性的自由文本，則必須使用INSDQualifier_value要素，且不得使用NonEnglishQualifier_value要素。

請注意：限定詞具有值格式為語言依賴性的「自由文本」的限定詞值，可能需因國家或區域程序要求進行翻譯。下表所列限定詞被認為具有語言依賴性的自由文本值：

表6：值為語言依賴性自由文本的胺基酸序列限定詞列表

項次	限定詞
8.2	note
8.3	organism

8.1	限定詞	mol_type
	定義	序列的體內分子類型
	必要值格式	蛋白質
	例示	<INSDQualifier_value>protein</INSDQualifier_value>
	Comment	在特徵關鍵source，限定詞「mol_type」為必填

8.2	限定詞	note
	定義	任何評論或附加資訊
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Heme (covalent)</INSDQualifier_value>
	Comment	在特徵關鍵BINDING、CARBOHYD、CROSSLNK；DISULFID、DNA_BIND、DOMAIN、LIPID、METAL、MOD_RES、NP_BIND、SITE及ZN_FING，限定詞「note」為必填

8.3	限定詞	organism
	定義	提供胜肽的生物體學名
	必要值格式	自由文本 語言依賴性：此值可能需因國家或區域程序要求進行翻譯
	例示	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
	Comment	在特徵關鍵source，限定詞「organism」為必填

21 - Trematode Mitochondrial Code	
AAs	= FFLSSSSYY**CCWLLLLPPPHHQRRRRIIMTTTTNNKSSSSVVVAAAADDEEGGG
Starts	= -----M-----M-----
Base1	= tttttttttttttttcccccccccccccaaaaaaaaaaaaaaggggggggggggggg
Base2	= tttccccaaaagggttttccccaaaagggttttccccaaaagggtttccccaaaagggg
Base3	= tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
22 - Scenedesmus obliquus Mitochondrial Code	
AAs	= FFLSS*SY*LCC*WLLLLPPPHHQRRRRIIMTTTTNNKSSRRVVVAAAADDEEGGG
Starts	= -----M-----
Base1	= tttttttttttttttcccccccccccccaaaaaaaaaaaaaaggggggggggggggg
Base2	= tttccccaaaagggttttccccaaaagggttttccccaaaagggtttccccaaaagggg
Base3	= tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
23 - Thraustochytrium Mitochondrial Code	
AAs	= FF*LSSSSYY**CC*WLLLLPPPHHQRRRRIIMTTTTNNKSSRRVVVAAAADDEEGGG
Starts	= -----M--M-----M-----
Base1	= tttttttttttttttcccccccccccccaaaaaaaaaaaaaaggggggggggggggg
Base2	= tttccccaaaagggttttccccaaaagggttttccccaaaagggtttccccaaaagggg
Base3	= tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
24 - Pterobranchia Mitochondrial Code	
AAs	= FFLSSSSYY**CCWLLLLPPPHHQRRRRIIMTTTTNNKSSKVVVAAAADDEEGGG
Starts	= --M-----M-----M-----M-----
Base1	= tttttttttttttttcccccccccccccaaaaaaaaaaaaaaggggggggggggggg
Base2	= tttccccaaaagggttttccccaaaagggttttccccaaaagggtttccccaaaagggg
Base3	= tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
25 - Candidate Division SR1 and Gracilibacteria Code	
AAs	= FFLSSSSYY**CCGWLLLLPPPHHQRRRRIIMTTTTNNKSSRRVVVAAAADDEEGGG
Starts	= ---M-----M-----M-----
Base1	= tttttttttttttttcccccccccccccaaaaaaaaaaaaaaggggggggggggggg
Base2	= tttccccaaaagggttttccccaaaagggttttccccaaaagggtttccccaaaagggg
Base3	= tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
26 - Pachysolen tannophilus Nuclear Code	
AAs	= FFLSSSSYY**CC*WLLLAPPPPHHQRRRRIIMTTTTNNKSSRRVVVAAAADDEEGGG
Starts	= -----M-----M-----
Base1	= tttttttttttttttcccccccccccccaaaaaaaaaaaaaaggggggggggggggg
Base2	= tttccccaaaagggttttccccaaaagggttttccccaaaagggtttccccaaaagggg
Base3	= tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
27 - Karyorelict Nuclear Code	
AAs	= FFLSSSSYYQCCWLLLLPPPHHQRRRRIIMTTTTNNKSSRRVVVAAAADDEEGGG
Starts	= -----*-----M-----
Base1	= tttttttttttttttcccccccccccccaaaaaaaaaaaaaaggggggggggggggg
Base2	= tttccccaaaagggttttccccaaaagggttttccccaaaagggtttccccaaaagggg
Base3	= tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag

附錄 II

序列表文檔類型定義(DTD)

版本 1.3

WIPO標準委員會(CWS)於2021年11月5日第九屆會議上批准修訂

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Annex II of WIPO Standard ST.26, Document Type Definition (DTD) for Sequence Listing

This entity may be identified by the PUBLIC identifier:
*****
****
PUBLIC "-//WIPO//DTD SEQUENCE LISTING 1.3//EN" "ST26SequenceListing_V1_3.dtd"
*****
****
* PUBLIC DTD URL

* https://www.wipo.int/standards/dtd/ST26SequenceListing_V1_3.dtd
*****

* Revision of Annex II to WIPO Standard ST.26 was approved by the Committee on WIPO
* Standards (CWS) at its eighth session.

*****
* CONTACTS
*****
*
* xml.standards@wipo.int
*
*
*****
* NOTES
*****

* The sequence data part is a subset of the complete INSDC DTD V.1.5 that only covers
* the requirements of WIPO Standard ST.26.

*****
* REVISION HISTORY
*****
2021-11-05: Revised Version 1.3 approved at CWS/9 (small edits to the comments)
2020-05-20: Version 1.3 approved at CWS/8.
Changes:
- Optional originalFreeTextLanguageCode attribute added to <ST26SequenceListing> to allow
applicants to indicate the language of the free text in the original sequence listing.
- Optional nonEnglishFreeTextLanguageCode attribute added to <ST26SequenceListing> to allow
applicants to indicate the language of the free text provided in the element
<NonEnglishQualifier_value>.
- Optional id attribute added to INSDQualifier to facilitate comparison of language-
dependent qualifier values between sequence listings.
- Optional element <NonEnglishQualifier_value> added to element <INSDQualifier> to allow
applicants to type language-dependent qualifiers in a non-English Language with the
characters set forth in paragraph 40(a) of the ST.26 main body document.

2018-10-19: Version 1.2 approved at CWS/6.
Changes:
```

<INSDQualifier*> changed to <INSDQualifier+> for alignment with business needs and advice from NCBI (an INSDFeature_qual element (if present) should have one or more INSDQualifier elements).

2017-06-02: Version 1.1 approved at the CWS/5

Changes:

Comments added to <INSDSeq_length>, <INSDSeq_division> and <INSDSeq_sequence> to clarify the reason of the differences between the INSDC DTD v.1.5 and ST26 Sequence Listing DTD V1_1.

2016-03-24: Version 1.0 adopted at the CWS/4Bis

2014-03-11: Final draft for adoption.

ST26SequenceListing

* ROOT ELEMENT

-->

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

<!--The elements ApplicantName and InventorName are optional in this DTD to facilitate the conversion between various encoding schemes-->

<!--originalFreeTextLanguageCode:

The language code (see reference in paragraph 9 to ISO 639-1:2002) for the single original language in which the language-dependent free text qualifiers (NonEnglishQualifier_value) were prepared.

-->

<!--nonEnglishFreeTextLanguageCode:

The language code (see reference in paragraph 9 to ISO 639-1:2002) for the language in which the language-dependent free text qualifiers (NonEnglishQualifier_value) currently correspond.

-->

<!ATTLIST ST26SequenceListing

dtdVersion CDATA #REQUIRED

fileName CDATA #IMPLIED

softwareName CDATA #IMPLIED

softwareVersion CDATA #IMPLIED

productionDate CDATA #IMPLIED

originalFreeTextLanguageCode CDATA #IMPLIED

nonEnglishFreeTextLanguageCode CDATA #IMPLIED

>

<!--ApplicantFileReference

Applicant's or agent's file reference, mandatory if application identification not provided.

-->

<!ELEMENT ApplicantFileReference (#PCDATA)>

<!--ApplicationIdentification

Application identification for which the sequence listing is submitted, when available.

-->

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

<!--EarliestPriorityApplicationIdentification

Identification of the earliest priority application, which contains IPOfficeCode, ApplicationNumberText and FilingDate elements.

-->

<!ELEMENT EarliestPriorityApplicationIdentification (IPOfficeCode, ApplicationNumberText,

FilingDate?)>

<!--ApplicantName

The name of the first mentioned applicant in characters set forth in paragraph 40(a) of the ST.26 main body document.

-->

```

<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of
names of languages - Part 1: Alpha-2
-->
<!ELEMENT ApplicantName (#PCDATA)>
<!ATTLIST ApplicantName
        languageCode CDATA #REQUIRED
>
<!--ApplicantNameLatin
Where ApplicantName is typed in characters other than those as set forth in paragraph
40(b), a translation or transliteration of the name of the first mentioned applicant must
also be typed in characters as set forth in paragraph 40(b) of the ST.26 main body
document.
-->
<!ELEMENT ApplicantNameLatin (#PCDATA)>
<!--InventorName
Name of the first mentioned inventor typed in the characters as set forth in paragraph
40(a).-->
<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of
names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventorName (#PCDATA)>
<!ATTLIST InventorName
        languageCode CDATA #REQUIRED
>
<!--InventorNameLatin
Where InventorName is typed in characters other than those as set forth in paragraph 40(b),
a translation or transliteration of the first mentioned inventor may also be typed in
characters as set forth in paragraph 40(b).
-->
<!ELEMENT InventorNameLatin (#PCDATA)>
<!--InventionTitle
Title of the invention typed in the characters as set forth in paragraph 40(a) in the
language of filing. A translation of the title of the invention into additional languages
may be typed in the characters as set forth in paragraph 40(a) using additional
InventionTitle elements. The title of invention should be between two to seven words.
-->
<!--languageCode: Appropriate language code from ISO 639-1 - Codes
for the representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventionTitle (#PCDATA)>
<!ATTLIST InventionTitle
        languageCode CDATA #REQUIRED
>
<!--SequenceTotalQuantity
Indicates the total number of sequences in the document.
Its purpose is to be quickly accessible for automatic processing.
-->
<!ELEMENT SequenceTotalQuantity (#PCDATA)>
<!--SequenceData
Data for individual Sequence.
For intentionally skipped sequences see the ST.26 main body document.
-->
<!ELEMENT SequenceData (INSDSeq)>
<!ATTLIST SequenceData
        sequenceIDNumber CDATA #REQUIRED
>
<!--IPOfficeCode
ST.3 code. For example, if the application identification is PCT/IB2013/099999, then
IPOfficeCode value will be "IB" for the International Bureau of WIPO.
-->
<!ELEMENT IPOfficeCode (#PCDATA)>
<!--ApplicationNumberText
The application identification as provided by the office of filing (e.g. PCT/IB2013/099999)
-->
<!ELEMENT ApplicationNumberText (#PCDATA)>
<!--FilingDate

```

The date of filing of the patent application for which the sequence listing is submitted in ST.2 format "CCYY-MM-DD", using a 4-digit calendar year, a 2-digit calendar month and a 2-digit day within the calendar month, e.g., 2015-01-31. For details, please see paragraphs 7 (a) and 11 of WIPO Standard ST.2.

```
-->
<!ELEMENT FilingDate (#PCDATA)>
<!--*****
* INSD Part
*****-->
```

The purpose of the INSD part of this DTD is to define a customized DTD for sequence listings to support the work of IP offices while facilitating the data exchange with the public repositories.

The INSD part is subset of the INSD DTD v1.5 and as such can only be used to generate an XML instance as it will not support the complete INSD structure.

This part is based on:

The International Nucleotide Sequence Database (INSD) collaboration.

INSDSeq provides the elements of a sequence as presented in the GenBank/EMBL/DDBJ-style flatfile formats. Not all elements are used here.

```
-->
<!--INSDSeq
Sequence data. Changed INSD V1.5 DTD elements, INSDSeq_division and INSDSeq_sequence from optional to mandatory per business requirements.
-->
<!ELEMENT INSDSeq (INSDSeq_length, INSDSeq_moltype, INSDSeq_division, INSDSeq_other-seqids?, INSDSeq_feature-table?, INSDSeq_sequence)>
<!--INSDSeq_length
The length of the sequence. INSDSeq_length allows only integer.
-->
<!ELEMENT INSDSeq_length (#PCDATA)>
<!--INSDSeq_moltype
Admissible values: DNA, RNA, AA
-->
<!ELEMENT INSDSeq_moltype (#PCDATA)>
<!--INSDSeq_division
Indication that a sequence is related to a patent application. Must be populated with the value PAT.
-->
<!ELEMENT INSDSeq_division (#PCDATA)>
<!--INSDSeq_other-seqids
In the context of data exchange with database providers, the IPOs should populate for each sequence the element INSDSeq_other-seqids with one INSDSeqid containing a reference to the corresponding published patent and the sequence identification.
-->
<!ELEMENT INSDSeq_other-seqids (INSDSeqid?)>
<!--INSDSeq_feature-table
Information on the location and roles of various regions within a particular sequence. Whenever the element INSDSeq_feature-table is used, it must contain at least one feature.
-->
<!ELEMENT INSDSeq_feature-table (INSDFeature+)>
<!--INSDSeq_sequence
The residues of the sequence. The sequence must not contain numbers, punctuation or whitespace characters.
-->
<!ELEMENT INSDSeq_sequence (#PCDATA)>
<!--INSDSeqid
Intended for the use of IPOs in data exchange only.
```

Format:
pat|{office code}|{publication number}|{document kind code}|{Sequence identification number}

where office code is the code of the IP office publishing the patent document, publication number is the publication number of the application or patent, document kind code is the letter codes to distinguish patent documents as defined in ST.16 and Sequence identification number is the number of the sequence in that application or patent

Example:

```
pat|WO|2013999999|A1|123456
```

This represents the 123456th sequence from WO patent publication No. 2013999999 (A1)

```
-->
<!ELEMENT INSDSeqid (#PCDATA)>
<!--INSDFeature
Description of one feature.
-->
<!ELEMENT INSDFeature (INSDFeature_key, INSDFeature_location, INSDFeature_qual?)>
<!--INSDFeature_key
A word or abbreviation indicating a feature.
-->
<!ELEMENT INSDFeature_key (#PCDATA)>
<!--INSDFeature_location
Region of the presented sequence which corresponds to the feature.
-->
<!ELEMENT INSDFeature_location (#PCDATA)>
<!--INSDFeature_qual
List of qualifiers containing auxiliary information about a feature.
-->
<!ELEMENT INSDFeature_qual (INSDQualifier+)>
<!--INSDQualifier
Additional information about a feature.
For coding sequences and variants see the ST.26 main body document.
-->
<!--id
Unique identifier for the INSDQualifier to facilitate comparison of versions of a sequence
listing specifically having language-dependent qualifier values in different languages.
-->
<!ELEMENT INSDQualifier (INSDQualifier name, INSDQualifier value?,
NonEnglishQualifier value?)>
<ATTLIST INSDQualifier
            id ID #IMPLIED
>
<!--INSDQualifier name
Name of the qualifier.
-->
<!ELEMENT INSDQualifier_name (#PCDATA)>
<!--INSDQualifier value
Value of the qualifier. Where the qualifier is language-dependent its value must be in the
English language and typed with the characters set forth in paragraph 40 (b).
-->
<!ELEMENT INSDQualifier_value (#PCDATA)>
<!--NonEnglishQualifier value
Value of a language-dependent qualifier in a language that is not English and typed with
the characters set forth in paragraph 40 (a). The language is indicated with the attribute
nonEnglishFreeTextLanguageCode.
-->
<!ELEMENT NonEnglishQualifier_value (#PCDATA)>
```

附錄 III

序列表例示(XML文件)

版本 1.4

WIPO標準委員會(CWS)於2021年11月5日第九屆會議上批准修訂

附錄III內容見於：https://www.wipo.int/standards/en/xml_material/st26/st26-annex-iii-sequence-listing-specimen.xml

附錄 IV

用於XML序列表的Unicode基本拉丁文代碼子集

版本 1.3

WIPO標準委員會(CWS) 於2020年12月4日第八屆會議上批准修訂

符號' & ' (Unicode編碼為0026)只能作為預定義實體的一部分使用。引號(Unicode編碼為0022)、撇號(Unicode編碼為0027)、小於符號(Unicode編碼為003C)及大於符號(Unicode編碼為003E)必須以他們的預定義實體表示。此外，在屬性值或要素內容中作為' & ' 符號使用時，必須使用其預定義的實體代替符號' & ' 。

Unicode code point	Character	Name
0020		SPACE
0021	!	EXCLAMATION MARK
0022	"	QUOTATION MARK
0023	#	NUMBER SIGN
0024	\$	DOLLAR SIGN
0025	%	PERCENT SIGN
0026	&	AMPERSAND
0027	'	APOSTROPHE
0028	(LEFT PARENTHESIS
0029)	RIGHT PARENTHESIS
002A	*	ASTERISK
002B	+	PLUS SIGN
002C	,	COMMA
002D	-	HYPHEN-MINUS
002E	.	FULL STOP
002F	/	SOLIDUS
0030	0	DIGIT ZERO
0031	1	DIGIT ONE
0032	2	DIGIT TWO
0033	3	DIGIT THREE
0034	4	DIGIT FOUR
0035	5	DIGIT FIVE
0036	6	DIGIT SIX
0037	7	DIGIT SEVEN
0038	8	DIGIT EIGHT
0039	9	DIGIT NINE
003A	:	COLON
003B	;	SEMICOLON
003C	<	LESS-THAN-SIGN
003D	=	EQUALS SIGN
003E	>	GREATER-THAN-SIGN
003F	?	QUESTION MARK
0040	@	COMMERCIAL AT
0041	A	LATIN CAPITAL LETTER A
0042	B	LATIN CAPITAL LETTER B
0043	C	LATIN CAPITAL LETTER C
0044	D	LATIN CAPITAL LETTER D
0045	E	LATIN CAPITAL LETTER E
0046	F	LATIN CAPITAL LETTER F
0047	G	LATIN CAPITAL LETTER G
0048	H	LATIN CAPITAL LETTER H
0049	I	LATIN CAPITAL LETTER I

Unicode code point	Character	Name
004A	J	LATIN CAPITAL LETTER J
004B	K	LATIN CAPITAL LETTER K
004C	L	LATIN CAPITAL LETTER L
004D	M	LATIN CAPITAL LETTER M
004E	N	LATIN CAPITAL LETTER N
004F	O	LATIN CAPITAL LETTER O
0050	P	LATIN CAPITAL LETTER P
0051	Q	LATIN CAPITAL LETTER Q
0052	R	LATIN CAPITAL LETTER R
0053	S	LATIN CAPITAL LETTER S
0054	T	LATIN CAPITAL LETTER T
0055	U	LATIN CAPITAL LETTER U
0056	V	LATIN CAPITAL LETTER V
0057	W	LATIN CAPITAL LETTER W
0058	X	LATIN CAPITAL LETTER X
0059	Y	LATIN CAPITAL LETTER Y
005A	Z	LATIN CAPITAL LETTER Z
005B	[LEFT SQUARE BRACKET
005C	\	REVERSE SOLIDUS
005D]	RIGHT SQUARE BRACKET
005E	^	CIRCUMFLEX ACCENT
005F	_	LOW LINE
0060	`	GRAVE ACCENT
0061	a	LATIN SMALL LETTER A
0062	b	LATIN SMALL LETTER B
0063	c	LATIN SMALL LETTER C
0064	d	LATIN SMALL LETTER D
0065	e	LATIN SMALL LETTER E
0066	f	LATIN SMALL LETTER F
0067	g	LATIN SMALL LETTER G
0068	h	LATIN SMALL LETTER H
0069	i	LATIN SMALL LETTER I
006A	j	LATIN SMALL LETTER J
006B	k	LATIN SMALL LETTER K
006C	l	LATIN SMALL LETTER L
006D	m	LATIN SMALL LETTER M
006E	n	LATIN SMALL LETTER N
006F	o	LATIN SMALL LETTER O
0070	p	LATIN SMALL LETTER P
0071	q	LATIN SMALL LETTER Q
0072	r	LATIN SMALL LETTER R
0073	s	LATIN SMALL LETTER S
0074	t	LATIN SMALL LETTER T
0075	u	LATIN SMALL LETTER U
0076	v	LATIN SMALL LETTER V
0077	w	LATIN SMALL LETTER W
0078	x	LATIN SMALL LETTER X
0079	y	LATIN SMALL LETTER Y
007A	z	LATIN SMALL LETTER Z
007B	{	LEFT CURLY BRACKET
007C		VERTICAL LINE
007D	}	RIGHT CURLY BRACKET
007E	~	TILDE

附錄 V

附加數據交換要件(僅供IPOs)

版本 1.4

WIPO標準委員會(CWS)於2021年11月5日第九屆會議上批准修訂

在與數據庫提供者(INSD成員)進行數據交換的背景下,建議智慧財產局為每個序列添加(populate)要素INSDSeq_other-seqids,其中包含一個INSDSeqid,記載引用(a reference to)對應的已公開專利及序列識別號,其格式如下:

`pat{office code}|{publication number}|{document kind code}|{sequence identification number}`

其中,office code是根據ST.3規定發布專利文件的智慧財產局代碼;
document kind code是用於識別不同種類專利文件的代碼,其格式由ST.16規定;
publication number是專利申請案或專利案的發布編號;
Sequence identification number是該專利申請案或專利案中該序列的編號。

實例:

`pat|WO|2013999999|A1|123456`

轉換為一個有效的XML實例,格式如下:

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

其中,“123456”是來自WO公開號 2013999999 (A1)的第123456個序列。

附錄 VI

例示指南文件

版本 1.5

WIPO標準委員會(CWS)於2021年11月5日第九屆會議上批准修訂

目錄

簡介.....	1
例示索引.....	8
例示.....	19
附件.....	89

簡介

本標準目的包含「允許申請人在專利申請中製作一份國際及國家或地區程序接受的序列表」。為此，以本附錄確保申請人及智慧財產局(IPO)對於序列揭露事項及表示方式要求均能理解並同意。

本附錄包括簡介、例示索引、序列揭露例示及包含例示序列的XML格式序列表的附錄，其中，「簡介」說明本附錄所使用的概念及術語；「例示」闡明本標準中特定段落的要求，並於每個例示標明其最相關段落的編號，當例示中進一步涉及其他段落時，於各該例示的文末指明其交叉引用；「索引」提供例示及其交叉引用的頁碼；「附錄」則以XML格式呈現本附錄的例示，且各例示均指定其專屬序列識別號(SEQ ID NO)。

在例示中，解釋性資訊與視為完整揭露的序列一起呈現，而所供答案僅考慮例示中明確提供的資訊。

本附錄的指導係針對專利申請日提交的序列表的規定。而在專利申請日後提交的序列表，由於另須考量所提供的資訊是否被IPO認為可加到原始揭露中，因此，本附錄的指導可能不適用於在專利申請日後提交的序列表。

序列表的準備

在準備專利申請案的序列表時須考慮下列問題：

- 一、 是否為ST. 26第7段要求揭露的序列？
- 二、 若非ST. 26要求揭露的序列，ST. 26是否允許包含該序列？

三、若為ST. 26要求或允許揭露的序列，該序列於序列表中建議如何表示？

關於問題一，ST. 26第7段要求揭露在專利申請案中列舉其殘基的序列，且該序列具有10個或更多具體定義核苷酸或4個或更多具體定義胺基酸。

關於問題二，ST. 26第8段禁止序列表包含任何少於10個具體定義核苷酸或少於4個具體定義胺基酸的序列。

要回答問題一及二，必須清楚瞭解「殘基列舉」及「具體定義」。

關於問題三，本附錄依據本標準提供例示序列，舉例說明各種序列態樣或一序列包含多個變異的「最具包容性的序列」，並討論各種序列態樣的序列表示方式。由於不可能列出所有的序列態樣，本附錄藉由闡明各個例示立意及ST. 26規定的應用，以利推論至其他未列舉的序列態樣。

殘基列舉

ST. 26第3(c)段將「殘基列舉」定義為透過依序列出的各殘基來揭露專利申請中的序列，其中(i)殘基由名稱、縮寫、符號或結構表示；或者(ii)多個殘基由簡寫式表示。在專利申請案中以「殘基列舉」揭露的序列應使用慣用符號，而慣用符號指ST. 26附錄I第1節表1所列核苷酸符號(小寫符號或其同義的大寫符號²)及ST. 26附錄I第3節表3所列胺基酸符號(大寫符號或其同義的小寫符號)。以下將這些表所列核苷酸及胺基酸符號稱為慣用符號，並以「非慣用」指稱未載於這些表的核苷酸及胺基酸。

在判定一序列是否為ST. 26第7段要求揭露在序列表或第8段禁止包含在序列表時，若其中的殘基被揭露為等同於慣用符號或縮寫(例如：「Z₁」係指「A」)或等同於慣用符號組成的特定序列(例如：「Z₁」係指「agga」)，則該序列被視為是以等同的慣用符號或縮寫公開。當非慣用核苷酸符號用作歧義符號(例如：X₁ = 肌苷或假尿苷)，且非表1第1節所列慣用歧義符號(即m、r、w、s、y、k、v、h、d、b或n)，則在判定序列是否為ST. 26第7段要求揭露或第8段禁止包含於序列表時，該殘基被視為「n」；同理，當非慣用胺基酸符號用作歧義符號(例如：Z₁為A、G、S或T)，且非第3節表3所列的慣用歧義符號(即B、Z、J或X)，則在判定序列是否為ST. 26第7段要求揭露或第8段禁止包含於序列表時，該殘基被視為「X」。

揭露序列應注意使用慣用符號，否則需檢視揭露內容以確認非慣用表示的含義。而在使用慣用符號的情況下，仍需檢視揭露內容對該序列的解釋，以確認是否以慣用方式使用符號。若以非慣用方式使用符號，則仍需判定該序列是否為ST. 26第7段要求揭露或第8段禁止包含於序列表的序列。

² 注意：雖然申請案中的序列揭露可用小寫或大寫符號表示核苷酸或胺基酸，但在序列表中只能以小寫字母表示核苷酸序列(見ST. 26第13段)及大寫字母表示胺基酸序列(見ST. 26第26段)。

具體定義

ST. 26第3(k)段界定「具體定義」為附錄I中除了以符號「n」所表示者外的任意核苷酸及除了以符號「X」所表示者以外的任意胺基酸，其中「n」及「X」分別列於第1節表1(即a、c、g、t/u、未知或其他)及第3節表3(即A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y、V、未知或其他)。在判定核苷酸或胺基酸是否為「具體定義」時，上述有關慣用符號、非慣用符號或縮寫及其以慣用或非慣用方式使用等都要加以考慮。

最包容的序列

當一申請案中揭露的序列符合第7段殘基列舉的要求，但在不同實施例有不同描述，例如：在一實施例中，其一或多個位置的「X」指任何胺基酸，但在另一實施例中，其「X」指特定幾個胺基酸，則ST. 26要求序列表中僅包含已列舉殘基的序列。根據前述第15、27段，若序列包含多個歧義符號「n」或「X」且未進一步定義，則「n」或「X」分別代表任何核苷酸或胺基酸。因此，需包含於序列表的序列為最具包容性的序列，而最具包容性的序列指包含於最多實施例，且具有以最具限制性歧義符號表示變異殘基的單一序列。

此外，當序列僅以殘基列舉揭露一次，但序列長度因重複次數不同而變化時，其最長的序列為最具包容性的序列。例如：一序列包含重複2-5次的重複區域，其具有5個次重複的例示即為最具包容性的序列，需包含於序列表中。然而，強烈鼓勵落實包含附加特定序列，例如：為發明關鍵部分的附加例示的序列。包含附加序列可使專利檢索更徹底，且可供作權利主張的公示。

歧義符號的使用

序列表中歧義符號「n」的正確使用

符號「n」

- (一) 不得用於表示單個核苷酸以外的物。
- (二) 若無進一步界定，則為a、c、g或t/u中之任一者。
- (三) 用於代表下列具有進一步描述的核苷酸：
 - 1、 不能用附錄I所列其他符號表示(參見第1節表1)的修飾核苷酸，例如：天然、合成或非天然存在。
 - 2、 「未知」核苷酸，即未確定、未揭露或不確定。
 - 3、 一殘基位點。
- (四) 用於表示序列變體，即替代、刪除、插入或替換，其中「n」是最嚴格的歧義符號。

序列表中歧義符號「X」的正確使用

符號「X」

- (一) 不得用於表示單個胺基酸以外的物；
- (二) 若無進一步界定，則為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V中之任一者；
- (三) 用於代表下列具有進一步的描述的胺基酸：
 - 1、 不能用附錄I所列其他符號表示(參見第3節表3)的修飾胺基酸，例如：天然、合成或非天然存在。
 - 2、 「未知」胺基酸，即未確定、未揭露或不確定。
- (四) 用於表示序列變體，即替代、刪除、插入或替換，其中「X」是最嚴格的歧義符號。

修飾殘基的註釋

本標準對於「修飾」殘基的要求包含第17段的核苷酸註釋及第30段的胺基酸註釋。另外，第3(e)段定義「修飾胺基酸」為第3(a)段所述L-丙胺酸、L-精胺酸、L-天冬醯胺酸、L-天冬胺酸、L-半胱胺酸、L-麩醯胺酸、L-麩胺酸、L-甘胺酸、L-組胺酸、L-白胺酸異白胺酸、L-白胺酸、L-離胺酸、L-甲硫氨酸、L-苯丙胺酸、L-脯胺酸、L-吡咯離胺酸、L-絲胺酸、L- 硒代半胱胺酸、L-蘇胺酸、L-色胺酸、L-酪胺酸或 L-纈胺酸以外的任意胺基酸；而第3(f)段定義「修飾核苷酸」為第3(g)段所述deoxyadenosine 3' - monophosphate、deoxyguanosine 3' - monophosphate、deoxycytidine 3' - monophosphate、deoxythymidine 3' - monophosphate、adenosine 3' - monophosphate、guanosine 3' - monophosphate、cytidine 3' - monophosphate或uridine 3' - monophosphate以外的任意核苷酸。

根據上述定義，修飾核酸的核鹼基或糖-磷酸骨架或修飾肽的胺基酸R基團或肽骨架產生一或多個「修飾核苷酸」或「修飾胺基酸」，因此，對此類核苷酸及胺基酸需加以註釋。骨架修飾的例示包括核苷酸類似物，例如：肽核酸(PNA)、乙二醇核酸(GNA)及D-胺基酸。

須注意的是，肽末端胺基酸或核酸末端核苷酸的修飾不一定導致「修飾核苷酸」或「修飾胺基酸」，需視該末端修飾並判定該修飾是否改變殘基的化學結構，從而使殘基不屬於第3(e)及3(f)段所述例外情況，例如：使用醯胺鍵連接C末端殘基至另一結構的肽(例如分支序列的一部分-見例示7(b)-3的第2個肽)不被視為「修飾殘基」，因此毋需註釋；同理，使用醯胺鍵結合N末端殘基與生物素的肽不被視為「修飾殘基」，亦毋需註釋。這兩種情形的C端或N端殘基，未改變本標

準第3(e)段所列慣用胺基酸的結構。

相對地，末端修飾若改變殘基的化學結構，則為「修飾殘基」，須加以註釋，例如：例示3(c)-1的C末端甲基化，以甲基取代通常存在於 α 羧基上的羥基，其已改變末端殘基的化學結構，因此，該甲基化離胺酸必須標註為「修飾殘基」。

請注意，申請人自行評估列舉序列的末端殘基修飾及判定末端殘基的結構是否發生變化。若修飾殘基的結構不同於標準第3(e)、3(f)段所示慣用胺基酸或核苷酸，則該修飾須加以註釋。

最後，建議申請人提供盡量多資訊於其序列表，以儘可能準確揭露，因此，即使是不需要註釋的修飾，仍建議包含於其中。

須注意的是，於變體列舉的註釋，其一級序列須符合ST. 26第93-100段之要求。修飾若被揭露為序列列舉的變體，則可能不要求包含於序列表。有關變體註釋的判定，請參見ST. 26第93-95段。

修飾殘基的表示

ST. 26 指出修飾的核苷酸及胺基酸，建議於序列表中以相應未修飾殘基表示（參見第16、29段）。請注意，此建議事項係「建議」，即「強烈鼓勵，但不要求」（參見第4(d)段），由申請人自行決定修飾殘基是以相應未修飾殘基表示或以「n」或「X」表示。

經驗上，當殘基透過添加部分進行修飾，例如甲基化或乙醯化，其相對應的未修飾殘基的基本結構通常未改變，則建議使用未修飾殘基表示，例如：甲基化的腺苷在序列表中建議以「a」表示。而當修飾的殘基在結構上不同於任何未修飾的殘基時，建議使用「n」或「X」。例如：正白胺酸為白胺酸的異構物，其中，正白胺酸的側鏈為4個碳的直鏈結構，而白胺酸的4碳側鏈於第二個碳上有分支，因此，正白胺酸並非單純對白胺酸修飾，而是完全不同結構，故建議在序列表中以「X」表示正白胺酸。

「具體定義」核苷酸係指除了以符號「n」所表示者外的任意核苷酸，而「具體定義」胺基酸係指除了以符號「X」所表示者以外的任意胺基酸（參見ST. 26第3(k)段），因此，序列中以「a」表示的2' O-methyl adenosine(2' O-甲基腺苷)為具體定義核苷酸，而序列中以「X」表示的norleucine(正白胺酸)則非具體定義胺基酸。

表A - 核苷酸符號及定義

符號	定義
a	Adenine (腺嘌呤)
c	Cytosine (胞嘧啶)
g	Guanine (鳥嘌呤)
t	在DNA序列中指胸腺嘧啶； t/u在RNA序列中指尿嘧啶
m	a或c
r	a或g
w	a或t/u
s	c或g
y	c或t/u
k	g或t/u
v	a或c或g；非t/u
h	a或c或t/u；非g
d	a或g或t/u；非c
b	c或g或t/u；非a
n	a或c或g或t/u；「未知」或「其他」

表B - 胺基酸符號、三字母代碼及定義

符號	三字母代碼	定義
A	Ala	Alanine (丙胺酸)
R	Arg	Arginine (精胺酸)
N	Asn	Asparagine (天冬醯胺酸)
D	Asp	Aspartic Acid (Aspartate) (天冬胺酸)
C	Cys	Cysteine (半胱胺酸)
Q	Gln	Glutamine (麩醯胺酸)
E	Glu	Glutamic Acid (Glutamate) (麩胺酸)
G	Gly	Glycine (甘胺酸)
H	His	Histidine (組胺酸)
I	Ile	Isoleucine (異白胺酸)
L	Leu	Leucine (白胺酸)
K	Lys	Lysine (離胺酸)
M	Met	Methionine (甲硫胺酸)
F	Phe	Phenylalanine (苯丙胺酸)
P	Pro	Proline (脯胺酸)
O	Pyl	Pyrrolysine (吡咯離胺酸)
S	Ser	Serine (絲胺酸)
U	Sec	Selenocysteine (硒代半胱胺酸)
T	Thr	Threonine (蘇胺酸)
W	Trp	Tryptophan (色胺酸)
Y	Tyr	Tyrosine (酪胺酸)
V	Val	Valine (纈胺酸)
B	Asx	Aspartic Acid或Asparagine
Z	Glx	Glutamine或Glutamic Acid
J	Xle	Leucine或Isoleucine
X	Xaa	A或R或N或D或C或Q或E或G或H或I或L或K或M或F或P或O或S或U或T或W或Y或V，「未知」或「其他」

例示索引

	頁碼
<u>第3(a)段「胺基酸」的定義</u>	
例示 3(a)-1：D-胺基酸.....	19
<u>交叉引用的例子</u>	
例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64
例示 29-2：使用相對應未修飾胺基酸.....	65
例示 30-1：特徵關鍵「CARBOHYD」.....	67
<u>第3(c)段—「列舉殘基」的定義</u>	
例示 3(c)-1：以化學結構列舉胺基酸.....	20
例示 3(c)-2：胺基酸序列的簡寫式.....	22
<u>交叉引用的例子</u>	
例示 27-1：胺基酸序列的簡寫式.....	57
例示 27-3：簡寫式—4個或更多具體定義胺基酸.....	60
<u>第3(f)段—「修飾核苷酸」的定義</u>	
<u>交叉引用的例子</u>	
例示 3(g)-4：核酸類似物.....	27
<u>第3(g)段「核苷酸」的定義</u>	
例示 3(g)-1：被C3間隔子中斷的核苷酸序列.....	23
例示 3(g)-2：具有替代殘基的核苷酸序列，包含C3間隔子.....	24
例示 3(g)-3：無鹼基位點.....	26
例示 3(g)-4：核酸類似物.....	27
<u>交叉引用的例子</u>	
例示 11(b)-1：雙股核苷酸序列——不同長度.....	52
例示 14-1：符號「t」代表RNA的尿嘧啶.....	55
<u>第3(k)段「具體定義」的定義</u>	
例示 3(k)-1：核苷酸歧義符號.....	28
例示 3(k)-2：以慣用及非慣用方式使用的歧義符號「n」.....	29
例示 3(k)-3：以非慣用方式使用的歧義符號「n」.....	31
例示 3(k)-4：「n」以外的歧義符號為「具體定義」.....	32

例示 3(k)-5：以非慣用方式使用的歧義縮寫「Xaa」.....	33
-----------------------------------	----

第7段—須列入序列表的序列

交叉引用的例子

例示 28-1：編碼核苷酸序列及編碼胺基酸序列.....	62
例示 55-1：DNA/RNA組合分子.....	74
例示 89-2：特徵位置超出揭露的序列.....	76
例示 92-1：由具有內含子的編碼序列編碼的胺基酸序列.....	78

第7(a)段—序列表中核苷酸序列的要求

例示 7(a)-1：分支核苷酸序列.....	34
例示 7(a)-2：具有二級結構的線性核苷酸序列.....	36
例示 7(a)-3：以非慣用方式使用的核苷酸歧義符號.....	37
例示 7(a)-4：以非慣用方式使用的核苷酸歧義符號.....	38
例示 7(a)-5：非慣用核苷酸符號.....	39
例示 7(a)-6：非慣用核苷酸符號.....	40

交叉引用的例子

例示 3(g)-1：被 C3間隔子中斷的核苷酸序列.....	23
例示 3(g)-2：具有替代殘基的核苷酸序列，包含C3間隔子.....	24
例示 3(g)-3：無鹼基位點.....	26
例示 3(g)-4：核酸類似物.....	27
例示 3(k)-1：核苷酸歧義符號.....	28
例示 3(k)-2：以慣用及非慣用方式使用的歧義符號「n」.....	29
例示 3(k)-3：以非慣用方式使用的歧義符號「n」.....	31
例示 3(k)-4：「n」以外的歧義符號為「具體定義」.....	32
例示 11(a)-1：雙股核苷酸序列——相同長度.....	51
例示 11(b)-1：雙股核苷酸序列——不同長度.....	52
例示 11(b)-2：雙股核苷酸序列——無鹼基配對片段.....	54
例示 14-1：符號「t」代表RNA的尿嘧啶.....	55
例示 89-1：編碼核苷酸序列及編碼胺基酸序列.....	75
例示 93-1：列舉變體的表示.....	80
例示 95(b)-1：以多個非獨立變變數表示獨立變體序列.....	87

第7(b)段—序列表中胺基酸序列的要求

例示 7(b)-1：4個或更多具體定義胺基酸.....	42
例示 7(b)-2：分支胺基酸序列.....	43
例示 7(b)-3：分支胺基酸序列.....	46
例示 7(b)-4：含有支鏈胺基酸序列的環肽.....	47
例示 7(b)-5：含有支鏈胺基酸序列的環肽.....	49

交叉引用的例子

例示 3(a)-1：D-胺基酸.....	19
例示 3(c)-1：以化學結構列舉胺基酸.....	20
例示 3(c)-2：胺基酸序列的簡寫式.....	22
例示 3(k)-5：以非慣用方式使用的歧義縮寫「Xaa」.....	33
例示 27-1：胺基酸序列的簡寫式.....	57
例示 27-3：簡寫式—4個或更多具體定義胺基酸.....	60
例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64
例示 29-2：使用相對應未修飾胺基酸.....	65
例示 30-1：特徵關鍵「CARBOHYD」..	67
例示 36-1：具有一個已知「X」殘基數量區域的序列表示為獨立序列.....	68
例示 37-1：具有未知數量「X」殘基區域的序列不得表示為獨立序列.....	72
例示 37-2：具有未知數量「X」殘基區域的序列不得表示為獨立序列.....	73
例示 89-1：編碼核苷酸序列及編碼胺基酸序列.....	75
例示 93-2：列舉變體的表示.....	81
例示 93-3：共有序列的表示.....	82
例示 94-1：含有列舉替代胺基酸的獨立序列的表示方式.....	84
例示 95(a)-1：透過主要序列的註釋表示變體序列.....	85

第8段—包含序列的閾值

交叉引用的例子

例示 3(k)-1：核苷酸歧義符號.....	28
例示 3(k)-2：以慣用及非慣用方式使用的歧義符號「n」.....	29
例示 7(a)-1：分支核苷酸序列.....	34
例示 7(a)-6：非慣用核苷酸符號.....	40
例示 7(b)-1：4個或更多具體定義胺基酸.....	42
例示 7(b)-2：分支胺基酸序列.....	43
例示 7(b)-4：含有支鏈胺基酸序列的環肽.....	47

例示 14-1：符號「t」代表RNA的尿嘧啶.....	55
例示 37-1：具有未知數量「X」殘基區域的序列不得表示為獨立序列.....	72
例示 37-2：具有未知數量「X」殘基區域的序列不得表示為獨立序列.....	73
例示 94-1：含有列舉替代胺基酸的獨立序列的表示方式.....	84

第11段—核苷酸序列的表示

交叉引用的例子

例示 3(g)-4：核酸類似物.....	27
例示 7(a)-1：分支核苷酸序列.....	34

第11(a)段—雙股核苷酸序列——完全互補

例示 11(a)-1：雙股核苷酸序列——相同長度.....	51
-------------------------------	----

第11(b)段—雙股核苷酸序列——不完全互補

例示 11(b)-1：雙股核苷酸序列——不同長度.....	52
例示 11(b)-2：雙股核苷酸序列——無鹼基配對片段.....	54

第13段—核苷酸的表示

交叉引用的例子

例示 3(k)-2：以慣用及非慣用方式使用的歧義符號「n」.....	29
例示 7(a)-1：分支核苷酸序列.....	34
例示 14-1：符號「t」代表RNA的尿嘧啶.....	55
例示 93-1：列舉變體的表示.....	80

第14段—符號「t」在RNA中解釋為尿嘧啶

例示 14-1：符號「t」代表RNA的尿嘧啶.....	55
-----------------------------	----

交叉引用的例子

例示 55-1：DNA/RNA組合分子.....	74
--------------------------	----

第15段—建議使用最嚴格的核苷酸歧義符號

交叉引用的例子

例示 3(g)-1：被C3間隔子中斷的核苷酸序列.....	23
例示 3(g)-2：具有替代殘基的核苷酸序列，包含C3間隔子.....	24
例示 3(k)-4：「n」以外的歧義符號為「具體定義」.....	32

例示 95(b)-1：以多個非獨立變變數表示獨立變體序列.....	87
-----------------------------------	----

第16段—修飾核苷酸的表示

交叉引用的例子

例示 3(g)-1：被C3間隔子中斷的核苷酸序列.....	23
例示 3(g)-4：核酸類似物.....	27

第17段—修飾核苷酸的註釋

交叉引用的例子

例示 3(g)-1：被C3間隔子中斷的核苷酸序列.....	23
例示 3(g)-3：無鹼基位點.....	26
例示 7(a)-1：分支核苷酸序列.....	34
例示 7(a)-2：具有二級結構的線性核苷酸序列.....	36
例示 7(a)-6：非慣用核苷酸符號.....	40

第18段—連續修飾核苷酸區域的註釋

交叉引用的例子

例示 3(g)-4：核酸類似物.....	27
例示 11(b)-1：雙股核苷酸序列——不同長度.....	52

第19段—DNA中的尿嘧啶或 RNA中的胸腺嘧啶的註釋

交叉引用的例子

例示 14-1：符號「t」代表RNA的尿嘧啶.....	55
-----------------------------	----

第25段—第1個胺基酸序列殘基

交叉引用的例子

例示 3(a)-1：D-胺基酸.....	19
例示 7(b)-4：含有支鏈胺基酸序列的環肽.....	47
例示 7(b)-5：含有支鏈胺基酸序列的環肽.....	49
例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64

第26段—胺基酸的表示

交叉引用的例子

例示 7(b)-2：分支胺基酸序列.....	43
------------------------	----

例示 7(b)-4：含有支鏈胺基酸序列的環肽.....	47
例示 7(b)-5：含有支鏈胺基酸序列的環肽.....	49
例示 36-1：具有一個已知「X」殘基數量區域的序列表示為獨立序列.....	68
例示 89-1：編碼核苷酸序列及編碼胺基酸序列.....	75
例示 92-1：由具有內含子的編碼序列編碼的胺基酸序列.....	78
例示 93-2：列舉變體的表示.....	81
例示 93-3：共有序列的表示.....	82

第27段—建議使用最嚴格胺基酸歧義符號

例示 27-1：胺基酸序列的簡寫式.....	57
例示 27-2：簡寫式—少於4個具體定義胺基酸.....	58
例示 27-3：簡寫式—4個或更多具體定義胺基酸.....	60

交叉引用的例子

例示 3(c)-2：胺基酸序列的簡寫式.....	22
例示 7(b)-1：4個或更多具體定義胺基酸.....	42
例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64
例示 36-1：具有一個已知「X」殘基數量區域的序列表示為獨立序列.....	68
例示 36-2：具有多個已知數量或範圍「X」殘基區域的序列表示為獨立序列.....	69
例示 36-3：具有多個已知數量或範圍「X」殘基區域的序列表示為獨立序列.....	70
例示 37-2：具有未知數量「X」殘基區域的序列不得表示為獨立序列.....	73
例示 93-3：共有序列的表示.....	82
例示 94-1：含有列舉替代胺基酸的獨立序列的表示方式.....	84
例示 95(a)-1：透過主要序列的註釋表示變體序列.....	85

第28段—由內部終止符分隔的胺基酸序列

例示 28-1：編碼核苷酸序列及編碼胺基酸序列.....	62
------------------------------	----

交叉引用的例子

例示 89-1：編碼核苷酸序列及編碼胺基酸序列.....	75
例示 92-1：由具有內含子的編碼序列編碼的胺基酸序列.....	78

第29段—「其他」胺基酸的表示

例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64
例示 29-2：使用相對應未修飾胺基酸.....	65

交叉引用的例子

例示 3(a)-1：D-胺基酸.....	19
例示 7(b)-2：分支胺基酸序列.....	43
例示 7(b)-3：分支胺基酸序列.....	46
例示 7(b)-4：含有支鏈胺基酸序列的環肽.....	47
例示 30-1：特徵關鍵「CARBOHYD」.....	67

第30段—修飾胺基酸的註釋

例示 30-1：特徵關鍵「CARBOHYD」.....	67
-----------------------------	----

交叉引用的例子

例示 3(a)-1：D-胺基酸.....	19
例示 3(c)-1：以化學結構列舉胺基酸.....	20
例示 7(b)-2：分支胺基酸序列.....	43
例示 7(b)-3：分支胺基酸序列.....	46
例示 7(b)-4：含有支鏈胺基酸序列的環肽.....	47
例示 7(b)-5：含有支鏈胺基酸序列的環肽.....	49
例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64
例示 29-2：使用相對應未修飾胺基酸.....	65

第31段—D-胺基酸的表示

交叉引用的例子

例示 3(a)-1：D-胺基酸.....	19
例示 3(c)-1：以化學結構列舉胺基酸.....	20
例示 7(b)-2：分支胺基酸序列.....	43
例示 7(b)-3：分支胺基酸序列.....	46
例示 7(b)-4：含有支鏈胺基酸序列的環肽.....	47
例示 7(b)-5：含有支鏈胺基酸序列的環肽.....	49

第32段—「未知」胺基酸的註釋

交叉引用的例子

例示 3(c)-1：以化學結構列舉胺基酸.....	20
---------------------------	----

第34段—連續「X」殘基區域的註釋

交叉引用的例子

例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64
------------------------------	----

第36段—包含精確數量連續「n」或「X」殘基區域的序列

例示 36-1：具有一個已知「X」殘基數量區域的序列表示為獨立序列..... 68

例示 36-2：具有多個已知數量或範圍「X」殘基區域的序列表示為獨立序列. 69

例示 36-3：具有多個已知數量或範圍「X」殘基區域的序列表示為獨立序列. 70

第37段—包含未知數目連續「n」或「X」殘基區域的序列

例示 37-1：具有未知數量「X」殘基區域的序列不得表示為獨立序列..... 72

例示 37-2：具有未知數量「X」殘基區域的序列不得表示為獨立序列..... 73

第41段—保留字元

交叉引用的例子

例示 89-2：特徵位置超出揭露的序列..... 76

第54段—INSDSeq_moltype 要素

交叉引用的例子

例示 14-1：符號「t」代表RNA的尿嘧啶..... 55

第55段—包含DNA及RNA片段的核苷酸序列

例示 55-1：DNA/RNA組合分子..... 74

第56段—例示說明包含DNA及RNA片段的核苷酸序列

交叉引用的例子

例示 55-1：DNA/RNA組合分子..... 74

第57段—INSDSeq_sequence 要素

交叉引用的例子

例示 28-1：編碼核苷酸序列及編碼胺基酸序列..... 62

例示 92-1：由具有內含子的編碼序列編碼的胺基酸序列..... 78

第65段—位置描述

交叉引用的例子

例示 3(g)-4：核酸類似物..... 27

例示 89-2：特徵位置超出揭露的序列..... 76

第66段—位置描述語法

交叉引用的例子

例示 3(g)-4：核酸類似物.....	27
例示 7(b)-4：含有支鏈胺基酸序列的環肽.....	47
例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64
例示 30-1：特徵關鍵「CARBOHYD」.....	67
例示 89-2：特徵位置超出揭露的序列.....	76

第67段—位置操作元件

交叉引用的例子

例示 92-1：由具有內含子的編碼序列編碼的胺基酸序列.....	78
----------------------------------	----

第70段—特徵位置

交叉引用的例子

例示 7(b)-4：含有支鏈胺基酸序列的環肽.....	47
例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64
例示 30-1：特徵關鍵「CARBOHYD」.....	67
例示 89-2：特徵位置超出揭露的序列.....	76

第71段—位置描述中字元「<」及「>」的表示

交叉引用的例子

例示 29-1：「其他」胺基酸的最嚴格歧義符號.....	64
例示 89-2：特徵位置超出揭露的序列.....	76

第83段—例示說明非天然存在核苷酸序列

交叉引用的例子

例示 55-1：DNA/RNA組合分子.....	74
--------------------------	----

第89段—特徵關鍵「CDS」

例示 89-1：編碼核苷酸序列及編碼胺基酸序列.....	75
例示 89-2：特徵位置超出揭露的序列.....	76

交叉引用的例子

例示 92-1：由具有內含子的編碼序列編碼的胺基酸序列.....	78
----------------------------------	----

第90段—限定詞「transl_table」及「translation」

交叉引用的例子

- 例示 28-1：編碼核苷酸序列及編碼胺基酸序列..... 62
例示 89-1：編碼核苷酸序列及編碼胺基酸序列..... 75
例示 92-1：由具有內含子的編碼序列編碼的胺基酸序列..... 78

第92段—由編碼序列編碼的胺基酸序列

- 例示 92-1：由具有內含子的編碼序列編碼的胺基酸序列..... 78

交叉引用的例子

- 例示 28-1：編碼核苷酸序列及編碼胺基酸序列..... 62
例示 89-1：編碼核苷酸序列及編碼胺基酸序列..... 75
例示 89-2：特徵位置超出揭露的序列 76

第93段—以殘基列舉的主要序列及其變體

- 例示 93-1：列舉變體的表示..... 80
例示 93-2：列舉變體的表示..... 81
例示 93-3：共有序列的表示..... 82

第94段—以單一序列含有列舉替代殘基方式揭露的變體序列

- 例示 94-1：含有列舉替代胺基酸的獨立序列的表示方式..... 84

第95(a)段—僅透過提及具有多個獨立變數的主要序列而揭露的變體序列

- 例示 95(a)-1：透過主要序列的註釋表示變體序列..... 85

第95(b)段—僅透過提及具有多個非獨立變數的主要序列而揭露的變體序列

- 例示 95(b)-1：以多個非獨立變變數表示獨立變體序列..... 87

第96段—變體序列的特徵關鍵及限定詞

交叉引用的例子

- 例示 29-1：「其他」胺基酸的最嚴格歧義符號..... 64

第97段—變體序列的註釋

交叉引用的例子

- 例示 29-1：「其他」胺基酸的最嚴格歧義符號 64

例示 93-3：共有序列的表示.....	82
例示 94-1：含有列舉替代胺基酸的獨立序列的表示方式	84

例示

第3(a)段「胺基酸」的定義

例示3(a)-1：D-胺基酸

專利申請案中描述下列序列：

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

問題1：ST. 26是否要求包含此序列？

是

標準第3(a)段將「胺基酸」定義為包含「D-胺基酸」及含有修飾或合成側鏈的胺基酸。根據此定義，列舉的胜肽包含5個具體定義胺基酸(D-Ala、D-Glu、Lys、Gly及 D-Met)。因此，依ST. 26第7(b)段要求，該序列必須包含在序列表。

問題3：此序列於序列表中建議如何表示？

第29段建議D-胺基酸在序列中表示為相對應的未修飾L-胺基酸。此外，不能用附錄I第3節表3之符號表示的修飾胺基酸必須用符號「X」表示。

在此例示中，序列包含3個D-胺基酸，可使用附錄I第3節表3中的未修飾L-胺基酸表示，一個L-胺基酸(Nle)及一個D-胺基酸(D-Nle)必須用符號「X」表示。第25段指出，當胺基酸序列呈環狀構型，且環僅由胺基酸殘基組成，並以肽鍵連接，申請人必須選擇殘基位置編號1的胺基酸。因此，序列可以表示為：

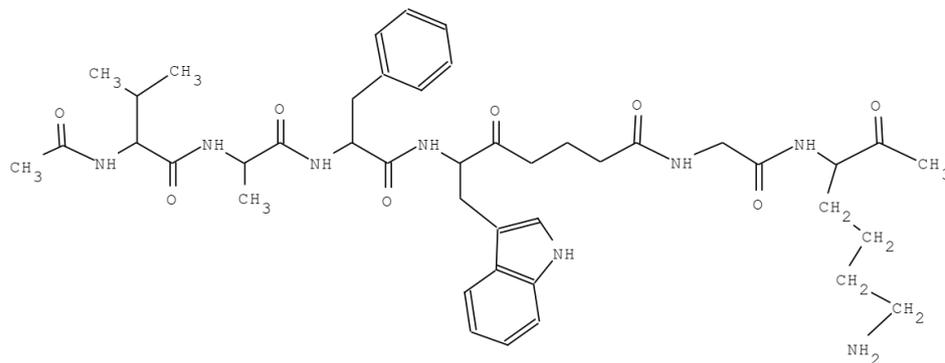
AEKXGMX (SEQ ID NO:1)

此外，此序列的任何其他胺基酸均可作為位置編號1。此序列須為每個D-胺基酸使用特徵關鍵「SITE」及限定詞「note」，以D-胺基酸的完整、非縮寫名稱(例如：D-丙胺酸及D-正白胺酸)作為限定詞值。此外，須使用特徵關鍵「SITE」及限定詞「note」，其值為L-正白胺酸的縮寫(即附錄I第4節表4中規定的「Nle」)。最後，建議使用特徵關鍵「REGION」及限定詞「note」表明肽是環狀的。

相關ST. 26段落：3(a)、7(b)、25、26、29、30及31。

第3(c)段—「列舉殘基」的定義

例示3(c)-1：以化學結構列舉胺基酸



問題1：ST. 26是否要求包含此序列？

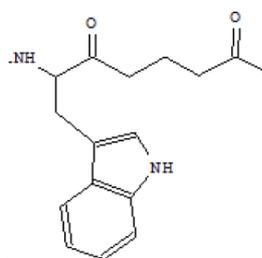
是

結構說明的列舉胜肽包含至少4個具體定義胺基酸。因此，該序列須包含在序列列表中。

問題3：此序列於序列表中建議如何表示？

該序列可以表示為：

VAFXGK (SEQ ID NO:2)



其中「X」代表「其他」修飾胺基酸：

使用特徵關鍵「SITE」及限定詞「note」，以限定詞「note」提供列舉胜肽位置4之修飾色胺酸的完整、非縮寫名稱，例如「6-amino-7-(1H-indol-3-yl)-5-oxoheptanoic acid」。

C端的甲基化因其末端-OH被-CH₃取代，改變末端賴胺酸的化學結構，由於結構有變化，序列中的賴胺酸為「修飾胺基酸」，因此，需使用特徵關鍵「SITE」及限定詞「note」來指示C端的甲基化。然而，賴胺酸添加乙醯基係以慣用的肽鍵鏈結，其乙醯化不改變賴胺酸的結構，故賴胺酸不被視為「修飾胺基酸」。因此，建議附加特徵關鍵「SITE」及限定詞「note」指示N端乙醯化。

另外，亦可將序列表示為：

VAFW (SEQ ID NO:3)

並使用特徵關鍵「SITE」及限定詞「note」指示在列舉胜肽位置4的色胺酸的修飾，其值為：「C端透過戊二醛橋接雙肽GK」。此外，建議位置1附加特徵關鍵「SITE」及限定詞「note」以指示N端乙醯化。

相關ST. 26段落：3(c)、7(b)、29、30及31。

例示3(c)-2：胺基酸序列的簡寫式



其中 G= Glycine，z=任何胺基酸，變量n可為任何整數。

問題1：ST. 26是否要求包含此序列？

是

該揭露表明「n」可以是「任何整數」；因此，涵蓋最廣泛例示的「n」是不確定的。由於「n」是不確定的，所以公式中的肽不能擴展到確定的長度，因此，必須考慮未展開的公式。

未展開式(「n」= 1)中的列舉肽提供4個具體定義的胺基酸，每個都是Gly，符號為「z」。習慣上「Z」是「麩胺醯胺或麩胺酸」的符號；然而例示將「z」定義為「任何胺基酸」。根據 ST. 26，未具體定義的胺基酸是用「X」表示。基於該分析，列舉的肽GGGGX包含4個被列舉且具體定義的甘胺酸殘基。因此，ST. 26第7(b)段要求將序列包含在序列表。

問題3：此序列於序列表中建議如何表示？

該序列使用非慣用符號「z」，其定義必須根據揭露內容確定（請參閱本附錄的簡介）。由於「z」被定義為任何胺基酸，用於代表這個胺基酸的慣用符號係「X」，因此，序列必須表示為單個序列，如下所示：

GGGGX (SEQ ID NO:4)

並建議使用特徵關鍵REGION進行註釋，特徵位置「>5」（對應 >5），限定詞note的值为「胺基酸 1-5的整個序列可以重複1次或多次」。

根據第27段，「X」可為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V中之任一者，除非在特徵表進一步描述。由於此例示中的「X」代表「任意胺基酸」，因此必須進行註釋，使用特徵關鍵「VARIANT」及限定詞「note」，其值为「X可以是任意胺基酸」。

在可行的情況下，各「X」都建議單獨註釋。然而，1個連續「X」殘基的區域，或大量的「X」殘基分散在整個序列中，可以用特徵關鍵「VARIANT」進行描述，使用語法「x..y」作為位置描述，其中x及y是第1個及最後1個「X」殘基的位置，以及值为「X可以是任何胺基酸」的限定詞「note」。

注意：上述序列的優選表示是針對在專利申請的申請日提供的序列表。相同的表示可能不適用於在專利申請的申請日之後提供的序列表，因為必須考慮提供的資訊是否可以被IPO考慮，以將權利主張添加到原始揭露中。

相關ST. 26段落：3(c)、7(b)及27。

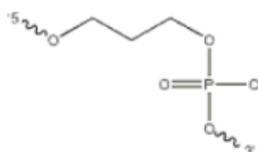
第3(g)段「核苷酸」的定義

例示3(g)-1：被 C3間隔子中斷的核苷酸序列

專利申請案描述下列序列：

atgcatgcatgcncggcatgcatgc

其中 n = 具有以下結構的C3間隔子：



問題1：ST. 26是否要求包含此序列？

是

列舉的序列包含2個具體定義的核苷酸片段，由C3間隔子隔開。

根據第3(g)段，C3間隔子不是核苷酸；慣用符號「n」被用於非慣用方式(參見本附錄簡介)。因此，每個片段都是單獨的核苷酸序列。由於各片段包含超過10個具體定義的核苷酸，因此兩者都必須包含在序列表中。

問題3：此序列於序列表中建議如何表示？

各片段必須作為單獨的序列包含在序列表中，並有其專屬序列識別號：

atgcatgcatgc (SEQ ID NO:5)

cggcatgcatgc (SEQ ID NO:6)

各片段中連接C3間隔子的胞嘧啶建議在特徵表中進一步使用特徵關鍵「misc_feature」及限定詞「note」，限定詞「note」的值「free text」建議指示間隔子的存在，其與另1個核酸相連，並透過任一完整非縮寫的化學名稱或其通用名稱(如C3間隔子)識別此間隔子

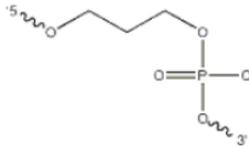
相關ST. 26段落：3(g)、7(a)及15。

例示3(g)-2：具有替代殘基的核苷酸序列，包含C3間隔子

專利申請案描述下列序列：

atgcatgcatgcncggcatgcatgc

其中 n = c、a、g或具有以下結構的C3間隔子：



問題1：ST. 26是否要求包含此序列？

是

列舉序列中有24個具體定義殘基，且有變數「n」插入，必須參考揭露內容對序列的解釋，以判定「n」是否以慣用或非慣用方式使用(參見本附錄簡介)。該揭露內容指出n=c、a、g或C3間隔子，由於「n」被描述為C3間隔子，不符合核苷酸的定義，因此「n」為非慣用方式使用的慣用符號。符號「n」亦被描述為「c」、「a」或「g」，因此，ST. 26要求序列表中包含此25個核苷酸序列。由於由C3間隔區分隔的2個片段是獨立於25個核苷酸序列的序列，序列表亦可包含此兩個12個核苷酸序列。

問題3：此序列於序列表建議如何表示？

該例示表明「n = c、a、g或C3間隔子」。如上所述，C3間隔子不是核苷酸。根據第15段，符號「n」不得用於代表核苷酸以外的任何事物；因此，符號「n」不能代表序列表中的C3間隔子。

第15段亦規定，在適用歧義符號的情況下，建議使用最嚴格的歧義符號。根據附錄I第1節表1，以符號「v」代表「a或c或g」較符號「n」更嚴格。若例示的變數「n」為c、a或g，單個序列所列舉的殘基包含揭露最多例示，因此最具包容性的序列(參見本附錄的簡介)必須包含在序列表中，如下所示：

atgcatgcatgcvcgcatgcatgc (SEQ ID NO:7)

如本附錄之簡介所述，強烈鼓勵包含本發明揭露或請求項中必要的任何附加序列。

例示中變數「n」為C3間隔子時，序列可視為在變數「n」任一側具有具體定義核苷酸的兩個獨立片段atgcatgcatgc(SEQ ID NO:8)及cggcatgcatgc(SEQ ID NO:9)。若對揭露或請求項必要，建議此2個序列亦包含於序列表中，並具有其專屬序列識別號。各片段中連接C3間隔子的胞嘧啶建議在特徵表中使用特徵關鍵「misc_feature」及限定詞「note」，以值為「free text」指示間隔子的存

在及間隔子連結至另一核酸，並以完整非縮寫的化學名稱或其通用名稱(如C3間隔子)識別此間隔子。

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：3(g)、7(a)及15。

例示3(g)-3：無鹼基位點

專利申請案描述以下序列：

gagcattgac-AP-taaggct

其中AP為無鹼基位點

問題1：ST. 26是否要求包含此序列？

是

列舉序列的具體定義殘基被無鹼基位點中斷。無鹼基位點的5'側包含10個核苷酸，而無鹼基位點的3'側包含7個核苷酸。第3(g)(ii)(2)段定義無鹼基位點為核苷酸序列的一部分時，其為「核苷酸」。因此，無鹼基位點在此例示中被認為是1個「核苷酸」，以判定是否必須以及如何將該序列包含在序列表中。據此，無鹼基位點的兩側殘基為單一系列序列的一部分，該序列共包含18個核苷酸，其中17個是具體定義的。因此，根據ST. 26第7(a)段的要求，此序列必須作為獨立序列包含在序列表中。

問題3：此序列於序列表中建議如何表示？

該序列須包含於序列表，並表示如下：

gagcattgacntaaggct (SEQ ID NO:10)

無鹼基位點必須用「n」表示，且必須在特徵表中進一步描述，較佳註釋方式為使用特徵關鍵「modified_base」及必要限定詞「mod_base」，其值為「OTHER」。且必須包含限定詞「note」，將修飾鹼基描述為無鹼基位點。

相關ST. 26段落：3(g)、7(a)及17。

例示3(g)-4：核酸類似物

專利申請案揭露下列乙二醇核酸(GNA)序列：

PO₄-tagttcattgactaaggctccccattgact-OH

其中該序列的左端模擬DNA序列的5'端。

問題1：ST. 26是否要求包含此序列？

是

根據ST. 26第3(g)(i)(2)段，構成GNA序列的單個殘基被視為核苷酸。據此，該序列有超過10個列舉且「具體定義」核苷酸，必須包含在序列表中。

問題3：此序列於序列表中建議如何表示？

GNA序列沒有5'端及3'端，而是3'端及2'端。其3'端指其具有末端磷酸基團的一端，對應DNA或RNA的5'端(注意其他核酸類似物可能不同地對應DNA及RNA的5'端及3'端)。根據第11段，其必須以「從左到右的方向仿5'到3'的方向」包含在序列表中，因此，該序列須包含於序列表，並表示如下：

tagttcattgactaaggctccccattgact (SEQ ID NO:11)

此序列須在特徵表中使用特徵關鍵「modified_base」及必要限定詞「modified_base」其值為縮寫「OTHER」，並須以限定詞「note」呈現修飾核苷酸的完整非縮寫名稱，例如「glycol nucleic acids」或「2,3-dihydroxypropyl nucleosides」。可使用單一INSDFeature要素將整個序列描述為GNA，其中INSDFeature_location的範圍為「1..30」。

相關ST. 26段落：3(d)、3(g)、7(a)、11、16、18、65及66。

第3(k)段「具體定義」的定義

例示3(k)-1：核苷酸歧義符號

5' NNG KNG KNG K 3'

N及K是IUPAC-IUB歧義代碼

問題1：ST. 26是否要求包含此序列？

否

IUPAC-IUB歧義代碼對應於附錄I第1節表1中定義的核苷酸符號列表。根據第3(k)段，具體定義的核苷酸是附錄I所列「n」符號所代表者以外的任意核苷酸，因此，「K」及「G」是具體定義核苷酸，而「N」不是具體定義核苷酸。

列舉序列未含10個或更多個具體定義的核苷酸，因此根據 ST. 26第7(a)段不須將其包含在序列表中。

問題2：ST. 26是否允許包含此序列？

否

根據第8段，「序列表不得包含任何少於10個具體定義核苷酸的序列……」，列舉的序列不含10個或更多個具體定義核苷酸，故不得包含在序列表中。

相關ST. 26段落：3(k)、7(a)、8及13。

例示3(k)-2：以慣用及非慣用方式使用的歧義符號「n」

專利申請案揭露人工序列：5'-AATGCCGGAN-3'。該揭露進一步指出：

- (i)在一例示中，N是任意核苷酸；
- (ii)在一例示中，N是任選的，但優選G；
- (iii)在一例示中，N是 K；
- (iv)在一例示中，N是C。

問題1：ST. 26是否要求包含此序列？

否

列舉序列包含9個具體定義的核苷酸及1個「N」，必須參考揭露內容對序列的解釋以判定符號「N」是否以慣用方式使用(參見本附錄的簡介)。

考慮揭露內容的例示(i)至(iv)中，「N」是「任意核苷酸」為序列所列舉的殘基包含揭露最多例示。在此最具包容性的例示中，列舉序列的「N」係以慣用方式使用。

在某些例示中，「N」被描述為具體定義殘基(如(iv)部分的「N是C」)。然而，當判斷序列是否須包含在序列表中時，只有最具包容性的例示(如「N是任意胺基酸」)必須被考慮。因此，必須評估的列舉序列為5'-AATGCCGGAN-3'。

依此分析，列舉序列(如AATGCCGGAN)不包含10個具體定義核苷酸，故ST. 26第7(a)段不要求將此序列包含在序列表中，儘管在部分例示將「n」定義為特定的核苷酸。

問題2：ST. 26是否允許包含此序列？

否

序列「AATGCCGGAN」不得包含在序列表中。但若將「N」替換為具體定義的核苷酸，所述替換的序列可能可以包含於序列表中。

問題3：此序列於序列表中建議如何表示？

強烈鼓勵包含本發明關鍵部分例示的序列，包含這些序列可供專利申請者進行更澈底的檢索及提供權利主張的公示。故針對上述例示，強烈建議將以下3個附加序列包含於序列表，且各別具有其專屬序列識別號：

aatgccggag (SEQ ID NO:12)

aatgccggak (SEQ ID NO:13)

aatgccggac (SEQ ID NO:14)

若少於上述所有3個序列被納入，則建議註釋替換「n」的核苷酸。例如：若序列列表只包含上述SEQ ID NO:12，則建議使用特徵位置「10」的特徵關鍵「misc_difference」，並與2個限定詞「replace」一起使用，其中一個值為「g」，另一個值為「c」。

注意：上述序列的優選表示是針對在專利申請日提交的序列列表。相同的陳述可能不適用於專利申請日後提交序列列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：3(k)、7(a)、8及13。

例示3(k)-3：以非慣用方式使用的歧義符號「n」

專利申請案揭露序列：5'-aatgttgga-3'

其中n為c。

問題1：ST. 26是否要求包含此序列？

是

根據第3(k)段，「具體定義」核苷酸係指除列於附錄I第1節表1中符號「n」所表示者外的任何核苷酸。

在此例示中，「n」以非慣用方式使用，僅表示「c」，並非以慣用方式使用「n」來表示「任何核苷酸」。因此，該序列必須解釋為序列中使用相當的慣用符號，即「c」（參見本文檔的介紹）。因此，列舉順序必須視為：

5'-aatgttggac-3'

該序列有10個具體定義核苷酸，ST. 26第7(a)段要求將其包含在序列表。

問題3：此序列於序列表中建議如何表示？

該序列須包含於序列表，並表示如下：

aatgttggac (SEQ ID NO:15)

相關ST. 26段落：3(k)及7(a)。

例示3(k)-4：「n」以外的歧義符號為「具體定義」

專利申請案描述以下序列：

5' NNG KNG KNG KAG VCR 3'

其中 N、K、V及R為IUPAC-IUB歧義碼。

問題1：ST. 26是否要求包含此序列？

是

IUPAC-IUB歧義碼相當於附錄I第1節表1所列的核苷酸符號。根據第3(k)段，「具體定義」的核苷酸係指除列於附錄I第1節表1中符號「n」所表示者外的任何核苷酸。因此，「K」、「V」及「R」為「具體定義」的核苷酸。

該序列列舉11個「具體定義」的核苷酸，ST. 26第7(a)段要求將其包含在序列表中。

問題3：此序列於序列表中建議如何表示？

該序列須包含於序列表，並表示如下：

nngkngkngkagvcr (SEQ ID NO:16)

相關ST. 26段落：3(k)、7(a)及15。

例示3(k)-5：以非慣用方式使用的歧義縮寫「Xaa」

專利申請案描述以下序列：

Xaa-Tyr-Glu-Xaa-Xaa-Xaa-Leu

其中第1位點的Xaa為任意胺基酸，第4位點的Xaa為Lys，第5位點的Xaa為Gly，第6位點的Xaa為白胺酸或異白胺酸。

問題1：ST. 26是否要求包含此序列？

是

式中列舉的肽在第2、3、7位點提供3個具體定義胺基酸。第1位點胺基酸用慣用縮寫Xaa代表任意胺基酸。而第4、5及6位點的胺基酸則將慣用縮寫以非慣用方式使用(參見本附錄的簡介)，因此，參考揭露內容對序列的解釋判定這些位置中「Xaa」的定義。由於第4-6位點的「Xaa」代表特定胺基酸，因此須將序列解釋為在序列中使用同義的慣用縮寫，即Lys、Gly及(Leu或Ile)。故而該序列包含4個或更多具體定義胺基酸，依ST. 26第7(b)段要求必須將其包含在序列表中。

問題3：此序列於序列表中建議如何表示？

該序列以非慣用方式使用慣用縮寫「Xaa」，因此，須參考揭露內容對序列的解釋判定第4、5及6位點的「Xaa」的定義。該揭露內容解釋「Xaa」定義為第4位點為離胺酸、第4位點為甘胺酸及第6位點為白胺酸或異白胺酸，而這些胺基酸的慣用符號分別為K、G及J，因此，該序列須包含於序列表，並表示如下：

XYEKGJL (SEQ ID NO:17)

根據第27段，「X」被解釋為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V，除非其於特徵表進一步描述。由於SEQ ID NO:17位置1的「X」表示「任意胺基酸」，因此必須用特徵關鍵「VARIANT」及限定詞「note」，其值為「X為任意胺基酸」。

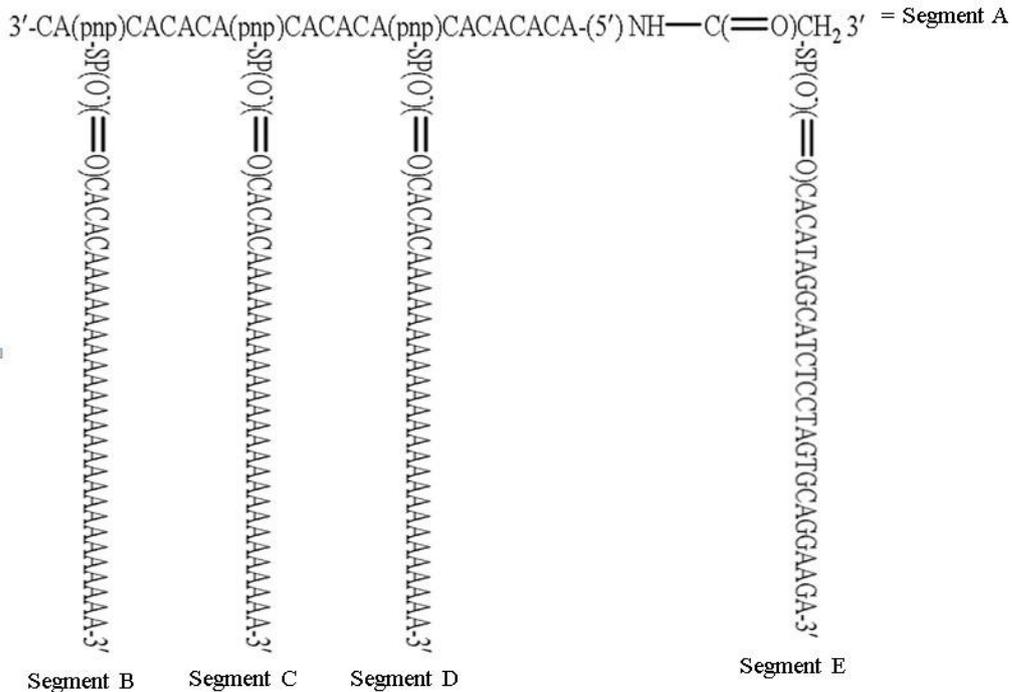
在可行的情況下，各「X」都建議單獨註釋。然而，連續的「X」殘基區域或大量的「X」殘基分散在整個序列中，可使用特徵關鍵「VARIANT」進行描述，使用語法「x..y」作為位置描述，其中x及y是第1個及最後1個「X」殘基的位置，並使用限定詞「note」及其值為「X可以是任意胺基酸」。

相關ST. 26段落：3(k)、7(b)、2及27。

第7(a)段—序列表中核苷酸序列的要求

例示7(a)-1：分支核苷酸序列

說明書揭露以下分支核苷酸序列：



其中「pnp」為含有溴乙醯氨基官能基的连接子或單體；

3' -CA(pnp)CACACA(pnp)CACACA(pnp)CACACACA-(5')NH—C(=O)CH₂ 3' 為片段A；

SP(O⁻)(=O)CACACAAAAAAAAAAAAAAAAAAAAAAAAA 3' 為片段B、C及D；及

SP(O⁻)(=O)CACATAGGCATCTCCTAGTGCAGGAAGA 3' 為片段 E。

問題1：ST. 26是否要求包含此序列？

是，4個垂直片段B-E必須包含在序列表中

否，水平片段A不須包含在序列表中

上圖例示為一「梳型」分支核酸序列，其包含5個線性片段：水平片段A及4個垂直片段B-E。

根據第7(a)段，分支核苷酸序列的線性區域包含10個或更多具體定義核苷酸，其相鄰核苷酸從3' 連接到5'，則該序列必須包含在序列表中。

4個垂直片段B-E各別包含超過10個具體定義核苷酸，其中相鄰核苷酸係從3'

到5' 連接，因此每個均必須包含在序列表中。

在水平區段A中，核苷酸序列的線性區域由非核苷酸部分“pnp”連接，且這些連接的線性區域均包含少於10個具體定義核苷酸。由於片段A的沒有包含10個或更多具體定義核苷酸且其中相鄰核苷酸從3' 到5' 連接的區域，因此非ST. 26第7(a)段要求包含在序列表的序列。

問題2：ST. 26是否允許包含此序列？

否

根據第8段，「序列表不得包含任何少於10個具體定義核苷酸的序列……」

片段A的任何區域均不包含10個或更多具體定義核苷酸，其中相鄰核苷酸從3' 連接到5'，因此，不得作為具有其序列識別號的單獨序列包含在序列表中。

然而，片段B、C、D及E會被註釋表明其與片段A連接。

問題3：此序列於序列表中建議如何表示？

片段B、C及D相同，必須以單個序列包含在序列表中：

cacacaaaaaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO:18)

序列中的第一個「c」建議使用特徵關鍵「misc_feature」及限定詞「note」進一步描述，其值例如為「This sequence is one of four branches of a branched polynucleotide」。

片段E必須以單個序列包含在序列表中：

cacataggcatctcctagtagcaggaaga (SEQ ID NO:19)

序列中之第一個「c」建議使用特徵關鍵「misc_feature」及限定詞「note」進一步描述，其值例如為「This sequence is one of four branches of a branched polynucleotide」。

相關ST. 26段落：7(a)、8、11、13及17。

例示7(a)-3：以非慣用方式使用的核苷酸歧義符號

專利申請案描述下列序列：

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

揭露內容解釋序列進一步指出：「一個『DR元件』由序列5' ATCAGCCAT 3' 組成。

突變的DR元件(或稱MDR)係指DR元件的中間5個核苷酸CAGCC突變為TTTTT」。

問題1：ST. 26是否要求包含此序列？

是

列舉序列使用符號「MDR」，但不清楚該符號於序列中係慣用符號(即附錄I第3節表3所列符號)或非慣用符號，必須參考揭露內容對序列的解釋才能判定(見本附錄的簡介)。根據表3，「MDR」可解釋為3個慣用符號(m = a或c, d = a或g或t/u, r = g或a)或其他結構的縮寫。

查閱揭露內容指出MDR元件相當於5' ATTTTTTAT 3'。由於「MDR」的字母係以非慣用方式使用的慣用符號，該序列必須被視為使用同等意義的慣用符號揭露。因此，其被視為以下列舉序列包含於序列表：

5' GATC ATTTTTTAT ATTTTTTAT ATTTTTTAT ATTTTTTAT GTAC 3'

該列舉序列有44個具體定義核苷酸，ST. 26第7(a)段要求將其包含在序列表中。

問題3：此序列於序列表中建議如何表示？

該序列須包含於序列表，並表示如下：

gatcattttttatattttttatattttttatattttttatgtac (SEQ ID NO:21)

相關ST. 26段落：7(a)及13。

例示7(a)-4：以非慣用方式使用的核苷酸歧義符號

專利申請案描述下列序列：

5' ATTC-N-N-N-N-GTAC 3'

揭露內容解釋序列進一步指出「N」由序列5' ATACGCACT 3' 組成。

問題1：ST. 26是否要求包含此序列？

是

列舉序列使用符號「N」，必須參考揭露內容對序列的解釋，以判定「N」是以慣用或非慣用的方式使用(參見本附錄的簡介)。

查閱揭露內容指出「N」相當於5' ATACGCACT 3'，因此，「N」是一個以非慣用方式使用的慣用符號，該序列必須被視為使用同等意義的慣用符號揭露：

5' ATTC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3'

該列舉序列有44個具體定義核苷酸，ST. 26第7(a)段要求將其包含在序列表中。

問題3：此序列於序列表中建議如何表示？

該序列須包含於序列表，並表示如下：

attcatacgcaactatacgcaactatacgcaactatacgcaactgtac (SEQ ID NO:22)

相關ST. 26段落：7(a)及13。

例示7(a)-5：非慣用核苷酸符號

專利申請案描述下列序列：

5' GATC-β-β-β-β-GTAC 3'

其揭露內容解釋序列進一步指出「β」由序列5' ATACGCACT 3' 組成。

問題1：ST. 26是否要求包含此序列？

是

列舉序列使用非慣用符號「β」，必須參考揭露內容對序列的解釋以判定「β」的含義(參見本附錄的簡介)。

查閱揭露內容指出「β」相當於5' ATACGCACT 3'，因此，「β」是一個非慣用符號，用於代表具有9個具體定義慣用符號的序列，該序列必須被視為使用同等意義的慣用符號揭露：

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

該列舉序列有44個具體定義核苷酸，ST. 26第7(a)段要求將其包含在序列表中。

問題3：此序列於序列表中建議如何表示？

該序列須包含於序列表，並表示如下：

gatcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO:23)

相關ST. 26段落：7(a)及13。

例示7(a)-6：非慣用核苷酸符號

專利申請案描述下列序列：

5' GATC- β - β - β - β -GTAC 3'

其揭露內容解釋序列進一步指出「 β 」相當於腺嘌呤、肌苷或假尿苷。

問題1：ST. 26是否要求包含此序列？

否

列舉序列使用非慣用符號「 β 」，必須參考揭露內容對序列的解釋以判定「 β 」的含義(參見本附錄的簡介)。

查閱揭露內容指出「 β 」相當於於腺嘌呤、肌苷或假尿苷。而唯一可用來表示「腺嘌呤、肌苷或假尿苷」的慣用符號是「n」，因此，「 β 」是一個非慣用符號，用來表示慣用符號「n」，必須將序列以4個「n」符號(以下顯示為「N」)取代4個「 β 」符號：

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

該列舉序列只有8個具體定義核苷酸，ST. 26第7(a)段不要求將其包含在序列表中。

問題2：ST. 26是否允許包含此序列？

否

列舉序列5' GATC-N-N-N-N-GTAC 3' 不得包含在序列表中。

然而，若其中至少2個「n」符號被腺嘌呤替代，從而產生具有至少10個或更多具體定義核苷酸的序列，則該揭露的替代序列可包含在序列表中。

問題3：此序列於序列表中建議如何表示？

一種可能允許表示的序列為：

gatcaaaagtac (SEQ ID NO:24)

在上述例示中，建議註釋以4個腺嘌呤核苷酸替換 β 符號，以備註這些位置可以用肌苷或假尿苷取代。

特徵關鍵「misc_difference」建議與特徵位置5-8及限定詞「note」併用，其值例如為「5-8位中任一位置核苷酸可被肌苷或假尿苷取代」。由於這些替代是修飾的核苷酸，因此要求使用特徵關鍵「modified_base」及限定詞「mod_base」，限定詞「mod_base」的值可為「OTHER」，且有限定詞「note」及其值為「i or p」。

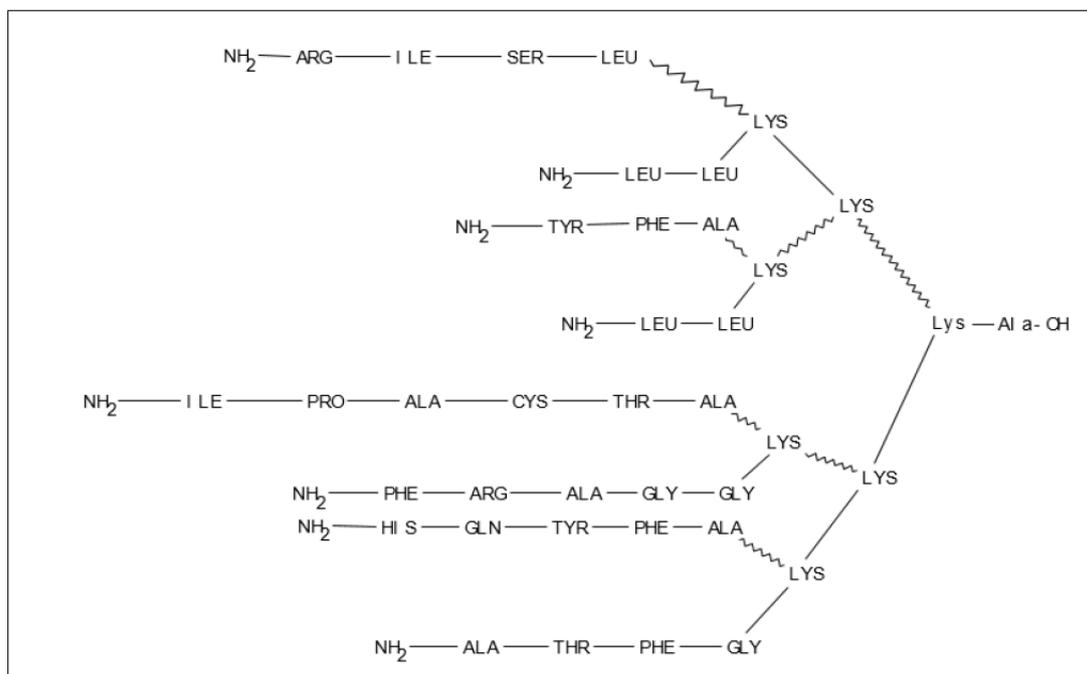
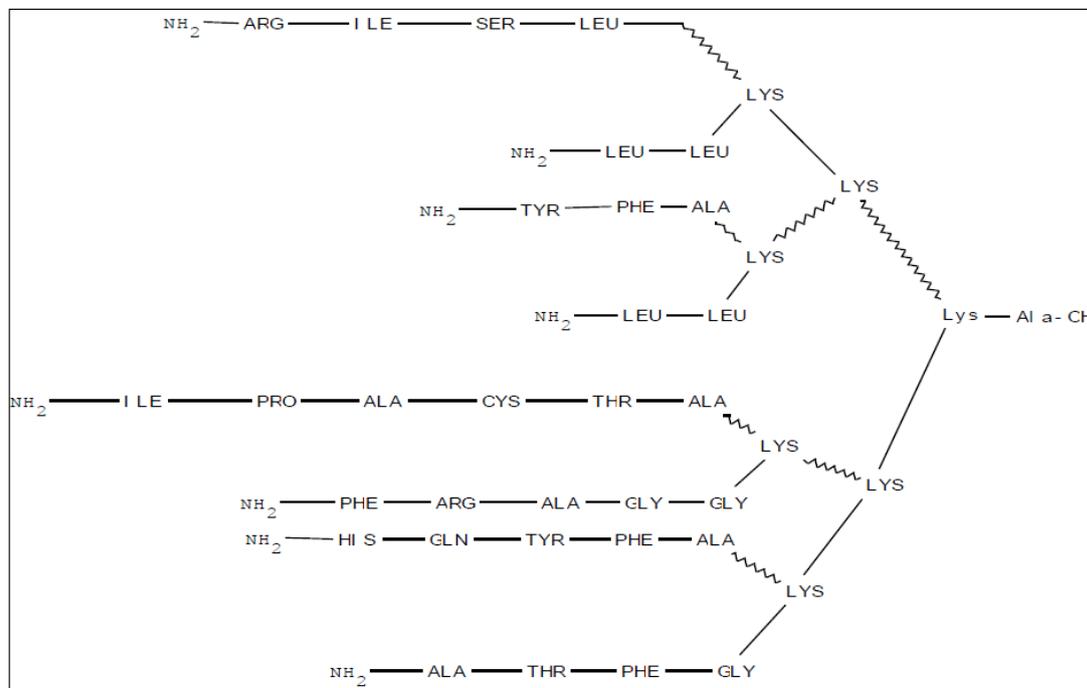
其他排列係可能的。

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：7(a)、8、13及17。

例示7(b)-2：分支胺基酸序列

申請案描述一分支序列，其以離胺酸殘基作為支架核心形成8個分支，並有多個線性肽鏈連接到各分支上。離胺酸為二元胺基酸，為肽鍵提供2個位點。該肽說明如下：

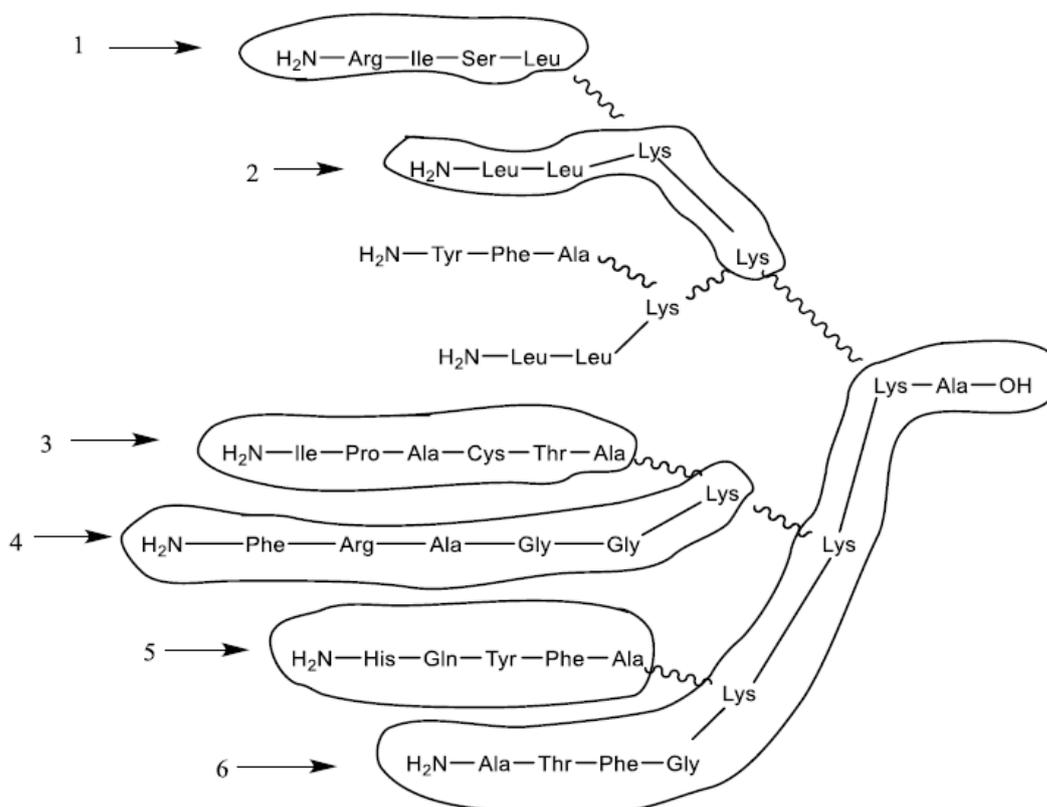


在上述分支肽中，離胺酸與另一個胺基酸的鍵結，以——代表由離胺酸的末端氨基與另一胺基酸的羧基形成的醯胺鍵；以~~~~代表由離胺酸的側鏈氨基與另一胺基酸的羧基形成的醯胺鍵。

問題1：ST. 26是否要求包含此序列？

是

該例示揭露一個以離胺酸殘基作為支架的分支序列，而第7(b)段要求序列中無論分支或線性區域，包含四個或更多具體定義胺基酸的序列須包含在序列表中。在上述例示中，分支肽的線性區域具有4個或更多具體定義胺基酸：



ST. 26第7(b)段要求上述肽1-6包含在序列表中。

不須包含在序列表中的肽係：

YFA

LLK

問題2：ST. 26是否允許包含此序列？

否

根據第8段，序列表不得包括任何少於4個具體定義胺基酸的序列。

YFA及LLK肽各別僅包含3個具體定義胺基酸，因此，不得作為具有其專屬序列識別號的獨立序列包含在序列表中。

問題3：此序列於序列表中建議如何表示？

肽1-6必須用單獨的序列識別號表示：

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)

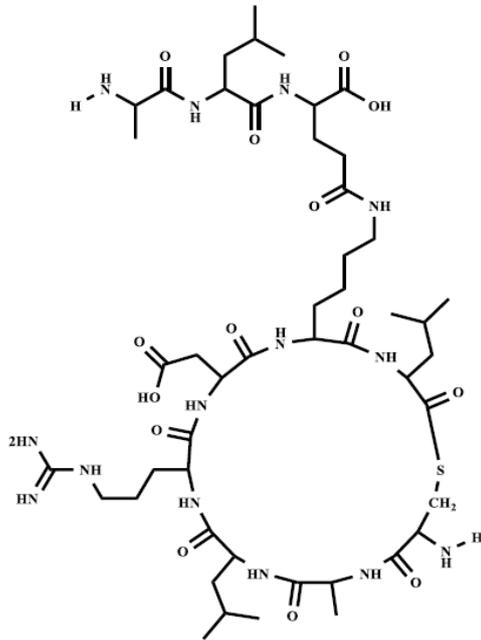
ATFGKKKA (SEQ ID NO:31)

交叉連結可使用特徵關鍵「SITE」及必要限定詞「note」，其值例如為「This sequence is one part of a branched amino acid sequence」。根據ST.26第30段，SEQ ID NOs:27、29及31必須為各個離胺酸註釋以表明其為修飾胺基酸，使用特徵關鍵「SITE」及限定詞「note」，描述離胺酸側鏈與另一序列透過醯胺鍵連接。SEQ ID NOs:26、28及30建議使用特徵關鍵「SITE」及限定詞「note」註釋，註明胺基酸的C端鏈接至另一序列。

相關ST. 26段落：7(b)、8、26、29、30及31。

例示7(b)-4：含有支鏈胺基酸序列的環肽

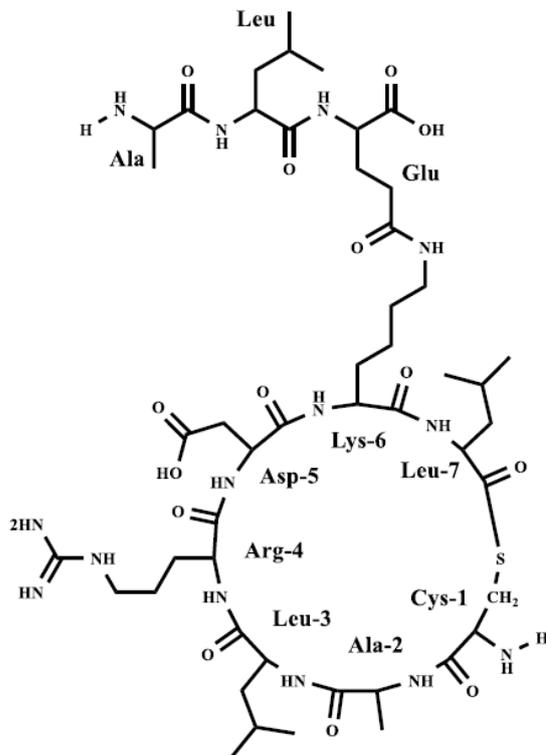
專利申請案揭露下列結構：



環狀結構的半胱胺酸及白胺酸以半胱胺酸的側鏈與白胺酸的末端羧基連接。

問題1：ST. 26是否要求包含此序列？

所示結構係1個分支的環狀胺基酸序列，其中包含以下胺基酸：



由於此環肽係以Cys的側鏈與Leu的羧基進行環化，故其N端位於Cys-1。

是——肽的環狀區域

ST. 26第7(b)段要求分支序列的線性區域包含4個或更多具體定義胺基酸，且胺基酸形成單獨肽主鏈，必須包含在序列表中。在上述例示中，分支肽的環狀區域有4個以上胺基酸，因此必須包含在序列表中。

否——肽的三肽分支

3肽分支 Ala-Leu-Glu 不須包含在序列表中。

問題2：ST. 26是否允許包含此序列？

否

根據第8段，序列表不得包括任何少於4個具體定義胺基酸的序列。

3肽分支僅包含3個具體定義胺基酸，因此，不得作為具有其序列識別號的獨立序列包含於序列表中。

問題3：此序列於序列表中建議如何表示？

雖然此例示說明一種環狀構型的肽，但如第25段所示，環並非僅由胺基酸殘基形成肽鍵。此肽由於其胺基酸序列係藉由半胱胺酸(Cys)的側鏈與白胺酸(Leu)的羧基產生環化，因此，半胱胺酸必須為肽環狀區域的第1位。此序列必須表示為：

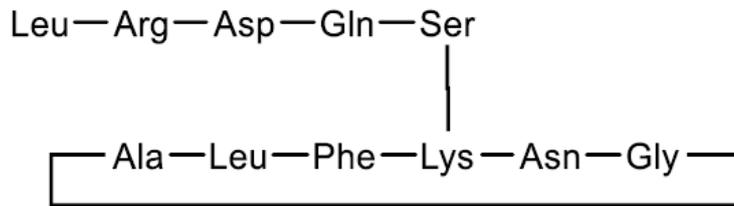
CALRDKL (SEQ ID NO: 89)

如上圖所示，胺基酸序列藉由半胱胺酸的側鏈及白胺酸羧基形成硫酯鍵結而環化。該修飾的半胱胺酸與白胺酸形成鏈內連接，須使用特徵關鍵「SITE」加以描述，特徵位置要素係交叉鏈接的胺基酸殘基編號，格式為「x.y」（即「1.7」）。必要限定詞「note」建議指明鍵結的性質，例如：「半胱胺酸白胺酸硫酯(Cys-Leu)」，以指定Cys-1與Leu-7 通過硫酯鍵連接。此外，位置編號6的離胺酸須註明被修飾，使用特徵關鍵「SITE」、必要限定詞「note」及限定詞值描述離胺酸側鏈連接3肽ALE。

相關ST. 26段落：7(b)、8、25、26、29、30、31、66(c)及70。

例示7(b)-5：含有支鏈胺基酸序列的環肽

專利申請案揭露下列支鏈環肽：

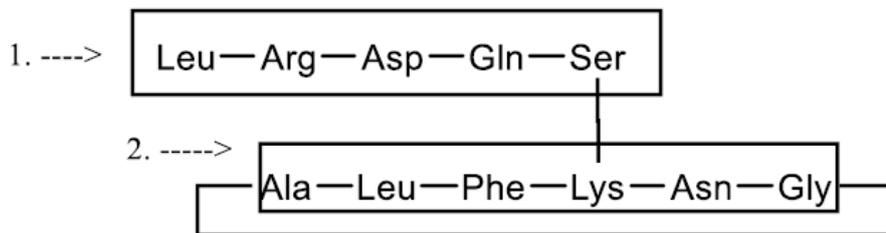


Ser及Lys係以絲胺酸的羧基及Lys側鏈的氨基形成的醯胺鍵連接。

問題1：ST. 26是否要求包含此序列？

是

第7(b)段要求在序列中表示的任何序列包含4個或更多具體定義胺基酸，且分支序列可表示為線性區域。在上述例示中，該肽包含以肽鍵連接胺基酸的環狀區域，及一分支區域其係連接到環狀區域的Lys的側鏈。該支鏈肽的2個區域均可以線性表示，且均包含4個或更多具體定義胺基酸：



ST. 26要求在序列中包含此環狀支鏈肽的序列1及2，且每個序列都有其專屬序列識別號。

問題3：此序列於序列表中建議如何表示？

序列1必須表示為：

LRDQS (SEQ ID NO:90)

序列1可使用特徵關鍵「SITE」及限定詞「note」來註釋，描述位置5的絲胺酸與另一序列透過Ser及另一個序列的Lys的側鏈形成的醯胺鍵連接。

序列2係環肽，第25段指出，當胺基酸序列呈環狀構型且沒有氨基及末端羧基時，申請人必須選擇位置編號為1的胺基酸殘基。因此，該序列可表示為：

ALFKNG (SEQ ID NO:91)

序列中任何其他胺基酸均可被指定為第1個殘基，該ALFKNG序列必須使用特徵關鍵「SITE」及限定詞「note」進一步描述，描述其第4個殘基Lys的側鏈透過

醯胺鍵連接到另一個序列。該側鏈連接修飾Lys，根據ST. 26第30段，修飾胺基酸必須在特徵表進一步描述，且建議使用特徵關鍵「REGION」及限定詞「note」指明肽ALFKNG為環形。

相關ST. 26段落：7(b)、25、26、30及31。

第11(a)段—雙股核苷酸序列——完全互補

例示11(a)-1：雙股核苷酸序列——相同長度

專利申請案描述下列雙股DNA序列：

3' -CCGGTTAACGCTA-5'

5' -GGCCAATTGCGAT-3'

問題1：ST. 26是否要求包含此序列？

是

各序列的列舉核苷酸均超過10個具體定義核苷酸。由於此雙股核苷酸序列的2股完全互補，故至少有1股必須包含在序列表中。

問題2：ST. 26是否允許包含此序列？

是

此序列雖然只要求其中1股必須包含在序列表中，仍可2股各別包括於序列表，每條鏈都有其專屬序列識別號。

問題3：此序列於序列表中建議如何表示？

雙股DNA序列在序列表中必須以單個序列或2個單獨的序列表示，每個包含於序列表的序列均必須以5'到3'的方向表示，並分配其專屬序列識別號。

Atcgcaattggcc(上方鏈)(SEQ ID NO:34)

及/或

ggccaattgcgat(下方鏈)(SEQ ID NO:35)

相關ST. 26段落：7(a)、11(a)及13。

第11(b)段—雙股核苷酸序列——不完全互補

例示11(b)-1：雙股核苷酸序列——不同長度

專利申請案包含以下附圖及說明文字：

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

以PNA探針(下方鏈)結合的人類基因ABC1啟動子區域(上方鏈),其中PNA中的「n」係選自5-硝基吲哚(5-nitroindole)及3-硝基吲哚(3-nitroindole)的通用PNA鹼基。

問題1：ST. 26是否要求包含此序列？

是——ABC1啟動子區域(上方鏈)

上方鏈有超過10個列舉且「具體定義」的核苷酸，須包含在序列表中。

是——PNA 探針(下方鏈)

由於此序列的2條鏈不完全互補，下方鏈亦須包含在序列表中，且具有其專屬序列識別號。根據ST. 26第3(g)段，構成PNA或「肽核酸」的各別殘基均視為核苷酸，因此，下方鏈有超過10個列舉且「具體定義」的核苷酸，須包含在序列表中。

問題3：此序列於序列表中建議如何表示？

上方鏈須包含於序列表，並表示如下：

tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc (SEQ ID NO:36)

而下方鏈為肽核酸，不具3'端及5'端，根據第11段，其必須以「從左到右方向，模仿5'到3'的方向」包含在序列表中。因此，其須包含於序列表，並表示如下：

cgcctnagtcaatggg (SEQ ID NO:37)

特徵關鍵「source」的限定詞「organism」的值須為「synthetic construct」，而必要性限定詞「mol_type」的值為「other DNA」。下方鏈在特徵表中的描述，須使用特徵關鍵「modified_base」及必要限定詞「mod_base」其值為縮寫「OTHER」，並須以限定詞「note」呈現修飾核苷酸的完整非縮寫名稱，例如N-(2-aminoethyl) glycine nucleosides。

「n」殘基在特徵表中須進一步描述，使用特徵關鍵「modified_base」及必要限定詞「mod_base」其值為縮寫「OTHER」，並須以限定詞「note」呈現修飾核

苷酸的完整非縮寫名稱：N-(2-aminoethyl) glycine 5-nitroindole或N-(2-aminoethyl) glycine 3-nitroindole。

相關ST. 26段落：3(g)、7(a)、11(b)、17及18。

例示11(b)-2：雙股核苷酸序列——無鹼基配對片段

專利申請案描述下列雙股DNA序列：

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

問題1：ST. 26是否要求包含此序列？

是

此雙股核苷酸序列所列舉的每一股均有超過10個具體定義核苷酸，由於2股彼此不完全互補，故2股皆須包含在序列表中，且每一股都有其專屬序列識別號。

問題3：此序列於序列表中建議如何表示？

每一股序列須以5'到3'的方向表示，並指定其專屬序列識別號：

atcgggatcgcatttcgattggcc(上方鏈)(SEQ ID NO:38)

及

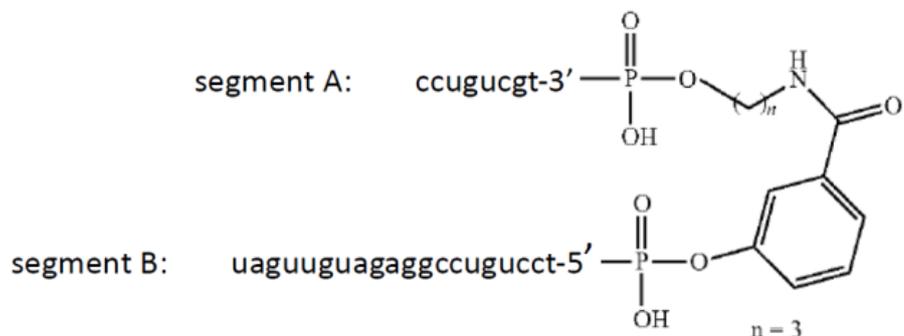
ggccaatatggcttgatcccgat(下方鏈)(SEQ ID NO:39)

相關ST. 26段落：7(a)、11(b)及13。

第14段—符號「t」在RNA中解釋為尿嘧啶

例示14-1：符號「t」代表RNA的尿嘧啶

專利申請案描述以下化合物：



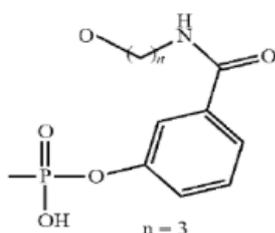
其中A片段及B片段係RNA序列。

問題1：ST. 26係否要求包含此序列？

是—B片段

否—A片段

列舉序列包含2個具體定義核苷酸片段，由以下「連接子(linker)」結構分隔：



根據第3(g)段，連接子結構非核苷酸；因此，各片段都必須被視為單獨的序列。B片段包含超過10個具體定義核苷酸，依ST. 26第7(a)段必須包含在序列表中。A片段僅包含8個具體定義核苷酸，因此不須包含在序列表中。

問題2：ST. 26 是否允許包含此序列？

否

A片段包含少於10個具體定義核苷酸，根據ST. 26第8段，不得包含在序列表中。

問題3：此序列於序列表中建議如何表示？

B片段為1個RNA分子；因此，「INSDSeq_moltype」要素必須為「RNA」。在序列表中，符號「u」不得用於表示RNA分子的尿嘧啶。根據第14段，符號「t」被解釋為RNA的尿嘧啶。因此，B片段必須以下列所示的形式包含在序列表中：

tcctgtccggagatggtgat (SEQ ID NO:40)

RNA中的胸腺嘧啶被視為修飾核苷酸，即修飾的尿嘧啶，必須在序列中表示為「t」，並在特徵表中進一步描述。因此，位置1的胸腺嘧啶必須進一步描述，使用特徵關鍵「modified_base」及限定詞「mod_base」，其值為「OTHER」，並使用限定詞「note」及其值為「胸腺嘧啶(thymine)」。

位置1的胸腺嘧啶，即修飾的尿嘧啶，也建議在特徵表中進一步使用特徵關鍵「misc_feature」及限定詞「note」，其值例如為「The 5' oxygen of the thymidine is attached through the linker (4-(3-hydroxybenzamido) butyl) phosphinic acid to another nucleotide sequence」。在可行情況下，其他序列的限定詞「note」的值可直接表示。

相關ST. 26段落：3(g)、7(a)、8、13、14、19及54。

第27段—建議使用最嚴格歧義符號

例示27-1：胺基酸的簡寫式

(GGGz)₂

其中 z 係任意胺基酸。

問題1：ST. 26是否要求包含此序列？

是

該序列以簡寫式的形式揭露，(GGGz)₂為GGGzGGGz序列的簡寫方式。通常先展開序列，而後確定每個變數(即「z」)的定義。

該序列使用非慣用符號「z」，「z」的定義必須從揭露序列的解釋中確定，其中此符號被定義為任意胺基酸(參見本附錄的簡介)，該例示未對「z」提供任何約束，例如：在每次出現時都相同。

例示中的肽具有8個列舉胺基酸，其中6個為具體定義的甘胺酸殘基，其餘2個變數「z」建議使用慣用符號「X」表示。ST. 26第7(b)段要求序列表中所包含的序列為具有其專屬序列識別號的獨立序列。

請注意，儘管列舉及具體定義的殘基不連續，但第7(b)段仍包含該序列。

問題3：此序列於序列表中建議如何表示？

該序列使用非慣用符號「z」，根據本揭露，其為任何胺基酸。而用於表示「任何胺基酸」的慣用符號為「X」。因此，序列必須表示為單個擴展序列：

GGGXGGGX (SEQ ID NO:41)

根據第27段，符號「X」可為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V中之任一者，除非在特徵表進一步描述，由於在此例示中「X」代表「任意胺基酸」，因此必須使用特徵關鍵「VARIANT」及限定詞「note」，其值為「X 可以是任意胺基酸」。

在可行的情況下，各「X」都建議單獨註釋。然而，連續的「X」殘基區域或大量的「X」殘基分散在整個序列中，可使用特徵關鍵「VARIANT」進行描述，使用語法「x.y」作為位置描述，其中x及y是第1個及最後1個「X」殘基的位置，並使用限定詞「note」及其值為「X 可以是任意胺基酸」。

此外，該例示未揭露擴展序列中2個位置的「z」為相同的胺基酸。若揭露「z」在2個位置為相同的胺基酸，則建議使用特徵關鍵「VARIANT」及限定詞「note」，說明第4個及第8個位置的「X」可為任意胺基酸，惟這2個位置為相同胺基酸。

相關ST. 26段落：3(c)、7(b)及27。

例示27-2：簡寫式一少於4個具體定義胺基酸

一胜肽的式子為 (Gly-Gly-Gly-z)_n

該揭露進一步指出，z是任何胺基酸，且

(i)變量n是任意長度；或

(ii)變量n為2-100，最佳為3

問題1：ST. 26是否要求包含此序列？

否

由於列舉的胜肽的式子所揭露的例示(i)及(ii)揭示「n」可為「任意長度」，因此，多數涵蓋的例示的「n」是不確定的。因為「n」是不確定的，簡寫式的胜肽不能擴展到確定長度，因此，必須考慮未展開式。

未展開式(即「n」= 1)中列舉的肽包含3個具體定義胺基酸(每個都是Gly)及符號「z」。習慣上「Z」為「麩醯胺酸或麩胺酸」的符號，然而，例示將「z」定義為「任意胺基酸」(參見本附錄的簡介)，而根據ST. 26，未具體定義的胺基酸以「X」表示。基於此分析，列舉的胜肽(即GGGX)不包含4個具體定義胺基酸，因此，ST. 26第7(b)段不須包含在內，儘管在部分例示中「n」被定義為特定數值。

問題2：ST. 26是否允許包含此序列？

是

該例示為變數「n」提供特定數值，即下限為2，上限為100，以及精確值3，序列表中可以包含至少4個具體定義胺基酸的任何序列。

問題3：此序列於序列表中建議如何表示？

優選的序列包含100個GGGX重複(SEQ ID NO:42)，並進一步註釋指出最多可刪除98個GGGX重複。強烈鼓勵包含作為本發明關鍵部分的進一步具體例示。

根據第27段，「X」可為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V中之任一者，除非在特徵表進一步描述。由於SEQ ID NO:42中的「X」代表「任意胺基酸」，因此必須進行註釋，使用特徵關鍵「VARIANT」及限定詞「note」，其值為「X可以是任意胺基酸」。

在可行的情況下，各「X」都建議單獨註釋。然而，連續的「X」殘基區域或大量的「X」殘基分散在整個序列中，可使用特徵關鍵「VARIANT」進行描述，使用語法「x..y」作為位置描述，其中x及y是第1個及最後1個「X」殘基的位置，並使用限定詞「note」及其值為「X可以是任意胺基酸」。

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：3(c)、7(b)、26及27。

例示27-3：簡寫式—4個或更多具體定義胺基酸

一胜肽的式子為(Gly-Gly-Gly-z)_n

其中z是任意胺基酸，變量n為2-100，最佳為3。

問題1：ST. 26是否要求包含此序列？

是

列舉胜肽的簡寫式提供3個具體定義胺基酸(每個都是Gly)及符號「z」。習慣上「Z」為「麩醯胺酸或麩胺酸」的符號，然而，例示將「z」定義為「任意胺基酸」(參見本附錄的簡介)，而根據ST. 26，未具體定義的胺基酸以「X」表示。基於此分析，列舉的重複胜肽不包含4個具體定義胺基酸，然而，其描述提供特定的變量數值「n」，即下限為2，上限為100。因此，該例示的揭露序列GGGzGGGz具有至少6個具體定義胺基酸，ST. 26要求將其包含在序列表中。

問題3：此序列於序列表中建議如何表示？

由於「z」代表任意胺基酸，所以用於表示第4、8個胺基酸的慣用符號為「X」。ST. 26要求在序列表中僅包含已列舉殘基的獨立序列，因此，序列表中必須包含至少1個具有2、3或100個GGGX重複的序列。然而，最具包容性的序列為包含100個GGGX重複的序列(SEQ ID NO: 42)(參見本附錄的簡介)，在該情況下，可進一步註釋表明最多可刪除98組GGGX。強烈鼓勵包含分別含有2個及3個GGGX重複的附加序列(SEQ ID NO:44-45)。

根據第27段，「X」可為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V中之任一者，除非在特徵表進一步描述。由於此例示中的「X」代表「任意胺基酸」，因此必須進行註釋，使用特徵關鍵「VARIANT」及限定詞「note」，其值為「X可以是任意胺基酸」。

在可行的情況下，各「X」都建議單獨註釋。然而，連續的「X」殘基區域或大量的「X」殘基分散在整個序列中，可使用特徵關鍵「VARIANT」進行描述，使用語法「x..y」作為位置描述，其中x及y是第1個及最後1個「X」殘基的位置，並使用限定詞「note」及其值為「X可以是任意胺基酸」。

此外，該例示未揭露在擴展序列中，變數「z」每一次都是相同胺基酸。若揭露「z」在所有位置都是相同胺基酸，則建議使用特徵關鍵「VARIANT」及限定詞「note」，說明所有位置的「X」可以是任意胺基酸，惟所有位置為相同胺基酸。

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO

認為可加到原始揭露中。

相關ST. 26段落：3(c)、7(b)、26及27。

第28段—由內部終止符分隔的胺基酸序列

例示28-1：編碼核苷酸序列及編碼胺基酸序列

專利申請案描述下列序列：

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

gcg ttg gcc 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

Protein A

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 gcc gtt acc caa ctt aat cgc
Leu Gln Arg Arg Asp Trp Glu Asn Pro Gly Val Thr Gln Leu Asn Arg

Protein B

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

Protein C

tcggatgcgg cgcgacgctt atccgaccaa catatcataa

問題1：ST. 26是否要求包含此序列？

是

該案描述1個包含終止密碼子的核苷酸序列，其可編碼3個不同的胺基酸序列。列舉的核苷酸序列包含超過10個具體定義核苷酸，須作為獨立序列包含在序列表中。

而編碼的胺基酸序列，第28段要求由內部終止符號(如空格)分隔的胺基酸序列必須作為獨立序列包含在序列表中。由於「A蛋白」、「B蛋白」及「C蛋白」均包含4個或更多個具體定義胺基酸，ST. 26第7(b)段要求皆須包含在序列表中，並分配其專屬序列識別號。

問題3：此序列於序列表中建議如何表示？

該核苷酸序列須包含於序列表，並表示如下：

caattcagggtggtgaatatggcgcccaatacgc aaaccgcctctccccgcgcggttgccgattc
attaatggaaagcgggcagtgaaatgaccatgattacggattcactggccgctgttttacaacgtc

gtgactgggaaaaccctggcggttacccaacttaatcgcttgcagcacattggtgtcaaaaataa
taataaccggatgtactatattatccctgatgctgcgctcaggtgaatgaagtcgcttaagcaa
tcaatgtcggatgcggcgcgacgcttatccgaccaacatatcataa (SEQ ID NO:46)

3種蛋白質的核苷酸序列建議使用特徵關鍵「CDS」進一步描述，且INSDFeature_location要素必須標識各編碼序列的位置，包含終止密碼子。

此外，對於各特徵關鍵「CDS」，限定詞「translation」建議包含在胺基酸序列中，限定詞的值為蛋白質。該申請沒有揭露用於該轉譯的遺傳密碼表(參見附錄I，第9節，表7)。若適用標準代碼表，則限定詞「transl_table」沒有必要；然而，若應用不同的遺傳密碼表，則限定詞「transl_table」必須自表7中使用適當的限定詞值。最後，必須包含限定詞「protein_id」，其值表示各轉譯胺基酸序列的序列識別號。

該胺基酸序列必須作為獨立序列包含在序列表內，且各序列都有其專屬序列識別號：

MAPNTQTASPRALADSLMQLARQVSRLESGQ (SEQ ID NO:47)

MTMITDSLAVVLQRRDWENPGVTQLNRLAAHWCQK (SEQ ID NO:48)

MLRRQVNEVA (SEQ ID NO:49)

注意：參見「例示90-1 由含內含子編碼序列編碼的胺基酸序列」說明轉譯的胺基酸序列表示為獨立序列。

相關ST. 26段落：7、26、28、57、89-92。

第29段—「其他」胺基酸的表示

例示29-1：「其他」胺基酸的最嚴格歧義符號

專利申請描述以下序列：

Ala-Hse-X₁-X₂-X₃-X₄-Tyr-Leu-Gly-Ser

其中，X₁ = Ala或Gly，

X₂ = Ala或Gly，

X₃ = Ala或Gly，

X₄ = Ala或Gly，及

Hse = Homoserine(高絲胺酸)

問題1：ST. 26是否要求包含此序列？

是

列舉的胜肽包含5個具體定義的胺基酸。符號「X」通常用於代表替代2個胺基酸(參見本附錄的簡介)。

因為有5個具體定義胺基酸(即Ala、Tyr、Leu、Gly及Ser)，ST. 26第7(b)段要求序列必須包含在序列表中。

問題3：此序列於序列表中建議如何表示？

第29段要求任何「其他」胺基酸必須用符號「X」表示。在例示中，序列在第2位含有胺基酸Hse，這在附錄I第3節表3中沒有發現。因此，Hse是「其他」胺基酸，必須用符號「X」表示。

X₁-X₄為變體位置，各位置都可以是A或G。替代A或G的最嚴格歧義符號是「X」。因此，序列可以表示為：

AXXXXXYLGS (SEQ ID NO:50)

如本附錄的簡介所述，強烈鼓勵包含本發明揭露或請求項中必要的任何附加序列。

由於胺基酸Hse未見於附錄I第4節表4，因此根據ST. 26第30段，必須使用特徵關鍵「SITE」及限定詞「note」提供完整非縮寫的高絲胺酸名稱。

根據第27段，因為X₁-X₄只代表2個胺基酸的替代，故須進一步描述。第96段指出，建議併用特徵關鍵「VARIANT」與限定詞「note」，其值為「A或G」。根據ST. 26第34段，由於這些位置相鄰且具有相同的描述，INSDFeature_location要素可以使用語法「3..6」作為位置描述。

相關 ST. 26段落：3(a)、7(b)、25-27、29、30、34、66、70、71及96-97。

例示29-2：使用相對應未修飾胺基酸

專利申請描述以下序列：

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

其中Hyl = hydroxylysine(羥離胺酸)(離胺酸的轉譯後修飾)，Nle = Norleucine(正白胺酸)，5ALA = δ -Aminolevulinic acid(δ -氨基乙醯丙酸)

問題1：ST. 26是否要求包含此序列？

是

列舉的胜肽包含超過4個具體定義胺基酸，因此，序列必須被包含在序列表中。

問題3：此序列於序列表中建議如何表示？

位置2的羥離胺酸、位置7的正白胺酸、位置10的 δ -氨基乙醯丙酸均為「修飾胺基酸」。首先，必須考慮每個修飾的胺基酸，並確定其是否建議由相對應的未修飾胺基酸或序列的變數「X」表示。第29段指出修飾的胺基酸「建議盡可能在序列中表示為相對應的未修飾胺基酸。」

由申請人自行決定修飾的胺基酸是以相對應的未修飾殘基或變數「X」表示。但建議考慮以下指導：若胺基酸透過添加部分進行修飾，例如甲基化或乙醯化，其相對應未修飾胺基酸的基本結構通常未改變，則建議使用未修飾胺基酸表示。若修飾胺基酸在結構上與相對應未修飾胺基酸有很大不同，則建議以「X」表示。

羥離胺酸的結構與離胺酸幾乎相同，僅R基團中第三個碳被羥基修飾，由於相對應的未修飾離胺酸殘基的基本結構未改變，羥離胺酸在序列中建議由賴胺酸（「K」）表示，而非「X」。

正白胺酸是白胺酸的異構體。白胺酸的R基團為4碳鏈，在第2個碳上分支。正白胺酸亦具4碳R基團，但其係線性無支鏈。因此，正白胺酸不僅是對白胺酸的修飾結果，而是一種完全不同（儘管相關）的結構。因此建議正白胺酸在序列表中以「X」表示。

δ -氨基乙醯丙酸在結構上與附錄I表3中列出的任何胺基酸都不相似。因此，建議 δ -氨基乙醯丙酸在序列表中用「X」表示。

因此，該序列建議包含在序列表中，如下所示：

AKYLG SXVSX (SEQ ID NO:51)

第30段要求對各修飾的胺基酸進一步註釋。

羥離胺酸係離胺酸的轉譯後修飾。因此，必須使用特徵關鍵「MOD_RES」以及限定詞「note」來描述修飾。請注意，「羥離胺酸」列在附錄I第4節表4的「修

飾胺基酸列表」，因此，限定詞「note」的值可為縮寫「Hyl」而非完整非縮寫名稱「hydroxylysine」。

正白胺酸不是轉譯後修飾的殘基，因此必須使用特徵關鍵「SITE」以及限定詞「note」來描述修飾。請注意，「正白胺酸」也列在附錄1第4節的表4中，因此，限定詞「note」的值可以包含縮寫「Nle」而非完整非縮寫的名稱「Norleucine」。

δ -氨基乙醯丙酸也不是轉譯後修飾的殘基，因此必須使用特徵關鍵「SITE」以及限定詞「note」來描述修飾。 δ -氨基乙醯丙酸未列在附錄1第4節的表4中，因此，限定詞「note」的值必須包含修飾殘基的完整非縮寫的名稱「 δ -Aminolevulinic acid」。

相關ST. 26段落：3(a)、3(e)、7(b)、29及30。

第30段—修飾胺基酸的註釋

例示30-1：特徵關鍵「CARBOHYD」

專利申請案描述一種具有特殊修飾胺基酸的多肽，含有糖基化側鏈，其特徵在於多肽的第4位置及第15位置對應的Cys形成二硫鍵，序列如下所示：

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

問題1：ST. 26是否要求包含此序列？

是

列舉的胜肽提供17個具體定義的胺基酸，其中天然胺基酸有16個，而第9位置(天冬醯胺酸)是糖基化的。因此，該序列必須依據ST. 26第(7)(b)段的要求包含在序列表中。

問題3：此序列於序列表中建議如何表示？

根據ST. 26第29段，修飾的胺基酸建議在序列中儘可能使用相對應的未修飾胺基酸表示。

因此，該序列必須包含在序列表中，如下所示：

LEYCLKRWNETISHCAW (SEQ ID NO:52)

須進一步描述修飾的胺基酸。特徵關鍵「CARBOHYD」連同(必要性)限定詞「note」建議用於表示糖鏈(去唾液酸寡糖)在第9位置與天冬醯胺酸連接。限定詞「note」描述連接的類型，例如N-連鎖(N-linked)。特徵位置要素中的位置描述為修飾的天冬醯胺酸的殘基位點編號。

此外，2個半胱胺酸殘基的間存在二硫鍵。因此，建議使用特徵關鍵「DISULFID」來描述鏈內交聯。特徵位置要素是連接的 Cys 殘基以「x.y」格式表示，即「4..15」。必要限定詞「Note」建議描述鏈內二硫鍵。

相關ST. 26段落：3(a)、7(b)、26、29、30、66(c)、70及附錄I第7節特徵關鍵7.4。

第36段—包含精確數量連續「n」或「X」殘基區域的序列

例示36-1：具有一個已知「X」殘基數量區域的序列表示為獨立序列

LL-100-KYMR

其中亮胺酸及賴胺酸的間的「-100-」反映序列中100個胺基酸的區域。

問題1：ST. 26是否要求包含此序列？

是

ST. 26第36段要求包含1個序列，該序列包含至少4個具體定義胺基酸，由1個或多個定義數量的「X」殘基組成的區域分隔。

揭露的序列使用非慣用符號，即「-100-」。「-100-」的定義必須從揭露中對序列的解釋來確定，將這個符號定義為亮胺酸及賴胺酸的間的100個胺基酸(參見本附錄的簡介)。因此，「-100-」是「X」殘基的定義區域。

由於序列中的106個胺基酸中有6個被具體定義，ST. 26第7(b)段要求序列必須包含在序列表中。

問題3：此序列於序列表中建議如何表示？

非慣用符號「-100-」表示為100個「X」殘基(因為用於表示胺基酸的任何符號都只相當於1個殘基)。因此，長度為106個胺基酸、在LL及KYMR的間含有100個「X」殘基的單個序列，必須包含在序列表中(SEQ ID NO:53)。

該序列包含LL及KYMR的間的100個變數「X」，若沒有進一步註釋，則ST. 26默認「X」為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V中之任一者(第27段)。若這100個變數「X」被定義為除此默認值以外的任何值，則必須為每個變數「X」提供適當的註釋。

相關ST. 26段落：7(b)、26、27及36。

並使用限定詞「note」及其值為「X可以是任意胺基酸」。

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：27及36。

第37段—包含未知數目連續「n」或「X」殘基區域的序列

例示37-1：具有未知數量「X」殘基區域的序列不得表示為獨立序列

Gly-Gly----Gly-Gly-Xaa-Xaa

其中符號----係序列中未定義的空位，Xaa是任何胺基酸，甘胺酸及Xaa殘基透過肽鍵相互連接。

問題1：ST. 26是否要求包含此序列？

否

ST. 26第37段禁止包含任何包含未定義分隔的序列；因此，不須列入整個序列。ST. 26第37段確實要求包含與未定義分隔相鄰的任何區域的序列，該序列含4個或更多具體定義的胺基酸。在上述例示中，不須包含與未定義分隔相鄰的任一區域，因為各區域僅包含2個具體定義的胺基酸。

問題2：ST. 26是否允許包含此序列？

否—不是整個序列

否—不是序列的任何區域

ST. 26第37段不允許包含整個序列。

ST. 26第8段不允許包含與未定義分隔相鄰的任何區域，因為各區域僅包含2個具體定義的胺基酸。

相關ST. 26段落：7(b)、8、26及37。

例示37-2：具有未知數量「X」殘基區域的序列不得表示為獨立序列

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

其中符號----係序列中未定義的空位，Xaa是任何胺基酸，甘胺酸及Xaa殘基透過肽鍵相互連接。

問題1：ST. 26是否要求包含此序列？

否—不是整個序列

是—序列的1個區域

ST. 26第37段禁止包含任何包含未定義分隔的序列，但與未定義間隔相鄰，且包含4個或更多具體定義胺基酸的任何序列區域須列入。

在上面的例示中，ST. 26不要求(並且禁止)包含未定義的間隔的整個序列及與未定義間隔相鄰的Gly-Gly區域，該區域僅包含2個具體定義的胺基酸。然而，ST. 26要求包含與未定義間隔相鄰的Gly-Gly-Ala-Gly-Xaa-Xaa區域，因為他包含至少4個具體定義的胺基酸。

問題2：ST. 26是否允許包含此序列？

否—不允許整個序列，及Gly-Gly區域

問題3：此序列於序列表中建議如何表示？

與包含4個具體定義胺基酸的未定義間隔相鄰的序列區域必須表示如下：

GGAGXX (SEQ ID NO:58)

該序列建議使用特徵關鍵「SITE」、特徵位置「1」及限定詞「note」的值進行註釋，例如「該殘基在N端連接到具有N端Gly-Gly及未定義長度間隔的胜肽」，以表明所表示的序列是包含未定義間隔的更大序列的一部分。

根據第27段，「X」可為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V中之任一者，除非在特徵表進一步描述。由於SEQ ID NO: 58中的「X」代表「任意胺基酸」，因此必須進一步註釋，使用特徵關鍵「VARIANT」及限定詞「note」，其值為「X可為任意胺基酸」。

在可行的情況下，各「X」都建議單獨註釋。然而，連續的「X」殘基區域或大量的「X」殘基分散在整個序列中，可使用特徵關鍵「VARIANT」進行描述，使用語法「x..y」作為位置描述，其中x及y是第1個及最後1個「X」殘基的位置，並使用限定詞「note」及其值為「X可以是任意胺基酸」。

相關ST. 26段落：7(b)、8、26、27及37。

第55段—包含DNA及RNA片段的核苷酸序列

例示55-1：DNA/RNA組合分子

一項專利申請描述寡核苷酸序列，如下所示：

AGACCTTcggagucuccuguugaacagauagucaaaaguagauC

其中大寫字母代表DNA殘基，小寫字母代表RNA殘基。

問題1：ST. 26是否要求包含此序列？

是

所揭露的序列具有超過10個列舉及具體定義的核苷酸；因此他必須包含在序列表中。

問題3：此序列於序列表中建議如何表示？

核苷酸序列必須包含在序列表中，如下所示：

agaccttcggagtctcctgttgaacagatagtagtagatc (SEQ ID NO:93)

請注意，尿嘧啶核苷酸必須在序列表中用符號「t」表示。

ST. 26第55段規定，包含DNA及RNA片段的核苷酸序列必須指明分子類型「DNA」，進一步使用特徵關鍵「source」、必要性限定詞「organism」，其值為

「synthetic construct」，以及必要限定詞「mol_type」，其值為「other

DNA」。此外，序列的每一段都必須進一步使用特徵關鍵「misc_feature」，包含片段的位置，以及限定詞「note」表示該片段是DNA或RNA。揭露的序列包含2個DNA片段(核苷酸位置1-7及43)及1個RNA片段(核苷酸位置8-42)。

相關ST. 26段落：7、14、55-56及83。

第89段—特徵關鍵「CDS」

例示89-1：編碼核苷酸序列及編碼胺基酸序列

一項專利申請描述以下核苷酸及其轉譯序列：

```
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
```

問題1：ST. 26是否要求包含此序列？

是

列舉的核苷酸序列具有超過10個具體定義的核苷酸。

列舉的胺基酸序列具有超於4個具體定義的胺基酸。

問題3：此序列於序列表中建議如何表示？

核苷酸序列必須表示為：

```
atgaccggaataaacctgaaaccgatgtttacgaaattttatga (SEQ ID NO:59)
```

核苷酸序列建議使用「CDS」特徵關鍵進一步描述，INSDFeature_location要素必須標識整個序列，包含終止密碼子(即位置1到45)。此外，限定詞「translation」建議包含限定詞值「MTGNKPETDVYEIL」。申請未揭露用於轉譯的遺傳密碼表(參見附錄I第9節表7)，若適用標準代碼表，則不須限定詞「transl_table」；然而，若適用不同的遺傳密碼表，則限定詞「transl_table」必須指明來自表7中適當的限定詞值。最後，限定詞「protein_id」必須包含限定詞值指明其轉譯的胺基酸序列的序列識別號。

胺基酸序列必須分別使用單字母代碼，並呈現其專屬的序列識別號，如下所示：

```
MTGNKPETDVYEIL (SEQ ID NO:60)
```

列舉胺基酸序列後面的STOP不得包含在序列表的胺基酸序列中。

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：7(a)、7(b)、26、28、89、90及92。

例示89-2：特徵位置超出揭露的序列

專利申請案包含下圖，揭露部分編碼序列及其轉譯的胺基酸序列：

```
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
```

圖1-智人 *ITCH1* 基因的部分編碼序列，其編碼442個胺基酸長的ITCH1蛋白的第20到91個胺基酸。

問題1：ST. 26是否要求包含此序列？

是

該申請揭露核苷酸序列及其轉譯的胺基酸序列。列舉的核苷酸序列包含超過10個具體定義的核苷酸，必須包含在序列表中。

胺基酸序列包含超過4個具體定義的胺基酸，並且還必須作為1個單獨的序列包含在序列表中，具有專屬序列識別號。

問題3：此序列於序列表中建議如何表示？

核苷酸序列必須包含在序列表中，如下所示：

```
catcacgcagcagaatgtggatTTTgtcctcaacaatggcaagttctacgtgggagtctgtg
catttgtgagggtccagctgaaggatggttcatatcaagatgttggttatggtgtgggcctc
aagtccaaggctttatcTTTggagaaggcaaggaaggaggcggtgacagacgggctgaagcg
agccctcaggagTTTTgggaatgcacttgg (SEQ ID NO:94)
```

核苷酸序列建議使用「CDS」特徵關鍵進一步描述，INSDFeature_location要素必須標示「CDS」特徵的位置，並且必須包含終止密碼子。

該圖描述不包含起始密碼子或終止密碼子的部分編碼序列。然而，序列的描述表明起始密碼子在第1個核苷酸的上游，終止密碼子位於第216位置的最後1個核苷酸的下游。

ST. 26規定位置描述不得包含超出INSDSeq_sequence要素的序列範圍的殘基編號。因此，在上述例示中，CDS特徵關鍵的位置描述不能包含1到216範圍之外的位置編號。INSDfeature_location要素中終止密碼子的位置必須使用符號「>」表示，以指示終止密碼子的位置位於第216位置的下游。同樣，符號「<」可用於指示起始密碼子的位置在第1位置的上游。因此，CDS 特徵關鍵的位置描述建議如下所示：

<1..>216

請注意，「<」及「>」是保留字元，在XML序列表的實例中將分別被替換為「<」及「>」。

「translation」限定詞建議包含在蛋白質的胺基酸序列中作為限定詞值。該圖未揭露適用於轉譯的遺傳密碼表(參見附錄I第9節表7)，若適用標準代碼表，則不須限定詞「transl_table」；然而，若適用不同的遺傳密碼表，則必須為限定詞「transl_table」指明ST. 26附錄I表7中的適當限定詞值。最後，限定詞「protein_id」必須包含在CDS特徵，限定詞的值為轉譯的胺基酸序列的序列識別號。

轉譯的胺基酸序列必須作為單個序列包含在內，並具其專屬的序列識別號：

HHAAECGFCPQQWQLRGSLCICEGPAEGWFI SRCWLWCGPQVQGF IFGEGKEGGDRRAEASP
QEFWECTW(SEQ ID NO:95)

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：7、41、65、66、70、71、89及92。

第92段—由編碼序列編碼的胺基酸序列

例示92-1：由具有內含子的編碼序列編碼的胺基酸序列

專利申請案包含下圖，揭露編碼及其轉譯序列：

```
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

ttgatacagag actctacgcc gagctgacgt gctaccgtat 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 acc
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 tgc 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
```

圖1 - 粗體顯示的核苷酸為內含子區域

問題1：ST. 26是否要求包含此序列？

是

所揭露的序列具有超過10個列舉及具體定義的核苷酸；因此他必須包含在序列表中。

該申請揭露一個核苷酸序列及其轉譯的胺基酸序列。列舉的核苷酸序列包含10個以上具體定義的核苷酸，必須列為單個序列包含在序列表中。

核苷酸序列包含由非編碼序列(內含子)分隔的編碼序列(外顯子)。此圖描繪將核苷酸序列轉譯為3個不連續的胺基酸序列。根據圖片說明，核苷酸的粗體區域為內含子序列，在轉譯成蛋白質之前，這些序列將從RNA轉錄物中剪接出來。因此，這3個胺基酸序列實際上是1個單一、連續且列舉的序列，其包含超過4個具體定義的胺基酸，必須作為單個序列包含在序列表中。

問題3：此序列於序列表中建議如何表示？

核苷酸序列必須包含在序列表中，如下所示：

```
atgaagacttttcgcagccttgctttccgctgtcactctcgcgctctcgggtgcgccccaggcgg
ctgtctggagtcaatgtaagtgccgctgcttttcatgtgatacagactctacgccgagctgacg
```

tgctaccgtataggtggcggtacaccgggttgacgggagaccacttgcgttgctggttcgg
tttgtacctccttgagctcagtgagcgactttcaatccgctcgtcattgctcctcatgtattgac
gattggccttcatagtcatactctcaatgcgttccgggctccgcaacgtccagcgtccggcg
ccccctcagcgacaacttcaggccccgcacctacggacggaacgtgctcggccagcggggcatg
gccgccattgacctga (SEQ ID NO:75)

核苷酸序列建議使用特徵關鍵「CDS」進一步描述，INSDFeature_location要素必須標示編碼序列的位置，包含由「Ter」指示的終止密碼子。CDS INSDFeature_location必須使用「join」位置運算符來指示由指定位置編碼的轉譯產物被連接，並使用「join(x1..y1, x2..y2, x3..y3)」格式形成單個連續的多肽，例如，「join(1..79, 142..212, 272..400)」。此外，限定詞「translation」建議包含在內，限定詞的值為蛋白質的胺基酸序列（注意終止符「Ter」在序列的最後位置不得包含在胺基酸序列中）。申請未揭露用於轉譯的遺傳密碼表（參見附錄I第9節表7）。若適用「標準代碼」表，則不須限定詞「transl_table」；然而，若適用不同的遺傳密碼表，則必須於限定詞「transl_table」指明適當的限定詞值來自表7。最後，限定詞「protein_id」必須以限定詞值表示轉譯的胺基酸序列之序列識別號。胺基酸序列必須作為單個序列包含在內：

MKTFAALLSAVTLALS VRAQAAVWSQCGGTPGWTGETTCVAGSVCTSLSSYSQCVPGSATSSA
PAAPSATTSGPAPT DGTCSASGAWPPLT (SEQ ID NO:76)

相關ST. 26段落：7、26、28、57、67及89-92。

第93段—以殘基列舉的主要序列及其變體

例示93-1：列舉變體的表示

專利申請案描述包含以下序列比對。

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

問題1：ST. 26是否要求包含此序列？

是

在本領域中常見序列比對中以「點」指示「此位置與其上面序列該位置者相同」，因此，*D. virilis*及*D. simulans*序列中的「點」被認為是列舉及具體定義的核苷酸，因為他們只係指示在特定位置具有與*D. melanogaster*相同核苷酸的簡寫方式。此外，序列比對經常顯示符號「-」表示殘基不存在，以最大化比對。因此，*D. melanogaster*及*D. simulans*的核苷酸序列包含22個列舉的及具體定義的核苷酸，而 *D. virilis*的核苷酸序列包含19個。因此，ST. 26第7(a)段要求將每個序列包含在序列表中，具有單獨序列識別號。

問題3：此序列於序列表中建議如何表示？

*Drosophila melanogaster*序列須包含於序列表，表示如下：

acattgaatctcataccacttt (SEQ ID NO:61)

*Drosophila virilis*序列須包含於序列表，表示如下：

acatggatcccacgacttt (SEQ ID NO:62)

*Drosophila simulans*序列須包含於序列表，表示如下：

gtatggcgtcgtatsgtagttt (SEQ ID NO:63)

相關ST. 26段落：7(a)、13及93。

例示93-2：列舉變體的表示

該描述包含下表的肽及其功能變體。下面表格中的空白區域表示變體中的胺基酸與「序列」中相對應的胺基酸相同，「-」表示「序列」中相對應的胺基酸缺失。

Position	1	2	3	4	5	6	7	8	9
Sequence	A	V	L	T	Y	L	R	G	E
Variant 1									A
Variant 2			P			P			
Variant 3			A	I	G	Y			
Variant 4							-		

問題1：ST. 26是否要求包含此序列？

是

如上所示，此表中的空格表示變體中的胺基酸與「序列」中相對應的胺基酸相同。因此，變體序列的胺基酸被列舉且具體定義。

由於4種變體序列皆包含超過4個列舉且具體定義的胺基酸，ST. 26第7(b)段要求每個序列都包含在序列表中，且具有其專屬序列識別號。

問題3：此序列於序列表中建議如何表示？

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)

相關ST. 26段落：7(b)、26及93。

例示93-3：共有序列的表示

專利申請案具有以下序列比對的圖1

<i>Consensus</i>	LEGnEQFINAakIIRHPkYnrkTlnNDImLIK
<i>Homo sapiens</i>	LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK
<i>Pongo abelii</i>	LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK
<i>Papio Anubis</i>	LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK
<i>Rhinopithecus roxellana</i>	LEGTEQFINAAKIIRHPNYNRITLDNDILLIK
<i>Pan paniscus</i>	LEGNEQFINAAKIIRHPKYNRITLNNDIMLIK
<i>Rhinopithecus bieti</i>	LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK
<i>Rhinopithecus roxellana</i>	LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK

共有序列包含大寫字母，代表保守胺基酸殘基，而小寫字母「n」、「a」、「k」、「r」、「l」及「m」代表比對序列中的主要胺基酸殘基。

問題1：ST. 26是否要求包含此序列？

是

共有序列中的小寫字母各自代表1個胺基酸殘基。因此，共有序列及圖1中其餘7個序列每個都包含至少4個具體定義的胺基酸。ST. 26第7(b)段要求在序列中包含所有8個序列。

問題3：此序列於序列表中建議如何表示？

共有序列中的小寫字母被用作歧義符號來表示可能變體特定位置上的主要胺基酸。因此，小寫字母「n」、「a」、「k」、「r」、「l」及「m」係以非慣用方式使用的慣用符號，共有序列必須使用歧義符號代替各小寫字母來表示。

建議使用最嚴格的歧義符號。對於共有序列中的大多數位置，「X」是最嚴格的歧義符號；但是，位置20的最嚴格歧義符號為「D」或「N」，位置25是「B」。共有序列建議包含在序列表中，如下所示：

LEGXEQFINAXXIIRHPXYBXXTXBNNDIXLIK (SEQ ID NO:82)

根據第27段，「X」被解釋為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V，除非其於特徵表進一步描述。因此共有序列中的「X」必須用特徵關鍵「VARIANT」及限定詞「note」，其值為「X為任意胺基酸」。

其餘7個序列必須包含在序列表中，如下所示：

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)

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：7(b)、26、27、93及97。

第94段—以單一序列含有列舉替代殘基方式揭露的變體序列

例示94-1：含有列舉替代胺基酸的獨立序列的表示方式

專利申請案請求一胜肽序列，如下：

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

問題1：ST. 26是否要求包含此序列？

是

該序列提供4個具體定義的胺基酸，且ST. 26第7(b)段要求序列表中包含該序列。

問題3：此序列於序列表中建議如何表示？

附錄I表3第3節將歧義符號「J」定義為異白胺酸或白胺酸。因此，序列的較佳表示方式為：

GGGJATX (SEQ ID NO:64)

建議使用特徵關鍵「VARIANT」及限定詞「note」在特徵表中進行進一步描述「X」為絲胺酸或蘇胺酸。

或者，序列可以表示為，例如：

GGGLATS (SEQ ID NO:65)

建議使用特徵關鍵「VARIANT」及限定詞「note」在特徵表中進行進一步描述L可以被I代替，S可以被T代替。

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：7(b)、8、26、27、94及97。

第95(a)段—僅透過提及具有多個獨立變數的主要序列而揭露的變體序列

例示95(a)-1：透過主要序列的註釋表示變體序列

專利申請案包含以下揭露內容：

「胜肽片段1為Gly-Leu-Pro-Xaa-Arg-Ile-Cys，其中Xaa可以是任何胺基酸……。」

另1個例示中，胜肽片段1為Gly-Leu-Pro-Xaa-Arg-Ile-Cys，其中Xaa可以是Val、Thr 或Asp...

另1個例示中，胜肽片段1為Gly-Leu-Pro-Xaa-Arg-Ile-Cys，其中Xaa可以是Val。」

問題1：ST. 26是否要求包含此序列？

是

3個揭露的例示中的每個「胜肽片段1」皆提供至少6個具體定義的胺基酸；因此，該序列必須根據ST. 26第7(b)段的要求包含在序列表中。

問題 3：此序列於序列表中建議如何表示？

在這個例子中，「胜肽片段1」的列舉序列被揭露3次，且3個不同的例示各自有其Xaa的替代描述。在此例示中，「X」是Xaa位置的最嚴格的歧義符號。

ST. 26要求僅包含一次揭露的列舉序列。在三者中最具包容性的例示中，Xaa是任何胺基酸(參見本附錄的簡介)。因此，序列表中必須包含的序列為：

GLPXRIC (SEQ ID NO:66)

根據第27段，「X」被解釋為A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、O、S、U、T、W、Y或V中之任一者，除非其於特徵表進一步描述。由於SEQ ID NO:66的「X」表示「任意胺基酸」，因此必須用特徵關鍵「VARIANT」及限定詞「note」，其值為「X為任意胺基酸」。

在可行的情況下，每個「X」都建議單獨註釋。然而，連續「X」殘基的區域或分散在整個序列中的大量「X」殘基，可使用特徵關鍵「VARIANT」以「x..y」作為位置描述，其中x及y是第一個及最後一個「X」殘基的位置，並使用限定詞「note」其值為「X為任意胺基酸」。

如本附錄的簡介所述，強烈鼓勵包含本發明揭露或請求項中必要的任何附加序列。

對於上述的例示，強烈建議將以下3個額外序列包含在序列表，且每個都有其專屬的序列識別號：

GLPVRIC (SEQ ID NO:67)

GLPTRIC (SEQ ID NO:68)

GLPDRIC (SEQ ID NO:69)

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：7(b)、26、27及95(a)。

第95(b)段—僅透過提及具有多個非獨立變數的主要序列而揭露的變體序列

例示95(b)-1：以多個非獨立變數表示獨立變體序列

專利申請案描述以下共有序列：

cgaatgn₁cccactacgaatgn₂cacgaatgn₃cccaca

其中n₁、n₂及n₃可以是a、t、g或c。

變體序列揭露如下：

若n₁是a，則n₂及n₃是t、g或c；

若n₁是t，則n₂及n₃是a、g或c；

若n₁是g，則n₂及n₃是t、a或c；

若n₁是c，則n₂及n₃是t、g或a。

問題1：ST. 26是否要求包含此序列？

是

此序列具有超過10個列舉及「具體定義」的核苷酸，並且依ST. 26第7(a)段要求包含在序列表中。

問題3：此序列於序列表中建議如何表示？

列舉的序列包含超過10個具體定義的核苷酸及3個「n」殘基。ST. 26要求包含揭露的列舉序列，且在適用歧義符號的地方，建議使用最嚴格的歧義符號。在此例示中，n₁、n₂及n₃可以是a、t、g或c，所以「n」是最嚴格的歧義符號。因此，該序列必須包含在序列表中，如下所示：

cgaatgncccactacgaatgncacgaatgncccaca (SEQ ID NO:70)

ST. 26第15段規定「符號「n」將被解釋為「a」、「c」、「g」或「t/u」中的任何1個，除非於特徵表中進一步描述。此序列中各「n」殘基的值相當於默認的「a」、「c」、「g」或「t」，不須進一步的註釋。

列舉序列包含3個不同位置的變化，並且變化的發生係相互依存。強烈建議包含代表發明的關鍵附加例示的附加序列，如同本文檔的介紹中所討論。因此，根據ST. 26第95(b)段，附加例示建議作為4個單獨序列包含在序列表中，各序列都有專屬序列識別號：

cgaatgaccactacgaatgbcacgaatgbccaca (SEQ ID NO:71)

cgaatgtcccactacgaatgvcacgaatgvcccaca (SEQ ID NO:72)

cgaatggcccactacgaatghcacgaatghcccaca (SEQ ID NO:73)

cgaatgccccactacgaatgdcacgaatgdcccaca (SEQ ID NO:74)

(注意b = t、g或c；v = a、g或c；h = t、a或c；d = t、g或a；參見附錄I

第1節表1)

根據ST. 26第15段，必須使用最嚴格的符號來表示可變位置。因此，在序列表中n₂及n₃必不能用「n」表示。

注意：上述序列的優選表示是針對在專利申請日提交的序列表。相同的陳述可能不適用於專利申請日後提交序列表，因為必須考量所提供的資訊是否被IPO認為可加到原始揭露中。

相關ST. 26段落：7(a)、15及95(b)。

附件

XML格式序列的指南文件(Guidance document sequence in XML)

附件可在以下網址取得：

https://www.wipo.int/standards/en/xml_material/st26/st26-annex-vi-appendix-guidance-document-sequences.xml

附錄 VII

ST. 25轉換到ST. 26序列列表之建議 可能增加或刪除的主題事項(SUBJECT MATTER)

版本 1.5

WIPO標準委員會(CWS)於2021年11月5日第九屆會議上批准修訂

介紹

WIPO ST. 25及ST. 26標準對於核苷酸及胺基酸序列的呈現要求有所不同。因此，引發了一個問題：是否需要在符合WIPO ST. 26標準的國際申請所提交的序列表中增加或刪除任何主題，這些主題可能無法被提出優先權的申請所支持。

本文件範圍

本文件探討ST. 26的強制要求以及可能產生的任何後果。本文件未涵蓋每種可能的情況；如果ST. 26中對於ST. 25序列列表資訊的表達方式不清楚，則可在申請說明書中包含該等資訊，以避免相關主題事項被刪除。

可能添加或刪除主題事項的建議

檢視本文件所包含的問題可以證明，從ST. 25轉換到ST. 26本質上(inherently)不應該導致增加或刪除主題事項，尤其是，當ST. 25 序列列表完全符合ST. 25標準時。然而，有些情況需要申請人謹慎處理。我們提供了一些建議以避免增加或刪除主題事項。

情境1

ST. 25使用數字識別號標記各種類型的數據(data)，例如<110>用於申請人姓名。ST. 26使用英語單詞作為要素名稱及屬性來標記數據。

建議：

ST. 26術語僅描述數據內容的類型，因此使用ST. 26要素名稱及屬性，不會構成添加主題事項。

情境2

ST. 26明確要求，包括：(a)分支序列；(b)帶有D-胺基酸的序列；(c)核苷酸類似

物;及(d)帶有缺失鹼基的序列。ST. 25對包含或禁止此類序列的要求並不明確。

建議：

即使可能沒有包含在ST. 25序列表中，應在申請說明書中有揭露足夠的相關資訊，以在ST. 26序列表中表示此類的序列。對於此類序列的資訊，必須注意不要添加超出已揭露範圍主題事項。例如，對於核苷酸序列的mol_type限定詞，請參閱下面的討論(在情境4中)。

情境3

ST. 26排除了少於10個具體定義的核苷酸(不包括“n”)及少於4個具體定義的胺基酸(不包括“X”)序列。

建議：

排除的序列可以記載在說明書中。

情境4

ST. 26有強制要求的特徵關鍵--所有核苷酸序列需要“source”特徵關鍵，所有胺基酸序列需要“source”特徵關鍵，每個特徵關鍵都有2個強制要求的限定詞。ST. 25為核苷酸序列雖有定義對應的特徵關鍵(但很少使用)，然而並沒有對應的限定詞，對於胺基酸序列則沒有相應的特徵關鍵。

核苷酸序列

ST. 26 - 特徵關鍵5.37 source；強制要求的限定詞6.45 organism及6.39 mol_type (請參見ST. 26第75段)

限定詞	限定詞的值
mol_type	genomic DNA
	genomic RNA
	mRNA
	tRNA
	rRNA
	other DNA (適用於合成分子)
	other RNA (適用於合成分子)
	transcribed RNA
	viral cRNA
	unassigned DNA (適用在體內分子未知情況下)
	unassigned RNA (適用在體內分子未知情況下)

胺基酸序列

ST. 26 - 特徵關鍵7.30 source；強制要求的限定詞8.3 organism及8.1 mol_type（請參見ST. 26第75段）

限定詞	限定詞的值
mol_type	protein

建議：

唯一需要關注的議題是，與核苷酸序列的mol_type限定詞相關的控制詞彙值。然而，上述某些值的選項，可能在說明書揭露中得不到足夠的支持。然而，可以透過使用對特定序列的最通用值(the most generic value)來避免增加主題事項。例如，對於一種合成分子使用“other DNA”及“other RNA”，對於體內分子使用“unassigned DNA”及“unassigned RNA”。

情境5

當序列中包含“Xaa”，ST. 25要求在<221>(名稱)及<222>(位置)關聯下的<223>(其他資訊)中，記載包含有關該殘基的進一步資訊。ST. 25未為“Xaa”(ST. 26中的“X”)提供默認值(default value)。但是，ST. 26提供了默認值，因此，不總是需要進一步資訊。對於胜肽序列，最常用於變量“Xaa”或“X”的詮釋是“任何胺基酸”或“任何天然胺基酸”，然而這樣的註釋可被解釋為，包含ST. 25或ST. 26中未有列出的胺基酸。ST. 26對於“X”的默認值是沒有進一步的注釋，“X”則是在附錄I(見第3節，表3)中所列出的22種胺基酸之一。ST. 26的默認值可能會構成添加或刪除主題事項，因此從ST. 25轉換到ST. 26時，可能會對專利申請的範圍產生不利影響。

建議：

- (a) 當ST. 25序列列表中包含<221>名稱、對應於Xaa的<222>位置以及<223>Xaa的進一步其他資訊，並且<221>名稱是有適當的ST. 26特徵關鍵，例如SITE、VARIANT或UNSURE，那麼應該使用ST. 26特徵關鍵。此外，為了避免潛在可能被刪除的主題事項，必須在相應的限定詞“note”中包含<223>所記載的其他資訊。
- (b) 當ST. 25序列列表中包含<221>名稱、對應於Xaa的<222>位置<223>Xaa的進一步其他資訊，並且<221>名稱不是ST. 26特徵關鍵，那麼應該使用適當的ST. 26特徵鍵，SITE或REGION。此外，為了避免潛在可能被刪除的主題事項，必

須在相應的限定詞“note”中包含<223>所記載的其他資訊，以及不適當的<221>名稱。例如，ST. 25序列表使用了一個不在ST. 25或ST. 26中的<221>名稱Variable，並附帶<223>Xaa是任何胺基酸的進一步其他資訊。在這個例子中，ST. 26中“note”限定詞的值會是記載“Variable - Xaa是任何胺基酸”。

- (c) 當ST. 25序列表中沒有對應於Xaa的<221>、<222>或<223>，或者是，包含與Xaa相對應的<221>及<222>，但是沒有對應的<223>。上述該等情況都不符合ST. 25要求，但實務上仍然會發生。對上述該等情況，ST. 26中“note”限定詞的值會是記載，任何包含在說明書中對於“Xaa”的記載資訊，並且給予適當的特徵關鍵，例如SITE、REGION或UNSURE，以及其特徵位置。

情境6

在ST. 25中，在序列中尿嘧啶用“u”表示、胸腺嘧啶用“t”表示。在ST. 26中，在序列中尿嘧啶及胸腺嘧啶都用“t”表示，並且沒有進一步的註釋；“t”表示RNA中的尿嘧啶及DNA中的胸腺嘧啶。

建議：

- (a) 當DNA序列中包含尿嘧啶，ST. 26將其視為一修飾核苷酸，要求使用“t”來表示尿嘧啶，並使用特徵關鍵“modified_base”、限定詞“mod_base”（其限定詞值為“OTHER”）及限定詞“note”（其限定詞值為“uracil”）進一步描述。當ST. 25中DNA序列包含“u”，則不會認為此ST. 26註釋是新增主題事項。
- (b) 當RNA序列中包含胸腺嘧啶，ST. 26將其視為一修飾核苷酸，要求使用“t”來表示胸腺嘧啶，並且使用特徵鍵“modified_base”、限定詞“mod_base”（其限定詞值為“OTHER”）及限定詞“note”（其限定詞值為“thymine”）進一步描述。當ST. 25中RNA序列包含“t”，則不會認為此ST. 26註釋新增主題事項。

情境7

在ST. 25及ST. 26中，修飾核苷酸或胺基酸必須有進一步描述。在ST. 26中，若合適，可以使用附錄I第2節第2表中的縮寫來表示修飾核苷酸的身分。否則，必須表示修飾核苷酸完整且非縮寫的名稱。同樣地，若合適，可以使用附錄I第4節第4表中的縮寫來表示修飾胺基酸的身份。否則，必須表示修飾胺基酸完整且非縮寫的名稱。相較之下，若修飾殘基不包含在ST. 25序列表中，則不需要使用完整且非縮寫的名稱，通常會使用縮寫代替完整名稱。

建議：

- (a) 當說明書及ST. 25序列表中使用，未列於附錄I第2節表2或第4段表4中修飾核苷酸或胺基酸的縮寫名稱，並且該技術領域者已知該縮寫名稱只會表示一種具體的修飾核苷酸或是修飾胺基酸，那麼ST. 26中使用完整且非縮寫名稱並不會構成添加主題事項。
- (b) 當說明書及ST. 25序列表中只使用，未列於附錄I第2節表2或第4段表4中修飾核苷酸或胺基酸的縮寫名稱(且說明書中不包含其化學結構)，並且該技術領域者不認為該縮寫名稱只會表示一種具體的修飾核苷酸或修飾胺基酸，即該縮寫名稱在該技術領域中可能代表多種不同的修飾核苷酸或修飾胺基酸。導致在不引入添加主題事項的情況下，是無法符合ST. 26的規定。當然這種情況可能是，優先權申請案及序列表都未有清楚記載。為避免潛在可能被刪除的主題事項，ST. 26中除了包含修飾核苷酸或修飾胺基酸的完整且非縮寫名稱以外，應該將ST. 25序列表中的縮寫名稱放在ST. 26的限定詞“note”中，加以描述。前述情況下，ST. 26序列表中對於修飾殘基所要求的完整且非縮寫名稱，不享有先前申請的優先權。因此，為避免未來可能發生的問題，應注意草擬原始的(ST. 25)序列表及申請揭露文件時，應記載有完整且非縮寫名稱。

情境8

ST. 25包含許多ST. 26中沒有的特徵關鍵，因此申請人必須以符合ST. 26規定納入先前ST. 25特徵關鍵中的資訊，才不會引入添加或刪除的主題事項。

建議：

以下表格提供關於以符合ST. 26規定納入先前ST. 25特徵關鍵中所包含的資訊，而不會引入添加或刪除的主題事項。第1-23編號是與核苷酸序列有關的特徵關鍵，第24-43編號是與胺基酸序列有關的特徵關鍵。

編號	ST. 25 特徵關鍵<221>	ST. 26等同(equivalent)描述		
		特徵關鍵	限定詞	限定詞值
1	allele	misc_feature	allele	<223>值
2	attenuator	Regulatory ¹	regulatory_class ¹	“attenuator”
			note (若有<223>)	<223>值
3	CAAT_signal	Regulatory ¹	regulatory_class ¹	“CAAT_signal”
			note(若有<223>)	<223>值
4	conflict	misc_feature	note	“conflict” 及 <223>值
5	enhancer	Regulatory ¹	regulatory_class ¹	“enhancer”
			note(若有<223>)	<223>值
6	GC_signal	Regulatory ¹	regulatory_class	“GC_signal”
			note(若有<223>)	<223>值
7	LTR	Mobile_element ¹	rpt_type ¹	“long_terminal_r repeat
			note(若有<223>)	<223>值
8	misc_signal	Regulatory ¹	regulatory_class ¹	“other”
			note(若有<223>)	<223>值
9	mutation	variation	note	“mutation” 及 <223>值
10	old_sequence	misc_feature	note	“old_sequence” 及<223>值
11	polyA_signal	Regulatory ¹	regulatory_class ¹	“polyA_signal_se quence”
			note(若有<223>)	<223>值
12	promoter	Regulatory ¹	regulatory_class ¹	“promoter”
			note(若有<223>)	<223>值
13	RBS	Regulatory ¹	regulatory_class ¹	“ribosome_bindin g_site”
			note(若有<223>)	<223>值

¹ ST. 26可能要求特定的ST. 25功能(例如TATA_signal)被更廣泛的特徵鍵/修飾語/值(例如regulatory/regulatory_class/TATA_box)替代。在這種情況下，較窄的ST. 25功能將獲得較早申請的優先權。但是，更廣泛的ST. 26特徵鍵/修飾語(例如regulatory/regulatory_class)的全部範圍將無法獲得較早申請的優先權。

編號	ST. 25 特徵關鍵<221>	ST. 26等同(equivalent)描述		
		特徵關鍵	限定詞	限定詞值
14	repeat_unit (a) when repeat_region not used	misc_feature	note	“repeat_unit” 及<223>值
	repeat_unit (b) when repeat_region used	repeat_region	rpt_unit_range	第一個殘基.. 最後一個殘基
			note(若有<223>)	<223>值
15	satellite	repeat_region	satellite	“satellite”、 “microsatellit e”或“minisate llite”-(若可 被支持)
			note(若有<223>)	<223>值
16	scRNA	ncRNA ¹	ncRNA_class ¹	“scRNA”
			note(若有<223>)	<223>值
17	snRNA	ncRNA ¹	ncRNA_class ¹	“snRNA”
			note(若有<223>)	<223>值
18	TATA_signal	Regulatory ¹	regulatory_class ¹	“TATA_box”
			note(若有<223>)	<223>值
19	terminator	Regulatory ¹	regulatory_class ¹	“terminator”
			note(若有<223>)	<223>值
20	3' clip	misc_feature	note	“3' clip”及 <223>值
21	5' clip	misc_feature	note	“5' clip”及 <223>值
22	-10_signal	Regulatory ¹	regulatory_class ¹	“minus_10_sigh al”
			note(若有<223>)	<223>值
23	-35_signal	Regulatory ¹	regulatory_class ¹	“minus_35_sigh al”
			note(若有<223>)	<223>值

編號	ST. 25 特徵關鍵<221>	ST. 26等同(equivalent)描述		
		特徵關鍵	限定詞	限定詞值
24	NON_CONS	這個特徵涉及一序列中未知殘基數量的空隙，這在ST. 25 (第22段)及ST. 26(第37段)中都是禁止的。因此，由ST. 26第7段具體定義的每個胺基酸殘基區域，都必須作為單獨的序列，包含在序列列表中，並且分配有專屬序列識別號。為避免添加/刪除主題事項，必須對每個這樣的序列進行註釋，以表示其為包含未定義空隙(undefined gap)的大序列的一部分		
		REGION	note	Description
		說明序列與何處及何物相連接。例如，該殘基與肽N端相連接，其中該肽具有N端Gly-Gly以及未定義長度的空隙		
25	SIMILAR	REGION	note	“SIMILAR” 及<223>值(若存在)
26	THIOETH	CROSSLNK	note	“THIOETH” 及<223>值(若存在)
		進一步的位置資訊指引，請參閱ST. 26附錄I中的CROSSLNK特徵關鍵建議		
27	THIOLEST	CROSSLNK	note	“THIOLEST” 及<223>值(若存在)
		進一步的位置資訊指引，請參閱ST. 26附錄I中的CROSSLNK特徵關鍵建議		
28	VARSP LIC	於下述情境13中討論		
29	ACETYLATION	MOD_RES	note	“ACETYLATION” 及<223>值(若存在)
			note	若可能，ST. 26附錄I MOD_RES特徵關鍵建議要求的資訊(不能添加主題事項)
30	AMIDATION	MOD_RES	note	“AMIDATION” 及<223>值(若存在)
			note	若可能，ST. 26附錄I MOD_RES特徵關鍵建議要求的資訊(不能添加主題事項)
31	BLOCKED	MOD_RES	note	“BLOCKED” 及<223>值(若存在)
			note	若可能，ST. 26附錄I MOD_RES特徵關鍵建議要求的資訊(不能添加主題事項)

編號	ST. 25 特徵關鍵<221>	ST. 26等同(equivalent)描述		
		特徵關鍵	限定詞	限定詞值
32	FORMYLATION	MOD_RES	note	“FORMYLATION” 及<223>值(若存在)
			note	若可能，ST. 26附錄I MOD_RES特徵關鍵建議要求的資訊(不能添加主題事項)
33	GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	MOD_RES	note	“GAMMA-CARBOXYLGLUTAMIC ACID HYDROXYLATION” 及<223>值(若存在)
			note	若可能，ST. 26附錄I MOD_RES特徵關鍵建議要求的資訊(不能添加主題事項)
34	METHYLATION	MOD_RES	note	“METHYLATION” 及<223>值(若存在)
			note	若可能，ST. 26附錄I MOD_RES特徵關鍵建議要求的資訊(不能添加主題事項)
35	PHOSPHORYLATION	MOD_RES	note	“PHOSPHORYLATION” 及<223>值(若存在)
			note	若可能，ST. 26附錄I MOD_RES特徵關鍵建議要求的資訊(不能添加主題事項)
36	PYRROLIDONE CARBOXYLIC ACID	MOD_RES	note	“PYRROLIDONE CARBOXYLIC ACID” 及<223>值(若存在)
			note	若可能，ST. 26附錄I MOD_RES特徵關鍵建議要求的資訊(不能添加主題事項)
37	SULFATATION	MOD_RES	note	“SULFATATION” 及<223>值(若存在)
			note	若可能，ST. 26附錄I MOD_RES特徵關鍵建議要求的資訊(不能添加主題事項)
38	MYRISTATE	LIPID	note	“MYRISTATE” 及<223>值(若存在)
			note	若可能，ST. 26附錄I LIPID特徵關鍵建議要求的資訊(不能添加主題事項)

編號	ST. 25 特徵關鍵<221>	ST. 26等同(equivalent)描述		
		特徵關鍵	限定詞	限定詞值
39	PALMITATE	LIPID	note	“PALMITATE” 及<223>值(若存在)
			note	若可能,ST. 26附錄I LIPID特徵關鍵建議要求的資訊(不能添加主題事項)
40	FARNESYL	LIPID	note	“FARNESYL” 及<223>值(若存在)
			note	若可能,ST. 26附錄I LIPID特徵關鍵建議要求的資訊(不能添加主題事項)
41	GERANYL-GERANYL	LIPID	note	“GERANYL-GERANYL” 及<223>值(若存在)
			note	若可能,ST. 26附錄I LIPID特徵關鍵建議要求的資訊(不能添加主題事項)
42	GPI-ANCHOR	LIPID	note	“GPI-ANCHOR” 及<223>值(若存在)
			note	若可能,ST. 26附錄I LIPID特徵關鍵建議要求的資訊(不能添加主題事項)
43	N-ACYL DIGLYCERIDE	LIPID	note	“N-ACYL DIGLYCERIDE” 及<223>值(若存在)
			note	若可能,ST. 26附錄I LIPID特徵關鍵建議要求的資訊(不能添加主題事項)

情境9

某些特徵關鍵在ST. 25及ST. 26中都存在,並且核苷酸序列及胺基酸序列都可以使用。該等特徵關鍵在ST. 26中具有其必要限定詞,如下所示。核苷酸序列特徵關鍵“modified_base”在ST. 25及ST. 26中都存在,在情境7中有對應的相關建議。在ST. 25中沒有限定詞,但是有一個<223>自由文本欄位(free text field)。當ST. 25中<223>欄位所包含的資訊,適合作為ST. 26中必要限定詞的值時,則應包含該等資訊。當ST. 25中<223>欄位未提供或是所包含資訊,不適合作為ST. 26中必要限定詞的值時,申請人必須注意以符合ST. 26規定納入ST. 25特徵關鍵<223>所包含資訊,避免引入添加或刪除的主題事項。

核苷酸序列²

特徵關鍵	必要限定詞
5.12 - misc_binding	5.12 - misc_binding
5.30 - protein_bind	5.30 - protein_bind

建議：

- (a) 當ST. 25 <223>欄位不存在或不適當，並且說明書揭露了可能與核酸特徵位置結合的分子/複合物名稱，則應該在限定詞“bound_moiety”中表示將該等分子/複合物名稱。
- (i) 當ST. 25 <223>欄位中所包含的任何資訊，不適合在限定詞“bound_moiety”中表示，則應該在特徵關鍵的適當可選限定詞中，例如“note”，插入表示該等資訊。
- (b) 當ST. 25 <223>欄位不存在或不適當，並且說明書未有揭露可能與核酸特徵位置結合的分子/複合物名稱，則應該使用ST. 26特徵關鍵“misc_feature”代替“misc_binding”或“protein_bind”，並且使用限定詞“note”。
- (i) 當ST. 25 <223>欄位不存在，則限定詞“note”的值應為ST. 25特徵關鍵名稱；
- (ii) 當ST. 25 <223>欄位包含不適當的資訊，則限定詞“note”的值應為ST. 25特徵關鍵名稱以及<223>欄位中的資訊。

胺基酸序列³

特徵關鍵	必要限定詞
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

²下表中的數字參考(numeric references)參見ST. 26附錄I控制詞彙的特徵關鍵和限定詞編號。

³下表中的數字參考(numeric references)參見ST. 26附錄I控制詞彙的特徵關鍵和限定詞編號。

特徵關鍵	必要限定詞
7.17 - METAL	8.2 - note
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

建議：

- (a) 當ST. 25 <223>欄位不存在或不適當，並且說明書揭露必要限定詞要求的特定資訊，則應該在必要限定詞“note”中表示該等資訊。
- (i) 當ST. 25 <223>欄位包含的任何資訊，不適合在必要限定詞“note”中（請參閱特徵關鍵定義及註釋）表示，則應該在第2個限定詞“note”中，插入記載相關資訊。
- (b) 當ST. 25 <223>欄位不存在或不適當，並且說明書未有揭露必要限定詞要求的特定資訊，則應該改用ST. 26特徵關鍵“SITE”（用於一個胺基酸）或“REGION”（用於一範圍的胺基酸），並且使用限定詞“note”。
- (i) 當ST. 25 <223>欄位不存在，則限定詞“note”的值應該為ST. 25特徵關鍵名稱；
- (ii) 當ST. 25 <223>欄位包含不適當的資訊，則限定詞“note”的值應為ST. 25特徵關鍵名稱及<223>欄位中的資訊。

情境10

ST. 25中的每個特定特徵關鍵都有一個<222>欄位，其用來表示特徵位置；然而，ST. 25並不要求對大多數特徵進行位置指示，而且位置訊息的格式並沒有標準化。此外，ST. 25沒有位置運算符，例如“join”。相較下，ST. 26具有標準化的位置描述符及運算符，每個特徵都必須包含至少一個位置描述符（CDS特徵是一個特殊情況，將在情境11中討論）。

建議：

- (a) 當ST. 25序列表中有<222>欄位，則直接導入或導入到ST. 26格式中不應該會引起任何添加的主題事項；
- (b) 當ST. 25序列表沒有<222>欄位，但是在說明書中有揭露位置資訊，則直接導入或導入到ST. 26格式中不應該會引起任何添加的主題事項；
- (c) 當ST. 25序列列及說明書都未揭露位置資訊時，該特徵可能適用在整個序列

(在說明書無法支持的情況下，指定少於整個序列的位置可能構成添加/刪除的主題事項)。應注意草擬原始(ST. 25)序列表及說明書揭露，儘可能包括位置資訊，以避免未來發生問題。

情境 11

在ST. 25中，一編碼序列其編碼單一且連續的多肽序列，但是其被1個或多個非編碼序列(例如內含子)所中斷。則該編碼序列會以多個各別的CDS特徵來表示，如下所示：

```
<220>  
<221> CDS  
<222> (1)..(571)  
  
<220>  
<221> CDS  
<222> (639)..(859)
```

相較下，ST. 26可使用1個join位置運算符，將所指定位置編碼的多肽序列連接起來，以形成1個單一且連續的多肽序列。(註：ST. 25及ST. 26都要求停止密碼子必須包含在CDS特徵位置中)

建議：

(a)當ST. 25序列表或說明書明確記載，由多個分別的CDS特徵所編碼的多肽序列，連結形成一個單一連續的多肽，則由內含子所中斷的編碼序列，應該要使用join位置運算子表示為單一的CDS特徵，以避免引入任何添加的主題事項：

```
<INSDFeature_key>CDS</INSDFeature_key>  
<INSDFeature_location>join(1..571,639..859)</INSDFeature_location>
```

(b)當ST. 25序列表或說明書未有指出，由兩個分別的CDS特徵所編碼的多肽序列，連結形成單一連續的多肽，則使用join位置運算子可能構成添加的主題事項。

情境12

ST. 25規定特徵名稱必須從表格5或表格6中選擇。然而，美國法規指出這些特徵名稱僅為建議，並不要求。因此，在符合美國法規的ST. 25序列表中，可能會有1個“自定義(custom)”特徵名稱，而在ST. 26中沒有對應的特徵名稱。情況也有可能是，<221>欄位未提供特徵名稱，或是根本不存在該欄位。可以以類似方式處理上述情況。

建議：

可將ST. 25中的“自定義”特徵名稱，在ST. 26序列表中表示為一個特徵，則不引

入任何添加的主題事項，具體表示方法如下：

類型	ST. 25特徵 關鍵<221>	可能之ST. 26等同描述		
		特徵關鍵	限定詞	限定詞值
核苷酸	“自定義” 特徵關鍵	misc_feature	note	自定義特徵關鍵名稱及 <223>值(若存在)
胺基酸	“自定義” 特徵關鍵	SITE或REGION	note	自定義特徵關鍵名稱及 <223>值(若存在)

情境13

ST. 25包含1個“VARSPPLIC”特徵關鍵，定義為「由可選擇剪接產生的序列變異描述」。在ST. 26中，“VARSPPLIC”被更廣泛的特徵關鍵“VAR_SEQ”所取代，其定義為「由替代剪接、替代啟動子使用、替代起始及核糖體移碼(ribosomal frameshifting)產生的序列變異描述」。也因此，ST. 26序列表不應該只是以“VAR_SEQ”替代“VARSPPLIC”，而未進一步指明究竟是哪種替代。

建議：

在ST. 26中，應該使用特徵關鍵“VAR_SEQ”並且加上限定詞“note”，其值應該包括ST. 25較狹義範圍的解釋，例如「由替代剪接產生的序列變異描述」。任何包含在ST. 25 <223>欄位中的其他資訊，也應該包含在“note”限定詞中。

情境14

當序列來源是人造，在ST. 25中<213>欄位需要使用詞語“Artificial Sequence”。在ST. 26中，特徵關鍵“source”需要限定詞“organism”，其值必須表示為“synthetic construct”，而不是“Artificial Sequence”。

建議：

ST. 26限定詞“organism”的值必須表示為“synthetic construct”。為避免可能刪除主題事項，ST. 25 <223>欄位中的任何說明資訊，都應該包含在特徵關鍵“source”的限定詞“note”中。

情境15

當一序列的來源生物體的學名是不清楚時，ST. 25 <213>生物學名欄位必須表示“Unknown”。在ST. 26中，特徵關鍵“source”必須表示限定詞“organism”，其值必須表示為“unidentified”而非“Unknown”。

建議：

在ST. 26中，限定詞“organism”的值必須表示為“unidentified”。為避免可

能刪除主題事項，ST. 25 <223>欄位中的任何說明資訊都應該包含在特徵關鍵 “source” 的限定詞 “note” 中。

情境16

ST. 25允許在列舉胺基酸時，可選擇包含負數，並從編號為1的胺基酸旁邊開始倒數計算，此適用於成熟蛋白質之前的胺基酸，例如前序列、前蛋白質、前原蛋白質及信號序列。然而，ST. 26不允許在特徵位置中使用負數表示。

建議：

- (a) 當ST. 25序列表中<221>及對應的<222>欄位中表示包含有正數及/或負數編號的特徵時，例如 “PROPEP” 及/或 “CHAIN” 時，在 ST. 26序列表中建議使用特徵關鍵，例如 “PROPEP” 及/或 “CHAIN” 加以描述。若存在的話，可以使用限定詞 “note” 表示<223>欄位中的資訊，作為其限定詞值。
- (b) 當ST. 25序列表中<221>及對應的<222>欄位中沒有記載特徵，但是在說明書揭露有正數及/或負數編號的特徵時，在ST. 26序列表中建議使用特徵關鍵，例如 “PROPEP” 及/或 “CHAIN” 加以描述。否則，可以使用特徵關鍵 “REGION” 。若存在的話，可以使用限定詞 “note” 表示說明書中的資訊，作為其限定詞值。
- (c) 當ST. 25序列表及說明書中都未揭露有解釋正數及/或負數編號的資訊，為避免在ST. 26序列表中出現可能被刪除的主題事項，建議使用特徵關鍵 “REGION” ，其特徵位置跨越ST. 25序列表的負數區域。此外，建議使用限定詞 “note” 表示主張優先權的申請案的ST. 25序列表中，胺基酸序列是帶負數編號。

情境17

ST. 25允許在<300>至<313>欄位中提供公開資訊(publication information)。ST. 26並不允許包含此資訊。

建議：

當ST. 25序列表中<300>至<313>欄位的公開資訊，未包含在申請書主文(application body)中，建議將其插入記載在申請書正文中。

情境18

ST. 25未提供一種標準方式，來表示一核苷酸序列的CDS區域使用除標準遺傳密碼表(the standard genetic code table)外的遺傳密碼表進行翻譯。相較下，ST. 26具有一限定詞 “transl_table” ，可與特徵關鍵 “CDS” 一起使用，以表示該

區域應使用替代遺傳密碼表進行翻譯。當不使用限定詞“transl_table”，則推定是使用標準遺傳密碼表。

建議：

- (a) 當ST. 25序列表或是說明書清楚地揭露使用替代遺傳密碼表來轉譯CDS區域，則必須使用限定詞“transl_table”，並且以對應的遺傳密碼表編號作為限定詞的值。若未使用限定詞“transl_table”，表示默認使用“標準代碼”，則通常會構成添加主題事項。若未在ST. 26序列表中表示，ST. 25序列表或是說明書中的替代遺傳密碼表的資訊，則通常會構成刪除主題事項。
- (b) 當ST. 25序列表或是說明書未揭露使用替代遺傳密碼表來翻譯CDS區域，則不建議使用限定詞“transl_table”，或是僅使用限定詞值“1”，即表示標準遺傳密碼表。使用任何限定詞值不等於“1”的限定符“transl_table”，通常會構成添加及刪除主題事項。

情境19

ST. 25未提供一種標準方式來表示特徵位置，特別是，1個位於超過特定殘基或殘基範圍的位點或區域的特徵位置。例如，1個超過所揭露序列一端或兩端的核苷酸序列的CDS區域。相較下，ST. 26的特徵位置描述符，提供一標準方式來表示這樣一個位點或區域的位置，通常使用“<”或“>”符號。例如，「CDS」的特徵位置必須包括停止密碼子，即使該停止密碼子未包含在揭露序列本身中，可以透過將位置表示為例如1..>321來進行指示。

建議：

- (a) 當ST. 25序列表未明確表示特徵位置超出序列範圍，但該位置可被揭露所支持或是從序列本身明確得知（例如，未包含在序列中的CDS特徵的停止密碼子），則ST. 26序列表記載可使用“<”或“>”符號，而無需添加主題事項。
- (b) 當ST. 25序列表未明確表示特徵位置超出序列範圍，並且該位置既未被揭露所支持，也無法從序列本身明確得知。則在不引入額外主題事項的情況下，可能無法符合ST. 26的要求。在這種情況下，優先權申請及序列記載本身可能會被認為是不完整的。在這種情況下，ST. 26序列表記載的特徵位置描述，將無法享有優先權。因此，在草擬原始ST. 25序列表記載及申請揭露內容時，應當注意要包括完整的特徵資訊。

情境20

ST. 25附錄I要求，當一核苷酸序列同時包含有DNA及RNA片段，則<212>欄位的記載應為“DNA”，並且必須在<220>至<223>欄位的特徵部分，進一步描述為組合

DNA/RNA分子。然而，對此情形，進一步描述內容並不明確，造成該要求未能常規遵循。相較下，ST. 26第55段要求，必須使用特徵關鍵字“misc_feature”，進一步描述組合的DNA/RNA分子中的每個DNA及RNA片段(ST. 26使用“segment”而不是“fragment”以保持內部一致性)，並且該描述包括片段位置，並且表示該片段是DNA或是RNA的限定語“note”。

建議：

- (a) 當ST. 25序列表在一個或多個特徵中使用<221> misc_feature，描述DNA及RNA片段，並且在<222>表示適當位置以及在<223>表示哪些片段是DNA或RNA。如此一來，將該等資訊納入ST. 26格式，對於每個DNA及RNA片段使用各別misc_feature，應該不會引入任何添加主題事項；
- (b) 當ST. 25序列表在一個或多個特徵中使用<221>表示其他特徵關鍵而非misc_feature，來描述DNA及RNA片段，並且在<222>表示適當位置以及在<223>表示哪些片段是DNA或RNA。如此一來，將該等資訊納入ST. 26格式，對於每個DNA及RNA片段使用各別misc_feature並且添加限定詞“note”，該限定詞的值為原始<221>特徵關鍵，應該不會引入任何添加或刪除主題事項；
- (c) 當ST. 25序列表在<223>欄位中表示每個片段的身份(DNA或RNA)及位置，但是沒有提供<221>及<222>欄位的資訊。例如，一人工序列的解釋。如此一來，將該等資訊納入ST. 26格式，對於每個DNA及RNA片段使用1個misc_feature，應該不會引入任何添加主題事項；
- (d) 當ST. 25序列表使用<221>misc_feature及<223>說明該分子是組合 DNA/RNA 分子，但是未提供每個片段的位置資訊，並且
 - (i) 當說明書揭露了每個DNA及RNA片段的位置，則在ST. 26格式中添加每個DNA及RNA片段的misc_feature，應該不會引入任何添加主題事項；
 - (ii) 當說明書沒有揭露每個DNA及RNA片段的位置資訊，則在不引入添加主題的情況下，可能無法符合ST. 26要求。在這種情況下，優先權申請及序列表本身可能是不完整的。在這種情況下，ST. 26序列表中的任何特徵位置描述，將不會享有優先權。因此，應當謹慎草擬原始ST. 25序列表及申請說明書內容，才能涵蓋完整的特徵資訊。
- (e) 當ST. 25序列表記載分子特徵，使用<221>表示其他特徵鍵而非misc_feature，並且在<223>中表示分子是組合的DNA/RNA分子，但是沒有提供每個片段的位置資訊，並且
 - (i) 當說明書揭露了每個DNA及RNA片段的位置，則將該等資訊納入ST. 26格式，使用一個misc_feature並且添加限定詞“note”，該限定詞的值為原始<221>特徵關鍵，應該不會引入任何添加或刪除主題事項；

- (ii) 當說明書未揭露每個DNA及RNA片段的位置資訊，則在不引入添加主題的情況下，可能無法符合ST. 26要求。在這種情況下，優先權申請及序列表本身可能是不完整的。在這種情況下，ST. 26序列表中的任何特徵位置描述，將不會享有優先權。因此，應當謹慎草擬原始ST. 25序列表及申請說明書內容，才能涵蓋完整的特徵資訊。
- (f) 當ST. 25序列表在<223>(例如，人工序列的解釋)表示該分子是組合的DNA/RNA分子，但是沒有提供每個片段的任何特徵關鍵或位置資訊，並且：
 - (i) 當說明書揭露了每個DNA及RNA片段的位置，則將該等資訊合納入ST. 26格式，對於每個DNA及RNA片段使用各別misc_feature，應該不會引入任何添加主題事項；
 - (ii) 當說明書未揭露每個DNA及RNA片段的位置資訊，則在不引入添加主題的情況下，可能無法符合ST. 26要求。在這種情況下，優先權申請及序列表本身可能是不完整的。在這種情況下，ST. 26序列表中的任何特徵位置描述，將不會享有優先權。因此，應當謹慎草擬原始ST. 25序列表及申請說明書內容，才能涵蓋完整的特徵資訊。