

核苷酸及胺基酸序列表記載格式

前言

1.依據專利法施行細則第十七條第六項規定，發明專利包含一個或多個核苷酸或胺基酸序列者，說明書應包含依專利專責機關訂定之格式單獨記載之序列表。其序列表得以專利專責機關規定之電子檔為之，並得檢送相符之電子資料。核苷酸及胺基酸序列表記載格式（以下簡稱「本序列表格式」）提供專利申請案中核苷酸及胺基酸序列表之記載方式，以及申請人製作序列表之格式。序列表不但加強申請案中描述核苷酸及胺基酸序列之準確性和品質，同時也便利序列資料之檢索、序列表電子資料之交換和電腦資料庫之輸入。

定義

2.本序列表標準中採用之名詞定義如下：

- (1)「序列表」(sequence listing)：為說明書的一部分，用以詳細記載核苷酸序列及/或胺基酸序列及其他資訊，須以紙本或檢送電腦可讀形式之序列表電子資料。
- (2)序列表所指的序列：包括不少於4個胺基酸的非支鏈胺基酸序列，或不少於10個核苷酸的非支鏈核苷酸序列。所述的序列不包括支鏈序列、少於4個特別定義的核苷酸或胺基酸序列，也不包括含有列於附錄2表1-4以外之核苷酸或胺基酸序列。
- (3)「核苷酸」(nucleotides)：只包括附錄2表1中列出的符號所表示之核苷酸。附錄2表2中的符號是用於表示核苷酸的修飾形式，例如甲基化鹼基。對於核苷酸的修飾形式，不得在核苷酸序列中直接使用表2的符號表示。
- (4)「胺基酸」(amino acids)：只包括附錄2表3中列出的存在於自然界蛋白質的L型胺基酸。亦即，只要是包含一個D型胺基酸之胺基酸序列即非本序列表所定義之「胺基酸」。附錄2表4中列出的符號用於表示胺基酸的修飾形式，例如羥基化(hydroxylation)和醣基化(glycosylation)形式。對於胺基酸的修飾形式，不得在胺基酸序列中直接使用表4中的符號表示。本序列表標準中定義之「胺基酸」包括以附錄2表3中的符號配合序列表外所載的說明，例如：非正常之連結(abnormal linkages)、交聯(cross-links)（如雙硫橋(disulfide bridge)）和末端蓋(end caps)、非肽鍵(non-peptidyl bonds)等。
- (5)「序列識別號」(sequence identifier)：對應於序列表中每一個 SEQ ID NO 唯一的正整數編號。
- (6)「數字識別號」(numeric identifier)：代表特定資料項目之三位數字編號。

- (7)「中性語言詞彙」(language-neutral vocabulary)：用於序列表中之規定詞彙，其代表科學性名詞（包括附錄 2 表 1-4 中列出之符號、科學名稱、描述詞及規定之數值，以及附錄 2 表 5-6 中列出之關鍵字）。

序列表

- 3.第 2(1)點定義之序列表必須於申請時與申請案一同檢送，應列於說明書之最後。此部分應以「序列表」為標題，列於新的一頁並獨立編頁碼。除第 36 點另有規定外，序列表不需要在說明書的其他部分重複記載。
- 4.序列表如未於申請時與申請案一起檢送，而於申請後補正或修正者，該文件應以「序列表」為標題並獨立編頁碼。申請後提出的序列表中必須使用申請案所載序列之原始編號（參考第 5 點規定）。
- 5.每個序列應有獨立的序列識別號(SEQ ID NO)。該序列識別號應以 1 開始並依序增加一個整數。如果一序列識別號無對應之序列，則在其數字識別號<400>的下一行開始的位置標上 000。不論數字識別號<400>標示的是序列或 000，數字識別號<160>必須指出所有 SEQ ID NO 的總數。
- 6.在申請案的說明書、申請專利範圍或圖式中，序列可直接引述序列表中之序列識別號，並在其前加上"SEQ ID NO"。
- 7.核苷酸和胺基酸序列至少必須以下列三種形式之一表示：
 - (i)純核苷酸序列；
 - (ii)純胺基酸序列；
 - (iii)核苷酸序列和與其對應之胺基酸序列。以上述第(iii)款的形式表示之序列，其胺基酸序列部分必須另以純胺基酸序列的形式表示，並有獨立之序列識別號。

核苷酸序列

使用的符號

- 8.核苷酸序列必須以單股表示，從左到右是 5'端到 3'端。序列中不應出現用語 5'端和 3'端。
- 9.核苷酸序列中的鹼基只能以單字母代碼來表示序列特性。而且只能以附錄 2 表 1 中的小寫字母表示。
- 10.如果核苷酸序列中有附錄 2 表 2 中經修飾的鹼基，其在序列中必須以對應之未修飾鹼基或“n”表示，並以附錄 2 表 2 之符號在序列表其他特徵部分加以描述。

附錄 2 表 2 中的符號可用於說明書或序列表的特徵部分，但不能用於序列本身（參考第 32 點規定）。符號“n”等同於唯一的一個未知或經修飾的核苷酸。

使用的形式

11. 核苷酸序列中每行最多不超過 60 個鹼基，且每 10 個鹼基一組，中間以空格間隔。
12. 除序列的編碼區域外，核苷酸序列中的鹼基（包括內含子(introns)）必須每 10 個為一組。序列的非編碼區域之末端鹼基剩餘不到 10 個時，則自成一組並以空格和鄰近鹼基組群間隔。
13. 核苷酸序列中，編碼區域的鹼基必須以 3 個為一組（密碼子）。
14. 核苷酸的計算是以最左邊開始，第一個鹼基為 1，由 5'端到 3'端接續計算。並在序列每一行最後一個鹼基的右端，標示該行最後鹼基的數字。同樣的核苷酸計算方式也用於環狀的核苷酸序列，不同的是環狀的核苷酸序列可以任何一個鹼基當作第一個核苷酸。
15. 不論是由大段序列的一個或多個非連續區段，或是來自不同序列的區段所組成的核苷酸序列，均必須以單獨的序列來計數，並給予單獨的序列識別號。序列中有一個或多個缺口(gap)，應視為多個單獨序列，並各給予單獨序列識別號，而其單獨序列的數目應與序列資料中的連續序列數目相等。

胺基酸序列

使用的符號

16. 蛋白質或肽序列的胺基酸應由左到右以胺端(amino)到羧端(carboxy)的方向表示。序列中不應出現胺基和羧基基團。
17. 胺基酸應以附錄 2 表 3 中的三字母代碼表示，其第一個字母應大寫。胺基酸序列中若出現空白(blank)或終止(terminator)符號（例如“Ter”或“*”或“.”），則不可視為單個胺基酸序列，而應視為分開的單獨胺基酸序列（參考第 22 點規定）。
18. 如果經修飾之胺基酸序列是附錄 2 表 4 中列出的胺基酸之一，則在序列中必須以對應之未修飾胺基酸或“Xaa”來表示該經修飾及非常見的胺基酸，並以附錄 2 表 4 之符號在序列表的特徵部分加以描述該修飾。附錄 2 表 4 中的符號可用於說明書或序列表的特徵部分，但不能用於序列本身（參考第 32 點規定）。“Xaa”的符號等同於唯一的一個未知或經修飾的胺基酸。

使用的形式

19. 胺基酸序列中每行最多不超過 16 個胺基酸，且每個胺基酸中間以空格間隔。
20. 對應於核苷酸序列編碼區域密碼子之胺基酸的標示必須直接在其對應密碼子的下方。如果密碼子被內含子切開，則胺基酸的符號需標示在含有 2 個核苷酸的密碼子部分之下方。
21. 胺基酸的編號是以序列的第一個胺基酸為 1。當成熟蛋白質之前存在有胺基酸，例如對於前-序列(pre-)、原-序列(pro-)、前-原-序列(pre-pro-)和訊號序列(signal sequence)時，則以成熟蛋白質編號 1 的胺基酸為開始，以負數往回編號。當胺基酸以負數編號區分成熟蛋白質時，不得使用數字 0。每 5 個胺基酸需在其下標示其編號。同樣的胺基酸序列之計數方式也用於環狀的胺基酸序列，不同的是，環狀的胺基酸序列可以任何一個胺基酸當作第一個胺基酸。
22. 不論是由大段序列的一個或多個非連續區段，或是來自不同序列的區段所組成的胺基酸序列，均必須以單獨的序列來計數，並給予單獨的序列識別號。序列中有一個或多個缺口，應視為多個單獨序列，並各給予單獨序列識別號，而其單獨序列的數目應與序列資料中的連續序列數目相等。

序列表 中其他資料

23. 序列表中資料項目的順序必須依循如附錄 1 資料項目中數字識別號項目的順序。
24. 序列表中之資料項目應以附錄 1 資料項目中的數字識別號表示，而不可以其對應說明表示。提供的資料緊接列於數字識別號後，但只有資料提供之數字識別號需出現在序列表中。對該種要求有二個例外，即分別代表「特徵」和「公開資訊」之數字識別號 <220> 和 <300>。數字識別號 <220> 和 <300> 分別與數字識別號 <221> 到 <223> 和 <301> 到 <313> 的資料相關聯。當序列表中提供 <221> 到 <223> 之特徵和 <301> 到 <313> 之公開資訊時，應分別標示數字識別號 <220> 和 <300>，但緊接是空白。當數字識別號之第一個或第二個位置的數字改變時，則在數字識別號間應當空一行，而前兩位數字相同的數字識別號之間，例如 <210> 到 <213> 之間和 <220> 到 <223> 之間，則無需空一行。此外，對於任一重覆的數字識別號，其前應當空一行。

必要的資料項目

25. 序列表除了包括先前的核苷酸及/或胺基酸序列外，還必須包括附錄 1 中的以下資訊項目（必要的資料項目）：

<110>	申請人姓名或名稱
<120>	發明名稱
<160>	SEQ ID NO之數目
<210>	SEQ ID NO:x
<211>	序列長度
<212>	序列類型
<213>	生物體(Organism)
<400>	序列

申請人的姓名或名稱（數字識別號<110>）須以中文表示，外國人須附記外文原文或英文或拉丁字母。

必要資料項目中，除了數字識別號<110>、<120>和<160>以外，其他都必須在序列表中每一個序列重複出現。如果一個序列識別號沒有對應的序列，則僅有數字識別號<210>和<400>必須註明（參考第5點及附錄3範例中的SEQ ID NO:4）。

26.除了上述第25點所規定必要記載的資料項目外，當序列表和申請案同時檢送時，該申請案尚未有申請案號者，而申請人認為有必要編號以為參考者，則下列資料項目必須包括在序列表中。若申請人認為無須編有參考編號(file number)，則不需要包括此資料項目。

<130>	參考編號
-------	------

27.除了上述第25點所規定必要記載的資料項目外，當檢送序列表時申請案已有申請案號者，則下列資料項目必須包括在序列表中：

<140>	本案申請案號
<141>	本案申請日

28.除了上述第25點所規定必要記載的資料項目外，當申請案主張優先權時，則下列資料項目必須包括在序列表中：

<150>	主張優先權之基礎案申請案號
<151>	優先權日

29.如果序列中有使用“n”或“Xaa”或一經修飾鹼基或經修飾/不常見之L型胺基酸，則下列資料項目必須包括在序列表中及載明內容：

<220>	特徵
<221>	名稱/關鍵字

<222>	位置
<223>	其他資訊

30.如果生物體（數字識別號<213>）為“人工序列”或“未知”，則下列資料項目必須包括在序列表中及載明內容：

<220>	特徵
<223>	其他資訊

非必要的資料項目

31.所有附錄 1 中的資料項目，如屬第 25 到 30 點未提及者，則為非必要的資料項目。

特徵的描述

32.當描述序列特徵時（數字識別號<220>），應以附錄 2 表 5 和 6 中的“特徵關鍵字”來描述。

非關鍵詞文字(Free text)

33.非關鍵詞文字是指在數字識別號<223>（其他資訊）中，用來描述序列的特性，但不是使用第 2(7)點中的中性語言詞彙(language-neutral vocabulary)。

34.非關鍵詞文字的使用必須簡短，且為了解序列所必需的。其長度不能超過 4 行，每行最多 20 個字。其他資訊的文字描述應包含在說明書中。

35.任何非關鍵詞文字應以中文書寫。

36.當序列表中有非關鍵詞文字時，必須在說明書中重複說明。建議說明書中可以一特定章節說明非關鍵詞文字，該章節標題為“序列表之非關鍵詞文字 (Sequence Listing Free Text)”。

序列表電子資料

37.除紙本式序列表外，申請人得另檢送電腦可讀形式的序列表電子資料，電子資料部分亦可以電子申請方式傳達。

38.序列表電子資料必須存於一適當之電子檔案中，如以光碟片(CD-R，CD-RW)一適當的電子媒體檢送。必要時並提供支援以閱讀序列表中所使用之文字。

39.序列表電子資料應以所提供的軟體進行編輯且須為個人電腦作業系統可解讀。

- 40.檢送的序列表電子資料檔案可為壓縮檔案，惟須為自我解壓縮(self-extracting)格式，可在個人電腦作業系統解壓縮者。
- 41.所檢送含序列表電子資料的電子媒體必須以手寫或打字註明申請人姓名、發明名稱、參考編號、記錄資料的日期、電腦作業系統以供辨別。
- 42.如果序列表電子資料是在申請案申請後才檢送，則所檢送含序列表電子資料的電子媒體上必須標示該申請案之申請日和申請案號。
- 43.檢送補正或修正之序列表電子資料時，必須在序列表電子資料中註明。

附錄

附錄 1：數字識別號 (Numeric Identifiers)

附錄 2：核苷酸和胺基酸符號和特徵表

表 1：核苷酸表

表 2：經修飾的核苷酸表

表 3：胺基酸表

表 4：經修飾和非常見的胺基酸表

表 5：與核苷酸序列相關的特徵關鍵字表

表 6：與蛋白質序列相關的特徵關鍵字表

附錄 3：序列表範例

附錄 1：數字識別號

序列表只可使用下表中所列的數字識別號。下列資料項目的文字部分並不包括在序列表中。

序列表必須記載之資料項目必須包括在所有的序列表中之數字識別號(亦即必要的資料項目，參考序列表記載格式中的第 25 點：<110>、<120>、<160>、<210>、<211>、<212>、<213>和<400>)，及其他依序列表記載格式在特定情形下包括之數字識別號(參考序列表記載格式中的第 26、27、28、29 和 30 點：<130>、<140>、<141>、<150>和<151>，和<220>到<223>)，都以符號”M”標示。數字識別號中，非必要的資料項目(參考序列表記載格式中的第 31 點)則以符號”O”標示。

數字識別號	數字識別號說明	必要(M)或非必要(O)	註釋
<110>	申請人姓名或名稱	M	申請人的姓名或名稱須以中文表示，外國人須附記外文原文或英文或拉丁字母
<120>	發明名稱	M	
<130>	參考編號	M，限於序列表記載格式中之第 26 點之特定情形	參考序列表記載格式第 26 點
<140>	本案申請案號	M，限於序列表記載格式中之第 27 點之特定情形	參考序列表記載格式第 27 點，例如，TW 090123456
<141>	本案申請日	M，限於序列表記載格式中之第 27 點之特定情形	參考序列表記載格式第 27 點；日期必須以 CCYY-MM-DD 表示，例如：2002-01-31
<150>	主張優先權之基礎案申請案號	M，限於序列表記載格式中之第 28 點之特定情形	參考序列表記載格式第 28 點，例如：US 08/999,999

數字識別號	數字識別號說明	必要(M)或非必要(O)	註釋
<151>	優先權日	M，限於序列表記載格式中之第 28 點之特定情形	參考序列表記載格式第 28 點；日期必須以 CCYY-MM-DD 表示
<160>	SEQ ID NO 之數目	M	
<170>	軟體(Software)	O	
<210>	SEQ ID NO:x	M	為一代表所示 SEQ ID NO 之整數
<211>	序列長度	M	序列長度是以鹼基對或胺基酸數目表示
<212>	序列類型	M	SEQ ID NO:x 序列之分子類型，有去氧核糖核酸(DNA)、核糖核酸(RNA)或蛋白質(PRT)；如果核苷酸序列同時包括 DNA 和 RNA 片段，則標示應為“DNA”；另外，該組合的 DNA/RNA 分子必須在<220>到<223>中作進一步說明
<213>	生物體	M	屬、種（科學命名）或“人工序列”或“未知”
<220>	特徵	M，限於序列表記載格式中之 SEQ ID NO 29 和 30 之特定情形	保留空白；參考序列表記載格式第 29 和 30 點；SEQ ID NO:x 序列之生物特性的重點說明（依所述特徵之項目而可重複）
<221>	名稱/關鍵字	M，序列表記載格式中之第 29 點之特定情形	參考序列表記載格式第 29 點；只有附錄 2 表 5 或表 6 中所述之關鍵字可使用

數字識別號	數字識別號說明	必要(M)或非必要(O)	註釋
<222>	位置	M，限於序列表記載格式中之第 29 點之特定情形	參考序列表記載格式第 29 點； - from (始於) (特徵之第一個鹼基/胺基酸的數字) - to (終於) (特徵之最後一個鹼基/胺基酸的數字) - 鹼基對(base pairs) (鹼基對在核苷酸序列中相對位置的數字) - 胺基酸 (胺基酸在胺基酸序列中相對位置之數字) - 不論特徵是否在於該序列之互補股 (complementary strand)
<223>	其他資訊	M，限於序列表記載格式中之第 29 及 30 點之特定情形	參考序列表記載格式第 29 及 30 點；任何其他有關的資訊，可以中性語言詞彙(language neutral vocabulary)或非關鍵詞 (free text, 最好是中文) 表示。任何非關鍵詞必須在說明書中曾出現 (參考序列表記載格式第 36 點)。序列中，任何經修飾的鹼基或經修飾/非常見的 L 型胺基酸，都必須以附錄 2 表 2 和 4 中的符號表示
<300>	公開的資訊	O	保留空白；在每個有關的公開資訊均會重複
<301>	作者	O	
<302>	標題	O	公開資訊之標題
<303>	期刊	O	公開資訊之期刊名
<304>	卷號	O	公開資訊之卷號
<305>	期號	O	公開資訊之期號
<306>	頁數	O	公開資訊之頁數
<307>	日期	O	公開資訊之期刊的出版日期；日期必須以 CCYY- MM- DD 表示
<308>	資料庫的登記號 (database accession number)	O	資料庫中之登記號，包括資料庫的名稱
<309>	登錄資料庫日期 (database entry date)	O	登錄入資料庫的日期；日期必須以 CCYY- MM- DD 表示

數字識別號	數字識別號說明	必要(M)或非必要(O)	註釋
<310>	(公開資訊之) 專利文件案號 (document number)	O	專利文件案號只用於專利類型(patent-type)之敘述
<311>	(公開資訊之) 申請日期	O	專利申請案之申請日期，只用於專利類型之敘述；日期必須以 CCYY-MM-DD 表示
<312>	(公開資訊之) 公開日期	O	專利申請案之公開日期，只用於專利類型之引用；日期必須以 CCYY-MM-DD 表示
<313>	(公開資訊之) 序列識別號 x 中之相關殘基：from (始於) ...to (終於)	O	
<400>	(公開資訊之) 序列	M	序列識別號：x 必須跟在數字識別號(numeric identifier)之後，且必須出現在序列之前一行(參考附錄 3)

附錄 2：核苷酸/胺基酸符號及特徵表

表 1：核苷酸表

符號	意義	對應來源
a	a	腺嘌呤(adenine)
g	g	鳥嘌呤(guanine)
c	c	胞嘧啶(cytosine)
t	t	胸腺嘧啶(thymine)
u	u	尿嘧啶(uracil)
r	g 或 a	嘌呤(purine)
y	t/u 或 c	嘧啶(pyrimidine)
m	a 或 c	胺基(amino)
k	g 或 t/u	酮基(keto)
s	g 或 c	強的交互作用；3 個氫鍵
w	a 或 t/u	弱的交互作用；2 個氫鍵
b	g 或 c 或 t/u	非 a
d	a 或 g 或 t/u	非 c
h	a 或 c 或 t/u	非 g
v	a 或 g 或 c	非 t；非 u
n	a 或 g 或 c 或 t/u； 未知；或其他	任何(any)

表 2：經修飾之核苷酸表

符號	意義
ac4c	4-乙醯胞苷(4-acetylcytidine)
chm5u	5-(羧羥甲基)尿苷(5-(carboxyhydroxymethyl)uridine)
Cm	2'-O-甲基胞苷(2'-O-methylcytidine)
Cmm5s2u	5-羧甲基胺甲基-2-硫代尿苷 (5-carboxymethylaminomethyl-2-thiouridine)
Cmm5u	5-羧甲基胺甲基尿苷(5-carboxymethylaminomethyluridine)
d	二氫尿苷(dihydrouridine)
fm	2'-O-甲基假尿苷(2'-O-methylpseudouridine)
gal q	β ,D-半乳糖 Q 核苷(beta, D-galactosylqueosine)
gm	2'-O-甲基鳥苷(2'-O-methylguanosine)
i	肌苷(inosine)
i6a	N6-異戊烯基腺苷(N6-isopentenyladenosine)
mla	1-甲基腺苷(1-methyladenosine)
mlf	1-甲基假尿苷(1-methylpseudouridine)
mlg	1-甲基鳥苷(1-methylguanosine)
mli	1-甲基肌苷(1-methylinosine)
m22g	2,2-二甲基鳥苷(2,2-dimethylguanosine)
m2a	2-甲基腺苷(2-methyladenosine)
m2g	2-甲基鳥苷(2-methylguanosine)
m3c	3-甲基胞苷(3-methylcytidine)
m5c	5-甲基胞苷(5-methylcytidine)
m6a	N6-甲基腺苷(N6-methyladenosine)
m7g	7-甲基鳥苷(7-methylguanosine)
mam5u	5-甲基胺基甲基尿苷(5-methylaminomethyluridine)
mam5s2u	5-甲氧基胺基甲基-2-硫代尿苷(5-methoxyaminomethyl-2-thiouridine)
man q	β , -D-甘露糖 Q 核苷(beta, D-mannosylqueosine)
mcm5s2u	5-甲氧基羰基甲基-2-硫代尿苷(5-methoxycarbonylmethyl-2-thiouridine)
mcm5u	5-甲氧基羰基甲基尿苷 (5-methoxycarbonylmethyluridine)
mo5u	5-甲氧基尿苷(5-methoxyuridine)
ms2i6a	2-甲硫基-N6-異戊烯基腺苷 (2-methylthio-N6-isopentenyladenosine)
ms2t6a	N-((9- β -D-呋喃核糖基-2-甲硫基嘌呤-6-基)胺甲醯基)蘇胺酸 (N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine)
mt6a	N-((9- β -呋喃核糖基嘌呤-6-基)N-甲基胺甲醯基)蘇胺酸 (N-((9-beta-D-ribofuranosylpurine-6-yl)N-methylcarbamoyl)threonine)

符號	意義
mv	尿苷-5-氧化乙酸-甲酯(Uridine-5-oxyacetic acid-methylester)
o5u	尿苷-5-氧化乙酸(uridine-5-oxyacetic acid)
osyw	Wybutoxosine
p	假尿苷(pseudouridine)
q	Q 核苷(queuosine)
s2c	2-硫代胞苷(2-thiocytidine)
s2t	5-甲基-2-硫代尿苷(5-methyl-2-thiouridine)
s2u	2-硫代尿苷(2-thiouridine)
s4u	4-硫代尿苷(4-thiouridine)
t	5-甲基尿苷(5-methyluridine)
t6a	N-((9-β-呋喃核糖醣嘌呤-6-基)-胺甲醯基)蘇胺酸 (N-((9-beta-D-ribofuranosylpurine-6-yl)-carbamoyl)threonine)
tm	2'-O-甲基-5-甲基尿苷(2'-O-methyl-5-methyluridine)
um	2'-O-甲基尿苷(2'-O-methyluridine)
yw	wybutosine
x	3-(3-胺基-3-羧基-丙基)尿苷,(acp3)u (3-(3-amino-3-carboxy-propyl)uridine,(acp3)u)

表 3：胺基酸表

符號	意義
A 或 Ala	丙胺酸(alanine)
C 或 Cys	半胱胺酸(cysteine)
D 或 Asp	天冬胺酸(aspartic acid)
E 或 Glu	麩胺酸(glutamic acid)
F 或 Phe	苯丙胺酸(phenylalanine)
G 或 Gly	甘胺酸(glycine)
H 或 His	組胺酸(histidine)
I 或 Ile	異白胺酸(isoleucine)
K 或 Lys	離胺酸(lysine)
L 或 Leu	白胺酸(leucine)
M 或 Met	甲硫胺酸(methionine)
N 或 Asn	天冬醯胺酸(asparagine)
P 或 Pro	脯胺酸(proline)
Q 或 Gln	麩醯胺酸(glutamine)
R 或 Arg	精胺酸(arginine)
S 或 Ser	絲胺酸(serine)
T 或 Thr	蘇胺酸(threonine)
V 或 Val	纈胺酸(valine)
W 或 Trp	色胺酸(tryptophan)
Y 或 Tyr	酪胺酸(tyrosine)
B 或 Asx	天冬胺酸或天冬醯胺酸(Asp 或 Asn)
Z 或 Glx	麩胺酸或麩醯胺酸(Glu 或 gln)
X 或 Xaa	未知或其他

表 4：經修飾和非常見之胺基酸表

符號	意義
Aad	2-胺基己二酸(2-aminoadipic acid)
bAad	3-胺基己二酸(3-aminoadipic acid)
bAla	β -丙胺酸， β -胺基丙酸(beta-alanine, beta-aminopropionic acid)
Abu	2-胺基丁酸(2-aminobutyric acid)
4Abu	4-胺基丁酸，哌啶酸(4-aminobutyric acid, piperidinic acid)
Acp	6-胺基己酸(6-aminocaproic acid)
Ahe	2-胺基庚酸(2-aminoheptanoic acid)
Aib	2-胺基異丁酸(2-aminoisobutyric acid)
bAib	3-胺基異丁酸(3-aminoisobutyric acid)
Apm	2-胺基庚二酸(2-aminopimelic acid)
Dbu	2,4 二胺基丁酸(2,4 diaminobutyric acid)
Des	賴胺素(desmosine)
Dpm	2,2'-二胺基庚二酸(2,2'-diaminopimelic acid)
Dpr	2,3-二胺基丙酸(2,3-diaminopropionic acid)
EtGly	N-乙基甘胺酸(N-ethylglycine)
EtAsn	N-乙基天冬醯胺酸(N-ethylasparagine)
Hyl	羥基離胺酸(hydroxylysine)
aHyl	別-羥基離胺酸(allo-hydroxylysine)
3Hyp	3-羥基脯胺酸(3-hydroxyproline)
4Hyp	4-羥基脯胺酸(4-hydroxyproline)
Ide	異賴胺素(isodesmosine)
alle	別-異白胺酸(allo-isoleucine)
MeGly	N-甲基甘胺酸，肌胺酸(N-methylglycine, sarcosine)
MeIle	N-甲基異白胺酸(N-methylisoleucine)
MeLys	6-N-甲基離胺酸(6-N-methyllysine)
MeVal	N-甲基纈胺酸(N-methylvaline)
Nva	降纈胺酸(norvaline)
Nle	降白胺酸(norleucine)
Orn	鳥胺酸(ornithine)

表 5、與核苷酸序列相關的特徵關鍵字表

關鍵字	說明
allele	相關的個體或菌株含有相同基因的穩定的其他形式，該形式不同於此一位置的現有的序列（和或許其他序列）
attenuator	(1)調節轉錄終止的 DNA 區域,其控制一些細菌操縱子(operon)的表現；(2)位於啟動子和第一個結構基因之間，能引起轉錄部分終止的序列區段
C_region	免疫球蛋白輕鏈和重鏈的恒定區，和 T 細胞受體 α 、 β 、 γ 鏈的恒定區；依特定的鏈而定，可包括一個或多個表現子(exon)
CAAT_signal	CAAT 區;位於真核生物轉錄起始位置上游約 75bp 的一段保留性序列，其可能與 RNA 聚合酶結合有關；一致序列=GG(C 或 T)CAATCT
CDS	編碼序列；對應於蛋白質中的胺基酸序列的核苷酸序列（包括終止密碼子）；特徵包括胺基酸概念上的轉譯
conflict	對於「相同」之序列在此一點或此一區域上之定序有差異
D-loop	置換環(displacement loop)；粒線體 DNA 內的一段區域，其中一段 DNA 鏈與一段短的 RNA 鏈配對,取代了原始配對的 DNA 鏈；也可用於描述在 RecA 蛋白質催化反應中，雙鏈 DNA 中的一條被一單鏈入侵子(invader)替代的區域
D-segment	免疫球蛋白重鏈,和 T-細胞受體 β 鏈的多變區(diversity segment)
enhancer	一段順式作用(cis-acting)序列，其可以增強(某些)真核生物啟動子的作用，並能在啟動子的任一方向和任何位置(上游或下游)起作用
exon	編碼剪接後的 mRNA 序列的基因體區域；可能含有 5'UTR，所有 CDS 和 3'UTR
GC_signal	GC 區；一段保留性之富含 GC 的區域，其位於真核生物轉錄單位起始點的上游，可以有多套(copy)或以任一方向存在；一致序列=GGGCGG
gene	在生物學上有意義的序列定義為基因，並已經命名
iDNA	介入(intervening)DNA；該 DNA 經由幾種重組中的任何一種方式將之去除
intron	被轉錄的 DNA 區段，但經由同時剪接位於其兩側的序列（表現子）即可從轉錄子內部將其去除
J_segment	免疫球蛋白輕鏈和重鏈的連接區段(joining segment),和 T 細胞受體 α 、 β 、 γ 鏈的連接區段
LTR	一段長的末端重複的序列，其在所定義序列的兩端直接重複，常見於逆轉錄病毒中

關鍵字	說明
mat_peptide	成熟肽或蛋白質的編碼序列；為轉譯後修飾之成熟或最終肽或蛋白質產物的編碼序列…；該序列不包括終止密碼子（與相應的 CDS 不同）
Misc_binding	在核酸中以共價或非共價的方式與另一部分結合的位置，且其不能用任何其他結合關鍵字(primer_bind 或 protein_bind)表述
Misc_difference	與記載中已存在的有所不同的特徵序列，且其不能用任何其他不同關鍵字(conflict, unsure, old_sequence, mutation, variation, allele 或 modified_base)表述
Misc_feature	具有生物學意義的區域，且其不能用任何其他其他的特徵關鍵字表述；一個新的或少見的特徵
Misc_recomb	任何一般性、位置專一或複製重組的位置，該位置中有不能用其他重組關鍵字(iDNA 和 virion)或來源關鍵字的描述詞 (/transposon, /proviral)表述的雙螺旋 DNA 的斷裂和重新組合
Misc_RNA	不能用其他 RNA 關鍵字(prim_transcript, precursor_RNA, mRNA, 5'clip, 3'clip, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, rRNA, tRNA, scRNA 和 snRNA)定義的任何轉錄子或 RNA 產物
Misc_signal	含有信號控制或改變基因功能或表達之任何區域，且其不能用其他信號關鍵字(promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator, 和 rep_origin)表述
Misc_structure	不能用其他結構關鍵字(stem_loop 和 D-loop)表述的任何二級或三級結構或構形
modified_base	所示的核苷酸是經修飾的核苷酸，並應由所示的分子（在 mod_base 描述詞中示出）所取代
mRNA	信使(messenger)RNA；包括 5'非轉譯區(5'UTR)，編碼序列(CDS，表現子)和 3'非轉譯區(3'UTR)
mutation	相關品系在此位置的序列有突然及可遺傳的改變
N_region	在重排的免疫球蛋白區段之間插入的額外核苷酸
old_sequence	現存的序列修改了此位置以前版本的序列
PolyA_signal	內切核酸酶裂解 RNA 轉錄子所必需的識別區域，其後接著是聚腺苷酸化；一致序列=AATAAA
PolyA_site	RNA 轉錄子上的位置，通過轉錄後聚腺苷酸化能使該位置被加上腺嘌呤殘基
Precursor_RNA	尚未成熟的 RNA 產物的任何 RNA 種類；可包括 5'剪切區(5'clip)，5'非轉譯區(5'UTR)，編碼序列(CDS，表現子)，介入序列(不表現子(intron))，3'非轉譯區(3'UTR)，和 3'剪切區(3'clip)

關鍵字	說明
Prim_transcript	初級（最初的，未加工的）轉錄子；包括 5'剪切區(5'clip)，5'非轉譯區(5'UTR)，編碼序列(CDS，表現子)，介入序列（不表現子），3'非轉譯區(3'UTR)和 3'剪切區(3'clip)
primer_bind	供啟動複製，轉錄或逆轉錄用的非共價的引子結合位置；其包括例如合成的 PCR 引子的結合位置
Promoter	參與 RNA 聚合酶的結合以啟動轉錄的 DNA 分子區域
protein_bind	核酸上非共價的蛋白質結合位置
RBS	核糖體結合位置
Repeat_region	含有重複單位的基因體區域
Repeat_unit	單個重複單位
rep_origin	複製起點；複製核酸以得到兩個相同套的起始位置
rRNA	成熟的核糖體 RNA；將胺基酸裝配成蛋白質的核糖核蛋白顆粒（核糖體）中的 RNA 成份
S_region	免疫球蛋白重鏈的開關區(switch region)；其參與重鏈 DNA 的重排，導致來自相同 B 細胞之不同免疫球蛋白類的表現
Satellite	許多短的基本重複單位的串聯重複(tandem repeats)（相同或相關的）；多數因具有與一般基因體不同的鹼基組成或其他性質，而使其能自大部分（主要帶區，main band）的基因體 DNA 中分離
scRNA	小的細胞質 RNA；任一存在於真核生物的細胞質（有時在細胞核）中之小細胞質 RNA 分子
sig_peptide	信號肽編碼序列；分泌蛋白 N 端結構域的編碼序列；此結構域參與新生多肽與膜的結合；前導序列
snRNA	小的核 RNA；任一局限於核中之多種小的 RNA 種類；部分 snRNA 參與剪接或其它 RNA 加工反應
source	指明序列中特定範圍的生物來源；此關鍵字是必須填寫的；每一項至少要有一個跨越整個序列的單一 source 關鍵字；每個序列可允許有一個以上的 source 關鍵字
stem_loop	髮夾結構；一段雙螺旋區域，由單鏈 RNA 或 DNA 的相鄰（反向）互補序列之間的鹼基配對所形成
STS	序列標記點(sequence tagged site)：短的，單套之 DNA 序列，可作為基因圖上之標記並能以 PCR 檢測；藉由測定一系列 STS 的順序即可決定出基因體區域的圖譜
TATA-signal	TATA 區；Goldberg—Hogness 區；一段保留性的富含 AT 的七聚體(septamer)，在每個真核生物 RNA 聚合酶 II 轉錄單位起點前約 25bp 處，其可能涉及使該酶定位以正確地起始；一致序列 =TATA(A 或 T)A(A 或 T)

關鍵字	說明
terminator	位於轉錄區的末端或者與啟動子區域相鄰的 DNA 序列，該序列可導致 RNA 聚合酶終止轉錄；也可以是阻抑蛋白(repressor protein)的結合位
transit_peptide	轉運肽編碼序列；核編碼的胞器蛋白質 N 末端區域的編碼序列；此區域參與將蛋白質轉譯後運送到胞器中
tRNA	成熟的轉移 RNA，小的 RNA 分子（75—85 個鹼基長），其參與將核酸序列轉譯成胺基酸序列
unsure	作者不能確定此區域的準確序列
V_region	免疫球蛋白之輕鏈和重鏈的可變區，和 T 細胞受體 α 、 β 、 γ 鏈的可變區；編碼可變的胺基末端部分；可由 V_segment，D_segment，N_region 和 J_segment 組成
V_segment	免疫球蛋白之輕鏈和重鏈，和 T 細胞受體 α 、 β 和 γ 鏈的可變區段；編碼大多數可變區(V_region)和前導肽的最後幾個胺基酸
variation	含有來自相同基因的穩定突變的相關品系（例如 RFLP，多形性等），其與其他序列在此位置（可能還有其他位置）不同
3'clip	在加工過程中被切下的前驅轉錄子 3'的最末端
3'UTP	不被轉譯成蛋白質的成熟轉錄子的 3'末端區域(終止密碼子之後)
5'clip	在加工過程中被切下的前驅轉錄子 5'的最末端
5'UTP	不被轉譯成蛋白質的成熟轉錄子的 5'末端區域(起始密碼子之前)
-10_signal	Pribnow 區；細菌轉錄單位起點上游約 10bp 處的保留性區域，其可能參與結合 RNA 聚合酶；一致序列=TAtAaT
-35_signal	細菌轉錄單位起點上游約 35bp 處的保留性六聚體(hexamer)；一致序列=TTGACa[]或 TGTTGACA[]

表 6、與蛋白質序列相關的特徵關鍵字表

關鍵字	說明
CONFLICT	不同的論文報導不同的序列
VARIANT	作者報導存在序列變異體
VARSLIC	由可選擇的剪接(alternative splicing)產生序列變異體的描述
MUTAGEN	經實驗操作已改變的位置
MOD_RES	殘基的轉譯後修飾(post-translational modification)
ACETYLATION	N 末端或其他
AMIDATION	通常位於成熟活性胜肽的 C 末端
BLOCKED	不能被測定的 N 或 C 末端封閉基團(blocking group)
FORMYLATION	在 N 末端甲硫胺酸的
GAMMA – CARBOXY GLUTAMIC ACID HYDROXYLATION	在天冬醯胺酸，天冬胺酸，脯胺酸或離胺酸的
METHYLATION	通常在離胺酸或精胺酸的
PHOSPHORYLATION	在絲胺酸，蘇胺酸，酪胺酸，天冬胺酸或組胺酸的
PYRROLIDONE CARBOXYLIC ACID	已形成內部環內醯胺(internal cyclic lactam)的 N 末端麩胺酸
SULFATATION	通常在酪胺酸的
LIPID	脂質組成成分的共價結合
MYRISTATE	經由醯胺鍵與成熟型蛋白質的 N 末端甘胺酸殘基或內部的離胺酸殘基結合的豆蔻酸基團(myristate group)
PALMITATE	經由硫醚鍵(thioether bond)與半胱胺酸殘基或經由酯鍵與絲胺酸或酥胺酸殘基結合的棕櫚酸基團(palmitate group)
FARNESYL	經由硫醚鍵(thioether bond)與半胱胺酸殘基結合的法尼基團(farnesyl group)
GERANYL-GERANYL	經由硫醚鍵(thioether bond)與半胱胺酸殘基結合的香葉基-香葉基基團(geranyl-geranyl group)
GPI-ANCHOR	糖基-磷脂醯肌醇(GPI)基團與成熟型蛋白質 C 末端殘基的 α -羧基相連
N-ACYL DIGLYCERIDE	成熟型原核生物脂蛋白的 N 末端半胱胺酸，該脂蛋白具有醯胺鍵聯(amide-linked)的脂肪酸和經由酯鍵連接了兩個脂肪酸的甘油基

關鍵字	說明
DISULFID	雙硫鍵；“FROM”和“TO”終點表示經由一個鏈內雙硫鍵連接的兩個殘基；如果“FROM”和“TO”終點是完全相同的，則雙硫鍵是鏈間鍵(interchain)，且說明書中指示出交互連結(cross-link)的性質
THIOLEST	硫醇酯鍵(thiolester bond)；“FROM”和“TO”終點表示經由硫醇酯鍵連接的兩個殘基
THIOETH	硫醚鍵(thioether bond)；“FROM”和“TO”終點表示經由硫醚鍵連接的兩個殘基
CARBOHYD	醣基化位置；其碳水化合物之性質於說明書中敘述
METAL	金屬離子的結合位置；該金屬之性質於說明書中敘述
BINDING	任何化學基團（輔酶，輔基等等）的結合位置；基團的化學性質於說明書中敘述
SIGNAL	信號序列的範圍（前肽）(prepeptide)
TRANSIT	轉運肽(transit peptide)的範圍(線粒體，葉綠體或微小體)
PROPEP	原肽(propeptide)的範圍
CHAIN	成熟蛋白質中多肽鏈的範圍
PEPTIDE	被釋放的活性肽的範圍
DOMAIN	序列中所述區域的範圍；該區域的性質於說明書中敘述
CA_BIND	鈣結合區域的範圍
DNA_BIND	DNA 結合區域的範圍
NP_BIND	核苷酸磷酸酯(nucleotide phosphate)結合區域的範圍；核苷酸磷酸酯的性質於說明書中敘述
TRANSMEM	跨膜(transmembrane)區域的範圍
ZN_FING	鋅指(zinc finger)區域的範圍
SIMILAR	與另一蛋白質序列具有相似性的區域；與該序列有關的精確資料於說明書中敘述
REPEAT	內部序列重複的範圍
HELIX	二級結構：螺旋，例如 α 螺旋，3 (10) 螺旋，或 Pi 螺旋
STRAND	二級結構： β 鏈，例如氫鍵連接的 β 鏈，或分離的 β 橋中的殘基
TURN	二級結構的轉折(turn)，例如 H 鍵的轉折（3 轉折，4 轉折或 5 轉折）
ACT_SITE	涉及酶活性的胺基酸
SITE	序列中任何其他所提及的位置
INIT_MET	以起始密碼子甲硫胺酸為開始的已知序列

關鍵字	說明
NON_TER	在序列極末端的殘基，但其不是末端殘基；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，表示此位置不是完整分子的 C-末端
NON_CONS	非連續殘基(consecutive residue)；表示序列中的兩個殘基不是連串的，在它們之間有很多未定序的殘基
UNSURE	序列的不確定性；用於表述一段作者不能確定序列排列的序列區域

附錄 3：序列表範例

<110> ××基因開發有限公司

<120> 序列表

<130> 01-00001

<140> TW 087012345

<141> 1998-10-14

<150> US 08/999,999

<151> 1997-10-15

<160> 4

<170> PatentIn Version 2.0

<210> 1

<211> 389

<212> DNA

<213> 草履蟲種 (Paramecium sp.)

<220>

<221> CDS

<222> (279)...(389)

<300>

<301> Doe, Richard

<302> Isolation and Characterization of a Gene Encoding a Protease from
Paramecium sp.

<303> Journal of Genes

<304> 1

<305> 4

<306> 1-7

<307> 1988-06-31

<308> 123456

<309> 1988-06-31

<400> 1

agctgtagtc attcctgtgt cctcttctct ctgggcttct caccctgcta atcagatctc 60

agggagagtg tcttgaccn cctctgcctt tgcagcttcn caggcaggca gncaggcagc 120

tgatgtggca attgctggca gtgccacagg ctttccagcc aggcttaggg tgggttccgc 180

cgcgggcgcg cggccctct cgcgctcctc tcgcgctct ctctcgctct cctctcgctc 240
 ggacctgatt aggtgagcag gaggaggggg cagttagc atg gtt tca atg ttc agc 296
 Met Val Ser Met Phe Ser
 1 5
 ttg tct ttc aaa tgg cct gga ttt tgt ttg ttt gtt tgt ttg ttc caa 344
 Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu Phe Val Cys Leu Phe Gln
 10 15 20
 tgt ccc aaa gtc ctc ccc tgt cac tca tca ctg cag ccg aat ctt 389
 Cys Pro Lys Val Leu Pro Cys His Ser Ser Leu Gln Pro Asn Leu
 25 30 35

 <210> 2
 <211> 37
 <212> PRT
 <213> 草履蟲種 (Paramecium sp.)

 <400> 2
 Met Val Ser Met Phe Ser Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu
 1 5 10 15
 Phe Val Cys Leu Phe Gln Cys Pro Lys Val Leu Pro Cys His Ser Ser
 20 25 30
 Leu Gln Pro Asn Leu
 35

 <210> 3
 <211> 11
 <212> PRT
 <213> 人工序列

 <220>
 <223> 根據大小和極性而設計的肽，其係用作 XYZ 蛋白的 α 和 β 鏈之間的連接子 (linker)。

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

 <210> 4

 <400> 4
 000