ZC

中华人民共和国知识产权行业标准

ZC 0003—2001

核苷酸和/或氨基酸序列表

和序列表电子文件标准

2001-11-01 发布　　2001-11-01 实施

中华人民共和国国家知识产权局 发布

核苷酸和/或氨基酸序列表和序列表电子文件标准

1 总则

根据专利法实施细则第 18 条第 4 款的规定，包含一个或多个核苷酸或者氨基酸序列的发明专利申请，说明书中应当包括符合国家知识产权局专利局规定的序列表，并按照国家知识产权局专利局的规定提交含有该序列表的计算机可读形式的副本。

为了使提交的纸件形式的核苷酸和/或氨基酸序列表及计算机可读形式的含有该序列表的电子文件规范化，以利于申请人提交；也为了使序列表电子文件可以快捷地输入国家知识产权局专利局的计算机数据库，并与其它的序列检索数据库交换数据，以利于公众检索；同时也利于专利局审查员加快审查，更好地为申请人服务；特制定本标准。

2 适用范围

本标准适用于所有向国家知识产权局专利局提交的包含核苷酸和/或氨基酸序列的发明专利申请，具体地说，适用于该申请提交的纸件形式的核苷酸和/或氨基酸序列表，以及含有核苷酸和/或氨基酸序列表的计算机可读形式的序列表电子文件。

3 术语和定义

在本标准中，采用下面术语和定义：

（1）序列表：是指以纸件形式提交的专利申请说明书的一部分，它公开了核苷酸和/或氨基酸序列的详细内容和其它有用信息。序列表中的序列是不少于 10 个核苷酸的非支链核苷酸序列，或者是不少于 4 个氨基酸的非支链氨基酸序列。所述的序列不包括支链序列；不包括具有少于 4 个特别定义的核苷酸或氨基酸的序列；也不包括含有列于附录 1 之表 1 - 4 以外的核苷酸或氨基酸的序列。

（2）序列表电子文件：是指包含核苷酸和/或氨基酸序列表的计算机可读形式的纯文本文件。

（3）核苷酸：只包括附录 1 之表 1 中列出的符号所表示的核苷酸。附录 1 之表 2 中列出的符号用于表述核苷酸的修饰形式，例如甲基化碱基。对于核苷酸的修饰形式，不得在核苷酸序列中直接使用表 2 中的符号表示，其具体的表述方式见本标准 4.4.7 节（1）和 4.4.5 节的内容。

（4）氨基酸：只包括列于附录 1 之表 3 中的存在于天然蛋白质中的 L - 氨基酸，不包括 D - 氨基酸。附录 1 之表 4 中列出的符号用于表述氨基酸的修饰形式，例如羟基化或糖基化形式。对于氨基酸的修饰形式，不得在氨基酸序列中直接使用表 4 中的符号表示，其具体的表述方式见本标准 4.4.7 节（2）和 4.4.5 节的内容。

（5）序列标识符：对应于序列表中每个序列的序列标识号的唯一的正整数。

（6）数字标识符：由尖括号<>括起来的代表特定内容数据项的三位数字。

4 序列表和序列表电子文件中的数字标识符、内容及其格式：

在核苷酸和/或氨基酸序列表和序列表电子文件中，应当有本标准中指出的数字标识符，在数字标识符之后（即在其之右，必要时还包括在其下面的若干行）是相应的具体内容，它们应当符合本标准规定的格式。附录 2 给出了一个说明数字标识符、其后内容及格式的序列表样例。

序列表和序列表电子文件中包括的数字标识符及相应内容和格式具体如下：

4.1、序列表和序列表电子文件中的著录项目：

下面 4.1.1 - 4.1.7 节中的内容应当与专利申请请求书中的相应内容一致。

4.1.1、申请人的姓名或名称：其数字标识符为<110>。

在数字标识符<110>之后，是该专利申请的所有申请人的姓名或名称。

外国申请人还应当在中文译名之后注明英文姓名或名称，并将其用圆括号括起来。

4.1.2、发明名称：其数字标识符为<120>

在数字标识符<120>之后，是该专利申请的发明名称。

4.1.3、案卷参考号：其数字标识符为<130>

在数字标识符<130>之后，是该专利申请的案卷参考号；没有案卷参考号的，无需包括此项内容。

4.1.4、专利申请号：其数字标识符为<140>

对于首次提交的专利申请，无需包括此项内容；当补交或提交修改时，在数字标识符<140>之后，是该专利申请的申请号。

4.1.5、专利申请日：其数字标识符为<141>

对于首次提交的专利申请，无需包括此项内容；当补交或提交修改时，在数字标识符<141>之后，是该专利申请的申请日，其格式为：YYYY - MM - DD，例如 2002 - 01 - 18。

4.1.6、优先权号：其数字标识符为<150>

没有优先权的专利申请，无需包括此项内容；如果有优先权的话，那么在数字标识符<150>之后，是该专利申请的优先权号，其格式为：世界知识产权组织（WIPO）标准 3（ST 3）的国家、地区和政府间组织代码 + 优先权号，例如，CN93112388.7。

4.1.7、优先权日：其数字标识符为<151>

没有优先权的专利申请，无需包括此项内容；如果有优先权的话，那么在数字标识符<151>之后，是该专利申请的优先权日，其格式为：YYYY - MM - DD，例如 2001 - 09 - 20。

4.2、序列表电子文件的软件版本信息：其数字标识符为<170>

当使用国家知识产权局专利局或其它专利组织（例如欧洲专利局）提供的软件形成核苷酸和/或氨基酸序列表电子文件时，在数字标识符<170>之后，是该软件的名称与版本号；未使用所述软件时，可以不包含此项内容。

4.3、序列表中序列的个数：其数字标识符为<160>。

在数字标识符<160>之后，是序列的总数，即与数值最大的序列标识符相对应的正整数。

4.4、序列中的各项内容：

4.4.1、序列标识符：其数字标识符为<210>。

在序列表中，每个序列应当有独立的、唯一的序列标识符，它应当从 1 开始并逐一增加。序列标识符表示每个序列在序列表中的序号。

在数字标识符<210>之后，是与一个序列相对应的序列标识符。

在一个序列标识符之后到下一个序列标识符之前是该序列的各项具体内容，即下面 4.4.2 - 4.4.7 节的内容。

在序列表中有多个序列的情况下，应当按照序列标识符数值从小到大的次序逐一填写每个序列的各项内容。

4.4.2、序列的长度：其数字标识符为<211>。

在数字标识符<211>之后，是以碱基或氨基酸的数目表示的该序列的长度。

4.4.3、序列的类型：其数字标识符为<212>。

在数字标识符<212>之后，应当指出该序列的分子类型，有 DNA、RNA 或 PRT 三种类型。如果核苷酸序列含有 DNA 和 RNA 片段的话，那么其类型

应该是 DNA；另外，对于 DNA/RNA 的结合分子，应该在该序列的特征部分（数字标识符<220> - <223>）进一步表述。

4.4.4、生物体：其数字标识符为<213>。

在数字标识符<213>之后，应当用中文和拉丁文（拉丁文应当放在中文之后并用圆括号括起来，例如，草履虫种(Paramecium sp.)）注明该序列来源的生物名称，即科学命名的生物属种；或者是“人工序列”或“未知”。

4.4.5、序列中特征部分的内容：数字标识符<220> - <223>

本节涉及到序列中与特征相关的内容的表述。

在核苷酸序列（数字标识符< 400> ）中含有“n”或修饰的碱基的情况下（参见本标准 4.4.7 节(1)的内容）,或者在氨基酸序列(数字标识符< 400> ）中含有“Xaa”或修饰的氨基酸或不常用的 L - 氨基酸的情况下(参见本标准 4.4.7 节（ 2 ）的内容），必须包括下面（ 1 ） -（ 4 ）项的内容。

在生物体（数字标识符< 213> ）是“人工序列”或“未知”的情况下，必须包括下面（ 1 ）和（ 4 ）项的内容。

在一个序列中有多个特征的情况下 ,应当按照这些特征在序列中出现的先后次序逐一地表述每个特征。

序列中特征部分的具体内容和数字标识符如下：

（ 1 ）特征：其数字标识符为<220>。

在数字标识符<220>之后，应当是空白。

（2）名称/关键词：其数字标识符为<221>。

在数字标识符<221>之后，是特征名称或关键词。使用关键词表述特征时，只能使用附录 1 之表 5 或表 6 中列出的关键词来表述。

（3）位置：其数字标识符为<222>。

在数字标识符<222>之后，应当标明特征的位置，标注的方式为：从特征中的第一个碱基或氨基酸的编号到特征的最后一个碱基或氨基酸的编号，编号圆括号括起来，两个编号中间是“...”，例如：(279)...(389)；当序列中使用了多个“n”或“Xaa”时，应当标明它们的所有位置，例如：(80,100,112)。参见附录 2 的序列表样例。

（4）其它信息：其数字标识符为<223>。

在数字标识符<223>之后，应当表述序列中与特征有关的其它相关信息。在表述修饰的碱基或修饰的氨基酸时，应该用附录 1 之表 2 或表 4 中给出的符号来表述。

4.4.6、出版公开信息：数字标识符<300> - <312>

出版公开信息是非强制性的内容，在序列表和序列表电子文件中，可以包含也可以不包含这些内容。

（1）公开出版信息：其数字标识符为<300>

在数字标识符<300>之后，应当是空白。

（2）作者：其数字标识符为<301>

在数字标识符<301>之后，是该文献作者的姓名。

（3）题目：其数字标识符为<302>

在数字标识符<302>之后，是出版物中该文献的题目。

（4）杂志名称：其数字标识符为<303>

在数字标识符<303>之后，是公开出版物的杂志名称。

（5）公开出版物的卷号：其数字标识符为<304>

在数字标识符<304>之后，是公开出版物的卷号。

（6）公开出版物的出版号：其数字标识符为<305>

在数字标识符<305>之后，是公开出版物的出版号。

（7）页码：其数字标识符为<306>

在数字标识符<306>之后，是该文献的起始－终止页码。

（8）出版日期：其数字标识符为<307>

在数字标识符<307>之后，是该公开出版物的出版日期，其格式为：YYYY－MM－DD，例如 1999?9?0。

（9）公开出版物的数据库登记号：其数字标识符为<308>

如果该文献被收入某个数据库的话，那么在数字标识符<308>之后，是该文献在该数据库中的登记号。

（10）录入数据库的日期：其数字标识符为<309>

如果该文献被收入某个数据库的话，那么在数字标识符<309>之后，是该文献录入该数据库的日期，其格式为：YYYY - MM - DD，例如 1999?9?0。

（11）专利公开号：其数字标识符为<310>

如果该公开出版物是专利文献的话，那么在数字标识符<310>之后，是该专利的公开号，其格式为：世界知识产权组织（WIPO）标准 3（ST 3）的国家、地区和政府间组织代码 + 标准 6（ST 6）的公开号 + 标准 16（ST 16）的文献类型，例如 CN1183117A。

（12）专利申请日：其数字标识符为<311>

如果该公开出版物是专利文献的话，那么在数字标识符<311>之后，是该专利的申请日，其格式为：YYYY - MM - DD，例如 1999?9?0。

（13）专利公开日：其数字标识符为<312>

如果该公开出版物是专利文献的话，那么在数字标识符<312>之后，是该专利的公开日，其格式为：YYYY - MM - DD，例如 1999?9?0。

4.4.7、核苷酸序列和/或氨基酸序列：其数字标识符为<400>。

在数字标识符<400>之后，是该序列的序列标识符；从下一行开始是该核苷酸和/或氨基酸序列。

该序列可以是纯核苷酸序列，或者是纯氨基酸序列，或者是核苷酸序列和与它对应的氨基酸序列。

（1）纯核苷酸序列：

核苷酸序列应当只用单链表示，从左到右是5’-末端至3’-末端的方向，序列中不应当出现术语5’和3’。

应当用单字母代码表示核苷酸序列的碱基来表述核苷酸序列的特征；只能使用与附录1之表1中给出的符号相一致的小写字母来表示。

在一个核苷酸序列中，如果经修饰的碱基是附录1之表2中列出的之一，那么在该序列本身中，应当用未修饰的碱基或“n”来表示该经修饰的碱基，符号“n”等同于唯一的一个未知的或经修饰的核苷酸；但在该序列的特征部分（数字标识符<220>-<223>）应当使用附录1之表2中给出的符号进一步表述该修饰（参见本标准4.4.5节）。附录1之表2中的符号可以用于说明书或序列的特征部分，但不得用于序列本身。

核苷酸序列中碱基的编号开始于序列中的第1个碱基，并从5’到3’方向连续地计数。该计数方法也用于构型为环状的核苷酸序列，在这种情况下，申请人可任意指定序列的第一个核苷酸。

来自大序列的一个或更多非邻接区段或来自不同序列的区段组成的核苷酸序列，应当作为带有单独序列标识符的单独序列来计数。带有一个缺口或多个缺

口的序列应当作为带有单独序列标识符的多个单独序列来计数，而单独序列的数目与序列数据的连续序列的数目相同。

核苷酸序列每行最多 60 个核苷酸碱基，每 10 个核苷酸碱基后空一格。该行的最后是该行最后一个碱基的编号。

（2）纯氨基酸序列：

对于氨基酸序列，蛋白质或肽序列中的氨基酸应当从左到右以氨基到羧基的方向列出；序列中不应当出现氨基或羧基基团。

氨基酸应当使用与附录 1 之表 3 中的符号相一致的、第一个字母大写的三字母符号表示。有空白或内部中止符号（例如“Ter”或“*”或“•”）的氨基酸序列不应当表示为单个氨基酸序列，而应当作为独立的氨基酸序列分别列出。

在一个氨基酸序列中，如果经修饰的氨基酸是附录 1 之表 4 中列出的氨基酸之一，那么在该序列本身中，应当用相应的未经修饰的氨基酸或“Xaa”来表示该经修饰的和不常用的氨基酸，符号“Xaa”等同于唯一的一个未知的或经修饰的氨基酸；但在该序列的特征部分（数字标识符<220>-<223>），应当使用附录 1 之表 4 中给出的符号进一步表述该修饰（参见本标准 4.4.5 节）。附录 1 之表 4 中的符号可以用于说明书或序列的特征部分，但不得用于序列本身。

氨基酸的编号开始于序列中的第 1 个氨基酸，以数字 1 表示并标注在该氨基酸的下面；以后每隔 5 个氨基酸在其下面标注上该氨基酸的编号。当成熟蛋白质之前存在氨基酸时，例如对于前-序列，原-序列，前-原-序列和信号序列而言，可以任选地从与成熟蛋白第一个氨基酸相邻的氨基酸开始以负数往回编

号。当氨基酸编号使用负数以区分成熟蛋白质时，不得使用数字 0。上述氨基酸序列的计数方法也适用于环状构型的氨基酸序列，申请人可以任意指定第一个氨基酸。

来自大序列的一个或更多非邻接区段或不同序列的区段组成的氨基酸序列，应当作为具有单独序列标识符的单独序列来计数。具有一个缺口或多个缺口的序列应当作为具有单独序列标识符的多个单独序列来计数，单独序列的数目与序列数据的连续序列的数目相同。

氨基酸序列每行最多 16 个氨基酸，每个氨基酸之间空一格。

（3）核苷酸序列和与它对应的氨基酸序列：

对于核苷酸序列和与它对应的氨基酸序列，对应于其编码的氨基酸的核苷酸序列的碱基应当以“三联体”密码子列出，每个密码子之间应当空一格；对应于核苷酸序列的编码部分的氨基酸可以直接列于相应密码子的下方；对于该氨基酸序列，应当在第一个氨基酸的下面标注上编号 1，然后每隔 5 个氨基酸在其下面标注上该氨基酸的编号。

对于这种核苷酸和其编码的氨基酸序列的混合形式，与核苷酸序列相对应的氨基酸序列还应当以纯氨基酸序列的形式另外给出。

4.5 数字标识符连同其后内容的排列格式

在本节中，“数字标识符及内容”指的是数字标识符连同其后的相应内容。

数字标识符及内容应当按照数字标识符的数值从小到大的次序排列在序列表中。

每个数字标识符及内容之间应当空一行，不过在前两位数字相同的数字标识符及内容之间，例如<210>到<213>之间和<220>到<223>之间，无需空一行，但对于一个序列中有多个特征的情况，在表述每个特征时，每个数字标识符<220>之前应当空一行。

对于序列表中有多个序列的情况，数字标识符及内容应当按照序列标识符的数值从小到大的次序排列。在每个序列中，应当按照数字标识符数值从小到大的次序列出仅仅与该序列有关的数字标识符及内容，即排列上从<210>到<400>的数字标识符及内容。

对于一个序列中有多个特征的情况，应当按照这些特征在序列中出现的先后次序逐一排列从<220>到<223>的数字标识符及内容。

5、序列表电子文件的格式

5.1、序列表电子文件是一个包含上述第 4 部分的数字标识符和内容，并符合上述第 4 部分格式要求的纯文本文件；该文件应当使用中华人民共和国颁布的信息交换用汉字编码字符集标准。

5.2、序列表电子文件应当记录在 CD-ROM 光盘或 3.5 英寸软盘上提交，或者按照国家知识产权局专利局规定的其它形式提交。当记录在 CD-ROM 光盘上时，该 CD-ROM 光盘应当采用 ISO9660 标准刻录；当记录在 3.5 英寸软盘上时，

该软盘应当符合 FAT 12 格式。该光盘或软盘的目录结构如下：在根目录下，有且仅有一个后缀名为“.SEQ”的纯文本文件。

6 其它事项

6.1、申请人应当保证提交的计算机可读形式的序列表电子文件中的内容与纸件形式的序列表完全相同。

6.2、申请人在形成符合本标准的序列表电子文件时，可以使用国家知识产权局专利局提供的序列表编辑软件来形成；也可以使用其它专利组织提供的软件（例如欧洲专利局提供的 Patentin）来形成；还可以使用任何纯文本文件编辑软件来形成。无论使用何种软件，所形成的电子文件都必须符合本标准的规定。

6.3、当申请人以光盘或软盘的形式提交序列表电子文件时，应当在提交的光盘或软盘上贴有永久性标记，注明申请人姓名或名称、发明名称、光盘或软盘中的文件名和提交日期；申请人委托了代理人的，也可以任选地标注上代理机构给该申请的案卷号。对于申请人补交或提交修改的情况，应当注明申请号并注明“补交”或“修改”。

注明申请人姓名或名称等项内容时，应当使用本标准中的数字标识符，即应当标注上数字标识符，并在其后注明具体内容，例如：<110> ××基因开发有限公司。注明提交日期的格式为：YYYY - MM - DD。

当序列表电子文件的字节数太大不能记录在一张软盘上时，应当将序列表电子文件记录在一张光盘上提交。

7 颁布和实施

本标准由中华人民共和国国家知识产权局颁布，自 2001 年 11 月 1 日起实施。

中华人民共和国国家知识产权局

二〇〇一年十一月一日

附：欧洲专利局的 Patentin 软件，点击下载（文件为 zip 压缩）

附录 1 核苷酸和氨基酸符号和特征关键词表

表 1 核苷酸表

符号	含义	名称的来源
a	A	腺嘌呤
g	G	鸟嘌呤
c	C	胞嘧啶
t	T	胸腺嘧啶
r	g 或 a	嘌呤
y	t/u 或 c	嘧啶
m	a 或 c	氨基
k	g 或 t/u	酮基
s	g 或 c	弱作用

		3 H键
w	a 或 t/u	强作用 2 H键
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，或其它	任何

表 2 经修饰的核苷酸表

符号	含义
ac4c	4 －乙酰胞苷
chm5u	5 －（羧羟甲基）尿苷
cm	2'－O－甲基胞苷
cmnm5s2u	5 －羧甲基氨甲基－ 2 －硫代尿苷
cmnm5u	5 －羧甲基氨甲基尿苷
d	二氢尿苷
fm	2'－O－甲基假尿苷
gal q	β，D－半乳糖 Q 核苷
gm	2'－O－甲基鸟苷
i	肌苷

i6a	Ｎ６－异戊烯基腺苷
mla	１－甲基腺苷
mlf	１－甲基假尿苷
mlg	１－甲基腺苷
mli	１－甲基肌苷
m22g	２'２－二甲基腺苷
m2a	２－甲基腺苷
m2g	２－甲基鸟苷
m3c	３－甲基胞苷
m5c	５－甲基胞苷
m6a	Ｎ６－甲基腺苷
m7g	７－甲基鸟苷
mam5u	５－甲基氨基甲基尿苷
mam5s2u	５－甲氧基氨基甲基－２－硫代尿苷
man q	β，Ｄ－甘露糖 Q 核苷
mcm5s2u	５－甲氧基羰基甲基－２－硫代尿苷
mcm5u	５－甲氧基羰基甲基尿苷
mo5u	５－甲氧基尿苷
ms2i6a	２－硫代甲基－Ｎ６－异戊烯基腺苷
ms2t6a	Ｎ－（（９－β－Ｄ－呋喃核糖基－２－硫代甲基嘌呤－６－Ｙｌ）氨基甲酰）苏氨酸
mt6a	Ｎ－（（９－β－Ｄ－呋喃核糖嘌呤－６－yl）Ｎ－甲基氨基甲酰）苏氨酸
mv	尿苷－５－氧化乙酸－甲基酯
o5u	尿苷－５－氧化乙酸
osyw	Wybutoxosine

p	假尿苷
q	Q 核苷
s2c	2 －硫代胞苷
s2t	5 －甲基－ 2 硫代尿苷
s2u	2 －硫代尿苷
s4u	4 －硫代尿苷
t	5 －甲基尿苷
t6a	N －（（ 9 －β－ D －呋喃核糖嘌呤－ 6 －基）－氨基甲酰）苏氨酸
tm	2 '－ O －甲基－ 5 －甲基尿苷
um	2 '－ O －甲基尿苷
yw	Wybutosine
x	3 －（ 3 －氨基－ 3 －羧基－丙其）尿苷，（acp3）u

表 3 三字母表示的氨基酸表

符号	含义
Ala	丙氨酸
Cys	半胱氨酸
Asp	天冬氨酸
Glu	谷氨酸
Phe	苯丙氨酸
Gly	甘氨酸
His	组氨酸

Ile	异亮氨酸
Lys	赖氨酸
Leu	亮氨酸
Met	蛋氨酸
Asn	天冬酰酸
Pro	脯氨酸
Gln	谷氨酰胺
Arg	精氨酸
Ser	丝氨酸
Thr	苏氨酸
Val	缬氨酸
Trp	色氨酸
Tyr	酪氨酸
Asx	天冬氨酸或天冬酰胺
Glx	谷氨酸或谷氨酰胺
Xaa	未知或其它

表 3 三字母表示的氨基酸表

符号	含义
Ala	丙氨酸
Cys	半胱氨酸
Asp	天冬氨酸

Glu	谷氨酸
Phe	苯丙氨酸
Gly	甘氨酸
His	组氨酸
Ile	异亮氨酸
Lys	赖氨酸
Leu	亮氨酸
Met	蛋氨酸
Asn	天冬酰酸
Pro	脯氨酸
Gln	谷氨酰胺
Arg	精氨酸
Ser	丝氨酸
Thr	苏氨酸
Val	缬氨酸
Trp	色氨酸
Tyr	酪氨酸
Asx	天冬氨酸或天冬酰胺
Glx	谷氨酸或谷氨酰胺
Xaa	未知或其它

表 4 经修饰的和不常用的氨基酸表

符号	含义
Aad	2－氨基已二酸
bAad	3－氨基已二酸
bAla	β一丙氨酸，β一氨基丙酸
Abu	2－氨基丁酸
4Abu	4－氨基丁酸，哌啶酸
Acp	6－氨基已酸
Ahe	2－氨基庚酸
Aib	2－氨基异丁酸
bAib	3－氨基异丁酸
Apm	2－氨基庚二酸
Dbu	2，4二氨基丁酸
Des	赖氨素
Dpm	2，2'一二氨基庚二酸
Dpr	2，3－二氨基丙酸
EtGly	N－乙基甘氨酸
EtAsn	N－乙基天冬氨酸
Hyl	羟基赖氨酸
aHyl	别－羟基赖氨酸
3Hyp	3－羟基脯氨酸
4Hyp	4－羟基脯氨酸
Ide	异赖氨素
aIle	别－异亮氨酸
MeGly	N－甲基甘氨酸，肌氨酸
MeIle	N－甲基异亮氨酸

MeLys	6 －N－甲基赖氨酸
MeVal	N－甲基缬氨酸
Nva	正缬氨酸
Nle	正亮氨酸
Orn	鸟氨酸

表 5 与核苷酸序列相关的特征关键词表

关键词	说明
allele	相关的个体或菌株含有相同基因的稳定的其它形式,该形式区别于这一位置的现有的序列(和或许其它序列)
attenuator	存在调节转录的终止的 DNA 区域,它控制了一些细菌操纵子的表达;(2)位于启动子和第一个结构基因之间 ,引起转录的部分终止的序列区段
C_region	免疫球蛋白轻和重链的恒定区,和 T-细胞受体α,β,和γ链；根据特定的链可包括一个或多个外显子
CAAT_signal	CAAT 盒;位于可能参与 RNA 聚合酶结合的真核生物转录单位的起始点的 75bp 上游的保守序列的一部分；共有序列 = GG(C 或 T)CAATCT
CDS	编码序列;对应于蛋白质中的氨基酸序列的核苷酸的序列(位置包括终

	止密码子) ;特征包括氨基酸概念上的翻译
Conflict	在这一位点或区域,单独确定的“相同”序列有所不同
D-loop	置换环；线粒体 DNA 内的一个区域,其中 RNA 的短的序列与 DNA 的一条链配对，代替了这一区域的原始配对 DNA 链;也用于说明在 RecA 蛋白质催化的反应中，侵入的单链替代双链 DNA 的一条链的区域
D-segment	免疫球蛋白重链的多变区,和 T-细胞受体的β链
Enhancer	顺式-作用序列,它增强了(一些)真核生物启动子的作用,并能在任一方向和与启动子相关的任何位置处（上游或下游)起作用
Exon	编码剪接 mRNA 部分的基因组区域;可以含有 5'UTR,所有 CDS,和 3'UTR
GC_signal	GC 盒;位于真核生物转录单位起始点上游的保守的富含 GC 区域,可以以多重拷贝或任一方向存在;共有序列=GGGCGG
gene	鉴定为基因的生物学意义的区域,并已经指定名称
iDNA	间插 DNA;通过几种重组中的任何一种能被消除的 DNA
intron	被转录的 DNA 区段,但通过同时剪接位于其两侧的序列(外显子)即可从转录本内部将其除去

J_segment	免疫球蛋白轻链和重链的连接区段,和 T-细胞受体α,β和γ链
LTR	长的末端重复,在确定序列的两端直接重复的序列，类型典型地见于逆转录病毒中
mat_peptide	成熟的肽或蛋白质的编码序列;翻译后修饰之后成熟的或最终的肽或蛋白质产物的编码序列;位置不包括终止密码子(与相应的 CDS 不同)
misc_binding	不能用任何其它 Binding 关键词(primer_bind 或 protein_bind)表述的与另一个组成成分共价或非-共价结合的核酸中的位点
misc_difference	特征序列与记载中存在的有所不同,并且不能用任何其它不同关键词(conflict,unsure,old_sequence,mutation,variation,allele 或 modified_base)表述
misc_feature	不能用任何其它的特征关键词表述的具有生物学意义的区域;新的或少见的特征
misc_recomb	任何一般性的,位点特异性的或复制的重组事件的位点,该位点中有不能用其它重组关键词(iDNA 和 virion)或来源关键词的修饰词(/transposon,/proviral)表述的双螺旋 DNA 的断裂和愈合

misc_RNA	不能用其他 RNA 关键词（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)限定的任何转录本或 RNA 产物
misc_signal	含有控制或改变基因功能或表达之信号的任何区域,所述信号不能用其他 Signal 关键词(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 Structure 关键词(stem_loop 和 D-loop)表述的任何二级或三级结构或构象
modified_base	被指示的核苷酸是经修饰的核苷酸,并应由被指示的分子(在 mod_base 修饰词意义中给出)所取代
mRNA	信使 RNA；包括 5'非翻译区(5'UTR)，编码序列(CDS，外显子)和 3 '非翻译区（ 3 'UTR ）
mutation	在此位置处，相关品系的序列中具有突然的，可遗传的变化
N_region	在重排的免疫球蛋白区段之间插入的额外的核苷酸

Old_sequence	在此位置处，所表述的序列修改了此序列以前的版本
PolyA_signal	聚腺苷酸化之后内切核酸酶裂解 RNA 转录本所必需的识别区域；共有序列 = AATAAA
PolyA_site	RNA 转录本上的位点,通过转录后聚腺苷酸化该位点将被加上腺嘌呤残基
Precursor_RNA	仍不是成熟的 RNA 产物的任何 RNA 种类；可包括 5'剪切区(5'clip)，5'非翻译区(5'UTR)，编码序列(CDS，外显子)，间插序列（内含子）3'非翻译区（3'UTR）,和 3'剪切区(3'clip)
prim_transcript	初级（最初的，未加工的）转录本；包括 5 '剪切区（ 5 'clip），5 '非翻译区（ 5 'UTR），编码序列(CDS,外显子),间插序列(内含子),3'非翻译区(3'UTR)和 3'剪切区(3'clip)
prim_bind	起始复制,转录或逆转录的非 - 共价的引物结合位点;包括合成的例如 PCR 引物元件的位点
Promoter	参与 RNA 聚合酶的结合以启动转录的 DNA 分子区域
protein_bind	核酸上非 - 共价的蛋白质结合位点
RBS	核糖体结合位点

repeat_region	含有重复单位的基因组区域
repeat_unit	单个重复元件
rep_origin	复制起点;复制核酸以得到两个相同拷贝的起始位点
RRNA	成熟的核糖体 RNA；将氨基酸装配成蛋白质的核糖核蛋白颗粒（核糖体）中的 RNA 成份
S_region	免疫球蛋白重链的开关区；它参与重链 DNA 的重排,导致来自相同 B - 细胞的不同免疫球蛋白类的表达
Satellite	短的基本重复单位的很多串联重复(相同或相关的);大多数具有的碱基组成或其它性质与基因组的一般水平不同,这使得它们与大部分(主带)的基因组 DNA 分离开来
ScRNA	小的细胞质 RNA;几个小的细胞质 R N A 分子中的任何一个存在于真核生物的细胞质和（有时）核中
sig_peptide	信号肽编码序列；被分泌的蛋白质的N - 末端结构域的编码序列；此结构域涉及新生多肽与膜的结合；前导序列
SnRNA	小的核 RNA；很多小的 RNA 种类中的任何一个都被局限于核中；几

	个 snRNA 参与剪接或其它 RNA 加工反应
source	鉴定序列中特定范围的生物来源；此关键词是强制性的；每一项至少要有一个跨越整个序列的单一来源关键词；每个序列可允许有一个以上的来源关键词
stem_loop	发卡结构；由 RNA 或 DNA 单链的相邻（反向）互补序列之间的碱基一配对形成的双螺旋区域
STS	序列标记位点：表述基因组上作图界标并能通过 PCR 检测的短的，单拷贝 DNA 序列；通过测定 STS 系列的次序即可作出图谱的基因组区域
TATA_signal	TATA 盒；Goldberg-Hogness 盒；在每个真核生物 RNA 聚合酶II转录单位起点前约 25bp 处发现的保守的富含 AT 的七聚体，它可能涉及使酶定位以正确地起始；共有序列 = TATA（ A或T ）A（ A或T ）
terminator	或者位于转录本的末端或者与启动子区域相邻的 DNA 序列，该序列可导致 RNA 聚合酶终止转录；也可以是阻抑蛋白的结合位点
transit_peptide	转运肽编码序列；核编码的细胞器蛋白质N - 末端结构域的编码序列；此结构域参与将蛋白质翻译后运送到细胞器中
tRNA	成熟的转移 RNA,，小的 RNA 分子（75 - 85 个碱基长），介导核酸序列翻译成氨基酸序列

unsure	作者不能确定此区域的准确序列
V_region	免疫球蛋白轻链和重链的可变区，和 T - 细胞受体α，β和γ链；编码可变的氨基末端部分；可由 V _ segment,D_segment,N_region 和 J_segment 组成
V_segment	免疫球蛋白轻链和重链的可变区段，和 T - 细胞受体α，β和γ链；编码大多数可变区（v_region）和前导肽的最后几个氨基酸
variation	含有来自相同基因的稳定突变的相关系列（例如 RFLP，多态性等），在此（和可能其它）位置处所述相同基因与被表述的不同
3’ clip	在加工过程中被切下的前体转录本 3'端大部分区域
3’ UTP	不被翻译成蛋白质的成熟转录本的 3 '末端区域（终止密码子之后）
5’ clip	在加工过程中被切下的前体转录本 5 '端大部分区域
5’ UTP	不被翻译成蛋白质的成熟转录本的 5 '末端区域（起始密码子之前）
_ 10 _signal	Pribnow 盒；细菌转录单位起点上游约 10bp 处的保守区域,它可能参与结合 RNA 聚合酶;共有序列=TatAaT
_ 35 _signal	细菌转录单位起点上游约 35bp 处的保守六聚体；共有序列 =TTGACa[]或 TGTTGACA[]

表 6 与蛋白质序列相关的特征关键词表

关键词	说明
CONFLICT	不同的论文报道了不同的序列
VARIANT	作者报道存在序列变体
VARSLIC	由可选择的剪接产生的序列变体的表述
MUTAGEN	经实验操作已改变的位点
MOD_RES	残基的翻译后修饰
ACETYLATION	N－末端或其它
AMIDATION	通常位于成熟的活性肽的 C－末端
BLOCKED	不能被测定的 N－或 C－末端封闭基团
FORMYLATION	N－末端甲硫氨酸的
GAMMA-CARBOXY-GLUTAMIC ACID HYDROXYLATION	天冬酰胺，天冬氨酸，脯氨酸或赖氨酸的
METHYLATION	通常为赖氨酸或精氨酸的
PHOSPHORYLATION	丝氨酸，苏氨酸，酪氨酸，天冬氨酸或组氨酸的
PYRROLIDONE CARBOXYLICACID	已形成内部环内酰胺的 N－末端谷氨酸
SULFATATION	通常为酪氨酸的
LIPID	脂质组成成分的共价结合
MYRISTATE	通过酰胺键与蛋白质成熟形式的 N－末端甘氨酸残基或内部的赖氨酸残基结合的豆蔻酸基团
PALMITATE	通过硫酯键与半胱氨酸残基或通过酯键与丝氨酸或苏氨酸残基结合的棕榈酸基团

FARNESYL	通过硫酯键与半胱氨酸残基结合的法尼基
GERANYL-GERANYL	通过硫酯键与半胱氨酸残基结合的香叶基－香叶基基团
GPI__ANCHOR	与蛋白质成熟形式C－末端残基的α－羧基相连的糖基－磷脂酰肌醇（GPI）基团
N__ACYL DIGLYCERIDE	原核生物脂蛋白成熟形式的N－末端半胱氨酸，所述脂蛋白具有酰胺－键联的脂肪酸和通过酯键连接了两个脂肪酸的甘油基
DISULFID	二硫键；“FROM”和“TO”终点表示通过一个链－内二硫键连接的两个残基；如果“FROM”和“TO”终点是完全相同的，则二硫键是链－间键，而说明书领域示出交联的性质
THIOLEST	硫醇酯键；“FROM”和“TO”终点表示通过硫醇酯键连接的两个残基
THIOETH	硫醚键；“FROM”和“TO”终点表示通过硫醚键连接的两个残基
CARBOHYD	糖基化位点；碳水化物（如果已知）的性质在说明书领域给出
METAL	金属离子的结合位点；说明书领域示出金属的性质
BINDING	任何化学基团（辅酶，辅基，等等）的结合位点；基团的化学性质在说明书领域给出
SIGNAL	信号序列的范围（前肽）
TRANSIT	运转肽的范围（线粒体，叶绿体或微体）
PROPEP	前肽的范围
CHAIN	成熟蛋白质中多肽链的范围
PEPTIDE	被释放的活性肽的范围
DOMAIN	序列中感兴趣的区域的范围；所述区域的特征在说明书领域给出
CA__BIND	钙－结合区域的范围
DNA__BIND	DNA-－结合区域的范围
NP_BIND	核苷酸磷酸酯结合区域；核苷酸磷酸酯的特征示于说明书领域

TRANSMEM	转膜区域的范围
ZN_FING	锌指区域的范围
SIMILAR	与另一个蛋白质序列具有相似性的区域；与那个序列有关的精确的资料在说明书领域给出
REPEAT	内部序列重复的范围
HELIX	二级结构；螺旋，例如α一螺旋， 3 （10）螺旋，或 Pi-螺旋
STRAND	二级结构；β一链，例如氢键连接的β一链，或分离的β一桥中的残基
TURN	二级结构转角，例如Ｈ－键连的转角（３－转角，４－转角或５－转角）
ACT_SITE	涉及酶活性的氨基酸
SITE	序列中任何其它感兴趣的位点
INIT_MET	已知序列以起始密码子甲硫氨酸开始
NON_TER	序列末端的残基不是末端残基；如果应用于位置１，这表示第一个位置不是完整分子的Ｎ－末端；如果应用于最后一个位置，这表示此位置不是完整分子的Ｃ－末端；对此关键词没有说明书领域
NON_CONS	非连串残基；表示序列中的两个残基不是连串的，在它们之间有很多未测序的残基
UNSURE	序列的不确定性；用于表述不能确定序列排列的序列区域

附录 2：

序列表样例

<110> ××基因开发有限公司

<120> 序列表样例

<160> 3

<170> PatentIn Version 2.1

<210> 1

<211> 389

<212> DNA

<213> 草履虫种（Paramecium sp.）

<220>

<221> misc_feature

<222> (80,100,112)

<223> n =a 或 g 或 c 或 t

<220>

<221> CDS

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

<400> 1

agctgtagtc attcctgtgt cctcttctct ctgggcttct caccctgcta atcagatctc 60

agggagagtg tcttgacccn cctctgcctt tgcagcttcn caggcaggca gncaggcagc 120

tgatgtggca attgctggca gtgccacagg cttttcagcc aggcttaggg tgggttccgc 180

cgcggcgcgg cggcccctct cgcgctcctc tcgcgcctct 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> 草履虫种（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> 根据大小和极性而设计，以用作 XYZ 蛋白的α和β链之间的接头的肽。

<400> 3

Met Val Asn Leu Glu Pro Met His Thr Glu Ile

1 5 10