**核苷酸和/或氨基酸序列表和序列表电子文件标准**

（2001年11月1日国家知识产权局令第15号公布）

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。