(4) The name and address of the depository.

(e) Any amendment required by paragraphs (d)(1), (d)(2) or (d)(4) of this section must be filed before or with the payment of the issue fee (see § 1.312).

[54 FR 34880, Aug. 22, 1989, as amended at 66 FR 21092, Apr. 27, 2001; 78 FR 62408, Oct. 21, 2013]

APPLICATION DISCLOSURES CONTAINING NUCLEOTIDE AND/OR AMINO ACID SEQUENCES

SOURCE: Sections 1.821 through 1.825 appear at 55 FR 18245, May 1, 1990, unless otherwise noted.

§ 1.821 Nucleotide and/or amino acid sequence disclosures in patent applications.

(a) Nucleotide and/or amino acid sequences, as used in §§ 1.821 through 1.825, are interpreted to mean an unbranched sequence of 4 or more amino acids or an unbranched sequence of 10 or more nucleotides. Branched sequences are specifically excluded from this definition. Sequences with fewer than four specifically defined nucleotides or amino acids are specifically excluded from this section. ''Specifically defined'' means those amino acids other than ''Xaa'' and those nucleotide bases other than ''n,'' defined in accordance with Appendices A through F to this subpart. Nucleotides and amino acids are further defined as follows:

(1) *Nucleotides.* Nucleotides are intended to embrace only those nucleotides that can be represented using the symbols set forth in Appendix A to this subpart. Modifications (*e.g.*, methylated bases) may be described as set forth in Appendix B to this subpart but shall not be shown explicitly in the nucleotide sequence.

(2) *Amino acids.* Amino acids are those L-amino acids commonly found in naturally occurring proteins and are listed in appendix C to this subpart. Those amino acid sequences containing D-amino acids are not intended to be embraced by this definition. Any amino acid sequence that contains post-translationally modified amino acids may be described as the amino acid sequence that is initially translated using the symbols shown in appendix C to this subpart, with the modified positions (*e.g.*, hydroxylations or glycosylations) being described as set forth in appendix D to this subpart, but these modifications shall not be shown explicitly in the amino acid sequence. Any peptide or protein that can be expressed as a sequence using the symbols in appendix C to this subpart, in conjunction with a description in the Feature section, to describe, for example, modified linkages, cross links and end caps, non-peptidyl bonds, etc., is embraced by this definition.

NOTE 1 TO PARAGRAPH (A): Appendices A through F to this subpart contain Tables 1–6 of the World Intellectual Property Organization (WIPO) Handbook on Industrial Property Information and Documentation, Standard ST.25: Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in Patent Applications (2009).

(b) Patent applications which contain disclosures of nucleotide and/or amino acid sequences, in accordance with the definition in paragraph (a) of this section, shall, with regard to the manner in which the nucleotide and/or amino acid sequences are presented and described, conform exclusively to the requirements of §§ 1.821 through 1.825.

(c) Patent applications that contain disclosures of nucleotide and/or amino acid sequences, as defined in paragraph (a) of this section, must contain a ''Sequence Listing,'' which is a separate part of the specification containing each of those nucleotide and/or amino acid sequences and associated information using the symbols and format in accordance with the requirements of §§ 1.822 and 1.823. The ''Sequence Listing'' must be submitted as follows, except for a national stage entry under § 1.495(b)(1), where the ''Sequence Listing'' has been previously communicated by the International Bureau or originally filed in the United States Patent and Trademark Office and complies with Patent Cooperation Treaty (PCT) Rule 5.2:

(1) As an ASCII plain text file, in compliance with § 1.824, submitted via the USPTO patent electronic filing system or on a read-only optical disc under § 1.52(e), accompanied by an incorporation by reference statement of the ASCII plain text file, in a separate paragraph of the specification, in accordance with § 1.77(b)(5);

(2) As a PDF file via the USPTO patent electronic filing system; or

(3) On physical sheets of paper.

(d) Where the description or claims of a patent application discuss a sequence that is set forth in the ''Sequence Listing,'' in accordance with paragraph (c) of this section, reference must be made to the sequence by use of a sequence identifier (§1.823(a)(5)), preceded by ''SEQ ID NO:'' or the like, in the text of the description or claims, even if the sequence is also embedded in the text of the description or claims of the patent application. Where a sequence is presented in a drawing, reference must be made to the sequence by use of the sequence identifier (§1.823(a)(5)), either in the drawing or in the Brief Description of the Drawings, where the correlation between multiple sequences in the drawing and their sequence identifiers (§1.823(a)(5)) in the Brief Description is clear.

(e)(1) If the ''Sequence Listing'' under paragraph (c) of this section is submitted in an application filed under 35 U.S.C. 111(a) as a PDF file (§1.821(c)(2)) via the USPTO patent electronic filing system or on physical sheets of paper (§1.821(c)(3)), then the following must be submitted:

(i) A CRF of the ''Sequence Listing,'' in accordance with the requirements of §1.824; and

(ii) A statement that the sequence information contained in the CRF submitted under paragraph (e)(1)(i) of this section is identical to the sequence information contained in the ''Sequence Listing'' under paragraph (c) of this section.

(2) If the ''Sequence Listing'' under paragraph (c) of this section in an application submitted under 35 U.S.C. 371 is a PDF file (paragraph (c)(2) of this section) or on physical sheets of paper (paragraph (c)(3) of this section), and not also as an ASCII plain text file, in compliance with §1.824 (paragraph (c)(1) of this section), then the following must be submitted:

(i) A CRF of the ''Sequence Listing,'' in accordance with the requirements of §1.824; and

(ii) A statement that the sequence information contained in the CRF submitted under paragraph (e)(2)(i) of this section is identical to the sequence information contained in the ''Sequence Listing'' under paragraph (c)(2) or (3) of this section.

(3) If a ''Sequence Listing'' in ASCII plain text format, in compliance with §1.824, has not been submitted for an international application under the PCT, and that application contains disclosures of nucleotide and/or amino acid sequences, as defined in paragraph (a) of this section, and is to be searched by the United States International Searching Authority or examined by the United States International Preliminary Examining Authority, then the following must be submitted:

(i) A CRF of the ''Sequence Listing,'' in accordance with the requirements of §1.824;

(ii) The late furnishing fee for providing a ''Sequence Listing'' in response to an invitation, as set forth in §1.445(a)(5); and

(iii) A statement that the sequence information contained in the CRF, submitted under paragraph (e)(3)(i) of this section, does not go beyond the disclosure in the international application as filed, or a statement that the information recorded in the ASCII plain text file, submitted under paragraph (e)(3)(i) of this section, is identical to the sequence listing contained in the international application as filed, as applicable.

(4) The CRF may not be retained as a part of the patent application file.

(f) [Reserved]

(g) If any of the requirements of paragraphs (b) through (e) of this section are not satisfied at the time of filing under 35 U.S.C. 111(a) or at the time of entering the national stage under 35 U.S.C. 371, the applicant will be notified and given a period of time within which to comply with such requirements in order to prevent abandonment of the application. Any amendment to add or replace a ''Sequence Listing'' and CRF copy thereof in reply to a requirement under this paragraph must be submitted in accordance with the requirements of §1.825.

(h) If any of the requirements of paragraph (e)(3) of this section are not satisfied at the time of filing an international application under the PCT, and the application is to be searched by the United States International

Searching Authority or examined by the United States International Preliminary Examining Authority, the applicant may be sent a notice necessitating compliance with the requirements within a prescribed time period. Where a ''Sequence Listing'' under PCT Rule 13*ter* is provided in reply to a requirement under this paragraph, it must be accompanied by a statement that the information recorded in the ASCII plain text file under paragraph (e)(3)(i) of this section is identical to the sequence listing contained in the international application as filed, or does not go beyond the disclosure in the international application as filed, as applicable. It must also be accompanied by the late furnishing fee, as set forth in §1.445(a)(5). If the applicant fails to timely provide the required CRF, the United States International Searching Authority shall search only to the extent that a meaningful search can be performed without the CRF, and the United States International Preliminary Examining Authority shall examine only to the extent that a meaningful examination can be performed without the CRF.

[63 FR 29634, June 1, 1998, as amended at 65 FR 54680, Sept. 8, 2000; 69 FR 18803, Apr. 9, 2004; 70 FR 10489, Mar. 4, 2005; 86 FR 57048, Oct. 14, 2021]

§ 1.822 Symbols and format to be used for nucleotide and/or amino acid sequence data.

(a) The symbols and format to be used for nucleotide and/or amino acid sequence data shall conform to the requirements of paragraphs (b) through (e) of this section.

(b) The code for representing the nucleotide and/or amino acid sequence characters shall conform to the code set forth in appendices A and C to this subpart. No code other than that specified in these sections shall be used in nucleotide and amino acid sequences. A modified base or modified or unusual amino acid may be presented in a given sequence as the corresponding unmodified base or amino acid if the modified base or modified or unusual amino acid is one of those listed in appendices B and D to this subpart, and the modification is also set forth in the Feature section. Otherwise, each occurrence of a base or amino acid not appearing in appendices A and C, shall be listed in a given sequence as ''n'' or ''Xaa,'' respectively, with further information, as appropriate, given in the Feature section, by including one or more feature keys listed in appendices E and F to this subpart.

NOTE 1 TO PARAGRAPH (B): Appendices A through F to this subpart contain Tables 1–6 of the World Intellectual Property Organization (WIPO) Handbook on Industrial Property Information and Documentation, Standard ST.25: Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in Patent Applications (2009).

(c) *Format representation of nucleotides.* (1) A nucleotide sequence shall be listed using the lowercase letter for representing the one-letter code for the nucleotide bases set forth in appendix A to this subpart.

(2) The bases in a nucleotide sequence (including introns) shall be listed in groups of 10 bases except in the coding parts of the sequence. Leftover bases, fewer than 10 in number, at the end of noncoding parts of a sequence shall be grouped together and separated from adjacent groups of 10 or 3 bases by a space.

(3) The bases in the coding parts of a nucleotide sequence shall be listed as triplets (codons). The amino acids corresponding to the codons in the coding parts of a nucleotide sequence shall be listed immediately below the corresponding codons. Where a codon spans an intron, the amino acid symbol shall be listed below the portion of the codon containing two nucleotides.

(4) A nucleotide sequence shall be listed with a maximum of 16 codons or 60 bases per line, with a space provided between each codon or group of 10 bases.

(5) A nucleotide sequence shall be represented, only by a single strand, in the 5 to 3 direction, from left to right.

(6) The enumeration of nucleotide bases shall start at the first base of the sequence with number 1. The enumeration shall be continuous through the whole sequence in the direction 5 to 3. The enumeration shall appear in the right margin, next to the line containing the one-letter codes for the bases and giving the number of the last base of that line.