AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

(h) If any of the requirements of paragraphs (b) through (f) of this section are not satisfied at the time of filing an international application under the Patent Cooperation Treaty (PCT), which application is to be searched by the United States International Searching Authority or examined by the United States International Preliminary Examining Authority, applicant will be sent a notice necessitating compliance with the requirements within a prescribed time period. Any submission in reply to a requirement under this paragraph must be accompanied by a statement that the submission does not include matter which goes beyond the disclosure in the international application as filed. If applicant fails to timely provide the required computer readable form, the United States International Searching Authority shall search only to the extent that a meaningful search can be performed without the computer readable form and the United States International Preliminary Examining Authority shall examine only to the extent that a meaningful examination can be performed without the computer readable form.

[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]

§ 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 the tables in WIPO Standard ST.25 (1998), Appendix 2, Tables 1 and 3. This incorporation by reference was approved by the Director of the Federal Register in accordance with 5 U.S.C. 552(a) and 1 CFR part 51. Copies of ST.25 may be obtained from the World Intellectual Property Organization; 34 chemin des Colombettes; 1211 Geneva 20 Switzerland. Copies may also be inspected at the National Archives and Records Administration (NARA). For information on the availability of this material at NARA, call 202–741–6030, or go to: *http://www.archives.gov/federal_register/code_of_federal_regulations/ibr_locations.html*. 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 WIPO Standard ST.25 (1998), Appendix 2, Tables 2 and 4, and the modification is also set forth in the Feature section. Otherwise, each occurrence of a base or amino acid not appearing in WIPO Standard ST.25 (1998), Appendix 2, Tables 1 and 3, shall be listed in a given sequence as ''n'' or ''Xaa,'' respectively, with further information, as appropriate, given in the Feature section, preferably by including one or more feature keys listed in WIPO Standard ST.25 (1998), Appendix 2, Tables 5 and 6.

(c) *Format representation of nucleotides.* (1) A nucleotide sequence shall be listed using the lower-case letter for representing the one-letter code for the nucleotide bases set forth in WIPO Standard ST.25 (1998), Appendix 2, Table 1.

(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 typed immediately below the corresponding codons. Where a codon spans an intron, the amino acid symbol shall be typed 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 presented, 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 be marked 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.

(7) For those nucleotide sequences that are circular in configuration, the enumeration method set forth in paragraph (c)(6) of this section remains applicable with the exception that the designation of the first base of the nucleotide sequence may be made at the option of the applicant.

(d) *Representation of amino acids.* (1) The amino acids in a protein or peptide sequence shall be listed using the three-letter abbreviation with the first letter as an upper case character, as in WIPO Standard ST.25 (1998), Appendix 2, Table 3.

(2) A protein or peptide sequence shall be listed with a maximum of 16 amino acids per line, with a space provided between each amino acid.

(3) An amino acid sequence shall be presented in the amino to carboxy direction, from left to right, and the amino and carboxy groups shall not be presented in the sequence.

(4) The enumeration of amino acids may start at the first amino acid of the first mature protein, with the number 1. When presented, the amino acids preceding the mature protein, *e.g.*, pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, shall have negative numbers, counting backwards starting with the amino acid next to number 1. Otherwise, the enumeration of amino acids shall start at the first amino acid at the amino terminal as number 1. It shall be marked below the sequence every 5 amino acids. The enumeration method for amino acid sequences that is set forth in this section remains applicable for amino acid sequences that are circular in configuration, with the exception that the designation of the first amino acid of the sequence may be made at the option of the applicant.

(5) An amino acid sequence that contains internal terminator symbols (*e.g.*, ''Ter'', ''*'', or ''.'', etc.) may not be represented as a single amino acid sequence, but shall be presented as separate amino acid sequences.

(e) A sequence with a gap or gaps shall be presented as a plurality of separate sequences, with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data. A sequence that is made up of one or more noncontiguous segments of a larger sequence or segments from different sequences shall be presented as a separate sequence.

[63 FR 29635, June 1, 1998, as amended at 69 FR 18803, Apr. 9, 2004; 70 FR 10489, Mar. 4, 2005]

§ 1.823 Requirements for nucleotide and/or amino acid sequences as part of the application.

(a)(1) If the ''Sequence Listing'' required by §1.821(c) is submitted on paper: The ''Sequence Listing,'' setting forth the nucleotide and/or amino acid sequence and associated information in accordance with paragraph (b) of this section, must begin on a new page and must be titled ''Sequence Listing.'' The pages of the ''Sequence Listing'' preferably should be numbered independently of the numbering of the remainder of the application. Each page of the ''Sequence Listing'' shall contain no more than 66 lines and each line shall contain no more than 72 characters. The sheet or sheets presenting a sequence listing may not include material other than part of the sequence listing. A fixed-width font should be used exclusively throughout the ''Sequence Listing.''

(2) If the ''Sequence Listing'' required by §1.821(c) is submitted on compact disc: The ''Sequence Listing'' must be submitted on a compact disc in compliance with §1.52(e). The compact disc may also contain table information if the application contains table information that may be submitted on a compact disc (§1.52(e)(1)(iii)). The specification must contain an incorporation-by-reference of the Sequence Listing as required by §1.52(e)(5). The presentation of the