(1) Examination will begin after the earliest of:

(i) Receipt of the amendment;

(ii) Receipt of applicant's statement that no amendment will be made; or

(iii) Expiration of the time period set in the notification.

(2) No international preliminary examination report will be established prior to issuance of an international search report.

*   *   *   *   *

17. Section 1.485 is amended by revising paragraph (a) to read as follows:

§ 1.485   Amendment by applicant during international preliminary examination.

(a) The applicant may make amendments at the time of filing the Demand. The applicant may also make amendments within the time limit set by the International Preliminary Examining Authority for reply to any notification under § 1.484(b) or to any written opinion. Any such amendments must:

(1) Be made by submitting a replacement sheet in compliance with PCT Rules 10 and 11.1 to 11.13 for every sheet of the application which differs from the sheet it replaces unless an entire sheet is cancelled; and

(2) Include a description of how the replacement sheet differs from the replaced sheet. Amendments that do not comply with PCT Rules 10 and 11.1 to 11.13 may not be entered.

*   *   *   *   *

18. Section 1.494 is amended by revising paragraph (c) to read as follows:

§ 1.494   Entering the national stage in the United States of America as a Designated Office.

*   *   *   *   *

(c) If applicant complies with paragraph (b) of this section before expiration of 20 months from the priority date but omits:

(1) A translation of the international application, as filed, into the English language, if it was originally filed in another language (35 U.S.C. 371(c)(2)) and/or

(2) The oath or declaration of the inventor (35 U.S.C. 371(c)(4); see § 1.497), applicant will be so notified and given a period of time within which to file the translation and/or oath or declaration in order to prevent abandonment of the application. The payment of the processing fee set forth in § 1.492(f) is required for acceptance of an English translation later than the expiration of 20 months after the priority date. The payment of the surcharge set forth in § 1.492(e) is required for acceptance of the oath or declaration of the inventor later than the

expiration of 20 months after the priority date. A ''Sequence Listing'' need not be translated if the ''Sequence Listing'' complies with PCT Rule 12.1(d) and the description complies with PCT Rule 5.2(b).

*   *   *   *   *

19. Section 1.495 is amended by revising paragraph (c) to read as follows:

§ 1.495   Entering the national stage in the United States of America as an Elected Office.

*   *   *   *   *

(c) If applicant complies with paragraph (b) of this section before expiration of 30 months from the priority date but omits:

(1) A translation of the international application, as filed, into the English language, if it was originally filed in another language (35 U.S.C. 371(c)(2)) and/or

(2) The oath or declaration of the inventor (35 U.S.C. 371(c)(4); see § 1.497), applicant will be so notified and given a period of time within which to file the translation and/or oath or declaration in order to prevent abandonment of the application. The payment of the processing fee set forth in § 1.492(f) is required for acceptance of an English translation later than the expiration of 30 months after the priority date. The payment of the surcharge set forth in § 1.492(e) is required for acceptance of the oath or declaration of the inventor later than the expiration of 30 months after the priority date.

A ''Sequence Listing'' need not be translated if the ''Sequence Listing'' complies with PCT Rule 12.1(d) and the description complies with PCT Rule 5.2(b).

*   *   *   *   *

Dated: May 22, 1998.

Bruce A. Lehman,

Assistant Secretary of Commerce and Commissioner of Patents and Trademarks.

[FR Doc. 98–14195 Filed 5–29–98; 8:45 am]

BILLING CODE 3510–16–P

DEPARTMENT OF COMMERCE

Patent and Trademark Office

37 CFR Part 1

[Docket No: 960828235–8109–02]

RIN: 0651–AA88

Requirements for Patent Applications Containing Nucleotide Sequence and/ or Amino Acid Disclosures

AGENCY: Patent and Trademark Office, Commerce.

ACTION: Final rule.

SUMMARY: The Patent and Trademark Office (PTO) is amending the rules for submitting nucleotide or amino acid sequences in computer readable form (CRF) for patent applications. These amendments simplify the requirements of the rules, rearrange portions of the rules for better understanding and establish consistent rules to permit a single internationally acceptable computer readable form. Sequence Listings will be presented in an international, language neutral format using numeric identifiers rather than the current subject headings. The Paper Sequence Listing will preferably be a separately numbered section of the patent application. Sequences which contain fewer than four specifically identified nucleotides or amino acids will no longer be required to be submitted in computer readable form.

DATES: *Effective date:* July 1, 1998. The incorporation by reference of certain publications listed in the regulations is approved by the Director of the Federal Register as of July 1, 1998.

Applicability date: Sections 1.821 through 1.825 as amended apply to applications filed on or after July 1, 1998, except for: (1) applications that claim the benefit of a prior application under 35 U.S.C. 120 filed before July 1, 1998, and which do not add subject matter involving a sequence listing subject to §§ 1.821 through 1.825; and (2) reissue applications in which the application for the patent sought to be reissued was filed before July 1, 1998. Sections 1.821 through 1.825 apply during a reexamination proceeding if the application for the patent sought to be reexamined was filed on or after July 1, 1998.

FOR FURTHER INFORMATION CONTACT: Esther M. Kepplinger, by telephone at (703) 308–1495; by mail addressed to: Box Comments—Patents, Assistant Commissioner for Patents, Washington, DC 20231 marked to her attention; by facsimile to (703) 305–3935; or by electronic mail at esther.kepplinger@uspto.gov.

SUPPLEMENTARY INFORMATION: Sections 1.821 through 1.825 of title 37 provide a standardized format for the description of nucleotide and amino acid sequence data in patent applications and require the submission of such sequences in computer readable form (CRF). Sections 1.821 through 1.825 provide the following benefits to the PTO: (1) Improved search capabilities; (2) improved interference detection; (3) more efficient examination; (4) cost savings for the input of the sequence data; (5) more

efficient and accurate printing of sequences in patents; (6) exchange of the sequence data with other patent offices electronically; and (7) improved public access to the sequences electronically.

Reasons for the Changes

In response to the needs of our customers, the procedural requirements found in former §§ 1.821 through 1.825 have been reduced. Sections 1.821 through 1.825 are being amended to be consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (signed in 1998 and effective July 1, 1998). ST.25 replaces WIPO Standards ST.23 and ST.24 which deal with paper and electronic submissions of sequence listings.

A Meeting of International Authorities (MIA) under the Patent Cooperation Treaty (PCT) was held in November of 1994 to discuss simplification of sequence listing submission requirements. Under the previous PCT Regulations, each International Searching Authority, each International Preliminary Examining Authority and each designated/elected office was free to set the requirements for submission of sequence listings in paper and electronic form. This imposed a burden on applicants by requiring them to prepare sequence listings in many different formats. In addition, sequence listings were required to be translated for consideration in the national stage at considerable cost to applicants and at the risk that the information could be inaccurately translated.

After the November 1994 MIA, the PTO, the European Patent Office (EPO) and the Japanese Patent Office (JPO) worked together with WIPO to create a new international standard which forms the basis of WIPO Standard ST.25 (1998). Sections 1.821 through 1.825 of 37 CFR, as amended herein, are consistent with WIPO Standard ST.25 (1998) and the PCT sequence listing requirements. Sequence listings prepared in accordance with §§ 1.821 through 1.825 as amended generally will be acceptable in all countries which adhere to WIPO Standard ST.25 (1998). In addition, a sequence listing prepared in accordance with the §§ 1.821 through 1.825 as amended will be acceptable for the national stage in all PCT member countries which require the submission of a sequence listing. As a result of this rule change, applicants will experience a reduction in cost since only one sequence listing in paper and electronic form will need to be prepared and translations of this listing will not be needed.

All necessary changes to the text of §§ 1.821 through 1.825 to reflect the new WIPO Standard ST.25 (1998), have been made. Each change is described below.

Overview of the Changes

The changes in this Final Rule include:

(1) Use of numeric identifiers to replace the language subject headings within the submission;

(2) Elimination of unnecessary and confusing data elements;

(3) Movement of the paper Sequence Listing to the end of the application, preferably with separately numbered pages;

(4) Elimination of the requirement to provide a submission for sequences with fewer than four specifically defined nucleotides or amino acids;

(5) Use of *lower-case* one-letter codes for nucleotide bases;

(6) Rearrangement of portions of the rules to improve their context;

(7) Clarification and simplification of the rules to aid in understanding; and

(8) Minor changes to accomplish harmonization with WIPO Standard ST.25 (1998) as well as the EPO and the JPO standards.

Amended §§ 1.821 through 1.825 are not mandatory for: (1) applications that claim the benefit of a prior application under 35 U.S.C. 120 filed before July 1, 1998, and which do not add subject matter involving a sequence listing subject to §§ 1.821 through 1.825; (2) reissue applications in which the application for the patent sought to be reissued was filed before July 1, 1998; and (3) reexamination proceedings if the application for the patent sought to be reexamined was filed before July 1, 1998. The PTO will accept and encourages the submission of sequence listings in compliance with amended §§ 1.821 through 1.825 for any application or reexamination proceeding. All sequence listings (including the entire computer readable form) must be submitted in compliance with either §§ 1.821 through 1.825 as amended in this Final Rule or (when permitted) former §§ 1.821 through 1.825.

If the CRF for a new application would be identical to a compliant CRF already on file in the PTO, the applicant may make reference to the other application and the CRF in lieu of filing a duplicate CRF in the new application by following the procedures set forth in § 1.821(e). If exceptional circumstances do arise and certain applicants experience specific hardships in attempting to comply with amended §§ 1.821 through 1.825, the PTO will

consider a petition under § 1.183 to waive certain requirements of §§ 1.821 through 1.825.

A Notice of Proposed Rulemaking entitled ''Changes Implementing Nucleotide and/or Amino Acid Sequence Listings'' (Notice of Proposed Rulemaking) was published in the **Federal Register** at 61 FR 51855 (October 4, 1996), and in the *Official Gazette of the Patent and Trademark Office*, at 1191 *Off. Gaz. Pat. Office* 168 (October 29, 1996). Sections 1.821 through 1.825 as adopted contain several changes from these sections. This Final Rule provides a discussion of the content of the specific rules being amended, description of the changes in the text of the proposed rules, and explanation of the reasons supporting the changes. In addition, comments received in response to the Notice of Proposed Rulemaking are analyzed.

Discussion of Specific Rules and Changes from the Proposed Rules:

Title 37 of the Code of Federal Regulations, Part 1, is amended as follows.

Section 1.77

The proposed change to 37 CFR 1.77 was previously adopted. *See Miscellaneous Changes to Patent Practice*; Final Rule, 61 FR 42790 (August 19, 1996), 1190 *Off. Gaz. Pat. Office* 67 (September 17, 1996).

Section 1.821

Section 1.821 incorporates by reference the World Intellectual Property Organization (WIPO) Handbook on Industrial Property Information and Documentation, Standard ST.25 (1998), including Tables 1 through 6 of Appendix 2, in accordance with 5 U.S.C. 552(a) and 1 CFR part 51. Copies may be obtained from the World Intellectual Property Organization; 34 chemin des Colombettes; 1211 Geneva 20 Switzerland. Copies may be inspected at the Patent Search Room; Crystal Plaza 3, Lobby Level; 2021 South Clark Place; Arlington, VA 22202. Copies may also be inspected at the Office of the Federal Register, 800 North Capitol Street, NW, Suite 700, Washington, DC 20408. These Tables are reproduced below.

WIPO Standard ST.25 (1998), Appendix 2, Table 1, provides that the bases of a nucleotide sequence should be represented using the following one-letter code for nucleotide sequence characters:

TABLE 1.—ONE LETTER CODES FOR NUCLEOTIDE SEQUENCES

Symbol	Meaning	Origin of designation
a	a	*a*denine.
g	g	*g*uanine.
c	c	*c*ytosine.
t	t	*t*hymine.
u	u	*u*racil.
r	g or a	*pu*rine.
y	t/u or c	p*y*rimidine.
m	a or c	a*m*ino.
k	g or t/u	*k*eto.
s	g or c	*s*trong interactions 3 H-bonds.
w	a or t/u	*w*eak interactions 2 H-bonds.
b	g or c or t/u	not a.
d	a or g or t/u	not c.
h	a or c or t/u	not g.
v	a or g or c	not t, not u.
n	(a or g or c or t/u) or (unknown or other)	a*n*y

WIPO Standard ST.25 (1998), Appendix 2, Table 2, provides that modified bases may be represented as the corresponding unmodified bases in the sequence itself, if the modified base is one of those listed below and the modification is further described in the Feature section of the Sequence Listing. The codes from the list below may be used in the description (i.e., the specification and drawings, or in the Sequence Listing) but these codes may not be used in the sequence itself.

TABLE 2.—MODIFIED BASES

Symbol	Meaning
ac4c	4-acetylcytidine.
chm5u	5-(carboxyhydroxylmethyl)uridine.
cm	2-O-methylcytidine.
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine.
cmnm5u	5-carboxymethylaminomethyluridine.
d	dihydrouridine.
fm	2-O-methylpseudouridine.
gal q	beta, D-galactosylqueuosine.
gm	2-O-methylguanosine.
I	inosine.
i6a	N6-isopentenyladenosine.
m1a	1-methyladenosine.
m1f	1-methylpseudouridine.
m1g	1-methylguanosine.
m1i	1-methylinosine.
m22g	2,2-dimethylguanosine.
m2a	2-methyladenosine.
m2g	2-methylguanosine.
m3c	3-methylcytidine.
m5c	5-methylcytidine.
m6a	N6-methyladenosine.
m7g	7-methylguanosine.
mam5u	5-methylaminomethyluridine.
mam5s2u	5-methoxyaminomethyl-2-thiouridine.
man q	beta, D-mannosylqueuosine.
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine.
mcm5u	5-methoxycarbonylmethyluridine.
mo5u	5-methoxyuridine.
ms2i6a	2-methylthio-N6-isopentenyladenosine.
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl) carbamoyl) threonine.
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methylcarbamoyl) threonine.
mv	uridine-5-oxyacetic acid-methylester.
o5u	uridine-5-oxyacetic acid.
osyw	wybutoxosine.
p	pseudouridine.
q	queuosine.
s2c	2-thiocytidine.
s2t	5-methyl-2-thiouridine.
s2u	2-thiouridine.
s4u	4-thiouridine.
t	5-methyluridine.
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)-carbamoyl)threonine.
tm	2-O-methyl-5-methyluridine.
um	2-O-methyluridine.

TABLE 2.—MODIFIED BASES—Continued

Symbol	Meaning
yw	wybutosine.
x	3-(3-amino-3-carboxy-propyl)uridine, (acp3)u.

WIPO Standard ST.25 (1998), Appendix 2, Table 3, provides that the amino acids should be represented using the following three-letter code with the first letter as a capital.

TABLE 3.—AMINO ACID THREE-LETTER CODES

Symbol	Meaning
Ala	Alanine.
Cys	Cysteine.
Asp	Aspartic Acid.
Glu	Glutamic Acid.
Phe	Phenylalanine.
Gly	Glycine.
His	Histidine.
Ile	Isoleucine.
Lys	Lysine.

TABLE 3.—AMINO ACID THREE-LETTER CODES—Continued

Symbol	Meaning
Leu	Leucine.
Met	Methionine.
Asn	Asparagine.
Pro	Proline.
Gln	Glutamine.
Arg	Arginine.
Ser	Serine.
Thr	Threonine.
Val	Valine.
Trp	Tryptophan.
Tyr	Tyrosine.
Asx	Asp or Asn.
Glx	Glu or Gln.
Xaa	Unknown or other.

WIPO Standard ST.25 (1998), Appendix 2, Table 4, provides that modified and unusual amino acids may be represented as the corresponding unmodified amino acids in the sequence itself if the modified or unusual amino acid is one of those listed below and the modification is further described in the Feature section of the Sequence Listing. The codes from the list below may be used in the description (i.e., the specification and drawings, or in the Sequence Listing) but these codes may not be used in the sequence itself.

TABLE 4.—MODIFIED AND UNUSUAL AMINO ACID CODES

Symbol	Meaning
Aad	2-Aminoadipic acid.
bAad	3-aminoadipic acid.
bAla	beta-Alanine, beta-Aminopropionic acid.
Abu	2-Aminobutyric acid.
4Abu	4-Aminobutyric acid, piperidinic acid.
Acp	6-Aminocaproic acid.
Ahe	2-Aminoheptanoic acid.
Aib	2-Aminoisobutyric acid.
bAib	3-Aminoisobutyric acid.
Apm	2-Aminopimelic acid.
Dbu	2,4-Diaminobutyric acid.
Des	Desmosine.
Dpm	2,2-Diaminopimelic acid.
Dpr	2,3-Diaminopropionic acid.
EtGly	N-Ethylglycine.
EtAsn	N-Ethylasparagine.
Hyl	Hydroxylysine.
aHyl	allo-Hydroxylysine.
3Hyp	3-Hydroxyproline.
4Hyp	4-Hydroxyproline.
Ide	Isodesmosine.
aIle	allo-Isoleucine.
MeGly	N-Methylglycine, sarcosine.
MeIle	N-Methylisoleucine.
MeLys	6-N-Methyllysine.
MeVal	N-Methylvaline.
Nva	Norvaline.
Nle	Norleucine.
Orn	Ornithine.

WIPO Standard ST.25 (1998), Appendix 2, Table 5 provides for feature keys related to DNA sequences.

BILLING CODE 3510–16–P

Table 5: Feature keys related to nucleotide sequences

Key	Description
allele	a related individual or strain contains stable, alternative forms of the same gene which differs from the presented sequence at this location (and perhaps others)
attenuator	1) region of DNA at which regulation of termination of transcription occurs, which controls the expression of some bacterial operons; 2) sequence segment located between the promoter and the first structural gene that causes partial termination of transcription
C_region	constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains. Includes one or more exons depending on the particular chain
CAAT_signal	CAAT box; part of a conserved sequence located about 75 bp up-stream of the start point of eukaryotic transcription units which may be involved in RNA polymerase binding; consensus=GG (C or T) CAATCT
CDS	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon). Feature includes amino acid conceptual translation
conflict	independent determinations of the "same" sequence differ at this site or region
D-loop	displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA by a single stranded invader in the reaction catalyzed by RecA protein
D-segment	diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain
enhancer	a cis-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter
exon	region of genome that codes for portion of spliced mRNA; may contain 5'UTR, all CDSs, and 3'UTR
GC_signal	GC box; a conserved GC-rich region located upstream of the start point of eukaryotic transcription units which may occur in multiple copies or in either orientation; consensus=GGGCGG
gene	region of biological interest identified as a gene and for which a name has been assigned
iDNA	Intervening DNA; DNA which is eliminated through any of several kinds of recombination
intron	a segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it
J_segment	joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma-chains
LTR	long terminal repeat, a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses
mat_peptide	mature peptide or protein coding sequence; coding sequence for the mature or final peptide or protein product following post-translational modification. The location does not include the stop codon (unlike the corresponding CDS)

misc_binding	site in nucleic acid which covalently or non-covalently binds another moiety that cannot be described by any other Binding key (primer_bind or protein_bind)
misc_difference	feature sequence is different from that presented in the entry and cannot be described by any other Difference key (conflict, unsure, old_sequence, mutation, variation, allele, or modified_base)
misc_feature	region of biological interest which cannot be described by any other feature key; a new or rare feature
misc_recomb	site of any generalized, site-specific or replicative recombination event where there is a breakage and reunion of duplex DNA that cannot be described by other recombination keys (iDNA and virion) or qualifiers of source key (/insertion_seq,/transposon,/proviral)
misc_RNA	any transcript or RNA product that cannot be defined by other RNA keys (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, and snRNA)
misc_signal	any region containing a signal controlling or altering gene function or expression that cannot be described by other Signal keys (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator, and rep_origin)
misc_structure	any secondary or tertiary structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)
modified_base	the indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)
mRNA	messenger RNA; includes 5'untranslated region (5'UTR), coding sequences (CDS, exon) and 3'untranslated region (3'UTR)
mutation	a related strain has an abrupt, inheritable change in the sequence at this location
N_region	Extra nucleotides inserted between rearranged immunoglobulin segments
old_sequence	the presented sequence revises a previous version of the sequence at this location
polyA_signal	recognition region necessary for endonuclease cleavage of an RNA transcript that is followed by polyadenylation; consensus=AATAAA
polyA_site	site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation
precursor_RNA	any RNA species that is not yet the mature RNA product; may include 5'clipped region (5'clip), 5'untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3'untranslated region (3'UTR), and 3'clipped region (3'clip)
prim_transcript	primary (initial, unprocessed) transcript; includes 5'clipped region (5'clip), 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3'untranslated region (3'UTR), and 3'clipped region (3'clip)
primer_bind	Non-covalent primer binding site for initiation of replication, transcription, or reverse transcription. Includes site(s) for synthetic e.g., PCR primer elements
promoter	region on a DNA molecule involved in RNA polymerase binding to initiate transcription
protein_bind	non-covalent protein binding site on nucleic acid
RBS	ribosome binding site
repeat_region	region of genome containing repeating units
repeat_unit	single repeat element

rep_origin	origin of replication; starting site for duplication of nucleic acid to give two identical copies
rRNA	mature ribosomal RNA; the RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins
S_region	Switch region of immunoglobulin heavy chains. Involved in the rearrangement of heavy chain DNA leading to the expression of a different immunoglobulin class from the same B-cell
satellite	many tandem repeats (identical or related) of a short basic repeating unit; many have a base composition or other property different from the genome average that allows them to be separated from the bulk (main band) genomic DNA
scRNA	small cytoplasmic RNA; any one of several small cytoplasmic RNA molecules present in the cytoplasm and (sometimes) nucleus of a eukaryote
sig_peptide	signal peptide coding sequence; coding sequence for an N-terminal domain of a secreted protein; this domain is involved in attaching nascent polypeptide to the membrane; leader sequence
snRNA	small nuclear RNA; any one of many small RNA species confined to the nucleus; several of the snRNAs are involved in splicing or other RNA processing reactions
source	identifies the biological source of the specified span of the sequence. This key is mandatory. Every entry will have, as a minimum, a single source key spanning the entire sequence. More than one source key per sequence is permissible
stem_loop	hairpin; a double-helical region formed by base-pairing between adjacent (inverted) complementary sequences in a single strand of RNA or DNA
STS	Sequence Tagged Site. Short, single-copy DNA sequence that characterizes a mapping landmark on the genome and can be detected by PCR. A region of the genome can be mapped by determining the order of a series of STSs
TATA_signal	TATA box; Goldberg-Hogness box; a conserved AT-rich septamer found about 25 bp before the start point of each eukaryotic RNA polymerase II transcript unit which may be involved in positioning the enzyme for correct initiation; consensus=TATA(A or T)A(A or T)
terminator	sequence of DNA located either at the end of the transcript or adjacent to a promoter region that causes RNA polymerase to terminate transcription; may also be site of binding of repressor protein
transit_peptide	transit peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organellar protein; this domain is involved in post-translational import of the protein into the organelle
tRNA	mature transfer RNA, a small RNA molecule (75-85 bases long) that mediates the translation of a nucleic acid sequence into an amino acid sequence
unsure	author is unsure of exact sequence in this region
V_region	Variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains. Codes for the variable amino terminal portion. Can be made up from V_segments, D_segments, N_regions, and J_segments
V_segment	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains. Codes for most of the variable region (V_region) and the last few amino acids of the leader peptide

variation	a related strain contains stable mutations from the same gene (e.g., RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others)
3'clip	3'-most region of a precursor transcript that is clipped off during processing
3'UTR	region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein
5'clip	5'-most region of a precursor transcript that is clipped off during processing
5'UTR	region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein
-10_signal	pribnow box; a conserved region about 10 bp upstream of the start point of bacterial transcription units which may be involved in binding RNA polymerase; consensus=TATAAT
-35_signal	a conserved hexamer about 35 bp upstream of the start point of bacterial transcription units; consensus=TTGACA [ ] or TGTTGACA [ ]

WIPO Standard ST.25 (1998), Appendix 2, Table 6 provide for feature keys related to protein sequences.

Table 6: Feature keys related to Protein sequences

Key	Description
CONFLICT	Different papers report differing sequences
VARIANT	Authors report that sequence variants exist
VARSPLIC	Description of sequence variants produced by alternative splicing
MUTAGEN	Site which has been experimentally altered
MOD_RES	Post-translational modification of a residue
ACETYLATION	N-terminal or other
AMIDATION	Generally at the C-terminal of a mature active peptide
BLOCKED	Undetermined N- or C-terminal blocking group
FORMYLATION	Of the N-terminal methionine
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	Of asparagine, aspartic acid, proline or lysine
METHYLATION	Generally of lysine or arginine
PHOSPHORYLATION	Of serine, threonine, tyrosine, aspartic acid or histidine
PYRROLIDONE CARBOXYLIC ACID	N-terminal glutamate which has formed an internal cyclic lactam
SULFATATION	Generally of tyrosine
LIPID	Covalent binding of a lipidic moiety
MYRISTATE	Myristate group attached through an amide bond to the N-terminal glycine residue of the mature form of a protein or to an internal lysine residue

PALMITATE	Palmitate group attached through a thioether bond to a cysteine residue or through an ester bond to a serine or threonine residue
FARNESYL	Farnesyl group attached through a thioether bond to a cysteine residue
GERANYL-GERANYL	Geranyl-geranyl group attached through a thioether bond to a cysteine residue
GPI-ANCHOR	Glycosyl-phosphatidylinositol (GPI) group linked to the alpha-carboxyl group of the C-terminal residue of the mature form of a protein
N-ACYL DIGLYCERIDE	N-terminal cysteine of the mature form of a prokaryotic lipoprotein with an amide-linked fatty acid and a glyceryl group to which two fatty acids are linked by ester linkages
DISULFID	Disulfide bond. The 'FROM' and 'TO' endpoints represent the two residues which are linked by an intra-chain disulfide bond. If the 'FROM' and 'TO' endpoints are identical, the disulfide bond is an interchain one and the description field indicates the nature of the cross-link
THIOLEST	Thiolester bond. The 'FROM' and 'TO' endpoints represent the two residues which are linked by the thiolester bond
THIOETH	Thioether bond. The 'FROM' and 'TO' endpoints represent the two residues which are linked by the thioether bond
CARBOHYD	Glycosylation site. The nature of the carbohydrate (if known) is given in the description field
METAL	Binding site for a metal ion. The description field indicates the nature of the metal
BINDING	Binding site for any chemical group (co-enzyme, prosthetic group, etc.). The chemical nature of the group is given in the description field
SIGNAL	Extent of a signal sequence (prepeptide)
TRANSIT	Extent of a transit peptide (mitochondrial, chloroplastic, or for a microbody)
PROPEP	Extent of a propeptide
CHAIN	Extent of a polypeptide chain in the mature protein
PEPTIDE	Extent of a released active peptide
DOMAIN	Extent of a domain of interest on the sequence. The nature of that domain is given in the description field
CA_BIND	Extent of a calcium-binding region
DNA_BIND	Extent of a DNA-binding region
NP_BIND	Extent of a nucleotide phosphate binding region. The nature of the nucleotide phosphate is indicated in the description field
TRANSMEM	Extent of a transmembrane region
ZN_FING	Extent of a zinc finger region
SIMILAR	Extent of a similarity with another protein sequence. Precise information, relative to that sequence is given in the description field
REPEAT	Extent of an internal sequence repetition
HELIX	Secondary structure - Helices, e.g., Alpha-helix, 3(10) helix, or Pi-helix

STRAND	Secondary structure - Beta-strand, e.g., Hydrogen bonded beta-strand, or Residue in an isolated beta-bridge
TURN	Secondary structure - Turns, e.g., H-bonded turn (3-turn, 4-turn, or 5-turn)
ACT_SITE	Amino acid(s) involved in the activity of an enzyme
SITE	Any other interesting site on the sequence
INIT_MET	The sequence is known to start with an initiator methionine
NON_TER	The residue at an extremity of the sequence is not the terminal residue. If applied to position 1, this signifies that the first position is not the N-terminus of the complete molecule. If applied to the last position, it signifies that this position is not the C-terminus of the complete molecule. There is no description field for this key
NON_CONS	Non consecutive residues. Indicates that two residues in a sequence are not consecutive and that there are a number of unsequenced residues between them
UNSURE	Uncertainties in the sequence. Used to describe region(s) of a sequence for which the authors are unsure about the sequence assignment

BILLING CODE 3510–16–C

Further in paragraph (a) of § 1.821, both occurrences of "Copies of ST.23" have been changed to "Copies of WIPO Standard ST.25 (1998)." This change is necessary to reflect the new standard number.

In paragraph (a)(1) of § 1.821, "ST.23 (April 1994), paragraph 8" has been changed to "ST.25 (1998), Appendix 2, Table 1." This change reflects the correct information with regard to the incorporated WIPO standard and the list of symbols to be used for nucleotide sequence characters.

Further in paragraph (a)(1) of § 1.821, "ST.23 (April 1994), paragraph 9" has been changed to "ST.25 (1998), Appendix 2, Table 2." This change reflects the correct information with regard to the incorporated WIPO standard and the list of modified bases which can be presented as unmodified nucleotide sequence characters.

In paragraph (a)(2) of § 1.821, all three occurrences of "ST.23 (April 1994), paragraph 11" have been changed to "ST.25 (1998), Appendix 2, Table 3." This change reflects the correct information with regard to the incorporated WIPO standard and the list of symbols to be used for amino acid sequence characters.

Further in paragraph (a)(2) of § 1.821, "ST.23 (April 1994), paragraph 12" has been changed to "ST.25 (1998), Appendix 2, Table 4." This change reflects the correct information with regard to the incorporated WIPO standard and the list of modified or unusual amino acids which can be presented as unmodified amino acid sequence characters.

In paragraph (c) of § 1.821, each of the three occurrences of the words "integer identifier" or "integer identifiers" has been changed to "sequence identifier" or "sequence identifiers" as appropriate. WIPO Standard ST.25 (1998), uses the term "sequence identifier" rather than "integer identifier." Thus, this change is necessary to achieve harmonization with the international standard.

In the last sentence of paragraph (c) of § 1.821, the phrase "The sequence omitted shall appear following the integer identifier" of the proposed rule has been replaced by "the code '000' shall be used in place of the sequence." The response for the numeric identifier <160> shall include the total number of SEQ ID NOs, whether followed by a sequence or by the code "000". The code "000" should be put into <400>. This change permits flexibility in the preparation and amendment of Sequence Listings. It also makes the rule language-neutral and is consistent with WIPO Standard ST.25 (1998).

In paragraph (d) of § 1.821, the words "integer identifier" have been changed to "sequence identifier." WIPO Standard ST.25 (1998) uses the term "sequence identifier" rather than "integer identifier." Thus, this change is necessary to achieve harmonization with the international standard.

In paragraphs (f), (g) and (h) of § 1.821, the sentence "Such a statement must be a verified statement if made by a person not registered to practice before the Office" has been deleted. The separate verification requirements in § 1.821 have been eliminated in view of the recent amendment to §§ 1.4(d) and 10.18. *See Changes to Patent Practice and Procedure;* Final Rule, 62 FR. 53131 (October 10, 1997), 1203 *Off. Gaz. Pat. Office* 63 (October 21, 1997). Paragraph (g) of § 1.821 has also been amended to provide that the Office will provide a "period of time" (rather than one month) within which the applicant must comply with the requirements of § 1.821(b) through (f) in order to avoid abandonment.

Further in paragraph (f) of § 1.821, the following has been added at the end of the first sentence, ", *e.g.,* the information recorded in computer readable form is identical to the written sequence listing." WIPO Standard ST.25 (1998), paragraph 39, requires the language which has been added as an acceptable example for phrasing the required statement that the computer readable form and the written sequence listing are the same.

Section 1.822

In paragraph (b) of § 1.822, both references to WIPO Standard ST.23 (April 1994), paragraphs 8 and 11, as proposed have been changed to "WIPO Standard ST.25 (1998), Appendix 2, Tables 1 and 3." These changes reflect the correct information with regard to the incorporated WIPO standard and the lists of symbols for nucleotide and amino acid sequence characters.

Further in paragraph (b) of § 1.822, "WIPO Standard ST.23 (April 1994), paragraphs 9 and 12" as proposed has been changed to "WIPO Standard ST.25 (1998), Appendix 2, Tables 2 and 4."

This change reflects the correct information with regard to the incorporated WIPO standard and the lists of modified bases and modified or unusual amino acids which can be depicted in the Sequence Listing via the symbols for a corresponding unmodified base or amino acid.

Further in paragraph (b) of § 1.822, the symbol designating an unknown nucleotide base or a nucleotide base other than those listed in the WIPO standard was proposed as an upper case letter "N." This symbol has been changed to a lower case letter "n." This change is consistent with the use of lower case letters for the symbols representing the nucleotide bases.

Further in paragraph (b) of § 1.822, the language has been clarified to specifically state that each "n" or "Xaa" represents only a single residue. Thus, for example, a single "Xaa" may not be used to designate a string of four amino acids, each of which is unknown. This represents a codification of existing practice.

Further in paragraph (b) of § 1.822, the information required in the Feature section to explain the use of "n" or "Xaa" in a given sequence is referred to "as appropriate." Additional instruction is added at the end of paragraph (b) of § 1.822 following "the Feature section" indicating", preferably by including one or more feature keys listed in WIPO Standard ST.25 (1998), Appendix 2, Tables 5 and 6." This change specifies the preference for using the feature keys listed in the WIPO standard in order to aid applicants in filing a CRF which will comply with WIPO Standard ST.25 (1998). These feature keys are controlled vocabulary and are considered language neutral. Their use is required in a PCT patent application or a patent application in a foreign country which has adopted WIPO Standard ST.25 (1998).

In paragraph (c)(1) of § 1.822, "WIPO Standard ST.23 (April 1994), paragraph 8" as proposed has been changed to "WIPO Standard ST.25 (1998), Appendix 2, Table 1." This change reflects the correct information with regard to the incorporated WIPO standard and the list of symbols to be used for nucleotide sequence characters.

In paragraph (d)(1) of § 1.822, "WIPO Standard ST.23 (April 1994), paragraph 11, as proposed has been changed to "WIPO Standard ST.25 (1998), Appendix 2, Table 3." This change reflects the correct information with regard to the incorporated WIPO standard and the list of symbols to be used for amino acid sequence characters.

In paragraph (d)(4) of § 1.822, the section notes that enumeration requirements are applicable to amino acid sequences that are circular in configuration. The following language has been added to the end of the paragraph ", with the exception that the designation of the first amino acid of the sequence may be made at the option of the applicant." This change is necessary to provide consistency with its counterpart of circular nucleotide sequences as provided in paragraph (c)(7) of § 1.822. This change is also consistent with WIPO Standard ST.25 (1998), paragraph 21.

In paragraph (e) of § 1.822, the words "integer identifiers" have been changed to "sequence identifiers ." WIPO Standard ST.25 (1998) uses the term "sequence identifier" rather than "integer identifier." Thus, this change is necessary to achieve harmonization with the international standard.

Section 1.823

In paragraph (a) of § 1.823, the entire second sentence which read "On a separate page of the application specification, immediately prior to the claims, there shall be a reference to the presence of the 'Sequence Listing' in a 'Sequence Listing Annex.'" has been eliminated. The designation of the Sequence Listing as an annex to the specification was initially proposed in an early version of the international standard. This terminology is not used in WIPO Standard ST.25 (1998), however, and so it has also been eliminated from paragraph (a) of § 1.823, as proposed. Simplification results as well by the elimination of the requirement that the Sequence Listing must be designated as an annex to the specification.

In paragraph (a) of § 1.823, the third sentence has been modified by deleting the words "shall appear in the 'Sequence Listing Annex,' which is." As explained above, the current version of the international standard does not require designating the Sequence Listing as an annex to the specification.

In paragraph (a) of § 1.823, the words "preferably should be" have been added to the third sentence, before "numbered independently of the numbering of the remainder of the application" to describe the independent page numbering of the Sequence Listing in paper copy form. The term "preferably" was added for purposes of harmonization with WIPO Standard ST.25 (1998).

In paragraph (a) of § 1.823, the last clause of the third sentence "and shall be placed in the application file" has been deleted as unnecessary and potentially confusing now that the reference to a "Sequence Listing Annex" has been removed from this paragraph.

In paragraph (a) of § 1.823, the fourth sentence has been eliminated in its entirety. As explained above, the current version of the international standard does not require designating the Sequence Listing as an annex to the specification.

In paragraph (a) of § 1.823, in both occurrences in the fifth sentence and in the single occurrence in the sixth sentence, the word "shall" has been changed to "should." These changes are necessary for purposes of achieving consistency with WIPO Standard ST.25 (1998).

In paragraph (b) of § 1.823, the first sentence has been modified by the deletion of the words "in addition to and immediately preceding." This change is consistent with WIPO Standard ST.25 (1998).

In paragraph (b) of § 1.823, the fifth sentence has been deleted, eliminating the prohibition of any item of information occupying more than one line. This change is consistent with WIPO Standard ST.25 (1998).

In paragraph (b) of § 1.823, the last sentence has been deleted to eliminate the "rep" designation for data elements of the "Sequence Listing." Certain data elements may still be repeated within the listing but this change was made for harmonization of the table with WIPO Standard ST.25 (1998).

In paragraph (b) of § 1.823, the eighth sentence has been modified to reflect the new numeric numbering scheme, for harmonization with WIPO Standard ST.25 (1998). Specifically, "<100> through <193>" of the proposed rule has been changed to "<110> through <170>."

The table in paragraph (b) of § 1.823, has been changed to reflect the revised numbering scheme and data elements used in WIPO Standard ST.25 (1998). The specific changes are as follows:

Numeric identifier "<100>, General Information," has been deleted from the proposed rules, as it is not present in WIPO Standard ST.25 (1998).

Numeric identifier "<110>, Applicant," in the proposed rule, has been changed to indicate that "preferably " a maximum of ten names may be indicated. This change allows for more than ten names in the Applicant field for those instances in which such would be appropriate. This change is consistent with WIPO Standard ST.25 (1998).

Numeric identifier "<120>, Title of Invention," in the proposed rule, has been changed to eliminate the limitation

that the title be a maximum of four lines. This change allows applicants more flexibility with respect to the title. This change is consistent with WIPO Standard ST.25 (1998).

Numeric identifier "<130>, Number of Sequences," in the proposed rule, has been changed to reflect "<130>, File Reference," as stated in WIPO Standard ST.25 (1998). This numeric identifier was indicated as "<183>, File Reference/Docket Number ", in the rule as proposed. As proposed this was an optional numeric identifier. The numeric identifier remains optional once the application has been assigned an application number, *e.g.,* a serial number. This numeric identifier is now MANDATORY when an application number has not yet been assigned to the application, such as on the day the application is initially filed. This change will assist in the matching of sequence information submissions with an application in the event that either the paper copy or the computer readable form were to become separated from the remainder of the application. This change is consistent with WIPO Standard ST.25 (1998).

The Number of Sequences field identified as "<130>" in the proposed rule is now numbered "<160>" in § 1.823 as adopted and redefined as "Number of SEQ ID NOs."

The information associated with numeric identifiers "<140>" through "<153>," "Correspondence Address" through "Operating System" of the proposed rule, has been eliminated to reduce the burden on the applicant and to harmonize with WIPO Standard ST.25 (1998). Some of these numeric identifiers have been used in the new numbering scheme and have been associated with different information as indicated herein and in the Table of § 1.823.

One remaining numeric identifier within the Computer Readable Form section, "<154>, Software," of the proposed rule, will remain, with the exception that it has been reassigned the numeric identifier of "<170>" to reflect the numbering scheme presented in WIPO Standard ST.25 (1998).

The main headings "<160>, Current Application Data" and "<170>, Prior Application Data," of the proposed rules, have been eliminated to harmonize with WIPO Standard ST.25 (1998) and reduce the number of fields in the Sequence Listing. The information that was to appear under these main headings remains in the rules but has been reassigned numeric identifiers <140> through <151>. The specific changes are as follows: "<160>" has been redefined as "Number of SEQ

ID NOs "; "<161>, Application Number," of the proposed rule is now numbered as "<140>," and is defined as "Current Application Number'; "<162>, Filing Date," of the proposed rule is now numbered "<141>," and is defined as "Current Filing Date"; "<170>" has been redefined as "Software "; "<171>, Application Number," of the proposed rule is now numbered as "<150>," and is defined as "Prior Application Number"; "<172>, Filing Date," of the proposed rule is now numbered as "<151>," and is defined as "Prior Application Filing Date."

The numeric identifiers now numbered "<150>, Prior Application Number,", and "<151>, Prior Application Filing Date," are now mandatory only in those instances in which a claim for priority with respect to those prior applications is being made under either 35 U.S.C. 119 or 120. This change will provide information in this regard when it is most useful and was necessary to harmonize these rules with WIPO Standard ST.25 (1998). Throughout the Sequence Listing, application numbers must be set forth as a combination of the two digit country code, as set forth in WIPO Standard ST.3, as well as an application number in accordance with WIPO Standard ST.13 or for an international application, the numbering system as set out in Section 307(a) of the Administrative Instructions under the PCT.

Numeric identifiers "<180>, Attorney/Agent Information," through "<182>, Registration Number," of the proposed rule, have been eliminated to harmonize with WIPO Standard ST.25 (1998) and reduce the number of fields in the Sequence Listing.

Numeric identifier "<183>, File Reference/Docket Number" of the proposed rule has been reassigned as numeric identifier "<130>," and redefined as "File Reference" in an effort to harmonize with WIPO Standard ST.25 (1998).

The Telecommunication Information section, "<190>" through "<193>" of the proposed rules, has been eliminated in order to reduce the number of fields in the Sequence Listing and harmonize with WIPO Standard ST.25 (1998).

Numeric identifier "<200>, Information for SEQ ID NO:#:", has been reassigned the numeric identifier "<210>, SEQ ID NO: #:" This numeric identifier indicates the integer, referred to in these final rules as the sequence identifier for both the sequence information and the actual sequence which follows the information.

Numeric identifier "<210>, Sequence Characteristics," of the proposed rule

has been eliminated in order to reduce the number of required elements in the Sequence Listing and harmonize with WIPO Standard ST.25 (1998).

The valid responses for the mandatory numeric identifier "<212>, Type," have been changed from "N" and "A", as stated in the proposed rule, to "DNA," "RNA," and "PRT" (protein) in order to harmonize with WIPO Standard ST.25 (1998). A compound that is a mixture of DNA and RNA should be represented by "DNA." This change is consistent with WIPO Standard ST.25 (1998).

Numeric identifier "<213>, Organism," has been added to the Sequence Listing of these final rules in an effort to harmonize with WIPO Standard ST.25 (1998). A response for the Organism identifier is MANDATORY. The valid responses are the scientific name, i.e. "Genus species", "Artificial Sequence", or "Unknown."

Numeric identifier "<214>, Topology," of the proposed rule, has been eliminated to harmonize with WIPO Standard ST.25 (1998), and to reduce the burden on the applicant.

Numeric identifier "<290>, Feature," has become numeric identifier "<220>, Feature." This numeric identifier has become MANDATORY for those sequences in which numeric identifier "<213>, Organism," is completed with either "Artificial Sequence" or "Unknown." This numeric identifier is also required if the compound sequence is a mixture of DNA and RNA. Numeric identifier "<220>, Feature" is a header only. No data are added immediately following this numeric identifier. These changes are required to achieve harmonization with WIPO Standard ST.25 (1998).

Numeric identifier "<291>, Name/ Key," has become numeric identifier "<221>, Name/Key." As proposed, the information provided was restricted to a maximum of four lines. The four line restriction has been removed to reduce the limitations on this field. The comment section of this numeric identifier has been changed in that it now indicates that the selection of a feature name or feature key is preferably made from those listed in Tables 5 and 6 of WIPO Standard ST.25 (1998). These tables are reproduced above and this preference for the listed feature names and keys is consistent with the requirement of WIPO Standard ST.25 (1998).

Numeric identifier "<292>, Location," has become "<222>, Location," so as to be consistent with the numeric identifiers contained in WIPO Standard ST.25 (1998).

Numeric identifier "<294>, Other Information," has become numeric identifier "<223>, Other Information," so as to be consistent with the numeric identifiers contained in WIPO Standard ST.25 (1998). This numeric identifier has become MANDATORY for those sequences in which numeric identifier "<213>, Organism," is completed with either "Artificial Sequence" or "Unknown". Numeric identifier "<223>, Other Information," should contain source information in those instances when the organism is unknown or is an artificial sequence. For example, the source may be unknown because the material was isolated from a mixed bacterial culture rather than a pure culture. In such a case, numeric identifier "<223>, Other Information," should be completed by explaining the mixed culture source of the sequenced material. If a sequence is completely synthesized this should be indicated in numeric identifier "<223>, Other Information," while numeric identifier "<213>, Organism," would indicate "Artificial Sequence." This change has been made to accomplish harmonization between these rules and WIPO Standard ST.25 (1998) which contains the same mandatory requirement in this regard.

Numeric identifiers "<308>" through "<310>," referring to the " Patent Document Number," "Filing Date" and " Publication Date," of the proposed rule, have been moved to numeric identifiers "<310>" to "<312>," respectively, of this Final Rule in order to harmonize with the numeric numbering scheme of WIPO Standard ST.25 (1998). Citations in the Sequence Listing must comply with WIPO Standard ST.6 for publication numbers and WIPO Standard ST.16 for document codes.

New numeric identifiers "<308>, Database Accession Number," and "<309> Database Entry Date," have been added to the final rules to harmonize with WIPO Standard ST.25 (1998). These fields were added to the publication information section of WIPO Standard ST.25 (1998) to give an applicant more opportunity to further identify a published citation.

Numeric identifier <400> "Sequence Description: SEQ ID NO:#:" has been changed to "Sequence " for clarity. Also for clarity, the explanation in the table has been changed to "SEQ ID NO shall follow the numeric identifier and should appear on the line preceding the sequence."

The format of the date fields has been changed throughout the Sequence Listing to accommodate for international conventions. All date fields referenced in the Sequence Listing shall conform to WIPO Standard ST.2. Because compliance with §§ 1.821 through 1.825 as amended should produce Sequence Listings that are acceptable to all receiving offices, a standardized date field convention was required.

Section 1.824

In paragraph (a)(6) of § 1.824, ", the date on which the data were recorded on the computer readable form" was added after "title of the invention" to harmonize with WIPO Standard ST.25 (1998) requirements. While this requirement of § 1.824 was proposed to be eliminated, that proposal is not adopted for purposes of harmonization with WIPO Standard ST.25 (1998). Also in paragraph (a)(6) of § 1.824, " name and type of computer and" was deleted to reduce the requirements.

Section 1.825

In paragraphs (a), (b), and (d) of § 1.825, the sentence "Such a statement must be a verified statement if made by a person not registered to practice before the Office" has been deleted. The separate verification requirements of § 1.825 have been eliminated in view of the recent amendment to §§ 1.4(d) and 10.18. *See Changes to Patent Practice and Procedure*; Final Rule, 62 FR. 53131 (October 10, 1997), 1203 *Off. Gaz. Pat. Office* 63 (October 21, 1997).

Response to and Analysis of Comments

Six written comments were received in response to the Notice of Proposed Rulemaking. Several of these comments address the three specific queries set forth in the Notice of Proposed Rulemaking.

The first query posed in the Notice of Proposed Rulemaking was: (1) Should the PTO accept voluntary submissions of computer readable forms and Sequence Listings where a D-amino acid is contained in the sequence? If such voluntary submissions are accepted, should there be a restriction on the choice of identifying a D-amino acid by an Xaa or by its L-amino acid counterpart abbreviation?

Comment: One comment indicated that not only should the PTO accept voluntary submissions under these rules where a D-amino acid is contained in the sequence, the Office should make such submissions mandatory and designated by an Xaa. One comment indicated that sequences containing D-amino acids should not be in the PTO databases.

Response: Upon careful consideration, the PTO has decided to accept voluntary submissions of protein sequences containing D-amino acids. The PTO strongly encourages anyone making such voluntary submissions to identify a D-amino acid with an Xaa, describing the D-amino acid in the Features section of the Sequence Listing. This section is indicated by numeric identifiers <220> through <223> in 37 CFR 1.823. Procedural concerns compel this acceptance of voluntary submissions. Computer readable forms are processed prior to examination. It is cumbersome to establish a viable procedure to redact any voluntary submissions out of the PTO database. The use of Xaa to indicate a D-amino acid, should such sequence information be submitted in accordance with these rules, is encouraged so as to alert anyone reviewing the sequence that a particular amino acid is other than a naturally occurring L-amino acid and to more accurately depict the extent of similarities between such a sequence and the L-amino acid containing sequences present in a database being searched for examination or other purposes.

Because the sequence databases do not currently include D-amino acids in sequences and thus are not searchable for such sequences, the submission of those sequences containing D-amino acids will not be made mandatory.

The second query posed in the proposed rules was: (2) Should the provisions of 37 CFR 1.821(c) be altered to exclude some prior art sequences from inclusion in the Sequence Listing even though they are presented in a patent application disclosure as sequences? Should the reference to an accession number of an admitted prior art sequence in a publicly available, electronic, sequence database suffice and exclude that sequence from the requirements of the sequence rules?

Comment: Four comments indicated that known "prior art" sequences should not be required in the Sequence Listing. A referral to a publicly available, electronic, sequence database for access to such "prior art" sequences would be an acceptable alternative to two of those commenting on this aspect; the other two did not address this point. The reasons given for excluding such sequences are the expense and time required by applicants and their representatives in the inclusion of "prior art" sequences that are considered to be "non-inventive". Reducing the bulk of the paper copy of the Sequence Listing was also mentioned.

Response: The requirement to submit all disclosed sequences in the format required by §§ 1.821 through 1.825 is

maintained. This point was discussed with officials from the JPO and EPO. The offices have considered the stated concerns with regard to costs to applicants. Sections 1.821 through 1.825 do not require any information to be disclosed in the form of a sequence, but rather require a particular format whenever information is presented in the form of a sequence. Those applicants for whom compliance with the rules remains a significant hardship may petition under § 1.183 for a waiver of the applicable requirement of §§ 1.821 through 1.825.

The technical and legal concerns mentioned in the Notice of Proposed Rulemaking still exist concerning the use of an alternative reference to a publicly available, electronic, sequence database. These concerns are: (1) What constitutes a publicly available, electronic, sequence database? (2) Would the USPTO and the other patent offices which have similar rules be required to produce a list of internationally accepted databases? (3) What would be the criteria for such acceptance? (4) An additional issue would exist involving electronic records maintenance: is there any assurance that once information is contained in a database that it will be retained and available indefinitely without alteration? Changes to the information in nucleic acid sequence databases resulting from the discovery of sequencing errors are well-known.

(5) Does the mere existence of the sequence information in such a record constitute reasonable means of retrieval? In other words, would one need some text basis or other identifier to retrieve the information?

Additional reasons for the inclusion of these prior art sequences remain relevant. These reasons are: (1) the assessment of whether a particular sequence falls within the requirements of the current rules is simple; (2) the general public is assured that all patents which contain any sequence information contain all of the sequence information in the Sequence Listing and all sequences are available in a computer accessible form; and (3) as a publication, the contextual association of new and old information is potentially unique to the patent and very valuable to anyone assessing the state of the art at the time of a patented invention, and thus are desirable to be present in electronic form in association with that patent.

The third query posed in the proposed rules was: (3) Should Sequence Listings filed in an international application filed under the PCT be published only electronically and made available for retrieval electronically by an accession number from several sequence repositories?

Comment: Two comments were received in response to this query, one in favor and one opposed to limiting the publication of the Sequence Listing to an electronic form for published PCT applications in the international phase.

Response: At this time paper copies of the Sequence Listings filed as part of the description will continue to be published in applications filed under PCT. The PTO together with the EPO, JPO and WIPO will continue to discuss the possibility of electronic publication. However, any implementation of such electronic publication in lieu of publication in paper form will not be undertaken until further study has been completed.

Comment: One comment suggested that informative English words be placed next to the numerical headings in the Sequence Listing as printed in a U.S. patent.

Response: The PTO will provide English words corresponding to the numeric identifiers in the printed U.S. patents.

Comment: One comment suggested addition of a descriptive comment line to the Sequence Listing.

Response: The ''Other Information'' line in the Features section, which is numeric identifier <223> in § 1.823, provides for a description of a sequence. While completion of this section is only mandatory when the sequence contains ''n'', ''Xaa'', a modified or unusual L-amino acid or a modified base, it is frequently completed in other circumstances.

Comment: One comment requested we harmonize §§ 1.821 through 1.825 with PCT, EPO and other authorities such that the differences in the requirements for Sequence Listing submissions are minimal.

Response: This change to §§ 1.821 through 1.825 is the result of such an effort to harmonize the PTO, PCT, EPO and JPO Sequence Listing requirements to the extent possible. The requirements of newly developed WIPO ST.25 are substantially identical to the requirements of amended §§ 1.821 through 1.825. PatentIn Version 2.0 software, now available, is drafted to meet all of the requirements of WIPO Standard ST.25 (1998). The requirements of §§ 1.821 through 1.825, however, are less stringent than the requirements of WIPO Standard ST.25 (1998). Thus, applicants who wish to file in countries which adhere to WIPO Standard ST.25 (1998) should consider the following when not using PatentIn Version 2.0:

1. The WIPO Standard ST.25 (1998) does not permit submissions using a Macintosh computer.

2. The WIPO Standard ST.25 (1998) does not accept the range of media permitted by amended §§ 1.821 through 1.825.

3. The answers in field <221> and <222> must use selections from Tables 5 and 6 of WIPO Standard ST.25 (1998) to comply with that standard. The terms from these Tables are considered language neutral vocabulary.

4. Any free text in numeric identifier <223> of a Sequence Listing will not be translated and thus must also appear in the specification of applications filed under WIPO Standard ST.25 (1998) for compliance.

5. A CRF filed after the filing of an application under the PCT does not form part of the disclosure and will not be published in the pamphlet.

6. Paragraph 39 of WIPO Standard ST.25 (1998) requires the specific wording ''the information recorded on the form is identical to the written sequence listing.''

7. WIPO Standard ST.25 (1998), paragraph 24, requires spaces between specified numeric identifiers in the Sequence Listing.

Comment: One comment requested a WINDOWS® based version of PatentIn.

Response: A WINDOWS® based version of PatentIn, PatentIn 2.0, has been developed through a Trilaterally-sponsored joint initiative and is being made available.

Comment: One comment expressed concern over application of the doctrine of equivalents by the courts to sequence-based claim language.

Response: Sections 1.821 through 1.825 do not establish a disclosure requirement, nor do they alter the requirements of 35 U.S.C. § 112. They merely require a particular format whenever information is presented in the form of a sequence. The use of sequence identification numbers (SEQ ID NO: #) only provides a shorthand way for applicants to refer to sequence information. These identification numbers do not in any way restrict the manner in which an invention can be claimed. Similarly, the use of this format does not impact the potential interpretations and legal determinations that could be made with respect to claims containing information in the form of a nucleotide or amino acid sequence.

Comment: One comment requested the flexibility to use single-letter amino acid codes.

Response: Sections 1.821 through 1.825 as amended do not constrain an applicant from using single letter codes

in the disclosure. The requirements of the sequence searching and the sequence storage mechanisms include only the three-letter codes, thus the need for the constraint on the Sequence Listing information. There is no such restriction on the sequence format in the body of the disclosure or in the figures imposed by §§ 1.821 through 1.825, or any of the rules of practice; only the format for the Sequence Listing is specified by §§ 1.821 through 1.825.

Review Under the Paperwork Reduction Act of 1995

Notwithstanding any other provision of law, no person is required to respond to nor shall a person be subject to a penalty for failure to comply with a collection of information subject to the requirements of the Paperwork Reduction Act (PRA) unless that collection of information displays a currently valid OMB control number.

This rule contains collections of information requirements subject to the PRA. The principal impact of this Final Rule is: (1) Elimination of certain requirements of §§ 1.821 through 1.825; and (2) revision of §§ 1.821 through 1.825 for consistency with WIPO Standard ST.25 (1998), which will permit Sequence Listings to be presented in an international, language neutral format.

The public reporting burden for these collections of information have been approved by the Office of Management and Budget (OMB) under OMB control number 0651–0024. The public reporting burden for this collection of information is estimated to average 80 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the information. Send comments regarding this burden estimate or any other aspect of the data requirements, including suggestions for reducing this burden, to Esther M. Kepplinger at the address specified above or to the Office of Information and Regulatory Affairs of OMB, New Executive Office Bldg., 725 17th St. NW, rm. 10235, Washington, DC 20230, Attn: Desk Officer for the Patent and Trademark Office.

Other Considerations

This Final Rule is in conformity with the requirements of the Regulatory Flexibility Act (5 U.S.C. 601 *et seq.*), Executive Order 12612 (October 26, 1987), and the Paperwork Reduction Act of 1995 (44 U.S.C. 3501 *et seq.*). It has been determined that this rulemaking is not significant for the purposes of Executive Order 12866 (September 30, 1993).

The Assistant General Counsel for Legislation and Regulation of the Department of Commerce has certified to the Chief Counsel for Advocacy, Small Business Administration that this Final Rule would not have a significant impact on a substantial number of small entities (Regulatory Flexibility Act, 5 U.S.C. 605(b)). The principal impact of this Final Rule is: (1) Elimination of certain requirements of §§ 1.821 through 1.825; and (2) revision of §§ 1.821 through 1.825 for consistency with WIPO Standard ST.25 (1998), which will permit Sequence Listings to be presented in an international, language neutral format.

The Office has determined that this Final Rule has no Federalism implications affecting the relationship between the National Government and the States as outlined in Executive Order 12612.

List of Subjects in 37 CFR Part 1

Administrative practice and procedure, Courts, Freedom of information, Inventions and patents, Incorporation by reference, Reporting and recordkeeping requirements, Small businesses.

For the reasons set forth in the preamble and under the authority granted to the Commissioner of Patents and Trademarks by 35 U.S.C. 6, Title 37 of the Code of Federal Regulations, part 1, is amended as follows:

PART 1—RULES OF PRACTICE IN PATENT CASES

1. The authority citation for 37 CFR part 1 continues to read as follows:

Authority: 35 U.S.C. 6, unless otherwise noted.

2. Section 1.821 is revised to read as follows:

§ 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 four or more amino acids or an unbranched sequence of ten 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 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 (1998), including Tables 1 through 6 in Appendix 2, herein incorporated by reference. (Hereinafter "WIPO Standard ST.25 (1998)"). 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 WIPO Standard ST.25 (1998) may be obtained from the World Intellectual Property Organization; 34 chemin des Colombettes; 1211 Geneva 20 Switzerland. Copies of ST.25 may be inspected at the Patent Search Room; Crystal Plaza 3, Lobby Level; 2021 South Clark Place; Arlington, VA 22202. Copies may also be inspected at the Office of the Federal Register, 800 North Capitol Street, NW, Suite 700, Washington, DC. 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 WIPO Standard ST.25 (1998), Appendix 2, Table 1. Modifications, *e.g.,* methylated bases, may be described as set forth in WIPO Standard ST.25 (1998), Appendix 2, Table 2, 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 WIPO Standard ST.25 (1998), Appendix 2, Table 3. 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 WIPO Standard ST.25 (1998), Appendix 2, Table 3 with the modified positions; *e.g.,* hydroxylations or glycosylations, being described as set forth in WIPO Standard ST.25 (1998), Appendix 2, Table 4, 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 WIPO Standard ST.25 (1998), Appendix 2, Table 3 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.

(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 which contain disclosures of nucleotide and/or amino acid sequences must contain, as a separate part of the disclosure, a paper copy disclosing the 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. This paper copy is hereinafter referred to as the ''Sequence Listing.'' Each sequence disclosed must appear separately in the ''Sequence Listing.'' Each sequence set forth in the ''Sequence Listing'' shall be assigned a separate sequence identifier. The sequence identifiers shall begin with 1 and increase sequentially by integers. If no sequence is present for a sequence identifier, the code ''000'' shall be used in place of the sequence. The response for the numeric identifier <160> shall include the total number of SEQ ID NOs, whether followed by a sequence or by the code ''000.''

(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 the sequence identifier, preceded by ''SEQ ID NO:'' 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.

(e) A copy of the ''Sequence Listing'' referred to in paragraph (c) of this section must also be submitted in computer readable form in accordance with the requirements of § 1.824. The computer readable form is a copy of the ''Sequence Listing'' and will not necessarily be retained as a part of the patent application file. If the computer readable form of a new application is to be identical with the computer readable form of another application of the applicant on file in the Patent and Trademark Office, reference may be made to the other application and computer readable form in lieu of filing a duplicate computer readable form in the new application if the computer readable form in the other application was compliant with all of the requirements of these rules. The new application shall be accompanied by a letter making such reference to the other application and computer readable form, both of which shall be completely identified. In the new application, applicant must also request the use of the compliant computer readable ''Sequence Listing'' that is already on file for the other application and must state that the paper copy of the

''Sequence Listing'' in the new application is identical to the computer readable copy filed for the other application.

(f) In addition to the paper copy required by paragraph (c) of this section and the computer readable form required by paragraph (e) of this section, a statement that the content of the paper and computer readable copies are the same must be submitted with the computer readable form, *e.g.,* a statement that ''the information recorded in computer readable form is identical to the written sequence listing.''

(g) If any of the requirements of paragraphs (b) through (f) 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, 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 submission in reply to a requirement under this paragraph must be accompanied by a statement that the submission includes no new matter.

(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.

3. Section 1.822 is revised to read as follows:

§ 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 of ST.25 may be inspected at the Patent Search Room; Crystal Plaza 3, Lobby Level; 2021 South Clark Place; Arlington, VA 22202. Copies may also be inspected at the Office of the Federal Register, 800 North Capitol Street, NW, Suite 700, Washington, DC. 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.

4. Section 1.823 is revised to read as follows:

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

(a) The "Sequence Listing" required by § 1.821(c), setting forth the nucleotide and/or amino acid sequences and associated information in accordance with paragraph (b) of this section, must begin on a new page and must be titled "Sequence Listing". The "Sequence Listing" preferably should be numbered independently of the numbering of the remainder of the application. Each page of the "Sequence Listing" should contain no more than 66 lines and each line should contain no more than 72 characters. A fixed-width font should be used exclusively throughout the "Sequence Listing."

(b) The "Sequence Listing" shall, except as otherwise indicated, include the actual nucleotide and/or amino acid sequence, the numeric identifiers and their accompanying information as shown in the following table. The numeric identifier shall be used only in the "Sequence Listing." The order and presentation of the items of information in the "Sequence Listing" shall conform to the arrangement given below. Each item of information shall begin on a new line and shall begin with the numeric identifier enclosed in angle brackets as shown. The submission of those items of information designated with an "M" is mandatory. The submission of those items of information designated with an "O" is optional. Numeric identifiers <110> through <170> shall only be set forth at the beginning of the "Sequence Listing." The following table illustrates the numeric identifiers.

Numeric identifier	Definition	Comments and format	Mandatory (M) or optional (O).
<110>	Applicant	Preferably max. of 10 names; one name per line; preferable format: Surname, Other Names and/ or Initials.	M.
<120>	Title of Invention	....	M.
<130>	File Reference	Personal file reference	M when filed prior to assignment of appl. number.
<140>	Current Application Number.	Specify as: US 07/999,999 or PCT/US96/99999	M, if available.
<141>	Current Filing Date	Specify as: yyyy–mm–dd	M, if available.
<150>	Prior Application Number.	Specify as: US 07/999,999 or PCT/US96/99999	M, if applicable include priority documents under 35 USC 119 and 120.
<151>	Prior Application Filing Date.	Specify as: yyyy–mm–dd	M, if applicable.
<160>	Number of SEQ ID NOs	Count includes total number of SEQ ID NOs	M.
<170>	Software	Name of software used to create the Sequence Listing.	O.
<210>	SEQ ID NO:#:	Response shall be an integer representing the SEQ ID NO shown.	M.
<211>	Length	Respond with an integer expressing the number of bases or amino acid residues.	M.

Numeric identifier	Definition	Comments and format	Mandatory (M) or optional (O).
<212>	Type	Whether presented sequence molecule is DNA, RNA, or PRT (protein). If a nucleotide sequence contains both DNA and RNA fragments, the type shall be "DNA." In addition, the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section.	M.
<213>	Organism	Scientific name, i.e. Genus/ species, Unknown or Artificial Sequence. In addition, the "Unknown" or "Artificial Sequence" organisms shall be further described in the <220> to <223> feature section.	M
<220>	Feature	Leave blank after <220>. <221–223> provide for a description of points of biological significance in the sequence..	M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is "Artificial Sequence" or "Unknown'; if molecule is combined DNA/RNA"
<221>	Name/Key	Provide appropriate identifier for feature, preferably from WIPO Standard ST.25 (1998), Appendix 2, Tables 5 and 6.	M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence.
<222>	Location	Specify location within sequence; where appropriate state number of first and last bases/ amino acids in feature.	M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence.
<223>	Other Information	Other relevant information; four lines maximum ...	M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is "Artificial Sequence" or "Unknown"; if molecule is combined DNA/RNA.
<300>	Publication Information ..	Leave blank after <300>	O.
<301>	Authors	Preferably max of ten named authors of publication; specify one name per line; preferable format: Surname, Other Names and/or Initials.	O.
<302>	Title	....	O.
<303>	Journal	....	O.
<304>	Volume	....	O.
<305>	Issue	....	O.
<306>	Pages	....	O.
<307>	Date	Journal date on which data published; specify as yyyy–mm–dd, MMM–yyyy or Season-yyyy.	O.
<308>	Database Accession Number.	Accession number assigned by database including database name.	O.
<309>	Database Entry Date	Date of entry in database; specify as yyyy–mm–dd or MMM–yyyy.	O.
<310>	Patent Document Number.	Document number; for patent-type citations only. Specify as, for example, US 07/999,999.	O.
<311>	Patent Filing Date	Document filing date, for patent-type citations only; specify as yyyy–mm–dd.	O.
<312>	Publication Date	Document publication date, for patent-type citations only; specify as yyyy–mm–dd.	O.
<313>	Relevant Residues	FROM (position) TO (position)	O.
<400>	Sequence	SEQ ID NO should follow the numeric identifier and should appear on the line preceding the actual sequence.	M.

5. Section 1.824 is revised to read as follows:

§ 1.824   Form and format for nucleotide and/or amino acid sequence submissions in computer readable form.

(a) The computer readable form required by § 1.821(e) shall meet the following specifications:

(1) The computer readable form shall contain a single "Sequence Listing" as either a diskette, series of diskettes, or other permissible media outlined in paragraph (c) of this section.

(2) The "Sequence Listing" in paragraph (a) (l) of this section shall be submitted in American Standard Code for Information Interchange (ASCII) text. No other formats shall be allowed.

(3) The computer readable form may be created by any means, such as word processors, nucleotide/amino acid sequence editors or other custom computer programs; however, it shall conform to all specifications detailed in this section.

(4) File compression is acceptable when using diskette media, so long as the compressed file is in a self-extracting format that will decompress on one of the systems described in paragraph (b) of this section.

(5) Page numbering shall not appear within the computer readable form version of the "Sequence Listing" file.

(6) All computer readable forms shall have a label permanently affixed thereto on which has been hand-printed or typed: the name of the applicant, the title of the invention, the date on which the data were recorded on the computer readable form, the operating system used, a reference number, and an application serial number and filing date, if known.

(b) Computer readable form submissions must meet these format requirements:

(1) Computer: IBM PC/XT/AT, or compatibles, or Apple Macintosh;

(2) Operating System: MS–DOS, Unix or Macintosh;

(3) Line Terminator: ASCII Carriage Return plus ASCII Line Feed;

(4) Pagination: Continuous file (no "hard page break" codes permitted);

(c) Computer readable form files submitted may be in any of the following media:

(1) Diskette : 3.50 inch, 1.44 Mb storage; 3.50 inch, 720 Kb storage; 5.25 inch, 1.2 Mb storage; 5.25 inch, 360 Kb storage.

(2) Magnetic tape: 0.5 inch, up to 24000 feet; Density: 1600 or 6250 bits per inch, 9 track; Format: Unix tar command; specify blocking factor (not "block size"); Line Terminator: ASCII Carriage Return plus ASCII Line Feed.

(3) 8mm Data Cartridge: Format: Unix tar command; specify blocking factor (not "block size"); Line Terminator: ASCII Carriage Return plus ASCII Line Feed.

(4) CD–ROM: Format: ISO 9660 or High Sierra Format

(5) Magneto Optical Disk: Size/ Storage Specifications: 5.25 inch, 640 Mb.

(d) Computer readable forms that are submitted to the Office will not be returned to the applicant.

6. Section 1.825 is revised to read as follows:

§ 1.825 Amendments to or replacement of sequence listing and computer readable copy thereof.

(a) Any amendment to the paper copy of the "Sequence Listing" (§ 1.821(c)) must be made by the submission of substitute sheets. Amendments must be accompanied by a statement that indicates support for the amendment in the application, as filed, and a statement that the substitute sheets include no new matter.

(b) Any amendment to the paper copy of the "Sequence Listing," in accordance with paragraph (a) of this section, must be accompanied by a substitute copy of the computer readable form (§ 1.821(e)) including all previously submitted data with the amendment incorporated therein, accompanied by a statement that the copy in computer readable form is the same as the substitute copy of the "Sequence Listing."

(c) Any appropriate amendments to the "Sequence Listing" in a patent; *e.g.*, by reason of reissue or certificate of correction, must comply with the requirements of paragraphs (a) and (b) of this section.

(d) If, upon receipt, the computer readable form is found to be damaged or unreadable, applicant must provide, within such time as set by the Commissioner, a substitute copy of the data in computer readable form accompanied by a statement that the substitute data is identical to that originally filed.

7. Appendix A To Subpart G to Part 1 is revised to read as follows:

BILLING CODE 3510–16–P

Appendix A To Subpart G to Part 1—Sample Sequence Listing

```
<110> Smith, John

      Smith, Jane



<120> Example of a Sequence Listing



<130> 01-00001



<140> US 08/999,999

<141> 1998-02-28



<150> EP 91000000

<151> 1997-12-31
```

```
<160> 2


<170> PatentIn ver. 2.0



<210> 1

<211> 403

<212> DNA

<213> Paramecium aurelia



<220>

<221> CDS

<222> 341..394



<300>

<301> Doe, Richard

<302> Isolation and Characterization of a Gene Encoding a

      Protease from Paramecium sp.

<303> Journal of Fictional Genes
```

<304> 1

<305> 4

<306> 1 - 7

<307> 1988-06-20

<400> 1

ctactctact ctactctcat ctactatctt ctttggatct ctgagtctgc ctgagtggta   60

ctcttgagtc ctggagatct ctcctctcac atgtgatcgt cgagactgac cgatagatcg   120

ctgactgact ctgagatagt cgagcccgta cgagacccgt cgagggtgac agagagtggg   180

cgcgtgcgcg cagagcgccg cgccggtgcg cgcgcgagtg cgcggtgggc cgcgcgaggg   240

ctttcgcggc agcggcggcg ctttccggcg cgcgcccgtc cgcccctaga cctgagaggt   300

cttctcttcc ctcctcttca ctagagaggt ctatatatac atg gtt tca atg ttc   355

                                            Met Val Ser Met Phe

```
                                                    1                   5

agc ttg tct ttc aaa tgg cct gga ttt tgt ttg ttt gtt tgtttgctc        403

Ser Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu Phe Val
                    10                  15
```

<210> 2

<211> 18

<212> PRT

<213> Paramecium aurelia

<400> 2

```
Met Val Ser Met Phe Ser Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu


1               5                  10                  15




Phe Val
```

Dated: May 22, 1998.

Bruce A. Lehman,

Assistant Secretary of Commerce and
Commissioner of Patents and Trademarks.

[FR Doc. 98–14194 Filed 5–29–98; 8:45 am]

BILLING CODE 3510–16–C