DEPARTMENT OF COMMERCE

Patent and Trademark Office

37 CFR Part 1

[Docket No: 960828235–6235–01]

RIN 0651–AA88

Changes Implementing Nucleotide and/or Amino Acid Sequence Listings

AGENCY: Patent and Trademark Office, Commerce.

ACTION: Notice of proposed rulemaking and request for comments.

SUMMARY: The Patent and Trademark Office (PTO) is proposing to amend the rules for submitting nucleic acid or amino acid sequences in computer readable form (CRF) for patent applications to simplify the requirements of the rules, to rearrange portions of the rules for better understanding and to establish consistent rules to permit a single internationally acceptable computer readable form. The Sequence Listing will be presented in an international, language neutral format using numeric identifiers rather than the current subject headings and the paper Sequence Listing will be a separately numbered section of the patent application. Sequences which contain fewer than four (4) specifically identified nucleotides or amino acids will no longer be required to be submitted in computer readable form.

DATE: Written comments must be received by December 3, 1996.

ADDRESSES: Address written comments to: Box Comments—Patents, Assistant Commissioner for Patents, Washington, DC 20231, Attention: Esther M. Kepplinger or by Fax to (703) 305–3601 to her attention. Comments may be sent by mail message over the Internet addressed to seqrule@uspto.gov. The written comments will be available for public inspection in Suite 520, Crystal Park One, 2011 Crystal Drive, Arlington, Virginia.

FOR FURTHER INFORMATION CONTACT: Esther M. Kepplinger, by telephone at (703) 308–2339 or by mail addressed to: Box Comments—Patents, Assistant Commissioner for Patents, Washington, DC 20231 marked to her attention or by Fax to (703) 305–3601 or by electronic mail at ekepplin@uspto.gov.

SUPPLEMENTARY INFORMATION: The existing sequence rules (37 CFR 1.821–1.825) 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). The existing sequence rules have provided 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.

In an effort to streamline and reduce the procedural requirements of the existing rules and to respond to the needs of our customers while establishing an internationally acceptable standard, the PTO proposes to modify the current rules requiring the submission of computer readable forms for nucleotide and amino acid sequences.

To decrease the burden on applicants who file applications containing nucleotide and amino acid sequence information under the Patent Cooperation Treaty (PCT), the PTO entered into discussions at the PCT Meeting of International Authorities (MIA) in November 1994 on changing the applicable rules for submission and transfer of Sequence Listings. Under the current PCT rules, each International Searching Authority and national Office may set the standard for submission of the paper and electronic Sequence Listing information. This may impose a burden on applicants of providing several different formats of Sequence Listings in different languages during the international and national phases of the PCT procedure.

Under the current PCT practice, the applicant serves as the data repository for requests during each stage of the PCT practice for new electronic copies of the Sequence Listings.

Under national practice, a Sequence Listing may be required to be translated into the national language at considerable cost and posing the danger that the data could be inadvertently altered.

At the November 1994 MIA to address these problems, rule changes were proposed to require a language neutral Sequence Listing submission which would suffice for PCT and national stage sequence information processing. Initial Trilateral meetings and correspondence suggest that such a sequence submission would be acceptable under European Patent Office (EPO) and Japanese Patent Office (JPO) procedures, thus further lessening the burden on applicants.

These sequence rules are proposed to be revised in concert with World Intellectual Property Organization

(WIPO) International Standards ST.23 and ST.24 for the paper and electronic submission of sequence information in patent applications, as well as PCT requirements. This should result in an applicant having to produce a single Sequence Listing that would satisfy the filing requirements in all countries, as well as permitting an applicant to submit only a single electronic Sequence Listing in PCT applications.

In an effort to profit from the experiences of the nucleotide database information providers which pioneered the electronic submission of sequence information, the PTO discussed with them the possible simplification of the PTO sequence submission rules. In response to their advice (which confirmed the PTO experience), the number of mandatory data elements is proposed to be reduced.

Thus, the proposed rule changes 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 as a section with separately numbered pages;

(4) Modification of 37 CFR § 1.77 to include the paper Sequence Listing as a part of the specification and to provide a place for the paper Sequence Listing in the printed patent;

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

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

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

(8) Clarification and simplification of the rules to aid in understanding of the requirements that they set forth.

Request for Comments

The PTO is particularly interested in receiving comments on three queries. Currently sequences containing D-amino acids need not be provided in the ''Sequence Listing'', but the PTO has accepted voluntary submissions of sequences containing D-amino acids.

The commercially available sequence searching software used to search prior art databases is not capable of discerning D-amino acids since they do not have distinct designators. It is for this reason that the rules do not require a computer readable form for the disclosure of sequences which contain D-amino acids.

Those seeking to volunteer the information in accordance with these

rules might be seeking assurance that a machine search of the closest prior art will be conducted by the PTO or they consider the information useful and wish it to be in the database. If the PTO does not accept voluntary submissions, that would exclude information from the databases that at least some applicants believe to be valuable information.

The potential conflict created by accepting these D-amino acid-containing sequences is that the published database will contain sequences with D-amino acids and those using the published database may be operating on the assumption that it does not, given the indication in § 1.821(a)(2) that D-amino acid-containing sequences are not intended to be included. For this reason, there may be an advantage to having the D-amino acids indicated by Xaa to alert the user that the Feature section must be consulted. A disadvantage of voluntary submissions is that they will result in the generation of a database which is incomplete and cannot be relied upon to provide a complete search of the U.S. patent literature including sequences containing D-amino acids.

The PTO seeks comments on the following query:

(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?

Section 1.821(c) will continue to require that all sequence information contained in a disclosure, including in the specification, drawings or claims, be presented in the Sequence Listing in accordance with §§ 1.821—1.825. This provision does not discriminate between prior art sequences and ''new'' sequences. The PTO has received comments in the past and is seeking additional comments on this issue. The suggestion has been made that sequences which are prior art, and/or are contained in a database at the time of filing, need not be provided to the PTO in computer readable form since the sequence information is obtainable by other means. Responsive to these public comments, the PTO is considering amending the rules to permit omission of some sequences from the Sequence Listing if these sequences are admitted prior art to applicant and are in a publicly available, electronic, sequence database and the database accession number is supplied.

The suggestion to exclude prior art sequences was made when §§ 1.821–1.825 were originally adopted. 55 FR 18230, 18237 (1990). The final rules, however, required the submission of all sequence information in computer readable form. The reasons for that decision include: (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; (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; and (4) these rules 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. These reasons continue to be relevant.

The PTO is concerned about how such a provision would be drafted without creating difficult questions. A provision which excludes sequences whenever a sequence is prior art and has previously been included in a publicly available, electronic, sequence database appears to be straightforward; however, many technical and legal issues would result. What constitutes a publicly available, electronic, sequence database? Would the USPTO and the other patent offices which have similar rules be required to produce a list of internationally accepted databases? What would be the criteria for such acceptance? 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. Does the mere existence of the sequence information in such a record constitute reasonable means of retrieval? Would not one need some text basis or other identifier to retrieve the information?

Concerns have been voiced that the redundancy of including old sequences in the PTO database creates electronic searching problems, such as increased cost and reduced speed. Upon investigation, it has been found that requiring all disclosed sequences to be included in the Sequence Listing does

not cause search processing problems at the PTO or incur increased costs. The PTO seeks comments on the following query:

(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?

At the November 1994 MIA, it was proposed that the Sequence Listings submitted in an international application filed under the PCT would no longer be published on paper. It was suggested that the Sequence Listings be published electronically and be available in the electronic form from several sequence repositories throughout the world. These repositories would have the Sequence Listings available in electronic form at the time of publication of the PCT pamphlet.

The PTO seeks comments on the following query:

(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?

Written comments will be available for public inspection and will be available on the Internet (address: www.uspto.gov). Commentators should note that since their comments will be made publicly available, information that is not desired to be made public, such as the address and phone number of the commentator, should not be included in the comments. A public hearing will not be conducted.

Discussion of Specific Rules

Section 1.77 is proposed to be amended by revising paragraph (g), which would provide for a reference to a Sequence Listing Annex, if any exists. In the application as filed, on a separate page immediately before the claims, reference would be made to a Sequence Listing Annex and the Sequence Listing would be provided as a separately numbered section or Annex to the application. In a printed patent the Sequence Listing would appear immediately before the claims.

Section 1.77 is proposed to be amended to redesignate existing paragraphs (g)–(j) as paragraphs (h)–(k) and add an additional paragraph (l) Sequence Listing Annex. In the application as filed, the Sequence Listing would be provided by applicants as a separately numbered section or

Annex of the application. The pages of the Sequence Listing Annex should be numbered independently from the specification using sequential integers preceded by "A" to identify them as a part of the Annex and to prevent any confusion which might arise from using numbers already used in the specification. In a printed patent the Sequence Listing would be printed immediately before the claims. In cases where the Sequence Listing is voluminous, the files are difficult to handle. This change would permit easier storage of very large Sequence Listings apart from the main part of the application during pendency. The presentation of the Sequence Listing as a separate Annex would also facilitate compliance with PCT requirements and other national patent office rules.

Sections 1.821(a)(1) and (2) are proposed to be amended by referring to sections in World Intellectual Property Organization (WIPO) Handbook on Industrial Property Information and Documentation, Standard ST.23, paragraphs 8 through 12, April 1994, herein incorporated by reference, rather than to paragraphs in § 1.822. The WIPO Standard ST. 23 (April 1994) is consistent with § 1.822 except for certain corrections which are noted herein and the requirement of the use of the lower case for the one-letter code for nucleotide bases. The proposed rule states that the incorporation has been approved. This language is required by the Federal Register. This incorporation by reference will be reviewed by the Director of the Federal Register in accordance with 5 U.S.C. 552(a) and 1 CFR part 51 before any Final Rule is adopted. 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; or at the Office of the Federal Register, 800 North Capitol Street, NW, Suite 700, Washington, DC 20408.

Section 1.821(a) is proposed to be amended so that sequences with fewer than four specifically defined amino acids or nucleotides would be expressly excluded from this rule. "Specifically defined" means those amino acids other than "Xaa" and those nucleotide bases other than "N" defined in accordance with WIPO Standard ST.23.

This change is being proposed to reduce the burden on applicants for those sequences that contain only a minimal amount of sequence information. For example, if an amino acid sequence is disclosed as being

entirely "Xaa" residues, the 1990 version of the sequence rules would require this sequence to be submitted in computer readable form. However, this sequence has no value as sequence information because each of the positions is represented as a "wild card." Such low-information sequences are not very useful in any sequence matching and alignment algorithm. In order to minimize the inclusion of such low-information-value sequence data in the database and to relieve the burden on applicants to submit low-information-value sequences, the Office proposes this change to the sequence rules. If applicants should wish to voluntarily submit a CRF for such sequences, they would be accepted and entered in the PTO's database.

It is not necessary that any of the non-N or non-Xaa residues be adjacent to any other non-N or non-Xaa residue in order for a sequence to be subject to § 1.821(a).

Sections 1.821(a)(2) and 1.822(b) are proposed to be amended by changing "elsewhere in the 'Sequence Listing' " to "in the Feature section." The purpose of this change is to enhance clarity of the rule. The only place in the "Sequence Listing" where additional information is permitted is in the Feature section. The current language implies that there are other acceptable portions of the "Sequence Listing" appropriate for additional information and thus is ambiguous and misleading.

Section 1.821(a)(2) will continue to indicate that sequences containing D-amino acids need not comply with the provisions of §§ 1.822–1.825. To date, the PTO has accepted voluntary submissions of sequences which contain D-amino acids. The sequence information has either indicated an Xaa at each occurrence of a D-amino acid or has indicated the amino acid (or imino acid) by abbreviation as if it were an L-amino acid (or imino acid) and explained the existence of the D-amino acid in the Feature section associated with that sequence.

Section 1.821(c) is proposed to be amended by clarifying and establishing a language neutral format sequence listing. Specifically, the use of integer identifiers is proposed for identifying sequences. Where a sequence integer identifier is intentionally omitted, it must be noted by applicant to avoid confusion in the published document.

Section 1.821(d) is proposed to be amended by changing "assigned identifier" to "integer identifier" to be consistent with the term used in § 1.821(c).

Section 1.821(d) is proposed to be amended by adding the phrase,

"preceded by 'SEQ ID NO:' ". This change is necessitated by the change to § 1.821(c). Since the integer identifier in the "Sequence Listing" would be defined now as a numeral only, it is necessary that any reference to a particular sequence in the specification and claims be preceded by "SEQ ID NO:". It is not acceptable to use only a numeric identifier, such as "<200>" or "<400>"—see infra Sequence Listing table, in the description or the claims because one reading a patent may not reasonably be presumed to be familiar with the meanings of numeric identifiers.

Section 1.821(e) is proposed to be amended by setting forth the procedure for transferring an accepted computer readable Sequence Listing from one application to a subsequently filed application. The existing rules did not adequately describe the process of transferring a computer readable Sequence Listing into a new application if an identical CRF was previously accepted by the PTO for another application. A further description of the intended procedures has been added for purposes of clarity. This section is intended to describe that if a computer readable Sequence Listing is identical to one that is error-free and already on file at the PTO, an applicant has two options. A new diskette may be submitted, or an applicant may submit a statement clearly directing the PTO to use the previously submitted CRF since they are identical, and that the paper copy of the Sequence Listing in the new application is identical to the disk in the previous application.

Section 1.821(g) is proposed to be amended by correcting the reference to 35 U.S.C. 111(a) applications. Section 1.821(h) is proposed to be amended by clarifying that this rule applies to all international applications searched and examined by the PTO. In addition to international applications filed in the United States Receiving Office, the United States is a competent International Searching Authority (ISA) for applications filed in receiving Offices of, or acting for, Brazil, Israel, Mexico, and Trinidad and Tobago. The United States is also a competent ISA for applications filed in the International Bureau where at least one of the applicants is a resident or national of the United States or a resident or national of Barbados. In addition, the United States acts as an International Preliminary Examining Authority for certain applications searched in the EPO. The language change regarding the time limit for compliance and statement accompanying the submission are

necessary to conform with the language found in PCT Rule 13ter.1.

Section 1.822 is proposed to be revised for clarity and better organization and to accommodate an international request for the use of *lower case* one-letter codes for nucleotide bases.

Section 1.822 (b) is proposed to be amended to refer to WIPO Standard ST.23 (April 1994) and incorporate the information therein. The reorganization groups all nucleotide and all amino acid formats together.

Section 1.822 (c)(1) is proposed to be amended by requiring the use of lower case one-letter code for the nucleotide bases. This change would put the PTO requirements in conformance with most large databases. Additionally, the use of lower case letters in a sequence makes the confusion of ''g'' for ''c'' and vice versa less likely.

Current paragraph (d) is proposed to be redesignated as a part of subparagraph (c)(3) and current paragraph (e) is proposed to be deleted with the substance of the paragraph being incorporated into (d)(1). Current paragraph (f) is proposed to be redesignated as subparagraph (c)(2); current paragraph (g) is proposed to be redesignated as subparagraph (c)(3) and amended to incorporate current paragraph (d). Current paragraph (h) is proposed to be redesignated as subparagraph (d)(2). Current paragraphs (i) and (j) are proposed to be redesignated as (c)(4) and (c)(5). Current paragraph (k) is proposed to be redesignated as (d)(3). Current paragraph (l) is proposed to be redesignated as (c)(6) and current paragraph (m) is proposed to be

redesignated as (d)(4). Current paragraph (n) is proposed to be redesignated as (c)(7) and amended to delete a sentence, the substance of which is incorporated into (d)(4).

Paragraph (d)(1) is proposed to be added to include a reference to WIPO Standard ST.23 (April 1994). Paragraphs (d)(2–4) incorporate the material from current paragraphs (h), (k), (m) and a sentence of (n). Paragraph (d)(5) is proposed to be added to clarify that the use of terminator symbols is not acceptable in amino acid sequences either as ''internal'' terminator symbols or following the carboxy terminal amino acid of a peptide or polypeptide.

Current paragraph (o) is proposed to be redesignated as paragraph (e) and amended to recite integer identifier to be consistent with § 1.821 (c) and to permit the language neutral submission.

Current paragraph (p) is proposed to be deleted.

The lists of nucleic acid and amino acid abbreviations and the lists of modified base controlled vocabulary and the modified and unusual amino acids would be replaced by reference to WIPO Standard ST.23 RECOMMENDATION FOR THE PRESENTATION OF NUCLEOTIDE AND AMINO ACID SEQUENCE LISTINGS IN PATENT APPLICATIONS AND IN PUBLISHED PATENT DOCUMENTS (April 1994) to simplify and shorten the rules. This information will also appear in an appropriate section of the Manual of Patent Examining Procedure to assist applicants in preparing Sequence Listings. For purposes of facilitating review of these proposed rule changes, appropriate corrected excerpts of

paragraphs 8, 9, 11 and 12 of WIPO Standard ST.23 are provided below.

WIPO Standard ST.23, paragraph 8, provides that the bases of a nucleotide sequence should be represented using the following one-letter code for nucleotide sequence characters.

Symbol	Meaning	Origin of designation
A	A	Adenine
G	G	Guanine.
C	C	Cytosine.
T	T	Thymine.
U	U	Uracil.
R	G or A	puRine.
Y	T/U or C	pYrimidine.
M	A or C	aMino.
K	G or T/U	Keto.
S	G or C	Strong interactions 3H-bonds.
W	A or T/U	Weak interactions 2H-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).	aNy.

WIPO Standard ST.23, paragraph 9, provides: 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 elsewhere in the Sequence Listing. The codes from the list below may be used in the description or the Sequence Listing but not in the sequence itself.

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-galactosylqueosine.
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.

Symbol	Meaning
man q	*beta, D-mannosylqueosine.
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 (v).
osyw	wybutoxosine.
p	pseudouridine.
q	*queosine.
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.
yw	wybutosine.
x	3-(3-amino-3-carboxy-propyl)uridine, (acp3)u.

*Indicates a correction of minor typographical errors.

WIPO Standard ST.23, paragraph 11, provides that the amino acids should be represented using the following three-letter code with the first letter as a capital.

Symbol	Meaning
Ala	Alanine.
Cys	Cysteine.
Asp	Aspartic Acid.
Glu	Glutamic Acid.
Phe	Phenylalanine.
Gly	Glycine.
His	Histidine.
Ile	Isoleucine.
Lys	Lysine.
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.23, paragraph 12, provides: Modified and unusual amino acids may be represented as the corresponding unmodified amino acids in the sequence itself if the modified amino acid is one of those listed below and the modification is further described elsewhere in the Sequence Listing. The codes from the list below may be used in the description or the Sequence Listing but not in the sequence itself.

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.
*alle	allo-Isoleucine.
MeGly	N-Methylglycine, sarcosine.
*MeIle	N-Methylisoleucine.
MeLys	6-N-Methyllysine.
MeVal	N-Methylvaline.
Nva	Norvaline.
Nle	Norleucine.
Orn	Ornithine.

* Indicates a correction of a minor typographical error.

Section 1.823(a) is proposed to be amended to provide for a reference to a Sequence Listing Annex in the application immediately before the claims and to provide the paper Sequence Listing as an Annex, which is a separately numbered section of the application. This is an internationally desired change and also would facilitate easier storage of very large Sequence Listings separate from the main part of the file during pendency of the application.

Section 1.823(b) is proposed to be amended to insert a table to depict items of information (data elements) which are to be included in the Sequence Listing and to indicate whether they are mandatory or optional. The proposed revisions reflect the change to a language neutral submission. The English language data elements headings would be replaced by numeric identifiers. The numeric identifiers are similar to INID codes (''Internationally agreed Numbers for the Identification of Data'' as per WIPO Standard ST.9, December 1990) already utilized internationally in patent documents. This change would facilitate a single international standard which would eliminate the need for translations into non-English languages.

Large portions of Section 1.823(b) are proposed to be deleted to lessen the burden on applicants and to eliminate collections of material which is of limited use to the Office. The following items are typical of material which would be deleted:

(1)(vi)(C) CLASSIFICATION;
(2)(i)(C) STRANDEDNESS;
(2)(ii) MOLECULE TYPE through (2)(vii)(C) UNITS; and
(2)(ix)(C) IDENTIFICATION METHOD.

In order to clarify the rule, the proposed change would identify specifically those items which can be enumerated once in a Sequence Listing. It is proposed that the recommended designation be eliminated, leaving only mandatory and optional elements. Accordingly, it is proposed to change element <140> Correspondence Address

and elements <150> through <154> from mandatory to optional. Elements <100> General Information, <200> Information for SEQ ID NO, and <400> Sequence Description: SEQ ID NO have been clarified as mandatory. In element <193>, it is proposed to change TELEX to Electronic mail address to be current with technology.

It is proposed to eliminate Strandedness because the information is of limited use to the Office. It is proposed to limit the response for Topology to linear or circular because any other response does not permit an adequate search. Because it is essential to the search to know whether the sequence is circular, providing one of these two responses to this data element is mandatory in the Sequence Listing. Consistent with the international desire for eliminating language in the Sequence Listing, Topology would be identified as L (linear) or C (circular), and sequence Type would be N (nucleotide) or A (amino acid).

It is proposed to change Feature from a recommended to a mandatory element if the sequence contains ''N'', ''Xaa'', a modified or unusual L-amino acid or a modified base. This change would highlight the presence of an unusual residue in the sequence which is important to anyone using Sequence Listing information.

Section 1.824 is proposed to be amended by revising the current paragraphs (a) through (h) into paragraphs (a) through (c).

Specifically, the following changes are proposed for § 1.824:

Current § 1.824, paragraph (a), is proposed to be redesignated as paragraph (a)(1). In addition, the term ''series of diskettes'' would be added to indicate the acceptability of receiving numerous disks for large submissions. Current paragraph (b) is proposed to be redesignated as paragraph (a)(2). Current paragraph (c) is proposed to be redesignated as paragraph (a)(3). Current paragraph (d) is proposed to be deleted because it is incorporated into subparagraph (a)(1). Current paragraph (e) is proposed to be deleted since the PTO has not found it to be necessary and feels it should not be a requirement placed on the applicant, although the applicant may optionally continue the practice of using write-protection if desired. In proposed paragraph (a)(4), a ''compressed file'' format would be introduced as an acceptable means to submit a large sequence listing, and in proposed paragraph (a)(5), directions on suppressing page numbering on the computer readable form version would be added for clarity.

The text of current paragraph (f) is proposed to be deleted, but the list of computer readable files is proposed to be redesignated as subparagraphs under new (b) and (c). In proposed paragraph (b), the explanation for ''pagination'' is proposed to be revised to reflect the correct format required. Proposed paragraph (b)(1) is proposed to be revised by deleting diskettes from PS/2 operating system as an accepted format. In proposed paragraph (c), the diskette requirements are proposed to be rearranged so that the most common diskette size used for submissions is at the top of the list. Also in proposed paragraph (c)(2), ''format'' is proposed to be amended to accommodate the current PTO equipment, and in proposed new paragraphs (c)(3), (4), and (5), additional items would be added to the list of acceptable media types due to the changes in available equipment at the PTO.

Current paragraph (g) is proposed to be redesignated as paragraph (d).

Current paragraph (h) is proposed to be deleted because the text is proposed to be incorporated into paragraph (a)(6). The label requirements would be rewritten more concisely than with the previous rules. In addition, fewer items would be required to be placed on the label under this proposed paragraph because the other items are no longer deemed necessary by the PTO.

Current Appendix A is proposed to be rewritten to reflect the correct format of a Sequence Listing. The proposed Appendix A is presented to provide a sample listing in the correct format as described in the Table of amended § 1.823(b). This sample includes the use of numeric identifiers which reflect the change to a language neutral submission. Current Appendix B is proposed to be deleted as the information it presents is no longer valid under changes in this proposed rule.

Review Under the Paperwork Reduction Act of 1995

This proposed rule change contains information collection requirements which are subject to review by the Office of Management and Budget (OMB) under the Paperwork Reduction Act of 1995, 44 U.S.C. 3501, *et seq.* The title, description and respondent description of the information collection is shown below with an estimate of the annual reporting burdens. Included in the estimate is the time for reviewing instructions, gathering and maintaining the data needed, and completing and reviewing the collection of information. With respect to the following collection of information, the PTO invites comments on: (1) Whether the proposed collection of information is necessary for the proper performance of the PTO's functions, including whether the information will have practical utility; (2) the accuracy of the PTO's estimate of the burden of the proposed collection of information, including the validity of the methodology and assumptions used; (3) ways to enhance the quality, utility, and clarity of the information to be collected; and (4) ways to minimize the burden of the collection of information on respondents, including through the use of automated collection techniques, when appropriate, and other forms of information technology.

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 unless that collection of information displays a currently valid OMB control number.

OMB Number: 0651–0024.

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

Form Numbers: None.

Type of Review: Revision of currently approved collection.

Affected Public: Individuals or households, business or other for-profit institutions, not-for-profit institutions, and Federal Government.

Estimated Number of Respondents: 4,600.

Estimated Time Per Response: 80 minutes.

Estimated Total Annual Burden Hours: 6,133.

Needs and Uses: The PTO requires biotechnology patent applicants to submit sequence information to enable the PTO to properly examine and process their applications.

As required by the Paperwork Reduction Act of 1995, 44 U.S.C. 3507(d), the PTO has submitted a copy of this proposed rulemaking to OMB for its review of this information collection. Interested persons are requested to send comments regarding this information collections, including suggestions for reducing this burden, to the Office of Information and Regulatory Affairs of OMB, New Executive Office Bldg., 725 17th Street, NW., Room 10235, Washington, D.C. 20503, Attn: Desk Officer for the Patent and Trademark Office.

OMB is required to make a decision concerning the collection of information in these proposed regulations between 30 and 60 days after the publication of this document in the Federal Register.

Therefore, a comment to OMB is best assured of having its full effect if OMB receives it within 30 days of publication. This does not affect the deadline for the public to comment to the PTO on the proposed regulations.

Other Considerations

This proposed rule change is in conformity with the requirements of the Regulatory Flexibility Act (5 U.S.C. 601 *et seq.*), Executive Order 12612, and the Paperwork Reduction Act of 1995, 44 U.S.C. 3501 *et seq.* It has been determined that this proposed rule is not significant for the purposes of Executive Order 12866.

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 proposed rule change would not have a significant economic impact on a substantial number of small entities (Regulatory Flexibility Act, 5 U.S.C. 601 *et seq.*). The principal effect of this rule change is to simplify and clarify the rules governing the submission of Sequence Listings for patent applications containing nucleic acid and/or amino acid sequences.

The PTO has also determined that this proposed rule change 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, Reporting and record-keeping 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, the PTO proposes to amend 37 CFR part 1 as set forth below. Removals are indicated by brackets ([]) and additions indicated by arrows (> <).

PART 1—RULES OF PRACTICE IN PATENT CASES

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

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

2. Section 1.77 is proposed to be amended by redesignating current paragraphs (g) through (j) as paragraphs (h) through (k) and by adding new paragraphs (g) and (l) to read as follows:

§ 1.77 Arrangement of application elements.

* * * * *

>(g) Reference to Sequence Listing Annex.<

[(g)]>(h)< Claim or claims.

[(h)]>(i)< Abstract of the disclosure.

[(i)]>(j)< Signed oath or declaration.

[(j)]>(k)< Drawings.

>(l) Sequence Listing Annex.<

3. Section 1.821 is proposed to be amended by revising paragraphs (a) and (c) through (h) 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 rule. ''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.23: Recommendation for the Presentation of Nucleotide and Amino Acid Sequence Listings in Patent Applications and in Published Patent Documents, paragraphs 8 through 12, April 1994, herein incorporated by reference. (Hereinafter ''WIPO Standard ST.23 (April, 1994)''). 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.23 may be obtained from the World Intellectual Property Organization; 34 chemin des Colombettes; 1211 Geneva 20 Switzerland. Copies of ST.23 may be inspected at the Patent Search Room; Crystal Plaza 3, Lobby Level; 2021 South Clark Place; Arlington, VA 22202; or 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* are intended to embrace only those nucleotides that can be represented using the symbols set forth in [§ 1.822(b)(1)] >WIPO Standard ST.23 (April 1994), paragraph 8<. Modifications, *e.g.*, methylated bases, may be described as set forth in [§ 1.822(b)] >WIPO Standard ST.23 (April 1994), paragraph 9< , but shall not be shown explicitly in the nucleotide sequence.

(2) *Amino acids* are those L-amino acids commonly found in naturally occurring proteins and are listed in [§ 1.822(b)(2)] >WIPO Standard ST.23 (April 1994), paragraph 11<. 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 [§ 1.822(b)(2)] >WIPO Standard ST.23 (April 1994), paragraph 11< with the modified positions; *e.g.*, hydroxylations or glycosylations, being described as set forth in [§ 1.822(b)] >WIPO Standard ST.23 (April 1994), paragraph 12<, 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 [§ 1.822(b)(2)] >WIPO Standard ST.23 (April 1994), paragraph 11< in conjunction with a description [elsewhere in the ''Sequence Listing''] >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) * * *

(c) Patent applications which contain disclosures of nucleotide and/or amino acid sequences must contain, as a separate part of the disclosure on paper copy, hereinafter referred to as the ''Sequence Listing,'' a disclosure of 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. Each sequence disclosed must appear separately in the ''Sequence Listing.'' Each sequence set forth in the ''Sequence Listing'' shall be assigned a separate >integer< identifier [written as SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, etc]. >The integer identifiers shall begin with 1 and increase sequentially by integers. If no sequence is present for an integer identifier, the words ''This sequence omitted'' shall appear following the integer identifier.<

(d) Where the description or claims of a patent application discuss a sequence listing 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 [assigned] >integer< 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 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. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(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 has one month from the date of a notice which will be sent requiring compliance with the requirements in order to prevent abandonment of the application. Any submission in response to a requirement under this paragraph must be accompanied by a statement that the submission includes no new matter. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(h) If any of the requirements of paragraphs (b) through (f) of this section are not satisfied at the time of filing [,in the United States Receiving Office,] an international application under the Patent Cooperation Treaty (PCT) [applicant has one month from the date

of a notice which] >, 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< requiring compliance with the requirements [,or such other time as may be set by the Commissioner, in which to comply] >within a prescribed time period<. Any submission in response to a requirement under this paragraph must be accompanied by a statement that the submission does not include [new] matter [or go] >which goes< beyond the disclosure in the international application as filed. Such a statement must be a verified statement if made by a person not registered to practice before the Office. 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 >and the United States International Preliminary Examining Authority shall examine only to the extent that a meaningful examination can be performed<.

*   *   *   *   *

4. Section 1.822 is proposed to be 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 [(p)] >(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 [paragraphs (b)(1) and (b)(2) of this section] >WIPO Standard ST.23 (April 1994), paragraphs 8 and 11. 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.23 may be obtained from the World Intellectual Property Organization; 34 chemin des Colombettes; 1211 Geneva 20 Switzerland. Copies of ST.23 may be inspected at the Patent Search Room; Crystal Plaza 3, Lobby Level; 2021 South Clark Place; Arlington, VA 22202; or at the Office of the Federal Register, 800 North Capitol Street, NW, Suite 700, Washington, DC<. No code other than that specified in [this section] >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 [paragraphs (p)(1) or (p)(2) of this section] >WIPO Standard ST.23 (April 1994), paragraphs 9 and 12< and the modification is also set forth [elsewhere in the Sequence Listing (for example, FEATURES § 1.823(b)(2)(ix))] >in the Feature section<. Otherwise, all bases or amino acids not appearing in paragraphs [(b)(1) or (b)(2) of this section] >8 and 11 of the WIPO Standard ST.23 (April 1994)< shall be listed in a given sequence as ''N'' or ''Xaa,'' respectively, with further information, as appropriate, given [elsewhere in the Sequence Listing] >in the Feature section<.

[(1) Base codes:

Symbol	Meaning
A	A; adenine.
C	C; cytosine.
G	G; guanine.
T	T; thymine.
U	U; uracil.
M	A or C.
R	A or G.
W	A or T/U.
S	C or G.
Y	C or T/U.
K	G or T/U.
V	A or C or G; not T/U.
H	A or C or T/U; not G.
D	A or G or T/U; not C.
B	C or G or T/U; not A.
N	(A or C or G or T/U) or (unknown or other).

(2) Amino acid three-letter abbreviations:

Abbreviation	Amino acid name
Ala	Alanine.
Arg	Arginine.
Asn	Asparagine.
Asp	Aspartic Acid.
Asx	Aspartic Acid or Asparagine.
Cys	Cysteine.
Glu	Glutamic Acid.
Gln	Glutamine.
Glx	Glutamine or Glutamic Acid.
Gly	Glycine.
His	Histidine.
Ile	Isoleucine.
Leu	Leucine.
Lys	Lysine.
Met	Methionine.
Phe	Phenylalanine.
Pro	Proline.
Ser	Serine.
Thr	Threonine.
Trp	Tryptophan.
Tyr	Tyrosine.
Val	Valine.
Xaa	Unknown or other].

(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[, as] >set forth< in [paragraph (b)(1) of this section] >WIPO Standard ST.23 (April 1994), paragraph 8<.

[(d) 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.

(e) 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 paragraph (b)(2) of this section.]

[(f)] >(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.

[(g)] >(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.<

[(h) 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.]

[(i)] >(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.

[(j)] >(5)< A nucleotide sequence shall be presented, only by a single strand, in the 5′ to 3′ direction, from left to right.

[(k) 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.]

[(l)] >(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.

[(m) The enumeration of amino acids may start at the first amino acid of the first mature protein, with the number 1. The amino acids preceding the mature protein, e.g., pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, when presented, 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.]

[(n)] >(7)< For those nucleotide sequences that are circular in configuration, the enumeration method set forth in paragraph [(l)] >(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. [The enumeration method for amino acid sequences that is set forth in paragraph (m) of this section remains applicable for amino acid sequences that are circular in configuration.]

>(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.23 (April 1994), paragraph 11.

(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. The amino acids preceding the mature protein, e.g., pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, when presented, 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.

(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)< [(o)] A sequence with a gap or gaps shall be presented as a plurality of separate sequences, with separate [sequence] >integer< 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.

[(p) The code for representing modified nucleotide bases and modified or unusual amino acids shall conform to the code set forth in the tables in paragraphs (p)(1) and (p)(2) of this section. The modified base controlled vocabulary in paragraph (p)(1) of this section and the modified and unusual amino acids in paragraph (p)(2) of this section shall not be used in the nucleotide and/or amino acid sequences; but may be used in the description and/or the "Sequence Listing" corresponding to, but not including, the nucleotide and/or amino acid sequence.

(1) Modified base controlled vocabulary:

Abbreviation	Modified base description
ac4c	4-acetylcytidine.
chm5u	5-(carboxyhydroxylmethyl)uridine.
cm	2′-O-methylcytidine.
cmnm5s2u	5-carboxymethylaminomethyl-2-thioridine.
cmnm5u	5-carboxymethylaminomethyluridine.
d	dihydrouridine.
fm	2′-O-methylpseudouridine.
galq	beta,D-galactosylqueosine.
gm	2′-O-methylguanosine.
i	inosine.
i6a	N6-isopentenyladenosine.
m1a	1-methyladenosine.
m1f	1-methylpseudouridine.

Abbreviation	Modified base description
m1g	1-methylguanosine.
m1l	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.
manq	beta,D-mannosylqueosine.
mcm5s2u	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-methyl-carbamoyl)threonine.
mv	uridine-5-oxyacetic acid methylester.
o5u	uridine-5-oxyacetic acid (v).
osyw	wybutoxosine.
p	pseudouridine.
q	queosine.
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.
yw	wybutosine.
x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u.

(2) Modified and unusual amino acids:

Abbreviation	Modified and unusual amino acid
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.
alle	allo-Isoleucine.
MeGly	N-Methylglycine, sarcosine.
MeIle	N-Methylisoleucine.
MeLys	N-Methylvaline.
Nva	Norvaline.
Nle	Norleucine.
Orn	Ornithine.]

5. Section 1.823 is proposed to be 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 be titled ''Sequence Listing'' [and appear] >. 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.'' The ''Sequence Listing'' shall appear in the ''Sequence Listing Annex,'' which is numbered independently of the numbering of the remainder of the application and shall be placed in the application file. Upon printing the application as a patent, the ''Sequence Listing Annex'' containing the paper ''Sequence Listing'' shall be printed immediately before the patented claims.< Each page of the ''Sequence Listing'' shall contain no more than 66 lines and each line shall contain no more than 72 characters. A fixed-width font shall be used exclusively throughout the ''Sequence Listing.''

(b) The ''Sequence Listing'' shall, except as otherwise indicated, include, in addition to and immediately preceding the actual nucleotide and/or amino acid sequence, the [following items of information.] > 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 [,except that parenthetical explanatory information following the headings (identifiers) is to be omitted]. Each item of information shall begin on a new line [, enumerated with the number/numeral/letter in parentheses as shown below, with the heading (identifier) in upper case characters, followed by a colon, and then followed by the information provided] > beginning with the numeric identifier enclosed in angle brackets as shown<. Except as allowed below, no item of information shall occupy more than one line. [Those items of information that are applicable for all sequences shall only be set forth once in the ''Sequence Listing.''] The submission of those items of information designated with an ''M'' is mandatory. [The submission of those items of information designated with an ''R'' is recommended, but not required.] The submission of those items of information designated with an ''O'' is optional. >Numeric identifiers <100>

through <193> shall only be set forth at the beginning of the ''Sequence Listing.''< Those items designated with ''rep'' may have multiple responses and, as such, the item may be repeated in the ''Sequence Listing.

[(1) GENERAL INFORMATION (Application, diskette/tape and publication information):

(i) APPLICANT (maximum of first ten named applicants; specify one name per line: SURNAME comma OTHER NAMES and/or INITIALS—M/rep):

(ii) TITLE OF INVENTION (title of the invention, as elsewhere in application, four lines maximum—M):

(iii) NUMBER OF SEQUENCES (number of sequences in the ''Sequence Listing'' (M):

(iv) CORRESPONDENCE ADDRESS (M):

(A) ADDRESSEE (name of applicant, firm, company or institution, as may be appropriate):

(B) STREET (correspondence street address, as elsewhere in application, four lines maximum):

(C) CITY (correspondence city address, as elsewhere in application):

(D) STATE (correspondence state, as elsewhere in application):

(E) COUNTRY (correspondence country, as elsewhere in application):

(F) ZIP (correspondence ZIP or postal code, as elsewhere in application):

(v) COMPUTER READABLE FORM (M):

(A) MEDIUM TYPE (type of diskette/tape submitted):

(B) COMPUTER (type of computer used with diskette/tape submitted):

(C) OPERATING SYSTEM (type of operating system used):

(D) SOFTWARE (type of software used to create computer readable form):

(vi) CURRENT APPLICATION DATA (M, if available):

(A) APPLICATION NUMBER (U.S application number, including a series code, a slash and a serial number, or U.S. PCT application number, including the letters PCT, a slash, a two-letter code indicating the U.S. as the Receiving Office, a two-digit indication of the year, a slash and a five-digit number, if available):

(B) FILING DATE (U.S. or PCT application filing date, if available; specify as dd-MMM-yyyy):

(C) CLASSIFICATION (IPC/US classification or F-term designation, where F-terms have been developed, if assigned, specify each designation, left justified, within an eighteen-position alpha numeric field—rep, to a maximum of ten classification designations):

(vii) PRIOR APPLICATION DATA (prior domestic, foreign priority or

international application data, if applicable—M/rep):

(A) APPLICATION NUMBER (application number; specify as two-letter country code and an eight-digit application number; or if a PCT application, specify as the letters PCT, a slash, a two-letter code indicating the Receiving Office, a two-digit indication of the year, a slash and a five-digit number):

(B) FILING DATE (document filing date, specify as dd-MMM-yyyy):

(viii) ATTORNEY/AGENT INFORMATION (O):

(A) NAME (attorney/agent name; SURNAME comma OTHER NAMES and/or INITIALS):

(B) REGISTRATION NUMBER (attorney/agent registration number):

(C) REFERENCE/DOCKET NUMBER (attorney/agent reference or docket number):

(ix) TELECOMMUNICATION INFORMATION (O):

(A) TELEPHONE (telephone number of applicant or attorney/agent):

(B) TELEFAX (telefax number of applicant or attorney/agent):

(C) TELEX (telex number of applicant or attorney/agent):

(2) INFORMATION FOR SEQ ID NO: X (rep):

(i) SEQUENCE CHARACTERISTICS (M):

(A) LENGTH (sequence length, expressed as number of base pairs or amino acid residues):

(B) TYPE (sequence type, i.e., whether nucleic acid or amino acid):

(C) STRANDEDNESS (if nucleic acid, number of strands of source organism molecule, i.e., whether single-stranded, double-stranded, both or unknown to applicant):

(D) TOPOLOGY (whether source organism molecule is circular, linear, both or unknown to applicant):

(ii) MOLECULE TYPE (type of molecule sequenced in SEQ ID NO:X (at least one of the following should be included with subheadings, if any, in Sequence Listing—R)):

—Genomic RNA;
—Genomic DNA;
—mRNA
—tRNA;
—rRNA;
—snRNA;
—scRNA;
—preRNA;
—cDNA to genomic RNA;
—cDNA to mRNA;
—cDNA to tRNA;
—cDNA to rRNA;
—cDNA to snRNA;
—cDNA to scRNA;
—Other nucleic acid;

(A) DESCRIPTION (four lines maximum):

—protein and
—peptide.

(iii) HYPOTHETICAL (yes/no—R):

(iv) ANTI-SENSE (yes/no—R):

(v) FRAGMENT TYPE (for proteins and peptides only, at least one of the following should be included in the Sequence Listing—R):

—N-terminal fragment;
—C-terminal fragment and
—internal fragment.

(vi) ORIGINAL SOURCE (original source of molecule sequenced in SEQ ID NO:X—R):

(A) ORGANISM (scientific name of source organism):

(B) STRAIN:

(C) INDIVIDUAL ISOLATE (name/number of individual/isolate):

(D) DEVELOPMENTAL STAGE (give developmental stage of source organism and indicate whether derived from germ-line or rearranged developmental pattern):

(E) HAPLOTYPE:

(F) TISSUE TYPE:

(G) CELL TYPE:

(H) CELL LINE:

(I) ORGANELLE:

(vii) IMMEDIATE SOURCE (immediate experimental source of the sequence in SEQ ID NO:X—R):

(A) LIBRARY (library -type, name):

(B) CLONE (clone(s)):

(viii) POSITION IN GENOME (position of sequence in SEQ ID NO:X in genome—R):

(A) CHROMOSOME/SEGMENT (chromosome/segment—name/number):

(B) MAP POSITION:

(C) UNITS (units for map position, i.e., whether units are genome percent, nucleotide number or other/specify):

(ix) FEATURE (description of points of biological significance in the sequence in SEQ ID NO:X -R/rep):

A) NAME/KEY (provide appropriate identifier for feature—four lines maximum):

(B) LOCATION (specify location according to syntax of DDBJ/EMBL/GenBank Feature Tables Definition, including whether feature is on complement of presented sequence; where appropriate state number of first and last bases/amino acids in feature—four lines maximum):

(C) IDENTIFICATION METHOD (method by which the feature was identified, i.e., by experiment, by similarity with known sequence or to an established consensus sequence, or by similarity to some other pattern—four lines maximum):

(D) OTHER INFORMATION (include information on phenotype conferred,

biological activity of sequence or its product, macromolecules which bind to sequence or its product, or other relevant information—four lines maximum):

(x) PUBLICATION INFORMATION (Repeat section for each relevant publication—O/rep):

(A) AUTHORS (maximum of first ten named authors of publication; specify one name per line: SURNAME comma OTHER NAMES and/or INITIALS—rep):

(B) TITLE (title of publication):

(C) JOURNAL (journal name in which data published):

(D) VOLUME (journal volume in which data published):

(E) ISSUE (journal issue number in which data published):

(F) PAGES (journal page numbers in which data published):

(G) DATE (journal date in which data published; specify as dd-MMM-yyyy, MMM-yyyy or Season-yyyy):

(H) DOCUMENT NUMBER (document number, for patent type citations only; specify as two-letter country code, eight-digit document number (right justified), one letter and as appropriate, one number or a space as a document type code; or if a PCT application specify as the letters PCT, a slash, a two-letter code indicating the Receiving Office, a two-digit indication of the year, a slash and

a five-digit number; or if a PCT publication, specify as the two letters WO, a two-digit indication of the year, a slash and a five-digit publication number):

(I) FILING DATE (document filing date, for patent-type citations only; specify as dd-MMM-yyyy):

(J) PUBLICATION DATE (document publication date; for patent-type citations only, specify as dd-MMM-yyyy):

(K) RELEVANT RESIDUES In SEQ ID NO:X (rep): FROM position) TO position)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:X:]

Numeric identifier	Definition	Comments and format	Mandatory (M) or Optional (O)
<100>	General Information	Leave blank after <100>	M
<110>	Applicant	Max. of 10 names; one name per line; use format: Surname, Other Names and/or Initials; rep.	M
<120>	Title of Invention	Four lines maximum	M
<130>	Number of Sequences	Use an integer as a response	M
<140>	Correspondence Address	140> must be present if subheadings <141>-<146> are used.	O
<141>	Addressee	....	O
<142>	Street	Four lines maximum	O
<143>	City	....	O
<144>	State or Province	....	O
<145>	Country	....	O
<146>	Zip or Postal Code	....	O
<150>	Computer Readable Form	Leave blank after <150>	O
<151>	Medium Type	Type of diskette/tape submitted	O
<152>	Computer	Type of computer used to create diskette/tape	O
<153>	Operating System	Type of operating system on computer	O
<154>	Software	Type of software used to create computer readable form.	O
<160>	Current Application Data	Leave blank after <160>; <160> must be present if subheadings <161> & <162> are used.	M, if available.
<161>	Application Number	Specify as: US 07/999,999 or PCT/US96/99999	M, if available.
<162>	Filing Date	Specify as: dd-MMM-yyyy	M, if available
<170>	Prior Application Data	Insert heading/subheadings only if applicable; leave blank after <170>; <170> must be present if subheadings <171> & <172> are used; rep.	M, if applicable
<171>	Application Number	Specify as: US 07/999,999 or PCT/US96/99999	M, if applicable.
<172>	Filing Date	Specify as: dd-MMM-yyyy	M, if applicable.
<180>	Attorney/Agent Information	Leave blank after <180>	O
<181>	Name	Use format: Surname, Other Names and/or Initials	O
<182>	Registration Number	....	O
<183>	File Reference/Docket Number	....	O
<190>	Telecommunication Information	Leave blank after <190>	O
<191>	Telephone	....	O
<192>	Telefax	....	O
<193>	Electronic mail address	....	O
<200>	Information for SEQ ID NO:#:	Response shall be an integer representing the SEQ ID NO shown; rep.	M
<210>	Sequence Characteristics	Leave blank after <210>	M
<211>	Length	Respond with an integer expressing the number of bases or amino acid residues.	M
<212>	Type	Whether presented sequence molecule is nucleotide or amino acid, indicated by N or A.	M
<214>	Topology	Whether presented sequence molecule is linear or circular, indicated as L or C.	M
<290>	Feature	Description of points of biological significance in the sequence; leave blank after <290>; rep.	M, if "N", "Xaa", or a modified or unusual L-amino acid or modified base was used in the sequence.

Numeric identifier	Definition	Comments and format	Mandatory (M) or Optional (O)
<291>	Name/Key	Provide appropriate identifier for feature; four lines maximum.	M, if "N", "Xaa", or a modified or unusual L-amino acid or modified base was used in the sequence.
<292>	Location	Specify location within sequence; where appropriate state number of first and last bases/amino acids in feature; four lines maximum.	M, if "N", "Xaa", or a modified or unusual L-amino acid or modified base was used in the sequence.
<294>	Other Information	Other relevant information; four lines maximum	M, if "N", "Xaa", or a modified or unusual L-amino acid or modified base was used in the sequence.
<300>	Publication Information	Leave blank after <300>; rep	O
<301>	Authors	Maximum of ten named authors of publication; specify one name per line; use 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 in which data published; specify as dd-MMM-yyyy, MMM-yyyy or Season-yyyy.	O
<308>	Patent Document Number	Document number; for patent-type citations only	O
<309>	Filing Date	Document filing date, for patent-type citations only; specify as dd-MMM-yyyy.	O
<310>	Publication Date	Document publication date, for patent-type citations only; specify as dd-MMM-yyyy.	O
<311>	Relevant Residues	FROM (position) TO (position)	O
<400>	Sequence Description: SEQ ID NO:#:	Response shall be an integer representing the SEQ ID NO shown; rep.	M

6. Section 1.824 is proposed to be 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 [contain a printable copy of the "Sequence Listing," as defined in §§ 1.821(c), 1.822 and 1.823, recorded as] >meet the following specifications:

(1) The computer readable form shall contain< a single [file on] >"Sequence Listing" as< either a diskette, [or a magnetic tape] >series of diskettes, or other permissible media outlined in § 1.824(c)<. [The computer readable form shall be encoded and formatted such that a printed copy of the "Sequence Listing" may be recreated using the print commands of the computer/operating-system configurations specified in paragraph (f) of this section.]

[(b)] >(2)< The [file] >"Sequence Listing"< in paragraph (a) >(l)< of this

section shall be [encoded in a subset of the] >submitted in< American Standard Code for Information Interchange (ASCII) >text<. [This subset shall consist of all printable ASCII characters including the ASCII space character plus line-termination, pagination and end-of-file characters associated with the computer/operating-system configurations specified in paragraph (f) of this section.] No other [characters] >formats< shall be allowed.

[(c)] >(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 [be readable by one of the computer/operating-system configurations specified in paragraph (f) of this section, and shall] conform to [the] >all< specifications [in paragraphs (a) and (b) of] >detailed in< this section.

[(d) The entire printable copy of the "Sequence Listing shall be contained within one file on a single diskette or magnetic tape unless it is shown to the

satisfaction of the Commissioner that it is not practical or possible to submit the entire printable copy of the "Sequence Listing" within one file on a single diskette or magnetic tape.

(e) The submitted diskette or tape shall be write-protected such as by covering or uncovering diskette holes, removing diskette write tabs or removing tape write rings.

(f) As set forth in paragraph (c), above, any means may be used to create the computer readable form, as long as the following conditions are satisfied. A submitted diskette shall be readable on one of the computer/operating-system configurations described in paragraphs (1) through (3), below. A submitted tape shall satisfy the format specifications described in paragraph (4), below.]

>(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 name and type of computer and operating system used, and application serial number and filing date, if known.

(b) Computer readable form files submitted must meet these format requirements:<

(1) Computer: IBM PC/XT/AT, >or compatibles< [IBM PS/2 or compatibles]>, or Apple Macintosh<;

[(i)]>(2)< Operating System: [PC–DOS or] MS–DOS [(Versions 2.1 or above)] >, Unix or Macintosh<;

[(ii)]>(3)< Line Terminator: ASCII Carriage Return plus ASCII Line Feed;

[(iii)]>(4)< Pagination: [ASCII Form Feed or Series of Line Terminators] >Continuous file (no ''hard page break'' codes permitted)<;

[(iv) End-of-File: ASCII SUB (Ctrl–Z);

(v) Media:]

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

[(A) Diskette—5.25 inch, 360 Kb storage;

(B) Diskette—5.25 inch, 1.2 Mb storage;

(C) Diskette—3.50 inch, 720 Kb storage;

(D) Diskette—3.5 inch, 1.44 Mb storage;]

>(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;<

[(vi) Print Command: PRINT filename.extension;

(2) Computer: IBM PC/XT/AT, IBM PS/2 or compatibles;

(i) Operating system: Xenix;

(ii) Line Terminator: ASCII Carriage Return;

(iii) Pagination: ASCII Form Feed or Series of Line Terminators;

(iv) End-of-File: None;

(v) Media:

(A) Diskette—5.25 inch, 360 Kb storage;

(B) Diskette—5.25 inch, 1.2 Mb storage;

(C) Diskette—3.50 inch, 720 Kb storage;

(D) Diskette—3.5 inch, 1.44 Mb storage;

(vi) Print Command: Ipr filename;

(3) Computer: Apple Macintosh;

(i) Operating System: Macintosh;

(ii) Macintosh File Type: text with line termination

(iii) Line Terminator: Pre-defined by text type file;

(iv) Pagination: Pre-defined by text type file;

(v) End-of-File: Pre-defined by text type file;

(vi) Media:

(A) Diskette—3.50 inch, 400 Kb storage;

(B) Diskette—3.50 inch, 800 Kb storage;

(C) Diskette—3.50 inch, 1.4 Mb storage;

(vii) Print Command: Use PRINT command from any Macintosh Application that processes text files, such as Mac-Write or TeachText;

(4) Magnetic tape: 0.5 inch, up to 2400 feet;

(i) Density: 1600 or 6250 bits per inch, 9 track;

(ii) Format: raw, unblocked;

(iii) Line Terminator: ASCII Carriage Return plus optional ASCII Line Feed;

(iv) Pagination: ASCII Form Feed or Series of Line Terminators;

(v) Print Command (Unix shell version given here as sample response— mt/dev/rmt0; 1pr/dev/rmt0):]

>(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<

[(g)]>(d)< Computer readable forms that are submitted to the Office will not be returned to the applicant.

[(h) All computer readable forms shall have a label permanently affixed thereto on which has been hand-printed or typed, a description of the format of the computer readable form as well as the name of the applicant, the title of the invention, the date on which the data were recorded on the computer readable form and the name and type of computer and operating system which generated the files on the computer readable form. If all this information cannot be printed on a label affixed to the computer readable form, by reason of size or otherwise, the label shall include the name of the applicant and the title of the invention and a reference number, and the additional information may be provided on a container for the computer readable form with the name

of the applicant, the title of the invention, the reference number and the additional information affixed to the container. If the computer readable form is submitted after the date of filing under 35 U.S.C. 111, after the date of entry in the national stage under 35 U.S.C. 371 or after the time of filing, in the United States Receiving Office, an international application under the PCT, the labels mentioned herein must also include the date of the application number, including series code and serial number.]

7. Section 1.825 is proposed to be amended by revising paragraphs (a), (b) and (d) 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. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(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.'' Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(c) * * *

(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. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

8. Appendix A to subpart G is proposed to be revised to read as follows:

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

[(1) GENERAL INFORMATION:
(i) APPLICANT: Doe, Joan X, Doe, John Q
(ii) TITLE OF INVENTION: Isolation and Characterization of a Gene Encoding a Protease from Paramecium sp.
(iii) NUMBER OF SEQUENCES: 2
(iv) CORRESPONDENCE ADDRESSES:
    (A) ADDRESSEE: Smith and Jones
    (B) STREET: 123 Main Street
    (C) CITY: Smalltown
    (D) STATE: Anystate
    (E) COUNTRY: USA
    (F) ZIP: 12345
(v) COMPUTER READABLE FORM:
    (A) MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb storage
    (B) COMPUTER: Apple Macintosh
    (C) OPERATING SYSTEM: Macintosh 5.0
    (D) SOFTWARE: MacWrite
(vi) CURRENT APPLICATION DATA:
    (A) APPLICATION NUMBER: 09/999,999

    (B) FILING DATE: 28–FEB–1989
    (C) CLASSIFICATION: 999/99
(vii) PRIOR APPLICATION DATA:
    (A) APPLICATION NUMBER: PCT/US88/99999
    (B) FILING DATE: 01–MAR–1988
(viii) ATTORNEY/AGENT INFORMATION:
    (A) NAME: Smith, John A
    (B) REGISTRATION NUMBER: 00001
    (C) REFERENCE/DOCKET NUMBER: 01–0001
(ix) TELECOMMUNICATIONS INFORMATION:
    (A) TELEPHONE: (909) 999–001
    (B) TELEFAX: (909) 999–0002
    (2) INFORMATION FOR SEQ ID NO: 1:
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 954 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA
(iii) HYPOTHETICAL: yes

(iv) ANTI-SENSE: no
(vi) ORIGINAL SOURCE:
    (A) ORGANISM: Paramecium sp
    (C) INDIVIDUAL/ISOLATE: XYZ2
    (G) CELL TYPE: unicellular organism
(vii) IMMEDIATE SOURCE:
    (A) LIBRARY: genomic
    (B) CLONE: Para-XYZ2/36
(x) PUBLICATION INFORMATION:
    (A) AUTHORS: Doe, Joan X, Doe, John Q
    (B) TITLE: Isolation and Characterization of a Gene Encoding a Protease from Paramecium sp.
    (C) JOURNAL: Fictional Genes
    (D) VOLUME: I
    (E) ISSUE: 1
    (F) PAGES: 1–20
    (G) DATE: 02–MAR–1988
    (K) RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 954

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
ATCGGGATAG  TACTGGTCAA  GACCGGTGGA  CACCGGTTAA  CCCCGGTTAA  GTACCGGTTA      60

TAGGCCATTT  CAGGCCAAAT  GTGCCCAACT  ACGCCAATTG  TTTTGCCAAC  GGCCAACGTT     120

ACGTTCGTAC  GCACGTATGT  ACCTAGGTAC  TTACGGACGT  GACTACGGAC  ACTTCCGTAC     180

GTACGTACGT  TTACGTACCC  ATCCCAACGT  AACCACAGTG  TGGTCGCAGT  GTCCCAGTGT     240

ACACAGACTG  CCAGACATTC  TTCACAGACA  CCCC ATG  ACA CCA CCT GAA CGT  CTC     295
                                         Met  Thr Pro Pro Glu Arg  Leu
                                                                    -30

TTC CTC CCA AGG GTG TGT GGC ACC ACC CTA CAC CTC  CTC  CTT  CTG  GGG       343
Phe Leu Pro Arg Val Cys Gly Thr Thr Leu His Leu  Leu  Leu  Leu  Gly
        -25               -20                -15

CTG CTG CTG GTT CTG CTG  CCT  GGG  GCC  CAT  GTGAGGCAGC AGGAGAATGG         393
Leu Leu Leu Val Leu Leu  Pro  Gly  Ala  His
    -10               -5

GGTGGCTCAG  CCAAACCTTG  AGCCCTAGAG  CCCCCCTCAA  CTCTGTTCTC   CTAG GGG      450
                                                                 Gly

CTC ATG CAT CTT GCC CAC AGC AAC CTC  AAA CCT  GCT  GCT  CAC  CTC  ATT     498
Leu Met His Leu Ala His Ser Asn Leu  Lys Pro  Ala  Ala  His  Leu  Ile
 1               5                    10                    15

GTAAACATCC  ACCTGACCTC  CCAGACATGT  CCCCACCAGC  TCTCCTCCTA  CCCCTGCCTC     558

AGGAACCCAA  GCATCCACCC  CTCTCCCCCA  ACTTCCCCCA  CGCTAAAAAA  AACAGAGGGA     618

GCCCACTCCT  ATGCCTCCCC  CTGCCATCCC  CCAGGAACTC  AGTTGTTCAG  TGCCCACTTC     678

TAC CCC AGC AAG CAG AAC TCA CTG CTC TGG AGA GCA  AAC  ACG  GAC  CGT       726
Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala  Asn  Thr  Asp  Arg
            20              25                    30

GCC TTC CTC CAG GAT GGT TTC TCC TTG AGC AAC AAT TCT CTC CTG  GTC         774
Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu Leu  Val
        35              40              45

TAGAAAAAAT  AATTGATTTC  AAGACCTTCT  CCCCATTCTG  CCTCCATTCT  GACCATTTCA     834

GGGGTCGTCA  CCACCTCTCC  TTTGGCCATT  CCAACAGCTC  AAGTCTTCCC  TGATCAAGTC     894

ACCGGAGCTT  TCAAAGAAGG  AATTCTAGGC  ATCCCAGGGG  ACCCACACCT  CCCTGAACCA     954
```

(2) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 82 amino acids
    (B) TYPE: amino acid
    (C) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(ix) FEATURE:
    (A) NAME/KEY: signal sequence
    (B) LOCATION: − 34 to − 1

(C) IDENTIFICATION METHOD: similarity to other signal sequences, hydrophobic
(D) OTHER INFORMATION: expresses protease
(x) PUBLICATION INFORMATION:
    (A) AUTHORS: Doe, Joan X, Doe, John Q
    (B) TITLE: Isolation and Characterization of a Gene Encoding a Protease from *Paramecium* sp.

(C) JOURNAL: Fictional Genes
(D) VOLUME: I
(E) ISSUE: 1
(F) PAGES: 1–20
(G) DATE: 02–MAR–1988
(H) RELEVANT RESIDUES IN SEQ ID NO:2: FROM -34 TO 48

BILLING CODE 3510–16–P

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
            -30             -25                 -20

Leu His Leu Leu Leu Leu Gly Leu Leu Leu Val Leu Leu Pro Gly Ala
        -15             -10                 -5

His Gly Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His
    1           5                   10

Leu Ile Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr
15              20              25                      30

Asp Arg Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu
            35                      40                  45

Leu Val ]
```

```
>
  <100>
  <110> Doe, Joan X, Doe, John Q
  <120> Isolation and Characterization of a
    Gene Encoding a Protease from
    Paramecium sp.
  <130> 2
  <140>
  <141> Smith and Jones
  <142> 123 Main Street
  <143> Smalltown
  <144> Anystate
  <145> USA
  <146> 12345
  <150>
  <151> Floppy disk
  <152> IBM PC compatible
  <153> PC-DOS/MS-DOS
```

```
<154> PatentIn Release #2.00
<160>
<161> 09/999,999
<162> 28–FEB–1989
<170>
<171> PCT/US/88/99999
<172> 01–MAR–1988
<180>
<181> Smith, John A
<182> REGISTRATION NUMBER: 00001
<183> 01–0001
<190>
<191> (909) 999–0001
<192> (909) 999–0002
<200> 1
<210>
<211> 954 base pairs
<212> N
<214> L
```

```
<290>
<291> CDS
<292> join(275..373, 448..498, 679..774)
<290>
<291> mat__peptide
<292> join(451..498, 679..774)
<300>
<301> Doe, Joan X, Doe, John Q
<302> Isolation and Characterization of a
    Gene Encoding a Protease from
    Paramecium sp.
<303> Fictional Genes
<304> 1
<305> 1
<306> 1–20
<307> 02–MAR–1988
<308> FROM 1 TO 957
<400> 1
```

BILLING CODE 3510–16–P

atcgggatag tactggtcaa gaccggtgga
caccggttaa ccccggttaa gtaccggtta 60
taggccattt caggccaaat gtgcccaact
acgccaattg ttttgccaac ggccaacgtt 120
acgttcgtac gcacgtatgt acctaggtac
ttacggacgt gactacggac acttccgtac 180
gtacgtacgt ttacgtaccc atcccaacgt
aaccacagtg tggtcgcagt gtcccagtgt 240
acacagactg ccagacattc ttcacagaca cccc atg
aca cca cct gaa cgt 292
Met Thr Pro Pro Glu Arg
-30
ctc ttc ctc cca agg gtg tgt ggc acc acc cta
cac ctc ctc ctt ctg 340
Leu Phe Leu Pro Arg Val Cys Gly Thr
Thr Leu His Leu Leu Leu Leu
-25 -20 -15
ggg ctg ctg ctg gtt ctg ctg cct ggg gcc cat
gtgaggcagc aggagaatgg 393

Gly Leu Leu Leu Val Leu Leu Pro Gly
Ala His
-10 -5
ggtggctcag ccaaaccttg agccctagag
ccccccctcaa ctctgttctc ctag ggg 450
Gly
ctc atg cat ctt gcc cac agc aac ctc aaa cct
gct gct cac ctc att 498
Leu Met His Leu Ala His Ser Asn Leu
Lys Pro Ala Ala His Leu Ile
1 5 10 15
gtaaacatcc acctgacctc ccagacatgt
ccccaccagc tctcctccta cccctgcctc 558
aggaacccaa gcatccaccc ctctcccca
acttcccca cgctaaaaaa aacagaggga 618
gcccactcct atgcctcccc ctgccatccc
ccaggaactc agttgttcag tgcccacttc 678
tac ccc agc aag cag aac tca ctg ctc tgg aga
gca aac acg gac cgt 726

Tyr Pro Ser Lys Gln Asn Ser Leu Leu
Trp Arg Ala Asn Thr Asp Arg
20 25 30
gcc ttc ctc cag gat ggt ttc tcc ttg agc aac
aat tct ctc ctg gtc 774
Ala Phe Leu Gln Asp Gly Phe Ser Leu
Ser Asn Asn Ser Leu Leu Val
35 40 45
tagaaaaaat aattgatttc aagaccttct ccccattctg
cctccattct gaccatttca 834
ggggtcgtca ccacctctcc tttggccatt
ccaacagctc aagtcttccc tgatcaagtc 894
accggagctt tcaaagaagg aattctaggc
atcccagggg acccacacct ccctgaacca 954

BILLING CODE 3510–16–P

```
atcgggatag tactggtcaa gaccggtgga caccggttaa ccccggttaa gtaccggtta      60

taggccattt caggccaaat gtgcccaact acgccaattg ttttgccaac ggccaacgtt     120

acgttcgtac gcacgtatgt acctaggtac ttacggacgt gactacggac acttccgtac     180

gtacgtacgt ttacgtaccc atcccaacgt aaccacagtg tggtcgcagt gtcccagtgt     240

acacagactg ccagacattc ttcacagaca cccc atg aca cca cct gaa cgt         292
                                        Met Thr Pro Pro Glu Arg
                                                              -30

ctc ttc ctc cca agg gtg tgt ggc acc acc cta cac ctc ctc ctt ctg      340
Leu Phe Leu Pro Arg Val Cys Gly Thr Thr Leu His Leu Leu Leu Leu
            -25               -20               -15

ggg ctg ctg ctg gtt ctg ctg cct ggg gcc cat gtgaggcagc aggagaatgg    393
Gly Leu Leu Leu Val Leu Leu Pro Gly Ala His
        -10               -5

ggtggctcag ccaaaccttg agccctagag cccccctcaa ctctgttctc ctag ggg      450
                                                                Gly

ctc atg cat ctt gcc cac agc aac ctc aaa cct gct gct cac ctc att      498
Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His Leu Ile
  1               5               10               15

gtaaacatcc acctgacctc ccagacatgt ccccaccagc tctcctccta cccctgcctc     558

aggaacccaa gcatccaccc ctctccccca acttcccca cgctaaaaaa aacagaggga      618

gcccactcct atgcctcccc ctgccatccc ccaggaactc agttgttcag tgcccacttc     678

tac ccc agc aag cag aac tca ctg ctc tgg aga gca aac acg gac cgt      726
Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr Asp Arg
              20               25               30

gcc ttc ctc cag gat ggt ttc tcc ttg agc aac aat tct ctc ctg gtc      774
Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu Leu Val
          35               40               45

tagaaaaaat aattgatttc aagaccttct ccccattctg cctccattct gaccatttca     834

ggggtcgtca ccacctctcc tttggccatt ccaacagctc aagtcttccc tgatcaagtc     894
```

<210>   <212> A

<200> 2

<211> 82 amino acids   <214> L

<400> 2
Met Thr Pro Pro Glu Arg Leu Phe Leu
Pro Arg Val Cys Gly Thr Thr
-30 -25 -20
Leu His Leu Leu Leu Leu Gly Leu Leu
Leu Val Leu Leu Pro Gly Ala

-15 -10 -5
His Gly Leu Met His Leu Ala His Ser
Asn Leu Lys Pro Ala Ala His
1 5 10
Leu Ile Tyr Pro Ser Lys Gln Asn Ser Leu
Leu Trp Arg Ala Asn Thr

15 20 25 30
Asp Arg Ala Phe Leu Gln Asp Gly Phe
Ser Leu Ser Asn Asn Ser Leu
35 40 45
Leu Val <
BILLING CODE 3510–16–P

<200> 2

<210>

<211> 82 amino acids

<212> A

<214> L

<400> 2

Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
          -30   -25   -20

Leu His Leu Leu Leu Leu Gly Leu Leu Leu Val Leu Leu Pro Gly Ala
        -15   -10   -5

His Gly Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His
          1   5   10

Leu Ile Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr
   15   20   25   30

Asp Arg Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu
                35   40   45

Leu Val <

9. Appendix B to Subpart G is proposed to be removed.

[Appendix B To Subpart G of Part 1— Headings For Information Items In § 1.823

(1) GENERAL INFORMATION:
(i) APPLICANT:
(ii) TITLE OF INVENTION:
(iii) NUMBER OF SEQUENCES:
(iv) CORRESPONDENCE ADDRESS:
  (A) ADDRESSEE:
  (B) STREET:
  (C) CITY:
  (D) STATE:
  (E) COUNTRY:
  (F) ZIP:
(v) COMPUTER READABLE FORM:
  (A) MEDIUM TYPE:
  (B) COMPUTER:
  (C) OPERATING SYSTEM:

(D) SOFTWARE
(vi) CURRENT APPLICATION DATA:
   (A) APPLICATION NUMBER:
   (B) FILING DATE:
   (C) CLASSIFICATION:
(vii) PRIOR APPLICATION DATA:
   (A) APPLICATION NUMBER:
   (B) FILING DATE:
(viii) ATTORNEY/AGENT INFORMATION:
   (A) NAME:
   (B) REGISTRATION NUMBER:
   (C) REFERENCE/DOCKET NUMBER:
(ix) TELECOMMUNICATIONS
      INFORMATION:
   (A) TELEPHONE:
   (B) TELEFAX:
   (C) TELEX:
   (2) INFORMATION FOR SEQ ID NO: X:
(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH:
   (B) TYPE:
   (C) STRANDEDNESS:
   (D) TOPOLOGY:
(ii) MOLECULE TYPE:
   —Genomic RNA;
   —Genomic DNA;
   —mRNA;
   —tRNA;
   —rRNA;
   —snRNA;
   —scRNA;
   —preRNA;
   —cDNA to genomic RNA;
   —cDNA to mRNA;
   —cDNA to tRNA;
   —cDNA to rRNA;
   —cDNA to snRNA;
   —cDNA to scRNA;
   —Other nucleic acid;
   (A) DESCRIPTION:
   —protein and
   —peptide.
(iii) HYPOTHETICAL:
(iv) ANTI-SENSE:
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
   (A) ORGANISM:
   (B) STRAIN:
   (C) INDIVIDUAL ISOLATE:
   (D) DEVELOPMENTAL STAGE:
   (E) HAPLOTYPE:
   (F) TISSUE TYPE:
   (G) CELL TYPE:
   (H) CELL LINE:
   (I) ORGANELLE:
(vii) IMMEDIATE SOURCE:
   (A) LIBRARY:
   (B) CLONE:
(viii) POSITION IN GENOME:
   (A) CHROMOSOME/SEGMENT:
   (B) MAP POSITION:
   (C) UNITS:
(ix) FEATURE:
   (A) NAME/KEY:
   (B) LOCATION:
   (C) IDENTIFICATION METHOD:
   (D) OTHER INFORMATION:
(x) PUBLICATION INFORMATION:
   (A) AUTHORS:
   (B) TITLE:
   (C) JOURNAL:
   (D) VOLUME:
   (E) ISSUE:
   (F) PAGES:
   (G) DATE:
   (H) DOCUMENT NUMBER:

   (I) FILING DATES:
   (J) PUBLICATION DATE:
   (K) RELEVANT RESIDUES:
(xi) SEQUENCE DESCRIPTION: SEQ ID
    NO:X:]

   Dated: September 23, 1996.

Bruce A. Lehman,
Assistant Secretary of Commerce and Commissioner of Patents and Trademarks.
[FR Doc. 96–25074 Filed 10–3–96; 8:45 am]
BILLING CODE 3510–16–P

ENVIRONMENTAL PROTECTION AGENCY

40 CFR Part 281

[FRL–5620–3]

Alabama; Approval of State Underground Storage Tank Program

AGENCY: Environmental Protection Agency.

ACTION: Notice of tentative determination on application of State of Alabama for final approval, public hearing and public comment period.

SUMMARY: The State of Alabama has applied for approval of its underground storage tank program for petroleum and hazardous substances under Subtitle I of the Resource Conservation and Recovery Act (RCRA). The Environmental Protection Agency (EPA) has reviewed Alabama's application and has made the tentative decision that Alabama's underground storage tank program for petroleum and hazardous substances satisfies all of the requirements necessary to qualify for approval. Alabama's application for approval is available for public review and comment. A public hearing will be held to solicit comments on the application, unless insufficient public interest is expressed.

DATES: A public hearing is scheduled for November 14, 1996, unless insufficient public interest is expressed in holding a hearing. EPA reserves the right to cancel the public hearing if sufficient public interest is not communicated to EPA in writing by November 4, 1996. EPA will determine by November 8, 1996, whether there is significant interest to hold the public hearing. The State of Alabama will participate in the public hearing held by EPA on this subject. Written comments on Alabama's approval application, as well as requests to present oral testimony, must be received by the close of business on November 4, 1996.

ADDRESSES: Copies of Alabama's approval application are available during the hours of 9:00 am to 5:00 pm

at the following addresses for inspection and copying:

Alabama Department of Environmental Management, Ground Water Branch, 1751 W. L. Dickinson Drive, Montgomery, AL 36130, phone: (334) 270–5655.

U.S. EPA Docket Clerk, Office of Underground Storage Tanks, 1235 Jefferson Davis Highway, Arlington, VA 22202, phone: (703) 603–9231; and,

U.S. EPA Region 4, Underground Storage Tank Section, Atlanta Federal Center, 15th Floor, 100 Alabama Street SW., Atlanta, GA 30303, phone: (404) 562–9438.

   Written comments should be sent to Mr. John K. Mason, Chief of Underground Storage Tank Section, U.S. EPA Region 4, 100 Alabama Street S.W., Atlanta, Georgia 30303, telephone (404) 562–9438.

   Unless insufficient public interest is expressed, EPA will hold a public hearing on the State of Alabama's application for program approval on November 14, 1996, at 7:00 pm at the Alabama Department of Environmental Management Hearing Room, 1751 W.L. Dickinson Drive, Montgomery, Alabama 36109–2608. Anyone who wishes to learn whether or not the public hearing on the State's application has been cancelled should telephone the following contacts after November 8, 1996:

Mr. John K. Mason, Chief, Underground Storage Tank Section, U.S. EPA Region 4, 100 Alabama Street SW., Atlanta, GA 30303, phone: (404) 562–9438, or

Ms. Sonja Massey, Chief, Ground Water Branch, Alabama Department of Environmental Management, 1751 Congressman W.L. Dickinson Drive, Montgomery, AL 36130, phone: (334) 270–5655.

FOR FURTHER INFORMATION CONTACT:
Mr. John K. Mason, Chief, Underground Storage Tank Section, U.S. EPA Region 4, 100 Alabama Street SW., Atlanta, GA 30303, phone: (404) 562–9438.

SUPPLEMENTARY INFORMATION:

A. Background

   Section 9004 of the Resource Conservation and Recovery Act (RCRA) authorizes EPA to approve State underground storage tank programs to operate in the State in lieu of the Federal underground storage tank (UST) program. Program approval may be granted by EPA pursuant to RCRA Section 9004(b), if the Agency finds that the State program is "no less stringent" than the Federal program for the seven elements set forth at RCRA Section