WIPO

PCT/MIA/VI/15
ORIGINAL: English
DATE: February 17, 1997

WORLD INTELLECTUAL PROPERTY ORGANIZATION
GENEVA

INTERNATIONAL PATENT COOPERATION UNION
(PCT UNION)

MEETING OF INTERNATIONAL AUTHORITIES UNDER THE PCT

Sixth Session
Canberra, February 17 to 21, 1997

ESTABLISHMENT OF A UNIFORM STANDARD FOR THE PRESENTATION OF NUCLEOTIDE AND/OR AMINO ACID SEQUENCE LISTINGS FOR INTERNATIONAL APPLICATIONS

Proposal by the European Patent Office, the Japanese Patent Office and the United States Patent and Trademark Office

1. The European Patent Office, the Japanese Patent Office and the United States Patent and Trademark Office (the "Trilateral Offices"), in the course of their trilateral cooperation, have agreed on a uniform standard for the presentation of nucleotide and/or amino acid sequence listings for international applications.

2. The Annex to this document contains a proposal by the Trilateral Offices for such a uniform standard.

3. The Meeting is invited to consider the proposal of the Trilateral Offices contained in the Annex to this document.

[Annex follows]

ANNEX

12-02-97 21:25 St1801r

Trilateral Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in Patent Applications (version 18/01/1997)

INTRODUCTION

1. This Standard has been elaborated so as to provide standardization of the presentation of nucleotide and amino acid sequence listings in national, regional and international patent applications. It is based upon the WIPO Standards ST. 23 and ST. 24, and includes provisions for the presentation by applicants of nucleotide and amino acid sequence listings both on paper and on electronic data carriers. The Standard is intended to allow the applicant to draw up a single sequence listing on paper and in machine-readable form which is acceptable to all receiving Offices and International Searching and Preliminary Examining Authorities, and to all designated and elected Offices for the purposes of the national phase. It is intended to enhance the accuracy and quality of presentations of nucleotide and amino acid sequences given in international applications, to make for easier interpretation of sequences by applicants, the public and examiners, to facilitate searching of sequence data and to allow the introduction of sequence data onto computerized databases and the exchange of sequence data in electronic form.

DEFINITIONS

2. For the purposes of this Standard:

   (i) the expression "Sequence Listing" means a separate part of the description of the application as filed or a separate document filed subsequently to the application, which gives a detailed disclosure of the nucleotide and/or amino acid sequences and other available information.

   (ii) the expressions "nucleotide sequence" and "amino acid sequence" mean an unbranched sequence of ten or more nucleotides or an unbranched sequence of four or more contiguous amino acids, respectively. Sequences comprising nucleotides or amino acids other than those listed in paragraph 8, 9 and 11,12 are specifically excluded from this definition. Nucleotides and amino acids are further defined as follows:

      (a) "Nucleotides" embrace only those nucleotides that can be represented using the symbols set forth in paragraph 8. Modifications, e.g. methylated bases, may be described as set forth in paragraph 9, but shall not be shown explicitly in the nucleotide sequence.

(b) "Amino acids" are those L-amino acids commonly found in naturally occurring proteins and are listed in 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 paragraph 11 with the modified positions, e.g. hydroxylations or glycosylations, being described as set forth in 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 paragraph 11 in conjunction with a description elsewhere to describe, for example, abnormal linkages, cross-links and end caps, non-peptidyl bonds, etc., is embraced by this definition.

SEQUENCE LISTING

3. The sequence listing as defined in paragraph 2(i) shall, where it is part of the description of the application as filed, be placed after the last sheet of the abstract or, where applicable, after the last sheet of the drawings. This part shall be entitled "Sequence Listing", shall begin on a new page and shall have independent page numbering. The sequence listing is an integrate part of the description.

4. Where the sequence listing as defined in paragraph 2(i) is a separate document filed subsequently to the application, it shall be entitled "Sequence Listing" and shall have independent page numbering. A subsequently filed sequence listing does not form part of the application. Therefore, it can not be used for the purpose of determining the content of the original disclosure of the application. This does not apply to the corrected version (e.g. filed under Rule 26 or 91 PCT) of a sequence listing which was part of the description of the application as filed.

5. Each sequence should be assigned a separate integer identifier. The integer identifiers should begin with 1 and increase sequentially by integers. If no sequence is present for an integer identifier, the words "This sequence is omitted" should appear following the integer identifier. The number of sequences presented in the Sequence Listing should be indicated.

6. In the description, claims or drawings of the application, the sequences represented in the Sequence Listing shall be referred to by the integer identifier.

7. Nucleotide and amino acid sequences should be represented by at least one of the following three possibilities:

(i) a pure nucleotide sequence;

(ii) a pure amino acid sequence;

(iii) a nucleotide sequence together with its corresponding amino acid sequence.

12-02-97 21:25 St1801r

SYMBOLS TO BE USED

8. A nucleotide sequence shall be presented only by a single strand, in the 5'-end to 3'-end direction from left to right. The 5'-end and 3'-end symbols shall not be represented in the sequence.

9. The bases of a nucleotide sequence shall be represented using the one-letter code for nucleotide sequence characters. Only lower case letters in conformity with the following list given in Annex 2 shall be used:

10. Modified bases shall be represented as the corresponding unmodified bases in the sequence itself and the modification shall be further described in an accompanying feature using the codes given below. The codes from the list given in Annex 2 may be used in the description or the Feature section of the Sequence Listing but not in the sequence itself.

    If a specific unmodified base symbol from paragraph 8 is not used to designate a modified or unusual base, then the symbol "n" should be listed in the sequence, with further information given in the Feature section of the Sequence Listing (see also para. 30). The symbol "n" is not to be used at the 5' and/or 3' end of the nucleotide sequence, with the exception of non-contiguous segments of a larger sequence described elsewhere in the sequence listing.

11. The amino acids in a protein or peptide sequence shall be listed in the amino acid to carboxy direction from left to right, and the amino and carboxy groups shall not be represented in the sequence.

12 The amino acids shall be represented using the three-letter code with the first letter as a capital and shall conform to the list given in Annex 2.

13. 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 also further described in the Feature section of the Sequence Listing. The codes from the list given in Annex 2 may be used in the description or the Sequence Listing but not in the sequence itself.If a specific unmodified amino acid symbol from table 4 in Annex 2 is not used to designate a modified or unusual amino acid, then the symbol "Xaa" should be listed in the sequence, with further information given in the Feature section of the Sequence Listing. The symbol "Xaa" is not to be used at the carboxy- and/or amino-terminus of the amino acid sequence, with the exception of non-contiguous segments of a larger sequence described elsewhere in the sequence listing.

FORMAT TO BE USED

14. A nucleotide sequence shall be listed with a maximum of 60 bases per line, with a space between each group of 10 bases.

15. A protein or peptide sequence shall be listed with a maximum of 15 amino acids per line, with a space provided between each amino acid.

16. The bases of a nucleotide sequence (including introns) should 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 non-coding parts of a sequence, should be grouped together and

separated from adjacent groups by a space (see Annex 3, Example 10, nucleotides 356-360).

17. The bases of the coding parts of a nucleotide sequence should be listed as triplets (codons).

18. Amino acids corresponding to the codons in the coding parts of a nucleotide sequence should be typed immediately under the corresponding codons. Where a codon is split by an intron, the amino acid symbol should be typed below the portion of the codon containing two nucleotides (see Annex 2, Example 11, position 120).

19. The enumeration of the nucleotide bases shall start at the first base of the sequence with number 1. It shall be continuous through the whole sequence in the direction 5' to 3'. It 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. The enumeration method for nucleotide sequences set forth above remains applicable to nucleotide sequences that are circular in configuration, with the exception that the designation of the first base of the nucleotide sequence may be made at the option of the applicant.

20. The enumeration of amino acids shall start at the first amino acid of the protein, with number 1. It shall be marked in the right margin, next to the line containing the three-letter codes for the amino acids, and giving the number of the last amino acids of that line. Optionally, the amino acids preceding the mature protein, for example pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, when present, may have negative numbers, counting backwards starting with the amino acid next to number 1. Zero (0) is not used when the numbering of amino acids uses negative numbers to distinguish the mature protein. The enumeration method for amino acid sequences set forth above remains applicable for amino acid sequences that are circular in configuration, with the exception that the designation of the first amino acid of the amino acid sequence may be made at the option of the applicant.

21. A partial sequence—made up of one or more non-contiguous segments of a larger sequence or of segments from different sequences—shall be numbered as a separate sequence, with a separate sequence identifier. A sequence with a gap or gaps shall be numbered as a plurality of separate sequences with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data.

OTHER AVAILABLE INFORMATION IN THE SEQUENCE LISTING

22. The order of the items of information in the Sequence Listings shall follow the order in which those items are listed in the list of Numeric identifiers as defined in Annex 1 to this Standard.

23. Only Numeric Identifiers of data elements as defined in Annex 1 to this Standard shall be used for the presentation of the items of information in the Sequence Listings. The equivalent Data Element Headings shall not be used. The provided information shall follow immediately after the Numeric Identifier while only those Numeric Identifiers for which information is given need appear on the Sequence Listing. Generally, a blank line shall be inserted between Numeric Identifiers when the digit in the first position of the Numeric Identifier changes. Additionally, a blank line shall precede any repeated Numeric Identifier.

MANDATORY DATA ELEMENTS

12-02-97 21:25 St1801r

24. The Sequence Listing shall include, in addition to and immediately preceding the actual nucleotide and/or amino acid sequence, the following items of information defined in Annex 1 to this Standard (mandatory data elements):

<110>	Applicant name
<120>	Title of invention
<160>	Number of sequences
<210>	Sequence ID Number
<212>	Length
<214>	Type
<215>	Organism
<400>	Sequence description

The data elements, except <110>,<120> and <160> shall be repeated for each sequence included in the sequence listing.

25. In addition to the data elements identified in paragraph 23 above, when a Sequence Listing is filed at the same time as the application to which it pertains or at any time prior to the assignment of an application number, the following data element shall be used:

<130>	File reference

26. In addition to the data elements identified in paragraph 23 above, when a Sequence Listing is filed in response to a request from an international authority or at any time following the assignment of an application number, the following data elements shall be used:

<140>	International application number
<141>	International filing date

27. In addition to the data elements identified in paragraph 23 above, when a Sequence Listing is filed relating to an application which claims the priority of an earlier application, the following data elements shall be used:

<150>	Earlier application number
<151>	Earlier application filing date

28. In addition to the data elements identified in paragraph 23 above, the following identifiers are mandatory only under the following conditions:
    1. if "N", "Xaa" or a modified or unusual L-amino acid or modified base was used in sequence
    2. if ORGANISM is Artificial Sequence or Unknown

<220>	FEATURE
<221>	NAME/KEY
<222>	LOCATION
<223>	OTHER INFORMATION:

OPTIONAL DATA ELEMENTS

29. All data elements defined in Annex 1 to this Standard, not mentioned in paragraph 23 to 27, above, are optional.

30 When such optional data elements are presented in machine readable form, they should be presented in accordance with paragraph 36 of this Standard.

31. When features of sequences including other information (i.e. identifier <223>) are presented, they shall be described in accordance with the DDBJ/EMBL/GenBank Feature Table (Table 5 and 6 of Annex 2 to this Standard), i.e. the "feature keys" in the controlled vocabulary set out in this table shall be used.

FREE TEXT

32. "Free text" is a wording describing characteristics of the sequence under the identifier <223>which are not included in the DDBJ/EMBL/GenBank Feature Table.

33. The use of free text shall be limited to a few short terms indispensable for the understanding of the sequence. It shall not exceed 3 lines with a maximum of 65 characters per line for each given data element.

34. Any free text should preferably be in the English language.

35. Where the Sequence Listing Annex is filed together with the international application and contains free text, and where the free text is in English but English is not the language in which the application is filed, the description and, where applicable, the claims shall contain that free text in the language of the description. This may be required by designated or elected offices or other national authorities as a condition that the free text be considered as part of the original disclosure of the invention in the application as filed.

COMPUTER-READABLE FORM OF THE SEQUENCE LISTING

36. A copy of the Sequence Listing shall also be submitted in computer-readable form, in addition to the Sequence Listing as contained in the typed patent application, whenever

this is required by the competent Authority. The Sequence Listing both on paper and in computer-readable form shall contain the Numeric Identifiers as shown in Annex 1 to this Recommendation. In the Sequence Listing, the provided information should follow immediately after the Numeric Identifier. Only those Numeric Identifiers for which information is given need appear in the Sequence Listing. Generally, a blank line should be inserted between Numeric Identifiers when the digit in the second position of the Numeric Identifier changes. An exception to this general rule is that no blank line should appear preceding Numeric Identifier <310>. Additionally, a blank line should precede any repeated Numeric Identifier. The examples as contained in Annex 3 to this Standard illustrate the use of Numeric Identifiers in several languages.

37. The entire printable copy of the Sequence Listing shall be contained within one file preferably on a single diskette or any other electronic medium that is acceptable to the International Searching Authority. The file recorded on the diskette or any other electronic medium that is acceptable to the International Searching Authority shall be encoded using IBM[2] Code Page 437, IBM Code Page 932[3] or a compatible code page, as long as no other character code set is specified in paragraph 36. A compatible code page, as would be required for Cyrillic, Arabic, Greek, Hebrew, etc., characters, is one that assigns the Roman alphabet and numerals to the same hexadecimal positions as do the specified code pages.

38. The computer-readable form may be created by any means, such as word processors, nucleotide/amino acid sequence editors, computer editors, dedicated software such as PatentIn or other custom computer programs;

39. 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 40 .

40. The diskette or any other electronic medium that is acceptable to the competent Authority shall have a label permanently affixed thereto on which has been hand printed in block capitals or typed, the name of the applicant, the title of the invention, a reference number and the date on which the data were recorded on the diskette or any other electronic medium that is acceptable to the competent Authority. If applicable, the selected International Searching Authority should be indicated.

41. If the diskette or any other electronic medium that is acceptable to the competent Authority is submitted after the date of filing of an application, the labels shall also include the international filing date of the application and the international application number.

42. As set forth in paragraph 36 above, any means may be used to create the computer-readable form, as long as the Sequence Listing on a submitted diskette or any other electronic medium that is acceptable to the competent Authority is readable under the DOS computer/operating-system.

REPLACEMENT OF THE COMPUTER-READABLE FORM

43. Any amendments to a Sequence Listing printed on paper shall be accompanied by a substitute copy of the computer-readable form, including all previously submitted data

---

[2] IBM is a registered trademark of International Business Machine Corporation, USA
[3] The specified code pages are de facto standards for personal computers

with the amendments incorporated therein, if an initial computer-readable form was submitted with the application.

STATEMENT RELATING TO ANY MACHINE READABLE FORM

44. Any machine readable form shall be accompanied by a declaration that *"the information recorded on the form is identical to the written sequence listing"*.

SUBSEQUENTLY FILED SEQUENCE LISTING

45. Any sequence listing furnished after the filing of the application shall be accompanied by a statement to the effect that the sequence listing does not include matter which goes beyond the disclosure in the application as filed. This means that a sequence listing furnished subsequently to the filing of the application shall contain only those sequences that have been disclosed in the application as filed . Moreover, in all cases the original numbering of the sequences in the application as filed shall be maintained in the subsequently filed sequence listing.

nnexes follow]

12-02-97 21:25 St1801r

ANNEX 1

NUMERIC IDENTIFIERS

Only Numeric Identifiers as defined below shall be used in Sequence Listings submitted in patent applications. The text of the Data Element Headings given below shall not be included in the Sequence Listings

Numeric Identifiers of mandatory data elements which, i.e., data shall be included in all Sequence Listings (see Sequence Listing Standard, paragraph 23, are marked by the symbol M,(optional elements by the symbol O) and include items 100, 110, 120, 130, 160, 210, 212, 214, 215 and 400.

Numeric Identifiers of data elements that are mandatory in the circumstances specified in the Sequence Listing Standard, paragraphs 9, 12, 25 and 26, are items 140, 141, 150 and 151 and 220 to 223.

NUMERIC IDENTIFIER	DATA ELEMENT HEADING	M/O	Comment
<100>	GENERAL INFORMATION	M	
<110>	APPLICANT NAME	M	
<120>	TITLE OF INVENTION	M	
<130>	FILE REFERENCE	M	
<140>	CURRENT APPLICATION NUMBER	M, if available	
<141>	CURRENT FILING DATE	M, if available	
<150>	EARLIER APPLICATION NUMBER	M, if applicable	
<151>	EARLIER FILING DATE	M, if applicable	
<160>	NUMBER OF SEQUENCES	M	
<170>	SOFTWARE	O	
<180>	PRIOR ART REFERENCE	O	list of prior art sequences referenced in the application
<210>	INFORMATION FOR SEQ ID NO: X	M	
<211>	SEQUENCE LOCATION REFERENCE	O	indicate where the sequence is described, e.g. example 1, claim 4, figure 8
<212>	LENGTH	M	sequence length expressed in number of base pairs or amino acids
<213>	TOPOLOGY	O	l/c
<214>	TYPE	M	type of molecule sequenced in SEQ ID NO:X, either DNA, RNA PRT
<215>	ORGANISM	M	Genus species(i.e. latin name)or Artificial Sequence or Unknown

12-02-97 21:25 St1801r

<220>	FEATURE	M, under the following conditions: 1. if "N", "Xaa" or a modified or unusual L-amino acid or modified base was used in sequence 2. if ORGANISM is Artificial Sequence or Unknown Otherwise O	description of points of biological significance in the sequence in SEQ ID NO:X) (may be repeated depending on the number of features indicated
<221>	NAME/KEY	M, under the following conditions: 1. if "N", "Xaa" or a modified or unusual L-amino acid or modified base was used in sequence 2. if ORGANISM is Artificial Sequence or Unknown Otherwise O	only those keys as described in Table 5 or 6 of Annex 2 shall be used.
<222>	LOCATION	M, under the following conditions: 1. if "N", "Xaa" or a modified or unusual L-amino acid or modified base was used in sequence 2. if ORGANISM is Artificial Sequence or Unknown Otherwise O	- from (number of first base/amino acid in the feature) - to (number of last base/amino acid in the feature) - base pairs (numbers refer to positions of base pairs in a nucleotide sequence) - amino acids (numbers refer to positions of amino acid residues in an amino acid sequence) - whether feature is located on the complementary strand to that filed in the Sequence Listing

12-02-97 21:25 St1801r

<223>	OTHER INFORMATION:	M, under the following conditions: 1. if "N", "Xaa" or a modified or unusual L-amino acid or modified base was used in sequence 2. if ORGANISM is Artificial Sequence or Unknown Otherwise O	any other relevant information
<300>	PUBLICATION INFORMATION	O	repeat section for each relevant publication
<301>	AUTHORS	O	
<302>	TITLE	O	title of publication
<303>	JOURNAL	O	journal name in which data published
<304>	VOLUME	O	journal volume in which data published
<305>	ISSUE	O	journal issue number in which data published
<306>	PAGES	O	journal page numbers in which data published
<307>	DATE	O	journal date in which data published
<308>	DATABASE ACCESSIONNUMBER	O	accession number assigned by database including database name
<309>	DOCUMENT NUMBER	O	document number, for patent type citations only; specify two-letter code, number, kind-of document code
<310>	FILING DATE	O	document filing date, for patent-type citations only
<311>	PUBLICATION DATE	O	document publication date; for patent-type citations only
<312>	RELEVANT RESIDUES IN SEQ ID NO:X: FROM TO	O	
<400>	SEQUENCE DESCRIPTION: SEQ ID NO	M	

12-02-97 21:25 St1801r

Annex 2

Table 1: List of nucleotides

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	unknown or other

12-02-97 21:25 St1801r

Table 2: List of modified nucleotides

Symbol	Meaning
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)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
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

Table 3: List of amino acids

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

12-02-97 21:25 St1801r

Table 4: Modified and unusual amino acids

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

12-02-97 21:25 St1801r

Table 5: Feature keys related to DNA sequences

Key	Description
allele	Related strain contains alternative gene form
attenuator	Sequence related to transcription termination
C_region	Constant region of immunoglobulin light and heavy chain, and T-cell receptor alpha, beta and gamma chains
CAAT_signal	'CAAT box' in eukaryotic promoters
CDS	Sequence coding for amino acids in protein (includes stop codon)
conflict	Independent determinations differ
D-loop	Displacement loop
D-segment	Diversity segment of immunoglobulin heavy chain and T-cell receptor beta-chain
enhancer	Cis-acting enhancer of promoter function
exon	Region that codes for part of spliced mRNA
gene	region of biological interest identified as a gene and for which a name has been assigned.
GC_signal	'GC box' in eukaryotic promoters
iDNA	Intervening DNA eliminated by recombination
intron	Transcribed region excised by mRNA splicing
J_segment	Joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta and gamma-chains
LTR	Long terminal repeat
mat_peptide	Mature peptide coding region (does not include stop codon)
misc_binding	Miscellaneous binding site
misc_difference	Miscellaneous difference feature
misc_feature	Region of biological significance that cannot be described by any other feature
misc_recomb	Miscellaneous recombination feature
misc_RNA	Miscellaneous transcript feature not defined by other RNA keys
misc_signal	Miscellaneous signal
misc_structure	Miscellaneous DNA or RNA structure
modified_base	The indicated base is a modified nucleotide
mRNA	Messenger RNA
mutation	A mutation alters the sequence here
N_region	Extra nucleotides inserted betwen rearranged immunoglobulin segments
old_sequence	Presented sequence revises a previous version
polyA_signal	Signal for cleavage & polyadenylation
polyA_site	Site at which polyadenine is added to mRNA
precursor_RNA	Any RNA species that is not yet the mature RNA product
prim_transcript	Primary (unprocessed) transcript
primer_bind	Non-covalent primer binding site
promoter	A region involved in transcription initiation
protein_bind	Non-covalent protein binding site on DNA or RNA
RBS	Ribosome binding site
rep_origin	Replication origin for duplex DNA
repeat_region	Sequence containing repeated subsequences
repeat_unit	One repeated unit of a repeat_region
rRNA	Ribosomal RNA
S_region	Switch region of immunoglobulin heavy chains
satellite	Satellite repeated sequence
scRNA	Small cytoplasmic RNA
sig_peptide	Signal peptide coding region
snRNA	Small nuclear RNA
source	Biological source of the specified span of sequence
stem_loop	Hair-pin loop structure in DNA or RNA
STS	Sequence Tagged Site
TATA_signal	'TATA box' in eukaryotic promoters
terminator	Sequence causing transcription termination
transit_peptide	Transit peptide coding region

12-02-97 21:25 St1801r

tRNA	Transfer RNA
unsure	Authors are unsure about the sequence in this region
V_region	Variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains.

12-02-97 21:25 St1801r

Table 6: Feature keys related to Protein sequences

Key	Description
CONFLICT	Different papers report differing sequences.
VARIANT	Authors report that sequence variants exist.
VARSPLIC	Description of sequence variants produced by alternative splicing.
MUTAGEN	Site which has been experimentally altered.
MOD_RES	Post-translational modification of a residue.
ACETYLATION	N-terminal or other.
AMIDATION	Generally at the C-terminal of a mature active peptide.
BLOCKED	Undetermined N- or C-terminal blocking group.
FORMYLATION	Of the N-terminal methionine.
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	Of asparagine, aspartic acid, proline or lysine.
METHYLATION	Generally of lysine or arginine.
PHOSPHORYLATION	Of serine, threonine, tyrosine, aspartic acid or histidine.
PYRROLIDONE CARBOXYLIC ACID	N-terminal glutamate which has formed an internal cyclic lactam.
SULFATATION	Generally of tyrosine.
LIPID	
MYRISTATE	Myristate group attached through an amide bond to the N-terminal glycine residue of the mature form of a protein [or to an internal lysine residue.
PALMITATE	Palmitate group attached through a thioether bond to a cysteine residue or through an ester bond to a serine or threonine residue.
FARNESYL	Farnesyl group attached through a thioether bond to a cysteine residue [
GERANYL-GERANYL	Geranyl-geranyl group attached through a thioether bond to a cysteine residue.
GPI-ANCHOR	Glycosyl-phosphatidylinositol (GPI) group linked to the alpha-carboxyl group of the C-terminal residue of the mature form of a protein.
N-ACYL DIGLYCERIDE	N-terminal cysteine of the mature form of a prokaryotic lipoprotein with an amide-linked fatty acid and a glyceryl group to which two fatty acids are linked by ester linkages.
DISULFID	Disulfide bond. The 'FROM' and 'TO' endpoints represent the two residues which are linked by an intra-chain disulfide bond. If the 'FROM' and 'TO' endpoints are identical, the disulfide bond is an interchain one and the description field indicates the nature of the cross-link.
THIOLEST	Thiolester bond. The 'FROM' and 'TO' endpoints represent the two residues which are linked by the thiolester bond.
THIOETH	Thioether bond. The 'FROM' and 'TO' endpoints represent the two residues which are linked by the thioether bond.
CARBOHYD	Glycosylation site. The nature of the carbohydrate (if known) is given in the description field.
METAL	Binding site for a metal ion. The description field indicates the nature of the metal.
BINDING	Binding site for any chemical group (co-enzyme, prosthetic group, etc.). The chemical nature of the group is given in the description field.
SIGNAL	Extent of a signal sequence (prepeptide).
TRANSIT	Extent of a transit peptide (mitochondrial, chloroplastic, or for a microbody).

PROPEP	Extent of a propeptide.
CHAIN	Extent of a polypeptide chain in the mature protein.
PEPTIDE	Extent of a released active peptide.
DOMAIN	Extent of a domain of interest on the sequence. The nature of that domain is given in the description field.
CA_BIND	Extent of a calcium-binding region.
DNA_BIND	Extent of a DNA-binding region.
NP_BIND	Extent of a nucleotide phosphate binding region. The nature of the nucleotide phosphate is indicated in the description field.
TRANSMEM	Extent of a transmembrane region.
ZN_FING	Extent of a zinc finger region.
SIMILAR	Extent of a similarity with another protein sequence. Precise information, relative to that sequence is given in the description field.
REPEAT	Extent of an internal sequence repetition.
Secondary structure	The feature table of sequence entries of proteins whose tertiary structure is known experimentally contains the secondary structure information corresponding to that protein. The secondary structure assignment is made according to DSSP (see Kabsch W., Sander C.; Biopolymers, 22:2577-2637(1983)) and the information is extracted from the coordinate data sets of the Protein Data Bank (PDB). In the feature table only three types of secondary structure are specified : helices (key HELIX), beta-strand (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a `loop' or `random-coil' structure). Because the DSSP assignment has more than the three common secondary structure classes, we have converted the following DSSP assignments to HELIX, STRAND, and TURN:

DSSP	DSSP definition	code SWISS-PROT assignment
H	Alpha-helix	HELIX
G	3(10) helix	HELIX
I	Pi-helix	HELIX
E	Hydrogen bonded beta-strand (extended strand)	STRAND
B	Residue in an isolated beta-bridge	STRAND
T	H-bonded turn (3-turn, 4-turn or 5-turn)	TURN
S	Bend (five-residue bend centered at residue i)	Not specified

One should be aware of the following facts:
a) Segment Length. For helices (alpha and 3-10), the residue just before and just after the helix as given by DSSP participates in the helical hydrogen bonding pattern with a single H-bond. For some practical purposes, one can therefore extend the HELIX range by one residue on each side. E.g. HELIX 25-35 instead of HELIX 26-34. Also, the ends of secondary structure segments are less well defined for lower resolution structures. A fluctuation of +/- one residue is common.
b) Missing segments. In low resolution structures, badly formed helices or strands may be omitted in the DSSP definition.
c) Special helices and strands. Helices of length three are 3-10 helices, those of length four and longer are either alpha-helices or 3-10 helices (pi helices are extremely rare). A strand of length one corresponds to a residue in an isolated beta-bridge. Such bridges can be structurally important.
d) Missing secondary structure. No secondary structure is currently given in the feature table in the following cases:
- No sequence data in the PDB entry.
- Structure for which only C-alpha coordinates are in PDB.
- NMR structure with more than one coordinate data set.

	- Model (i.e. theoretical) structure.
ACT_SITE	Amino acid(s) involved in the activity of an enzyme.
SITE	Any other interesting site on the sequence.
INIT_MET	The sequence is known to start with an initiator methionine.
NON_TER	The residue at an extremity of the sequence is not the terminal residue. If applied to position 1, this signifies that the first position is not the N-terminus of the complete molecule. If applied to the last position, it signifies that this position is not the C-terminus of the complete molecule. There is no description field for this key.
NON_CONS	Non consecutive residues. Indicates that two residues in a sequence are not consecutive and that there are a number of unsequenced residues between them.
UNSURE	Uncertainties in the sequence. Used to describe region(s) of a sequence for which the authors are unsure about the sequence assignment.

12-02-97 21:25 St1801r

Annex 3

SPECIMEN SEQUENCE LISTING

<110> Applicant Plc
<120> Example of a Sequence Listing<130> A1235

<140>EP 91000000

<150> US 07/123456
<151> 30-AUG-1991

<160>1

<170> PatentIn Release #1.0, Version #1.25

<210>: 1
<212> 2654
<214> DNA
<215> Homo Sapiens

<221> TATA-signal
<222> 339..344
<221> signal peptide
<222> 824..1009
<221> Misc_feature
<224> human lymphotoxin gene

<400>

```
aagggtgcag          agatgttata   tatgattgct   cttcagggaa   ccggcctcca   gctcacaccc    60
cagctgctca          accgcctcct   ctctgaattg   actgtccctt   ctttggaact   ctaggcctga   120
cccctctccc          tggcccttcc   agcccaacgt   tcccctgacc   cgactccctt   tccaggaact   180
cagtcgcctg          aaccccccagc  ctgtggttct   ctcctaggcc   tcagccttte   ctgcctttga   240
ctgaaacagc          agtatcttct   acacgctggg   gattcccgcg   gcccagcccc   gacctagaac   300
ccgcccgctg          cctgccaccc   tgccactgcc   gattcctcta   taaagggacc   tgagcgtccg   360
cgccgcaggg          ctccacaccag  caggtgaggc   tctcctgccc   catctccttg   ggctgcccgt   420
gattcgtgct          ttggactacc   gcccccgagtg  tcctgccctc   tgcctgggcc   tcggtccctc   480
ctgcacctgc          tgcctggatc   cccggcctgc   ctgggcctgg   gcctttggtgg  gtttggtttt   540
ggtttcattc          tctgtctctg   actctccatc   tgtcagtctc   attgtctctg   tcacacattc   600
tctgtttctg          ccatggttcc   tctctgttcc   attcctgtct   ctctctgtct   ccctctgctc   660
accttggggt          ttctctgact   gcatcttgtc   cccttctctg   tccgatctct   ctctcggggg   720
tcgggggctg          ctgtctccca   gggcgggagg   tctgtcttcc   gccgcgtgcc   acgccccgct   780
cactgtctct          ctctctctct   ctctttctct   gcaggtttct   ccc atg aca   cca cct     835

gaa   cgt  ctc  ttc  ctc  cca  agg  gtg  tgt  ggc  acc  acc  cta  cac ctc  ctc
           B83
Glu   Arg  Leu  Phe  Leu  Pro  Arg  Val  Cys  Gly  Thr  Thr  Leu  His Leu  Leu          20
-30                 -25                      -20                  -15

ctt  ctg   ggg  ctg  ctg  ctg  gtt  ctg  ctg  cct   ggg  gcc   cag   gtgaggcagc          932
Leu  Leu   Gly  Leu  Leu  Leu  Val  Leu  Leu  Pro   Gly  Ala   Gln
       33
               -10                 -5

aggagaatgg       gggctgctgg          ggtggctcag    ccaaacattg    agccctagag    cccccctcaa    992

ctctgttctc     ctag ggg   ctc  cct  ggt   gtt  ggc  ctc  aca  cct  tca  gct  gcc
       1042
                                Gly  Leu   Pro  Gly  Val  Gly  Leu  Thr  Pro  Ser  Ala
          Ala                45
                              1                   5                   10

cag   act  gcc  cgt  cag  cac  ccc  aag  atg  cat  ctt  gcc  cac  agc  aac  ctc
       1090
Gln   Thr  Ala  Arg  Gln  His  Pro  Lys  Met  His  Leu  Ala  His  Ser  Asn  Leu
       60
```

```
        15                  20                  25

aaa cct     ggt ggt cac ctc att g   gtaaacatcc     acctgacctc  ccagacatgt
Lys Pro     Ala Ala His Leu Ile                                            67
            30

ccccaccagc      tctcctccta      ccccctgcctc  aggaaacccaa  gcatccaccc  ctctcccccа  1202
acttcccccа      cgctaaaaaa      aacagaggga   gcccactcct   atgcctcccc  ctgccatccc  1262
ccaggaactc      agttgttcag       tgcccaattc   ctcagggatt   gagacctctg  atcccagacc  1322
ctgatctccc      acccccatcc       cctatggctc   ttcctag ga   gac ccc agc aag  cag  1376
                                             Gly  Asp     Pro Ser Lys  Gln
                                                   35                  40

aac tca ctg ctc tgg  aga  gca  aac  acg  gac  agt  gcc  ttc  ctc cag gat        1424
Asn Ser Leu Leu Trp  Arg  Ala  Asn  Thr  Asp  Arg  Ala  Phe  Leu Gln Asp        89
                45                        50                   55

ggt ttc tcc ttg agc  aac  aat  tct  ctc  ctg  gtc  ccc  acc  agt ggc atc        1472
Gly Phe Ser Leu Ser  Asn  Asn  Ser  Leu  Leu  Val  Pro  Thr  Ser Gly Ile       105
            60                        65                   70

tac ttc gtc tac tcc  cag  gtg  gtc  ttc  tct  ggg  aaa  gcc  tac tct c        1520
Tyr Phe Val Tyr Ser  Gln  Val  Val  Phe  Ser  Gly  Lys  Ala  Tyr Ser Pro      121
        75                        80                   85

aag gcc acc tcc tcc  cca  ctc  tac  ctg  gcc  cat  gag  gct  cag ctc ttc        1568
Lys Ala Thr Ser Ser  Pro  Leu  Tyr  Leu  Ala  His  Glu  Val  Gln Leu Phe      137
      90                        95                  100

tcc tcc cag tac ccc  ttc  cat  gtg  cct  ctc  ctc  agc  tcc  cag aag atg        1616
Ser Ser Gln Tyr Pro  Phe  His  Val  Pro  Leu  Leu  Ser  Ser  Gln Lys Met      151
105             110                       115                  120

gtg tat cca ggg ctg  cag  gaa  ccc  tgg  ctg  cac  tcg  atg  tac cac ggg        1664
Val Tyr Pro Gly Leu  Gln  Glu  Pro  Trp  Leu  His  Ser  Met  Tyr His Gly      167
                125                       130                  135

gct gcg ttc cag ctc  acc  cag  gga  gac  cag  cta  tcc  acc  cac aca gat        1712
Ala Ala Phe Gln Leu  Thr  Gln  Gly  Asp  Gln  Leu  Ser  Thr  His Thr Asp      183
            140                       145                  150

ggc atc ccc cac cta  gtc  ctc  agc  cct  agt  act  gtc  ttc  ttt gga ggc        1760
Gly Ile Pro His Leu  Val  Leu  Ser  Pro  Ser  Thr  Val  Phe  Phe Gly Ala      199
        155                       160                  165

ttc gct ctg tagaacttgg aaaaatccag   aaagaaaaaa    taattgattt   c                1810
Phe Ala  Leu                                                                 202
      170

aagaccttct   ccccattctg            cctccattct   gaccatttca   ggggtcgtca   ccaactctcc
             1870
tttggccatt   ccaacagctc            aagtcttccc   tgatcaagtc   accggagctt   tcaaagaagg  1930
aattctaggc   atcccagggg            accccacacct  ccctgaacca   tccctgatgt   ctgtctggct  1990
gaggatttca   agcctgccta            ggaattccca   gcccaaaget   gttggtcttg   tcccaccagc  2050
tagytggggc   ctagatccac            acacagagga   agagcaggca   catggaggag   cttgggggat  2110
gactagaggc   agggagggga            ctatttatga   aggccaaaaa   attaaattat   ttatttatgg  2170
aggatggaga   gaggggaata            atagaagaac   atccaaggag   aaacagagac   aggcccaaga  2230
gatgaagagt   gagagggcat            gcgcacaagg   ctgaccaaga   gagaaagaag   taggcatgag  2290
ggatcacagg   gccccagaag            gcagggaaag   gctctgaaag   ccagctgccg   accagagccc  2350
cacacggagg   catctgcacc            ctcgatgaag   cccaataaac   atcttttctc   tgaaatgctg  2410
tctgcttgtg\  tgtgtgtgtc            tgggagtgag   aacttcccag   tctatctaag   gaatggaggg  2470
agggacagag   ggctcaaagg            gacgaagagc   tgtgggggga   acaaaaggat   aagggctcga  2530
gagagcttca   aggatatgtg            atggatcacc   aggtgaggcc   gccagactgc   tgcagggyaa  2590
gcaaaggaga   agctgagaag            atgaaggaaa   agtcagggtc   tggaggyyyy   ggggtcaggg  2650
agct                                                                             2654
```

[End of Annex and of Standard]

Report drawn up by the EPO

Report of the **Video Conference** between the JPO, USPTO and EPO held January 21, 1997, 13.00- 14.55 (European time)

Copies to: **Participants to the Video Conference**
Mr. J. Michel, Mr. C. Jonckheere (DG1)
Ms. C. Acéti, Mr. Y. Busse, Mr. A. Clarke (DG5)

Participants: Mr. S. Kunin, Mr. A. Purcell, Ms. E. Kepplinger, Ms. J. Knoblock, Mr. J. Doll, Mr. M. Moore (all USPTO); Mr. N. Takebayashi, Mr. J. Yuita, Ms. S. Kato (all JPO); Mrs. L. Gruszow, Ms. N. Bosard, Mr. J. Descamps, Mr. S. Nauche, Mr. A.J. van Putten (all EPO).

Introduction

JPO and USPTO agreed that the meeting was chaired by Arnold van Putten of the EPO. After the introduction of the different participants the chairman stated that both the EPO and the USPTO had prepared an agenda for this meeting. However, a comparison of both agenda did not reveal many differences. The JPO did not receive the agenda as prepared by the USPTO. Moreover, the addendum to the EPO agenda that had been faxed just prior to the meeting was not received by the JPO and the USPTO. It was agreed that the meeting would be based on the EPO agenda. Basically 4 issues need to be addressed in the broad context of this subject, namely:

1. ST. 23/ST. 24: Trilateral standard for sequence listings
2. PatentIn/Receipt system
3. Framework issues
4. Suggested next steps

The latter three subjects are discussed in the USPTO agenda. In addition, the EPO proposed to add the issue of the mega-sequences to the agenda. The chairman feared that due to time constraints it would not be possible to discuss all points, and suggested to start with discussing the different paragraph of the Trilateral standard for sequence listings. The discussions are summarised in the table below:

Report drawn up by the EPO

Paragraph	Proposal	Comments/ Action to be taken	Status
1	The present Standard is drafted for a situation where the applicants files both a paper and a computer readable copy of the sequence listing. The JPO already has a system in place where the applicant files only the computer readable copy of the sequence listing (the EPO and USPTO are in the process of implementing such a system). What additions need to be made to the text of para. 1 and what are the consequences for 36-42(JPO).	JPO will draft wording to be included in paragraphs 1, 36, 37 and 43. USPTO will draft text for paragraph 37 and 42 **JPO comments 31/1/1997:** Replace "on paper and on electronic data carrier" in the 5th line, by "on application medium and in computer readable form". Replace "on paper and in machine readable form" in the 6th line, by "on application medium and in computer readable form".	Pending
2	Proposal to replace the text " four or more contiguous amino acids" by " Sequences with fewer than 4 specifically defined nucleotides or amino acids are specifically excluded" (USPTO, see rationale in Note Art Purcell). See also the last parts of par. 10 and 13.	JPO would like to reflect on this and will answer later **JPO comments 31/1/1997:** USPTO's proposal is basically acceptable. However, if our understanding is correct, the phrase "Sequence with fewer than 4 specifically defined nucleotides or amino acids" should be added after the phrase "four or more contiguous amino acids (Ref. US Federal Register/Vol. 61).	Pending
2	Proposal to replace "Sequence Listing Annex" by "Sequence Listing" as in the current practice (EPO).	None	Accepted
2	Proposal to replace: "D-amino acids" by "at least 1 D-amino acid"(EPO)	None	Accepted
3	Insert "as a last separate part of the description"	In contrast with the current practice (WIPO ST. 23), it	Pending

	between "placed" and "after" (JPO). Delete the phrase "shall have independent numbering".(JPO).	had been proposed previously to have separate numbering for the sequence listings (MIA Meeting, EPO and USPTO). USPTO drew the attention to the facts are closely related and that these could not be seen separately. Advantages of separate numbering (treatment of sequence listings similar to drawings) are that complete re-numbering of the description is avoided if number of sequences changes and that a limited amount of text in the language other than the application could be allowed. JPO indicated that this might cause difficulties under the JPO law. However, it is not clear whether the JPO proposal is compatible with the current wording that the sequence listing would anyhow be placed after the last sheet of the abstract, or where applicable , the drawings. The JPO indicated that it would like to reflect on this matter. **JPO comments 31/1/1997:** Insert "as a last separate part of the description or" between "placed" and "after in 2nd line. Insert "preferably" between "shall" and "have" in 4th line or Insert ",when the competent Authority requires," between "and" and "shall" in the 4th line. **EPO comments 4/2/1997:** The EPO has suggested to add the phrase "It is therefore unnecessary to describe the sequences elsewhere in the description, claims or, where	

		applicable, drawings."	
4	In the last sentence, insert "or amended" between "corrected" and "version" (JPO). The EPO proposes either to delete entirely the wording "This does not apply...as filed." or to complement the sentence by "; the applicant's right to file amendments to the description (e.g. under Art. 34 (2) (b) PCT) is not affected."	USPTO is in favour of deleting all text related to corrections or amendment. JPO is in favour of adding the phrase on amendments proposed by the EPO. Both JPO and USPTO will submit wording. **JPO comments 31/1/1997:** We prefer to add the phrase on amendments proposed by the EPO. However, we will accept any other wording of the same meaning.	Pending
10/13	Should Xaa and n be defined as 1 modified nucleotide or amino acid to avoid that applicants define Xaa as a string of amino acids (EPO question?)	EPO will propose the wording for the definition that Xaa and n represent 1 amino acid and 1 nucleotide respectively. With this addition, terminal Xaa's and n's will be acceptable, hence parts of parts of par. 10 and 13 need to be deleted. **JPO comments 31/1/1997:** EPO proposal is acceptable.	Pending
15	Proposal to have 16 amino acids per line (JPO, supported by USPTO).	None	Accepted
16	Delete "(see Annex 3, Example 10, nucleotides 356-360)	None	Accepted
18	Delete "(see Annex 2, Example 11, position 120)	None	Accepted
19/20		**Additional note from EPO:** It is observed that the database producers have a numbering in the left margin. In order to stay close to the database producers standard, left margin numbering might be considered.	
20	Proposal to make negative numbering of amino acids	None	Accepted

Report drawn up by the EPO

	(pre-sequences etc.) optional (EPO)		
20	Proposal to mark numbering of amino acids in the right margin (EPO)	USPTO would like to reflect on this while JPO was to the opinion that this will put an unnecessary burden on the applicant. **EPO comments 4/2/1997:** Additional note from EPO: It is observed that the database producers have a numbering in the left margin. In order to stay close to the database producers standard, left margin numbering might be considered.	Pending
24	In the current draft the amount of general information is limited to: Applicant name Title of Invention File reference Application data (date and number) Prior application data (date and number) JPO proposal to further reduce this to: File reference Application data (date and number) Prior application data (date and number)	EPO indicated their wish to maintain applicant name and title of invention since these can be used as checking items, to make sure that the right floppy disk is processed. JPO will reflect on this taking into account any result from the public hearings. **EPO comments 4/2/1997:** The EPO is not in favour of having data on a label exclusively as suggested by the USPTO. We experienced bad data on labels. Presently we use the title and applicant's name to check the authenticity electronically. Hence the EPO is in favour of having the applicant's name and title as mandatory "general information" items **JPO comments 5/2/1997:** We agree to make "applicant name" and "title of invention" mandatory, providing that these elements are language dependent and can be described in languages other than English (e.g. French, German or Japanese).	Pending

24	Proposal to make Organism mandatory item (EPO)	The Offices agreed to make Organism a mandatory item. However, USPTO is not in favour of making a feature mandatory if Organism is artificial because of the risk of introducing lengthy free text necessary tom explain the artificial organism. Both JPO and EPO are in favour of this proposal. USPTO will send comments on this (e.g. example legal sequences).	Pending
24	Proposal to make Topology mandatory item (USPTO)	USPTO will inquire whether this is still a requirement for the search algorithms used.	Pending
25	How do applicants describe file reference (question JPO)	None	Clarified
33	Proposal to have 4 lines of free text (USPTO proposal)	The Offices agreed that it should be avoided the applicants use the free text field for writing long parts of the description. Hence, a limitation on the amount of free text is desirable., also in view of the fact that the applicant will (if the free text is in English) have to repeat the text in the description). USPTO proposed to change the present wording as follows:: ....indispensable for the understanding of the sequence, normally not exceeding 3 lines with a maximum .... JPO expressed concerns about the amount of characters since this is different for Japanese and Latin alphabets. The Offices will provide proposals for new wording for this paragraph. **EPO comments 4/2/1997:** The EPO believes that it is important to give the message to the applicants that the text in field in	Pending

		<223> shall be "limited to a few short terms indispensable for understanding of the sequence". Moreover, we believe that this should be feasible in the four lines as proposed since it should be kept in mind that the applicants , for the same sequence and even for the same amino acid or nucleotide, can give as many <220> FEATURE fields as the applicant wants. **JPO comments 5/2/1997:** We support the EPO proposal.	
35	Free text is only the wording that appears under the number <223>. The current draft proposes that if the free text of the sequence listing is in English but English is not the language of the application, the free text should also appear elsewhere in the description to ensure that the free text is considered as part of the original disclosure of the invention in the application as filed. However, it is not specified where this free text should appear in the description.	The EPO indicated that the use of the word "Glossary" as proposed by the JPO would not be desirable in view of the use of the word Glossary in earlier drafts of the standard. The EPO proposed to add as a **recommendation** that the free text information in the language of the application should be placed in a special section of the description called "Sequence Listing Free Text". This was accepted by the three Offices. The EPO will propose modified wording for this paragraph. **Proposal**: The free text in the language of the application may be put in a specific section of the description called "Sequence Listing Free Text". **JPO comments 31/1/1997:** The EPO proposal is acceptable.	
35	Why limitation to International?(JPO question)	EPO comment: International shall be deleted	Accepted
35	Delete "and, where applicable the claims" (EPO proposal)	None	Accepted

Report drawn up by the EPO

35	4th line: Insert "also" between "shall" and "contain" and replace "description" by "application".	None	Accepted
36	".	**JPO comments 31/1/1997:** Replace "on paper and in computer readable form" in the 3rd line, by "on application medium and in computer readable form	
37	Is file recorded on diskette text file only (JPO question)	USPTO replied that the sequence listing is always in ASCII format.	Accepted
37	Do the compatible code pages include Japanese (JPO question)	USPTO referred to discussions that took place during the drafting of ST. 24 and confirmed that the code pages include Japanese **EPO suggestion:** add Japanese to the languages listed in this paragraph.	Pending
38	It would be desirable if the applicants could be encouraged to use the Patentin software. A proposed wording could be " The computer-readable form may be created by any means, but preferably by dedicate software such as Patentin or other custom computer programs." (EPO)	**JPO comments 31/1/1997:** We support the EPO proposal to recommend Patentin or other custom computer programs.	To be discussed
40	Insert "Name and type of computer and operating system" on the label (JPO suggestion)	The size of the label should be considered None	Accepted
40	Insert "Correspondence address etc." on the label	None	Accepted

Report drawn up by the EPO

	(USPTO suggestion)		
42	Replace "DOS" by more modern terminology (EPO)	**JPO comments 31/1/1997:** To use more modern terminology is preferable.	Not yet discussed
44	Replace "declaration" by "statement" (JPO proposal supported by EPO).	This corresponds to the WIPO text. None	Accepted
Annex 1	<100> Should these identifiers without information be required?	The offices agreed that these could be omitted	Accepted
Annex 1	<180> Is this solution for the inclusion of prior art sequences feasible?	**JPO comments 31/1/1997:** We support the EPO proposal to refer Prior art by accession numbers.	Not yet discussed
Annex 1	<211> Should the Sequence Location Reference be maintained (e.g. Claims, Description, Figures)	The USPTO argued that if the references are made too specific (e.g. page nr.) and that if changes are made to the description or number of pages, the sequence listing has to be adapted in subsequent procedures. However, if the wording chosen is too general , its use becomes questionable. The JPO would consider this element but only as optional. The EPO will provide further information on this topic. **EPO comments 4/2/1997:** Identifier <211> is of valuable information for the examiners in particular when a document has been retrieved from the database search. It is worth mentioning that GENESEQ (Derwent's patent sequence database) already uses such a notification	Pending
Examples		Need to be improved	Pending

Report drawn up by the EPO

Annex 1	Add "strandedness" as optional item		Not yet discussed
Examples	Need to be improved, e.g. to include an example of free text	Not done yet, the EPO will prepare appropriate examples.	
General	Should a limitation on sequence length be put in the standard since search algorithms can only deal with a maximum length.		Not yet discussed
General	Lay-out	A review of the lay-out will be done by the EPO	Pending

Conclusions

The chairman noticed with satisfaction that good progress had been made during the video conference on the different open issues for the Trilateral Standard for sequence listings. The most important issue that could not be discussed was the requirement to include prior art sequences in the sequence listings. However, it was felt that this could be discussed by E-mail exchange. The USPTO emphasised that several issues need to be addressed once the standard has been established, examples are the PatentIn/Receipt system and the implementation of the Standard. It was suggested to take up these issues immediately once agreement on the Standard has been reached. The Offices agreed that the requested information will be provided within 1 week - 10 days, which will allow submission of the draft Trilateral proposal to the International Bureau around February 1, 1997 in order to be able to discuss the paper at MIA6 in Canberra. Before closing the meeting the EPO asked the USPTO whether the requested information on Mega-sequences could be sent to the other partners by E-mail. Before closing the meeting the chairman thanked all participants and those who were involved in the preparations of the video conference. The meeting was closed at 14.55 (European time).

[End of annex and of document]