NOTICE: This file contains information that was previously published in the WIPO *Handbook on Industrial Property Information and Documentation*, but that has become outdated.

STANDARD ST.24

RECOMMENDATION CONCERNING THE FILING OF NUCLEOTIDE AND AMINO ACID SEQUENCE LISTINGS IN COMPUTER-READABLE FORM

INTRODUCTION

1. WIPO Standard ST.23--Recommendation for the Presentation of Nucleotide and Amino Acid Sequence Listings in Patent Applications and in Published Patent Documents, provides specific recommendations, including the standardized format, for the disclosing of nucleotide and amino acid sequences in patent applications in a "Sequence Listing." However, WIPO Standard ST.23 makes no provision for the presentation by applicants of Sequence Listings on electronic data carriers, which is advisable for industrial property offices wishing to introduce an electronic filing system and to use computerized search systems.

2. This Recommendation has been elaborated in order to provide standardization in the filing of Sequence Listings in computer-readable form so as to facilitate searching of sequence data, the introduction of sequence data into computerized databases and the exchange of sequence data in electronic form.

DEFINITIONS

3. For the purposes of this Recommendation:

(i) the expression "patent application" means a document filed by an applicant, or by an agent (representative) on his (her) behalf, requesting the grant of a patent. It usually contains a detailed description of the invention, the claims, and drawings when necessary for the understanding of the invention;

(ii) the expression "Sequence Listing" means a last, separate part of the description in the typed patent application, which gives a detailed disclosure of the nucleotide and/or amino acid sequences and other available information. Each sequence disclosed appears separately in the Sequence Listing and is assigned a separate identifier number. Symbols and format used for sequences and other available information follow the recommendations given in WIPO Standard ST.23.

COMPUTER-READABLE FORM OF THE SEQUENCE LISTING

4. A copy of the Sequence Listing referred to in paragraph 3(ii) above should also be submitted in computer-readable form, in addition to the Sequence Listing as contained in the typed patent application. The Sequence Listing should include the Data Element Headings or the Numeric Identifiers as contained in Annex 1 to this Recommendation. When a Sequence Listing is prepared using the Numeric Identifiers, the provided information may 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 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 Numeric Identifiers of Annex 1 may also be used in a paper Sequence Listing as recommended in WIPO Standard ST.23. The examples as contained in Annex 2 to this Recommendation illustrate the alternative use of Numeric Identifiers and Data Element Headings in several languages.

5. The computer-readable form should comprise a printable copy of the Sequence Listing recorded on diskette. The computer-readable form should 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 configuration specified in paragraph 12 of this Recommendation.

6. The entire printable copy of the Sequence Listing should preferably be contained within one file on a single diskette. The file recorded on the diskette should be encoded using IBM Code Page 437*, IBM Code Page 932 or a compatible code page, as long as no other character code set is specified in paragraph 12. 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. The specified code pages are de facto standards for personal computers.

7. 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; however, it should be readable by one of the computer/operating-system configurations specified in paragraph 12 and should conform to the specifications given in paragraphs 5 and 6 of this Recommendation.

8. The submitted diskette should be write-protected.

9. The diskette should have a label permanently affixed thereto on which has been hand printed in block capitals or typed, a description of the format of the diskette as well as the name of the applicant, the title of the invention, a reference number, the date on which the data were recorded on the diskette and the name and type of computer and operating-system which generated the file on the diskette. The code page used should also be identified when a code page other than that specified in paragraphs 6 and 12 is used.

10. If all of this information cannot be printed on a label affixed to the diskette, by reason of size or otherwise, the label should 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 diskette with the name of the applicant, the title of the invention, the reference number, and the additional information affixed to the container.

11. If the diskette is submitted after the date of filing of an application, the labels should also include the filing date of the application and the application number.

12. As set forth in paragraph 7, above, any means may be used to create the computer-readable form, as long as the Sequence Listing on a submitted diskette is readable on one of the computer/operating-system configurations or comply with the specified format described in subparagraphs (a), (b) and (c), below. If the Sequence Listing is in the Japanese language, the computer/operating-system, media, character codes and file specifications of paragraph (c) should apply.

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

Operating system: PC-DOS or MS-DOS** (Versions 2.1 or above);
Line Terminator: Carriage Return plus Line Feed;
Pagination: Form Feed or Series of Line Terminators;
End-of-File: Ctrl-Z;
Media: Diskette - 5.25 inch, 360 Kb storage;
Diskette - 5.25 inch, 1.2 Mb storage;
Diskette - 3.50 inch, 720 Kb storage;
Diskette - 3.50 inch, 1.44 Mb storage;
Print Command: PRINT filename.extension;

---

* IBM is a registered trademark of International Business Machine Corporation, US.
** MS-DOS is a registered trademark of Microsoft Corporation, US.

WIPO FOR OFFICIAL USE ONLY

(b) Computer: Apple Macintosh*;

Operating system: Macintosh;
Macintosh File Type: Text with line termination;
Line Terminator: Pre-defined by text type file;
Pagination: Pre-defined by text type file;
End-of-File: Pre-defined by text type file;
Media: Diskette - 3.50 inch, 400 Kb storage;
Diskette - 3.50 inch, 800 Kb storage;
Diskette - 3.50 inch, 1.4 Mb storage;
Print Command: Use PRINT command from any Macintosh application that processes text files, such as MacWrite** or TeachText.

(c) Computer: In accordance with the electronic application standards of the Japanese Patent Office (JPO);

Operating system: In accordance with the JPO's electronic application standards;

Software: In accordance with the JPO's electronic application standards;

File Specification: File specification for Japanese Documents Interchange (JIS X4004-1988);

Character codes: Code of Japanese graphic character set for information interchange (JIS X 0208-1983);

Media: Diskette - 8.00 inch, both-sided double density (2d) (JISX6201);
Diskette - 5.25 inch, high density (2HD) (JIS X6211);
Diskette - 3.50 inch, high density (2HD) (JIS X6223).

REPLACEMENT OF THE COMPUTER-READABLE FORM

13. Any amendments to the paper copy of the Sequence Listing should be accompanied by a substitute copy of the computer-readable form, including all previously submitted data with the amendments incorporated therein, if an initial computer-readable form was submitted with the patent application.

IMPLEMENTATION OF THIS RECOMMENDATION

14. Each office or other entity choosing to adopt this Recommendation should publish an announcement to that effect in its official journal or gazette. Such an announcement should include an identification of the Sequence Listing Data Elements which the office holds to be mandatory, a specification of the format (as provided in paragraph 12) and of the code page(s) (as provided in paragraph 6) that are to be used, whether the office will accept a Sequence Listing in Data Element Heading form, in Numeric Identifier form or in either form and, if not previously published, the language(s) to be used.

[Annexes follow]

* Apple and Macintosh are registered trademarks of Apple Computer Inc., US.
** MacWrite is a registered trademark of Claris Corporation, US.

WIPO FOR OFFICIAL USE ONLY

ANNEX I

DATA ELEMENT HEADINGS AND NUMERIC IDENTIFIERS

Data Element Heading	Numeric Identifier	ST.23 Number
GENERAL INFORMATION:	<100>	
APPLICANT:	<110>	
NAME:	<111>	
STREET:	<112>	
CITY:	<113>	
STATE OR PROVINCE:	<114>	
COUNTRY:	<115>	
POSTAL CODE:	<116>	
TELEPHONE:	<117>	
TELEFAX;	<118>	
TELEX OR ELECTRONIC MAIL:	<119>	
TITLE OF INVENTION:	<120>	
NUMBER OF SEQUENCES:	<130>	
CORRESPONDENCE ADDRESS:	<140>	
ADDRESSEE:	< 141>	
STREET:	<142>	
CITY:	<143>	
STATE OR PROVINCE:	<144>	
COUNTRY:	<145>	
POSTAL CODE:	<146>	
COMPUTER READABLE FORM:	<150>	
MEDIUM TYPE:	<151>	
COMPUTER:	<152>	
OPERATING SYSTEM:	<153>	
SOFTWARE:	<154>	
CURRENT APPLICATION DATA:	<160>	
APPLICATION NUMBER:	<161>	
FILING DATE:	<162>	
CLASSIFICATION:	<163>	
PRIOR APPLICATION DATA:	<170>	
APPLICATION NUMBER:	<171>	
FILING DATE:	<172>	
CLASSIFICATION:	<173>	
ATTORNEY/AGENT INFORMATION:	<180>	
NAME:	<181>	
REGISTRATION NUMBER:	<182>	
REFERENCE/DOCKET NUMBER:	<183>	
TELECOMMUNICATION INFORMATION:	<190>	
TELEPHONE:	<191>	
TELEFAX:	<192>	
TELEX OR ELECTRONIC MAIL:	<193>	
INFORMATION FOR SEQ ID NO: X	<200>	1
SEQUENCE CHARACTERISTICS:	<210>	1.i
LENGTH:	<211>	1.i.A
TYPE:	<212>	1.i.B
STRANDEDNESS:	<213>	1.i.C
TOPOLOGY:	<214>	1.i.D

MOLECULE TYPE:	<220>	1.ii
HYPOTHETICAL:	<230>	1.iii
ANTI-SENSE:	<240>	1.iv
FRAGMENT TYPE:	<250>	1.v
ORIGINAL SOURCE:	<260>	1.vi
ORGANISM:	<261>	1.vi.A
STRAIN:	<262>	1.vi.B
INDIVIDUAL ISOLATE:	<263>	1.vi.C
DEVELOPMENTAL STAGE:	<264>	1.vi.D
HAPLOTYPE:	<265>	1.vi.E
TISSUE TYPE:	<266>	1.vi.F
CELL TYPE:	<267>	1.vi.G
CELL LINE:	<268>	1.vi.H
ORGANELLE:	<269>	1.vi.I
IMMEDIATE SOURCE:	<270>	1.vii
LIBRARY:	<271>	1.vii.A
CLONE:	<272>	1.vii.B
POSITION IN GENOME:	<280>	1.viii
CHROMOSOME/SEGMENT:	<281>	1.viii.A
MAP POSITION:	<282>	1.viii.B
UNITS:	<283>	1.viii.C
FEATURE:	<290>	1.ix
NAME/KEY:	<291>	1.ix.A
LOCATION:	<292>	1.ix.B
IDENTIFICATION METHOD:	<293>	1.ix.C
OTHER INFORMATION:	<294>	1.ix.D
PUBLICATION INFORMATION:	<300>	
AUTHORS:	<301>	
TITLE:	<302>	
JOURNAL:	<303>	
VOLUME:	<304>	
ISSUE:	<305>	
PAGES:	<306>	
DATE:	<307>	
DOCUMENT NUMBER:	<308>	
FILING DATE:	<309>	
PUBLICATION DATE:	<310>	
RELEVANT RESIDUES IN SEQ ID NO:X: FROM TO	<311>	
SEQUENCE DESCRIPTION: SEQ ID NO:	<400>	1.xi

[Annex 2 follows]

WIPO FOR OFFICIAL USE ONLY

ANNEX 2

EXAMPLES OF SEQUENCE LISTINGS IN DIFFERENT LANGUAGES

Example 1

German Language Sequence Listing using Numeric Identifiers

Example 2

French Language Sequence Listing using Numeric Identifiers

Example 3

English Language Sequence Listing using Numeric Identifiers

Example 4

Japanese Language Sequence Listing using Numeric Identifiers

Example 5

German Language Sequence Listing using Data Element Headings

Example 6

French Language Sequence Listing using Data Element Headings

Example 7

English Language Sequence Listing using Data Element Headings

Example 8

Japanese Language Sequence Listing using Data Element Headings

WIPO FOR OFFICIAL USE ONLY

Annex 2
EXAMPLE 1, PAGE 1

SEQUENZPROTOKOLL

<111> Anmelder GmbH
<112> Erhardtstrasse 27
<113 > Miinchen
<115> DEUTSCHLAND
<116> W-8000
<117> 089-23990
<118> 089-23994465
<119> 523656

<120> Beispiel eines Sequenzprotokolls

<130> 2

<151> Floppy disk
<152> IBM PC compatible
<153> PC-DOS/MS-DOS
<154> Patentin Release #1.0, Version #1.25 (EPA)

<161> EP 91000000

<200> 1

<211> 954 Basenpaare
<212> Nukleinsâure
<213> einzel
<214> linear

<230> JA

<240> NEIN

<261> Paramecium sp
<263> XYZ2
<267> unicellular organism

<271> genomic
<272> Para-XyZ2/36

<291> CDS
<292> join(275,.373, 448..498, 679..774)

<291> mat_peptide
<292> join(451..498, 679..774)

<291> sig_peptide
<292> 275..448

WIPO FOR OFFICIAL USE ONLY

Annex 2
Example 1, page 2

```
<301> Doe, Joan X
<301> 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
<3Q7> MAR-2-1988
<311> RELEVANTE RESTE IN SEQ ID NO: 1: VON 1 BIS 957

<400> 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      292
                                      Met Thr Pro Pro Glu Arg
                                      -34             -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
                                                            -1

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 Gin 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 Gin Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu Leu Val
        35                  40                  45
```


WIPO FOR OFFICIAL USE ONLY

Annex 2

Example 1, page 3

```
TAGAAAAAAT AATTGATTTC AAGACCTTCT CCCCATTCTG CCTCCATTCT GACCATTTCA   834

GGGGTCGTCA CCACCTCTCC TTTGGCCATT CCAACAGCTC AAGTCTTCCC TGATCAAGTC   894

ACCGGAGCTT TCAAAGAAGG AATTCTAGGC ATCCCAGGGG ACCCACACCT CCCTÇAACCA   954

<200> 2

<211> 82 Aminosâuren

<212> Aminosâure

<214> linear

<220> Protein

<400> 2

Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
-34             -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 Gin Asn Ser Leu Leu Trp Arg Ala Asn Thr
 15                  20                  25                  30

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

Leu Val
```


WIPO FOR OFFICIAL USE ONLY

Annex 2
Example 2, page 1

LISTAGE DE SEQUENCES

<111> Déposant SA
<112> 26bis, rue de St.-Petersbourg
<113> Paris Cedex 08
<115> France
<116> 75800
<117> 331-42 94 52 52
<119> 290 368 INPI Paris

<120> Exemple de listage de séquences

<130> 2

<151> Floppy disk
<152> IBM PC compatible
<153> PC-DOS/MS-DOS
<154> Patentin Release #1.0, Version #1.25 (OEB)

<161> EP 91000000

<200> 1

<211> 954 paires de bases
<212> acide nucléique ^
<213> simple I
<214> linéaire

<230> OUI

<240> NON

<261> Paramecium sp
<263> XYZ2
<267> unicellular organism

<271> genomic
<272> Para-XYZ2/36

<291> CDS
<292> join(275..373, 448..498, 679..774)

<291> mat_peptide
<292> join(451..498, 679..774)

<291> sig_peptide
<292> 275..448

WIPO FOR OFFICIAL USE ONLY

Annex 2
Exemple 2, page 2

```
<301> Doe, Joan X
<301> 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> MAR-2-1988
<311> RESIDUS PERTINENTS DANS LA SEQ ID NO: 1: DE 1 A 957

<400> 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     292
                                      Met Thr Pro Pro Glu Arg
                                      -34             -30

CTC TTC CTO CCA AGG GTG TGT GGC ACC ACC CTA CAO 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
                                                             -1

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 Gin 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 Gin Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu Leu Val
         35                  40                  45
```


WIPO FOR OFFICIAL USE ONLY

Annex 2
Example 2, page 3

```
TAGAAAAAAT AATTGATTTC AAGACCTTCT CCCCATTCTG CCTCCATTCT GACCATTTCA   834

GGGGTCGTCA CCACCTCTCC TTTGGCCATT CCAACAGCTC AAGTCTTCCC TGATCAAGTC   894

ACCGGAGCTT TCAAAGAAGG AATTCTAGGC ATCCCAGGGG ACCCACACCT CCCTGAACCA   954

<200> 2
<211> 82 acides aminés
<212> acide aminé
<214> linéaire

<220> protéine

<400> 2

Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
-34             -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 Gin Asn Ser Leu Leu Trp Arg Ala Asn Thr
 15                  20                  25                  30

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

Leu Val
```

Annex 2
Example 3, page 1

SEQUENCE LISTING

<111> Applicant Pic
<112> Patentlaan 2
<113> Rijswijk
<114> Zuid-Holland
<115> The Netherlands
<116> 2288 EE
<117> 070-3402040
<118> 340-3403016
<119> 31651 epo ni

<120> Example of a Sequence Listing

<130> 2

<151> Floppy disk
<152> IBM PC compatible
<153> PC-DOS/MS-DOS
<154> Patentln Release #1.0, Version #1.25 (EPO)

<161> EP 91000000

<200> 1
<211> 954 base pairs
<212> nucleic acid
<213> single
<214> linear

<230> YES

<240> NO

<261> Paramecium sp
<263> XYZ2
<267> unicellular organism

<271> genomic
<272> Para-XYZ2/36

<291> CDS
<292> join(275..373, 448,.498, 679.-774)

<291> mat_peptide
<292> join(451..498, 679..774)

<291> sig_peptide
<292> 275..448

WIPO FOR OFFICIAL USE ONLY

Annex 2
Example 3, page 2

<301> Doe, Joan X
<301> 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> MAR-2-1988
<311> RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 957

<400> 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       292
                                      Met Thr Pro Pro Glu Arg
                                      -34             -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
                                                             -1

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 Gin 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 Gin Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu Leu Val
         35                  40                  45
```


WIPO FOR OFFICIAL USE ONLY

Annex 2

Example 3, page 3

```
TAGAAAAAAT AATTGATTTC AAGACCTTCT CCCCATTCTG CCTCCATTCT GACCATTTCA   834

GGGGTCGTCA CCACCTCTCC TTTGGCCATT CCAACAGCTC AAGTCTTCÇC TGATCAAGTC   894

ACCGGAGCTT TCAAAGAAGG AATTCTAGGC ATCCCAGGGG ACCCACACCT CCCTGAACCA   954

<200> 2
<211> 82 amino acids
<212> amino acid
<214> linear

<220> protein

<400> 2

Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
-34             -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 Gin Asn Ser Leu Leu Trp Arg Ala Asn Thr
 15                  20                  25                  30

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

Leu Val
```

Annex 2

Example 4, page 1

```
<111>
<112>
<113>
<115>
<116> 100
<117> 3581-1101
<118> 3595-2735
<120>

<130> 2

<151>
<152> IBM PC
<153> MS-DOS 932
<154>

<161> 4-400000

<200> 1

<211> 954
<212>
<213>
<214>
<220> Genomic DNA

<230> Yes

<240> No

<261>
<263> XYZ2
<267>

<271>   (genomic)
<272> Para-XYZ2/36

<291> CDS
<292> join(275. . 373, 448.. 498. 679.. 774)

<293> E

<291> mat_peptlde
<292> joln(451.. 498. 679.. 774)

<293> E

<291> sig_peptlde
<292> 275. .448
<293> P

<301> (Doe. Jean X)
```

Annex 2

Example 4, page 2

<katakana> (Doe, Joan Q)

<302> <katakana/kanji>
(Isolation and Characterization of a Gene Encodlng a Protease from ParaDeciun sp.)
<303> 7 -f ^ 3 y - V X (Fictional Genes)
<304> 1
<305> 1
<306> 1-20
<3a7> 1988 3 2
<311> 957

<400> 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       292
                                     Met Thr Pro Pro Glu Arg
                                     -34             -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 Hls 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 Hls
        -10                 -5

GGTGGCTCAG CCAAACCTTG AGCCCTAGAG CCCCCCTCAA CTCTGTTCTC CTAG GGG     450
                                                           Gly
                                                            -1

CTC ATG CAT CTT GCC CAC AGC AAC CTC AAA CCT GCT GCT CAC CTC ATT     498
Leu Met Hls Leu Ala Hls Ser Asn Leu Lys Pro Ala Ala Hls 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 GAG CGT     726
Tyr Pro Ser Lys Gin 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 Gin 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
```


WIPO FOR OFFICIAL USE ONLY

Annex 2

Example 4, page 3

<200> 2

<211> 82

<212>

<214>

<220>

<400>

```
Met Thr Pro Fro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
-34             -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 Gin Asn Ser Leu Leu Trp Arg Ala Asn Thr
  15                 20                  25                  30

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

Leu Val
```


WIPO FOR OFFICIAL USE ONLY

Annex 2
Example 5, page 1

SEQUENZPROTOKOLL

ALGEMEINE INFORMATION:

ANMELDER:
NAME: Anmelder GmbH
STRASSE: Erhardtstrasse 27
ORT: MUnchen
LAND: DEUTSCHLAND
POSTLEITZAHL: W-8000
TELEPHON: 089-23990
TELEFAX; 089-23994465
TELEX ODER ELEKTRONISCHE POST: 523656

ANMELDETITEL: Beispiel eines Sequenzprotokolls

ANZAHL DER SEQUENZEN: 2

COMPUTERLESBARE FORM:
DATENTRAGER: Floppy disk
COMPUTER: IBM PC compatible
BETRIEBSSYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPA)

INFORMATION ZU SEQ ID NO: 1:

SEQUENZCHARAKTERISTIKA:
LANGE: 954 Basenpaare
ART: NukleinsSure
STRANGFORM: Einzel
TOPOLOGIE: linear

HYPOTHETISCH: JA

ANTISENSE: NEIN

URSPRUNGLICHE HERKUNFT:
ORGANISMUS: Paramecium sp
INDIVIDUUM ISOLAT: XYZ2
ZELLTYP: unicellular organism

UNMITTELBARE HERKUNFT:
BIBLIOTHEK: genomic
CLON: Para-XYZ2/36

MERKMALE:
NAME/SCHLUSSEL: CDS
LAGE: join(275..373, 448..498, 679..774)

MERKMALE:
NAME/SCHLUSSEL: mat_peptide
LAGE: join(451..498, 679..774)

WIPO FOR OFFICIAL USE ONLY

Annex 2

Example 5, page 2

```
      MERKMALE;
            NAME/SCHLÛSSEL: sig_peptide
            LAGE: 275..448

      VEROFFENTLICHUNGSINFORMATION:
            AUTOREN: Doe, Joan X
                     Doe, John Q
            TITEL: Isolation and Characterization of a Gene
                   Encoding a Protease from Paramecium sp.
            ZEITSCHRIFT: Fictional Genes
            BAND: 1
            AUSGABE: 1
            SEITEN: 1-20
            DATUM: MAR-2-1988
            RELEVANTE RESTE IN SEQ ID NO: 1: VON 1 BIS 957

      SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

ATCGGGATAG TACTGGTCAA GACCGGTGGA CACCGGTTAA CCCCGGTTAA GTACCGGTTA   60

TAGGCCÀTTT CAGGCCAAAT GTGCCCAACT ACGCCAÀTTG 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
                                      -34             -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
                                                            -1

CTC ATG CAT CTT GCC CAC ÀGC 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 Gin Asn Ser Leu Leu Trp Arg Ala Asn Thr Asp Arg
            20                  25                  30
```


WIPO FOR OFFICIAL USE ONLY

Annex 2

Example 5, page 3

```
GCC TTC CTC GAG GAT GGT TTC TCC TTG AGC AAC AAT TCT CTC CTG GTC    774
Ala Phe Leu Gin 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
```

INFORMATION ZU SEQ ID NO: 2:

SEQUENZCHARAKTERISTIKA
LANGE: 82 Aminosâuren
ART: Aminosâure
TOPOLOGIE; linear

ART DES MOLEKÛLS: Protein

SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

```
Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
-34             -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 Gin Asn Ser Leu Leu Trp Arg Ala Asn Thr
  15                 20                  25                  30

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

Leu Val
```