E

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

PCT/WG/4/9
ORIGINAL: ENGLISH
DATE: APRIL 21, 2011

Patent Cooperation Treaty (PCT)
Working Group

Fourth Session
June 6 to 10, 2011, Geneva

Presentation of Sequence Listings Under the PCT

Document prepared by the International Bureau

SUMMARY

1. Discussions have commenced in WIPO's Committee on WIPO Standards (CWS) aimed at establishing a new WIPO XML Sequence Listing Standard. The present document contains a proposal to take the opportunity, arising from the need to modify the present WIPO Sequence Listing Standard ST.25 and the PCT Sequence Listing Standard so as to move both into an XML environment, to review the relationship between these two Standards. The aim of such a review would be to establish, on the one hand, a general "filing-route neutral" WIPO Standard (or Standards) on the presentation, contents and structure of sequence listings in national, regional and international applications (be they submitted on paper, in text format or in XML format) and, on the other hand, Administrative Instructions under the PCT (it is proposed to no longer refer to a PCT "Standard") dealing only with PCT-specific issues.

BACKGROUND

2. At its first session, held from October 25 to 29, 2010, the Committee on WIPO Standards (CWS) discussed a proposal by the European Patent Office (EPO) for a new standard on the presentation of nucleotide and amino acid sequence listings based on eXtensible Markup Language (XML). The discussions of the Committee are set out in paragraphs 26 to 29 of the draft report of that session (document CWS/1/10 Prov.), reproduced in the following paragraphs:

"26. Discussions were based on document CWS/1/5, which contained a proposal for a new standard on the presentation of nucleotide and amino acid sequence listings based on eXtensible Markup Language (XML). The CWS noted the request made by the EPO on the preparation of the said new WIPO standard. For a number of technical and practical reasons, WIPO Standard ST.25 should be replaced, or at least supplemented, by a new standard based on XML format. Such new standard would mitigate the shortcomings of WIPO Standard ST.25 and provide additional advantages for both applicants and IPOs since the drafting and submission of high quality sequence listings would enable more efficient downstream processes.

"27. The CWS also noted that WIPO Standard ST.25 recommended that, *mutatis mutandis*, offices should apply the provisions set out in the Annex C to the Administrative Instructions under the PCT. Therefore, when discussing the proposal for a new standard, due consideration should be given to the impact of the future standard on the current WIPO Standard ST.25, entailing changes to WIPO Standard ST.25 and Annex C to the PCT Administrative Instructions.

"28. The CWS agreed to:

"(a) create the following Task: "Prepare a recommendation on the presentation of nucleotide and amino acid sequence listings based on eXtensible Markup Language (XML) for adoption as a WIPO standard. The proposal of the new WIPO standard should be presented along with a report on the impact of the said standard on the current WIPO Standard ST.25, including the proposed necessary changes to Standard ST.25";

"(b) establish a Task Force to handle the Task;

"(c) request the Task Force to liaise with the appropriate PCT body with regard to the possible impact of such standard on Annex C to the Administrative Instructions under the PCT; and

"(d) request the Task Force to present the proposal of the new WIPO standard and necessary changes to Standard ST.25 for consideration and approval by the CWS at its session to be held in 2011.

"29. The CWS welcomed the offer of the EPO, which was designated as the Task Force Leader".

3. Following the first session of the CWS, the International Bureau, in a Circular dated December 10, 2010 (Circular C.CWS 10), invited offices wishing to participate actively in the Task Force to nominate a representative having the required knowledge of XML technologies and WIPO Standard ST.25 to join the Task Force. In response to Circular C. CWS 10, 10 Offices have expressed an interest in joining the Task Force. In the meantime, work of that Task Force has begun on the establishment of a possible new WIPO XML sequence listing standard, based on a first draft by the EPO as the Task Force Leader.

RELATIONSHIP BETWEEN WIPO STANDARD ST.25 AND THE PCT SEQUENCE LISTING STANDARD

4. As referred to in paragraph 27 of the draft report of the first session of the CWS (reproduced in paragraph 1, above), the present WIPO Standard ST.25 consists of a single paragraph, recommending "that Offices apply the provisions of the "Standard for the Presentation of Nucleotide and Amino Acid Sequence Listings in International Patent Applications Under the Patent Cooperation Treaty (PCT)" as set out in Annex C to the Administrative Instructions under the PCT, *mutatis mutandis*, to all patent applications other than the PCT international applications, noting that certain provisions specific to the PCT procedures and requirements may not be applicable to patent applications other than PCT international applications." In other words, all of the substantive provisions making up the international sequence listing standard are contained in Annex C to the Administrative Instructions under the PCT rather than, as one would perhaps expect, in WIPO Standard ST.25.

5. Looking back, one of the main reasons for this somewhat unusual relationship between the two Standards (with the WIPO Standard referring to the PCT Standard) appears to be that Member States, at the time of the establishment of the first common sequence listing standard back in 1997/1998, found it preferable "to get things off the ground" within the framework of the PCT. The applicability of the PCT Standard was then broadened to applications filed outside of the PCT by referring to the PCT Standard in WIPO Standard ST.25. As a consequence, WIPO Standard ST.25 contains not only details of the presentation of sequence listings but also processing arrangements which are specific to the PCT system only. Despite the fact that WIPO Standard ST.25 expressly states that "certain provisions specific to the PCT procedures and requirements may not be applicable to patent applications other than PCT international applications", the focus of WIPO Standard ST.25 on the PCT adds complexity and can appear confusing.

OPPORTUNITY FOR REVIEW

6. As noted in paragraph 27 of the draft report of the first session of the CWS (reproduced in paragraph 1, above), the establishment of a new WIPO XML Sequence Listing Standard will have an impact on the current WIPO Standard ST.25 and thus also on the PCT Standard for presentation of sequence listings. Both will therefore have to be modified to take account of the fact that applicants will be permitted, and indeed encouraged, to submit sequence listings in XML format.

7. It is in this context that the International Bureau would like to propose to take the opportunity arising from the need to modify WIPO Standard ST.25 and the PCT Sequence Listing Standard to review the relationship between these two Standards.

8. Rather than simply moving WIPO Standard ST.25 into an XML environment and thus "cementing" the PCT focus of the WIPO Standard for the foreseeable future, it proposed to consider whether the relationship between the PCT Standard and the WIPO Standard should be "turned around". This approach would aim to establish, on the one hand, a general "filing-route neutral" WIPO Standard (or Standards) on the presentation, contents and structure of sequence listings in national, regional and international applications (be they submitted on paper, in text format or in XML format) and, on the other hand, Administrative Instructions under the PCT (it is proposed to no longer refer to a PCT "Standard") dealing only with PCT-specific issues but otherwise, with regard to the presentation, contents and structure of sequence listings, referring to the WIPO Standard (or Standards).

9. The substantive contents of a revised Annex C of the Administrative Instructions would thus be limited to:

(a) details of how and where sequence listings should be presented in relation to the remainder of the application body (where the listing is filed in paper form) and in what formats they should be submitted if forming part of the application as filed;

(b) details of whether, how and in what formats other copies of the sequence listing may be submitted, for example, to deal with corrections, rectifications or copies for the purpose of international search or preliminary examination; and

(c) the provisions from existing paragraph 42 of Annex C relating to treatment of the sequence listings before designated and elected Offices.

DISCUSSIONS AT THE EIGHTEENTH MEETING OF INTERNATIONAL AUTHORITIES

10. The issue of a possible modification of the PCT Sequence Listing Standard was discussed at the eighteenth Meeting of International Authorities, held in Moscow from March 15 to 17, 2011. The discussions of the Meeting of International Authorities are outlined in document PCT/MIA/18/16, paragraphs 88 to 92, reproduced in the following paragraphs:

"88. Discussions were based on document PCT/MIA/18/13.

"89. A representative from the European Patent Office, noting that the Office acted as leader of the task force established by the CWS to prepare a recommendation on the establishment of a new XML WIPO sequence listing standard, stated that a first draft of a possible new standard had been posted on the task force's wiki just last week, with comments by task force members expected to be submitted by April 8. The task force's aim was to finalize its discussions by the end of June 2011 with a view to adopting a proposal at the CWS November 2011 meeting. The representative further stated that a software tool had been developed by the European Patent Office to support the new XML standard, for use by applicants and Offices, and that it was envisaged to make that tool available for applicants filing direct European patent applications as of April 2011; that tool could also, if so wished, be made available to applicants and Offices for use in respect of international applications, once the new WIPO XML sequence listing standard and a revised PCT sequence listing standard had been approved and implemented.

"90. All Authorities which took the floor on the matter supported the proposal to review the relationship between the PCT Sequence Listing Standard and WIPO Standard St.25 (and any future WIPO XML sequence listing standard) with a view to establishing, on the one hand, a general "filing-route neutral" WIPO Standard (or Standards) on the presentation of sequence listings in national, regional and international applications (be they in text format or in the XML format) and, on the other hand, a PCT Standard dealing only with PCT-specific issues.

"91. One Authority stated that, while it fully supported the development of a new XML sequence listing standard and the consequential modification of WIPO Standard ST.25 and of the PCT sequence listing standard, it wondered whether the envisaged time table for discussion and adoption of the envisaged new XML standard was overly ambitious. The substantive changes envisaged for WIPO Standard ST.25 and the PCT sequence listing standard required a thorough review and careful consideration. New software had to be developed for applicants to prepare and validate sequence submissions. Offices would need new software for internal validation of submissions, as well as modifications to internal handling

systems. New style sheets were required to render XML submissions readable by humans. Offices needed to change national and regional legal frameworks to ensure that only one standard was effective in a particular Office regardless of whether the sequence listing was filed in a national, a regional or an international application. As undesirable as they might be, transitional provisions could most likely not be avoided; as a result of making substantive changes to WIPO Standard ST.25, a simple conversion to XML would not render an ST.25 sequence listing compliant with the new XML standard.

"92. In response to a comment made by one Authority as to the need for style sheets to make any XML sequence listing readable to humans, the representative of the European Patent Office as the CWS task force leader stated that such style sheets had been prepared and had been posted on the CWS task force's WIKI in the form of an Annex to the main proposal for a new XML sequence listing standard."

PROPOSED DRAFT MODIFICATIONS TO ANNEX C OF THE PCT ADMINISTRATIVE INSTRUCTIONS

11. The Annex to this document contains a preliminary draft of proposed modifications to Annex C of the Administrative Instructions. Those proposed preliminary draft modifications are being submitted to the Working Group only to form the basis for a first round of preliminary discussions as to how to best address the issues set out in paragraphs 6 to 9, above. As usual in the case of proposed modifications of the PCT Administrative Instructions, a formal consultation procedure on a set of proposed modifications will follow (by way of one or more Circulars) at a later stage, depending on the progress achieved in the CWS Task Force.

12. The proposed draft modifications set out in the Annex are certainly incomplete. In particular, the current standard defines a format which is laid out for human viewing in a sufficiently standardized way so that it can be read into a machine. The new format, on the other hand, is mainly intended for machine processing. While it will be possible to print out a sequence listing established in the new format and for a human to understand it directly, it is expected that human inspection would normally be performed with the aid of a computer transformation. Elements of at least paragraphs 3, 5, 6 and 7 of the draft as renumbered include aspects of presentation which only make sense in the context of sequence listings presented for human reading. A final proposal will need to clearly separate aspects of processing requirements which apply to sequence listings filed in an essentially paper/image-based form, those in a text-based format according to the current standard and those in the new XML-based format. The details of this will need to take into account the quality of the tools which are made available in conjunction with the standard for effective viewing of the listings, including the EPO's new free software tool entitled "BISSAP" and any style sheets which are adopted as part of the standard for transforming the XML information into a view arranged for human reading.

FURTHER STEPS

13. As indicated in paragraph 28(c) of document CWS/1/10 Prov. (reproduced in paragraph 2, above), the CWS Task Force is requested to liaise with "the appropriate PCT body" with regard to the possible impact of the new XML Sequence Listing Standard on the PCT Standard. The International Bureau (represented by staff from both the PCT and the WIPO Standards areas) will work closely with the Task Force in developing the new XML Standard and the consequential modifications to WIPO Standard ST.25 and the PCT Standard, taking into account any feedback received from Member States and users during the discussions at the present meeting and at later stages of the discussions.

14. The Working Group is invited to comment on the issues raised in the present document.

[Annex follows]

ANNEX

PRELIMINARY DRAFT OF
PROPOSED MODIFICATIONS TO
ANNEX C OF THE PCT ADMINISTRATIVE INSTRUCTIONS

INSTRUCTIONS RELATING TO ~~STANDARD FOR~~ THE PRESENTATION OF NUCLEOTIDE AND AMINO ACID SEQUENCE LISTINGS IN INTERNATIONAL PATENT APPLICATIONS UNDER THE PCT

INTRODUCTION

1. These Instructions have ~~This Standard has~~ been elaborated so as to provide, together with WIPO Standard ST.XX, standardization of the presentation of nucleotide and amino acid sequence listings in international patent applications. The Instructions are ~~The Standard is~~ intended to allow the applicant to draw up a single sequence listing which is acceptable to all receiving Offices, International Searching and Preliminary Examining Authorities for the purposes of the international phase, 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 presentation and dissemination of sequences for the benefit of applicants, the public and examiners, to facilitate searching of sequence data and to allow the exchange of sequence data in electronic form and the introduction of sequence data onto computerized databases.~~

DEFINITIONS

2. For the purposes of these Instructions ~~this Standard~~:

(i) the expressions ~~expression~~ "sequence listing", "sequences", "nucleotides", "amino acids", "sequence identifiers" and "controlled vocabulary" have the same meaning as defined in WIPO Standard ST.XX; ~~means a nucleotide and/or amino acid sequence listing which gives a detailed disclosure of the nucleotide and/or amino acid sequences and other available information;~~

(ii)~~(i-bis)~~ the expression "sequence listing forming part of the international application" means a sequence listing contained in the international application as filed (as referred to in paragraph 3), including any sequence listing or part thereof which is included in the international application under Rule 20.5(b) or (c), which is considered to have been contained in the international application under Rule 20.6(b), or which has been corrected under Rule 26, rectified under Rule 91 or amended under Article 34(2); or a sequence listing included in the international application by way of a rectification under Rule 91 or an amendment under Article 34(2)(b) of the description in relation to sequences contained in the international application as filed (as referred to in paragraphs 4 and 5 ~~paragraphs 3bis and 3ter~~);

(iii)~~(i-ter)~~ the expression "sequence listing not forming part of the international application" means a sequence listing which does not form part of the international application but is furnished for the purposes of the international search or international preliminary examination (as referred to in paragraphs 6 and 7 ~~paragraphs 4 and 4bis~~);

~~(ii) sequences which are included are any unbranched sequences of four or more amino acids or unbranched sequences of ten or more nucleotides. Branched sequences, sequences with fewer than four specifically defined nucleotides or amino acids as well as sequences comprising nucleotides or amino acids other than those listed in Appendix 2, Tables 1, 2, 3 and 4, are specifically excluded from this definition;~~

~~(iii) "nucleotides" embrace only those nucleotides that can be represented using the symbols set forth in Appendix 2, Table 1. Modifications, for example, methylated bases, may be described as set forth in Appendix 2, Table 2, but shall not be shown explicitly in the nucleotide sequence;~~

~~(iv) "amino acids" are those L-amino acids commonly found in naturally occurring proteins and are listed in Appendix 2, Table 3. Those amino acid sequences containing at least one D-amino acid 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 Appendix 2, Table 3, with the modified positions, for example, hydroxylations or glycosylations, being described as set forth in Appendix 2, Table 4, but these modifications shall not be shown explicitly in the amino acid sequence. Any peptide or protein that can be expressed as a sequence using the symbols in Appendix 2, Table 3, in conjunction with a description elsewhere to describe, for example, abnormal linkages, cross-links (for example, disulfide bridge) and end caps, non-peptidyl bonds, etc., is embraced by this definition;~~

~~(v) "sequence identifier" is a unique integer that corresponds to the SEQ ID NO assigned to each sequence in the listing;~~

~~(vi) "numeric identifier" is a three-digit number which represents a specific data element;~~

~~(vii) "language-neutral vocabulary" is a controlled vocabulary used in the sequence listing that represents scientific terms as prescribed by sequence database providers (including scientific names, qualifiers and their controlled-vocabulary values, the symbols appearing in Appendix 2, Tables 1, 2, 3 and 4, and the feature keys appearing in Appendix 2, Tables 5 and 6);~~

(iv)~~(viii)~~ "competent Authority" is the International Searching Authority that is to carry out the international search and to establish the written opinion of the International Searching Authority on the international application, or the International Preliminary Examining Authority that is to carry out the international preliminary examination on the international application.

SEQUENCE LISTINGS

Sequence Listing Forming Part of the International Application

3. A sequence listing which is contained in the international application as filed:

(i) shall be presented as a separate part of the description, be placed at the end of the application, preferably be entitled "Sequence Listing", begin on a new page and have independent page numbering[1]; preferably, the sequence listing shall not be reproduced in any other part of the application.~~; subject to paragraph 36, it~~ It is unnecessary to describe the sequences elsewhere in the description, except where the sequence listing forming part of the international application contains free text which, in accordance with WIPO Standard ST.XX, is to be repeated in the main part of the description in the language thereof;

---

[1] Editor's Note: No independent page numbering is required where the sequence listing is contained in an international application filed in electronic form and is in the electronic document format referred to in paragraph 12(i) or (ii) ~~paragraph 40~~.

(ii) shall present the sequences represented in the sequence listing and other available information in the sequence listing in accordance with paragraph 8 ~~paragraphs 5 to 35~~;

(iii) if contained in an international application filed in electronic form, shall be in an electronic document format and filed by a means of transmittal in accordance with paragraph 9 ~~paragraph 37~~.

4.~~3bis.~~ Any correction under Rule 26, rectification under Rule 91 or amendment under Article 34(2) of the description submitted in relation to a sequence listing contained in the international application filed on paper and any sequence listing included in the international application by way of a rectification under Rule 91 or an amendment under Article 34(2)(b) of the description in relation to sequences contained in the international application filed on paper shall be submitted in accordance with Rule 26.4, Rule 91 or Rule 66.8, respectively.

5.~~3ter.~~ Any correction under Rule 26, rectification under Rule 91 or amendment under Article 34(2)(b) of the description submitted in relation to a sequence listing contained in the international application filed in electronic form and any sequence listing included in the international application by way of a rectification under Rule 91 or an amendment under Article 34(2)(b) of the description in relation to sequences contained in the international application filed in electronic form shall be submitted in the form of a sequence listing in electronic form comprising the entire listing with the relevant correction, rectification or amendment. Any such sequence listing:

(i) shall preferably be entitled “Sequence Listing – Correction”, “Sequence Listing – Rectification” or “Sequence Listing – Amendment”, as the case may be, and have independent page numbering[1];

(ii) shall present the sequences represented in the sequence listing and other available information in the sequence listing in accordance with paragraph 8 ~~paragraphs 5 to 35~~; where applicable, the original numbering of the sequences in the international application as filed (in accordance with paragraph 8 ~~as referred to in paragraph 5~~) shall be maintained; otherwise, the sequences shall be numbered in accordance with paragraph 8 ~~paragraphs 5~~;

(iii) shall be in an electronic document format and filed by a means of transmittal in accordance with paragraph 10 ~~paragraph 38~~.

Sequence Listing Not Forming Part of the International Application

6.~~4.~~ A sequence listing furnished under Rule 13*ter* for the purposes of the international search or international preliminary examination:

(i) shall preferably be entitled “Sequence Listing – Rule 13*ter*”;

(ii) shall present the sequences represented in the sequence listing and other available information in the sequence listing in accordance with paragraph 8 ~~paragraphs 5 to 35~~; where applicable, the original numbering of the sequences in the international application as filed (in accordance with paragraph 8 ~~as referred to in paragraph 5~~) shall be maintained; otherwise, the sequences shall be numbered in accordance with paragraph 8 ~~paragraphs 5~~;

(iii) if furnished on paper in accordance with Rule 13*ter*.1(b), shall have independent page numbering;

(iv) if furnished in electronic form, shall be in an electronic document format and filed by a means of transmittal in accordance with paragraph 11 ~~paragraph 39~~;

(v) if furnished in electronic form together with the international application, shall be identical to the sequence listing as contained in the application and be accompanied by a statement that "the information recorded in electronic form furnished under Rule 13*ter* is identical to the sequence listing as contained in the international application";

(vi) if furnished subsequently to the filing of the international application, shall not go beyond the disclosure in the international application as filed and be accompanied by a statement to that effect; any such sequence listing shall contain only those sequences that were disclosed in the international application as filed.

7.~~4bis.~~ Any correction under Rule 26, rectification under Rule 91 or amendment under Article 34(2)(b) of the description submitted in relation to a sequence listing contained in the international application as filed and any sequence listing included in the international application by way of a rectification under Rule 91 or an amendment under Article 34(2)(b) of the description in relation to sequences contained in the international application as filed shall be accompanied, for the purposes of the international search or international preliminary examination, by a sequence listing in electronic form in an electronic document format in accordance with paragraph 11 ~~paragraph 39~~, comprising the entire listing including any such correction, rectification or amendment, whenever this is required by the competent authority, unless such listing in electronic form is already available to that authority in a form and manner acceptable to it. Any such sequence listing in electronic form:

(i) shall preferably be entitled "Sequence Listing – Correction – Rule 13*ter*", "Sequence Listing – Rectification – Rule 13*ter*" or "Sequence Listing – Amendment – Rule 13*ter*", as the case may be;

(ii) shall present the sequences represented in the sequence listing and other available information in the sequence listing in accordance with paragraph 8 ~~paragraphs 5 to 35~~; where applicable, the original numbering of the sequences in the international application as filed (in accordance with paragraph 8 ~~as referred to in paragraph 5~~) shall be maintained; otherwise, the sequences shall be numbered in accordance with paragraph 8 ~~paragraphs 5~~;

(iii) shall be filed by a means of transmittal in accordance with paragraph 11 ~~paragraph 39~~;

(iv) shall be identical to the corrected, rectified or amended sequence listing and be accompanied by a statement that "the information recorded in electronic form furnished under Rule 13*ter* is identical to the corrected sequence listing" (or to the "rectified sequence listing" or the "amended sequence listing", as the case may be).

Where such sequence listing in electronic form and, where applicable, such statement is not available to the competent authority, any such correction, rectification or amendment need only be taken into account by that authority for the purposes of the international search or preliminary examination to the extent that a meaningful search or preliminary examination can be carried out without such sequence listing in electronic form.

PRESENTATION OF SEQUENCES; CONTENTS AND STRUCTURE OF SEQUENCE LISTINGS

8.~~5.~~ If the sequence listing is filed:

(i) on paper, sequences shall be presented, and the contents and structure of the sequence listing shall be in compliance with, WIPO Standard ST.XX, paragraphs AA to BB and paragraphs XX to XX;

(ii) in electronic form in the electronic document format referred to in paragraph 12(i), sequences shall be presented, and the contents and structure of the sequence listing shall be in compliance with, WIPO Standard ST.XX, paragraphs AA to BB and paragraphs YY to YY;

(iii) in electronic form in the electronic document format referred to in paragraph 12(ii), sequences shall be presented, and the contents and structure of the sequence listing shall be in compliance with, WIPO Standard ST.XX, paragraphs AA to BB and paragraphs ZZ to ZZ.

~~Each sequence shall be assigned a separate sequence identifier. The sequence identifiers shall begin with 1 and increase sequentially by integers. If no sequence is present for a sequence identifier, the code 000 should appear under numeric identifier <400>, beginning on the next line following the SEQ ID NO. The response for numeric identifier <160> shall include the total number of SEQ ID NOs, whether followed by a sequence or by the code 000.~~

~~6. In the description, claims or drawings of the application, the sequences represented in the sequence listing shall be referred to by the sequence identifier and preceded by "SEQ ID NO:".~~

~~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.~~

~~For those sequences disclosed in the format specified in option (iii), above, the amino acid sequence must be disclosed separately in the sequence listing as a pure amino acid sequence with a separate integer sequence identifier.~~

~~**Nucleotide Sequences**~~

~~*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 terms 3' and 5' 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 list given in Appendix 2, Table 1, shall be used.~~

~~10. Modified bases shall be represented as the corresponding unmodified bases or as "n" in the sequence itself if the modified base is one of those listed in Appendix 2, Table 2, and the modification shall be further described in the feature section of the sequence listing, using the codes given in Appendix 2, Table 2. These codes may be used in the description or the feature section of the sequence listing but not in the sequence itself (see also paragraph 32). The symbol "n" is the equivalent of only one unknown or modified nucleotide.~~

Format to Be Used

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

12. The bases of a nucleotide sequence (including introns) shall be listed in groups of 10 bases, except in the coding parts of the sequence. Leftover bases, fewer than 10 in number at the end of non-coding parts of a sequence, should be grouped together and separated from adjacent groups by a space.

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

14. The enumeration of the nucleotide 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 nucleotide of the sequence may be made at the option of the applicant.

15. A nucleotide sequence that is 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.

Amino Acid Sequences

Symbols to Be Used

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

17. 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 Appendix 2, Table 3. An amino acid sequence that contains a blank or internal terminator symbols (for example, "Ter" or "*" or ".") may not be represented as a single amino acid sequence, but shall be presented as separate amino acid sequences (see paragraph 22).

18. Modified and unusual amino acids shall be represented as the corresponding unmodified amino acids or as "Xaa" in the sequence itself if the modified amino acid is one of those listed in Appendix 2, Table 4, and the modification shall be further described in the feature section of the sequence listing, using the codes given in Appendix 2, Table 4. These codes may be used in the description or the feature section of the sequence listing but not in the sequence itself (see also paragraph 32). The symbol "Xaa" is the equivalent of only one unknown or modified amino acid.

Format to Be Used

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

20. Amino acids corresponding to the codons in the coding parts of a nucleotide sequence shall be placed immediately under the corresponding codons. Where a codon is split by an intron, the amino acid symbol should be given below the portion of the codon containing two nucleotides.

21. The enumeration of amino acids shall start at the first amino acid of the sequence, with number 1. 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. It shall be marked under the sequence every five amino acids. 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 sequence may be made at the option of the applicant.

22. An amino acid sequence that is 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

23. 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 of data elements as defined in Appendix 1.

24. Only numeric identifiers of data elements as defined in Appendix 1 shall be used for the presentation of the items of information in the sequence listing. The corresponding numeric identifier descriptions 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. Two exceptions to this requirement are numeric identifiers <220> and <300>, which serve as headers for "Feature" and "Publication Information," respectively, and are associated with information in numeric identifiers <221> to <223> and <301> to <313>, respectively. When feature and publication information is provided in the sequence listing under those numeric identifiers, numeric identifiers <220> and <300>, respectively, should be included, but left blank. Generally, a blank line shall be inserted between numeric identifiers when the digit in the first or 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 shall precede any repeated numeric identifier.

Mandatory Data Elements

25. 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 Appendix 1 (mandatory data elements):

<110>	Applicant name
<120>	Title of invention
<160>	Number of SEQ ID NOs
<210>	SEQ ID NO: x
<211>	Length
<212>	Type
<213>	Organism
<400>	Sequence

Where the name of the applicant (numeric identifier <110>) is written in characters other than those of the Latin alphabet, it shall also be indicated in characters of the Latin alphabet either as a mere transliteration or through translation into English.

The data elements, except those under numeric identifiers <110>, <120> and <160>, shall be repeated for each sequence included in the sequence listing. Only the data elements under numeric identifiers <210> and <400> are mandatory if no sequence is present for a sequence identifier (see paragraph 5, above, and SEQ ID NO: 4 in the example depicted in Appendix 3 of this Standard).

26. In addition to the data elements identified in paragraph 25, above, when a sequence listing is furnished at any time prior to the assignment of an application number, the following data element shall be included in the sequence listing:

<130>	File reference

27. In addition to the data elements identified in paragraph 25, above, when a sequence listing is furnished at any time following the assignment of an application number, the following data elements shall be included in the sequence listing:

<140>	Current patent application
<141>	Current filing date

28. In addition to the data elements identified in paragraph 25, 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 included in the sequence listing:

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

29. If “n” or “Xaa” or a modified base or modified/unusual L-amino acid is used in the sequence, the following data elements are mandatory:

<220>	Feature
<221>	Name/key
<222>	Location
<223>	Other information

30. If the organism (numeric identifier <213>) is "Artificial Sequence" or "Unknown," the following data elements are mandatory:

<220>	Feature
<223>	Other information

Optional Data Elements

31. All data elements defined in Appendix 1, not mentioned in paragraphs 25 to 30, above, are optional (optional data elements).

Presentation of Features

32. When features of sequences are presented (that is, numeric identifier <220>), they shall be described by the "feature keys" set out in Appendix 2, Tables 5 and 6.2

Free Text

33. "Free text" is a wording describing characteristics of the sequence under numeric identifier <223> (Other information) which does not use language-neutral vocabulary as referred to in paragraph 2(vii).

34. The use of free text shall be limited to a few short terms indispensable for the understanding of the sequence. It shall not exceed four lines with a maximum of 65 characters per line for each given data element, when written in English. Any further information shall be included in the main part of the description in the language thereof.

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

REPETITION OF FREE TEXT IN MAIN PART OF DESCRIPTION

36. Where the sequence listing forming part of the international application contains free text, any such free text shall be repeated in the main part of the description in the language thereof. It is recommended that the free text in the language of the main part of the description be put in a specific section of the description called "Sequence Listing Free Text".

---

2 Editor's Note: These tables contain extracts from the DDBJ/EMBL/GenBank Feature Table (nucleotide sequences) and the SWISS-PROT Feature Table (amino acid sequences).

SEQUENCE LISTINGS IN ELECTRONIC FORM

9.~~37.~~ Any sequence listing referred to in paragraph 3 contained in an international application filed in electronic form shall be in an electronic document format and be filed by a means of transmittal that has been specified by the receiving Office for the purposes of filing of international applications in electronic form, provided that any such sequence listing shall preferably be in the electronic document format specified in paragraph 12(i) or, if so accepted by the competent authority, paragraph 12(ii) ~~paragraph 40~~ and be filed, if possible, by a means of transmittal which has been specified by both the receiving Office and the competent authority.[3], [4]

10.~~38.~~ Any sequence listing in electronic form referred to in paragraph 5 ~~paragraph 3ter~~ shall be in an electronic document format that has been specified by the receiving Office (in the case of a correction) or by the competent authority (in the case of a rectification or an amendment) for the purposes of filing of international applications in electronic form, provided that any such listing shall preferably be in the electronic document format specified in paragraph 12 ~~paragraph 40~~. Any such listing shall be filed by a means of transmittal which has been specified by the receiving Office or the competent authority, as applicable, for the purposes of this paragraph; if possible, it shall preferably be filed by a means of transmittal which has been specified by both the receiving Office and the competent authority.[5]

11.~~39.~~ Any sequence listing in electronic form referred to in paragraphs 6 and 7 ~~paragraphs 4 and 4bis~~ furnished for the purposes of the international search or international preliminary examination shall be in the electronic document format specified in paragraph 12 ~~paragraph 40~~ and be filed by a means of transmittal which has been specified by the competent authority for the purposes of this paragraph.

12.~~40.~~ For the purposes of the international search and international preliminary examination, any sequence listing in electronic form shall:

---

3 *Editor's Note:* Where a sequence listing in electronic form complying with this Standard is not available to the competent authority in a form and manner acceptable to it (that is, in particular, where it is not available to it in the electronic document format specified in paragraph 12(i) or (ii), as applicable ~~paragraph 40~~), the competent authority may invite the applicant to furnish to it such a sequence listing in electronic form (see Rule 13*ter*).

4 *Editor's Note:* Irrespective of the electronic document format of the sequence listing, the spatial relationship (e.g., columns and rows) of the data elements included in the sequence listing and the format of the actual nucleotide and/or amino acid sequences, as specified in WIPO Standard ST.XX ~~this Annex~~, shall be maintained.

5 *Editor's Note:* Where a replacement sequence listing in electronic form including any correction, rectification or amendment is not available to the competent authority in a form and manner acceptable to it (that is, in particular, where it is not available to it in the electronic document format specified in paragraph 12(i) or (ii), as applicable ~~paragraph 40~~), any such correction, rectification or amendment need only be taken into account by that authority for the purposes of the international search or preliminary examination to the extent that a meaningful search or preliminary examination can be carried out without the replacement sequence listing (see paragraph 7 ~~paragraph 4bis~~, above). See also Editor's Note 38, which equally applies to any replacement sequence listing in electronic form referred to in paragraph 5 ~~paragraph 3*ter*~~.

(i) be in the electronic document format specified in WIPO Standard ST.XX, paragraph XX, and preferably be created by dedicated software such as PatentIn ~~contained within one electronic file encoded using IBM6 Code Page 437, IBM Code Page 9327 or a compatible code page to represent the sequence listing as set out in paragraph 8 paragraphs 5 to 36 with no other codes included; a A compatible code page, as would be required for, for example, Japanese, Chinese, Cyrillic, Arabic, Greek or Hebrew characters, is one that assigns the Roman alphabet and numerals to the same hexadecimal positions as do the specified code pages;~~ or

(ii) if so accepted by the competent authority, preferably be the electronic document format specified in WIPO Standard ST.XX, paragraph YY, and preferably be created by dedicated software such as BISSAP.

~~41. Any sequence listing in the electronic document format specified in paragraph 40 shall preferably be created by dedicated software such as PatentIn.~~

PROCEDURE BEFORE DESIGNATED AND ELECTED OFFICES

13.~~42.~~ For the purposes of the procedure before a designated or elected Office before which the processing of an international application which contains the disclosure of one or more nucleotide and/or amino acid sequences has started (see Rule 13*ter*.3):

(i) any reference to the receiving Office or the competent authority shall be construed as a reference to the designated or elected Office concerned;

(ii) any reference to a sequence listing which is included in the international application by way of a rectification under Rule 91 or an amendment under Article 34(2)(b) of the description in relation to sequences contained in the application as filed shall be construed to also include any sequence listing included in the application, under the national law applied by the designated or elected Office concerned, by way of a rectification (of an obvious mistake) or amendment of the description in relation to sequences contained in the application as filed;

(iii) any reference to a sequence listing furnished for the purposes of international search or international preliminary examination shall be construed to also include any such listing furnished to the designated or elected Office concerned for the purposes of national search or examination by that Office;

(iv) the designated or elected Office concerned may invite the applicant to furnish to it, within a time limit which shall be reasonable under the circumstances, for the purposes of national search and/or examination, a sequence listing in electronic form complying with this Standard, unless such listing in electronic form is already available to that Office in a form and manner acceptable to it.

~~6 Editor's Note: IBM is a registered trademark of International Business Machine Corporation, United States of America.~~

~~7 Editor's Note: The specified code pages are de facto standards for personal computers.~~

~~**Appendices**~~

~~*Appendix 1: Numeric Identifiers*~~

~~*Appendix 2: Nucleotide and Amino Acid Symbols and Feature Table*~~

~~Table 1: List of Nucleotides~~

~~Table 2: List of Modified Nucleotides~~

~~Table 3: List of Amino Acids~~

~~Table 4: List of Modified and Unusual Amino Acids~~

~~Table 5: List of Feature Keys Related to Nucleotide Sequences~~

~~Table 6: List of Feature Keys Related to Protein Sequences~~

~~*Appendix 3: Specimen Sequence Listing*~~

[End of Annex and of document]