



Version 1.5

Standards **ST.26**

| 2021년 11월 5일 개정 |

2022. 5.



UNOFFICIAL TRANSLATION

This document is an unofficial translation of WIPO Standard ST.26 in Korean language.
The International Bureau of WIPO is not responsible for this document and has not verified its accuracy. Refer to the official versions of WIPO Standard ST.26 in English, Spanish, or French for verified contents.
Korean Translation provided by the Korean Intellectual Property Office (KIPO).

비공식 번역

해당 문서는 WIPO 표준 ST.26에 대한 비공식적 한국어 번역문입니다.
WIPO 국제사무국은 해당 문서에 대해 책임이 없으며 해당 문서의 정확성에 대해서도 검증하지 않았습니다. 검증된 내용을 확인하고자 하는 경우 WIPO 표준 ST.26의 공식 버전들(영어, 서반아어, 또는 불어)을 참조해주시기 바랍니다.

한국어 번역은 한국특허청이 제공하였습니다.

STANDARD ST.26

RECOMMENDED STANDARD FOR THE PRESENTATION OF NUCLEOTIDE AND AMINO ACID SEQUENCE LISTINGS
USING XML (EXTENSIBLE MARKUP LANGUAGE)

Version 1.5

*Revision approved by the Committee on WIPO Standards (CWS)
at its ninth session on November 5, 2021*

Editorial Note prepared by the International Bureau

The Fifty-Fourth (25th Ordinary) Session of WIPO General Assembly approved the new big-bang implementation date of WIPO Standard ST.26, July 1, 2022, at national, regional and international levels. Meanwhile, Standard ST.25 should continue to be used.

The Standard is published for information purposes of industrial property offices and other interested parties.

표준 ST.26

XML(EXTENSIBLE MARKUP LANGUAGE)을 사용한 핵산염기 및 아미노산 서열목록 제시를 위한 권장 표준

버전 1.5

2021년 11월 5일 WIPO 표준위원회(CWS) 제9차 회의에서 승인된 개정판

국제사무국이 작성한 편집 노트

WIPO 총회의 제54회(제25회 일반) 세션에서는 WIPO 표준 ST.26의 새로운 시행일로 2022년 7월 1일을 국가, 지역 및 국제 차원에서 승인하였다. 그 동안에, 표준 ST.25는 계속 사용되어야 한다.

본 표준은 특허청 및 기타 이해당사자에게 정보를 제공할 목적으로 간행되었다.

TABLE OF CONTENTS

INTRODUCTION	3.26.3
DEFINITIONS	3.26.3
SCOPE	3.26.4
REFERENCES	3.26.5
REPRESENTATION OF SEQUENCES	3.26.5
Nucleotide sequences	3.26.5
Amino acid sequences	3.26.8
Presentation of special situations	3.26.9
STRUCTURE OF THESEQUENCELISTING INXML	3.26.9
Root element	3.26.10
General information part	3.26.11
Sequence data part	3.26.14
Feature table	3.26.16
Feature keys	3.26.17
Mandatory feature keys	3.26.17
Feature location	3.26.17
Feature qualifiers	3.26.19
Mandatory feature qualifiers	3.26.19
Qualifier elements	3.26.19
Free text	3.26.21
Coding sequences	3.26.23
Variants	3.26.23

ANNEXES

Annex I - Controlled vocabulary	3.26.i.1~57
Annex II - DocumentType Definition(DTD) forSequence Listing	3.26.ii.1~5
AnnexIII - Sequence Listing Specimen (XML file)	3.26.iii.1
Annex IV - Character Subset from the Unicode Basic Latin Code Tablefor Use in an XML Instance of a Sequence Listing ..	3.26.iv.1~2
Annex V -Additional data exchange requirements (forIPOs only)	3.26.v.1
Annex VI -Guidance document with illustrated examples	3.26.vi.1~76
Appendix -Guidance document sequences in XML	
Annex VII -Recommendation for the transformation of a sequence listing from ST.25 to ST.26:	
potential added or deleted subject matter	3.26.vii.1~13

목차

서론	3.26.3
정의	3.26.3
적용범위	3.26.4
참조	3.26.5
서열의 표현	3.26.5
<i>핵산염기 서열</i>	3.26.5
<i>아미노산 서열</i>	3.26.8
<i>특수 상황의 표현</i>	3.26.9
XML로 표시한 서열목록의 구조	3.26.9
<i>루트 요소</i>	3.26.10
<i>일반 정보 부분</i>	3.26.11
<i>서열 데이터 부분</i>	3.26.14
<i>특징 표</i>	3.26.16
<i>특징 기호</i>	3.26.17
<i>필수 특징 기호</i>	3.26.17
<i>특징 위치</i>	3.26.17
<i>특징 한정자</i>	3.26.19
<i>필수 특징 한정자</i>	3.26.19
<i>한정자 요소</i>	3.26.19
<i>자유텍스트</i>	3.26.21
<i>코딩 서열</i>	3.26.23
<i>변이</i>	3.26.23

별첨

별첨 I 통제 어휘	3.26.i.1~57
별첨 II 서열목록에 대한 문서 타입 정의(DTD)	3.26.ii.1~5
별첨 III 서열목록 견본(XML 파일)	3.26.iii.1
별첨 IV 서열목록의 XML 인스턴스 사용을 위해 유니코드 기본 라틴 코드 표 유래의 문자 서브세트	3.26.iv.1~2
별첨 V 추가 데이터 교환 요건(특허청 전용)	3.26.v.1
별첨 VI 실시예를 포함한 지침서	3.26.vi.1~76
(부록) XML로 작성된 지침 서열	
별첨 VII ST.25에서 ST.26으로 서열목록 변환을 위한 권장 사항: 잠정적인 추가 또는 삭제된 주제	3.26.vii.1~13

STANDARD ST.26RECOMMENDED STANDARD FOR THE PRESENTATION OF NUCLEOTIDE AND AMINO ACID SEQUENCE LISTINGS
USING XML (EXTENSIBLE MARKUP LANGUAGE)*Version 1.5**Revision approved by the Committee on WIPO Standards (CWS)
at its ninth session on November 5, 2021*

INTRODUCTION

1. This Standard defines the nucleotide and amino acid sequence disclosures in a patent application required to be included in a sequence listing, the manner in which those disclosures are to be represented, and the Document Type Definition (DTD) for a sequence listing in XML (eXtensible Markup Language). It is recommended that industrial property offices accept any sequence listing compliant with this Standard filed as part of a patent application or in relation to a patent application.
2. The purpose of this Standard is to:
 - (a) allow applicants to draw up a single sequence listing in a patent application acceptable for the purposes of both international and national or regional procedures;
 - (b) enhance the accuracy and quality of presentations of sequences for easier dissemination, benefiting applicants, the public and examiners;
 - (c) facilitate searching of the sequence data; and
 - (d) allow sequence data to be exchanged in electronic form and introduced into computerized databases.

DEFINITIONS

3. For the purpose of this Standard, the expression:
 - (a) "amino acid" means any amino acid that can be represented using any of the symbols set forth in Annex I (see Section 3, Table 3). Such amino acids include, inter alia, D-amino acids and amino acids containing modified or synthetic side chains. Amino acids will be construed as unmodified L-amino acids unless further described in the feature table as modified according to paragraph 30. For the purpose of this standard, a peptide nucleic acid (PNA) residue is not considered an amino acid, but is considered a nucleotide as set forth in paragraph 3(g)(i)(2).
 - (b) "controlled vocabulary" is the terminology contained in this Standard that must be used when describing the features of a sequence, i.e., annotations of regions or sites of interest as set forth in Annex I.
 - (c) "enumeration of its residues" means disclosure of a sequence in a patent application by listing, in order, each residue of the sequence, wherein:
 - (i) the residue is represented by a name, abbreviation, symbol, or structure (e.g., HHHHHHQ or HisHisHisHisHisHisGln); or
 - (ii) multiple residues are represented by a shorthand formula (e.g., His₆Gln).
 - (d) "intentionally skipped sequence", also known as an empty sequence, refers to a placeholder to preserve the numbering of sequences in the sequence listing for consistency with the application disclosure, for example, where a sequence is deleted from the disclosure to avoid renumbering of the sequences in both the disclosure and the sequence listing.
 - (e) "modified amino acid" means any amino acid as described in paragraph 3(a) other than L-alanine, L-arginine, L-asparagine, L-aspartic acid, L-cysteine, L-glutamine, L-glutamic acid, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-pyrrolysine, L-serine, L-selenocysteine, L-threonine, L-tryptophan, L-tyrosine, or L-valine.
 - (f) "modified nucleotide" means any nucleotide as described in paragraph 3(g) other than deoxyadenosine 3'-monophosphate, deoxyguanosine 3'-monophosphate, deoxycytidine 3'-monophosphate, deoxythymidine 3'-monophosphate, adenosine 3'-monophosphate, guanosine 3'-monophosphate, cytidine 3'-monophosphate, or uridine 3'-monophosphate.

표준 ST.26

XML(EXTENSIBLE MARKUP LANGUAGE)을 사용한 핵산염기 및 아미노산 서열목록 제시를 위한 권장 표준

버전 1.5

2021년 11월 5일 WIPO 표준위원회(CWS) 제9차 회의에서 승인된 개정판

서론

1. 본 표준은 서열목록으로서 특허출원에 포함되어야 하는 핵산염기 및 아미노산 서열의 내용, 이러한 내용이 표현되는 방식, 및 XML(eXtensible Markup Language)인 서열목록에 대한 '문서 타입 정의'(DTD, Document Type Definition)를 정의한다. 특허출원의 일부 또는 특허출원과 관련하여 제출된 모든 서열목록이 이 표준을 준수하는 경우, 특허청이 수용할 것을 권장한다.

2. 본 표준의 목적은 다음과 같다:

- (a) 출원인이 국제 및 국내 또는 지역 절차에 적합한 특허출원에서 단일 서열목록을 작성하는 것을 허용하고;
- (b) 서열 표현의 정확성 및 품질을 향상시켜, 보다 쉬운 보급과 출원인, 일반인 및 심사관에게 도움을 주고;
- (c) 서열 데이터의 검색을 용이하게 하고;
- (d) 서열 데이터가 전자 양식으로 교환되고 전산화된 데이터베이스에 도입될 수 있게 한다.

정의

3. 본 표준에서 사용되는 용어의 정의는 다음과 같다:

(a) "아미노산(amino acid)"이란 별첨 I에 명시된 기호를 사용하여 표현되는 아미노산을 의미한다(섹션 3, 표 3 참조). 이러한 아미노산은 D-아미노산 및 변형된 또는 합성된 측쇄(side chain)를 가진 아미노산을 포함한다. 아미노산은, 30절에 따라 특정 표에 추가로 변형 아미노산이라고 기술되지 않는 한, 미변형 L-아미노산으로 해석된다. 본 표준에서는, 펩티드핵산(PNA) 잔기는 아미노산이 아니라, 3(g)(i)(2)절에 기재된 핵산염기로 간주된다.

(b) "통제 어휘(controlled vocabulary)"란 본 표준에 포함된 용어로, 별첨 I에 기재된 관심 영역 또는 위치의 주석과 같은 서열의 특징을 기술할 경우 사용하여야 한다.

(c) "잔기(殘基, residues)의 나열"이란 서열의 각 잔기를 순서대로 나열함으로써 특허출원에서 서열을 공개하는 것을 의미하며, 여기에서:

- (i) 잔기는 명칭, 약어, 기호 또는 구조로 표현되고(예를 들면, HHHHHHQ 또는 HisHisHisHisHisGln); 또는
- (ii) 다수의 잔기는 축약식으로 표현된다(예: His₆Gln).

(d) "의도적으로 생략된 서열(intentionally skipped sequence)"이란 빈(empty) 서열을 의미하며, 출원 개시 내용과의 일관성을 위해 서열목록에서 서열의 번호를 보존하기 위한 자리 표시자(placeholder)를 의미한다. 예를 들면, 출원 개시 내용 및 서열목록 모두에서 서열이 삭제된 경우, 서열의 번호를 다시 부여하는 것을 방지한다.

(e) "변형 아미노산(modified amino acid)"이란 L-알라닌(L-alanine), L-아르기닌(L-arginine), L-아스파라긴(L-asparagine), L-아스파르트산(L-aspartic acid), L-시스테인(L-cysteine), L-글루탐인(L-glutamine), L-글루탐산(L-glutamic acid), L-글리신(L-glycine), L-히스티딘(L-histidine), L-이소류신(L-isoleucine), L-류신(L-leucine), L-리신(L-lysine), L-메티오닌(L-methionine), L-페닐알라닌(L-phenylalanine), L-프롤린(L-proline), L-피롤리신(L-pyrrolysine), L-세린(L-serine), L-셀레노시스테인(L-selenocysteine), L-트레오닌(L-threonine), L-트립토판(L-tryptophan), L-티로신(L-tyrosine) 또는 L-발린(L-valine) 이외의 3(a)절에서 설명된 임의의 아미노산을 의미한다.

(f) "변형 핵산염기(modified nucleotide)"란 디옥시아데노신 3'-모노포스페이트(deoxyadenosine 3'-monophosphate), 디옥시구아노신 3'-모노포스페이트(deoxyguanosine 3'-monophosphate), 디옥시시티딘 3'-모노포스페이트(deoxycytidine 3'-monophosphate), 디옥시티미딘 3'-모노포스페이트(deoxythymidine 3'-monophosphate), 아데노신 3'-모노포스페이트(adenosine 3'-monophosphate), 구아노신 3'-모노포스페이트(guanosine 3'-monophosphate), 시티딘 3'-모노포스페이트(cytidine 3'-monophosphate) 또는 우리딘 3'-모노포스페이트(uridine 3'-monophosphate) 이외의 3(g)절에 설명된 임의의 핵산염기를 의미한다.

(g) “nucleotide” means any nucleotide or nucleotide analogue that can be represented using any of the symbols set forth in Annex I (see Section 1, Table 1) wherein the nucleotide or nucleotide analogue contains:

(i) a backbone moiety selected from:

- (1) 2' deoxyribose 5' monophosphate (the backbone moiety of a deoxyribonucleotide) or ribose 5' monophosphate (the backbone moiety of a ribonucleotide); or
- (2) an analogue of a 2' deoxyribose 5' monophosphate or ribose 5' monophosphate, which when forming the backbone of a nucleic acid analogue, results in an arrangement of nucleobases that mimics the arrangement of nucleobases in nucleic acids containing a 2' deoxyribose 5' monophosphate or ribose 5' monophosphate backbone, wherein the nucleic acid analogue is capable of base pairing with a complementary nucleic acid; examples of backbone moieties include amino acids as in peptide nucleic acids, glycol molecules as in glycol nucleic acids, threofuranosyl sugar molecules as in threose nucleic acids, morpholine rings and phosphorodiamidate groups as in morpholinos, and cyclohexenyl molecules as in cyclohexenyl nucleic acids.

and

(ii) the backbone moiety is either:

- (1) joined to a nucleobase, including a modified or synthetic purine or pyrimidine nucleobase; or
- (2) lacking a purine or pyrimidine nucleobase when the nucleotide is part of a nucleotide sequence, referred to as an “AP site” or an “abasic site”.

(h) “residue” means any individual nucleotide or amino acid or their respective analogues in a sequence.

(i) “sequence identification number” means a unique number (integer) assigned to each sequence in the sequence listing.

(j) “sequence listing” means a part of the description of the patent application as filed or a document filed subsequently to the application, which includes the disclosed nucleotide and/or amino acid sequence(s), along with any further description, as prescribed by this Standard.

(k) “specifically defined” means any nucleotide other than those represented by the symbol “n” and any amino acid other than those represented by the symbol “X”, listed in Annex I (see Section 1, Table 1, and Section 3, Table 3, respectively).

(l) “unknown” nucleotide or amino acid means that a single nucleotide or amino acid is present but its identity is unknown or not disclosed.

(m) “variant sequence” means a nucleotide or amino acid sequence that contains one or more differences with respect to a primary sequence. These differences may include alternative residues (see paragraphs 15 and 27), modified residues (see paragraphs 3(g), 3(h), 16, and 29), deletions, insertions, and substitutions. See paragraphs 93 to 95.

(n) “free text” is a type of value format for certain qualifiers, presented in the form of a descriptive text phrase or other specified format (as indicated in Annex I). See paragraph 85.

(o) “language-dependent free text” means the free text value of certain qualifiers, which may require translation for national, regional or international procedures. See paragraph 87.

4. For the purpose of this Standard, the word(s):

- (a) “may” refers to an optional or permissible approach, but not a requirement.
- (b) “must” refers to a requirement of the Standard; disregard of the requirement will result in noncompliance.
- (c) “must not” refers to a prohibition of the Standard.
- (d) “should” refers to a strongly encouraged approach, but not a requirement.
- (e) “should not” refers to a strongly discouraged approach, but not a prohibition.

SCOPE

5. This Standard establishes the requirements for the presentation of nucleotide and amino acid sequence listings of sequences disclosed in patent applications.

(g) "핵산염기(nucleotide)"란 별첨 I(색선 1, 표 1 참조)에 제시된 기호 중 하나를 사용하여 표현될 수 있는 임의의 핵산염기 또는 핵산염기 유사체를 의미하며, 여기서 핵산염기 또는 핵산염기 유사체는 다음을 포함한다:

(i) 하기(下記)로부터 선택되는 골격 부분:

(1) 2' 디옥시리보스 5' 모노포스페이트(2' deoxyribose 5' monophosphate)(디옥시리보 핵산염기 (deoxyribonucleotide)의 골격 부분) 또는 리보스 5' 모노포스페이트(ribose 5' monophosphate)(리보 핵산염기 (ribonucleotide)의 골격부분); 또는

(2) 골격을 형성할 때 2' 디옥시리보스 5' 모노포스페이트(2' deoxyribose 5' monophosphate) 또는 리보스 5' 모노포스페이트(ribose 5' monophosphate) 골격을 함유하는 핵산에서의 핵염기 배열을 모방하는 핵염기 배열을 초래하고, 상보적 핵산과 염기쌍을 이룰 수 있는, 2' 디옥시리보스 5' 모노포스페이트(2' deoxyribose 5' monophosphate) 또는 리보스 5' 모노포스페이트(ribose 5' monophosphate)의 유사체; 핵산염기 골격 부분의 예는 펩타이드 핵산에서와 같은 아미노산, 글리콜(glycol) 핵산에서와 같은 글리콜(glycol) 분자, 트레오스(threose) 핵산에서와 같은 트레오후라노실(threofuranosyl) 당 분자, 모르폴리노에서와 같은 모르폴린(morpholine) 고리(rings) 및 포스포로디아미데이트(phosphorodiamidate) 기(基) 및 시클로헥세닐(cyclohexenyl) 핵산에서와 같은 시클로헥세닐(cyclohexenyl) 분자를 포함한다.

(ii) 골격 부분(moiety)은:

(1) 변형 또는 합성 퓨린(purine) 또는 피리미딘(pyrimidine) 핵염기를 포함하는 핵염기에 결합되어 있거나; 또는

(2) 핵산염기가 핵산염기 서열의 일부면서 퓨린(purine) 또는 피리미딘(pyrimidine) 핵염기가 결합되는 경우, 즉 "AP 부위" 또는 "탈염기 부위(abasic site)"인 경우

(h) "잔기(residue)"란 서열에서 임의의 개별 핵산염기 또는 아미노산 또는 이들 각각의 유사체를 의미한다.

(i) "서열식별번호(sequence identification number)"란 서열목록의 각 서열에 부여된 특정 번호(정수)를 의미한다.

(j) "서열목록(sequence listing)"이란 출원된 특허출원의 발명의 설명의 일부든지 출원 후에 제출되는 문서의 일부를 의미하는데, 여기에는 본 표준에 의해 규정된 대로, 추가적인 설명(description)과 함께 공개된 핵산염기 및/또는 아미노산 서열을 포함한다.

(k) "구체적으로 정의된(specifically defined)"이란 별첨 I에 나열된, 기호 "n"으로 표현되는 것을 제외한 임의의 핵산염기 및 기호 "X"로 표현되는 것을 제외한 임의의 아미노산을 의미한다(각각 색선1, 표 1 및 색선 3, 표 3을 참조).

(l) "불명(unknown)" 핵산염기 또는 아미노산이란 단일 핵산염기 또는 아미노산이 존재하지만 이에 대한 정체성(identity)이 알려지지 않았거나 공개되지 않았음을 의미한다.

(m) "변이 서열(variant sequence)"은 1차 서열과 관련하여 하나 이상의 차이를 포함하는 핵산염기 또는 아미노산 서열을 의미한다. 이러한 차이점에는 대체(alternative) 잔기(15 및 27절 참조), 변형 잔기(3(g), 3(h), 16 및 29절 참조), 삭제(deletion), 삽입(insertion) 및 치환(substitution)이 포함될 수 있다. (93-95절 참조)

(n) "자유텍스트(free text)"란 설명 텍스트 구문 또는 기타 지정된 형식(별첨 I에 표시됨)의 양식으로 제공되는 특정 한정자에 대한 값 형식의 한 유형이다. (85절 참조)

(o) "언어의존적 자유텍스트(language-dependent free text)"란 어떤 한정자의 자유텍스트 값을 의미하며, 국가, 지역 또는 국제 절차에 대한 번역이 필요할 수 있다. (87절 참조)

4. 이 표준에서는, 하기의 어휘는:

(a) "~일 수 있다(may)"는 선택적 또는 허가된 접근법을 말하지만, 필수 요건은 아니다.

(b) "(반드시) ~하여야 한다(must)"는 이 표준의 필수 요건을 말한다; 요건을 무시하면 '미준수'의 결과가 생긴다.

(c) "(반드시) ~해서는 아니된다(must not)"는 본 표준의 금지 규정을 말한다.

(d) "~하여야 한다(should)"는 강력히 권장되지만, 필수 요건은 아니다.

(e) "~ 하여서는 아니된다(should not)"는 강력하게 비추천하지만, 금지하는 것은 아니다.

적용범위

5. 본 표준은 특허출원에 개시된 서열의 핵산염기 및 아미노산 서열목록을 제시하기 위한 요건을 확립한다.

6. A sequence listing complying with this Standard (hereinafter sequence listing) contains a general information part and a sequence data part. The sequence listing must be presented as a single file in XML using the Document Type Definition (DTD) presented in Annex II. The purpose of the bibliographic information contained in the general information part is solely for association of the sequence listing to the patent application for which the sequence listing is submitted. The sequence data part is composed of one or more sequence data elements each of which contain information about one sequence. The sequence data elements include various feature keys and subsequent qualifiers based on the International Nucleotide Sequence Database Collaboration (INSDC) and UniProt specifications.

7. For the purpose of this Standard, a sequence for which inclusion in a sequence listing is required is one that is disclosed anywhere in an application by enumeration of its residues and can be represented as:

(a) an unbranched sequence or a linear region of a branched sequence containing ten or more specifically defined nucleotides, wherein adjacent nucleotides are joined by:

(i) a 3' to 5' (or 5' to 3') phosphodiester linkage; or

(ii) any chemical bond that results in an arrangement of adjacent nucleobases that mimics the arrangement of nucleobases in naturally occurring nucleic acids; or

(b) an unbranched sequence or a linear region of a branched sequence containing four or more specifically defined amino acids, wherein the amino acids form a single peptide backbone, i.e. adjacent amino acids are joined by peptide bonds.

8. A sequence listing must not include, as a sequence assigned its own sequence identification number, any sequences having fewer than ten specifically defined nucleotides, or fewer than four specifically defined amino acids.

REFERENCES

9. References to the following Standards and resources are of relevance to this Standard:

International Nucleotide Sequence Database Collaboration (INSDC)	http://www.insdc.org/ ;
International Standard ISO 639-1:2002	Codes for the representation of names of languages - Part 1: Alpha-2 code;
UniProt Consortium	http://www.uniprot.org/ ;
W3C XML 1.0	http://www.w3.org/ ;
WIPO Standard <u>ST.2</u>	Standard manner for designating calendar dates by using Gregorian calendar;
WIPO Standard <u>ST.3</u>	Recommended standard on two-letter codes for the representation of states, other entities and intergovernmental organizations;
WIPO Standard <u>ST.16</u>	Recommended standard code for the identification of different kinds of patent documents;
WIPO Standard <u>ST.25</u>	Standard for the presentation of nucleotide and amino acid sequence listings in patent applications.

REPRESENTATION OF SEQUENCES

10. Each sequence encompassed by paragraph 7 must be assigned a separate sequence identification number, including a sequence which is identical to a region of a longer sequence. The sequence identification numbers must begin with number 1, and increase consecutively by integers. Where no sequence is present for a sequence identification number, i.e. an intentionally skipped sequence, "000" must be used in place of a sequence (see paragraph 58). The total number of sequences must be indicated in the sequence listing and must equal the total number of sequence identification numbers, whether followed by a sequence or by "000."

Nucleotide sequences

11. A nucleotide sequence must be represented only by a single strand, in the 5' to 3' direction from left to right, or in the direction from left to right that mimics the 5' to 3' direction. The designations 5' and 3' or any other similar designations must not be included in the sequence. A double-stranded nucleotide sequence disclosed by enumeration of the residues of both strands must be represented as:

(a) a single sequence or as two separate sequences, each assigned its own sequence identification number, where the two separate strands are fully complementary to each other, or

6. 본 표준을 따르는 서열목록(이하, 서열목록)은 '일반 정보' 부분과 '서열 데이터' 부분을 포함한다. 서열목록은 별첨 II에 설명된 DTD를 사용하여 XML 형식의 단일 파일로 제공되어야 한다. '일반 정보' 부분에 포함된 서지 정보의 목적은 서열목록이 제출된 특허출원에 서열목록을 연관시키기 위한 것이다. '서열 데이터' 부분은 하나 이상의 서열 데이터 요소로 이루어져야 하며, 각각의 서열 데이터 요소는 하나의 서열에 대한 정보를 포함하고 있어야 한다. 서열 데이터 요소는 '국제 핵산염기 서열 데이터베이스 연합'(INSDC, International Nucleotide Sequence Database Collaboration) 및 UniProt 규정에 기초한 다양한 특징 기호 및 후속 한정자를 포함한다.

7. 본 표준에서는, 서열목록에 포함되어야 하는 서열은 그 잔기를 나열함으로써 출원서의 어느 곳에서나 개시되고 다음과 같이 표현될 수 있다:

(a) 10개 이상의 구체적으로 정의된 핵산염기를 함유하는 분지형(branched) 서열의 선형(linear) 부위 또는 비분지형(unbranched) 서열로서, 여기에서 인접된 핵산염기는:

(i) 3'-말단 → 5'-말단(또는 5'-말단 → 3'-말단) 포스포디에스테르(phosphodiester) 연결되거나; 또는

(ii) 자연적으로 발생하는 핵산들의 핵염기(nucleobases) 배열과 유사한 인접한 핵염기들의 배열을 초래하는 화학 결합을 통해 연결되거나; 또는

(b) 4개 이상의 구체적으로 정의된 아미노산을 함유하는 분지형 서열의 선형 부위 또는 비분지형 서열로서, 여기에서 아미노산들은 단일 펩티드 골격을 형성하며, 인접한 아미노산은 펩티드 결합에 의해 연결된다.

8. 서열목록은 자신의 서열식별번호가 할당된 서열로서 10개 미만의 구체적으로 정의된 핵산염기, 또는 4개 미만의 구체적으로 정의된 아미노산을 갖는 서열을 포함해서는 아니된다.

참조

9. 다음 표준 및 자원에 대한 참조 목록은 본 표준과 관련이 있다:

국제 핵산염기 서열 데이터베이스 연합(INSDC)	http://www.insdc.org/ ;
국제 표준 ISO 639-1: 2002	언어 명칭 표현을 위한 코드-1부: 알파-2 코드;
UniProt 컨소시엄	http://www.uniprot.org/ ;
W3C XML 1.0	http://www.w3.org/ ;
WIPO 표준 ST.2	그레고리력을 사용하여 달력 날짜를 지정하기 위한 표준 방법;
WIPO 표준 ST.3	국가, 기타 단체 및 정부간 기구를 표현하는 2 문자 코드에 대한 권고 표준;
WIPO 표준 ST.16	다른 종류의 특허 문헌 식별에 대한 권고 표준;
WIPO 표준 ST.25	특허출원에 기재된 핵산염기 및 아미노산 서열목록의 제시를 위한 표준

서열의 표현

10. 더 긴 서열의 영역과 동일한 서열을 포함하는 7절에 포함된 각 서열에는 별도의 서열식별번호를 부여하여야 한다. 서열식별번호는 숫자 1로 시작하고 정수로 연속해서 증가하여야 한다. 서열이 존재하지 않는 서열식별번호의 경우, 즉 의도적으로 생략된 서열은 서열 대신 "000"을 사용하여야 한다(58절 참조). 총 서열 수는 서열목록에 표시되어야 하며, 서열 또는 "000"을 포함하는 서열식별번호의 총 수와 같아야 한다.

핵산염기 서열

11. 핵산염기 서열은 단일 가닥으로, 왼쪽에서 오른쪽으로 5'-말단에서 3'-말단 방향 또는 5'-말단에서 3'-말단 방향을 모방하는 방향으로 왼쪽에서 오른쪽으로 표현되어야 한다. 서열에 5'-말단 및 3'-말단을 지정하거나 다른 유사한 지정 표시가 포함되면 아니된다.

이중가닥(double-stranded)의 핵산염기 서열은 양쪽가닥의 잔기를 아래와 같이 나열하여 표현하여야 한다:

(a) 단일 서열 또는 2개의 개별 서열로 각각 고유 서열식별번호가 부여되며, 2개의 개별 가닥은 서로 완전히 상보적인 것, 또는

(b) two separate sequences, each assigned its own sequence identification number, where the two strands are not fully complementary to each other.

12. For the purpose of this Standard, the first nucleotide presented in the sequence is residue position number 1. When nucleotide sequences are circular in configuration, applicant must choose the nucleotide in residue position number 1. Numbering is continuous throughout the entire sequence in the 5' to 3' direction, or in the direction that mimics the 5' to 3' direction. The last residue position number must equal the number of nucleotides in the sequence.

13. All nucleotides in a sequence must be represented using the symbols set forth in Annex I (see Section 1, Table 1). Only lower case letters must be used. Any symbol used to represent a nucleotide is the equivalent of only one residue.

14. The symbol "t" will be construed as thymine in DNA and uracil in RNA. Uracil in DNA or thymine in RNA is considered a modified nucleotide and must be further described in the feature table as provided by paragraph 19.

15. Where an ambiguity symbol (representing two or more alternative nucleotides) is appropriate, the most restrictive symbol should be used, as listed in Annex I (section 1, Table 1). For example, if a nucleotide in a given position could be "a" or "g", then "r" should be used, rather than "n". The symbol "n" will be construed as any one of "a", "c", "g", or "t/u" except where it is used with a further description in the feature table. The symbol "n" must not be used to represent anything other than a nucleotide. A single modified or "unknown" nucleotide may be represented by the symbol "n", together with a further description in the feature table, as provided in paragraphs 16, 17, 21, or 93-96. For representation of sequence variants, i.e., alternatives, deletions, insertions, or substitutions, see paragraphs 93 to 100.

16. Modified nucleotides should be represented in the sequence as the corresponding unmodified nucleotides, i.e., "a", "c", "g" or "t" whenever possible. Any modified nucleotide in a sequence that cannot otherwise be represented by any other symbol in Annex I (see Section 1, Table 1), i.e., an "other" nucleotide, such as a non-naturally occurring nucleotide, must be represented by the symbol "n". The symbol "n" is the equivalent of only one residue.

17. A modified nucleotide must be further described in the feature table (see paragraph 60 *et seq.*) using the feature key "modified_base" and the mandatory qualifier "mod_base" in conjunction with a single abbreviation from Annex I (see Section 2, Table 2) as the qualifier value; if the abbreviation is "OTHER", the complete unabbreviated name of the modified nucleotide must be provided as the value in a "note" qualifier. For a listing of alternative modified nucleotides, the qualifier value "OTHER" may be used in conjunction with a further "note" qualifier (see paragraphs 97 and 98). The abbreviations (or full names) provided in Annex I (see Section 2, Table 2) referred to above must not be used in the sequence itself.

18. A nucleotide sequence including one or more regions of consecutive modified nucleotides that share the same backbone moiety (see paragraph 3(g)(i)(2)), must be further described in the feature table as provided by paragraph 17. The modified nucleotides of each such region may be jointly described in a single INSDFeature element as provided by paragraph 22. The most restrictive unabbreviated chemical name that encompasses all of the modified nucleotides in the range or a list of the chemical names of all the nucleotides in the range must be provided as the value in the "note" qualifier. For example, a glycol nucleic acid sequence containing "a", "c", "g", or "t" nucleobases may be described in the "note" qualifier as "2,3-dihydroxypropyl nucleosides." Alternatively, the same sequence may be described in the "note" qualifier as "2,3-dihydroxypropyladenine, 2,3-dihydroxypropylthymine, 2,3-dihydroxypropylguanine, or 2,3-dihydroxypropylcytosine." Where an individual modified nucleotide in the region includes an additional modification, then the modified nucleotide must also be further described in the feature table as provided in paragraph 17.

19. Uracil in DNA or thymine in RNA are considered modified nucleotides and must be represented in the sequence as "t" and be further described in the feature table using the feature key "modified_base", the qualifier "mod_base" with "OTHER" as the qualifier value and the qualifier "note" with "uracil" or "thymine", respectively, as the qualifier value.

20. The following examples illustrate the representation of modified nucleotides according to paragraphs 16 to 18 above:

Example 1: Modified nucleotide using an abbreviation from Annex I (see Section 2, Table 2)

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>15</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Example 2: Modified nucleotide using "OTHER" from Annex I (see Section 2, Table 2)

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>4</INSDFeature_location>
  <INSDFeature_qual>
```

(b) 서로 완전히 상보적이지 않는, 각각의 서열식별번호가 부여된 2개의 개별 서열

12. 본 표준에서는, 서열에 설명된 첫 번째 핵산염기는 잔기 위치 번호 1이다. 핵산염기 서열이 원형인 경우, 출처원은 잔기 위치 번호 1의 핵산염기를 선택하여야 한다. 5'-말단 → 3'-말단 방향으로 또는 5'-말단 → 3'-말단 방향을 모방하는 방향으로 전체 서열의 위치 번호가 연속된다. 마지막 잔기 위치 번호는 서열에서 핵산염기의 개수와 같아야 한다.

13. 서열의 모든 핵산염기는 별첨 I에 설명된 기호를 사용하여 표현하여야 한다(섹션 1, 표 1 참조). 오직 소문자만 사용하여야 한다. 핵산염기를 표현에 사용된 임의의 기호는 오직 하나의 잔기와 같아야 한다.

14. 기호 "t"는 DNA에서 티민(thymine) 및 RNA에서 우라실(uracil)로 해석될 것이다. DNA의 우라실(uracil) 또는 RNA의 티민(thymine)은 변형 핵산염기로 간주되며 19절에서 제공된 특징 표에 추가로 설명되어 있어야 한다.

15. 다중핵산염기(ambiguity) 기호(2개 이상의 대체 핵산염기 표현)가 적절한 경우, 별첨 I(섹션 1, 표 1 참조)에 나열된 대로 가장 제한적인 기호를 사용하여야 한다. 예를 들면, 주어진 위치의 핵산염기가 "a" 또는 "g"일 수 있다면, "n" 대신 "r"이 사용되어야 한다. 기호 "n"은 특징 표에서 추가 설명과 함께 사용되는 경우를 제외하고, "a", "c", "g" 또는 "t/u" 중의 하나로 해석된다. "n" 기호는 핵산염기 이외의 것을 표현하는 곳에 사용해서는 아니된다. 16절, 17절, 21절 또는 93절~96절에 제공된 바와 같이, 단일 변형 또는 "불명" 핵산염기는 특징 표의 추가 설명과 함께 "n" 기호로 표시될 수 있다. 서열 변이 즉, 대체, 삭제, 삽입 또는 치환의 표현에 대해서는 94절~100절을 참조하면 된다.

16. 변형 핵산염기는 이에 대응하는 일반 핵산염기, 즉 "a", "c", "g" 또는 "t" 로 서열에 표현되어야 한다. 별첨 I(섹션 1, 표 1 참조)의 다른 기호로 표현할 수 없는 서열의 변형 핵산염기, 즉 자연적으로 발생하지 않는 핵산염기와 같은 "other" 핵산염기는 기호 "n"으로 표현하여야 한다. 기호 "n"이 변형 핵산염기를 나타내는데 사용되는 경우 이는 하나의 잔기와 동등하다.

17. 변형 핵산염기는 특징 기호 "modified_base"와 필수 한정자 "mod_base"를 별첨 I의 단일 약어와 함께 사용하여 특징 표(60절 이하 참조)에 한정자의 값으로서 추가로 제시하여야 한다(섹션 2, 표 2 참조); 약어가 "OTHER"인 경우, 변형 핵산염기의 단축되지 않고 완전한 이름을 "note" 한정자의 값으로 제공하여야 한다. 대체 변형 핵산염기를 목록에 기재하기 위해서는 한정자 값 "OTHER"를 추가 "note" 한정자와 함께 사용할 수 있다(97절 및 98절 참조). 위에서 언급한 별첨 I(2절, 표 2 참조)에 제공된 약어(또는 전체 이름)는 서열 자체에 사용해서는 아니된다.

18. 동일한 골격 부분을 공유하는 연속 변형 핵산염기의 하나 이상의 영역을 포함하는 핵산염기 서열(3(g)(i)(2)절 참조)은 17절에 의해 제공된 특징 표에 추가로 기술되어야 한다. 각각의 이러한 영역의 변형 핵산염기는 22절에 의해 제공된 바와 같이, 단일 INSDFeature 요소에 공동으로 기술될 수 있다. 범위 내의 모든 변형 핵산염기를 포함하는 가장 제한적인 생략되지 않은 화학명 또는 해당 범위에 있는 모든 핵산염기의 화학명 목록은 "note" 한정자의 값으로 제공되어야 한다. 예를 들면, "a", "c", "g" 또는 "t" 핵염기를 함유하는 글리콜(glycol) 핵산염기 서열은 "note" 한정자에서 "2,3-dihydroxypropyl nucleosides"로 기술될 수 있다. 또는 "note" 한정자에서 동일한 서열은 "2,3-dihydroxypropyladenine, 2,3-dihydroxypropylthymine, 2,3-dihydroxypropylguanine 또는 2,3-dihydroxypropylcytosine"으로 설명될 수 있다. 해당 범위에 있는 개별 변형 핵산염기가 추가적인 변형을 포함하는 경우 변형 핵산염기는 17절에 제공된 것처럼 특징 표에 추가로 기술되어야 한다.

19. DNA의 우라실(uracil) 또는 RNA의 티민(thymine)은 변형 핵산염기로 간주되며, 서열에서는 "t"로 표시되어야 하며, 이에 대해 특징 기호 "modified_base"와 한정자 값으로 "OTHER"를 가진 한정자 "mod_base", 및 한정자 값으로 각각 "uracil" 또는 "thymine"을 가진 한정자 "note"를 사용하여 특징 표에서 추가로 표현되어야 한다.

20. 아래의 실시예는 상기 16절~18절에 따른 변형 핵산염기의 표현을 나타낸다:

실시예 1: 별첨 I의 약어를 사용한 변형 핵산염기(섹션 2, 표 2 참조)

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>15</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

실시예 2: 별첨 I의 "OTHER"를 사용한 변형 핵산염기(섹션 2, 표 2 참조)

```
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>4</INSDFeature_location>
  <INSDFeature_qual>
```

```

<INSDQualifier>
  <INSDQualifier_name>mod_base</INSDQualifier_name>
  <INSDQualifier_value>OTHER</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier>
  <INSDQualifier_name>note</INSDQualifier_name>
  <INSDQualifier_value>xanthine</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>

```

Example 3: A nucleotide sequence composed of modified nucleotides encompassed by paragraph 3(g)(i)(2) with two individual nucleotides that include a further modification

```

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>1..954</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>2,3-dihydroxypropyl nucleosides</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>439</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>684</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>xanthine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

21. Any “unknown” nucleotide must be represented by the symbol “n” in the sequence. An “unknown” nucleotide should be further described in the feature table (see paragraph 60 *et seq.*) using the feature key “unsure”. The symbol “n” is the equivalent of only one residue.

22. A region containing a known number of contiguous “a”, “c”, “g”, “t”, or “n” residues for which the same description applies may be jointly described using a single INSDFeature element with the syntax “x..y” as the location descriptor in the element INSDFeature_location (see paragraphs 64 to 71). For representation of sequence variants, i.e., alternatives, deletions, insertions or substitutions, see paragraphs 93 to 100.

23. The following example illustrates the representation of a region of modified nucleotides for which the same description applies, according to paragraph 22 above:

```

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>358..485</INSDFeature_location>
  <INSDFeature_qual>

```

```

    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>xanthine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

실시예 3: 추가 변형을 포함하는 2개의 개별 핵산염기를 갖는 3(g)(i)(2)절에 포함된 변형 핵산염기로 이루어진 핵산염기 서열

```

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>1..954</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>2,3-dihydroxypropyl nucleosides</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>439</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>i</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>684</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>xanthine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

21. 임의의 "불명(unknown)" 핵산염기는 서열에서 기호 "n"으로 표현되어야 한다. "불명" 핵산염기는 특징 기호 "unsure"를 사용하여 특징 표(60절 이하 참조)에 추가로 기술하여야 한다. 기호 "n"은 하나의 잔기만을 나타낸다.

22. 알려진 수의 연속된 "a", "c", "g", "t" 또는 "n" 잔기를 포함하는 영역에 동일한 설명이 적용되는 경우, INSDFeature_location 요소와 구분 "x.y:"를 가진 INSDFeature 요소를 사용하여 공동으로 설명될 수 있다(64절~71절 참조). 서열 변이, 즉 대체, 삭제, 삽입 또는 치환의 표현에 대해서는 93절~100절을 참조하면 된다.

23. 다음 실시예는 상기 22절에 따라 동일한 설명이 적용되는 변형 핵산염기 영역의 표현을 예시한다:

```

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>358..485</INSDFeature_location>
  <INSDFeature_qual>

```



```

<INSDQualifier>
  <INSDQualifier_name>mod_base</INSDQualifier_name>
  <INSDQualifier_value>OTHER</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier>
  <INSDQualifier_name>note</INSDQualifier_name>
  <INSDQualifier_value>isoguanine</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>

```

Amino acid sequences

24. The amino acids in an amino acid sequence must be represented in the amino to carboxy direction from left to right. The amino and carboxy groups must not be represented in the sequence.

25. For the purpose of this Standard, the first amino acid in the sequence is residue position number 1, including amino acids preceding the mature protein, for example, pre-sequences, pro-sequences, pre-pro-sequences and signal sequences. When an amino acid sequence is circular in configuration and the ring consists solely of amino acid residues linked by peptide bonds, i.e., the sequence has no amino and carboxy termini, applicant must choose the amino acid in residue position number 1. Numbering is continuous through the entire sequence in the amino to carboxy direction.

26. All amino acids in a sequence must be represented using the symbols set forth in Annex I (see Section 3, Table 3). Only upper case letters must be used. Any symbol used to represent an amino acid is the equivalent of only one residue.

27. Where an ambiguity symbol (representing two or more amino acids in the alternative) is appropriate, the most restrictive symbol should be used, as listed in Annex I (Section 3, Table 3). For example, if an amino acid in a given position could be aspartic acid or asparagine, the symbol "B" should be used, rather than "X". The symbol "X" will be construed as any one of "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", or "V", except where it is used with a further description in the feature table. The symbol "X" must not be used to represent anything other than an amino acid. A single modified or "unknown" amino acid may be represented by the symbol "X", together with a further description in the feature table, e.g., as provided by paragraphs 29, 30, 32, or 93-98. For representation of sequence variants, i.e., alternatives, deletions, insertions, or substitutions, see paragraphs 93 to 100.

28. Disclosed amino acid sequences separated by internal terminator symbols, represented for example by "Ter" or asterisk "*" or period "." or a blank space, must be included as separate sequences for each amino acid sequence that contains at least four specifically defined amino acids and is encompassed by paragraph 7. Each such separate sequence must be assigned its own sequence identification number. Terminator symbols and spaces must not be included in sequences in a sequence listing (see paragraph 57).

29. Modified amino acids, including D-amino acids, should be represented in the sequence as the corresponding unmodified amino acids whenever possible. Any modified amino acid in a sequence that cannot otherwise be represented by any other symbol in Annex I (see Section 3, Table 3), i.e., an "other" amino acid, must be represented by "X". The symbol "X" is the equivalent of only one residue.

30. A modified amino acid must be further described in the feature table (see paragraph 60 *et seq.*). Where applicable, the feature keys "CARBOHYD" or "LIPID" should be used together with the qualifier "note". The feature key "MOD_RES" should be used for other post-translationally modified amino acids together with the qualifier "note"; otherwise the feature key "SITE" together with the qualifier "note" should be used. The value for the qualifier "note" must either be an abbreviation set forth in Annex I (see Section 4, Table 4), or the complete, unabbreviated name of the modified amino acid. The abbreviations set forth in Table 4 referred to above or the complete, unabbreviated names must not be used in the sequence itself.

31. The following examples illustrate the representation of modified amino acids according to paragraph 30 above:

Example 1: Post-translationally modified amino acid

```

<INSDFeature>
  <INSDFeature_key>MOD_RES</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>3Hyp</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

```

<INSDQualifier>
  <INSDQualifier_name>mod_base</INSDQualifier_name>
  <INSDQualifier_value>OTHER</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier>
  <INSDQualifier_name>note</INSDQualifier_name>
  <INSDQualifier_value>isoguanine</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>

```

아미노산 서열

24. 아미노산 서열의 아미노산은 왼쪽에서 오른쪽으로 아미노(amino)에서 카르복시(carboxy) 방향으로 표현되어야 한다. 아미노기 및 카르복시기는 서열에 표현되면 아니된다.
25. 본 표준에서는, 성숙(mature) 단백질에 선행하는 아미노산, 예를 들면 프리서열(pre-sequence), 프로서열(pro-sequence), 프리프로서열(pre-pro-sequence) 및 시그널서열(signal sequence)을 포함하여 서열의 첫 번째 아미노산은 잔기 위치 번호 1이다. 아미노산 서열이 원형 구조이고, 고리가 펩티드 결합에 의해 연결된 아미노산 잔기뿐만 아니라 이루어진 경우, 즉, 서열에 아미노 및 카르복시 말단이 없는 경우, 출원인은 잔기 위치 번호 1 아미노산을 선택하여야 한다. 위치 번호는 아미노에서 카르복시 방향으로 전체 서열을 통해 연속적으로 한다.
26. 서열의 모든 아미노산은 별첨 I에 설명된 기호를 사용하여 표현되어야 한다(섹션 3, 표 3 참조). 대문자만을 사용하여야 한다. 아미노산 표현을 위해 사용되는 모든 기호는 하나의 잔기와 동등하다.
27. 다중아미노산(ambiguity) 기호(선택적으로 둘 이상의 아미노산을 표현)로 표기해야 할 경우, 별첨 I(표 3, 섹션 3)에서 가장 제한적인 기호를 사용하여야 한다. 예를 들면, 주어진 위치의 아미노산이 아스파르트산(aspartic acid) 또는 아스파라긴(asparagine)일 경우 "X" 대신 "B" 기호를 사용하여야 한다. "X" 기호는 "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석되는데, 특징 표에서 추가 설명과 함께 사용되는 경우는 예외이다. "X" 기호는 아미노산 이외의 것을 나타내는 곳에 사용해서는 아니된다. 단일 변형 또는 "불명" 아미노산은 29절, 30절, 32절, 또는 93절~98절에 의해 제공되는 바와 같이 특징 표의 추가 설명과 함께 "X" 기호로 표현될 수 있다. 서열 변이, 즉 대체, 삭제, 삽입 또는 치환의 표현에 대해서는 93절~100절을 참조하면 된다.
28. "Ter" 또는 별표 "*" 또는 마침표 "." 또는 공백으로 표현되는 내부 종결 기호(internal terminator symbol)로 분리된 아미노산 서열은 적어도 4개 이상의 구체적으로 정의된 아미노산을 포함하고 7절에 규정되어 있는 각각의 아미노산 서열에 대한 개별 서열로 포함되어야 한다. 이러한 각각의 개별 서열에는 독립된 서열식별번호를 부여하여야 한다. 종결 기호 및 공백은 서열목록의 서열에 포함되면 안 된다(57절 참조).
29. D-아미노산을 포함한 변형 아미노산은 가능한 이에 해당하는 미변형 아미노산으로 서열에 표시되어야 한다. 별첨 I(섹션 3, 표 3 참조)에서 다른 기호로 표현될 수 없는 서열의 변형 아미노산, 즉 "기타" 아미노산은 "X"로 표현되어야 한다. "X" 기호는 하나의 잔기와 동일하다.
30. 변형 아미노산은 특징 표에 추가로 기술되어야 한다(60절 이하 참조). 해당되는 경우에 특징 기호 "CARBOHYD" 또는 "LIPID"를 한정자 "note"와 함께 사용하여야 한다. 특징 기호 "MOD_RES"는 한정자 "note"와 함께 번역 후(post-translationally) 변형된 아미노산을 위해 사용하여야 한다; 그렇지 않은 경우 한정자 "note"와 함께 특징 기호 "SITE"를 사용하여야 한다. 한정자 "note"의 값은 별첨 I(섹션 4, 표 4 참조)에 설명된 약어이거나 변형 아미노산의 단축되지 않은 완전한 명칭이어야 한다. 위에 언급된 표 4에 설명된 약어 또는 단축되지 않은 완전한 명칭은 서열 자체에 사용해서는 안 된다.
31. 하기 실시예는 상기 30절에 따른 변형 아미노산의 표현을 예시한다:

실시예 1: 번역 후 변형된 아미노산

```

<INSDFeature>
  <INSDFeature_key>MOD_RES</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>3Hyp</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

Example 2: Non post-translationally modified amino acid

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>Orn</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Example 3: D-amino acid

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>9</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>D-Arginine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

32. Any “unknown” amino acid must be represented by the symbol “X” in the sequence. An “unknown” amino acid designated as “X” must be further described in the feature table (see paragraph 60 *et seq.*) using the feature key “UNSURE” and optionally the qualifier “note.” The symbol “X” is the equivalent of only one residue.

33. The following example illustrates the representation of an “unknown” amino acid according to paragraph 32 above:

```
<INSDFeature>
  <INSDFeature_key>UNSURE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>A or V</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

34. A region containing a known number of contiguous “X” residues for which the same description applies may be jointly described using the syntax “x..y” as the location descriptor in the element `INSDFeature_location` (see paragraphs 64 to 70). For representation of sequence variants, i.e., alternatives, deletions, insertions, or substitutions, see paragraphs 93 to 100.

Presentation of special situations

35. A sequence disclosed by enumeration of its residues that is constructed as a single continuous sequence from one or more non-contiguous segments of a larger sequence or of segments from different sequences must be included in the sequence listing and assigned its own sequence identification number.

36. A sequence that contains regions of specifically defined residues separated by one or more regions of contiguous “n” or “X” residues (see paragraphs 15 and 27, respectively), wherein the exact number of “n” or “X” residues in each region is disclosed, must be included in the sequence listing as one sequence and assigned its own sequence identification number.

37. A sequence that contains regions of specifically defined residues separated by one or more gaps of an unknown or undisclosed number of residues must not be represented in the sequence listing as a single sequence. Each region of specifically defined residues that is encompassed by paragraph 7 must be included in the sequence listing as a separate sequence and assigned its own sequence identification number.

STRUCTURE OF THE SEQUENCE LISTING IN XML

38. In accordance with paragraph 6 above, an XML instance of a sequence listing file according to this Standard is composed of:

실시예 2: 번역 후 변형된 경우가 아닌 변형 아미노산

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_qualifiers>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>Orn</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qualifiers>
</INSDFeature>
```

실시예 3: D-아미노산

```
<INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>9</INSDFeature_location>
  <INSDFeature_qualifiers>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>D-Arginine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qualifiers>
</INSDFeature>
```

32. “불명(unknown)” 아미노산은 서열에서 기호 “X”로 표현되어야 한다. “X”로 지정된 “불명” 아미노산은 특정 기호 “UNSURE” 및 선택적으로 한정자 “note”를 사용하여 특징 표(60절 이하 참조)에 추가로 기술하여야 한다. “X” 기호는 단 하나의 잔기와 동일하다.

33. 다음 실시예는 상기 32절에 따른 “불명” 아미노산의 표현을 예시한다:

```
<INSDFeature>
  <INSDFeature_key>UNSURE</INSDFeature_key>
  <INSDFeature_location>3</INSDFeature_location>
  <INSDFeature_qualifiers>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>A or V</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qualifiers>
</INSDFeature>
```

34. 동일한 설명이 적용되는 연속적인 “X” 잔기를 포함하는 영역은 INSDFeature_location 요소(64절~70절 참조)에서 위치 설명자로 “x.y” 구문을 사용하여 함께 설명될 수 있다. 서열 변이, 즉 삭제, 대체, 삽입 또는 치환의 표현에 대해서는 93절~100절을 참조하면 된다.

특수 상황의 표현

35. 길이가 긴 서열에서 유래한 하나 이상의 불연속 단편(non-contiguous segment) 또는 서로 다른 서열의 단편으로부터 유래한 단일 연속 서열로 이루어진 잔기의 나열로 표시된 서열은 서열목록에 포함되고 독립된 서열식별번호가 부여되어야 한다.

36. 연속 “n” 또는 “X” 잔기(각각 15절 및 27절 참조)의 하나 이상의 영역에 의해 분리된 구체적으로 정의된 잔기의 영역을 포함하는 서열은(각 영역에서 정확한 수의 “n” 또는 “X” 잔기가 표시됨) 하나의 서열로 서열목록에 포함되어야 하고, 독립된 서열식별번호가 부여되어야 한다.

37. 알려지지 않거나 표시되지 않은 수의 잔기인 하나 이상의 갭(gap)에 의해 분리된 구체적으로 정의된 잔기의 영역들을 포함하는 서열은 단일 서열로서 서열목록에 표현되지 않아야 한다. 7절에 포함되는 구체적으로 정의된 잔기의 각 영역은 별도의 서열로 서열목록에 포함되어야 하고, 독립된 서열식별번호가 부여되어야 한다.

XML로 표시한 서열목록의 구조

38. 상기 6절에 따라, 본 표준에 따른 서열목록 파일의 XML 인스턴스는 다음과 같이 이루어진다:

(a) a general information part, w hich contains information concerning the patent application to w hich the sequence listing is directed; and

(b) a sequence data part, w hich contains one or more sequence data elements, each of w hich, in turn, contain information about one sequence.

An example of a sequence listing is provided in Annex III.

39. The sequence listing must be presented in XML 1.0 using the DTD presented in the Annex II “Document Type Definition (DTD) for Sequence Listing”.

(a) The first line of the XML instance must contain the XML declaration:

```
<?xml version="1.0" encoding="UTF-8"?>.
```

(b) The second line of the XML instance must contain a document type (DOCTYPE) declaration:

```
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">.
```

40. The entire electronic sequence listing must be contained w ithin one file. The file must be encoded using Unicode UTF-8, w ith the follow ing restrictions:

(a) the information contained in the elements `ApplicantName`, `InventorName` and `InventionTitle` of the general information part, and the `NonEnglishQualifier` value of the sequence data part, may be composed of any valid Unicode characters indicated in the XML 1.0 specification except the Unicode Control code points 0000-001F and 007F-009F. The reserved characters “, &, ‘, <, and > (Unicode code points 0022, 0026, 0027, 003C and 003E respectively), must be replaced as set forth in paragraph 41; and

(b) the information contained in all other elements and attributes of the general information part and in all other elements and attributes of the sequence data part must be composed of printable characters (including the space character) from the Unicode Basic Latin code table (i.e., limited to Unicode code points 0020 through 007E – see Annex IV). The reserved characters “, &, ‘, <, and > (Unicode code points 0022, 0026, 0027, 003C and 003E respectively), must be replaced as set forth in paragraph 41.

41. In an XML instance of a sequence listing, numeric character references¹ must not be used and the follow ing reserved characters must be replaced by the corresponding predefined entities when used in a value of an attribute or content of an element:

Reserved Character	Pre defined Entities
<	<
>	>
&	&
“	"
’	'

See paragraph 71 for an example. The only character entity references permitted are the predefined entities set forth in this paragraph.

42. All mandatory elements must be populated (except as provided for in paragraph 58 for an intentionally skipped sequence). Optional elements for w hich content is not available should not appear in the XML instance (except as provided for in paragraph 97 for representation of a deletion in a sequence in the value for the qualifier “replace”).

Root element

43. The root element of an XML instance according to this Standard is the element `ST26SequenceListing`, having the follow ing attributes:

¹ A numeric character reference refers to a character by its Universal Character Set/Unicode code point, and uses the format: “&#nnnn;” or “&#xhhh;”, where “nnnn” is the code point in decimal form, and “hhh” is the code point in hexadecimal form.

- (a) 서열목록이 표시된 특허출원에 관한 정보를 포함하는 일반 정보 부분; 및
- (b) 하나 이상의 서열 데이터 요소를 포함하는 서열 데이터 부분으로, 각각은 차례로 하나의 서열에 대한 정보를 포함한다.

서열목록의 실시예는 별첨 III에 제공되어 있다.

39. 서열목록은 별첨 II “서열목록을 위한 문서 타입 정의(DTD)”에 설명된 DTD를 사용하여 XML 1.0으로 제시하여야 한다.

(a) XML 인스턴스의 첫 줄에는 XML 선언을 포함하여야 한다.

`<?xml version="1.0" encoding="UTF-8"?>`.

(b) XML 인스턴스의 두 번째 줄에는 문서 타입(DOCTYPE) 선언을 포함하여야 한다:

`<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">`.

40. 전체 전자 서열목록은 하나의 파일 내에 포함되어야 한다. 파일은 다음 제한 사항과 함께 유니코드 UTF-8을 사용하여 암호화(encoded) 되어야 한다.

(a) 일반 정보 부분의 ApplicantName, InventorName 및 InventionTitle 요소 및 서열 데이터 부분의 NonEnglishQualifier_value에 포함된 정보는 유니코드 제어 코드 포인트 0000-001F 및 007F-009F를 제외한 XML 1.0 사양에 표시된 모든 유효한 유니코드 문자로 구성될 수 있다. 예약 문자(reserved character) “, &, ‘, < 및 >(각각 유니코드 코드 포인트 0022, 0026, 0027, 003C 및 003E)는 41절에 명시된 대로 대체되어야 한다; 그리고

(b) 일반 정보 부분의 모든 다른 요소와 속성에 포함된 정보 및 서열 데이터 부분의 모든 다른 요소와 속성에 포함된 정보는 유니코드 기본 라틴 코드 표(예: 유니코드 코드 포인트 0020-007E로 한정 - 별첨 IV 참조)의 인쇄 가능한 문자(공백 문자 포함)로 구성되어야 한다. 예약 문자 “, &, ‘, < 및 >(각각 유니코드 코드 포인트 0022, 0026, 0027, 003C 및 003E 각각)는 41절에 명시된 대로 대체되어야 한다.

41. 서열목록의 XML 인스턴스에서 수치 문자 참조(numeric character reference)¹는 사용되어서는 안되며, 다음 예약 문자는 요소의 속성 또는 콘텐츠 값에 사용될 때 대응되는 사전 정의된 개체로 대체되어야 한다:

예약 문자	사전 정의된 개체
<	<
>	>
&	&
"	"
'	'

예로써 71절을 참조하면 된다. 유일하게 허용된 문자 엔티티(entity) 참조는 이 절에서 사전 정의된 엔티티(entities)이다.

42. 모든 필수 요소를 채워야 한다(58절에 설명된 ‘의도적으로 생략된 서열’의 경우는 제외). 콘텐츠를 사용할 수 없는 선택적 요소는 XML 인스턴스에 표시되지 않아야 한다(97절에 제공된 “replace” 한정자 값의 서열에서 삭제의 표현을 위해 경우는 제외).

루트 요소

43. 이 표준에 따른 XML 인스턴스의 루트 요소는 다음 속성을 가진 ‘ST26SequenceListing’ 요소이다:

1. 수치 문자 참조는 Universal Character Set/유니코드 코드 포인트에 의한 문자를 의미하며 그 포맷을 사용한다: “&#nnnn;” 또는 “&#xhhhh;”. 여기서 “nnnn”은 십진법 형태의 코드 포인트이며 “hhhh”가 16진법 형태의 코드 포인트이다

Attribute	Description	Mandatory/Optional
dtdVersion	Version of the DTD used to create this file in the format "V#.#", e.g., "V1_3".	Mandatory
fileName	Name of the sequence listing file.	Optional
softwareName	Name of the software that generated this file.	Optional
softwareVersion	Version of the software that generated this file.	Optional
productionDate	Date of production of the sequence listing file (format "CCYY-MM-DD").	Optional
originalFreeTextLanguageCode	The language code (see reference in paragraph 9 to ISO 639-1:2002) for the single original language in which the language-dependent free text qualifiers were prepared.	Optional
nonEnglishFreeTextLanguageCode	The language code (see reference in paragraph 9 to ISO 639-1:2002) for the NonEnglishQualifier_value elements	Mandatory when a NonEnglishQualifier_value element is present in the sequence listing

44. The following example illustrates the root element `ST26SequenceListing`, and its attributes, of an XML instance as per paragraph 43 above:

```
<ST26SequenceListing dtdVersion="V1_3" fileName="US11_405455_SEQ1.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="de" nonEnglishFreeTextLanguageCode="fr">
  {...}*
</ST26SequenceListing>
```

*{...} represents the general information part and the sequence data part that have not been included in this example.

General information part

45. The elements of the general information part relate to patent application information, as follows:

Element	Description	Mandatory/Optional
ApplicationIdentification	The application identification for which the sequence listing is submitted	Mandatory when a sequence listing is furnished at any time following the assignment of the application number
The ApplicationIdentification is composed of:		
IPOfficeCode	ST.3 Code of the office of filing	Mandatory
ApplicationNumberText	The application identification as provided by the office of filing (e.g., PCT/IB2013/099999)	Mandatory

속성	설명	필수 / 선택 사항
dtdVersion	"V#.#" 형식(예: "V1_3")에서 이 파일을 생성하는 데에 사용된 DTD 버전.	필수
fileName	서열목록 파일의 명칭.	선택
softwareName	이 파일을 생성한 소프트웨어의 명칭.	선택
SoftwareVersion	이 파일을 생성한 소프트웨어의 버전.	선택
ProductionDate	서열목록 파일의 생산 날짜 (형식 "CCYY-MM-DD").	선택
OriginalFreeTextLanguageCode	언어의존적 자유텍스트 한정자가 준비된 하나의 최초 언어에 대한 언어 코드(9 절 ISO 639-1:2002 참조)	선택
non-EnglishFreeTextLanguageCode	NonEnglishQualifier_value 요소에 대한 언어 코드(9 절 ISO 639-1:2002 참조)	NonEnglishQualifier_value 요소가 서열목록에 제시된 경우, 필수

44. 다음 예제는 상기 43절에 따라 XML 인스턴스의 루트 요소 'ST26SequenceListing' 및 해당 속성을 보여준다.

```
<ST26SequenceListing dtdVersion="V1_3" fileName="US11_405455_SEQ1.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="de" nonEnglishFreeTextLanguageCode="fr">
{...} *
</ST26SequenceListing>
```

* {...}는 해당 예에 포함되지 않은 일반 정보 부분과 서열 데이터 부분을 표현한다.

일반 정보 부분

45. 일반 정보 부분의 요소는 다음과 같이 특허출원 정보와 관련이 있다.

요소	설명	필수 / 선택
ApplicationIdentification	서열목록이 제출된 출원서 식별	출원번호 부여 후 언제라도 서열목록을 제공할 때에는 필수
'ApplicationIdentification'는 다음과 같이 구성된다:		
IPOfficeCode	출원 청의 ST.3 코드	필수
ApplicationNumberText	출원 청에서 제공한 출원서 식별 (예: PCT/IB2013/099999)	필수

Element	Description	Mandatory/ Optional
FilingDate	The date of filing of the patent application for which the sequence listing is submitted (ST.2 format "CCYY-MM-DD", using a 4-digit calendar year, a 2-digit calendar month and a 2-digit day within the calendar month, e.g., 2015-01-31)	Mandatory when a sequence listing is furnished at any time following the assignment of a filing date
ApplicantFileReference	A single unique identifier assigned by applicant to identify a particular application, typed in the characters as set forth in paragraph 40 (b)	Mandatory when a sequence listing is furnished at any time prior to assignment of the application number; otherwise, Optional
EarliestPriorityApplicationIdentification	The identification of the earliest priority application (also contains IPOfficeCode, ApplicationNumberText and FilingDate, see ApplicationIdentification above)	Mandatory where priority is claimed
ApplicantName	Name of the first mentioned applicant typed in the characters as set forth in paragraph 40 (a). This element includes the mandatory attribute languageCode as set forth in paragraph 47.	Mandatory
ApplicantNameLatin	Where ApplicantName is typed in characters other than those as set forth in paragraph 40 (b), a translation or transliteration of the name of the first mentioned applicant must also be typed in characters as set forth in paragraph 40 (b)	Mandatory where ApplicantName contains non-Latin characters
InventorName	Name of the first mentioned inventor typed in the characters as set forth in paragraph 40 (a). This element includes the mandatory attribute languageCode as set forth in paragraph 47.	Optional
InventorNameLatin	Where InventorName is typed in characters other than those as set forth in paragraph 40 (b), a translation or transliteration of the first mentioned inventor may also be typed in characters as set forth in paragraph 40 (b)	Optional
InventionTitle	Title of the invention typed in the characters as set forth in paragraph 40 (a) in the language of filing. A translation of the title of the invention into additional languages may be typed in the characters as set forth in paragraph 40 (a) using additional InventionTitle elements. This element includes the mandatory attribute languageCode as set forth in paragraph 48. The title of invention should be between two to seven words.	Mandatory in the language of filing. Optional for additional languages.

요소	설명	필수 / 선택
FilingDate	서열목록이 제출된 특허출원의 출원 날짜(4자리 연도, 2자리 월 및 2자리 날짜(예: 2015-01-31)를 사용한 ST.2 형식 "CCYY-MM-DD")	출원일 부여 후 언제라도 서열목록을 제공할 때에 필수
ApplicantFileReference	40(b)절에 규정된 문자로 입력된 특정 출원서를 식별하기 위해 출원인이 부여한 단일 고유 식별자	출원번호를 부여하기 전에 언제라도 서열목록을 제공해야하는 경우에 필수; 그렇지 않으면, 선택
EarliestPriorityApplicationIdentification	가장 이른 우선권 주장 출원서 식별 ('IPOfficeCode', 'ApplicationNumberText' 및 'FilingDate' 포함, 위의 'ApplicationIdentification' 참조)	우선권이 주장되는 경우에 필수
ApplicantName	40(a)절에 규정된 문자로 입력된, 첫 번째로 언급된 출원인 이름. 해당 요소는 47절에서 제시된 필수적인 속성 languageCode를 포함한다.	필수
ApplicantNameLatin	'ApplicantName'은 40(b)절에 규정된 문자와 다른 문자로 입력된 경우, 최초로 언급된 출원인 이름의 번역 또는 음역은 40(b)절에 명시된 문자로 입력되어야 한다.	'ApplicantName'에 비-라틴 문자가 포함된 경우에 필수
InventorName	40(a)절에 규정된 문자로 입력된, 첫 번째로 언급된 발명자의 이름. 해당 요소에는 47절에 명시된 필수 속성 'languageCode'가 포함된다.	선택
InventorNameLatin	40(b)절에 규정된 것과 다른 문자로 'InventorName'을 입력하는 경우에, 최초 언급된 발명자의 번역 또는 음역도 40(b)절에 명시된 문자로 입력 할 수 있다.	선택
InventionTitle	출원언어로서 40(a)절에 규정된 문자로 입력된 발명의 명칭. 본 발명의 명칭을 추가 언어로 번역하는 것은 추가 InventionTitle 요소를 사용하여 40(a)절에 명시된 문자로 입력될 수 있다. 본 요소는 48절에 제시된 필수 속성 'languageCode'를 필수적으로 포함한다. 발명의 명칭은 2~7개 사이의 단어를 사용하여야 한다.	출원 언어로 필수. 추가 언어의 경우에 선택.

Element	Description	Mandatory/ Optional
SequenceTotalQuantity	The total number of all sequences in the sequence listing including intentionally skipped sequences (also known as empty sequences) (see paragraph 10).	Mandatory

46. The following examples illustrate the presentation of the general information part of the sequence listing as per paragraph 45 above:

Example 1: Sequence listing filed prior to assignment of the application identification and filing date

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="Invention_SEQ1.xml"
softwareName="WIPO Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="ja">
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2013/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode="en">Keiko Nakamura</InventorName>
  <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
  <SequenceTotalQuantity>9</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1"> {...} * </SequenceData>
  <SequenceData sequenceIDNumber="2"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="3"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="4"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="5"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="6"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="7"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="8"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>
```

*{...} represents relevant information for each sequence that has not been included in this example.

Example 2: Sequence listing filed after assignment of the application identification and filing date

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="1_3" fileName="Invention_SEQ1.xml" softwareName="WIPO
Sequence" softwareVersion="1.0" productionDate="2022-05-10"
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="ja">
  <ApplicationIdentification>
    <IPOfficeCode>US</IPOfficeCode>
    <ApplicationNumberText>14/999,999</ApplicationNumberText>
    <FilingDate>2015-01-05</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode="en">Keiko Nakamura</InventorName>
```

요소	설명	필수 / 선택
SequenceTotalQuantity	의도적으로 생략된 서열(공 서열이라고도 함)을 포함하여 서열목록에 있는 모든 서열의 총 수(10 절 참조).	필수

46. 다음 실시예는 위의 45절에 따른 서열목록의 일반 정보 부분을 보여준다.

실시예 1: 출원 식별 및 출원일의 부여 이전에 제출된 서열목록

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="Invention_SEQ1.xml" softwareName="WIPO
Sequence" softwareVersion="1.0" productionDate="2022-05-10">
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="ja">
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2013/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode="en">Keiko Nakamura</InventorName>
  <InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
  <SequenceTotalQuantity>9</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1"> {...} * </SequenceData>
  <SequenceData sequenceIDNumber="2"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="3"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="4"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="5"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="6"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="7"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="8"> {...} </SequenceData>
  <SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>
```

* {...}는 이 실시예에 포함되지 않은 각 서열에 대한 관련 정보를 표현한다.

실시예 2: 출원 식별 및 출원일 부여 이후에 제출된 서열목록

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="1_3" fileName="Invention_SEQ1.xml" softwareName="WIPO
Sequence" softwareVersion="1.0" productionDate="2022-05-10">
originalFreeTextLanguageCode="en" nonEnglishFreeTextLanguageCode="ja">
  <ApplicationIdentification>
    <IPOfficeCode>US</IPOfficeCode>
    <ApplicationNumberText>14/999,999</ApplicationNumberText>
    <FilingDate>2015-01-05</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>AB123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/099999</ApplicationNumberText>
    <FilingDate>2014-07-10</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">GENOS Co., Inc.</ApplicantName>
  <InventorName languageCode="en">Keiko Nakamura</InventorName>
```

```

<InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
<SequenceTotalQuantity>9</SequenceTotalQuantity>
<SequenceData sequenceIDNumber="1"> {...}</SequenceData>
<SequenceData sequenceIDNumber="2"> {...}</SequenceData>
<SequenceData sequenceIDNumber="3"> {...}</SequenceData>
<SequenceData sequenceIDNumber="4"> {...}</SequenceData>
<SequenceData sequenceIDNumber="5"> {...}</SequenceData>
<SequenceData sequenceIDNumber="6"> {...}</SequenceData>
<SequenceData sequenceIDNumber="7"> {...}</SequenceData>
<SequenceData sequenceIDNumber="8"> {...}</SequenceData>
<SequenceData sequenceIDNumber="9"> {...}</SequenceData>
</ST26SequenceListing>*
{...} represents relevant information for each sequence that has not been included in
this example.

```

47. The name of the applicant and, optionally, the name of the inventor must be indicated in the element `ApplicantName` and `InventorName`, respectively, as they are generally referred to in the language in which the application is filed. The appropriate language code (see reference in paragraph 9 to ISO 639-1:2002) must be indicated in the `languageCode` attribute for each element. Where the applicant name indicated contains characters other than those of the Latin alphabet as set forth in paragraph 40 (b), a transliteration or translation of the applicant name must also be indicated in characters of the Latin alphabet in the element `ApplicantNameLatin`. Where the inventor name indicated contains characters other than those of the Latin alphabet, a transliteration or a translation of the inventor name may also be indicated in characters of the Latin alphabet in the element `InventorNameLatin`.

48. The title of the invention must be indicated in the element `InventionTitle` in the language of filing and may also be indicated in additional languages using multiple `InventionTitle` elements (see table in paragraph 45). The appropriate language code (see reference in paragraph 9 to ISO 639-1:2002) must be indicated in the `languageCode` attribute of the element.

49. The following example illustrates the presentation of names and title of the invention as per paragraphs 47 and 48 above:

Example: Applicant name and inventor name are each presented in Japanese and Latin characters and the title of the invention is presented in Japanese, English and French

```

<ApplicantName languageCode="ja">出願製薬株式会社</ApplicantName>
<ApplicantNameLatin>Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantNameLatin>
<InventorName languageCode="ja">特許太郎</InventorName>
<InventorNameLatin>Taro Tokkyo</InventorNameLatin>
<InventionTitle languageCode="ja">efg タンパク質をコードするマウス abcd-1 遺伝子
</InventionTitle>
<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg
protein</InventionTitle>
<InventionTitle languageCode="fr">Gène abcd-1 de Mus musculus pour protéine
efg</InventionTitle>

```

Sequence data part

50. The sequence data part must be composed of one or more `SequenceData` elements, each element containing information about one sequence.

51. Each `SequenceData` element must have a mandatory attribute `sequenceIDNumber`, in which the sequence identification number (see paragraph 10) for each sequence is contained. For example:

```
<SequenceData sequenceIDNumber="1">
```

52. The `SequenceData` element must contain a dependent element `INSDSeq`, consisting of further dependent elements as follows:

```
<InventionTitle languageCode="en">SIGNAL RECOGNITION PARTICLE RNA AND
PROTEINS</InventionTitle>
<SequenceTotalQuantity>9</SequenceTotalQuantity>
<SequenceData sequenceIDNumber="1"> {...}* </SequenceData>
<SequenceData sequenceIDNumber="2"> {...} </SequenceData>
<SequenceData sequenceIDNumber="3"> {...} </SequenceData>
<SequenceData sequenceIDNumber="4"> {...} </SequenceData>
<SequenceData sequenceIDNumber="5"> {...} </SequenceData>
<SequenceData sequenceIDNumber="6"> {...} </SequenceData>
<SequenceData sequenceIDNumber="7"> {...} </SequenceData>
<SequenceData sequenceIDNumber="8"> {...} </SequenceData>
<SequenceData sequenceIDNumber="9"> {...} </SequenceData>
</ST26SequenceListing>
```

* {...}는 이 실시예에 포함되지 않은 각 서열에 대한 관련 정보를 표현한다.

47. 출원인 이름 및 발명자 이름이 출원된 언어로 언급되는 것처럼, 출원인 이름 및 선택사항인 발명자 이름은 ApplicantName 및 InventorName 요소에 각각 제시되어야 한다. 적절한 언어 코드(9절의 ISO 639-1: 2002 참조)는 각 요소에 대한 languageCode 속성에 표시되어야 한다. 명시된 출원인 명에 40(b)절에 명시된 라틴 알파벳 이외의 문자가 포함된 경우에, 출원인 이름의 음역 또는 번역은 ApplicantNameLatin 요소에 라틴 알파벳 문자로도 제시되어야 한다. 표시된 발명자 명에 라틴 알파벳 이외의 문자가 포함된 경우에 발명자 명의 음역 또는 번역은 InventorNameLatin 요소에 라틴 알파벳 문자로도 제시될 수도 있다.

48. 본 발명의 명칭은 출원의 언어로 InventionTitle 요소에 표시되어야 하며 다수의 InventionTitle 요소를 사용하여 추가 언어로 표시될 수 있다(45절의 표 참조). 요소의 languageCode 속성에 적절한 언어 코드(9절의 ISO 639-1: 2002 참조)가 표시되어야 한다.

49. 다음 예는 상기 47절 및 48절 등에 대한 발명의 명칭 및 발명자 이름을 표시하는 방법을 보여주고 있다.:

실시예: 출원인 이름 및 발명자 이름은 각각 일본어 및 라틴 문자로 표시되며 본 발명의 명칭은 일본어, 영어 및 프랑스어로 표기된다.

```
<ApplicantName languageCode="ja"> 出願製薬株式会社 </ApplicantName>
<ApplicantNameLatin>Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantNameLatin>
<InventorName languageCode="ja"> 特許 太郎 </InventorName>
<InventorNameLatin>Taro Tokkyo</InventorNameLatin>
<InventionTitle languageCode="ja">efg タンパク質をコードするマウスabcd-1遺伝子</InventionTitle>
<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg
protein</InventionTitle>
<InventionTitle languageCode="fr">Gène abcd-1 de Mus musculus pour protéine
efg</InventionTitle>
```

서열 데이터 부분

50. 서열 데이터 부분은 하나 이상의 서열 데이터 요소로 구성되어야 하며, 각 요소는 하나의 서열에 대한 정보를 포함한다.

51. 각 SequenceData 요소는 필수 속성 sequenceIDNumber를 가져야 하며, 각 서열의 서열식별번호(10절 참조)가 포함된다. 예를 들면 다음과 같다:

```
<SequenceData sequenceIDNumber="1">
```

52. SequenceData 요소는 다음과 같은 추가적인 종속 요소로 이루어진 종속 요소 INSDSeq를 포함하여야 한다:

Element	Description	Mandatory/NotIncluded	
		Sequences	Intentionally Skipped Sequences
INSDSeq_length	Length of the sequence	Mandatory	Mandatory with no value
INSDSeq_moltype	Molecule type	Mandatory	Mandatory with no value
INSDSeq_division	Indication that a sequence is related to a patent application	Mandatory with the value "PAT"	Mandatory with no value
INSDSeq_feature-table	List of annotations of the sequence	Mandatory	Must NOT be included
INSDSeq_sequence	Sequence	Mandatory	Mandatory with the value "000"

53. The element `INSDSeq_length` must disclose the number of nucleotides or amino acids of the sequence contained in the `INSDSeq_sequence` element. For example:

```
<INSDSeq_length>8</INSDSeq_length>
```

54. The element `INSDSeq_moltype` must disclose the type of molecule that is being represented. For nucleotide sequences, including nucleotide analogue sequences, the molecule type must be indicated as DNA or RNA. For amino acid sequences, the molecule type must be indicated as AA. (This element is distinct from the qualifier "mol_type" discussed in paragraphs 55 and 84). For example:

```
<INSDSeq_moltype>AA</INSDSeq_moltype>
```

55. For a nucleotide sequence that contains both DNA and RNA segments of one or more nucleotides, the molecule type must be indicated as DNA. The combined DNA/RNA molecule must be further described in the feature table, using the feature key "source" and the mandatory qualifier "organism" with the value "synthetic construct" and the mandatory qualifier "mol_type" with the value "other DNA". Each DNA and RNA segment of the combined DNA/RNA molecule must be further described with the feature key "misc_feature" and the qualifier "note", which indicates whether the segment is DNA or RNA.

56. The following example illustrates the description of a nucleotide sequence containing both DNA and RNA segments as per paragraph 55 above:

```
<INSDSeq>
  <INSDSeq_length>120</INSDSeq_length>
  <INSDSeq_moltype>DNA</INSDSeq_moltype>
  <INSDSeq_division>PAT</INSDSeq_division>
  <INSDSeq_feature-table>
    <INSDFeature>
      <INSDFeature_key>source</INSDFeature_key>
      <INSDFeature_location>1..120</INSDFeature_location>
      <INSDFeature_qual>
        <INSDQualifier>
          <INSDQualifier_name>organism</INSDQualifier_name>
          <INSDQualifier_value>synthetic construct</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
          <INSDQualifier_name>mol_type</INSDQualifier_name>
          <INSDQualifier_value>other DNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_qual>
    </INSDFeature>
    <INSDFeature>
      <INSDFeature_key>misc_feature</INSDFeature_key>
      <INSDFeature_location>1..60</INSDFeature_location>
      <INSDFeature_qual>
        <INSDQualifier>
          <INSDQualifier_name>note</INSDQualifier_name>
          <INSDQualifier_value>DNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_qual>
    </INSDFeature>
  </INSDSeq_feature-table>
</INSDSeq>
```

요소	설명	필수적 / 포함되지 않음	
		서열	의도적으로 생략된 서열
INSDSeq_length	서열의 길이	필수	필수적이거나 값(value)이 존재하지 않음
INSDSeq_moltype	분자 타입	필수	필수적이거나 값(value)이 존재하지 않음
INSDSeq_division	서열이 특허출원과 관련되어 있다는 표시	필수이며 "PAT"값을 가지고 있어야 함	필수적이거나 값이 존재하지 않음
INSDSeq_feature-table	서열의 특징 표	필수	포함되지 않아야 한다
INSDSeq_sequence	서열	필수	필수적이며 값 "000"을 가지고 있어야 함

53. INSDSeq_length 요소는 INSDSeq_sequence 요소에 포함된 서열의 핵산염기 또는 아미노산의 수를 공개하여야 한다. 예를 들면 다음과 같다:

```
<INSDSeq_length>8</INSDSeq_length>
```

54. INSDSeq_moltype 요소는 대표되는 분자의 타입을 표현하여야 한다. 핵산염기 유사체 서열을 포함하는 핵산염기 서열의 경우, 분자 타입은 DNA 또는 RNA로 표시되어야 한다. 아미노산 서열의 경우, 분자 타입은 AA로 표시되어야 한다.(이 요소는 55절 및 84절에서 논의된 한정자 "mol_type"과는 구분된다). 예를 들면 다음과 같다:

```
<INSDSeq_moltype>AA</INSDSeq_moltype>
```

55. 하나 이상의 핵산염기의 DNA 및 RNA 단편을 모두 포함하는 핵산염기 서열의 경우에, 분자 타입은 DNA로 제시되어야 한다. 결합된 DNA/RNA 분자는 특징 기호 "source"와 "synthetic construct" 값을 가진 필수 한정자 "organism"과 "other DNA" 값을 가진 필수 한정자 "mol_type"을 사용하여 특징 표에 추가로 기술되어야 한다. 결합된 DNA/RNA 분자의 각 DNA 및 RNA 단편은 DNA 또는 RNA인지를 분명히 하기 위해 특징 기호 "misc_feature" 및 한정자 "note"를 사용하여 추가로 설명되어야 한다.

56. 다음의 예는 상기 55절에 따라, DNA 및 RNA 단편들 모두를 포함하는 핵산염기 서열의 설명을 보여준다:

```
<INSDSeq>
  <INSDSeq_length>120</INSDSeq_length>
  <INSDSeq_moltype>DNA</INSDSeq_moltype>
  <INSDSeq_division>PAT</INSDSeq_division>
  <INSDSeq_feature-table>
    <INSDFeature>
      <INSDFeature_key>source</INSDFeature_key>
      <INSDFeature_location>1..120</INSDFeature_location>
      <INSDFeature_qual>
        <INSDQualifier>
          <INSDQualifier_name>organism</INSDQualifier_name>
          <INSDQualifier_value>synthetic construct</INSDQualifier_value>
        </INSDQualifier>
        <INSDQualifier>
          <INSDQualifier_name>mol_type</INSDQualifier_name>
          <INSDQualifier_value>other DNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_qual>
    </INSDFeature>
    <INSDFeature>
      <INSDFeature_key>misc_feature</INSDFeature_key>
      <INSDFeature_location>1..60</INSDFeature_location>
      <INSDFeature_qual>
        <INSDQualifier>
          <INSDQualifier_name>note</INSDQualifier_name>
          <INSDQualifier_value>DNA</INSDQualifier_value>
        </INSDQualifier>
      </INSDFeature_qual>
    </INSDFeature>
  </INSDSeq_feature-table>
</INSDSeq>
```



```

        </INSDQualifier>
    </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
    <INSDFeature_key>misc_feature</INSDFeature_key>
    <INSDFeature_location>61..120</INSDFeature_location>
    <INSDFeature_qual>
        <INSDQualifier>
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>RNA</INSDQualifier_value>
        </INSDQualifier>
    </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>cgacccaacgcgcgcgactcattgctcctcgtaccgctcgagcggc</INSDSeq_sequence>
</INSDSeq>

```

57. The element `INSDSeq_sequence` must disclose the sequence. Only the appropriate symbols set forth in Annex I (see Section 1, Table 1 and Section 3, Table 3) must be included in the sequence. The sequence must not include numbers, punctuation or whitespace characters.

58. An intentionally skipped sequence must be included in the sequence listing and represented as follows:

- (a) the element `SequenceData` and its attribute `sequenceIDNumber`, with the sequence identification number of the skipped sequence provided as the value;
- (b) the elements `INSDSeq_length`, `INSDSeq_moltype`, `INSDSeq_division`, present but with no value provided;
- (c) the element `INSDSeq_feature-table` must not be included; and
- (d) the element `INSDSeq_sequence` with the string "000" as the value.

59. The following example illustrates the representation of an intentionally skipped sequence as per paragraph 58 above:

```

<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length/>
    <INSDSeq_moltype/>
    <INSDSeq_division/>
    <INSDSeq_sequence>000</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

Feature table

60. The feature table contains information on the location and roles of various regions within a particular sequence. A feature table is required for every sequence, except for any intentionally skipped sequence, in which case it must not be included. The feature table is contained in the element `INSDSeq_feature-table`, which consists of one or more `INSDFeature` elements.

61. Each `INSDFeature` element describes one feature, and consists of dependent elements as follows:

Element	Description	Mandatory/Optional
<code>INSDFeature_key</code>	A word or abbreviation indicating a feature	Mandatory
<code>INSDFeature_location</code>	Region of the sequence which corresponds to the feature	Mandatory
<code>INSDFeature_qual</code>	Qualifier containing auxiliary information about a feature	Mandatory where the feature key requires one or more qualifiers, e.g., source; otherwise, Optional

```

        </INSDQualifier>
      </INSDFeature_qual>
    </INSDFeature>
  <INSDFeature>
    <INSDFeature_key>misc_feature</INSDFeature_key>
    <INSDFeature_location>61..120</INSDFeature_location>
    <INSDFeature_qual>
      <INSDQualifier>
        <INSDQualifier_name>note</INSDQualifier_name>
        <INSDQualifier_value>RNA</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>cgaccacgcgctccgaggaaccaaccatcacgtttgaggacttcgtgaaggaattggataataaccgctcc
taccaaaatggcgagcgcgactcattgctcctcgtaccgctcgagcggc</INSDSeq_sequence>
</INSDSeq>

```

57. INSDSeq_sequence 요소는 서열을 표시하여야 한다. 별첨 1에 명시된 적절한 기호(섹션 1, 표 1 및 섹션 3, 표 3 참조) 만 서열에 포함되어야 한다. 서열에는 숫자, 구두점, 또는 공백 문자가 포함되지 않아야 한다.

58. 의도적으로 생략된 서열은 서열목록에 포함되어야 하며 다음과 같이 표현되어야 한다:

(a) 생략된 서열의 서열식별번호가 값으로 제공된 SequenceData 요소 및 그 속성인 sequenceIDNumber ;

(b) 존재하지만 값이 제공되지 않은 INSDSeq_length, INSDSeq_moltype, INSDSeq_division 요소;

(c) INSDSeq_feature-table 요소는 포함되지 않아야 한다; 그리고

(d) 문자열 "000"을 값으로 하는 INSDSeq_sequence 요소.

59. 다음 예는 위의 58절에 따라 의도적으로 생략된 서열을 표현한다:

```

<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length/>
    <INSDSeq_moltype/>
    <INSDSeq_division/>
    <INSDSeq_sequence>000</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

특징 표

60. 특징 표에는 특정 서열 내에서 다양한 영역의 위치와 역할에 대한 정보가 포함된다. 특징 표는 이를 포함하면 안되는 의도적으로 생략된 서열을 제외한 모든 서열에 대해서 요구된다. 특징 표는 하나 이상의 INSDFeature 요소로 이루어진 INSDSeq_feature-table 요소에 포함되어 있다.

61. 각 INSDFeature 요소는 하나의 특징을 설명하며 다음과 같이 종속 요소로 이루어진다.

요소	설명	필수 / 선택
INSDFeature_key	특징을 나타내는 단어 또는 약어	필수
INSDFeature_location	특징에 해당하는 서열 영역	필수
INSDFeature_qual	특징에 대한 보조 정보를 포함하는 한정자	특징 기호에 하나 이상의 한정자가 필요한 경우(예를 들면, source), 필수; 그 이외의 경우 선택

Feature keys

62. Annex I contains an exclusive listing of feature keys that must be used under this Standard, along with an exclusive listing of associated qualifiers and an indication as to whether those qualifiers are mandatory or optional. Section 5 of Annex I provides the exclusive listing of feature keys for nucleotide sequences and Section 7 provides the exclusive listing of feature keys for amino acid sequences.

Mandatory feature keys

63. The “source” feature key is mandatory for all nucleotide sequences and for all amino acid sequences, except for any intentionally skipped sequence. Each sequence must have a single “source” feature key spanning the entire sequence. Where a sequence originates from multiple sources, those sources may be further described in the feature table, using the feature key “misc_feature” and the qualifier “note” for nucleotide sequences, and the feature key “REGION” and the qualifier “note” for amino acid sequences.

Feature location

64. The mandatory element `INSDFeature_location` must contain at least one location descriptor, which defines a site or a region corresponding to a feature of the sequence in the `INSDSeq_sequence` element. Amino acid sequences must contain one and only one location descriptor in the mandatory `INSDFeature_location` element. Nucleotide sequences may have more than one location descriptor in the mandatory `INSDFeature_location` element when used in conjunction with one or more location operator(s) (see paragraphs 67 to 70).

65. The location descriptor can be a single residue number, a region delimiting a contiguous span of residue numbers, or a site or region that extends beyond the specified residue or span of residues. The location descriptor must not include numbering for residues beyond the range of the sequence in the `INSDSeq_sequence` element. For nucleotide sequences only, a location descriptor can be a site between two adjacent residue numbers. Multiple location descriptors must be used in conjunction with a location operator when a feature corresponds to discontinuous sites or regions of a nucleotide sequence (see paragraphs 67 to 70).

66. The syntax for each type of location descriptor is indicated in the table below, where x and y are residue numbers, indicated as positive integers, not greater than the length of the sequence in the `INSDSeq_sequence` element, and x is less than y.

(a) Location descriptors for nucleotide and amino acid sequences:

Location descriptor type	Syntax	Description
Single residue number	x	Points to a single residue in the sequence.
Residue numbers delimiting a sequence span	x..y	Points to a continuous range of residues bounded by and including the starting and ending residues.
Residues before the first or beyond the last specified residue number	<x >x <x..y x..>y <x..>y	Points to a region including a specified residue or span of residues and extending beyond a specified residue. The '<' and '>' symbols may be used with a single residue or the starting and ending residue numbers of a span of residues to indicate that a feature extends beyond the specified residue number.

(b) Location descriptors for nucleotide sequences only:

Location descriptor type	Syntax	Description
A site between two adjoining nucleotides	x^y	Points to a site between two adjoining nucleotides, e.g., endonucleolytic cleavage site. The position numbers for the adjacent nucleotides are separated by a caret (^). The permitted formats for this descriptor are x^x+1 (for example 55^56), or, for circular nucleotides, x^1, where “x” is the full length of the molecule, i.e. 1000^1 for circular molecule with length 1000.

특징 기호

62. 별첨 1에는 본 표준에 따라 사용해야 하는 특징 기호의 전용 목록과 관련 한정자 목록 및 해당 한정자가 필수인지 선택인지에 대한 표시가 포함되어 있다. 별첨 1의 섹션 5는 핵산염기 서열에 대한 특징 기호의 전용 목록을 제공하고, 섹션 7은 아미노산 서열에 대한 특징 기호의 목록을 제공한다.

필수 특징 기호

63. 의도적으로 생략된 서열을 제외하고, "source" 특징 기호는 모든 핵산염기 서열과 모든 아미노산 서열에 필수이다. 각 서열에는 전체 서열에 걸쳐있는 단일 "source" 특징 기호가 있어야 한다. 서열이 다수의 공급원으로부터 유래되는 경우에, 이들 공급원은 핵산염기 서열에 대해 특징 기호 "misc_feature" 및 한정자 "note", 아미노산 서열에 대해 특징 기호 "REGION" 및 한정자 "note"를 사용하여 특징 표에 추가로 기술될 수 있다.

특징 위치

64. 필수 요소 INSDFeature_location은 INSDSeq_sequence 요소의 서열 특징에 해당하는 부위 또는 영역을 정의하는 위치 설명자를 하나 이상 포함하여야 한다. 아미노산 서열은 필수 요소 INSDFeature_location에 하나의 위치 설명자만을 포함하여야 한다. 핵산염기 서열의 경우 하나 이상의 위치 연산자와 함께 사용할 경우, 필수 INSDFeature_location 요소에 하나 이상의 위치 설명자를 포함할 수 있다(67절~70절 참조).

65. 위치 설명자는 단일 잔기 번호, 잔기 수의 연속 범위를 구분하는 영역 또는 지정된 잔기 또는 잔기 범위를 벗어나는 부위 또는 영역일 수 있다. 위치 설명자는 INSDSeq_sequence 요소의 서열 범위를 벗어나는 잔기에 대한 번호를 포함하지 않아야 한다. 핵산염기 서열의 경우에만 위치 설명자는 두 개의 인접한 잔기 번호 사이의 부위가 될 수 있다. 다수의 위치 설명자들은 특징이 핵산염기 서열의 불연속적인 부위 또는 영역에 해당하는 경우 위치 연산자와 함께 사용되어야 한다(67절~70절 참조).

66. 각 타입의 위치 설명자에 대한 구문은 아래 표에 표시되어 있다. 여기서 x와 y는 INSDSeq_sequence 요소의 서열 길이보다 크지 않은 양수인 정수로 표시되는 잔류 번호이며 x는 y보다 작다.

(a) 핵산염기 및 아미노산 서열에 대한 위치 설명자:

위치 설명자 타입	구문	설명
단일 잔기 번호	x	서열에서 단일 잔기를 가리킨다.
서열 범위를 한정하는 잔기 번호	x..y	시작 및 말단 잔기에 의해 묶이고 이를 포함하는 연속적인 범위의 잔기를 가리킨다.
첫 번째로 지정된 잔기 수 이전의 또는 마지막으로 지정된 잔기 수 이후의 잔기들	<x >x <x..y x..>y <x..>y	특정 잔기 또는 잔기 범위를 포함하고 특정 잔기를 넘어 확장되는 영역을 가리킨다. '<' 및 '>'기호를 단일 잔기 또는 서열 범위의 시작 및 말단과 함께 사용하여 특징이 지정된 잔기 수 이상으로 확장될 수 있음을 명시한다.

(b) 핵산염기 서열만을 위한 위치 설명자:

위치 설명자 타입	구문	설명
두 개의 인접한 핵산염기 사이의 부위	x^y	2개의 인접한 핵산 사이의 부위를 가리킨다(예: 내부핵산서열 절단 부위). 인접한 핵산의 위치 번호는 캐럿(^)으로 구분된다. 이 설명자에 허용되는 형식은 x^x+1(예: 55^56)이다. 또는 원형 핵산염기에 대해 허용되는 형식은 x^1으로 여기서 "x"는 분자의 전체 길이, 즉 길이가 1000인 원형 분자의 경우에 1000^1이다.

(c) Location descriptors for amino acid sequences only:

Location descriptor type	Syntax	Description
Residue numbers joined by an intrachain cross-link	x..y	Points to amino acids joined by an intrachain linkage when used with a feature that indicates an intrachain cross-link, such as "CROSSLNK" or "DISULFID".

67. The INSDFeature_location element of nucleotide sequences may contain one or more location operators. A location operator is a prefix to either one location descriptor or a combination of location descriptors corresponding to a single but discontinuous feature, and specifies where the location corresponding to the feature on the indicated sequence is found or how the feature is constructed. A list of location operators is provided below with their definitions. Location operators can be used for nucleotides only.

Location syntax	Location description
join(location, location, ..., location)	The indicated locations are joined (placed end-to-end) to form one contiguous sequence.
order(location, location, ..., location)	The elements are found in the specified order but nothing is implied about whether joining those elements is reasonable.
complement(location)	Indicates that the feature is located on the strand complementary to the sequence span specified by the location descriptor, when read in the 5' to 3' direction or in the direction that mimics the 5' to 3' direction.

68. The join and order location operators require that at least two comma-separated location descriptors be provided. Location descriptors involving sites between two adjacent residues, i.e. x^y, must not be used within a join or order location. Use of the join location operator implies that the residues described by the location descriptors are physically brought into contact by biological processes (for example, the exons that contribute to a coding region feature).

69. The location operator "complement" can be used in combination with either "join" or "order" within the same location. Combinations of "join" and "order" within the same location must not be used.

70. The following examples illustrate feature locations, as per paragraphs 64 to 69 above:

(a) locations for nucleotide and amino acid sequences:

Location Example	Description
467	Points to residue 467 in the sequence.
340..565	Points to a continuous range of residues bounded by and including residues 340 and 565.
<1	Points to a feature location before the first residue.
<345..500	Indicates that the exact lower boundary point of a feature is unknown. The location begins at some residue previous to 345 and continues to and includes residue 500.
<1..888	Indicates that the feature starts before the first sequenced residue and continues to and includes residue 888.
1..>888	Indicates that the feature starts at the first sequenced residue and continues beyond residue 888.
<1..>888	Indicates that the feature starts before the first sequenced residue and continues beyond residue 888.

(c) 아미노산 서열만을 위한 위치 설명자:

위치 설명자 타입	구문	설명
사슬간 가교 결합에 의해 결합된 잔기 번호	x..y	"CROSSLINK" 또는 "DISULFID"와 같은 사슬간 가교 결합을 나타내는 특징과 함께 사용될 때 사슬간 결합에 의해 결합되는 아미노산을 가리킨다.

67. 핵산염기 서열의 INSDFeature_location 요소는 하나 이상의 위치 연산자(location operator)를 포함할 수 있다. 위치 연산자는 하나의 위치 설명자 또는 단일이지만 불연속적인 특징에 대응하는 위치 설명자의 조합에 대한 접두사이며, 표시된 서열에서 특징에 대응하는 위치가 발견되는 부위나 특징이 구성되는 방법을 지정한다. 위치 연산자 목록이 아래 정의와 함께 제공된다. 위치 연산자는 핵산염기 서열에만 사용할 수 있다.

위치 구문	위치 설명
join(location,location, ... location)	표시된 위치는 하나의 연속적인 서열을 형성하기 위해 결합되어 있다(말단에서 말단까지 배치).
order(location,location, ... location)	본 요소는 지정된 순서대로 발견되지만 해당 요소를 결합하는 것이 합리적인지에 대해서는 암시하지 않는다.
complement(location)	5'-말단 → 3'-말단 방향으로 읽거나, 5'-말단 → 3'-말단 방향을 모방하는 방향으로 읽을 경우에 특징은 위치 설명자가 지정한 서열 범위에 상보적인 가닥에 위치해 있음을 나타낸다.

68. 결합(join) 및 순서(order) 위치 연산자는 적어도 두 개의 선택표로 구분된 위치 설명자가 제공되어야 한다. 인접한 두 잔기 사이의 부위, 즉 x^y를 포함하는 위치 설명자는 결합 또는 순서 위치에서 사용되어서는 안 된다. 결합 위치 연산자의 사용은 위치 설명자에 의해 기술된 잔기가 생물학적 프로세스(예를 들면, 코딩 영역 특징에 기여하는 엑손)에 의해 물리적으로 접촉된다는 것을 의미한다.

69. 위치 연산자 "complement"는 동일한 위치에서 "join" 또는 "order"와 결합하여 사용될 수 있다. 동일한 위치에서 "join"과 "order"의 결합을 사용해서는 안 된다.

70. 다음 예는 상기의 64절~69절에 따른 특징 위치를 보여준다.

(a) 핵산염기 및 아미노산 서열의 위치:

위치 예	설명
467	서열에서 잔기 467을 가리킨다.
340..565	잔기 340 및 565를 경계로 하는 잔기의 연속 범위를 가리킨다.
<1	첫 번째 잔기 이전의 특징 위치를 가리킨다.
<345..500	특징의 정확한 하한(下限)은 알려져 있지 않음을 명시한다. 위치는 345 이전의 일부 잔기에서 시작하여 잔기 500까지 계속되며 잔기 500을 포함한다.
<1..888	특징이 첫 번째 서열 잔기보다 먼저 시작하고 잔기 888로 계속되며 이를 포함함을 나타낸다.
1..>888	특징이 첫 번째 시퀀싱 된 잔기에서 시작하여 잔기 888 이상으로 계속됨을 나타낸다.
<1..>888	특징이 첫 번째 시퀀싱 된 잔기 전에 시작하고 잔기 888 이상으로 계속됨을 나타낸다.

(b) locations for nucleotide sequences only:

Location example	Description
123^124	Points to a site between residues 123 and 124.
join(12..78,134..202)	Indicates that regions 12 to 78 and 134 to 202 should be joined to form one contiguous sequence.
complement(34..126)	Starts at the nucleotide complementary to 126 and finishes at the nucleotide complementary to nucleotide 34 (the feature is on the strand complementary to the presented strand).
complement(join(2691..4571,4918..5163))	Joins nucleotides 2691 to 4571 and 4918 to 5163, then complements the joined segments (the feature is on the strand complementary to the presented strand).
join(complement(4918..5163),complement(2691..4571))	Complements regions 4918 to 5163 and 2691 to 4571, then joins the complemented segments (the feature is on the strand complementary to the presented strand).

(c) locations for amino acid sequences only:

Location example	Description
340..565	Indicates that the amino acids at positions 340 and 565 are joined by an intrachain linkage when used with a feature that indicates an intrachain cross-link, such as "CROSSLNK" or "DISULFID".

71. In an XML instance of a sequence listing, the characters "<" and ">" in a location descriptor must be replaced by the appropriate predefined entities (see paragraph 41). For example:

Feature location "<1":

```
<INSDFeature_location>&lt;1</INSDFeature_location>
```

Feature location "1..>888":

```
<INSDFeature_location>1..&gt;888</INSDFeature_location>
```

Feature qualifiers

72. Qualifiers are used to supply information about features in addition to that conveyed by the feature key and feature location. There are three types of value formats to accommodate different types of information conveyed by qualifiers, namely:

- (a) free text (see paragraphs 85 to 87);
- (b) controlled vocabulary or enumerated values (e.g., a number or date); and
- (c) sequences.

73. Section 6 of Annex I provides the exclusive listing of qualifiers and their specified value formats, if any, for each nucleotide sequence feature key and Section 8 provides the exclusive listing of qualifiers for each amino acid sequence feature key.

74. Any sequence encompassed by paragraph 7 which is provided as a qualifier value must be separately included in the sequence listing and assigned its own sequence identification number (see paragraph 10).

Mandatory feature qualifiers

75. One mandatory feature key, i.e., "source" for nucleotide sequences and amino acid sequences, requires two mandatory qualifiers, "organism" and "mol_type". Some optional feature keys also require mandatory qualifiers.

Qualifier elements

76. The element `INSDFeature_qual` contains one or more `INSDQualifier` elements. Each `INSDQualifier` element represents a single qualifier and consists of three dependent elements as follows:

(b) 핵산염기 서열만을 위한 위치:

위치 예	설명
123^124	잔기 123 및 124 사이의 위치를 가리킨다.
join(12..78,134..202)	영역 12-78과 134-202가 하나의 연속 서열을 형성하기 위해 결합되어야 함을 나타낸다.
complement(34..126)	126에 상보적인 핵산염기에서 시작하여 핵산염기 34에 상보적인 핵산염기에서 마무리된다(특징은 제시된 가닥에 상보적인 가닥 상에 있다).
complement(join(2691..4571, 4918..5163))	핵산염기 2691~4571 및 4918~5163을 연결한 다음, 결합된 단편에 대해 상보적이다(특징은 제시된 가닥에 상보적인 가닥에 있다).
join(complement(4918..5163), complement(2691..4571))	영역 4918~5163 및 2691~4571에 대해 상보적인 단편을 만든 다음, 각각의 상보적인 단편을 결합시킨다(특징은 제시된 가닥에 상보적인 가닥에 있다).

(c) 아미노산 서열만을 위한 위치:

위치 예	설명
340..565	340 및 565 위치에 있는 아미노산은 "CROSSLNK" 또는 "DISULFID"와 같은 사슬간 가교 결합을 나타내는 특징과 함께 사슬간 결합에 의해 결합된다는 것을 나타낸다.

71. 서열목록의 XML 인스턴스에서, 위치 설명자의 "<" 및 ">" 문자는 사전 정의된 적절한 개체로 대체되어야 한다(41절 참조). 예를 들면 다음과 같다:

Feature location "<1":
<INSDFeature_location><1</INSDFeature_location>

Feature location "1.>888":
<INSDFeature_location>1.>888</INSDFeature_location>

특징 한정자

72. 한정자는 특징 기호와 특징 위치에 추가로 특징에 대한 정보를 제공하기 위해 사용된다. 한정자가 전달하는 서로 다른 타입의 정보를 수용하기 위한 세 가지 타입의 값 형식은 다음과 같다.

- (a) 자유텍스트(85절~87절 참조);
- (b) 통제 어휘 또는 나열된 값(예: 숫자 또는 날짜); 및
- (c) 서열.

73. 별첨 1의 섹션 6은 각 핵산염기 서열 특징 기호에 대한 전용 한정자 목록과 지정된 값 형식(있는 경우)을 제공하고, 섹션 8은 각 아미노산 서열 특징 기호에 대한 전용 한정자 목록을 제공한다.

74. 한정자 값으로 제공되는 7절에 포함된 모든 서열은 서열목록에 별도로 포함되고 독립된 서열식별번호를 부여하여야 한다(10절 참조).

필수 특징 한정자

75. 하나의 필수 특징 기호, 즉 핵산염기 서열 및 아미노산 서열의 "source"는 2개의 필수 한정자로, "organism" 및 "mol_type"이 필요하다. 일부 선택 특징 기호에는 필수 한정자가 필요하다.

한정자 요소

76. INSDFeature_qual 요소는 하나 이상의 INSDQualifier 요소를 포함한다. 각 INSDQualifier 요소는 단일 한정자를 표현하며 다음과 같이 세 개의 종속 요소로 이루어진다:

Element	Description	Mandatory/Optional
INSDQualifier_name	Name of the qualifier (see Annex I, Sections 6 and 8)	Mandatory
INSDQualifier_value	Value of the qualifier, if any, in the specified format (see Annex I, Sections 6 and 8) and composed in the characters as set forth in paragraph 40(b).	Mandatory, when specified (see paragraph 87 and Annex I, Sections 6 and 8)
NonEnglishQualifier_value	Value of the qualifier, if any, in the specified format (see Annex I, Sections 6 and 8) and composed in the characters as set forth in paragraph 40(a).	Mandatory, when specified (see paragraph 87 and Annex I, Sections 6 and 8)

77. The organism qualifier, i.e., “organism” for nucleotide sequences (see Annex I, Section 6) and “organism for amino acid sequences (see Annex I, Section 8) must disclose the source, i.e., a single organism or origin, of the sequence. Organism designations should be selected from a taxonomy database.

78. If the sequence is naturally occurring and the source organism has a Latin genus and species designation, that designation must be used as the qualifier value. The preferred English common name may be specified using the qualifier “note” for nucleotide sequences and amino acid sequences, but must not be used in the organism qualifier value.

79. The following examples illustrate the source of a sequence as per paragraphs 77 and 78 above:

Example 1: Source for a nucleotide sequence

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..5164</INSDFeature_location>
    <INSDFeature_qualifiers>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>Solanum lycopersicum</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>note</INSDQualifier_name>
        <INSDQualifier_value>common name: tomato</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>genomic DNA</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qualifiers>
  </INSDFeature>
</INSDSeq_feature-table>
```

Example 2: Source for an amino acid sequence

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..174</INSDFeature_location>
    <INSDFeature_qualifiers>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>protein</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qualifiers>
  </INSDFeature>
</INSDSeq_feature-table>
```

요소	설명	필수 / 선택 사항
INSDQualifier_name	한정자 명칭(별첨 I, 섹션 6 및 8)	필수
INSDQualifier_value	지정된 형식(별첨 I, 섹션 6 및 8 참조) 및 40(b)절에 제시된 문자로 구성된 한정자의 값(있는 경우)	지정된 경우 필수(87절, 및 별첨 I의 섹션 6 및 8 참조)
NonEnglishQualifier_value	지정된 형식(별첨 I, 섹션 6 및 8 참조) 및 40(a)절에 제시된 문자로 구성된 한정자의 값(있는 경우)	지정된 경우 필수(87절, 및 별첨 I의 섹션 6 및 8 참조)

77. 생물체 한정자, 즉 핵산염기 서열(별첨 I, 섹션 6 참조) 및 아미노산 서열(별첨 I, 섹션 8 참조)에 대한 "organism"은 기원(source), 즉 서열의 단일 생물체 또는 기원을 명시하여야 한다. 생물체 지정은 생물 분류(taxonomy) 데이터베이스에서 선택하여야 한다.

78. 서열이 자연적으로 발생하고 기원 생물체(source organism)에 라틴어 속명(genus)과 종명(species)이 있는 경우에, 그 명칭을 한정자의 값으로 사용하여야 한다. 일반적인 영어 생물명칭은 핵산염기 서열 및 아미노산 서열의 한정자 "note"를 사용하여 지정할 수 있지만 organism 한정자 값에는 사용해서는 아니된다.

79. 하기 실시예는 상기 77절 및 78절에 따른 서열의 기원을 예시한다:

실시예 1: 핵산염기 서열의 기원

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..5164</INSDFeature_location>
    <INSDFeature_qual>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>Solanum lycopersicum</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>note</INSDQualifier_name>
        <INSDQualifier_value>common name: tomato</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>genomic DNA</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>
</INSDSeq_feature-table>
```

실시예 2: 아미노산 서열의 기원

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..174</INSDFeature_location>
    <INSDFeature_qual>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>Homo sapiens</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>protein</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>
</INSDSeq_feature-table>
```

80. If the sequence is naturally occurring and the source organism has a known Latin genus, but the species is unspecified or unidentified, then the organism qualifier value must indicate the Latin genus followed by “sp.” For example:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Bacillus sp.</INSDQualifier_value>
```

81. If the sequence is naturally occurring, but the Latin organism genus and species designation is unknown, then the organism qualifier value must be indicated as “unidentified”. Any known taxonomic information should be indicated in the qualifier “note” for nucleotide sequences and the qualifier “note” for amino acid sequences. For example:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>unidentified</INSDQualifier_value>
<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>bacterium B8</INSDQualifier_value>
```

82. If the sequence is naturally occurring and the source organism does not have a Latin genus and species designation, such as a virus, then another acceptable scientific name (e.g., “Canine adenovirus type 2”) must be used as the organism qualifier value. For example:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Canine adenovirus type 2</INSDQualifier_value>
```

83. If the sequence is not naturally occurring, the organism qualifier value must be indicated as “synthetic construct”. Further information with respect to the way the sequence was generated may be specified using the qualifier “note” for nucleotide sequences and the qualifier “note” for amino acid sequences. For example:

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..40</INSDFeature_location>
    <INSDFeature_qual>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>synthetic construct</INSDQualifier_value>
      </INSDQualifier>
    </INSDQualifier>
    <INSDQualifier_name> mol_type</INSDQualifier_name>
    <INSDQualifier_value>protein</INSDQualifier_value>
  </INSDQualifier>
  <INSDQualifier_name>note</INSDQualifier_name>
  <INSDQualifier_value>synthetic peptide used as assay for
antibodies</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
```

84. The “mol_type” qualifier for nucleotide sequences (see Annex I, Section 6) and “mol_type” qualifier for amino acid sequences (see Annex I, Section 8) must disclose the type of molecule represented in the sequence. These qualifiers are distinct from the element `INSDSeq_moltype` discussed in paragraph 54:

(a) For a nucleotide sequence, the “mol_type” qualifier value must be one of the following: “genomic DNA”, “genomic RNA”, “mRNA”, “tRNA”, “rRNA”, “other RNA”, “other DNA”, “transcribed RNA”, “viral cRNA”, “unassigned DNA”, or “unassigned RNA”. If the sequence is not naturally occurring, i.e. the value of the “organism” qualifier is “synthetic construct”, the “mol_type” qualifier value must be either “other RNA” or “other DNA”;

(b) For an amino acid sequences, the “mol_type” qualifier value is “protein”.

Free text

85. Free text, as indicated in paragraph 3 (n), is a type of value format for certain qualifiers presented in the form of a descriptive text phrase or other specified format (as indicated in Annex I).

86. The use of free text must be limited to a few short terms indispensable for the understanding of a characteristic of the sequence. For each qualifier, the free text must not exceed 1000 characters.

80. 서열이 자연적으로 발생하고 기원 생물체에 공지된 라틴어 속(genus)이 있지만, 종(species)을 지정하지 않았거나 확인 불가능한 경우, organism 한정자 값은 속(genus) 뒤에 "sp."를 표시하여야 한다. 예를 들면 다음과 같다:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Bacillus sp.</INSDQualifier_value>
```

81. 서열이 자연적으로 발생하지만 라틴어의 생물체 속(genus)과 종(species) 명칭이 공지되지 않은 경우, 생물체 한정자 값은 "식별되지 않음(identified)"으로 표시되어야 한다. 공지된 분류학적 정보는 핵산염기 서열의 한정자 "note" 및 아미노산 서열의 한정자 "note"에 표시되어야 한다. 예를 들면 다음과 같다:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>unidentified</INSDQualifier_value>
<INSDQualifier_name>note</INSDQualifier_name>
<INSDQualifier_value>bacterium B8</INSDQualifier_value>
```

82. 서열이 자연적으로 발생하고 기원 생물체에 바이러스와 같은 라틴어 속(genus)과 종(species) 지정이 없는 경우, 허용 가능한 또 다른 과학적 명칭(예: "Canine Adenovirus type 2")을 organism 한정자 값으로 사용하여야 한다. 예를 들면 다음과 같다:

```
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>Canine adenovirus type 2</INSDQualifier_value>
```

83. 서열은 자연적으로 발생하지 않은 경우에, 생물체 한정자 값은 "synthetic construct"로 표시되어야 한다. 서열이 생성된 방식에 관한 추가 정보는 핵산염기 서열에 대한 한정자 "note" 및 아미노산 서열에 대한 한정자 "note"를 사용하여 특정될 수 있다. 예를 들면 다음과 같다:

```
<INSDSeq_feature-table>
  <INSDFeature>
    <INSDFeature_key>source</INSDFeature_key>
    <INSDFeature_location>1..40</INSDFeature_location>
    <INSDFeature_qual>
      <INSDQualifier>
        <INSDQualifier_name>organism</INSDQualifier_name>
        <INSDQualifier_value>synthetic construct</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>mol_type</INSDQualifier_name>
        <INSDQualifier_value>protein</INSDQualifier_value>
      </INSDQualifier>
      <INSDQualifier>
        <INSDQualifier_name>note</INSDQualifier_name>
        <INSDQualifier_value>synthetic peptide used as assay for antibodies</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>
</INSDSeq_feature-table>
```

84. 핵산염기 서열에 대한 "mol_type" 한정자(별첨 I, 섹션 6 참조) 및 아미노산 서열에 대한 "mol_type" 한정자(별첨 I, 섹션 8 참조)는 서열에 표현된 분자 타입을 공개하여야 한다. 이러한 한정자들은 54절에서 논의된 INSDSeq_moltype 요소와 구별된다:

(a) 핵산염기 서열의 경우에, "mol_type" 한정자 값은 "genomic DNA", "genomic RNA", "mRNA", "rRNA", "rRNA", "other RNA", "other DNA", "transcribed RNA", "viral cRNA", "unassigned DNA", or "unassigned RNA" 중 하나이어야 한다. 서열이 자연적으로 발생하지 않으면, 즉 "organism" 한정자의 값이 "synthetic construct"인 경우, "mol_type" 한정자 값은 "other RNA" 또는 "other DNA"이어야 하고;

(b) 아미노산 서열의 경우에, "mol_type" 한정자 값은 "protein"이다.

자유텍스트

85. 3(n)절에 표시된 것과 같은 자유텍스트는 설명적 텍스트 구 또는 기타 특정 형식의 형태로 표시된 특정 한정자 값의 한 형식이다 (별첨 I에 표시됨).

86. 자유텍스트의 사용은 서열의 특징을 이해하는 데에 없어서는 안 될 몇 가지 짧은 용어로 제한되어야 한다. 각 한정자에 대해 자유텍스트는 1000자를 초과하지 않아야 한다.

87. Language-dependent free text, as indicated in paragraph 3 (o), is the free text value of certain qualifiers that is language-dependent in that it may require translation for national, regional or international procedures. Qualifiers for nucleotide sequences with a language-dependent free text value format are identified in Annex I, Section 6, Table 5. Qualifiers for amino acid sequences with a language-dependent free text value format are identified in Annex I, Section 8, Table 6.

(a) Language-dependent free text must be presented in the `INSDQualifier_value` element in English, or in the `NonEnglishQualifier_value` element in a language other than English, or in both elements. Note that if an organism name is a Latin genus and species name, no translation is required. Technical terms and proper names originating from non-English words that are used internationally are considered English for the purpose of the value of the `INSDQualifier_value` element (e.g., 'in vitro', 'in vivo').

(b) If a `NonEnglishQualifier_value` element is present in a sequence listing, the appropriate language code (see reference in paragraph 9 to ISO 639-1:2002) must be indicated in the `nonEnglishFreeTextLanguageCode` attribute in the root element (see paragraph 43). All `NonEnglishQualifier_value` elements in a single sequence listing must have values in the language indicated in the `nonEnglishFreeTextLanguageCode` attribute. The `NonEnglishQualifier_value` element is permitted only for qualifiers that have a language-dependent free text value format.

(c) Where `NonEnglishQualifier_value` and `INSDQualifier_value` are both present for a single qualifier, the information contained in the two elements must be equivalent. One of the following conditions must be true: `NonEnglishQualifier_value` contains a translation of the value of `INSDQualifier_value`; or, `INSDQualifier_value` contains a translation of the value of `NonEnglishQualifier_value`; or, both elements contain a translation of the qualifier value from the language specified in the `originalFreeTextLanguageCode` attribute (see paragraph 43).

(d) For language-dependent qualifiers, the `INSDQualifier` element may include an optional attribute `id`. The value of this attribute must be in the format "q" followed by a positive integer, e.g. "q23", and must be unique to one `INSDQualifier` element, i.e. the attribute value must only be used once in a sequence listing file.

88. The following examples illustrate the presentation of language-dependent free text as discussed in paragraph 87.

Example 1: language-dependent free text in an `INSDQualifier_value` element:

```
<INSDFeature>
  <INSDFeature_key>regulatory</INSDFeature_key>
  <INSDFeature_location>1..60</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q1">
      <INSDQualifier_name>function</INSDQualifier_name>
      <INSDQualifier_value>binds to regulatory protein Est3</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Example 2: language-dependent free text in an `INSDQualifier_value` element and a `NonEnglishQualifier_value` element:

```
<INSDFeature>
  <INSDFeature_key>ACT_SITE</INSDFeature_key>
  <INSDFeature_location>51..64</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q45">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>cleaves carbohydrate chain</INSDQualifier_value>
      <NonEnglishQualifier_value>clive la chaîne glucidique</NonEnglishQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Example 3: language-dependent free text in a `NonEnglishQualifier_value` element:

```
<INSDFeature>
  <INSDFeature_key>ACT_SITE</INSDFeature_key>
  <INSDFeature_location>51..64</INSDFeature_location>
  <INSDFeature_qual>
```

87. 3(o)절에서 표시한 바와 같이 언어의존적 자유텍스트는 국내, 지역 또는 국제 절차를 위해 번역이 필요할 수 있다는 점에서 언어에 의존하는 특정 한정자의 자유텍스트 값이다. 언어의존적 자유텍스트 값 형식을 가진 핵산염기 서열에 대한 한정자는 별첨 I, 섹션 6, 표 5에서 확인 가능하다. 언어의존적 자유텍스트 값 형식을 가진 아미노산 서열에 대한 한정자는 별첨 I, 섹션 8, 표 6에서 확인 가능하다.

(a) 언어의존적 자유텍스트는 영어로 `INSDQualifier_value` 요소 또는 영어가 아닌 `NonEnglishQualifier_value` 요소 또는 두 요소 모두로 제시되어야 한다. 유기체 이름이 라틴어 속(`genus`)과 종(`species`) 이름인 경우라면 번역은 필요하지 않다. 국제적으로 사용되는, 영어가 아닌 단어에서 유래한 기술적 용어와 적절한 이름(예: 'in vitro', 'in vivo')은 `INSDQualifier_value` 요소의 값을 목적으로 영어로 간주된다.

(b) 서열목록에 `NonEnglishQualifier_value` 요소가 있는 경우, 루트 요소의 `nonEnglishFreeTextLanguageCode` 속성에 적절한 언어 코드(9절 ISO 639-1:2002를 참조)를 표시하여야 한다(43절 참조). 단일 서열목록의 모든 `NonEnglishQualifier_value` 요소들은 `nonEnglishFreeTextLanguageCode` 속성에 표시된 언어로 값을 가져야 한다. `NonEnglishQualifier_value` 요소는 언어의존적 자유텍스트 값 형식을 가진 한정자들에 대해서만 허용된다.

(c) 단일 한정자에 대해 `NonEnglishQualifier_value` 및 `INSDQualifier_value` 모두가 존재하는 경우, 두 요소에 포함된 정보는 동일하여야 한다. 다음 조건 중 하나는 참이어야 한다:

`NonEnglishQualifier_value`은 `INSDQualifier_value`의 값에 대한 번역을 포함; 또는
`INSDQualifier_value`은 `NonEnglishQualifier_value`의 값에 대한 번역을 포함; 또는
두 요소 모두 해당 `originalFreeTextLanguageCode` 속성에 지정된 언어의 한정자 값에 대한 번역을 포함(43절 참조).

(d) 언어의존적 한정자의 경우 `INSDQualifier` 요소는 선택적 속성 `id`를 포함할 수 있다. 해당 속성의 값은 "q" 형식 다음에 "q23"과 같은 양의 정수여야 하며, 하나의 `INSDQualifier` 요소에 고유(unique)하여야 한다. 즉, 속성 값은 서열목록 파일에 한 번만 사용하여야 한다.

88. 다음 예제들은 87절에서 논의된 것과 같은 언어의존적 자유텍스트의 예제이다.

실시예 1: `INSDQualifier_value` 요소에서 언어의존적 자유텍스트:

```
<INSDFeature>
  <INSDFeature_key>regulatory</INSDFeature_key>
  <INSDFeature_location>1..60</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q1">
      <INSDQualifier_name>function</INSDQualifier_name>
      <INSDQualifier_value>binds to regulatory protein Est3</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

실시예 2: `INSDQualifier_value` 요소와 `NonEnglishQualifier_value` 요소에서 언어의존적 자유텍스트:

```
<INSDFeature>
  <INSDFeature_key>ACT_SITE</INSDFeature_key>
  <INSDFeature_location>51..64</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier id="q45">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>cleaves carbohydrate chain</INSDQualifier_value>
      <NonEnglishQualifier_value>clive la chaîne glucidique
    </NonEnglishQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

실시예 3: `NonEnglishQualifier_value` 요소에서 언어의존적 자유텍스트:

```
<INSDFeature>
  <INSDFeature_key>ACT_SITE</INSDFeature_key>
  <INSDFeature_location>51..64</INSDFeature_location>
  <INSDFeature_qual>
```

```
<INSDQualifier id="q1034">
  <INSDQualifier_name>note</INSDQualifier_name>
  <NonEnglishQualifier_value>clive la chaîne glucidique
</NonEnglishQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

Coding sequences

89. The “CDS” feature key may be used to identify coding sequences, i.e., sequences of nucleotides which correspond to the sequence of amino acids in a protein and the stop codon. The location of the “CDS” feature in the mandatory element `INSDFeature_location` must include the stop codon.

90. The “`transl_table`” and “translation” qualifiers may be used with the “CDS” feature key (see Annex I). Where the “`transl_table`” qualifier is not used, the use of the Standard Code Table (see Annex I, Section 9, Table 7) is assumed.

91. The “`transl_except`” qualifier must be used with the “CDS” feature key and the “translation” qualifier to identify a codon that encodes either pyrrolysine or selenocysteine.

92. An amino acid sequence encoded by the coding sequence and disclosed in a “translation” qualifier that is encompassed by paragraph 7 must be included in the sequence listing and assigned its own sequence identification number. The sequence identification number assigned to the amino acid sequence must be provided as the value in the qualifier “`protein_id`” with the “CDS” feature key. The “organism” qualifier of the “source” feature key for the amino acid sequence must be identical to that of its coding sequence. For example:

```
<INSDFeature>
  <INSDFeature_key>CDS</INSDFeature_key>
  <INSDFeature_location>1..507</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>transl_table</INSDQualifier_name>
      <INSDQualifier_value>11</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>translation</INSDQualifier_name>
      <INSDQualifier_value>MLVHLERTT IMFDFSSL INLPLIWGLLIAI AVLLEYI LMDGFDLGIGILL
      PFA PSDKCRDHMSIA PFWDGNETWLVLGGGGLFAAF PLAYS ILMPAFYI PII IMLLGLIVRGVSFEFR
      FKAEGKYRRLWDYA FHFGLGA AFCQGMILGAF IHGVEVNGRNFSGQLM
    </INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>protein_id</INSDQualifier_name>
      <INSDQualifier_value>89</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Variants

93. A primary sequence and any variant of that sequence, each disclosed by enumeration of their residues and encompassed by paragraph 7, must each be included in the sequence listing and assigned their own sequence identification number.

94. Any variant sequence, disclosed as a single sequence with enumerated alternative residues at one or more positions, must be included in the sequence listing and should be represented by a single sequence, wherein the enumerated alternative residues are represented by the most restrictive ambiguity symbol (see paragraphs 15 and 27).

95. Any variant sequence, disclosed only by reference to deletion(s), insertion(s), or substitution(s) in a primary sequence in the sequence listing, should be included in the sequence listing. Where included in the sequence listing, such a variant sequence:

(a) may be represented by annotation of the primary sequence, where it contains variation(s) at a single location or multiple distinct locations and the occurrence of those variations are independent;

(b) should be represented as a separate sequence and assigned its own sequence identification number, where it contains variations at multiple distinct locations and the occurrence of those variations are interdependent; and

```
<INSDQualifier id="q1034">
  <INSDQualifier_name>note</INSDQualifier_name>
  <NonEnglishQualifier_value>clive la chaîne glucidique
</NonEnglishQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

코딩 서열

89. "CDS" 특징 기호는 코딩 서열, 즉 단백질의 아미노산 서열 및 정지 코돈에 대응하는 핵산염기의 서열을 확인하는 데에 사용될 수 있다. 필수적 요소 INSDFeature_location에서 "CDS" 특징의 위치에는 정지 코돈이 포함되어야 한다.

90. "transl_table" 및 "translation" 한정자는 "CDS" 특징 기호와 함께 사용할 수 있다(별첨 I 참조). "transl_table" 한정자가 사용되지 않는 경우, 표준 코드 표(별첨 I, 섹션 9, 표 7 참조)를 사용하는 것으로 가정한다.

91. "transl_except" 한정자는 "CDS" 특징 기호 및 "translation" 한정자와 함께 피롤리신(pyrrrolysine) 또는 셀레노시스테인(selenocysteine)을 인코딩하는 코돈을 식별하기 위해 사용되어야 한다.

92. 코딩 서열에 의해 암호화되고(encoded) "translation" 한정자에 개시된 7절에 포함되는 아미노산 서열은 서열목록에 포함되고 독립된 서열식별번호를 부여하여야 한다. 아미노산 서열에 부여된 서열식별번호는 "CDS" 특징 기호를 사용하여 한정자 "protein_id"의 값으로 제공되어야 한다. 아미노산 서열에 대한 "source" 특징 기호의 "organism" 한정자는 코딩 서열의 것과 동일하여야 한다. 예를 들면 다음과 같다:

```
<INSDFeature>
  <INSDFeature_key>CDS</INSDFeature_key>
  <INSDFeature_location>1..507</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>transl_table</INSDQualifier_name>
      <INSDQualifier_value>11</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>translation</INSDQualifier_name>
      <INSDQualifier_value>MLVHLERTTIMFDFSSLINLPLIWGLLTAIAVLLYIILMDGFDLGI
      GILLPFAPSDKCRDHMISSIAPFWDGNETWLVLGGGGLFAAFPLAYSILMPAFYIPIIIMLLGLIV
      RGVSEFRFKAEKGYRRLWDYAFHFGLGA AFCQGMILGAFIHGVEVNGRNFSGGQLM
      </INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>protein_id</INSDQualifier_name>
      <INSDQualifier_value>89</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

변이

93. 잔기 나열에 의해 표시되고 7절에 포함되는 1차 서열 및 이 서열의 임의의 변이체는 각각 서열목록에 포함되고, 각각의 고유 서열식별번호를 부여하여야 한다.

94. 하나 이상의 위치에 나열된 대체 잔기를 갖는 단일 서열로서 공개된 임의의 변이 서열은 서열목록에 포함되어야 하고 단일 서열로 표현되어야 하며, 여기서 나열된 대체 잔기는 가장 제한적인 다중 기호로 표현되어야 한다(15절 및 27절 참조).

95. 서열목록의 1차 서열에 대한 삭제, 삽입 또는 치환의 언급을 통해서만 개시된 임의의 변이 서열은 서열목록에 포함되어야 한다. 서열목록에 포함되는 경우에, 이러한 변이 서열은:

(a) 1차 서열의 주석으로 표현될 수 있으며, 여기서 단일 위치 또는 다수의 별개의 위치에서 변이를 포함하고 이러한 변이의 발생은 독립적이다;

(b) 별도의 서열로 표현되어야 하고, 고유한 서열식별번호가 부여되어야 하며, 여기서 다수의 다른 위치에서 변이를 포함하고 이러한 변이의 발생은 상호 의존적이다; 및

(c) must be represented as a separate sequence and assigned its own sequence identification number, where it contains an inserted or substituted sequence that contains in excess of 1000 residues (see paragraph 86).

96. The table below indicates the proper use of feature keys and qualifiers for nucleic acid and amino acid sequence variants:

Type of sequence	Feature Key	Qualifier	Use
Nucleic acid	variation	replace or note	Naturally occurring mutations and polymorphisms, e.g., alleles, RFLPs.
Nucleic acid	misc_difference	replace or note	Variability introduced artificially, e.g., by genetic manipulation or by chemical synthesis.
Amino acid	VAR_SEQ	note	Variant produced by alternative splicing, alternative promoter usage, alternative initiation and ribosomal frameshifting.
Amino acid	VARIANT	note	Any type of variant for which VAR_SEQ is not applicable.

97. Annotation of a sequence for a specific variant must include a feature key and qualifier, as indicated in the table above, and the feature location. The value for the "replace" qualifier must be only a single alternative nucleotide or nucleotide sequence using only the symbols in set forth Section 1, Table 1, or empty. A listing of alternative residues may be provided as the value in the "note" qualifier. In particular, a listing of alternative amino acids must be provided as the value in the "note" qualifier where "X" is used in a sequence, and represents a value other than "any one of 'A', 'R', 'N', 'D', 'C', 'Q', 'E', 'G', 'H', 'I', 'L', 'K', 'M', 'F', 'P', 'O', 'S', 'U', 'T', 'W', 'Y', or 'V'" (see paragraph 27). A deletion must be represented by an empty qualifier value for the "replace" qualifier or by an indication in the "note" qualifier that the residue may be deleted. An inserted or substituted residue(s) must be provided in the "replace" or "note" qualifier. The value format for the "replace" and "note" qualifiers is free text and must not exceed 1000 characters, as provided in paragraph 86. See paragraph 100 for sequences encompassed by paragraph 7 that are provided as an insertion or a substitution in a qualifier value.

98. The symbols set forth in Annex I (see Sections 1 to 4, Tables 1 to 4, respectively) should be used to represent variant residues where appropriate. For the "note" qualifier, where the variant residue is a modified residue not set forth in Tables 2 or 4 of Annex I, the complete unabbreviated name of the modified residue must be provided as the qualifier value. Modified residues must be further described in the feature table as provided in paragraph 17 or 30.

99. The following examples illustrate the representation of variants as per paragraphs 95 to 98 above:

Example 1: Feature key "misc_difference" for enumerated alternative nucleotides. The "n" at position 53 of the sequence can be one of five alternative nucleotides.

```
<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>53</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>w, cmnm5s2u, mam5u, mcm5s2u, or
p</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>53</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
```

(c) 별도의 서열로 표현되어야 하고, 고유한 서열식별번호가 부여되어야 하며, 여기에는 1000개 초과와 잔기를 포함하는 삽입 또는 치환된 서열이 포함된다(86절 참조).

96. 아래 표를 통해 핵산염기 및 아미노산 서열 변이에 대한 특징 기호와 한정자의 적절한 사용을 나타낸다.

서열 타입	특징 기호	한정자	용도
핵산	variation	replace 또는 note	자연 발생 돌연변이 및 다형성, 예: allele, RFLP.
핵산	misc_difference	replace 또는 note	유전자 조작 또는 화학적 합성 등에 의해 인위적으로 도입된 변이.
아미노산	VAR_SEQ	note	선택적 스펙트라이싱, 선택적 프로모터(promoter) 사용, 선택적 개시 및 리보솜 프레임시프팅에 의해 생성된 변이.
아미노산	VARIANT	note	VAR_SEQ가 적용되지 않는 타입의 변이.

97. 특정 변이에 대한 서열의 주석에는 위 표에 표시된 특징 기호와 한정자 및 특징 위치가 포함되어야 한다. "replace" 한정자에 대한 값은 색선 1, 표 1의 기호만 사용하거나 비어있는 단일 대체 핵산 또는 핵산염기 서열이어야 한다. 선택적인 잔기의 목록은 "note" 한정자 값으로 제공될 수 있다. 특히, 대체 아미노산의 목록은 "note" 한정자 값으로 제공되어야 한다. 여기서 "X"는 서열 내에 사용되며, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", or "V" 중 어떤 아미노산(27절 참조) 이외의 값을 표현한다. 삭제는 "replace" 한정자에 대한 빈 한정자 값 또는 "note" 한정자에 잔기가 삭제될 수 있다는 표시로 표현되어야 한다. 삽입되거나 치환된 잔기는 "replace", 또는 "note" 한정자로 제공되어야 한다. "replace"와 "note" 한정자에 대한 값 형식은 자유텍스트며, 86절에 설명된 바와 같이 1000자를 초과할 수 없다. 한정자 값에서 삽입 또는 대체로 제공된 7절에 포함된 서열에 대해서는 100절을 참조하면 된다.

98. 별첨 10에 설명된 기호(각각 색선 1~4, 표 1~4 참조)는 적절한 경우에 변이 잔기를 표현하기 위해 사용하여야 한다. 변이 잔기가 별첨 1의 표 2 또는 4에 제시되지 않은 변형된 잔기인 것에 대한 "note" 한정자의 경우, 변형된 잔기의 완전한 축약되지 않은 명칭이 한정자 값으로 제공되어야 한다. 변형된 잔기는 17절 또는 30절에 제공된 특징 표에 추가로 기술되어야 한다.

99. 다음의 실시예들은 상기 95절~98절에 따른 변형의 표현을 예시한다:

실시예 1: 나열된 선택적 변이 핵산염기에 대한 특징 기호 "misc_difference".
서열의 53번 위치의 "n"은 5개의 선택적인 핵산염기 중 하나일 수 있다.

```
<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>53</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>w, cmnm5s2u, mam5u, mcm5s2u, or p</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>53</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>mod_base</INSDQualifier_name>
      <INSDQualifier_value>OTHER</INSDQualifier_value>
    </INSDQualifier>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
```

```
        <INSDQualifier_value>cmnm5s2u, mam5u, mcm5s2u, or p</INSDQualifier_value>
    </INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
```

Example 2: Feature key “misc_difference” for a deletion in a nucleotide sequence.
The nucleotide at position 413 of the sequence is deleted.

```
<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>413</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value></INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Example 3: Feature key “misc_difference” for an insertion in a nucleotide sequence.
The sequence “atgccaaatat” is inserted between positions 100 and 101 of the primary sequence.

```
<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>100^101</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value>atgccaaatat</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Example 4: Feature key “variation” for a substitution in a nucleotide sequence.
A cytosine replaces the nucleotide given in position 413 of the sequence.

```
<INSDFeature>
  <INSDFeature_key>variation</INSDFeature_key>
  <INSDFeature_location>413</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value>c</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
```

Example 5: Feature key “VARIANT” for a substitution in an amino acid sequence.
The amino acid given in position 100 of the sequence can be replaced by I, A, F, Y, aIle, MeIle, or Nle.

```
<INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>I, A, F, Y, aIle, MeIle, or Nle
    </INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>MOD_RES</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
```

```

        <INSDQualifier_value>cmm5s2u, mam5u, mcm5s2u, or p</INSDQualifier_value>
    </INSDQualifier>
</INSDFeature_qual>
</INSDFeature>

```

실시예 2: 핵산염기 서열에서 삭제에 대한 특징 기호 "misc_difference".

서열의 413 위치 핵산염기가 삭제된다.

```

<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>413</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value></INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

실시예 3: 핵산염기 서열에서의 삽입을 위한 특징 기호 "misc_difference".

"atgccaaatat" 서열은 1차 서열의 위치 100과 101 사이에 삽입된다.

```

<INSDFeature>
  <INSDFeature_key>misc_difference</INSDFeature_key>
  <INSDFeature_location>100^101</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value>atgccaaatat</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

실시예 4: 핵산염기 서열에서의 치환에 대한 특징 기호 "변이". 시토신은 서열의 413번째 위치에서 주어진 핵산염기를 대체한다.

```

<INSDFeature>
  <INSDFeature_key>variation</INSDFeature_key>
  <INSDFeature_location>413</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>replace</INSDQualifier_name>
      <INSDQualifier_value>c</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

실시예 5: 아미노산 서열에서의 치환을 위한 특징 기호 "VARIANT".

서열의 100번 위치에 제공된 아미노산은 I, A, F, Y, alele, MeIle 또는 Nle로 대체될 수 있다.

```

<INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>I, A, F, Y, ale, MeIle or Nle
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key> MOD_RES </INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>

```

```

        <INSDQualifier_value>aIle, MeIle, or Nle</INSDQualifier_value>
    </INSDQualifier>
</INSDFeature_qual>
</INSDFeature>

```

Example 6: Feature key "VARIANT" for a substitution in an amino acid sequence.
The amino acid given in position 100 of the sequence can be replaced by any amino acid except for Lys, Arg or His.

```

<INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>not K, R, or H</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

100. A sequence encompassed by paragraph 7 that is provided as an insertion or a substitution in a qualifier value for a primary sequence annotation must also be included in the sequence listing and assigned its own sequence identification number.

[Annex I follow s]

```

        <INSDQualifier_value>ale, meIle or Nle</INSDQualifier_value>
      </INSDQualifier>
    </INSDFeature_qual>
  </INSDFeature>

```

실시예 6: 아미노산 서열에서의 치환을 위한 특징 기호 "VARIANT".

서열의 100번 위치에 제공된 아미노산은 Lys, Arg 또는 His를 제외한 임의의 아미노산으로 대체될 수 있다.

```

<INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>100</INSDFeature_location>
  <INSDFeature_qual>
    <INSDQualifier>
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>not K, R or H</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

100. 1차 서열 주석에 대한 한정자 값의 삽입 또는 대체로 제공되는 7절에 포함된 서열은 서열목록에 포함되고 고유한 서열식별번호가 부여되어야 한다.

[별첨 1로 이어짐]

ANNEX I

CONTROLLED VOCABULARY

Version 1.5

*Revision approved by the Committee on WIPO Standards (CWS)
at its ninth session on November 5, 2021*

TABLE OF CONTENTS

SECTION : LIST OF NUCLEOTIDES	3.26.i.2
SECTION : LIST OF MODIFIED NUCLEOTIDES.2	3.26.i.2
SECTION : LIST OF AMINO ACIDS	3.26.i.4
SECTION : LIST OF MODIFIED AMINO ACIDS	3.26.i.5
SECTION : FEATURE KEYS FOR NUCLEOTIDE SEQUENCES	3.26.i.6
SECTION : QUALIFIERS FOR NUCLEOTIDE SEQUENCES	3.26.i.24
SECTION : FEATURE KEYS FOR AMINO ACID SEQUENCES	3.26.i.46
SECTION : QUALIFIERS FOR AMINO ACID SEQUENCES	3.26.i.53
SECTION : GENETIC CODE TABLES	3.26.i.54

별첨 I

통제 어휘

버전 1.5

2021년 11월 5일 WIPO 표준위원회(CWS) 제9차 회의에서 승인된 개정판

목차

섹션 1: 핵산염기 목록	3.26.i.2
섹션 2: 변형 핵산염기 목록	3.26.i.2
섹션 3: 아미노산 목록	3.26.i.4
섹션 4: 변형 아미노산 목록	3.26.i.5
섹션 5: 핵산염기 서열에 대한 특징 기호	3.26.i.6
섹션 6: 핵산염기 서열에 대한 한정자	3.26.i.24
섹션 7: 아미노산 서열에 대한 특징 기호	3.26.i.46
섹션 8: 아미노산 서열에 대한 한정자	3.26.i.53
섹션 9: 유전코드 표	3.26.i.54

SECTION 1: LIST OF NUCLEOTIDES

The nucleotide base symbols to be used in sequence listings are presented in Table 1. The symbol “t” will be construed as thymine in DNA and uracil in RNA when it is used with no further description. Where an ambiguity symbol (representing two or more bases in the alternative) is appropriate, the most restrictive symbol should be used. For example, if a base in a given position could be “a or g,” then “r” should be used, rather than “n”. The symbol “n” will be construed as “a or c or g or t/u” when it is used with no further description.

Table 1: List of nucleotides symbols

Symbol	Definition
a	adenine
c	cytosine
g	guanine
t	thymine in DNA/uracil in RNA (t/u)
m	a or c
r	a or g
w	a or t/u
s	c or g
y	c or t/u
k	g or t/u
v	a or c or g; not t/u
h	a or c or t/u; not g
d	a or g or t/u; not c
b	c or g or t/u; not a
n	a or c or g or t/u; “unknown” or “other”

SECTION 2: LIST OF MODIFIED NUCLEOTIDES

The abbreviations listed in Table 2 are the only permitted values for the mod_base qualifier. Where a specific modified nucleotide is not present in the table below, then the abbreviation “OTHER” must be used as its value. If the abbreviation is “OTHER”, then the complete unabbreviated name of the modified base must be provided in a note qualifier. The abbreviations provided in Table 2 must not be used in the sequence itself.

Table 2: List of modified nucleotides

Abbreviation	Definition
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
dhu	dihydrouridine
fm	2'-O-methylpseudouridine
gal q	beta-D-galactosylqueuosine
gm	2'-O-methylguanosine
i	inosine
i6a	N6-isopentenyladenosine
m1a	1-methyladenosine
m1f	1-methylpseudouridine
m1g	1-methylguanosine
m1i	1-methylinosine
m22g	2,2-dimethylguanosine
m2a	2-methyladenosine
m2g	2-methylguanosine
m3c	3-methylcytidine

섹션 1: 핵산염기 목록

서열목록에 사용될 핵산염기 기호는 표 1에 제시되어 있다. "t" 기호는 추가 설명 없이 사용될 때 DNA에서 티민(thymine) 및 RNA에서 우라실(uracil)로 해석될 것이다. 다중핵산 기호(선택에서 둘 이상의 염기를 표현)가 적절한 경우에 가장 제한적인 기호를 사용하여야 한다. 예를 들면, 주어진 위치의 염기가 "a 또는 g"일 수 있으면 "n"보다 "r"을 사용하여야 한다. 기호 "n"은 추가 설명 없이 사용되는 경우 "a 또는 c 또는 g 또는 t/u"로 해석된다.

표 1: 핵산염기 기호 목록

기호	정의
a	아데닌(adenine)
c	시토신(cytosine)
g	구아닌(guanine)
t	DNA의 티민(thymine)/RNA의 우라실(uracil)(t/u)
m	a 또는 c
r	a 또는 g
w	a 또는 t/u
s	c 또는 g
y	c 또는 t/u
k	g 또는 t/ u
v	a 또는 c 또는 g; t/u가 아님
h	a 또는 c 또는 t/u; g가 아님
d	a 또는 g 또는 t/u; c가 아님
b	c 또는 g 또는 t/u; a가 아님
n	a 또는 c 또는 g 또는 t/u; "불명" 또는 "기타"

섹션 2: 변형 핵산염기 목록

표 2에 나열된 약어는 mod_base 한정자에 허용되는 유일한 값이다. 아래 표에 특정 변형 핵산염기가 없는 경우, 약어 "OTHER"를 그 값으로 사용해야만 한다. 약어가 "OTHER"인 경우, 변형된 염기의 단축되지 않고 완전한 이름을 note 한정자에 제공하여야 한다. 표 2에 제공된 약어는 핵산염기 서열 자체에 사용해서는 아니된다.

표 2: 변형 핵산염기 목록

약어	정의
ac4c	4-아세틸시티딘(4-acetylcytidine)
chm5u	5-(카르복시히드록실메틸)우리딘(5-(carboxyhydroxylmethyl)uridine)
cm	2'-O-메틸시티딘(2'-O-methylcytidine)
cmnm5s2u	5-카르복시메틸아미노메틸-2-티오우리딘(5-carboxymethylaminomethyl-2-thiouridine)
cmnm5u	5-카르복시메틸아미노메틸우리딘(5-carboxymethylaminomethyluridine)
dhu	디하이드로우리딘(dihydrouridine)
fm	2'-O-메틸슈도우리딘(2'-O-methylpseudouridine)
gal q	베타-D-갈락토실큐에오신(beta-D-galactosylqueosine)
gm	2'-O-메틸구아노신(2'-O-methylguanosine)
i	이노신(inosine)
i6a	N6-이소펜텐일아데노신(N6-isopentenyladenosine)
m1a	1-메틸아데노신(1-methyladenosine)
m1f	1-메틸슈도우리딘(1-methylpseudouridine)
m1g	1-메틸구아노신(1-methylguanosine)
m1i	1-메틸이노신(1-methylinosine)
m22g	2,2-디메틸구아노신(2,2-dimethylguanosine)
m2a	2-메틸아데노신(2-methyladenosine)
m2g	2-메틸구아노신(2-methylguanosine)
m3c	3-메틸시티딘(3-methylcytidine)

Abbreviation	Definition
m4c	N4-methylcytosine
m5c	5-methylcytidine
m6a	N6-methyladenosine
m7g	7-methylguanosine
mam5u	5-methylaminomethyluridine
mam5s2u	5-methylaminomethyl-2-thiouridine
man q	beta-D-mannosylqueuosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine
mv	uridine-5-oxoacetic acid-methylester
o5u	uridine-5-oxoacetic acid (v)
osyw	w ybutoxosine
p	pseudouridine
q	queuosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
m5u	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	w ybutosine
x	3-(3-amino-3-carboxypropyl)uridine, (acp3)u
OTHER	(requires note qualifier)

약어	정의
m4c	N4-메틸시토신(N4-methylcytosine)
m5c	5-메틸시티딘(5-methylcytidine)
m6a	N6-메틸아데노신(N6-methyladenosine)
m7g	7-메틸구아노신(7-methylguanosine)
mam5u	5-메틸아미노메틸우리딘(5-methylaminomethyluridine)
mam5s2u	5-메틸아미노메틸-2-티오우리딘(5-methylaminomethyl-2-thiouridine)
man q	베타-D-만노실큐에오신(beta-D-mannosylqueuosine)
mcm5s2u	5-메톡시카르보닐메틸-2-티오우리딘(5-methoxycarbonylmethyl-2-thiouridine)
mcm5u	5-메톡시카르보닐메틸우리딘(5-methoxycarbonylmethyluridine)
mo5u	5-메톡시우리딘(5-methoxyuridine)
ms2i6a	2-메틸티오-N6-이소펜텐테닐아데노신(2-methylthio-N6-isopentenyladenosine)
ms2t6a	N-((9-베타-D-리보푸라노실-2-메틸티오퀴린-6-일)카바모일)트레오닌 (N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine)
mt6a	N-((9-베타-D-리보푸라노실푸린-6-일)N-메틸-카르바모일)트레오닌 (N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine)
mv	우리딘-5-옥소아세트산-메틸에스테르(uridine-5-oxoacetic acid-methylester)
o5u	우리딘-5-옥시아세트산(v)(uridine-5-oxycetic acid(v))
osyw	와이부톡소신(wybutoxosine)
p	슈도우리딘(pseudouridine)
q	큐에오신(queosine)
s2c	2-티오시티딘(2-thiocytidine)
s2t	5-메틸-2-티오우리딘(5-methyl-2-thiouridine)
s2u	2-티오우리딘(2-thiouridine)
s4u	4-티오우리딘(4-thiouridine)
m5u	5-메틸우리딘(5-methyluridine)
t6a	N-((9-베타-D-리보푸라노실푸린-6-일)카바모일)트레오닌 (N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine)
tm	2'-O-메틸-5-메틸우리딘(2'-O-methyl-5-methyluridine)
um	2'-O-메틸우리딘(2'-O-methyluridine)
yw	와이부토신(wybutosine)
x	3-(3-아미노-3-카복시프로필)우리딘(3-(3-amino-3-carboxypropyl)uridine), (acp3)u
OTHER	(note 한정자 필요)

SECTION 3: LIST OF AMINO ACIDS

The amino acid symbols to be used in sequence are presented in Table 3. Where an ambiguity symbol (representing two or more amino acids in the alternative) is appropriate, the most restrictive symbol should be used. For example, if an amino acid in a given position could be aspartic acid or asparagine, the symbol "B" should be used, rather than "X". The symbol "X" will be construed as any one of "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", or "V", when it is used with no further description.

Table 3: List of amino acids symbols

Symbol	Definition
A	Alanine
R	Arginine
N	Asparagine
D	Aspartic acid (Aspartate)
C	Cysteine
Q	Glutamine
E	Glutamic acid (Glutamate)
G	Glycine
H	Histidine
I	Isoleucine
L	Leucine
K	Lysine
M	Methionine
F	Phenylalanine
P	Proline
O	Pyrrolysine
S	Serine
U	Selenocysteine
T	Threonine
W	Tryptophan
Y	Tyrosine
V	Valine
B	Aspartic acid or Asparagine
Z	Glutamine or Glutamic acid
J	Leucine or Isoleucine
X	A or R or N or D or C or Q or E or G or H or I or L or K or M or F or P or O or S or U or T or W or Y or V; "unknown" or "other"

섹션 3: 아미노산 목록

서열에 사용될 아미노산 기호는 표 3에 제시되어 있다. 다중아미노산(ambiguity) 기호(선택에서 둘 이상의 아미노산을 표현)가 적절한 경우에, 가장 제한적인 기호를 사용하여야 한다. 예를 들면, 주어진 위치의 아미노산이 아스파르트산(aspartic acid) 또는 아스파라긴(asparagine)일 수 있는 경우에, "X" 보다 "B" 기호를 사용하여야 한다. "X" 기호는 추가 설명 없이 사용되는 경우 "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", 또는 "V" 중 하나로 해석된다.

표 3: 아미노산 기호 목록

기호	정의
A	알라닌(Alanine)
R	아르기닌(Arginine)
N	아스파라긴(Asparagine)
D	아스파르트산(Aspartic acid (Aspartate))
C	시스테인(Cysteine)
Q	글루타민(Glutamine)
E	글루탐산(Glutamic acid (Glutamate))
G	글리신(Glycine)
H	히스티딘(Histidine)
I	이소류신(Isoleucine)
L	류신(Leucine)
K	라이신(Lysine)
M	메티오닌(Methionine)
F	페닐알라닌(Phenylalanine)
P	프롤린(Proline)
O	피롤리신(Pyrolysine)
S	세린(Serine)
U	셀레노시스테인(Selenocysteine)
T	트레오닌(Threonine)
W	ทริป토판(Tryptophan)
Y	티로신(Tyrosine)
V	발린(Valine)
B	아스파르트산 또는 아스파라긴(Aspartic acid or Asparagine)
Z	글루타민 또는 글루탐산(Glutamine or Glutamic acid)
J	류신 또는 이소류신(Leucine or Isoleucine)
X	A 또는 R 또는 N 또는 D 또는 C 또는 Q 또는 E 또는 G 또는 H 또는 I 또는 L 또는 K 또는 M 또는 F 또는 P 또는 O 또는 S 또는 U 또는 T 또는 W 또는 Y 또는 V; "불명" 또는 "기타"

SECTION 4: LIST OF MODIFIED AMINO ACIDS

Table 4 lists the only permitted abbreviations for a modified amino acid in the mandatory qualifier “note” for feature keys “MOD_RES” or “SITE”. The value for the qualifier “note” must be either an abbreviation from this table, where appropriate, or the complete, unabbreviated name of the modified amino acid. The abbreviations (or full names) provided in this table must not be used in the sequence itself.

Table 4: List of modified amino acids

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

섹션 4: 변형 아미노산 목록

표 4에는 특징 기호 "MOD_RES" 또는 "SITE"에 대한 필수 한정자 "note"의 변형 아미노산에 대해 유일하게 허용되는 약어가 나열되어 있다. 한정자 "note"의 값은 본 표의 해당하는 약어 또는 변형 아미노산의 단축되지 않고 완전한 이름을 사용하여야 한다. 이 표에 제공된 약어(또는 전체 명칭)를 아미노산 서열 자체에 사용해서는 아니된다.

표 4: 변형 아미노산 목록

약어	변형 아미노산
Aad	2-아미노아디프산(2-Aminoadipic acid)
bAad	3-아미노아디프산(3-Aminoadipic acid)
bAla	베타-알라닌, 베타-아미노프로피온산 (beta-Alanine, beta-Aminopropionic acid)
Abu	2-아미노부티르산(2-Aminobutyric acid)
4Abu	4-아미노부티르산, 피페리딘산(4-Aminobutyric acid, piperidinic acid)
Acp	6-아미노카프로산(6-Aminocaproic acid)
Ahe	2-아미노헵타노산(2-Aminoheptanoic acid)
Aib	2-아미노이소부티르산(2-Aminoisobutyric acid)
bAib	3-아미노이소부티르산(3-Aminoisobutyric acid)
Apm	2-아미노피멜산(2-Aminopimelic acid)
Dbu	2,4-디아미노부티르산(2,4-Diaminobutyric acid)
Des	데스모신(Desmosine)
Dpm	2,2'-디아미노피멜산(2,2'-Diaminopimelic acid)
Dpr	2,3-디아미노프로피온산(2,3-Diaminopropionic acid)
EtGly	N-에틸글리신(N-Ethylglycine)
EtAsn	N-에틸라스파라긴(N-Ethylasparagine)
Hyl	히드록시리신(Hydroxylysine)
aHyl	알로히드록시리신(allo-Hydroxylysine)
3Hyp	3-히드록시프롤린(3-Hydroxyproline)
4Hyp	4-히드록시프롤린(4-Hydroxyproline)
Ide	이소데모신(Isodesmosine)
aIle	알로이소류신(allo-Isoleucine)
MeGly	N-메틸글리신, 사르코신(N-Methylglycine, sarcosine)
Melle	N- 메틸이소류신(N-Methylisoleucine)
MeLys	6-N-메틸라이신(6-N-Methyllysine)
MeVal	N-메틸발린(N-Methylvaline)
Nva	노르발린(Norvaline)
Nle	노르류신(Norleucine)
Orn	오르니틴(Omithine)

SECTION 5: FEATURE KEYS FOR NUCLEOTIDE SEQUENCES

This section contains the list of allowed feature keys to be used for nucleotide sequences, and lists mandatory and optional qualifiers. The feature keys are listed in alphabetic order. The feature keys can be used for either DNA or RNA unless otherwise indicated under "Molecule scope". Certain Feature Keys may be appropriate for use with artificial sequences in addition to the specified "organism scope".

Feature key names must be used in the XML instance of the sequence listing exactly as they appear following "Feature key" in the descriptions below, except for the feature keys 3'UTR and 5'UTR. See "Comment" in the description for the 3'UTR and 5'UTR feature keys.

5.1.	Feature Key	C_region
	Definition	constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; includes one or more exons depending on the particular chain
	Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes
5.2.	Feature Key	CDS
	Definition	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature may include amino acid conceptual translation
	Optional qualifiers	allele circular_RNA codon_start EC_number exception function gene gene_synonym map note number operon product protein_id pseudo pseudogene ribosomal_slippage standard_name translation transl_except transl_table trans_splicing

섹션 5: 핵산염기 서열에 대한 특징 기호

이 섹션에는 핵산염기 서열에 사용할 수 있는 특징 기호 목록과 필수 및 선택 한정자가 나열되어 있다. 특징 기호는 알파벳 순서로 나열된다. 특징 기호는 "분자 범위(Molecule scope)"에서 달리 지시되지 않는 한, DNA 또는 RNA에 사용될 수 있다. 어떤 특징 기호는 지정된 "생물체 범위(organism scope)"에 뿐만 아니라 인공 서열에 사용하기에 적합할 수 있다.

특징 기호 명칭은 특징 기호 3'UTR 및 5'UTR을 제외하고, 아래 설명의 "특징 기호(Feature key)"에 나타나는 것처럼 정확하게 서열목록의 XML 인스턴스에서 사용되어야 한다. 3'UTR 및 5'UTR 특징 기호에 대한 설명에서 "참고(comment)"를 참조하십시오.

5.1.	특징 기호	C_region
	정의	면역글로불린(immunoglobulin) 경쇄(light chain) 및 중쇄(heavy chain) 및 T-세포 수용체의 $\alpha/\beta/\gamma$ 사슬의 불변 영역(constant region); 특정 사슬에 따라 하나 이상의 역손을 포함
	선택 한정자	allele gene gene_synonym map note product pseudo pseudogene standard_name
	생물체 적용범위	진핵생물
5.2.	특징 기호	CDS
	정의	코딩 서열: 단백질에서 아미노산의 서열에 대응하는 핵산염기의 서열(정지 코돈 위치도 포함); 아미노산 개념적인(conceptual) 번역(translation)을 포함할 수 있음
	선택 한정자	allele circular_RNA codon_start EC_number exception function gene gene_synonym map note number operon product protein_id pseudo pseudogene ribosomal_slippage standard_name translation transl_except transl_table trans_splicing

Comment	codon_start qualifier has valid value of 1 or 2 or 3, indicating the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature; transl_table defines the genetic code table used if other than the Standard or universal genetic code table; genetic code exceptions outside the range of the specified tables are reported in transl_except qualifier; only one of the qualifiers translation, pseugogene or pseudo are permitted with a CDS feature key; when the translation qualifier is used, the protein_id qualifier is mandatory if the translation product contains four or more specifically defined amino acids
<hr/>	
5.3. Feature Key	centromere
Definition	region of biological interest identified as a centromere and which has been experimentally characterized
Optional qualifiers	note standard_name
Comment	the centromere feature describes the interval of DNA that corresponds to a region where chromatids are held and a kinetochore is formed
<hr/>	
5.4. Feature Key	D-loop
Definition	displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA by a single stranded invader in the reaction catalyzed by RecA protein
Optional qualifiers	allele gene gene_synonym map note
Molecule scope	DNA
<hr/>	
5.5. Feature Key	D_segment
Definition	Diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain
Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
Organism scope	eukaryotes

참고 codon_start 한정자는 1 또는 2 또는 3의 유효값을 가지며, 이는 해당 특징의 첫 번째 염기에 대한 코딩 특징의 첫 번째 완전한 코돈이 발견될 수 있는 오프셋을 나타낸다; transl_table은 표준 또는 범용 유전코드 표가 아닌 경우에 사용되는 유전코드 표를 정의한다. 규정된 표의 범위를 벗어난 유전코드 예외는 transl_except 한정자에 보고된다; translation, pseudogene 또는 pseudo 한정자 중 하나만 CDS 특징 기호와 같이 허용된다. translation 한정자가 사용될 때, 번역 산물에 4개 이상의 특정 아미노산이 포함된 경우에 protein_id 한정자는 필수이다.

5.3.	특징 기호	centromere
	정의	동원체(動原體)로 식별되고 실험적으로 확인된 생물학적 관심 영역
	선택 한정자	note standard_name
	참고	centromere 특징은 염색질이 유지되고 키네토코어(kinetochore)가 형성되는 영역에 해당하는 DNA 구간을 설명한다.
5.4.	특징 기호	D-loop
	정의	치환(displacement) 루프; RNA의 짧은 부위가 한 가닥의 DNA와 쌍을 이루는 미토콘드리아 DNA 내의 영역으로, 원래 파트너 DNA 가닥을 대체(displacing)함; 또한 RecA 단백질에 의해 촉매되는 반응에서 단일 가닥 침입자에 의한 이중 DNA 한 가닥의 영역의 대체를 설명하기 위해 사용
	선택 한정자	allele gene gene_synonym map note
	분자 적용범위	DNA
5.5.	특징 기호	D-segments
	정의	면역글로불린(immunoglobulin) 중쇄(heavy chain) 및 T-세포 수용체 베타 사슬의 다양성 단편(diversity segment)
	선택 한정자	allele gene gene_synonym map note product pseudo pseudogene standard_name
	생물체 적용범위	진핵생물

5.6.	Feature Key	exon
	Definition	region of genome that codes for portion of spliced mRNA, rRNA and tRNA; may contain 5'UTR, all CDSs and 3' UTR
	Optional qualifiers	allele EC_number function gene gene_synonym map note number product pseudo pseudogene standard_name trans_splicing

5.6.	특징 기호	exon
	정의	스플라이싱된(spliced) mRNA, rRNA, 및 tRNA의 부분을 코딩하는 유전체 영역; 5'UTR, 모든 CDS 및 3'UTR을 포함할 수 있음.
	선택 한정자	allele EC_number function gene gene_synonym map note number product pseudo pseudogene standard_name trans_splicing

5.7.	Feature Key	gene
	Definition	region of biological interest identified as a gene and for which a name has been assigned
	Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene phenotype standard_name trans_splicing
	Comment	the gene feature describes the interval of DNA that corresponds to a genetic trait or phenotype; the feature is, by definition, not strictly bound to its positions at the ends; it is meant to represent a region where the gene is located.
5.8.	Feature Key	iDNA
	Definition	intervening DNA; DNA which is eliminated through any of several kinds of recombination
	Optional qualifiers	allele function gene gene_synonym map note number standard_name
	Molecule scope	DNA
	Comment	e.g., in the somatic processing of immunoglobulin genes.
5.9.	Feature Key	intron
	Definition	a segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it
	Optional qualifiers	allele function gene gene_synonym map note number pseudo pseudogene standard_name trans_splicing

5.7.	특징 기호	gene
	정의	유전자로 식별되고 명칭이 지정된 생물학적 관심 영역
	선택 한정자	allele function gene gene_synonym map note operon product pseudo pseudogene phenotype standard_name trans_splicing
	참고	gene 특징은 유전 형질 또는 표현형에 대응하는 DNA의 영역을 기술한다; 정의에 따라 본 특징은 유전자 말단의 위치를 명확하게 정하지 않는다; 이는 유전자가 위치한 영역을 표현하는 것을 의미한다.
5.8.	특징 기호	iDNA
	정의	개재(intervening) DNA; 여러 종류의 재조합을 통해 제거되는 DNA
	선택 한정자	allele function gene gene_synonym map note number standard_name
	분자 적용범위	DNA
	참고	예를 들면, 번역글로불린 유전자의 체세포 재조합 공정에서 보여지는 현상
5.9.	특징 기호	intron
	정의	전사는 되지만, 양쪽 exon 서열의 splicing에 의해 전사체에서 제거되는 DNA 단편
	선택 한정자	allele function gene gene_synonym map note number pseudo pseudogene standard_name trans_splicing

Ref.: Standards - ST.26

page: 3.26.i.10

5.10.	Feature Key	J_segment
	Definition	joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains
	Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes
5.11.	Feature Key	mat_peptide
	Definition	mature peptide or protein coding sequence; coding sequence for the mature or final peptide or protein product following post-translational modification; the location does not include the stop codon (unlike the corresponding CDS)
	Optional qualifiers	allele EC_number function gene gene_synonym map note product pseudo pseudogene standard_name
5.12.	Feature Key	misc_binding
	Definition	site in nucleic acid which covalently or non-covalently binds another moiety that cannot be described by any other binding key (primer_bind or protein_bind)
	Mandatory qualifiers	bound_moiety
	Optional qualifiers	allele function gene gene_synonym map note
	Comment	note that the regulatory feature key and regulatory_class qualifier with the value "ribosome_binding_site" must be used for describing ribosome binding sites

5.10.	특징 기호	J_segment
	정의	면역글로블린 경쇄, 중쇄 및 T-세포 수용체의 $\alpha/\beta/\gamma$ 사슬의 연결 분절(joining segment)
	선택 한정자	allele gene gene_synonym map note product pseudo pseudogene standard_name
	생물체 적용범위	진핵생물
5.11.	특징 기호	mat_peptide
	정의	성숙(mature) 펩티드/단백질을 코딩하는 서열; 번역 후 조절(post-translational modification)에 따른 성숙 또는 최종 펩티드/단백질 산물; (CDS와는 달리) 정지 코돈(stop codon)을 포함하지 않음
	선택 한정자	allele EC_number function gene gene_synonym map note product pseudo pseudogene standard_name
5.12.	특징 기호	misc_binding
	정의	다른 결합 기호(primer_bind 또는 protein_bind)에 의해 표현될 수 없는 또 다른 잔기에 공유 또는 비공유 결합하는 핵산 부위
	필수 한정자	bound_moiety
	선택 한정자	allele function gene gene_synonym map note
	참고	regulatory 특징 기호와 "ribosome_bind ing_site" 값을 가진 regulatory_class 한정자는 리보솜 결합 부위를 설명하기 위해 반드시 사용되어야 함을 주의할 것

5.13. Feature Key	misc_difference
Definition	featured sequence differs from the presented sequence at this location and cannot be described by any other Difference key (variation, or modified_base)
Optional qualifiers	allele clone compare gene gene_synonym map note phenotype replace standard_name
Comment	the misc_difference feature key must be used to describe variability introduced artificially, e.g., by genetic manipulation or by chemical synthesis; use the replace qualifier to annotate a deletion, insertion, or substitution. The variation feature key must be used to describe naturally occurring genetic variability.
5.14. Feature Key	misc_feature
Definition	region of biological interest which cannot be described by any other feature key; a new or rare feature
Optional qualifiers	allele function gene gene_synonym map note number phenotype product pseudo pseudogene standard_name
Comment	this key should not be used when the need is merely to mark a region in order to comment on it or to use it in another feature's location
5.15. Feature Key	misc_recomb
Definition	site of any generalized, site-specific or replicative recombination event where there is a breakage and reunion of duplex DNA that cannot be described by other recombination keys or qualifiers of source key (proviral)
Optional qualifiers	allele gene gene_synonym map note recombination_class standard_name
Molecule scope	DNA

5.13.	특징 기호	misc_difference
	정의	특정 서열이 해당 위치에 제시된 서열과 다르고, 다른 차이 기호(Difference key: variation 또는 modified_base)로는 표시할 수 없는 경우
	선택 한정자	allele clone compare gene gene_synonym map note phenotype replace standard_name
	참고	misc_difference 특징 기호는 인위적으로 생성된 다양성 설명에 사용되어야 한다(예: 유전자 조작 또는 화학적 합성에 의한 것); replace 한정자를 사용하여 삭제, 삽입, 또는 치환에 대한 주석을 달 수 있다. variation 특징 기호는 자연적으로 발생하는 유전적 다양성을 설명하는 데에 사용되어야 한다.

5.14.	특징 기호	misc_feature
	정의	다른 특징 기호로는 설명할 수 없는 생물학적 관심 영역; 새롭거나 희귀한 특징
	선택 한정자	allele function gene gene_synonym map note number phenotype product pseudo pseudogene standard_name
	참고	이 기호는 주석을 달거나 다른 특징의 위치에서 사용하기 위해 단순히 영역을 표시해야 하는 경우에는 사용하면 안 된다.

5.15.	특징 기호	misc_recomb
	정의	일반적, 부위-특이적 또는 복재적 재조합 발생 영역으로, duplex DNA의 절단 및 재결합이 일어나며, 다른 recombination 기호 또는 source기호의 한정자(proviral)로는 기술될 수 없는 영역
	선택 한정자	allele gene gene_synonym map note recombination_class standard_name
	분자 적용범위	DNA

5.16.	Feature Key	misc_RNA
	Definition	any transcript or RNA product that cannot be defined by other RNA keys (prim_transcript, precursor_RNA, mRNA, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, ncRNA, rRNA and tRNA)
	Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
5.17.	Feature Key	misc_structure
	Definition	any secondary or tertiary nucleotide structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)
	Optional qualifiers	allele function gene gene_synonym map note standard_name
5.18.	Feature Key	mobile_element
	Definition	region of genome containing mobile elements
	Mandatory qualifiers	mobile_element_type
	Optional qualifiers	allele function gene gene_synonym map note rpt_family rpt_type standard_name
5.19.	Feature Key	modified_base
	Definition	the indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)
	Mandatory qualifiers	mod_base
	Optional qualifiers	allele frequency gene gene_synonym map note

5.16.	특징 기호	misc_RNA
	정의	다른 RNA 기호(prim_transcript, precursor_RNA, mRNA, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, ncRNA, rRNA 및 tRNA)에 의해 정의될 수 없는 전사체 또는 RNA 산물(product)
	선택 한정자	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
5.17.	특징 기호	misc_structure
	정의	다른 Structure 기호(stem_loop 및 D-loop)로 기술될 수 없는 2차 또는 3차 핵산염기 구조 또는 형태(conformation)
	선택 한정자	allele function gene gene_synonym map note standard_name
5.18.	특징 기호	mobile_element
	정의	이동 요소(mobile elements)를 포함하는 유전체 영역
	필수 한정자	mobile_element_type
	선택 한정자	allele function gene gene_synonym map note rpt_family rpt_type standard_name
5.19.	특징 기호	modified_base
	정의	지시된 핵산염기는 변형 핵산염기며 지시된 분자(mod_base 한정자 값으로 제공됨)에 의해 치환되어야 한다.
	필수 한정자	mod_base
	선택 한정자	allele frequency gene gene_synonym map note

Comment	value for the mandatory mod_base qualifier is limited to the restricted vocabulary for modified base abbreviations in Section 2 of this Annex.
<hr/>	
5.20. Feature Key	mRNA
Definition	messenger RNA; includes 5' untranslated region (5'UTR), coding sequences (CDS, exon) and 3' untranslated region (3'UTR)
Optional qualifiers	allele circular_RNA function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
<hr/>	
5.21. Feature Key	ncRNA
Definition	a non-protein-coding gene, other than ribosomal RNA and transfer RNA, the functional molecule of which is the RNA transcript
Mandatory qualifiers	ncRNA_class
Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
Comment	the ncRNA feature must not be used for ribosomal and transfer RNA annotation, for which the rRNA and tRNA feature keys must be used, respectively
<hr/>	
5.22. Feature Key	N_region
Definition	extra nucleotides inserted between rearranged immunoglobulin segments
Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
Organism scope	eukaryotes

참고 필수 한정자 mod_base에 대한 값은 본 별첨의 섹션 2의 변형 핵산염기 약어에 대한 제한된 용어로 한정된다.

5.20.	특징 기호	mRNA
	정의	messenger RNA; 5'-말단 비번역 영역(5'UTR), 코딩 서열(CDS, 엑손) 및 3'-말단 비번역 영역(3'UTR) 포함
	선택 한정자	allele circular_RNA function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
5.21.	특징 기호	ncRNA
	정의	ribosomal RNA 및 messenger RNA 이외의 비-단백질-코딩 유전자, RNA 전사체인 기능성 분자
	필수 한정자	ncRNA_class
	선택 한정자	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
	참고	ncRNA 특징은 ribosome 및 messenger RNA 주석에 사용되어서는 안 되며, rRNA 및 tRNA 특징 기호가 각각 사용되어야 한다.
5.22.	특징 기호	N_region
	정의	재배열된 면역글로불린(immunoglobulin) 단편 사이에 삽입된 여분의 핵산염기
	선택 한정자	allele gene gene_synonym map note product pseudo pseudogene standard_name
	생물체 적용범위	진핵생물

5.23.	Feature Key	operon
	Definition	region containing polycistronic transcript including a cluster of genes that are under the control of the same regulatory sequences/promoter and in the same biological pathway
	Mandatory qualifiers	operon
	Optional qualifiers	allele function map note phenotype pseudo pseudogene standard_name
5.24.	Feature Key	oriT
	Definition	origin of transfer; region of a DNA molecule where transfer is initiated during the process of conjugation or mobilization
	Optional qualifiers	allele bound_moiety direction gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq standard_name
	Molecule Scope	DNA
	Comment	rep_origin must be used to describe origins of replication; direction qualifier has permitted values left, right, and both, however only left and right are valid when used in conjunction with the oriT feature; origins of transfer can be present in the chromosome; plasmids can contain multiple origins of transfer
5.25.	Feature Key	polyA_site
	Definition	site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation
	Optional qualifiers	allele gene gene_synonym map note
	Organism scope	eukaryotes and eukaryotic viruses

5.23.	특징 기호	operon
	정의	동일한 조절서열(regulatory sequence)/프로모터(promoter)의 제어 및 동일한 생물학적 경로(pathway)에 있는 유전자 클러스터를 포함하는 폴리시스트로닉(polycistronic) 전사체를 함유하는 영역
	필수 한정자	operon
	선택 한정자	allele function map note phenotype pseudo pseudogene standard_name
5.24.	특징 기호	oriT
	정의	전달 원점(origin of transfer); 컨주게이션(conjugation) 또는 동원(mobilization) 과정에서 전달이 시작되는 DNA 분자 영역
	선택 한정자	allele bound_moiety direction gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq standard_name
	분자 적용범위	DNA
	참고	rep_origin을 사용하여 복제 원점을 기술하는데 사용되어야 한다. direction 한정자는 왼쪽, 오른쪽 및 모두의 값을 허용하지만, oriT 특징과 함께 사용할 때는 왼쪽 및 오른쪽만 유효하다. 전달 원점은 염색체에 존재할 수 있으며; 플라스미드는 다수의 전달 원점을 포함할 수 있다.
5.25.	특징 기호	polyA_site
	정의	전사후 폴리아데닐화(polyadenylation)에 의해 아데닌(adenine) 잔기가 첨가될 RNA 전사체 상의 부위.
	선택 한정자	allele gene gene_synonym map note
	생물체 적용범위	진핵 생물 및 진핵 바이러스.

5.26.	Feature Key	precursor_RNA
	Definition	any RNA species that is not yet the mature RNA product; may include ncRNA, rRNA, tRNA, 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR)
	Optional qualifiers	allele function gene gene_synonym map note operon product standard_name trans_splicing
	Comment	used for RNA which may be the result of post-transcriptional processing; if the RNA in question is known not to have been processed, use the prim_transcript key
5.27.	Feature Key	prim_transcript
	Definition	primary (initial, unprocessed) transcript; may include ncRNA, rRNA, tRNA, 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron) and 3' untranslated region (3'UTR)
	Optional qualifiers	allele function gene gene_synonym map note operon standard_name
5.28.	Feature Key	primer_bind
	Definition	non-covalent primer binding site for initiation of replication, transcription, or reverse transcription; includes site(s) for synthetic e.g., PCR primer elements
	Optional qualifiers	allele gene gene_synonym map note standard_name
	Comment	used to annotate the site on a given sequence to which a primer molecule binds - not intended to represent the sequence of the primer molecule itself; since PCR reactions most often involve pairs of primers, a single primer_bind key may use the order(location,location) operator with two locations, or a pair of primer_bind keys may be used
5.29.	Feature Key	propeptide
	Definition	propeptide coding sequence; coding sequence for the domain of a proprotein that is cleaved to form the mature protein product.
	Optional qualifiers	allele

5.26.	특징 기호	precursor_RNA
	정의	아직 성숙(mature) RNA 생성물이 아닌 임의의 RNA 종; ncRNA, rRNA, tRNA, 5' 미번역 영역(5'UTR), 코딩 서열(CDS, 엑손), 개재 서열(인트론), 3'-말단의 미번역 영역(3'UTR)을 포함할 수 있음.
	선택 한정자	allele function gene gene_synonym map note operon product standard_name trans_splicing
	참고	전사 후 프로세싱(post-transcriptional processing)의 결과일 수 있는 RNA에 사용되며; 해당 RNA가 처리되지 않은 것으로 알려진 경우에 prim_transcript 기호를 사용한다.
5.27.	특징 기호	prim_transcript
	정의	1차(초기, 미가공) 전사체; ncRNA, rRNA, tRNA, 5' 미번역 영역(5'UTR), 코딩 서열(CDS, 엑손), 개재 서열(인트론), 3'-말단 미번역 영역(3'UTR)을 포함할 수 있음.
	선택 한정자	allele function gene gene_synonym map note operon standard_name
5.28.	특징 기호	primer_bind
	정의	복제, 전사 또는 역전사를 개시를 위한 비공유 프라이머 결합 부위; 합성 PCR 프라이머의 결합 부위를 포함.
	선택 한정자	allele gene gene_synonym map note standard_name
	참고	프라이머 분자가 결합하는 주어진 서열상의 부위에 주석을 달기 위해 사용됨-프라이머 분자 자체의 서열을 표현하기 위한 것은 아니다; PCR 반응에는 대부분 프라이머 쌍이 포함되기 때문에 단일 primer_bind 기호는 두 위치와 함께 order(location, location) 연산자를 사용하거나 primer_bind 기호 쌍을 사용할 수 있다.
5.29.	특징 기호	propeptide
	정의	프로펩타이드 코딩 서열; 성숙(mature) 단백질 생성물을 형성하도록 절단되는 프로단백질(proprotein)의 도메인에 대한 코딩 서열.
	선택 한정자	allele

function
gene
gene_synonym
map
note
product
pseudo
pseudogene
standard_name

5.30. Feature Key	protein_bind
Definition	non-covalent protein binding site on nucleic acid
Mandatory qualifiers	bound_moiety
Optional qualifiers	allele function gene gene_synonym map note operon standard_name
Comment	note that the regulatory feature key and regulatory_class qualifier with the value "ribosome_binding_site" must be used to describe ribosome binding sites

5.31. Feature Key	regulatory
Definition	any region of a sequence that functions in the regulation of transcription, translation, replication or chromatin structure;
Mandatory qualifiers	regulatory_class
Optional qualifiers	allele bound_moiety function gene gene_synonym map note operon phenotype pseudo pseudogene standard_name

		function gene gene_synonym map note product pseudo pseudogene standard_name
5.30.	특징 기호	protein_bind
	정의	핵산 상의 단백질 비공유 결합 부위
	필수 한정자	bound_moiety
	선택 한정자	allele function gene gene_synonym map note operon standard_name
	참고	regulatory 특징 기호 및 "ribosome_binding_site" 값을 가진 regulatory_class 한정자를 사용하여 ribosome 결합 부위를 설명하여야 한다.
5.31.	특징 기호	regulatory
	정의	전사, 번역, 복제 또는 염색질 구조의 조절에 작용하는 서열의 영역;
	필수 한정자	regulatory_class
	선택 한정자	allele bound_moiety function gene gene_synonym map note operon phenotype pseudo pseudogene standard_name

Ref.: Standards - ST.26

page: 3.26.i.17

5.32.	Feature Key	repeat_region
	Definition	region of genome containing repeating units
	Optional qualifiers	allele function gene gene_synonym map note rpt_family rpt_type rpt_unit_range rpt_unit_seq satellite standard_name

5.33.	Feature Key	rep_origin
	Definition	origin of replication; starting site for duplication of nucleic acid to give two identical copies
	Optional Qualifiers	allele direction function gene gene_synonym map note standard_name
	Comment	direction qualifier has valid values: left, right, or both

5.34.	Feature Key	rRNA
	Definition	mature ribosomal RNA; RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins
	Optional qualifiers	allele function gene gene_synonym map note operon product pseudo pseudogene standard_name
	Comment	rRNA sizes should be annotated with the product qualifier

5.32. 특징 기호 repeat_region

정의 반복 단위를 포함하는 유전체 영역

선택 한정자 allele
function
gene
gene_synonym
map
note
rpt_family
rpt_type
rpt_unit_range
rpt_unit_seq
satellite
standard_name

5.33. 특징 기호 rep_origin

정의 복제의 원점; 두 개의 동일한 사본을 제공하기 위한 핵산 복제의 시작 부위

선택 한정자 allele
direction
function
gene
gene_synonym
map
note
standard_name

참고 direction 한정자는 다음의 유효한 값을 갖는다: left, right, 또는 both

5.34. 특징 기호 rRNA

정의 성숙 ribosome RNA; 아미노산을 단백질로 조립하는 리보핵산단백질 입자(리보솜)의 RNA 성분

선택 한정자 allele
function
gene
gene_synonym
map
note
operon
product
pseudo
pseudogene
standard_name

참고 rRNA 크기는 product 한정자로 주석을 달아야 한다.

5.35. Feature Key	S_region
Definition	switch region of immunoglobulin heavy chains; involved in the rearrangement of heavy chain DNA leading to the expression of a different immunoglobulin class from the same B-cell
Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
Organism scope	eukaryotes

5.36. Feature Key	sig_peptide
Definition	signal peptide coding sequence; coding sequence for an N-terminal domain of a secreted protein; this domain is involved in attaching nascent polypeptide to the membrane leader sequence
Optional qualifiers	allele function gene gene_synonym map note product pseudo pseudogene standard_name

5.35.	특징 기호	S_region
	정의	면역글로불린(immunoglobulin) 중쇄(heavy chains)의 스위치 영역; 동일한 B-세포로부터 다른 면역글로불린(immunoglobulin) 클래스의 발현을 유도하는 중쇄 DNA의 재배열에 관여함.
	선택 한정자	allele gene gene_synonym map note product pseudo pseudogene standard_name
	생물체 적용범위	진핵생물

5.36.	특징 기호	sig_peptide
	정의	신호 펩티드 코딩 서열; 분비된 단백질의 N-말단 도메인에 대한 코딩 서열; 이 도메인은 초기 폴리펩티드를 막 리더(membrane leader) 서열에 부착시키는 데에 관여함.
	선택 한정자	allele function gene gene_synonym map note product pseudo pseudogene standard_name

5.37.	Feature Key	source
	Definition	identifies the source of the sequence; this key is mandatory; every sequence will have a single source key spanning the entire sequence
	Mandatory qualifiers	organism mol_type
	Optional qualifiers	cell_line cell_type chromosome clone clone_lib collected_by collection_date cultivar dev_stage ecotype environmental_sample germline haplogroup haplotype host identified_by isolate isolation_source lab_host lat_lon macronuclear map mating_type note organelle PCR_primers plasmid pop_variant proviral rearranged segment serotype serovar sex strain sub_clone sub_species sub_strain tissue_lib tissue_type variety
	Molecule scope	any
5.38.	Feature Key	stem_loop
	Definition	hairpin; a double-helical region formed by base-pairing between adjacent (inverted) complementary sequences in a single strand of RNA or DNA
	Optional qualifiers	allele function gene gene_synonym map note operon standard_name

5.37.	특징 기호	source
	정의	서열의 기원을 식별함; 이 기호는 필수 기재 사항임; 모든 서열은 전체 서열에 걸쳐 하나의 source 기호를 가질 것임
	필수 한정자	organism mol_type
	선택 한정자	cell_line cell_type chromosome clone clone_lib collected_by collection_date cultivar dev_stage ecotype environmental_sample germline haplogroup haplotype host identified_by isolate isolation_source lab_host lat_lon macronuclear map mating_type note organelle PCR_primers plasmid pop_variant proviral rearranged segment serotype serovar sex strain sub_clone sub_species sub_strain tissue_lib tissue_type variety
	분자 적용범위	모든 것
5.38.	특징 기호	stem_loop
	정의	hairpin 모양; 단일 가닥의 RNA 또는 DNA에서 인접한(반전된) 상보적 서열 사이의 염기쌍에 의해 형성된 이중 나선 영역
	선택 한정자	allele function gene gene_synonym map note operon standard_name

5.39. Feature Key	STS
Definition	sequence tagged site; short, single-copy DNA sequence that characterizes a mapping landmark on the genome and can be detected by PCR; a region of the genome can be mapped by determining the order of a series of STSS
Optional qualifiers	allele gene gene_synonym map note standard_name
Molecule scope	DNA
Comment	STS location to include primer(s) in primer_bind key or primers
5.40. Feature Key	telomere
Definition	region of biological interest identified as a telomere and which has been experimentally characterized
Optional qualifiers	note rpt_type rpt_unit_range rpt_unit_seq standard_name
Comment	the telomere feature describes the interval of DNA that corresponds to a specific structure at the end of the linear eukaryotic chromosome which is required for the integrity and maintenance of the end; this region is unique compared to the rest of the chromosome and represents the physical end of the chromosome
5.41. Feature Key	tmRNA
Definition	transfer messenger RNA; tmRNA acts as a tRNA first, and then as an mRNA that encodes a peptide tag; the ribosome translates this mRNA region of tmRNA and attaches the encoded peptide tag to the C-terminus of the unfinished protein; this attached tag targets the protein for destruction or proteolysis
Optional qualifiers	allele function gene gene_synonym map note product pseudo pseudogene standard_name tag_peptide

5.39.	특징 기호	STS
	정의	서열 태그(tagged) 부위; 유전체 상의 맵핑 랜드마크(landmark)를 특징 짓고 PCR에 의해 검출될 수 있는 짧은 단일-카피 DNA 서열; 유전체의 영역은 일련의 STS의 순서를 결정함으로써 맵핑될 수 있음.
	선택 한정자	allele gene gene_synonym map note standard_name
	분자 적용범위	DNA
	참고	primer_bind 기호 또는 프라이머에서 프라이머(들)를 포함하기 위한 STS 위치
5.40.	특징 기호	telomere
	정의	텔로미어로 식별되고 실험적으로 확인되는 생물학적 관심 영역
	선택 한정자	note rpt_type rpt_unit_range rpt_unit_seq standard_name
	참고	텔로미어 특징은 말단의 완전성 및 말단의 유지를 위해 필요한 선형 진핵 염색체의 말단의 특정 구조에 해당하는 DNA 부위를 기술하고; 이 영역은 나머지 염색체와 비교하여 독특하며 염색체의 물리적 말단을 표현한다.
5.41.	특징 기호	tmRNA
	정의	transfer messenger RNA; tmRNA는 우선 tRNA의 역할을 수행한 다음, 펩티드 태그를 암호화하는 mRNA의 역할을 한다. 리보솜은 tmRNA의 mRNA 영역을 번역하고, 부호화된 펩티드 태그를 미완성 단백질의 C-말단에 부착시키고; 이 부착된 태그는 파괴 또는 단백질 분해를 위해 단백질을 표적으로 한다.
	선택 한정자	allele function gene gene_synonym map note product pseudo pseudogene standard_name tag_peptide

Ref.: Standards - ST.26

page: 3.26.i.21

5.42.	Feature Key	transit_peptide
	Definition	transit peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organellar protein; this domain is involved in post-translational import of the protein into the organelle
	Optional qualifiers	allele function gene gene_synonym map note product pseudo pseudogene standard_name
5.43.	Feature Key	tRNA
	Definition	mature transfer RNA, a small RNA molecule (75-85 bases long) that mediates the translation of a nucleic acid sequence into an amino acid sequence
	Optional qualifiers	allele circular_RNA anticodon function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
5.44.	Feature Key	unsure
	Definition	a small region of sequenced bases, generally 10 or fewer in its length, which could not be confidently identified. Such a region might contain called bases (a, t, g, or c), or a mixture of called-bases and uncalled-bases ('n').
	Optional qualifiers	allele compare gene gene_synonym map note replace
	Comment	use the replace qualifier to annotate a deletion, insertion, or substitution.

5.42.	특징 기호	transit_peptide
	정의	수송 펩티드 코딩 서열: 핵-암호화된 소기관 단백질의 N-말단 도메인에 대한 코딩 서열; 이 도메인은 번역된 단백질을 소기관으로 가져오는 것에 관여함.
	선택 한정자	allele function gene gene_synonym map note product pseudo pseudogene standard_name
5.43.	특징 기호	tRNA
	정의	성숙 transfer RNA, 핵산염기 서열의 아미노산 서열로의 번역을 매개하는 작은 RNA 분자(75-85 염기 길이)
	선택 한정자	allele circular_RNA anticodon function gene gene_synonym map note operon product pseudo pseudogene standard_name trans_splicing
5.44.	특징 기호	unsure
	정의	확실하게 식별할 수 없는, 길이가 일반적으로 10 이하인 염기 서열의 작은 영역; 이러한 영역에는 알려진 핵산염기(a, t, g 또는 c), 또는 알려진 핵산염기와 다중 핵산염기('n')가 혼합된 형태가 포함될 수 있다.
	선택 한정자	allele compare gene gene_synonym map note replace
	참고	삭제, 삽입 또는 대체 주석을 달기 위해 replace 한정자를 사용한다.

Ref.: Standards - ST.26

page: 3.26.i.22

5.45.	Feature Key	V_region
	Definition	variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for the variable amino terminal portion; can be composed of V_segments, D_segments, N_regions, and J_segments
	Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes
5.46.	Feature Key	V_segment
	Definition	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for most of the variable region (V_region) and the last few amino acids of the leader peptide
	Optional qualifiers	allele gene gene_synonym map note product pseudo pseudogene standard_name
	Organism scope	eukaryotes
5.47.	Feature Key	variation
	Definition	a related strain contains stable mutations from the same gene (e.g., RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others)
	Optional qualifiers	allele compare frequency gene gene_synonym map note phenotype product replace standard_name
	Comment	used to describe alleles, RFLP's, and other naturally occurring mutations and polymorphisms; use the replace qualifier to annotate a deletion, insertion, or substitution; variability arising as a result of genetic manipulation (e.g., site directed mutagenesis) must be described with the misc_difference feature

5.45.	특징 기호	V_region
	정의	면역글로불린 경쇄, 중쇄 및 T-세포 수용체의 $\alpha/\beta/\gamma$ 사슬의 가변 영역(variable region)
	선택 한정자	allele gene gene_synonym map note product pseudo pseudogene standard_name
	생물체 적용범위	진핵생물
5.46.	특징 기호	V_segment
	정의	면역글로불린 경쇄, 중쇄 및 T-세포 수용체의 $\alpha/\beta/\gamma$ 사슬의 가변 단편(variable segment) 및 leader 펩티드의 마지막 몇 아미노산에 대한 코드
	선택 한정자	allele gene gene_synonym map note product pseudo pseudogene standard_name
	생물체 적용범위	진핵생물
5.47.	특징 기호	variation
	정의	관련 균주는 동일한 유전자(예: RFLP, 다형성 등)에서 이 위치(및 가능하면 다른)에서 제시된 서열과 다른 안정적인 돌연변이를 포함한다.
	선택 한정자	allele compare frequency gene gene_synonym map note phenotype product replace standard_name
	참고	대립 유전자(allele), RFLP 및 기타 자연 발생 돌연변이 및 다형성을 기술하는 데에 사용된다. replace 한정자를 사용하여 삭제, 삽입 또는 대체에 주석을 단다. 유전자 조작의 결과로 발생하는 변동성(예: 부위 지정 돌연변이 유발)은 misc_difference 특징으로 기술되어야 한다.

5.48.	Feature Key	3'UTR
	Definition	1) region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein; 2) region at the 3' end of an RNA virus (following the last stop codon) that is not translated into a protein;
	Optional qualifiers	allele function gene gene_synonym map note standard_name trans_splicing
	Comment	The apostrophe character has special meaning in XML, and must be substituted with "'" in the value of an element. Thus "3'UTR" must be represented as "3'UTR" in the XML file, i.e., <INSDFeature_key>3'UTR</INSDFeature_key>.
5.49.	Feature Key	5'UTR
	Definition	1) region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein; 2) region at the 5' end of an RNA virus (preceding the first initiation codon) that is not translated into a protein;
	Optional qualifiers	allele function gene gene_synonym map note standard_name trans_splicing
	Comment	The apostrophe character has special meaning in XML, and must be substituted with "'" in the value of an element. Thus "5'UTR" must be represented as "5'UTR" in the XML file, i.e., <INSDFeature_key>5'UTR</INSDFeature_key>.

5.48.	특징 기호	3'UTR
	정의	1) 단백질로 번역되지 않는 성숙 전사체(정지 코돈에 이어)의 3'-말단의 영역; 2) 단백질로 번역되지 않는 RNA 바이러스의 전사체(마지막 정지 코돈에 이어)의 3'-말단의 영역;
	선택 한정자	allele function gene gene_synonym map note standard_name trans_splicing
	참고	아포스트로피(') 문자는 XML에서 특정 의미를 가지며, 요소 값에서 "'"로 대체되어야 한다. 따라서 XML 파일에서 "3'UTR"은 "3'UTR"로 표현되어야 한다(예: <INSDFeature_key>3'UTR</INSDFeature_key>).
5.49.	특징 기호	5'UTR
	정의	1) 단백질로 번역되지 않는 성숙 전사체(개시 코돈앞)의 5'-말단에 있는 영역; 2) 단백질로 번역되지 않는 RNA 바이러스의 전사체(첫번째 개시 코돈 앞)의 5'-영역;
	선택 한정자	allele function gene gene_synonym map note standard_name trans_splicing
	참고	아포스트로피(') 문자는 XML에서 특정 의미를 가지며, 요소 값에서 "'"로 대체되어야 한다. 따라서 XML 파일에서 "5'UTR"은 "5'UTR"로 표현되어야 한다(예: <INSDFeature_key>5'UTR</INSDFeature_key>).

SECTION 6: QUALIFIERS FOR NUCLEOTIDE SEQUENCES

This section contains the list of qualifiers to be used for features in nucleotide sequences. The qualifiers are listed in alphabetic order.

Where the value format is “none”, the `INSDQualifier_value` element must not be used and the `NonEnglishQualifier_value` element must not be used.

Where the value format is free text that is identified as language-dependent, one of the following must be used:

- 1) the `INSDQualifier_value` element; or
- 2) the `NonEnglishQualifier_value` element; or
- 3) both the `INSDQualifier_value` element and the `NonEnglishQualifier_value` element.

Where the value format is something other than “none” but not identified as language-dependent free text, the `INSDQualifier_value` element must be used and the `NonEnglishQualifier_value` element must not be used.

PLEASE NOTE: Any qualifier value provided for a qualifier with a language-dependent “free text” value format may require translation for national or regional procedures. The qualifiers listed in the following table are considered to have language-dependent free text values:

Table 5: List of qualifier values for nucleotide sequences with language-dependent free-text values

Section	Language-Dependent Free Text Value
6.3	bound_moiety
6.5	cell_type
6.8	clone
6.9	clone_lib
6.11	collected_by
6.14	cultivar
6.15	dev_stage
6.18	ecotype
6.22	function
6.24	gene_synonym
6.26	haplogroup
6.28	host
6.29	identified_by
6.30	isolate
6.31	isolation_source
6.32	lab_host
6.36	mating_type
6.41	note
6.45	organism
6.47	phenotype
6.49	pop_variant
6.50	product
6.66	serotype
6.67	serovar
6.68	sex
6.69	standard_name
6.70	strain
6.71	sub_clone
6.72	sub_species
6.73	sub_strain
6.75	tissue_lib
6.76	tissue_type
6.81	variety

섹션 6: 핵산염기 서열에 대한 한정자

이 섹션에는 핵산염기 서열의 특징에 사용되는 한정자 목록이 포함되어 있다. 한정자는 알파벳 순서로 나열된다.

값 형식이 "none"이면, INSDQualifier_value 요소가 사용되어야 하며 NonEnglishQualifier_value 요소는 사용되어서는 안 된다.

값 형식이 언어에 의존에 의하여 식별되는 자유텍스트인 경우, 다음 중의 하나가 사용되어야 한다:

- 1) INSDQualifier_value 요소; 또는
- 2) NonEnglishQualifier_value 요소; 또는
- 3) INSDQualifier_value 요소와 NonEnglishQualifier_value 요소 모두.

값 형식이 "none"이 아니나 언어의존적 자유텍스트로 식별되지 않는 경우, INSDQualifier_value 요소는 사용되어야 하며 NonEnglishQualifier_value 요소는 사용되어서는 안 된다.

참고: 언어의존적 "자유텍스트" 값 형식을 가진 한정자에게 제공되는 모든 한정자 값은 국가/지역 절차에 대한 번역이 필요할 수 있다. 다음 테이블에 나열된 한정자는 언어의존적 자유텍스트 값을 갖는 것으로 간주된다.

표 5: 언어의존적 자유텍스트 값을 갖는 핵산염기 서열의 한정자 값 목록

섹션	언어의존적 자유텍스트 값
6.3	bound_moiety
6.5	cell_type
6.8	clone
6.9	clone_lib
6.11	collected_by
6.14	cultivar
6.15	dev_stage
6.18	ecotype
6.22	function
6.24	gene_synonym
6.26	haplogroup
6.28	host
6.29	identified_by
6.30	isolate
6.31	isolation_source
6.32	lab_host
6.36	mating_type
6.41	note
6.45	organism
6.47	phenotype
6.49	pop_variant
6.50	product
6.66	serotype
6.67	serovar
6.68	sex
6.69	standard_name
6.70	strain
6.71	sub_clone
6.72	sub_species
6.73	sub_strain
6.75	tissue_lib
6.76	tissue_type
6.81	variety

6.1.	Qualifier	allele
	Definition	name of the allele for the given gene
	Mandatory value format	free text
	Example	<INSDQualifier_value>adh1-1</INSDQualifier_value>
	Comment	all gene-related features (exon, CDS etc) for a given gene should share the same allele qualifier value; the allele qualifier value must, by definition, be different from the gene qualifier value; when used with the variation feature key, the allele qualifier value should be that of the variant.
6.2.	Qualifier	anticodon
	Definition	location of the anticodon of tRNA and the amino acid for which it codes
	Mandatory value format	(pos:<location>,aa:<amino_acid>,seq:<text>) where <location> is the position of the anticodon and <amino_acid> is the three letter abbreviation for the amino acid encoded and <text> is the sequence of the anticodon
	Example	<INSDQualifier_value>(pos:34..36,aa:Phe,seq:aaa)</INSDQualifier_value> <INSDQualifier_value>(pos:join(5,495..496),aa:Leu,seq:taa)</INSDQualifier_value> <INSDQualifier_value>(pos:complement(4156..4158),aa:Glu,seq:ttg)</INSDQualifier_value>
6.3.	Qualifier	bound_moiety
	Definition	name of the molecule/complex that may bind to the given feature
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>GAL4</INSDQualifier_value>
	Comment	A single bound_moiety qualifier is permitted on the "misc_binding", "orit" and "protein_bind" features.
6.4.	Qualifier	cell_line
	Definition	cell line from which the sequence was obtained
	Mandatory value format	free text
	Example	<INSDQualifier_value>MCF7</INSDQualifier_value>
6.5.	Qualifier	cell_type
	Definition	cell type from which the sequence was obtained
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>leukocyte</INSDQualifier_value>

6.1.	한정자	allele
	정의	주어진 유전자에 대한 allele의 명칭
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>adh1-1</INSDQualifier_value>
	참고	주어진 유전자에 대한 모든 유전자 관련 특징(exon, CDS 등)은 동일한 allele 한정자 값을 공유하여야 한다. allele 한정자 값은 정의에 의해 gene 한정자 값과 달라야 한다. variation 특징 기호와 함께 사용하는 경우에, allele 한정자 값은 변형의 값이어야 한다.
6.2.	한정자	ant icodon
	정의	tRNA의 안티코돈(anticodon) 및 이를 코딩하는 아미노산의 위치
	필수값 포맷	(pos:<location>,aa:<amino_acid>,seq:<text>) 여기서 <location>은 안티코돈의 위치이고, <amino_acid>는 암호화된 아미노산의 3-문자 약어이다. <text>는 안티코돈의 서열이다.
	예	<INSDQualifier_value>(pos:34..36,aa:Phe,seq:aaa)</INSDQualifier_value> <INSDQualifier_value>(pos:join(5,495..496),aa:Leu,seq:taa)</INSDQualifier_value> <INSDQualifier_value>(pos:complement(4156..4158),aa:Glu,seq:ttg)</INSDQualifier_value>
6.3.	한정자	bound_moiety
	정의	주어진 특징에 결합할 수 있는 분자/복합체의 명칭
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>GAL4</INSDQualifier_value>
	참고	단일 bound_moiety 한정자가 "misc_binding", "oriT" 및 "protein_bind" 특징에는 허용된다.
6.4.	한정자	cell_line
	정의	서열을 얻은 세포주(cell line)
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>MCF7</INSDQualifier_value>
6.5.	한정자	cell_type
	정의	서열을 얻은 세포 타입
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>leukocyte</INSDQualifier_value>

6.6.	Qualifier	chromosome
	Definition	chromosome (e.g., Chromosome number) from which the sequence was obtained
	Mandatory value format	free text
	Example	<INSDQualifier_value>1</INSDQualifier_value> <INSDQualifier_value>X</INSDQualifier_value>
6.7.	Qualifier	circular_RNA
	Definition	indicates that exons are out-of-order or overlapping because this spliced RNA product is a circular RNA (circRNA) created by backsplicing, for example when a downstream exon in the gene is located 5' of an upstream exon in the RNA product
	Value format	none
	Comment	should be used on features such as CDS, mRNA, tRNA and other features that are produced as a result of a backsplicing event. This qualifier should be used only when the splice event is indicated in the "join" operator, eg join(complement(69611..69724),139856..140087)
6.8.	Qualifier	clone
	Definition	clone from which the sequence was obtained
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>lambda-HIL7.3</INSDQualifier_value>
	Comment	a source feature must not contain more than one clone qualifier; where the sequence was obtained from multiple clones it may be further described in the feature table using the feature key misc_feature and a note qualifier to specify the multiple clones.
6.9.	Qualifier	clone_lib
	Definition	clone library from which the sequence was obtained
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>lambda-HIL7</INSDQualifier_value>
6.10.	Qualifier	codon_start
	Definition	indicates the offset at which the first complete codon of a coding feature can be found, relative to the first base of that feature.
	Mandatory value format	1 or 2 or 3
	Example	<INSDQualifier_value>2</INSDQualifier_value>

6.6.	한정자	chromosome
	정의	서열을 얻은 염색체(예: 염색체 번호)
	필수값 포맷	자유텍스트
	예	<INSDQUALIFIER_value>1</INSDQualifier_value> <INSDQualifier_value>X</INSDQualifier_value>
6.7.	한정자	circular_RNA
	정의	엑손(exon)들이 순서가 바뀌거나 중첩(overlapping)이 된 것을 나타내며, 이는 이러한 스플라이싱된 RNA 산물이 백스플라이싱(backsplicing)에 의해서 원형(circular) RNA(circRNA)이 되었기 때문이다. 예를 들면, 유전자의 아래쪽(downstream) 엑손이 DNA상에서 위쪽(upstream) 엑손의 5'에 위치하는 것이다.
	값 포맷	없음
	참고	CDS, mRNA, tRNA와 같은 특징이나 backsplicing 사건의 결과로서 생성되는 다른 특징에 사용되어야 한다. 이 한정자는 'join' 연산자(예를 들면 join(complement(69611..69724).139856..40087))에서 스플라이싱(splicing) 사건을 나타낼만 사용되어야 한다.
6.8.	한정자	clone
	정의	서열을 얻은 클론
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>lambda-hIL7.3</INSDQualifier_value>
	참고	source 특징은 둘 이상의 clone 한정자를 포함해서는 아니된다. 서열이 다중 클론으로부터 유래된 경우에, 특징 기호 misc_feature 및 note 한정자를 사용하여 특징 표에 추가로 기술되어 다중 클론을 특정할 수 있다.
6.9.	한정자	clone_lib
	정의	서열을 얻은 클론 라이브러리
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>lambda-h IL7</INSDQualifier_value>
6.10.	한정자	codon_start
	정의	해당 특징의 첫번째 염기에 대해서 coding 특징의 첫번째 완전한 코돈(codon)이 발견될 수 있는 오프셋(offset)을 나타낸다.
	필수값 포맷	1 또는 2 또는 3
	예	<INSDQualifier_value>2</INSDQualifier_value>

6.11. Qualifier	collected_by
Definition	name of persons or institute who collected the specimen
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Dan Janzen</INSDQualifier_value>
6.12. Qualifier	collection_date
Definition	date that the specimen was collected.
Mandatory value format	YYYY-MM-DD, YYYY-MM or YYYY
Example	<INSDQualifier_value>1952-10-21</INSDQualifier_value> <INSDQualifier_value>1952-10</INSDQualifier_value> <INSDQualifier_value>1952</INSDQualifier_value>
Comment	'YYYY' is a four-digit value representing the year. 'MM' is a two-digit value representing the month. 'DD' is a two-digit value representing the day of the month.
6.13. Qualifier	compare
Definition	Reference details of an existing public INSD entry to which a comparison is made
Mandatory value format	[accession-number.sequence-version]
Example	<INSDQualifier_value>AJ634337.1</INSDQualifier_value>
Comment	This qualifier may be used on the following features: misc_difference, unsure, and variation. Multiple compare qualifiers with different contents are allowed within a single feature. This qualifier is not intended for large-scale annotation of variations, such as SNPs.
6.14. Qualifier	cultivar
Definition	cultivar (cultivated variety) of plant from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Nipponbare</INSDQualifier_value> <INSDQualifier_value>Tenuifolius</INSDQualifier_value> <INSDQualifier_value>Candy Cane</INSDQualifier_value> <INSDQualifier_value>IR36</INSDQualifier_value>
Comment	'cultivar' is applied solely to products of artificial selection; use the variety qualifier for natural, named plant and fungal varieties.

6.11.	한정자	collected_by
	정의	표본을 수집한 사람 또는 기관의 명칭
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Dan Janzen</INSDQualifier_value>
6.12.	한정자	collection_date
	정의	표본이 수집된 날짜.
	필수값 포맷	YYYY-MM-DD, YYYY-MM 또는 YYYY
	예	<INSDQualifier_value>1952-10-21</INSDQualifier_value> <INSDQualifier_value>1952-10</INSDQualifier_value> <INSDQualifier_value>1952</INSDQualifier_value>
	참고	'YYYY'는 연도를 표현하는 네 자리 숫자이다. 'MM'은 월을 표현하는 두 자리 숫자 값이다. 'DD'는 월의 일을 표현하는 두 자리 숫자 값이다.
6.13.	한정자	compare
	정의	비교가 이루어지는 기존 공개 INSD 항목에 대한 참조 세부 사항
	필수값 포맷	[accession-number.sequence-version]
	예	<INSDQualifier_value>AJ634337.1</INSDQualifier_value>
	참고	이 한정자는 misc_difference, unsure 및 variation 특징에서 사용할 수 있다. 다른 콘텐츠를 가진 여러 'compare' 한정자는 단일 특징으로 허용된다. 이 한정자는 SNP와 같은 변형의 대규모 주석에는 적합하지 않다.
6.14.	한정자	cultivar
	정의	서열을 얻은 식물의 품종(재배 품종)
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Nipponbare</INSDQualifier_value> <INSDQualifier_value>Tenuifolius</INSDQualifier_value> <INSDQualifier_value>Candy Cane</INSDQualifier_value> <INSDQualifier_value>IR36</INSDQualifier_value>
	참고	'cultivar'는 인공 선택 제품에만 적용된다. 천연 식물 및 곰팡이 품종에 variety 한정자를 사용하십시오.

6.15. Qualifier	dev_stage
Definition	if the sequence was obtained from an organism in a specific developmental stage, it is specified with this qualifier
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>fourth instar larva</INSDQualifier_value>
6.16. Qualifier	direction
Definition	direction of DNA replication
Mandatory value format	left, right, or both where left indicates toward the 5' end of the sequence (as presented) and right indicates toward the 3' end
Example	<INSDQualifier_value>left</INSDQualifier_value>
Comment	The values left, right, and both are permitted when the direction qualifier is used to annotate a rep_origin feature key. However, only left and right values are permitted when the direction qualifier is used to annotate an oriT feature key.
6.17. Qualifier	EC_number
Definition	Enzyme Commission number for enzyme product of sequence
Mandatory value format	free text
Example	<INSDQualifier_value>1.1.2.4</INSDQualifier_value> <INSDQualifier_value>1.1.2.-</INSDQualifier_value> <INSDQualifier_value>1.1.2.n</INSDQualifier_value> <INSDQualifier_value>1.1.2.n1</INSDQualifier_value>
Comment	valid values for EC numbers are defined in the list prepared by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) (published in Enzyme Nomenclature 1992, Academic Press, San Diego, or a more recent revision thereof).The format represents a string of four numbers separated by full stops; up to three numbers starting from the end of the string may be replaced by dash "-" to indicate uncertain assignment. Symbols including an "n", e.g., "n", "n1" and so on, may be used in the last position instead of a number where the EC number is awaiting assignment. Please note that such incomplete EC numbers are not approved by NC-IUBMB.
6.18. Qualifier	ecotype
Definition	a population within a given species displaying genetically based, phenotypic traits that reflect adaptation to a local habitat
Mandatory value Format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Columbia</INSDQualifier_value>
Comment	an example of such a population is one that has adapted hairier than normal leaves as a response to an especially sunny habitat. 'Ecotype' is often applied to standard genetic stocks of Arabidopsis thaliana, but it can be applied to any sessile organism.

6.15.	한정자	dev_stage
	정의	서열이 특정 발생 단계에 있는 생물체로부터 얻어진 경우에, 이 한정자로 지정된다.
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>fourth instar larva</INSDQualifier_value>
6.16.	한정자	direction
	정의	DNA 복제의 방향
	필수값 포맷	왼쪽, 오른쪽 또는 양쪽 방향 여기에서, 왼쪽은(표시된 대로) 서열의 5'-말단 쪽으로 향하고, 오른쪽은 서열의 3'-말단 쪽으로 표시한다.
	예	<INSDQualifier_value>left</INSDQualifier_value>
	참고	direction 한정자를 사용하여 rep_origin 특징 기호에 주석을 달 때는 왼쪽, 오른쪽 및 모두의 값이 허용된다. 그러나 direction 한정자를 사용하여 oriT 특징 기호에 주석을 달 때는 왼쪽 및 오른쪽 값만 허용된다.
6.17.	한정자	EC_number
	정의	서열의 효소 생성물에 대한 효소 번호
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>1.1.2.4</INSDQualifier_value> <INSDQualifier_value>1.1.2.-</INSDQualifier_value> <INSDQualifier_value>1.1.2.n</INSDQualifier_value> <INSDQualifier_value>1.1.2.n1</INSDQualifier_value>
	참고	EC 번호에 대한 유효한 값은 국제 생화학 및 분자 생물학 협회(NC-IUBMB)의 명명 위원회(Enzyme Nomenclature 1992, Academic Press, San Diego, 또는 그보다 최근 개정판에 발간)에 의해 작성된 목록에 정의되어 있다. 형식은 마침표로 구분된 4개의 숫자로 이루어진 문자열을 표현한다. 문자열의 말단에서 시작하여 최대 3개의 숫자는 불확실한 할당을 표시하기 위해 "-"로 대체될 수 있다. EC 번호가 할당을 기다리는 번호 대신 마지막 위치에 "n", "n1" 등과 같은 "n"을 포함한 기호 등을 사용할 수 있다. 이러한 불완전한 EC 번호는 NC-IUBMB에서 승인되지 않는다
6.18.	한정자	ecotype
	정의	지역 서식지에 대한 적응을 반영하는 유전자에 기반한, 표현형 특징을 나타내는, 주어진 종(種) 내에서의 집단
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Columbia</INSDQualifier_value>
	참고	그러한 집단의 예는 특히 햇볕이 잘 드는 서식지에 대한 반응으로 정상 잎보다 털이 많은 집단이다. 'Ecotype'은 종종 <i>Arabidopsis thaliana</i> 의 표준 유전자 스톡(stock)에 적용되지만, 모든 고착(sessile) 생물체에 적용할 수 있다.

6.19. Qualifier	environmental_sample
Definition	identifies sequences derived by direct molecular isolation from a bulk environmental DNA sample (by PCR with or without subsequent cloning of the product, DGGE, or other anonymous methods) with no reliable identification of the source organism. Environmental samples include clinical samples, gut contents, and other sequences from anonymous organisms that may be associated with a particular host. They do not include endosymbionts that can be reliably recovered from a particular host, organisms from a readily identifiable but uncultured field sample (e.g., many cyanobacteria), or phytoplasmas that can be reliably recovered from diseased plants (even though these cannot be grown in axenic culture)
Value format	none
Comment	used only with the source feature key; source feature keys containing the environmental_sample qualifier should also contain the isolation_source qualifier; a source feature including the environmental_sample qualifier must not include the strain qualifier.
6.20. Qualifier	exception
Definition	indicates that the coding region cannot be translated using standard biological rules
Mandatory value format	One of the following controlled vocabulary phrases: RNA editing rearrangement required for product annotated by transcript or proteomic data
Example	<INSDQualifier_value>RNA editing</INSDQualifier_value> <INSDQualifier_value>rearrangement required for product</INSDQualifier_value>
Comment	only to be used to describe biological mechanisms such as RNA editing; protein translation of a CDS with an exception qualifier will be different from the corresponding conceptual translation; must not be used where transl_except qualifier would be adequate, e.g., in case of stop codon completion use.
6.21. Qualifier	frequency
Definition	frequency of the occurrence of a feature
Mandatory value format	free text representing the proportion of a population carrying the feature expressed as a fraction
Example	<INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value>
6.22. Qualifier	function
Definition	function attributed to a sequence
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>essential for recognition of cofactor </INSDQualifier_value>

6.19.	한정자	environmental_sample
	정의	소스(source) 생물체의 확실한 확인없이 대량 환경 DNA 샘플(생성물의 후속 클로닝에 관계 없는 PCR 방법, DGGE 방법 또는 기타 익명 방법에 의해)로부터 직접 분리에 의해 나온 서열을 식별한다. 환경 샘플에는 임상 샘플, 내장 내용물 및 특정 숙주와 연관될 수 있는 익명의 생물체의 다른 서열이 포함된다. 여기에는 특정 숙주에서 안정적으로 회수할 수 있는 내생 공생생물(endosymbiont), 그리고 쉽게 식별할 수는 있지만 배양되지 않은 현장 시료(예: 많은 시아노박테리아)의 생물체 또는 병든 식물에서 안정적으로 회수할 수 있는 파이토플라스마(phytoplasma)는 포함되어 있지 않다(비록 무균 배양으로 자랄 수는 없지만).
	값 포맷	없음
	참고	오직 source 특징 기호와 사용된다. environment_sample 한정자를 포함하는 source 특징 기호에는 isolation_source 한정자가 포함되어야 한다. environment_sample 한정자를 포함하는 source 특징에는 strain 한정자를 포함해서는 아니된다.
6.20.	한정자	exception
	정의	표준 생물학적 규칙을 사용하여 코딩 영역을 번역할 수 없음을 나타낸다.
	필수값 포맷	다음의 연속 어휘 구문 중 하나: RNA 편집(editing); 생성물에 필요한 재배열(rearrangement); 전사체 또는 단백질 데이터에 의해 주석이 달린
	예	<INSDQualifier_value>RNA editing</INSDQualifier_value> <INSDQualifier_value>rearrangement required for product</INSDQualifier_value>
	참고	RNA 편집과 같은 생물학적 메커니즘을 설명하는 데만 사용된다. exception 한정자를 갖는 CDS의 단백질 번역은 대응하는 개념적 번역과 다를 것이다; transl_except 한정자가 적절한 위치에 사용해서는 아니된다(예: 정지 코돈 완료 사용의 경우)
6.21.	한정자	frequency
	정의	특징의 발생 빈도
	필수값 포맷	분율(fraction)로 표현되는 특징을 수행하는 population의 비율을 표현하는 자유텍스트
	예	<INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value>
6.22.	한정자	function
	정의	서열에 기인한 특징
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>essential for recognition of cofactor</INSDQualifier_value>

Comment	The function qualifier is used when the gene name and/or product name do not convey the function attributable to a sequence.
6.23. Qualifier	gene
Definition	symbol of the gene corresponding to a sequence region
Mandatory value format	free text
Example	<INSDQualifier_value>ilv</INSDQualifier_value>
Comment	Use gene qualifier to provide the gene symbol; use standard_name qualifier to provide the full gene name.
6.24. Qualifier	gene_synonym
Definition	synonymous, replaced, obsolete or former gene symbol
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Hox-3.3</INSDQualifier_value> in a feature where the gene qualifier value is Hoxc6
Comment	used where it is helpful to indicate a gene symbol synonym; when the gene_synonym qualifier is used, a primary gene symbol must always be indicated in a gene qualifier
6.25. Qualifier	germline
Definition	the sequence presented has not undergone somatic rearrangement as part of an adaptive immune response; it is the unrearranged sequence that was inherited from the parental germline
Value format	none
Comment	germline qualifier must not be used to indicate that the source of the sequence is a gamete or germ cell; germline and rearranged qualifiers must not be used in the same source feature; germline and rearranged qualifiers must only be used for molecules that can undergo somatic rearrangements as part of an adaptive immune response; these are the T-cell receptor (TCR) and immunoglobulin loci in the jawed vertebrates, and the unrelated variable lymphocyte receptor (VLR) locus in the jawless fish (lampreys and hagfish); germline and rearranged qualifiers should not be used outside of the Craniata (taxid=89593)
6.26. Qualifier	haplogroup
Definition	name for a group of similar haplotypes that share some sequence variation. Haplogroups are often used to track migration of population groups.
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>H*</INSDQualifier_value>

참고	function	한정자는 유전자 명칭 및/또는 생성물 명칭이 서열에 기인한 특징을 전달하지 않을 때 사용된다.
6.23.	한정자	gene
	정의	서열 영역에 대응하는 유전자의 상징
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>ilvE</INSDQualifier_value>
	참고	gene 한정자를 사용하여 유전자 기호를 제공하십시오. 전체 유전자 명칭을 제공하려면 standard_name 한정자를 사용하십시오.
6.24.	한정자	gene_synonym
	정의	동언어, 대체, 구식 또는 이전 유전자 심볼(symbol)
	필수값 포맷	자유텍스트 (언어 의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Hox-3.3</INSDQualifier_value>
	참고	유전자 기호 동언어를 나타내는 것이 도움이 되는 곳에 사용된다; gene_synonym 한정자가 사용될 때 기본 유전자 기호는 항상 gene 한정자에 표시되어야 한다.
6.25.	한정자	germline
	정의	설명된 서열은 적응성(adaptive) 면역 반응의 일부로서 체세포 재배열을 받지 않았다; 이것은 부모의 생식계열(germline)에서 물려받은 재배열되지 않은 서열이다
	값 포맷	없음
	참고	germline 한정자를 사용하여 서열의 출처가 배우자(gamete) 또는 생식 세포(germ cell)임을 나타내지 않아야 한다. germline과 rearranged 한정자를 동일한 source 특징에 사용해서는 안된다. germline과 rearranged 한정자는 적응성 면역 반응의 일부로 체세포 재배열을 받을 수 있는 분자에만 사용되어야 한다. 이들은 턱이 있는 척추 동물에서 T-세포 수용체(TCR) 및 면역글로블린 유전자좌(locus)이고, 턱없는 물고기(lampreys 및 hagfish)에서의 관련되지 않은 가변 림프구 수용체(VLR) 유전자좌(locus)이다; germline과 rearranged 한정자를 Craniata(taxid = 89593) 이외에서 사용해서는 아니된다.
6.26.	한정자	haplogroup
	정의	일부 서열 변이를 공유하는 유사한 haplotype의 그룹의 명칭. haplogroup은 종종 개체군의 이동을 추적하는 데에 사용된다.
	필수값 포맷	자유텍스트 (언어 의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>H*</INSDQualifier_value>

6.27. Qualifier	haplotype
Definition	name for a specific set of alleles that are linked together on the same physical chromosome. In the absence of recombination, each haplotype is inherited as a unit, and may be used to track gene flow in populations.
Mandatory value format	free text
Example	<INSDQualifier_value>Dw3 B5 Cw1 A1</INSDQualifier_value>
6.28. Qualifier	host
Definition	natural (as opposed to laboratory) host to the organism from which sequenced molecule was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Homo sapiens</INSDQualifier_value> <INSDQualifier_value>Homo sapiens 12 year old girl</INSDQualifier_value> <INSDQualifier_value>Rhizobium NGR234</INSDQualifier_value>
6.29. Qualifier	identified_by
Definition	name of the expert who identified the specimen taxonomically
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>John Burns</INSDQualifier_value>
6.30. Qualifier	isolate
Definition	individual isolate from which the sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Patient #152</INSDQualifier_value> <INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>
6.31. Qualifier	isolation_source
Definition	describes the physical, environmental and/or local geographical source of the biological sample from which the sequence was derived
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Examples	<INSDQualifier_value>rumen isolates from standard Pelleted ration-fed steer #67</INSDQualifier_value> <INSDQualifier_value>permanent Antarctic sea ice</INSDQualifier_value> <INSDQualifier_value>denitrifying activated sludge from carbon_limited continuous reactor</INSDQualifier_value>

6.27.	한정자	haplotype
	정의	동일한 물리적 염색체에서 서로 연결된 특정 대립 유전자의 명칭. 재조합이 없는 경우에, 각 haplotype은 하나의 단위로 유전되며, 집단의 유전자 흐름을 추적하는 데에 사용될 수 있다.
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>Dw3 B5 Cw1 A1</INSDQualifier_value>
6.28.	한정자	host
	정의	염기 서열 분자가 얻어진 생물체에 대한 자연(실험실과 대조적인) 숙주
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Homo sapiens</INSDQualifier_value> <INSDQualifier_value>Homo sapiens 12 year old girl</INSDQualifier_value> <INSDQualifier_value>Rhizobium NGR234</INSDQualifier_value>
6.29.	한정자	identified_by
	정의	분류학적으로 표본을 식별한 전문가의 명칭
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>John Burns</INSDQualifier_value>
6.30.	한정자	isolate
	정의	서열을 얻은 개별 분리물(individual isolate)
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Patient #152</INSDQualifier_value> <INSDQualifier_value>DGGE band PSBAC-13</INSDQualifier_value>
6.31.	한정자	isolation_source
	정의	서열이 유래된 생물학적 시료의 물리적, 환경적 및/또는 지역적인 지리적 근원을 기술한다.
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>rumen isolates from standard Pelletedration-fed steer #67</INSDQualifier_value> <INSDQualifier_value>permanent Antarctic sea ice</INSDQualifier_value> <INSDQualifier_value>denitrifying activated sludge from carbon_limited continuous reactor</INSDQualifier_value>

Comment	used only with the source feature key; source feature keys containing an environmental_sample qualifier should also contain an isolation_source qualifier
6.32. Qualifier	lab_host
Definition	scientific name of the laboratory host used to propagate the source organism from which the sequenced molecule was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Gallus gallus</INSDQualifier_value> <INSDQualifier_value>Gallus gallus embryo</INSDQualifier_value> <INSDQualifier_value>Escherichia coli strain DH5 alpha</INSDQualifier_value> <INSDQualifier_value>Homo sapiens HeLa cells</INSDQualifier_value>
Comment	the full binomial scientific name of the host organism should be used when known; extra conditional information relating to the host may also be included
6.33. Qualifier	lat_lon
Definition	geographical coordinates of the location where the specimen was collected
Mandatory value format	free text - degrees latitude and longitude in format "d[d.ddd] N S d[dd.ddd] W E"
Example	<INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value> <INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value>
6.34. Qualifier	macronuclear
Definition	if the sequence shown is DNA and from an organism which undergoes chromosomal differentiation between macronuclear and micronuclear stages, this qualifier is used to denote that the sequence is from macronuclear DNA
Value format	none
6.35. Qualifier	map
Definition	genomic map position of feature
Mandatory value format	free text
Example	<INSDQualifier_value>8q12-q13</INSDQualifier_value>
6.36. Qualifier	mating_type
Definition	mating type of the organism from which the sequence was obtained; mating type is used for prokaryotes, and for eukaryotes that undergo meiosis without sexually dimorphic gametes
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Examples	<INSDQualifier_value>MAT-1</INSDQualifier_value> <INSDQualifier_value>plus</INSDQualifier_value>

	참고	오직 source 특징 기호와 함께 사용된다; environment_sample 한정자를 포함하는 source 특징 기호에는 isolation_source 한정자가 포함되어야 한다.
6.32.	한정자	lab_host
	정의	시퀀싱된 분자가 얻어진 기원 생물체의 전파에 사용된 실험실 숙주의 학명
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Gallus gallus</INSDQualifier_value> <INSDQualifier_value>Gallus gallus embryo</INSDQualifier_value> <INSDQualifier_value>Escherichia coli strain DH5 alpha</INSDQualifier_value> <INSDQualifier_value>Homo sapiens HeLa cells</INSDQualifier_value>
	참고	숙주 생물체의 완전한 과학적 학명은 알려진 경우에 사용하여야 한다. 호스트와 관련된 추가 조건 정보도 포함될 수 있다.
6.33.	한정자	lat_lon
	정의	표본이 수집된 위치의 지리적 좌표
	필수값 포맷	자유텍스트 - "d[.ddd] N S d[.ddd] W E" 형식의 위도 및 경도
	예	<INSDQualifier_value>47.94 N 28.12 W</INSDQualifier_value> <INSDQualifier_value>45.0123 S 4.1234 E</INSDQualifier_value>
6.34.	한정자	macronuclear
	정의	표시된 서열이 DNA이고 거대 핵 단계와 미세 핵 단계 사이에서 염색체 분화를 겪는 생물체에서 유래한 경우에, 이 한정자는 서열이 거대 핵 DNA에서 유래되었음을 나타내는 데에 사용된다.
	값 포맷	없음
6.35.	한정자	map
	정의	특정의 유전체지도 위치
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>8q12-q13</INSDQualifier_value>
6.36.	한정자	mating_type
	정의	서열이 얻어진 생물체의 교배 타입; 짝짓기 타입은 원핵생물 및 성적으로 이형적인 생식 세포없이 감수 분열을 겪는 진핵 생물에 사용된다.
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>MAT-1</INSDQualifier_value> <INSDQualifier_value>plus</INSDQualifier_value>

	<pre><INSDQualifier_value>-</INSDQualifier_value> <INSDQualifier_value>odd</INSDQualifier_value> <INSDQualifier_value>even</INSDQualifier_value>"</pre>
Comment	mating_type qualifier values male and female are valid in the prokaryotes, but not in the eukaryotes; for more information, see the entry for the sex qualifier.
6.37. Qualifier	mobile_element_type
Definition	type and name or identifier of the mobile element which is described by the parent feature
Mandatory value format	<pre><mobile_element_type>[:<mobile_element_name>] where <mobile_element_type> is one of the following: transposon retrotransposon integron insertion sequence non-LTR retrotransposon SINE MITE LINE other</pre>
Example	<pre><INSDQualifier_value>transposon:Tnp9</INSDQualifier_value></pre>
Comment	mobile_element_type is permitted on mobile_element feature key only. Mobile element should be used to represent both elements which are currently mobile, and those which were mobile in the past. Value "other" for <mobile_element_type> requires a <mobile_element_name>
6.38. Qualifier	mod_base
Definition	abbreviation for a modified nucleotide base
Mandatory value format	modified base abbreviation chosen from this Annex, Section 2
Example	<pre><INSDQualifier_value>m5c</INSDQualifier_value> <INSDQualifier_value>OTHER</INSDQualifier_value></pre>
Comment	specific modified nucleotides not found in Section 2 of this Annex are annotated by entering OTHER as the value for the mod_base qualifier and including a note qualifier with the full name of the modified base as its value
6.39. Qualifier	mol_type
Definition	molecule type of sequence
Mandatory value format	<pre>One chosen from the following: genomic DNA genomic RNA mRNA tRNA rRNA other RNA other DNA transcribed RNA viral cRNA unassigned DNA unassigned RNA</pre>

		<p><INSDQualifier_value>-</INSDQualifier_value> <INSDQualifier_value>odd</INSDQualifier_value> <INSDQualifier_value>even</INSDQualifier_value>"</p>
참고		mating_type 한정자 값은 원핵 생물에서는 유효하지만 진핵 생물에서는 그렇지 않다. 자세한 내용은 sex 한정자 항목을 참조하십시오.
6.37.	한정자	mobile_element_type
	정의	parent 특징으로 설명되는 이동 요소(mobile element)의 타입 및 명칭 또는 식별자
	필수값 포맷	<p><mobile_element_type>[:<mobile_element_name>] 여기에서, <mobile_element_type>은 다음 중 하나이다: transposon retrotransposon integron insertion sequence non-LTR retrotransposon SINE MITE LINE other</p>
	예	<INSDQualifier_value>transposon:Tnp9</INSDQualifier_value>
	참고	mobile_element_type은 mobile_element 특징 기호에서만 허용된다. 모바일 요소는 현재 모바일 요소와 과거에 모바일 요소를 모두 표현하는 것에 사용하여야 한다. <mobile_element_type>의 값 "other"에는 <mobile_element_name>이 필요하다.
6.38.	한정자	mod_base
	정의	변형 핵산염기 염기에 대한 약어
	필수값 포맷	이 별칭, 섹션 2에서 선택된 변형 염기 약어
	예	<p><INSDQualifier_value>m5c</INSDQualifier_value> <INSDQualifier_value>OTHER</INSDQualifier_value></p>
	참고	이 별칭의 섹션 2에서 찾을 수 없는 특정 변형 핵산염기는 mod_base 한정자에 대한 값으로 OTHER를 입력하고 변형 염기의 전체 명칭을 그 값으로 하는 note 한정자를 포함한다.
6.39.	한정자	mol_type
	정의	분자 타입의 서열
	필수값 포맷	<p>다음 중으로부터 하나가 선택된다: genomic DNA genomic RNA mRNA tRNA rRNA other RNA other DNA transcribed RNA viral cRNA unassigned DNA unassigned RNA</p>

Example	<INSDQualifier_value>genomic DNA</INSDQualifier_value> <INSDQualifier_value>other RNA</INSDQualifier_value>
Comment	mol_type qualifier is mandatory on the source feature key; the value "genomic DNA" does not imply that the molecule is nuclear (e.g., organelle and plasmid DNA must be described using "genomic DNA"); ribosomal RNA genes must be described using "genomic DNA"; "rRNA" must only be used if the ribosomal RNA molecule itself has been sequenced; values "other RNA" and "other DNA" must be applied to synthetic molecules, values "unassigned DNA", "unassigned RNA" must be applied where in vivo molecule is unknown.
<hr/>	
6.40. Qualifier	ncRNA_class
Definition	a structured description of the classification of the non-coding RNA described by the ncRNA parent key
Mandatory value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: antisense_RNA autocatalytically_spliced_intron circRNA ribozyme hammerhead_ribozyme lncRNA RNase_P_RNA RNase_MRP_RNA telomerase_RNA guide_RNA sgRNA asiRNA scrRNA scarRNA sirRNA pre_mirRNA mirRNA pirRNA snRNA snRNA SRP_RNA vault_RNA Y_RNA other
Example	<INSDQualifier_value>autocatalytically_spliced_intron </INSDQualifier_value> <INSDQualifier_value>asiRNA</INSDQualifier_value> <INSDQualifier_value>scrRNA</INSDQualifier_value> <INSDQualifier_value>other</INSDQualifier_value>
Comment	specific ncRNA types not yet in the ncRNA_class controlled vocabulary must be annotated by entering "other" as the ncRNA_class qualifier value, and providing a brief explanation of novel ncRNA_class in a note qualifier
<hr/>	
6.41. Qualifier	note
Definition	any comment or additional information
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>A comment about the feature</INSDQualifier_value>

예	<INSDQualifier_value>genomic DNA</INSDQualifier_value> <INSDQualifier_value>other RNA</INSDQualifier_value>
참고	mol_type 한정자는 source 특징 기호에서 필수이다. "genomic DNA"값은 분자가 핵임을 의미하지는 않는다(예를 들면, 세포 소기관 및 플라스미드 DNA는 "genomic DNA"를 사용하여 기술되어야 함); 리보솜 RNA 유전자는 "genomic DNA"를 사용하여 설명하여야 한다. "rRNA"는 리보솜 RNA 분자 자체가 시퀀싱 된 경우에만 사용하여야 한다. 값 "other RNA" 및 "other DNA"는 합성 분자에 적용되어야 하고, 값 "unassigned DNA", "unassigned RNA"는 생체내(in vivo) 분자가 알려지지 않은 경우 적용되어야 한다.
6.40.	한정자 ncRNA_class
정의	ncRNA 모(母) 기호(parent key)에 의해 기술된 비-코딩 RNA의 분류의 구조화된 설명
필수값 포맷	TYPE 여기에서 TYPE은 다음과 같은 통제 어휘 용어 또는 구문 중 하나이다: ant isense_RNA autocatalytically_spliced_intron circRNA ribozyme hammerhead_ribozyme lncRNA RNase_P_RNA RNase_MRP_RNA telomerase_RNA guide_RNA sgRNA rasiRNA scRNA scaRNA siRNA pre_m_iRNA miRNA piRNA snoRNA snRNA SRP_RNA vaul_t_RNA Y_RNA other
예	<INSDQualifier_value> autocatalytically_spliced_intron</INSDQualifier_value> <INSDQualifier_value> siRNA</INSDQualifier_value> <INSDQualifier_value> scRNA</INSDQualifier_value> <INSDQualifier_value> other</INSDQualifier_value>
참고	ncRNA_class 통제 어휘에 아직 없는 특정 ncRNA 타입은 ncRNA_class 한정자 값으로 "other"를 입력하고 note 한정자에 새로운 ncRNA_class에 대한 간략한 설명을 제공하여야 한다.
6.41.	한정자 note
정의	모든 의견 또는 추가 정보
필수값 포맷	자유텍스트 (언어 의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
예	<INSDQualifier_value>A comment about the feature</INSDQualifier_value>

6.42. Qualifier	number
Definition	a number to indicate the order of genetic elements (e.g., exons or introns) in the 5' to 3' direction
Mandatory value format	free text (with no whitespace characters)
Example	<INSDQualifier_value>4</INSDQualifier_value> <INSDQualifier_value>6B</INSDQualifier_value>
Comment	text limited to integers, letters or combination of integers and/or letters represented as a data value that contains no whitespace characters; any additional terms should be included in a standard_name qualifier. Example: a number qualifier with a value of 2A and a standard_name qualifier with a value of "long"
6.43. Qualifier	operon
Definition	name of the group of contiguous genes transcribed into a single transcript to which that feature belongs
Mandatory value format	free text
Example	<INSDQualifier_value>lac</INSDQualifier_value>
6.44. Qualifier	organelle
Definition	type of membrane-bound intracellular structure from which the sequence was obtained
Mandatory value format	One of the following controlled vocabulary terms and phrases: chromatophore hydrogenosome mitochondrion nucleomorph plastid mitochondrion:kinetoplast plastid:chloroplast plastid:apicoplast plastid:chromoplast plastid:cyanelle plastid:leucoplast plastid:proplastid
Examples	<INSDQualifier_value>chromatophore</INSDQualifier_value> <INSDQualifier_value>hydrogenosome</INSDQualifier_value> <INSDQualifier_value>mitochondrion</INSDQualifier_value> <INSDQualifier_value>nucleomorph</INSDQualifier_value> <INSDQualifier_value>plastid</INSDQualifier_value> <INSDQualifier_value>mitochondrion:kinetoplast</INSDQualifier_value> <INSDQualifier_value>plastid:chloroplast</INSDQualifier_value> <INSDQualifier_value>plastid:apicoplast</INSDQualifier_value> <INSDQualifier_value>plastid:chromoplast</INSDQualifier_value> <INSDQualifier_value>plastid:cyanelle</INSDQualifier_value> <INSDQualifier_value>plastid:leucoplast</INSDQualifier_value> <INSDQualifier_value>plastid:proplastid</INSDQualifier_value>
6.45. Qualifier	organism
Definition	scientific name of the organism that provided the sequenced genetic material, if known, or the available taxonomic information if the organism is unclassified; or an indication that the sequence is a synthetic construct

6.42.	한정자	number
	정의	5'-말단에서 3'-말단 방향으로 유전 요소(예: 엑손 또는 인트론)의 순서를 나타내는 숫자
	필수값 포맷	자유텍스트(공백 문자가없는)
	예	<INSDQualifier_value>4</INSDQualifier_value> <INSDQualifier_value>6B</INSDQualifier_value>
	참고	공백 문자를 포함하지 않는 데이터 값으로 표현되는 정수, 문자 또는 정수 및/또는 문자의 조합으로 제한되는 텍스트; 추가 용어는 standard_name 한정자에 포함되어야 한다. 예: 2A 값을 가진 number 한정자 및 "long"의 값을 가진 standard_name 한정자
6.43.	한정자	operon
	정의	해당 특징이 속하는 단일 전사체로 전사되는 연속 유전자 그룹의 명칭
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value> lac</INSDQUALIFIER_value>
6.44.	한정자	organelle
	정의	서열을 얻은 막성 세포내 구조의 종류
	필수값 포맷	통제 어휘 또는 문구 중 하나이다: chromatophore hydrogenosome mitochondrion nucleomorph plastid mitochondrion: kinetoplast plastid: chloroplast plastid: apicoplast plastid: chromoplast plastid: cyanelle plastid: leucoplast plastid: proplastid
	예	<INSDQualifier_value> chromatophore</INSDQualifier_value> <INSDQualifier_value> hydrogenosome</INSDQualifier_value> <INSDQualifier_value> mitochondrion</INSDQualifier_value> <INSDQualifier_value> nucleomorph</INSDQualifier_value> <INSDQualifier_value> plastid</INSDQualifier_value> <INSDQualifier_value> mitochondrion:kinetoplast</INSDQualifier_value> <INSDQualifier_value> plastid:chloroplast</INSDQualifier_value> <INSDQualifier_value> plastid:apicoplast</INSDQualifier_value> <INSDQualifier_value> plastid:chromoplast</INSDQualifier_value> <INSDQualifier_value> plastid:cyanelle</INSDQualifier_value> <INSDQualifier_value> plastid:leucoplast</INSDQualifier_value> <INSDQualifier_value> plastid:proplastid</INSDQualifier_value>
6.45.	한정자	organism
	정의	서열의 유전 물질을 제공한 생물체가 알려진 경우의 과학적 명칭 또는 생물체가 분류되지 않은 경우에 이용 가능한 분류학적 정보; 또는 서열이 합성 제작물(synthetic construct)이라는 표시

Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
6.46. Qualifier	PCR_primers
Definition	PCR primers that were used to amplify the sequence. A single PCR_primers qualifier should contain all the primers used for a single PCR reaction. If multiple forward or reverse primers are present in a single PCR reaction, multiple sets of fwd_name/fwd_seq or rev_name/rev_seq values will be present
Mandatory value format	[fwd_name: XXX1,]fwd_seq: xxxxx1,[fwd_name: XXX2,]fwd_seq: xxxxx2, [rev_name: YYY1,]rev_seq: yyyyy1,[rev_name: YYY2,]rev_seq: yyyyy2
Example	<INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgattttttggtcayccwgaagt,rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value> <INSDQualifier_value>fwd_name: hoge1, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<i><i>>gtgtatcttact</INSDQualifier_value> <INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgattttttggtcayccwgaagt, fwd_name: CO1P2, fwd_seq: gatacacaggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwvytardcctarraartgttg</INSDQualifier_value>
Comment	fwd_seq and rev_seq are both mandatory; fwd_name and rev_name are both optional. Both sequences must be presented in 5' to 3' order. The sequences must be given in the symbols from Section 1 of this Annex, except for the modified bases, which must be enclosed within angle brackets < >. In XML, the angle brackets < and > must be substituted with < and > since they are reserved characters in XML.
6.47. Qualifier	phenotype
Definition	phenotype conferred by the feature, where phenotype is defined as a physical, biochemical or behavioural characteristic or set of characteristics
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>erythromycin resistance</INSDQualifier_value>
6.48. Qualifier	plasmid
Definition	name of naturally occurring plasmid from which the sequence was obtained, where plasmid is defined as an independently replicating genetic unit that cannot be described by chromosome or segment qualifiers
Mandatory value format	free text
Example	<INSDQualifier_value>pC589</INSDQualifier_value>
6.49. Qualifier	pop_variant
Definition	name of subpopulation or phenotype of the sample from which the sequence was derived
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures

	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
6.46.	한정자	PCR_primers
	정의	서열을 증폭시키기 위해 사용한 PCR 프라이머. 단일 PCR_primers 한정자는 단일 PCR 반응에 사용되는 모든 프라이머를 포함하여야 한다. 다수의 정방향 또는 표적 프라이머가 존재하는 단일 PCR 반응인 경우에는 여러 세트의 fwd_name/fwd_seq 또는 rev_name/rev_seq 값이 존재한다.
	필수값 포맷	[fwd_name: XXX1,]fwd_seq: xxxxx1,[fwd_name: XXX2,]fwd_seq: xxxxx2, [rev_name: YYY1,]rev_seq : yyyyy1, [rev_name: YYY2,]rev_seq : yyyyy2
	예	INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt, rev_name: CO1R4, rev_seq: ccwytardcctarraartgttg</INSDQualifier_value> <INSDQualifier_value>fwd_name: hoge1, fwd_seq: cgkgtgtatcttact, rev_name: hoge2, rev_seq: cg<< >gtgtatttact</INSDQualifier_value> <INSDQualifier_value>fwd_name: CO1P1, fwd_seq: ttgatttttggtcayccwgaagt, fwd_name: CO1P2, fwd_seq: gatacacagtcayccwgaagt, rev_name: CO1R4, rev_seq : ccwytardcctarraartgttg</INSDQualifier_value>
	참고	fwd_seq 및 rev_seq는 모두 필수이다. fwd_name 및 rev_name은 모두 선택 사항이다. 두 순서 모두 5' > 3'순서로 제시하여야 한다. 순서는 변형 염기를 제외하고 이 별첨의 섹션 1의 기호로 제공되어야 하며 꺾쇠 괄호 < >로 묶어야 한다. XML에서 꺾쇠 괄호 < 및 >는 < >로 대체되어야 한다; 그것들은 XML로 예약된 문자이기 때문이다.
6.47.	한정자	phenotype
	정의	특징에 의해 부여된 표현형, 여기서 표현형은 물리적, 생화학적 또는 행동적 특징 또는 특징 세트로 정의됨.
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>erythromycin resistance</INSDQualifier_value>
6.48.	한정자	plasmid
	정의	서열이 얻어진 자연 발생 플라스미드의 명칭, 여기서 플라스미드는 염색체 또는 분절 한정자에 의해 설명될 수 없는 독립적으로 복제되는 유전자 단위로 정의됨.
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>pC589</INSDQualifier_value>
6.49.	한정자	pop_variant
	정의	서열이 유래된 샘플의 소집단 또는 표현형의 명칭
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)

Example	<INSDQualifier_value>pop1</INSDQualifier_value> <INSDQualifier_value>Bear Paw</INSDQualifier_value>
6.50. Qualifier	product
Definition	name of the product associated with the feature, e.g., the mRNA of an mRNA feature, the polypeptide of a CDS, the mature peptide of a mat_peptide, etc.
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>trypsinogen</INSDQualifier_value> (when qualifier appears in CDS feature) <INSDQualifier_value>trypsin</INSDQualifier_value> (when qualifier appears in mat_peptide feature) <INSDQualifier_value>XYZ neural-specific transcript</INSDQualifier_value> (when qualifier appears in mRNA feature)
6.51. Qualifier	protein_id
Definition	protein sequence identification number, an integer used in a sequence listing to designate the protein sequence encoded by the coding sequence identified in the corresponding CDS feature key and translation qualifier
Mandatory value format	an integer greater than zero
Example	<INSDQualifier_value>89</INSDQualifier_value>
6.52. Qualifier	proviral
Definition	this qualifier is used to flag sequence obtained from a virus or phage that is integrated into the genome of another organism
Value format	none
6.53. Qualifier	pseudo
Definition	indicates that this feature is a non-functional version of the element named by the feature key
Value format	none
Comment	The qualifier pseudo should be used to describe non-functional genes that are not formally described as pseudogenes, e.g., CDS has no translation due to other reasons than pseudogenization events. Other reasons may include sequencing or assembly errors. In order to annotate pseudogenes the qualifier pseudogene must be used, indicating the TYPE of pseudogene.
6.54. Qualifier	pseudogene
Definition	indicates that this feature is a pseudogene of the element named by the feature key
Mandatory value format	TYPE where TYPE is one of the following controlled vocabulary terms or phrases: processed unprocessed unitary

	예	<INSDQualifier_value>pop1</INSDQualifier_value> <INSDQualifier_value>Bear Paw</INSDQualifier_value>
6.50.	한정자	product
	정의	특징과 관련된 제품 명칭(예: mRNA 특징의 mRNA, CDS의 폴리펩티드, mat_peptide의 성숙 펩티드 등).
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>trypsinogen</INSDQualifier_value>(한정자가 CDS 특징에 나타날 때) <INSDQualifier_value>trypsin</INSDQualifier_value>(한정자가 mat_peptide 특징에 나타날 때) <INSDQualifier_value>XYZ neural- specific t ranscript</INSDQualifier_value>(한정자가 mRNA 특징에 나타날 때)
6.51.	한정자	protein_id
	정의	단백질 서열식별 번호; 대응하는 CDS 특징 기호 및 translation 한정자에서 식별된 코딩 서열에 의해 암호화된(encoded) 단백질 서열을 지정하기 위해 서열목록에 사용되는 정수
	필수값 포맷	0보다 큰 정수
	예	<INSDQualifier_value>89</INSDQualifier_value>
6.52.	한정자	proviral
	정의	이 한정자는 다른 생물체의 유전체에 통합된 바이러스 또는 파지에서 얻은 서열을 표시하는 데에 사용.
	값 포맷	없음
6.53.	한정자	pseudo
	정의	이 특징은 특징 기호로 명명된 요소의 비기능(non-functional) 버전이라는 것을 나타냄.
	값 포맷	없음
	참고	한정자 pseudo는 유사유전자(pseudogene)로서 공식적으로 기술되지 않은 비기능성 유전자, 예를 들면 유사유전자 생성과정(pseudogenization) 이벤트 이외의 다른 이유로 인해 번역되지 않는 CDS. 상기 다른 이유는 시퀀싱 또는 어셈블리 오류를 포함할 수 있다. 유사유전자에 주석을 달려면, 한정자 pseudogene을 사용해야 하여야 하며 상기 한정자는 유사유전자의 타입을 나타냄.
6.54.	한정자	pseudogene
	정의	이 특징은, 특징 기호로 명칭이 지정된 요소의 유사유전자(pseudogene) 임을 나타낸다.
	필수값 포맷	TYPE 여기에서 TYPE은 다음의 통제 어휘 또는 문구 중 하나이다: processed unprocessed unitary

	allelic unknown
Example	<pre><INSDQualifier_value>processed</INSDQualifier_value> <INSDQualifier_value>unprocessed</INSDQualifier_value> <INSDQualifier_value>unitary</INSDQualifier_value> <INSDQualifier_value>allelic</INSDQualifier_value> <INSDQualifier_value>unknown</INSDQualifier_value></pre>
Comment	<p>Definitions of TYPE values:</p> <p>processed - the pseudogene has arisen by reverse transcription of a mRNA into cDNA, followed by reintegration into the genome. Therefore, it has lost any intron/exon structure, and it might have a pseudo-polyA-tail.</p> <p>unprocessed - the pseudogene has arisen from a copy of the parent gene by duplication followed by accumulation of random mutations. The changes, compared to their functional homolog, include insertions, deletions, premature stop codons, frameshifts and a higher proportion of non-synonymous versus synonymous substitutions.</p> <p>unitary - the pseudogene has no parent. It is the original gene, which is functional in some species but disrupted in some way (indels, mutation, recombination) in another species or strain.</p> <p>allelic - a (unitary) pseudogene that is stable in the population but importantly it has a functional alternative allele also in the population. i.e., one strain may have the gene, another strain may have the pseudogene. MHC haplotypes have allelic pseudogenes.</p> <p>unknown - the submitter does not know the method of pseudogenization.</p>
6.55. Qualifier	rearranged
Definition	the sequence presented in the entry has undergone somatic rearrangement as part of an adaptive immune response; it is not the unrearranged sequence that was inherited from the parental germline
Value format	none
Comment	The rearranged qualifier must not be used to annotate chromosome rearrangements that are not involved in an adaptive immune response; germline and rearranged qualifiers must not be used in the same source feature; germline and rearranged qualifiers must only be used for molecules that can undergo somatic rearrangements as part of an adaptive immune response; these are the T-cell receptor (TCR) and immunoglobulin loci in the jawed vertebrates, and the unrelated variable lymphocyte receptor (VLR) locus in the jawless fish (lampreys and hagfish); germline and rearranged qualifiers should not be used outside of the Craniata (taxid=89593)
6.56. Qualifier	recombination_class
Definition	a structured description of the classification of recombination hotspot region within a sequence
Mandatory value format	<p>TYPE</p> <p>where TYPE is one of the following controlled vocabulary terms or phrases:</p> <pre>meiotic mitotic non_allelic_homologous chromosome_breakpoint other</pre>
Example	<pre><INSDQualifier_value>meiotic</INSDQualifier_value> <INSDQualifier_value>chromosome_breakpoint</INSDQualifier_value></pre>
Comment	specific recombination classes not yet in the recombination_class controlled vocabulary must be annotated by entering "other" as the recombination_class qualifier value and providing a brief explanation of the novel recombination_class in a note qualifier

		allelic unknown
예		<INSDQualifier_value>processed</INSDQualifier_value> <INSDQualifier_value>unprocessed</INSDQualifier_value> <INSDQualifier_value>unitary</INSDQualifier_value> <INSDQualifier_value>allelic</INSDQualifier_value> <INSDQualifier_value>unknown</INSDQualifier_value>
참고		TYPE 값의 정의: processed - 유사유전자는 mRNA를 cDNA로 역전사시킨 후 유전체로 재통합함으로써 발생하였다. 따라서 인트론/엑손 구조가 없어지고 pseudo-polyA-tail이 있을 수 있다. unprocessed - 유사유전자는 복제 후 무작위 돌연변이의 축적에 의해 모 유전자의 카피로부터 발생하였다. 이들의 기능적 상동 유전자에 비해 삽입, 삭제, 조기 정지 코돈, 프레임시프트 및 비-동의 대 동의어 치환의 더 높은 비율을 포함한다. unitary - 유사유전자에는 parent가 없다. 이것은 고유의 유전자이며, 기능적으로 어떤 종이지만 다른 종이나 균주에서 어떤 방식(Indels, 돌연변이, 재조합)으로 변경된 것이다. allelic - 집단에서 안정한(단일) 유사유전자(pseudogene)는 집단에서 안정적이나 중요한 점은 이 유전자는 집단에서 기능적으로 상이한 대립유전자를 가진다는 것이다. 즉, 하나의 strain은 유전자이고 유사유전자 일 수 있다. MHC 일배체는 유사유전자인 대립 유전자를 가진다. unknown - 제출자가 유사유전자 생성과정(pseudogenization)을 모른다.
6.55.	한정자	rearranged
	정의	엔트리에 설명된 서열은 적응성 면역 반응의 일부로서 체세포 재배열을 겪은 것으로, 상기 서열은 부모 생식 계통에서 물려받은 재배열된 서열이다.
	값 포맷	없음
	참고	rearranged 한정자는 적응 면역 반응에 관여하지 않는 염색체 재배열을 나타내기 위해 사용되어서는 아니된다. germline과 rearranged 한정자를 동일한 source 특징에 사용해서는 아니된다. germline과 rearranged 한정자는 적응성 면역 반응의 일부로 체세포 재배열을 받을 수 있는 분자에만 사용하여야 한다. 이들은 턱이 있는 척추 동물에서 T- 세포 수용체(TCR) 및 면역글로블린 유전자좌이고, 턱이 없는 어류(칠성장어류 및 해파리)에서 관련없는 가변 림프구 수용체(VLR) 유전자좌이고; germline과 rearranged 한정자를 Craniata(taxid = 89593) 이외에서 사용해서는 아니된다.
6.56.	한정자	recombination_class
	정의	서열 내에서 재조합 핫스팟 영역의 분류에 대한 구조적 설명
	필수값 포맷	TYPE 여기에서 TYPE은 통제 어휘 또는 문구 중 하나이다: meiotic mitotic non_allelic_homologous chromosome_breakpoint other
	예	<INSDQualifier_value>meiotic</INSDQualifier_value> <INSDQualifier_value>chromosome_breakpoint</INSDQualifier_value>
	참고	recombination_class 통제 어휘에 아직없는 특정 재조합 클래스는 recombination_class 한정자 값으로 "other"를 입력하고 note 한정자에서 신규 recombination_class에 대한 간략한 설명을 제공하여 주석을 달아야 한다.

6.57. Qualifier	regulatory_class
Definition	a structured description of the classification of transcriptional, translational, replicational and chromatin structure related regulatory elements in a sequence
Mandatory value format	<p>TYPE</p> <p>where TYPE is one of the following controlled vocabulary terms or phrases:</p> <p>attenuator CAAT_signal DNase_I_hypersensitive_site enhancer enhancer_blocking_element GC_signal imprinting_control_region insulator locus_control_region matrix_attachment_region minus_35_signal minus_10_signal polyA_signal_sequence promoter recoding_stimulatory_region recombination_enhancer replication_regulatory_region response_element ribosome_binding_site riboswitch silencer TATA_box terminator transcriptional_cis_regulatory_region uORF other</p>
Example	<p><INSDQualifier_value>promoter</INSDQualifier_value></p> <p><INSDQualifier_value>enhancer</INSDQualifier_value></p> <p><INSDQualifier_value>ribosome_binding_site</INSDQualifier_value></p>
Comment	specific regulatory classes not yet in the regulatory_class controlled vocabulary must be annotated by entering "other" as the regulatory_class qualifier value and providing a brief explanation of the novel regulatory_class in a note qualifier
6.58. Qualifier	replace
Definition	indicates that the sequence identified in a feature's location is replaced by the sequence shown in the qualifier's value; if no sequence (i.e., no value) is contained within the qualifier, this indicates a deletion
Mandatory value format	free text
Example	<p><INSDQualifier_value>a</INSDQualifier_value></p> <p><INSDQualifier_value></INSDQualifier_value> - for a deletion</p>
6.59. Qualifier	ribosomal_slippage
Definition	during protein translation, certain sequences can program ribosomes to change to an alternative reading frame by a mechanism known as ribosomal slippage
Value format	none

6.57.	한정자	regulatory_class
	정의	전사, 번역, 복제 및 염색질 구조 관련 서열 조절 요소의 분류에 대한 구조적 설명
	필수값 포맷	TYPE 여기에서 TYPE은 통제 어휘 또는 문구 중 하나이다: attenuator CAAT_signal DNase_I_hypersensitive_site enhancer enhancer_blocking_element GC_signal imprinting_control_region insulator locus_control_region matrix_attachment_region minus_35_signal minus_0_signal polyA_signal_sequence promoter recoding_stimulatory_region recombination_enhancer replication_regulatory_region response_element ribosome_binding_site riboswitch silencer TATA_box terminator transcriptional_cis_regulatory_region uORF other
	예	<INSDQualifier_value>promoter</INSDQualifier_value> <INSDQualifier_value> enhancer</INSDQualifier_value> <INSDQualifier_value>ribosome_binding site</INSDQualifier_value>
	참고	regulatory_class 통제 어휘에 아직 속하지 않은 특정 규제 클래스는 regulatory_class 한정자 값으로 "other"를 입력하고 note 한정 자에 새로운 regulatory_class에 대한 간략한 설명을 제공하여 주석을 달아야 한다.
6.58.	한정자	replace
	정의	특정의 위치에서 식별된 서열이 한정자 값에 표시된 서열로 대체됨을 나타낸다. 한정자 값에 표시된 서열이 없는 경우(즉, 값이 없는 경우), 이는 서열의 삭제를 나타낸다.
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value> a</INSDQualifier_value> <INSDQualifier_value> </INSDQualifier_value> - 삭제를 나타내는 경우
6.59.	한정자	ribosomal_slippage
	정의	단백질 번역 동안, 특정 서열은 리보솜 미끄러짐(ribosomal slippage)으로 알려진 메커니즘에 의해 리보솜이 대체 판독 프레임으로 변경되도록 프로그래밍할 수 있다.
	값 포맷	없음

Comment	a join operator, e.g., [join(486..1784,1787..4810)] must be used in the CDS feature location to indicate the location of ribosomal_slippage
6.60. Qualifier	rpt_family
Definition	type of repeated sequence; "Alu" or "Kpn", for example
Mandatory value format	free text
Example	<INSDQualifier_value>Alu</INSDQualifier_value>
6.61. Qualifier	rpt_type
Definition	structure and distribution of repeated sequence
Mandatory value format	One of the following controlled vocabulary terms or phrases: tandem direct inverted flanking nested terminal dispersed long_terminal_repeat non_ltr_retrotransposon_polymeric_tract centromeric_repeat telomeric_repeat x_element_combinatorial_repeat y_prime_element other
Example	<INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value>
Comment	Definitions of the values: tandem - a repeat that exists adjacent to another in the same orientation; direct - a repeat that exists not always adjacent but is in the same orientation; inverted - a repeat pair occurring in reverse orientation to one another on the same molecule; flanking - a repeat lying outside the sequence for which it has functional significance (eg. transposon insertion target sites); nested - a repeat that is disrupted by the insertion of another element; dispersed - a repeat that is found dispersed throughout the genome; terminal - a repeat at the ends of and within the sequence for which it has functional significance (eg. transposon LTRs); long_terminal_repeat - a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses; non_ltr_retrotransposon_polymeric_tract - a polymeric tract, such as poly(dA), within a non LTR retrotransposon; centromeric_repeat - a repeat region found within the modular centromere; telomeric_repeat - a repeat region found within the telomere; x_element_combinatorial_repeat - a repeat region located between the x element and the telomere or adjacent y' element; y_prime_element - a repeat region located adjacent to telomeric repeats or x element combinatorial repeats, either as a single copy or tandem repeat of two to four copies; other - a repeat exhibiting important attributes that cannot be described by other values.

참고 join 연산자(예: [join(486..1784,1787..4810)]을 CDS 특징 위치에서 사용하여 ribosomal_slippage의 위치를 나타낼 수 있다.

6.60.	한정자	rpt_family
	정의	반복되는 서열의 타입; 예를 들면 "Alu" 또는 "Kpn"
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>Alu</INSDQualifier_value>
6.61.	한정자	rpt_type
	정의	반복된 서열의 구조 및 분포
	필수값 포맷	다음과 같은 통제 어휘 또는 문구 중 하나이다. tandem direct inverted flanking nested terminal dispersed long_terminal_repeat non_ltr_retrotransposon_polymeric_tract centromeric_repeat telomeric_repeat x_element_combinatorial_repeat y_prime_element other
	예	<INSDQualifier_value>inverted</INSDQualifier_value> <INSDQualifier_value>long_terminal_repeat</INSDQualifier_value>
	참고	값의 정의: tandem - 동일한 방향으로 다른 것에 인접하여 존재하는 반복; direct - 항상 인접하지 않으며 존재하지만 같은 방향으로 반복; inverted - 같은 분자 내에서 서로 반대 방향으로 반복되는 쌍; flanking - 기능적 중요성을 가지는 서열의 외부에 있는 반복(예를 들면, 트랜스포손 삽입 표적 부위); nested - 다른 요소를 삽입하여 중단된 반복; dispersed - 유전체 전체에 분산된 것으로 확인 된 반복; terminal - 기능적 중요성을 갖는 서열의 말미 및 내부의 반복(예: 트랜스포손 LTR); long_terminal_repeat - 정의된 서열의 양쪽 말단에서 직접 반복되는 서열로, 레트로바이러스에서 일반적으로 발견됨; non_ltr_retrotransposon_polymeric_tract - 비 LTR 레트로트랜스포손 내에서 poly(dA)와 같은 폴리머 트랙; centromeric_repeat - 모듈식 센트로미어(modular centromere) 내에서 발견되는 반복 영역; telomeric_repeat - 텔로미어(telomere) 내에서 발견되는 반복 영역; x_element_combinatorial_repeat - X 요소와 텔로미어(telomere) 또는 인접한 Y' 요소 사이에 위치한 반복 영역; y_prime_element - 텔로미어(telomere) 반복 또는 X 요소 재조합 반복 인근에 위치한 반복 영역으로, 단일 사본 또는 2~4 사본의 탠덤(tandem) 반복; other - 다른 값으로 설명할 수 없는 중요한 속성을 나타내는 반복.

6.62.	Qualifier	rpt_unit_range
	Definition	location of a repeating unit expressed as a range
	Mandatory value format	<base_range> - where <base_range> is the first and last base (separated by two dots) of a repeating unit
	Example	<INSDQualifier_value>202..245</INSDQualifier_value>
	Comment	used to indicate the base range of the sequence that constitutes a repeating unit within the region specified by the feature keys oriT and repeat_region.
6.63.	Qualifier	rpt_unit_seq
	Definition	identity of a repeat sequence
	Mandatory value format	free text
	Example	<INSDQualifier_value>aagggc</INSDQualifier_value> <INSDQualifier_value>ag(5)tg(8)</INSDQualifier_value> <INSDQualifier_value>(AAAGA)6(AAAA)1(AAAGA)12</INSDQualifier_value>
	Comment	used to indicate the literal sequence that constitutes a repeating unit within the region specified by the feature keys oriT and repeat_region
6.64.	Qualifier	satellite
	Definition	identifier for a satellite DNA marker, composed of many tandem repeats (identical or related) of a short basic repeated unit
	Mandatory value format	<satellite_type>[:<class>][<identifier>] - where <satellite_type> is one of the following: satellite microsatellite minisatellite
	Example	<INSDQualifier_value>satellite: Sla</INSDQualifier_value> <INSDQualifier_value>satellite: alpha</INSDQualifier_value> <INSDQualifier_value>satellite: gamma III</INSDQualifier_value> <INSDQualifier_value>microsatellite: DC130</INSDQualifier_value>
	Comment	many satellites have base composition or other properties that differ from those of the rest of the genome that allows them to be identified.
6.65.	Qualifier	segment
	Definition	name of viral or phage segment sequenced
	Mandatory value format	free text
	Example	<INSDQualifier_value>6</INSDQualifier_value>
6.66.	Qualifier	serotype
	Definition	serological variety of a species characterized by its antigenic properties
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional

6.62.	한정자	rpt_unit_range
	정의	정수의 범위로 표현된 반복 단위의 위치
	필수값 포맷	<base_range> - 여기서 <base_range>는 반복 단위의 첫 번째 및 마지막 염기이다(두 개의 점으로 구분).
	예	<INSDQualifier_value>202..245</INSDQualifier_value>
	참고	특징 기호 oriT 및 repeat_region으로 지정된 영역 내에서 반복 단위를 구성하는 서열의 염기 영역의 범위를 나타내는 데에 사용된다.
6.63.	한정자	rpt_unit_seq
	정의	반복 서열 그 자체
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>aagggc</INSDQualifier_value> <INSDQualifier_value>ag(5)tg(8)</INSDQualifier_value> <INSDQualifier_value>(AAAGA)6(AAAA)1(AAAGA)12</INSDQualifier_value>
	참고	특징 기호 oriT 및 repeat_region으로 지정된 영역 내에서 반복 단위를 구성하는 문자 그대로의 서열을 나타내는 데에 사용된다.
6.64.	한정자	satellite
	정의	짧은 기본 반복 단위의 많은 tandem 반복(동일 또는 관련)인 satellite DNA marker의 식별자.
	필수값 포맷	<satellite_type>[:<class>][<identifier>] - 여기에서 <satellite_type>은 다음 중 하나: satellite microsatellite minisatellite
	예	<INSDQualifier_value>satellite: Sla</INSDQualifier_value> <INSDQualifier_value>satellite: alpha</INSDQualifier_value> <INSDQualifier_value>satellite: gamma III</INSDQualifier_value> <INSDQualifier_value>microsatellite: DC130</INSDQualifier_value>
	참고	다수의 satellite는 기본 구성이나 다른 유전체와 다른 특징을 가지고 있어 이를 식별할 수 있다.
6.65.	한정자	segment
	정의	시퀀싱된 바이러스 또는 파아지 단편의 명칭
	필수값 포맷	자유텍스트
	예	<INSDQualifier_value>6</INSDQualifier_value>
6.66.	한정자	serotype
	정의	항원 성질에 의해서 구분되는 종의 혈청학적 다양성
	필수값 포맷	자유텍스트

	procedures
Example	<INSDQualifier_value>B1</INSDQualifier_value>
Comment	used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for the prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms".
<hr/>	
6.67. Qualifier	serovar
Definition	serological variety of a species (usually a prokaryote) characterized by its antigenic properties
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>O157:H7</INSDQualifier_value>
Comment	used only with the source feature key; the Bacteriological Code recommends the use of the term 'serovar' instead of 'serotype' for prokaryotes; see the International Code of Nomenclature of Bacteria (1990 Revision) Appendix 10.B "Infraspecific Terms".
<hr/>	
6.68. Qualifier	sex
Definition	sex of the organism from which the sequence was obtained; sex is used for eukaryotic organisms that undergo meiosis and have sexually dimorphic gametes
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Examples	<INSDQualifier_value>female</INSDQualifier_value> <INSDQualifier_value>male</INSDQualifier_value> <INSDQualifier_value>hermaphrodite</INSDQualifier_value> <INSDQualifier_value>unisexual</INSDQualifier_value> <INSDQualifier_value>bisexual</INSDQualifier_value> <INSDQualifier_value>asexual</INSDQualifier_value> <INSDQualifier_value>monoecious</INSDQualifier_value> [or monocious] <INSDQualifier_value>dioecious</INSDQualifier_value> [or diecious]
Comment	The sex qualifier should be used (instead of mating_type qualifier) in the Metazoa, Embryophyta, Rhodophyta & Phaeophyceae; mating_type qualifier should be used (instead of sex qualifier) in the Bacteria, Archaea & Fungi; neither sex nor mating_type qualifiers should be used in the viruses; outside of the taxa listed above, mating_type qualifier should be used unless the value of the qualifier is taken from the vocabulary given in the examples above
<hr/>	
6.69. Qualifier	standard_name
Definition	accepted standard name for this feature
Mandatory value format	free text this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>dotted</INSDQualifier_value>
Comment	use standard_name qualifier to give full gene name, but use gene qualifier to give

		(언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
예		<INSDQualifier_value>B1</INSDQualifier_value>
참고		항상 source 특징 기호와 함께 사용된다. Bacteriological Code에서는 원핵생물에 대해 'serotype' 대신에 'serovar' 용어를 사용할 것을 권장한다. 국제 박테리아 명명법(1990 개정판) 별첨 10.B "인프라 특정 용어(Infraspecific Terms)"를 참조하십시오.
6.67.	한정자	serovar
	정의	항원 특징으로 구분되는 종(일반적으로 원핵생물)의 혈청학적 다양성
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>O157:H7</INSDQualifier_value>
	참고	source 특징 기호와 함께 사용된다. Bacteriological Code에서는 원핵생물에 대해 'serotype' 대신에 'serovar' 용어를 사용할 것을 권장한다. 국제 박테리아 명명법(1990 개정판) 별첨 10.B "인프라 특정 용어(Infraspecific Terms)"를 참조하십시오.
6.68.	한정자	sex
	정의	서열을 얻은 생물체의 성별; 성(sex)은 감수 분열을 겪고 성적으로 이형적인 생식 세포가 있는 진핵생물에 사용된다.
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>female</INSDQualifier_value> <INSDQualifier_value>male</INSDQualifier_value> <INSDQualifier_value>hermaphrodite</INSDQualifier_value> <INSDQualifier_value>unisexual</INSDQualifier_value> <INSDQualifier_value>bisexual</INSDQualifier_value> <INSDQualifier_value>asexual</INSDQualifier_value> <INSDQualifier_value>monoecious</INSDQualifier_value> [or monocious] <INSDQualifier_value>dioecious</INSDQualifier_value> [or diecious]
	참고	sex 한정자는 Metazoa, Embryophyta, Rhodophyta 및 Phaeophyceae에서 mating_type 한정자 대신 사용되어야 한다. 박테리아, Archaea 및 Fungi에서는 mating_type 한정자를 사용하여야 한다(sex 한정자 대신에). 바이러스에는 sex 한정자나 mating_type 한정자를 사용해서는 아니된다. 위에 나열된 분류군 이외의 경우에 한정자 값이 위 예에 제공된 어휘를 가져오지 않는 한 mating_type 한정자를 사용하여야 한다.
6.69.	한정자	standard_name
	정의	이 특징에 허용되는 표준 명칭
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>dotted</INSDQualifier_value>
	참고	standard_name 한정자를 사용하여 전체 유전자 명칭을 제공하고, gene 한정자를 사용하여 유전자 심볼(symbol)를 제공해야 함(위의 예에서 gene 한정자 값은 Dt 임).

gene symbol (in the above example gene qualifier value is Dt).

6.70. Qualifier	strain
Definition	strain from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>BALB/c</INSDQualifier_value>
Comment	feature entries including a strain qualifier must not include the environmental_sample qualifier
6.71. Qualifier	sub_clone
Definition	sub-clone from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>lambda-hIL7.20g</INSDQualifier_value>
Comment	a source feature must not contain more than one sub_clone qualifier; to indicate that the sequence was obtained from multiple sub_clones, multiple sources may be further described using the feature key "misc_feature" and the qualifier "note"
6.72. Qualifier	sub_species
Definition	name of sub-species of organism from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>lactis</INSDQualifier_value>
6.73. Qualifier	sub_strain
Definition	name or identifier of a genetically or otherwise modified strain from which sequence was obtained, derived from a parental strain (which should be annotated in the strain qualifier). sub_strain from which sequence was obtained
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>abis</INSDQualifier_value>
Comment	must be accompanied by a strain qualifier in a source feature; if the parental strain is not given, the modified strain should be annotated in the strain qualifier instead of sub_strain. For example, either a strain qualifier with the value K-12 and a substrain qualifier with the value MG1655 or a strain qualifier with the value MG1655

6.70.	한정자	strain
	정의	서열을 얻은 균주
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>BALB /c</INSDQualifier_value>
	참고	strain 한정자를 포함한 특징 항목은 environment_sample 한정자를 포함하지 않아야 한다.
6.71.	한정자	sub_clone
	정의	서열을 얻은 서브 클론
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>lambdahl7.20g</INSDQualifier_value>
	참고	source 특징에는 둘 이상의 sub_clone 한정자가 포함되어서는 안 된다. 서열이 다수의 서브 클론으로부터 얻었다는 것을 나타내기 위해, 특징 기호 "misc_feature" 및 한정자 "note"를 사용하여 다수의 소스를 추가로 기술할 수 있다.
6.72.	한정자	sub_species
	정의	서열을 얻은 생물체의 아종(sub-species) 명칭
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>lactis</INSDQualifier_value>
6.73.	한정자	sub_strain
	정의	부모 균주(strain 한정자에 주석 필요)에서 유래하고, 서열을 얻은 균주(유전적으로 또는 변형된)의 명칭 또는 식별자. 서열을 얻은 sub_strain
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>abis</INSDQualifier_value>
	참고	source 특징에서 strain 한정자를 동반하여야 한다. 부모 균주가 주어지지 않은 경우에 변형된 균주는 sub_strain 대신에 strain 한정자에 주석을 달아야 한다. 예를 들면, 값이 K-12 값을 지닌 strain 한정자 및 값이 MG1655 값을 지닌 substrain한정자 또는 값이 MG1655 값을 지닌 strain 한정자

6.74.	Qualifier	tag_peptide
	Definition	base location encoding the polypeptide for proteolysis tag of tmRNA and its termination codon
	Mandatory value format	<base_range> - where <base_range> provides the first and last base (separated by two dots) of the location for the proteolysis tag
	Example	<INSDQualifier_value>90..122</INSDQualifier_value>
	Comment	it is recommended that the amino acid sequence corresponding to the tag_peptide be annotated by describing a 5' partial CDS feature; e.g., CDS with a location of <90..122
6.75.	Qualifier	tissue_lib
	Definition	tissue library from which sequence was obtained
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>tissue library 772</INSDQualifier_value>
6.76.	Qualifier	tissue_type
	Definition	tissue type from which the sequence was obtained
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>liver</INSDQualifier_value>
6.77.	Qualifier	transl_except
	Definition	translational exception: single codon the translation of which does not conform to genetic code defined by organism or transl_table.
	Mandatory value format	(pos:<location>,aa:<amino_acid>) where <amino_acid> is the three letter abbreviation for the amino acid coded by the codon at the base_range position
	Example	<INSDQualifier_value>(pos:213..215,aa:Trp) </INSDQualifier_value> <INSDQualifier_value>(pos:462..464,aa:OTHER) </INSDQualifier_value> <INSDQualifier_value>(pos:1017,aa:TERM) </INSDQualifier_value> <INSDQualifier_value>(pos:2000..2001,aa:TERM) </INSDQualifier_value>
	Comment	if the amino acid is not one of the specific amino acids listed in Section 3 of this Annex, use OTHER as <amino_acid> and provide the name of the unusual amino acid in a note qualifier; for modified amino-acid selenocysteine use three letter abbreviation 'Sec' (one letter symbol 'U' in amino-acid sequence) for <amino_acid>; for modified amino-acid pyrrolysine use three letter abbreviation 'Pyl' (one letter symbol 'O' in amino-acid sequence) for <amino_acid>; for partial termination codons where TAA stop codon is completed by the addition of 3' A residues to the mRNA either a single base_position or a base_range is used for the location, see the third and fourth examples above, in conjunction with a note qualifier indicating 'stop codon completed by the addition of 3' A residues to the mRNA'.

6.74.	한정자	tag_peptide
	정의	tmRNA와 그의 종료 코돈의 단백질 분해(proteolysis) 태그(tag)를 위한 폴리펩티드를 코딩하는 기본 위치
	필수값 포맷	<base_range> - 여기서 <base_range>는 단백질 분해(proteolysis) 태그(tag) 의 첫 번째 및 마지막 염기의 위치(두 개의 점으로 구분)를 제공한다.
	예	<INSDQualifier_value>90..122</INSDQualifier_value>
	참고	5'-말단 부분 CDS 특징을 기술하여 tag_peptide에 대응하는 아미노산 서열에 주석을 달 것을 권장한다; 예: 위치가 <90..122>인 CDS.
6.75.	한정자	tissue_lib
	정의	서열을 얻은 조직 라이브러리
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>tissue library 772</INSDQualifier_value>
6.76.	한정자	tissue_type
	정의	서열을 얻은 조직 타입
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>liver</INSDQualifier_value>
6.77.	한정자	transl_except
	정의	번역 예외: 번역이 생물체 또는 transl_table에 의해 정의된 유전코드와 일치하지 않는 단일 코돈.
	필수값 포맷	(pos:<location>,aa:<amino_acid>) 여기서 <amino_acid>는 base_range 위치에서 코돈에 의해 암호화된 아미노산에 대한 3문자 약어이다.
	예	<INSDQualifier_value>(pos:213..215,aa:Trp)</INSDQualifier_value> <INSDQualifier_value>(pos:462..464,aa:OTHER)</INSDQualifier_value> <INSDQualifier_value>(pos:1017,aa:TERM)</INSDQualifier_value> <INSDQualifier_value>(pos:2000..2001,aa:TERM)</INSDQualifier_value>
	참고	아미노산이 본 별첨의 섹션 3에 나열된 특정 아미노산 중 하나가 아닌 경우, OTHER를 <amino_acid>로 사용하고 note 한정자에 특이한 아미노산의 명칭을 제공하십시오. 변형 아미노산 셀레노시스테인(selenocysteine)의 경우에, <amino_acid>에 대해 3 문자 약어 'Sec'(아미노산 서열에서 하나 문자 기호 'U')를 사용한다. 변형 아미노산 피롤리신(pyrrrolysine)의 경우, <amino_acid>에 대해 3 문자 약어 'Pyl'(아미노산 서열에서 하나 문자 기호 'O')을 사용하십시오. 3'-말단 A 잔기를 mRNA에 추가하여 TAA 정지 코돈이 완성되는 부분 종료 코돈의 경우, 단일 base-position 또는 base_range가 위치를 나타내기 위해 사용된다. 'mRNA에 3'-말단의 A 잔기들을 첨가하여 완료된 정지 코돈'을 나타내는 note 한정자와 함께 사용, 위의 세 번째 및 네 번째 예를 참조하십시오.

6.78. Qualifier	transl_table
Definition	definition of genetic code table used if other than universal or standard genetic code table. Tables used are described in this Annex
Mandatory value format	<integer> where <integer> is the number assigned to the genetic code table
Example	<INSDQualifier_value>3</INSDQualifier_value> - example where the yeast mitochondrial code is to be used
Comment	if the transl_table qualifier is not used to further annotate a CDS feature key, then the CDS is translated using the Standard Code (i.e. Universal Genetic Code). Genetic code exceptions outside the range of specified tables are reported in transl_except qualifiers.
6.79. Qualifier	trans_splicing
Definition	indicates that exons from two RNA molecules are ligated in intermolecular reaction to form mature RNA
Value format	none
Comment	should be used on features such as CDS, mRNA and other features that are produced as a result of a trans-splicing event. This qualifier must be used only when the splice event is indicated in the "join" operator, e.g., join(complement(69611..69724),139856..140087) in the feature location
6.80. Qualifier	translation
Definition	one-letter abbreviated amino acid sequence derived from either the standard (or universal) genetic code or the table as specified in a transl_table qualifier and as determined by an exception in the transl_except qualifier
Mandatory value format	contiguous string of one-letter amino acid abbreviations from Section 3 of this Annex, "X" is to be used for AA exceptions.
Example	<INSDQualifier_value>MASTFPWPYRGCSTPSLKGLIMCTW</INSDQualifier_value>
Comment	to be used with CDS feature only; must be accompanied by protein_id qualifier when the translation product contains four or more specifically defined amino acids; see transl_table for definition and location of genetic code Tables; only one of the qualifiers translation, pseudo and pseudogene are permitted to further annotate a CDS feature.
6.81. Qualifier	variety
Definition	variety (= varietas, a formal Linnaean rank) of organism from which sequence was derived.
Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
Example	<INSDQualifier_value>insularis</INSDQualifier_value>
Comment	use the cultivar qualifier for cultivated plant varieties, i.e., products of artificial selection; varieties other than plant and fungal varietas should be annotated via a note qualifier, e.g., with the value <INSDQualifier_value>breed:Cukorova</INSDQualifier_value>

6.78.	한정자	transl_table
	정의	범용 또는 표준 유전코드 표 이외의 경우에 사용되는 유전코드 표의 정의. 사용된 표는 이 별첨에 설명되어 있다.
	필수값 포맷	<integer> 여기에서 <integer>는 유전코드 표에 지정된 숫자이다.
	예	<INSDQualifier_value>3</INSDQualifier_value> - 효모 미토콘드리아 코드가 사용되는 예.
	참고	transl_table 한정자가 CDS 특징 기호를 추가로 정의하는데 사용되지 않는 경우에, CDS는 표준 코드(예: 범용 유전코드)를 사용하여 변환된다. 지정된 표 범위를 벗어난 유전코드 예외는 transl_except 한정자에 보고된다.
6.79.	한정자	trans_splicing
	정의	2개 RNA 분자로부터의 엑손이 분자간 반응에서 라이게이션(ligation)되어 성숙(mature) RNA를 형성함을 나타낸다.
	값 포맷	없음
	참고	CDS, mRNA와 같은 특징 및 trans-splicing 이벤트의 결과로 생성되는 다른 특징에 사용되어야 한다. 이 한정자는 splice 이벤트가 "join"연산자에 표시될 때만 사용되어야 한다(예: 특징 위치에서 join(complement(69611..69724),139856..140087)).
6.80.	한정자	translation
	정의	표준(또는 범용) 유전코드 또는 transl_table 한정자에 지정되고 transl_except 한정자에서 예외로 결정된 표에서 유래한 한 글자 아미노산 서열
	필수값 포맷	본 별첨의 섹션 3에서 한 글자로 된 연속적인 아미노산 약어이며, "X"는 AA 예외로 사용된다.
	예	<INSDQualifier_value>MASTFPPWYRGCASTPSLKGL IMCTW</INSDQualifier_value>
	참고	CDS 특징에만 사용된다. 번역(translation) 생산물(product)에 4개 이상의 구체적으로 정의된 아미노산이 포함된 경우에 protein_id 한정자를 동반하여야 한다. 유전코드 표의 정의 및 위치는 transl_table를 참조하십시오. 한정자 translation, pseudo 및 pseudogene 중 하나만 CDS 특징에 추가적으로 주석을 달 수 있다.
6.81.	한정자	variety
	정의	서열이 유래된 생물체의 다양성(= varietas, 공식적인 린네식(Linnaean) 분류 등급).
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>insularis</INSDQualifier_value>
	참고	재배 식물 품종, 즉 인공 개량 제품에 대해 cultivar 한정자를 사용하십시오. 식물 및 곰팡이 변이 이외의 품종은 note 한정자를 통해 주석을 달아야 한다(예: <INSDQualifier_value>breed:Cukorova</INSDQualifier_value> 값으로).

SECTION 7: FEATURE KEYS FOR AMINO ACID SEQUENCES

This section contains the list of allowed feature keys to be used for amino acid sequences. The feature keys are listed in alphabetic order.

7.1.	Feature Key	ACT_SITE
	Definition	Amino acid(s) involved in the activity of an enzyme
	Optional qualifiers	note
	Comment	Each amino acid residue of the active site must be annotated separately with the ACT_SITE feature key. The corresponding amino acid residue number must be provided as the location descriptor in the feature location element.
7.2.	Feature Key	BINDING
	Definition	Binding site for any chemical group (co-enzyme, prosthetic group, etc.). The chemical nature of the group is indicated in the note qualifier
	Mandatory qualifiers	note
	Comment	Examples of values for the "note" qualifier: "Heme (covalent)" and "Chloride." Where appropriate, the features keys CA_BIND, DNA_BIND, METAL, and NP_BIND should be used rather than BINDING.
7.3.	Feature Key	CA_BIND
	Definition	Extent of a calcium-binding region
	Optional qualifiers	note
7.4.	Feature Key	CARBOHYD
	Definition	Glycosylation site
	Mandatory qualifiers	note
	Comment	This key describes the occurrence of the attachment of a glycan (mono- or polysaccharide) to a residue of the protein. The type of linkage (C-, N- or O-linked) to the protein is indicated in the "note" qualifier. If the nature of the reducing terminal sugar is known, its abbreviation is shown between parentheses. If three dots '...' follow the abbreviation this indicates an extension of the carbohydrate chain. Conversely no dots means that a monosaccharide is linked. Examples of values used in the "note" qualifier: N-linked (GlcNAc...); O-linked (GlcNAc); O-linked (Glc...); C-linked (Man) partial; O-linked (Ara...).
7.5.	Feature Key	CHAIN
	Definition	Extent of a polypeptide chain in the mature protein
	Optional qualifiers	note

섹션 7: 아미노산 서열에 대한 특징 기호

이 섹션에는 아미노산 서열에 사용할 수 있는 특징 기호 목록이 포함되어 있다. 특징 기호는 알파벳 순서로 나열되어 있다.

7.1.	특징 기호	ACT_SITE
	정의	효소의 활성에 관여하는 아미노산
	선택 한정자	note
	참고	활성 부위의 각 아미노산 잔기는 ACT_SITE 특징 기호를 사용하여 별도로 주석을 달아야 한다. 해당 아미노산 잔기 번호는 특징 location 요소에서 location 설명자로 제공되어야 한다.
7.2.	특징 기호	BINDING
	정의	모든 화학 그룹(조효소, 보결분자족 등)의 결합 부위. 그룹의 화학적 성질은 note 한정자에 표시되어 있다.
	필수 한정자	note
	참고	"note" 한정자에 대한 값의 예: "Heme(covalent)" 및 "Chloride". 적절한 경우에, 특징 기호 CA_BIND, DNA_BIND, METAL 및 NP_BIND가 BINDING 대신 사용하는 것이 바람직하다.
7.3.	특징 기호	CA_BIND
	정의	칼슘 결합 영역의 범위
	선택 한정자	note
7.4.	특징 기호	CARBOHYD
	정의	당화(糖化) 부위
	필수 한정자	note
	참고	이 기호는 글리칸(단- 또는 다당류)이 단백질의 잔기에 부착된 경우를 나타낸다. 단백질에 대한 연결 타입(C-, N- 또는 O- 연결)은 "note" 한정자에 표시된다. 환원성 말단 당의 성질이 알려진 경우, 약어는 괄호 안에 표시된다. 약어에 3개의 점 '...'이 따르는 경우, 탄수화물 사슬이 확장되었음을 나타낸다. 반대로 점이 없다는 것은 단당류가 연결되어 있음을 의미한다. "note" 한정자에 사용된 값의 예: N-linked(GlcNAc...); O-linked(GlcNAc); O-linked(Glc.); C-linked Man) 부분; O-linked(Ara...).
7.5.	특징 기호	CHAIN
	정의	성숙(mature) 단백질에서 폴리펩티드 사슬의 고유 식별번호
	선택 한정자	note

7.6.	Feature Key	COILED
	Definition	Extent of a coiled-coil region
	Optional qualifiers	note
7.7.	Feature Key	COMPBIAS
	Definition	Extent of a compositionally biased region
	Optional qualifiers	note
7.8.	Feature Key	CONFLICT
	Definition	Different sources report differing sequences
	Optional qualifiers	note
	Comment	Examples of values for the "note" qualifier: Missing; K -> Q; GSDSE -> RIRLR; V -> A.
7.9.	Feature Key	CROSSLNK
	Definition	Post translationally formed amino acid bonds
	Mandatory qualifiers	note
	Comment	Covalent linkages of various types formed between two proteins (interchain cross-links) or between two parts of the same protein (intrachain cross-links); except for cross-links formed by disulfide bonds, for which the "DISULFID" feature key is to be used. For an interchain cross-link, the location descriptor in the feature location element is the residue number of the amino acid cross-linked to the other protein. For an intrachain cross-link, the location descriptor in the feature location element is the residue numbers of the cross-linked amino acids in "x..y" format, e.g. "42..50". The note qualifier indicates the nature of the cross-link; at least specifying the name of the conjugate and the identity of the two amino acids involved. Examples of values for the "note" qualifier: "Isoglutamyl cysteine thioester (Cys-Gln);" "Beta-methylanthionine (Cys-Thr);" and "Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin)"
7.10.	Feature Key	DISULFID
	Definition	Disulfide bond
	Mandatory qualifiers	note
	Comment	For an interchain disulfide bond, the location descriptor in the feature location element is the residue number of the cysteine linked to the other protein. For an intrachain cross-link, the location descriptor in the feature location element is the residue numbers of the linked cysteines in "x..y" format, e.g. "42..50". For interchain disulfide bonds, the note qualifier indicates the nature of the cross-link, by identifying the other protein, for example, "Interchain (between A and B chains)"
7.11.	Feature Key	DNA_BIND
	Definition	Extent of a DNA-binding region

7.6.	특징 기호	COILED
	정의	또 꼬인 나선(coiled-coil) 영역의 범위
	선택 한정자	note
7.7.	특징 기호	COMBIAS
	정의	구성적으로 편향된 영역의 범위
	선택 한정자	note
7.8.	특징 기호	CONFLICT
	정의	다른 소스는 다른 서열을 보고하는 경우 사용됨
	선택 한정자	note
	참고	"note" 한정자에 대한 값의 예: Missing; K -> Q; GSDSE -> RIRLR; V -> A.
7.9.	특징 기호	CROSSLNK
	정의	번역 후(Post translationally) 형성된 아미노산 결합
	필수 한정자	note
	참고	2개의 단백질(사슬간 가교 결합) 사이 또는 동일한 단백질 내의 2개 부분(사슬간 가교 결합) 사이에 형성된 다양한 타입의 공유 결합("DISULFID" 특징 기호가 사용되는 이황화 결합에 의해 형성된 가교를 제외). 사슬간 가교의 경우, 특징 location 요소의 location 서술자는 다른 단백질에 가교된 아미노산의 잔기 번호이다. 사슬간 가교의 경우, 특징 location 요소의 location 설명자는 "x.y" 형식(예 "42..50")의 가교된 아미노산의 잔기 번호이다. note 한정자는 가교 결합의 성질을 나타내며, 최소한 결합물의 이름과 관련된 두 아미노산의 잔기의 이름을 명시한다. "note" 한정자에 대한 값의 예: "이소글루타미드 시스테인 티오에스테르(Cys-Gln);" "베타-메틸란 티오닌(Cys-Thr);" 및 "글리실라이신 이소펩티드(Lys-Gly)(유비퀴틴의 G-Cter와의 가교)"
7.10.	특징 기호	DISULFID
	정의	이황화 결합
	필수 한정자	note
	참고	사슬간 이황화 결합의 경우, 특징 location 요소의 location 설명자는 다른 단백질에 연결된 시스테인 잔기 번호이다. 사슬내 가교의 경우, 특징 location 요소의 location 서술자는 연결된 시스테인(cysteine) 잔기 수를 "x.y" 형식(예: "42..50")으로 나타낸 것이다. 사슬간 이황화물 결합의 경우, note 한정자는 다른 단백질, 예를 들면 "사슬간(A 와 B 체인 간)"을 식별함으로써 가교 결합의 특성을 나타낸다.
7.11.	특징 기호	DNA_BIND
	정의	DNA 결합 영역의 범위

Mandatory qualifiers	note
Comment	The nature of the DNA-binding region is given in the note qualifier. Examples of values for the "note" qualifier: "Homeobox" and "Myb 2"
<hr/>	
7.12. Feature Key	DOMAIN
Definition	Extent of a domain, which is defined as a specific combination of secondary structures organized into a characteristic three-dimensional structure or fold
Mandatory qualifiers	note
Comment	The domain type is given in the note qualifier. Where several copies of a domain are present, the domains are numbered. Examples of values for the "note" qualifier: "Ras-GAP" and "Cadherin 1"
<hr/>	
7.13. Feature Key	HELIX
Definition	Secondary structure: Helices, for example, Alpha-helix; 3(10) helix; or Pi-helix
Optional qualifiers	note
Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a 'loop' or 'random-coil' structure.
<hr/>	
7.14. Feature Key	INIT_MET
Definition	Initiator methionine
Optional qualifiers	note
Comment	The location descriptor in the feature location element is "1". This feature key indicates the N-terminal methionine is cleaved off. This feature is not used when the initiator methionine is not cleaved off.
<hr/>	
7.15. Feature Key	INTRAMEM
Definition	Extent of a region located in a membrane without crossing it
Optional qualifiers	note
<hr/>	
7.16. Feature Key	LIPID
Definition	Covalent binding of a lipid moiety
Mandatory qualifiers	note
Comment	The chemical nature of the bound lipid moiety is given in the note qualifier, indicating at least the name of the lipidated amino acid. Examples of values for the "note" qualifier: "N-myristoyl glycine"; "GPI-anchor amidated serine" and "S-diacylglycerol cysteine."

	필수 한정자	note
	참고	DNA 결합 영역의 성질은 note 한정자에 나와 있다. "note" 한정자 값의 예: "Homeobox" 및 "Myb 2"
7.12.	특징 기호	DOMAIN
	정의	특징적인 3차원 구조 또는 집합으로 이루어진 2차 구조의 특정 조합으로 정의되는 영역의 범위
	필수 한정자	note
	참고	도메인 타입은 note 한정자에 나와 있다. 도메인의 사본이 여러 개인 경우에 도메인에 번호가 지정된다. "note" 한정자 값의 예: "Ras-GAP" 및 "Cadherin 1"
7.13.	특징 기호	HELIX
	정의	2차 구조: 나선, 예를 들면 알파-헬릭스(Alpha-helix); 3(10) 나선; 또는 Pi-나선
	선택 한정자	note
	참고	이 특징은 3차 구조가 알려진 단백질에만 사용된다. helix(기호 HELIX), beta-strands(기호 STRAND) 및 turns(기호 TURN)이라는 세 가지 타입의 보조 구조만 지정된다. 상기 클래스 중 어느 하나에 지정되지 않은 잔기는 '루프(loop)' 는 '랜덤 코일(random-coil)' 구조이다.
7.14.	특징 기호	INIT_MET
	정의	개시(initiator) 메티오닌(methione)
	선택 한정자	note
	참고	특징 location 요소의 location 설명자는 "1"이다. 이 특징 기호는 N-말단 메티오닌(methione)이 절단되었음을 나타낸다. 이 특징은 개시자(initiator) 메티오닌(methione)이 분리되지 않은 경우에는 사용되지 않는다.
7.15.	특징 기호	INTRAMEM
	정의	막(膜)을 가로 지르지 않고 막 안에 위치한 영역의 범위
	선택 한정자	note
7.16.	특징 기호	LIPID
	정의	지질 부분(moiety)의 공유 결합
	필수 한정자	note
	참고	결합된 지질 부분(moiety)의 화학적 성질은 적어도 한정된 지질화된 아미노산의 명칭을 나타내는 note 한정자에 주어진다. "note" 한정자에 대한 값의 예: "N-미리스토일 글리신(N-myristoyl glycine)"; "GPI-앵커 아미드화된 세린(GPI-anchor amidated serine)" 및 "S-디아실 글리세롤 시스테인(S-diacylglycerol cysteine)".

Ref.: Standards - ST.26

page: 3.26.i.49

7.17.	Feature Key	METAL
	Definition	Binding site for a metal ion.
	Mandatory qualifiers	note
	Comment	The note qualifier indicates the nature of the metal. Examples of values for the "note" qualifier: "Iron (heme axial ligand)" and "Copper".
7.18.	Feature Key	MOD_RES
	Definition	Posttranslational modification of a residue
	Mandatory qualifiers	note
	Comment	The chemical nature of the modified residue is given in the note qualifier, indicating at least the name of the post-translationally modified amino acid. If the modified amino acid is listed in Section 4 of this Annex, the abbreviation may be used in place of the the full name. Examples of values for the "note" qualifier: "N-acetylalanine"; "3-Hyp"; and "MeLys" or "N-6-methyllysine"
7.19.	Feature Key	MOTIF
	Definition	Short (up to 20 amino acids) sequence motif of biological interest
	Optional qualifiers	note
7.20.	Feature Key	MUTAGEN
	Definition	Site which has been experimentally altered by mutagenesis
	Optional qualifiers	note
7.21.	Feature Key	NON_STD
	Definition	Non-standard amino acid
	Optional qualifiers	note
	Comment	This key only describes the occurrence of non-standard amino acids selenocysteine (U) and pyrrolysine (O) in the amino acid sequence.
7.22.	Feature Key	NON_TER
	Definition	The residue at an extremity of the sequence is not the terminal residue
	Optional qualifiers	note
	Comment	If applied to position 1, this means that the first position is not the N-terminus of the complete molecule. If applied to the last position, it means that this position is not the C-terminus of the complete molecule.

7.17.	특징 기호	METAL
	정의	금속 이온의 결합 부위.
	필수 한정자	note
	참고	note 한정자는 금속의 성질을 나타낸다. "note" 한정자에 대한 값의 예: "철(heme axial ligand)" 및 "구리"
7.18.	특징 기호	MOD_RES
	정의	잔기의 번역 후 변형(Post-translational modification)
	필수 한정자	note
	참고	변형된 잔기의 화학적 성질은 적어도 번역후 변형 아미노산의 명칭을 나타내는 note 한정자에 주어진다. 변형 아미노산이 이 별첨의 섹션 4에 나열된 경우에, 약어가 전체 명칭 대신 사용될 수 있다. "note" 한정자에 대한 값의 예: "N-acetylalanine"; "3- 하이드"; 및 "MeLys" 또는 "N-6-methyllysine"
7.19.	특징 기호	MOTIF
	정의	생물학적으로 관심이 있는 짧은(최대 20개 아미노산) 서열 모티프(motif)
	선택 한정자	note
7.20.	특징 기호	MUTAGEN
	정의	돌연변이 유발에 의해 실험적으로 변형된 부위
	선택 한정자	note
7.21.	특징 기호	NON_STD
	정의	비표준 아미노산
	선택 한정자	note
	참고	이 기호는 단지 아미노산 서열에서 비표준 아미노산 셀레노시스테인(selenocysteine)(U) 및 롤리신(pyrrolysine)(O)의 발생을 설명한다.
7.22.	특징 기호	NON_TER
	정의	서열의 말단에 있지만 말단 잔기가 아님을 나타냄.
	선택 한정자	note
	참고	1번 위치에 적용하면 첫 번째 위치는 완전한 분자의 N-말단이 아니다. 마지막 위치에 적용되면, 이 위치가 완전한 분자의 C-말단이 아님을 의미한다.

Ref.: Standards - ST.26

page: 3.26.i.50

7.23.	Feature Key	NP_BIND
	Definition	Extent of a nucleotide phosphate-binding region
	Mandatory qualifiers	note
	Comment	The nature of the nucleotide phosphate is indicated in the note qualifier. Examples of values for the "note" qualifier: "ATP" and "FAD".
7.24.	Feature Key	PEPTIDE
	Definition	Extent of a released active peptide
	Optional qualifiers	note
7.25.	Feature Key	PROPEP
	Definition	Extent of a propeptide
	Optional qualifiers	note
7.26.	Feature Key	REGION
	Definition	Extent of a region of interest in the sequence
	Optional qualifiers	note
7.27.	Feature Key	REPEAT
	Definition	Extent of an internal sequence repetition
	Optional qualifiers	note
7.28.	Feature Key	SIGNAL
	Definition	Extent of a signal sequence (prepeptide)
	Optional qualifiers	note
7.29.	Feature Key	SITE
	Definition	Any interesting single amino-acid site on the sequence that is not defined by another feature key. It can also apply to an amino acid bond which is represented by the positions of the two flanking amino acids
	Mandatory qualifier	note
	Comment	When SITE is used to annotate a modified amino acid the value for the qualifier "note" must either be an abbreviation set forth in Section 4 of this Annex, or the complete, unabbreviated name of the modified amino acid.

7.23.	특징 기호	NP_BIND
	정의	핵산염기 인산염 결합 영역의 범위
	필수 한정자	note
	참고	핵산염기 인산염의 성질은 note 한정자에 표시되어 있다. "note" 한정자 값의 예: "ATP" 및 "FAD".
7.24.	특징 기호	PEPTIDE
	정의	방출된 활성 펩티드의 범위
	선택 한정자	note
7.25.	특징 기호	PROPEP
	정의	프로펩타이드(propeptide)의 범위
	선택 한정자	note
7.26.	특징 기호	REGION
	정의	서열에서 관심 영역의 범위
	선택 한정자	note
7.27.	특징 기호	REPEAT
	정의	내부 서열 반복 범위
	선택 한정자	note
7.28.	특징 기호	SIGNAL
	정의	신호 서열(프리펩티드, prepeptide)의 범위
	선택 한정자	note
7.29.	특징 기호	SITE
	정의	다른 특징 기호에 의해 정의되지 않은 서열상의 임의의 관심 단일 아미노산 부위. 두 개의 flanking 아미노산의 위치로 표현되는 아미노산 결합에도 적용할 수 있다.
	필수 한정자	note
	참고	SITE가 변형 아미노산에 주석을 달기 위해 사용될 경우, 한정자 "note"의 값은 본 별첨의 섹션 4에 제시된 약어이거나 변형 아미노산의 완전하고 축약되지 않는 명칭이어야 한다.

7.30.	Feature Key	source
	Definition	Identifies the source of the sequence; this key is mandatory; every sequence will have a single source feature spanning the entire sequence
	Mandatory qualifiers	mol_type organism
	Optional qualifiers	note
7.31.	Feature Key	STRAND
	Definition	Secondary structure: Beta-strand; for example Hydrogen bonded beta-strand or residue in an isolated beta-bridge
	Optional qualifiers	note
	Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a 'loop' or 'random-coil' structure.
7.32.	Feature Key	TOPO_DOM
	Definition	Topological domain
	Optional qualifiers	note
7.33.	Feature Key	TRANSMEM
	Definition	Extent of a transmembrane region
	Optional qualifiers	note
7.34.	Feature Key	TRANSIT
	Definition	Extent of a transit peptide (mitochondrion, chloroplast, thylakoid, cyanelle, peroxisome etc.)
	Optional qualifiers	note
7.35.	Feature Key	TURN
	Definition	Secondary structure Turns, for example, H-bonded turn (3-turn, 4-turn or 5-turn)
	Optional qualifiers	note
	Comment	This feature is used only for proteins whose tertiary structure is known. Only three types of secondary structure are specified: helices (key HELIX), beta-strands (key STRAND) and turns (key TURN). Residues not specified in one of these classes are in a 'loop' or 'random-coil' structure.

7.30.	특징 기호	source
	정의	서열의 소스를 식별한다. 이 기호는 필수이다. 모든 서열에는 전체 서열에 걸쳐 단일 source 제공되어야 함.
	필수 한정자	mol_type organism
	선택 한정자	note
7.31.	특징 기호	STRAND
	정의	2차 구조: 베타 가닥(Beta-strand); 예를 들면, 분리된 베타-브리지의 잔기 또는 수소 결합된 베타-가닥 잔기
	선택 한정자	note
	참고	이 특징은 3 차 구조가 알려진 단백질에만 사용된다. helices(기호 HELIX), Beta-strands(기호 STRAND) 및 turns(기호 TURN)의 세 가지 타입의 2차 구조만 지정된다. 이 클래스에 지정되지 않은 잔기는 '루프(loop)' 또는 '랜덤 코일(random-coil)' 구조이다.
7.32.	특징 기호	TOPO_DOM
	정의	위상 영역
	선택 한정자	note
7.33.	특징 기호	TRANSMEM
	정의	막 횡단 영역의 범위
	선택 한정자	note
7.34.	특징 기호	TRANSIT
	정의	수송(transit) 펩티드의 범위(미토콘드리아, 엽록체, 틸라코이드, 시아넬, 퍼옥시솜 등)
	선택 한정자	note
7.35.	특징 기호	TURN
	정의	2차 구조 회전, 예: H-본딩 턴(turn)(3-턴, 4-턴 또는 5-턴).
	선택 한정자	note
	참고	이 특징은 3차 구조가 알려진 단백질에만 사용된다. helices(기호 HELIX), Beta-strands(기호 STRAND) 및 turns(기호 TURN)의 세 가지 타입의 2차 구조만 지정된다. 이 클래스에 지정되지 않은 잔기는 '루프(loop)' 또는 '랜덤 코일(random-coil)' 구조이다.

Ref.: Standards - ST.26

page: 3.26.i.52

7.36. Feature Key	UNSURE
Definition	Uncertainties in the sequence
Optional qualifiers	note
Comment	Used to describe region(s) of an amino acid sequence for which the authors are unsure about the sequence presentation.

7.37. Feature Key	VARIANT
Definition	Authors report that sequence variants exist
Optional qualifiers	note

7.38. Feature Key	VAR_SEQ
Definition	Description of sequence variants produced by alternative splicing, alternative promoter usage, alternative initiation and ribosomal frameshifting
Optional qualifiers	note

7.39. Feature Key	ZN_FING
Definition	Extent of a zinc finger region
Mandatory qualifiers	note
Comment	The type of zinc finger is indicated in the note qualifier. For example: "GATA-type" and "NR C4-type"

7.36.	특징 기호	UNSURE
	정의	서열의 불확실성
	선택 한정자	note
	참고	저자들이 서열 제시에 대해 확신이 없는 아미노산 서열의 영역을 기술하는 데에 사용된다.
7.37.	특징 기호	VARIANT
	정의	저자들이 보고하는 서열 변이.
	선택 한정자	note
7.38.	특징 기호	VAR_SEQ
	정의	선택적 스플라이싱(alternative splicing), 선택적 프로모터(alternative promoter) 사용, 선택적 개시(initiation) 및 리보솜 프레임시프팅(frameshifting)에 의해 생성된 서열 변이에 대한 설명
	선택 한정자	note
7.39.	특징 기호	ZN_FING
	정의	징크핑거 영역의 범위
	필수 한정자	note
	참고	징크핑거의 타입은 note 한정자에 표시되어 있다. 예: "GATA-type" 및 "NR C4-type"

SECTION 8: QUALIFIERS FOR AMINO ACID SEQUENCES

This section contains the list of allowed qualifiers to be used for amino acid sequences.

Where the value format is free text that is identified as language-dependent, one of the following must be used:

- 1) the INSDQualifier_value element; or
- 2) the NonEnglishQualifier_value element; or
- 3) both the INSDQualifier_value element and the NonEnglishQualifier_value element.

Where the value format is not identified as language-dependent free text, the INSDQualifier_value element must be used and the NonEnglishQualifier_value element must not be used.

PLEASE NOTE: Any qualifier value provided for a qualifier with a "free text" value format may require translation for national or regional procedures. The qualifiers listed in the following table are considered to have language-dependent free text values:

Table 6: List of qualifiers for amino acid sequences with language-dependent free text values

Section	Language-Dependent Free Text Value
8.2	note
8.3	organism

8.1.	Qualifier	mol_type
	Definition	In vivo molecule type of sequence
	Mandatory value format	protein
	Example	<INSDQualifier_value>protein</INSDQualifier_value>
	Comment	The "mol_type" qualifier is mandatory on the source feature key.
8.2.	Qualifier	note
	Definition	Any comment or additional information
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>Heme (covalent)</INSDQualifier_value>
	Comment	The "note" qualifier is mandatory for the feature keys: BINDING; CARBOHYD; CROSSLNK; DISULFID; DNA_BIND; DOMAIN; LIPID; METAL; MOD_RES; NP_BIND; SITE and ZN_FING
8.3.	Qualifier	organism
	Definition	Scientific name of the organism that provided the peptide
	Mandatory value format	free text Language-dependent: this value may require translation for National/Regional procedures
	Example	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
	Comment	The "organism" qualifier is mandatory for the source feature key.

섹션 8: 아미노산 서열에 대한 한정자

이 섹션에는 아미노산 서열에 사용할 수 있는 한정자 목록이 포함되어 있다.

값 형식이 언어에 의존하여 식별되는 자유텍스트인 경우 다음 중 하나를 사용하여야 한다:

- 1) INSDQualifier_value element; 또는
- 2) NonEnglishQualifier_value element; 또는
- 3) INSDQualifier_value element와 NonEnglishQualifier_value 모두.

값 형식이 언어 의존적 자유텍스트로 식별되지 않는 경우, INSDQualifier_value 요소가 사용되어야 하며 NonEnglishQualifier_value 요소는 사용되어서는 안 된다.

참고: "자유텍스트"값 형식의 한정자에 제공된 한정자 값은 국가/지역 절차에 대한 번역이 필요할 수 있다. 다음 표에 나열된 한정자는 언어의존적 자유텍스트 값을 가지는 것으로 간주된다:

표 6: 언어의존적 자유텍스트 값을 가지는 아미노산 서열의 한정자 목록

섹션	언어의존적 자유텍스트 값
8.2	note
8.3	organism

8.1.	한정자	mol_type
	정의	서열의 생체내 분자 타입
	필수값 포맷	protein
	예	<INSDQualifier_value>protein</INSDQualifier_value>
	참고	"mol_type" 한정자는 source 특징 기호에서 필수이다.
8.2.	한정자	note
	정의	모든 의견 또는 추가 정보
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Heme(covalent)</INSDQualifier_value>
	참고	BINDING; CARBOHYD; CROSSLNK; ISULF ID; DNA_BIND; DOMAIN; LIPID; METAL; MOD_RES; NP_BIND; SITE 및 ZN_FING 특징 기호는 "note" 한정자가 필수임:
8.3.	한정자	organism
	정의	펩티드를 획득한 생물체의 학명
	필수값 포맷	자유텍스트 (언어의존적: 이 값은 국가/지역 절차에 대한 번역이 필요할 수 있다.)
	예	<INSDQualifier_value>Homo sapiens</INSDQualifier_value>
	참고	source 특징 기호에 "organism" 한정자는 필수임.

9 - Echinoderm and Flatworm Mitochondrial Code	
AAS =	FLLSSSSYY**CCWLLLLPPPHHQRRRIIMTTTTNKKSSSSVVVAAAADDEEGGG
Starts =	-----M-----M-----
Base1 =	ttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggg
Base2 =	ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
10 - Euplotid Nuclear Code	
AAS =	FLLSSSSYY**CCCWLLLLPPPHHQRRRIIMTTTTNKKSSRRVVVAAAADDEEGGG
Starts =	-----M-----
Base1 =	ttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggg
Base2 =	ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
11 – Bacterial, Archaeal, and Plant Plastid Code	
AAS =	FLLSSSSYY**CC*WLLLLPPPHHQRRRIIMTTTTNKKSSRRVVVAAAADDEEGGG
Starts =	---M-----M-----MMM-----M-----
Base1 =	ttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggg
Base2 =	ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
12 - Alternative Yeast Nuclear Code	
AAS =	FLLSSSSYY**CC*WLLLSPPPHHQRRRIIMTTTTNKKSSRRVVVAAAADDEEGGG
Starts =	-----M-----M-----
Base1 =	ttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggg
Base2 =	ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
13 – Ascidian Mitochondrial Code	
AAS =	FLLSSSSYY**CCWLLLLPPPHHQRRRIIMTTTTNKKSSGGVVVAAAADDEEGGG
Starts =	---M-----MM-----M-----
Base1 =	ttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggg
Base2 =	ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
14 - Alternative Flatworm Mitochondrial Code	
AAS =	FLLSSSSYYY*CCWLLLLPPPHHQRRRIIMTTTTNKKSSSSVVVAAAADDEEGGG
Starts =	-----M-----
Base1 =	ttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggg
Base2 =	ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag
16 - Chlorophycean Mitochondrial Code	
AAS =	FLLSSSSYY*LCC*WLLLSPPPHHQRRRIIMTTTTNKKSSRRVVVAAAADDEEGGG
Starts =	-----M-----
Base1 =	ttttttttttttttcccccccccccccccaaaaaaaaaaaaaaaaaaggggggggggggggg
Base2 =	ttttccccaaaaggggttttccccaaaaggggttttccccaaaaggggttttccccaaaaggg
Base3 =	tcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag

ANNEX II

DOCUMENT TYPE DEFINITION (DTD) FOR SEQUENCE LISTING

Version 1.3

*Revision approved by the Committee on WIPO Standards (CWS)
at its ninth session on November 5, 2021*

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Annex II of WIPO Standard ST.26, Document Type Definition (DTD) for Sequence Listing

This entity may be identified by the PUBLIC identifier:
*****
****
PUBLIC "-//WIPO//DTD SEQUENCE LISTING 1.3//EN" "ST26SequenceListing_V1_3.dtd"
*****
****
* PUBLIC DTD URL

* https://www.wipo.int/standards/dtd/ST26SequenceListing_V1_3.dtd
*****

* Revision of Annex II to WIPO Standard ST.26 was approved by the Committee on WIPO
* Standards (CWS) at its eighth session.

*****
* CONTACTS
*****
*
* xml.standards@wipo.int
*
*****
* NOTES
*****

* The sequence data part is a subset of the complete INSDC DTD V.1.5 that only covers
* the requirements of WIPO Standard ST.26.

*****
* REVISION HISTORY
*****
2021-11-05: Revised Version 1.3 approved at CWS/9 (small edits to the comments)
2020-05-20: Version 1.3 approved at CWS/8.

Changes:
- Optional originalFreeTextLanguageCode attribute added to <ST26SequenceListing> to allow
  applicants to indicate the language of the free text in the original sequence listing.
- Optional nonEnglishFreeTextLanguageCode attribute added to <ST26SequenceListing> to allow
  applicants to indicate the language of the free text provided in the element
  <NonEnglishQualifier_value>.
- Optional id attribute added to INSDQualifier to facilitate comparison of language-
  dependent qualifier values between sequence listings.
- Optional element <NonEnglishQualifier_value> added to element <INSDQualifier> to allow
  applicants to type language-dependent qualifiers in a non-English Language with the
  characters set forth in paragraph 40(a) of the ST.26 main body document.

2018-10-19: Version 1.2 approved at CWS/6.
Changes:
```

별첨 II

서열목록에 대한 문서 타입 정의(DTD)

버전 1.3

2021년 11월 5일 WIPO 표준위원회(CWS) 제9차 회의에서 승인된 개정판

```
<?xml version="1.0" encoding="UTF-8"?>
<!--WIPO 표준 ST.26의 별첨 II, 서열목록에 대한 DTD(Document Type)
```

이 개체(entity)는 PUBLIC 식별자로 식별될 수 있다:

```
*****
PUBLIC "-//WIPO//DTD SEQUENCE LISTING 1.3//EN" "ST26SequenceListing_V1_3.dtd"
*****
```

* PUBLIC DTD URL

* https://www.wipo.int/standards/dtd/ST26SequenceListing_V1_3.dtd

```
*****
```

WIPO 표준 ST.26에 대한 별첨 II의 개정은 WIPO 표준위원회(CWS) 6차 회의에서 승인되었다.

```
*****
```

* CONTACTS

```
*****
```

xml.standards@wipo.int

```
*****
```

* NOTES

```
*****
```

서열 데이터 부분은 WIPO 표준 ST.26의 요건만 다루는 완전한 INSDC DTD V.1.5의 부분 집합이다.

```
*****
```

* REVISION HISTORY

```
*****
```

2021-11-05: CWS/9에서 승인된 개정 버전 1.3(참고에 대한 약간의 수정)

2020-05-20: CWS/8에서 승인된 버전 1.3

변경 사항:

- 선택적 originalFreeTextLanguageCode 속성이 <ST26SequenceListing>에 추가되어, 출원인이 원 서열목록에서 자유텍스트의 언어를 제시할 수 있음.

- 선택적 nonEnglishFreeTextLanguageCode 속성이 <ST26SequenceListing>에 추가되어, 출원인이 <NonEnglishQualifier_value> 요소에 제공된 자유텍스트의 언어를 제시할 수 있음

- 선택적 id 속성이 INSDQualifier에 추가되어, 서열목록 간의 언어 의존적 한정자 값의 비교를 용이하게 함

- 선택적 요소 <NonEnglishQualifier_value>가 <INSDQualifier> 요소에 추가되어, 출원인이 ST.26 본문의 40(a)절에 명시된 문자로 비영어(non-English)로 된 언어 의존적 한정자를 입력할 수 있음

2018-10-19: CWS/6에서 승인된 버전 1.2.

변경사항:

<INSDQualifier*> changed to <INSDQualifier+> for alignment with business needs and advice from NCBI (an INSDFeature_qual element (if present) should have one or more INSDQualifier elements).

2017-06-02: Version 1.1 approved at the CWS/5

Changes:

Comments added to <INSDSeq_length>, <INSDSeq_division> and <INSDSeq_sequence> to clarify the reason of the differences between the INSDC DTD v.1.5 and ST26 Sequence Listing DTD V1_1.

2016-03-24: Version 1.0 adopted at the CWS/4Bis

2014-03-11: Final draft for adoption.

ST26SequenceListing

* ROOT ELEMENT

-->

<!ELEMENT ST26SequenceListing ((ApplicantFileReference | (ApplicationIdentification, ApplicantFileReference?)), EarliestPriorityApplicationIdentification?, (ApplicantName, ApplicantNameLatin?)?, (InventorName, InventorNameLatin?)?, InventionTitle+, SequenceTotalQuantity, SequenceData+)>

<!--The elements ApplicantName and InventorName are optional in this DTD to facilitate the conversion between various encoding schemes-->

<!--originalFreeTextLanguageCode:

The language code (see reference in paragraph 9 to ISO 639-1:2002) for the single original language in which the language-dependent free text qualifiers (NonEnglishQualifier_value) were prepared.

-->

<!--nonEnglishFreeTextLanguageCode:

The language code (see reference in paragraph 9 to ISO 639-1:2002) for the language in which the language-dependent free text qualifiers (NonEnglishQualifier_value) currently correspond.

-->

<!ATTLIST ST26SequenceListing

 dtdVersion CDATA #REQUIRED

 fileName CDATA #IMPLIED

 softwareName CDATA #IMPLIED

 softwareVersion CDATA #IMPLIED

 productionDate CDATA #IMPLIED

 originalFreeTextLanguageCode CDATA #IMPLIED

 nonEnglishFreeTextLanguageCode CDATA #IMPLIED

>

<!--ApplicantFileReference

Applicant's or agent's file reference, mandatory if application identification not provided.

-->

<!ELEMENT ApplicantFileReference (#PCDATA)>

<!--ApplicationIdentification

Application identification for which the sequence listing is submitted, when available.

-->

<!ELEMENT ApplicationIdentification (IPOfficeCode, ApplicationNumberText, FilingDate?)>

<!--EarliestPriorityApplicationIdentification

Identification of the earliest priority application, which contains IPOfficeCode, ApplicationNumberText and FilingDate elements.

-->

<!ELEMENT EarliestPriorityApplicationIdentification (IPOfficeCode, ApplicationNumberText, FilingDate?)>

<!--ApplicantName

The name of the first mentioned applicant in characters set forth in paragraph 40 (a) of the ST.26 main body document.

-->

비즈니스 필요와 NCBI의 권고안에 따라서 <INSDQualifier*>를 <INSDQualifier+>로 변경함 (INSDFeature_qual 요소(있는 경우)는 하나 이상의 INSDQualifier 요소를 가져야 함).

2017-06-02: CWS/5에서 승인된 버전 1.1

변경사항:

INSDC DTD v.1.5와 ST26 Sequence Listing DTD V1_1 사이의 차이 이유를 명확하게 하기 위해서 <INSDSeq_length>, <INSDSeq_division> 및 <INSDSeq_sequence>에 코멘트가 추가됨.

2016-03-24: CWS/4Bis에서 채택된 버전 1.0

2014-03-11: 채택을 위한 최종 초안.

ST26SequenceListing

* ROOT ELEMENT

-->

<!ELEMENT ST26SequenceListing ((ApplicantFileReference | (ApplicationIdentification, ApplicantFileReference?)), EarliestPriorityApplicationIdentification?, (ApplicantName, ApplicantNameLatin?)), (InventorName, InventorNameLatin?)), InventionTitle+, SequenceTotalQuantity, SequenceData+)>

<!-- ApplicantName 및 InventorName 요소는 이 DTD에서 선택 사항으로, 다양한 암호화(encoding) 체계 간의 변환을 용이하게 한다 -->

<!--originalFreeTextLanguageCode:

언어 의존적 자유텍스트 한정자에서 단일 원(original) 언어에 대한 언어 코드(9절 ISO 639-1:2002 참조)

-->

<!--nonEnglishFreeTextLanguageCode:

언어 의존적 자유텍스트 한정자(NonEnglishQualifier_value)가 현재 대응하는 언어 코드(9절 ISO 639-1:2002 참조)

-->

<!ATTLIST ST26SequenceListing

dtdVersion CDATA #REQUIRED

fileName CDATA #IMPLIED

softwareName CDATA #IMPLIED

softwareVersion CDATA #IMPLIED

productionDate CDATA #IMPLIED

originalFreeTextLanguageCode CDATA #IMPLIED

nonEnglishFreeTextLanguageCode CDATA #IMPLIED

>

<!--ApplicantFileReference

출원인 또는 대리인의 파일 참조, 출원 식별이 제공되지 않은 경우 필수임.

-->

<!ELEMENT ApplicantFileReference (#PCDATA)>

<!--ApplicationIdentification

가능한 경우, 서열목록이 제출되는 것에 대한 출원 식별.

-->

<!ELEMENT ApplicationIdentification (IPOfficeCode, ApplicationNumberText, FilingDate?)>

<!--EarliestPriorityApplicationIdentification

IPOfficeCode, ApplicationNumberText 및 FilingDate 요소를 포함하는 최초 우선권 주장 출원의 식별.

-->

<!ELEMENT EarliestPriorityApplicationIdentification (IPOfficeCode, ApplicationNumberText, FilingDate?)>

<!--ApplicantName

ST.26 본문 문서의 40(a) 절에 명시된 문자로 처음 언급된 출원인의 이름.

-->

```
<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of
names of languages - Part 1: Alpha-2
-->
<!ELEMENT ApplicantName (#PCDATA)>
<!ATTLIST ApplicantName
    languageCode CDATA #REQUIRED
>
<!--ApplicantNameLatin
Where ApplicantName is typed in characters other than those as set forth in paragraph
40(b), a translation or transliteration of the name of the first mentioned applicant must
also be typed in characters as set forth in paragraph 40(b) of the ST.26 main body
document.
-->
<!ELEMENT ApplicantNameLatin (#PCDATA)>
<!--InventorName
Name of the first mentioned inventor typed in the characters as set forth in paragraph
40(a).-->
<!--languageCode: Appropriate language code from ISO 639-1-Codes for the representation of
names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventorName (#PCDATA)>
<!ATTLIST InventorName
    languageCode CDATA #REQUIRED
>
<!--InventorNameLatin
Where InventorName is typed in characters other than those as set forth in paragraph 40(b),
a translation or transliteration of the first mentioned inventor may also be typed in
characters as set forth in paragraph 40(b).
-->
<!ELEMENT InventorNameLatin (#PCDATA)>
<!--InventionTitle
Title of the invention typed in the characters as set forth in paragraph 40(a) in the
language of filing. A translation of the title of the invention into additional languages
may be typed in the characters as set forth in paragraph 40(a) using additional
InventionTitle elements. The title of invention should be between two to seven words.
-->
<!--languageCode: Appropriate language code from ISO 639-1 - Codes
for the representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventionTitle (#PCDATA)>
<!ATTLIST InventionTitle
    languageCode CDATA #REQUIRED
>
<!--SequenceTotalQuantity
Indicates the total number of sequences in the document.
Its purpose is to be quickly accessible for automatic processing.
-->
<!ELEMENT SequenceTotalQuantity (#PCDATA)>
<!--SequenceData
Data for individual Sequence.
For intentionally skipped sequences see the ST.26 main body document.
-->
<!ELEMENT SequenceData (INSDSeq)>
<!ATTLIST SequenceData
    sequenceIDNumber CDATA #REQUIRED
>
<!--IPOfficeCode
ST.3 code. For example, if the application identification is PCT/IB2013/099999, then
IPOfficeCode value will be "IB" for the International Bureau of WIPO.
-->
<!ELEMENT IPOfficeCode (#PCDATA)>
<!--ApplicationNumberText
The application identification as provided by the office of filing (e.g. PCT/IB2013/099999)
-->
<!ELEMENT ApplicationNumberText (#PCDATA)>
<!--FilingDate
```

<!--languageCode: 언어 명칭의 표현을 위해서 ISO 639-1-Codes에서의 적절한 언어 코드 - Part 1: Alpha-2

-->
<!ELEMENT ApplicantName (#PCDATA)>
<!ATTLIST ApplicantName
 languageCode CDATA #REQUIRED

>
<!--ApplicantNameLatin

>
<!--InventorNameLatin
ApplicantName이 40 (b)절에 명시된 것과 다른 문자로 입력되는 경우, 첫 번째 언급된 출원인 이름의 번역 또는 음역은 ST.26 문서 본문
40 (b)절에 명시된 문자로도 입력하여야 한다.

-->
<!ELEMENT ApplicantNameLatin (#PCDATA)>
<!--InventorName
40 (a)절에 명시된 문자로 입력된 첫 번째 발명자의 이름

-->
<!--languageCode: 언어 명칭의 표현을 위해서 ISO 639-1-Codes에서의 적절한 언어 코드 - Part 1: Alpha-2

-->
<!ELEMENT InventorName (#PCDATA)>
<!ATTLIST InventorName
 languageCode CDATA #REQUIRED

ApplicantName이 40 (b)절에 명시된 것과 다른 문자로 입력되는 경우, 첫 번째 언급된 출원인 이름의 번역 또는 음역은 ST.26 문서 본문
40 (b)절에 명시된 문자로도 입력하여야 한다.

-->
<!ELEMENT InventorNameLatin (#PCDATA)>
<!--InventionTitle

출원의 언어로 40 (a)절에 명시된 문자로 입력된 발명의 명칭. 본 발명의 명칭을 추가 언어로 번역하는 것은 추가적인 발명 명칭 요소를
사용하여 40 (a)절에 명시된 문자로 입력될 수 있다. 발명의 타이틀은 2개에서 7개 사이의 단어를 사용하여야 한다.

-->
<!--languageCode: 언어 명칭의 표현을 위해서 ISO 639-1-Codes에서의 적절한 언어 코드 - Part 1: Alpha-2

-->
<!ELEMENT InventionTitle (#PCDATA)>
<!ATTLIST InventionTitle
 languageCode CDATA #REQUIRED

>
<!--SequenceTotalQuantity
문서의 총 서열 수를 제시한다.
이것의 목적은 자동적인 처리를 위해 빠르게 액세스할 수 있도록 하는 것이다.

-->
<!ELEMENT SequenceTotalQuantity (#PCDATA)>
<!--SequenceData
개별 서열에 대한 데이터.
의도적으로 생략된 서열은 ST.26 문서 본문을 참조하십시오.
-->
<!ELEMENT SequenceData (INSDSeq)>
<!ATTLIST SequenceData
 sequenceIDNumber CDATA #REQUIRED

>
<!--IPOfficeCode
ST.3 코드. 예를 들면, 출원서 식별이 PCT/IB2013/099999인 경우, IPOfficeCode 값은 WIPO 국제 사무국의 "IB"가 된다.

-->
<!ELEMENT IPOfficeCode (#PCDATA)>
<!--ApplicationNumberText
출원 관청에서 제공되는 출원서 식별 (예: PCT/IB2013/099999)

-->
<!ELEMENT ApplicationNumberText (#PCDATA)>
<!--FilingDate

The date of filing of the patent application for which the sequence listing is submitted in ST.2 format "CCYY-MM-DD", using a 4-digit calendar year, a 2-digit calendar month and a 2-digit day within the calendar month, e.g., 2015-01-31. For details, please see paragraphs 7 (a) and 11 of WIPO Standard ST.2.

```
-->
<!ELEMENT FilingDate (#PCDATA)>
<!--*****
* INSD Part
*****-->
```

The purpose of the INSD part of this DTD is to define a customized DTD for sequence listings to support the work of IP offices while facilitating the data exchange with the public repositories.

The INSD part is subset of the INSD DTD v1.5 and as such can only be used to generate an XML instance as it will not support the complete INSD structure.

This part is based on:

The International Nucleotide Sequence Database (INSD) collaboration.

INSDSeq provides the elements of a sequence as presented in the GenBank/EMBL/DDBJ-style flatfile formats. Not all elements are used here.

```
-->
<!--INSDSeq
Sequence data. Changed INSD V1.5 DTD elements, INSDSeq_division and INSDSeq_sequence from
optional to mandatory per business requirements.
-->
<!ELEMENT INSDSeq (INSDSeq_length, INSDSeq_moltype, INSDSeq_division, INSDSeq_other-
seqids?, INSDSeq_feature-table?, INSDSeq_sequence)>
<!--INSDSeq_length
The length of the sequence. INSDSeq_length allows only integer.
-->
<!ELEMENT INSDSeq_length (#PCDATA)>
<!--INSDSeq_moltype
Admissible values: DNA, RNA, AA
-->
<!ELEMENT INSDSeq_moltype (#PCDATA)>
<!--INSDSeq_division
Indication that a sequence is related to a patent application. Must be populated with the
value PAT.
-->
<!ELEMENT INSDSeq_division (#PCDATA)>
<!--INSDSeq_other-seqids
In the context of data exchange with database providers, the IPOs should populate for each
sequence the element INSDSeq_other-seqids with one INSDSeqid containing a reference to the
corresponding published patent and the sequence identification.
-->
<!ELEMENT INSDSeq_other-seqids (INSDSeqid?)>
<!--INSDSeq_feature-table
Information on the location and roles of various regions within a particular sequence.
Whenever the element INSDSeq_feature-table is used, it must contain at least one feature.
-->
<!ELEMENT INSDSeq_feature-table (INSDFeature+)>
<!--INSDSeq_sequence
The residues of the sequence. The sequence must not contain numbers, punctuation or
whitespace characters.
-->
<!ELEMENT INSDSeq_sequence (#PCDATA)>
<!--INSDSeqid
Intended for the use of IPOs in data exchange only.
```

Format:

```
pat|{office code}|{publication number}|{document kind code}|{Sequence identification
number}
```

달력에서 4자리 연도, 2자리 월 및 2자리 날짜를 사용하여, ST.2 형식 "CCYY-MM-DD"로 제출된 서열목록에 대한 특허출원의 출원일(예: 2015-01-31). 자세한 내용은 7절 및 WIPO 표준 ST.2의 11을 참조하십시오.

```
-->
<!ELEMENT FilingDate (#PCDATA)>
<!--*****
* INSD Part
*****
```

본 DTD의 INSD 부분의 목적은 공용 저장소(repository)에서 데이터 교환을 용이하게 하면서 특허 관청(IP office)의 업무를 지원하기 위해서 개별요구에 맞춘(customized) DTD를 정의하는 것이다.

INSD 부분은 INSD DTD v1.5의 부분집합이며, 완전한 INSD 구조를 지원하지 않으므로 XML 인스턴스를 생성하는 데만 사용할 수 있다.

이 부분은 다음을 기반으로 한다.

INSD(International Nucleotide Sequence Database) 공동 작업.

INSDSeq는 GenBank/EMBL/DDBJ 양식의 플랫폼파일(Flatfile) 형식으로 표시되는 서열의 요소들을 제공한다. 여기에서는 모든 요소가 사용되는 것은 아니다.-->

```
<!--INSDSeq
서열 데이터. INSD V1.5 DTD 요소와 INSDSeq_division 및 INSDSeq_sequence는 비즈니스 요건에 따라 선택에서 필수로 변경됨.-->

<!ELEMENT INSDSeq (INSDSeq_length, INSDSeq_moltype, INSDSeq_division, INSDSeq_other-seqids?, INSDSeq_feature-table?,
INSDSeq_sequence)> <!--
INSDSeq_length
```

서열의 길이. INSDSeq_length는 정수만 허용한다.

```
-->
<!ELEMENT INSDSeq_length (#PCDATA)>
<!--INSDSeq_moltype
허용되는 값: DNA, RNA, AA
-->
<!ELEMENT INSDSeq_moltype (#PCDATA)>
<!--INSDSeq_division
```

서열이 특허출원과 관련되어 있음을 나타낸다. PAT 값으로 채워져야 한다.

```
-->
<!ELEMENT INSDSeq_division (#PCDATA)>
<!--INSDSeq_other-seqids
```

데이터베이스 제공자와 데이터 교환과 관련하여, IPO는 각 서열에 대한 것을 공개 특허에 대한 참조와 서열 식별을 포함하는 하나의 INSDSeqid를 가진 INSDSeq_other-seqids 요소로 채워야 한다.

```
-->
<!ELEMENT INSDSeq_other-seqids (INSDSeqid?)>
<!--INSDSeq_feature-table
```

특정 서열 내에서 다양한 지역의 위치 및 역할에 대한 정보. INSDSeq_feature-table 요소가 사용될 때마다 하나 이상의 특징을 포함하여야 한다.

```
-->
<!ELEMENT INSDSeq_feature-table (INSDFeature+)>
<!--INSDSeq_sequence
```

서열의 잔기. 서열은 숫자, 구두점 또는 공백 문자를 포함하지 않아 한다.

```
-->
<!ELEMENT INSDSeq_sequence (#PCDATA)>
<!--INSDSeqid
```

데이터 교환 시에 한하여 IPO의 사용을 위한 것이다.

Format:

```
pat|{office code}|{publication number}|{document kind code}|{Sequence identification number}
```


where office code is the code of the IP office publishing the patent document, publication number is the publication number of the application or patent, document kind code is the letter codes to distinguish patent documents as defined in ST.16 and Sequence identification number is the number of the sequence in that application or patent

Example:

```
pat|WO|2013999999|A1|123456
```

This represents the 123456th sequence from WO patent publication No. 2013999999 (A1)

```
-->
<!ELEMENT INSDSeqid (#PCDATA)>
<!--INSDFeature
Description of one feature.
-->
<!ELEMENT INSDFeature (INSDFeature_key, INSDFeature_location, INSDFeature_qual?)>
<!--INSDFeature_key
A word or abbreviation indicating a feature.
-->
<!ELEMENT INSDFeature_key (#PCDATA)>
<!--INSDFeature_location
Region of the presented sequence which corresponds to the feature.
-->
<!ELEMENT INSDFeature_location (#PCDATA)>
<!--INSDFeature_qual
List of qualifiers containing auxiliary information about a feature.
-->
<!ELEMENT INSDFeature_qual (INSDQualifier+)>
<!--INSDQualifier
Additional information about a feature.
For coding sequences and variants see the ST.26 main body document.
-->
<!--id
Unique identifier for the INSDQualifier to facilitate comparison of versions of a sequence
listing specifically having language-dependent qualifier values in different languages.
-->
<!ELEMENT INSDQualifier (INSDQualifier_name, INSDQualifier_value?,
NonEnglishQualifier_value?)>
<!ATTLIST INSDQualifier
                id ID #IMPLIED
>
<!--INSDQualifier_name
Name of the qualifier.
-->
<!ELEMENT INSDQualifier_name (#PCDATA)>
<!--INSDQualifier_value
Value of the qualifier. Where the qualifier is language-dependent its value must be in the
English language and typed with the characters set forth in paragraph 40 (b).
-->
<!ELEMENT INSDQualifier_value (#PCDATA)>
<!--NonEnglishQualifier_value
Value of a language-dependent qualifier in a language that is not English and typed with
the characters set forth in paragraph 40 (a). The language is indicated with the attribute
nonEnglishFreeTextLanguageCode.
-->
<!ELEMENT NonEnglishQualifier_value (#PCDATA)>
```

[Annex III follow s]

여기에서, office code(관청코드)는 특허 문헌을 공고하는 IP 청의 코드이고, publication number(공고 번호)는 출원 또는 특허의 공고 번호이며, document kind code(문서 종류 코드)는 ST.16에 정의된 특허 문헌을 구별하기 위한 문자 코드이며, Sequence identification number(서열식별번호)는 해당 출원 또는 특허의 서열 번호이다.

Example:

```
pat|WO|2013999999|A1|123456
```

이는 WO 특허 공개번호 2013999999 (A1)의 123456 번째 서열을 표현한다.

```
-->
```

```
<!ELEMENT INSDSeqid (#PCDATA)>
```

```
<!--INSDFeature
```

하나의 특징에 대한 설명.

```
-->
```

```
<!ELEMENT INSDFeature (INSDFeature_key, INSDFeature_location, INSDFeature_qual?)>
```

```
<!--INSDFeature_key
```

특징을 나타내는 단어 또는 약어.

```
-->
```

```
<!ELEMENT INSDFeature_key (#PCDATA)>
```

```
<!--INSDFeature_location
```

특징에 대응하는 제시된 서열의 영역.

```
-->
```

```
<!ELEMENT INSDFeature_location (#PCDATA)>
```

```
<!--INSDFeature_qual
```

특징에 대한 보조 정보가 포함된 한정자 목록.

```
-->
```

```
<!ELEMENT INSDFeature_qual (INSDQualifier+)>
```

```
<!--INSDQualifier
```

특징에 대한 추가 정보.

코딩 서열 및 변이에 대해서는 ST.26 본문 문서를 참조하십시오.

```
-->
```

```
<!--id
```

다른 언어로 언어 의존적인 한정자 값을 가진 서열목록 버전 간의 비교를 용이하게 하기 위한 INSDQualifier에 대한 고유한 식별자

```
-->
```

```
<!ELEMENT INSDQualifier (INSDQualifier_name, INSDQualifier_value?, NonEnglishQualifier_value?)>
```

```
<!ATTLIST INSDQualifier
```

```
id ID #IMPLIED
```

```
>
```

```
<!--INSDQualifier_name
```

한정자의 명칭.

```
-->
```

```
<!ELEMENT INSDQualifier_name (#PCDATA)>
```

```
<!--INSDQualifier_value
```

한정자의 값. 한정자가 언어 의존적인 경우 그 값은 영어로 되어야 하며, 40(b)절에 규정된 문자로 입력되어야 한다.

```
-->
```

```
<!ELEMENT INSDQualifier_value (#PCDATA)>
```

```
<!--NonEnglishQualifier_value
```

영어가 아니며 40(a)절에 규정된 문자로 입력된 언어 의존적인 한정자의 값. 그 언어는 nonEnglishFreeTextLanguageCode 속성으로 제시된다.

```
-->
```

```
<!ELEMENT NonEnglishQualifier_value (#PCDATA)>
```

[별첨 III으로 이어짐]

ANNEX III

SEQUENCE LISTING SPECIMEN (XML file)

Version 1.4

*Revision approved by the Committee on WIPO Standards (CWS)
at its ninth session on November 5, 2021*

The Annex III is available at: https://www.wipo.int/standards/en/xml_material/st26/st26-annex-iii-sequence-listing-specimen.xml

[Annex IV follow s]

별첨 III

서열목록 견본(XML 파일)

버전 1.4

2021년 11월 5일 WIPO 표준위원회(CWS) 제9차 회의에서 승인된 개정판

별첨 III는 https://www.wipo.int/standards/en/xml_material/st26/st26-annex-iii-sequence-listing-specimen.xml 에서 검색할 수 있습니다.

[ST.26의 별첨 IV로 이어짐]

ANNEX IV

CHARACTER SUBSET FROM THE UNICODE BASIC LATIN CODE TABLE
FOR USE IN AN XML INSTANCE OF A SEQUENCE LISTING

Version 1.3

*Revision approved by the Committee on WIPO Standards (CWS)
at its eighth session on December 4, 2020*

The ampersand character (0026) is only permitted as part of a predefined entity. The quotation mark (0022), the apostrophe (0027), the less-than sign (003C), and the greater-than sign (003E) must be represented by their predefined entities. In addition, the ampersand character (0026) must be represented by its predefined entity when used as an ampersand in a value of an attribute or content of an element.

Unicode code point	Character	Name
0020		SPACE
0021	!	EXCLAMATION MARK
0022	"	QUOTATION MARK
0023	#	NUMBER SIGN
0024	\$	DOLLAR SIGN
0025	%	PERCENT SIGN
0026	&	AMPERSAND
0027	'	APOSTROPHE
0028	(LEFT PARENTHESIS
0029)	RIGHT PARENTHESIS
002A	*	ASTERISK
002B	+	PLUS SIGN
002C	,	COMMA
002D	-	HYPHEN-MINUS
002E	.	FULL STOP
002F	/	SOLIDUS
0030	0	DIGIT ZERO
0031	1	DIGIT ONE
0032	2	DIGIT TWO
0033	3	DIGIT THREE
0034	4	DIGIT FOUR
0035	5	DIGIT FIVE
0036	6	DIGIT SIX
0037	7	DIGIT SEVEN
0038	8	DIGIT EIGHT
0039	9	DIGIT NINE
003A	:	COLON
003B	;	SEMICOLON
003C	<	LESS-THAN-SIGN
003D	=	EQUALS SIGN
003E	>	GREATER-THAN-SIGN
003F	?	QUESTION MARK
0040	@	COMMERCIAL AT
0041	A	LATIN CAPITAL LETTER A
0042	B	LATIN CAPITAL LETTER B
0043	C	LATIN CAPITAL LETTER C
0044	D	LATIN CAPITAL LETTER D
0045	E	LATIN CAPITAL LETTER E
0046	F	LATIN CAPITAL LETTER F
0047	G	LATIN CAPITAL LETTER G
0048	H	LATIN CAPITAL LETTER H
0049	I	LATIN CAPITAL LETTER I

별첨 IV

서열목록의 XML 인스턴스 사용을 위해
유니코드 기본 라틴 코드 표 유래의 문자 서브세트

버전 1.3

2020년 12월 4일 WIPO 표준위원회(CWS) 제8차 회의에서 승인된 개정판

앰퍼샌드 문자(0026)는 사전 정의된 개체의 일부로만 허용된다. 인용 부호(0022), 아포스트로피(0027), (작음 부호(003C) 및 큼 부호(003E)는 사전에 정의된 개체로 표현되어야 한다. 또한, 앰퍼샌드 문자(0026)는 속성값 또는 요소의 내용에서 앰퍼샌드로 사용될 때 사전에 정의된 개체로 표현되어야 한다.

유니 코드 코드 포인트	캐릭터	명칭
0020		여백
0021	!	느낌표
0022	"	인용 부호
0023	#	숫자 부호
0024	\$	달러 표시
0025	%	퍼센트 부호
0026	&	앰퍼샌드
0027	'	아포스트로피
0028	(왼쪽 괄호
0029)	오른쪽 괄호
002A	*	별표
002B	+	더하기 기호
002C	,	coma
002D	-	하이픈 - 마이너스
002E	.	마침표
002F	/	슬래시
0030	0	숫자 0
0031	1	숫자 1
0032	2	숫자 2
0033	3	숫자 3
0034	4	숫자 4
0035	5	숫자 5
0036	6	숫자 6
0037	7	숫자 7
0038	8	숫자 8
0039	9	숫자 9
003A	:	콜론
003B	;	세미콜론
003C	<	작음 부호
003D	=	등호
003E	>	큼 부호
003F	?	물음표
0040	@	상업적 "at"
0041	A	라틴 대문자 A
0042	B	라틴 대문자 B
0043	C	라틴 대문자 C
0044	D	라틴 대문자 D
0045	E	라틴 대문자 E
0046	F	라틴 대문자 F
0047	G	라틴 대문자 G
0048	H	라틴 대문자 H
0049	I	라틴 대문자 I

Unicode code point	Character	Name
004A	J	LATIN CAPITAL LETTER J
004B	K	LATIN CAPITAL LETTER K
004C	L	LATIN CAPITAL LETTER L
004D	M	LATIN CAPITAL LETTER M
004E	N	LATIN CAPITAL LETTER N
004F	O	LATIN CAPITAL LETTER O
0050	P	LATIN CAPITAL LETTER P
0051	Q	LATIN CAPITAL LETTER Q
0052	R	LATIN CAPITAL LETTER R
0053	S	LATIN CAPITAL LETTER S
0054	T	LATIN CAPITAL LETTER T
0055	U	LATIN CAPITAL LETTER U
0056	V	LATIN CAPITAL LETTER V
0057	W	LATIN CAPITAL LETTER W
0058	X	LATIN CAPITAL LETTER X
0059	Y	LATIN CAPITAL LETTER Y
005A	Z	LATIN CAPITAL LETTER Z
005B	[LEFT SQUARE BRACKET
005C	\	REVERSE SOLIDUS
005D]	RIGHT SQUARE BRACKET
005E	^	CIRCUMFLEX ACCENT
005F	_	LOW LINE
0060	`	GRAVE ACCENT
0061	a	LATIN SMALL LETTER A
0062	b	LATIN SMALL LETTER B
0063	c	LATIN SMALL LETTER C
0064	d	LATIN SMALL LETTER D
0065	e	LATIN SMALL LETTER E
0066	f	LATIN SMALL LETTER F
0067	g	LATIN SMALL LETTER G
0068	h	LATIN SMALL LETTER H
0069	i	LATIN SMALL LETTER I
006A	j	LATIN SMALL LETTER J
006B	k	LATIN SMALL LETTER K
006C	l	LATIN SMALL LETTER L
006D	m	LATIN SMALL LETTER M
006E	n	LATIN SMALL LETTER N
006F	o	LATIN SMALL LETTER O
0070	p	LATIN SMALL LETTER P
0071	q	LATIN SMALL LETTER Q
0072	r	LATIN SMALL LETTER R
0073	s	LATIN SMALL LETTER S
0074	t	LATIN SMALL LETTER T
0075	u	LATIN SMALL LETTER U
0076	v	LATIN SMALL LETTER V
0077	w	LATIN SMALL LETTER W
0078	x	LATIN SMALL LETTER X
0079	y	LATIN SMALL LETTER Y
007A	z	LATIN SMALL LETTER Z
007B	{	LEFT CURLY BRACKET
007C		VERTICAL LINE
007D	}	RIGHT CURLY BRACKET
007E	~	TILDE

[Annex V follows]

유니 코드 코드 포인트	캐릭터	명칭
004A	J	라틴 대문자 J
004B	K	라틴 대문자 K
004C	L	라틴 대문자 L
004D	M	라틴 대문자 M
004E	N	라틴 대문자 N
004F	O	라틴 대문자 O
0050	P	라틴 대문자 P
0051	Q	라틴 대문자 Q
0052	R	라틴 대문자 R
0053	S	라틴 대문자 S
0054	T	라틴 대문자 T
0055	U	라틴 대문자 U
0056	V	라틴 대문자 V
0057	W	라틴 대문자 W
0058	X	라틴 대문자 X
0059	Y	라틴 대문자 Y
005A	Z	라틴 대문자 Z
005B	[왼쪽 사각형 브래킷
005C	\	역슬래시
005D]	오른쪽 사각형 브래킷
005E	^	곡절 악센트(CIRCUMFLEX ACCENT)
005F	_	로우 라인
0060	`	그레이브 악센트
0061	a	라틴어 소문자 A
0062	b	라틴어 소문자 B
0063	c	라틴어 소문자 C
0064	d	라틴어 소문자 D
0065	e	라틴어 소문자 E
0066	f	라틴어 소문자 F
0067	g	라틴어 소문자 G
0068	h	라틴어 소문자 H
0069	i	라틴어 소문자 I
006A	j	라틴어 소문자 J
006B	k	라틴어 소문자 K
006C	l	라틴어 소문자 L
006D	m	라틴어 소문자 M
006E	n	라틴어 소문자 N
006F	o	라틴어 소문자 O
0070	p	라틴어 소문자 P
0071	q	라틴어 소문자 Q
0072	r	라틴어 소문자 R
0073	s	라틴어 소문자 S
0074	t	라틴어 소문자 T
0075	u	라틴어 소문자 U
0076	v	라틴어 소문자 V
0077	w	라틴어 소문자 W
0078	x	라틴어 소문자 X
0079	y	라틴어 소문자 Y
007A	z	라틴어 소문자 Z
007B	{	왼쪽 쉼표 브래킷
007C		수직선
007D	}	오른쪽 쉼표 브래킷
007E	~	틸데

ANNEX V

ADDITIONAL DATA EXCHANGE REQUIREMENTS (FOR IPOs ONLY)

Version 1.4

*Revision approved by the Committee on WIPO Standards (CWS)
at its ninth session on November 5, 2021*

In the context of data exchange with database providers (INSD members), the Intellectual Property Offices should populate for each sequence the element `INSDSeq_other-seqids` with one `INSDSeqid` containing a reference to the corresponding published patent and the sequence identification number in the following format:

`pat{{office code}}{publication number}{{document kind code}}{sequence identification number}`

where office code is the code of the IP office publishing the patent document as set forth in ST.3; document kind code is the code for the identification of different kinds of patent documents as set forth in ST.16; publication number is the publication number of the application or patent; and Sequence identification number is the number of the sequence in that application or patent.

Example:

`pat|WO|2013999999|A1|123456`

Which would be translated into a valid XML instance as:

```
<INSDSeq_other-seqids>  
  <INSDSeqid>pat|WO|2013999999|A1|123456</INSDSeqid>  
</INSDSeq_other-seqids>
```

Where "123456" is the 123456th sequence from the WO publication no. 2013999999 (A1).

[Annex VI follows]

별첨 V

추가 데이터 교환 요건(특허청 전용)

버전 1.4

2021년 11월 5일 WIPO 표준위원회(CWS) 제9차 회의에서 승인된 개정판

데이터베이스 제공자(INSD 멤버)와의 데이터 교환과 관련하여, 특허 관청은 각 서열마다 INSDSeq_other-seqids 요소에 해당하는 공개 특허 및 서열식별번호에 대한 참조를 포함하는 하나의 다음 형식의 INSDSeqid를 채워야 한다.

pat{|office code|}{publication number|}{document kind code|}{sequence identification number}

여기서 office code(관청 코드)는 ST.3에 기술된 특허 문헌을 게시하는 IP office의 코드이다. Document kind code(문서 종류 코드)는 ST.16에 기술된 다양한 종류의 특허 문헌을 식별하기 위한 코드이다. Publication number(공개번호)는 출원 또는 특허의 공개번호이며; 그리고 sequence identification number(서열식별번호)는 그 출원 또는 특허에서의 서열 번호이다.

예:

pat|WO|2013999999|A1|123456

다음과 같이 유효한 XML 인스턴스로 변환된다.

```
<INSDSeq_other-seqids>  
  <INSDSeqid>pat|WO|2013999999|A1|123456</INSDSeqid>  
</INSDSeq_other-seqids>
```

여기에서 "123456"은 WO 공개번호 2013999999(A1)로부터 123456번째 서열이다.

[별첨 VI로 이어짐]

ANNEX VI

GUIDANCE DOCUMENT WITH ILLUSTRATED EXAMPLES

Version 1.5

*Revision approved by the Committee on WIPO Standards (CWS)
at its ninth session on November 5, 2021*

TABLE OF CONTENTS

INTRODUCTION.....	3.26.vi.1
EXAMPLE INDEX.....	3.26.vi.8
EXAMPLES.....	3.26.vi.18
APPENDIX.....	3.26.vi.76

INTRODUCTION

This Standard indicates as one of its purposes, to “allow applicants to draw up a single sequence listing in a patent application acceptable for the purposes of both international and national or regional procedures.” The purpose of this Guidance Document is to ensure that all applicants and Intellectual Property Offices (IPOs) understand and agree on the requirements for inclusion and representation of sequence disclosures, such that this purpose is realized.

This guidance document consists of this introduction, an example index, examples of sequence disclosures, and an appendix containing a sequence listing in XML with sequences from the examples. This introduction explains certain concepts and terminology used in the remainder of this document. The examples illustrate the requirements of specific paragraphs of the Standard and each example has been designated with the most relevant paragraph number. Some examples further illustrate other paragraphs and appropriate cross-references are indicated at the end of each example. The index provides page numbers for the examples and any indicated cross-references. Each sequence in an example that either must or may be included in a sequence listing has been assigned a sequence identification number (SEQ ID NO) and appears in XML format in the [Appendix](#) to this document.

For each example, any explanatory information presented with a sequence is intended to be considered as the entirety of the disclosure concerning that sequence. The given answers take into account only the information explicitly presented in the example.

The guidance provided in this document is directed to the preparation of a sequence listing for provision on the filing date of a patent application. Preparation of a sequence listing for provision subsequent to the filing date of a patent application must take into account whether the information provided could be considered by an IPO to add subject matter to the original disclosure. Therefore, it is possible that the guidance provided in this document may not be applicable to a sequence listing provided subsequent to the filing date of a patent application.

Preparation of a sequence listing

Sequence listing preparation for a patent application requires consideration of the following questions:

1. Does ST.26 paragraph 7 require inclusion of a particular disclosed sequence?
2. If inclusion of a particular disclosed sequence is not required, is inclusion of that sequence permitted by ST.26?
3. If inclusion of a particular disclosed sequence is required or permitted by ST.26, how should that sequence be represented in the sequence listing?

Regarding the first question, ST.26 paragraph 7 (with certain restrictions) requires inclusion of a sequence disclosed in a patent application by enumeration of its residues, where the sequence contains ten or more specifically defined nucleotides or four or more specifically defined amino acids.

Regarding the second question, ST.26 paragraph 8 prohibits inclusion of any sequences having fewer than ten specifically defined nucleotides or four specifically defined amino acids.

별첨 VI

실시예를 포함한 지침서

버전 1.5

2021년 11월 5일 WIPO 표준위원회(CWS) 제9차 회의에서 승인된 개정판

목차

서론.....	3.26.vi.1
예제 색인.....	3.26.vi.8
예제.....	3.26.vi.18
부록.....	3.26.vi.76

서론

이 표준의 목적 중 하나는 "출원인이 국내외 및 지역 절차의 목적에 부합하는 특허출원에서 단일의 서열목록을 작성할 수 있게 하는" 것으로 되어있다. 이러한 목적을 달성하기 위하여 본 지침 문서에서는 모든 출원인과 특허청(Intellectual Property Offices)이 서열 개시 포함 및 표현의 요건을 제대로 이해하고 동의할 수 있도록 하는 것을 그 목적으로 하고 있다.

이 지침 문서는 본 서론, 예제 색인, 서열 개시 예제, 예제에 나온 서열과 함께 XML 형식으로 된 서열목록을 포함하는 부록으로 구성되어 있다. 본 서론에서는 이 문서의 나머지 부분에서 사용되는 특정 개념과 용어에 대해 설명한다. 예제는 본 표준의 특정 절(paragraph)의 요건을 설명하는데, 가장 관련성이 높은 절의 번호가 각 예제에 지정되었다. 일부 예제는 다른 절을 추가로 설명하기도 하며, 적절한 상호 참조는 각 예제의 끝에 표시된다. 색인에서는 예제와 표시된 상호 참조에 대한 페이지 번호가 제공된다. 예제에서 서열목록에 포함되거나 포함될 수 있는 서열에는 각각 서열식별번호(SEQ ID NO)가 지정되었으며 이 문서의 부록에 XML 형식으로 표시해 두었다.

각 예제에서, 서열과 함께 제시된 설명 정보는 모두 그 목적이 해당 서열에 대한 개시 내용 전체로 이해되도록 하였다. 제공된 답에서는 예제에서 명시적으로 주어진 정보만 고려하였다.

본 문서에 제공된 지침은 특허출원일에 제공하기 위한 서열목록 작성의 준비 방법을 제공해 준다. 특허출원일 이후에 제공하기 위한 서열목록 작성을 준비할 때는, 최초 공개 내용에 주제를 추가하기 위하여 특허청이 제공한 정보를 고려할 수 있을지도 생각해봐야 한다. 따라서, 본 문서에 제공된 지침은 특허출원일 이후에 제공된 서열목록에는 적용되지 않을 수 있다.

서열목록 준비

특허출원을 위한 서열목록을 준비하려면 다음 질문을 고려하여야 한다.

1. ST.26의 7절에 특정하게 개시된 서열을 요구하는가?
2. 만약 개시된 특정 서열의 포함이 필요하지 않는 경우, 해당 서열의 포함이 ST.26에 따라 허용되는 것인가?
3. 만약 개시된 특정 서열이 ST.26에 따라 요구되거나 허용되는 경우, 서열목록에서 해당 서열을 어떻게 표현해야 하는가?

첫 번째 질문과 관련하여, ST.26의 7절(특정 제한사항 있음)에 따르면, 특허출원서에 구체적으로 정의된 핵산염기 10개 이상 또는 구체적으로 정의된 아미노산 4개 이상을 함유하는 서열의 잔기를 나열하는 방식으로 개시되어야 한다.

두 번째 질문과 관련하여, ST.26의 8절에 따르면, 어느 서열이든지 그 서열이 함유한 '구체적으로 정의된' 핵산염기가 10개 미만이거나 아미노산이 4개 미만일 경우 서열목록에의 포함이 금지된다.

A clear understanding of “enumeration of its residues” and “specifically defined” is necessary to answer these two questions.

Regarding the third question, this document provides sequence disclosures which exemplify a variety of scenarios together with a complete discussion of the preferred means of representation of each sequence, or where a sequence contains multiple variations - the “most encompassing sequence”, in accordance with this Standard. Since it is impossible to address every possible unusual sequence scenario, this guidance document attempts to set forth the reasoning behind the approach to each example and the manner in which ST.26 provisions are applied, such that the same reasoning can be applied to other sequence scenarios not exemplified.

Enumeration of its residues

ST.26 paragraph 3(c) defines “enumeration of its residues” as disclosure of a sequence in a patent application by listing, in order, each residue of the sequence, wherein (i) the residue is represented by a name, abbreviation, symbol, or structure; or (ii) multiple residues are represented by a shorthand formula. A sequence should be disclosed in a patent application by “enumeration of its residues” using conventional symbols, which are the nucleotide symbols set forth in Section 1, Table 1 of ST.26 Annex 1 (i.e., the lower case symbols or their upper case equivalents¹) and the amino acid symbols set forth in Section 3, Table 3 of ST.26 Annex 1 (i.e., the upper case symbols or their lower case equivalents¹). Hereinafter, these nucleotide and amino acid symbols are referred to as conventional symbols. Representations of nucleotides and amino acids that are other than those set forth in these tables are hereinafter referred to as “nonconventional”.

Where a representation of a residue is disclosed as equivalent to a conventional symbol or abbreviation (e.g., “Z₁” means “A”), or to a specific sequence of conventional symbols (e.g., “Z₁” means “agga”), then the sequence is interpreted as though it were disclosed using the equivalent conventional symbol(s) or abbreviation(s), to determine whether ST.26 paragraph 7 requires inclusion in the sequence listing or whether paragraph 8 prohibits inclusion. Where a nonconventional nucleotide symbol is used as an ambiguity symbol (e.g., X₁ = inosine or pseudouridine), but is not equivalent to one of the conventional ambiguity symbols in Section 1, Table 1 (i.e., “m”, “r”, “w”, “s”, “y”, “k”, “v”, “h”, “d”, “b”, or “n”), then the residue is interpreted as an “n” residue to determine whether ST.26 Paragraph 7 requires inclusion of the sequence in the sequence listing or whether ST.26 Paragraph 8 prohibits inclusion. Similarly, where a nonconventional amino acid symbol is used as an ambiguity symbol (e.g., “Z₁” means “A”, “G”, “S” or “T”), but is not equivalent to one of the conventional ambiguity symbols in Section 3, Table 3 (i.e., B, Z, J, or X), then the residue is interpreted as an “X” residue to determine whether ST.26 paragraph 7 requires inclusion of the sequence in the sequence listing or whether ST.26 paragraph 8 prohibits inclusion.

Care should be taken to disclose sequences using conventional symbols; however, where sequences are otherwise disclosed, it may be necessary to consult the disclosure for an explanation to determine the meaning of the nonconventional representation.

Where a conventional symbol is used, the explanation of the sequence in the disclosure must still be consulted to confirm that the symbol is used in a conventional manner. If the symbol is used in a nonconventional manner, this explanation is necessary to determine whether ST.26 paragraph 7 requires inclusion of the sequence in the sequence listing or whether paragraph 8 prohibits inclusion.

Specifically defined

ST.26 paragraph 3(k) defines “specifically defined” as any nucleotide other than those represented by the symbol “n” and any amino acid other than those represented by the symbol “X”, listed in Annex I, wherein “n” and “X” are used in a conventional manner as described in Section 1, Table 1 (i.e., “a or c or g or t/u; ‘unknown’ or ‘other’”) and Section 3, Table 3 (i.e., “A or R or N or D or C or Q or E or G or H or I or L or K or M or F or P or O or S or U or T or W or Y or V; ‘unknown’ or ‘other’”), respectively. The discussion above concerning conventional symbols or nonconventional symbols or abbreviations and their use in a conventional or nonconventional manner will be taken into account to determine whether a nucleotide or an amino acid is “specifically defined”.

¹ NOTE: While an application disclosure may represent nucleotides or amino acids with either lower case or upper case symbols, for a sequence included in a sequence listing, only lower case letters must be used for representation of a nucleotide sequence (see ST.26 paragraph 13) and only upper case letters must be used for representation of an amino acid sequence (see ST.26 paragraph 26).

사실 위의 두 가지 질문에 답하기 위해서는 “잔기의 나열”과 “구체적으로 정의된”에 대한 명확한 이해가 필요하다.

세 번째 질문과 관련하여, 본 문서에서는 다음과 같은 서열 개시를 제공한다. 각 서열을 표현할 때 선호되는 방법에 대해 최대한 논의하면서, 다양한 경우의 수를 예제로 제시할 것이고, 하나의 서열이 다수의 변이형을 포함하는 경우, 본 표준에 부합하는 ‘가장 포괄적인 서열’을 개시할 것이다. 모든 특이한 서열의 모든 가능한 경우의 수를 다룰 수는 없기 때문에, 본 지침 문서에서는 각 예제의 접근법에 사용된 논리와, 예제로 포함되지 않은 다른 서열의 경우 적용할 수 있도록, ST.26의 조항이 적용되는 방식을 제시하고자 했다.

잔기의 나열

ST.26의 3(c)절에 따르면 “잔기의 나열”이란 특허출원에서 각 서열의 (i) 잔기를 명칭, 약어, 기호, 구조로 표현하거나, (ii) 잔기 여러 개를 축약식으로 표현하여 순서대로 목록화하는 방식으로 이루어진 서열 개시로 정의된다. 서열은 특허출원에서 ST.26의 별첨 1의 1절, 표 1에 제시된 핵산염기 기호(즉, 소대문자 기호¹) 및 ST.26의 별첨 1의 3절, 표 3에 제시된 아미노산 기호(즉, 대소문자 기호¹) 등 통상적인 기호를 사용하여 ‘잔기를 나열’로 개시하여야 한다. 아래에서는 이들 핵산 및 아미노산 심볼을 ‘통상적인(conventional)’ 심볼 symbols라고 부른다. 위의 표에 제시된 것들 이외의 핵산 및 아미노산 표현들은 아래에서는 “비통상적(nonconventional)”이라고 부른다.

잔기의 표현이 통상적인 기호나 약어와 동등한 것으로 개시된 경우(예: “Z₁”은 “A”를 의미), 또는 통상적인 기호의 특정 서열에 해당하는 것으로 기재된 경우(예: “Z₁”은 “agga”를 의미), 그 서열은 ST.26의 7절에 따라 서열목록에 포함되어야 하는지, 8항에 따라 포함해서는 안되는지 여부를 결정하기 위해서 마치 그에 해당하는 통상적인 기호-약어를 사용하여 개시된 것처럼 해석할 수 있다. 비통상적 핵산염기 기호가 다중핵산염기 기호로 쓰였지만(예: X₁ = 이노신 또는 슈도유리딘) 제1절, 표1의 통상적인 다중핵산염기 기호(즉, “m”, “r”, “w”, “s”, “y”, “k”, “v”, “h”, “d”, “b”또는 “n”) 중 하나에 해당하지 않는 경우, ST.26의 7절에 따라 서열목록에 포함되어야 하는지, 아니면 8항에 따라 포함해서는 안되는지를 결정하기 위해서 해당 잔기를 ‘n’잔기로 해석한다. 이와 비슷하게, 비통상적 아미노산 기호가 다중아미노산 기호로 쓰였지만(예: “Z₁”은 “A”, “G”, “S”또는 “T”를 의미) 제3절 표3의 통상적인 다중아미노산 기호(즉, B, Z, J 또는 X) 중 하나에 해당하지 않는 경우, ST.26의 7절에 따라 서열목록에 포함되어야 하는지, 아니면 8항에 따라 포함해서는 안되는지를 결정하기 위해서 해당 잔기를 ‘X’잔기로 해석한다.

통상적인 방식 방식으로 서열이 개시된 경우 주의가 요구된다. 그러나, 서열이 다른 방식으로 공개되는 경우, 설명에 비통상적 표현을 의미하는 것으로 결정하는 것은 협의과정이 필요할 수 있다.

통상적인 기호가 사용되는 경우도, 이 기호가 통상적인 방식으로 사용된 것을 확정하기 위해 개시된 서열에 대한 설명에 대한 협의가 여전히 필요하다. 기호가 비통상적인 방식으로 사용된 경우, ST.26의 7절에 따라 서열목록에 포함되어야 하는지, 아니면 8항에 따라 포함해서는 안되는지 결정하기 위한 설명이 필요하다.

구체적 정의

ST.26의 3(k)항에서 “구체적으로 정의된”란 기호 ‘n’으로 표현되지 않은 핵산염기와, 기호 ‘X’로 표현되지 않은 아미노산 일체를 말하는데, 이는 별첨 1에 나열되어 있다. 여기서 ‘n’과 ‘X’는 섹션 1, 표 1(즉, “a 또는 c 또는 g 또는 t/u; ‘불명’ 또는 ‘기타’)이나, 섹션 3, 표 3(즉, “A 또는 R 또는 N 또는 D 또는 C 또는 Q 또는 E 또는 G 또는 H 또는 I 또는 L 또는 K 또는 M 또는 F 또는 P 또는 O 또는 S 또는 U 또는 T 또는 Y 또는 V; ‘불명’ 또는 ‘기타’)”에서와 같이 통상적인 방식으로 사용된다. 위에서 논의된 통상적인 기호 또는 비통상적 기호나 약어 및 그들의 비통상적 방식의 사용 여부는 어느 핵산염기나 아미노산이 “구체적으로 정의된” 것인가를 결정할 때 고려될 것이다.

1. 출원서를 개시할 때는 대소문자 중 한 쪽을 자유롭게 사용하여 핵산염기나 아미노산을 표현할 수 있지만, 서열목록에 포함된 서열의 경우, 핵산염기 서열의 표현에는 소문자만 사용해야 하고(ST.26의 13 절 참조), 아미노산 서열의 표현에는 대문자만 사용해야 한다.(ST.26의 26 절 참조)

Most encompassing sequence

Where a sequence that meets the requirements of paragraph 7 is disclosed by enumeration of its residues only once in an application, but is described differently in multiple embodiments, e.g., one embodiment "X" in one or more locations could be any amino acid, but in further embodiments, "X" could be only a limited number of amino acids, ST.26 requires inclusion in a sequence listing of only the single sequence that has been enumerated by its residues. As per paragraphs 15 and 27, where such a sequence contains multiple "n" or "X" ambiguity symbols, "n" or "X" is construed to represent any nucleotide or amino acid, respectively, in the absence of further annotation. Consequently, the single sequence required to be included is the most encompassing sequence disclosed. The most encompassing sequence is the single sequence having variant residues that are represented by the most restrictive ambiguity symbols that include the most disclosed embodiments. Likewise, where a sequence is disclosed by enumeration of its residues only once, but the length of the sequence may vary due to copy number variation, the longest embodiment of the sequence is considered the most encompassing sequence. For example, consider a sequence containing a repeated region that can vary from 2 to 5 copies as enumerated. The embodiment with 5 copies of the repeat is the most encompassing sequence and should be included in the sequence listing. However, inclusion of additional specific sequences is strongly encouraged where practical, e.g., those that represent additional embodiments that are a key part of the invention. Inclusion of the additional sequences allows for a more thorough search and provides public notice of the subject matter for which a patent is sought.

Usage of Ambiguity Symbol

Proper Usage of the Ambiguity Symbol "n" in a Sequence Listing

The symbol "n"

- a. must not be used to represent anything other than a single nucleotide;
- b. will be construed as any one of "a", "c", "g", or "t/u" except where it is used with a further description;
- c. should be used to represent any of the following nucleotides together with a further description:
 - i. modified nucleotide, e.g., natural, synthetic, or non-naturally occurring, that cannot otherwise be represented by any other symbol in Annex I (see Section 1, Table 1);
 - ii. "unknown" nucleotide, i.e., not determined, not disclosed, or unsure;
 - iii. an abasic site; or
- d. may be used to represent a sequence variant, i.e., alternatives, deletions, insertions, or substitutions, where "n" is the most restrictive ambiguity symbol.

Proper Usage of the Ambiguity Symbol "X" in a Sequence Listing

The symbol "X"

- a. must not be used to represent anything other than a single amino acid;
- b. will be construed as any one of "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", or "V", except where it is used with a further description;
- c. should be used to represent any of the following amino acids together with a further description:
 - i. modified amino acid, e.g., natural, synthetic, or non-naturally occurring, that cannot otherwise be represented by any other symbol in Annex I (see Section 3, Table 3);
 - ii. "unknown" amino acid, i.e., not determined, not disclosed, or unsure; or
- d. may be used to represent a sequence variant, i.e., alternatives, deletions, insertions, or substitutions, where "X" is the most restrictive ambiguity symbol.

Annotation of Modified Residues

This Standard requires that "modified" residues are annotated per paragraph 17 for nucleotides, and per paragraph 30 for amino acids.

가장 포괄적인 서열

7절의 요건을 충족하는 서열이 한 출원서에서 단 한번 그 잔기를 나열하여 개시되었으나, 다수의 실시예에서 그와 다르게 기술되어 있을 경우, 예컨대 한 실시예에서 하나 이상의 경우 "X"가 불특정한 아미노산 전부가 될 수 있지만, 그 외의 실시예에서 "X"는 제한된 수의 아미노산만이 될 수 있는 경우, ST.26에서는 잔기로 나열된 단일한 서열만을 서열목록에 포함시킬 것을 요구한다. 15절, 27절에 따라, 이들 서열이 다수의 "n"이나 "X"라는 다중(ambiguity) 기호를 포함한 경우, "n"이나 "X"는 추가 주석이 없을 경우 각각 임의의 핵산염기 및 아미노산을 표현하는 것으로 해석될 수 있다. 따라서, 포함되어야 할 단일 서열은 개시된 중에 가장 포괄적인 서열이다. 가장 포괄적인 서열은 다양한 잔기를 가지는 단일 서열로, 가장 많이 개시된 실시예를 포함하는 가장 제한적인 다중(ambiguity) 기호로 표현한다. 비슷하게, 서열이 잔기들의 단 한번 나열에 의해서 개시되면, 복제수변이(copy number variation)에 따라서 그 서열의 길이는 다양할 수 있다. 그 가장 많이포괄하는 서열을 고려하여 구체적으로 가장 긴 서열이어야 한다. 예를 들면, 서열이 2개에서 5개 반복이 다양한 반복부위를 포함하는 경우, 5개의 반복으로 구체화하는 것이 가장 많이 포괄하는 서열이며 이를 서열목록에 포함하여야 한다. 그러나, 추가적인 특정 서열의 포함은 실시 가능한 경우, 예컨대 발명의 핵심부인 추가적인 실시예를 표현하는 경우 강력히 권장된다. 추가 서열을 포함하면 더 철저한 조사가 가능하고, 특허가 청구된 주제를 공지할 수 있다.

다중핵산 및 아미노산 기호의 사용

서열목록에서 다중염기 기호 "n"의 적절한 사용법

기호 "n"은:

- a. 단일 핵산염기 이외의 것을 표현하는 데에 사용되어서는 아니된다.
- b. 추가 설명과 함께 사용되는 경우 외에는 "a", "c", "g" 또는 "t/u"중 하나로 해석한다.
- c. 추가 설명과 함께 다음 핵산염기 중 임의의 것을 표현하는 데에 사용되어야 한다:
 - i. 별첨 I의 다른 기호(1절, 표 1 참조)로 표시될 수 없는 변형 핵산염기, 예; 자연적, 합성된, 또는 비자연적 발생된 것;
 - ii. "불명" 핵산염기, 즉 파악되지 않았거나, 공개되지 않았거나, 또는 불확실한 것;
 - iii. 탈염기 부위(abasic site); 또는
- d. 서열 변이, 즉 대체, 삭제, 삽입, 또는 치환을 표현하는 데에 "n"이 가장 제한적인 다중염기 기호인 경우 사용될 수 있다

서열목록에서 다중아미노산 기호 "X"의 적절한 사용법

기호 "X"는:

- a. 단일 아미노산 이외의 것을 표현하는 데에 사용해서는 아니된다.
- b. 추가 설명과 함께 사용되는 경우를 제외하고, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중의 하나로 해석될 것이다.
- c. 추가 설명과 함께 다음 아미노산 중 하나를 표현하는 데에 사용하여야 한다.
 - i. 별첨 I에서 다른 기호(3절, 표 3 참조)로 표현될 수 없는 변형 아미노산, 예; 자연적, 합성된, 또는 비자연적 발생된 것
 - ii. "불명" 아미노산, 즉 파악되지 않았거나, 공개되지 않았거나, 불확실한; 또는
- d. 서열 변이, 즉 대체, 삭제, 삽입, 또는 치환을 표현하는 데에 "X"가 가장 제한적인 다중아미노산 기호인 경우

변형 잔기의 주석(annotation)

본 표준에 따르면 "변형된(modified)" 잔기는 핵산은 17절 그리고 아미노산은 30절에 따라 주석이 필요하다.

ST.26 paragraph 3(e) defines “modified amino acid” as any amino acid as described in paragraph 3(a) other than L-alanine, L-arginine, L-asparagine, L-aspartic acid, L-cysteine, L-glutamine, L-glutamic acid, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-pyrrolysine, L-serine, L-selenocysteine, L-threonine, L-tryptophan, L-tyrosine, or L-valine. Similarly, the Standard defines “modified nucleotide” as any nucleotide as described in paragraph 3(g) other than deoxyadenosine 3'-monophosphate, deoxyguanosine 3'-monophosphate, deoxycytidine 3'-monophosphate, deoxythymidine 3'-monophosphate, adenosine 3'-monophosphate, guanosine 3'-monophosphate, cytidine 3'-monophosphate, or uridine 3'-monophosphate (ST.26, paragraph 3(f)).

Based on the definitions above, modifications to the nucleobases or sugar-phosphate backbone of a nucleic acid and modifications to the amino acid R groups or peptide backbone of a peptide result in one or more “modified nucleotides” or “modified amino acids,” respectively. Therefore, such nucleotides and amino acids must be annotated. Examples of backbone modifications include nucleotide analogs such as peptide nucleic acids (PNAs) and glycol nucleic acids (GNAs), and D-amino acids.

Note that modification of a terminal amino acid of a peptide or a terminal nucleotide of a nucleic acid does not necessarily result in a “modified amino acid” or “modified nucleotide”. One must look at the terminal modification and determine whether the modification changes the chemical structure of residue such that residue falls outside the exceptions set forth within paragraph 3(e) and 3(f). For example, a peptide in which the C-terminal residue is linked to a structure (such as part of a branched sequence – see peptide #2 in example 7(b)-3) via a conventional amide linkage is not considered a “modified residue” and therefore is not required to be annotated. Similarly, a peptide in which the N-terminal residue is amide bonded to biotin is not considered a “modified residue” and therefore is not required to be annotated. In both scenarios, the structure of the residue involved in the C-terminal or N-terminal linkage is not changed from the conventional amino acids recited in paragraph 3(e) of the Standard.

In contrast, terminal modifications that change the chemical structure of the residue are considered “modified residues” and must be annotated. For example, the methylation of the C-terminus in Example 3(c)-1 does change the chemical structure of the terminal residue, since the methyl group replaces the hydroxyl normally found at the alpha carboxyl group. Therefore, this methylated lysine must be annotated as a “modified residue”.

Note that it will be up to the applicant to evaluate each terminal residue modification within an enumerated sequence and make a determination as to whether or not the structure of the terminal residue is changed. If the modified residue structure is different from the conventional amino acids or nucleotides indicated in paragraph 3(e) and 3(f) of the Standard, then the modification must be annotated.

Finally, it is always recommended that applicants include as much information as reasonable in their sequence listings to represent their disclosures as accurately as possible. Therefore, even if a modification isn't required to be annotated, it should preferably be included.

Note however that annotation of variants of an enumerated, primary sequence must comply with the requirements of in ST.26 paragraphs 93-100. Modifications that are disclosed as variants of an enumerated sequence may not be required to be included in the sequence listing. For the definition of annotation of variants, see ST.26 paragraphs 93-95.

Representation of Modified Residues

ST.26 indicates that modified nucleotides and amino acids should be represented in the sequence listing as the corresponding unmodified residue whenever possible (see paragraphs 16 and 29). Note that this recommendation is a “should” – a “strongly encouraged approach, but not a requirement” (see paragraph 4(d)). It is up to the discretion of the applicant to decide if a modified residue will be represented by the corresponding unmodified residue or the variables “n” or “X”.

As a general rule of thumb – if a residue is modified by the addition of a moiety, such as methylation or acetylation, and the structure of the unmodified residue is generally unchanged, then representation by the unmodified residue is recommended. For example, a methylated adenosine should preferably be represented by “a” in the sequence listing. However, when the modified residue is structurally different from any unmodified residue, then an “n” or an “X” is recommended. For example, norleucine is an isomer of leucine, and its side chain is a linear structure of 4 carbons. Leucine also has a 4 carbon side chain, but it is branched at the second carbon. Therefore, norleucine isn't simply the result of a modification added to a leucine, but a completely different (although related) structure. It is therefore recommended that Norleucine be represented by an “X” in a sequence listing.

ST.26 3(e)절에는 "변형 아미노산(modified amino acid)"이란 L-알라닌(L-alanine), L-아르기닌(L-arginine), L-아스파라긴(L-asparagine), L-아스파르트산 (L-aspartic acid), L-시스테인(L-cysteine), L-글루타민(L-glutamine), L-글루탐산(L-glutamic acid), L-글리신(L-glycine), L-히스티딘(L-histidine), L-이소류신(L-isoleucine), L-류신(L-leucine), L-리신(L-lysine), L-메티오닌(L-methionine), L-페닐알라닌(L-phenylalanine), L-프롤린(L-proline), L-피롤리신(L-pyrrolysine), L-세린(L-serine), L-셀레노시스테인(L-selenocysteine), L-트레오닌(L-threonine), L-트립토판(L-tryptophan), L-티로신(L-tyrosine) 또는 L-발린(L-valine) 이외의 3(a)절에서 설명된 임의의 아미노산으로 정의한다. 유사하게, 본 표준에서는 "변형 핵산염기(modified nucleotide)"란 디옥시아데노신 3'-모노포스페이트(deoxyadenosine 3'-monophosphate), 디옥시구아노신 3'-모노포스페이트(deoxyguanosine 3'-monophosphate), 디옥시시티딘 3'-모노포스페이트(deoxycytidine 3'-monophosphate), 디옥시티미딘 3'-모노포스페이트(deoxythymidine 3'-monophosphate), 아데노신 3'-모노포스페이트(adenosine 3'-monophosphate), 구아노신 3'-모노포스페이트(guanosine 3'-monophosphate), 시티딘 3'-모노포스페이트(cytidine 3'-monophosphate) 또는 우리딘 3'-모노포스페이트(uridine 3'-monophosphate) 이외의 3(g)절에 설명된 임의의 핵산염기로 정의한다.

위의 정의를 기반으로 핵산염기 또는 핵산의 당-인산(sugar-phosphate) 백본(backbone)의 변형과 아미노산 R-그룹 또는 펩타이드의 펩타이드 백본의 변형은 각각 한 개 이상의 "변형 핵산" 또는 "변형 아미노산"을 발생시킨다. 그러므로, 이러한 핵산 및 아미노산은 반드시 주석이 필요하다. 백본 변형의 예는 펩타이드 핵산(PNAs), 글라이콜 핵산(GNA), 그리고 D-아미노산 같은 핵산 유사체가 있다.

펩타이드 말단 아미노산 또는 핵산의 말단 핵산의 변형이 반드시 "변형 아미노산" 또는 "변형 핵산"의 결과를 가져오는 것은 아니다. 말단의 변형을 보고 이러한 변형이 잔기의 화학적 구조를 변경하는지를 결정하여야 한다. 이러한 잔기는 3(e)절과 3(f)절 내의 나타난 예외 밖에 있다. 예를 들면, 펩타이드의 C-말단 잔기는 통상적인 아미이드(amide) 결합을 통한 구조(분기된 서열 - 실시 예 7(b)-3의 펩타이드 #2를 보세요)는 "변형 잔기"로 고려되지 않는다. 그러므로 주석을 요구하지 않는다. 유사하게, 펩타이드의 N-말단 잔기가 바이오틴(biotin)과 아미이드 결합한 경우도 "변형 잔기"로 고려되지 않는다. 그러므로 주석을 요구하지 않는다. 두 시나리오에서 C-말단 또는 N-말단 결합과 관련된 잔기의 구조는 본 표준의 3(3)절에 언급된 통상적인 아미노산을 변경시키지 않는다.

대조적으로 잔기의 화학적 구조를 변경하는 말단 변형은 "변형 잔기"로 고려되어야 하며 반드시 주석이 필요하다. 예를 들면, 실시 예 3(c)-1의 C-말단의 메틸화(methylation)은 말단 잔기의 화학적 구조를 변경시킨다. 메틸기는 알파 카르복실기(carboxyl group)에 일반적으로 발견되는 수산기(hydroxyl)를 대체하기 때문에 메틸화된 라이신(lysine)은 반드시 "변형 잔기"로서 주석이 필요하다.

나열된 잔기 중에서 각 말단 잔기 변형을 평가하고 말단 잔기의 구조가 바뀌었는지를 결정하는 것은 출원인에 달려있다. 만약 변경된 잔기 구조가 본 표준의 3(e), 3(f)절에 나타난 통상적인 핵산 또는 아미노산과 다르다면 이러한 변형은 반드시 주석이 요구된다.

마지막으로, 출원인은 가능한 정확한 개시를 표현하기 위해서 서열목록에 타당한 많은 정보를 포함하는 것을 권장한다. 그러므로, 변형이 주석을 요구하지 않더라도 우선적으로 포함되어야 한다. 그러나 나열된 일차 서열의 변이의 주석은 ST.26 93-100절의 요구사항을 준수하여야 한다. 나열된 서열의 변이로서 개시되는 변형은 서열목록에 포함되는 것을 요구하지 않을 수 있다. 변이의 주석 정의는 ST.26 93-95절을 참고하세요.

변형 잔기의 표현(representation)

ST.26은 변형된 핵산 및 아미노산은 가능하다면 상응하는 변형되지 않은 잔기로서 서열목록에 표현되어야 한다는 것을 나타낸다(16, 29절 참고). 이러한 권고는 "강하게 권장되는 방법이지만 필수적인 것은 아니다"(4(d)절 참고). 변형된 잔기가 사용하는 변형되지 않은 잔기 또는 "n" 혹은 "X"에 의해서 표현될지는 출원인의 결정에 달려있다.

통상적인 경험으로는 만약 잔기가 메틸화(methylation) 또는 아세틸화(acetylation) 같은 일부의 추가에 의해서 변형된 경우, 일반적으로 변형되지 않은 잔기의 구조는 변경되지 않는다. 그래서 변경되지 않은 잔기로 표현하는 것을 추천한다. 예를 들면, 메틸화된 아데노신(methylated adenosine)은 서열목록에서 "a"로 우선적으로 표현되어야 한다. 그러나, 변형된 잔기가 어떤 변형되지 않은 잔기와 구조적으로 다르다면, "n" 또는 "X"를 권장한다. 예를 들면, norleucine은 루이신(leucine)의 이성질체(isomer)이며, 이것의 사이드체인(side chain)은 4개 탄소로 된 직선 구조이다. 루이신 역시 4개 탄소 구조이지만, 두 번째 탄소에 가지가 있다. 그러므로 norleucine은 단지 루이신의 추가된 변형의 결과가 아니라, 완전히 다른(비록 관련된) 구조이다. 그러므로, norleucine은 서열목록에서 "X"로 표현되는 것이 바람직하다.

A nucleotide is “specifically defined” when it is represented by anything other than ‘n’, and an amino acid is “specifically defined” when it is represented by anything other than ‘X’ (see ST.26, paragraph 3(k)). Therefore, a 2’ O-methyl adenosine represented by an ‘a’ in the sequence is specifically defined, whereas norleucine represented by ‘X’ in the sequence is not specifically defined.

핵산은 'n' 이외 어떤 것으로 표현될 때는 “명확하게 정의(specifically defined)”되어야 하며, 아미노산은 'X' 이외 어떤 것으로 표현될 때는 “명확하게 정의(specifically defined)”되어야 한다(ST.26 3(k)절 참고). 그러므로 서열에서 'a'로 표현된 2' O-methyl adenosine은 구체적으로 정의되어야 하는 반면, 서열에서 “X”로 표현된 norleucine은 명확하게 정의되지 않는다.

Table A – Conventional Nucleotide Symbols, and Definitions

Symbol	Definition
a	adenine
c	cytosine
g	guanine
t	thymine in DNA uracil in RNA (t/u)
m	a or c
r	a or g
w	a or t/u
s	c or g
y	c or t/u
k	g or t/u
v	a or c or g; not t/u
h	a or c or t/u; not g
d	a or g or t/u; not c
b	c or g or t/u; not a
n	a or c or g or t/u; “unknown” or “other”

표 A - 통상적인 핵산염기 기호 및 정의

기호	정의
a	아데닌(adenine)
c	시토신(cytosine)
g	구아닌(guanine)
t	DNA의 티민(thymine) RNA의 우라실(uracil)(t/u)
m	a 또는 c
r	a 또는 g
w	a 또는 t/u
s	c 또는 g
y	c 또는 t/u
k	g 또는 t/u
v	a 또는 c 또는 g; t/u가 아님
h	a 또는 c 또는 t/u; g가 아님
d	a 또는 g 또는 t/u; c가 아님
b	c 또는 g 또는 t/u; a가 아님
n	a 또는 c 또는 g 또는 t/u; "불명" 또는 "기타"

Table B – Conventional Amino Acid Symbols, Threeletter Codes, and Definitions

Symbol	3-Letter Code	Definition
A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartic Acid (Aspartate)
C	Cys	Cysteine
Q	Gln	Glutamine
E	Glu	Glutamic Acid (Glutamate)
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
O	Pyl	Pyrolysine
S	Ser	Serine
U	Sec	Selenocysteine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine
B	Asx	Aspartic Acid or Asparagine
Z	Glx	Glutamine or Glutamic Acid
J	Xle	Leucine or Isoleucine
X	Xaa	A or R or N or D or C or Q or E or G or H or I or L or K or M or F or P or O or S or U or T or W or Y or V, “unknown” or “other”

표 B – 통상적인 아미노산 기호, 세 글자 코드 및 정의

기호	세 글자 코드	정의
A	Ala	알라닌(Alanine)
R	Arg	아르기닌(Arginine)
N	Asn	아스파라긴(Asparagine)
D	Asp	아스파르트 산(아스파르트산염) (Aspartic Acid (Aspartate))
C	Cys	시스테인(Cysteine)
Q	Gln	글루타민(Glutamine)
E	Glu	글루타민산(글루탐산염)(Glutamic Acid (Glutamate))
G	Gly	글리신(Glycine)
H	His	히스티딘(Histidine)
I	Ile	이소류신(Isoleucine)
L	Leu	류신(Leucine)
K	Lys	라이신(Lysine)
M	Met	메티오닌(Methionine)
F	Phe	페닐알라닌(Phenylalanine)
P	Pro	프롤린(Proline)
O	Pyl	피롤리신(Pyrrolysine)
S	Ser	세린(Serine)
U	Sec	셀레노시스테인(Selenocysteine)
T	Thr	트레오닌(Threonine)
W	Trp	트립토판(Tryptophan)
Y	Tyr	티로신(Tyrosine)
V	Val	발린(Valine)
B	Asx	아스파르트산 또는 아스파라긴(Aspartic Acid or Asparagine)
Z	Glx	글루타민 또는 글루탐산(Glutamine or Glutamic Acid)
J	Xle	류신 또는 이소류신(Leucine or Isoleucine)
X	Xaa	A 또는 R 또는 N 또는 D 또는 C 또는 Q 또는 E 또는 G 또는 H 또는 I 또는 L 또는 K 또는 M 또는 F 또는 P 또는 O 또는 S 또는 U 또는 T 또는 W 또는 Y 또는 V, "불명" 또는 "기타"

EXAMPLE INDEX

	Page
<u>Paragraph 3(a) – Definition of “amino acid”</u>	
Example 3(a)-1: D-amino acids	18
<u>Cross-referenced examples</u>	
Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid.....	56
Example 29-2: Use of the corresponding unmodified amino acid.....	57
Example 30-1: Feature key “CARBOHYD”	58
<u>Paragraph 3(c) – Definition of “enumeration of its residues”</u>	
Example 3(c)-1: Enumeration of amino acids by chemical structure	19
Example 3(c)-2: Shorthand formula for an amino acid sequence	20
<u>Cross-referenced examples</u>	
Example 27-1: Shorthand formula for an amino acid sequence.....	51
Example 27-3: Shorthand formula - four or more specifically defined amino acids	53
<u>Paragraph 3(f) – Definition of “modified nucleotide”</u>	
<u>Cross-referenced examples</u>	
Example 3(g)-4: Nucleic Acid Analogues.....	24
<u>Paragraph 3(g) – Definition of “nucleotide”</u>	
Example 3(g)-1: Nucleotide sequence interrupted by a C3 spacer	21
Example 3(g)-2: Nucleotide sequence with residue alternatives, including a C3 spacer	22
Example 3(g)-3: Abasic site	23
Example 3(g)-4: Nucleic Acid Analogues	24
<u>Cross-referenced examples</u>	
Example 11(b)-1: Double-stranded nucleotide sequence – different lengths	47
Example 14-1: The symbol “t” represents uracil in RNA.....	49
<u>Paragraph 3(k) – Definition of “specifically defined”</u>	
Example 3(k)-1: Nucleotide ambiguity symbols	25
Example 3(k)-2: Ambiguity symbol “n” used in both a conventional and nonconventional manner	26
Example 3(k)-3: Ambiguity symbol “n” used in a nonconventional manner	27
Example 3(k)-4: Ambiguity symbols other than “n” are “specifically defined”	28
Example 3(k)-5: Ambiguity abbreviation “Xaa” used in a nonconventional manner	29
<u>Paragraph 7 – Sequences for which inclusion in a sequence listing is required</u>	
<u>Cross-referenced examples</u>	

예제 색인

3.26.vi의 쪽번호

3(a)절 – “아미노산”의 정의

실시예 3(a)-1: D-아미노산 18

상호참조 예제

실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중 아미노산 기호 56

실시예 29-2: 상응하는 변형되지 않은 아미노산의 사용 57

실시예 30-1: 특징 기호 “CARBOHYD” 58

3(c)절 – “잔기의 나열”의 정의

실시예 3(c)-1: 화학구조에 의한 아미노산 나열 19

실시예 3(c)-2: 아미노산 서열의 축약식 20

상호참조 예제

실시예 27-1: 아미노산 서열의 축약식 51

실시예 37-3: 축약식 – 네 개 이상의 구체적으로 정의된 아미노산 53

3(f)절 – “변형 핵산염기”의 정의

상호참조 예제

실시예 3(g)-4: 핵산 유사체 24

3(g)절 – “핵산염기”의 정의

실시예 3(g)-1: C3 스페이서에 의해 단절된 핵산염기 서열 21

실시예 3(g)-2: C3 스페이서를 포함하여, 잔기 대체물(alternative)을 갖는 핵산염기 서열 22

실시예 3(g)-3: 탈염기 부위(abasic site) 23

실시예 3(g)-4: 핵산 유사체 24

상호참조 예제

실시예 11(b)-1: 이중가닥 핵산염기 서열 – 길이가 다를 경우 47

실시예 14-1: 기호 “t”는 RNA에서 우라실을 표현한다 49

3(k)절 – “구체적으로 정의된”의 정의

실시예 3(k)-1: 다중핵산염기 기호 25

실시예 3(k)-2: 통상적·비통상적 방식으로 사용되는 다중핵산염기 기호 “n” 26

실시예 3(k)-3: 비통상적 방식으로 사용되는 다중핵산염기 기호 “n” 27

실시예 3(k)-4: “n” 이외의 다중핵산염기 기호가 “구체적으로 정의”됨 28

실시예 3(k)-5: 비통상적 방식으로 사용되는 다중아미노산 약어 “Xaa” 29

7절 – 서열목록에 포함되어야 하는 서열

상호참조 예제

Example 28-1: Encoding nucleotide sequence and encoded amino acid sequence	54
Example 55-1: Combined DNA/RNA Molecule	64
Example 89-2: Feature location extends beyond the disclosed sequence	66
Example 92-1: Amino acid sequence encoded by a coding sequence with introns	68

Paragraph 7(a) – Nucleotide sequences required in a sequence listing

Example 7(a)-1: Branched nucleotide sequence.....	30
Example 7(a)-2: Linear nucleotide sequence having a secondary structure.....	32
Example 7(a)-3: Nucleotide ambiguity symbols used in a nonconventional manner	33
Example 7(a)-4: Nucleotide ambiguity symbols used in a nonconventional manner	34
Example 7(a)-5: Nonconventional nucleotide symbols	35
Example 7(a)-6: Nonconventional nucleotide symbols	36

Cross-referenced examples

Example 3(g)-1: Nucleotide sequence interrupted by a C3 spacer	21
Example 3(g)-2: Nucleotide sequence with residue alternatives, including a C3 spacer	22
Example 3(g)-3: Abasic site.....	23
Example 3(g)-4: Nucleic Acid Analogues.....	24
Example 3(k)-1: Nucleotide ambiguity symbols.....	25
Example 3(k)-2: Ambiguity symbol “n” used in both a conventional and nonconventional manner.....	26
Example 3(k)-3: Ambiguity symbol “n” used in a nonconventional manner	27
Example 3(k)-4: Ambiguity symbols other than “n” are “specifically defined”	28
Example 11(a)-1: Double-stranded nucleotide sequence – same lengths.....	46
Example 11(b)-1: Double-stranded nucleotide sequence – different lengths	47
Example 11(b)-2: Double-stranded nucleotide sequence – no base-pairing segment.....	48
Example 14-1: The symbol “t” represents uracil in RNA.....	49
Example 89-1: Encoding nucleotide sequence and encoded amino acid sequence	65
Example 93-1: Representation of enumerated variants	70
Example 95(b)-1: Representation of individual variant sequences with multiple interdependent variations	75

Paragraph 7(b) – Amino acid sequences required in a sequence listing

Example 7(b)-1: Four or more specifically defined amino acids.....	37
Example 7(b)-2: Branched amino acid sequence.....	38
Example 7(b)-3: Branched amino acid sequence.....	41
Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence	42

실시예 28-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	54
실시예 55-1: 결합된 DNA/RNA 분자	64
실시예 89-2: 특정 위치가 개시된 서열 이상으로 확장됨	66
실시예 92-1: 인트론(intron)을 가진 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열	68
7(a)절 – 서열목록에 필요한 핵산염기 서열	
실시예 7(a)-1: 분지형(branched) 핵산염기 서열	30
실시예 7(a)-2: 이차구조를 갖는 선형(linear) 핵산염기 서열	32
실시예 7(a)-3: 비통상적 방식으로 사용되는 다중핵산염기 기호	33
실시예 7(a)-4: 비통상적 방식으로 사용되는 다중핵산염기 기호	34
실시예 7(a)-5: 비통상적 핵산염기 기호	35
실시예 7(a)-6: 비통상적 핵산염기 기호	36
<u>상호참조 예제</u>	
실시예 3(g)-1: C3 스페이서에 의해 단절된 핵산염기 서열	21
실시예 3(g)-2: C3 스페이서를 포함하여, 잔기 대체물(alternative)을 갖는 핵산염기 서열	22
실시예 3(g)-3: 탈염기 부위(abasic site)	23
실시예 3(g)-4: 핵산 유사체	24
실시예 3(k)-1: 다중핵산염기 기호	25
실시예 3(k)-2: 통상적-비통상적 방식으로 사용되는 다중핵산염기 기호 "n"	26
실시예 3(k)-3: 비통상적 방식으로 사용되는 다중핵산염기 기호 "n"	27
실시예 3(k)-4: "n" 이외의 다중핵산염기 기호가 "구체적으로 정의"됨	28
실시예 11(a)-1: 이중가닥 핵산염기 서열 – 길이가 같을 경우	46
실시예 11(b)-1: 이중가닥 핵산염기 서열 – 길이가 다를 경우	47
실시예 11(b)-2: 이중가닥 핵산염기 서열 – 염기쌍(base-pairing)을 이루지 않는 단편	48
실시예 14-1: 기호 "t"는 RNA에서 우라실을 표현한다	49
실시예 89-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	65
실시예 93-1: 나열된 변이의 표현	70
실시예 95(b)-1: 다수의 상호 의존적 변이를 갖는 개별 변이 서열의 표현	75
7(b)절 – 서열목록에 필요한 아미노산 서열	
실시예 7(b)-1: 네 개 이상의 구체적으로 정의된 아미노산	37
실시예 7(b)-2: 분지형(branched) 아미노산 서열	38
실시예 7(b)-3: 분지형(branched) 아미노산 서열	41
실시예 7(b)-4: 분지형(branched) 아미노산 서열을 가진 고리모양(cyclic) 펩티드 서열	42

Example 7(b)-5: Cyclic peptide containing a branched amino acid sequence.....45

Cross-referenced examples

Example 3(a)-1: D-amino acids.....	18
Example 3(c)-1: Enumeration of amino acids by chemical structure.....	19
Example 3(c)-2: Shorthand formula for an amino acid sequence.....	20
Example 3(k)-5: Ambiguity abbreviation “Xaa” used in a nonconventional manner.....	29
Example 27-1: Shorthand formula for a nucleotide sequence.....	51
Example 27-3: Shorthand formula - four or more specifically defined amino acids.....	53
Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid.....	56
Example 29-2: Use of the corresponding unmodified amino acid.....	57
Example 30-1: Feature key “CARBODHYD”.....	58
Example 36-1: Sequence with a region of a known number of “X” residues represented as a single sequence.....	59
Example 37-1: Sequence with regions of an unknown number of “X” residues must not be represented as a single sequence.....	62
Example 37-2: Sequence with regions of an unknown number of “X” residues must not be represented as a single sequence.....	63
Example 89-1: Encoding nucleotide sequence and encoded amino acid sequence.....	65
Example 93-2: Representation of enumerated variants.....	71
Example 93-3: Representation of a consensus sequence.....	72
Example 94-1: Representation of single sequence with enumerated alternative amino acids.....	73
Example 95(a)-1: Representation of a variant sequence by annotation of the primary sequence.....	74

Paragraph 8 – Threshold for inclusion of sequences

Cross-referenced examples

Example 3(k)-1: Nucleotide ambiguity symbols.....	25
Example 3(k)-2: Ambiguity symbol “n” used in both a conventional and nonconventional manner.....	26
Example 7(a)-1: Branched nucleotide sequence.....	30
Example 7(a)-6: Nonconventional nucleotide symbols.....	36
Example 7(b)-1: Four or more specifically defined amino acids.....	37
Example 7(b)-2: Branched amino acid sequence.....	38
Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence.....	42
Example 14-1: The symbol “t” represents uracil in RNA.....	49
Example 37-1: Sequence with regions of an unknown number of “X” residues must not be represented as a single sequence.....	62
Example 37-2: Sequence with regions of an unknown number of “X” residues must not be represented as a single sequence.....	63

실시예 7(b)-5: 분지형(branched) 아미노산 서열을 가진 고리모양(cyclic) 펩티드 서열 45

상호참조 예제

실시예 3(a)-1: D-아미노산	18
실시예 3(c)-1: 화학구조에 의한 아미노산 나열	19
실시예 3(c)-2: 아미노산 서열의 축약식	20
실시예 3(k)-5: 비통상적 방식으로 사용되는 다중아미노산 약어 "Xaa"	29
실시예 27-1: 핵산염기 서열의 축약식	51
실시예 27-3: 축약식 – 네 개 이상의 구체적으로 정의된 아미노산	53
실시예 29-1: "기타" 아미노산에 대한 가장 제한적인 다중 아미노산 기호	56
실시예 29-2: 상응하는 변형되지 않은 아미노산의 사용	57
실시예 30-1: 특징 기호 "CARBOHYD"	58
실시예 36-1: 단일 서열로 표시된 알려진 개수의 "X" 잔기 영역을 가진 서열	59
실시예 37-1: 알려지지 않은 개수의 "X" 잔기 영역을 가진 서열은 단일 서열로 표현되어서는 안된다	62
실시예 37-2: 알려지지 않은 개수의 "X" 잔기 영역을 가진 서열은 단일 서열로 표현되어서는 안된다	63
실시예 89-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	65
실시예 93-2: 나열된 변이의 표현	71
실시예 93-3: 일치(consensus) 서열 표현	72
실시예 94-1: 나열된 선택적 아미노산을 가진 단일 서열의 표현	73
실시예 95(a)-1: 일차 서열의 주석에 의한 변이 서열의 표현	74

8절 – 서열 포함의 임계점(threshold)

상호참조 예제

실시예 3(k)-1: 다중핵산염기 기호	25
실시예 3(k)-2: 통상적-비통상적 방식으로 사용되는 다중핵산염기 기호 "n"	26
실시예 7(a)-1: 분지형(branched) 핵산염기 서열	30
실시예 7(a)-6: 비통상적 핵산염기 기호	36
실시예 7(b)-1: 네 개 이상의 구체적으로 정의된 아미노산	37
실시예 7(b)-2: 분지형(branched) 아미노산 서열	38
실시예 7(b)-4: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	42
실시예 14-1: 기호 "t"는 RNA에서 우라실을 표현한다	49
실시예 37-1: 알려지지 않은 개수의 "X" 잔기 영역을 가진 서열은 단일 서열로 표현되어서는 안된다	62
실시예 37-2: 알려지지 않은 개수의 "X" 잔기 영역을 가진 서열은 단일 서열로 표현되어서는 안된다	63

Example 94-1: Representation of single sequence with enumerated alternative amino acids..... [73](#)

Paragraph 11 – Representation of a nucleotide sequence

Cross-referenced examples

Example 3(g)-4: Nucleic Acid Analogues..... [24](#)

Example 7(a)-1: Branched nucleotide sequence..... [30](#)

Paragraph 11(a) – Double-stranded nucleotide sequence - fully complementary

Example 11(a)-1: Double-stranded nucleotide sequence – same lengths..... [46](#)

Paragraph 11(b) – Double-stranded nucleotide sequence – not fully complementary

Example 11(b)-1: Double-stranded nucleotide sequence – different lengths..... [47](#)

Example 11(b)-2: Double-stranded nucleotide sequence – no base-pairing segment..... [48](#)

Paragraph 13 – Representation of nucleotides

Cross-referenced examples

Example 3(k)-2: Ambiguity symbol “n” used in both a conventional and nonconventional manner..... [26](#)

Example 7(a)-1: Branched nucleotide sequence..... [30](#)

Example 14-1: The symbol “t” represents uracil in RNA..... [49](#)

Example 93-1: Representation of enumerated variants [70](#)

Paragraph 14 – Symbol “t” construed as uracil in RNA

Example 14-1: The symbol “t” represents uracil in RNA..... [49](#)

Cross-referenced examples

Example 55-1: Combined DNA/RNA Molecule [64](#)

Paragraph 15 – The most restrictive nucleotide ambiguity symbols should be used

Cross-referenced examples

Example 3(g)-1: Nucleotide sequence interrupted by a C3 spacer [21](#)

Example 3(g)-2: Nucleotide sequence with residue alternatives, including a C3 spacer..... [22](#)

Example 3(k)-4: Ambiguity symbols other than “n” are “specifically defined” [28](#)

Example 95(b)-1: Representation of individual variant sequences with multiple interdependent variations... [75](#)

Paragraph 16 – Representation of a modified nucleotide

Cross-referenced examples

Example 3(g)-1: Nucleotide sequence interrupted by a C3 spacer [21](#)

Example 3(g)-4: Nucleic Acid Analogues..... [24](#)

Paragraph 17 – Annotation of a modified nucleotide

Cross-referenced examples

실시예 94-1: 나열된 선택적 아미노산을 가진 단일 서열의 표현	73
11절 – 핵산염기 서열의 표현	
<u>상호참조 예제</u>	
실시예 3(g)-4: 핵산 유사체	24
실시예 7(a)-1: 분지형(branched) 핵산염기 서열	30
11(a)절 – 이중가닥 핵산염기 서열 - 완전 상보적	
실시예 11(a)-1: 이중가닥 핵산염기 서열 - 길이가 같을 경우	46
11(b)절 – 이중가닥 핵산염기 서열 - 불완전 상보적	
실시예 11(b)-1: 이중가닥 핵산염기 서열 - 길이가 다를 경우	47
실시예 11(b)-2: 이중가닥 핵산염기 서열 - 염기쌍(base-pairing)을 이루지 않는 단편 포함	48
13절 – 핵산염기의 표현	
<u>상호참조 예제</u>	
실시예 3(k)-2: 통상적-비통상적 방식으로 사용되는 다중핵산염기 기호 “n”	26
실시예 7(a)-1: 분지형(branched) 핵산염기 서열	30
실시예 14-1: 기호 “t”는 RNA에서 우라실을 표현한다	49
실시예 93-1: 나열된 변이의 표현	70
14절 – RNA에서 우라실로 해석되는 기호 “t”	
실시예 14-1: 기호 “t”는 RNA에서 우라실을 표현한다	49
<u>상호참조 예제</u>	
실시예 55-1: 결합된 DNA/RNA 분자	64
15절 – 가장 제한적인 다중핵산염기 기호를 사용하여야 한다	
<u>상호참조 예제</u>	
실시예 3(g)-1: C3 스페이서에 의해 단절된 핵산염기 서열	21
실시예 3(g)-2: C3 스페이서를 포함하여, 잔기 대체물(alternative)을 갖는 핵산염기 서열	22
실시예 3(k)-4: “n” 이외의 다중핵산염기 기호가 “구체적으로 정의”됨	28
실시예 95(b)-1: 다수의 상호 의존적 변이를 갖는 개별 변이 서열의 표현	75
16절 – 변형 핵산염기의 표현	
<u>상호참조 예제</u>	
실시예 3(g)-1: C3 스페이서에 의해 단절된 핵산염기 서열	21
실시예 3(g)-4: 핵산 유사체	24
17절 – 변형 핵산염기의 주석	
<u>상호참조 예제</u>	

Example 3(g)-1: Nucleotide sequence interrupted by a C3 spacer	21
Example 3(g)-3: Abasic site.....	23
Example 7(a)-1: Branched nucleotide sequence.....	30
Example 7(a)-2: Linear nucleotide sequence having a secondary structure.....	32
Example 7(a)-6: Nonconventional nucleotide symbols.....	36

Paragraph 18 – Annotation of regions of consecutive modified nucleotides

Cross-referenced examples

Example 3(g)-4: Nucleic Acid Analogues.....	24
Example 11(b)-1: Double-stranded nucleotide sequence – different lengths	47

Paragraph 19 – Annotation of uracil in DNA or thymine in RNA

Cross-referenced examples

Example 14-1: The symbol “t” represents uracil in RNA.....	49
--	--------------------

Paragraph 25 – Amino acid sequence residue position number 1

Cross-referenced examples

Example 3(a)-1: D-amino acids.....	18
Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence.....	42
Example 7(b)-5: Cyclic peptide containing a branched amino acid sequence.....	45
Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid.....	56

Paragraph 26 – Representation of amino acids

Cross-referenced examples

Example 7(b)-2: Branched amino acid sequence.....	38
Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence.....	42
Example 7(b)-5: Cyclic peptide containing a branched amino acid sequence.....	45
Example 36-1: Sequence with a region of a known number of “X” residues represented as a single sequence	59
Example 89-1: Encoding nucleotide sequence and encoded amino acid sequence	65
Example 92-1: Amino acid sequence encoded by a coding sequence with introns.....	68
Example 93-2: Representation of enumerated variants	71
Example 93-3: Representation of a consensus sequence	72

Paragraph 27 – The most restrictive amino acid ambiguity symbols should be used

Example 27-1: Shorthand formula for an amino acid sequence	51
Example 27-2: Shorthand formula - less than four specifically defined amino acids	52
Example 27-3: Shorthand formula - four or more specifically defined amino acids	53

실시예 3(g)-1: C3 스페이서에 의해 단절된 핵산염기 서열	21
실시예 3(g)-3: 탈염기 부위(abasic site)	23
실시예 7(a)-1: 분지형(branched) 핵산염기 서열	30
실시예 7(a)-2: 이차구조를 갖는 선형(linear) 핵산염기 서열	32
실시예 7(a)-6: 비통상적 핵산염기 기호	36
18절 – 연속적인 변형 핵산염기 영역의 주석	
<u>상호참조 예제</u>	
실시예 3(g)-4: 핵산 유사체	24
실시예 11(b)-1: 이중가닥 핵산염기 서열 – 길이가 다를 경우	47
19절 – DNA에서의 우라실 또는 RNA에서의 타민의 주석	
<u>상호참조 예제</u>	
실시예 14-1: 기호 “t”는 RNA에서 우라실을 표현한다	49
25절 – 아미노산 서열 잔기 위치 번호 1	
<u>상호참조 예제</u>	
실시예 3(a)-1: D-아미노산	18
실시예 7(b)-4: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	42
실시예 7(b)-5: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	45
실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중아미노산 기호	56
26절 – 아미노산의 표현	
<u>상호참조 예제</u>	
실시예 7(b)-2: 분지형(branched) 아미노산 서열	38
실시예 7(b)-4: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	42
실시예 7(b)-5: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	45
실시예 36-1: 단일 서열로 표시된 알려진 개수의 “X” 잔기 영역을 가진 서열	59
실시예 89-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	65
실시예 92-1: 인트론(intron)을 가진 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열	68
실시예 93-2: 나열된 변이의 표현	71
실시예 93-3: 일치(consensus) 서열의 표현	72
27절 – 가장 제한적인 다중아미노산 기호를 사용하여야 한다	
실시예 27-1: 아미노산 서열의 축약식	51
실시예 27-2: 축약식 – 네 개 미만으로 구체적으로 정의된 아미노산	52
실시예 27-3: 축약식 – 네 개 이상으로 구체적으로 정의된 아미노산	53

Cross-referenced examples

Example 3(c)-2: Shorthand formula for an amino acid sequence.....	20
Example 7(b)-1: Four or more specifically defined amino acids.....	37
Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid.....	56
Example 36-1: Sequence with a region of a known number of “X” residues represented as a single sequence.....	59
Example 36-2: Sequence with multiple regions of a known number or range of “X” residues represented as a single sequence.....	60
Example 36-3: Sequence with multiple regions of a known number or range of “X” residues represented as a single sequence.....	61
Example 37-2: Sequence with regions of an unknown number of “X” residues must not be represented as a single sequence.....	63
Example 93-3: Representation of a consensus sequence.....	72
Example 94-1: Representation of single sequence with enumerated alternative amino acids.....	73
Example 95(a)-1: Representation of a variant sequence by annotation of the primary sequence.....	74

Paragraph 28 – Amino acid sequences separated by internal terminator symbols

Example 28-1: Encoding nucleotide sequence and encoded amino acid sequence	54
---	---------------------------

Cross-referenced examples

Example 89-1: Encoding nucleotide sequence and encoded amino acid sequence.....	65
Example 92-1: Amino acid sequence encoded by a coding sequence with introns.....	68

Paragraph 29 – Representation of an “other” amino acid

Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid	56
Example 29-2: Use of the corresponding unmodified amino acid.....	57

Cross-referenced examples

Example 3(a)-1: D-amino acids.....	18
Example 7(b)-2: Branched amino acid sequence.....	38
Example 7(b)-3: Branched amino acid sequence.....	41
Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence.....	42
Example 30-1: Feature key “CARBODHYD”	58

Paragraph 30 – Annotation of a modified amino acid

Example 30-1: Feature key “CARBODHYD”	58
--	---------------------------

Cross-referenced examples

Example 3(a)-1: D-amino acids.....	18
Example 3(c)-1: Enumeration of amino acids by chemical structure.....	19
Example 7(b)-2: Branched amino acid sequence.....	38

상호참조 예제

실시예 3(c)-2: 아미노산 서열의 축약식	20
실시예 7(b)-1: 네 개 이상의 구체적으로 정의된 아미노산	37
실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중아미노산 기호	56
실시예 36-1: 단일 서열로 표시된 알려진 개수의 “X” 잔기 영역을 가진 서열	59
실시예 36-2: 단일 서열로 표시된 알려진 개수 또는 범위의 “X” 잔기의 다중 영역이 있는 서열	60
실시예 36-3: 단일 서열로 표시된 알려진 개수 또는 범위의 “X” 잔기의 다중 영역이 있는 서열	61
실시예 37-2: 알려지지 않은 개수의 “X” 잔기 영역을 가진 서열은 단일 서열로 표현되어서는 안된다	63
실시예 93-3: 일치(consensus) 서열 표현	72
실시예 94-1: 나열된 선택적 아미노산을 가진 단일 서열의 표현	73
실시예 95(a)-1: 일차 서열의 주석에 의한 변이 서열의 표현	74

28절 – 내부 종결 기호로 분리된 아미노산 서열

실시예 28-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	54
--	-----------

상호참조 예제

실시예 89-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	65
실시예 92-1: 인트론(intron)을 가진 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열	68

29절 – “기타” 아미노산의 표현

실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중 아미노산 기호	56
실시예 29-2: 상응하는 변형되지 않은 아미노산의 사용	57

상호참조 예제

실시예 3(a)-1: D-아미노산	18
실시예 7(b)-2: 분지형(branched) 아미노산 서열	38
실시예 7(b)-3: 분지형(branched) 아미노산 서열	41
실시예 7(b)-4: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	42
실시예 30-1: 특징 기호 “CARBOHYD”	58

30절 – 변형 아미노산의 주석

실시예 30-1: 특징 기호 “CARBOHYD”	58
---	-----------

상호참조 예제

실시예 3(a)-1: D-아미노산	18
실시예 3(c)-1: 화학구조에 의한 아미노산 나열	19
실시예 7(b)-2: 분지형(branched) 아미노산 서열	38

Example 7(b)-3: Branched amino acid sequence.....	41
Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence.....	42
Example 7(b)-5: Cyclic peptide containing a branched amino acid sequence.....	45
Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid.....	56
Example 29-2: Use of the corresponding unmodified amino acid.....	57

Paragraph 31 – Representation of a D-amino acid

Cross-referenced examples

Example 3(a)-1: D-amino acids.....	18
Example 3(c)-1: Enumeration of amino acids by chemical structure.....	19
Example 7(b)-2: Branched amino acid sequence.....	38
Example 7(b)-3: Branched amino acid sequence.....	41
Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence.....	42
Example 7(b)-5: Cyclic peptide containing a branched amino acid sequence.....	45

Paragraph 32 – Annotation of an “unknown” amino acid

Cross-referenced examples

Example 3(c)-1: Enumeration of amino acids by chemical structure.....	19
---	--------------------

Paragraph 34 – Annotation of a contiguous region of “X” residues

Cross-referenced examples

Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid.....	56
--	--------------------

Paragraph 36 – Sequences containing regions of an exact number of contiguous “n” or “X” residues

Example 36-1: Sequence with a region of a known number of “X” residues represented as a single sequence.....	59
Example 36-2: Sequence with multiple regions of a known number or range of “X” residues represented as a single sequence.....	60
Example 36-3: Sequence with multiple regions of a known number or range of “X” residues represented as a single sequence.....	61

Paragraph 37 – Sequences containing regions of an unknown number of contiguous “n” or “X” residues

Example 37-1: Sequence with regions of an unknown number of “X” residues must not be represented as a single sequence.....	62
Example 37-2: Sequence with regions of an unknown number of “X” residues must not be represented as a single sequence.....	63

Paragraph 41 – Reserved characters

Cross-referenced examples

Example 89-2: Feature location extends beyond the disclosed sequence.....	66
---	--------------------

실시예 7(b)-3: 분지형(branched) 아미노산 서열	41
실시예 7(b)-4: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	42
실시예 7(b)-5: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	45
실시예 29-1: "기타" 아미노산에 대한 가장 제한적인 다중아미노산 기호	56
실시예 29-2: 상응하는 변형되지 않은 아미노산의 사용	57

31절 – D-아미노산의 표현

상호참조 예제

실시예 3(a)-1: D-아미노산	18
실시예 3(c)-1: 화학구조에 의한 아미노산 나열	19
실시예 7(b)-2: 분지형(branched) 아미노산 서열	38
실시예 7(b)-3: 분지형(branched) 아미노산 서열	41
실시예 7(b)-4: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	42
실시예 7(b)-5: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열	45

32절 – "불명" 아미노산의 주석

상호참조 예제

실시예 3(c)-1: 화학구조에 의한 아미노산 나열	19
------------------------------------	----

34절 – "X" 잔기의 연속 영역의 주석

상호참조 예제

실시예 29-1: "기타" 아미노산에 대한 가장 제한적인 다중아미노산 기호	56
---	----

36절 – 정확한 수의 연속된 "n" 또는 "X" 잔기 영역을 포함하는 서열

실시예 36-1: 단일 서열로 표시된 알려진 개수의 "X" 잔기 영역을 가진 서열	59
실시예 36-2: 단일 서열로 표시된 알려진 개수 또는 범위의 "X" 잔기의 다중 영역이 있는 서열	60
실시예 36-3: 단일 서열로 표시된 알려진 개수 또는 범위의 "X" 잔기의 다중 영역이 있는 서열	61

37절 – 알려지지 않은 수의 연속된 "n" 또는 "X" 잔기 영역을 포함하는 서열

실시예 37-1: 알려지지 않은 개수의 "X" 잔기 영역을 가진 서열은 단일 서열로 표현되어서는 안된다	62
실시예 37-2: 알려지지 않은 개수의 "X" 잔기 영역을 가진 서열은 단일 서열로 표현되어서는 안된다	63

41절 – 예약어

상호참조 예제

실시예 89-2: 특정 위치가 개시된 서열 이상으로 확장됨	66
--	----

Paragraph 54 – The element INSDSeq moltype

Cross-referenced examples

Example 14-1: The symbol “t” represents uracil in RNA [49](#)

Paragraph 55 – A nucleotide sequence that contains both DNA and RNA segments

Example 55-1: Combined DNA/RNA Molecule [64](#)

Paragraph 56 – Example illustrating a nucleotide sequence that contains both DNA and RNA segments

Cross-referenced examples

Example 55-1: Combined DNA/RNA Molecule [64](#)

Paragraph 57 – The element INSDSeq sequence

Cross-referenced examples

Example 28-1: Encoding nucleotide sequence and encoded amino acid sequence [54](#)

Example 92-1: Amino acid sequence encoded by a coding sequence with introns [68](#)

Paragraph 65 – Location descriptor

Cross-referenced examples

Example 3(g)-4: Nucleic Acid Analogues [24](#)

Example 89-2: Feature location extends beyond the disclosed sequence [66](#)

Paragraph 66 – Location descriptor syntax

Cross-referenced examples

Example 3(g)-4: Nucleic Acid Analogues [24](#)

Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence [42](#)

Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid [56](#)

Example 30-1: Feature key “CARBODHYD” [58](#)

Example 89-2: Feature location extends beyond the disclosed sequence [66](#)

Paragraph 67 – Location operator

Cross-referenced examples

Example 92-1: Amino acid sequence encoded by a coding sequence with introns [68](#)

Paragraph 70 – Feature locations

Cross-referenced examples

Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence [42](#)

Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid [56](#)

Example 30-1: Feature key “CARBODHYD” [58](#)

Example 89-2: Feature location extends beyond the disclosed sequence [66](#)

Paragraph 71 – Representation of the characters “<” and “>” in a location descriptor

54절 – INSDSeq_moltype 요소

상호참조 예제

실시예 14-1: 기호 “t”는 RNA에서 우라실을 표현한다 49

55절 – DNA와 RNA 단편을 모두 포함하는 핵산염기 서열

실시예 55-1: 결합된 DNA/RNA 분자 64

56절 – DNA와 RNA 단편을 모두 포함하는 핵산염기 서열을 보여주는 예제

상호참조 예제

실시예 55-1: 결합된 DNA/RNA 분자 64

57절 – INSDSeq_서열 요소

상호참조 예제

실시예 28-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열 54

실시예 92-1: 인트론(intron)을 가진 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열 68

65절 – 위치 설명자

상호참조 예제

실시예 3(g)-4: 핵산 유사체 24

실시예 89-2: 특정 위치가 개시된 서열 이상으로 확장됨 66

66절 – 위치 설명자 구문

상호참조 예제

실시예 3(g)-4: 핵산 유사체 24

실시예 7(b)-4: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열 42

실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중아미노산 기호 56

실시예 30-1: 특정 기호 “CARBOHYD” 58

실시예 89-2: 특정 위치가 개시된 서열 이상으로 확장됨 66

67절 – 위치 연산자

상호참조 예제

실시예 92-1: 인트론(intron)을 가진 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열 42

70절 – 특정 위치

상호참조 예제

실시예 7(b)-4: 분지형(branched) 아미노산을 가진 고리모양(cyclic) 펩티드 서열 56

실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중아미노산 기호 58

실시예 30-1: 특정 기호 “CARBOHYD” 58

실시예 89-2: 특정 위치가 개시된 서열 이상으로 확장됨 66

71절 – 위치 설명자에서 기호 ‘<’와 ‘>’의 표시

Cross-referenced examples

Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid..... [56](#)

Example 89-2: Feature location extends beyond the disclosed sequence..... [66](#)

Paragraph 83 – Example illustrating a nucleotide sequence that is not naturally occurring

Cross-referenced examples

Example 55-1: Combined DNA/RNA Molecule [64](#)

Paragraph 89 – “CDS” Feature key

Example 89-1: Encoding nucleotide sequence and encoded amino acid sequence [65](#)

Example 89-2: Feature location extends beyond the disclosed sequence [66](#)

Cross-referenced examples

Example 92-1: Amino acid sequence encoded by a coding sequence with introns..... [68](#)

Paragraph 90 – The qualifiers “transl table” and “translation”

Cross-referenced examples

Example 28-1: Encoding nucleotide sequence and encoded amino acid sequence [54](#)

Example 89-1: Encoding nucleotide sequence and encoded amino acid sequence [65](#)

Example 92-1: Amino acid sequence encoded by a coding sequence with introns..... [68](#)

Paragraph 92 – Amino acid sequence encoded by a coding sequence

Example 92-1: Amino acid sequence encoded by a coding sequence with introns [68](#)

Cross-referenced examples

Example 28-1: Encoding nucleotide sequence and encoded amino acid sequence [54](#)

Example 89-1: Encoding nucleotide sequence and encoded amino acid sequence [65](#)

Example 89-2: Feature location extends beyond the disclosed sequence..... [66](#)

Paragraph 93 – Primary sequence and a variant, each enumerated by its residue

Example 93-1: Representation of enumerated variants [70](#)

Example 93-2: Representation of enumerated variants [71](#)

Example 93-3: Representation of a consensus sequence [72](#)

Paragraph 94 – Variant sequence disclosed as a single sequence with enumerated alternative residues

Example 94-1: Representation of single sequence with enumerated alternative amino acids [73](#)

Paragraph 95(a) – A variant sequence disclosed only by reference to a primary sequence with multiple independent variations

Example 95(a)-1: Representation of a variant sequence by annotation of the primary sequence..... [74](#)

Paragraph 95(b) – A variant sequence disclosed only by reference to a primary sequence with multiple interdependent variations

<u>상호참조 예제</u>	
실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중아미노산 기호	56
실시예 89-2: 특정 위치가 개시된 서열 이상으로 확장됨	66
83절 – 자연적으로 생기지 않는 핵산염기 서열을 나타내는 예제	
<u>상호참조 예제</u>	
실시예 55-1: 결합된 DNA/RNA 분자	64
89절 – “CDS” 특징 기호	
실시예 89-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	65
실시예 89-2: 특정 위치가 개시된 서열 이상으로 확장됨	66
<u>상호참조 예제</u>	
실시예 92-1: 인트론(intron)을 가진 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열	68
90절 – 한정자 “transl_table”과 “translation”	
<u>상호참조 예제</u>	
실시예 28-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	54
실시예 89-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	65
실시예 92-1: 인트론(intron)을 가진 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열	68
92절 – 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열	
실시예 92-1: 인트론(intron)을 가진 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열	68
<u>상호참조 예제</u>	
실시예 28-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	54
실시예 89-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열	65
실시예 89-2: 특정 위치가 개시된 서열 이상으로 확장됨	66
93절 – 일차 서열 및 변이(각각 잔기로 나열됨)	
실시예 93-1: 나열된 변이의 표현	70
실시예 93-2: 나열된 변이의 표현	71
실시예 93-3: 일치(consensus) 서열 표현	72
94절 – 나열된 선택 잔기를 갖는 단일 서열로서 개시된 변이 서열	
실시예 94-1: 나열된 선택적 아미노산을 가진 단일 서열의 표현	73
95(a)절 – 다수의 독립적 변이를 갖는 일차 서열의 참조만으로 개시된 변이 서열	
실시예 95(a)-1: 일차 서열의 주석에 의한 변이 서열의 표현	74
95(b)절 – 다수의 상호 의존적 변이를 갖는 일차 서열의 참조만으로 개시된 변이 서열	

Example 95(b)-1: Representation of individual variant sequences with multiple interdependent variations [75](#)

Paragraph 96 – Feature keys and qualifiers for a variant sequence

Cross-referenced examples

Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid..... [56](#)

Paragraph 97 – Annotation of a variant sequence

Cross-referenced examples

Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid..... [56](#)

Example 93-3: Representation of a consensus sequence [72](#)

Example 94-1: Representation of single sequence with enumerated alternative amino acids..... [73](#)

실시예 95(b)-1: 다수의 상호 의존적 변이를 갖는 개별 변이 서열의 표현 75

96절 – 변형 서열의 특징 기호 및 한정어

상호참조 예제

실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중아미노산 기호 56

97절 – 변형 서열의 주석

상호참조 예제

실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중아미노산 기호 56

실시예 93-3: 일치(consensus) 서열 표현 72

실시예 94-1: 나열된 선택적 아미노산을 가진 단일 서열의 표현 73

EXAMPLES

Paragraph 3(a) Definition of “amino acid”

Example 3(a)-1: D-amino acids

A patent application describes the following sequence:

Cyclo (D-Ala-D-Glu-Lys-Nle-Gly-D-Met-D-Nle)

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

Paragraph 3(a) of the Standard defines “amino acid” as including “D-amino acids” and amino acids containing modified or synthetic side chains. Based on this definition, the enumerated peptide contains five amino acids that are specifically defined (D-Ala, D-Glu, Lys, Gly, and D-Met). Therefore, the sequence must be included in a sequence listing as required by ST.26 paragraph 7(b).

Question 3: How should the sequence(s) be represented in the sequence listing?

Paragraph 29 requires that D-amino acids should be represented in the sequence as the corresponding unmodified L-amino acid. Further, any modified amino acid that cannot be represented by any other symbol in Annex I, Section 3, Table 3, must be represented by the symbol “X”.

In this example, the sequence contains three D-amino acids that can be represented by an unmodified L-amino acid in Annex I, Section 3, Table 3, one L-amino acid (Nle), and one D-amino acid (D-Nle) that must be represented by the symbol “X”.

Paragraph 25 indicates that when amino acid sequences are circular in configuration and the ring consists solely of amino acid residues linked by peptide bonds, applicant must choose the amino acid in residue position number 1. Accordingly, the sequence may be represented as:

AEKXGMX (SEQ ID NO: 1)

or otherwise, with any other amino acid in the sequence in residue position number 1. A feature key “SITE” and a qualifier “note” must be provided for each D-amino acid with the complete, unabbreviated name of the D-amino acid as the qualifier value, e.g., D-alanine and D-norleucine. Further, a feature key “SITE” and a qualifier “note” must be provided with the abbreviation for L-norleucine as the qualifier value, i.e. “Nle”, as set forth in Annex I, Section 4, Table 4. Finally, a feature key “REGION” and a qualifier “note” should be provided to indicate that the peptide is circular.

Relevant ST.26 paragraphs: 3(a), 7(b), 25, 26, 29, 30, and 31

실시예들

3(a) 절 "아미노산"의 정의

실시예 3(a)-1: D-아미노산

특허출원에 다음 서열이 기술되어 있다:

사이클로(D-Ala-D-Glu-Lys-Nle-Gly-D-Met-D-Nle)

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

본 표준의 3(a)절에 따르면 "아미노산"의 정의에는 "D-아미노산" 및 변형 또는 합성 측쇄를 함유하는 아미노산이 포함된다. 이러한 정의에 기초하여, 나열된 펩티드는 구체적으로 정의된 5개의 아미노산(D-Ala, D-Glu, Lys, Gly 및 D-Met)을 함유한다. 그러므로 이 서열은 본 ST.26의 7(b)절에 따라 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

29절에서는 서열에서 D-아미노산을 동가의 미변형 L-아미노산으로 표현할 것을 요구한다. 또한, 별첨 I, 3절, 표3에 있는 다른 기호로 표현할 수 없는 변형 아미노산은 기호 "X"로 표현해야 한다.

이 예제에서, 해당 서열은 별첨 I, 3절, 표 3에서 미변형 L-아미노산에 의해 표현될 수 있는 3개의 D-아미노산, 기호 "X"로 표현해야 하는 하나의 L-아미노산(Nle) 및 하나의 D-아미노산(D-Nle)을 포함한다.

25절에서는 아미노산 서열 구조가 원형이고 고리가 펩티드 결합에 의해 연결된 아미노산 잔기로만 이루어진 경우, 출원인은 잔기 위치 번호 1의 아미노산을 선택하여야 한다. 따라서, 본 서열은 다음과 같이 표현될 수 있다.

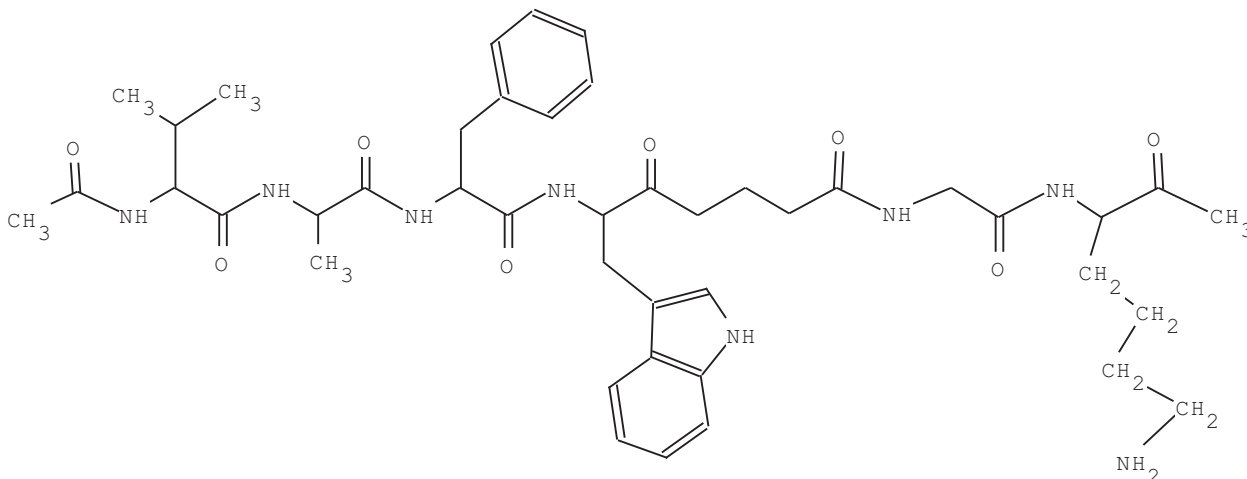
AEKXGMX (SEQ ID NO: 1)

또는 달리 임의의 다른 아미노산을 서열 번호의 잔기 위치 1번으로 표현될 수 있다. 특징 기호 "SITE" 및 한정자 "note"에는 한정자 값으로 D-아미노산(예: D-알라닌 및 D-노르류신)의 단축되지 않은 완전한 이름으로 각 D-아미노산을 제공되어야 한다. 또한, 특징 기호 "SITE" 및 한정자 "noteE"에는 별첨1, 4절, 표 4에 기술된 것처럼 한정자 값으로 L-norleucine의 약어(즉 "Nle")가 제공되어야 한다. 끝으로, 펩티드가 원형임을 나타내는 특징 기호 "REGION" 및 한정자 "note"를 제공하여야 한다.

관련 ST.26 절: 3(a), 7(b), 25, 26, 29, 30 및 31

Paragraph 3(c) – Definition of “enumeration of its residues”

Example 3(c)-1: Enumeration of amino acids by chemical structure



Question 1: Does ST.26 require inclusion of the sequence(s)?

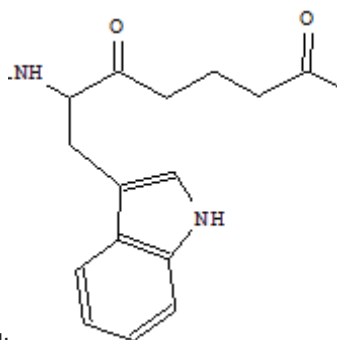
YES

The enumerated peptide, illustrated as a structure, contains at least four specifically defined amino acids. Therefore, the sequence must be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence may be represented as:

VAFXGK (SEQ ID NO: 2)



wherein “X” represents an “other” modified amino acid: *6-amino-7-(1H-indol-3-yl)-5-oxoheptanoic acid*, which requires a feature key “SITE” together with the qualifier “note”. The qualifier “note” provides the complete, unabbreviated name of the modified tryptophan in position 4 of the enumerated peptide, e.g., “6-amino-7-(1H-indol-3-yl)-5-oxoheptanoic acid”. The methylation of the C terminus changes the chemical structure of the terminal lysine since the -OH on the terminal end is replaced by -CH₃. Due to this structural change, the lysine within the sequence is considered a “modified amino acid.” Accordingly, a feature key “SITE” and qualifier “note” are required to indicate the methylation of the C-terminus. Valine, however, is not considered a “modified amino acid” since the addition of the acetyl group to the valine involves a conventional peptide linkage. The acylation does not alter the structure of the valine. Accordingly, an additional feature key “SITE” and qualifier “note” should be included to indicate the acetylation of the N terminus.

Alternatively, the sequence may be represented as:

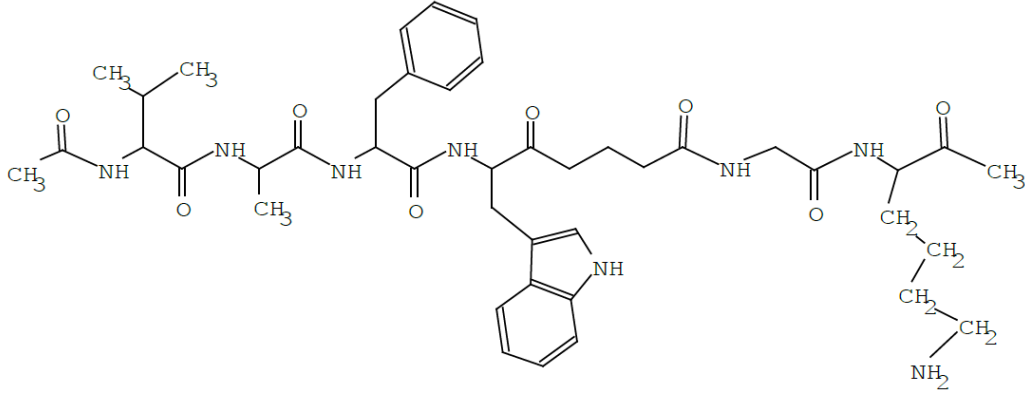
VAFW (SEQ ID NO: 3)

A feature key “SITE” and qualifier “note” are required to indicate modification of tryptophan in position 4 of the enumerated peptide with the value: “C-terminus linked via a glutaraldehyde bridge to dipeptide GK”. Further, an additional feature key “SITE” at location 1 and qualifier “note” should be included to indicate the acetylation of the N-terminus.

Relevant ST.26 paragraph(s): 3(c), 7(b), 29, 30, and 31

3(c) 절 – “잔기의 나열”의 정의

실시예 3(c)-1: 화학구조에 의한 아미노산 나열



질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

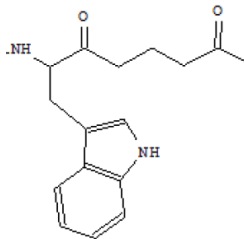
예

구조로 나타낸, 나열된 펩티드는 최소한 4개의 구체적으로 정의된 아미노산을 함유한다. 따라서 이 서열은 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

이 서열은 다음과 같이 표현될 수 있다:

VAFXGK (SEQ ID NO: 2)



여기에서 "X"는 "other" 변형 아미노산: 을 표현하며, 이는 한정자 "note"와 함께 특징 기호 "SITE"를 필요로 한다. 한정자 "note"는 나열된 펩티드의 위치 4에서 변형된 트립토판의 단축되지 않는 완전한 명칭, 예를 들면 "6-amino-7-(1H-indol-3-yl)-5-oxoheptanoic acid"을 제공한다. 말단의 -OH기가 -CH₃기로 치환되었으므로 C-말단의 메틸화(methylation)가 라이신의 화학적 구조를 변경시킨다. 따라서, 특징 기호 "SITE" 및 한정자 "note"가 C-말단의 메틸화(methylation)를 나타내기 위해 필요하다. 그러나, 발린(valine)은 발린에 아세틸기(acetyl group) 추가가 통상적인 펩타이드 결합에 해당하므로 "변형 아미노산"으로 고려되지 않는다. 아세틸화(acetylation)는 발린의 구조를 변경시키지는 않는다. 따라서 N-말단의 아세틸화(acetylation)를 나타내기 위해서 추가 특징 기호 "SITE" 및 한정자 "note"가 필요하다.

선택적으로, 서열은 다음과 같이 표현될 수 있다:

VAFW (SEQ ID NO: 3)

특징 기호 "SITE" 및 한정자 "note"는 값으로 "C-terminus linked via a glutaraldehyde bridge to dipeptide GK"을 필요로 하며, 이는 나열된 펩티드의 위치 4에서 트립토판의 변형을 나타낸다. 또한, 부가적으로 위치 1에 대해서 특징 기호 "SITE"와 N-말단의 아세틸화(acetylation)를 나타내기 위한 한정자 "note"가 포함되어야 한다.

관련 ST.26 절: 3(c), 7(b), 29, 30 및 31

Example 3(c)-2: Shorthand formula for an amino acid sequence $(G_4z)_n$

Where G= Glycine, z = any amino acid and variable n can be any whole integer.

Question 1: Does ST.26 require inclusion of the sequence(s)?**YES**

The disclosure indicates that “n” can be “any whole integer”; therefore, the most encompassing embodiment of “n” is indeterminate. Since “n” is indeterminate, the peptide of the formula cannot be expanded to a definite length, and therefore, the unexpanded formula must be considered.

The enumerated peptide in the unexpanded formula (“n” = 1) provides four specifically defined amino acids, each of which is Gly, and the symbol “z”. Conventionally “Z” is the symbol for “glutamine or glutamic acid”; however, the example defines “z” as “any amino acid”. Under ST.26, an amino acid that is not specifically defined is represented by “X”. Based on this analysis, the enumerated peptide, i.e. GGGGX, contains four glycine residues that are enumerated and specifically defined. Thus, ST.26 paragraph 7(b) requires inclusion of the sequence in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence uses a nonconventional symbol “z”, the definition of which must be determined from the disclosure (see Introduction to this document). Since “z” is defined as any amino acid, the conventional symbol used to represent this amino acid is “X”. Therefore, the sequence must be represented as a single sequence:

GGGGX (SEQ ID NO: 4)

and should be annotated with the feature key REGION, feature location “>5” (corresponds to >5), with a note qualifier with the value “The entire sequence of amino acids 1-5 can be repeated one or more times.”

According to paragraph 27, “X” will be construed as any one of “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, or “V”, except where it is used with a further description in the feature table. Since in this example “X” represents “any amino acid”, it must be annotated with the feature key VARIANT and a note qualifier with the value, “X can be any amino acid”.

Where practicable, each “X” should be annotated individually. However, a region of contiguous “X” residues, or a multitude of “X” residues dispersed throughout the sequence, may be jointly described with the feature key VARIANT using the syntax “x..y” as the location descriptor, where x and y are the positions of the first and last “X” residues, and a note qualifier with the value, “X can be any amino acid”.

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraph(s): 3(c), 7(b) and 27.

실시예 3(c)-2: 아미노산 서열의 축약식

$(G_z)_n$

여기에서, G는 글리신(Glycine), z는 임의의 아미노산 및 변수 "n"은 정수가 될 수 있다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

본 개시는 "n"이 "임의의 정수"일 수 있음을 나타내고; 따라서, "n"의 가장 포괄적인 구체적 예는 미정(未定)이다. "n"은 미정이기 때문에, 식의 펩티드는 한정된 길이로 확장될 수 없으므로, 확장되지 않은 식이 고려되어야 한다.

확장되지 않은 식("n"= 1)에서 나열된 펩티드는 각각 Gly인 4개의 구체적으로 정의된 아미노산 및 기호 "z"를 제공한다. 통상적으로 "Z"는 "글루타민(glutamine) 또는 글루탐산(glutamic acid)"의 상징이다. 그러나, 이 예는 "z"를 "임의의 아미노산"으로 정의한다. ST.26에서, 구체적으로 정의되지 않은 아미노산은 "X"로 표현된다. 이 분석에 기초하여, 나열된 펩티드, 즉 GGGGX는 나열되고 구체적으로 정의된 4개의 글리신(Glycine) 잔기를 함유한다. 따라서 ST.26의 7(b)절에 따라 이 서열을 서열목록에 포함시켜야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

이 순서는 비통상적인 기호 "z"를 사용하며, 그 정의는 반드시 출원공개에서 결정하여야 한다(본 문서의 서론 참조). "z"는 임의의 아미노산으로 정의되므로, 이 아미노산을 표현하는 데에 사용되는 통상적인 기호는 "X"이다. 따라서, 서열은 단일 서열로 표현하여야 한다:

GGGGX (SEQ ID NO: 4)

특징 기호 REGION, 특징 위치 ">5에 상응함" 및 "아미노산 1-5의 전체 서열은 1회 이상 반복될 수 있다"는 값을 갖는 주석이 note 한정자에 주석으로 달려야 한다.

27절에 따르면, "X"는 특징 표에서 추가 설명과 함께 사용되는 경우를 제외하고는, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다. 본 실시예에서 "X"가 "임의의 아미노산"을 표현하므로 VARIANT 특징 기호와 "X can be any amino acid"라는 값을 가진 note 한정자로 반드시 주석을 달아야 한다.

가능한 경우, 각 "X"는 독립적으로 주석을 달아야 한다. 그러나, 연속 "X" 잔기의 영역, 또는 서열 전체에 분산된 다수의 "X" 잔기는 위치 설명자로서 "x..y" 구문을 사용하는 특징 기호 'VARIANT'와 "X can be any amino acid"라는 값을 가진 'note' 한정자를 이용하여 함께 기술될 수 있다. 여기서 x와 y는 첫 번째 및 마지막 "X" 잔기의 위치이다.

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 3(c), 7(b) 및 (27).

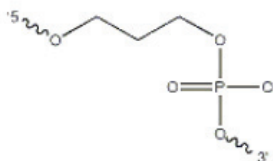
Paragraph 3(g) Definition of "nucleotide"

Example 3(g)-1: Nucleotide sequence interrupted by a C3 spacer

A patent application describes the following sequence:

atgcatgcatgcnⁿcgcatgcatgc

where n = a C3 spacer with the following structure:



Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated sequence contains two segments of specifically defined nucleotides separated by a C3 spacer.

The C3 spacer is not a nucleotide according to paragraph 3(g); the conventional symbol "n" is being used in a nonconventional manner (see Introduction to this document). Consequently, each segment is a separate nucleotide sequence. Since each segment contains more than 10 specifically defined nucleotides, both must be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

Each segment must be included in a sequence listing as a separate sequence, each with their own sequence identification number:

atgcatgcatgc (SEQ ID NO: 5)

cgcatgcatgc (SEQ ID NO: 6)

The cytosine in each segment that is attached to the C3 spacer should be further described in a feature table using the feature key "misc_feature" and the qualifier "note". The "note" qualifier value, which is "free text", should indicate the presence of the spacer, which is joined to another nucleic acid and identify the spacer by either its complete unabbreviated chemical name, or by its common name, e.g., C3 spacer.

Relevant ST.26 paragraphs: 3(g), 7(a), and 15

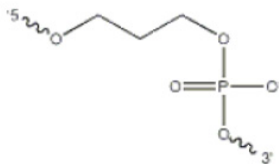
3(g) 절: "핵산염기"의 정의

실시예 3(g)-1: C3 스페이서(spacer)에 의해 단절된 핵산염기 서열

특허출원에 다음 서열이 기술되어 있다:

atgcatgcatgcnccggcatgcatgc

여기에서 n = 다음 구조의 C3 스페이서(spacer):



질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

나열된 서열은 C3 스페이서에 의해 분리된 구체적으로 정의된 핵산염기의 2개의 단편을 보유한다.

C3 스페이서는 3(g)절에 따른 핵산염기가 아니고; 기존의 기호 "n"은 비통상적인 방식으로 사용된다(이 문서의 서론 참조). 결과적으로, 각 단편은 별개의 핵산염기 서열이다. 각 단편은 10개가 넘는 구체적으로 정의된 핵산염기를 포함하므로 모두 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

각 단편은 별도의 서열, 자신의 서열식별번호와 함께 같은 서열목록에 포함되어야 한다:

atgcatgcatgc (SEQ ID NO: 5)

cgcatgcatgc (SEQ ID NO: 6)

C3 스페이서(spacer)에 부착된 각 단편의 시토신(cytosine)은 특징 기호 "misc_feature"와 한정자 "note"를 사용하여 특징 표에 추가적으로 설명되어야 한다. "자유텍스트"인 "note" 한정자 값에 다른 핵산에 결합된 스페이서의 존재를 나타내며, 단축되지 않은 완전한 화학 명칭이나 일반적인 명칭(예: C3 스페이서)으로 나타내야 한다.

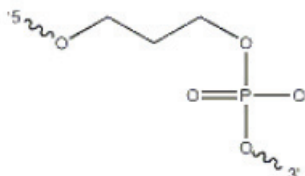
관련 ST.26 절: 3(g), 7(a) 및 15

Example 3(g)-2: Nucleotide sequence with residue alternatives, including a C3 spacer

A patent application describes the following sequence:

atgcatgcatgcnⁿcgcatgcatgc

where n = c, a, g, or a C3 spacer with the following structure:



Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

There are 24 specifically defined residues in the enumerated sequence interrupted by the variable “n.” The explanation of the sequence in the disclosure must be consulted to determine if the “n” is used in a conventional or nonconventional manner (see Introduction to this document).

The disclosure indicates that n = c, a, g, or a C3 spacer. The “n” is a conventional symbol used in a nonconventional manner, since it is described as including a C3 spacer, which does not meet the definition of a nucleotide. The symbol “n” is also described as including “c”, “a”, or “g”; therefore, ST.26 requires inclusion of the 25 nucleotide sequence in a sequence listing. Since two segments separated by the C3 spacer are distinct sequences from the 25 nucleotide sequence, the two 12 nucleotide sequences may also be included.

Question 3: How should the sequence(s) be represented in the sequence listing?

The example indicates that “n = c, a, g, or a C3 spacer”. As discussed above, a C3 spacer is not a nucleotide. According to paragraph 15, the symbol “n” must not be used to represent anything other than a nucleotide; therefore, the symbol “n” cannot represent a C3 spacer in a sequence listing.

Paragraph 15 also states that where an ambiguity symbol is appropriate, the most restrictive symbol should be used. The symbol “v” represents “a or c or g” according to Annex I, Section 1, Table 1, which is more restrictive than “n”.

Where variable “n” in the example is c, a, or g, the single sequence enumerated by its residues that includes the most disclosed embodiments, and is therefore, the most encompassing sequence (see Introduction to this document) that must be included in a sequence listing is:

atgcatgcatgcvⁿcgcatgcatgc (SEQ ID NO: 7)

Inclusion of any additional sequences essential to the disclosure or claims of the invention is strongly encouraged, as discussed in the introduction to this document.

Where variable “n” in the example is a C3 spacer, the sequence can be considered two separate segments of specifically defined nucleotides on either side of the variable “n”, i.e. atgcatgcatgc (SEQ ID NO: 8); and cgcatgcatgc (SEQ ID NO: 9). If essential to the disclosure or claims, these two sequences should also be included in the sequence listing, each with their own sequence identification number.

The cytosine in each segment that is attached to the C3 spacer should be further described in a feature table using the feature key “misc_feature” and the qualifier “note”. The “note” qualifier value, which is “free text”, should indicate the presence of the spacer, which is joined to another nucleic acid and identify the spacer by either its complete unabbreviated chemical name, or by its common name, e.g., C3 spacer.

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

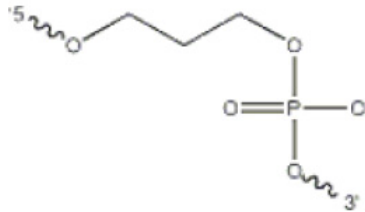
Relevant ST.26 paragraphs: 3(g), 7(a), and 15

실시예 3(g)-2: C3 스페이서(spacer)를 포함하여, 잔기 대체물(alternative)을 갖는 핵산염기 서열

특허출원에 다음 서열이 기술되어 있다:

atgcatgcatgcncggcatgcatgc

여기에서 n = c, a, g 또는 다음 구조의 C3 스페이서(spacer):



질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

변수 "n"에 의해 중단되어 나열된 서열에는 24개의 구체적으로 정의된 잔기가 존재한다. 본 개시에서 서열에 대한 설명은 "n"이 통상적인 또는 비통상적인 방식으로 사용되는지를 결정하기 위해 참조되어야 한다(본 문서의 서론 참조).

본 개시에서는 n = c, a, g 또는 C3 스페이서임을 나타낸다. "n"은 핵산염기의 정의를 충족시키지 않는 C3 스페이서를 포함하는 것으로 기술되기 때문에 비통상적인 방식으로 사용되는 통상적인 기호이다. 기호 "n"은 또한 "c", "a" 또는 "g"를 포함하는 것으로 설명되며; 따라서, ST.26은 서열목록에는 25개 핵산염기 서열이 포함될 필요로 한다. C3 스페이서(spacer)에 의해 분리된 2개의 단편은 25의 핵산염기 서열과 별개의 서열이므로, 2개의 12개의 핵산염기 서열도 포함될 수 있다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

예는 "n = c, a, g 또는 C3 스페이서"를 나타낸다. 위에서 논의된 바와 같이, C3 스페이서는 핵산염기가 아니다. 15절에 따르면, 기호는 "n"은 핵산염기가 아닌 다른 것을 표현하는 데에 사용할 수 없다; 따라서, 기호 "n"은 서열목록에서 C3 스페이서를 표현할 수 없다.

15절은 또한 다중핵산염기 기호가 적절한 경우 가장 제한적인 기호를 사용하여야 한다고 명시하고 있다. "v" 기호는 별첨 1, 섹션 1, 표 1에 따라 "a" 또는 "c" 또는 "g"를 표현하며 "n"보다 더 제한적이다.

실시예에서 변수 "n"이 c, a 또는 g인 경우, 가장 개시된 실시 양태를 포함하는 잔기에 의해 나열된 하나의 서열이고, 따라서 서열목록에 포함되어야 하는 가장 포괄적인 서열(본 문서의 서론 참조) 다음과 같다.

atgcatgcatgcvcgcatgcatgc (SEQ ID NO: 7)

본 문서의 서두에서 논의된 바와 같이, 본 발명의 개시 또는 청구 범위에 필수적인 임의의 추가 서열의 포함하는 것이 강력히 권장된다.

실시예에서 변수 "n"이 C3 스페이서인 경우, 서열은 변수 "n"의 어느 한 측면, 즉 atgcatgcatgc(SEQ ID NO: 8); 및 cgcatgcatgc(SEQ ID NO: 9)이다. 개시 또는 청구 범위에 필수적인 경우, 이들 2개의 서열은 또한 각각 자신의 서열식별번호를 갖으며 서열목록에 포함되어야 한다.

C3 스페이서에 부착된 각 단편의 시토신(cytosine)은 특징 기호 "misc_feature"와 한정자 "note"를 사용하여 특징 표에 추가로 설명되어야 한다. "자유텍스트"인 "note" 한정자 값은 다른 핵산에 결합된 스페이서의 존재를 나타내며 단축되지 않는 완전한 화학 명칭 또는 일반적인 명칭(예: C3 스페이서)으로 나타낸다.

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 3(g), 7(a) 및 15

Example 3(g)-3: Abasic site

A patent application describes the following sequence:

gagcattgac-AP-taaggct

Wherein AP is an abasic site

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The specifically defined residues of the enumerated sequence are interrupted by an abasic site. The 5' side of the abasic site contains 10 nucleotides and the 3' side of the abasic site contains 7 nucleotides. Paragraph 3(g)(ii)(2) defines an abasic site as a "nucleotide" when it is part of a nucleotide sequence. Consequently, the abasic site in this example is considered a "nucleotide" for the purposes of determining if and how the sequence is required to be included in a sequence listing. Accordingly, the residues on each side of the abasic site are part of a single enumerated sequence containing 18 nucleotides total, 17 of which are specifically defined. Therefore, the sequence must be included as a single sequence in a sequence listing as required by ST.26 paragraph 7(a).

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence must be included in a sequence listing as:

gagcattgacntaaggct (SEQ ID NO: 10)

The abasic site must be represented by an "n" and must be further described in a feature table. The preferred means of annotation is the feature key "modified_base" and the mandatory qualifier "mod_base" with the value "OTHER". A "note" qualifier must be included that describes the modified base as an abasic site.

Relevant ST.26 paragraphs: 3(g), 7(a), and 17

실시예 3(g)-3: 탈염기 부위(abasic site)

특허출원에 다음 서열이 기술되어 있다:

gagcattgac-AP-taaggct

여기서, AP는 탈염기 부위(abasic site) 이다

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?**예**

나열된 서열의 구체적으로 정의된 잔기들은 탈염기 부위에 의해 단절된다. 탈염기 부위의 5'-말단 쪽은 10개의 핵산염기가 있고, 탈염기 부위의 3'-말단 쪽은 7개의 핵산염기가 있다. 3(g)(ii)(2)절은 탈염기 부위가 핵산염기 서열의 일부일 때 "핵산염기"로 정의된다. 결과적으로, 본 사례의 탈염기 부위는 서열이 서열목록에 포함되어야 하는지 여부 및 방법을 결정할 목적으로 "핵산염기"로 고려된다. 따라서, 탈염기 부위의 각 측면의 잔기는 총 18개의 핵산염기를 함유하는 단일 나열된 서열의 일부이며, 그 중 17개가 구체적으로 특정된다. 그러므로, 서열은 ST.26의 7(a)절에 따라 서열목록에 단일 서열로 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

서열은 다음과 같이 서열목록에 포함되어야 한다.

gagcattgacntaaggct (SEQ ID NO: 10)

탈염기 부위는 "n"으로 표현되어야 하며 특징 표에 자세히 설명되어야 한다. 추천되는 주석 방법은 특징 기호 "modified_base"와 "OTHER" 값을 갖는 필수 한정자 "mod_base"이다. 변형 핵산염기를 탈염기 부위로 설명하는 "note" 한정자를 포함하여야 한다.

관련 ST.26 절: 3(g), 7(a) 및 17

Example 3(g)-4: Nucleic Acid Analogues

A patent application discloses the following glycol nucleic acid (GNA) sequence:

PO₄-tagttcattgactaaggctcccattgact-OH

Wherein the left end of the sequence mimics the 5' end of a DNA sequence.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES – The individual residues that comprise a GNA sequence are considered nucleotides according to ST.26 paragraph 3(g)(i)(2). Accordingly, the sequence has more than ten enumerated and “specifically defined” nucleotides and is required to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

GNA sequences do not have a 5'-end and a 3'-end, but rather, a 3'-end and a 2'-end. The 3'-end, which is routinely depicted as having a terminal phosphate group, corresponds to the 5'-end of DNA or RNA. (Note that other nucleic acid analogues may correspond differently to the 5'-end and 3'-end of DNA and RNA.) According to paragraph 11, it must be included in a sequence listing “in the direction from left to right that mimics the 5'-end to 3'-end direction.” Therefore, it must be included in a sequence listing as:

tagttcattgactaaggctcccattgact (SEQ ID NO: 11)

The sequence must be described in a feature table using the feature key “modified_base” and the mandatory qualifier “mod_base” with the abbreviation “OTHER”. A “note” qualifier must be included with the complete unabbreviated name of the modified nucleotides, such as “glycol nucleic acids” or “2,3-dihydroxypropyl nucleosides”. A single INSDFeature element can be used to describe the entire sequence as a GNA where the INSDFeature_location has the range “1..30”.

Relevant ST.26 paragraphs: 3(d), 3(g), 7(a), 11, 16, 18, 65, and 66

실시예 3(g)-4: 핵산 유사체

특허출원에 다음의 글리콜(glycol) 핵산(GNA) 서열을 개시하고 있다:

PO₄-tagttcattgactaaggctccccattgact-OH

여기에서, 상기 서열의 좌측 말단은 DNA 서열의 5'-말단과 유사하다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예 – GNA 서열을 포함하는 개별 잔기는 ST.26 3(g)(i)(2)절에 따라 핵산염기로 간주된다. 따라서, 서열은 10개 초과인 나열되고 "구체적으로 정의된" 핵산염기를 가지며, 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

GNA 서열은 5'-말단 및 3'-말단이 아니라 3'-말단 및 2'-말단을 갖는다. 일반적으로 말단 포스페이트기(phosphate group)를 갖는 것으로 설명되는 3'-말단은 DNA 또는 RNA의 5'-말단에 대응한다(다른 핵산 유사체는 DNA와 RNA의 5'-말단과 3'-말단과 다르게 대응할 수 있다). 11절에 따르면, 본 서열은 "5'-말단에서 3'-말단 방향으로 왼쪽에서 오른쪽으로" 서열목록에 포함되어야 한다. 따라서, 서열은 다음과 같이 서열목록에 포함되어야 한다.

tagttcattgactaaggctccccattgact (SEQ ID NO: 11)

서열은 특징 기호 "modified_base"와 "OTHER" 약자를 가진 필수 한정자 "mod_base"를 사용하여 특징 표에 설명되어야 한다. "note" 한정자는 "glycol nucleic acids" 또는 "2,3-dihydroxypropyl nucleosides"와 같은 변형 핵산염기의 단축되지 않은 완전한 명칭으로 표시되어야 한다. 단일 INSDFeature 요소는 INSDFeature_location의 범위 "1..30"을 갖는 GNA 처럼 전체 서열을 설명하는 데에 사용될 수 있다.

관련 ST.26 절: 3(d), 3(g), 7(a), 11, 16, 18, 65 및 66

Paragraph 3(k) Definition of “specifically defined”

Example 3(k)-1: Nucleotide ambiguity symbols

5' NNG KNG KNG K 3'

N and K are IUPAC-IUB ambiguity codes

Question 1: Does ST.26 require inclusion of the sequence(s)?

NO

IUPAC-IUB ambiguity codes correspond to the list of nucleotide symbols defined in Annex I, Section 1, Table 1. According to paragraph 3(k), a specifically defined nucleotide is any nucleotide other than those represented by the symbol “n” listed in Annex I. Therefore, “K” and “G” are specifically defined nucleotides and “N” is not a specifically defined nucleotide.

The enumerated sequence does not have ten or more specifically defined nucleotides and therefore is not required by ST.26 paragraph 7(a) to be included in a sequence listing.

Question 2: Does ST.26 permit inclusion of the sequence(s)?

NO

According to paragraph 8, “A sequence listing must not include any sequences having fewer than ten specifically defined nucleotides....” The enumerated sequence does not have ten or more specifically defined nucleotides; therefore, it must not be included in a sequence listing.

Relevant ST.26 paragraphs: 3(k), 7(a), 8, and 13

3(k)절 “구체적으로 정의된”의 정의

실시예 3(k)-1: 다중핵산염기 기호

5'NNG KNG KNG K 3'

여기서, N과 K는 IUPAC-IUB 다중핵산염기 코드이다

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

아니오

IUPAC-IUB 다중핵산염기 코드는 별첨 I, 섹션 1, 표 1에 정의된 핵산염기 기호 목록에 해당한다. 3(k)절에 따르면, 구체적으로 정의된 핵산염기는 별첨 I에 나열된 기호 "n"으로 표현되는 것 이외의 핵산염기이다. 따라서, "K" 및 "G"는 구체적으로 정의된 핵산염기이고, "N"은 구체적으로 정의된 핵산염기가 아니다.

나열된 서열은 10개 이상의 구체적으로 정의된 핵산염기가 없고, 따라서 ST.26 7절에 의해 서열목록에 포함되지 아니한다.

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가

아니오

8절에 따르면, "서열목록은 10개 미만의 구체적으로 정의된 핵산염기를 갖는 임의의 서열을 포함해서는 아니된다...." 나열된 서열은 10개 이상의 구체적으로 정의된 핵산염기를 갖지 않으며; 따라서 서열목록에 포함되어서는 아니된다.

관련 ST.26 절: 3(k), 7(a), 8 및 13

Example 3(k)-2: Ambiguity symbol “n” used in both a conventional and nonconventional manner

An application discloses the artificial sequence: 5'-AATGCCGGAN-3'. The disclosure further states:

- (i) in one embodiment, N is any nucleotide;
- (ii) in one embodiment, N is optional but is preferably G;
- (iii) in one embodiment, N is K;
- (iv) in one embodiment, N is C.

Question 1: Does ST.26 require inclusion of the sequence(s)?**NO**

The enumerated sequence contains 9 specifically defined nucleotides and an “N.” The explanation of the sequence in the disclosure must be consulted to determine if the symbol “N” is used in a conventional manner (see Introduction to this document).

Consideration of disclosed embodiments (i) through (iv) of the enumerated sequence reveals that the most encompassing embodiment of “N” is “any nucleotide”. In the most encompassing embodiment, “N” in the enumerated sequence is used in a conventional manner.

In certain embodiments “N” is described as specifically defined residues (i.e., “N is C” in part (iv)). However, only the most encompassing embodiment (i.e., “N is any nucleotide”) is considered when determining if a sequence must be included in a sequence listing. Thus, the enumerated sequence that must be evaluated is 5'-AATGCCGGAN-3'.

Based on this analysis, the enumerated sequence, i.e. AATGCCGGAN, does not contain ten specifically defined nucleotides. Therefore, ST.26 paragraph 7(a) does not require inclusion of the sequence in a sequence listing, despite the fact that “n” is also defined as specific nucleotides in some embodiments.

Question 2: Does ST.26 permit inclusion of the sequence(s)?**NO**

The sequence “AATGCCGGAN” must not be included in a sequence listing.

However, a described alternative sequence may be included in a sequence listing if the “N” is replaced with a specifically defined nucleotide.

Question 3: How should the sequence(s) be represented in the sequence listing?

Inclusion of sequences which represent embodiments that are a key part of the invention is **strongly** encouraged. Inclusion of these sequences allows for a more thorough search and provides public notice of the subject matter for which a patent is sought.

For the above example, it is highly recommended that the following three additional sequences are included in the sequence listing, each with their own sequence identification number:

aatgccggag (SEQ ID NO: 12)

aatgccggak (SEQ ID NO: 13)

aatgccggac (SEQ ID NO: 14)

If less than all three of the above sequences are included, the nucleotide that replaces the “n” should be annotated to describe the alternatives. For example, if only SEQ ID NO: 12 above is included in the sequence listing, the feature key “misc_difference” with feature location “10” should be used together with two “replace” qualifiers where the value for one would be “g” and the second would be “c”.

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraphs: 3(k), 7(a), 8, and 13

실시예 3(k)-2: 통상적·비통상적 방식으로 사용된 다중핵산염기 기호 "n"

출원에 인공 서열 5'-AATGCCGGAN-3'이 개시되어 있다. 본 개시는 또한 다음과 같이 명시한다:

- (i) 실시예에서, N은 임의의 핵산염기이고;
- (ii) 실시예에서, N은 선택적이지만 가급적 G이고;
- (iii) 실시예에서, N은 K이고;
- (iv) 실시예에서, N은 C이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

아니오

나열된 서열은 9개의 구체적으로 정의된 핵산염기 및 "N"을 함유한다. 기호 "N"이 통상적인 방식으로 사용되는지를 결정하기 위해 본 개시의 서열 설명을 참조하여야 한다(본 문서의 서론 참조).

나열된 서열의 개시된 실시예 (i) 내지 (iv)를 살펴보면, "N"의 가장 포괄적인 실시예는 "임의의 핵산염기"이다. 가장 포괄적인 실시예에서 나열된 서열의 "N"은 통상적인 방식으로 사용된다.

특정 실시예에서 "N"은 구체적으로 정의된 잔기(즉, 부분 (iv)에서 "N은 C")로 기술된다. 그러나, 서열이 서열목록에 포함되어야 하는지 여부를 결정할 때, 가장 포괄적인 실시예(즉, "N은 임의의 핵산염기이다")만이 고려된다. 따라서 평가해야 하는 나열된 서열은 5'-AATGCCGGAN-3'이다.

이 분석에 기초하여, 나열된 서열, 즉 AATGCCGGAN은 10개의 구체적으로 정의된 핵산염기를 함유하지 않는다. 따라서, "n"이 일부 실시예에서 특정 핵산염기로 정의된다는 사실에도 불구하고, ST.26 7(a)절에 따라 서열은 서열목록에 포함되지 아니한다.

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가?

아니오

"AATGCCGGAN"서열은 서열목록에 포함되어서는 아니된다.

그러나, "N"이 구체적으로 정의된 핵산염기로 대체된다면, 기술된 선택적 서열은 서열목록에 포함될 수 있다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

발명의 핵심 부분인 실시예를 표현되는 서열을 포함하도록 강력히 권장한다. 서열을 포함하면 보다 철저한 검색이 가능해지며 특허가 추구하는 특허대상을 공개적으로 알릴 수 있다.

위 사례에 대해, 서열목록에 다음과 같은 각각 고유한 서열식별번호가 있는 세 가지 추가 서열을 포함할 것을 적극 권고한다.

aatgccggag (SEQ ID NO: 12)

aatgccggak (SEQ ID NO: 13)

aatgccggac (SEQ ID NO: 14)

위의 서열 중 3 개 미만이 모두 포함되는 경우, "n"을 대체하는 핵산염기에 대해서 대체 핵산을 설명하기 위해 주석을 달아야 한다. 예를 들면, 위의 서열 번호: 12만 서열목록에 포함된 경우, 특정 위치가 "10"인 특징 기호 "misc_difference"는 두 개의 "replace" 한정자와 함께 사용되어야 한다. 여기에서 하나의 값은 "g"이고, 두 번째는 "c"가 될 수 있다.

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 3(k), 7(a), 8 및 13

Example 3(k)-3: Ambiguity symbol “n” used in a nonconventional manner

An application discloses the sequence: 5'-aatgttggan-3'

Wherein n is c

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

According to paragraph 3(k), a “specifically defined” nucleotide is any nucleotide other than those represented by the symbol “n” listed in Annex I, Section 1, Table 1.

In this example “n” is used in a nonconventional manner to represent only “c”. The disclosure does not indicate that “n” is used in the conventional manner to represent “any nucleotide”. Therefore, the sequence must be interpreted as if the equivalent conventional symbol, i.e. “c”, had been used in the sequence (see Introduction to this document). Accordingly, the enumerated sequence that must be considered is:

5'-aatgttgac-3'

This sequence has ten specifically defined nucleotides and is required by ST.26 paragraph 7(a) to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence must be included in a sequence listing as: aatgttgac (SEQ ID NO: 15)

Relevant ST.26 paragraphs: 3(k) and 7(a)

실시예 3(k)-3: 비통상적인 방식으로 사용된 다중핵산염기 기호 "n"

출원에 다음 서열이 개시되어 있다: 5'-aatgttggan-3'

여기에서 n은 c이다

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

3(k)절에 따르면, " 구체적으로 정의된 "핵산염기는 별첨 1, 섹션 1, 표 10에 나열된 기호 "n"으로 표현되는 것 이외의 핵산염기이다.

이 사례에서, "n"은 "c"만을 표현하기 위해 비통상적인 방식으로 사용된다. 본 개시에서 "n"은 "임의의 핵산염기"를 표현하기 위해 통상적인 방식으로 사용되지 않고 있다. 따라서 서열은 동등한 기존 기호, 즉 "c"가 서열에서 사용된 것처럼 해석하여야 한다(이 문서의 서론 참조). 따라서 고려해야 할 나열된 서열은 다음과 같다:

5'-aatgttggac-3'

이 서열은 10개의 구체적으로 정의된 핵산염기를 가지며 ST.26 7절(a)에 의해 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

서열은 다음과 같이 서열목록에 포함되어야 한다. aatgttggac (SEQ ID NO: 15)

관련 ST.26 절: 3(k) 및 7(a)

Example 3(k)-4: Ambiguity symbols other than “n” are “specifically defined”

A patent application describes the following sequence:

5' NNG KNG KNG KAG VCR 3'

w herein N, K, V, and R are IUPAC-IUB ambiguity codes

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

IUPAC-IUB ambiguity codes correspond to the list of nucleotide symbols defined in Annex I, Section 1, Table 1. According to paragraph 3(k), a “specifically defined” nucleotide is any nucleotide other than those represented by the symbol “n” listed in Annex I, Section 1, Table 1. Therefore, “K”, “V”, and “R” are “specifically defined” nucleotides.

The sequence has eleven enumerated and “specifically defined” nucleotides and is required by ST.26 paragraph 7(a) to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence must be included in a sequence listing as:

nngkngkngkagvcr (SEQ ID NO: 16)

Relevant ST.26 paragraphs: 3(k), 7(a) and 15

실시예 3(k)-4: "n" 이외의 다중핵산염기 기호는 "구체적으로 정의된"이다

특허출원에 다음 서열이 기술되어 있다:

5'NNG KNG KNG KAG VCR 3'

여기서 N, K, V 및 R은 IUPAC-IUB 다중핵산염기 코드이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

IUPAC-IUB 다중핵산염기 코드는 별첨 I, 섹션 1, 표 1에 정의된 핵산염기 기호 목록에 해당한다. 3(k)절에 따르면, "구체적으로 정의된" 핵산염기는 "n" 기호로 표현되는 것 이외의 핵산염기이다. 따라서, "K", "V" 및 "R"은 "구체적으로 정의된" 핵산염기이다.

서열은 "구체적으로 정의된" 11개의 나열된 핵산염기를 가지며 ST.26의 7절(a)에 의해 서열목록에 포함된다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

서열은 다음과 같이 서열목록에 포함되어야 한다:

Nngkngkngkagvcr (SEQ ID NO: 16)

관련 ST.26 절: 3(k), 7(a) 및 15

Example 3(k)-5: Ambiguity abbreviation “Xaa” used in a nonconventional manner

A patent application describes the following sequence:

Xaa-Tyr-Glu-Xaa-Xaa-Xaa-Leu

Wherein Xaa in position 1 is any amino acid, Xaa in position 4 is Lys, Xaa in position 5 is Gly and Xaa in position 6 is Leucine or Isoleucine.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated peptide in the formula provides three specifically defined amino acids in positions 2, 3 and 7. The first amino acid is represented by a conventional abbreviation, i.e., Xaa, representing any amino acid. However, the 4th, 5th and 6th amino acids are represented by a conventional abbreviation used in a nonconventional manner (see Introduction to this document). Therefore, the explanation of the sequence in the disclosure is consulted to determine the definition of “Xaa” in these positions. Since “Xaa” in positions 4-6 are indicated as a specific amino acid, the sequence must be interpreted as if the equivalent conventional abbreviations had been used in the sequence, i.e. Lys, Gly, and (Leu or Ile). Consequently, the sequence contains four or more specifically defined amino acids and must be included in a sequence listing as required by ST.26 paragraph 7(b).

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence uses a conventional abbreviation “Xaa” in a nonconventional manner. Therefore, the explanation of the sequence in the disclosure must be consulted to determine the definition of “Xaa” in positions 4, 5 and 6. The explanation defines “Xaa” as a lysine in position 4, a glycine in position 5 and a leucine or isoleucine in position 6. The conventional symbols for these amino acids are K, G, and J respectively. Therefore, the sequence should be represented as in the sequence listing as:

XYEKGJL (SEQ ID NO: 17)

According to paragraph 27, “X” will be construed as any one of A, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, or “V”, except where it is used with a further description in the feature table. Since “X” at position 1 of SEQ ID NO: 17 represents “any amino acid”, it must be annotated with the feature key VARIANT and a note qualifier with the value, “X can be any amino acid”.

Where practicable, each “X” should be annotated individually. However, a region of contiguous “X” residues, or a multitude of “X” residues dispersed throughout the sequence, may be jointly described with the feature key VARIANT using the syntax “x..y” as the location descriptor, where x and y are the positions of the first and last “X” residues, and a note qualifier with the value, “X can be any amino acid”.

Relevant ST.26 paragraphs: 3(k), 7(b), 26, and 27

실시예 3(k)-5: 비통상적인 방식으로 사용되는 다중아미노산 약어 "Xaa"

특허출원에 다음 서열이 기술되어 있다:

Xaa-Tyr-Glu-Xaa-Xaa-Xaa-Leu

여기에서, 위치 1의 Xaa는 임의의 아미노산이고, 위치 4의 Xaa는 Lys이고, 위치 5의 Xaa는 Gly이고, 위치 6의 Xaa는 류신(Leucine) 또는 이소류신(Isoleucine)이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

위의 식에서 나열된 펩티드는 위치 2, 3 및 7에 3개의 구체적으로 정의된 아미노산을 제공한다. 제 1 아미노산은 임의의 아미노산을 표현하는 통상적인 약어, 즉 Xaa로 표현된다. 그러나, 4번째, 5번째 및 6번째 아미노산은 비통상적인 방식으로 사용되는 통상적인 약어로 표현된다(본 문서의 서론 참조). 따라서, 이들 위치에서 "Xaa"의 정의를 결정하기 위해 본 개시의 서열 설명을 참고한다. 위치 4-6의 "Xaa"는 특정 아미노산으로 표시되므로, 서열은 동등한 통상적인 약어, 즉 Lys, Gly 및 (Leu 또는 Ile)가 서열에서 사용된 것처럼 해석되어야 한다. 결과적으로, 서열은 4개 이상의 구체적으로 정의된 아미노산을 가지며 ST.26의 7절(b)에 따라 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

이 서열은 기존의 약어 "Xaa"를 비통상적인 방식으로 사용한다. 따라서, 위치 4, 5 및 6에서 "Xaa"의 정의를 결정하기 위해 본 개시의 서열 설명을 참고하여야 한다. 설명에서는 "Xaa"를 위치 4의 라이신(lysine), 위치 5의 글리신(glycine) 및 위치 6의 류신(Leucine) 또는 이소류신(Isoleucine)으로 정의한다. 이들 아미노산에 대한 통상적인 기호는 각각 K, G 및 J이다. 따라서 서열은 다음과 같이 서열목록에 표현되어야 한다.

XYEKGJL (SEQ ID NO: 17)

27절에 따르면, "X"는 특징 표에서 추가 설명과 함께 사용되는 경우를 제외하고는, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다. 서열 번호 17의 첫 번째 위치의 "X"는 "임의의 아미노산"을 표현하므로 VARIANT 특징 기호와 "X can be any amino acid"라는 값을 가진 note 한정자로 주석을 달아야 한다.

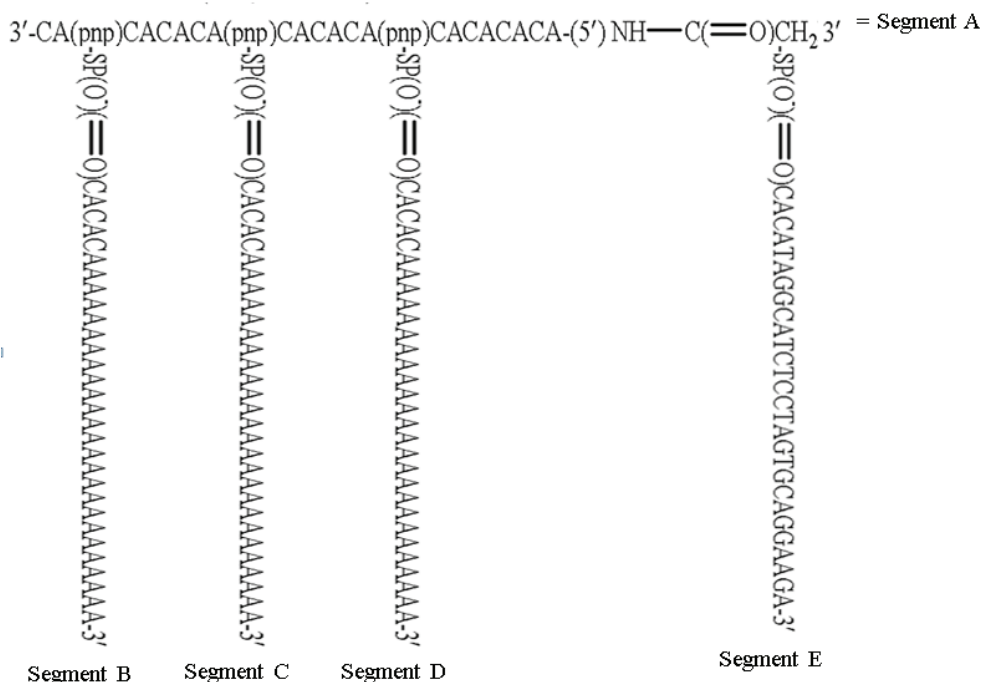
가능한 경우, 각 "X"는 독립적으로 주석을 달아야 한다. 그러나, 연속 "X" 잔기의 영역, 또는 서열 전체에 분산된 다수의 "X" 잔기는 위치 설명자로서 "x..y" 구문을 사용하는 특징 기호 'VARIANT'와 "X can be any amino acid"라는 값을 가진 'note' 한정자를 이용하여 함께 기술될 수 있다. 여기서 x와 y는 첫 번째 및 마지막 "X" 잔기의 위치이다.

관련 ST.26 절: 3(k), 7(b), 26 및 27

Paragraph 7(a) – Nucleotide sequences required in a sequence listing

Example 7(a)-1: Branched nucleotide sequence

The description discloses the following branched nucleotide sequence:



w herein "pnp" is a linkage or monomer containing an bromoacetyl amino functionality;
 3'-CA(pnp)CACACA(pnp)CACACA(pnp)CACACACA-(5')NH—C(=O)CH₂ 3' is segment A;
 SP(O)(=O)CACACAAAAAAAAAAAAAAAAAAAAA 3' is segments B, C, and D; and
 SP(O)(=O)CACATAGGCATCTCTAGTGCAGGAAGA 3' is segment E.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES – the four vertical segments B-E must be included in a sequence listing

NO – the horizontal segment A must not be included in a sequence listing

The above figure is an example of a "comb-type" branched nucleic acid sequence containing five linear segments: the horizontal segment A and the four vertical segments B-E.

According to paragraph 7(a), the linear regions of branched nucleotide sequences containing ten or more specifically defined nucleotides, wherein adjacent nucleotides are joined 3' to 5', must be included in a sequence listing.

The four vertical segments B-E each contain more than ten specifically defined nucleotides, wherein adjacent nucleotides are joined 3' to 5', and therefore each is required to be included in a sequence listing.

In horizontal segment A, the linear regions of the nucleotide sequence are linked by the non-nucleotide moiety "pnp" and each of these linked linear regions contains fewer than ten specifically defined nucleotides. Therefore,

since no region of segment A contains ten or more specifically defined nucleotides wherein adjacent nucleotides are joined 3' to 5', they are not required by ST.26 paragraph 7(a) to be included in a sequence listing.

Question2: Does ST.26 permit inclusion of the sequence(s)?**NO**

According to paragraph 8, "A sequence listing must not include any sequences having fewer than ten specifically defined nucleotides...."

No region of Segment A contains ten or more specifically defined nucleotides wherein adjacent nucleotides are joined 3' to 5'; therefore, it must not be included in a sequence listing as a separate sequence with its own sequence identification number.

How ever, segments B, C, D, and E may be annotated to indicate that they are linked to segment A.

Question3: How should the sequence(s) be represented in the sequence listing?

Segments B, C, and D are identical and must be included in a sequence listing as a single sequence:

cacacaaaaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO: 18)

The first "c" in the sequence should be further described using the feature key "misc_feature" and the qualifier "note" with the value e.g., "This sequence is one of four branches of a branched polynucleotide".

Segment E must be included in a sequence listing as a single sequence:

cacataggcatctcctagtcaggaaga (SEQ ID NO: 19)

The first "c" in the sequence should be further described using the feature key "misc_feature" and the qualifier "note" with the value e.g., "This sequence is one of four branches of a branched polynucleotide."

Relevant ST.26 paragraph(s): 7(a), 8, 11, 13, and 17

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가?아니오

8절의 "서열목록에는 구체적으로 정의된 핵산염기가 10개 미만인 서열이 포함되어서는 아니된다..."에 따라

단편 A의 어떠한 영역도 인접한 핵산염기가 3'-말단에서 5'-말단으로 연결된 10 개 이상의 구체적으로 정의된 핵산염기를 함유하지 않으며, 따라서, 고유한 서열식별번호를 갖는 별도의 서열로서 서열목록에 포함되어서는 아니된다.

그러나, 단편 B, C, D 및 E가 단편 A에 연결되어 있다고 표시하기 위해 주석이 달릴 수 있다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

단편 B, C 및 D는 동일하며 단일 서열로 서열목록에 포함되어야 한다:

cacacaaaaaaaaaaaaaaaaaaaaaaaaa (SEQ ID NO: 18)

서열에서 첫 번째 "c"는 특징 기호 "misc_feature"와 "This sequence is one of four branches of a branched polynucleotide"와 같은 값을 가진 한정자 "note"를 사용하여 추가로 기술되어야 한다.

단편 E는 단일 서열로 서열목록에 포함되어야 한다:

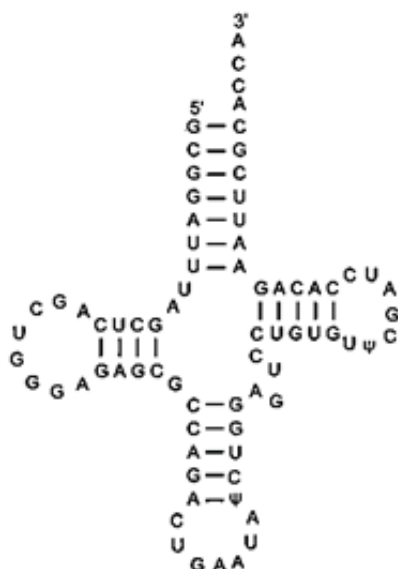
cacataggcatcctcctagtcaggaaga (SEQ ID NO: 19)

서열에서 첫 번째 "c"는 특징 기호 "misc_feature"와 "This sequence is one of four branches of a branched polynucleotide"와 같은 값을 가진 한정자 "note"를 사용하여 추가로 기술되어야 한다.

관련 ST.26 절: 7(a), 8, 11, 13 및 17

Example 7(a)-2: Linear nucleotide sequence having a secondary structure

A patent application describes the following sequence:



Wherein Ψ is pseudouridine.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The nucleotide sequence contains seventy-three enumerated and specifically defined nucleotides. Thus, the example has ten or more “specifically defined” nucleotides, and as required by ST.26 paragraph (7)(a), must be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

Consultation of the disclosure indicates that “Ψ” is equivalent to pseudouridine. The only conventional symbol that can be used to represent pseudouridine is “n”; therefore, the “Ψ” is a nonconventional symbol used to represent the conventional symbol “n” (see Introduction to this document). Accordingly, the sequence must be interpreted to have two “n” symbols in place of the two “Ψ” symbols.

The symbol “u” must not be used to represent uracil in an RNA molecule in the sequence listing. According to paragraph 14, the symbol “t” will be construed as uracil in RNA. The sequence must be included as:

gcggtattagctcagctggagagcgccagactgaatanctggagctctgtgncgatccacagaattgcacca (SEQ ID NO: 20)

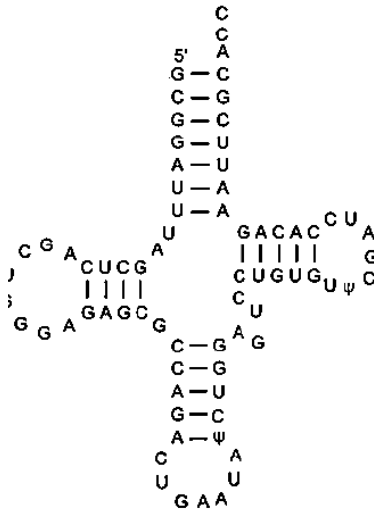
The value of the mandatory “mol type” qualifier of the mandatory “source” feature key is “tRNA”. Additional information may be provided with feature key “tRNA” and any appropriate qualifier(s).

The “n” residues must be further described in a feature table using the feature key “modified base” and the mandatory qualifier “mod_base” with the abbreviation “p” for pseudouridine as the qualifier value (see Annex 1, Table 2).

Relevant ST.26 paragraph(s): 7(a), 11, 13, 14, 17, 62, 84 and Annex I, sections 2 and 5, feature key 5.43

실시예 7(a)-2: 이차구조를 갖는 선형(linear) 핵산염기 서열

특허출원에 다음 서열이 기술되어 있다:



여기에서, ψ는 슈도우리딘(pseudouridine)이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

핵산염기 서열은 73개의 나열되고 구체적으로 정의된 핵산염기를 함유한다. 따라서, 이 사례는 10개 이상의 "구체적으로 정의된" 핵산염기를 갖고, ST.26의 7(a)절에 따라 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

본 개시를 참고하면, "ψ"가 슈도우리딘(pseudouridine)과 동등함을 나타낸다. 슈도우리딘을 표현하는데 사용될 수 있는 유일한 통상적인 기호는 "n"이고; 따라서 "ψ"는 일반적인 기호 "n"을 표현하는 목적으로 사용되는 비통상적인 기호이다(이 문서의 서론 참조) 따라서 서열은 2개의 "ψ" 기호 대신 2개의 "n" 기호를 갖는 것으로 해석되어야 한다.

"u" 기호를 서열목록에서 RNA 분자의 우라실을 표현하기 위해 사용해서는 아니된다. 14절에 따르면, 기호 "t"는 RNA에서 우라실로 해석된다. 서열은 다음과 같이 포함되어야 한다.

cgggattagctcagctgggagagcgcagactgaatanctggagtcctgtgncgatccacagaattcgcacca (SEQ ID NO: 20)

"source" 필수 특징 기호의 "mol_type" 필수 한정자 값은 "tRNA"이다. 특징 기호 "tRNA"와 적절한 한정자(들)에 추가 정보가 제공될 수 있다.

"n" 잔기는 특징 기호 "modified_base"와 필수 한정자 "mod_base"의 한정자 값으로 슈도우리딘에 대해 "p" 약어를 사용하여 특징 표에서 추가적으로 설명되어야 한다(별첨 1, 표 2 참조).

관련 ST.26 절: 7(a), 11, 13, 14, 17, 62, 84 및 별첨 1, 섹션 2 및 5, 특징 기호 5.43

Example 7(a)-3: Nucleotide ambiguity symbols used in a nonconventional manner

A patent application describes the following sequence:

5' GATC-MDR-MDR-MDR-MDR-GTAC 3'

The explanation of the sequence in the disclosure further indicates: "A "DR Element" consists of the sequence 5' ATCA GCCAT 3'. A mutant DR Element, or MDR, is a DR element wherein the middle 5 nucleotides, CAGCC, are mutated to TTTTT."

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated sequence uses the symbol "MDR". Where it is unclear if a symbol used in a sequence is intended to be a conventional symbol, i.e., a symbol set forth in Annex 1, Section 3, Table 3, or a nonconventional symbol, the explanation of the sequence in the disclosure must be consulted to make a determination (see Introduction to this document). According to Table 3, "MDR" could be interpreted as three conventional symbols (m = a or c, d = a or g or t/u, r = g or a) or as an abbreviation that is short-hand notation for some other structure.

Consultation of the disclosure indicates that an MDR element is equivalent to 5' ATTTTTTAT 3'. The letters "MDR" are considered conventional symbols used in a nonconventional manner; therefore, the sequence must be interpreted as though it were disclosed using the equivalent conventional symbols. Accordingly, the enumerated sequence that is considered for inclusion in a sequence listing is:

5' GATC ATTTTTTAT ATTTTTTAT ATTTTTTAT ATTTTTTAT GTAC 3'

The enumerated sequence has 44 specifically defined nucleotides and is required by ST.26 paragraph 7(a) to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence must be included in a sequence listing as:

gatcatttttatatttttatatttttatatttttatgtac (SEQ ID NO: 21)

Relevant ST.26 paragraphs: 7(a) and 13

실시예 7(a)-3: 비통상적 방식으로 사용된 다중핵산염기 기호

특허출원에 다음 서열이 기술되어 있다:

5'GATC-MDR-MDR-MDR-MDR-GTAC 3'

본 개시의 서열 설명은 다음과 같다: "하나의 "DR Element"가 서열 5'ATCAGCCAT 3'로 구성된다고 추가로 기재하고 있다. 돌연변이 DR 엘리먼트 또는 MDR은 중간 5개 핵산염기 CAGCC가 TTTT로 돌연변이되는 DR 엘리먼트이다."

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

나열된 서열은 "MDR" 기호를 사용한다. 서열에 사용된 기호가 종래의 기호, 즉 별첨 1, 섹션 3, 표 3에 설명된 기호 또는 비통상적인 기호인지의 여부가 불분명한 경우, 결정을 위해 본 개시의 서열 설명을 참고하여야 한다(이 문서의 서론 참조). 표 3에 따르면, "MDR"은 3개의 통상적인 기호(m=a 또는 c, d = a 또는 g 또는 t/u, r = g 또는 a)나 어떤 다른 구조에 대한 간략식 표기법으로 해석될 수 있다.

본 개시를 참고하면, MDR 요소가 5'ATTTTTTAT 3'와 동일함을 나타낸다. 문자 "MDR"은 비통상적인 방식으로 사용되는 일반적인 기호로 간주된다. 그러므로, 서열은 그것이 동등한 종래의 기호를 사용하여 개시된 것처럼 해석되어야 한다. 따라서, 서열목록에 포함되는 것으로 간주되는 나열된 서열은 다음과 같다:

5'GATC ATTTTTTAT ATTTTTTAT ATTTTTTAT ATTTTTTAT GTAC 3'

나열된 서열은 44개의 구체적으로 정의된 핵산염기를 가지며 ST.26의 7(a) 절에 의해 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

서열은 다음과 같이 서열목록에 포함되어야 한다.

gatcatttttatattttatattttatattttatgtac (SEQ ID NO: 21)

관련 ST.26 절: 7(a) 및 13

Example 7(a)-4: Nucleotide ambiguity symbols used in a nonconventional manner

A patent application describes the following sequence:

5' ATTC-N-N-N-N-GTAC 3'

The explanation of the sequence in the disclosure further indicates that "N" consists of the sequence 5' ATACGCACT 3'.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated sequence uses the symbol "N". The explanation of the sequence in the disclosure must be consulted to determine if the "N" is used in a conventional or nonconventional manner (see Introduction to this document).

Consultation of the disclosure indicates that "N" is equivalent to 5' ATACGCACT 3'. Thus, the "N" is a conventional symbol used in a nonconventional manner. Accordingly, the sequence must be interpreted as though it were disclosed using the equivalent conventional symbols:

5' ATTC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3'

The enumerated sequence has 44 specifically defined nucleotides and is required by ST.26 paragraph 7(a) to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence must be included in a sequence listing as:

atctacatcgactatcgactatcgactatcgactatcgactgtac (SEQ ID NO: 22)

Relevant ST.26 paragraphs: 7(a) and 13

실시예 7(a)-4: 비통상적 방식으로 사용된 다중핵산염기 기호

특허출원에 다음 서열이 기술되어 있다:

5'ATTC-N-N-N-N-GTAC 3'

본 개시의 서열 설명은 "N"이 서열 5'ATACGCACT 3'로 구성된다고 추가로 기재하고 있다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

나열된 서열은 "N"기호를 사용한다. "N"이 통상적 방식 또는 비통상적인 방식으로 사용되는지를 결정하기 위해 본 개시의 서열의 설명을 참조하여야 한다(본 문서의 서론 참조).

본 개시를 참고하면, "N"이 5'ATACGCACT 3'와 동일함을 나타낸다. 그러므로, "N"은 비통상적인 방식으로 사용되는 종래의 기호이다. 따라서 서열은 동등한 기존 기호를 사용하여 공개된 것처럼 해석하여야 한다.

5'ATTC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3'

나열된 서열은 44개의 구체적으로 정의된 핵산염기를 가지며 ST.26의 7(a)절에 의해 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

서열은 서열목록에 다음과 같이 포함되어야 한다:

attcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO: 22)

관련 ST.26 절: 7(a) 및 13

Example 7(a)-5: Nonconventional nucleotide symbols

A patent application describes the following sequence:

5' GATC-β-β-β-β-GTAC3'

The explanation of the sequence in the disclosure further indicates that "β" consists of the sequence 5' ATACGCACT 3'.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated sequence uses the nonconventional symbol "β". The explanation of the sequence in the disclosure must be consulted to determine the meaning of "β" (see Introduction to this document).

Consultation of the disclosure indicates that "β" is equivalent to 5' ATACGCACT 3'. Thus, the "β" is a nonconventional symbol used to represent a sequence of nine specifically defined, conventional symbols. Accordingly, the sequence must be interpreted as though it were disclosed using the equivalent conventional symbols:

5' GATC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3'

The enumerated sequence has 44 specifically defined nucleotides and is required by ST.26 paragraph 7(a) to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence must be included in a sequence listing as:

gatcatagcactatagcactatagcactatagcactatagcactgtac (SEQ ID NO: 23)

Relevant ST.26 paragraphs: 7(a) and 13

실시예 7(a)-5: 비통상적 핵산염기 기호

특허출원에 다음 서열이 기술되어 있다:

5'GATC-β-β-β-GTAC 3'

본 개시의 서열 설명은 "β"가 서열 5' ATACGCACT 3'로 구성된다고 추가로 기재하고 있다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

나열된 서열은 비통상적인 기호 "β"를 사용한다. "β"의 의미를 결정하기 위해 본 개시의 서열 설명을 참고하여야 한다(본 문서의 서론 참조).

본 개시를 참고하면, "β"가 5'ATACGCACT 3'와 동일함을 제시한다. 따라서, "β"는 9개의 구체적으로 정의된 통상적인 기호의 서열을 표현하는 데에 사용되는 비통상적인 기호이다. 따라서 서열은 동등한 기존 기호를 사용하여 공개된 것처럼 해석하여야 한다.

5'GATC-ATACGCACT-ATACGCACT-ATACGCACT-ATACGCACT-GTAC 3'

나열된 서열은 44개의 구체적으로 정의된 핵산염기를 가지며 ST.26의 7(a)절에 의해 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

서열은 다음과 같이 서열목록에 포함되어야 한다.

gatcatacgcactatacgcactatacgcactatacgcactgtac (SEQ ID NO: 23)

관련 ST.26 절: 7(a) 및 13

Example 7(a)-6: Nonconventional nucleotide symbols

A patent application describes the following sequence:

5' GATC-β-β-β-GTAC3'

The explanation of the sequence in the disclosure further indicates that "β" is equal to adenine, inosine, or pseudouridine.

Question 1: Does ST.26 require inclusion of the sequence(s)?

NO

The enumerated sequence uses the nonconventional symbol "β". The explanation of the sequence in the disclosure must be consulted to determine the meaning of "β" (see Introduction to this document).

Consultation of the disclosure indicates that "β" is equivalent to adenine, inosine, or pseudouridine. The only conventional symbol that can be used to represent "adenine, inosine, or pseudouridine" is "n"; therefore, the "β" is a nonconventional symbol used to represent the conventional symbol "n". Accordingly, the sequence must be interpreted to have four "n" symbols (shown as "N" below) in place of the four "β" symbols:

5' GATC-N-N-N-N-GTAC 3'

The enumerated sequence has only eight specifically defined nucleotides and is not required by ST.26 paragraph 7(a) to be included in a sequence listing.

Question 2: Does ST.26 permit inclusion of the sequence(s)?

NO

The enumerated sequence, 5' GATC-N-N-N-N-GTAC 3' must not be included in a sequence listing.

However, a disclosed alternative sequence may be included in a sequence listing if at least 2 of the "n" symbols are replaced by adenine, resulting in a sequence with at least 10 or more specifically defined nucleotides.

Question 3: How should the sequence(s) be represented in the sequence listing?

One possible permitted representation is:

gatcaaaagtac (SEQ ID NO: 24)

In the above example, the four adenine nucleotides that replace the β symbols should be annotated to note that these positions could be substituted with inosine or pseudouridine.

The feature key "misc_difference" should be used with a feature location 5-8 and a qualifier "note" with the value, e.g., "A nucleotide in any of positions 5-8 may be replaced with inosine or pseudouridine". Since these alternatives are modified nucleotides, then the feature key "modified_base" together with the qualifier "mod_base" would be required. The value for the "mod_base" qualifier can be "OTHER" with a "note" qualifier and the value of "i or p".

Other permutations are possible.

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraphs: 7(a), 8, 13, and 17

실시예 7(a)-6: 비정상적 핵산염기 기호

특허출원에 다음 서열이 기술되어 있다:

5'GATC-β-β-β-GTAC 3'

본 개시의 서열 설명은 "β"가 아데닌(adenine), 이노신(nosine) 또는 슈도우리딘(pseudouridine)과 동일하다고 추가로 기재하고 있다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

아니오

나열된 서열은 비정상적인 기호 "β"를 사용한다. "β"의 의미를 결정하기 위해 본 개시의 서열 설명을 참고하여야 한다(본 문서의 서론 참조).

본 개시를 참고하면, "β"는 아데닌(adenine), 이노신(nosine) 또는 슈도우리딘(pseudouridine)과 동등하다. "아데닌, 이노신 또는 슈도우리딘"을 표현하는 데에 사용될 수 있는 유일한 통상적인 기호는 "n"이고; 따라서, "β"는 종래의 기호 "n"을 표현하는 데에 사용되는 비정상적인 기호이다. 따라서, 서열은 4개의 "β" 기호 대신에 4개의 "n" 기호(하기 "N"으로 표시)를 갖는 것으로 해석되어야 한다:

5'GATC-N-N-N-N-GTAC 3'

나열된 서열은 8개의 구체적으로 정의된 핵산염기를 가지며 ST.26의 7(a)절에 의해 서열목록에 포함되지 아니한다.

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가?

아니오

나열된 서열 5'GATC-N-N-N-N-GTAC 3'은 서열목록에 포함되지 않아야 한다.

그러나, "n" 기호 중 2개 이상이 아데닌(adenine)으로 대체되어 10개 이상의 구체적으로 정의된 핵산염기를 갖는 서열이 된다면, 개시된 선택적 서열이 서열목록에 포함될 수 있다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

한 가지 가능한 허용된 표현은 다음과 같다:

gatcaaaagtac (SEQ ID NO: 24)

위 사례에서, β 기호를 대체하는 4개의 아데닌 핵산염기는 주석 처리되어야 한다.

이들 위치는 이노신(nosine) 또는 슈도우리딘(pseudouridine)으로 치환될 수 있다.

특징 기호 "misc_difference"는 특징 위치 5-8과 한정자 값을 가진 "note" 한정자와 같이 사용되어야 한다. 예를 들면, 한정자 값으로 "A nucleotide in any of positions 5-8 may be replaced with inosine or pseudouridine"이다. 이들 대체는 변형 핵산염기이기 이므로, 특징 기호 "modified_base"와 한정자 "mod_base"가 필요할 수 있다. "mod_base" 한정자 값은 "note" 한정자와 "i 또는 p"값을 가진 "OTHER"일 수 있다.

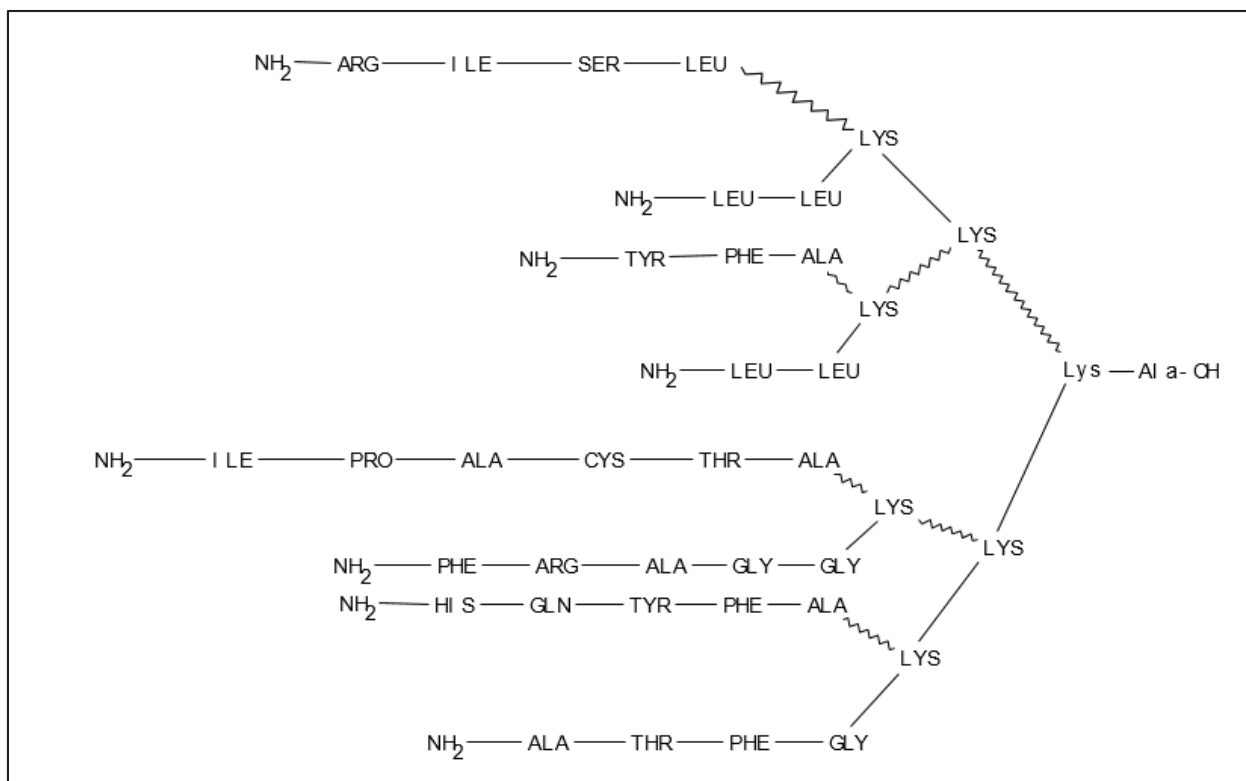
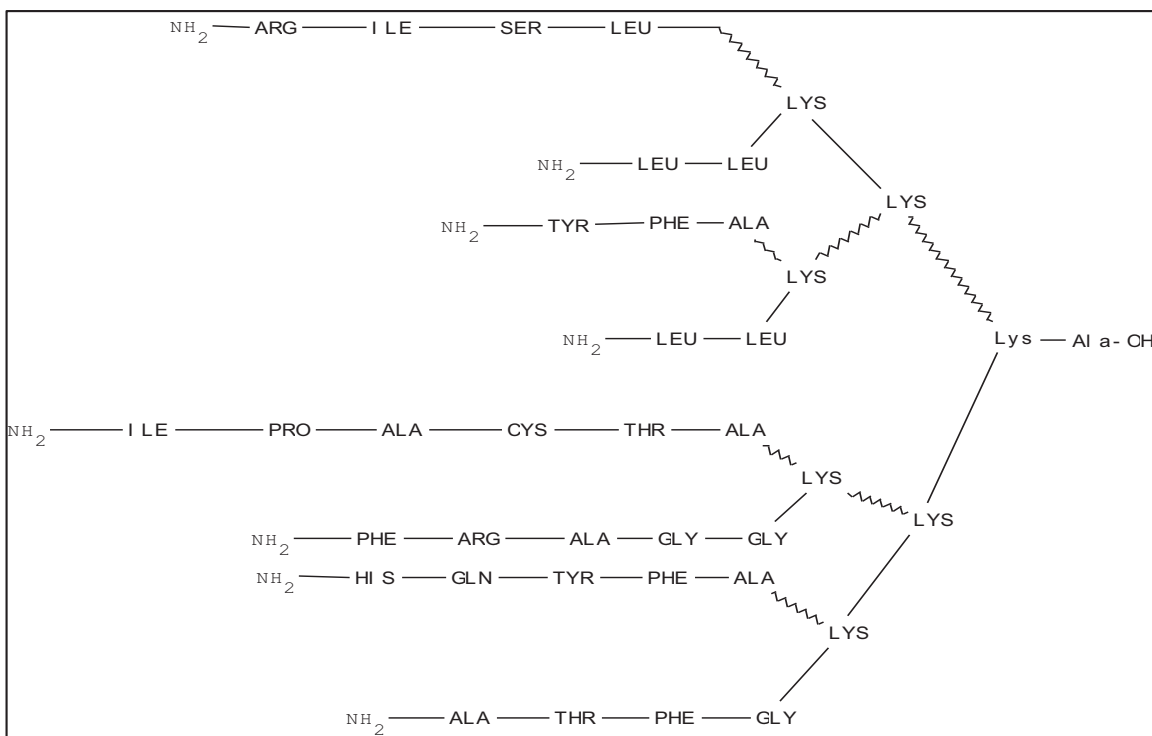
다른 서열도 가능하다.

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 7(a), 8, 13 및 17

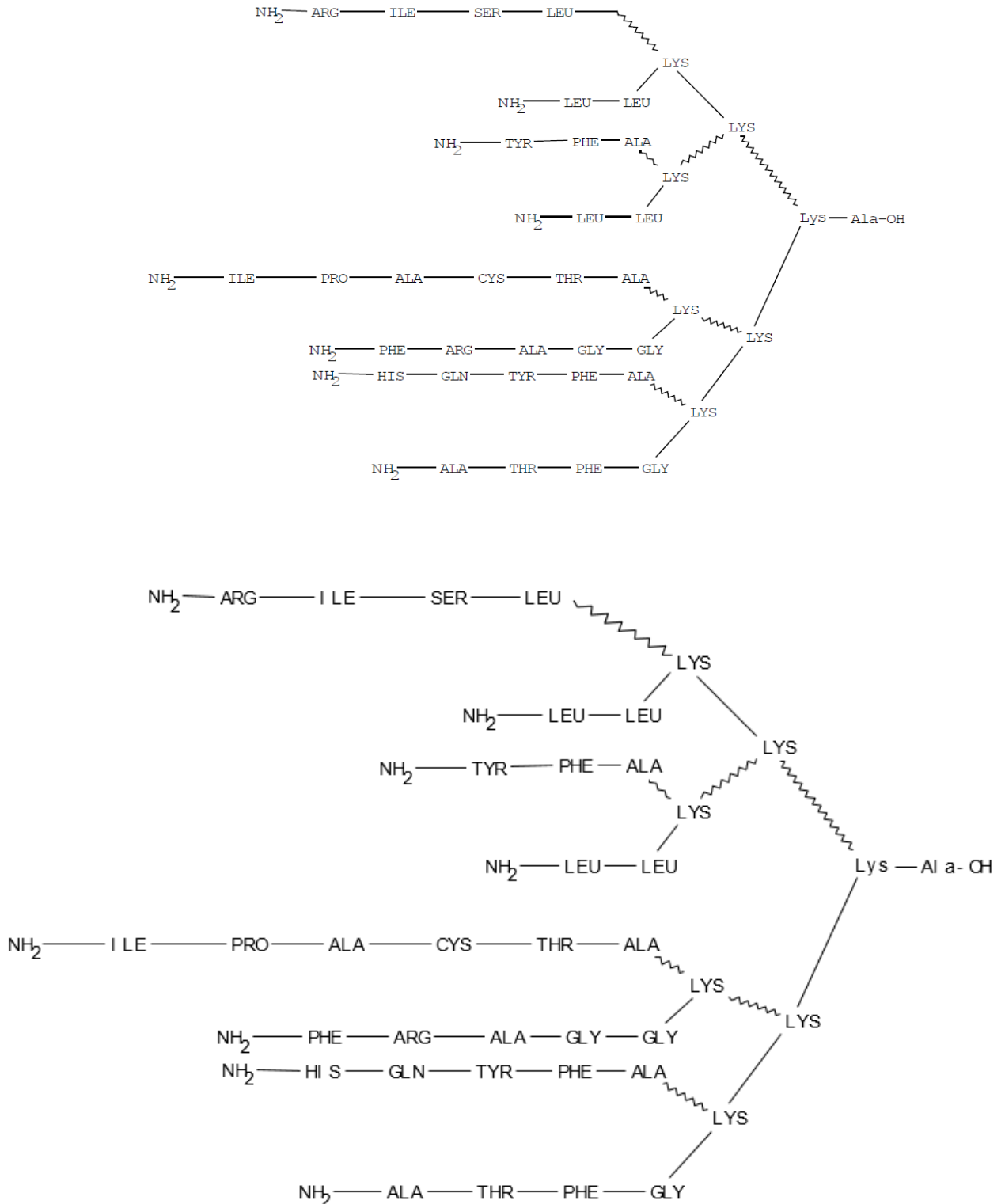
Example 7(b)-2: Branched amino acid sequence

The application describes a branched sequence where the Lysine residues are used as a scaffolding core to form eight branches to which multiple linear peptide chains are attached. Lysine is a dibasic amino acid, providing it with two sites for peptide-bonding. The peptide is illustrated as follows:



실시예 7(b)-2: 분지형(branched) 아미노산 서열

본 출원은 라이신(Lysine) 잔기가 스캐폴딩(scaf folding) 코어로서 사용되어 다중 선형 펩티드 사슬이 부착된 8의 분지형을 형성하는 분지형 서열을 기술한다. 라이신(Lysine)은 이염기성(dibasic) 아미노산으로서, 펩티드 결합을 위한 2개의 부위를 제공한다. 펩티드는 다음과 같이 예시된다:

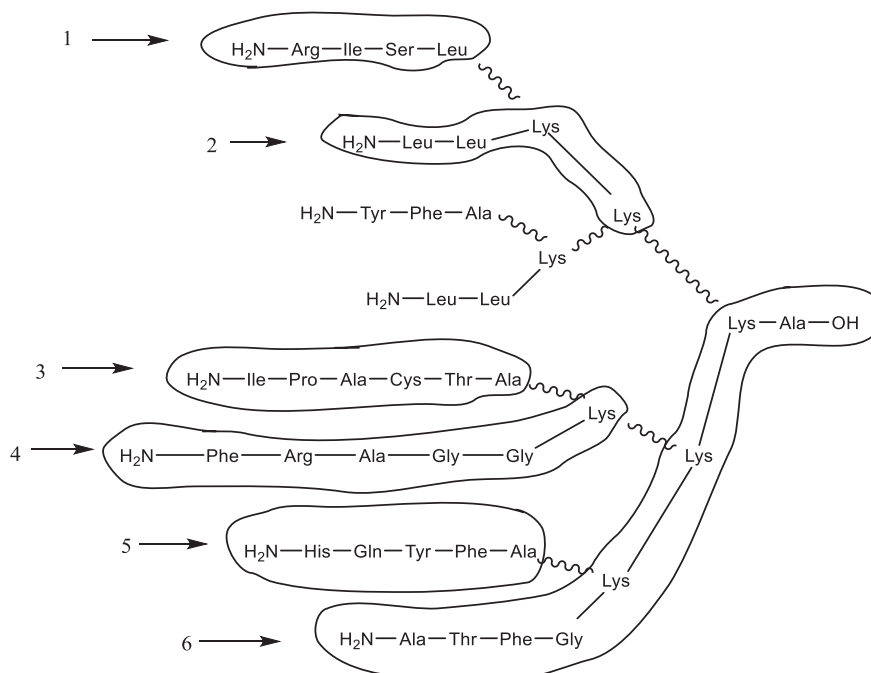


In the above branched peptide, the bonds between a lysine and another amino acid depicted by — represent an amide linkage between the terminal amine of the lysine and the carboxyl end of the bonded amino acid. The bonds depicted by ~ represent an amide linkage between the side chain amine of the lysine and the carboxyl end of the bonded amino acid.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The example discloses a branched sequence where the lysine residues are used as a scaffolding. Paragraph 7(b) requires that the unbranched or linear region of the sequence, containing four or more specifically defined amino acids, be included in a sequence listing. In the above example, the linear regions of the branched peptide that have four or more specifically defined amino acids are encircled:



ST.26 paragraph 7(b) requires inclusion of peptides 1-6 above in a sequence listing.

Peptides which are not required to be included in the sequence listing are:

YFA

LLK

Question 2: Does ST.26 permit inclusion of the sequence(s)?

NO

According to paragraph 8, a sequence listing must not include any sequences having fewer than four specifically defined amino acids.

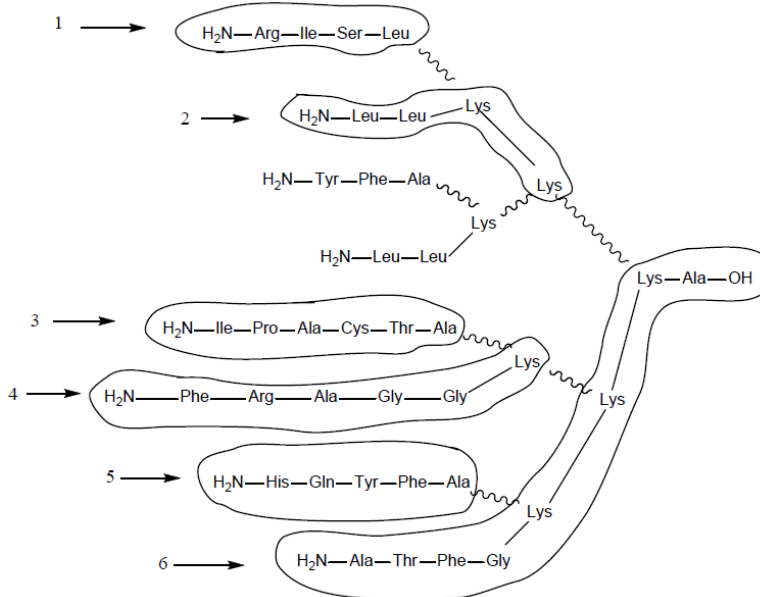
The peptides YFA and LLK each contain only three specifically defined amino acids and therefore, they must not be included in a sequence listing as separate sequences with their own sequence identification numbers.

상기 분지형 펩티드에서, —————로 표현되는 라이신(Lysine)과 다른 아미노산 사이의 결합은 라이신(Lysine)의 아민 말단과 결합된 아미노산의 카르복실 말단 사이의 아마이드 결합을 나타낸다. ~~~~~로 표현된 결합(Lysine)은 라이신의 측쇄(side chain)인 아민과 카르복실 말단의 결합된 아미노산 사이의 아마이드 결합을 나타낸다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

이 사례에는 라이신(Lysine) 잔기가 발판으로서 사용되는 분지형 서열이 개시되어 있다. 7(b)절에 따라 네 개 이상의 구체적으로 정의된 아미노산을 함유하는 서열의 분지형 또는 선형 영역은 서열목록에 포함되어야 한다. 위의 사례에서, 4개 이상의 구체적으로 정의된 아미노산을 갖는 분지형 펩티드의 선형 영역은 다음과 같이 둘러싸여 있다:



ST.26 7(b)절에 따라 서열목록에 위의 펩티드 1-6을 포함하여야 한다.

서열목록에 포함될 필요가 없는 펩티드는 다음과 같다:

YFA
LLK

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가?

아니오

8절에 따르면, 서열목록은 구체적으로 정의된 4개 미만의 아미노산을 갖는 서열을 포함해서는 아니된다.

펩티드 YFA 및 LLK는 각각 3개의 구체적으로 정의된 아미노산만을 함유하므로, 이들은 고유 서열식별번호를 갖는 별도의 서열로서 서열목록에 포함되어서는 아니된다.

Question 3: How should the sequence(s) be represented in the sequence listing?

Peptides 1-6 must be represented with separate sequence identifiers:

RISL (SEQ ID NO: 26)

LLKK (SEQ ID NO: 27)

IPACTA (SEQ ID NO: 28)

FRAGGK (SEQ ID NO: 29)

HQYFA (SEQ ID NO: 30)

ATFGKKKA (SEQ ID NO: 31)

The cross linkage may be annotated using the feature key "SITE" and the mandatory qualifier "note" with the value e.g., "This sequence is one part of a branched amino acid sequence". According to ST.26 paragraph 30, SEQ ID Nos 27, 29, and 31, must include an annotation for each lysine to indicate that it is a modified amino acid, using the feature key "SITE" together with the qualifier "note" describing that the side chain of the lysine is linked via an amide linkage to another sequence. Each of the SEQ ID Nos 26, 28, and 30 should include an annotation to indicate that the C-terminal amino acid is linked to another sequence, using the feature key "SITE" together with the qualifier "note".

Relevant ST.26 paragraph(s): 7(b), 8, 26, 29, 30, and 31

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

펩티드 1-6은 별도의 서열 식별자로 표현되어야 한다:

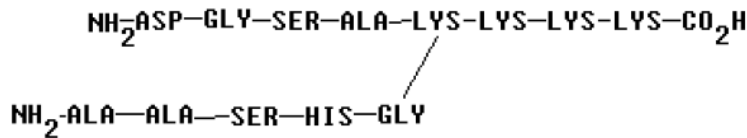
RISL (SEQ ID NO: 26)
LLKK (SEQ ID NO: 27)
IPACTA (SEQ ID NO: 28)
FRAGGK (SEQ ID NO: 29)
HQYFA (SEQ ID NO: 30)
ATFGKKKA (SEQ ID NO: 31)

가교(cross linkage)는 특징 기호 "SITE" 및 그 값(예를 들면, "This sequence if one part of a branched amino acid sequence")을 갖는 필수 한정자 "note"를 사용하여 주석이 되 수 있다. ST.26의 30절에 따르면, 서열 번호 27, 29 및 31은, 라이신(Lysine)의 분지가 아마이드 연결을 통해 다른 서열에 연결됨을 기술하는 한정자 "note"와 함께 특징 기호 "SITE"를 사용하여 라이신(Lysine)이 변형 아미노산임을 나타내는 각 라이신(Lysine)에 대한 주석을 포함하여야 한다. 서열 번호 26, 28 및 30 각각은 한정자 "note"와 함께 특징 기호 "SITE"를 사용하여 C-말단 아미노산이 다른 서열에 연결되어 있음을 표시하는 주석을 포함하여야 한다.

관련 ST.26 절: 7(b), 8, 26, 29, 30 및 31

실시예 7(b)-3: 분지형(branched) 아미노산 서열

다음 서열의 펩티드:

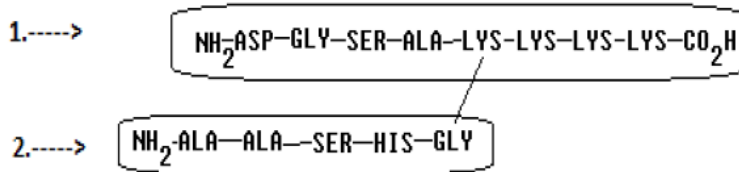


아래 서열의 말단 글리신(Glycine) 잔기와 위 서열의 라이신(Lysine) 사이의 연결은 글리신(Glycine)의 카르복시 말단과 라이신(Lysine)의 아미노 말단 측쇄 사이의 아마이드 결합을 통해 이루어진다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

4개 이상의 구체적으로 정의된 아미노산을 함유하는 서열의 비분지형 또는 선형 영역은 서열목록에 포함되어야 한다. 위의 사례에서 네 개 이상의 아미노산을 갖는 펩티드 분지의 선형 영역은 다음과 같다:



ST.26의 7(b)절에 따라 서열 1 및 2는 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

서열 1과 2는 별도의 서열 식별자로 표현하여야 한다:

DGSAK KKK (SEQ ID NO: 32)

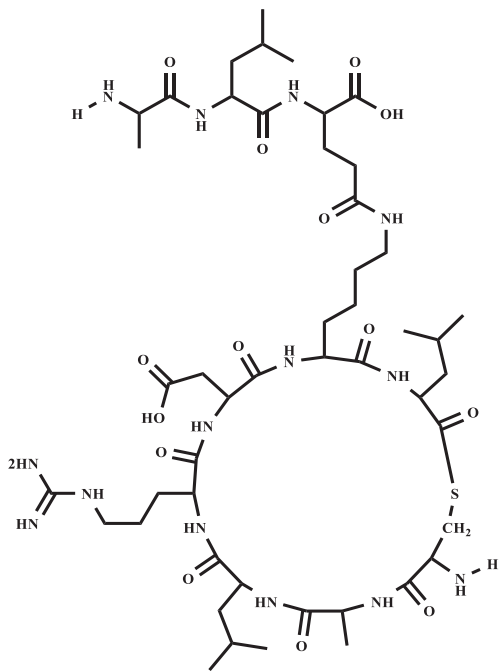
AASHG (SEQ ID NO: 33)

서열 DGSAK KKK는 위치 번호 5의 라이신(Lysine)이 변형 아미노산임을 표시하는 주석을 포함해야 하며, 특징 기호 "SITE"와 한정자 "note"를 함께 사용하여 라이신(Lysine)의 분지가 아마이드 연결을 통해 다른 서열로 연결되어 있음을 설명한다. 서열 AASHG는 위치 번호 5의 글리신(Glycine)이 한정자 "note"와 함께 특징 기호 "SITE"를 사용하여 다른 서열에 연결되어 있음을 표시하는 주석을 포함하여야 한다.

관련 ST.26 절: 7(b), 26, 29, 30 및 31

Example 7(b)-4: Cyclic peptide containing a branched amino acid sequence

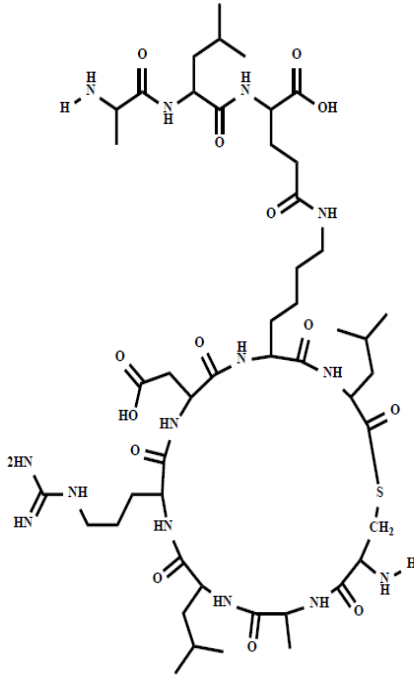
A patent application discloses the following structure:



The Cysteine and Leucine in the cyclic structure are linked through the side chain of the Cys and carboxy terminus of the Leu.

실시예 7(b)-4: 분지형(branched) 아미노산 서열을 가진 고리모양(cyclic) 펩티드 서열

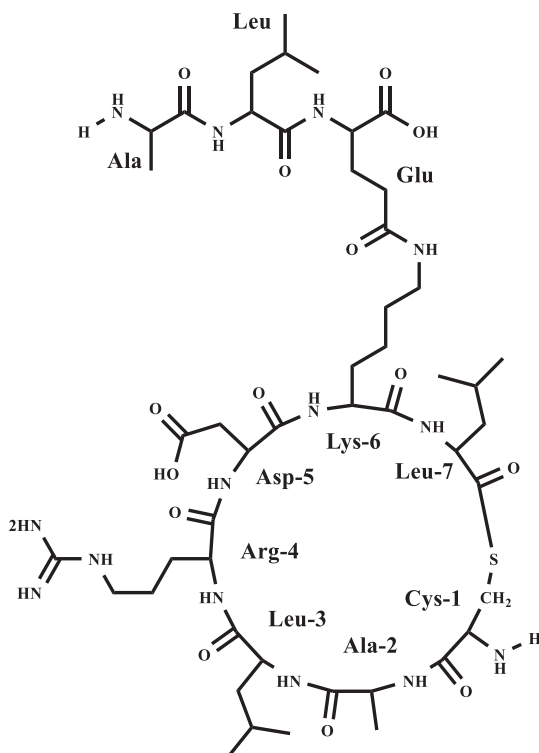
특허출원에 다음 구조가 개시되어 있다:



고리모양 구조에서 시스테인(Cysteine) 및 루이신(Leucine)은 Cys의 측쇄(side chain)와 Leu의 카르복시(carboxy) 말단을 통해 연결된다.

Question 1: Does ST.26 require inclusion of the sequence(s)?

The structure shown is a branched cyclic amino acid sequence which contains the following amino acids:



Since the side chain of the Cys and carboxy terminus of the Leu are involved in the cyclization, the N-terminus of the cyclic peptide is located at Cys-1.

YES – the cyclic region of the peptide

ST.26 paragraph 7(b) requires that the linear region of a branched sequence containing four or more specifically defined amino acids, wherein the amino acids form a single peptide backbone, must be included in a sequence listing. In the above example, the cyclic region of the branched peptide has more than four amino acids, and therefore, must be included in a sequence listing.

NO – the tripeptide branch of the peptide

The tripeptide branch Ala-Leu-Glu is not required to be in the sequence listing.

Question 2: Does ST.26 permit inclusion of the sequence(s)?

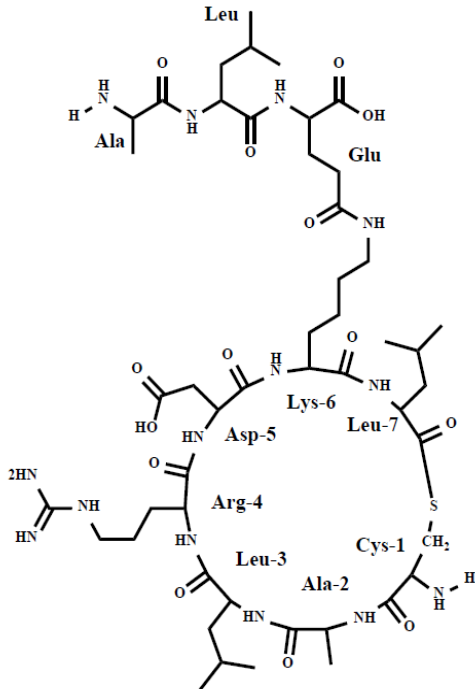
NO

According to paragraph 8, a sequence listing must not include any sequences having fewer than four specifically defined amino acids.

The tripeptide branch contains only three specifically defined amino acids and therefore, it must not be included in a sequence listing as a separate sequence with its own sequence identification number.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

도시된 구조는 다음의 아미노산을 가지는 분지형의 고리모양(cyclic) 아미노산 서열이다:



Cys의 측쇄(side chain) 및 Leu의 카르복시 말단이 고리화에 관여하기 때문에, 고리모양(cyclic) 펩티드의 N-말단은 Cys-1에 위치한다.

예 - 펩티드의 고리 부분

ST.26의 7(b)절에 따라, 아미노산이 단일 펩티드 골격을 형성하는 4개 이상의 구체적으로 정의된 아미노산을 함유하는 분지형 서열의 선형 영역은 서열목록에 포함되어야 한다. 위의 사례에서, 분지형 펩티드의 고리모양(cyclic) 부분은 4개 이상의 아미노산을 가지므로, 서열목록에 포함되어야 한다.

아니오- 펩티드의 트리펩티드 측쇄

트리펩티드 측쇄 Ala-Leu-Glu는 서열목록에 있을 필요는 없다.

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가?

아니오

8절에 따르면, 서열목록은 구체적으로 정의된 4개 미만의 아미노산을 갖는 서열을 포함해서는 아니된다.

트리 펩타이드 측쇄는 3개의 구체적으로 정의된 아미노산만을 함유하므로, 고유 서열식별번호를 갖는 별도의 서열로서 서열목록에 포함되어서는 아니된다.

Question 3: How should the sequence(s) be represented in the sequence listing?

While this example illustrates a peptide that is circular in configuration, the ring does not consist solely of amino acid residues in peptide linkages, as indicated in paragraph 25. Since the cyclization of the amino acid sequence occurs through the side chain of cysteine (Cys) and the carboxy terminus of the leucine (Leu), the cysteine must be assigned position number 1 within the cyclic region of the peptide. Accordingly, the sequence must be represented as:

CALRDKL (SEQ ID NO: 89)

As indicated in the figure above, the amino acid sequence is cyclized through a thioester conjugation between the cysteine side chain and the carboxy terminus of the leucine. The feature key "SITE" must be used to describe the modified cysteine, which forms the intrachain linkage with leucine. The feature location element is the residue numbers of the cross-linked amino acids in "x..y" format, i.e., "1..7". The mandatory qualifier "note" should indicate the nature of the linkage, e.g., "cysteine leucine thioester (Cys-Leu)", to specify that Cys-1 and Leu-7 are linked through a thioester bond. Further, the lysine in position number 6 must be annotated to indicate that it is modified, by using the feature key "SITE" together with the mandatory qualifier "note", where the qualifier value describes that the lysine side chain links the tripeptide ALE.

Relevant ST.26 paragraphs: 7(b), 8, 25, 26, 29, 30, 31, 66(c), and 70

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

본 사례는 구성이 원형인 펩티드를 예시하지만, 25절에 기재된 바와 같이 고리가 펩티드 연결인 아미노산 잔기만으로 구성되지는 않는다. 아미노산 서열의 고리화는 시스테인(Cysteine)의 측쇄(side chain)와 루이신(Leuine)의 카르복시 말단을 통해 발생하기 때문에, 시스테인(Cysteine)은 펩티드의 원형(cyclic) 부분 내 위치 1에 할당되어야 한다. 따라서 서열은 다음과 같이 표현되어야 한다:

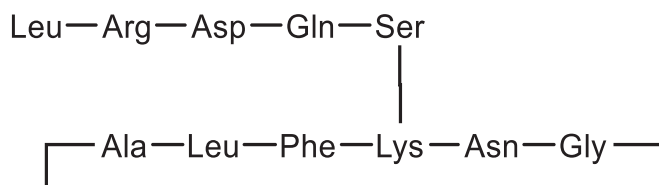
CALRDKL (SEQ ID NO: 89)

상기 도에 표시된 바와 같이, 아미노산 서열은 시스테인(Cysteine) 측쇄(side chain)와 루이신(Leuine)의 카르복시 말단 사이의 티오에스테(thioester) 접합을 통해 고리화된다. 특정 기호 "SITE"는 루이신(Leuine)과의 사슬내 연결을 형성하는 변형된 시스테인(Cysteine)을 설명하는 데에 사용하여야 한다. 특정 위치 요소는 "x...y" 포맷(예: "1...7" 포맷)에 교차 결합된 아미노산의 잔기 번호이다. 필수 한정자 "note"는, Cys-1과 Leu-7이 티오에스테르(thioester) 결합을 통해 연결된다는 것을 명시하기 위해서, 연결 성질(예를 들면, "Cysteine leucine thioester (Cys-Leu)")을 기재하여야 한다. 또한, 위치 번호 6의 라이신(lysine)은 필수 한정자 "note"와 함께 특정 기호 "SITE"를 사용하여 변형됨을 나타내기 위해 주석을 달아야 한다. 여기서 한정자 값은 라이신(lysine) 측쇄가 트리펩티드 ALE을 연결한다고 설명한다.

관련 ST.26 절: 7(b), 8, 25, 26, 29, 30, 31, 66(c) 및 70

Example 7(b)-5: Cyclic peptide containing a branched amino acid sequence

A patent application discloses the following branched cyclic peptide:

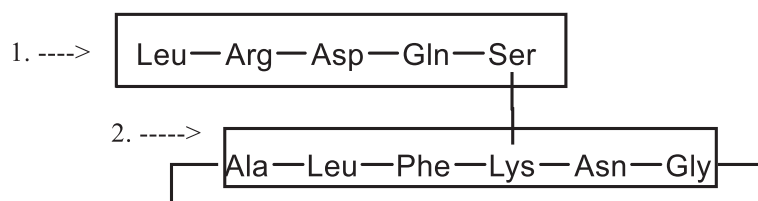


The Ser and the Lys are linked through an amide bond between the carboxy terminus of the serine and amine in the side chain of the Lys.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

Paragraph 7(b) requires inclusion of any sequence that contains four or more specifically defined amino acids and which can be represented as a linear region of a branched sequence in a sequence listing. In the above example, the peptide contains a cyclic region wherein the amino acids are joined by peptide bonds, and a branched region which is joined to a side chain of the Lys in the cyclic region. The regions of this branched peptide which can be represented as linear and which contain four or more specifically defined amino acids are:



ST.26 requires inclusion of sequences 1 and 2 of this cyclic branched peptide in a sequence listing, each with their own sequence identification number.

Question 3: How should the sequence(s) be represented in the sequence listing?

Sequence 1 must be represented as:

LRDQS (SEQ. ID. NO: 90)

Sequence 1 may be annotated by using the feature key "SITE" together with the qualifier "note" to describe that the serine in position 5 is linked to another sequence through an amide linkage between Ser and a side chain of a Lys in the other sequence.

Sequence 2 is a cyclic peptide. Paragraph 25 indicates that when an amino acid sequence is circular in configuration and has no amino and carboxy termini, applicant must choose the amino acid residue in position number 1. Accordingly, the sequence may be represented as:

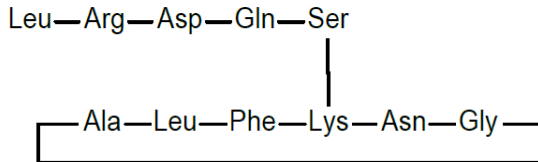
ALFKNG (SEQ. ID. NO: 91)

Alternatively, any other amino acid in the sequence could be designated as residue position number 1. The sequence ALFKNG must be further described using the feature key "SITE" together with the qualifier "note" to describe that the side chain of the Lys in residue position number 4 is linked via an amide linkage to another sequence. This side chain linkage modifies the Lys, and according to ST.26 paragraph 30, a modified amino acid must be further described in the feature table. Moreover, a feature key "REGION" and a qualifier "note" should be provided to indicate that the peptide ALFKNG is circular.

Relevant ST.26 paragraphs: 7(b), 25, 26, 30, and 31

실시예 7(b)-5: 분지형(branched) 아미노산 서열을 가진 고리모양(cyclic) 펩티드 서열

특허출원에 다음의 분지형의 원형(cyclic) 펩티드가 개시되어 있다:

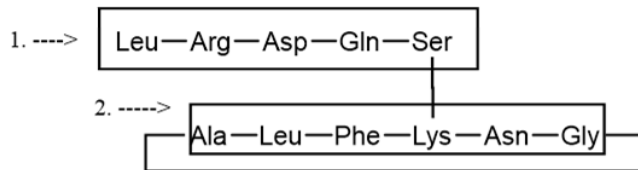


Ser 및 Lys는 세린(serine)의 카르복시 말단과 Lys의 측쇄(side chain)인 아민기 사이의 아마이드 결합을 통해 연결된다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

7(b)절에 따라 4개 이상의 구체적으로 정의된 아미노산을 함유하고 분지형 서열의 선형 영역으로 표현될 수 있는 모든 서열은 서열목록에 포함되어야 한다. 상기 사례에서의 펩티드는 아미노산이 펩티드 결합에 의해 연결된 고리형 영역 및 고리형 영역에서 Lys의 측쇄에 연결된 분지된 영역을 포함한다. 선형으로 표현될 수 있고 4개 이상의 구체적으로 정의된 아미노산을 함유하는 이 측쇄 펩티드의 영역은 다음과 같다:



ST.26에 따라 원형의 분지형 펩티드의 서열 1 및 2는 각각 자신의 서열식별번호와 함께 서열목록에 포함되어야 한다

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

서열 1은 다음과 같이 표현되어야 한다:

LRDQS (SEQ ID NO: 90)

서열 1은 위치 5의 세린(serine)이 Ser과 다른 Lys의 측쇄 사이의 아마이드 연결을 통해 다른 서열에 연결됨을 설명하는 한정자 "note"와 함께 특징 기호 "SITE"를 사용하여 주석을 달수 있다.

서열 2는 고리형 펩티드이다. 25절에 따르면, 아미노산 서열이 배열상 고리형이고 아미노 및 카르복시 말단이 없는 경우, 출원인은 위치 번호 1의 아미노산 잔기를 선택하여야 한다. 따라서, 서열은 다음과 같이 표현될 수 있다:

ALFKNG (SEQ ID NO: 91)

선택적으로, 서열에서 임의의 다른 아미노산은 잔기 위치 번호 1로 지정될 수 있다. 서열 ALFKNG는 특징 기호 "SITE"를 한정자 "note"와 함께 사용하여 위치 번호 4의 Lys잔기의 측쇄는 아마이드 연결을 통해 다른 서열에 연결되는 점을 설명하여야 한다. 이러한 측쇄 연결은 Lys를 변형시키므로, ST.26의 30절에 따라 변형 아미노산은 특징 표에 추가로 기술되어야 한다. 또한, 펩티드 ALFKNG이 고리형인 것을 표시하기 위해 특징 기호 "REGION"과 한정자 "note"가 제공되어야 한다.

관련 ST.26절: 7(b), 25, 26, 30 및 31

Paragraph 11(a) – Double-stranded nucleotide sequence – fully complementary

Example 11(a)-1: Double-stranded nucleotide sequence – same lengths

A patent application describes the following double-stranded DNA sequence:

3' – CCGGTAAACGCTA – 5'

5' – GGCCAATTGCGAT – 3'

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

Each enumerated nucleotide sequence has more than 10 specifically defined nucleotides. At least one strand must be included in the sequence listing, because the two strands of this double-stranded nucleotide sequence are fully complementary to each other.

Question 2: Does ST.26 permit inclusion of the sequence(s)?

YES

While the sequence of only one strand must be included in the sequence listing, the sequences of both strands may be included, each with its own sequence identification number.

Question 3: How should the sequence(s) be represented in the sequence listing?

The double-stranded DNA sequence must be represented either as a single sequence or as two separate sequences. Each sequence included in the sequence listing must be represented in the 5' to 3' direction and assigned its own sequence identification number.

atcgcaattgcc (top strand) (SEQ ID NO: 34)

and/or

ggccaattgcat (bottom strand) (SEQ ID NO: 35)

Relevant ST.26 paragraphs: 7(a), 11(a), and 13

11(a)절 - 이중가닥 핵산염기 서열 - 완전 상보적

실시에 11(a)-1: 이중가닥 핵산염기 서열 - 길이가 같을 경우

특허출원에 다음의 이중가닥 DNA 서열이 기술되어 있다:

3' -CCGGTTAACGCTA-5'

5' -GGCCAATTGCGAT-3'

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

나열된 각각의 핵산염기 서열은 10개 초과로 구체적으로 정의된 핵산염기를 갖는다. 이 이중가닥 핵산염기 서열의 2개의 가닥이 서로 완전히 상보적이기 때문에, 적어도 하나의 가닥이 서열목록에 포함되어야 한다.

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가?

예

하나의 가닥의 서열이 서열목록에 포함되어야 하지만, 각각의 가닥 식별 번호를 갖는 두 가닥의 서열이 포함될 수도 있다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

이중가닥 DNA 서열은 단일 서열 또는 두 개의 분리된 서열 중 하나로 표현되어야 한다. 서열목록에 포함된 각각의 서열은 5'-말단에서 3'-말단 방향으로 표현되어야 하며 고유한 서열식별번호를 부여하여야 한다.

atcgcaattggcc(상단 가닥) (SEQ ID NO: 34)

및/또는

ggccaattgcat(하단 가닥) (SEQ ID NO: 35)

관련 ST.26절: 7(a), 11(a) 및 13

11(b)절 - 이중가닥 핵산염기 서열 - 불완전 상보적

실시에 11(b)-1: 이중가닥 핵산염기 서열 - 길이가 다를 경우

특허출원에 다음의 도면과 설명이 포함되어 있다.

```

5' -tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc-3'
      |||||||||||||||
      gggtactgantccgc
  
```

PNA 프로브(probe)(하단 가닥)에 의해 결합된 인간 유전자 ABC1 프로모터(promoter) 영역(상단 가닥). 여기서 PNA 프로브(probe)의 "n"은 5-니트로인돌(5-nitroindole) 및 3-니트로인돌(3-nitroindole)로 구성된 그룹으로부터 선택된 범용 PNA염기이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예 - ABC1 프로모터(promoter) 영역(상단 가닥)

상단 가닥은 10개 이상의 나열된 „구체적으로 특정된“ 핵산염기를 가지므로 서열목록에 포함되어야 한다.

예 - PNA 프로브(probe)(하단 가닥)

두 가닥이 서로 완전히 상보적이지 않기 때문에 하단 가닥도 독립된 서열식별번호와 함께 서열목록에 포함되어야 한다. PNA 또는 "펩티드 핵산"을 포함하는 개별 잔기는 ST.26의 3(g)절에 따라 핵산염기로 간주된다. 따라서, 하부 가닥은 10개 이상의 나열된 "구체적으로 특정된" 핵산염기를 가지므로 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

상단 가닥은 다음과 같이 서열목록에 포함되어야 한다:

tagttcattgactaaggctccccattgactaaggcgactagcattgactaaggcaagc (SEQ ID NO: 36)

하부 가닥은 펩타이드 핵산이므로 3'-말단 및 5'-말단을 갖지 않는다. 11절에 따르면, "5'-말단에서 3'-말단 방향을 모방하도록 왼쪽에서 오른쪽으로 서열목록에 포함되어야 한다. 따라서 다음과 같이 서열목록에 포함되어야 한다:

cgcctnagtcaatggg (SEQ ID NO: 37)

특징 기호 "source"의 "organism"한정자는 "synthetic construct"값과 "other DNA"값을 갖는 필수 한정자 "mol_type"을 가져야 한다.

하단 가닥은 특징 기호 "modified_base"와 약어 "OTHER"가 있는 필수 한정자 "mod_base"를 사용하여 특징 표에 설명되어야 한다.

"note" 한정자는 "N-(2-aminoethyl) glycine nucleosides"와 같은 변형 핵산염기의 단축되지 않는 완전한 명칭과 함께 포함되어야 한다.

"n" 잔기는 특징 기호 "modified_base"와 필수 한정자 "mod_base"를 약어 "OTHER"로 사용하여 특징 표에 추가로 설명되어야 한다.

한다. "note" 한정자는 변형 핵산염기의 단축되지 않는 완전한 명칭 "N-(2-aminoethyl) glycine 5-nitroindole or N-(2-aminoethyl) glycine 3-nitroindole"과 함께 포함되어야 한다.

관련 ST.26 절: 3(g), 7(a), 11(b), 17 및 18

Example 11(b)-2: Double-stranded nucleotide sequence – no base-pairing segment

A patent application describes the following double-stranded DNA sequence:

```

3' -CCGGTTAGCTTATACGCTAGGGCTA- 5'
   |||||      |||||
5' -GGCCAATATGGCTTGCGATCCCGAT- 3'

```

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

Each strand of the enumerated, double-stranded nucleotide sequence has more than 10 specifically defined nucleotides. Both strands must be included in the sequence listing, each with its own sequence identification number, because the two strands are not fully complementary to each other.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence of each strand must be represented in the 5' to 3' direction and assigned its own sequence identification number:

atcgggatcgcattcgtggcc (top strand) (SEQ ID NO: 38)

and

ggccaatatggcttgatccgat (bottom strand) (SEQ ID NO: 39)

Relevant ST.26 paragraphs: 7(a), 11(b), and 13

실시예 11(b)-2: 이중가닥 핵산염기 서열 – 염기쌍(base-pairing)을 이루지 않는 단편 포함

특허출원에 다음의 이중 가닥 DNA 서열이 기술되어 있다:

```

3' -CCGGTTAGCTTATACGCTAGGGCTA-5'
      ||| ||| ||| ||| ||| ||| |||
5' -GGCCAATATGGCTTGCGATCCCGAT-3'

```

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

나열된 이중가닥 핵산염기 서열의 각 가닥은 10개 초과로 특정하게 정의된 핵산염기를 갖는다. 두 가닥이 서로 완전히 상보적이지 않기 때문에, 두 가닥 모두가 각각 고유한 서열식별번호를 갖는 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

각 가닥의 순서는 5'-말단에서 3'-말단 방향으로 표현되어야 하며 고유한 서열식별번호가 지정되어야 한다.

atcgggatcgcatattcgattggcc (상단 가닥) (SEQ ID NO: 38)

및

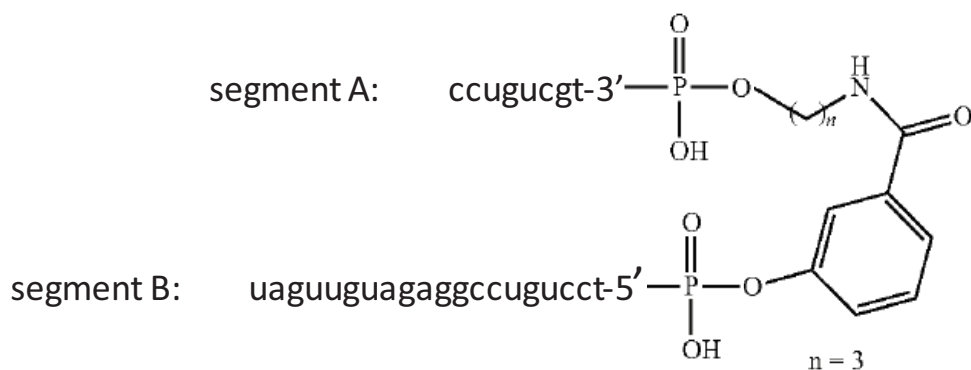
ggccaatatggcttgatcccgat (하단 가닥) (SEQ ID NO: 39)

관련 ST.26 절: 7(a), 11(b) 및 13

Paragraph 14 – Symbol “t” construed as uracil in RNA

Example 14-1: The symbol “t” represents uracil in RNA

A patent application describes the following compound:



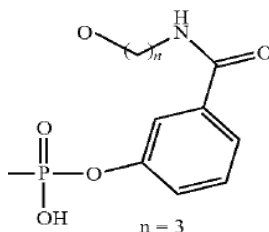
wherein segment A and segment B are RNA sequences.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES – segment B

NO – segment A

The enumerated sequence contains two segments of specifically defined nucleotides separated by the following “linker” structure:



The linker structure is not a nucleotide according to paragraph 3(g); therefore, each segment must be considered a separate sequence. Segment B contains more than 10 specifically defined nucleotides and ST.26 paragraph 7(a) requires inclusion in a sequence listing. Segment A contains only eight specifically defined nucleotides and therefore is not required to be included in a sequence listing.

Question 2: Does ST.26 permit inclusion of the sequence(s)?

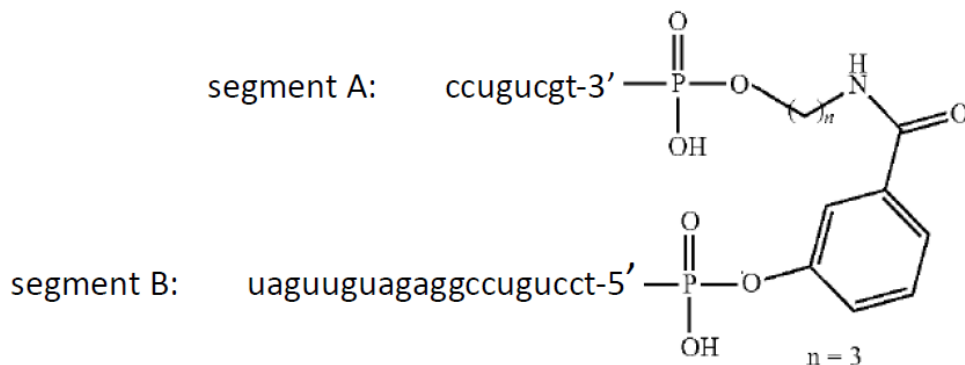
NO

Segment A contains fewer than 10 specifically defined nucleotides, and as per ST.26 paragraph 8, it must not be included in a sequence listing.

14절 – RNA에서 우라실(uracil)로 해석되는 기호 "t"

실시예 14-1: 기호 "t"는 RNA의 우라실(uracil)을 표현한다

특허출원에 다음 화합물이 기술되고 있다.



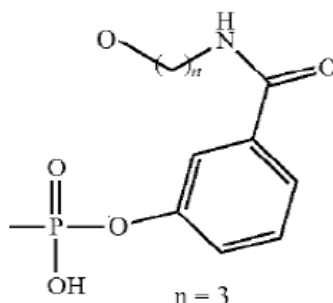
단편 A 및 단편 B는 RNA 서열이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예 _ 단편 B

아니오 – 단편 A

나열된 서열은 하기 "링커(linker)"구조에 의해 분리된 구체적으로 정의된 핵산염기의 2개의 단편을 함유한다:



링커(linker) 구조는 3(g)절에 따른 핵산염기가 아니다; 따라서 각 단편은 별도의 서열로 간주하여야 한다. 단편 B는 10개 초과로 특정하게 정의된 핵산염기를 함유하며 ST.26의 7(a)절에 따라 서열목록에 포함되어야 한다. 단편 A는 8개의 구체적으로 정의된 핵산염기만을 가지므로 서열목록에 포함되지 않는다.

질문 2: ST.26은 서열의 포함을 허용하는가?

아니오

단편 A는 10개 미만의 구체적으로 정의된 핵산염기를 함유하며, ST.26의 8절에 따라, 서열목록에 포함되어서는 아니된다.

Question 3: How should the sequence(s) be represented in the sequence listing?

Segment B is an RNA molecule; therefore, the element "INSDSeq_moltype" must be "RNA." The symbol "u" must not be used to represent uracil in an RNA molecule in a sequence listing. According to paragraph 14, the symbol "t" will be construed as uracil in RNA. Accordingly, segment B must be included in the sequence listing as:

tctgtccggagatgtgat (SEQ ID NO: 40)

Thymine in RNA is considered a modified nucleotide, i.e. modified uracil, and must be represented in the sequence as "t" and be further described in a feature table. Accordingly, the thymine in position 1 must be further described using the feature key "modified_base", the qualifier "mod_base" with "OTHER" as the qualifier value, and a qualifier "note" with "thymine" as the qualifier value.

The thymine, i.e. modified uracil, in position 1 should also be further described in a feature table using the feature key "misc_feature" and a qualifier "note" with the value e.g., "The 5' oxygen of the thymidine is attached through the linker (4-(3-hydroxybenzamido)butyl) phosphinic acid to another nucleotide sequence." Where practicable, the other sequence may be directly indicated as the value in the qualifier "note".

Relevant ST.26 paragraphs: 3(g), 7(a), 8, 13, 14, 19, and 54

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

단편 B는 RNA 분자이며; 따라서 "INSDSeq_moltype" 요소는 "RNA"이어야 한다. "u" 기호는 서열목록에서 RNA 분자의 우라실을 표현하는 데에 사용해서는 아니된다. 14절에 따르면, 기호 "t"는 RNA에서 우라실로 해석될 것이다. 따라서 단편 B는 다음과 같이 서열목록에 포함되어야 한다:

tcctgtccggagatggtgat (SEQ ID NO: 40)

RNA의 티민(Thymine) 은 변형 핵산염기(변형 우라실)로 간주된다. 그리고, "t"와 같이 서열에 표현되어야 하고 특징 표에 추가적인 설명이 되어야 한다. 따라서 위치 1의 티민은 특징 기호 "modified_base", 한정자 값으로 "OTHER"를 가진 한정자 "mod_base", 그리고 한정자 값으로 "thymine"을 가진 한정자 "note"를 사용하여 추가로 설명되어야 한다.

위치 1의 티민, 즉 변형 우라실은 특징 기호 "misc_feature"와 값("The 5' oxygen of the thymidine is attached through the linker (4-(3-hydroxybenzamido)butyl) phosphinic acid to another nucleotide sequence")을 사용하여 추가로 설명되어야 한다. 가능하다면, 다른 서열은 한정자 "note"의 값으로 직접 표시될 수 있다.

관련 ST.26 절: 3(g), 7(a), 8, 13, 14, 19 및 54

Paragraph 27 – The most restrictive ambiguity symbol should be used

Example 27-1: Shorthand formula for an amino acid

(GGGz)₂

Where z is any amino acid.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The sequence is disclosed as a formula. (GGGz)₂ is simply a shorthand way of representing the sequence GGGzGGGz. Conventionally, a sequence is expanded first, and the definition of any variable, i.e. “z”, is determined thereafter.

The sequence uses the nonconventional symbol “z”. The definition of “z” must be determined from the explanation of the sequence in the disclosure, which defines this symbol as any amino acid (see Introduction to this document). The example does not provide any constraint on “z”, e.g., that it is the same in each occurrence.

The peptide in the example has eight enumerated amino acids, six of which are specifically defined glycine residues, and the remaining two are the “z” variable that should be represented in this sequence using the conventional symbol “X”. ST.26 paragraph 7(b) requires inclusion of the sequence in a sequence listing as a single sequence with a single sequence identification number.

Note that the sequence is still encompassed by Paragraph 7(b) despite the fact that the enumerated and specifically defined residues are not contiguous.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence uses the nonconventional symbol “z”, which according to the disclosure is any amino acid. The conventional symbol used to represent “any amino acid” is “X”. Therefore, the sequence must be represented as the single expanded sequence:

GGGXGGGX (SEQ ID NO: 41)

According to paragraph 27, “X” will be construed as any one of “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, or “V”, except where it is used with a further description in the feature table. Since in this example “X” represents “any amino acid”, it must be annotated with the feature key VARIANT and a note qualifier with the value, “X can be any amino acid”.

Where practicable, each “X” should be annotated individually. However, a region of contiguous “X” residues, or a multitude of “X” residues dispersed throughout the sequence, may be jointly described with the feature key VARIANT using the syntax “x..y” as the location descriptor, where x and y are the positions of the first and last “X” residues, and a note qualifier with the value, “X can be any amino acid”.

Further, the example does not disclose that “z” is the same amino acid in both positions in the expanded sequence. However, if “z” is disclosed as the same amino acid in both positions, then a feature key “VARIANT” and a qualifier “note” should be provided stating that “X” in position 4 and 8 can be any amino acid, as long as they are the same in both positions.

Relevant ST.26 paragraph(s): 3(c), 7(b) and 27

27절 – 가장 제한적인 다중아미노산 기호를 사용하여야 한다

실시예 27-1: 아미노산 서열의 축약식

(GGGz)₂

여기에서, z는 아미노산이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

서열은 식으로 공개된다. (GGGz)₂는 단순히 서열 GGGzGGGz를 표현하는 간단한 방법이다. 일반적으로, 먼저 서열을 확장하고 그 후에 임의의 변수, 즉 "z"가 결정된다.

이 서열에는 비정상적인 기호 "z"가 있다. "z"의 정의는 이 기호를 임의의 아미노산으로 정의하는 본 개시 내용의 서열의 설명으로부터 결정되어야 한다(본 문서의 서론 참조). 이 실시예는 "z"에 대한 제한을 제공하지 않는다(예: 각 경우 동일 함).

본 실시예에서 펩타이드는 8개의 나열된 아미노산으로 되어있다. 그중에 6개는 구체적으로 정의된 glycine 잔기이며, 나머지 2개는 "z" 변수이며, 이는 본 서열에서 통상적인 심볼인 "X"로 표현되어야 한다. ST.26의 7(b)절은 단일 서열식별번호를 갖는 단일 서열로서 서열목록에 서열을 포함시킬 것을 요구한다.

나열된 구체적으로 정의된 잔기가 연속적이지 않다는 사실에도 불구하고, 서열은 여전히 7(b)절에 포함되는 점을 주목해야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

상기 서열은 본 발명에 따른 임의의 아미노산인 비정상적 기호 "z"를 사용한다. "임의의 아미노산"을 표현하는 데에 사용되는 일반적인 기호는 "X"이다. 따라서 서열은 단일 확장 서열로 표현되어야 한다.

GGGXGGGX (SEQ ID NO: 41)

27절에 따르면, "X"는 특징 표에서 추가 설명과 함께 사용되는 경우를 제외하고는, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다. 본 실시예에서 "X"는 "임의의 아미노산"을 나타내므로, VARIANT 특징 기호와 "X can be any amino acid"라는 값을 가진 note 한정자로 주석을 달아야 한다.

가능한 경우, 각 "X"는 독립적으로 주석을 달아야 한다. 그러나, 연속 "X" 잔기의 영역, 또는 서열 전체에 분산된 다수의 "X" 잔기는 위치 설명자로서 "x..y" 구문을 사용하는 특징 기호 'VARIANT'와 "X can be any amino acid"라는 값을 가진 'note' 한정자를 이용하여 함께 기술될 수 있다. 여기서 x와y는 첫 번째 및 마지막 "X" 잔기의 위치이다.

또한, 실시예는 "z"가 확장된 서열에서 두 위치에서 동일한 아미노산이라는 것을 개시하지 않았다. 그러나, "z"가 두 위치에서 동일한 아미노산으로 개시되는 경우, 위치 기호 4 및 8의 "X"가 양위치에서 동일한 임의의 아미노산일 수 있음을 나타내기 위해 특징 기호 "VARIANT" 및 한정자 "note"가 제공되어야 한다.

관련 ST.26절: 3(c), 7(b) 및 27

Example 27-2: Shorthand formula - less than four specifically defined amino acids

A peptide of the formula (Gly-Gly-Gly-z)_n

The disclosure further states, that z is any amino acid and

- (i) variable n is any length; or
- (ii) variable n is 2-100, preferably 3

Question 1: Does ST.26 require inclusion of the sequence(s)?

NO

Consideration of both disclosed embodiments (i) and (ii) of the enumerated peptide of the formula reveals that “n” can be “any length”; therefore, the most encompassing embodiment of “n” is indeterminate. Since “n” is indeterminate, the peptide of the formula cannot be expanded to a definite length, and therefore, the unexpanded formula must be considered.

The enumerated peptide in the unexpanded formula (“n” = 1) provides three specifically defined amino acids, each of which is Gly, and the symbol “z”. Conventionally “Z” is the symbol for “glutamine or glutamic acid”; however, the example defines “z” as “any amino acid” (see Introduction to this document). Under ST.26, an amino acid that is not specifically defined is represented by “X”. Based on this analysis, the enumerated peptide, i.e. GGGX, does not contain four specifically defined amino acids. Therefore, ST.26 paragraph 7(b) does not require inclusion, despite the fact that “n” is also defined as specific numerical values in some embodiments.

Question 2: Does ST.26 permit inclusion of the sequence(s)?

YES

The example provides a specific numerical value for variable “n,” i.e., a lower limit of 2, an upper limit of 100, and an exact value 3. Any sequence containing at least four specifically defined amino acids may be included in the sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

A sequence containing 100 copies of GGGX is preferred (SEQ ID NO: 42). A further annotation should indicate that up to 98 copies of GGGX could be deleted. Inclusion of further specific embodiments that are a key part of the invention is strongly encouraged.

According to paragraph 27, “X” will be construed as any one of “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, or “V”, except where it is used with a further description in the feature table. Since “X” in SEQ ID NO: 42 represents “any amino acid”, it must be annotated with the feature key VARIANT and a note qualifier with the value, “X can be any amino acid”.

Where practicable, each “X” should be annotated individually. However, a region of contiguous “X” residues, or a multitude of “X” residues dispersed throughout the sequence, may be jointly described with the feature key VARIANT using the syntax “x..y” as the location descriptor, where x and y are the positions of the first and last “X” residues, and a note qualifier with the value, “X can be any amino acid”.

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraph(s): 3(c), 7(b), 26, and 27

실시예 27-2: 축약식 - 네 개 미만의 구체적으로 정의된 아미노산

식(Gly-Gly-Gly-z)_n의 펩티드

본 개시 내용은 또한 z는 임의의 아미노산이고

- (i) 변수 n은 임의의 길이이며; 또는
- (ii) 변수 n은 2-100, 가급적 3

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

아니오

위의 식으로 나열된 펩티드의 개시된 실시 양태 (i) 및 (ii) 모두를 고려하면 "n"은 "임의의 길이"일 수 있으며; 따라서, "n"의 가장 포괄적인 구체 예는 미정이다. "n"은 불명확하기 때문에, 위의 식에서 펩티드는 한정된 길이로 확장될 수 없으므로, 확장되지 않은 축약식이 고려되어야 한다.

확장되지 않은 축약식("n"= 1)에서 나열된 펩티드는 각각 Gly인 3개의 구체적으로 정의된 아미노산 및 기호 "z"를 제공한다. 통상적으로 "Z"는 "글루타민 또는 글루탐산(glutamine or glutamic acid)"의 상징이지만, 이 예에서는 "z"를 "임의의 아미노산"으로 정의한다(이 문서의 서론 참조). ST.26에서 구체적으로 정의되지 않은 아미노산은 "X"로 표현된다. 이 분석에 기초하여, 나열된 펩티드, 즉 GGGX는 4개의 구체적으로 정의된 아미노산을 함유하지 않는다. 따라서, ST.26의 7(b)절은 일부 실시예에서 "n"이 특정 수치로 정의된다는 사실에도 불구하고 포함을 요구하지 않는다.

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가?

예

이 예는 변수 "n"에 대한 특정 수치, 즉 하한 2, 상한 100 및 정확한 값 3을 제공한다. 4개 이상의 구체적으로 정의된 아미노산을 함유하는 임의의 서열이 서열목록에 포함될 수 있다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

100개의 GGGX 사본을 함유하는 서열이 바람직하다(SEQ ID NO: 42). 추가 주석은 최대 98개의 GGGX 사본을 삭제할 수 있음을 나타낸다. 본 발명의 중요한 부분인 추가 특정 실시예를 포함할 것을 강력히 권장한다.

27절에 따르면, "X"는 특징 표에서 추가 설명과 함께 사용되는 경우를 제외하고는, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다. 본 실시예에서 "X"는 "임의의 아미노산"을 나타내므로, VARIANT 특징 기호와 "X can be any amino acid"라는 값을 가진 note 한정자로 주석을 달아야 한다.

가능한 경우, 각 "X"는 독립적으로 주석을 달아야 한다. 그러나, 연속 "X" 잔기의 영역, 또는 서열 전체에 분산된 다수의 "X" 잔기는 위치 설명자로서 "x..y" 구문을 사용하는 특징 기호 'VARIANT'와 "X can be any amino acid"라는 값을 가진 'note' 한정자를 이용하여 함께 기술될 수 있다. 여기서 x와 y는 첫 번째 및 마지막 "X" 잔기의 위치이다.

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 3(c), 7(b), 26 및 27

Example 27-3: Shorthand formula - four or more specifically defined amino acids

A peptide of the formula (Gly-Gly-Gly-z)_n

Where z is any amino acid and variable n is 2-100, preferably 3.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated peptide of the formula provides three specifically defined amino acids, each of which is Gly, and the symbol "z". Conventionally, "Z" is the symbol for "glutamine or glutamic acid"; however, the description in this example defines "z" as "any amino acid" (see Introduction to this document). Under ST.26, an amino acid that is not specifically defined is represented by "X". Based on this analysis, the enumerated repeat peptide does not contain four specifically defined amino acids. However, the description provides a specific numerical value for variable "n," i.e., a lower limit of 2 and an upper limit of 100. Therefore, the example discloses a peptide having at least six specifically defined amino acids in the sequence GGGzGGGz, which is required by ST.26 to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

Since "z" represents any amino acid, the conventional symbol used to represent the fourth and eighth amino acids is "X."

ST.26 requires inclusion in a sequence listing of only the single sequence that has been enumerated by its residues. Therefore, at least one sequence containing any of 2, 3, or 100 copies of GGGX must be included in the sequence listing; however, the most encompassing sequence containing 100 copies of GGGX is preferred (SEQ ID NO: 42) (see Introduction to this document). In the latter case, a further annotation could indicate that up to 98 copies of GGGX could be deleted. Inclusion of two additional sequences containing 2 and 3 copies of GGGX, respectively (SEQ ID NO: 44-45), is strongly encouraged.

According to paragraph 27, "X" will be construed as any one of "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", or "V", except where it is used with a further description in the feature table. Since "X" in this example represents "any amino acid", it must be annotated with the feature key VARIANT and a note qualifier with the value, "X can be any amino acid".

Where practicable, each "X" should be annotated individually. However, a region of contiguous "X" residues, or a multitude of "X" residues dispersed throughout the sequence, may be jointly described with the feature key VARIANT using the syntax "x..y" as the location descriptor, where x and y are the positions of the first and last "X" residues, and a note qualifier with the value, "X can be any amino acid".

Further, the example does not disclose that the "z" variable is the same in each of the two occurrences in the expanded sequence. However, if "z" is disclosed as the same amino acid in all locations, then a feature Key VARIANT and a Qualifier note should indicate that "X" in all positions can be any amino acid, as long as they are the same in all locations.

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraph(s): 3(c), 7(b), 26, and 27

실시예 27-3: 축약식 - 네 개 이상의 구체적으로 정의된 아미노산

식(Gly-Gly-Gly-z)_n의 펩티드

여기서, z는 임의의 아미노산이고 가변 n은 2-100, 가급적 3이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

위의 식에 나열된 펩티드 식은 각각 Gly인 3개의 구체적으로 정의된 아미노산과이고, 상징 "z"를 제공한다. 통상적으로, "Z"는 "글루타민 또는 글루탐산(glutamine or glutamic acid)"의 상징이나 본 예제의 설명은 "z"를 "임의의 아미노산"으로 정의한다(이 문서의 서론 참조). ST.26에서 구체적으로 정의되지 않은 아미노산은 "X"로 표현된다. 따라서 나열된 반복 펩티드는 4개의 구체적으로 정의된 아미노산을 함유하지 않는다. 그러나, 발명의 설명은 변수 "n"에 대한 특정 수치, 즉 하한 2 및 상한 100을 제공한다. 따라서, 본 실시예는 서열 GGGzGGGz에서 6개 이상의 구체적으로 정의된 아미노산을 갖는 펩티드를 개시하고, ST.26은 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

"z"는 임의의 아미노산을 표현하므로, 제 4 및 제 8 아미노산을 표현하는 데에 사용되는 통상적인 기호는 "X"이다.

ST.26은 그 잔기에 의해 나열된 단일 서열의 서열목록에 포함을 요구한다. 따라서 GGGX의 2, 3 또는 100개 사본 중 하나를 포함하는 하나 이상의 서열이 서열목록에 포함되어야 한다. 그러나 100개의 GGGX 사본을 함유하는 가장 포괄적인 서열이 바람직하다(SEQ ID NO: 42)(이 문서의 서론 참조). 후자의 경우 추가 주석은 최대 98개의 GGGX 사본을 삭제할 수 있음을 나타낼 수 있다. 각각 2개 및 3개의 GGGX 사본을 함유하는 2개의 추가 서열(SEQ ID NO: 44-45)을 포함하는 것이 바람직하다.

27절에 따르면, "X"는 특징 표에서 추가 설명과 함께 사용되는 경우를 제외하고는, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다. 본 실시예에서 "X"는 "임의의 아미노산"을 나타내므로, VARIANT 특징 기호와 "X can be any amino acid"라는 값을 가진 note 한정자로 주석을 달아야 한다.

가능한 경우, 각 "X"는 독립적으로 주석을 달아야 한다. 그러나, 연속 "X" 잔기의 영역, 또는 서열 전체에 분산된 다수의 "X" 잔기는 위치 설명자로서 "x..y" 구문을 사용하는 특징 기호 'VARIANT'와 "X can be any amino acid"라는 값을 가진 'note' 한정자를 이용하여 함께 기술될 수 있다. 여기서 x와 y는 첫 번째 및 마지막 "X" 잔기의 위치이다.

또한, 실시예는 확장된 서열에서 "z" 변수가 2개의 발생(occurrences) 각각에서 동일하다는 것을 개시하지 않는다. 그러나 만일 "z"가 모든 위치에서 동일한 아미노산으로 공개되는 경우, 특징 기호 'VARIANT'와 한정자 'note'로 모든 위치의 "X"가 모든 지점에서 동일한 임의의 아미노산일 수 있음을 나타내야 한다.

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 3(c), 7(b), 26 및 27

Paragraph 28 – Amino acid sequences separated by internal terminator symbols

Example 28-1: Encoding nucleotide sequence and encoded amino acid sequence

A patent application describes the following sequences:

caattcaggg tggatgaat atg gcg ccc aat acg caa acc gcc tct ccc cgc
Met Ala Pro Asn Thr Gln Thr Ala Ser Pro Arg

gcg ttg gcc gat tca tta atg cag ctg gca cga cag gtt tcc cga ctg
Ala Leu Ala Asp Ser Leu Met Gln Leu Ala Arg Gln Val Ser Arg Leu

Protein A

gaa agc ggg cag tga atg acc atg att acg gat tca ctg gcc gtc gtt
Glu Ser Gly Gln Met Thr Met Ile Thr Asp Ser Leu Ala Val Val

tta caa cgt cgt gac tgg gaa aac cct ggc gtt acc caa ctt aat cgc
Leu Gln Arg Arg Asp Trp Glu Asn Pro Gly Val Thr Gln Leu Asn Arg

Protein B

ctt gca gca cat tgg tgt caa aaa taa taataaccgg atgtactatt
Leu Ala Ala His Trp Cys Gln Lys

tatccctg atg ctg cgt cgt cag gtg aat gaa gtc gct taa gcaatcaatg
Met Leu Arg Arg Gln Val Asn Glu Val Ala

Protein C

tcggatgagg cgcgacgctt atccgaccaa catatcataa

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The application describes a nucleotide sequence, containing termination codons, which encodes three distinct amino acid sequences.

The enumerated nucleotide sequence contains more than 10 specifically defined nucleotides and must be included in a sequence listing as a single sequence.

Regarding the encoded amino acid sequences, paragraph 28 requires that amino acid sequences separated by an internal terminator symbol such as a blank space, must be included as separate sequences. Since each of "Protein A", "Protein B", and "Protein C" contain four or more specifically defined amino acids, ST.26 paragraph 7(b) requires that each must be included in a sequence listing and must be assigned its own sequence identification number.

28절 - 내부 종결 기호로 분리된 아미노산 서열

실시예 28-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열

특허출원에 다음과 같은 서열이 기술되고 있다:

caattcaggg tggtgaat atg gcg ccc aat acg caa acc gcc tct ccc cgc
Met Ala Pro Asn Thr Gln Thr Ala Ser Pro Arg

gcg ttg gca gat tca tta atg cag ctg gca cga cag gtt tcc cga ctg
Ala Leu Ala Asp Ser Leu Met Gln Leu Ala Arg Gln Val Ser Arg Leu

Protein A

gaa agc ggg cag tga atg acc atg att acg gat tca ctg gcc gtc gtt
Glu Ser Gly Gln Met Thr Met Ile Thr Asp Ser Leu Ala Val Val

tta caa cgt cgt gac tgg gaa aac cct ggc gtt acc caa ctt aat cgc
Leu Gln Arg Arg Asp Trp Glu Asn Pro Gly Val Thr Gln Leu Asn Arg

Protein B

ctt gca gca cat tgg tgt caa aaa taa taataaccgg atgtactatt
Leu Ala Ala His Trp Cys Gln Lys

tatccctg atg ctg cgt cgt cag gtg aat gaa gtc gct taa gcaatcaatg
Met Leu Arg Arg Gln Val Asn Glu Val Ala

Protein C

tcggatgcgcgacgctt atccgaccaa catatcataa

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

이 출원은 세 가지 아미노산 서열을 코딩하는 정지 코돈을 포함하는 핵산염기 서열을 설명한다.

나열된 핵산염기 서열은 10개 이상의 구체적으로 정의된 핵산염기를 가지므로 단일 서열로서 나열된 서열에 포함되어야 한다.

암호화된(encoded) 아미노산 서열과 관련하여, 28절은 공백과 같은 내부 종결 기호로 분리된 아미노산 서열을 별도의 서열로 포함해야 한다. "단백질 A", "단백질 B" 및 "단백질 C" 각각은 4개 이상의 구체적으로 정의된 아미노산을 함유하기 때문에, ST.26의 7(b)절에 따라 각각이 서열목록에 포함되어야 하고 자신의 서열식별번호가 부여되어야 한다.

Question 3: How should the sequence(s) be represented in the sequence listing?

The nucleotide sequence must be included in a sequence listing as:

```
caattcaaggtggtgaatatgacgccaatacgcgaaccgctctcccgcgctgacgattcattaatgaaaagcggcagtgaaatgaccatgattacggat  
tcactggccgtcatttacaacgtcgtgactgggaaaaccctgacgacccaactaatacgcctgcaqcaattggtgtcaaaaataataaaccggatgtacta  
ttatccctgatgctgctgctgaggtgaatgaagtcgcttaagcaatcaatgctggatgcgcgcgacgcttatccgaccaacatatcataa (SEQ ID NO:  
46)
```

The nucleotide sequence should further be described using a “CDS” feature key for each of the three proteins and the element INSDFeature_location must identify the location of each coding sequence, including the stop codon. In addition, for each “CDS” feature key, the “translation” qualifier should be included with the amino acid sequence of the protein as the qualifier value. The application does not disclose the genetic code table that applies to the translation (see Annex 1, Section 9, Table 7). If the Standard Code table applies, then the qualifier “transl_table” is not necessary; however, if a different genetic code table applies, then the appropriate qualifier value from Table 7 must be indicated for the qualifier “transl_table”. Finally, the qualifier “protein_id” must be included with the qualifier value indicating the sequence identification number of each of the translated amino acid sequences.

The amino acid sequences must be included as separate sequences, each assigned its own sequence identification number:

MAPNTQTASPRALADSLMQLARQVSRLESGQ (SEQ ID NO: 47)

MTMITDSLAVVLQRRDWENPGVTQLNRLAAHWCQK (SEQ ID NO: 48)

MLRRQVNEVA (SEQ ID NO: 49)

NOTE: See “Example 90-1 Amino acid sequence encoded by a coding sequence with introns” for an illustration of a translated amino acid sequence represented as a single sequence.

Relevant ST.26 paragraphs: 7, 26, 28, 57, 89-92

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

핵산염기 서열은 다음과 같은 서열목록에 포함되어야 한다:

```
caattcagggttgatggaatggcgccaatacgcaaaccgcctctccccgcgcgttgccgattcattaatggaagcgggcagtgatgaccatgattacggat
tcactggccgctgtttacaacgtcgtgactgggaaaaccctggcgttaccacaactaatcgccctgcagcacattggtgtcaaaaaataataaacggatgacta
ttatccctgatgctgcgtcagtgatgaatgaagtcgcttaagcaatcaatgctggatgcggcgcagccttatccgaccaacatata. (SEQ ID NO: 46)
```

핵산염기 서열은 3개의 단백질 각각에 대해 "CDS" 특징 기호를 사용하여 추가로 기술되어야 하며, INSDFeature_location 요소는 정지 코돈을 포함하여 각각의 코딩 서열의 위치를 식별하여야 한다. 또한 각 "CDS" 특징 기호에 대해 "translation" 한정자는 한정자 값으로 단백질의 아미노산 서열과 함께 포함되어야 한다. 본 출원은 번역에 적용되는 유전자 코드 표를 공개하지 않았다(부록 1, 섹션 9, 표 7 참조). 표준 코드 표가 적용되는 경우 한정자 "transl_table"은 필요하지 않다. 그러나 다른 유전자 코드 표가 적용되는 경우, 한정자 "transl_table" 값으로 표 7의 해당 한정자 값을 표시하여야 한다. 최종적으로, 한정자 "protein_id"는 각각의 번역된 아미노산 서열의 서열식별번호를 나타내는 한정자 값과 함께 포함되어야 한다.

아미노산 서열은 별도의 서열로 포함되어야 하며, 각각 고유의 서열식별번호가 부여된다:

MAPNTQTASPRALADSLMQLARQVSRLESGQ (SEQ ID NO: 47)

MTMITDSLAVLVQRDWDENPGVTQLNRLAAHWQCQK (SEQ ID NO: 48)

MLRRQVNEVA (SEQ ID NO: 49)

주의: 단일 서열로 표현되는 번역된 아미노산 서열에 대한 설명은 "예제 90-1 인트론(intron)을 갖는 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열" 참조.

관련 ST.26 절: 7, 26, 28, 57, 89-92

Paragraph 29 – Representation of an “other” amino acid

Example 29-1: Most restrictive ambiguity symbol for an “other” amino acid

A patent application describes the following sequence:

Ala-Hse-X₁-X₂-X₃-X₄-Tyr-Leu-Gly-Ser

Wherein, X₁= Ala or Gly,

X₂= Ala or Gly,

X₃= Ala or Gly,

X₄= Ala or Gly, and

Hse = Homoserine

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated peptide contains five specifically defined amino acids. The symbol “X” is used conventionally to represent two amino acids in the alternative (see Introduction to this document).

Because there are five specifically defined amino acids, i.e., Ala, Tyr, Leu, Gly and Ser, ST.26 paragraph 7(b) requires that the sequence must be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

Paragraph 29 requires any “other” amino acid must be represented by the symbol “X”. In the example, the sequence contains the amino acid Hse in position 2 which is not found in Annex I, Section 3, Table 3. Accordingly, Hse is an “other” amino acid and must be represented by the symbol “X”.

X₁-X₄ are variant positions, each of which can be A or G. The most restrictive ambiguity symbol for alternatives A or G is “X”. Therefore, the sequence may be represented as:

AXXXXXYLGS (SEQ ID NO: 50)

Inclusion of any specific sequences essential to the disclosure or claims of the invention is strongly encouraged, as discussed in the introduction to this document.

Since amino acid Hse is not found in Annex I, Section 4, Table 4, a feature key “SITE” and a qualifier “note” must be provided with the complete, unabbreviated name of homoserine as per ST.26 paragraph 30.

According to paragraph 27, because X₁-X₄ represent an alternative of only 2 amino acids, then further description is required. Paragraph 96 indicates that the feature key “VARIANT” should be used with the qualifier “note” and qualifier value “A or G”. According to ST.26 paragraph 34, since these positions are adjacent and have the same description, they may be jointly described using the syntax “3..6” as the location descriptor in the element INSDFeature_location.

Relevant ST.26 paragraphs: 3(a), 7(b), 25-27, 29, 30, 34, 66, 70, 71, and 96-97

29절 – “기타” 아미노산의 표현

실시예 29-1: “기타” 아미노산에 대한 가장 제한적인 다중아미노산 기호

어떤 특허출원에 다음 서열이 기술되어 있다:

Ala-Hse-X₁-X₂-X₃-X₄-Tyr-Leu-Gly-Ser

Wherein, X₁= Ala or Gly,

X₂= Ala or Gly,

X₃= Ala or Gly,

X₄= Ala or Gly, and

Hse = Homoserine

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

나열된 펩티드는 5개의 구체적으로 정의된 아미노산을 함유한다. “X” 기호는 선택으로 두 개의 아미노산을 표현하는 데에 일반적으로 사용된다(본 문서 서론 참조).

5개의 구체적으로 정의된 아미노산, 즉 Ala, Tyr, Leu, Gly 및 Ser이 있기 때문에, ST.26의 7(b)절에 따라 서열이 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

29절에서는 “기타” 아미노산이 “X” 기호로 표현되어야 한다. 본 실시예에서 서열은 2번 위치의 아미노산 Hse를 포함하며, 이는 별첨 1의 섹션 3, 표 3에서 찾아볼 수 없다. 따라서, Hse는 “기타” 아미노산이고 기호 “X”로 표시되어야 한다.

X₁-X₄는 각각 A 또는 G일 수 있는 변형 위치이다. 선택 A 또는 G에 대한 가장 제한적인 다중아미노산 기호는 “X”이다. 따라서 서열은 다음과 같이 표현될 수 있다:

AXXXXXYLGS (SEQ ID NO: 50)

본 문서의 서두에서 논의된 바와 같이, 본 발명의 개시 또는 청구 범위에 필수적인 임의의 특정 서열의 포함이 강력히 권장된다.

아미노산 Hse는 별첨 1, 표 4, 섹션 4에서 찾을 수 없으므로, 특징 기호 “SITE” 및 한정자 “note”에는 ST.26의 30절에 따라 완전하고 단축되지 않은 호모세린(homoserine) 명칭이 제공되어야 한다.

27절에 따라, X₁-X₄가 단지 2개의 아미노산의 대체를 표현하기 때문에 설명이 추가로 요구된다. 96절에 특징 기호 “VARIANT”를 한정자 “note” 및 한정자 값 “A or G”와 함께 사용해야 함이 기재되어 있다. ST.26의 34절에 따르면, 이들 위치는 인접하고 동일한 설명을 갖기 때문에, INSDFeature_location 요소에서 위치 설명자로서 “3..6” 구문을 사용하여 공동으로 설명될 수 있다.

관련 ST.26 절: 3(a), 7(b), 25-27, 29, 30, 34, 66, 70, 71 및 96-97

Example 29-2: Use of the corresponding unmodified amino acid

A patent application describes the following sequence:

Ala-Hyl-Tyr-Leu-Gly-Ser-Nle-Val-Ser-5ALA

Wherein Hyl = hydroxylysine (post-translational modification of lysine), Nle = Norleucine, and 5ALA = δ -Aminolevulinic acid

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated peptide contains more than four specifically defined amino acids; therefore, the sequence must be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The hydroxylysine in position 2, the norleucine in position 7, and the δ -aminolevulinic acid in position 10 are all "modified amino acids". First, we must consider each modified amino acid and determine if it should be represented by the corresponding unmodified amino acid or the variable "X" in the sequence. Paragraph 29 states that a modified amino acid "should be represented in the sequence as the corresponding unmodified amino acid whenever possible."

It is up to the discretion of the applicant to decide if a modified amino acid will be represented by the corresponding unmodified residue or the variables "X". However, the following guidance should be taken into account: If an amino acid is modified by the addition of a moiety, such as methylation or acetylation, and the basic structure of the corresponding unmodified amino acid is generally unchanged, then representation by the unmodified amino acid is recommended. If the modified amino acid is structurally very different from the corresponding unmodified amino acid, then representation by an "X" is recommended.

The structure of hydroxylysine is nearly identical to lysine, except that the third carbon in the R-group is modified with a hydroxyl group. Since the basic structure of the corresponding unmodified lysine residue is intact, hydroxylysine should be represented in the sequence by lysine ("K"), not by "X".

Norleucine is an isomer of leucine. The R-group of leucine is a 4 carbon chain, branched at the second carbon. Norleucine also has a 4 carbon R-group, but it is linear and not branched. Therefore, norleucine isn't simply the result of a modification added to a leucine, but a completely different (although related) structure. Therefore, it is recommended that norleucine be represented by an "X" in a sequence listing.

δ -Aminolevulinic acid is not structurally similar to any of the amino acids listed in Annex I, Table 3. Therefore, it is recommended that δ -aminolevulinic acid be represented by an "X" in a sequence listing.

Accordingly, the sequence should be included in a sequence listing as:

AKYLGXSVSX (SEQ ID NO: 51)

Paragraph 30 requires the further annotation of each modified amino acid.

Hydroxylysine is a post-translational modification of lysine. Therefore, it must be described using the feature key "MOD_RES" together with a qualifier "note" that describes the modification. Note that "hydroxylysine" is listed in Annex 1, Section 4, Table 4, "List of Modified Amino Acids." Therefore, the value of the qualifier "note" can contain the abbreviation "Hyl" instead of the complete, unabbreviated name "hydroxylysine."

Norleucine is not a post-translationally modified residue, therefore it must be described using the feature key "SITE" together with a qualifier "note" that describes the modification. Note that "norleucine" is also listed in Annex 1, Section 4, Table 4. Therefore, the value of the qualifier "note" can contain the abbreviation "Nle" instead of the complete, unabbreviated name "norleucine."

δ -Aminolevulinic acid is also not a post-translationally modified residue, therefore it must be described using the feature key "SITE" together with a qualifier "note" that describes the modification. δ -Aminolevulinic acid is not listed in Annex 1, Section 4, Table 4, therefore, the value of the qualifier "note" must contain the complete, unabbreviated name of the modified residue, " δ -aminolevulinic acid."

Relevant ST.26 paragraphs: 3(a), 3(e), 7(b), 29, and 30

실시예 29-2: “기타” 아미노산에 대한 가장 제한적인 다중아미노산 기호

어떤 특허출원에 다음 서열이 기술되어 있다:

Ala-Hyl-Tyr-Leu-Gly-Ser-Nle-Val-Ser-5ALA

여기서 Hyl = hydroxylysine (post-translational modification of lysine), Nle = Norleucine, and 5ALA = δ -Aminolevulinic acid

질문 1: 이 서열 ST.26은 이 서열(들)의 포함을 요구하는가?

예

나열된 펩티드는 4개 이상의 구체적으로 정의된 아미노산을 함유한다. 그러므로, 서열은 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

2번 위치의 hydroxylysine, 2번 위치의 hydroxylysine, 그리고 10번 위치에 δ -Aminolevulinic acid는 모두 “변형된 아미노산”이다. 우선, 각 변형된 아미노산을 고려하여 그것에 해당하는 변형되지 않은 아미노산에 의해서 표현될 수 있는지 또는 서열에서 “X”로 표현되어야 하는지를 결정하여야 한다. 29절은 “변형된 아미노산은 가능하면 상응하는 변형되지 않은 아미노산으로 표현되어야 한다”라고 규정하고 있다.

변형된 아미노산이 해당하는 변형되지 않은 아미노산에 의해서 표현될 수 있는지 또는 “X”로 표현되어야 하는지를 결정하는 것은 출처의 재량에 달려있다. 그러나 다음 가이드라인을 고려하여야 한다. 만약 아미노산 메틸레이션(methylation) 또는 아세틸레이션(acetylation) 같은 부분적인 기질이 추가되어서 변형이 되고 변형되지 않은 아미노산의 기본 구조가 일반적으로 변경되지 않은 경우 변형되지 않은 아미노산으로 표현하는 것을 추천한다. 만약 변형된 아미노산이 대응하는 변형되지 않은 아미노산과 비교하여 구조적으로 매우 다른 경우 “X”로 표현하는 것을 추천한다.

Hydroxylysine의 구조가 R-그룹의 3번째 탄소가 hydroxyl 그룹으로 변경된 것을 제외하면 lysine과 거의 동일하다. 해당하는 변형되지 않은 lysine 잔기의 기본 구조가 그대로 이므로 hydroxylysine은 서열에서 “X”가 아닌 lysine(“K”)로 표현이 되어야 한다.

Norleucine은 leucine의 이성질체(isomer)이다. Leucine의 R-그룹이 4개 탄소 체인이며 2번째 탄소에 가지를 가지고 있다. Norleucine 역시 4개 탄소가 된 R-그룹을 가지고 있지만, 직선형으로 가지를 가지고 있지 않다. 그러므로, norleucine은 단순히 leucine에 추가된 변형의 결과가 아니며 완전히 다른 구조를 가지고 있다. 따라서 norleucine은 서열목록에서 “X”로 표현하는 것을 권장한다.

δ -Aminolevulinic acid는 구조적으로 별첨1, 테이블3에 나열된 어떤 아미노산과도 유사하지 않다. 그러므로, δ -Aminolevulinic acid는 서열목록에서 “X”로 표현하는 것을 권장한다.

따라서 서열은 다음과 같이 서열목록에 포함되어야 한다.

AKYLGSXVSX (SEQ ID NO: 51)

20절은 각 변형된 아미노산의 추가적인 주석을 요구한다.

Hydroxylysine은 lysine의 전사후 변형(post-translational modification)의 결과이다. 그러므로, 변형을 설명하는 “note” 한정자와 함께 “MOD_RES” 특징 기호를 사용하여 설명되어야 한다. “hydroxylysine”은 별첨1, 섹션4, 테이블4의 “변형된 아미노산 목록”에 나열되어 있다는 것을 주목하여야 한다. 그러므로, “note” 한정자의 값으로 완전하고 축약되지 않은 이름인 “hydroxylysine” 대신 축약어인 “Hyl”이 포함되어야 한다.

Norleucine은 전사후 변형(post-translational modification)의 결과가 아니다. 그러므로, 변형을 설명하는 “note” 한정자와 함께 “SITE” 특징 기호를 사용하여 설명되어야 한다. “Norleucine”은 또한 별첨1, 섹션4, 테이블4에 나열되어 있다. 그러므로 “note” 한정자의 값으로 완전하고 축약되지 않은 이름인 norleucine” 대신 축약어인 “Nle”이 포함되어야 한다.

δ -Aminolevulinic acid 역시 전사후 변형(post-translational modification)의 결과가 아니다. 그러므로, 변형을 설명하는 “note” 한정자와 함께 “SITE” 특징 기호를 사용하여 설명되어야 한다. “ δ -Aminolevulinic acid”은 별첨1, 섹션4, 테이블4에 나열되어 있지 않다. 그러므로 “note” 한정자의 값으로 완전하고 축약되지 않은 변형 잔기의 이름인 “ δ -Aminolevulinic acid”이 포함되어야 한다.

관련 ST.26 절: 3(a), 3(e), 7(b), 29, 및 30

Paragraph 30 – Annotation of a modified amino acid

Example 30-1 – Feature key “CARBOHYD”

A patent application describes a polypeptide with a specifically modified amino acid, containing a glycosylated side chain, characterized in that Cys corresponding to positions 4 and 15 of the polypeptide forms a disulfide bond, according to the following sequence:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Asn(asialyloligosaccharide)-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated peptide provides 17 specifically defined amino acids. There are 16 natural amino acids, wherein the ninth (asparagine) is glycosylated. Therefore, the sequence must be included in a sequence listing as required by ST.26 paragraph (7)(b).

Question 3: How should the sequence(s) be represented in the sequence listing?

According to ST.26 paragraph 29, a modified amino acid should be represented in the sequence as the corresponding unmodified amino acid whenever possible.

Therefore the sequence must be included in a sequence listing as:

LEYCLKRWNETISHCAW (SEQ ID NO: 52)

A further description of the modified amino acid is required. The feature key “CARBOHYD” together with the (mandatory) qualifier “note” should be used to indicate the occurrence of the attachment of a sugar chain (asialyloligosaccharide) to asparagine in position 9. The qualifier “note” describes the type of linkage, e.g., N-linked. The location descriptor in the feature location element is the residue position number of the modified asparagine.

In addition, there is a disulfide bond between the two Cys residues. Therefore the feature key “DISULFID” should be used to describe an intrachain crosslink. The feature location element is the residue position numbers of the linked Cys residues in “x..y” format, i.e., “4..15”. The mandatory qualifier note should describe the intrachain disulfide bond.

Relevant ST.26 paragraph(s): 3(a), 7(b), 26, 29, **30**, 66(c), 70, and Annex I, section 7, feature key 7.4

30절 _ 변형 아미노산의 주석

실시예 30-1 – 특징 기호 “CARBOHYD”

특허출원이 아래 서열에 따라 폴리펩티드의 위치 4 및 15에 대응하는 Cys가 이황화(disulfide) 결합을 형성하는 것을 특징으로 하는, 글리코실화된(glycosylated) 측쇄를 함유하는 특정 변형 아미노산을 갖는 폴리펩티드를 기술한다:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Asn(asialyloligosaccharide)-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

나열된 펩티드는 17개의 특정하게 정의된 아미노산을 제공한다. 9 번째(아스파라긴)가 글리코실화된 16개의 천연 아미노산이 있다. 그러므로, 서열은 ST.26의 7(b)절에 따라 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

ST.26의 29절에 따르면, 변형 아미노산은 가능할 때마다 이에 대응하는 미변형 아미노산으로 서열에 표현되어야 한다.

따라서 서열은 다음과 같은 서열목록에 포함되어야 한다:

LEYCLKRWNETISHCAW (SEQ ID NO: 52)

변형 아미노산에 대한 추가 설명이 필요하다. (필수) 한정자 "note"와 함께 특징 기호 "CARBOHYD"를 사용하여 9번 위치의 아스파라긴에 당 사슬(asialyloligosaccharide)이 부착된 것을 나타내어야 한다. 한정자 "note"는 연결 타입을 나타낸다(예: N 연결). 특징 위치 요소의 위치 설명자는 변형 아스파라긴의 잔기 위치 번호이다.

또한, 2개의 Cys 잔기 사이에 이황화(disulfide) 결합이 존재한다. 따라서 특징 기호 "DISULFID"는 체인 내 가교를 설명하는 데에 사용되어야 한다. 특징 위치 요소(feature location element)는 "x.y 포맷(예, "4..15")으로 연결된 Cys 잔기의 잔기 위치 번호를 나타낸다. 필수 한정자 note에는 체인내(intrachain)의 이황화(disulfide) 결합에 대해서 설명이 있어야 한다.

관련 ST.26 절: 3(a), 7(b), 26, 29, 30, 66(c), 70 및 별첨 I(7), 특징 기호 7.4

Paragraph 36 – Sequences containing regions of an exact number of contiguous “n” or “X” residues

Example 36-1: Sequence with a region of a known number of “X” residues represented as a single sequence

LL-100-KYMR

Where the “-100-” between amino acids Leucine and Lysine reflects a 100 amino acid region in the sequence.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

ST.26 paragraph 36 requires inclusion of a sequence that contains at least four specifically defined amino acids separated by one or more regions of a defined number of “X” residues.

The disclosed sequence uses a nonconventional symbol, i.e. “-100-”. The definition of “-100-” must be determined from the explanation of the sequence in the disclosure, which defines this symbol as 100 amino acids between leucine and lysine (see Introduction to this document). Therefore, “-100-” is a defined region of “X” residues.

Since six of the 106 amino acids in the sequence are specifically defined, ST.26 paragraph 7(b) requires that the sequence must be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The nonconventional symbol “-100-” is represented as 100 “X” residues (since any symbol used to represent an amino acid is equivalent to only one residue). Therefore, a single sequence of 106 amino acids in length, containing 100 “X” residues between LL and KYMR, must be included in a sequence listing (SEQ ID NO: 53).

This sequence contains 100 “X” variables between LL and KYMR. The ST.26 default value for “X” with no further annotation, is any one of “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “P”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, or “V” (paragraph 27). If these 100 “X” variables are defined as anything other than this default value, then a proper annotation for each “X” variable must be provided.

Relevant ST.26 paragraph(s): 7(b), 26, 27, and 36

36절 – 정확한 수의 연속된 "n" 또는 "X" 잔기 영역을 포함하는 서열

실시예 36-1: 단일 서열로 표시된 알려진 개수의 "X" 잔기 영역을 가진 서열

LL-100-KYMR

여기서 아미노산 루이신(Leucine)과 라이신(Lysine) 사이의 "-100-"는 서열에서 100개의 아미노산 영역을 반영한다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

ST.26의 36절에서는 정의된 수의 "X" 잔기의 하나 이상의 영역에 의해 분리된 적어도 네개의 구체적으로 정의된 아미노산을 포함하는 서열이 포함될 것을 요구한다.

개시된 서열은 비통상적인 상징, 즉 "-100-"를 사용한다. "-100-"의 정의는 루이신(Leucine)과 라이신(Lysine) 사이의 100개의 아미노산으로 정의하는 본 개시 내용의 서열의 설명으로부터 결정되어야 한다(본 문서의 서론 참조). 따라서, "-100-"은 "X" 잔기로 특정된 영역이다.

서열에서 106개의 아미노산 중 6개가 구체적으로 정의되어 있기 때문에, ST.26 7(b)절은 서열이 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

비통상적 기호 "-100-"는 100개의 "X" 잔기(아미노산을 표현하는 데에 사용되는 기호가 하나의 잔기에 해당하기 때문에)를 표현한다. 따라서, LL과 KYMR 사이에 100개의 "X" 잔기를 함유하는 106개 아미노산 길이의 단일 서열이 서열목록(SEQ ID NO: 53)에 포함되어야 한다.

본 서열은 LL과 KYMR사이에 100개의 "X"를 가지고 있다. ST.26에서는 더 이상의 설명이 없는 "X"는 "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다(27절). 만약 이들 100개 "X"가 기본값 이외 다른 어떤 것으로 정의된다면, 각 "X"에 대해서 적절한 주석이 있어야 한다.

관련 ST.26 절: 7(b), 26, 27 및 36

Example 36-2: Sequence with multiple regions of a known number or range of "X" residues represented as a single sequence

Lys-z₂-Lys-z_m-Lys-z₃-Lys-z_n-Lys-z₂-Lys

Where z is any amino acid, m=20, n=19-20, z₂ means that the pairs of Lysines are separated by any two amino acids, and z₃ means the pairs of Lysines are separated by any three amino acids.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The disclosed sequence uses a nonconventional symbol, i.e. "z." Therefore, the disclosure must be consulted to determine the definition; "z" is defined as any amino acid (see Introduction to this document). The conventional symbol used to represent any amino acid is "X". Considering the presence of "X" variables, the peptide contains six lysine residues that are enumerated and specifically defined, which is required to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The sequence uses a nonconventional symbol "z", the definition of which must be determined from the disclosure. Since "z" is defined as any amino acid, the conventional symbol is "X."

The preferred and most encompassing means of representation is (see Introduction to this document):

KXXKXX (SEQ ID NO: 54)

Wherein z_n is equal to 20 "X's", with a further description that the "X" variable corresponding to position 30 can be deleted.

Alternatively, or in addition to the above, the sequence may be represented as:

KXXKXX (SEQ ID NO: 55)

Wherein z_n is equal to 19 "X's", with a further description that an "X" variable between position numbers 29 and 30 can be inserted.

According to paragraph 27, "X" will be construed as any one of "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y", or "V", except where it is used with a further description in the feature table. Since "X" in SEQ ID Nos 54 and 55 represents "any amino acid", it must be annotated with the feature key VARIANT and a note qualifier with the value, "X can be any amino acid".

Where practicable, each "X" should be annotated individually. However, a region of contiguous "X" residues, or a multitude of "X" residues dispersed throughout the sequence, may be jointly described with the feature key VARIANT using the syntax "x..y" as the location descriptor, where x and y are the positions of the first and last "X" residues, and a note qualifier with the value, "X can be any amino acid".

Relevant ST.26 paragraph(s): 26, 27, and 36

실시예 36-2: 단일 서열로 표시된 알려진 개수 또는 범위의 "X" 잔기의 다중 영역을 갖는 서열

Lys-z₂-Lys-z_m-Lys-z₃-Lys-z_n-Lys-z₂-Lys

z는 임의의 아미노산이고, m = 20, N = 19 ~ 20, z₂는 라이신(Lysine) 쌍이 임의의 두 아미노산에 의해 분리되어 있음을 의미하고, z₃는 라이신(Lysine) 쌍이 임의의 세 개의 아미노산에 의해 분리되는 것을 의미한다.

질문 1: ST.26에 서열을 포함시켜야 하는가?

예

개시된 서열은 비통상적인 기호, 즉 "z"를 사용한다. 따라서, 정의를 결정하기 위해 개시 내용을 참조하여야 한다; "z"는 임의의 아미노산으로 정의된다(서론 참조). 임의의 아미노산을 표현하는데 사용되는 통상적인 기호는 "X"이다. "X" 변수의 존재를 고려하여, 펩티드는 나열되고 구체적으로 정의된 6개의 라이신(Lysine) 잔기를 함유하며, 이는 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 서열을 어떻게 표현해야 하는가?

이 서열은 비통상적인 기호 "z"를 사용하며, 그 정의는 개시 내용으로부터 결정되어야 한다. "z"는 임의의 아미노산으로 정의되므로, 통상적인 기호는 "X"이다.

선호하는 가장 포괄적인 표현 방법은 다음과 같다(이 문서의 서론 참조).

KXXK (SEQ ID NO: 54)

z_n은 20개의 "X"와 동일하고, 위치 30에 대응하는 "X" 변수는 삭제될 수 있다는 추가 설명이 있다.

선택적으로, 또는 상기에 추가하여, 서열은 다음과 같이 표현될 수 있다:

KXXK (SEQ ID NO: 55)

여기서, z_n은 19개의 "X"와 동일하고, 위치 번호 29와 30 사이의 "X" 변수가 삽입될 수 있다는 추가 설명이 있다.

27절에 따르면, "X"는 특징 표에서 추가 설명과 함께 사용되는 경우를 제외하고는, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다. 서열 번호 54와 55의 "X"는 "임의의 아미노산"을 표현하므로, 특징 기호 VARIANT와 "X can be any amino acid"라는 값을 가진 note 한정자로 주석을 달아야 한다.

가능한 경우, 각 "X"는 독립적으로 주석을 달아야 한다. 그러나, 연속 "X" 잔기의 영역, 또는 서열 전체에 분산된 다수의 "X" 잔기는 위치 설명자로서 "x.y" 구문을 사용하는 특징 기호 'VARIANT'와 "X can be any amino acid"라는 값을 가진 'note' 한정자를 이용하여 함께 기술될 수 있다. 여기서 x와 y는 첫 번째 및 마지막 "X" 잔기의 위치이다.

관련 ST.26 절: 26, 27 및 36

Paragraph 37 – Sequences containing regions of an unknown number of contiguous “n” or “X” residues

Example 37-1: Sequence with regions of an unknown number of “X” residues must not be represented as a single sequence

Gly-Gly----Gly-Gly-Xaa-Xaa

w here the symbol ---- is an undefined gap w ithin the sequence, w here Xaa is any amino acid, and the Glycine and Xaa residues are connected to one another through peptide bonds.

Question 1: Does ST.26 require inclusion of the sequence(s)?

NO

ST.26 paragraph 37 prohibits the inclusion of any sequence that contains an undefined gap; therefore, inclusion of the entire sequence is not required.

ST.26 paragraph 37 does require inclusion of any region of a sequence adjacent to an undefined gap that contains four or more specifically defined amino acids. In the example above, inclusion of either region adjacent to the undefined gap is not required, since each region contains only tw o specifically defined amino acids.

Question 2: Does ST.26 permit inclusion of the sequence(s)?

NO – not the entire sequence

NO – not any region of the sequence

ST.26 paragraph 37 does not permit inclusion of the entire sequence.

ST.26 paragraph 8 does not permit inclusion of either region adjacent to the undefined gap, since each region contains only tw o specifically defined amino acids.

Relevant ST.26 paragraphs: 7(b), 8, 26, and 37

37절 – 알려지지 않은 수의 연속된 "n" 또는 "X" 잔기 영역을 포함하는 서열

실시예 37-1: 알려지지 않은 개수의 "X" 잔기 영역을 가진 서열은 단일 서열로 표현되어서는 안된다

Gly-Gly----Gly-Gly-Xaa-Xaa

여기서 기호 ----는 서열내 정의되지 않은 갭(Gap)이며, 여기서 Xaa는 임의의 아미노산이며, 글리신 및 Xaa 잔기는 펩티드 결합을 통해 서로 연결된다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

아니오

ST.26의 37절은 정의되지 않은 구간(gap)을 포함하는 서열의 포함을 금지한다. 그러므로, 전체 서열의 포함은 필요하지 않다.

ST.26의 37절은 정의되지 않은 갭(Gap)에 인접한 4개 이상의 구체적으로 정의된 아미노산을 포함하는 서열의 임의의 영역의 포함을 요구한다. 상기 예에서, 각각의 영역은 단지 2개의 구체적으로 정의된 아미노산을 포함하기 때문에, 정의되지 않은 갭(Gap)에 인접한 어느 하나의 영역의 포함은 필요하지 않다.

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가?

아니오 – 전체 서열은 아님

아니오 – 서열의 어떤 영역도 아님

ST.26의 37절은 전체 서열의 포함을 허용하지 않는다.

각 영역은 두 개의 구체적으로 정의된 아미노산만을 포함하기 때문에, ST.26의 8절은 정의되지 않은 갭(Gap)에 인접한 어느 한 영역의 포함도 허용하지 않는다.

관련 ST.26 절: 7(b), 8, 26 및 37

Example 37-2: Sequence with regions of an unknown number of “X” residues must not be represented as a single sequence

Gly-Gly----Gly-Gly-Ala-Gly-Xaa-Xaa

w herein the symbol ---- is an undefined gap w ithin the sequence, w here Xaa is any amino acid, and the Glycine and Xaa residues are connected to one another through peptide bonds.

Question 1: Does ST.26 require inclusion of the sequence(s)?

NO – not the entire sequence

YES – a region of the sequence

ST.26 paragraph 37 prohibits the inclusion of any sequence that contains an undefined gap, but requires inclusion of any region of a sequence adjacent to an undefined gap that contains four or more specifically defined amino acids.

In the example above, ST.26 does not require (and prohibits) inclusion of both the entire sequence, w hich contains an undefined gap, and the Gly-Gly region adjacent to the undefined gap, w hich contains only two specifically defined amino acids. How ever, ST.26 requires inclusion of the Gly-Gly-Ala-Gly- Xaa-Xaa region adjacent to the undefined gap, since it contains at least four specifically defined amino acids.

Question 2: Does ST.26 permit inclusion of the sequence(s)?

NO – not the entire sequence and not the Gly-Gly region

Question 3: How should the sequence(s) be represented in the sequence listing?

The region of the sequence adjacent to the undefined gap that contains four specifically defined amino acids must be represented as:

GGAGXX (SEQ ID NO: 58)

The sequence should be annotated to indicate that the represented sequence is part of a larger sequence that contains an undefined gap by using the feature key “SITE”, the feature location “1” and the qualifier “note” w ith the value, e.g., “This residue is linked N-terminally to a peptide having an N-terminal Gly-Gly and a gap of undefined length”.

According to paragraph 27, “X” w ill be construed as any one of “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, or “V”, except where it is used w ith a further description in the feature table. Since “X” in SEQ ID NO: 58 represents “any amino acid”, it must be annotated w ith the feature key VARIANT and a note qualifier w ith the value, “X can be any amino acid”.

Where practicable, each “X” should be annotated individually. How ever, a region of contiguous “X” residues, or a multitude of “X” residues dispersed throughout the sequence, may be jointly described w ith the feature key VARIANT using the syntax “x..y” as the location descriptor, where x and y are the positions of the first and last “X” residues, and a note qualifier w ith the value, “X can be any amino acid”.

Relevant ST.26 paragraph(s): 7(b), 8, 26, 27, and 37

실시예 37-2: 알려지지 않은 수의 "X" 잔기 영역을 가진 서열은 단일 서열로 표현되어서는 안된다

Gly-Gly----Gly-Gly-Ala-Gly-Xaa-Xaa

여기서 ---- 기호는 서열 내 정의되지 않은 갭(Gap)이며, 여기서 Xaa는 임의의 아미노산이고, 글리신(Glycine) 및 Xaa 잔기는 펩티드 결합을 통해 서로 연결된다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

아니오 – 전체 서열이 아님

예 – 서열의 일부

ST.26의 37절에 따라 정의되지 않은 갭(gap)을 포함하는 임의의 서열의 포함을 금지하지만, 정의되지 않은 갭(gap)에 인접한 4개 이상의 구체적으로 정의된 아미노산을 포함하는 서열의 임의의 부분의 포함을 요구한다.

상기 예에서, ST.26에 따라 정의되지 않은 갭(gap)을 포함하는 전체 서열과 정의되지 않은 갭(gap)에 인접한 두개의 특정한 아미노산만을 포함하는 Gly-Gly 부분은 포함이 요구되지 않는다(또한 금지되어 있다). 그러나, 정의되지 않은 갭(gap)에 인접한 Gly-Gly-Ala-Gly-Xaa-Xaa 부분은 적어도 4개의 특정하게 정의된 아미노산을 포함하기 때문에 ST.26에 따라 포함이 요구된다.

질문 2: ST.26은 이 서열(들)의 포함을 허용하는가?

아니오 – 전체 서열도 아니며 Gly-Gly 부분도 아님

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

4개의 구체적으로 정의된 아미노산을 함유하는 정의되지 않은 갭(gap)에 인접한 서열의 부분은 다음과 같이 표현되어야 한다:

GGAGXX (SEQ ID NO: 58)

서열에는 특징 기호 "SITE", 특징 위치 "1" 및 한정자 "note"를 사용하여 정의되지 않은 갭(gap)을 포함하는 길이가 긴 서열의 일부임을 표현하도록 주석을 달아야 한다. 한정자 "note"의 예는 다음과 같다. " This residue is linked N-terminally to a peptide having an N-terminal Gly-Gly and a gap of undefined length."

27절에 따르면, "X"는 특징 표에서 추가 설명과 함께 사용되는 경우를 제외하고는, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다. 서열 번호 58의 "X"는 "임의의 아미노산"을 표현하므로, 특징 기호 VARIANT와 "X can be any amino acid"라는 값을 가진 note 한정자로 주석을 달아야 한다.

가능한 경우, 각 "X"는 독립적으로 주석을 달아야 한다. 그러나, 연속 "X" 잔기의 영역, 또는 서열 전체에 분산된 다수의 "X" 잔기는 위치 설명자로서 "x.y" 구문을 사용하는 특징 기호 'VARIANT'와 "X can be any amino acid"라는 값을 가진 'note' 한정자를 이용하여 함께 기술될 수 있다. 여기서 x와 y는 첫 번째 및 마지막 "X" 잔기의 위치이다.

관련 ST.26 절: 7(b), 8, 26, 27 및 37

Paragraph 55 – A nucleotide sequence that contains both DNA and RNA segments

Example 55-1: Combined DNA/RNA Molecule

A patent application describes the following oligonucleotide sequence:

AGACCTTcggagucuccuguuacagauagucuaaaguagauC

Wherein the upper-case letters represent DNA residues and lower-case letters represent RNA residues.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The disclosed sequence has more than ten enumerated and specifically defined nucleotides; therefore, it is required to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The nucleotide sequence must be included in a sequence listing as:

agaccttcggagtctcctgtgaacagatagtc aaagtagatc (SEQ ID NO: 93)

Note that the uracil nucleotides must be represented by the symbol "t" in the sequence listing.

ST.26 paragraph 55 dictates that a nucleotide sequence containing both DNA and RNA segments must be indicated as molecule type "DNA" and must be further described using the feature key "source" and the mandatory qualifier "organism" with the value "synthetic construct" and the mandatory qualifier "mol_type" with the value "other DNA". In addition, each segment of the sequence must be further described with the feature key "misc_feature," which includes the location of the segment, and the qualifier "note," which indicates whether the segment is DNA or RNA. The disclosed sequence contains two DNA segments (nucleotide positions 1-7 and 43) and one RNA segment (nucleotide positions 8-42).

Relevant ST.26 paragraphs: 7, 14, 55-56, and 83

55절 – DNA와 RNA 단편을 모두 포함하는 핵산염기 서열

실시예 55-1: 결합된 DNA/RNA 분자

특허출원에 다음과 같은 올리고핵산염기(oligonucleotide) 서열이 기술되어 있다:

AGACCTTcggagucuccuguugaacagauagucaagaugauC

여기서, 대문자는 DNA 잔기를 표현하고 소문자는 RNA 잔기를 표현한다.

질문 1: ST.26에 서열을 포함시켜야 하는가?

예

개시된 서열은 10개 이상의 나열된 구체적으로 정의된 핵산염기를 가지므로 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 서열을 어떻게 표현해야 하는가?

핵산염기 서열은 다음과 같은 서열목록에 포함되어야 한다:

agaccttcggagtcctcctgtgaacagatagtagcaagtagatc (SEQ ID NO: 93)

우라실(uracil) 핵산염기는 서열목록에서 "t"기호로 표현되어야 한다.

ST.26의 55절에 따르면, DNA 및 RNA 단편을 모두 포함하는 핵산염기 서열은 분자 타입 "DNA"로 제시되어야 하고, 특징 기호 "source", "synthetic construct"를 값으로 가진 필수 한정자 "organism", "other DNA" 값으로 가진 필수 한정자 "mol_type"을 사용하여 설명되어야 한다. 또한, 서열의 각 단편은 단편의 위치를 포함하는 특징 기호 "misc_feature"와 단편이 DNA인지 RNA인지를 나타내는 한정자 "note"로 추가로 기술되어야 한다. 개시된 서열은 2개의 DNA 단편(핵산염기 위치 1-7 및 43) 및 하나의 RNA 단편(핵산염기 위치 8-42)을 가진다.

관련 ST.26 절: 7, 14, 55-56 및 83

*Paragraph 89 – “CDS” Feature key***Example 89-1: Encoding nucleotide sequence and encoded amino acid sequence**

A patent application describes the following nucleotide sequence and its translation:

```
atg acc gga aat aaa cct gaa acc gat gtt tac gaa att tta tga
```

```
Met Thr Gly Asn Lys Pro Glu Thr Asp Val Tyr Glu Ile Leu STOP
```

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated nucleotide sequence has more than ten specifically defined nucleotides.

The enumerated amino acid sequence has more than four specifically defined amino acids.

Question 3: How should the sequence(s) be represented in the sequence listing?

The nucleotide sequence must be presented as:

```
atgaccggaataaacctgaaaccgatgttacgaaatttatga (SEQ ID NO: 59)
```

The nucleotide sequence should further be described using the “CDS” feature key and the element `INSDFeature_location` must identify the entire sequence, including the stop codon (i.e., position 1 through 45). In addition, the “translation” qualifier should be included with the qualifier value “MTGNKPETDVYEIL”. The application does not disclose the genetic code table that applies to the translation (see Annex 1, Section 9, Table 7). If the Standard Code table applies, then the qualifier “transl_table” is not necessary; however, if a different genetic code table applies, then the appropriate qualifier value from Table 7 must be indicated for the qualifier “transl_table”. Finally, the qualifier “protein_id” must be included with the qualifier value indicating the sequence identification number of the translated amino acid sequence.

The amino acid sequence must be separately presented with its own sequence identification number using single letter codes as follows:

```
MTGNKPETDVYEIL (SEQ ID NO: 60)
```

The STOP following the enumerated amino acid sequence must not be included in the amino acid sequence in the sequence listing.

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraphs: 7(a), 7(b), 26, 28, 89, 90, and 92

89절 –“CDS”특징 기호

실시예 89-1: 암호화(encoding) 핵산염기 서열 및 암호화된(encoded) 아미노산 서열

특허출원에 다음과 같은 핵산염기 서열과 해당 핵산염기 서열의 번역이 기술되어 있다:

```
atg acc gga aat aaa cct gaa acc gat gtt tac gaa att tta tga  
Met Thr Gly Asn Lys Pro Glu Thr Asp Val Tyr Glu Ile Leu STOP
```

질문 1: ST.26에 서열을 포함시켜야 하는가?

예

나열된 핵산염기 서열은 10개 이상의 구체적으로 정의된 핵산염기를 갖는다.

나열된 아미노산 서열은 4개 이상의 구체적으로 정의된 아미노산을 갖는다.

질문 3: 서열목록에서 서열을 어떻게 표현해야 하는가?

핵산염기 서열은 다음과 같이 제시되어야 한다:

```
atgaccggaataaacctgaaaccgatgttacgaaatttatga (SEQ ID NO: 59)
```

핵산염기 서열은 "CDS" 특징 기호를 사용하여 추가로 기술되어야 하고, INSDFeature_location 요소는 정지 코돈을 포함하여 전체 서열(즉, 위치 1 내지 45)을 식별하여야 한다. 또한, "translation" 한정자는 한정자 값 "MTGNKPETDVYEIL"을 포함되어야 한다. 본 출원은 번역에 적용되는 유전자 코드 표를 공개하지 않았다(별첨 1, 섹션 9, 표 7 참조). 표준 코드 표가 적용되는 경우, 한정자 "transl_table"은 필요하지 않는다. 그러나 다른 유전자 코드 표가 적용되는 경우, 한정자 "transl_table"에 표 7의 해당 한정자 값을 제시하여야 한다. 마지막으로, 한정자 "protein_id"의 한정자 값으로 번역되는 아미노산 서열의 서열식별번호를 나타내어야 한다.

아미노산 서열은 다음과 같이 단일 문자 코드를 사용하여 독립된 서열식별번호를 별도로 제시하여야 한다.

```
MTGNKPETDVYEIL (SEQ ID NO: 60)
```

나열된 아미노산 서열 다음의 STOP은 서열목록에서 아미노산 서열에 포함되어서는 아니된다.

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 7(a), 7(b), 26, 28, 89, 90 및 92

Example 89-2: Feature location extends beyond the disclosed sequence

A patent application contains the following figure disclosing a partial coding sequence and its translated amino acid sequence:

```

cat cac gca gca gaa tgt gga ttt tgt cct caa caa tgg caa gtt cta      48
His His Ala Ala Glu Cys Gly Phe Cys Pro Gln Gln Trp Gln Val Leu
1           5           10           15

cgt ggg agt ctg tgc att tgt gag ggt cca gct gaa gga tgg ttc ata      96
Arg Gly Ser Leu Cys Ile Cys Glu Gly Pro Ala Glu Gly Trp Phe Ile
          20           25           30

tca aga tgt tgg tta tgg tgt ggg cct caa gtc caa ggc ttt atc ttt      144
Ser Arg Cys Trp Leu Trp Cys Gly Pro Gln Val Gln Gly Phe Ile Phe
          35           40           45

gga gaa ggc aag gaa gga ggc ggt gac aga cgg gct gaa gcg agc cct      192
Gly Glu Gly Lys Glu Gly Gly Gly Asp Arg Arg Ala Glu Ala Ser Pro
          50           55           60

cag gag ttt tgg gaa tgc act tgg      216
Gln Glu Phe Trp Glu Cys Thr Trp
65           70

```

Figure 1 – partial coding sequence of the *Homo sapiens ITCH1* gene, which encodes amino acids 20 through 91 of the 442 amino acid long ITCH1 protein.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The application discloses a nucleotide sequence and its translated amino acid sequence. The enumerated nucleotide sequence contains more than 10 specifically defined nucleotides and must be included in a sequence listing.

The amino acid sequence contains more than 4 specifically defined amino acids and also must be included in a sequence listing as a separate sequence with its own sequence identification number.

Question 3: How should the sequence(s) be represented in the sequence listing?

The nucleotide sequence must be included in a sequence listing as:

```
catcacgcagcagaatgtggattttgtcctcaacaatggcaagttctacgtgggagtcctgcatcttgagggtccagcgaaggatggtcatatcaagatggtggt  
tatggtgtgggcctcaagtcgaaggctttatcttggagaaggaaggaaggaggcggtgacagacgggctgaagcagccctcaggagtttgggaatgcact  
tgg (SEQ ID NO: 94)
```

The nucleotide sequence should further be described using a "CDS" feature key. The element INSDFeature_location must identify the location of the "CDS" feature in the sequence and must include the stop codon.

The figure describes a partial coding sequence that does not include the start codon or the stop codon. However, the description of the sequence indicates that the start codon is upstream of the nucleotide in position 1 and the stop codon is downstream of the last nucleotide in position 216.

ST.26 dictates that the location descriptor must not include numbering for residues beyond the range of the sequence in the INSDSeq_sequence element. Consequently, in the above example, the location descriptor for the CDS feature key cannot include position numbers outside the range of 1 through 216. The location of the stop codon in the element INSDFeature_location must be represented using the symbol ">" to indicate that the stop codon is located downstream of position 216. Likewise, the symbol "<" can be used to indicate that the location of the start codon is upstream of position 1. Thus, the location descriptor for the CDS feature key should appear as follows:

<1..>216

Note that "<" and ">" are reserved characters and will be replaced by "<" and ">," respectively, in the XML instance of the sequence listing.

The "translation" qualifier should be included with the amino acid sequence of the protein as the qualifier value. The figure does not disclose the genetic code table that applies to the translation (see Annex 1, Section 9, Table 7). If the Standard Code table applies, then the qualifier "transl_table" is not necessary; however, if a different genetic code table applies, then the appropriate qualifier value from Table 7 of ST.26 Annex I must be indicated for the qualifier "transl_table". Finally, the qualifier "protein_id" must be included in the CDS feature with the qualifier value indicating the sequence identification number of the translated amino acid sequence.

The translated amino acid sequence must be included as a separate sequence with its own sequence identification number:

```
HHA AECGF C PQWQVLRGSLCICEGPAEGWFISRCWLWCGPQVQGFIFGEGKEGGDRRAEAS PQEFWECTW  
(SEQ ID NO: 95)
```

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraphs: 7, 41, 65, 66, 70, 71, 89, and 92

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

핵산염기 서열은 다음과 같은 서열목록에 포함되어야 한다:

```
catcacgcagcagaatgtggattttgcctcaacaatggcaagtctacgtgggagctgtgcattgtgagggtccagctgaaggatggtcatatcaagatgttgg  
tatggtgtggcctcaagccaaggcttatcttggagaaggcaaggaaggagcggtgacagacgggctgaagcagagccctcaggagtttgggaatgcacttg  
(SEQ ID NO: 94)
```

핵산염기 서열은 "CDS" 특징 기호를 사용하여 추가로 기술되어야 한다. "CDS" 특징의 위치를 INSDFeature_location 요소에서 지정하여야 하며 정지 코돈을 포함하여야 한다.

도면은 개시 코돈 또는 정지 코돈을 포함하지 않는 부분 코딩 서열을 설명한다. 그러나, 서열의 설명이 개시 코돈이 위치 1에서 염기의 상류(upstream) 및 정지 코돈은 위치(216)의 마지막 염기의 하류(downstream)임을 나타낸다.

ST.26은 위치 설명자가 INSDSeq_sequence 요소에서 서열의 범위를 벗어난 잔기에 대한 넘버링을 포함해서는 안된다고 규정하고 있다. 결과적으로 위의 실시예에서 CDS 특징 기호의 위치 설명자는 1-216 범위를 벗어난 위치 번호를 포함할 수 없다. INSDFeature_location 요소에서 위치 216의 하류(downstream)에 위치한 정지 코돈의 위치는 ">"기호를 사용하여 표현되어야 한다. 마찬가지로, "<"기호는 개시 코돈의 위치가 위치 1의 상류(upstream)임을 나타내는 데에 사용될 수 있다. 따라서, CDS 특징 기호의 위치 설명자는 다음과 같이 나타나야 한다:

<1..>216

"<" 및 ">"는 예약된 문자이며 서열목록의 XML 인스턴스에서 각각 "<" 및 ">"로 바뀐다.

"translation" 한정자는 한정자 값으로서 단백질의 아미노산 서열과 함께 포함되어야 한다. 이 그림에는 번역에 적용되는 유전자 코드 표가 나와 있지 않다(부록 1, 섹션 9, 표 7 참조). 표준 코드 표가 적용되는 경우 한정자 "transl_table"은 필요하지 않다. 그러나, 다른 유전자 코드 표가 적용되는 경우, 한정자 "transl_table"에 대해 ST.26 별첨 I의 표 7의 적절한 한정자 값을 표시하여야 한다. 마지막으로, 한정자 "protein_id"는 번역된 아미노산 서열의 서열식별번호를 나타내는 한정자 값과 함께 CDS 특징에 포함되어야 한다.

번역된 아미노산 서열은 독립된 서열식별번호를 갖는 별도의 서열로서 포함되어야 한다:

```
HHAACGFCPQQWQVLRGSLCICEGPAEGWFISRCWLWCGPQVQGFIFGEGKEGGDRRAEASPQEFWECTW (SEQ ID NO: 95)
```

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 7, 41, 65, 66, 70, 71, 89 및 92

Paragraph 92 – Amino acid sequence encoded by a coding sequence

Example 92-1: Amino acid sequence encoded by a coding sequence with introns

A patent application contains the following figure disclosing a coding sequence and its translation:

```

atg aag act ttc gca gcc ttg ctt tcc gct gtc act ctc gcg ctc tcg
Met Lys Thr Phe Ala Ala Leu Leu Ser Ala Val Thr Leu Ala Leu Ser

gtg cgc gcc cag gcg gct gtc tgg agt caa t gtaagtgccg ctgcttttca
Val Arg Ala Gln Ala Ala Val Trp Ser Gln

ttgatacgag actctacgcc gagctgacgt gctaccgtat ag gt ggc ggt aca
Cys Gly Gly Thr

ccg ggt tgg acg gcc gag acc act tgc gtt gct ggt tcg gtt tat acc
Pro Gly Trp Thr Gly Glu Thr Thr Cys Val Ala Gly Ser Val Cys Thr

tcc ttg agc tca gtgagcgact ttcaatccgt cgtcattgct cctcatgtat
Ser Leu Ser Ser

tgacgattgg ccttcatag tca tac tct caa tgc gtt ccg gcc tcc gca acg
Ser Tyr Ser Gln Cys Val Pro Gly Ser Ala Thr

tcc agc gct ccg gcg gcc ccc tca gcg aca act tca gcc ccc gca cct
Ser Ser Ala Pro Ala Ala Pro Ser Ala Thr Thr Ser Gly Pro Ala Pro

acg gac gga acg tgc tcg gcc agc ggg gca tgg ccg cca ttg acc tga
Thr Asp Gly Thr Cys Ser Ala Ser Gly Ala Trp Pro Pro Leu Thr Ter

```

Figure 1 – nucleotides shown in bold-face are intron regions.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The application discloses a nucleotide sequence and its amino acid translation. The enumerated nucleotide sequence contains more than 10 specifically defined nucleotides and must be included in a sequence listing as a single sequence.

The nucleotide sequence contains coding sequence (exons) separated by noncoding sequence (introns). The figure depicts the translation of the nucleotide sequence as three non-contiguous amino acid sequences. According to the figure caption, the bolded regions of nucleotides are intron sequences that will be spliced out of an RNA transcript before translation into a protein. Accordingly, the three amino acid sequences are actually a single, contiguous, enumerated sequence, which contains more than four specifically defined amino acids and must be included in a sequence listing as a single sequence.

92절 - 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열

실시예 92-1: 인트론(intron)을 가진 코딩(coding) 서열에 의해 암호화된(encoded) 아미노산 서열

특허출원에 코딩 서열 및 그 번역을 개시하는 다음 도면이 포함된다:

```

atg aag act ttc gca gcc ttg ctt tcc gct gtc act ctc gcg ctc tcg
Met Lys Thr Phe Ala Ala Leu Leu Ser Ala Val Thr Leu Ala Leu Ser

gtg cgc gcc cag gcg gct gtc tgg agt caa t gtaagtgccg ctgctttca
Val Arg Ala Gln Ala Ala Val Trp Ser Gln

ttgatacgag actctacgcc gagctgacgt gctaccgtat ag gt ggc ggt aca
Cys Gly Gly Thr

ccg ggt tgg acg gcc gag acc act tgc gtt gct ggt tcg gtt tgt acc
Pro Gly Trp Thr Gly Glu Thr Thr Cys Val Ala Gly Ser Val Cys Thr

tcc ttg agc tca gtgagcgact ttcaatccgt cgtcattgct cctcatgtat
Ser Leu Ser Ser

tgacgattgg ccttcatag tca tac tct caa tgc gtt ccg gcc tcc gca acg
Ser Tyr Ser Gln Cys Val Pro Gly Ser Ala Thr

tcc agc gct ccg gcg gcc ccc tca gcg aca act tca gcc ccc gca cct
Ser Ser Ala Pro Ala Ala Pro Ser Ala Thr Thr Ser Gly Pro Ala Pro

acg gac gga acg tgc tcg gcc agc ggg gca tgg ccg cca ttg acc tga
Thr Asp Gly Thr Cys Ser Ala Ser Gly Ala Trp Pro Pro Leu Thr Ter

```

그림 1 - 굵은 체로 표시된 핵산염기는 인트론 영역이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

본 출원은 핵산염기 서열 및 그의 아미노산 번역을 개시한다. 나열된 핵산염기 서열은 10개 이상의 구체적으로 정의된 핵산염기를 함유하고 단일 서열로서 나열된 서열에 포함되어야 한다.

핵산염기 서열은 비코딩 서열(인트론)에 의해 분리된 코딩 서열(엑손)을 함유한다. 도면은 3개의 비연속 아미노산 서열로서 핵산염기 서열의 번역을 도시한다. 도면 주석에 따르면, 핵산염기의 굵은 글씨체 영역은 단백질로 번역하기 전에 RNA 전사체로부터 스플라이싱될 인트론 서열이다. 따라서, 3개의 아미노산 서열은 실제로 단일의 연속적으로 나열된 서열이며, 이는 4개 이상의 구체적으로 정의된 아미노산을 함유하고 단일 서열로서 나열된 서열에 포함되어야 한다.

Question 3: How should the sequence(s) be represented in the sequence listing?

The nucleotide sequence must be included in a sequence listing as:

```
atgaagactttcgcagccttgctttccgctgtcactctcgcgctctcgggtgcgcgccaggcggctgtctggagtcaatgtaagtccgctgcttttcattgatac gaga  
ctctacgcgagctgacgtgtaccgtatagggtggcgggtacaccgggttgacgggcgagaccacttgcgtgtcgtggttgcctccttgagctcagtgag  
cgactttcaatccgctgctcctcctatgtattgacgattggcctcatagtcatactcctaatgcgttccgggtccgcacagctccagcgtccggcggccccctc  
agcgacaacttcaggccccgcacctacggacggaacgtgctcggccagcggggcatggccgacctgacctga (SEQ ID NO: 75)
```

The nucleotide sequence should further be described using a "CDS" feature key and the element `INSDFeature_location` must identify the location of the coding sequence, including the stop codon indicated by "Ter". The CDS `INSDFeature_location` must use the "join" location operator to indicate that the translation products encoded by the indicated locations are joined and form a single, contiguous polypeptide using the format "join(x1..y1,x2..y2,x3..y3)", e.g., "join(1..79,142..212,272..400)". In addition, the "translation" qualifier should be included, with the amino acid sequence of the protein as the qualifier value. (Note that the terminator symbol "Ter" in the last position of the sequence must not be included in the amino acid sequence.) The application does not disclose the genetic code table that applies to the translation (see Annex 1, Section 9, Table 7). If the "Standard Code" table applies, then the qualifier "transl_table" is not necessary; however, if a different genetic code table applies, then the appropriate qualifier value from Table 7 must be indicated for the qualifier "transl_table". Finally, the qualifier "protein_id" must be included with the qualifier value indicating the sequence identification number of the translated amino acid sequence. The amino acid sequence must be included as a single sequence:

```
MKTFAALLSAVTLALSVRAQAAVWSQCGGTPGWTGETTCVAGSVCTSLSSYSQCVPGSATSSAPAAPSATTSG  
PAPTDGTCSASGAWPPLT (SEQ ID NO: 76)
```

Relevant ST.26 paragraphs: 7, 26, 28, 57, 67, and 89-92

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

핵산염기 서열은 다음과 같은 서열목록에 포함되어야 한다:

```
atgaagacttctgcagccttgcttccgctgtcactctcgcgctcgggctcgcgcccaggcggctgtctggagtcactgtaagtgccgctgctttcattgatacggaga
ctctacgcccagctgacgtgtaccgtataggtggcgggtacaccgggtggacggcgagaccactgctgtgctggttcggttgatcctcctgagctcagtgag
cgacttcaatccgctgctcattgctcctcatgtattgacgattggcctcatagtcatactcactcaatgcttccgggctccgcaacgtccagcgtccggcgccccctc
agcgacaactcaggccccgcactacggacggaacgtgctcggccagcggggcatggccgattgacctga (SEQ ID NO: 75)
```

핵산 서열은 "CDS" 특징 기호를 사용하여 추가로 설명되어야 하며, 요소 INSDFeature_location은 "Ter"로 표시된 정지 코돈을 포함하여 코딩 서열의 위치를 나타내야 한다. CDS INSDFeature_location에서 지정된 위치에 의해 번역된 단백질들이 결합되었음을 제시하려면 "join" 위치 연산자를 사용해야 하며, "join(x1..y1,x2..y2,x3..y3)", 즉, "join(1..79,142..212,272..400)" 과 같은 형식을 사용하여 단일 연속 폴리펩타이드를 형성하여야 한다. 또한, 단백질의 아미노산 서열을 한정자 값으로 하는 "translation" 한정자를 포함하여야 한다(서열의 마지막 위치에있는 종결 기호 "Ter"는 아미노산 서열에 포함되어서는 아니된다). 본 출원은 번역에 적용되는 유전자 코드 표를 공개하지 않았다(별첨 1, 섹션 9, 표 7 참조). 표준 코드 표가 적용되는 경우, 한정자 "transl_table"은 필요하지 않는다. 그러나 다른 유전자 코드 표가 적용되는 경우, 한정자 "transl_table"에 표 7의 해당 한정자 값을 제시하여야 한다. 마지막으로, 한정자 "protein_id"의 한정자 값으로 번역되는 아미노산 서열의 서열식별번호를 나타내어야 한다.

아미노산 서열은 단일 서열로 포함되어야 한다:

```
MKTFALLLSAVTLALSVRAQAAVWSQCGGTPGWTGETTCVAGSVCTSLSSSYSQCVPGSATSSAPAAPSATTSG
PAPTDGTCASAGAWPPLT (SEQ ID NO: 76)
```

관련 ST.26 절: 7, 26, 28, 57, 67 및 89-92

Paragraph 93 – Primary sequence and a variant, each enumerated by its residues

Example 93-1: Representation of enumerated variants

The description includes the following sequence alignment.

<i>D. melanogaster</i>	ACATTGAATCTCATACCACTTT
<i>D. virilis</i>	. . . - . . G . . C . . - - . G
<i>D. simulans</i>	GT . . G . CG . . GT . . SGT . G . . .

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

It is common in the art to include “dots” in a sequence alignment to indicate “this position is the same as the position above it.” Therefore, the “dots” in *D. virilis* and *D. simulans* sequences are considered enumerated and specifically defined nucleotides, as they are simply a short-hand way of indicating that a given position is the same nucleotide as in *D. melanogaster*. In addition, sequence alignments frequently display the symbol “-” to indicate the absence of a residue in order to maximize the alignment.

Accordingly, the nucleotide sequences of *D. melanogaster* and *D. simulans* contain twenty-two enumerated and specifically defined nucleotides, whereas the nucleotide sequence of *D. virilis* contains nineteen. Thus, each sequence is required by ST.26 paragraph 7(a) to be included in a sequence listing with separate sequence identification numbers.

Question 3: How should the sequence(s) be represented in the sequence listing?

Drosophila melanogaster sequence must be included in a sequence listing as:

acattgaatctcataccacttt (SEQ ID NO: 61)

Drosophila virilis sequence must be included in a sequence listing as:

acatggatcccacgacttt (SEQ ID NO: 62)

Drosophila simulans sequence must be included in a sequence listing as:

gtatggcgtcgtatsgtagttt (SEQ ID NO: 63)

Relevant ST.26 paragraphs: 7(a), 13, and 93

93절 – 일차 서열 및 변이(각각 잔기로 나열됨)

실시예 93-1: 나열된 변이의 표현

발명의 설명에 다음과 같은 서열 배열이 포함되어 있다:

<i>D. melanogaster</i>	ACATTGAATCTCATACCACTTT
<i>D. virilis</i>	...-..G...C...-.G.....
<i>D. simulans</i>	GT..G.CG..GT..SGT.G...

질문 1: ST.26에 서열을 포함시켜야 하는가?

예

"이 위치는 그 위의 위치와 동일하다"는 것을 나타내기 위해 서열 정렬 상에 "점들(dots)"을 포함하는 것이 일반적이다. 따라서, *D. virilis* 및 *D. simulans* 서열의 "점"은 나열되고, 구체적으로 정의된 핵산염기로 간주되는데, 그 이유는 주어진 위치가 *D. melanogaster*에서와 같은 핵산염기임을 나타내는 간단한 방법이기 때문이다. 또한, 서열 정렬은 정렬을 최대화하기 위해 잔기가 없음을 나타내는 기호 "-"를 종종 표시한다.

따라서, *D. melanogaster* 및 *D. simulans*의 핵산염기 서열은 22개는 나열된 구체적으로 정의된 핵산염기를 함유하는 반면, *D. virilis*의 핵산염기 서열은 19개를 보유한다. 따라서 각 서열은 ST.26의 7(a)절에 의해 별도의 서열식별번호를 갖는 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 서열을 어떻게 표현해야 하는가?

Drosophila melanogaster 서열은 다음과 같이 서열목록에 포함되어야 한다:

acattgaatctcataccacttt (SEQ ID NO: 61)

Drosophila virilis 서열은 다음과 같이 서열목록에 포함되어야 한다:

acatggatcccacgacttt (SEQ ID NO: 62)

Drosophila simulans 서열은 다음과 같이 서열목록에 포함되어야 한다:

gtatggcgtcgtatsgtagttt (SEQ ID NO: 63)

관련 ST.26 절: 7(a), 13 및 93

Example 93-2: Representation of enumerated variants

The description includes the following table of a peptide and functional variants thereof. A blank space in the table below indicates that an amino acid in the variant is the same as the corresponding amino acid in the "Sequence" and a "-" indicates deletion of the corresponding amino acid in the "Sequence".

Position	1	2	3	4	5	6	7	8	9
Sequence	A	V	L	T	Y	L	R	G	E
Variant 1									A
Variant 2			P			P			
Variant 3			A	I	G	Y			
Variant 4							-		

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

As indicated, a blank space in this table indicates that an amino acid in the variant is the same as the corresponding amino acid in the "Sequence". Therefore, the amino acids of the variant sequences are enumerated and specifically defined.

Since the four variant sequences each contain more than four enumerated and specifically defined amino acids, each sequence is required by ST.26 paragraph 7(b) to be included in a sequence listing with separate sequence identification numbers.

Question 3: How should the sequence(s) be represented in the sequence listing?

AVLTYLRGE (SEQ ID NO: 77)

AVLTYLRGA (SEQ ID NO: 78)

AVPTYPRGE (SEQ ID NO: 79)

AVAIGYRGE (SEQ ID NO: 80)

AVLTYLGE (SEQ ID NO: 81)

Relevant ST.26 paragraphs: 7(b), 26, and 93

실시예 93-2: 나열된 변이의 표현

발명의 설명에 펩티드 및 이의 기능적 변이를 나타내는 아래 표가 포함되어 있다. 아래 표의 빈 공간은 변이 아미노산이 "Sequence"의 해당 아미노산과 동일함을 나타내고, "-"는 "Sequence"의 해당 아미노산의 삭제를 나타낸다.

Position	1	2	3	4	5	6	7	8	9
Sequence	A	V	L	T	Y	L	R	G	E
Variant 1									A
Variant 2			P			P			
Variant 3			A	I	G	Y			
Variant 4							-		

질문 1: ST.26에 서열을 포함시켜야 하는가?

예

나타낸 바와 같이, 이 표에서 빈 공간은 변이체의 아미노산이 "Sequence"의 대응하는 아미노산과 동일함을 나타낸다. 따라서, 변이 서열은 나열된 구체적으로 정의된 아미노산이다.

4개의 변이 서열은 각각 4개 이상의 나열된 구체적으로 정의된 아미노산을 함유하기 때문에, ST.26의 7(b)절에 의해 각각의 서열은 별도의 서열식별번호를 가지며 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

AVLTYLRGE (SEQ ID NO: 77)

AVLTYLRGA (SEQ ID NO: 78)

AVPTYPRGE (SEQ ID NO: 79)

AVAIGYRGE (SEQ ID NO: 80)

AVLTYLGE (SEQ ID NO: 81)

관련 ST.26 절: 7(b), 26 및 93

Example 93-3: Representation of a consensus sequence

A patent application includes Figure 1 with the following multiple sequence alignment.

<i>Consensus</i>	LEGnEQFINAAkIIRHPkYnrkTlnNDImL IK
<i>Homo sapiens</i>	LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK
<i>Pongo abelii</i>	LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK
<i>Papio Anubis</i>	LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK
<i>Rhinopithecus roxellana</i>	LEGTEQFINAAKIIRHPNYNRI TLDNDILLIK
<i>Pan paniscus</i>	LEGNEQFINAAKIIRHPKYNRI TLNNDIMLIK
<i>Rhinopithecus bieti</i>	LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK
<i>Rhinopithecus roxellana</i>	LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK

The consensus sequence includes upper case letters to represent conserved amino acid residues, while the lower case letters “n”, “a”, “k”, “r”, “l” and “m” represent the predominant amino acid residues among the aligned sequences.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The lower case letters in the consensus sequence each represent a single amino acid residue. Consequently, the consensus sequence, as well as each of the remaining seven sequences in Figure 1, includes at least four specifically defined amino acids. ST.26 paragraph 7(b) requires inclusion of all eight sequences in the sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The lower case letters in the consensus sequence are being used as ambiguity symbols to represent the predominant amino acid among the possible variants for a specific position. Therefore, the lower case letters “n”, “a”, “k”, “r”, “l” and “m” are conventional symbols used in a nonconventional manner and the consensus sequence must be represented using an ambiguity symbol in place of each of the lower case letters.

The most restrictive ambiguity symbol should be used. For most positions in the consensus sequence, “X” is the most restrictive ambiguity symbol; however, the most restrictive ambiguity symbol for “D” or “N” in positions 20 and 25 is “B”. The consensus sequence should be included in the sequence listing as:

LEGXEQFINAXXIIRHPXYBXXTXBNDXLIK (SEQ ID NO: 82)

According to paragraph 27, the symbol “X” will be construed as any one of “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, or “V”, except where it is used with a further description in the feature table. Therefore, each “X” in the consensus sequence must be further described in a feature table using the feature key “VARIANT” and the qualifier “note” to indicate the possible variants for each position.

The remaining seven sequences must be included in the sequence listing as:

LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK (SEQ ID NO: 83)

LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK (SEQ ID NO: 84)

LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK (SEQ ID NO: 85)

LEGTEQFINAAKIIRHPNYNRI TLDNDILLIK (SEQ ID NO: 86)

LEGNEQFINAAKIIRHPKYNRI TLNNDIMLIK (SEQ ID NO: 87)

LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 88)

LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 89)

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraphs: 7(b), 26, 27, 93, and 97

실시예 93-3: 일치(consensus) 서열의 표현

특허출원에 다음과 같은 다중 서열 배열을 갖는 그림 1이 포함되어 있다.

<i>Consensus</i>	LEGNEQFINAAkIIRHPkYnrkTlnNDImLIK
<i>Homo sapiens</i>	LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK
<i>Pongo abelii</i>	LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK
<i>Papio Anubis</i>	LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK
<i>Rhinopithecus roxellana</i>	LEGTEQFINAAKIIRHPNYNRI TLDNDILLIK
<i>Pan paniscus</i>	LEGNEQFINAAKIIRHPKYNRI TLNNDIMLIK
<i>Rhinopithecus bieti</i>	LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK
<i>Rhinopithecus roxellana</i>	LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK

일치(consensus) 서열은 보존된 아미노산 잔기를 표현하기 위해 대문자를 표기하는 반면, 소문자 "n", "a", "k", "r", "l" 및 "m"은 정렬된 서열 중에 우세한(predominant) 아미노산 잔기를 표현한다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

일치(consensus) 서열 소문자 각각은 하나의 아미노산 잔기를 표현한다. 결과적으로, 일치(consensus) 서열 및 그림 1의 나머지 7개의 서열 각각은 4개 이상의 구체적으로 정의된 아미노산을 포함한다. ST.26 7(b)절에 따라 서열목록에 8개의 서열 모두가 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

일치(consensus) 서열에서 소문자는 특정 위치에 대한 가능한 변이 중에서 우세한 아미노산을 표현하기 위해 다중아미노산 기호로 사용된다. 따라서 소문자 "n", "a", "k", "r", "l" 및 "m"은 비통상적인 방식으로 사용되는 일반적인 기호이며, 일치 서열은 각 소문자 대신 다중아미노산 기호를 사용하여 표현하여야 한다.

가장 제한적인 다중아미노산 기호를 사용하여야 한다. 일치(consensus) 서열의 대부분의 위치에서 "X"는 가장 제한적인 다중아미노산 기호이다. 그러나, 위치 20 및 25에서의 "D" 또는 "N"에 대한 가장 제한적인 다중아미노산 기호는 "B"이다. 일치 서열은 서열목록에 다음과 같이 포함되어야 한다.

LEGXEQFINAXXIIIRHPXYBXXTXBNDIXLIK (SEQ ID NO: 82)

27절에 따라 "X"는 특정 표에서 추가 설명과 함께 사용되는 경우를 제외하고는, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다. 따라서 일치 서열의 "X"는 각 위치의 가능한 변이라는 것을 나타내려면 "VARIANT" 특징 기호와 "noteE" 한정자를 사용하여 추가로 특징표에 설명되어야 한다.

나머지 7개의 서열은 서열목록에 다음과 같이 포함되어야 한다:

LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK (SEQ ID NO: 83)

LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK (SEQ ID NO: 84)

LEGTEQFINAAKIIRHPDYDRKTLNNDILLIK (SEQ ID NO: 85)

LEGTEQFINAAKIIRHPNYNRI TLDNDILLIK (SEQ ID NO: 86)

LEGNEQFINAAKIIRHPKYNRI TLNNDIMLIK (SEQ ID NO: 87)

LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 88)

LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK (SEQ ID NO: 89)

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 7(b), 26, 27, 93 및 97

Paragraph 94 – Variant sequence disclosed as a single sequence with enumerated alternative residues

Example 94-1: Representation of single sequence with enumerated alternative amino acids

A patent application claims a peptide of the sequence:

(i) Gly-Gly-Gly-[Leu or Ile]-Ala-Thr-[Ser or Thr]

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The sequence provides four specifically defined amino acids and ST.26 paragraph 7(b) requires inclusion of the sequence in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

Table 3 of Annex I, Section 3 defines the ambiguity symbol “J” as isoleucine or leucine. Therefore, the preferred representation of the sequence is:

GGGJATX (SEQ ID NO: 64)

which requires a further description in a feature table using the feature key “VARIANT” and the qualifier “note” to indicate that the “X” is serine or threonine.

Alternatively, the sequence may be represented, for example, as:

GGGLATS (SEQ ID NO: 65)

which requires a further description in a feature table using the feature key “VARIANT” and the qualifier “note” to indicate that L can be replaced by I, and S can be replaced by T.

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraph(s): 7(b), 8, 26, 27, 94, and 97

94절 - 나열된 선택 잔기를 갖는 단일 서열로서 개시된 변이 서열

실시예 94-1: 나열된 선택적 아미노산을 가진 단일 서열의 표현

특허출원에 다음과 같은 서열의 펩티드가 특허청구된다:

(i) Gly-Gly-Gly-[Leu 또는 Ile]-Ala-Thr-[Ser 또는 Thr]

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

본 서열은 4개의 구체적으로 정의된 아미노산을 제공하므로 ST.26 7(b)절에 따라 서열이 서열목록에 포함된다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

별첨 I, 섹션 3의 표 3은 다중아미노산 기호 "J"를 이소류신(isoleucine) 또는 류신(leucine)으로 정의한다. 따라서 서열의 바람직한 표현은 다음과 같다.

GGGJATX (SEQ ID NO: 64)

이 경우 특징 기호 "VARIANT"와 한정자 "note"를 사용하여 "X"가 세린(serine) 또는 트레오닌(threonine)임을 나타내는 특징 표에 대한 추가 설명이 필요하다.

선택적으로, 서열은 다음의 예처럼 표현될 수 있다.

GGGLATS (SEQ ID NO: 65)

이 경우 특징 표에는 L을 I로 대체할 수 있고 S를 T로 대체할 수 있음을 제시하기 위한 특징 기호 "VARIANT"와 한정자 "note"를 사용하여 추가적인 설명이 필요하다.

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 7(b), 8, 26, 27, 94 및 97

Paragraph 95(a) – A variant sequence disclosed only by reference to a primary sequence with multiple independent variations

Example 95(a)-1: Representation of a variant sequence by annotation of the primary sequence

An application contains the following disclosure:

“Peptide fragment 1 is Gly-Leu-Pro-Xaa-Arg-Ile-Cys wherein Xaa can be any amino acid....”

In another embodiment, peptide fragment 1 is Gly-Leu-Pro-Xaa-Arg-Ile-Cys wherein Xaa can be Val, Thr, or Asp....

In another embodiment, peptide fragment 1 is Gly-Leu-Pro-Xaa-Arg-Ile-Cys wherein Xaa can be Val.”

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

“Peptide fragment 1” in each of the three disclosed embodiments provides at least six specifically defined amino acids; therefore, the sequence must be included in a sequence listing as required by ST.26 paragraph 7(b).

Question 3: How should the sequence(s) be represented in the sequence listing?

In this example, the enumerated sequence of “Peptide fragment 1” is disclosed three times, as three different embodiments, each with an alternative description of Xaa. In this example, “X” is the most restrictive ambiguity symbol for the Xaa position.

ST.26 requires inclusion of the disclosed enumerated sequence only once. In the most encompassing of the three embodiments, Xaa is any amino acid (see Introduction to this document). Therefore, the sequence that must be included in the sequence listing is:

GLPXRIC (SEQ ID NO: 66)

According to paragraph 27, “X” will be construed as any one of “A”, “R”, “N”, “D”, “C”, “Q”, “E”, “G”, “H”, “I”, “L”, “K”, “M”, “F”, “P”, “O”, “S”, “U”, “T”, “W”, “Y”, or “V”, except where it is used with a further description in the feature table. Since “X” in SEQ ID NO: 66 represents “any amino acid”, it must be annotated with the feature key VARIANT and a note qualifier with the value, “X can be any amino acid”.

Where practicable, each “X” should be annotated individually. However, a region of contiguous “X” residues, or a multitude of “X” residues dispersed throughout the sequence, may be jointly described with the feature key VARIANT using the syntax “x..y” as the location descriptor, where x and y are the positions of the first and last “X” residues, and a note qualifier with the value, “X can be any amino acid”.

Inclusion of any additional sequences essential to the disclosure or claims of the invention is strongly encouraged, as discussed in the introduction to this document.

For the above example, it is strongly encouraged that the following additional three sequences are included in the sequence listing, each with their own sequence identification number:

GLPV RIC (SEQ ID NO: 67)

GLPTRIC (SEQ ID NO: 68)

GLPDRIC (SEQ ID NO: 69)

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraph(s): 7(b), 26, 27, and 95(a)

95(a)절 - 다수의 독립적 변이를 지닌 일차 서열의 참조만으로 개시된 변이 서열

실시예 95(a)-1: 일차 서열의 주석에 의한 변이 서열의 표현

출원서에 다음과 같은 내용이 포함된다:

"펩타이드 단편 1은 Gly-Leu-Pro-Xaa-Arg-Ile-Cys이며, 여기서 Xaa는 임의의 아미노산일 수 있다.

다른 구체적 실시예에서, 펩타이드 단편 1은 Gly-Leu-Pro-Xaa-Arg-Ile-Cys이며, 여기서 Xaa는 Val, Thr 또는 Asp 일 수 있다.

다른 구체적 실시예에서, 펩타이드 단편 1은 Gly-Leu-Pro-Xaa-Arg-Ile-Cys이며, 여기서 Xaa는 Val 일 수 있다."

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

3개의 개시된 구체적 실시예 각각에서 "펩타이드 단편 1"은 6개 이상의 구체적으로 정의된 아미노산을 제공한다; 그러므로 이 서열은 ST.26의 7(b)절에 따라 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

이 예에서, "펩타이드 조각 1"의 나열된 서열은 각각 Xaa의 대안적 설명을 갖는 3개의 서로 다른 실시 양태로서 3회 개시되어 있다. 이 실시예에서 "X"는 Xaa 위치에 대해 가장 제한적인 다중아미노산 기호이다.

ST.26은 공개된 나열된 서열을 한 번만 포함하여야 한다. 3가지 구체적 실시예 중 가장 포괄적인 것으로, Xaa는 임의의 아미노산이다(본 문서의 서론 참조). 따라서 서열목록에 포함되어야 하는 서열은 다음과 같다:

GLPXRIC (SEQ ID NO: 66)

27절에 따르면, "X"는 특징 표에서 추가 설명과 함께 사용되는 경우를 제외하고는, "A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "O", "S", "U", "T", "W", "Y" 또는 "V" 중 하나로 해석된다. 서열 번호 66의 "X"는 "임의의 아미노산"을 표현하므로, 특징 기호 VARIANT와 "X can be any amino acid"라는 값을 가진 note 한정자로 주석을 달아야 한다.

가능한 경우, 각 "X"는 독립적으로 주석을 달아야 한다. 그러나, 연속 "X" 잔기의 영역, 또는 서열 전체에 분산된 다수의 "X" 잔기는 위치 설명자로서 "x..y" 구문을 사용하는 특징 기호 'VARIANT'와 "X can be any amino acid"라는 값을 가진 'note' 한정자를 이용하여 함께 기술될 수 있다. 여기서 x와 y는 첫 번째 및 마지막 "X" 잔기의 위치이다.

본 문서의 서두에서 논의된 바와 같이, 본 발명의 개시 또는 청구 범위에 필수적인 임의의 추가 서열의 포함이 강력히 권장된다.

위의 예에서는 다음과 같은 추가 3개의 서열은 각각의 고유한 서열식별번호를 가지며, 서열목록에 포함되는 것을 강력히 권장한다.

GLPVVIC (SEQ ID NO: 67)

GLPTRIC (SEQ ID NO: 68)

GLPDRIC (SEQ ID NO: 69)

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 7(b), 26, 27 및 95(a)

Paragraph 95(b) – A variant sequence disclosed only by reference to a primary sequence with multiple interdependent variations

Example 95(b)-1: Representation of individual variant sequences with multiple interdependent variations

A patent application describes the following consensus sequence:

cgaatg_{n1}cccactacgaatg_{n2}cacgaatg_{n3}cccaca

wherein n_1 , n_2 , and n_3 can be a, t, g, or c.

Several variant sequences are disclosed as follows:

if n_1 is a, then n_2 and n_3 are t, g, or c;

if n_1 is t, then n_2 and n_3 are a, g, or c;

if n_1 is g, then n_2 and n_3 are t, a, or c;

if n_1 is c, then n_2 and n_3 are t, g, or a.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The sequence has more than ten enumerated and “specifically defined” nucleotides and is required by ST.26 paragraph 7(a) to be included in a sequence listing.

Question 3: How should the sequence(s) be represented in the sequence listing?

The enumerated sequence contains more than ten specifically defined nucleotides and three “n” residues. ST.26 requires inclusion of the disclosed enumerated sequence and where an ambiguity symbol is appropriate, the most restrictive symbol should be used. In this example, n_1 , n_2 , and n_3 can be a, t, g, or c, so “n” is the most restrictive ambiguity symbol. Therefore, the sequence that must be included in the sequence listing is:

cgaatg_ncccactacgaatg_ncacgaatg_ncccaca (SEQ ID NO: 70)

ST.26 paragraph 15 states that “the symbol “n” will be construed as any one of “a”, “c”, “g”, or “t/u” except where it is used with a further description in the feature table. Since the value of every “n” residue in this sequence is equivalent to the default “a”, “c”, “g”, or “t”, no further annotation is required.

The enumerated sequence contains variations at three distinct locations and the occurrence of the variations is interdependent. Inclusion of additional sequences which represent additional embodiments that are a key part of the invention is **strongly** encouraged, as discussed in the introduction to this document. Therefore, according to ST.26 paragraph 95(b), the additional embodiments should be included in a sequence listing as four separate sequences, each with its own sequence identification number:

cgaatg_acccactacgaatg_bcacgaatg_bcccaca (SEQ ID NO: 71)

cgaatg_tcccactacgaatg_vcacgaatg_vcccaca (SEQ ID NO: 72)

cgaatg_gcccactacgaatg_hcacgaatg_hcccaca (SEQ ID NO: 73)

cgaatg_ccccactacgaatg_dcacgaatg_dcccaca (SEQ ID NO: 74)

(Note that b = t, g, or c; v = a, g, or c; h = t, a, or c; and d = t, g, or a; see Annex I, Section 1, Table 1)

According to ST.26 paragraph 15, the most restrictive symbol must be used to represent variable positions. Consequently, n_2 and n_3 must not be represented by “n” in the sequence.

CAUTION: The preferred representation of the sequence indicated above is directed to the provision of a sequence listing on the filing date of a patent application. The same representation may not be applicable to a sequence listing provided subsequent to the filing date of a patent application, since consideration must be given to whether the information provided could be considered by an IPO to add subject matter to the original disclosure.

Relevant ST.26 paragraphs: 7(a), 15, and 95(b)

[Appendix to Annex VI follows]

95(b)절 - 다수의 상호 의존적 변이를 갖는 일차 서열의 참조만으로 개시된 변이 서열

실시예 95(b)-1: 다수의 상호 의존적 변이를 갖는 개별 변이 서열의 표현

특허출원에 다음과 같은 공통 서열이 기재되어 있다:

cgaatgn₁cccactacgaatgn₂cacgaatgn₃cccaca

여기서, n₁, n₂ 및 n₃ 은 a, t, g 또는 c일 수 있다.

여러 변형 서열이 다음과 같이 개시된다:

n₁이 a이면, n₂ 및 n₃ 은 t, g 또는 c이고;

n₁이 t이면, n₂ 및 n₃ 은 a, g 또는 c이고;

n₁이 g이면, n₂ 및 n₃ 은 t, a 또는 c이고;

n₁이 c이면, n₂ 및 n₃ 은 t, g 또는 a이다.

질문 1: ST.26은 이 서열(들)의 포함을 요구하는가?

예

서열은 10개 이상의 나열된 "구체적으로 정의된" 핵산염기를 가지며 ST.26의 7(a)절에 의해 서열목록에 포함되어야 한다.

질문 3: 서열목록에서 이 서열(들)을 어떻게 표현해야 하는가?

나열된 서열은 10개 이상의 구체적으로 정의된 핵산염기 및 3개의 "n" 잔기를 보유한다. ST.26은 개시되고 나열된 서열의 포함하는 것을 요구하며, 다중핵산염기 기호가 적절한 경우 가장 제한적인 기호가 사용되어야 한다. 본 예에서, n₁, n₂ 및 n₃ 은 a, t, g 또는 c 일 수 있으므로 "n"이 가장 제한적인 다중핵산염기 기호이다. 따라서 서열목록에 포함되어야 하는 서열은 다음과 같다:

cgaatgncccactacgaatgncacgaatgncccaca (SEQ ID NO: 70)

ST.26 절15는 "심볼 n"은 특징 표에 추가적인 설명으로 사용된 경우를 제외하고 "a", "c", "g", "t/u" 중 어떤 하나로 표현된다. 이 서열의 모든 "n" 잔기의 값은 기본값인 "a", "c", "g", "t/u"와 같으므로 추가적인 주석이 필요하지 않다.

나열된 서열은 3개의 별개의 위치에서 변이를 포함하고 변이의 발생은 상호 의존적이다. 본 문서의 서두에서 논의된 바와 같이, 본 발명의 주요 부분인 추가 실시 양태를 표현하는 추가 서열의 포함이 강력히 권장된다. 따라서 ST.26의 95(b)절에 따라 부가적인 실시예는 독립된 서열식별번호를 가지는 4개의 개별 서열이 서열목록에 포함되어야 한다:

cgaatgaccactacgaatgbcacgaatgbcaccaca (SEQ ID NO: 71)

cgaatgtcccactacgaatgvcacgaatgvccaccaca (SEQ ID NO: 72)

cgaatgccccactacgaatghcacgaatghccaccaca (SEQ ID NO: 73)

cgaatgccccactacgaatgdcacgaatgdccaccaca (SEQ ID NO: 74)

(b = t, g 또는 c; v = a, g 또는 c; h = t, a 또는 c; 및 d = t, g 또는 a; 별첨 1, 섹션 1, 표 1 참조)

ST.26의 15절에 따르면 변이 위치를 나타내는 데에 가장 제한적인 기호를 사용하여야 한다. 결과적으로, n₂ 및 n₃은 서열에서 "n"으로 표현하지 않아야 한다.

주의: 위에 제시된 서열의 선호 표현은 특허출원의 출원일에 서열목록을 제공하기 위한 것이다. 특허출원의 출원일 이후에 제공된 서열목록에는 동일한 표현이 적용되지 않을 수 있다. 이는 제공된 정보가 IPO에 의해 원본 개시 내용에 주제를 추가한 것으로 간주될 수 있기 때문이다.

관련 ST.26 절: 7(a), 15 및 95(b)

[ST.26의 별첨 VI의 부록으로 이어짐]

APPENDIX

GUIDANCE DOCUMENT SEQUENCES IN XML

The Appendix is available at:

https://www.wipo.int/standards/en/xml_material/st26/st26-annex-vi-appendix-guidance-document-sequences.xml

[Annex VII follow s]

부록

XML로 작성된 지침 서열

본 부록은 아래 사이트에서 볼 수 있음:

<https://www.wipo.int/standards/en/docs/st26-annex-vi-appendix-guidance-document-sequences.xml>

[별첨 VII로 이어짐]

ANNEX VII

RECOMMENDATION FOR THE TRANSFORMATION OF A SEQUENCE LISTING FROM ST.25 TO ST.26: POTENTIAL ADDED OR DELETED SUBJECT MATTER

Version 1.5

*Revision approved by the Committee on WIPO Standards (CWS)
at its ninth session on November 5, 2021*

Introduction

The requirements for the presentation of nucleotide and amino acid sequences differ between WIPO Standards ST.25 and ST.26. Consequently, the question has been raised as to whether Standard ST.26 would require addition or deletion of any subject matter in a sequence listing submitted as part of an international application under Standard ST.26 that may not be supported by an application from which priority is claimed.

Scope of the Document

This document addresses the mandatory requirements of ST.26, and any potential consequences of those requirements. This document does not address every possible scenario; if the means of representation in ST.26, of information contained in an ST.25 sequence listing, is not clear, then the information may always be included in the application description to avoid deleted subject matter.

Recommendations for Potential Added or Deleted Subject Matter

Review of the issues contained in this document demonstrates that transformation from ST.25 to ST.26 by itself should not inherently result in added or deleted subject matter, in particular, where the ST.25 sequence listing was fully compliant with Standard ST.25. However, there are certain scenarios that will require applicant caution. Recommendations have been provided to avoid added or deleted subject matter.

Scenario 1

ST.25 uses numeric identifiers to tag various types of data, e.g., <110> for Applicant Name. ST.26 uses terms in the English language, as element names and attributes, for data tagging.

Recommendation:

The ST.26 terms simply describe the type of data content; therefore, the use of the ST.26 element names and attributes does not constitute added subject matter.

Scenario 2

ST.26 explicitly requires inclusion of: (a) branched sequences; (b) sequences with D-amino acids; (c) nucleotide analogues; and (d) sequences with abasic sites. Under ST.25, the requirement for inclusion or the prohibition of such sequences is not clear.

Recommendation:

The disclosure contained in the application should be sufficient to represent these sequences in an ST.26 sequence listing, when they may not have been included in an ST.25 sequence listing. For certain types of information required by ST.26, care must be taken not to add subject matter beyond that disclosed, e.g., see discussion below (in Scenario 4) on the mol_type qualifier for nucleotide sequences.

Scenario 3

ST.26 excludes sequences with less than 10 specifically defined nucleotides (not including "n") and less than 4 specifically defined amino acids (not including "X").

Recommendation:

The excluded sequences may be included in the application body, where those sequences have not already been included therein.

별첨 VII

ST.25에서 ST.26으로 서열목록 변환을 위한 권장 사항:
잠정적으로 추가 또는 삭제된 주제

버전 1.5

2021년 11월 5일 WIPO 표준위원회(CWS) 제9차 회의에서 승인된 개정판

서론

핵산염기 및 아미노산 서열의 제시 요건은 WIPO 표준 ST.25와 ST.26에서 다르다. 결과적으로, 표준 ST.26이 표준 ST.26 하에서 국제 출원의 일부로 제출된 서열목록에서 우선권이 주장되는 출원서에 의해 뒷받침되지 않을 수 있는 임의의 주제의 추가 또는 삭제를 요구해야 하는지에 대한 의문이 제기되어 왔다.

문서의 범위

본 문서는 ST.26 필수 요건과 해당 요건의 잠정적 결과에 대해 설명한다. 본 문서는 모든 가능한 시나리오를 다루지는 않는다. 만약 ST.25 서열목록에 포함된 정보를 ST.26에 따라서 어떻게 표현해야 할지 명확하지 않은 경우, 이 정보는 삭제된 주제가 되는 것을 피하기 위해 발명의 설명에 항상 포함될 수 있다.

잠정적으로 추가 또는 삭제된 주제에 대한 권장 사항

본 문서에 포함된 이슈(issue) 검토를 통해 특히 ST.25 서열목록이 표준 ST.25를 완전히 준수하는 경우, ST.25에서 ST.26으로의 변환 자체가 본질적으로 내용이 추가되거나 삭제되지 않아야 함을 보여준다. 그러나 출원인의 주의가 필요한 특정 시나리오가 있다. 다음의 권장사항은 추가되거나 삭제된 주제를 막기 위해 제공된다.

시나리오 1

ST.25는 숫자식별기호를 사용하여 다양한 타입의 데이터(예: 출원인 명칭 <110>)에 태그를 지정한다. ST.26은 데이터 태그 지정을 위해 요소 명칭 및 속성으로 영어로 된 용어를 사용한다.

권고사항:

ST.26 용어는 단순히 데이터 내용의 타입을 설명한다. 그러므로 ST.26 요소 이름과 속성의 사용이 추가된 주제를 구성하지는 않는다.

시나리오 2

ST.26에는 다음 사항이 명시적으로 포함된다: (a) 분지형 서열; (b) D-아미노산을 갖는 서열; (c) 핵산염기 유사체; 및 (d) 탈염기 부위(abasic site)를 갖는 서열. ST.25 규정 하에서는 이러한 서열의 포함 또는 금지 요건은 명확하지 않다.

권고사항:

ST.25 서열목록에는 포함되지 않았을지라도, ST.26 서열목록에서는 이들 서열을 표현하려면 출원에 포함된 개시 내용이 충분하여야 한다. ST.26에 의해 요구되는 특정 타입의 정보에 대해, 개시된 것 이외의 주제를 부가하지 않도록 주의하여야 한다. 예를 들면, 핵산염기 서열에 대한 mol_type 한정자에 대한(시나리오 4) 논의를 참조하시오.

시나리오 3

ST.26은 10개 미만의 구체적으로 정의된 핵산염기("n"제외) 및 4개 미만의 구체적으로 정의된 아미노산("X"제외)을 갖는 서열을 배제한다.

권고사항:

배제된 서열은 아직 이 서열이 포함되지 않은 발명의 설명 본문에 포함될 수 있다.

Scenario 4

ST.26 has the mandatory feature keys – “source” for all nucleotide sequences and “source” for all amino acid sequences, each with two mandatory qualifiers. ST.25 has a corresponding feature key for nucleotide sequences (which is rarely used) with no corresponding qualifiers and there is no corresponding feature key for amino acid sequences. **Nucleotide sequences**

ST.26 – feature key 5.37 source; mandatory qualifiers 6.45 organism and 6.39 mol_type (see ST.26 paragraph 75)

Qualifier	Value
mol_type	genomic DNA
	genomic RNA
	mRNA
	tRNA
	rRNA
	other DNA (applies to synthetic molecules)
	other RNA (applies to synthetic molecules)
	transcribed RNA
	viral cRNA
	unassigned DNA (applies where <i>in vivo</i> molecule is unknown)
	unassigned RNA (applies where <i>in vivo</i> molecule is unknown)

Amino acid sequences

ST.26 – feature key 7.30 source; mandatory qualifiers 8.3 organism and 8.1 mol_type (see ST.26 paragraph 75)

Qualifier	Value
mol_type	protein

Recommendation:

The only issue of concern is the controlled vocabulary values associated with the mol_type qualifier for nucleotide sequences. Some of the value choices listed above may not be sufficiently supported in the disclosure. Added subject matter may be avoided, however, by use of the most generic value for a particular sequence, e.g., “other DNA” and “other RNA” for a synthetic molecule and “unassigned DNA” and “unassigned RNA” for an *in vivo* molecule.

Scenario 5

Where a sequence includes “Xaa”, ST.25 requires that further information concerning that residue be included in field <223>, which accompanies fields <221> (feature name) and <222> (feature location). ST.25 does not provide a default value for “Xaa” (“X” in ST.26). However, ST.26 does provide such a default value, and therefore, further information is not always required. Two of the most frequently used annotations in peptide sequences is “any amino acid” or “any naturally occurring amino acid” for variable “Xaa” or “X”. This language could be interpreted to include amino acids other than those listed in the amino acid tables contained in either ST.25 or ST.26. The ST.26 default value for “X” with no further annotation, is any of the 22 individual amino acids listed in Annex I (see Section 3, Table 3). This ST.26 default value may itself constitute added or deleted subject matter, and therefore, adversely affect the scope of a patent application when transitioning from ST.25 to ST.26.

Recommendations:

(a) Where the ST.25 sequence listing includes a <221> feature name, <222> feature location corresponding to the Xaa, and <223> further information on Xaa, and the <221> feature name is also an appropriate ST.26 feature key, e.g., SITE, VARIANT, or UNSURE, then the ST.26 feature key should be used. Furthermore, to avoid potential deleted subject matter, the information in field <223> must be included in an accompanying qualifier “note”.

(b) Where the ST.25 sequence listing includes a <221> feature name, <222> feature location corresponding to the Xaa, and <223> further information on Xaa, and the <221> feature name is not an ST.26 feature key, then ST.26 feature keys SITE or REGION, as appropriate, should be used. Furthermore, to avoid potential deleted subject matter, the information in field <223>, as well as the inappropriate <221> feature name, must be included in an accompanying qualifier “note”. For example, an ST.25 listing used a feature name that is not in ST.25 or ST.26, <221> Variable, together with further information <223> Xaa is any amino acid. In this example, the value of the ST.26 qualifier note would be “Variable – Xaa is any amino acid”.

시나리오 4

ST.26에는 필수 특징 기호(모든 핵산염기 서열에 대한 "source" 및 임의의 아미노산 서열에 대한 "source")가 있으며 각각 2개의 필수 한정자가 있다. ST.25는 대응하는 한정자가 없이 핵산염기 서열에 대해 대응하는 특징 기호를 가지며(거의 사용되지 않음) 아미노산 서열에 대한 대응하는 특징 기호는 없다.

핵산염기 서열

ST.26 – 특징 기호 5.37 source; 필수 한정자 6.45 organism 및 6.39 mol_type(ST.26, 75절 참조)

한정자	값
mol_type	genomic DNA
	genomic RNA
	mRNA
	tRNA
	rRNA
	other DNA (synthetic molecules 에 적용)
	other RNA (synthetic molecules 에 적용)
	transcribed RNA
	viral cRNA
	unassigned DNA (<i>in vivo</i> molecule 이 알려지지 않은 곳에 적용)
unassigned RNA (<i>in vivo</i> molecule 이 알려지지 않은 곳에 적용)	

아미노산 서열

ST.26 – 특징 기호 7.30 source; 필수 한정자 8.3 organism 및 8.1 mol_type(ST.26, 75절 참조)

한정자	값
mol_type	protein

권고사항:

유일한 이슈(issue)는 핵산염기 서열에 대한 mol_type 한정자와 관련된 통제 어휘 값이다. 위에 나열된 선택 가능한 값 중 일부는 본 개시에서 충분히 뒷받침되지 않을 수 있다. 그러나, 특정 서열에 대한 가장 일반적인 값, 예를 들면 합성 분자에 대해 "other DNA" 및 "other RNA" 및 생체 내 분자에 대해 "unassigned DNA" 및 "unassigned RNA"를 사용함으로써 신규사항의 추가를 피할 수 있다.

시나리오 5

서열이 "Xaa"를 포함하는 경우, ST.25는 잔기에 관한 추가 정보가 <223> 필드에 포함되도록 요구하며, 이는 <221> 필드(특징 명칭) 및 <222>(특징 위치)와 함께 제공된다. ST.25는 "Xaa"(ST.26의 "X")에 대한 기본값을 제공하지 않는다. 그러나 ST.26은 이러한 기본값을 제공하므로 추가 정보가 항상 필요한 것은 아니다. 펩티드 서열에서 가장 빈번하게 사용되는 주석 중 2개는 가변 "Xaa" 또는 "X"에 대한 "임의의 아미노산" 또는 "임의의 자연 발생 아미노산"이다. 본 언어는 ST.25 또는 ST.26에 포함된 아미노산 표에 나열된 것 이외의 아미노산을 포함하는 것으로 해석될 수 있다. 추가 주석이 없는 "X"의 ST.26 기본값은 별첨 1에 나열된 22개의 개별 아미노산 중 하나이다(섹션 3, 표 3 참조). 본 ST.26 기본값은 그 자체가 추가되거나 삭제된 주제를 구성할 수 있으므로 ST.25에서 ST.26으로 전환할 때 특허출원의 범위에 부정적인 영향을 미친다.

권고사항:

(a) ST.25 서열목록은 Xaa의 <221> 특징 명칭과 <222> 특징 위치 및 Xaa에 대한 <223> 추가 정보를 포함하며, <221> 특징 명칭이 적절한 ST.26 특징 기호(예: SITE, VARIANT 또는 UNSURE)가 있는 경우, ST.26 특징 기호를 사용하여야 한다. 또한, 잠재적인 삭제된 주제를 피하기 위해서 <223> 필드의 정보는 상응하는 한정자 "note"에 포함되어야 한다.

(b) ST.25 서열목록은 Xaa의 <221> 특징 명칭과 <222> 특징 위치 및 Xaa에 대한 <223> 추가 정보를 포함하며, <221> 특징 명칭이 ST.26 특징 기호가 아닌 경우, ST.26 특징 기호 SITE 또는 REGION을 적절하게 사용하여야 한다. 또한 잠재적인 삭제된 주제를 방지하려면 <223> 필드의 정보와 부적합한 <221> 특징 이름이 동반되는 한정자 "note"에 포함되어야 한다. 예를 들면, 어떤 ST.25 목록에 ST.25 또는 ST.26에 없는 특징 이름을 사용하면, 추가 정보 <223> Xaa와 함께 <221> 변수는 임의의 아미노산이다. 본 예에서는 ST.26 한정자 note의 값은 "Variable – Xaais any amino acid"이 될 것이다.

(c) Where the ST.25 sequence listing provides no <221>, <222>, or <223> field corresponding to the Xaa or where fields <221> and <222> corresponding to the Xaa are included, but no information is included in a corresponding <223> field (neither scenario is compliant with ST.25, but has occurred nonetheless), any information contained in the application body to describe "Xaa" should be included in the ST.26 qualifier "note" together with an appropriate feature key, e.g., SITE, REGION, or UNSURE, and location.

Scenario 6

In ST.25, uracil is represented in the sequence by "u" and thymine is represented by "t". In ST.26, uracil and thymine are both represented in the sequence by "t" and without further annotation; "t" represents uracil in RNA and thymine in DNA.

Recommendations:

(a) Where a DNA sequence contains uracil, ST.26 considers it to be a modified nucleotide, and requires that uracil must be represented as a "t" and be further described using the feature key "modified_base", the qualifier "mod_base" with "OTHER" as the qualifier value and the qualifier "note" with "uracil" as the qualifier value. This ST.26 annotation is not considered added subject matter where the ST.25 DNA sequence contained a "u".

(b) Where an RNA sequence contains thymine, ST.26 considers it to be a modified nucleotide, and requires that thymine must be represented as a "t" and be further described using the feature key "modified_base", the qualifier "mod_base" with "OTHER" as the qualifier value and the qualifier "note" with "thymine" as the qualifier value. This ST.26 annotation is not considered added subject matter where the ST.25 RNA sequence contained a "t".

Scenario 7

In both ST.25 and ST.26, modified nucleotides or amino acids must have a further description. In ST.26, the identity of a modified nucleotide may be indicated using an abbreviation from Annex I, Section 2, Table 2, where applicable. Otherwise, the complete unabbreviated name of the modified nucleotide must be indicated. Similarly, the identity of a modified amino acid may be indicated using an abbreviation from Annex I, Section 4, Table 4, where applicable. Otherwise, the complete unabbreviated name of the modified amino acid must be indicated. In contrast, if a modified residue is not contained in an ST.25 table, use of the complete, unabbreviated name is not required, and not infrequently, an abbreviation is used instead.

Recommendations:

(a) Where only an abbreviated name, which is not in Annex I, Section 2, Table 2 or Section 4, Table 4, was used both in the application and in an ST.25 sequence listing for either a modified nucleotide or a modified amino acid, and the abbreviated name is known in the art to reference only one specific modified nucleotide or modified amino acid, then use of the full, unabbreviated name would not itself constitute added subject matter.

(b) Where only an abbreviated name, which is not in Annex I, Section 2, Table 2 or Section 4, Table 4, was used both in the application and in an ST.25 sequence listing for either a modified nucleotide or a modified amino acid (and the application contains no chemical structure), and the abbreviated name is not known in the art to reference one specific modified nucleotide or modified amino acid, i.e., the abbreviation is either not known at all in the art, or could possibly represent multiple different modified nucleotides or modified amino acids, then compliance with ST.26, without introduction of added subject matter, is not possible in this situation. Of course in this case, the priority application and sequence listing are themselves vague. To avoid potential deleted subject matter, the abbreviated name from the ST.25 sequence listing should be placed in an ST.26 "note" qualifier in addition to the value of the complete unabbreviated name of the modified nucleotide or modified amino acid. The complete unabbreviated name of the modified nucleotide or modified amino acid required in an ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include the unabbreviated name to avoid future issues.

Scenario 8

ST.25 contains a number of feature keys that are not contained in ST.26. Therefore, applicants must take care to capture the information contained in those ST.25 feature keys in a manner compliant with ST.26 without the introduction of added or deleted subject matter.

(c) ST.25 서열목록이 Xaa에 해당하는 <221>, <222> 또는 <223> 필드를 제공하지 않거나 Xaa에 해당하는 <221> 및 <222> 필드가 포함된 경우이지만, 해당 <223> 필드에 정보가 포함되어 있지 않은 경우(시나리오가 ST.25를 준수하지는 않지만 그럼에도 불구하고 발생), "Xaa"를 설명하기 위해 출원서 본문에 포함된 모든 정보는 적절한 특징 기호(예: SITE, REGION, 또는 UNSURE 및 location)와 함께 ST.26 한정자 "note"에 포함되어야 한다.

시나리오 6

ST.25에서, 우라실(uracil)은 서열에서 "u"로 표현되고 티민(thymine)은 "t"로 표현된다. ST.26에서, 우라실(uracil) 및 티민(thymine)은 모두 "t"로 추가 주석 없이 서열로 표현된다; "t"는 RNA의 우라실(uracil) 및 DNA의 티민(thymine)으로 표현된다.

권고사항:

(a) DNA 서열이 우라실(uracil)을 보유하는 경우, ST.26은 이를 변형 핵산염기로 간주하고, 우라실(uracil)은 "t"로 표현되어야 하고, 특징 기호 "modified_base", 한정자 값으로 "OTHER"인 한정자 "mod_base" 및 한정자 값으로 "uracil"인 한정자 "note"를 사용하여 추가로 설명하여야 한다. 본 ST.26 주석은 ST.25 DNA 서열이 "u"를 포함하는 경우 추가된 주제로 간주되지 않는다.

(b) RNA 서열이 티민(thymine)을 보유하는 경우, ST.26은 이를 변형 핵산염기로 간주하고 티민(thymine)을 "t"로 표현되어야 하고, 특징 기호 "modified_base", 한정자 값으로 "OTHER"인 한정자 "mod_base" 및 한정자 값으로 "thymine"인 한정자 "note"를 사용하여 추가로 설명하여야 한다. 본 ST.26 주석은 ST.25 RNA 서열이 "t"를 포함하는 경우 추가된 주제로 간주되지 않는다.

시나리오 7

ST.25 및 ST.26 모두에서, 변형 핵산염기 또는 아미노산은 추가의 설명을 가져야 한다. ST.26에서, 변형 핵산염기의 식별(identity)은 표 2, 섹션 2의 별첨 1의 약어를 사용하여 표시될 수 있다. 그렇지 않으면, 미변형 핵산염기의 축약되지 않은 완전한 명칭을 표시하여야 한다. 비슷하게, 변형 아미노산의 식별은 별첨 1의 표 4, 섹션 4의 약어를 사용하여 표시될 수 있다. 그렇지 않으면, 미변형 아미노산의 축약되지 않은 완전한 명칭을 표시하여야 한다. 반대로, 변형된 잔기가 ST.25 표에 포함되어 있지 않은 경우, 축약되지 않은 완전한 명칭을 요구하지 않으며, 대신 약자가 사용된다.

권고사항:

(a) 별첨 1, 섹션 2, 표 2 또는 섹션 4, 표 4에 없는 약식 명칭만 변형 핵산염기 또는 변형 아미노산에 대해 출원서 및 ST.25 서열목록 모두에 사용되며, 그 약식 명칭이 하나의 특정 변형 핵산염기 또는 변형 아미노산만을 지칭하는 것으로 공지된 경우, 축약되지 않은 명칭 자체가 추가된 주제를 구성하지 않을 수 있다.

(b) 별첨 1, 섹션 2, 표 2 또는 섹션 4, 표 4에 없는 약식 명칭만 변형 핵산염기 또는 변형 아미노산에 대한 출원서 및 ST.25 서열목록 모두에 사용되며(또한 출원서는 화학 구조를 포함하지 않음), 약식 명칭이 하나의 특정 변형 핵산염기 또는 변형 아미노산을 언급하는 것으로 알려지지 않고, 즉 다시 말해 약식 명칭이 전혀 알려지지 않았거나 또는 다수의 다른 변형 핵산염기 또는 변형 아미노산을 표현하는 경우, 발명의 기술적 구성을 추가하지 않고 ST.26을 준수하는 것은 불가능하다. 물론 이 경우 우선권 출원과 서열목록 자체는 모호하다. 잠재적으로 삭제된 주제를 피하기 위해, ST.25 서열목록의 약식 명칭은 변형 핵산염기 또는 변형 아미노산의 단축되지 않은 완전한 명칭 값 외에 ST.26 "note" 한정자에 배치하여야 한다. ST.26 서열목록에서 요구되는 변형 핵산염기 또는 변형 아미노산의 단축되지 않은 완전한 명칭은 이전의 출원에 우선권이 부여되지 않을 것이다. 향후 문제를 피하기 위해 단축되지 않은 명칭을 포함하도록 원본(ST.25) 서열목록 및 출원서 공개를 작성하는 데에 주의를 기울여야 한다.

시나리오 8

ST.25에는 ST.26에 포함되지 않은 여러 특징 기호가 포함되어 있다. 따라서 출원인은 추가되거나 삭제되는 주제를 도입없이 ST.26을 준수하는 방식으로 ST.25 특징 기호에 포함된 정보를 표시하여야 한다.

Recommendations:

The following table provides guidance as to the manner in which the information contained in a former ST.25 feature key may be included in compliance with ST.26 without the introduction of added or deleted subject matter. Numbers 1-23 are feature keys related to nucleotide sequences and numbers 24-43 are feature keys related to amino acid sequences.

권고사항:

다음 표는 이전 ST.25 특징 기호에 포함된 정보가 주제 추가 또는 삭제 없이 ST.26에 따라 포함될 수 있는 방식에 대한 지침을 제공한다.

1-23은 핵산염기 서열과 관련된 특징 기호이고 24-43은 아미노산 서열과 관련된 특징 기호이다.

No.	ST.25 Feature key <221>	ST.26 equivalent		
		Feature key	Qualifier	Qualifier value
1	allele	misc_feature	allele	<223> value
2	attenuator	regulatory ¹	regulatory_class ¹	"attenuator"
			note (if <223> present)	<223> value
3	CAAT_signal	regulatory ¹	regulatory_class ¹	"CAAT_signal"
			note (if <223> present)	<223> value
4	conflict	misc_feature	note	"conflict" and <223> value
5	enhancer	regulatory ¹	regulatory_class ¹	"enhancer"
			note (if <223> present)	<223> value
6	GC_signal	regulatory ¹	regulatory_class ¹	"GC_signal"
			note (if <223> present)	<223> value
7	LTR	mobile_element ¹	rpt_type ¹	"long_terminal_repeat"
			note (if <223> present)	<223> value
8	misc_signal	regulatory ¹	regulatory_class ¹	"other"
			note (if <223> present)	<223> value
9	mutation	variation	note	"mutation" and <223> value
10	old_sequence	misc_feature	note	"old_sequence" and <223> value
11	polyA_signal	regulatory ¹	regulatory_class ¹	"polyA_signal_sequence"
			note (if <223> present)	<223> value
12	promoter	regulatory ¹	regulatory_class ¹	"promoter"
			note (if <223> present)	<223> value
13	RBS	regulatory ¹	regulatory_class ¹	"ribosome_binding_site"
			note (if <223> present)	<223> value
14	repeat_unit (a) w hen repeat_region not used	misc_feature	note	"repeat_unit" and <223> value
	repeat_unit (b) w hen repeat_region used	repeat_region	rpt_unit_range	1 st residue..last residue
15	satellite	repeat_region	satellite	"satellite" (or "microsatellite" or "minisatellite" – if supported)
			note (if <223> present)	<223> value
16	scRNA	ncRNA ¹	ncRNA_class ¹	"scRNA"
			note (if <223> present)	<223> value
17	snRNA	ncRNA ¹	ncRNA_class ¹	"snRNA"
			note (if <223> present)	<223> value
18	TATA_signal	regulatory ¹	regulatory_class ¹	"TATA_box"
			note (if <223> present)	<223> value
19	terminator	regulatory ¹	regulatory_class ¹	"terminator"
			note (if <223> present)	<223> value
20	3'clip	misc_feature	note	"3'clip" and <223> value
21	5'clip	misc_feature	note	"5'clip" and <223> value
22	-10_signal	regulatory ¹	regulatory_class ¹	"minus_10_signal"
			note (if <223> present)	<223> value
23	-35_signal	regulatory ¹	regulatory_class ¹	"minus_35_signal"
			note (if <223> present)	<223> value

¹ ST.26 may require that a specific ST.25 feature, e.g., TATA_signal, be replaced by a broader feature key/qualifier/value, e.g., regulatory/regulatory_class/TATA_box. In such a case, the narrower ST.25 feature will be afforded priority to the earlier application. However, the full breadth of the broader ST.26 feature key/qualifier, e.g., regulatory/regulatory_class, will not be afforded priority to the earlier application.

No.	ST.25 특징 기호 <221>	ST.26 대응		
		특징 기호	한정자	한정자 값
1	allele	misc_feature	allele	<223> 값
2	attenuator	regulatory ¹	regulatory_class ¹	"attenuator"
			note (<223>이 있으면)	<223> 값
3	CAAT_signal	regulatory ¹	regulatory_class ¹	"CAAT_signal"
			note (<223>이 있으면)	<223> 값
4	conflict	misc_feature	note	"conflict" 및 <223> 값
5	enhancer	regulatory ¹	regulatory_class ¹	"enhancer"
			note (<223>이 있으면)	<223> 값
6	GC_signal	regulatory ¹	regulatory_class ¹	"GC_signal"
			note (<223>이 있으면)	<223> 값
7	LTR	mobile_element	rpt_type	"long_terminal_repeat"
			note (<223>이 있으면)	<223> 값
8	misc_signal	regulatory ¹	regulatory_class ¹	"other"
			note (<223>이 있으면)	<223> 값
9	mutation	variation	note	"mutation" 및 <223> 값
10	old_sequence	misc_feature	note	"old_sequence" 및 <223> 값
11	polyA_signal	regulatory ¹	regulatory_class ¹	"polyA_signal_sequence"
			note (<223>이 있으면)	<223> 값
12	promoter	regulatory ¹	regulatory_class ¹	"promoter"
			note (<223>이 있으면)	<223> 값
13	RBS	regulatory ¹	regulatory_class ¹	"ribosome_binding_site"
			note (<223>이 있으면)	<223> 값
14	repeat_unit (a) repeat_region이 미사용된 경우	misc_feature	note	"repeat_unit" 및 <223> 값
	repeat_unit (b) repeat_region이 사용된 경우	repeat_region	rpt_unit_range note (<223>이 있으면)	1 st residue..last residue <223> 값
15	satellite	repeat_region	satellite	"satellite" (또는 "microsatellite" 또는 "minisatellite" – 뒷받침되는 경우)
			note (<223>이 있으면)	<223> 값
16	scRNA	ncRNA ¹	ncRNA_class ¹	"scRNA"
			note (<223>이 있으면)	<223> 값
17	snRNA	ncRNA ¹	ncRNA_class ¹	"snRNA"
			note (<223>이 있으면)	<223> 값
18	TATA_signal	regulatory ¹	regulatory_class ¹	"TATA_box"
			note (<223>이 있으면)	<223> 값
19	terminator	regulatory ¹	regulatory_class ¹	"terminator"
			note (<223>이 있으면)	<223> 값
20	3'clip	misc_feature	note	"3'clip" 및 <223> 값
21	5'clip	misc_feature	note	"5'clip" 및 <223> 값
22	-10_signal	regulatory ¹	regulatory_class ¹	"minus_10_signal"
			note (<223>이 있으면)	<223> 값
23	-35_signal	regulatory ¹	regulatory_class ¹	"minus_35_signal"
			note (<223>이 있으면)	<223> 값

¹ ST.26은 특정 ST.25의 특징(예: TATA_signal)이 더 넓은 특징 기호/한정자/값(예: regulatory/regulatory_class/TATA_box)으로 대체될 것을 요구할 수 있다. 이 경우 더 좁은 ST.25 특징에 이전 출원에 대한 우선권이 부여된다. 그러나 보다 광범위한 ST.26 특징 기호/한정자(예: regulatory/regulatory_class)의 전체 범위는 이전 출원에 우선권이 부여되지 않는다.

No.	ST.25 Feature key <221>	ST.26 equivalent		
		Feature key	Qualifier	Qualifier value
24	NON_CONS	This feature relates to a gap of an unknown number of residues in a single sequence, which is prohibited in both ST.25 (paragraph 22) and ST.26 (paragraph 37). Consequently, each region of specifically defined residues that is encompassed by ST.26 paragraph 7 must be included in the sequence listing as a separate sequence and assigned its own sequence identification number. To avoid added/deleted subject matter, each such sequence must be annotated to indicate that it is part of a larger sequence that contains an undefined gap.		
		REGION	note	Description
		Description - as to where and to what the sequence is linked, e.g., this residue is linked N-terminally to a peptide having an N-terminal Gly-Gly and a gap of undefined length.		
25	SIMILAR	REGION	note	"SIMILAR" and <223> value if present
26	THIOETH	CROSSLNK	note	"THIOETH" and <223> value if present
		For further location information guidance, see ST.26 Annex I, CROSSLNK Feature Key Comment		
27	THIOLEST	CROSSLNK	note	"THIOLEST" and <223> value if present
		For further location information guidance, see ST.26 Annex I, CROSSLNK Feature Key Comment		
28	VARSP LIC	Discussed in a Scenario 13 below		
29	ACETYLATION	MOD_RES	note	"ACETYLATION" and <223> value if present
			note	Information required by ST.26 Annex I MOD_RES Feature Key Comment, if possible (without added subject matter)
30	AMIDATION	MOD_RES	note	"AMIDATION" and <223> value if present
			note	Information required by ST.26 Annex I MOD_RES Feature Key Comment, if possible (without added subject matter)
31	BLOCKED	MOD_RES	note	"BLOCKED" and <223> value if present
			note	Information required by ST.26 Annex I MOD_RES Feature Key Comment, if possible (without added subject matter)
32	FORMYLATION	MOD_RES	note	"FORMYLATION" and <223> value if present
			note	Information required by ST.26 Annex I MOD_RES Feature Key Comment, if possible (without added subject matter)
33	GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	MOD_RES	note	"GAMMA-CARBOXYLGLUTAMIC ACID HYDROXYLATION" and <223> value if present
			note	Information required by ST.26 Annex I MOD_RES Feature Key Comment, if possible (without added subject matter)
34	METHYLATION	MOD_RES	note	"METHYLATION" and <223> value if present
			note	Information required by ST.26 Annex I MOD_RES Feature Key Comment, if possible (without added subject matter)
35	PHOSPHORYLATION	MOD_RES	note	"PHOSPHORYLATION" and <223> value if present
			note	Information required by ST.26 Annex I MOD_RES Feature Key Comment, if possible (without added subject matter)
36	PYRROLIDONE CARBOXYLIC ACID	MOD_RES	note	"PYRROLIDONE CARBOXYLIC ACID" and <223> value if present
			note	Information required by ST.26 Annex I MOD_RES Feature Key Comment, if possible (without added subject matter)

No.	ST.25 특징 기호 <221>	ST.26 대응		
		특징 기호	한정자	한정자 값
24	NON_CONS	본 특징은 단일 서열에서 알려지지 않은 수의 잔기 갭(gap)과 관련이 있으며, 이것은 ST.25(22절) 및 ST.26(37절) 둘 다에서 금지되어 있다. 결과적으로, ST.26의 7절에 포함되는 구체적으로 정의된 잔기(residues)의 각 영역은 별도의 서열로 서열목록에 포함되고 독립된 서열식별번호가 할당되어야 한다. 추가/삭제된 주제를 피하려면 이러한 각 서열이 주석을 달아서 정의되지 않은 갭(gap)이 포함된 더 큰 서열의 일부임을 제시하여야 한다.		
		SITE	note	Description
		서열이 연결된 위치와 대상에 대한 발명의 설명: 예: 이 잔기는 N-말단 Gly-Gly 및 정의되지 않은 길이의 갭(gap)을 갖는 펩티드에 N말단 방향으로 연결된다.		
25	SIMILAR	REGION	note	"SIMILAR" 및 <223> 값(있다면)
26	THIOETH	CROSSLNK	note	"THIOETH" 및 <223> 값(있다면)
		자세한 위치 정보 지침은 ST.26 별첨 I, CROSSLNK 특징 기호 참고를 참조하십시오.		
27	THIOLEST	CROSSLNK	note	"THIOLEST" 및 <223> 값(있다면)
		자세한 위치 정보 지침은 ST.26 별첨 I, CROSSLNK 특징 기호 참고를 참조하십시오.		
28	VARSP LIC	아래의 시나리오 13에서 논의		
29	ACETYLATION	MOD_RES	note	"ACETYLATION" 및 <223> 값(있다면)
			note	가능하면, ST.26 별첨 I의 MOD_RES 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
30	AMIDATION	MOD_RES	note	"AMIDATION" 및 <223> 값(있다면)
			note	가능하면, ST.26 별첨 I의 MOD_RES 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
31	BLOCKED	MOD_RES	note	"BLOCKED" 및 <223> 값(있다면)
			note	가능하면, ST.26 별첨 I의 MOD_RES 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
32	FORMYLATION	MOD_RES	note	"FORMYLATION" 및 <223> 값(있다면)
			note	가능하면, ST.26 별첨 I의 MOD_RES 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
33	GAMMA-CARBOXY GLUTAMIC ACID HYDROXY LATION	MOD_RES	note	"GAMMA-CARBOXYLGLUTAMIC ACID HYDROXYLATION" 및 <223> 값(있다면)
			note	가능하면, ST.26 별첨 I의 MOD_RES 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
34	METHYLATION	MOD_RES	note	"METHYLATION" 및 <223> 값(있다면)
			note	가능하면, ST.26 별첨 I의 MOD_RES 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
35	PHOSPHORYLA TION	MOD_RES	note	"PHOSPHORYLATION" 및 <223> 값(있다면)
			note	가능하면, ST.26 별첨 I의 MOD_RES 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
36	PYRROL IDONE CARBO XYLIC ACID	MOD_RES	note	"PYRROLIDONE CARBOXYLIC ACID" 및 <223> 값(있다면)
			note	가능하면, ST.26 별첨 I의 MOD_RES 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)

No.	ST.25 Feature key <221>	ST.26 equivalent		
		Feature key	Qualifier	Qualifier value
37	SULFATATION	MOD_RES	note	"SULFATATION" and <223> value if present
			note	Information required by ST.26 Annex I MOD_RES Feature Key Comment, if possible (w ithout added subject matter)
38	MYRISTATE	LIPID	note	"MYRISTATE" and <223> value if present
			note	Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (w ithout added subject matter)
39	PALMITATE	LIPID	note	"PALMITATE" and <223> value if present
			note	Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (w ithout added subject matter)
40	FARNESYL	LIPID	note	"FARNESYL" and <223> value if present
			note	Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (w ithout added subject matter)
41	GERANYL-GERANYL	LIPID	note	"GERANYL-GERANYL" and <223> value if present
			note	Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (w ithout added subject matter)
42	GPI-ANCHOR	LIPID	note	"GPI-ANCHOR" and <223> value if present
			note	Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (w ithout added subject matter)
43	N-ACYL DIGLYCERIDE	LIPID	note	"N-ACYL DIGLYCERIDE" and <223> value if present
			note	Information required by ST.26 Annex I LIPID Feature Key Comment, if possible (w ithout added subject matter)

Scenario 9

Certain feature keys present in both ST.25 and in ST.26, both for nucleotide sequences and amino acid sequences, have mandatory qualifiers in ST.26, as indicated below. The nucleotide sequence feature key "modified_base" is also present in both ST.25 and ST.26; however, Scenario 7 contains appropriate recommendations. ST.25 did not have any qualifiers, but did have a <223> free text field. When the information contained in an ST.25 <223> field is appropriate as the value for the ST.26 mandatory qualifier, then the information should be included as such. When an ST.25 <223> field has either not been provided or contains information that is not appropriate as the value for the ST.26 mandatory qualifier, then applicants must take care to capture the information contained in the ST.25 feature key/<223> field in a manner compliant with ST.26 without the introduction of added or deleted subject matter.

Nucleotide sequences²

Feature Key	Mandatory Qualifier
5.12 - misc_binding	6.3 - bound_moiety
5.30 - protein_bind	6.3 - bound_moiety

² The numeric references in the table below refer to the Feature key and Qualifier numbers of ST.26, Annex I Controlled Vocabulary.

No.	ST.25 특징 기호 <221>	ST.26 대응		
		특징 기호	한정자	한정자 값
37	SULFATATION	MOD_RES	note	"SULFATATION" 및 <223> 값(있다면)
			note	가능하다면, ST.26 별첨 I의 MOD_RES 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
38	MYRISTATE	LIPID	note	"MYRISTATE" 및 <223> 값(있다면)
			note	가능하다면, ST.26 별첨 I의 LIPID 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
39	PALMITATE	LIPID	note	"PALMITATE" 및 <223> 값(있다면)
			note	가능하다면, ST.26 별첨 I의 LIPID 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
40	FARNESYL	LIPID	note	"FARNESYL" 및 <223> 값(있다면)
			note	가능하다면, ST.26 별첨 I의 LIPID 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
41	GERANYL-GERANYL	LIPID	note	"GERANYL-GERANYL" 및 <223> 값(있다면)
			note	가능하다면, ST.26 별첨 I의 LIPID 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
42	GPI-ANCHOR	LIPID	note	"GPI-ANCHOR" 및 <223> 값(있다면)
			note	가능하다면, ST.26 별첨 I의 LIPID 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)
43	N-ACYL DIGLYCERIDE	LIPID	note	"N-ACYL DIGLYCERIDE" 및 <223> 값(있다면)
			note	가능하다면, ST.26 별첨 I의 LIPID 특징 기호 참고에 의해 필요한 정보(추가된 주제 없이)

시나리오 9

ST.25 및 ST.26 모두에 존재하고, 핵산염기 서열 및 아미노산 서열 모두에 존재하는 특정한 특징 기호는 아래에 제시된 바와 같이 ST.26에서 필수 한정자를 갖는다. 핵산염기 서열 특징 기호 "modified_base"는 또한 ST.25 및 ST.26 모두에 존재하지만, 시나리오 7에는 적절한 권장 사항이 포함되어 있다. ST.25에는 한정자가 없었지만, <223> 자유텍스트 필드가 있었다. ST.26 필수 한정자에 대한 값으로 ST.25 <223> 필드에 포함된 정보가 적절한 경우, 해당 정보도 포함되어야 한다. ST.25 <223> 필드가 제공되지 않았거나, ST.26 필수 한정자의 값으로 적합하지 않은 정보를 포함하는 경우, 출원인은, 주제의 추가나 삭제 없이 ST.26을 준수하는 방식으로 ST.25 특징 기호/<223> 필드에 포함된 정보를 포착하는 것을 고려하여야 한다.

핵산염기 서열²

특징 기호	필수 한정자
5.12 - misc_binding	6.3 - bound_moiety
5.30 - protein_bind	6.3 - bound_moiety

² 아래 표의 숫자 참조는 ST.26, '별첨 I 통제 어휘'의 특징 기호와 한정자 번호를 나타낸다.

Recommendations:

- (a) If the ST.25 <223> field is absent or inappropriate, and the application description disclosed the name of the molecule/complex that may bind to the feature location of the nucleic acid, then that name should be included in the qualifier "bound_moiety".
- (i) Any information contained in the ST.25 <223> field that is inappropriate for inclusion in the qualifier "bound_moiety" should be inserted into an appropriate optional qualifier of the feature key, e.g., "note".
- (b) If the ST.25 <223> field is absent or inappropriate, and the application description did not disclose the name of the molecule/complex that may bind to the feature location of the nucleic acid, then the ST.26 feature key "misc_feature" should be used instead of misc_binding or protein_bind, with the qualifier "note".
- (i) If the ST.25 <223> field was absent, then the value of the qualifier "note" should be the name of the ST.25 feature key;
- (ii) If the ST.25 <223> field contained inappropriate information, then the value of the qualifier "note" should be the name of the ST.25 feature key and the information from the <223> field.

Amino acid sequences³

Feature Key	Mandatory Qualifier
7.2 – BINDING	8.2 – note
7.4 – CARBOHYD	8.2 – note
7.10 – DISULFID	8.2 – note
7.11 – DNA_BIND	8.2 – note
7.12 – DOMAIN	8.2 – note
7.16 – LIPID	8.2 – note
7.17 – METAL	8.2 – note
7.18 – MOD_RES	8.2 – note
7.23 – NP_BIND	8.2 – note
7.29 – SITE	8.2 – note
7.39 – ZN_FING	8.2 – note

Recommendations:

- (a) If the ST.25 <223> field is absent or inappropriate, and the application description disclosed the specific information required in the mandatory qualifier, then that information should be included in the mandatory qualifier "note".
- (i) Any information contained in the ST.25 <223> field that is inappropriate for inclusion in the mandatory qualifier "note" (see feature key definition and comment) should be inserted into a second qualifier "note".
- (b) If the ST.25 <223> field is absent or inappropriate, and the application description did not disclose the specific information required in the mandatory qualifier, then the ST.26 feature key "SITE" (for one amino acid) or "REGION" (for a range of amino acids) should be used instead, with the qualifier "note".
- (i) If the ST.25 <223> field is absent, then the value of the qualifier "note" should be the name of the ST.25 feature key;
- (ii) If the ST.25 <223> field contained inappropriate information, then the value of the qualifier "note" should be the name of the ST.25 feature key and the information from the <223> field.

Scenario 10

Each specific feature key in ST.25 has a <222> field to indicate a feature location; however, ST.25 does not require an indication of the location for most features and the format of the location information is not standardized. Furthermore, ST.25 does not have location operators, e.g., "join". ST.26 has standardized location descriptors and operators and each feature must contain at least one location descriptor. (CDS features are a special case and are discussed below in Scenario 11).

³ The numeric references in the table below refer to the Feature key and Qualifier numbers of ST.26, Annex I Controlled Vocabulary.

권고사항:

(a) 만일 ST.25 <223> 필드가 없거나 부적절하고 발명의 설명에서 핵산의 특징 위치에 결합할 수 있는 분자/복합체의 명칭이 공개된 경우, 해당 명칭은 한정자 "bound_moiety"에 포함되어야 한다.

(i) 한정자 "bound_moiety"에 포함되기에 부적합한 ST.25 <223> 필드에 포함된 모든 정보는 특징 기호의 적절한 선택 한정자(예: "note")에 삽입되어야 한다.

(b) 만일 ST.25 <223> 필드가 없거나 부적절하고 발명의 설명에서 핵산의 특징 위치에 결합할 수 있는 분자/복합체의 명칭이 공개되지 않은 경우, 한정자 "note" 와 함께 ST.26 특징 기호 "misc_feature"가 misc_binding 또는 protein_bind 대신 사용되어야 한다.

(i) 만일 ST.25 <223> 필드가 없는 경우 한정자 "note"의 값은 ST.25 특징 기호의 명칭이어야 한다.

(ii) 만일 ST.25 <223> 필드에 부적절한 정보가 포함된 경우 한정자 "note"의 값은 ST.25 특징 기호의 명칭과 <223> 필드의 정보이어야 한다.

아미노산 서열³

특징 기호	필수 한정자
7.2 – BINDING	8.2 – note
7.4 – CARBOHYD	8.2 – note
7.10 – DISULFID	8.2 – note
7.11 – DNA_BIND	8.2 – note
7.12 – DOMAIN	8.2 – note
7.16 – LIPID	8.2 – note
7.17 – METAL	8.2 – note
7.18 – MOD_RES	8.2 – note
7.23 – NP_BIND	8.2 – note
7.29 – SITE	8.2 – note
7.39 – ZN_FING	8.2 – note

권고사항:

(a) 만일 ST.25 <223> 필드가 없거나 부적절하고, 발명의 설명에 필수 한정자에 필요한 특정 정보를 공개한 경우, 해당 정보는 필수 한정자 "note"에 포함되어야 한다.

(i) 필수 한정자 "note"에 포함하기에 부적합한 ST.25 <223> 필드에 포함된 모든 정보는 두 번째 한정자 "note"에 삽입하여야 한다.

(b) 만일 ST.25 <223> 필드가 없거나 부적절하고, 발명의 설명에 필수 한정자에 필요한 특정 정보가 공개되지 않은 경우, ST.26 특징 기호 "SITE"(1개의 아미노산) 또는 한정자 "noteE"와 함께 "REGION"(아미노산 범위)을 대신에 사용하여야 한다.

(i) 만일 ST.25 <223> 필드가 없으면 한정자 "note"의 값은 ST.25 특징 기호의 명칭이어야 한다.

(ii) 만일 ST.25 <223> 필드에 부적절한 정보가 포함된 경우, 한정자 "note"의 값은 ST.25 특징 기호의 명칭과 <223> 필드의 정보이어야 한다.

시나리오 10

ST.25의 각 특정한 특징 기호에는 특징 위치를 나타내는 <222> 필드가 있다. 그러나 ST.25는 대부분의 특징에 대한 위치 표시를 요구하지 않으며 위치 정보의 형식은 표준화되어 있지 않다. 더욱이, ST.25에는 "join"과 같은 위치 연산자가 없다. ST.26에는 표준화된 위치 설명자와 연산자가 있으며 각 특징에는 하나 이상의 위치 설명자가 있어야 한다(CDS 특징은 특별한 경우이며 아래 시나리오 11에서 설명한다).

³ 아래 표의 숫자 참조는 ST.26, '별첨 I 통제 어휘'의 특징 기호와 한정자 번호를 나타낸다.

Recommendations:

- (a) If the ST.25 sequence listing had a <222> field, direct importation or importation into ST.26 format should not raise any added subject matter consideration;
- (b) If the ST.25 sequence listing did not have a <222> field, but location information was contained in the application description, then direct importation or importation into ST.26 format should not raise any added subject matter consideration;
- (c) If neither the ST.25 sequence listing, nor the application description contained location information, then presumably, the feature applies to the entire sequence. (Indicating a location that is less than the entire sequence without support in the application description would likely constitute added/deleted subject matter.) Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include location information to the extent possible to avoid future issues.

Scenario 11

In ST.25, a coding sequence that encoded a single, contiguous polypeptide but that was interrupted by one or more non-coding sequence(s), e.g., introns, was indicated as multiple separate CDS features, as illustrated below:

```
<220>
<221> CDS
<222> (1)..(571)

<220>
<221> CDS
<222> (639)..(859)
```

In contrast, ST.26 has a join location operator that specifies that the polypeptides encoded by the indicated locations are joined and form a single, contiguous polypeptide. (Note: both ST.25 and ST.26 require that the stop codon be included in the CDS feature location.)

Recommendations:

- (a) If the ST.25 sequence listing or the application description clearly indicated that the polypeptide sequences encoded by the multiple separate CDS features form a single, contiguous polypeptide, then a coding sequence interrupted by an intron in a single CDS feature must be represented with the join location operator, as illustrated below, such that no added subject matter is introduced:
- ```
<INSDFeature_key>CDS</INSDFeature_key>
<INSDFeature_location>join(1..571,639..859)</INSDFeature_location>
```
- (b) If the ST.25 sequence listing or the application description did not indicate that the polypeptide sequences encoded by the two separate CDS features form a single, contiguous polypeptide, then use of the join location operator would likely constitute added subject matter.

*Scenario 12*

ST.25 specifies that feature names must be one from Table 5 or 6. However, U.S. regulations indicated that these feature names were recommended, but not required. Therefore, a sequence in an ST.25 sequence listing (compliant with U.S. regulations) might have a "custom" feature key name with no corresponding feature key in ST.26. It is also possible that no feature name was provided for the <221> field or the <221> field is absent. These scenarios may be handled in a similar manner.

Recommendation:

The "custom" feature key name from ST.25 may be represented in an ST.26 sequence listing with no added subject matter as follows:

권고사항:

- (a) 만일 ST.25 서열목록에 <222> 필드가 있는 경우, 직접 가져오거나 ST.26 형식으로 가져오면 신규사항 추가로 간주하면 안된다.
- (b) 만일 ST.25 서열목록에 <222> 필드가 없지만 위치 정보가 발명의 설명에 포함된 경우, 직접 가져오거나 ST.26 형식으로 가져오면 신규사항 추가로 간주하면 안된다.
- (c) 만일 ST.25 서열목록이나 발명의 설명 중 어디에도 위치 정보가 포함되어 있지 않은 경우, 이 특징은 전체 서열에 적용되는 것으로 간주된다.(발명의 설명에서 지원하지 않고 전체 서열보다 작은 위치를 표시하면 주제가 추가/삭제될 것일 수 있다.) 향후 발생할 수 있는 문제를 피하기 위해, 원본 서열목록(ST.25) 및 출원서가 위치정보를 포함하도록 주의를 기울여야 한다.

**시나리오 11**

ST.25에서는 코딩(coding) 서열이 하나의 연속적인 폴리펩티드로 번역되지만, 인트론(intron) 같은 번역되지 않는 서열에 의해 단절된 경우, 인트론에 의해 중단된 코딩 서열은 아래에 예시된 바와 같이 다수의 개별 CDS 특징으로 표시되었다:

```
<220>
<221> CDS
<222> (1)..(571)

<220>
<221> CDS
<222> (639)..(859)
```

대조적으로, ST.26은 지시된 위치에 의해 번역된 폴리펩티드가 연결되어 단일의 연속적인 폴리펩티드를 형성하도록 특징하는 "join" 위치 연산자를 갖는다.(참고: ST.25 및 ST.26 모두 정지 코돈이 CDS 특징 위치에 포함되어야 한다.)

권고사항:

- (a) 만일 ST.25 서열목록 또는 발명의 설명이 다수의 개별 CDS 특징들에 의해 번역된 폴리펩티드 서열이 단일의 연속 폴리펩티드를 형성함을 명확하게 나타내는 경우, 단일 CDS 특징에서 인트론에 의해 중단된 코딩 서열은 신규사항이 추가되지 않도록 아래에 설명된 바와 같이 join 위치 연산자로 표현하여야 한다.

```
<INSDFeature_key>CDS</INSDFeature_key>
<INSDFeature_location>join(1..571,639..859)</INSDFeature_location>
```

- (b) 만일 ST.25 서열목록 또는 발명의 설명이 2개의 개별 CDS 특징에 의해 번역된 폴리펩티드 서열이 단일의 연속 폴리펩티드를 형성한다는 것을 나타내지 않는다면, join 위치 연산자의 사용은 신규사항 추가로 간주될 수 있다.

**시나리오 12**

ST.25는 특징 명칭이 표 5 또는 6 중 하나이어야 한다고 지정하지만 미국 규정에 따르면 이 특징 명칭은 권장되었지만 필수적인 것은 아니다. 따라서 ST.25 서열목록의 서열(미국 규정 준수)에는 ST.26에 해당 특징 기호가 없는 "custom" 특징 기호 명칭이 있을 수 있다. <221> 필드에 어떤 특징 이름이 제공되지 않았거나 <221> 필드가 없을 수 있는 것도 또한 가능하다. 이러한 시나리오는 비슷한 방식으로 처리될 수 있다.

권고사항:

ST.25의 "custom" 특징 기호 이름은 다음과 같이 신규사항 추가로 간주되지 않고 ST.26 서열목록으로 표현될 수 있다.

| Type | ST.25 Feature Key <221> | Potential ST.26 Equivalent |           |                                                      |
|------|-------------------------|----------------------------|-----------|------------------------------------------------------|
|      |                         | Feature key                | Qualifier | Qualifier value                                      |
| NA   | "Custom" feature key    | misc_feature               | note      | "custom" feature key name and <223> value if present |
| AA   | "Custom" feature key    | SITE or REGION             | note      | "custom" feature key name and <223> value if present |

*Scenario 13*

ST.25 contains a feature key "VARSPPLIC" defined as "description of sequence variants produced by alternative splicing". In ST.26, "VARSPPLIC" has been replaced with the broader feature key VAR\_SEQ defined as "description of sequence variants produced by alternative splicing, alternative promoter usage, alternative initiation and ribosomal frameshifting". Therefore, the ST.26 sequence listing should not use "VAR\_SEQ" as a replacement of "VARSPPLIC" without a further explanation.

Recommendation:

In ST.26 the feature "VAR\_SEQ" should be used with the qualifier "note", whose value should include an explanation of the ST.25 narrower scope, e.g., "sequence variant produced by alternative splicing". Any additional information contained in an accompanying ST.25 <223> field should also be included in the qualifier "note".

*Scenario 14*

If the source of a sequence was artificial, the ST.25 <213> Organism field requires the phrase "Artificial Sequence". In ST.26, the feature key "source" requires the qualifier "organism", whose value must be indicated as "synthetic construct", rather than "Artificial Sequence".

Recommendation:

The value for the ST.26 qualifier "organism" must be indicated as "synthetic construct". To avoid potential deleted subject matter, any explanatory information contained in the required ST.25 <223> field should be included in a qualifier "note" (of the feature key "source").

*Scenario 15*

If the scientific name of the source organism of a sequence is unknown, the ST.25 <213> Organism field requires the term "Unknown". In ST.26, the feature key "source" requires the qualifier "organism", whose value must be indicated as "unidentified", rather than "Unknown".

Recommendation:

The value for the ST.26 qualifier "organism" must be indicated as "unidentified". To avoid potential deleted subject matter, any explanatory information contained in the required ST.25 <223> field should be included in a qualifier "note" (of the feature key "source").

*Scenario 16*

ST.25 allows for the enumeration of amino acids to optionally include negative numbers, counting backwards starting with the amino acid next to number 1, for the amino acids preceding the mature protein, for example pre-sequences, pro-sequences, pre-pro-sequences and signal sequences. ST.26 does not allow for negative numbers in the feature location.

Recommendations:

- (a) If the ST.25 sequence listing had a feature or features represented in a <221> and an accompanying <222> field which contained negative and/or positive numbering, e.g., "PROPEP" and/or "CHAIN", then in the ST.26 sequence listing, the appropriate feature key, e.g., "PROPEP" and/or "CHAIN", should be used. A qualifier "note" may be used with the information in a <223> field, if any, as the qualifier value;
- (b) If the ST.25 sequence listing did not have a feature or features represented in a <221> and accompanying <222> field, but information was contained in the application description regarding the negative and/or positive numbering, then in the ST.26 sequence listing, the appropriate feature key, e.g., "PROPEP" and/or

| Type | ST.25 특징 기호 <221> | 잠정적인 ST.26 대응  |      |                                  |
|------|-------------------|----------------|------|----------------------------------|
|      |                   | 특징 기호          | 한정자  | 한정자 값                            |
| NA   | "Custom" 특징 기호    | misc_feature   | note | "custom" 특징 기호 이름 및 <223> 값(있다면) |
| AA   | "Custom" 특징 기호    | SITE 또는 REGION | note | "custom" 특징 기호 이름 및 <223> 값(있다면) |

**시나리오 13**

ST.25에는 "선택적 스플라이싱(alternative splicing)에 의해 생성된 서열 변이에 대한 설명"으로 정의된 특징 기호 "VARSPPLIC"이 포함되어 있다. "VARSPPLIC"은 ST.26에서 "선택적 스플라이싱(alternative splicing), 선택적 프로모터(alternative promoter) 사용, 선택적 개시 및 리보솜 프레임시프팅(frameshifting)에 의해 생성된 서열 변이에 대한 설명"으로 정의된 더 광범위한 특징 기호 VAR\_SEQ로 대체되었다. 따라서 ST.26 서열목록은 추가 설명 없이 "VAR\_SEQ"를 "VARSPPLIC" 대신에 사용해서는 아니된다.

**권고사항:**

ST.26에서 "VAR\_SEQ" 특징은 한정자 "note"와 함께 사용해야 하며, 그 값은 ST. 25에서처럼 더 좁은 범위로, 예를 들면 "sequence variant produced by alternative splicing"과 같은 설명을 포함하여야 한다. ST.25 <223> 필드에 포함된 추가 정보는 한정자 "note"에도 포함되어야 한다.

**시나리오 14**

서열의 출처가 인공적인 경우 ST.25 <213> Organism 필드에는 "Artificial Sequence"이라는 문구가 필요하다. ST.26에서 특징 기호 "source"에는 한정자 "organism"이 필요하며, 이 값은 "Artificial Sequence"가 아닌 "synthetic construct"로 표시되어야 한다.

**권고사항:**

ST.26 한정자 "organism"의 값은 "synthetic construct"로 표시되어야 한다. 잠재적으로 삭제된 주제를 피하려면 필수 ST.25 <223> 필드에 포함된 설명 정보가 한정자 "note"(특징 기호 "source")에 포함되어야 한다.

**시나리오 15**

서열의 source 생물체의 학명이 알려지지 않은 경우, ST.25 <213> Organism 필드에는 "Unknown"이라는 문구가 필요하다. ST.26에서 특징 기호 "source"에는 한정자 "organism"이 필요하며, 이 값은 "Unknown"이 아닌 "unidentified"로 표시되어야 한다.

**권고사항:**

ST.26 한정자 "organism"의 값은 "unidentified"로 표시되어야 한다. 잠재적으로 삭제된 주제를 피하려면 필수 ST.25 <223> 필드에 포함된 설명 정보가 한정자 "note"(특징 기호 "source")에 포함되어야 한다.

**시나리오 16**

ST.25는 성숙(mature) 단백질 앞의 아미노산, 예를 들면 프리-서열(pre-sequence), 프로-서열(pro-sequence), 프리-프로-서열(pre-pro-sequence) 및 신호 서열(signal sequence)을 위해, 아미노산의 나열이 1번 아미노산 다음으로 거꾸로 세면서 선택적으로 음수를 가지는 것을 허용한다. ST.26은 특징 위치에서 음수를 허용하지 않는다.

**권고사항:**

- (a) 만일 ST.25 서열목록에 음수 및/또는 양수(예: "PROPEP" 및/또는 "CHAIN")가 포함된 <221> 및 <222> 필드에 표현된 특징이 있는 경우, ST.26 서열목록에는 적절한 특징 기호, 예를 들면 "PROPEP" 및/또는 "CHAIN"이 사용되어야 한다. <223> 필드의 정보와 함께 한정자 "note"를 한정자 값으로 사용할 수 있다.
- (b) 만일 ST.25 서열목록에는 <221> 및 <222> 필드에 표현된 특징이 없으나 음수 및/또는 양수와 관련된 정보가 발명의 설명에 포함된 경우, ST.26 서열목록에는 적절한 특징 기호, 예를 들면 "PROPEP" 및/또는 "CHAIN"을 사용하여야 한다. 그렇지 않으면 특징 기호 "REGION"이 사용될 수 있다. 한정자 "note"는 출처서 설명의 정보와 함께 한정자 값으로 사용될 수 있다.

“CHAIN”, should be used. Otherwise, the feature key “REGION” may be used. A qualifier “note” may be used with information in the application description, if any, as the qualifier value;

(c) If neither the ST.25 sequence listing, nor the application description, contains information explaining the negative and/or positive numbering, then to avoid potential deleted subject matter in the ST.26 sequence listing, the “REGION” feature key should be used, where the feature location spans the negatively numbered region of the ST.25 sequence. Also, a qualifier “note” should be used to indicate that the amino acid sequence was negatively numbered in the ST.25 sequence listing of the application to which priority is claimed.

#### Scenario 17

ST.25 provides for publication information in fields <300> to <313>. ST.26 does not provide for inclusion of such information.

#### Recommendation:

The information contained in ST.25 fields <300> to <313> should be inserted into the accompanying application body, if not already contained therein.

#### Scenario 18

ST.25 does not provide a standardized way to indicate that a CDS region of a nucleotide sequence was to be translated using a genetic code table other than the standard genetic code table. In contrast, ST.26 has a “transl\_table” qualifier that can be used with the “CDS” feature key to indicate that the region is to be translated using an alternative genetic code table. If the “transl\_table” qualifier is not used, the use of the standard genetic code table is assumed.

#### Recommendations:

(a) If the ST.25 sequence listing or the application description clearly indicated that a CDS region is to be translated using an alternative genetic code table, then the “transl\_table” qualifier must be used with the appropriate genetic code table number as the qualifier value. Failure to use the “transl\_table” qualifier would likely constitute added subject matter, as the default “Standard Code” table would be assumed. Failure to include, in the ST.26 sequence listing, the alternative genetic code table information from the ST.25 sequence listing or from the application description would likely constitute deleted subject matter.

(b) If the ST.25 sequence listing or the application description did not indicate that a CDS region is to be translated using an alternative genetic code table, then the “transl\_table” qualifier should not be used, or should be used only with the qualifier value “1,” i.e., the Standard Code table. Use of the “transl\_table” qualifier with any qualifier value other than “1” would likely constitute added and deleted subject matter.

#### Scenario 19

ST.25 does not provide a standardized way to indicate the location of a feature, in particular, one contained in a site or region that extends beyond a specified residue or span of residues, e.g., a CDS region of a nucleotide sequence that extends beyond one or both ends of a disclosed sequence. In contrast, the ST.26 feature location descriptor provides a standardized way to indicate the location of such a site or region by using the “<” or “>” symbols. For example, the “CDS” feature location must include the stop codon, even when the stop codon is not included in the disclosed sequence itself, by indicating the location as e.g., 1..>321.

#### Recommendations:

(a) Where the ST.25 sequence listing did not explicitly indicate that the location of a feature extended beyond the sequence, but such a location is either supported by the disclosure or is clear from the sequence itself, e.g., the stop codon of a CDS feature that is not contained in the sequence, then the “<” or “>” symbols may be used in the ST.26 sequence listing without addition of subject matter.

(b) Where the ST.25 sequence listing did not explicitly indicate that the location of a feature extended beyond the sequence, and such a location is neither supported by the disclosure, nor is clear from the sequence itself, then compliance with ST.26, without introduction of added subject matter, may not be possible in this situation. In this case, the priority application and sequence listing are themselves arguably incomplete. In this situation, the location description of the feature in the ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include complete feature information.

(c) ST.25 서열목록이나 발명의 설명 모두 음수 및/또는 양수를 설명하는 정보를 포함하지 않는 경우, ST.26 서열목록에서 잠재적으로 주제가 삭제되는 것을 피하려면 특징 위치는 음수가 ST.25 서열 영역에 걸쳐 있는 "REGION" 특징 기호를 사용하여야 한다. 또한, 한정자 "note"는 아미노산 서열이 우선권을 주장하는 출원의 ST.25 서열목록에서 음수로 표시되었음을 나타내기 위해 사용되어야 한다.

#### 시나리오 17

ST.25는 <300>에서 <313> 필드에 발행 정보(publication information)를 제공한다. ST.26은 그러한 정보를 포함하지 않는다.

#### 권고사항:

ST.25 필드 <300>에서 <313>에 포함된 정보는, 아직 포함되어 있지 않았다면, 반드시 첨부된 출원서 본문에 삽입하여야 한다.

#### 시나리오 18

ST.25는 핵산염기 서열의 CDS 영역이 표준 유전자 코드 표 이외의 유전자 코드 표를 사용하여 번역되었다는 것을 나타내는 표준화된 방법을 제공하지 않는다. 반대로 ST.26에는 "CDS" 특징 기호와 함께 사용할 수 있는 "transl\_table" 한정자가 있어 대체 유전자 코드 표를 사용하여 해당 영역을 번역해야 함을 나타낸다. "transl\_table" 한정자를 사용하지 않으면 표준 유전자 코드 표를 사용하는 것으로 가정한다.

#### 권고사항:

(a) 만일 ST.25 서열목록 또는 발명의 설명에 CDS 영역이 대체 유전자 코드 표를 사용하여 번역되어야 한다고 명확하게 표시된 경우, "transl\_table" 한정자와 한정자 값으로 적절한 유전자 코드 표 번호를 함께 사용하여야 한다. "transl\_table" 한정자를 사용하지 않으면 기본 "표준 코드" 표가 자동으로 고려되므로 신규사항 추가가 될 수 있다. ST.26 서열목록에 ST.25 서열목록 또는 출원서 설명의 대체 유전자 코드 표 정보를 포함시키지 않으면 기술적 구성이 삭제된 것으로 간주될 수 있다.

(b) 만일 ST.25 서열목록 또는 발명의 설명에 CDS 영역이 대체 유전자 코드 표를 사용하여 번역되어야 한다고 표시되어 있지 않은 경우, "transl\_table" 한정자를 사용하지 않든지, 또는 단지 한정자 값 "1"(즉 표준 코드 표)과 함께 사용되어야 한다. "1" 이외의 한정자 값과 함께 "transl\_table" 한정자를 사용하면, 기술적 구성이 추가되거나 삭제된 것으로 간주될 수 있다.

#### 시나리오 19

ST.25는 특징의 위치, 특히 특정 잔기 또는 잔기 범위, 예를 들면, 핵산염기 서열의 CDS 영역을 개시된 서열의 한쪽 또는 양쪽 말단 넘어 연장되는 부위 또는 영역에 포함된 특징의 위치를 나타내는 표준화된 방법을 제공하지 않는다. 반대로 ST.26 특징 위치 설명자는 "<" 또는 ">"기호를 사용하여 이러한 사이트 또는 영역의 위치를 표시하는 표준화된 방법을 제공한다. 예를 들면, "CDS"특징의 위치는 정지 코돈이 개시된 서열 자체에 포함되지 않은 경우에도 그 위치를 예를 들면 1..> 321로 표시함으로써 정지 코돈을 포함하여야 한다.

#### 권고사항:

(a) ST.25 서열목록에 특징의 위치가 서열을 넘어 확장되었음을 명시적으로 나타내지 않았지만, 그러한 위치가 본 발명의 설명에 의해 뒷받침되거나 서열 자체에서 명확하다면(예: 서열에 포함되지 않은 CDS 특징의 정지코돈), 신규사항 추가없이 "<" 또는 ">"기호를 ST.26 서열목록에 사용할 수 있다.

(b) ST.25 서열목록에 특징의 위치가 서열을 넘어 확장되었음을 명시적으로 나타내지 않았으며, 그러한 위치는 본 개시에 의해 뒷받침되지도 않고 서열 자체에서도 명확하지 않은 경우, 신규사항 추가 없이 ST.26을 준수하는 것은 불가능하다. 이 경우, 선출원과 서열목록 자체는 불완전하다. 이 상황에서 ST.26 서열목록의 특징에 대한 위치 설명은 이전 출원에 대하여 우선권을 부여받지 못할 것이다. 완전한 특징 정보를 포함하도록 원본(ST.25) 서열목록과 출원서를 작성하도록 주의하여야 한다.



*Scenario 20*

ST.25 Appendix I requires that where a nucleotide sequence contains both DNA and RNA fragments, the value in <212> shall be "DNA" and the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section; however, the exact nature of the further description is not clear and this requirement is not routinely followed. ST.26, paragraph 55, requires that each DNA and RNA segment (ST.26 uses "segment" rather than "fragment" for internal consistency) of the combined DNA/RNA molecule must be further described with the feature key "misc\_feature", which includes the location of the segment, and the qualifier "note", which indicates whether the segment is DNA or RNA.

Recommendations:

- (a) If the ST.25 sequence listing described the DNA and RNA segments in one or more features using <221> misc\_feature, appropriate locations in <222>, and indications in <223> as to which segments were DNA or RNA, then incorporating that information into ST.26 format, using a misc\_feature for each DNA and RNA segment, should not raise any added subject matter consideration;
- (b) If the ST.25 sequence listing described the DNA and RNA segments in one or more features using a feature key in <221> other than misc\_feature, appropriate locations in <222>, and indications in <223> identifying which segments are DNA or RNA, then incorporating that information into ST.26 format, using a misc\_feature for each DNA and RNA segment and an additional "note" qualifier with the original <221> feature key as the value, should not raise any added or deleted subject matter consideration;
- (c) If the ST.25 sequence listing provides the identity (DNA or RNA) and location of each segment in a <223> field that is not associated with a <221> and <222> field, e.g., the explanation for an Artificial Sequence, then incorporating that information into ST.26 format using a misc\_feature for each DNA and RNA segment, should not raise any added subject matter consideration;
- (d) If the ST.25 sequence listing described the molecule in a feature using a <221> misc\_feature and a <223> noting that the molecule is a combined DNA/RNA molecule, but did not provide location information for each segment, and
  - (i) If the description provided the locations of each DNA and RNA segment, then incorporating that information into ST.26 format using a misc\_feature for each DNA and RNA segment, should not raise any added subject matter consideration;
  - (ii) If the description does not contain the location information of each DNA and RNA segment, then compliance with ST.26, without introduction of added subject matter, may not be possible in this situation. In this case, the priority application and sequence listing are themselves arguably incomplete. In this situation, any location descriptions of the features in the ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include complete feature information.
- (e) If the ST.25 sequence listing described the molecule in a feature using a feature key in <221> other than misc\_feature and a <223> noting that the molecule is a combined DNA/RNA molecule, but did not provide location information for each segment, and
  - (i) If the description provided the locations of each DNA and RNA segment, then incorporating that information into ST.26 format using a misc\_feature for each DNA and RNA segment and an additional "note" qualifier with the original <221> feature key as the value, should not raise any added or deleted subject matter consideration;
  - (ii) If the description does not contain the location information of each DNA and RNA segment, then compliance with ST.26, without introduction of added subject matter, may not be possible in this situation. In this case, the priority application and sequence listing are themselves arguably incomplete. In this situation, any location descriptions of the features in the ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include complete feature information.
- (f) If the ST.25 sequence listing noted that the molecule is a combined DNA/RNA molecule in a <223> field, e.g., the explanation for an Artificial Sequence, but did not provide any feature key or location information of each segment, and

**시나리오 20**

ST.25 부록 I은 핵산염기 서열에 DNA와 RNA 조각이 모두 포함되어 있는 경우, <212> 값은 "DNA"이어야 하고, 결합된 DNA/RNA 분자는 <220>에서 <223> 특징 섹션까지 자세히 설명하여야 한다. 그러나 추가 설명의 정확한 성질은 명확하지 않으며 이 요건은 일반적으로 따르지 않는다. ST.26의 55절은 결합된 DNA/RNA 분자의 DNA 및 RNA 단편 각각(ST.26은 내부 일관성을 위해 "조각(fragment)"이 아닌 "단편(segment)"을 사용함)은 단편의 위치를 포함하는 특징 기호 "misc\_feature"로, 그리고 단편이 DNA인지 RNA인지를 나타내는 한정자 "note"로 추가 설명되도록 요구한다.

**권고사항:**

- (a) 만약 ST.25 서열목록이 <221> misc\_feature, <222>의 적절한 위치 및 <223>의 어느 단편이 DNA 또는 RNA인지에 대한 하나 이상의 특징으로 DNA 및 RNA 단편을 기술한 경우, 각 DNA 및 RNA 단편에 대해, misc\_feature를 사용하여 해당 정보를 ST.26 형식으로 통합하는 것은 신규사항이 추가에 대한 고려 대상이 아니다.
- (b) 만약 ST.25 서열목록에 misc\_feature 이외의 <221>, <222>의 적절한 위치 및 어떤 단편이 DNA 또는 RNA인지 식별한 <223>의 표시를 사용하여 하나 또는 하나 이상의 특징에 있어서 DNA 및 RNA 단편을 기술한 경우, 각 DNA 및 RNA 단편에 대한 misc\_feature와 원래 <221> 특징 기호를 값으로 사용하는 "note" 한정자를 사용하여 해당 정보를 ST.26 형식으로 통합하는 것은 주제의 삭제나 추가에 대한 고려대상이 아니다.
- (c) 만약 ST.25 서열목록이 <221> 및 <222> 필드와 관련되지 않은 <223> 필드에서 각 단편의 동일성(DNA 또는 RNA) 및 위치를 제공하는 경우, 예를 들면 인공서열에 대한 설명을 제공하는 경우, 각 DNA 및 RNA 단편에 대해 misc\_feature를 사용하여 ST.26 형식으로 해당 정보를 통합하는 것은 신규사항에 대한 고려 대상이 아니다.
- (d) 만약 ST.25 서열목록에 <221> misc\_feature와 분자가 결합된 DNA/RNA 분자라는 것을 나타내는 <223>을 사용하여 특징의 분자를 설명하지만, 각 단편에 대한 위치 정보를 제공하지 않은 경우; 그리고,
  - (i) 만약 상기 발명의 설명에 각 DNA 및 RNA 단편의 위치가 제공된 경우, 그러면 각 DNA 및 RNA 단편에 대한 misc\_feature를 사용하여 해당 정보를 ST.26 형식으로 통합하는 것은 신규사항 추가에 대한 고려 대상이 아니다.
  - (ii) 만약 상기 발명의 설명에 각 DNA 및 RNA 단편의 위치 정보가 포함되어 있지 않다면, 이 경우 신규사항 추가 없이 ST.26을 준수하는 것이 불가능하다. 이 경우, 선출원과 서열목록 자체는 불완전해진다. 이런 상황에서는 ST.26 서열목록의 특징에 대한 위치 설명은 선출원에 대항하여 우선권이 인정되지 않는다. 원본(ST.25) 서열목록 및 출원서 개시사항에 완전한 특징 정보가 포함하도록 주의를 기울여야 한다.
- (e) 만약 ST.25 서열목록에 misc\_feature를 제외한 <221> 특징 기호와 분자가 결합된 DNA/RNA 분자라는 것을 나타내는 <223>을 사용하여 특징에서 분자를 설명하지만, 각 단편에 대한 위치 정보를 제공하지 않은 경우; 그리고,
  - (i) 만약 발명의 설명에서 각 DNA 및 RNA 단편의 위치를 제공한 경우, 각 DNA 및 RNA 단편에 대해 misc\_feature 및 원래 <221> 특징 기호가 있는 추가 "note" 한정자를 사용하여 해당 정보를 ST.26 형식으로 통합하는 것은 주제의 추가나 삭제에 대한 고려 대상이 아니다.
  - (ii) 만약 발명의 설명에 각 DNA 및 RNA 단편의 위치 정보가 포함되어 있지 않다면, 이 경우 신규사항 추가 없이 ST.26을 준수하는 것이 불가능하다. 이 경우, 선출원과 서열목록 자체는 불완전하다. 이런 상황에서는 ST.26 서열목록의 특징에 대한 위치 설명은 선출원에 대하여 우선권이 인정되지 않는다. 원본(ST.25) 서열목록 및 출원서 개시사항에 완전한 특징 정보가 포함하도록 주의를 기울여야 한다.
- (f) ST.25 서열목록에서 분자가 <223> 필드에서 결합된 DNA/RNA 분자이지만(예: 인공서열에 대한 설명), 각 단편의 특징 기호 또는 위치 정보를 제공하지 않은 경우; 또한,

- (i) If the description provided the locations of each DNA and RNA segment, then incorporating that information into ST.26 format using a misc\_feature for each DNA and RNA segment, should not raise any added subject matter consideration;
- (ii) If the description does not contain the location information of each DNA and RNA segment, then compliance with ST.26, without introduction of added subject matter, may not be possible in this situation. In this case, the priority application and sequence listing are themselves arguably incomplete. In this situation, any location descriptions of the features in the ST.26 sequence listing will not be afforded priority to the earlier application. Care should be taken to draft the original (ST.25) sequence listing and application disclosure to include complete feature information.

[End of Annex VII and of Standard]

- (i) 만약 발명의 설명이 각 DNA 및 RNA 단편의 위치를 제공한 경우, 그러면 각 DNA 및 RNA 단편에 대해 `misc_feature`를 사용하여 해당 정보를 ST.26 형식으로 통합하는 것은 신규사항을 추가에 대한 고려 대상이 아니다.
- (ii) 만약 발명의 설명에 각 DNA 및 RNA 단편의 위치 정보가 포함되어 있지 않다면, 이 경우 신규사항 추가 없이 ST.26을 준수하는 것이 불가능하다. 이 경우, 선출원과 서열목록 자체는 불완전하다. 이런 상황에서는 ST.26 서열목록의 특징에 대한 위치 설명은 선출원에 대하여 우선권이 인정되지 않는다. 원본(ST.25) 서열목록 및 출원서 개시사항에 완전한 특징 정보가 포함하도록 주의를 기울여야 한다.

[별첨 VII 및 표준의 마지막 장]