

# ■ WIPO 표준 ST.26 고급

웨비나 교육

참고: 이는 후속 모듈이며 참여자가 ST.26 Basics에 이미 참여한 것으로 가정한다.

ST.26 Basics 코스는 [https://www.wipo.int/meetings/en/details.jsp?meeting\\_id=62848](https://www.wipo.int/meetings/en/details.jsp?meeting_id=62848)에서 액세스할 수 있다.

# 오늘 다뤄야 할 것

- 일반적으로 사용되는 특징 키 및 한정자
- 특징 위치 형식
- 한정자 값 형식 및 비영어 언어 한정자 값
- 특수 상황 - DNA의 우라실 및 RNA의 티민; DNA/RNA 하이브리드 분자
- 뉴클레오티드 유사체, D-아미노산, 및 분지된 서열
- 서열 변이체

# 일반적으로 사용되는 특징 키 및 한정자

# 특징 키 및 한정자

- 특징 키는 위치에 의해 확인된 서열의 하나 이상의 잔기를 추가로 설명하기 위해 사용될 수 있다
  - 뉴클레오티드 서열에 대한 특징 키가 부록 I, 섹션 5에 열거되어 있다
  - 뉴클레오티드 서열에 대한 특징 키는 소문자이다; 예를 들어, “misc\_binding”
  - 아미노산 서열에 대한 특징 키가 부록 I, 섹션 7에 열거되어 있다
  - 아미노산 서열에 대한 특징 키는 대문자이다; 예를 들어, “REGION”
  
- 한정자를 사용하여 특징을 추가로 설명할 수 있다
  - 뉴클레오티드 서열에 대한 한정자가 부록 I, 섹션 6에 열거되어 있다
  - 뉴클레오티드 서열에 대한 한정자는 소문자이다; 예를 들어, “allele”
  - 아미노산 서열에 대한 한정자가 부록 I, 섹션 8에 열거되어 있다
  - 아미노산 서열에 대한 한정자는 대문자이다; 예를 들어, “NOTE”

# 특징 키 및 한정자

```

<SequenceData sequenceIDNumber="17">
- <INSDSeq>
  <INSDSeq_length>7</INSDSeq_length>
  <INSDSeq_moltype>AA</INSDSeq_moltype>
  <INSDSeq_division>PAT</INSDSeq_division>
- <INSDSeq_feature-table>
  - <INSDFeature>
    <INSDFeature_key>SOURCE</INSDFeature_key>
    <INSDFeature_location>1..7</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>
- <INSDFeature>
  <INSDFeature_key>VARIANT</INSDFeature_key>
  <INSDFeature_location>1</INSDFeature_location>
- <INSDFeature_qual>
  - <INSDQualifier>
    <INSDQualifier_name>NOTE</INSDQualifier_name>
    <INSDQualifier_value>X can be any amino acid</INSDQualifier_value>
  </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
  <INSDSeq_sequence>XYEKGJL</INSDSeq_sequence>
</INSDSeq>
</SequenceData>

```

# 특징 키 및 한정자

- 특징 키는 모든 서열에 필수인 “source”/“SOURCE” 특징을 제외하고는 선택적이다.
- 각 특징 키는 특징을 추가로 설명하는 데 사용할 수 있는 한정자 목록을 가질 수 있다. 대부분의 한정자는 선택적이거나; 일부 특징 키는 필수 한정자를 갖는다.

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

- “source”/“SOURCE” 특징에는 한정자 “mol\_type”/“MOL\_TYPE” 및 “organism”/“ORGANISM”가 필수이다.

# 특징 키 및 한정자

- 일부 특징 키는 추가 제한사항을 갖는다
  - 유기체 범위; 예를 들어, "C\_region"은 진핵생물로 제한된다
  - 분자 범위; 예를 들어, "D-loop"는 DNA 서열로 제한된다

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

# 특징 키 및 한정자

## 뉴클레오티드 서열: “modified\_base” 특징 키

- 특징 키 "modified\_base" 및 이의 필수 한정자 "mod\_base"는 변형된 뉴클레오티드를 설명하기 위해 사용되어야 한다 (ST.26, 단락 16)
- "변형된 뉴클레오티드"는 다음을 제외한 임의의 뉴클레오티드이다:
  - 데옥시-[a, g, c, 또는 t] 3'-모노포스페이트
  - [a, g, c, 또는 u] 3'-모노포스페이트 (ST.26, 단락 3(f))
- "변형된 뉴클레오티드"는 가능할 때마다 대응하는 변형되지 않은 뉴클레오티드에 의해 표현되어야 한다 (부록 1, 섹션 1, 표 1). 그렇지 않으면, 이는 "n"으로 표현될 수 있다. 예를 들어, "2'-O-메틸시티딘"은 서열 내의 "c"로 표현되어야 한다. "큐오신"은 "n"에 의해 표현되어야 한다. 기호 "n"은 단 하나의 잔기에 해당한다.
- 필수 한정자 'mod\_base'의 값은 부록 1, 섹션 2, 표 2의 값으로부터 선택되어야 한다. "기타"가 값이면, 추가 "참고" 한정자는 변형된 잔기의 완전하고 축약되지 않은 이름을 포함해야 한다.



# 특징 키 및 한정자

뉴클레오티드 서열: “modified\_base” 특징 키

- 예: 위치 15에서 이노신을 갖는 뉴클레오티드 서열

```
<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>
```

- "이노신"은 약어 "i"로 부록 1, 섹션 2, 표 2에 수록되어 있다.

# 특징 키 및 한정자

뉴클레오티드 서열: “modified\_base” 특징 키

- 예: 위치 22에서 잔틴을 갖는 뉴클레오티드 서열

```

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>22</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>

```

- "잔틴"은 부록 I, 섹션 2, 표 2에 나열되어 있지 않으며; 따라서, 필수 "mod\_base" 한정자에 대한 값은 "OTHER"이어야 하며, 추가 "참고" 한정자는 값 "잔틴"과 함께 포함되어야 한다.

# 특징 키 및 한정자

## 뉴클레오티드 서열: "modified\_base" 특징 키

- "modified\_base"는 또한 무염기 부위를 설명하는 데 사용될 수 있다:

```

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>11</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>abasic site</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>

```

- 무염기 부위는 서열에서 "n"으로 표현될 수 있으며, 값 "OTHER"을 갖는 "mod\_base" 한정자 및 값 "무염기 부위"를 갖는 추가 "참고" 한정자와 함께 "modified\_base" 특징 키를 사용하여 추가로 설명될 수 있다.

# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 특징 키 "CDS"는 코딩 서열을 확인하는 데 사용될 수 있다. CDS 특징의 위치는 정지 코돈을 포함해야 한다. (ST.26, pgh. 89)
- "CDS" 특징 키에 대한 필수 한정자가 없다
- 일반적으로 사용되는 한정자는 다음을 포함한다:

“pseudo”

“pseudogene”

“translation”

“transl\_table”

“codon\_start”

“transl\_except”

“protein\_id”

# 특징 키 및 한정자

뉴클레오티드 서열: "CDS" 특징 키

■ "CDS" 특징은 다음 한정자 중 하나만 가질 수 있다:

한정자	설명	값
의사	CDS 특징이 비-기능적이며 번역을 갖지 않지만, 의사유전자가 아니라는 것을 나타낸다	없음
의사유전자	CDS 특징이 의사유전자이며 번역을 갖지 않음을 나타낸다	처리 완료 비처리 단일 대립유전자 공지되지 않음
번역	CDS 번역으로부터 유래된 아미노산 서열을 나타낸다	1-글자 아미노산 약어

# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 코딩 서열에 의해 인코딩되고 "번역" 한정자에 개시된 4개 이상의 구체적으로 정의된 아미노산을 함유하는 아미노산 서열은 별개의 서열로서 서열 목록에 포함되어야 하고, 그 자신의 서열 확인 번호로 지정되어야 한다.
- 아미노산 서열로 지정된 서열 확인 번호는 "CDS" 특징 키 내의 한정자 "protein\_id"의 값으로 제공되어야 한다.
- 아미노산 서열에 대한 "SOURCE" 특징 키의 "ORGANISM" 한정자는 이의 코딩 서열의 것과 동일해야 한다.(ST.26, pgh. 92)

# 특징 키 및 한정자

뉴클레오티드 서열: "CDS" 특징 키

■ 번역된 서열을 변형시킬 수 있는 한정자:

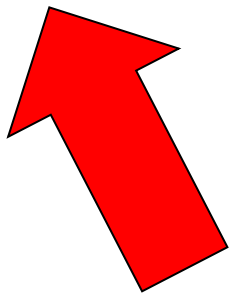
한정자	설명	값
transl_table	CDS를 번역하는 데 사용되는 유전 코드를 나타내고; 내정값은 "1-표준 코드"이다	부록 I, 섹션 9의 번역 표에 해당하는 숫자
transl_except	"transl_table"에 정의된 유전 코드에 따르지 않는 코돈의 번역을 나타낸다	(pos:<location>, aa:<amino_acid>)
codon_start	첫 번째 염기에 대한 CDS의 판독 프레임을 나타낸다	1, 2, 또는 3

# 특징 키 및 한정자

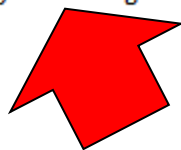
## 뉴클레오티드 서열: "CDS" 특징 키

- 예 - SEQ ID NO:1은 효모 미토콘드리아 유전자로부터의 코딩 서열의 단편이다 (유전 코드 표 "3-효모 미토콘드리아 코드"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```



서열은 부분 코돈으로 시작한  
다



위치 30 내지 32에서의 코돈  
은 셀레노시스테인을 인코딩  
한다 (Sec)

- 이러한 서열을 정확하게 나타내기 위해 CDS 특징에 어떠한 정보가 포함되어야 하는가?



# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 예 - SEQ ID NO:1은 효모 미토콘드리아 유전자로부터의 코딩 서열의 단편이다 (유전 코드 표 "3-효모 미토콘드리아 코드"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- 서열은 다음과 같은 서열 목록에 표시된다:

```
tggataatga agaagttaac gaagaatgta tgagattatt ttcaagaac gctcgtcatc taacatcaag gttgacataa
```

# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 예 - SEQ ID NO:1은 효모 미토콘드리아 유전자로부터의 코딩 서열의 단편이다 (유전 코드 표 "3-효모 미토콘드리아 코드"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- "CDS" 특징 키
- 특징 위치: <1..80

위치는 위치 78 내지 80에  
서 정지 코돈을 포함한다

“<”는 코딩 영역이 위치 1  
이전에 시작됨을 나타낸다

# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 예 - SEQ ID NO:1은 효모 미토콘드리아 유전자로부터의 코딩 서열의 단편이다 (유전 코드 표 "3-효모 미토콘드리아 코드"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- "CDS" 특징 키
- 특징 위치: <1..80
- 값 "3"을 갖는 "codon\_start" 한정자 ←

3의 "codon\_start" 한정자 값은 첫 번째 전체 코돈이 위치 내의 세 번째 위치에서 시작함을 나타낸다

# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 예 - SEQ ID NO:1은 효모 미토콘드리아 유전자로부터의 코딩 서열의 단편이다 (유전 코드 표 "3-효모 미토콘드리아 코드"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- "CDS" 특징 키
- 특징 위치: <1..80
- 값 "3"을 갖는 "codon\_start" 한정자
- 값 "3"을 갖는 "transl\_table" 한정자

부록 1, 섹션 9의 유전 코드 표는 "transl\_table" 한정자의 값을 결정하는 데 사용된다.

# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 예 - SEQ ID NO:1은 효모 미토콘드리아 유전자로부터의 코딩 서열의 단편이다 (유전 코드 표 "3-효모 미토콘드리아 코드"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa 80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- "CDS" 특징 키
- 특징 위치: <1..80
- 값 "3"을 갖는 "codon\_start" 한정자
- 값 "3"을 갖는 "transl\_table" 한정자
- 값 "(pos:30..32,aa:Sec)"을 갖는 "transl\_except" 한정자

“transl\_except” 한정자는 아미노산 셀레노시스테인에 넣을 것이다

# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 예 - SEQ ID NO:1은 효모 미토콘드리아 유전자로부터의 코딩 서열의 단편이다 (유전 코드 표 "3-효모 미토콘드리아 코드"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- "CDS" 특징 키
- 특징 위치: <1..80
- 값 "3"을 갖는 "codon\_start" 한정자
- 값 "3"을 갖는 "transl\_table" 한정자
- 값 "(pos:30..32,aa:Sec)"을 갖는 "transl\_except" 한정자
- 값을 갖는 "번역" 한정자 "**DNEEVNEECURLFFKNARHTTSSLT** "

정지 코돈은 번역 한정자에  
제시되지 않는다!

# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 예 - SEQ ID NO:1은 효모 미토콘드리아 유전자로부터의 코딩 서열의 단편이다 (유전 코드 표 "3-효모 미토콘드리아 코드"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- "CDS" 특징 키
- 특징 위치: <1..80
- 값 "3"을 갖는 "codon\_start" 한정자
- 값 "3"을 갖는 "transl\_table" 한정자
- 값 "(pos:30..32,aa:Sec)"을 갖는 "transl\_except" 한정자
- 값 "DNEEVNEECURLFFKNARHTTSRLT"를 갖는 "번역" 한정자
- 번역을 위한 별도의 단백질 서열

# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- 예 - SEQ ID NO:1은 효모 미토콘드리아 유전자로부터의 코딩 서열의 단편이다 (유전 코드 표 "3-효모 미토콘드리아 코드"):

```
...tg gat aat gaa gaa gtt aac gaa gaa tgt atg aga tta ttt ttc aag aac gct cgt cat cta aca tca agg ttg aca taa      80
... Asp Asn Glu Glu Val Asn Glu Glu Cys Sec Arg Leu Phe Phe Lys Asn Ala Arg His Thr Thr Ser Arg Leu Thr *
```

- "CDS" 특징 키
- 특징 위치: <1..80
- 값 "3"을 갖는 "codon\_start" 한정자
- 값 "3"을 갖는 "transl\_table" 한정자
- 값 "(pos:30..32,aa:Sec)"을 갖는 "transl\_except" 한정자
- 값 "DNEEVNEECURLFFKNARHTTSRLT"를 갖는 "번역" 한정자
- 번역을 위한 별도의 단백질 서열
- 번역된 단백질의 SEQ ID 번호를 갖는 "protein\_id" 한정자



# 특징 키 및 한정자

## 뉴클레오티드 서열: "CDS" 특징 키

- "CDS" 특징 위치는 서열의 불연속 세그먼트를 단일 코딩 영역으로 연결하기 위해 "연결" 위치 연산자를 사용할 수 있다

연결(위치1,위치2)

- "CDS" 특징 위치는 특징이 위치 설명자에 의해 지정된 서열에 상보적인 가닥 상에 위치되는 것을 나타내기 위해 "보완" 연산자를 사용할 수 있다

보완(위치)

# 특징 키 및 한정자

아미노산 서열: 일반적으로 사용되는 특징 키

특징 기호	설명	필수 한정자?
SITE	흥미로운 단일 아미노산 부위를 나타낸다	필수 참고 한정자는 부위를 설명해야 한다
REGION	관심 영역을 나타낸다	없음; 참고는 선택적이다
BINDING	화학기에 대한 결합 부위를 나타낸다	필수 참고 한정자는 화학기의 명칭을 함유해야 한다
UNSURE	서열에서의 불확실성 영역을 설명한다	없음; 참고는 선택적이다

# 특징 키 및 한정자

## 아미노산 서열: 변형된 아미노산

- "변형된 아미노산"은 다음을 제외한 임의의 아미노산이다:

L-알라닌	L-아르기닌	L-아스파라긴	L-아스파르트산
L-시스테인	L-글루타민	L-글루탐산	L-글리신
L-히스티딘	L-이소류신	L-류신	L-리신
L-메티오닌	L-페닐알라닌	L-프롤린	L-피롤리신
L-세린	L-셀레노시스테인	L-트레오닌	L-트립토판
	L-티로신	L-발린	

(ST.26, pgh. 3(e))

- "변형된 아미노산"은 가능할 때마다 대응하는 변형되지 않은 아미노산에 의해 표시되어야 한다. 그렇지 않으면, 이는 "X"로 표시될 수 있다. 예를 들어, "하이드록시리신"은 서열에서 "K"로 표시되어야 한다. "오르니틴"은 "X"로 표시되어야 한다. (ST.26, pgh. 29).
- 기호 "X"는 단지 하나의 잔기에 해당한다.

# 특징 키 및 한정자

## 아미노산 서열: 변형된 아미노산

- 여러 특징 키는 "변형된 아미노산"을 나타내는 데 사용될 수 있다.

특징 기호	설명
SITE	번역 후 변형되지 않은 아미노산을 나타낸다
MOD_RES	번역 후 변형된 아미노산을 나타낸다
CARBOHYD	당화된 아미노산을 나타낸다
LIPID	아미노산에 대한 지질 모이어티의 공유 결합을 나타낸다

- 필수 참고 한정자는 변형을 설명하기 위한 값과 함께 상기 특징 키 각각과 함께 포함되어야 한다.













# 특징 위치 형식

# 위치 형식

## 모든 분자 유형에 대한 위치 설명자

- 위치 설명자는 서열에서 특징의 위치를 확인하는 데 사용된다
- ST.26은 위치 설명자의 형식에 대한 필수 요구사항을 갖는다
- 다음 위치 설명자 형식은 뉴클레오타이드 및 아미노산 서열에 사용될 수 있다:

Location descriptor type	Syntax	Description
Single residue number	x	Points to a single residue in the sequence.
Residue numbers delimitating 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.

WIPO 표준 ST.26, 단락 66(a)

# 위치 형식

## 모든 분자 유형에 대한 위치 설명자

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.

WIPO 표준 ST.26, 단락 70(a)

# 위치 형식

## 뉴클레오티드 서열에 대해 독점적인 위치 설명자

- 다음 위치 설명자 형식은 DNA 및 RNA 서열에만 사용할 수 있다:

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 <sup>56</sup> ), or, for circular nucleotides, $x^1$ , where “x” is the full length of the molecule, i.e. 1000 <sup>1</sup> for circular molecule with length 1000.

WIPO 표준 ST.26, 단락 66(b)

# 위치 형식

## 뉴클레오티드 서열에 대해 독점적인 위치 설명자

- $x^y$  위치 형식을 사용하는 경우,  $x$  및  $y$ 는 인접한 잔기여야 한다

Location example	Description
123 <sup>124</sup>	Points to a site between residues 123 and 124.
867 <sup>1</sup>	In a circular molecule with 867 residues, points to a site between the residue indicated as position 1 and the residue indicated as position 867

WIPO 표준 ST. 26, 단락 70(b)

# 위치 형식

## 위치 설명자 – 아미노산 서열에 대한 특별한 경우

- 위치 설명자 x..y는 “CROSSLNK” 또는 “DISULFID” 특징 키와 함께 사용되는 경우 2개의 표시된 잔기 사이의 사슬간 가교결합을 나타낸다

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”.

WIPO 표준 ST.26, 단락 66(c)

# 위치 형식

## 위치 설명자 – 아미노산 서열에 대한 특수한 경우

```

<SequenceData sequenceIDNumber="4">
  <INSDSeq>
    <INSDSeq_length>81</INSDSeq_length>
    <INSDSeq_moltype>AA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>SOURCE</INSDFeature_key>
        <INSDFeature_location>1..81</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>MOL_TYPE</INSDQualifier_name>
            <INSDQualifier_value>protein</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q25">
            <INSDQualifier_name>ORGANISM</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
      <INSDFeature>
        <INSDFeature_key>DISULFID</INSDFeature_key>
        <INSDFeature_location>30..50</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier id="q22">
            <INSDQualifier_name>NOTE</INSDQualifier_name>
            <INSDQualifier_value>disulfide bond</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>QKKMIQFFKITHRYYYDIIIEHLCAKYDMNSVISNALFAKLNLMQYTDGLSTHEKIIINTSNPLTFSIVISLQRCVINLGGST</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```



# 위치 형식

## 뉴클레오티드 서열에 대한 위치 연산자

- 3개의 위치 연산자가 DNA 및 RNA 서열에서 사용할 수 있다: "연결", "순서" 및 "보완"
- 연결 위치 연산자의 사용은 위치 설명자에 의해 설명된 뉴클레오티드 잔기가 생물학적 과정에 의해 물리적으로 접촉되는 것을 의미한다 (ST.26, 단락 68)
- "연결" 및 "순서"를 사용하는 위치는 적어도 두 개의 심표로 구분된 위치 설명자들을 가져야 한다
- "보완"은 "연결" 또는 "순서와 조합하여 사용될 수 있다

Location syntax	Location description
<code>join(location, location, ..., location)</code>	The indicated locations are joined (placed end-to-end) to form one contiguous sequence.
<code>order(location, location, ..., location)</code>	The elements are found in the specified order but nothing is implied about whether joining those elements is reasonable.
<code>complement(location)</code>	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.

# 위치 형식

## 뉴클레오티드 서열에 대한 위치 연산자

Location example	Description
<code>join(12..78,134..202)</code>	Indicates that regions 12 to 78 and 134 to 202 should be joined to form one contiguous sequence.
<code>order(15..228,341..502)</code>	Indicates that regions 15 to 228 and 341 to 502 are present in the specified order
<code>complement(34..126)</code>	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).
<code>complement(join(2691..4571,4918..5163))</code>	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).
<code>join(complement(4918..5163),complement(2691..4571))</code>	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).

WIPO 표준 ST. 26, 단락 70(b)

# 궁금한 점

# 한정자 값 형식

# 한정자 값

- 한정자 추가 정의 특징

# 한정자 값

- 한정자 추가 정의 특징
- 한정자는 한정자 명칭 및 종종 한정자 값으로 구성된다

```
<INSDQualifier>  
  <INSDQualifier_name>transl_table</INSDQualifier_name>  
  <INSDQualifier_value>12</INSDQualifier_value>  
</INSDQualifier>
```

# 한정자 값

- 한정자 추가 정의 특징
- 한정자는 한정자 명칭 및 종종 한정자 값으로 구성된다

```
<INSDQualifier>  
  <INSDQualifier_name>transl_table</INSDQualifier_name>  
  <INSDQualifier_value>12</INSDQualifier_value>  
</INSDQualifier>
```

- 각각의 특징 키는 상기 특징에 허용되는 한정자 목록을 가질 것이다. 일부 특징 키는 필수 한정자를 갖는다.

# 한정자 값

- 한정자 추가 정의 특징
- 한정자는 한정자 명칭 및 종종 한정자 값으로 구성된다

```
<INSDQualifier>
  <INSDQualifier_name>transl_table</INSDQualifier_name>
  <INSDQualifier_value>12</INSDQualifier_value>
</INSDQualifier>
```

- 각각의 특징 키는 상기 특징에 허용되는 한정자 목록을 가질 것이다. 일부 특징 키는 필수 한정자를 갖는다.
- 특징 키 및 이들의 허용되는 한정자는 ST.26, 부록 I, 섹션 5 (뉴클레오티드 서열) 및 섹션 7 (아미노산 서열)에 열거되어 있다.



# 한정자 값

- 예 - 특징 키 “misc\_binding”는 하나의 필수 한정자, “bound\_moiety,” 및 6개의 선택적 한정자를 갖는다:

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

(ST.26, 부록 1, 섹션 5.12)

# 한정자 값

- 한정자 추가 정의 특징
- 한정자는 한정자 명칭 및 종종 한정자 값으로 구성된다

```
<INSDQualifier>
  <INSDQualifier_name>transl_table</INSDQualifier_name>
  <INSDQualifier_value>12</INSDQualifier_value>
</INSDQualifier>
```

- 각각의 특징 키는 상기 특징에 허용되는 한정자 목록을 가질 것이다. 일부 특징 키는 필수 한정자를 갖는다.
- 특징 키 및 이들의 허용되는 한정자는 ST.26, 부록 I, 섹션 5 (뉴클레오티드 서열) 및 섹션 7 (아미노산 서열)에 나열되어 있다.
- 설명과 함께 뉴클레오티드 서열에 대한 한정자는 ST.26, 부록 I, 섹션 6 (뉴클레오티드 서열) 및 섹션 8 (아미노산 서열)에 열거되어 있다.

# 한정자 값

---

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.

---

6.22. Qualifier	gene
Definition	symbol of the gene corresponding to a sequence region
Mandatory value format	free text
Example	<INSDQualifier_value>ilvE</INSDQualifier_value>
Comment	Use gene qualifier to provide the gene symbol; use standard_name qualifier to provide the full gene name.

(ST.26, 부록 I, 섹션 6.3 및 6.22)

# 한정자 값 형식 유형

## ■ 한정자 값은 여러 형식 유형을 갖는다:

1. 미리 정의된 값 선택을 갖는 한정자;
2. 정의된 값 형식을 갖는 한정자;
3. 값이 서열인 한정자;
4. 값을 갖지 않는 한정자;
5. "자유 텍스트" 값을 갖는 한정자

- "자유 텍스트" 한정자 값의 서브셋은 “언어 종속”으로 분류된다

# 한정자 값

## 형식 유형 – 미리 정의된 값 선택

- 미리 결정된 값 선택을 갖는 한정자
- 예:

“codon\_start” – 값은 “1”, “2”, 또는 “3”일 수 있다.

---

6.9. 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>

# 한정자 값

## 형식 유형 – 미리 정의된 값 선택

- 예: “rpt\_type”은 제한된 세트의 값 선택을 갖는다.

6.60. 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;

# 한정자 값

## 형식 유형 – 미리 정의된 값 선택

- 미리 정의된 값 선택을 갖는 한정자의 경우, WIPO 서열은 미리 채워진 드롭-다운 목록에 허용되는 모든 값을 표시할 것이다:

▼ QUALIFIERS

Add qualifier

Qualifier Name *	Qualifier Value
rpt_type	<ul style="list-style-type: none"> <li>tandem</li> <li>direct</li> <li>inverted</li> <li>flanking</li> <li>nested</li> <li>terminal</li> <li><b>dispersed</b></li> <li>long_terminal_repeat</li> <li>non_ltr_retrotransposon_polymeric_tract</li> <li>centromeric_repeat</li> <li>telomeric_repeat</li> <li>x_element_combinatorial_repeat</li> <li>y_prime_element</li> <li>other</li> </ul>

▼ SEQUENCE

tgccgatatt	atccataagc	atgggctct	ggaagtgccg	ctgacggccc	agaagggcgt
cttactgtta	caagaagagt	atcatgacgc	gcatcatagc	cacttagacc	cgcacttatg
gttgccgocg	caaaatatca	ttgcocctgt	gcaaggattg	gataaacata	ctgctgagtt

180

# 한정자 값

## 형식 유형 – 정의된 값 형식

- 정의된 값 형식을 갖는 한정자

- 예: "항코돈" – 값은 형식으로 존재해야 한다

"(pos:<location>, aa:<amino\_acid>, seq:<text>)"

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>



# 한정자 값

## 형식 유형 - 서열

- 값이 서열인 한정자
- 예: "번역" – 값은 하나의 문자 아미노산 약어를 사용하는 서열이어야 한다

6.79. 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>MASTFPPWYRGCSTPSLKGLIMCTW</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.

# 한정자 값

## 형식 유형 - 서열

- 한정자 "대체" – 값은 단일 뉴클레오티드 잔기이거나, 일련의 잔기이거나, 비어 있을 수 있다
- "대체"에 대한 빈 값은 해당 특징에 표시된 잔기의 결실을 의미한다

---

6.57. 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	<pre>&lt;INSDQualifier_value&gt;a&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;&lt;/INSDQualifier_value&gt; - for a deletion</pre>

# 한정자 값

## 형식 유형 - 서열

- 한정자 "대체" – 값은 단일 뉴클레오티드 잔기이거나, 일련의 잔기이거나, 비어 있을 수 있다.
- "대체"에 대한 빈 값은 해당 특징에 표시된 잔기의 결실을 의미한다

6.57. 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	<pre>&lt;INSDQualifier value&gt;a&lt;/INSDQualifier value&gt; &lt;INSDQualifier_value&gt;&lt;/INSDQualifier_value&gt; - for a deletion</pre>

"비어 있는 값"

# 한정자 값

## 형식 유형 - 값 없음

- 값이 없는 한정자

- 예:

“environmental\_sample”    “germline”    “macronuclear”    “proviral”

6.51. 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

- WIPO 서열은 값이 없는 한정자에 대한 값의 추가를 허용하지 않을 것이다
- 이들 한정자는 비어 있는 “INSDQualifier\_value” 요소를 갖지 않아야 한다

# 한정자 값

## 형식 유형 – “Free Text”

- 많은 한정자는 “자유 텍스트” 값 형식을 갖는다.
- ST.26, 단락 3(n)은 "자유 텍스트를" ... *특정 한정자에 대한 값 형식의 유형으로 정의하며, 이는 설명 텍스트 구문 또는 다른 지정된 형식 (부록 I에 표시됨)의 형태로 제시된다.*
- 자유 텍스트 한정자 값은 1000자로 제한된다 (ST.26, pgh. 86)
- "자유 텍스트" 값 형식을 갖는 한정자의 서브셋은 "언어 종속적"
- "언어 종속 "자유 텍스트"는 국가, 지역 또는 국제 절차에 대한 번역을 필요로 할 수 있다." ST.26, 단락 3(o)

# 한정자 값

## 형식 유형 – “자유 텍스트”

- "언어 종속 자유 텍스트" 값을 갖는 한정자는 번역을 필요로 할 수 있다
- 자유 텍스트 값 형식을 갖는 한정자가 언어 종속적인지의 여부를 어떻게 결정할 수 있는가?
  - ST.26, 부록 I, 섹션 6, 표 5는 언어 종속 자유 텍스트 값 형식을 갖는 모든 뉴클레오티드 서열 한정자를 열거한다
  - ST.26, 부록 I, 섹션 8, 표 6은 언어 종속 자유 텍스트 값 형식을 갖는 모든 아미노산 서열 한정자를 열거한다
  - 한정자 설명에서 "필수 값 형식"을 참조한다

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>

# 한정자 값

## 형식 유형 – “자유 텍스트”

6.20. Qualifier	frequency
Definition	frequency of the occurrence of a feature
Mandatory value format	<b>free text</b> representing the proportion of a population carrying the feature expressed as a fraction
Example	<pre>&lt;INSDQualifier_value&gt;23/108&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;1 in 12&lt;/INSDQualifier_value&gt; &lt;INSDQualifier_value&gt;0.85&lt;/INSDQualifier_value&gt;</pre>
6.21. Qualifier	function
Definition	function attributed to a sequence
Mandatory value format	<b>free text</b> Language-dependent: this value may require translation for National/Regional procedures
Example	<pre>&lt;INSDQualifier_value&gt;essential for recognition of cofactor &lt;/INSDQualifier_value&gt;</pre>
Comment	The function qualifier is used when the gene name and/or product name do not convey the function attributable to a sequence.

# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

- "언어 종속 자유 텍스트" 한정자 값은 서열 목록 XML에서 2개의 언어로 제공될 수 있다: 영어 및 다른 하나의 비-영어 언어 (ST.26, 단락 87)



# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

- "언어 종속 자유 텍스트" 한정자 값은 서열 목록 XML에서 2개의 언어로 제공될 수 있다: 영어 및 다른 하나의 비-영어 언어 (ST.26, 단락 87)
- 영어의 언어 종속 한정자 값은 INSDQualifier\_value 요소에 제공되어야 한다

# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

- "언어 종속 자유 텍스트" 한정자 값은 서열 목록 XML에서 2개의 언어로 제공될 수 있다: 영어 및 다른 하나의 비-영어 언어 (ST.26, 단락 87)
- 영어의 언어 종속 한정자 값은 INSDQualifier\_value 요소에 제공되어야 한다
- 영어 이외의 임의의 언어의 언어 종속 한정자 값은 NonEnglishQualifier\_value 요소에 제공되어야 한다

# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

- "언어 종속 자유 텍스트" 한정자 값은 서열 목록 XML에서 2개의 언어로 제공될 수 있다: 영어 및 다른 하나의 비-영어 언어 (ST.26, 단락 87)
- 영어의 언어 종속 한정자 값은 INSDQualifier\_value 요소에 제공되어야 한다
- 영어 이외의 임의의 언어의 언어 종속 한정자 값은 NonEnglishQualifier\_value 요소에 제공되어야 한다
- 언어 종속 자유 텍스트 값 형식을 갖는 한정자에 대해서만 NonEnglishQualifier\_value 요소가 허용된다 (ST.26, 단락 87(b))

# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

```

<SequenceData sequenceIDNumber="2">
  <INSDSeq>
    <INSDSeq_length>29</INSDSeq_length>
    <INSDSeq_moltype>AA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>SOURCE</INSDFeature_key>
        <INSDFeature_location>1..29</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 id="q1">
            <INSDQualifier_name>NOTE</INSDQualifier_name>
            <INSDQualifier_value>Synthetic peptide antigen fragment</INSDQualifier_value>
            <NonEnglishQualifier_value>Synthetisches Peptidantigenfragment</NonEnglishQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>GSLSDVRKDVVEKRIDKALEAFKNKMDKEK</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="3">

```

# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

- 서열 목록 XML이 비-영어 한정자 값을 함유하는 경우:

1. 서열 목록 XML의 루트 요소는 적절한 2-문자 언어 코드 약어가 있는 “nonEnglishFreeTextLanguageCode” 속성을 함유해야 한다 (ST.26 단락 43 및 87(b));

```
<ST26SequenceListing dtdVersion="V1_3" fileName="st26-annex-iii-sequence-listing-specimen.xml" softwareName="WIPO  
Sequence" softwareVersion="1.0" productionDate="2022-01-01" originalFreeTextLanguageCode="de"  
nonEnglishFreeTextLanguageCode="de">
```

# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

### ■ 서열 목록 XML이 비-영어 한정자 값을 함유하는 경우:

1. 서열 목록 XML의 루트 요소는 적절한 2-문자 언어 코드 약어가 있는 “nonEnglishFreeTextLanguageCode 속성을 함유해야 한다 (ST.26 단락 43 및 87(b));
2. 서열 목록의 모든 언어 종속 한정자는 “nonEnglishFreeTextLanguageCode” 속성에 표시된 언어로 된 값을 가져야 한다;

# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

### ■ 서열 목록 XML이 비-영어 한정자 값을 함유하는 경우:

1. 서열 목록 XML의 루트 요소는 적절한 2-문자 언어 코드 약어가 있는 “nonEnglishFreeTextLanguageCode 속성을 함유해야 한다 (ST.26 단락 43 및 87(b));
2. 서열 목록의 모든 언어 종속 한정자는 “nonEnglishFreeTextLanguageCode” 속성에 표시된 언어로 된 값을 가져야 한다;
3. NonEnglishQualifier\_value 및 INSDQualifier\_value 둘 모두가 단일 한정자에 존재하는 경우, 두 요소를 함유한 정보는 동일해야 한다 (ST.26 단락 87(c)).

# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

- INSDQualifier “id” 속성 – 무엇인가?
- ST.26 단락 87(d)는 “언어 종속 한정자의 경우, INSDQualifier 요소는 선택적 속성 ID를 포함할 수 있다. 이러한 속성의 값은 “q” 형식이어야 하며 그 뒤에 양의 정수, 예를 들어, “q23”이 있어야 하며, 하나의 INSDQualifier 요소에 고유해야 하며, 즉, 속성 값은 서열 목록 파일에서 한 번만 사용되어야 한다고 명시한다.

```
<INSDQualifier id="q2">
  <INSDQualifier_name>note</INSDQualifier_name>
  <INSDQualifier_value>common name: tomato</INSDQualifier_value>
  <NonEnglishQualifier_value>gemeinsamen Namen: Tomate</NonEnglishQualifier_value>
</INSDQualifier>
```



# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

- INSDQualifier “id” 속성:
  - WIPO 서열에 의해 XLIFF 파일로의 내보내기를 위해 번역을 요구할 수 있는 한정자 값을 고유하게 확인한다;
  - 선택적;
  - 언어 종속 한정자에 대해서만 허용됨;
  - 서열 목록에서 고유해야 한다;
  - WIPO 서열에 의해 생성된 XML에 자동으로 추가된다.

# 한정자 값

## 형식 유형 – “언어 종속 자유 텍스트”

### ■ “originalFreeTextLanguageCode” 속성

```
<ST26SequenceListing dtdVersion="V1_3" fileName="st26-annex-iii-sequence-listing-specimen.xml" softwareName="WIPO  
Sequence" softwareVersion="1.0" productionDate="2022-01-01" originalFreeTextLanguageCode="de"  
nonEnglishFreeTextLanguageCode="de">
```

- ST.26 단락 43은 “originalFreeTextLanguageCode” 속성을 “언어 종속 자유 텍스트 한정자가 준비된 단일 원본 언어에 대한 언어 코드...”로 정의한다.
- 이러한 속성은 선택적이다

특수 상황:  
DNA에서 DNA/RNA 하이브리  
드 분자 우라실 및  
RNA에서 티민

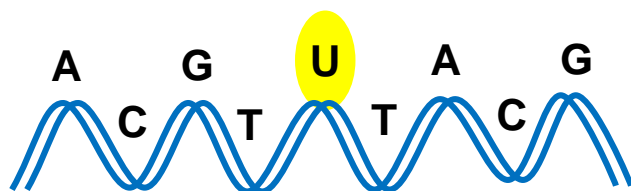
# 특수 상황

- 리마인더: 우라실의 "u" 기호는 ST.26 에서 허용되지 않는다
- DNA에서, "t"는 티민이다
- RNA에서, "t"는 우라실이다
- 고려할 두 가지 시나리오:
  1. 우라실 핵염기를 갖는 DNA 분자 또는 티민 핵염기를 갖는 RNA 분자;
  2. DNA/RNA 하이브리드 분자

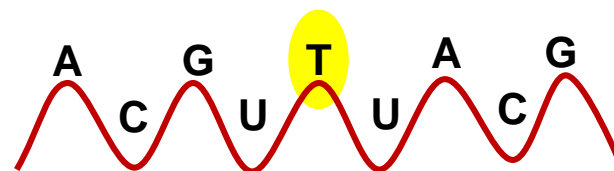
# 특수 상황

## DNA에서 우라실 및 RNA에서 티민

- 서열이 우라실 핵염기를 갖는 DNA 골격 또는 티민 핵염기를 갖는 RNA 골격을 갖는 경우 -은 "변형된 뉴클레오티드"로 설명된다



DNA



RNA

### ST.26 단락 14적용:

"14. 기호 "t"는 DNA에서 티민 및 RNA에서 우라실로 해석될 것이다. DNA의 우라실 또는 RNA의 티민은 **변형된 뉴클레오티드**로 간주되며, 단락 19에 의해 제공된 바와 같은 특징 표에 추가로 설명되어야 한다."

# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 RNA 서열을 개시한다:

5' -cgucccacgugtccgaggua-3'

- 위치 12의 "티민"을 참고한다. 이 잔기는 변형된 뉴클레오티드로 주석을 달아야 한다.
- ST.26 단락 19는 DNA의 우라실 또는 RNA의 티민은 변형된 뉴클레오티드로 간주되며 서열에서 "t"로 표현되어야 하고, 특징 키 "modified\_base", 한정자 값으로서 "OTHER"를 갖는 한정자 "mod\_base" 및 한정자 값으로서 "우라실" 또는 "티민"을 각각 갖는 한정자 "참조"를 사용하여 특징 표에 추가로 설명되어야 한다고 명시한다.

# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 RNA 서열을 개시한다:

5' -cgucccacgugtccgaggua-3'

- 위치 12의 "티민"을 참고한다. 이 잔기는 변형된 뉴클레오티드로 주석을 달아야 한다.
- ST.26 단락 19는 DNA의 우라실 또는 RNA의 티민은 변형된 뉴클레오티드로 간주되며 "t"와 같은 서열로 표현되어야 하고, 특징 키 "**modified\_base**", 한정자 값으로서 "OTHER"를 갖는 한정자 "**mod\_base**" 및 한정자 값으로서 "우라실" 또는 "티민"을 각각 갖는 한정자 "**참조**"를 사용하여 특징 표에 추가로 설명되어야 한다고 명시한다.

# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 RNA 서열을 개시한다:

5' -cgucccacgugtccgaggua - 3'

- ✓ 모든 우라실 잔기는 기호 "t" 로 표시되어야 한다. 따라서, 서열은 다음과 같이 서열 목록에 표시되어야 한다:

cgtcccacgtgtccgaggta



# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 RNA 서열을 개시한다:

**5' -cgucccacgugtccgaggua - 3'**

- ✓ 모든 우라실 잔기는 기호 "t" 로 표시되어야 한다. 따라서, 서열은 다음과 같이 서열 목록에 표시되어야 한다:

**cgtccccacgtgtccgaggta**

- ✓ 위치 "12"를 갖는 특징 키 "modified\_base"

# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 RNA 서열을 개시한다:

**5' -cgucccacgugtccgaggua - 3'**

- ✓ 모든 우라실 잔기는 기호 "t" 로 표시되어야 한다. 따라서, 서열은 다음과 같이 서열 목록에 표시되어야 한다:

**cgtccccacgtgtccgaggta**

- ✓ 위치 "12"를 갖는 특징 키 "modified\_base"
- ✓ 값 "OTHER"를 갖는 한정자 "mod\_base"

# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 RNA 서열을 개시한다:

**5' -cgucccacgugtccgaggua - 3'**

- ✓ 모든 우라실 잔기는 기호 "t" 로 표시되어야 한다. 따라서, 서열은 다음과 같이 서열 목록에 표시되어야 한다:

**cgtccccacgtgtccgaggta**

- ✓ 위치 "12"를 갖는 특징 키 "modified\_base"
- ✓ 값 "OTHER"를 갖는 한정자 "mod\_base"
- ✓ 값 "티민"을 갖는 한정자 "참고"

# 특수 상황

## DNA 및 RNA 서열

```

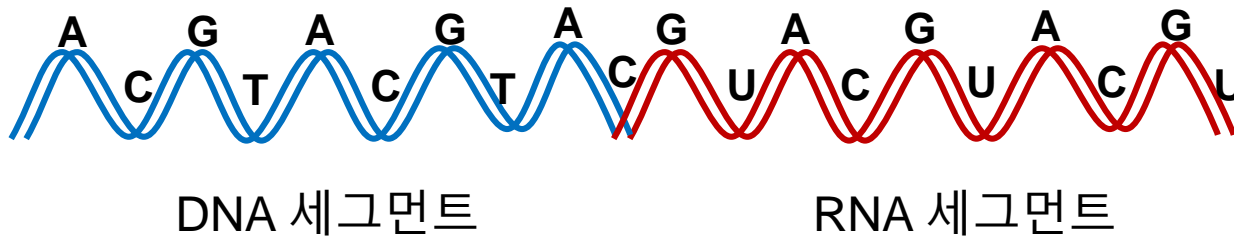
<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length>20</INSDSeq_length>
    <INSDSeq_moltype>RNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..20</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>other RNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q7">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
      <INSDFeature>
        <INSDFeature_key>modified_base</INSDFeature_key>
        <INSDFeature_location>12</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mod_base</INSDQualifier_name>
            <INSDQualifier_value>OTHER</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q8">
            <INSDQualifier_name>note</INSDQualifier_name>
            <INSDQualifier_value>thymine</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>cgtcccacgtgtcgcgaggtta</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

# 특수 상황

## DNA/RNA 하이브리드 서열

- 서열이 하이브리드 분자인 경우; 즉, 골격의 일부는 DNA이고, 골격의 일부는 RNA이다:



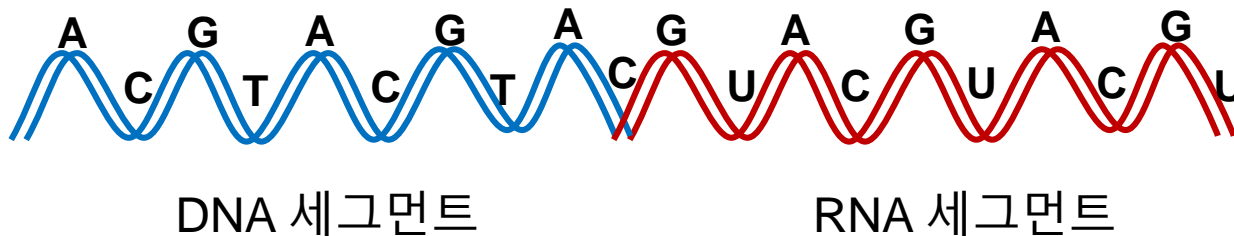
### ST.26 단락 55적용:

"55. 하나 이상의 뉴클레오타이드의 DNA 및 RNA 세그먼트 둘 모두를 포함하는 뉴클레오타이드 서열의 경우, 분자 유형은 DNA로 표시되어야 한다. 결합된 DNA/RNA 분자는 특징 키 "공급원" 및 값 "합성 작제물"을 갖는 필수 한정자 "유기체" 및 값 "다른 DNA"를 갖는 필수 한정자 "mol\_type"을 사용하여 특징 표에 추가로 설명되어야 한다. 결합된 DNA/RNA 분자의 각각의 DNA 및 RNA 세그먼트는 세그먼트가 DNA이거나 RNA인지를 나타내는 특징 키 "misc\_feature" 및 한정자 "참고"로 추가로 설명되어야 한다.

# 특수 상황

## DNA/RNA 하이브리드 서열

- 서열이 하이브리드 분자인 경우; 즉, 골격의 일부는 DNA이고, 골격의 일부는 RNA이다:



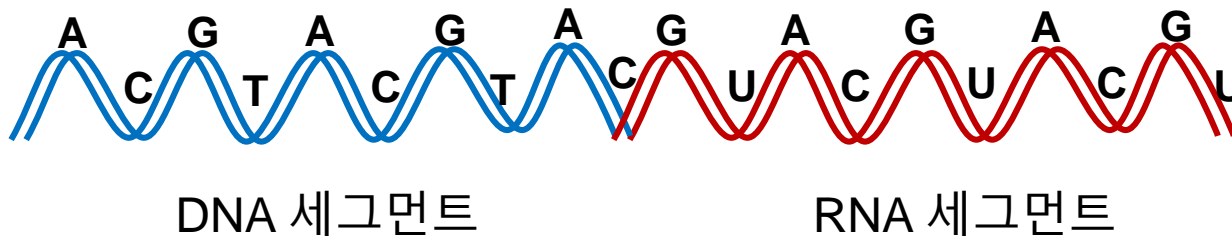
### ST.26 단락 55적용:

"55. 하나 이상의 뉴클레오타이드의 DNA 및 RNA 세그먼트 둘 모두를 포함하는 뉴클레오타이드 서열의 경우, 분자 유형은 DNA로 표시되어야 한다. 결합된 DNA/RNA 분자는 특징 키 "공급원" 및 값 "합성 작제물"을 갖는 필수 한정자 "유기체" 및 값 "다른 DNA"를 갖는 필수 한정자 "mol\_type"을 사용하여 특징 표에 추가로 설명되어야 한다. 결합된 DNA/RNA 분자의 각각의 DNA 및 RNA 세그먼트는 세그먼트가 DNA이거나 RNA인지를 나타내는 특징 키 "misc\_feature" 및 한정자 "참고"로 추가로 설명되어야 한다.

# 특수 상황

## DNA/RNA 하이브리드 서열

- 서열이 하이브리드 분자인 경우; 즉, 골격의 일부는 DNA이고, 골격의 일부는 RNA이다:



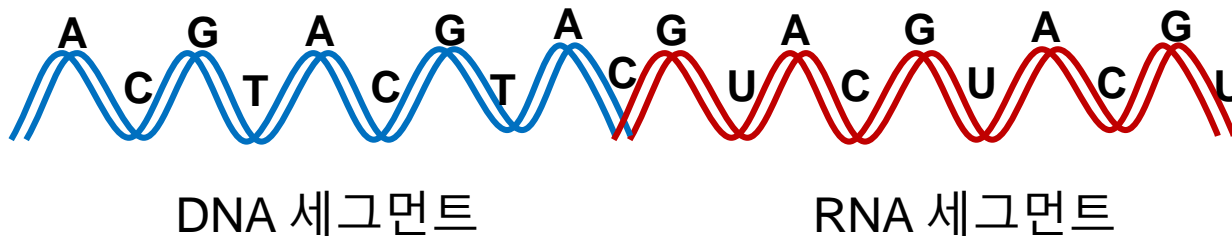
### ST.26 단락 55적용:

"55. 하나 이상의 뉴클레오타이드의 DNA 및 RNA 세그먼트 둘 모두를 포함하는 뉴클레오타이드 서열의 경우, 분자 유형은 DNA로 표시되어야 한다. 결합된 DNA/RNA 분자는 특징 키 "공급원" 및 값 "합성 작제물"을 갖는 필수 한정자 "유기체" 및 값 "다른 DNA"를 갖는 필수 한정자 "mol\_type"을 사용하여 특징 표에 추가로 설명되어야 한다. 결합된 DNA/RNA 분자의 각각의 DNA 및 RNA 세그먼트는 세그먼트가 DNA이거나 RNA인지를 나타내는 특징 키 "misc\_feature" 및 한정자 "참고"로 추가로 설명되어야 한다.

# 특수 상황

## DNA/RNA 하이브리드 서열

- 서열이 하이브리드 분자인 경우; 즉, 골격의 일부는 DNA이고, 골격의 일부는 RNA이다:



### ST.26 단락 55적용:

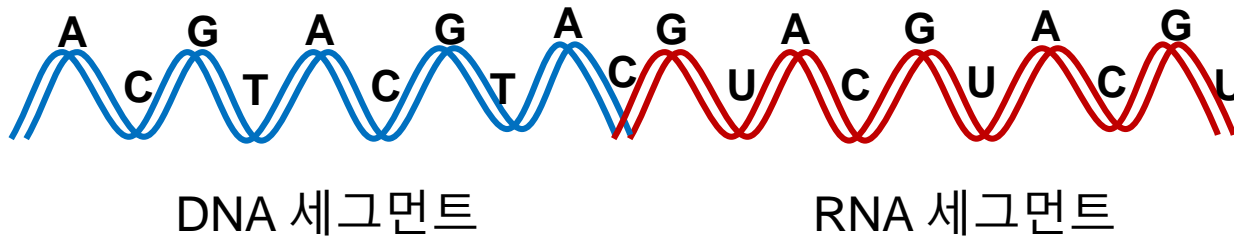
"55. 하나 이상의 뉴클레오타이드의 DNA 및 RNA 세그먼트 둘 모두를 포함하는 뉴클레오타이드 서열의 경우, 분자 유형은 DNA로 표시되어야 한다. 결합된 DNA/RNA 분자는 특징 키 "공급원" 및 값 "합성 작제물"을 갖는 필수 한정자 "유기체" 및 값 "다른 DNA"를 갖는 필수 한정자 "mol\_type"을 사용하여 특징 표에 추가로 설명되어야 한다. 결합된 DNA/RNA 분자의 각각의 DNA 및 RNA 세그먼트는 세그먼트가 DNA이거나 RNA인지를 나타내는 특징 키 "misc\_feature" 및 한정자 "참고"로 추가로 설명되어야 한다.



# 특수 상황

## DNA/RNA 하이브리드 서열

- 서열이 하이브리드 분자인 경우; 즉, 골격의 일부는 DNA이고, 골격의 일부는 RNA이다:



### ST.26 단락 55적용:

"55. 하나 이상의 뉴클레오타이드의 DNA 및 RNA 세그먼트 둘 모두를 포함하는 뉴클레오타이드 서열의 경우, 분자 유형은 DNA로 표시되어야 한다. 결합된 DNA/RNA 분자는 특징 키 "공급원" 및 값 "합성 작제물"을 갖는 필수 한정자 "유기체" 및 값 "다른 DNA"를 갖는 필수 한정자 "mol\_type"을 사용하여 특징 표에 추가로 설명되어야 한다. 결합된 DNA/RNA 분자의 각각의 DNA 및 RNA 세그먼트는 **세그먼트가 DNA이거나 RNA인지를 나타내는 특징 키 "misc\_feature" 및 한정자 "참고"로 추가로 설명되어야 한다.**

# 특수 상황

## DNA/RNA 하이브리드 서열

본 출원은 다음의 DNA/RNA 하이브리드 서열을 개시한다:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

여기서, 대문자 기호는 DNA 부분을 나타내고, 소문자 기호는 RNA 부분을 나타낸다.

- 잔기 1 내지 6 및 27 내지 32는 DNA이고, 잔기 7 내지 26은 RNA이다
- 고려 대상:
  1. 유기체 지정
  2. 분자 유형 및 mol\_type
  3. DNA 및 RNA 세그먼트의 확인

# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 DNA/RNA 하이브리드 서열을 개시한다:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

여기서, 대문자 기호는 DNA 부분을 나타내고, 소문자 기호는 RNA 부분을 나타낸다.

- ST.26 단락 55는 "...분자 유형은 DNA로 표시되어야 한다. 결합된 DNA/RNA 분자는 특징 키 "공급원" 및 값 "합성 작제물"을 갖는 필수 한정자 "유기체" 및 값 "다른 DNA"를 갖는 필수 한정자 "mol\_type"을 사용하여 특징 표에 추가로 기술되어야 한다"고 명시한다.

- ✓ 분자 유형 = "DNA"
- ✓ 한정자 유기체 명칭 = "합성 작제물"
- ✓ 한정자 mol\_type = "다른 DNA"

# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 DNA/RNA 하이브리드 서열을 개시한다:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

여기서, 대문자 기호는 DNA 부분을 나타내고, 소문자 기호는 RNA 부분을 나타낸다.

- ✓ 모든 우라실 잔기는 기호 "t" 로 표시되어야 한다. 따라서, 서열은 다음과 같이 서열 목록에 표시되어야 한다:

**acctgccgtcccacgtgtccgaggtagcatta**

# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 DNA/RNA 하이브리드 서열을 개시한다:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

여기서, 대문자 기호는 DNA 부분을 나타내고, 소문자 기호는 RNA 부분을 나타낸다.

- ST.26 단락 55는 "결합된 DNA/RNA 분자의 각각의 DNA 및 RNA 세그먼트는 세그먼트가 DNA 이거나 RNA 인지를 나타내는 특징 키 "misc\_feature" 및 한정자 "참고"로 추가로 기술되어야 한다" 고 명시한다.

✓ 3개의 세그먼트 = 3개의 "misc\_feature" 특징 키

# 특수 상황

## DNA 및 RNA 서열

본 출원은 다음의 DNA/RNA 하이브리드 서열을 개시한다:

**5' -ACCTGCcgucccacguguccgagguaGCATTA-3'**

여기서, 대문자 기호는 DNA 부분을 나타내고, 소문자 기호는 RNA 부분을 나타낸다.

- ✓ 세그먼트 1, 잔기 1 내지 6:  
위치 "1..6"을 갖는 "misc\_feature" 특징 키 값 "DNA"를 갖는 한정자 "참조"
- ✓ 세그먼트 2, 잔기 7 내지 26:  
위치 "7..26"을 갖는 "misc\_feature" 특징 키 값 "RNA"를 갖는 한정자 "참조"
- ✓ 세그먼트 3, 잔기 27 내지 32  
위치 "27..32"를 갖는 "misc\_feature" 특징 키 값 "DNA"를 갖는 한정자 "참조"

# 특수 상황

## DNA 및 RNA 서열

```

<SequenceData sequenceIDNumber="4">
  <INSDSeq>
    <INSDSeq_length>32</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..32</INSDFeature_location>
        <INSDFeature_quals>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>other DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q10">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>synthetic construct</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_quals>
      </INSDFeature>
    </INSDSeq_feature-table>
  </INSDSeq>
</SequenceData>

```

# 특수 상황

## DNA 및 RNA 서열

```

<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>1..6</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier id="q11">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>DNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>7..26</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier id="q12">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>RNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
<INSDFeature>
  <INSDFeature_key>misc_feature</INSDFeature_key>
  <INSDFeature_location>27..32</INSDFeature_location>
  <INSDFeature_quals>
    <INSDQualifier id="q13">
      <INSDQualifier_name>note</INSDQualifier_name>
      <INSDQualifier_value>DNA</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_quals>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>acctgccgtccccacgtgtccgaggtagcatta</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
</ST26SequenceListing>

```

DNA 세그먼트 1

RNA 세그먼트 2

DNA 세그먼트 3



# 궁금한 점

뉴클레오티드 유사체,  
D-아미노산, 및 분지  
된 서열

# 새로운 분자 유형

## 뉴클레오티드 유사체

- 하나 이상의 뉴클레오티드 유사체를 함유하는 핵산 서열은 ST.26 규칙에 적용된다
- 뉴클레오티드 유사체는 ST.26, 단락 3(g)(2) 아래의 "뉴클레오티드"의 정의에 포함된다:

*"핵산 유사체의 골격을 형성하는 경우 2' 데옥시리보스 5' 모노포스페이트 또는 리보스 5' 모노포스페이트의 유사체는 2' 데옥시리보스 5' 모노포스페이트 또는 리보스 5' 모노포스페이트 골격을 함유하는 핵산에서 핵염기의 배열을 모방하는 핵염기의 배열을 발생시키며, 여기서 핵산 유사체는 상보적 핵산과 염기쌍을 형성할 수 있다"*

- 일반적인 뉴클레오티드 유사체는 펩티드 핵산(PNA), 글리콜 핵산(GNA), 트레오스 핵산, 및 모르폴리노를 포함한다
- 5'에서 3' 방향을 모방하는 좌측에서 우측 방향으로 표시되어야 한다. (ST.26, 단락 11)

# 새로운 분자 유형

## 뉴클레오티드 유사체

본 특허 출원은 다음의 글리콜 핵산 (GNA) 서열을 개시한다:

$PO_4$ -tagttcattgactaaggctccccattgact-OH

여기서,  $PO_4$  서열의 말단은 DNA 서열의 5' 말단을 모방한다.

✓ 이 서열은 필수이며 서열 목록에 포함된다

# 새로운 분자 유형

## 뉴클레오티드 유사체

본 특허 출원은 다음의 글리콜 핵산 (GNA) 서열을 개시한다:

$$\text{PO}_4\text{-tagttcattgactaaggctccccattgact-OH}$$

여기서,  $\text{PO}_4$  서열의 말단은 DNA 서열의 5' 말단을 모방한다.

- ✓ 이 서열은 서열 목록에 포함되어야 한다
- ✓  $\text{PO}_4$  말단은 5' 말단을 모방하므로 서열은 다음에 표시되어야 한다  
제시된 방향

# 새로운 분자 유형

## 뉴클레오티드 유사체

본 특허 출원은 다음의 글리콜 핵산 (GNA) 서열을 개시한다:

$PO_4$ -tagttcattgactaaggctccccattgact-OH

여기서,  $PO_4$  서열의 말단은 DNA 서열의 5' 말단을 모방한다.

- ✓ 이 서열은 서열 목록에 포함되어야 한다
- ✓  $PO_4$  말단은 5' 말단을 모방하므로 서열은 다음에 표시되어야 한다  
제시된 방향
- ✓ 전체 서열에 "modified\_base" 특징 키, 값 "OTHER"를 갖는 "mod\_base" 한정자, 및 "글리콜 핵산"과 같은 변형된 뉴클레오티드의 완전하고 약칭되지 않은 명칭을 포함하는 참고 한정자가 주석이 달려야 한다

# 새로운 분자 유형

## 뉴클레오티드 유사체

```

- <INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>1..30</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 (glycol nucleic acids)</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>tagttcattgactaaggctccccattgact</INSDSeq_sequence>
</INSDSeq>

```

참고: 이 예의 대한 확장된 논의는 다음에서 찾을 수 있다  
WIPO 표준 ST.26, 부록 VI, 예 3(g)-4.

# 새로운 분자 유형

## D-아미노산

- 하나 이상의 D-아미노산을 함유하는 아미노산 서열은 ST.26 규칙에 적용된다
- D-아미노산은 ST.26, 단락 3(a) 아래의 "아미노산"의 정의에 포함된다:

*"아미노산"은 부록 1 (섹션 3, 표 3 참조)에 제시된 임의의 기호를 사용하여 표시될 수 있는 임의의 아미노산을 의미한다. 상기 아미노산은 특히 D-아미노산 및 변형된 또는 합성 측쇄를 함유하는 아미노산을 포함한다."*

- D-아미노산은 상응하는 변형되지 않은 L 아미노산 기호(가능한 경우)로서 서열 내에 표시되어야 한다
- 특징 표에서 변형된 아미노산으로 설명되어야 한다



# 새로운 분자 유형

## D-아미노산

본 특허 출원은 다음과 같은 서열을 기술한다:

D-Ala-D-Glu-Lys-Leu-Gly-D-Met

✓ 이 서열은 서열 목록에 포함되어야 한다

# 새로운 분자 유형

## D-아미노산

본 특허 출원은 다음과 같은 서열을 설명한다:

D-Ala-D-Glu-Lys-Leu-Gly-D-Met

- ✓ 이 서열은 서열 목록에 포함되어야 한다
- ✓ AGKLGGM으로 표현되어야 한다
- ✓ 위치 1의 알라닌, 위치 2의 글루탐산 및 위치 6의 메티오닌에는 각각 "SITE" 특징 키 및 "참고" 한정자가 주석이 달려야 하며, 이는 상응하는 아미노산의 완전하고 약칭되지 않은 명칭을 갖는다

# 새로운 분자 유형

## D-아미노산

```

- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>1</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q4">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>D-alanine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>2</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q5">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>D-glutamic acid</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>6</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q6">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>D-methionine</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>AEKLG</INSDSeq_sequence>

```

참고: 유사한 예의 확장된 논의는 다음에서 찾을 수 있다  
WIPO 표준 ST.26, 부록 VI, 예 3(a)-1.

# 새로운 분자 유형

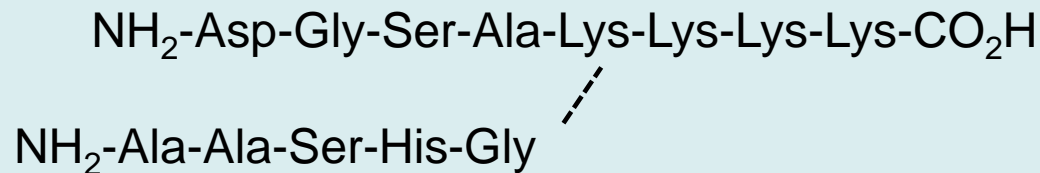
## 분지된 서열

- 분지된 핵산 서열 및 분지된 아미노산 서열은 ST.26 규칙에 적용된다
- 분지된 서열의 선형 영역은 10개 이상의 구체적으로 정의된 뉴클레오티드 또는 4개 이상의 구체적으로 정의된 아미노산을 함유하는 경우 서열 목록에 포함되어야 한다. (WIPO 표준 ST.26, 단락 7)
- 최소 길이 요건을 충족시키는 분지된 서열의 각각의 선형 영역은 그 자신의 SEQ ID 번호를 갖는 별도의 서열로서 포함되어야 한다
- 각각 개별 선형 영역에서 구체적으로 정의된 잔기의 수가 고려되어야 하며, 이는 구조 내의 구체적으로 정의된 잔기의 총 수가 아니다

# 새로운 분자 유형

## 분지된 서열

본 특허 출원은 다음과 같은 서열로 펩티드를 설명한다:.



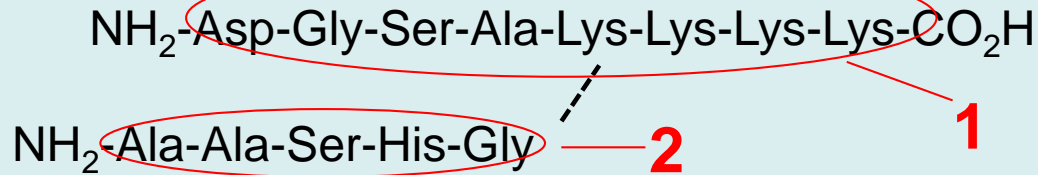
여기서, ----- 는 글리신의 카복시 말단과 리신의 측쇄 사이의 아미드 결합을 나타낸다

- ✓ 둘 모두의 선형 영역은  $\geq 4$ 개의 구체적으로 정의된 아미노산을 함유하므로, 둘 다 서열 목록에 포함될 필요가 있다

# 새로운 분자 유형

## 분지된 서열

본 특허 출원은 다음과 같은 서열로 펩티드를 설명한다:



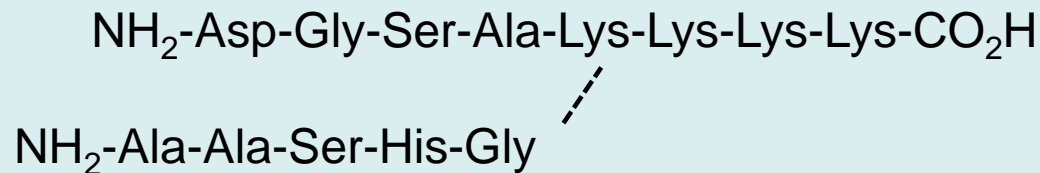
여기서, ----- 는 글리신의 카르복시 말단과 리신의 측쇄 사이의 아미드 결합을 나타낸다

- ✓ 둘 모두의 선형 영역은  $\geq 4$ 개의 구체적으로 정의된 아미노산을 포함하므로, 둘 다 서열 목록에 포함될 필요가 있다
- ✓ 각각의 선형 영역은 그 자신의 SEQ ID 번호를 갖는 별도의 서열로서 포함되어야 한다

# 새로운 분자 유형

## 분지된 서열

본 특허 출원은 다음과 같은 서열로 펩티드를 설명한다:



여기서, ----- 는 글리신의 카르복시 말단과 리신의 측쇄 사이의 아미드 결합을 나타낸다

- ✓ 둘 모두의 선형 영역은  $\geq 4$ 개의 구체적으로 정의된 아미노산을 포함하므로, 둘 다 서열 목록에 포함될 필요가 있다
- ✓ 각각의 선형 영역은 그 자신의 SEQ ID 번호를 갖는 별도의 서열로서 포함되어야 한다
- ✓ 아미드 결합 연결의 위치 및 특성을 표시하기 위해 둘 모두의 서열에 주석을 달아야 한다

# 새로운 분자 유형

## 분지된 서열

1

```

- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>5</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q4">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>Lysine side chain is amide bonded to the carboxy
        terminus of the glycine in SEQ ID NO:2</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>DGS AKKKK</INSDSeq_sequence>

```

2

```

- <INSDFeature>
  <INSDFeature_key>SITE</INSDFeature_key>
  <INSDFeature_location>5</INSDFeature_location>
  - <INSDFeature_qual>
    - <INSDQualifier id="q11">
      <INSDQualifier_name>NOTE</INSDQualifier_name>
      <INSDQualifier_value>glycine carboxy terminus is amide bonded to the side
        chain of the lysine in SEQ ID NO:1, position 5</INSDQualifier_value>
    </INSDQualifier>
  </INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>AASHG</INSDSeq_sequence>

```

참고: 유사한 예의 확장된 논의는 WIPO 표준 ST.26, 부록 VI, 예 7(b)-3에서 찾을 수 있다.



# 서열 변이체

# 서열 변이체

- ST.26 단락 3(m)은 "**변이체 서열**"을 다음과 같이 정의한다

*“일차 서열에 대한 하나 이상의 차이를 함유하는 뉴클레오티드 또는 아미노산 서열. 이들 차이는 대안적 잔기(단락 15 및 27 참조), 변형된 잔기(단락 3(g), 3(h), 16, 및 29 참조), 결실, 삽입, 및 치환을 포함할 수 있다. 단락 93 내지 95를 참조한다.”*

- 변이체 서열이 개시되는 방식은 서열 목록에 어떻게 표시되어야 하는지를 결정할 것이다.
- ST.26 단락 93 내지 95는 변이체가 표시되어야 하는 방법을 제어한다.

# 서열 변이체

## 단락 93

단락 93: 일차 서열 및 해당 서열의 임의의 변이체, 이는 각각 이들의 잔기의 열거에 의해 개시되고 단락 7에 포함되며, 각각 서열 목록에 포함되어야 하고, 이들 자신의 서열 확인 번호가 지정되어야 한다.



각각의 변이체가 개별적으로 열거된 경우, 각각의 변이체는 그 자신의 SEQ ID 번호를 가져야 한다!

# 서열 변이체

## 단락 93

본 특허 출원은 다음의 다중 서열 정렬을 갖는 도면을 포함한다:

컨센서스

Homo sapiens

Pongo abelii

Pan paniscus

Rhinopithecus bieti

Rhinopithecus roxellana

```

LEGnEQFINAakIIRHPknrkTLnNDImLIK
LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIK
LEGNEQFINAAKIIRHPQYDRKTVNNDIMLIK
LEGNEQFINAAKIIRHPKYNRITLNNDIMLIK
LEGNEQFINATKIIRHPKYNGNTLNNDIMLIK
LEGNEQFINATQIIRHPKYNGNTLNNDIMLIK
  
```

소문자는 정렬된 서열 중 우세한 아미노산 잔기를 나타낸다.

- ✓ 6개의 열거된 서열 각각은 그 자신의 SEQ ID 번호를 갖는 별도의 서열로서 서열 목록에 포함되어야 한다.

참고: 유사한 예의 확장된 논의는 다음에서 찾을 수 있다  
WIPO 표준 ST.26, 부록 VI, 예 93-3.

# 서열 변이체

## 적절한 주석

- 변형체에 대한 주석을 작성하는 적절한 기능 키는 분자 유형 및 변이의 특성에 좌우된다:

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.

WIPO 표준 ST.26, 단락 96

# 서열 변이체

## 가장 제한적인 모호성 기호

- ST.26 단락 15 및 27은 모호성 기호가 필요한 경우 "가장 제한적인 기호를 사용해야 한다..."라고 명시한다.

그게 무슨 의미인가?

예를 들어:

뉴클레오티드 서열에서, 위치가 "a" 또는 "c"일 수 있는 경우, "n" 대신에 모호성 기호 "m"을 사용한다.

아미노산 서열에서, 위치가 "L" 또는 "I"일 수 있는 경우, "X" 대신에 모호성 기호 "J"를 사용한다.

"n" 및 "X"는 내정값을 갖고 있기 때문에 "n" 또는 "X"가 내정값이 아닌 다른 값에 사용되는 경우에는 주석이 필요하다는 것을 기억한다.

# 서열 변이체

## 단락 94

*단락 94: 임의의 변이체 서열은 하나 이상의 위치에서 열거된 대안적 잔기를 갖는 단일 서열로 개시되며, 서열 목록에 포함되어야 하며, 단일 서열로 표시되어야 하며, 여기서 열거된 대안적 잔기는 가장 제한적인 모호성 기호로 표시된다(단락 15 및 27 참조).*



변이체가 별도로 열거되지 않고, 단순히 일차 서열 내의 가변 잔기로서 제시되는 경우, 이들은 별도의 SEQ ID 번호를 갖는 것이 요구되지 않는다!

# 서열 변이체

## 단락 94

본 특허 출원은 다음과 같은 서열의 펩티드를 개시한다:

Gly-Gly-Gly-[Leu 또는 Ile]-Ala-Thr-[Ser 또는 Thr]

- ✓ 단일 서열로 서열 목록에 포함될 수 있다
- ✓ 서열의 바람직한 표현은 다음과 같다: GGGJATX
- ✓ [Leu 또는 Ile]은 가장 제한적인 모호성 기호 "J"로 표시되어야 한다
- ✓ [Ser 또는 Thr]은 X가 세린 또는 트레오닌이라는 것을 나타내는 한정자 참조와 함께 특징 키 "VARIANT"와 함께 기호 "X"로 표시되어야 한다

참고: 이 예의 확장된 논의는 다음에서 찾을 수 있다  
WIPO 표준 ST.26, 부록 VI, 예 94-1.



# 서열 변이체

## 단락 95

단락 95: 임의의 변이체 서열은 서열 목록의 일차 서열에서 결실(들), 삽입(들), 또는 치환(들)에 대한 참조에 의해서만 개시되고, , 서열 목록에 포함되어야 한다. 서열 목록에 포함되는 경우, 상기 변이체 서열:

- (a)는 일차 서열의 주석에 의해 표시될 수 있으며, 여기서 이는 단일 위치 또는 여러 개의 구별되는 위치에 변이(들)를 함유하며, 이들 변이의 발생은 독립적이다;
- (b)는 별도의 서열로 표시되어야 하며, 이의 자신의 서열 확인 번호가 지정되어야 하고, 여기서 이는 다수의 별개 위치에서 변이를 함유하며, 해당 변이의 발생은 상호의존적이다;
- (c)는 별도의 서열로 표시되어야 하며, 이의 자신의 서열 확인 번호가 지정되어야 하고, 여기서 이는 1000개 초과와 잔기를 함유하는 삽입되거나 치환된 서열을 함유한다 (단락 86 참조).



"결실(들), 삽입(들), 또는 치환(들)에 대한 언급"은 변이체가 산문으로 개시된다는 것을 의미한다.

# 서열 변이체

## 단락 95(a)

본 출원은 다음과 같은 개시를 함유한다:

펩티드 단편 1: Gly-Leu-Pro-Xaa-Arg-Ile-Cys

여기서, Xaa는 임의의 아미노산이다

\* \* \*

... 또 다른 구현예에서, 펩티드 단편 1은 Gly-Leu-Pro-Xaa-Arg-Ile-Cys이고, 여기서 Xaa는 Val, Thr, 또는 Asp일 수 있다.....

\* \* \*

... 또 다른 구현예에서, 펩티드 단편 1은 Gly-Leu-Pro-Xaa-Arg-Ile-Cys이고, 여기서 Xaa는 Val일 수 있다...

*일차 서열, Gly-Leu-Pro-Xaa-Arg-Ile-Cys, 은 단일 위치에 변이(들)을 포함하며, 해당 변이의 발생은 독립적이다*

# 서열 변이체

## 단락 95(a)

- ✓ 일차 서열은 서열 목록에 포함되어야 하며, 변이체는 일차 서열의 주석에 의해 표시될 수 있다
- ✓ 가장 포괄적인 구현예는 서열 목록에 포함된 버전이어야 하며 – 이 예에서, 이는 Xaa가 "임의의 아미노산"인 버전이다
- ✓ 서열은 GLPXRIC로 표시되어야 하며, 위치 4에서의 특징 키 "VARIANT" 및 Xaa가 "임의의 아미노산" 임을 나타내는 한정자 "참고"를 필요로 한다
- ✓ 필요하지는 않지만, 다음 3개의 변이체가 서열 목록에 별도의 서열로 포함되는 것이 권장된다:

GLPVRIC

GLPIRIC

GLPDRIC

참고: 이 예의 확장된 논의는 WIPO 표준 ST.26, 부록 VI에서 찾을 수 있다,  
예 95(a)-1.

# 서열 변이체

## 단락 95(b)

본 특허 출원은 다음과 같은 컨센서스 서열을 기술한다:

aatg $n_1$ cccacgaatg $n_2$ cac

여기서,  $n_1$  및  $n_2$ 는 a, t, g, 또는 c일 수 있다.

여러 변이체 서열이 다음과 같이 개시된다:

$n_1$ 이 a인 경우,  $n_2$ 는 t, g 또는 c이다

$n_1$ 이 t인 경우,  $n_2$ 는 a, g 또는 c이다

$n_1$ 이 g인 경우,  $n_2$ 는 t, a 또는 c이다

$n_1$ 이 c인 경우,  $n_2$ 는 t, g 또는 a이다

*컨센서스 서열은 다수의 별개의 위치에서의 변이를 함유하고, 해당 변이의 발생은 상호의존적이다*

# 서열 변이체

## 단락 95(b)

- ✓ 컨센서스 서열은 서열 목록에 포함되어야 하며, 변이체는 별도의 서열로 표시되어야 한다
- ✓ 가장 포괄적인 구현에는 서열 목록에 포함된 버전이어야 하며: 여기서  $n_1$  및  $n_2$ "는 a, t, g 또는 c일 수 있다"
- ✓ 서열은 `aatgncccacgaatgncac`로 표시되어야 한다
- ✓ 특징 키의 부재하에서 "n"이 "a", "c", "g" 또는 "t" 중 하나로 해석되기 때문에 n에 대한 주석이 필요하지 않다 (ST.26 단락 15 참조)
- ✓ 필요하지는 않지만, 다음 4개의 변이체가 서열 목록에 별도의 서열로 포함되는 것이 강력히 권장된다:

`aatgacccacgaatgbcac`

(b = t, g, 또는 c)

`aatgtcccacgaatgvcac`

(v = a, g, 또는 c)

`aatggcccacgaatghcac`

(h = t, a, 또는 c)

`aatgccccacgaatgdcac`

(d = t, g, 또는 a)

참고: 유사한 예의 확장된 논의는 다음과 같을 수 있다  
WIPO 표준 ST.26, 부록 VI, 예 95(b)에서 발견되었다.

# 서열 변이체

## 단락 95(c)

본 출원은 다음과 같은 개시를 함유한다:

... -Met-Gly-Leu-Pro-Arg-Xaa-Arg-Ile-Cys-Lys- ...

여기서, Xaa는 Gly 또는 서열의 삽입이고

Cys-Tyr-Ile-Lys-Ser-(1000 아미노산)-Leu-Thr-Pro-Lys

*하나의 변이체 서열은 1000개 잔기를 초과하는 삽입된 또는 치환된 서열을 함유한다*

# 서열 변이체

## 단락 95(c)

- ✓ Xaa = 1000개가 넘는 잔기의 삽입인 변이체가 그 자신의 SEQ ID 번호를 갖는 별개의 서열로서 서열 목록에 포함되어야 한다.
- ✓ Xaa = Gly인 변이체가 또한 그 자신의 SEQ ID 번호를 갖는 별개의 서열로서 서열 목록에 포함될 것이다.

...-MGLPRGRICK-...

# 서열 변이체

**Q:** 다음의 개시에 적용되는 단락은 무엇인가?

**G-L-P-T-R-I-C- [L or I]-A-V- [G or A]**



# 서열 변이체

**Q:** 다음의 개시에 적용되는 단락은 무엇인가?

**G-L-P-T-R-I-C- [L or I] -A-V- [G or A]**

**A:** 단락 94:

"임의의 변이체 서열은 하나 이상의 위치에서 열거된 대안적 잔기가 있는 단일 서열로 개시되며, 서열 목록에 포함되어야 하며, 단일 서열로 표시되어야 하고, 여기서 열거된 대안적 잔기는 가장 제한적인 모호성 기호로 표시된다."

# 서열 변이체

**Q:** 다음의 개시에 적용되는 단락은 무엇인가?

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							-		

표의 빈 공간은 변이체 내의 아미노산이 "서열" 내의 상응하는 아미노산과 동일하다는 것을 나타내고, "-" 는 "서열" 내의 상응하는 아미노산의 결실을 나타낸다.

# 서열 변이체

**Q:** 다음의 개시에 적용되는 단락은 무엇인가?

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							-		

표의 빈 공간은 변이체 내의 아미노산이 "서열" 내의 상응하는 아미노산과 동일하다는 것을 나타내고, "-" 는 "서열" 내의 상응하는 아미노산의 결실을 나타낸다.

**A:** 단락 93

*"일차 서열 및 해당 서열의 임의의 변이체는 각각 이들의 잔기의 열거에 의해 개시되고, 단락 7에 포함되며, 각각 서열 목록에 포함되어야 하고, 이들 자신의 서열 확인 번호가 지정되어야 한다."*

# 서열 변이체

**Q:** 다음의 개시에 적용되는 단락은 무엇인가?

도면은 하기 서열을 개시한다:

Met-Gly-Ala-Ile-Pro-Asp-Val-Lys-Arg-Ala-Cys-Trp (서열 1)

본 명세서는 서열 1과 관련된 다음 정보를 함유한다:

... 특정 구현예에서, 서열 1의 위치 7의 발린은 알라닌으로 대체된다...

# 서열 변이체

**Q:** 다음의 개시에 적용되는 단락은 무엇인가?

도면은 하기 서열을 개시한다:

Met-Gly-Ala-Ile-Pro-Asp-Val-Lys-Arg-Ala-Cys-Trp (서열 1)

본 명세서는 서열 1과 관련된 다음 정보를 함유한다:

... 특정 구현예에서, 서열 1의 위치 7의 발린은 알라닌으로 대체된다...

**A:** 단락 95 (a): "임의의 변이체 서열은 서열 목록의 일차 서열에서 결실(들), 삽입(들), 또는 치환(들)에 대한 참조에 의해서만 개시되고, 서열 목록에 포함되어야 한다. 서열 목록에 포함되는 경우, 상기 변이체 서열:

(a)는 일차 서열의 주석에 의해 표시될 수 있으며, 여기서 이는 단일 위치 또는 다수의 구별되는 위치에 변이(들)를 함유하며, 해당 변이의 발생은 독립적이다;"

# 서열 변이체

**Q:** 다음의 개시에 적용되는 단락은 무엇인가?

도면은 하기 서열을 개시한다:

Met-Gly-Ala-Ile-Pro-Asp-Val-Lys-Arg-Ala-Cys-Trp (서열 1)

본 명세서는 서열 1과 관련된 다음 정보를 함유한다:

... 서열 1의 위치 7의 발린이 알라닌으로 대체되는 경우, 위치 10의 알라닌은 발린으로 대체된다....

# 서열 변이체

**Q:** 다음의 개시에 적용되는 단락은 무엇인가?

도면은 하기 서열을 개시한다:

Met-Gly-Ala-Ile-Pro-Asp-Val-Lys-Arg-Ala-Cys-Trp (서열 1)

본 명세서는 서열 1과 관련된 다음 정보를 함유한다:

... 서열 1의 위치 7의 발린이 알라닌으로 대체되는 경우, 위치 10의 알라닌은 발린으로 대체된다....

**A:** 단락 95 (b): 임의의 변이체 서열은 서열 목록의 일차 서열에서 결실(들), 삽입(들), 또는 치환(들)에 대한 참조에 의해서만 개시되고, 서열 목록에 포함되어야 한다. 서열 목록에 포함되는 경우, 상기 변이체 서열:

(b)는 별도의 서열로 표시되어야 하며, 이의 자신의 서열 확인 번호가 지정되어야 하고, 여기서 이는 다수의 별개 위치에서 변이를 함유하며, 해당 변이의 발생은 상호의존적이다;

# 궁금한 점