

■ WIPO 표준 ST.26 고급과정(advanced)

웨бина 교육

참고: 이는 후속 과정으로 사용자가 ST.26 입문 과정에 이미 참여한 것으로 가정합니다.

ST.26 입문과정은 https://www.wipo.int/meetings/en/details.jsp?meeting_id=65769에서 액세스 할 수 있습니다.

이 과정에서 다룰 내용

■ 특징 기호 및 한정자

- 특징 위치 형식
- 한정자 값 형식 및 비영어 언어 한정자 값

■ 특수 상황

- DNA의 우라실 및 RNA의 티민; DNA/RNA 하이브리드 분자
- 핵산 유사체, D-아미노산, 및 분지된 서열
- 서열 변이체

일반적으로 사용되는
특징 기호 (Feature key)
및 한정자(qualifier)

특징 기호 및 한정자

- **특징(feature) 기호:** 위치에 의해 확인된 서열의 하나 이상 잔기를 설명하기 위해 사용될 수 있다
 - 핵산 서열에 대한 특징 기호가 부록 I, 섹션 5에 열거되어 있다
 - 핵산 서열에 대한 특징 기호는 **소문자**이다; 예를 들어, “misc_binding”
 - 아미노산 서열에 대한 특징 기호가 부록 I, 섹션 7에 열거되어 있다
 - 아미노산 서열에 대한 특징 기호는 **대문자**이다; 예를 들어, “REGION”

- **한정자(qualifier)**를 사용하여 특징을 추가로 설명할 수 있다
 - 핵산 서열에 대한 한정자가 부록 I, 섹션 6에 열거되어 있다
 - 핵산 서열에 대한 한정자는 **소문자**이다; 예를 들어, “allele”
 - 아미노산 서열에 대한 한정자가 부록 I, 섹션 8에 열거되어 있다
 - 아미노산 서열에 대한 한정자는 **대문자**이다; 예를 들어, “NOTE”



특징 기호 및 한정자

- **특징(Feature):**
 - **국가**
- **한정자(qualifier):**
 - **명칭: 대한민국**
 - **위치: 아시아 극동**
 - **인구: 5000만명**
 - **수도: 서울**
 - **대표음식: 김치, 된장**
 - **공산품: 반도체**

...ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTTCGATCTCTTGATAGATCTGTTCTCTAAA
CGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAAC
TAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTG
TTGCAGCCGATCATCAGCACATCTAGGTTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTC
CCTGGTTTTCAACGAGAAAACACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTTTCGCGACGTGCTCGTAC
GTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGG
CTTAGTAGAAGTTGAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTGTTTCATCAAACGTTCCGGAT
GCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAACTCGAAGGCATTTCAGTACGGTC
GTAGTGGTGAGACACTTGGTGTCCCTCATGTGGGCGAAATACCAGTGGCTTACCGCAAGGTTCT
TCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTGACTTA
GGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAACCTGGAACACTAACATAGCAGTGGTG
TTACCCGTGAACTCATGCGTGAGCTTAACGGAGGGGCATACACTCGCTATGTCGATAACAACCTTCTGTGG
CCCTGATGGCTACCCTCTTGAGTGCATTAAAGACCTTCTAGCACGTGCTGGTAAAGCTTCATGCACTTTG
TCCGAACAACCTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTG
CTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTGAATTAATTTGGCAAAGAA
ATTTGACACCTTCAATGGGGAATGTCCAAATTTGTATTTCCCTTAAATTCCATAATCAAGACTATTCAA
CCAAGGGTTGAAAAGAAAAAGCTTGATGGCTTTATGGGTAGAATTCGATCTGTCTATCCAGTTGCGTCAC
CAAATGAATGCAACCAAATGTGCCTTTCAACTCTCATGAAGTGTGATCATTGTGGTGAACTTCATGGCA
GACGGGCGATTTTGTAAAGCCACTTGCGAATTTTGTGGCACTGAGAATTTGACTAAAGAAGGTGCCACT
ACTTGTGGTTACTTACCCCAAATGCTGTTGTTAAATTTATTGTCCAGCATGTCACAATTCAGAAGTAG...

...CGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAAC
 TAATTACTGTTCGTTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTG
 TTGCAGCCGATCATCAGCACATCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTC
 CCTGGTTTCAACGAGAAAACACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTTCCGCGACGTGCTCGTAC
 GTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGG

- **특징: 유전자(gene)**

- **한정자 :**

- 유전자명: 인슐린 유전자
- 위치: 420~600
- 조절지역(promoter): 420-450
- 조절 효소 결합부위: 430-440
- 단백질 coding 부위: 450~600

**CTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTGTTTCATCAAACGTTCCGGAT
 GCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAACTCGAAGGCATTACAGTACGGTC
 GTAGTGGTGAGACACTTGGTGTCTTGTCCCTCATGTGGGCGAAATACCAGTGGCTTACCGCAAGTTCT
 TCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTGACTTA
 GGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAACCTGGAACACTAAACATAGCAGTGGTG**

TTACCCGTGAACTCATGCGTGAGCTTAACGGAGGGGCATACACTCGCTATGTCGATAACAACCTTCTGTGG
 CCCTGATGGCTACCCTCTTGAGTGCATTAAGACCTTCTAGCACGTGCTGGTAAAGCTTCATGCACTTTG
 TCCGAACAACCTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTG
 CTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTGAATTAATTGGCAAAGAA
 ATTTGACACCTTCAATGGGGAATGTCCAAATTTTGTATTTCCCTTAAATTCCATAATCAAGACTATTCAA
 CCAAGGGTTGAAAAGAAAAAGCTTGATGGCTTTATGGGTAGAATTCGATCTGTCTATCCAGTTGCGTCAC...

핵산 특징(49개)

| | |
|-----------------|-----------------|
| C_region | polyA_site |
| CDS | precursor_RNA |
| centromere | primer_bind |
| D-loop | propeptide |
| D_segment | protein_bind |
| exon | regulatory |
| gene | repeat_region |
| iDNA | rep_origin |
| intron | rRNA |
| J_segment | S_region |
| mat_peptide | sig_peptide |
| misc_binding | source |
| misc_difference | stem_loop |
| misc_feature | STS |
| misc_feature | telomere |
| misc_RNA | tmRNA |
| misc_structure | transit_peptide |
| mobile_element | tRNA |
| modified_base | unsure |
| mRNA | V_region |
| ncRNA | V_segment |
| N_region | Variation |
| operon | 3'UTR |
| oriT | 5'UTR |

단백질 특징(39개)

| | |
|----------|---------------|
| ACT_SITE | NON_TER |
| BINDING | NP_BIND |
| CA_BIND | PEPTIDE |
| CARBOHYD | PROPEP |
| CHAIN | REGION |
| COILED | REPEAT |
| COMPBIAS | SIGNAL |
| CONFLICT | SITE |
| CROSSLNK | SOURCE |
| DISULFID | STRAND |
| DNA_BIND | TOPO_DOM |
| DOMAIN | TRANSMEM |
| HELIX | TRANSIT |
| INIT_MET | TURN |
| INTRAMEM | UNSURE |
| LIPID | VARIANT |
| METAL | VAR_SEQ |
| MOD_RES | ZN_FING |
| MOTIF | |
| MUTAGEN | |
| NON_STD | |

특징 기호 및 한정자

- 모든 서열에 필수인 **“source”/“SOURCE”** 특징을 제외하고는 선택적이다.
- 각 특징 기호는 특징을 추가로 설명하는 데 사용할 수 있는 **한정자 목록** 가질 수 있다. 대부분의 한정자는 **선택적**이나; 일부 특징 기호는 **필수 한정자**를 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”** 특징의 필수 한정자(2개)
 - **“mol_type”/“MOL_TYPE”**
 - **“organism”/“ORGANISM”**

특징 기호 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>
```

필수 특징

선택 특징

특징 기호 및 한정자

- 일부 특징 기호는 추가 제한사항을 갖는다
 - **Organism(생물체) 범위**; 예를 들어, "C_region"은 **진핵생물**로 제한된다
 - **분자 범위**; 예를 들어, "D-loop"는 **DNA 서열**로 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 기호 및 한정자

핵산 서열: “**modified_base**(변형된 염기)” 특징 기호

- 특징 기호 "**modified_base**" 및 이의 필수 한정자 "**mod_base**"는 변형된 핵산을 설명하기 위해 사용되어야 한다 (ST.26, 단락 16)
- "변형된 핵산"는 다음을 제외한 임의의 핵산이다:
 - 데옥시-[**a, g, c, t**] 3'-모노포스페이트
 - [**a, g, c, u**] 3'-모노포스페이트 (ST.26, 단락 3(f))
- "변형된 핵산"는 가능할 때마다 대응하는 변형되지 않은 핵산에 의해 표현되어야 한다 (부록 1, 섹션 1, 표 1). 그렇지 않으면, 이는 "n" 으로 표현될 수 있다.
 - 예를 들어, "**2'-O-methylcytidine**"은 서열 내의 "**c**"로 표현되어야 한다.
 - "**queuosine**"은 "n"에 의해 표현되어야 한다. 기호 "n"은 단 하나의 잔기에 해당.
- 필수 한정자 '**mod_base**'의 값은 부록 1, 섹션 2, 표 2의 값으로부터 선택되어야 한다. "**other**"가 값이면, 추가 "**note**" 한정자는 변형된 잔기의 완전하고 축약되지 않은 이름을 포함해야 한다.

Table 2: List of modified nucleotides

| Abbreviation | Modified Nucleotide |
|--------------|--|
| 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 |
| qm | 2'-O-methylguanosine |
| i | inosine |
| ib6a | 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 |
| m4c | N4-methylcytosine |

| Abbreviation | Modified Nucleotide |
|--------------|--|
| 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 | wybutosine |
| 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 | wybutosine |
| x | 3-(3-amino-3-carboxypropyl)uridine, (acp3)u |
| OTHER | (requires note qualifier) |

특징 기호 및 한정자

hexan serial: "modified_base" 특징 기호

- 예: 위치 15에서 **inosine**을 갖는 hexan serial

```
<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 I, 섹션 2, 표 2에 수록되어 있다.

특징 기호 및 한정자

hexan serial: "modified_base" 특징 기호

- 예: 위치 22에서 **xanthine**을 갖는 hexan serial

```

<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>

```

- "xanthine"은 부록 I, 섹션 2, 표 2에 나열되어 있지 않으며; 따라서, 필수 "mod_base" 한정자에 대한 값은 "OTHER"이어야 하며, 추가 "note" 한정자는 값 "xanthine"과 함께 포함되어야 한다.

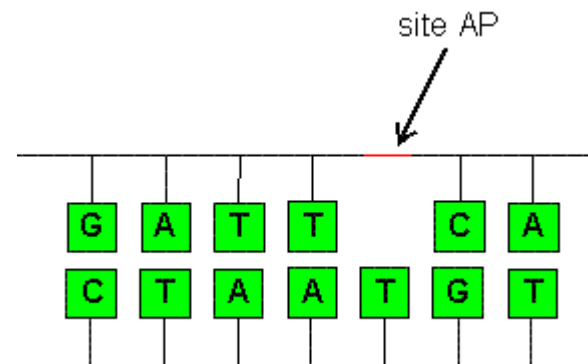
특징 기호 및 한정자

핵산 서열: “modified_base” 특징 기호

- “modified_base”는 또한 무염기 부위(abasic site)를 설명에 사용 될 수 있다:

```

<INSDFeature>
  <INSDFeature_key>modified_base</INSDFeature_key>
  <INSDFeature_location>11</INSDFeature_location>
  <INSDFeature_quals>
    <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_quals>
</INSDFeature>
  
```



- 무염기 부위는 서열에서 "n"으로 표현될 수 있으며, 값 “OTHER”을 갖는 “mod_base” 한정자 및 값 “abasic site”를 갖는 추가 “note” 한정자와 함께 “modified_base” 특징 기호를 사용하여 추가로 설명될 수 있다.

특징 기호 및 한정자

핵산 서열: "CDS" 특징 기호(Coding Sequence)

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

“pseudo”

“pseudogene”

“translation”

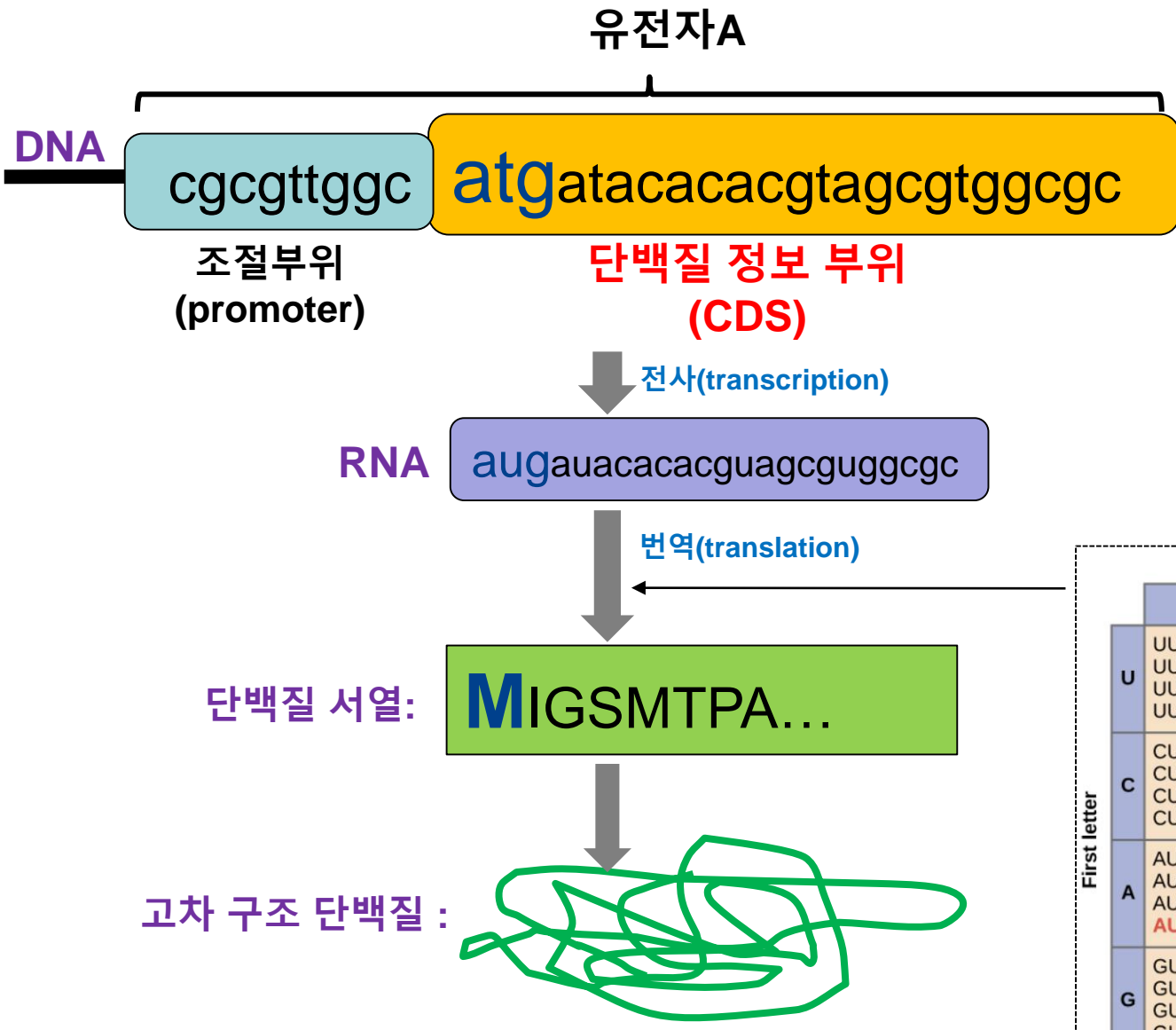
“transl_table”

“codon_start”

“transl_except”

“protein_id”

CDS(Coding Sequence): 실제로 Protein으로 번역되는 서열



Translation-table

| | | Second letter | | | | |
|---|----------------|---------------|-----------------|-----------------|---|--|
| | | U | C | A | G | |
| U | UUU } Phe | UCU } Ser | UAU } Tyr | UGU } Cys | U | |
| | UUC } Leu | UCC } Ser | UAC } Tyr | UGC } Cys | C | |
| | UUA } Leu | UCA } Ser | UAA Stop | UGA Stop | A | |
| | UUG } Leu | UCG } Ser | UAG Stop | UGG } Trp | G | |
| C | CUU } Leu | CCU } Pro | CAU } His | CGU } Arg | U | |
| | CUC } Leu | CCC } Pro | CAC } His | CGC } Arg | C | |
| | CUA } Leu | CCA } Pro | CAA } Gln | CGA } Arg | A | |
| | CUG } Leu | CCG } Pro | CAG } Gln | CGG } Arg | G | |
| A | AUU } Ile | ACU } Thr | AAU } Asn | AGU } Ser | U | |
| | AUC } Ile | ACC } Thr | AAC } Asn | AGC } Ser | C | |
| | AUA } Ile | ACA } Thr | AAA } Lys | AGA } Arg | A | |
| | AUG Met | ACG } Thr | AAG } Lys | AGG } Arg | G | |
| G | GUU } Val | GCU } Ala | GAU } Asp | GGU } Gly | U | |
| | GUC } Val | GCC } Ala | GAC } Asp | GGC } Gly | C | |
| | GUA } Val | GCA } Ala | GAA } Glu | GGA } Gly | A | |
| | GUG } Val | GCG } Ala | GAG } Glu | GGG } Gly | G | |

특징 기호 및 한정자

서열: "CDS" 특징 기호

■ "CDS" 특징은 다음 한정자 중 하나만 가질 수 있다:

| 한정자 | 설명 | 값 |
|-----------------------|---|---|
| 의사 (pseudo) | CDS 특징이 비-기능적이며 번역(translation)을 갖지 않지만, 의사 유전자가 아니라는 것을 나타낸다 | 없음 |
| 의사유전자 (pseudogene) | CDS 특징이 의사유전자이며 번역(translation)되지 않는다 | <ul style="list-style-type: none"> processed unprocessed unitary allelic unknown |
| 번역 (translation) | CDS 번역(translation)으로부터 유래된 아미노산 서열을 나타낸다 | 1-문자 아미노산 약어 |

특징 기호 및 한정자

핵산 서열: "CDS" 특징 기호

- 코딩 서열에 의해 인코딩되고 “**translation**” 한정자에 개시된 **4개 이상의 구체적으로 정의된 아미노산**을 포함하는 아미노산 서열은 **별개의 서열로서 서열 목록에 포함되어야 하고, 그 자신의 서열 확인 번호로 지정되어야 한다.**
- 아미노산 서열로 지정된 서열 확인 번호는 "CDS" 특징 기호 내의 한정자 “**protein_id**”의 값으로 제공되어야 한다.
- 아미노산 서열에 대한 “**SOURCE**” 특징 기호의 “**ORGANISM**” 한정자는 **이의 코딩 서열(DNA, RNA)의 것과 동일해야 한다.**(ST.26, pgh. 92)

핵산 서열식별번호(SEQ ID): 1

- 특징: source
- 한정자: mol_type: DNA
- 한정자: organism: **Homo sapiens**
- 특징: CDS
- 한정자: protein_id: **2**
translation: MMACC

아미노산 서열식별번호(SEQ ID): 2

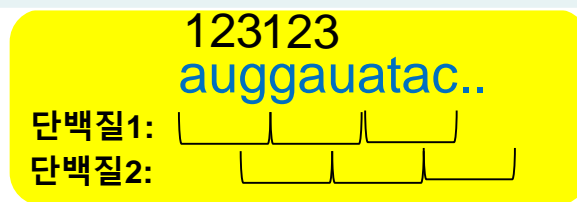
- 특징: SOURCE
- 한정자: MOL_TYPE: AA
- 한정자: ORGANISM: **Homo sapiens**
- 서열: MMACC

특징 기호 및 한정자

핵산 서열: "CDS" 특징 기호

- 번역된 서열을 변형시킬 수 있는 한정자:

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

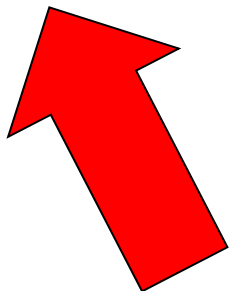


특징 기호 및 한정자

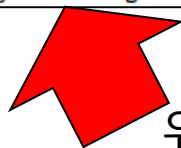
핵산 서열: "CDS" 특징 기호

- 예 - **SEQ ID NO:1**은 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 코돈**으로 시작한다



위치 **30 내지 32**에서의 코돈은 **셀레노시스테인(selenocysteine)**을 인코딩한다 (Sec)

- 이런 서열을 정확히 나타내기 위해 CDS 특징에 어떤 정보가 포함되어야 하나?

특징 기호 및 한정자

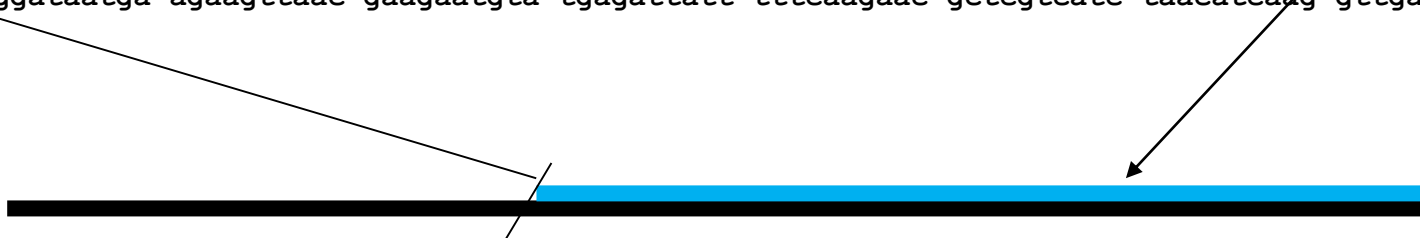
핵산 서열: "CDS" 특징 기호

- 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 tttcaagaac gctcgtcatc taacatcaag gttgacataa
```



특징 기호 및 한정자

핵산 서열: "CDS" 특징 기호

- 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 기호
- 특징 위치: <1..80

위치는 위치 78 내지 80에서 정지 코돈을 포함한다

“<”는 코딩 영역이 위치 1 이전에 시작됨을 나타낸다

특징 기호 및 한정자

핵산 서열: "CDS" 특징 기호

- 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 기호
- 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 기호 및 한정자

핵산 서열: "CDS" 특징 기호

- 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 기호
- 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 결정하는데 사용된다.

특징 기호 및 한정자

핵산 서열: "CDS" 특징 기호

- 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 기호
- 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 셀레노시스테인
(selenocysteine)에 넣을 것
이다

특징 기호 및 한정자

핵산 서열: "CDS" 특징 기호

- 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 기호
- 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 "translation" 한정자 "DNEEVNEECURLFFKNARHTTSRLT "

정지 코돈은 번역 한정자에
제시되지 않는다!

특징 기호 및 한정자

핵산 서열: "CDS" 특징 기호

- 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 기호
- 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 "translation" 한정자
- 번역(translation)된 별도의 단백질 서열

특징 기호 및 한정자

핵산 서열: "CDS" 특징 기호

- 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 *
```

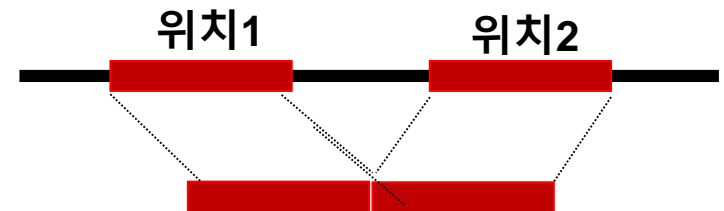
- "CDS" 특징 기호
- 특징 위치: <1..80
- 값 "3"을 갖는 "codon_start" 한정자
- 값 "3"을 갖는 "transl_table" 한정자
- 값 "(pos:30..32,aa:Sec)"을 갖는 "transl_except" 한정자
- 값 "DNEEVNEECURLFFKNARHTTSRLT"를 갖는 "translation" 한정자
- 번역을 위한 별도의 단백질 서열
- 번역된 단백질의 서열식별번호(SEQ ID)를 갖는 "protein_id" 한정자

특징 기호 및 한정자

hexan 서열: "CDS" 특징 기호

- "CDS" 특징 위치는 서열의 **불연속 세그먼트**를 단일 코딩 영역으로 연결하기 위해 "**join**" 위치 연산자를 사용할 수 있다

`join(위치1,위치2)`



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

`complement(위치)`

5' -accgtgcccgt-3'
3' -**tg**gcacgggca-5'

특징 기호 및 한정자

아미노산 서열: 일반적으로 사용되는 특징 기호

| 특징 기호 | 설명 | 필수 한정자? |
|----------------|-----------------------|---------------------------------------|
| SITE | 흥미로운 단일 아미노산 부위를 나타낸다 | 필수 NOTE 한정자는 부위를 설명 해야 한다 |
| REGION | 관심 영역을 나타낸다 | 없음; NOTE는 선택적이다 |
| BINDING | 화학기에 대한 결합 부위를 나타낸다 | 필수 NOTE 한정자는 화학기의 명칭 을 포함해야 한다 |
| UNSURE | 서열에서의 불확실성 영역을 설명한다 | 없음; NOTE는 선택적이다 |

특징 기호 및 한정자

아미노산 서열: modified amino acid(변형된 아미노산)

- " **modified amino acid**"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))

- " **modified amino acid** "은 가능하면 대응하는 **변형되지 않은 아미노산**으로 표시되어야 한다. 그렇지 않으면, 이는 "**X**"로 표시될 수 있다. 예를 들어,
 - "**hydroxylysine**"은 서열에서 "K"로 표시되어야 한다.(K에 -OH기가 붙음)
 - "**ornithine**"은 "X"로 표시되어야 한다. (ST.26, pgh. 29).
- 기호 "**X**"는 단지 하나의 잔기에 해당한다.

특징 기호 및 한정자

아미노산 서열: 변형된 아미노산

- 여러 특징 기호는 “modified amino acid”을 표기에 사용될 수 있다.

| 특징 기호 | 설명 |
|-----------------|--|
| SITE | ‘non post-translationally modified 아미노산’을 나타낸다 |
| MOD_RES | ‘post-translationally modified 아미노산’을 나타낸다 |
| CARBOHYD | 당화된(glycosylated) 아미노산을 나타낸다 |
| LIPID | 아미노산에 대한 지질(lipid)이 공유결합 된 것을 나타낸다 |

- 필수 NOTE 한정자는 변형을 설명하기 위한 값과 함께 상기 특징 기호 각각에 포함되어야 한다.

특징 위치 형식

위치 형식

모든 분자 유형에 대한 위치 설명자

- 위치 설명자는 서열에서 특징의 위치를 확인하는데 사용된다
- ST.26은 위치 설명자의 형식에 대한 필수 요구사항을 갖는다
- 다음 위치 설명자 형식은 핵산 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)

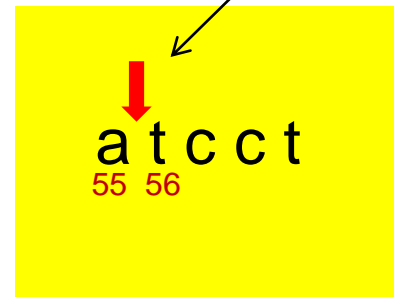
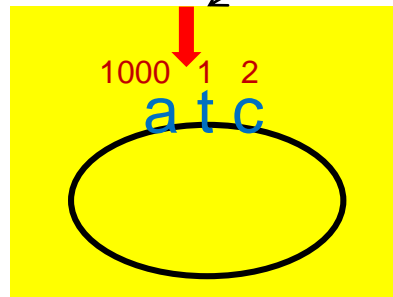
위치 형식

핵산 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^{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. |

WIPO 표준 ST.26, 단락 66(b)



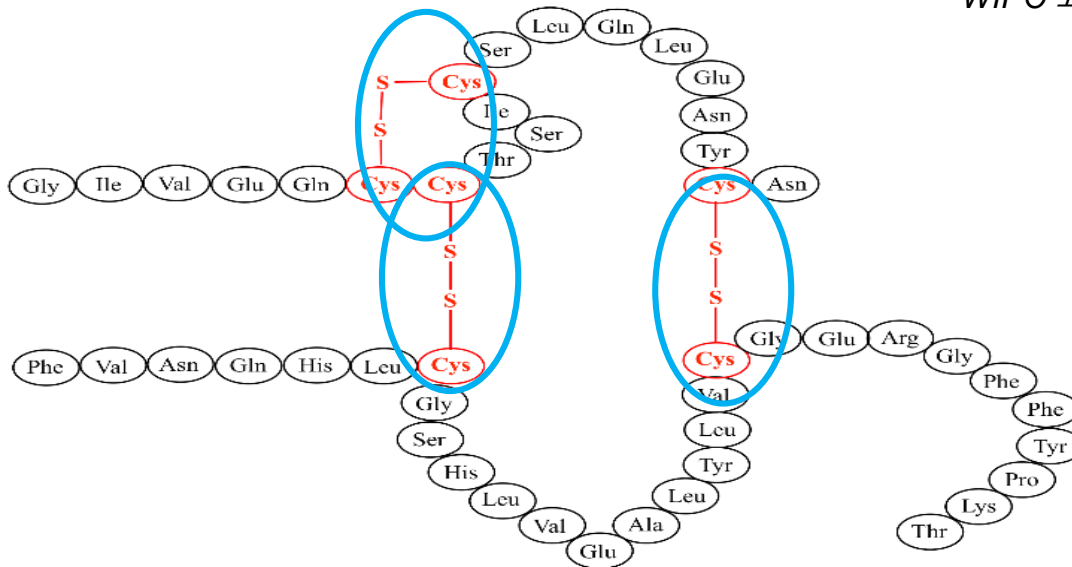
위치 형식

위치 설명자 – 아미노산 서열에 대한 특별한 경우

- 위치 설명자 $x..y$ 는 “**CROSSLNK**” 또는 “**DISULFID**” 특징 기호와 함께 사용되는 경우 **2개의 표시된 잔기 사이의 사슬 간 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

핵산 서열에 대한 위치 연산자

- 3개의 위치 연산자가 DNA 및 RNA 서열에서 사용할 수 있다: “**join**”, “**order**” 및 “**complement**”
- 연결 위치 연산자의 사용은 위치 설명자에 의해 설명된 핵산 잔기가 생물학적 과정에 의해 물리적으로 접촉되는 것을 의미한다 (ST.26, 단락 68)
- “**join**” 및 “**order**”를 사용하는 위치는 적어도 두 개의 심표로 구분된 위치 설명자들을 가져야 한다
- “**complement**”은 “**join**” 또는 “**order**”와 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

핵산 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(qualifier) 값(value) 형식

한정자 값

- 한정자 추가 정의 특징
- 한정자는 **한정자 명칭** 및 **종종 한정자 값**으로 구성된다

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

- 각각의 특징 기호는 상기 특징에 허용되는 한정자 목록을 가질 것이다. 일부 특징 기호는 필수 한정자를 갖는다.
- **특징 기호 및 이들의 허용되는 한정자**는 ST.26, 부록 I, 섹션 5 (핵산 서열) 및 섹션 7 (아미노산 서열)에 나열되어 있다.
- 설명과 함께 **핵산 서열에 대한 한정자**는 ST.26, 부록 I, 섹션 6 (핵산 서열) 및 섹션 8 (아미노산 서열)에 나열되어 있다.

한정자 값

- 예 - 특징 기호 “**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 I, 섹션 5.12)

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 선택**을 갖는 한정자;

2. 정의된 **값 형식**을 갖는 한정자;

3. 값이 **서열**인 한정자;

4. **값을 갖지 않는** 한정자;

5. "**자유 텍스트(free text)**" 값을 갖는 한정자

- "자유 텍스트" 한정자 값의 서브셋은 “언어 종속적(language dependent)”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 Sequence 프로그램에서 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"> 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 |

▼ SEQUENCE

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| tgcggatatt | attcataagc | atgggctct | ggaagtgccg | ctgacggccc | agaagggcgt |
| cttactgtta | caagaagagt | atcatgacgc | gcatcatagc | cacttagacc | cgcacttatg |
| gttggcgccg | caaaatatca | ttgcctctgt | gcaaggattg | gataaacata | ctgctgagtt |

180

한정자 값

형식 유형 – 정의된 값 형식

- 정의된 값 형식을 갖는 한정자

- 예: “anticodon” – 값의 형식이 존재한다

"(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: “**translation**” – 값은 아미노산 1-문자 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>MASTFPPWYRGCASTPSLKGLIMCTW</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 " **replace** " – 값은 단일 hexan 잔기이거나, 일련의 잔기들이거나, 비어 있을 수 있다.
- " **replace** "에 대한 **빈 값**은 해당 특징에 표시된 잔기의 결실(deletion)을 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><INSDQualifier value>a</INSDQualifier value> <INSDQualifier_value></INSDQualifier_value> - 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 Sequence는 값이 없는 한정자에 값을 추가할 수 없도록 할 것이다
- 이들 한정자는 비어 있는 “**INSDQualifier_value**” 요소를 가져서는 안된다

한정자 값

형식 유형 – “Free Text”

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

한정자 값

형식 유형 – “free text”

- " language-dependent free text(언어종속적 자유텍스트) " 값을 갖는 한정자는 번역을 필요로 할 수 있다
- Free text값 형식을 갖는 한정자가 언어 종속적인지의 여부를 어떻게 결정할 수 있는가?
 - ST.26, 부록 I, 섹션 6, 표 5는 언어 종속적 자유 텍스트 값 형식을 갖는 모든 핵산 서열 한정자를 열거한다
 - ST.26, 부록 I, 섹션 8, 표 6은 언어 종속적 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 | free text representing the proportion of a population carrying the feature expressed as a fraction |
| Example | <pre><INSDQualifier_value>23/108</INSDQualifier_value> <INSDQualifier_value>1 in 12</INSDQualifier_value> <INSDQualifier_value>0.85</INSDQualifier_value></pre> |
| 6.21. 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 | <pre><INSDQualifier_value>essential for recognition of cofactor </INSDQualifier_value></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 – “language-dependent free text ”

- " **language-dependent free text** " 한정자 값은 서열 목록 XML에서 2개의 언어로 제공될 수 있다: 영어 및 다른 하나의 비-영어 언어 (ST.26, 단락 87)
- 영어의 언어 종속적 한정자 값은 **INSDQualifier_value** 요소에 제공되어야 한다
- 영어가 아닌 언어의 언어 종속적 한정자 값은 **NonEnglishQualifier_value** 요소에 제공되어야 한다
- 언어 종속적 자유 텍스트 값 형식을 갖는 한정자에 대해서만 **NonEnglishQualifier_value** 요소가 허용된다 (ST.26, 단락 87(b))

한정자 값

형식 유형 – “language-dependent free text ”

```

<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 – “language-dependent free text ”

- 서열 목록 XML이 비-영어 한정자 값을 포함하는 경우:

1. 서열 목록 XML의 루트 요소는 적절한 2-문자 언어 코드 약어가 있는 “nonEnglishFreeTextLanguageCode” 속성을 포함해야 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 – “language-dependent free text”

■ 서열 목록 XML이 비-영어 한정자 값을 포함하는 경우:

1. 서열 목록 XML의 루트 요소는 적절한 2-문자 언어 코드 약어가 있는 “nonEnglishFreeTextLanguageCode” 속성을 포함해야 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 포함한 정보는 동일해야 한다 (ST.26 단락 87(c)).

한정자 값

형식 유형 – “language-dependent free text ”

- 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 한다”고 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 – “language-dependent free text ”

- INSDQualifier “id” 속성:
 - WIPO Sequence가 **XLIFF 파일**로 내보내기를 하기 위해 번역을 요구 할 수 있는 한정자 값을 고유하게 확인한다;
 - 선택적;
 - 언어 종속적 한정자에 대해서만 허용됨;
 - 서열 목록에서 고유해야 한다;
 - WIPO Sequence에 의해 생성된 XML에 자동으로 추가된다.

한정자 값

형식 유형 – “language-dependent free text ”

■ “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 종속적 자유 텍스트 한정자가 준비된 단일 원본 언어에 대한 언어 코드..."로 정의한다.
- 이러한 속성은 선택적이다

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

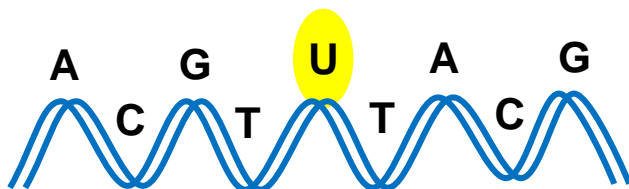
특수 상황

- 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**

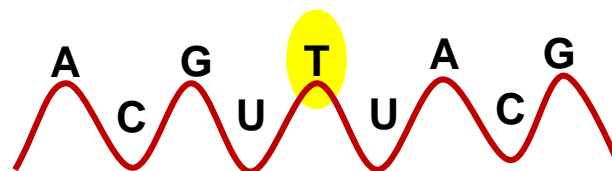
특수 상황

DNA에서 우라실 및 RNA에서 티민

- 서열이 우라실 핵염기를 갖는 DNA 골격 또는 티민 핵염기를 갖는 RNA 골격을 갖는 경우는 “변형된 핵산”으로 설명된다



DNA (acgt)



RNA (acgu)

ST.26 단락 14적용:

"14. 기호 "t"는 DNA에서 티민 및 RNA에서 우라실로 해석될 것이다. DNA의 우라실(u) 또는 RNA의 티민(t)은 **변형된 핵산(modified nucleotide)**으로 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 핵산로 주석을 달아야 한다.
- ST.26 단락 19는 DNA의 우라실 또는 RNA의 티민은 변형된 핵산으로 간주되며 "t"와 같은 서열로 표현되어야 하고, 특징 기호 **"modified_base"**, 한정자 값으로서 **"OTHER"**를 갖는 한정자 **"mod_base"** 및 한정자 값으로서 "uracil" 또는 "thymine"을 각각 갖는 한정자 **"note"**를 사용하여 특징 표에 추가로 설명되어야 한다고 명시한다.

특수 상황

DNA 및 RNA 서열

본 출원은 다음의 RNA 서열을 개시한다:

5' -cguccacgugtccgaggua-3'

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

cgtcccacgtgtccgaggta

- ✓ 위치 "12"를 갖는 특징 기호 "modified_base"
- ✓ 값 "OTHER"를 갖는 한정자 "mod_base"
- ✓ 값 "thymine"을 갖는 한정자 "note"

특수 상황

DNA 및 RNA 서열

```

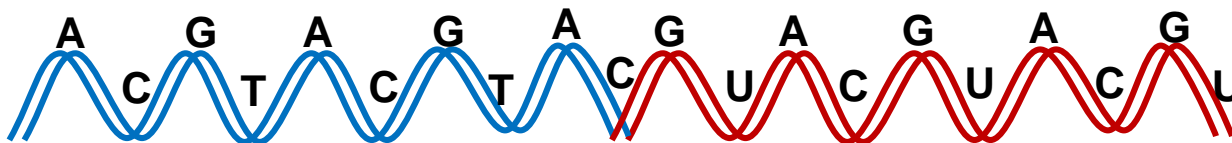
<SequenceData sequenceIDNumber="3">
  <INSDSeq>
    <INSDSeq_length>20</INSDSeq_length>
    <INSDSeq_moltype>RNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..20</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>cggtcccacgtgtcgcgaggt</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: DNA 세그먼트

RNA 세그먼트

"55. 하나 이상의 핵산의 DNA 및 RNA 세그먼트 둘 모두를 포함하는 핵산 서열의 경우, 분자 유형은 **DNA**로 표시되어야 한다.

- 특징 기호 **"source"**
 - 필수 한정자 **"organism"**: **"synthetic construct"**
 - 필수 한정자 **"mol_type"**: **"other DNA"**
- 특징 기호 **"misc_feature"**
 - 한정자 **"note"**로 추가로 설명: 결합된 DNA/RNA 분자의 각각의 DNA 및 RNA 세그먼트는 세그먼트가 DNA이거나 RNA 인지를 나타냄.

특수 상황

DNA/RNA 하이브리드 서열

본 출원은 다음의 DNA/RNA 하이브리드 서열을 개시한다:

5' -**ACCTGC**cgucccacguguccgaggua**GCATTA**-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. 생물체 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 **DNA**로 표시되어야 한다. 결합된 DNA/RNA 분자는 특징 기호 "source" 및 값 "synthetic construct"을 갖는 필수 한정자 "organism" 및 값 "other DNA"를 갖는 필수 한정자 "mol_type"을 사용하여 특징 표에 추가로 기술되어야 한다"고 명시한다.

- ✓ 분자 유형 = "DNA"
- ✓ 한정자 생물체 명칭 = "synthetic construct"
- ✓ 한정자 mol_type = "other 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(u)** 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 기호 "**misc_feature**" 및 한정자 "**note**"로 추가로 기술되어야 한다" 고 명시한다.



3개의 세그먼트 = 3개의 "misc_feature" 특징 기호

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 기호
값 "DNA"를 갖는 한정자 "note"
- ✓ 세그먼트 2, 잔기 7 내지 26:
위치 "7..26"을 갖는 "misc_feature" 특징 기호
값 "RNA"를 갖는 한정자 "note"
- ✓ 세그먼트 3, 잔기 27 내지 32
위치 "27..32"를 갖는 "misc_feature" 특징 기호
값 "DNA"를 갖는 한정자 "note"

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>

```


유사체, D-아미노산, 및 분지된 서열

새로운 분자 유형

핵산 유사체(Nucleotide Analogs)

- 하나 이상의 핵산 유사체를 포함하는 핵산 서열은 ST.26 규칙에 적용된다
- 핵산 유사체는 ST.26, 단락 3(g)(2) 아래의 “핵산”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 포함하는 핵산에서 핵염기의 배열을 모방하는 핵염기의 배열을 발생시키며, 여기서 핵산 유사체는 상보적 핵산과 염기쌍을 형성할 수 있다"

- 일반적인 핵산 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

핵산 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 제시된 방향으로 표시되어야 한다
- ✓ 전체 서열에 "modified_base" 특징 기호, 값 "OTHER"를 갖는 "mod_base" 한정자, 및 "glycol nucleic acid"와 같은 변형된 핵산의 완전하고 약칭되지 않은 명칭을 포함하는 note 한정자가 주석으로 달려야 한다

새로운 분자 유형

핵산 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 포함하는 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 표시 될 수 있는 임의의 아미노산을 의미한다. 상기 아미노산은 특히 D-아미노산 및 변형된 또는 합성 측쇄를 포함하는 아미노산을 포함한다."

- D-아미노산은 가능한 경우 상응하는 변형되지 않은 L 아미노산 기호로서 서열 내에 표시되어야 한다
- 특징 표에서 변형된 아미노산으로 설명되어야 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 **1**의 알라닌, 위치 **2**의 글루탐산 및 위치 **6**의 메티오닌에는 각각 "SITE" 특징 기호 및 "NOTE" 한정자가 주석으로 달려야 하며, 이 "NOTE" 한정자에는 상응하는 아미노산의 완전하고 약칭되지 않은 명칭이 있어야 한다

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

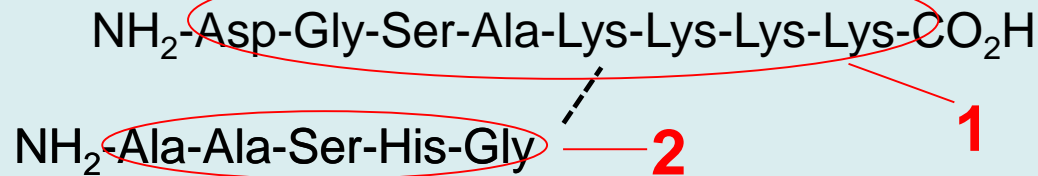
분지된 서열

- 분지된 핵산 서열 및 분지된 아미노산 서열은 **ST.26 규칙에 적용된다**
- 분지된 서열의 선형 영역은 **10개 이상의 구체적으로 정의된 핵산 또는 4개 이상의 구체적으로 정의된 아미노산**을 포함하는 경우 서열 목록에 포함되어야 한다. (WIPO 표준 ST.26, 단락 7)
- 최소 길이 요건을 충족시키는 분지된 서열의 **각각의 선형 영역은 그 자신의 서열식별번호(SEQ ID)를 갖는 별도의 서열**로서 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 카르복시 말단과 리신의 측쇄 사이의 아미드 결합을 나타낸다

- ✓ 둘 모두의 선형 영역은 ≥ 4 개의 구체적으로 정의된 아미노산을 포함하므로, 둘 다 서열 목록에 포함될 필요가 있다
- ✓ 각각의 선형 영역은 그 자신의 서열식별번호(SEQ ID)를 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 (variant sequence, 변이 서열)

- ST.26 단락 3(m)은 “**변이 서열**”을 다음과 같이 정의한다.

“1차 서열과 관련하여 하나 이상의 차이를 포함하는 핵산염기 또는 아미노산 서열을 의미한다. 이러한 차이점에는 대체(alternative) 잔기(단락 15 및 27 참조), 변형 잔기(단락 3(g), 3(h), 16 및 29 참조), 결실, 삽입 및 치환이 포함될 수 있다.(단락 93-95 참조)”

- 변이 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: 단락 7에 포함되고 각각 잔기 나열에 의해 표시되는 1차 서열 및 이 서열의 임의의 변이에 대해 각각의 고유 서열식별번호를 부여하여야 한다.



각각의 변이체가 개별적으로 열거된 경우, 각각의 변이체는 그 자신의 서열식별번호(SEQ ID)를 가져야 한다!

서열 변이체

단락 93

본 특허 출원은 다음의 다중 서열 정렬을 갖는 도면을 포함한다:

컨센서스

| | | | | | | | | | |
|-------------------------|----------|--------|-----------|-------|----------|------------|-------|----------|-----|
| LEG | n | EQFINA | ak | IIRHP | k | nrk | TLNND | m | LIK |
| Homo sapiens | | | | | | | | | |
| Pongo abelii | | | | | | | | | |
| Pan paniscus | | | | | | | | | |
| Rhinopithecus bieti | | | | | | | | | |
| Rhinopithecus roxellana | | | | | | | | | |

소문자는 정렬된 서열 중 우세한 아미노산 잔기를 나타낸다.

- ✓ 6개의 열거된 서열 각각은 그 자신의 서열식별번호(SEQ ID)를 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

- 변이체에 대한 주석을 작성하는 알맞은 특징 기호는 분자 유형 및 변이의 특성에 좌우된다:

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

WIPO 표준 ST.26, 단락 96

서열 변이체

가장 제한적인 다중아미노산 또는 다중핵산염기 기호

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

이것은 무슨 의미인가?

예를 들어:

핵산 서열에서, 위치가 **"a" 또는 "c"**일 수 있는 경우, **"n"** 대신에 다중핵산염기 기호 **"m"**을 사용한다.

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

"n" 및 "X"는 내정값(default value)을 갖고 있기 때문에 "n" 또는 "X"가 내정값(default value)이 아닌 **다른 값에 사용되는 경우에는 주석이 필요하다**는 것을 기억해야 한다.

서열 변이체

단락 94

단락 94: 하나 이상의 위치에 나열된 대체 잔기를 갖는 단일 서열으로써 공개된 임의의 변이 서열은 서열목록에 포함되어야 하고 단일 서열로 표현되어야 하며, 여기서 나열된 대체 잔기는 가장 제한적인 다중아미노산 또는 다중핵산염기 기호로 표현된다(단락 15 및 27 참조).



변이체가 별도로 열거되지 않고, 단순히 1차 서열 내의 가변 잔기로 표시되는 경우, 이들은 **별도의 서열식별번호(SEQ ID)**를 가질 필요가 없다!

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 다중아미노산 기호 "J"로 표시되어야 한다
- ✓ [Ser 또는 Thr]은 X가 세린 또는 트레오닌이라는 것을 나타내는 한정자 참조와 함께 특징 기호 "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: 서열목록의 1차 서열에서 결실(deletion(s)), 삽입(insertion(s)) 또는 치환(substitution(s))에 대한 언급을 통해서만 개시된 임의의 변이 서열은 서열목록에 포함되어야 한다. 서열목록에 포함되는 경우, 이러한 변형 서열은:

- (a) 1차 서열의 주석으로 표현될 수 있으며, 여기서 단일 위치 또는 다수의 별개의 위치에서 변형을 포함하고, 이러한 변형의 발생은 독립적이다;
- (b) 별도의 서열로 표현되어야 하고, 고유한 서열식별번호(SEQ ID)가 부여되어야 하며, 여기서 다수의 다른 위치에서 변이를 포함하고 이러한 변이의 발생은 상호 의존적이다;
- (c) 별도의 서열로 표현되어야 하고, 고유한 서열식별번호(SEQ ID)가 부여되어야 하며, 여기에는 1000개를 초과하는 잔기를 포함하는 삽입 또는 치환된 서열이 포함된다(단락 86).



“결실, 삽입 또는 치환에 대한 언급”은 변이체가 산문으로 개시되어 있다는 것을 의미한다.

서열 변이체

단락 95(a)

본 출원은 다음과 같은 개시를 포함한다:

펩티드 단편 1: Gly-Leu-Pro-**Xaa**-Arg-Ile-Cys

여기서, Xaa는 임의의 아미노산이다

* * *

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

* * *

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

1차 서열, Gly-Leu-Pro-Xaa-Arg-Ile-Cys, 은 단일 위치에 변이(들)을 포함하며, 해당 변이의 발생은 독립적이다

서열 변이체

단락 95(a)

- ✓ 1차 서열은 서열 목록에 포함되어야 하며, 변이체는 1차 서열의 주석에 의해 표시될 수 있다
- ✓ 가장 포괄적인 구현예는 서열 목록에 포함된 버전이어야 하며 – 이 예에서, 이는 **Xaa**가 " **any amino acid** "인 버전이다
- ✓ 서열은 **GLPXRIC**로 표시되어야 하며, 위치 4에서의 특징 기호 "VARIANT" 및 Xaa가 "any amino acid" 임을 나타내는 한정자 "note"를 필요로 한다
- ✓ 필요하지는 않지만, 다음 **3개의 변이체가 서열 목록에 별도의 서열로 포함** 되는 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 변이 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 포함하고, 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 aatgncccacgaatgncac로 표시되어야 한다
- ✓ 특징 기호의 부재하에서 "n"은 "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)

서열 변이체

단락 95(c)

본 출원은 다음과 같은 개시를 포함한다:

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

여기서, **Xaa**는 **Gly** 또는 서열의 삽입이고

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

하나의 변이 서열은 1000개 잔기를 초과하는 삽입된 또는 치환된 서열을 포함한다

서열 변이체

단락 95(c)

- ✓ Xaa = 1000개가 넘는 잔기의 삽입인 변이체가 그 자신의 서열식별번호(SEQ ID)를 갖는 별개의 서열로서 서열 목록에 포함되어야 한다.
- ✓ Xaa = Gly인 변이체 역시 그 자신의 서열식별번호(SEQ ID)를 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:

“하나 이상의 위치에 나열된 대체 잔기를 갖는 단일 서열로써 공개된 임의의 변이 서열은 서열목록에 포함되어야 하고 단일 서열로 표현되어야 하며, 여기서 나열된 대체 잔기는 가장 제한적인 다중아미노산 또는 다중핵산염기 기호로 표현된다.”

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

“단락 7에 포함되고 각각 잔기 나열에 의해 표시되는 1차 서열 및 이 서열의 임의의 변이에 대해 각각의 고유 서열식별번호를 부여하여야 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 포함한다:

... 특정 구현예에서, 서열 1의 위치 7의 **발린(Val)**은 **알라닌(Ala)**으로 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 포함한다:

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

A: 단락 95 (a): “서열목록의 1차 서열에서 결실(deletion(s)), 삽입(insertion(s)) 또는 치환(substitution(s))에 대한 언급을 통해서만 개시된 임의의 변이 서열은 서열목록에 포함되어야 한다. 서열목록에 포함되는 경우, 이러한 변형 서열은:

(a) 1차 서열의 주석으로 표현될 수 있으며, 여기서 단일 위치 또는 다수의 별개의 위치에서 변형을 포함하고, 이러한 변형의 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 포함한다:

... 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 포함한다:

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

A: 단락 95 (b): “서열목록의 1차 서열에서 결실(deletion(s)), 삽입(insertion(s)) 또는 치환(substitution(s))에 대한 언급을 통해서만 개시된 임의의 변이 서열은 서열목록에 포함되어야 한다. 서열목록에 포함되는 경우, 이러한 변형 서열은:

(b) 별도의 서열로 표현되어야 하고, 고유한 서열식별번호(SEQ ID)가 부여되어야 하며, 여기서 다수의 다른 위치에서 변이를 포함하고 이러한 변이의 발생은 상호 의존적이다”

공금한 점