



WIPO

WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

WIPO Standard ST.26 **INTRODUCTION**

Webinar training

To be covered today

- What is a sequence listing?
- INSDC Sequence Databases
- Why a new Standard?
- Benefits of WIPO ST.26
- ST.25 vs. ST.26 – what are the differences?
- Transition to ST.26
- XML Basics
- ST.26 Basics – parts of a sequence listing
- Contents of WIPO ST.26
- Introduction to WIPO Sequence

What is a sequence listing?

A sequence listing...

- Contains nucleotide and/or amino acid sequences disclosed in a patent application and forms part of the description
- Includes descriptive information about each sequence known as annotations
- Conforms to the requirements of the relevant WIPO Standard (ST.25 or ST.26)
- Allows for the sequence data of an invention to be searchable:
 - Inside an IP Office
 - In publically available databases (INSDC databases)

Background - INSDC

- INSDC: International Nucleotide Sequence Database Collaboration:
 - DDBJ: DNA Databank of Japan
 - EMBL-EBI: The European Bioinformatics Institute
 - NCBI: National Center for Biotechnology Information (GenBank)

- IP Offices that submit published/issued application sequence data to INSDC databases include:
 - European Patent Office
 - Japanese Patent Office
 - Korean Intellectual Property Office
 - United States Patent and Trademark Office

- INSDC databases are publicly searchable

Why a new Standard?

- Currently sequence listings are filed compliant with WIPO ST.25

However...

- ST.25 format is not compliant with INSDC requirements, so data is lost when entered into public databases
- ST.25 rules are not clear, and IP Offices worldwide interpret and enforce the rules differently
- Sequence types that are common today are not covered by ST.25 rules (nucleotide analogs, D-amino acids, branched sequences) and therefore are not present in searchable databases
- Data is unstructured – ST.25 format is difficult to use for automated validation and data exchange

Transition to WIPO ST.26

- CWS/5 agreed to a transition date of **January 1, 2022**, referred to as the '**big-bang**' date
- **All** Intellectual Property Offices (IPOs) will transition simultaneously at the international (PCT), national, and regional levels
- The international filing date will be the reference date that determines if an application falls under ST.25 or ST.26 sequence rules, **NOT** the priority date
- Note: ST.25 will remain effective for applications with a filing date prior to January 1, 2022

WIPO ST.26 Benefits (1)

- Acceptance of a single sequence listing worldwide*
- Standard serves as guidance to ensure agreement amongst IP Offices on application of sequence rules
- Clarifies what sequence disclosures are required or permitted to be included in a sequence listing, and how these sequences must be represented
- Enhanced submission quality due to the structure of XML sequence listings
- Increased automation of data validation and streamlined processing by IP Offices

* except for required translations of language dependent free text qualifiers into the language of filing for certain IP Offices, which may require replacement sequence listings

WIPO ST.26 Benefits (2)

- Data compatibility with INSDC database provider requirements - DDBJ, EBI, and NCBI – Sequence annotations (feature keys and qualifiers) will be included in publically searchable databases

- Standardization of:
 - Feature annotations
 - Feature locations
 - Qualifiers and qualifier values
 - Sequence variant presentation

- Requirement for inclusion of additional types of sequences (nucleotide analogs, D-amino acids, branched sequences) means more sequence data will be searchable

WIPO ST.25 versus ST.26

ST.25	ST.26
ASCII .txt with numeric identifiers	XML with elements and attributes
<p><u>Not required</u> to include:</p> <ul style="list-style-type: none"> - D-amino acids - Linear portions of branched sequences - Nucleotide analogs 	<p><u>Must</u> include:</p> <ul style="list-style-type: none"> - D-amino acids - Linear portions of branched sequences - Nucleotide analogs
<p>Annotation of sequences:</p> <ul style="list-style-type: none"> - Feature keys only 	<p>Annotation of sequences:</p> <ul style="list-style-type: none"> - Feature keys and qualifiers
<p><u>Permitted</u> to include sequences:</p> <ul style="list-style-type: none"> - < 10 specifically defined nucleotides - < 4 specifically defined amino acids 	<p><u>Prohibited</u> sequences:</p> <ul style="list-style-type: none"> - < 10 specifically defined nucleotides - < 4 specifically defined amino acids

WIPO ST.25 versus ST.26

...for General information

ST.25	ST.26
ALL priority application information may be included	ONLY the earliest priority application can be included
ALL applicant and inventor names may be included	ONLY one applicant AND optionally ONE inventor may be included
One invention title permitted	Multiple invention titles permitted, each one in a different language
Applicant/inventor names and invention titles must be in basic Latin characters	Applicant/inventor names may be included using any valid Unicode character along with a basic Latin translation or transliteration

WIPO ST.25 versus ST.26

...for Sequence data (1)

ST.25	ST.26
Sequences identified as DNA, RNA, or PRT only	Sequences identified as DNA, RNA, or AA along with a mandatory mol_type qualifier to further describe the molecule
Organism names: <ul style="list-style-type: none"> - Latin genus/species - Virus name - “artificial sequence” - “unknown” 	Organisms names: <ul style="list-style-type: none"> - Latin genus/species - Virus name - “synthetic construct” - “unidentified”
“u” represents uracil in nucleotide sequences	“t” represents uracil in RNA sequences and thymine in DNA sequences
Amino acid sequences represented by three letter abbreviations	Amino acid sequences represented by one letter abbreviations

WIPO ST.25 versus ST.26

...for Sequence data (2)

ST.25	ST.26
“n” and “Xaa” variables must have a definition provided in a feature	Default value assumed for “n” and “X” variables with no definition
Feature location format not clearly defined	Strictly defined feature location formats; permits use of “<” and “>” in all sequence types, and “^”, “join”, “order”, and “complement” in nucleotide sequences
“Mixed mode” sequences permitted – nucleotide sequence with amino acid translation shown below	NO “mixed mode”; nucleotide translations are included in “translation” qualifiers only

WIPO ST.26

...what must be included?

- Nucleotide sequences:
 - 10 or more “specifically defined”* and “enumerated”* residues
 - include sequences with nucleotide analogs such as peptide nucleic acids (PNAs) and glycol nucleic acids (GNAs)
- Amino acid sequences:
 - 4 or more “specifically defined” and “enumerated” residues
 - include sequences with D-amino acids
 - linear regions of branched sequences are required to be included in a sequence listing

**See ST.26 Annex VI for definitions*

WIPO ST.26

...what must be included?

- What is a “specifically defined” nucleotide or amino acid?
- “specifically defined” means any nucleotide other than those represented by the symbol “n” and any amino acid other than those represented by the symbol “X”, listed in Annex I
- only “specifically defined” residues count towards the minimum length requirement:
 - 10 or more specifically defined nucleotides; or,
 - 4 or more specifically defined amino acids

5'- anctggcaan – 3'

only 8 specifically defined nucleotides; must not be included in a sequence listing

5'- agctggcaat – 3'

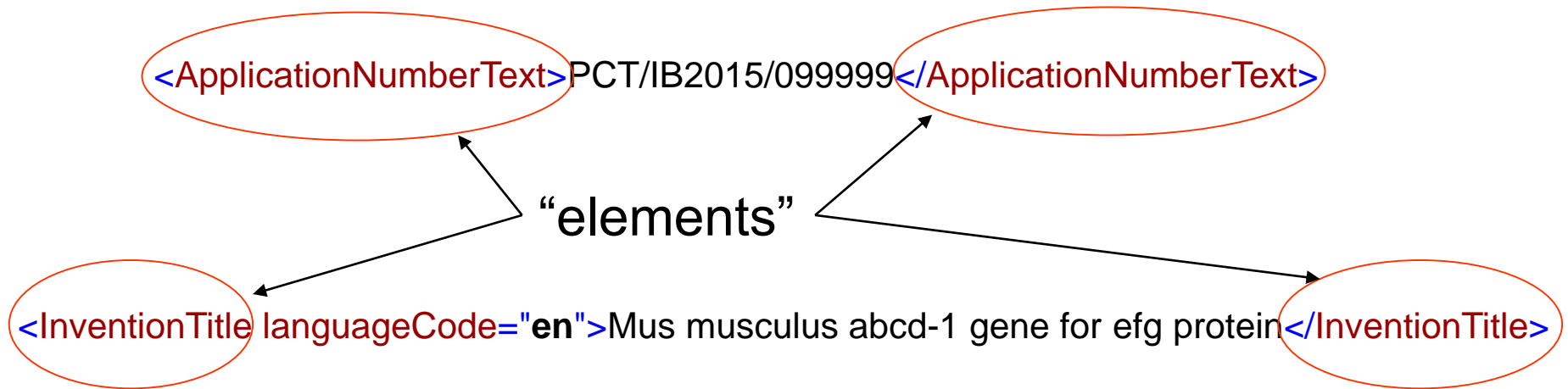
ten specifically defined nucleotides; must be included in a sequence listing

XML Basics

- XML = eXtensible Markup Language
- Information is ‘tagged’ using descriptive elements and attributes
- Standardized means of data exchange that is both human and machine readable
- DTD = Document Type Definition - defines the structure and the legal elements and attributes of an XML document

XML Basics

Elements, attributes, and values (1)



XML Basics

Elements, attributes, and values (2)

`<ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>`

“element values”

`<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>`

XML Basics

Elements, attributes, and values (3)

```
<ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
```

“attribute”

```
<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
```

“attribute value”


XML Basics


Reserved characters

Must be replaced by their predefined entities in an element value.

Reserved Character	Predefined Entities
<	<
>	>
&	&
“	"
'	'

Example: Desired feature location is “<50..62”

`<INSDFeature_location><50..62</INSDFeature_location>` 

`<INSDFeature_location><50..62</INSDFeature_location>` 

WIPO ST.26: XML document

- Must be provided as one XML 1.0 format file
- Must validate against the WIPO ST.26 DTD (Annex II) and business rules derived from the content of the Standard
- Must be encoded using Unicode UTF-8
- Structure of the ST.26 sequence listing:
 - XML declaration:

```
<?xml version="1.0" encoding="UTF-8"?>
```
 - Document type (DOCTYPE) declaration:

```
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD  
Sequence Listing 1.3//EN" "ST26SequenceListing_V1_3.dtd">
```
 - Root element
 - General information part
 - Sequence data part

WIPO ST.26:Example

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN" "ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="for WIPO Training" softwareName="WIPO Sequence" softwareVersion="1.1.0-beta4" productionDate="2021-03-24">
  <ApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
    <FilingDate>2015-01-31</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>ABC123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
    <FilingDate>2014-01-31</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantName>
  <InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
  <SequenceTotalQuantity>1</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1">
    <INSDSeq>
      <INSDSeq_length>52</INSDSeq_length>
      <INSDSeq_moltype>DNA</INSDSeq_moltype>
      <INSDSeq_division>PAT</INSDSeq_division>
      <INSDSeq_feature-table>
        <INSDFeature>
          <INSDFeature_key>source</INSDFeature_key>
          <INSDFeature_location>1..52</INSDFeature_location>
          <INSDFeature_qual>
            <INSDQualifier>
              <INSDQualifier_name>mol_type</INSDQualifier_name>
              <INSDQualifier_value>genomic DNA</INSDQualifier_value>
            </INSDQualifier>
            <INSDQualifier id="q2">
              <INSDQualifier_name>organism</INSDQualifier_name>
              <INSDQualifier_value>Mus musculus</INSDQualifier_value>
            </INSDQualifier>
          </INSDFeature_qual>
        </INSDFeature>
      </INSDSeq_feature-table>
      <INSDSeq_sequence>atgaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
    </INSDSeq>
  </SequenceData>
</ST26SequenceListing>

```

WIPO ST.26: Components (1)

Line 1 – the XML declaration
(ST.26 paragraph 39(a))

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN" "ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="for WIPO Training" softwareName="WIPO Sequence"
softwareVersion="1.1.0-beta4" productionDate="2021-03-24">
  <ApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
    <FilingDate>2015-01-31</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>ABC123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
    <FilingDate>2014-01-31</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantName>
  <InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
  <SequenceTotalQuantity>1</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1">
```

WIPO ST.26: Components (2)

Line 2 – the document type (DOCTYPE) declaration
(ST.26 paragraph 39(b))

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN" "ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="for WIPO Training" softwareName="WIPO Sequence"
softwareVersion="1.1.0-beta4" productionDate="2021-03-24">
  <ApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
    <FilingDate>2015-01-31</FilingDate>
  </ApplicationIdentification>
  <ApplicantFileReference>ABC123</ApplicantFileReference>
  <EarliestPriorityApplicationIdentification>
    <IPOfficeCode>IB</IPOfficeCode>
    <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
    <FilingDate>2014-01-31</FilingDate>
  </EarliestPriorityApplicationIdentification>
  <ApplicantName languageCode="en">Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantName>
  <InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
  <SequenceTotalQuantity>1</SequenceTotalQuantity>
  <SequenceData sequenceIDNumber="1">

```

WIPO ST.26: Components (3)

Line 3 – the root element
(ST.26 paragraph 43)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN" "ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="for WIPO Training" softwareName="WIPO Sequence"
softwareVersion="1.1.0-beta4" productionDate="2021-03-24">
```

```
<ApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
  <FilingDate>2015-01-31</FilingDate>
</ApplicationIdentification>
<ApplicantFileReference>ABC123</ApplicantFileReference>
<EarliestPriorityApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
  <FilingDate>2014-01-31</FilingDate>
</EarliestPriorityApplicationIdentification>
<ApplicantName languageCode="en">Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantName>
<InventionTitle languageCode="en">Mus musculus abcd-1 gene for efg protein</InventionTitle>
<SequenceTotalQuantity>1</SequenceTotalQuantity>
<SequenceData sequenceIDNumber="1">
```

**General
Information:**

(ST.26 paragraphs. 38(a),
45-49)

WIPO ST.26: Components (4)

```

<SequenceData sequenceIDNumber="1">
  <INSDSeq>
    <INSDSeq_length>52</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..52</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>genomic DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q2">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>Mus musculus</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>atgaaattaaaacataaaaaggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
</ST26SequenceListing>

```

Sequence data

(ST.26 paragraphs. 38(b),
50-100)

WIPO ST.26: General information (1)

- The “Application Identification” section
 - application number, filing date, and IP office code are mandatory if known;
 - otherwise, just the applicant file reference is sufficient

```
<ApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
  <FilingDate>2015-01-31</FilingDate>
</ApplicationIdentification>
<ApplicantFileReference>ABC123</ApplicantFileReference>
<EarliestPriorityApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
  <FilingDate>2014-01-31</FilingDate>
</EarliestPriorityApplicationIdentification>
```

WIPO ST.26: General information (2)

- The “Priority Application” section
 - only one priority application can be included in the sequence listing, and it must be the earliest priority application;
 - mandatory where priority is claimed

```
<ApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2015/099999</ApplicationNumberText>
  <FilingDate>2015-01-31</FilingDate>
</ApplicationIdentification>
<ApplicantFileReference>ABC123</ApplicantFileReference>
<EarliestPriorityApplicationIdentification>
  <IPOfficeCode>IB</IPOfficeCode>
  <ApplicationNumberText>PCT/IB2014/111111</ApplicationNumberText>
  <FilingDate>2014-01-31</FilingDate>
</EarliestPriorityApplicationIdentification>
```

WIPO ST.26: General information (3)

- The “Applicant and Inventor Name” section
 - only one applicant name and one inventor name may be included in the sequence listing, and they must be the “primary” applicant and inventor;
 - applicant name is mandatory; inventor name is optional;
 - a language code for applicant and inventor names is mandatory;
 - if the applicant and/or inventor name contains non-Unicode Basic Latin characters, then a transliteration or translation into Basic Latin characters must be included

```

<ApplicantName languageCode="ja">出願製薬株式会社</ApplicantName>
<ApplicantNameLatin>Shutsugan Pharmaceuticals Kabushiki Kaisha</ApplicantNameLatin>
<InventorName languageCode="ja">特許 太郎</InventorName>
<InventorNameLatin>Taro Tokkyo</InventorNameLatin>

```

WIPO ST.26: General Information (4)

- The “Invention Title” section
 - at least one invention title in the language of filing is mandatory;
 - additional titles in other languages may be included;
 - a language code is mandatory for each title;

```
<InventionTitle languageCode="en">My spectacular invention</InventionTitle>  
<InventionTitle languageCode="de">Meine spektakuläre Erfindung</InventionTitle>  
<InventionTitle languageCode="lv">Manš iespaidīgais izgudrojums</InventionTitle>  
<InventionTitle languageCode="ru">Мое зрелищное изобретение</InventionTitle>
```

WIPO ST.26: General Information (5)

- The “Sequence Total Quantity” element
 - mandatory;
 - the total must include skipped sequences

`<SequenceTotalQuantity>6</SequenceTotalQuantity>`

WIPO ST.26: Sequences (1)

- Mandatory elements

```

<SequenceData sequenceIDNumber="1">
  <INSDSeq>
    <INSDSeq_length>52</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..52</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>genomic DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q2">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>Mus musculus</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>atgaaattaaaacataaaaaggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

The "Sequence Identification Number" or "SEQ ID NO:"

The sequence length

The molecule type (DNA, RNA, or AA)

The division (always "PAT")

WIPO ST.26: Sequences (2)

- The mandatory “source” or “SOURCE” feature

```

<SequenceData sequenceIDNumber="1">
  <INSDSeq>
    <INSDSeq_length>52</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..52</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>genomic DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q2">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>Mus musculus</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>atgaaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

One and only one “source” feature is required per sequence

The “source” location must span the entire sequence

Two mandatory qualifiers: “mol_type” and “organism”

WIPO ST.26: Sequences (3)

ST.26 Sequence Data:

```

<SequenceData sequenceIDNumber="1">
  <INSDSeq>
    <INSDSeq_length>52</INSDSeq_length>
    <INSDSeq_moltype>DNA</INSDSeq_moltype>
    <INSDSeq_division>PAT</INSDSeq_division>
    <INSDSeq_feature-table>
      <INSDFeature>
        <INSDFeature_key>source</INSDFeature_key>
        <INSDFeature_location>1..52</INSDFeature_location>
        <INSDFeature_qual>
          <INSDQualifier>
            <INSDQualifier_name>mol_type</INSDQualifier_name>
            <INSDQualifier_value>genomic DNA</INSDQualifier_value>
          </INSDQualifier>
          <INSDQualifier id="q2">
            <INSDQualifier_name>organism</INSDQualifier_name>
            <INSDQualifier_value>Mus musculus</INSDQualifier_value>
          </INSDQualifier>
        </INSDFeature_qual>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>atgaaattaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

WIPO ST.26: Sequences (4)

- The mandatory “mol_type” and “MOL_TYPE” qualifiers – value choices:

DNA

genomic DNA

other DNA

unassigned DNA

RNA

genomic RNA

mRNA

tRNA

rRNA

other RNA

transcribed RNA

viral cRNA

unassigned RNA

AA

protein

WIPO ST.26: Sequences (5)

- The mandatory “organism” and “ORGANISM” qualifiers – value choices:
 - Latin genus and species name e.g., “Mus musculus”
 - Genus name followed by “sp.” e.g., “Mus sp.”
 - Virus name e.g., “Torque teno virus 1”
 - “unidentified”
 - “synthetic construct”
- Common names, such as “mouse,” must not be used as the organism name. If desired, common names can be included in the sequence listing in a note qualifier.

WIPO ST.26: Sequences (6)

Feature Keys and Qualifiers

In addition to the mandatory “source” or “SOURCE” feature, applicants can add multiple optional features to further describe the sequence

- different feature keys for nucleotide sequences and amino acid sequences;
- each feature may have one or more optional qualifiers, and may have a mandatory qualifier

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

(WIPO ST.26, Annex I, Sections 5-8)

WIPO ST.26: Sequences (7)

- Nucleotide sequences:
 - all lower case symbols;
 - no spaces, no numbering;
 - no “u” symbols; “t” represents uracil in RNA
 - “n” has a default value of “any one of ‘a’, ‘c’, ‘g’, or ‘t/u’”

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

WIPO ST.26: Sequences (8)

- Amino acid sequences:
 - all single letter, upper case symbols;
 - no spaces, no numbering;
 - “X” has a default value of “any one of ‘A’, ‘R’, ‘N’, ‘D’, ‘C’, ‘Q’, ‘E’, ‘G’, ‘H’, ‘I’, ‘L’, ‘K’, ‘M’, ‘F’, ‘P’, ‘O’, ‘S’, ‘U’, ‘T’, ‘W’, ‘Y’, or ‘V’”

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

WIPO ST.26: Sequences (8)

- Skipped sequences: allow an applicant to delete sequence data from a sequence listing without the need to renumber subsequence sequences.
 - INSDSeq_length, INSDSeq_moltype, INSDSeq_division present, but with no value;
 - No feature table and no source feature;
 - Sequence element must have the value “000”

```
<SequenceData sequenceIDNumber="7">  
  <INSDSeq>  
    <INSDSeq_length/>  
    <INSDSeq_moltype/>  
    <INSDSeq_division/>  
    <INSDSeq_sequence>000</INSDSeq_sequence>  
  </INSDSeq>  
</SequenceData>
```

WIPO ST.26: Contents

- **Main Body** – Requirements for inclusion/representation
- **Annex I** – Controlled vocabulary based on INSDC
- **Annex II** – ST.26 Document Type DTD
- **Annex III** – example ST.26 Sequence Listing XML file
- **Annex IV** – Character Subset of Basic Latin Code for ST.26 XML Instance
- **Annex V** – INDS Data Exchange Requirements (IPOs only)
- **Annex VI** – Guidance Document with Examples
- **Appendix to Annex VI** –XML file including all of the sequence disclosures exemplified in Annex VI
- **Annex VII** – Recommendation for the Transformation of a Sequence Listing from ST.25 to ST.26

WIPO ST.26:

Main Body

Paragraph reference	Content
1-9	Introduction; Definitions; Scope; References
10-37	Representation of sequences
38-49	Structure of the sequence listing in XML
50-71	Sequence data part; Feature table; Feature keys; Mandatory feature keys; Feature location
72-84	Feature qualifiers; Mandatory feature qualifiers
85-100	Free text; Coding sequences; Variants

WIPO ST.26 Annex I:

Controlled Vocabulary

Section	Content
1	List of Nucleotides (lower case, single letter symbols)
2	List of Modified Nucleotides
3	List of Amino Acids (upper case, single letter symbols)
4	List of Modified Amino Acids
5	Feature Keys for Nucleotide Sequences
6	Qualifiers for Nucleotide Sequences
7	Feature Keys for Amino Acids (adapted from UniProt)
8	Qualifiers for Amino Acid Sequences
9	Genetic Code Tables

WIPO ST.26 Annex II:

ST.26 DTD

- Details of ST.26 DTD: Current version 1.3

- General Information Part
 - Elements related to patent application information

- Sequence Data Part
 - Subset of the INSDC DTD
 - One or more sequence data elements where each element contains information about one sequence

WIPO ST.26 Annex VI:

Guidance Document

- Contains 49 real-world examples of sequence disclosures and an explanation of how ST.26 rules apply to each example;
- Each example addresses:
 1. Whether the sequence is required, permitted, or prohibited from inclusion in a sequence listing;
 2. If the sequence is required or permitted to be included in a sequence listing, how must it be represented.
- The Appendix to Annex VI is an ST.26 XML sequence listing that includes all of the examples presented in the Guidance Document.

WIPO ST.26 Annex VII:

Recommendation for the Transformation of a Sequence Listing from ST.25 to ST.26

- ...or how to avoid the inclusion of new matter.
- Requirements of ST.26 differ from ST.25 – ST.26 rules requires information that was not required by ST.25 rules
- Transformation of a sequence listing from ST.25 format to ST.26 format will always require input from the applicant
- Transformation of a compliant ST.25 sequence listing into ST.26 format will not result in new matter if the recommendations in Annex VII are followed
- Twenty transformation scenarios presented with recommendations and examples

WIPO Sequence (1)

- Desktop tool developed by WIPO to support authoring, validation, and generation of ST.26 compliant sequence listings
- Member states requested WIPO develop this common tool for all Offices and applicants at international, national and regional level
- Use of WIPO Sequence simplifies ST.26 XML creation with a user friendly interface: no need to ever directly edit an XML file
- Download the latest version for free from:
<https://www.wipo.int/standards/en/sequence/index.html>

WIPO Sequence (2)

- Sequence information can be saved in a project, validated, and then a sequence listing in ST.26 format can be generated
- Data can be imported from: ST.26 sequence listings, ST.26 projects, ST.25 sequence listings, multi-sequence format files, raw format files, and FASTA format files
- Validation of sequence listings in XML format as well
- Relevant feature keys, qualifiers, and organism names can be easily selected from drop-down menus
- Applicant and Inventor information can be stored in a “Persons and Organizations” database
- Supports export and import of XLIFF files used by translators

WIPO Sequence: Projects home page

WIPO | Sequence PROJECTS PERSONS & ORGANIZATIONS ORGANISMS HELP PREFERENCES ENGLISH

NEW PROJECT IMPORT PROJECT IMPORT SEQUENCE LISTING VALIDATE SEQUENCE LISTING

PROJECTS

Search project by name

Project name	Applicant file reference	Applicant name	Invention title	Status	Creation Date
16079428 no source sequence	H0075.70243US00	Ferring B.V.	STABLE LIQUID GONADOTROPIN FORMULATION	invalid	2021-02-01
All features and qualifiers	123abc	Simple Healthkit, Inc.	Compositions and Methods for Treating Cancer	modified	2021-02-01
Ark three letter aa code import from ST.25	abc123	Joe, Smith	Improper 3-letter AA code for import	new	2021-02-26
Beta 4 Test Project	abc123	Stephenson, Adam	Beta 4 Test Project	modified	2021-03-11
DNA RNA with u and t	uspto	Biotech, Inc.	Beta 4 test import ST25	modified	2021-03-15

WIPO Sequence: Project Detail

General Information Section

WIPO | Sequence
 FOR WIPO TRAINING
VERIFICATION REPORT
FREE TEXT QUALIFIERS
IMPORT REPORT
DISPLAY THE SEQUENCE LISTING
HELP

 PREFERENCES ENGLISH
 Return to project home

GENERAL INFORMATION SEQUENCES

GENERAL INFORMATION

APPLICATION IDENTIFICATION

Application Identified Before the assignment of the application number Application filed Applicant file reference ABC123	IP Office IB - International Bureau of the World Intellectual Property Organization (WIPO) Application number PCT/IB2015/099999 Filing date 2015-01-30	
--	--	--

PRIORITY IDENTIFICATION

IP Office	Application Number	Filing date	Selected Earliest Priority Application
IB - International Bureau of the World Intellectual Property Organization (WIPO)	PCT/IB2014/111111	2014-01-30	Yes

APPLICANT & INVENTOR

Applicant name Shutsugan Pharmaceuticals Kabushiki Kaisha	Primary applicant	
---	-------------------	--

INVENTION TITLE

Invention title Mus musculus abcd-1 gene for efg protein	Language en - English	
--	-----------------------	--

WIPO Sequence: Project Detail

Sequences Section

WIPO | Sequence
 FOR WIPO TRAINING
VERIFICATION REPORT
FREE TEXT QUALIFIERS
IMPORT REPORT
DISPLAY THE SEQUENCE LISTING
HELP

 PREFERENCES ENGLI
 Return to project home

GENERAL INFORMATION SEQUENCES

SEQUENCE 1

Sequence Number (ID) 1	Molecule Type DNA	✎
Sequence Name test	Organism Mus musculus	
Length 52		

▼ FEATURES

Add feature

Feature Key	Location	Qualifiers
<u>source</u>	1..52	mol_type = genomic DNA organism = Mus musculus

▼ SEQUENCE

atgaaattaa aacataaaaar ggatgataaa atgagatttg atataaaaaa gg	52	✎
< 1/1 >		

↑

What's next?

standards@wipo.int

Q&A session

Glossary: Acronyms

- CWS: Committee on WIPO Standards
- DDBJ: DNA Databank of Japan
- EMBL-EBI: The European Bioinformatics Institute
- EPO: European Patent Office
- INSDC: International Nucleotide Sequence Database Collaboration
- IPO: Intellectual Property Office
- NCBI: National Center for Biotechnology Information
- WIPO: World Intellectual Property Organization