



■ 产权组织标准ST.26

简介

网络研讨会培训

今天涉及的内容有哪些

- 什么是序列表？
- INSDC 序列数据库
- 为什么要制定新标准？
- 产权组织 ST.26的优势
- ST.25 和 ST.26 对比 – 有何差别？
- 向 ST.26 过渡
- XML 基本知识
- ST.26 基本知识 – 序列表的组成部分
- 产权组织 ST.26 的内容
- WIPO Sequence 简介

什么是序列表？

序列表

- 包含在专利申请中公开的、并构成说明书组成部分的核苷酸序列和/或氨基酸序列
- 包括关于每个序列的说明信息（称为注解）
- 符合产权组织相关标准（ST.25 或 ST.26）的要求
- 允许一项发明的序列数据可被检索
 - 在知识产权局
 - 在公开可用的数据库（INSDC 数据库）中

背景 - INSDC

- INSDC : 国际核苷酸序列数据库合作联盟
 - DDBJ : 日本DNA数据库
 - EMBL-EBI : 欧洲生物信息学研究所
 - NCBI : 国家生物技术信息中心 (GenBank)

- 向INSDC数据库提交已公布申请/已授权申请的序列数据的知识产权局包括：
 - 欧洲专利局
 - 日本特许厅
 - 韩国特许厅
 - 美国专利商标局

- INSDC数据库可供公开检索

为什么要制定新标准？

- 目前，提交的序列列表要符合产权组织ST.25的要求
但是.....
- ST.25 格式不符合INSDC的要求，因此在公共数据库中输入数据时会丢失数据
- ST.25 规则不够明确，世界各地的知识产权局对规则的解释和执行不尽相同
- 当前的常见序列类型（核苷酸类似物、D-氨基酸、支链序列）未被ST.25 规则所涵盖，因此不在可检索的数据库中
- 数据是非结构化数据 - ST.25 格式难以用于自动验证和数据交换

产权组织 ST.26 的优势 (1)

- 作为单一序列列表，可在全世界接受*
- 该标准作为指南，可确保各知识产权局在应用序列规则时保持一致
- 澄清了哪些序列需要或允许公开，才能纳入序列列表，并澄清了这些序列必须如何表示
- XML序列列表的结构提高了提交质量
- 提高了数据验证的自动化程度，简化了知识产权局的处理过程

* 需要将依赖语言的自由文本限定符翻译成申请语言的某些知识产权局除外，这些知识产权局可能需要替换序列列表

产权组织 ST.26 的优势 (2)

- 数据与INSDC数据库提供方的要求兼容 - DDBJ、EBI和NCBI - 序列注解（特征键和限定符）将被纳入可公开检索的数据库中
- 以下方面实现标准化：
 - 特征注解
 - 特征位置
 - 限定符和限定符值
 - 对序列变体的显示
- 纳入更多类型序列（核苷酸类似物、D-氨基酸、支链序列）的要求，意味着更多的序列数据可被检索

产权组织 ST.25 和 ST.26对比

ST.25	ST.26
带有数字标识符的ASCII .txt	带有元素和属性的XML
<p>不要求包括：</p> <ul style="list-style-type: none"> - D-氨基酸 - 支链序列的线性部分 - 核苷酸类似物 	<p>必须包括：</p> <ul style="list-style-type: none"> - D-氨基酸 - 支链序列的线性部分 - 核苷酸类似物
<p>序列注解：</p> <ul style="list-style-type: none"> - 仅特征键 	<p>序列注解：</p> <ul style="list-style-type: none"> - 特征键和限定符
<p>允许包括以下序列：</p> <ul style="list-style-type: none"> - < 10个专门定义的核苷酸 - < 4个专门定义的氨基酸 	<p>禁止以下序列：</p> <ul style="list-style-type: none"> - < 10个专门定义的核苷酸 - < 4个专门定义的氨基酸

产权组织 ST.25 和 ST.26对比

.....常规信息

ST.25	ST.26
可以包括 所有 优先权申请的信息	只能包括最早的 优先权申请
可以包括 所有 申请人和发明人名称	只能包括一个 申请人， 可以包括一个 发明人
允许一个发明名称	允许多个发明名称， 每个名称使用一种不同的语言
申请人/发明人名称和发明名称必须使用基本拉丁字符	可以包括使用任何有效的Unicode字符（带基本拉丁文翻译或音译）的申请人/发明人名称

产权组织 ST.25 和 ST.26 对比

.....序列数据 (1)

ST.25	ST.26
序列仅被识别为DNA、RNA或PRT	序列可被识别为DNA、RNA或AA，并带有必填的mol_type限定符，以便进一步描述分子
生物体名称： <ul style="list-style-type: none"> - 拉丁属/种 - 病毒名称 - “人工序列” - “未知” 	生物体名称： <ul style="list-style-type: none"> - 拉丁属/种 - 病毒名称 - “合成构建体” - “未识别”
“u”代表核苷酸序列中的尿嘧啶	“t”代表RNA序列中的尿嘧啶和DNA序列中的胸腺嘧啶
氨基酸序列用三个字母的缩写来表示	氨基酸序列用一个字母的缩写来表示

产权组织 ST.25 和 ST.26对比

.....序列数据 (2)

ST.25	ST.26
“n” 和 “Xaa” 变量必须具有在特征中提供的定义	对没有定义的 “n” 和 “X” 变量假设默认值
特征位置格式没有明确定义	严格定义的特征位置格式；允许在所有序列类型中使用 “<” 和 “>”，并允许在核苷酸序列中使用 “^”（脱字符）、“join”（连接）、“order”（排序）和“complement”（补充）
允许 “混合模式” 序列——带氨基酸翻译的核苷酸序列，如下所示	不允许 “混合模式”；核苷酸翻译仅包括在 “translation”（翻译）限定符中

向产权组织 ST.26 过渡

- CWS/5商定将**2022年1月22日**作为过渡日，称作“**大爆炸**”日
- **所有**知识产权局将在国际（PCT）、**国家和地区**层面同时过渡
- 国际申请日（IFD），**而不是**优先权日，将是决定一件申请是适用 ST.25 还是 ST.26 序列规则的参考日期
- **注**：ST.25 将继续对申请日在2022年1月1日前的申请有效

产权组织 ST.26

.....必须包括什么？

- 核苷酸序列：
 - 10个或更多“专门定义”并“枚举的”*残基
 - 包括具有核苷酸类似物（如肽核酸（PNA）和二醇核酸（GNA））的序列
- 氨基酸序列：
 - 4个或更多“专门定义”并“枚举的”*残基
 - 包括具有D-氨基酸的序列
 - 要求将支链序列的线性区域包括在序列表中

产权组织 ST.26

.....必须包括什么？

- 什么是“专门定义的”核苷酸或氨基酸？
- “专门定义”是指附件一中所列的除用符号“n”代表的核苷酸和除用符号“X”代表的氨基酸以外的任何核苷酸（产权组织标准ST.26, 第3 (m) 段)
- 只有“专门定义”的残基在计算最低长度时作数：
 - 10个或更多专门定义的核苷酸；或者
 - 4个或更多专门定义的氨基酸

5'- anctggcaan – 3' 只有8个专门定义的核苷酸；不得纳入序列表

5'- agctggcaat – 3' 10个专门定义的核苷酸；必须纳入序列表

产权组织 ST.26 : XML文件

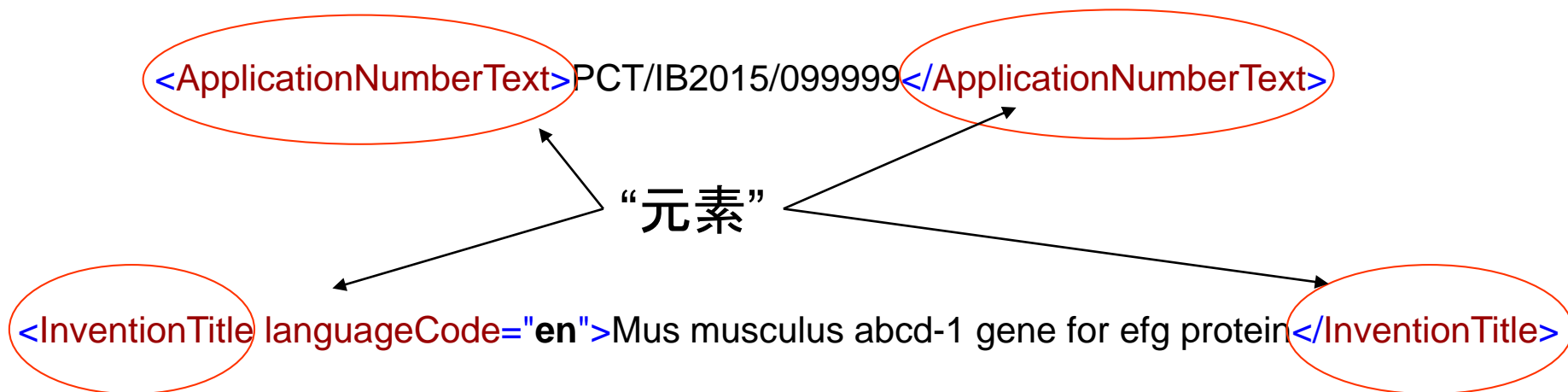
- 必须作为一个XML 1.0格式文件提供
- 必须根据产权组织ST.26 DTD（附件二）和从该标准内容中得出的业务规则进行验证
- 必须使用Unicode UTF-8进行编码
- 序列列表的结构：
 - XML声明
 - 文档类型（DOCTYPE）声明
 - 根元素
 - 常规信息部分
 - 序列数据部分

XML 基本知识

- XML = eXtensible (可扩展) Markup (标记) Language (语言)
- 使用描述性元素和属性对信息进行“标记”
- 标准化的数据交换方式，人和机器均可读取
- DTD = Document (文档) Type (类型) Definition (定义) - 对XML文档的结构和合法元素及属性进行定义

XML 基本知识

元素、属性和值 (1)



XML 基本知识

元素、属性和值 (2)

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

“元素值”

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

XML 基本知识

元素、属性和值 (3)

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

“属性”

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

“属性值”

XML 基本知识

保留字符

在元素值中必须用其预定义实体来代替。

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

例如：需要设定的特征位置为 “<50..62”

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



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



产权组织 ST.26 : 范例

```

<?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>atgaaattaaaacataaaaarggatgataaaatgagatttgatataaaaagg</INSDSeq_sequence>
    </INSDSeq>
  </SequenceData>
</ST26SequenceListing>

```

产权组织 ST.26 : 组件 (1)

第 1 行 - XML 声明
(ST.26 第 39(a)段)

```
<?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">
```

产权组织 ST.26 : 组件 (2)

第 2 行 – 文档类型 (DOCTYPE) 声明
(ST.26 第 39(b)段)

```
<?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">
```

产权组织 ST.26 : 组件 (3)

第 3 行- 根元素
(ST.26 第 43 段)

```
<?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">
```

常规信息：

(ST.26 第 38(a)段、第
45-49段)

产权组织 ST.26 : 组件 (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>atgaaattaaaacataaaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
</ST26SequenceListing>

```

序列数据

(ST.26 第 38(b)段、第
50-100段)

产权组织 ST.26 : 常规信息 (1)

- “Application Identification” (申请识别) 节
 - 如果已知的话, 申请号、申请日、知识产权局代码是必填项 ;
 - 否则, 只需提供申请人档案号

```
<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>
```

产权组织 ST.26：常规信息（2）

- “Priority Application”（优先权申请）节
 - 序列中只能包括一个优先权申请，而且必须是最早的优先权申请；
 - 在要求优先权的情况下是必填项

```
<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>
```

产权组织 ST.26：常规信息（3）

- “Applicant and Inventor Name”（申请人和发明人名称）节
 - 序列表中只能包括一个申请人名称和一个发明人名称，而且必须是“主要”申请人和发明人；
 - 申请人名称是**必填项**；发明人名称是**选填项**；
 - 申请人名称和发明人名称的语言代码是**必填项**；
 - 如果申请人和/或发明人名称包含非Unicode基本拉丁字符，则必须包括基本拉丁字符的音译或翻译

```

<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>
  
```

产权组织 ST.26：常规信息（4）

- “Invention Title”（发明名称）节
 - 必须以申请语言提供至少一个发明名称；
 - 可以包括其他语言的更多名称；
 - 每个名称的语言代码是必填项；

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

产权组织 ST.26：常规信息（5）

- “Sequence Total Quantity”（序列总数）元素
 - 必填；
 - 总数必须包括跳过的序列

<SequenceTotalQuantity>**6**</SequenceTotalQuantity>

产权组织 ST.26 : 序列 (1)

- 必填元素

“Sequence Identification Number” (序列识别号)
或 “SEQ ID NO:”

```

<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_quals>
          <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_quals>
      </INSDFeature>
    </INSDSeq_feature-table>
    <INSDSeq_sequence>atgaaattaaaacataaaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>
  
```

序列长度

分子类型
(DNA、RNA、或 AA)

分割
(通常是 “PAT”)

产权组织 ST.26 : 序列 (2)

- 必填的“source”或“SOURCE”特征

```

<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>atgaaattaaacataaaarggatgataaatgagatttgatataaaaaagg</INSDSeq_sequence>
  </INSDSeq>
</SequenceData>

```

要求每个序列有且只有一个“source”特征

“source”位置必须跨整个序列

两个必填限定符：
“mol_type”（分子类型）和
“organism”（生物体）

产权组织 ST.26 : 序列 (3)

ST.26 序列数据 :

```
<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>
```

产权组织 ST.26 : 序列 (4)

- 必填的“mol_type”和“MOL_TYPE”限定符 – 可选值：

DNA

基因组 DNA
其他 DNA
未分配的 DNA

RNA

基因组 RNA
mRNA
tRNA
rRNA
其他 RNA
转录 RNA
病毒 cRNA
未分配的 RNA

AA

蛋白质

产权组织 ST.26：序列（5）

- 必填的“organism”和“ORGANISM”限定符 – 可选值：
 - 拉丁文属名和种名，例如：“Mus musculus”
 - 属名后接“sp.”，例如：“Mus sp.”
 - 病毒码，例如：“Torque teno virus 1”
 - “unidentified”（未识别）
 - “synthetic construct”（合成构建体）
- 常用名，例如“老鼠”，不得用作生物体名称。如有需要，可以在序列表中的注释限定符中包括常用名。

产权组织 ST.26 : 序列表 (6)

特征键和限定符

除了必填的“source”或“SOURCE”特征外，申请人可以添加多个可选特征，来进一步描述序列

- 用于核苷酸序列和氨基酸序列的不同特征键；
- 每个特征可能有一个或多个可选限定符，并且可能有一个必填限定符

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

(产权组织 ST.26, 附件一, 第5-8节)

产权组织 ST.26 : 序列 (7)

- 核苷酸序列 :
 - 均为小写的符号 ;
 - 无空格, 无编号 ;
 - 没有 “u” 符号 ; “t” 代表RNA的尿嘧啶
 - “n” 的默认值是 ‘a’、‘c’、‘g’ 或 ‘t/u’ 中的任何一个

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”

产权组织 ST.26 : 序列 (8)

- 氨基酸序列 :
 - 均为单字母大写符号 ;
 - 无空格, 无编号 ;
 - “X” 的默认值为 ‘A’、‘R’、‘N’、‘D’、‘C’、‘Q’、‘E’、‘G’、‘H’、‘I’、‘L’、‘K’、‘M’、‘F’、‘P’、‘O’、‘S’、‘U’、‘T’、‘W’、‘Y’、或 ‘V’ 中的任何一个

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"

产权组织 ST.26 : 序列 (8)

- 跳过的序列：允许申请人从序列表中删除序列数据，而无需对后续序列重新编号。
 - INSDSeq_length, INSDSeq_moltype, INSDSeq_division 存在，但没有任何值；
 - 没有特征表，也没有源特征；
 - 序列元素必须有值“000”

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

产权组织 ST.26：内容

- **正文** – 对纳入/表示的要求
- **附件一** – 基于 INSDC的受控词汇表
- **附件二** – ST.26文档类型 DTD
- **附件三** – ST.26序列表 XML文件示例
- **附件四** – ST.26 XML 示例的基本拉丁代码字符子集
- **附件五** – INDS数据交换要求（仅涉及知识产权局）
- **附件六** – 带有范例的指导文件
- **附件六附录** – 附件六中列举的包括所有序列公开信息的XML文件
- **附件七** –关于将序列表从ST.25转换为ST.26的建议

产权组织 ST.26 : 正文

段号	内容
1-9	导言；定义；范围；参考
10-37	序列的表示
38-49	XML序列列表结构
50-71	序列数据组成部分；特征表；特征键；必填特征键；特征位置
72-84	特征限定符；必填特征限定符
85-100	自由文本；编码序列；变体

产权组织 ST.26 附件一： 受控词汇表

节	内容
1	核苷酸列表（小写，单字母符号）
2	修饰的核苷酸列表
3	氨基酸列表（大写，单字母符号）
4	修饰的氨基酸列表
5	用于核苷酸序列的特征键
6	用于核苷酸序列的限定符
7	用于氨基酸的特征键（改编自UniProt）
8	用于氨基酸序列的限定符
9	遗传密码表

产权组织 ST.26 附件二： ST.26 DTD

- ST.26 DTD详细信息：当前 1.3版
- 常规信息部分
 - 与专利申请信息有关的元素
- 序列数据组成部分
 - INSDC DTD的子集
 - 一个或多个序列数据元素，每个元素包含一个序列的信息

产权组织 ST.26附件六： 指导文件

- 包含49个关于序列公开的真实范例，并针对每个范例解释了ST.26规则如何适用；
- 每个范例都涉及：
 1. 是否要求、允许或禁止序列被纳入序列表；
 2. 如果要求或允许序列被纳入序列表，应该怎样表示。
- 附件六的附录是一个ST.26 XML序列表，包括了指导文件中列出的所有范例。

产权组织 ST.26 附件七： 关于将序列列表从ST.25转换为ST.26的建议

-或者如何避免增加新内容。
- ST.26的要求与ST.25不同 - ST.26规则要求提供ST.25规则没有要求提供的信息
- 将序列列表从ST.25格式转换为ST.26格式，通常需要申请人的意见
- 如果遵循附件七中的建议，将符合ST.25要求的序列列表转换为ST.26格式，不会导致新内容
- 给出了20种转换情形，并附有建议和范例

WIPO Sequence (1)


- 是产权组织开发的桌面工具，用于支持编写、验证和生成符合ST.26要求的序列列表
- 成员国要求产权组织在国际、国家和地区层面为所有主管局和申请人开发这一通用工具
- 使用 WIPO Sequence可以通过友好的界面，简化ST.26 XML的创建：无需直接编辑XML文件
- 在以下网址免费下载最新版本：

<https://www.wipo.int/standards/en/sequence/index.html>

WIPO Sequence (2)


- 序列信息可以在项目中保存、验证，然后可以生成ST.26格式的序列表
- 数据可以从以下文件导入：ST.26序列表、ST.26项目、ST.25序列表、multi-sequence格式文件、raw格式文件和FASTA格式文件
- 也可以验证XML格式的序列表
- 相关特征键、限定符和生物体名称可以从下拉菜单中轻松选择
- 申请人和发明人的信息可以存储在“人员和组织”数据库中
- 支持笔译员使用的XLIFF文件的导出和导入

WIPO Sequence : 项目主页

 WIPO | Sequence
PROJECTS PERSONS & ORGANIZATIONS ORGANISMS HELP
PREFERENCES ENGLISH

[NEW PROJECT](#) [IMPORT PROJECT](#) [IMPORT SEQUENCE LISTING](#) [VALIDATE SEQUENCE LISTING](#)

PROJECTS



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 : 项目详细信息

常规信息节

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 : 项目详细信息

序列节

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

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 >

✎

↑

下一步是什么？
standards@wipo.int

问答环节

词汇表：缩略语

- CWS：产权组织标准委员会
- DDBJ：日本 DNA 数据库
- EMBL-EBI：欧洲生物信息学研究所
- EPO：欧洲专利局
- INSDC：国际核苷酸序列数据库合作联盟
- IPO：知识产权局
- NCBI：国家生物技术信息中心
- WIPO：世界知识产权组织