



ÚŘAD
PRŮMYSLOVÉHO
VLASTNICTVÍ
Česká republika

www.upv.gov.cz

Sekvence primerů a sond podle standardu WIPO ST.26

ÚŘAD PRŮMYSLOVÉHO VLASTNICTVÍ
Ing. Marie Kleinová, Dr.
mkleinova@upv.gov.cz



Obsah přednášky

- Co je sekvenční protokol?
 - Sekvenční databáze INSDC
 - Proč nový standard?
 - Výhody WIPO ST.26
 - Rozdíly mezi ST.25 a ST.26
 - Přejechod na ST.26
 - Formát XML
 - Obsah sekvenčního protokolu
 - Úvod do nástroje WIPO Sekvence
-

Co je sekvenční protokol?

- Obsahuje nukleotidové a/nebo aminokyselinové sekvence popsané v přihlášce vynálezu nebo užitného vzoru a tvoří součást popisu
 - Zahrnuje detailní popisné informace o každé sekvenci
 - Splňuje požadavky příslušné normy WIPO (ST.25 nebo ST.26)
 - Umožňuje prohledávat sekvenční data:
 - v rámci patentového úřadu
 - ve veřejně dostupných databázích (databáze INSDC)
-

Co se skrývá za INSDC?

INSDC: International Sequence Database Collaboration
(Mezinárodní spolupráce databází nukleotidových sekvencí)

- soubor veřejně dostupných databází zahrnující:

- **DDBJ:** Japonskou databanku DNA
- **EMBL-EBI:** Evropský bioinformatický institut
- **NBCI:** Národní centrum pro biotechnologické informace (GenBank)

- Mezi patentové instituce, které poskytují sekvenční data do INSDC patří:

- Evropský patentový úřad
 - Japonský patentový úřad
 - Korejský úřad pro duševní vlastnictví
 - Úřad pro patenty a ochranné známky USA
-

Proč nový standard?

V současné době jsou výpisy sekvencí uloženy v souladu s WIPO ST.25, **ale:**

- Formát ST.25 není v souladu s požadavky INSDC, takže data se ztrácí při zadávání do veřejných databází
 - Pravidla ST.25 nejsou jasná, národní patentové úřady po celém světě je vykládají a prosazují jinak
 - Pravidla ST.25 nepokrývají některé typy sekvencí, které jsou dnes běžné (např. analogy nukleotidů, D-aminokyseliny, rozvětvené sekvence), a proto aktuálně nejsou v databázích zahrnuty
 - Data jsou nestrukturovaná – formát ST.25 je obtížně použitelný pro automatickou validaci a výměnu dat
-

Přechod na WIPO ST.26

- Po mnoha letech příprav byl jako datum „**velkého třesku**“ stanoven **1.leden 2022**
 - Přechod je koordinován na mezinárodní, národní i regionální úrovni - v České republice byl za datum přechodu určen **1. červenec 2022**
 - **Datum mezinárodního podání** (PCT) je referenčním datem a určuje, zda sekvence v přihlášce spadají do pravidel ST.25 nebo ST.26, **NE** datum priority
 - **ST.25 zůstane v platnosti pro PCT přihlášky s datem podání před 1. lednem 2022 (pro národní přihlášky v ČR před 1.7.2022)**
-

Výhody ST.26

- Umožňuje sjednocení sekvenčních protokolů v celosvětovém měřítku
 - Slouží jako vodítko pro zajištění shody v uplatňování pravidel pro výpis sekvencí mezi jednotlivými národními úřady
 - Objasňuje, jaký formát sekvencí v sekvenčním protokolu je požadován či povolen a jak tyto sekvence musí být zastoupeny
 - Zlepšuje kvalitu sdílení sekvenčních dat díky formátu XML
 - Umožňuje automatickou validaci zadaných dat a zefektivňuje jejich zpracování v národních úřadech
 - Zajišťuje kompatibilitu s databázemi INSDC (základní charakteristiky, identifikátory, zástupné symboly) a usnadňuje vyhledávání
-

WIPO ST.25 versus ST.26

ASCII .txt s číselnými identifikátory

Nezahrnuje:

- D-aminokyseliny
- lineární části rozvětvených sekvencí
- analogy nukleotidů

Anotace sekvencí obsahují:

- pouze základní (klíčové) charakteristiky

Sekvence mohou zahrnovat:

- méně než 10 specificky definovaných nukleotidů
- méně než 4 specificky definované aminokyseliny

Formát .xml s prvky a atributy

Musí zahrnovat:

- D-aminokyseliny
- lineární části rozvětvených sekvencí
- analogy nukleotidů

Anotace sekvencí obsahují:

- Základní (klíčové) charakteristiky a kvalifikátoky

Sekvence nesmí zahrnovat:

- méně než 10 specificky definovaných nukleotidů
- méně než 4 specificky definované aminokyseliny

WIPO ST.25 versus ST.26

rozdíly ve všeobecných informacích

Může zahrnovat informace ze všech prioritních přihlášek

Může obsahovat údaje o všech přihlašovatelích a původcích

Je povolen jen jeden název vynálezu

Jména přihlašovatele/původce a název vynálezu musí být uvedeny v základní latině

Zahrnuje pouze informace z přihlášky s nejstarší prioritou

Smí obsahovat data jen jednoho přihlašovatele a volitelně jednoho původce

Je povoleno více názvů vynálezu, každý v jiném jazyce

Jména přihlašovatele/původce mohou obsahovat jakýkoli platný znak Unicode spolu s překladem nebo přepisem do latinky

WIPO ST.25 versus ST.26

rozdíly ve výpisu sekvencí

Sekvence jsou identifikovány jako DNA, RNA nebo PRT

Používaná pojmenování organismů:

- latinský rod/druh
- název viru
- „umělá sekvence“
- „neznámý“

„u“ představuje uracil v nukleotidové sekvenci

Sekvence aminokyselin se uvádí třípísmennou zkratkou

Sekvence jsou identifikovány jako DNA, RNA nebo AA spolu s povinným kvalifikátorem „mol_type“, který molekulu blíže popisuje

Používaná pojmenování organismů:

- latinský rod/druh
- název viru
- „syntetický konstrukt“
- „neidentifikovaný“

„t“ představuje uracil v sekvenci RNA a thymin v sekvenci DNA

Sekvence aminokyselin se uvádí jednopísmennou zkratkou

WIPO ST.26 – co musí protokol obsahovat?

Nukleotidové sekvence obsahující:

- 10 nebo více „specificky definovaných“ a „vyjmenovaných“ nukleotidů
- včetně sekvencí s nukleotidových analogů, jako jsou peptidové nukleové kyseliny (PNAs) a glykolové nukleové kyseliny (GNAs)

Sekvence aminokyselin obsahující:

- 4 nebo více „specificky definovaných“ a „vyjmenovaných“ aminokyselin
 - včetně sekvencí s D-aminokyselin
 - včetně lineárních oblastí rozvětvených sekvencí
-

WIPO ST.26 – co musí protokol obsahovat?

Co je „specificky definovaný“ nukleotid nebo aminokyselina?

- „specificky definovaný“ znamená jakýkoli jiný nukleotid než ty, které jsou reprezentovány symbolem „n“, a jakoukoli jinou aminokyselinou než ty které jsou reprezentovány symbolem „X“, jak jsou uvedeny v příloze 1 (Annex I)
- do minimální délky sekvence se započítávají pouze „specificky definované“ zbytky, tedy:

≥ 10 specificky definovaných nukleotidů nebo

≥ 4 specificky definované aminokyseliny

Příklad: 5'- anctggcaan - 3' pouze 8 specificky definovaných nukleotidů; nesmí být zahrnuty v sekvenčním protokolu

5'- agctggcaat – 3' deset specificky definovaných nukleotidů; musí být zahrnuty v sekvenčním protokolu

Co představuje formát XML?

XML = e**X**tensible **M**arkup **L**anguage

- informace jsou „označeny“ pomocí popisných prvků a atributů
 - používá standardní kódování Unicode UTF-8 a předdefinované znaky
 - zajišťuje standardizovaný způsob výměny dat, který je čitelný jak člověkem, tak strojem
 - umožňuje snadný přenos dat do sekvenčních databází
-

Co obsahuje sekvenční protokol?

- XML deklarace (XML declaration)

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

- deklarace typu dokumentu (DOCTYPE declaration)

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

- Základní (kořenové) informace (root elements)
 - část obsahující všeobecné informace
 - část obsahující sekvenční data

Příklad:



Sekvence primerů a sond podle standardu WIPO ST.26

```
<?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>atgaaattaaaaacataaaarggatgataaaatgagatttgatataaaaaagg</INSDSeq_sequence>
    </INSDSeq>
  </SequenceData>
</ST26SequenceListing>
```

WIPO ST.26 – obsah kořenových částí

Část „Všeobecné informace“ (general information part)

- 1) údaje o přihlášce
- 2) údaje o prioritě
- 3) údaje o přihlašovatelci a původci
- 4) název vynálezu
- 5) počet sekvencí

Část „Sekvenční data“ (sequence data part)

- 6) identifikační číslo sekvence
- 7) délka sekvence
- 8) typ molekuly
- 9) původ (organismus)
- 10) vlastní výpis sekvence



1) Údaje o přihlášce (Application Identification)

- číslo přihlášky
- datum podání
- kód patentového úřadu (je-li znám)
- jsou-li tyto údaje známy, jinak postačuje odkaz na spis přihlašovatele

2) Údaje o prioritě (Priority Application Ident.)

- do sekvenčního protokolu může být zařazena pouze jedna prioritní přihláška a musí to být přihláška s nejstarší prioritou
 - údaj o prioritě je povinný, pokud je uplatňována
-

3) Údaje o přihlašovatele a původce (Applicant and Inventor Name)

- může být zahrnuto pouze jedno jméno přihlašovatele a jedno jméno původce, a to toho, který je na seznamu uveden první v pořadí
 - uvedení jména přihlašovatele je povinné, jméno původce lze uvést volitelně
 - pro jména přihlašovatelů/původců je povinné uvedení jazykového kódu
 - pokud jméno přihlašovatele a/nebo původce obsahuje jiné než základní znaky latinky (Unicode), pak musí být uveden přepis nebo překlad do základní latinky
-

4) Název vynálezu (Invention Title)

- je nutné uvést alespoň jeden název vynálezu v jazyce podání
- mohou být uvedeny i další názvy v jiných jazycích
- pro každý název je nutné uvést kód daného jazyka

5) Počet sekvencí (Sequence Total Quantity)

- je povinné uvést celkový počet sekvencí v protokolu
 - uvedený celkový počet musí zahrnovat i vypuštěné sekvence
-

6) Identifikační číslo sekvence (Sequence Identification Number)

- SEQ ID NO: 1, 2, 3, ...
- v češtině možné též SEQ ID Č.: nebo sekv. ID č.:

7) Délka sekvence (Sequence Length)

- počet jednotlivých členů, tj. nukleotidů nebo aminokyselin, popř. jejich analogů v sekvenci
-



8) Typ molekuly (mol_type / MOL_TYPE)

DNA

genomic DNA

other DNA

unassigned DNA

RNA

genomic RNA

mRNA

tRNA

rRNA

other RNA

transcribed RNA

viral cRNA

unassigned RNA

AA

protein



9) Původ (Organism)

- latinské rodové a druhové jméno, např. „Mus musculus“
 - rodové jméno následované „sp.“ např. „Mus sp.“
 - název viru, např. „Torque teno virus 1“
 - „unidentified“
 - „synthetic construct“
 - obecné názvy, jako například „mouse“, se jako název organismu používat nesmí. V případě potřeby mohou být obecné názvy v sekvenčním protokolu uvedeny pod kvalifikátorem „note“ - poznámka
-

10) Výpis sekvence (Sequence)

Nukleotidové sekvence:

- smí se používat jen malá písmena
- nepoužívat žádné mezery, žádné číslování
- nepoužívat symbol **u**, **t** představuje uracil v RNA
- symbol „n“ označuje jakékoli z **a**, **c**, **g** nebo **t/u**

Aminokyselinové sekvence:

- smí se používat pouze jednopísmenné kódy AK
 - smí se používat jen velká písmena
 - nepoužívat žádné mezery, žádné číslování
 - symbol „X“ označuje jakékoli z **A**, **R**, **N**, **D**, **C**, **Q**, **E**, **G**, **H**, **I**, **L**, **K**, **M**, **F**, **P**, **O**, **S**, **U**, **T**, **W**, **Y** nebo **V**
-

Vypuštění sekvence

- přihlašovatel má možnost vymazat sekvenci ze sekvenčního protokolu bez nutnosti přečíslování následujících sekvencí
- kvalifikátory INSSeq_length, INSSeq_moltype, INSSeq_division atd. budou v protokolu uvedeny, ale bez konkrétní hodnoty
- na místě výpisu sekvence bude uvedeno „000“

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

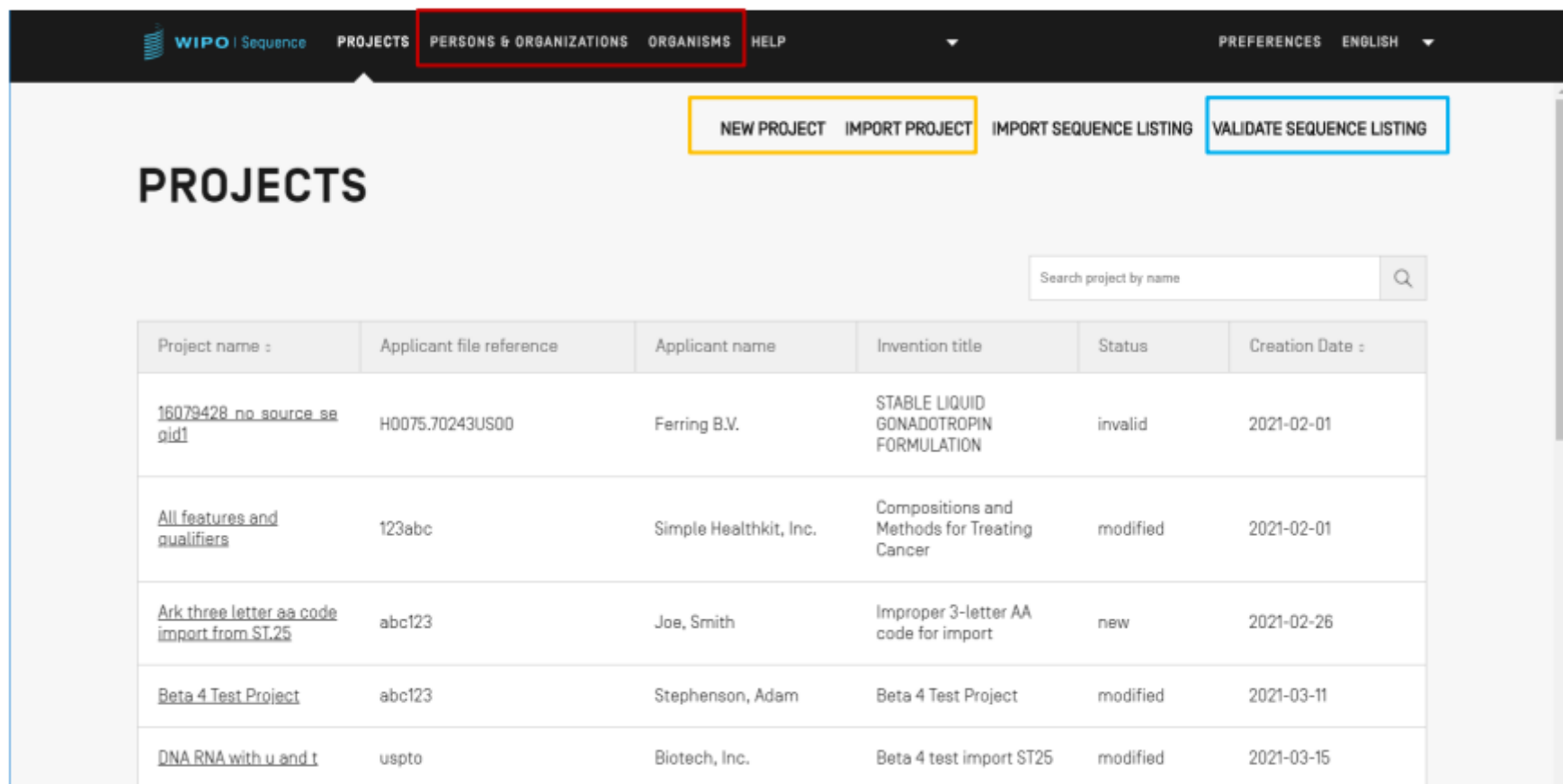

Kde hledat a stahovat?

- [WIPO Standard ST.26](#)
- <https://www.wipo.int/standards/en/sequence>

Kde přikládat?

- www.upv.gov.cz → E-PORTÁL
→ nové podání → příloha:
„Sekvence nukleotidů/aminokyselin dle ST.26“
-

Nástroj WIPO Sequence – úvodní strana



The screenshot displays the main interface of the WIPO Sequence tool. At the top, there is a navigation bar with the following menu items: WIPO | Sequence, PROJECTS, PERSONS & ORGANIZATIONS, ORGANISMS, and HELP. On the right side of the navigation bar, there are links for PREFERENCES and ENGLISH. Below the navigation bar, there are four main action buttons: NEW PROJECT, IMPORT PROJECT, IMPORT SEQUENCE LISTING, and VALIDATE SEQUENCE LISTING. The main heading is "PROJECTS". To the right of the heading is a search box labeled "Search project by name" with a magnifying glass icon. Below the search box is a table with the following columns: Project name, Applicant file reference, Applicant name, Invention title, Status, and Creation Date. The table contains five rows of project data.

Project name :	Applicant file reference	Applicant name	Invention title	Status	Creation Date :
16079428_no_source_se qid1	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



Nástroj WIPO Sequence – výstup z validace

WIPO | Sequence ZKOUŠKA 2 VERIFICATION REPORT LANGUAGE DEPENDENT QUALIFIERS IMPORT REPORT DISPLAY THE SEQUENCE LISTING HELP ▾ PREFERENCES ENGLISH ▾ [Return to project home](#)

ERROR: After project verification, some errors or warnings have been detected. ✕

Report generated on 2022-11-11

[Print Report](#)

Severity ▾	Data Element ▾	Message Text	Detected Value	Detected Sequence
ERROR	File Name	The file name field includes non-permitted characters: š. Only printable characters (including the space character) from the Unicode Basic Latin code table (excluding the reserved characters) are permitted.	Zkouška 2.xml	
WARNING	Application Identification	An application identification information has not been entered. If the application number has been assigned, it must be entered.	-	
WARNING	Applicant File Reference	The applicant file reference field includes non-permitted characters: š,ř,á. Only printable characters (including the space character) from the Unicode Basic Latin code table (excluding the reserved characters) are permitted.	naše přihláška	
WARNING	Earliest Priority Application Identifications	Earliest priority application information has not been entered. It must be entered when a priority claim is made to an earlier application.	-	



**Děkuji za pozornost
a přeji krásný den...**

Ing. Marie Kleinová, Dr.
