MUNI | RECETOX

E2020 – Soft skills II – Information Literacy
4. Databases: Practical

01.03.2022

Mgr. Ludovic Mayer
ludovic.mayer@recetox.muni.cz

Content

- Lecture divided into 2 parts:
  - 1st part: Theoretical – to understand how is the information stored
    – how to access it
  - 2nd part: Practical – **you are at the helm and you find the information!**

Content

– 2nd part: Practical – you are at the helm and you find the information!

- Publication databases
- Properties databases
- Spectral databases
- Registration (legal) databases
- Protein and DNA databases
- Patent database

This week

Next week

Publication database

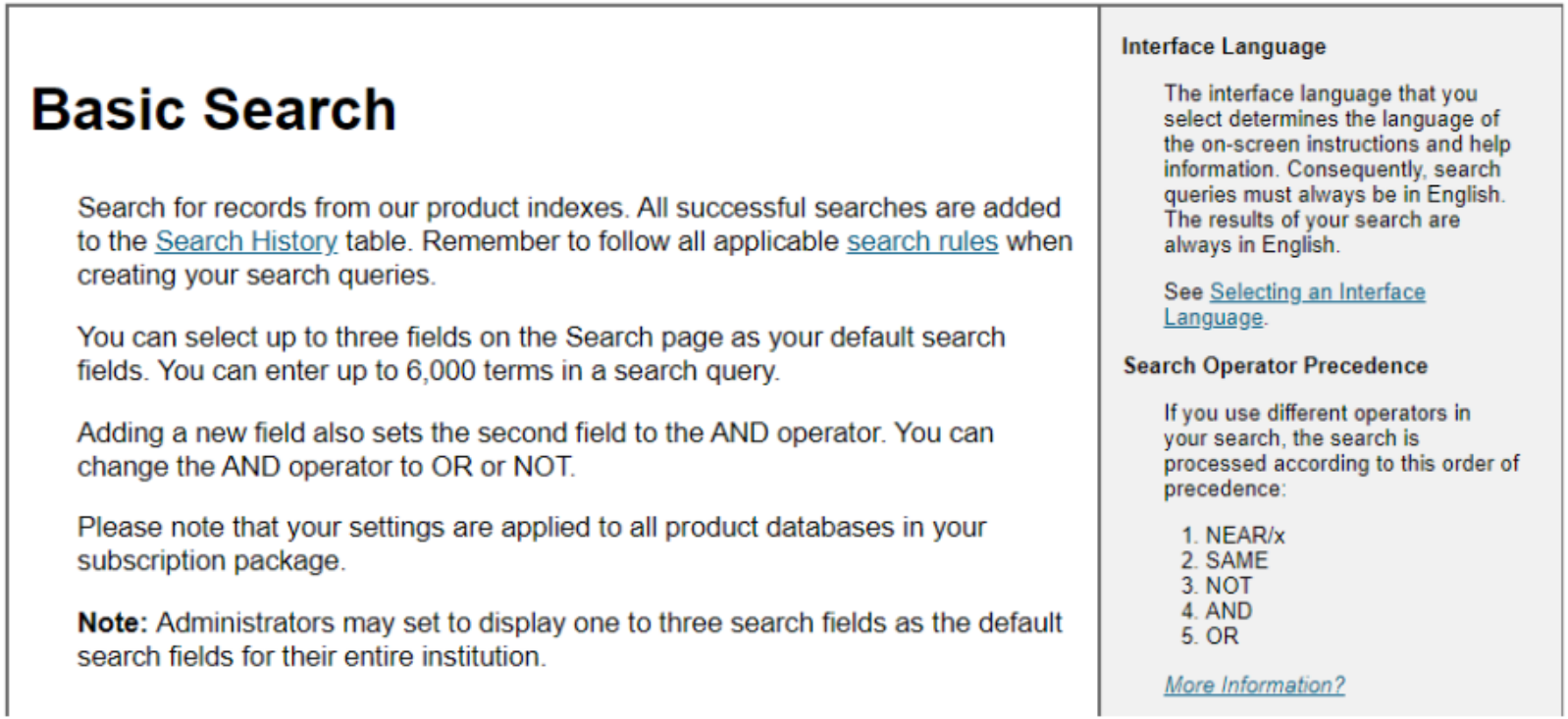
- https://www.scopus.com/search/form.uri?display=basic#basic
- https://www.webofscience.com/wos/woscc/basic-search
- https://scholar.google.com/

- **Task 0**: Let's look together for a research paper
  - Topic?

Proficient search

- How to build an advance request?
  - By using logical operator:

    AND, OR, etc.



Basic Search

Search for records from our product indexes. All successful searches are added to the Search History table. Remember to follow all applicable search rules when creating your search queries.

You can select up to three fields on the Search page as your default search fields. You can enter up to 6,000 terms in a search query.

Adding a new field also sets the second field to the AND operator. You can change the AND operator to OR or NOT.

Please note that your settings are applied to all product databases in your subscription package.

Note: Administrators may set to display one to three search fields as the default search fields for their entire institution.

Interface Language

The interface language that you select determines the language of the on-screen instructions and help information. Consequently, search queries must always be in English. The results of your search are always in English.

See Selecting an Interface Language.

Search Operator Precedence

If you use different operators in your search, the search is processed according to this order of precedence:

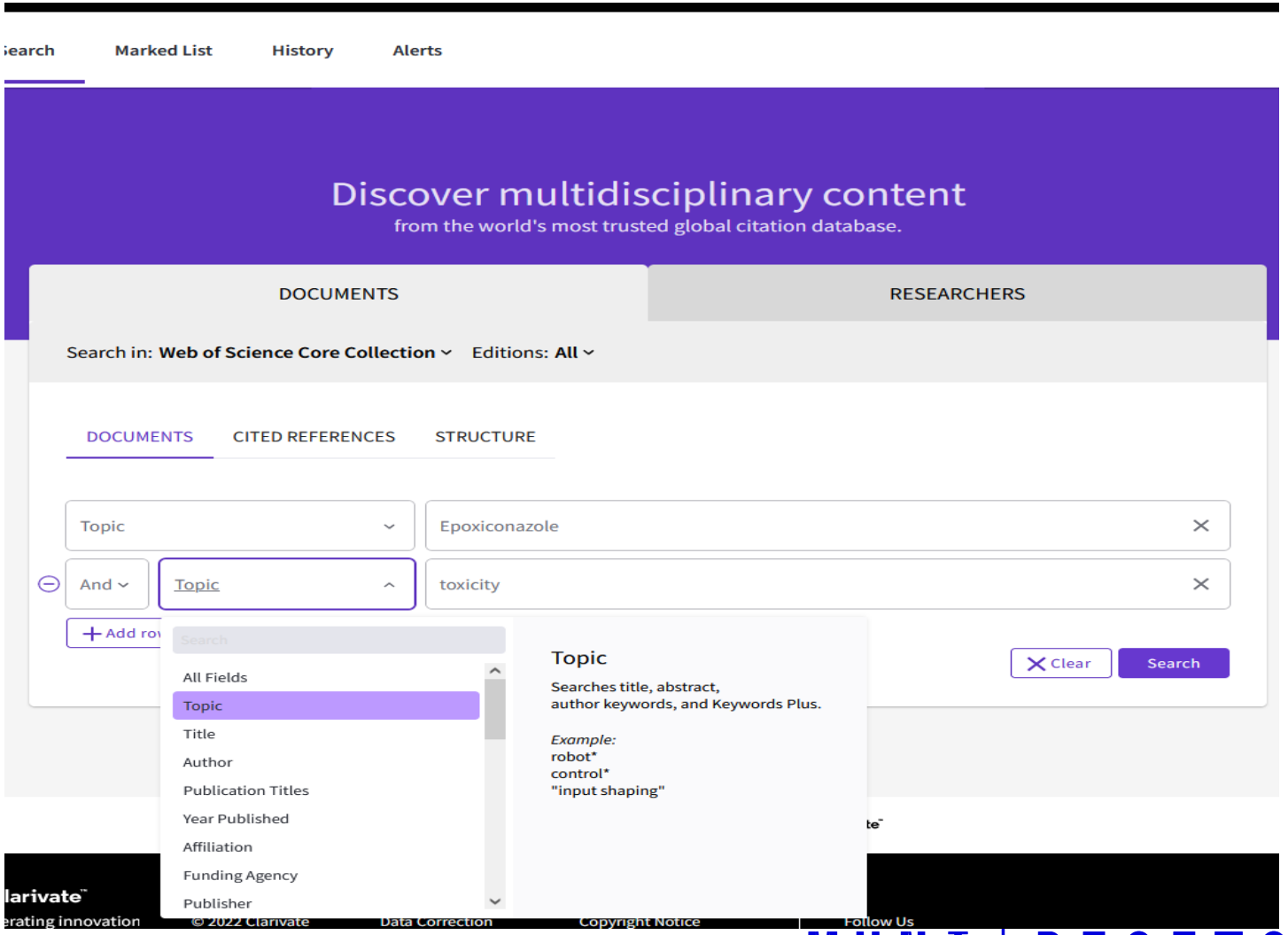
1. NEAR/x
2. SAME
3. NOT
4. AND
5. OR

More Information?

Proficient search

- https://www.webofscience.com/wos/woscc/basic-search

- Filter a results
  - By adding a keyword
  - Restrictions by publication type: article, review, etc.
  - Time limit: eg the last 10 years

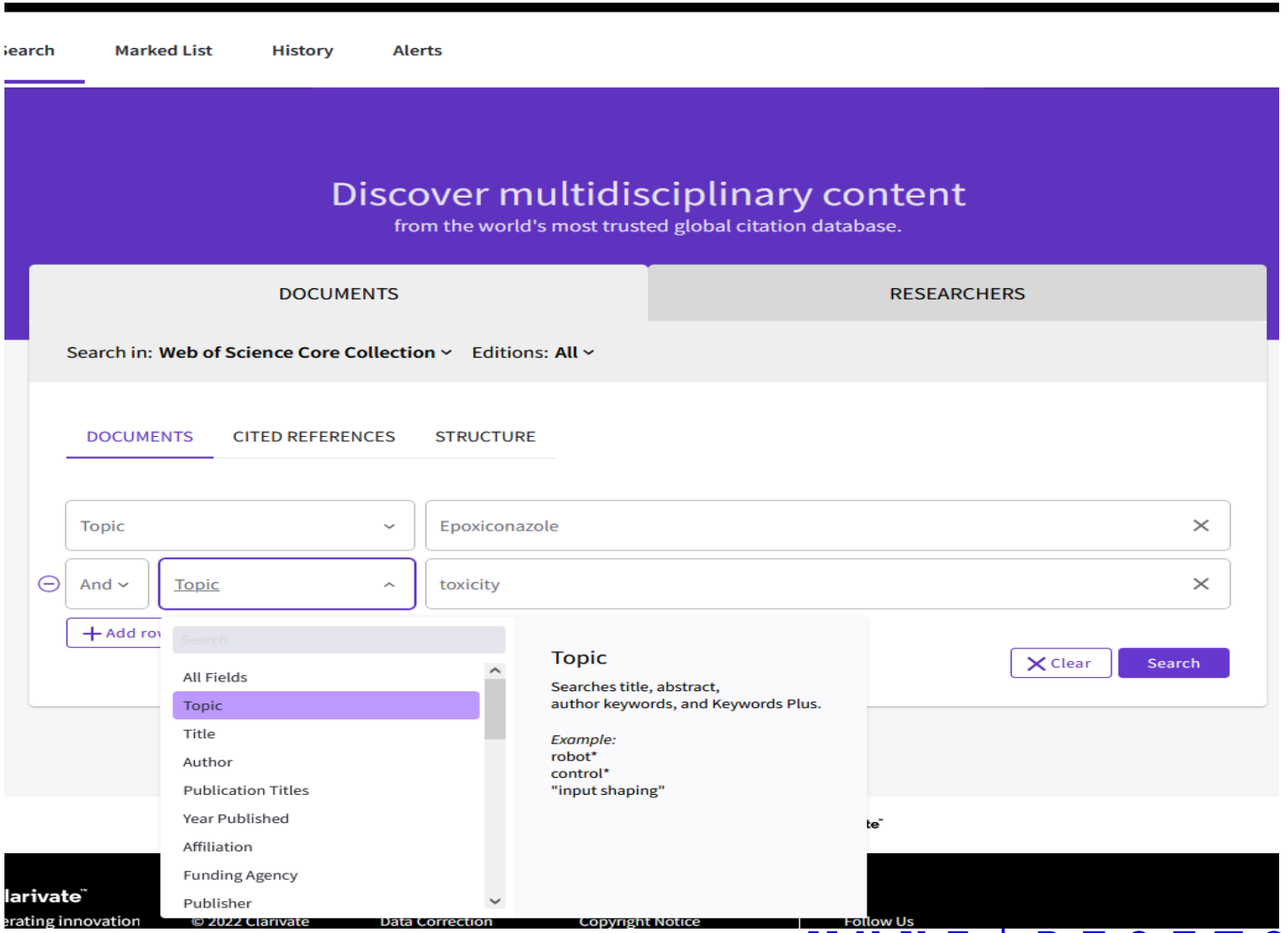


MUNI | RECETOX

Proficient search

- https://www.webofscience.com/wos/woscc/basic-search

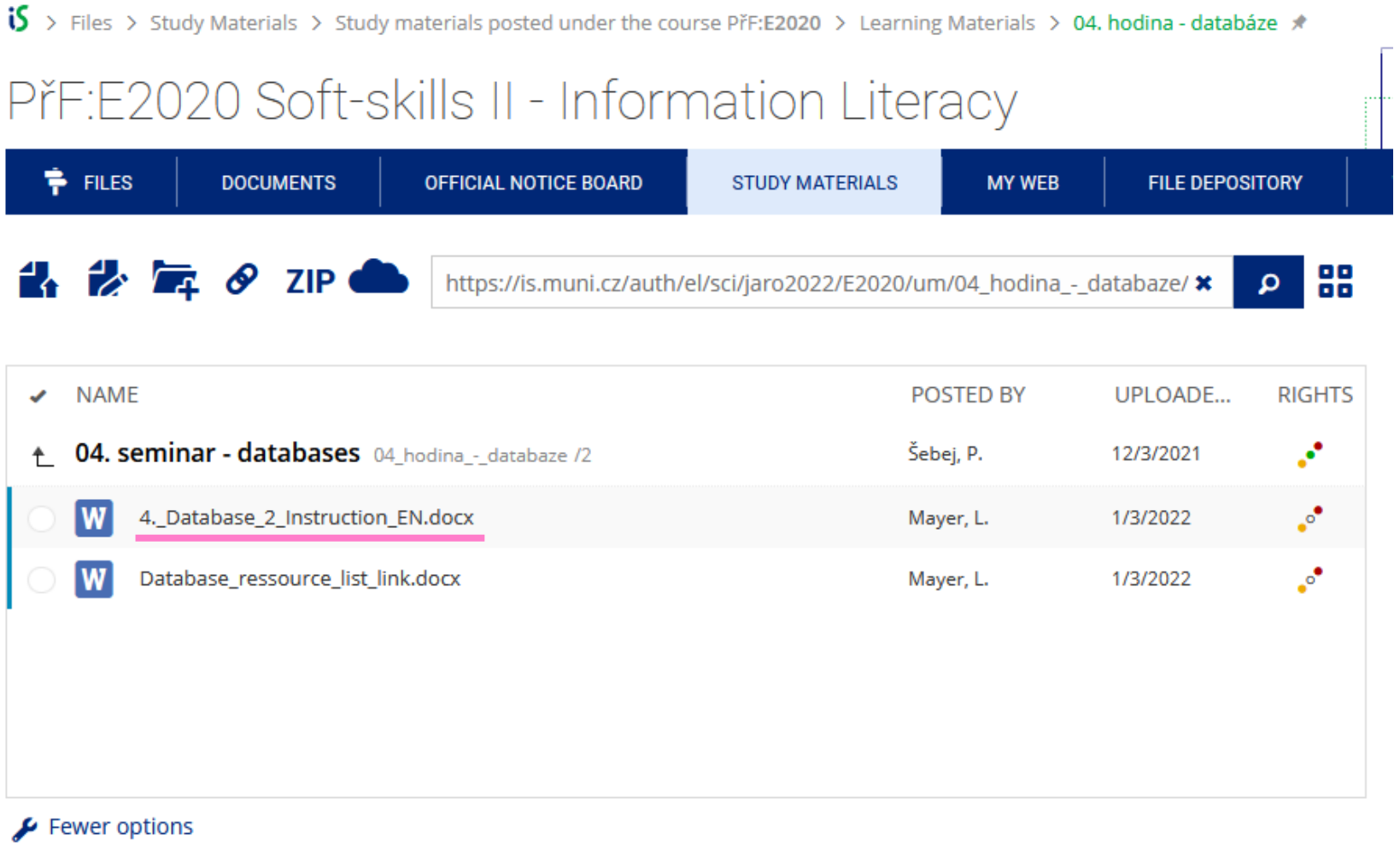
- Filter a results
  - By adding a keyword
  - Restrictions by publication type: article, review, etc.
  - Time limit: eg the last 10 years



MUNI | RECETOX

For the rest of the lecture

- IS: open word .doc
  - with list of tasks



MUNI | RECETOX

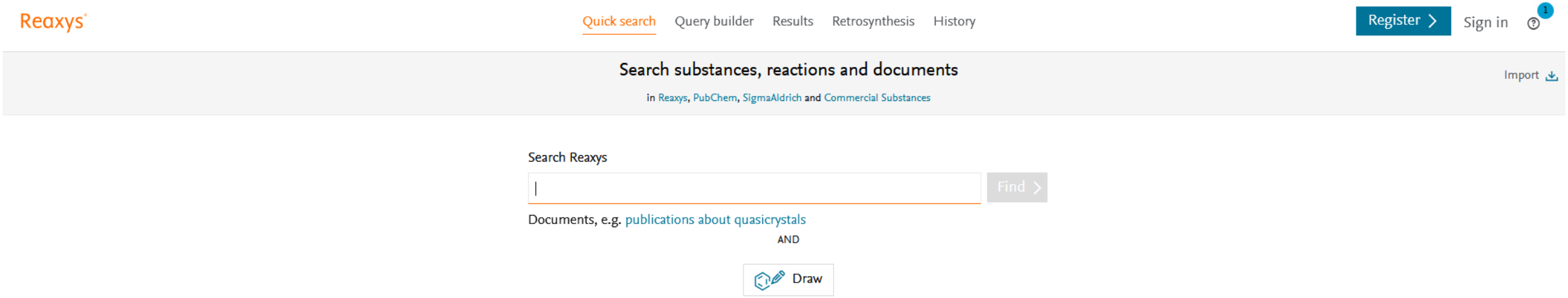
Properties databases

Task 1: Follow the link below, for Chlorpyrifos, what kind of information can we find on that website?

https://www.reaxys.com/#/structure-editor :

Properties databases

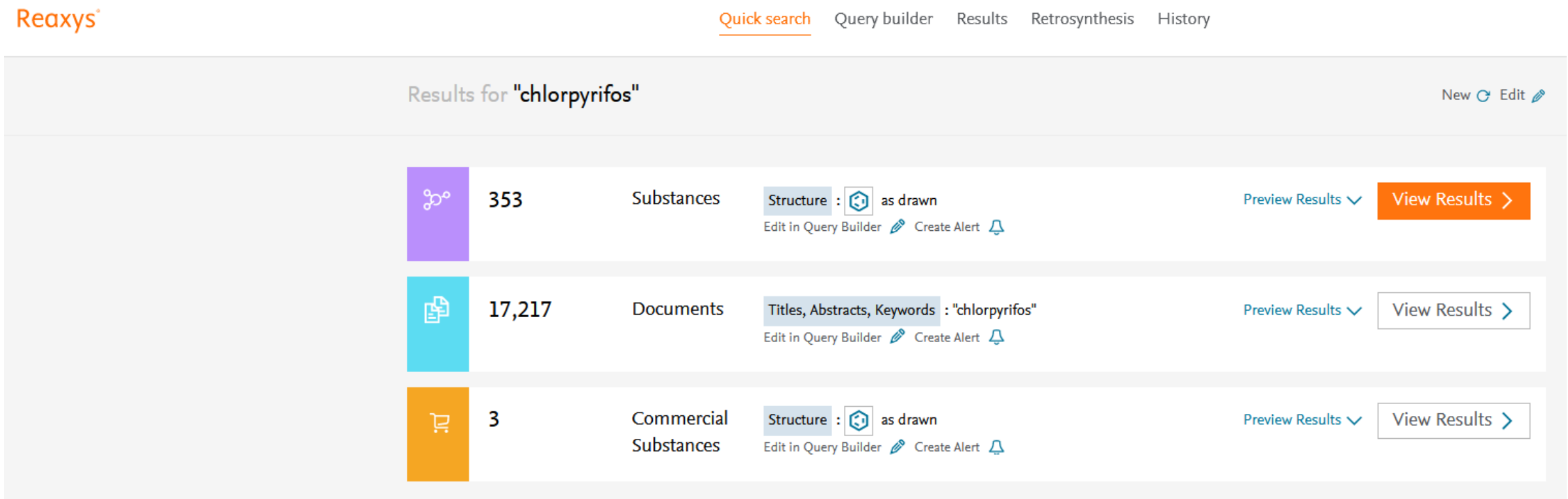
https://www.reaxys.com/#/structure-editor :



MUNI | RECETOX

Properties databases

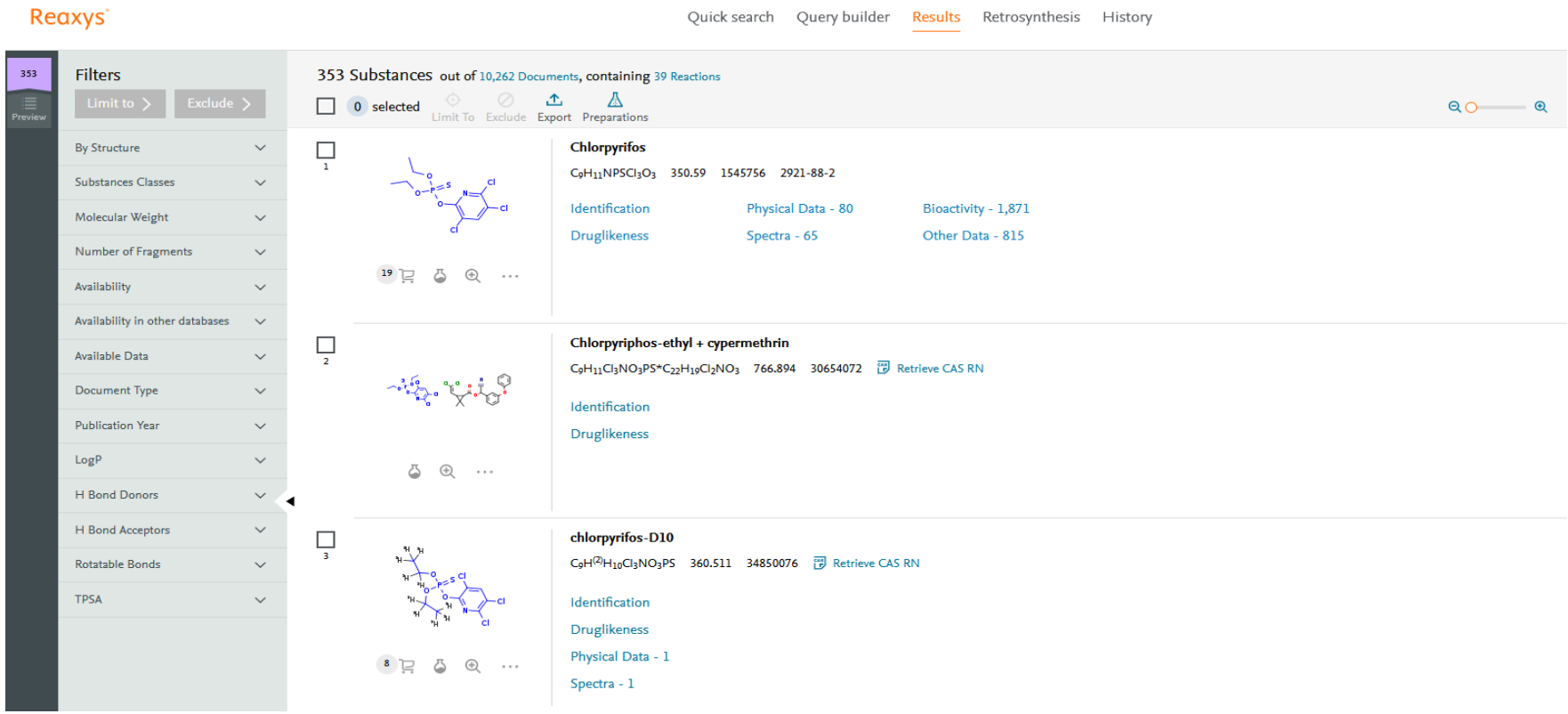
https://www.reaxys.com/#/structure-editor :



MUNI | RECETOX

Properties databases

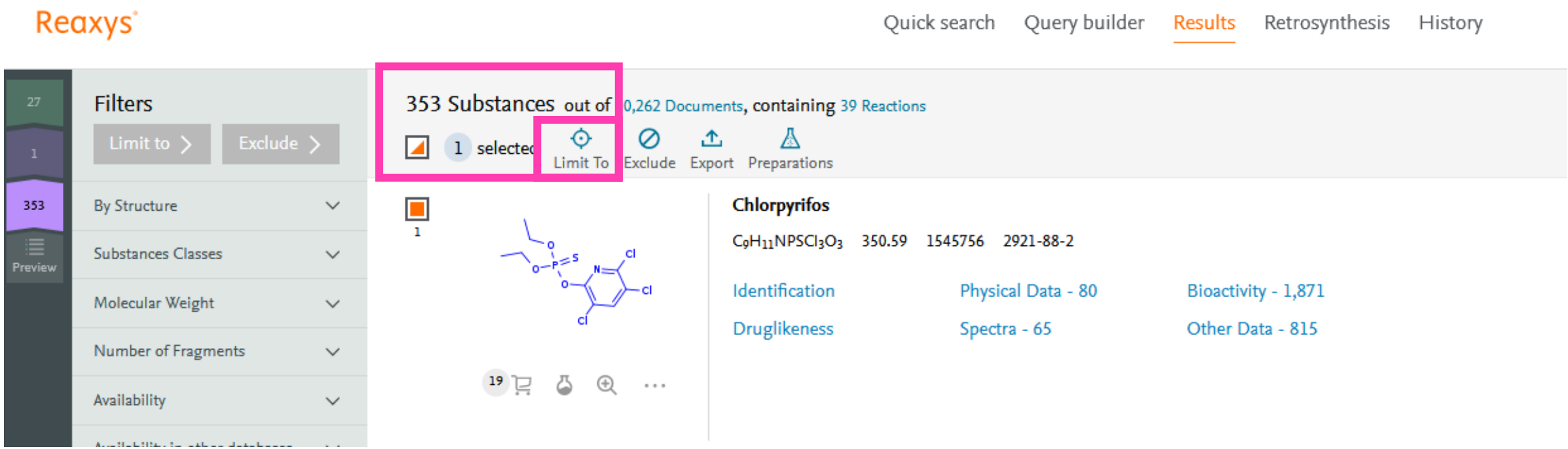
https://www.reaxys.com/#/structure-editor :



MUNI | RECETOX

Properties databases

https://www.reaxys.com/#/structure-editor :



MUNI | RECETOX

Properties databases

Task 1: Follow the link below, for Chlorpyrifos, what kind of information can we find on that website?

https://www.reaxys.com/#/structure-editor :

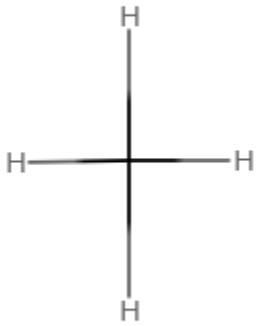
- Physical properties (molecular weight, density, melting point, log K_{ow} (log P)
- CAS RN
- Concentration in Environment
- Transport through matrices
- Distribution in organisms
- Toxicology data
- Existing reaction

Main advantage: contains references

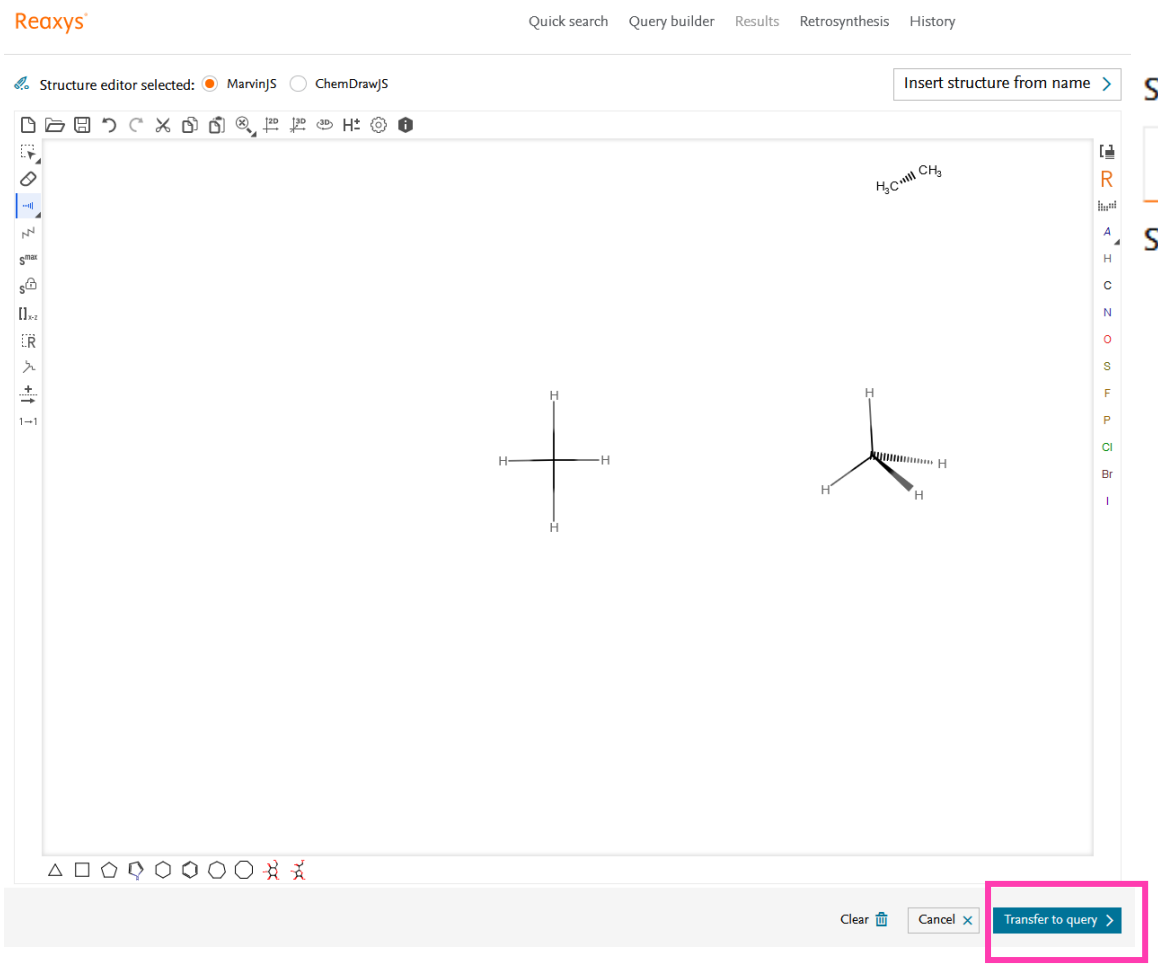
Properties databases

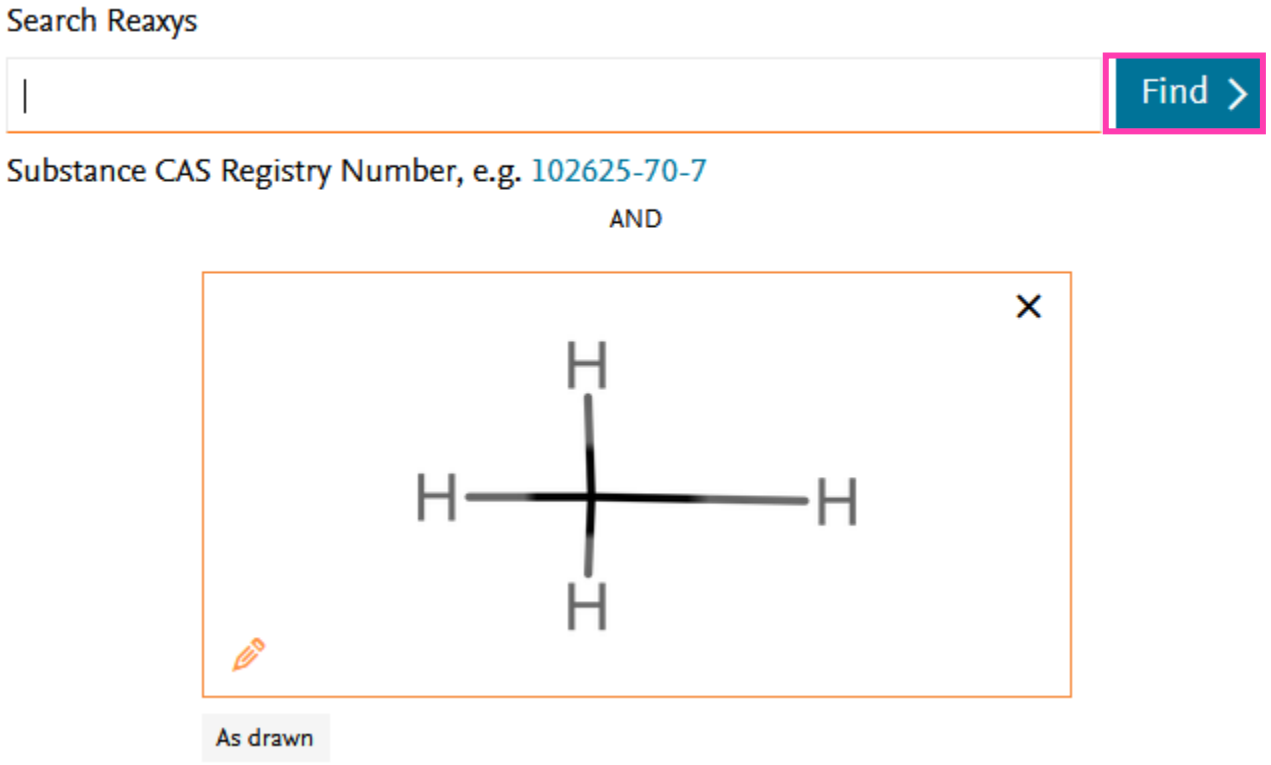
Task 2: draw a methane (CH_4) molecule with apparent bonds, in 2D and in 3D

https://www.reaxys.com/#/structure-editor :



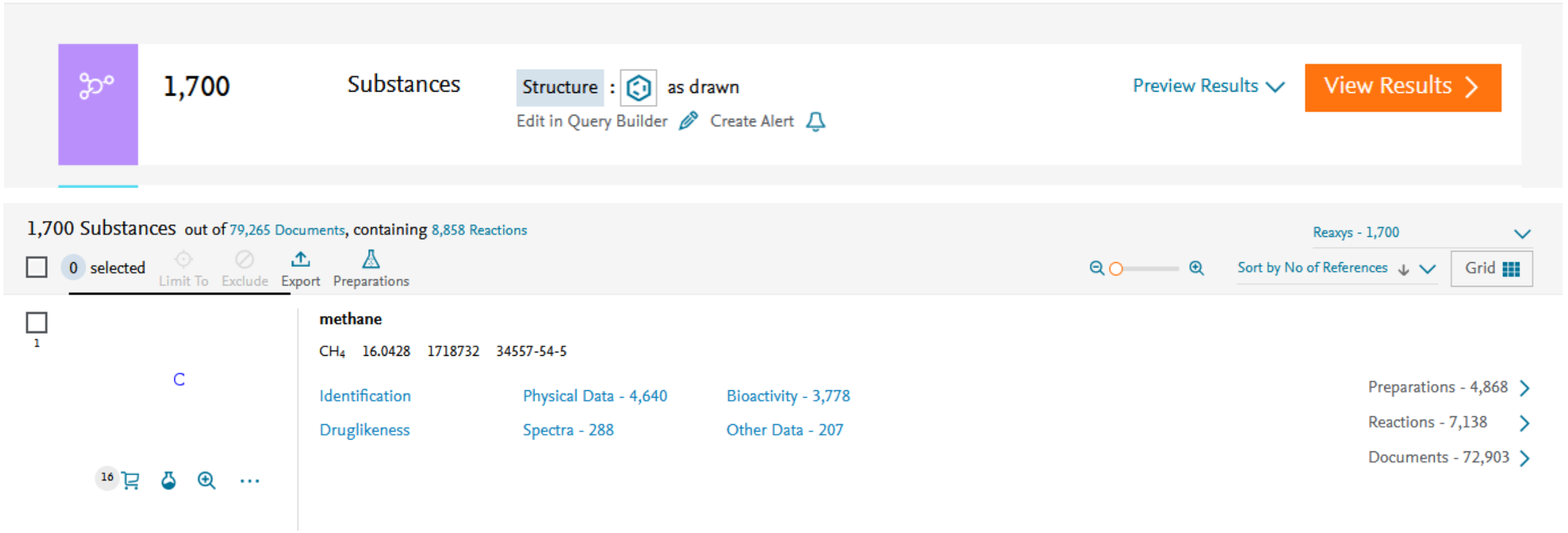
Properties databases





MUNI | RECETOX

Properties databases



MUNI | RECETOX

Properties databases

Task 3: Follow the link below and look for properties information on Fonofos (pesticide) using EpiSuite values

- if experimental value exists: give experimental, if not, give predicted values

http://www.chemspider.com/

Give information for:

- Vapor pressure:
- Log K_{ow}:
- Solubility in water:
- Half life in air:

Properties databases

Task 3: Follow the link below and look for properties information on **Fonofos (pesticide)** using EpiSuite values

- if experimental value exists: give experimental, if not, give predicted values

http://www.chemspider.com/

Why Episuite ?

Software developped by US-EPA:

- predicting physico chemical properties
- based on experimental data for compound
- and predicted values based on modelling

Properties databases

Task 3: Follow the link below and look for properties information on Fonofos (pesticide) using EpiSuite values

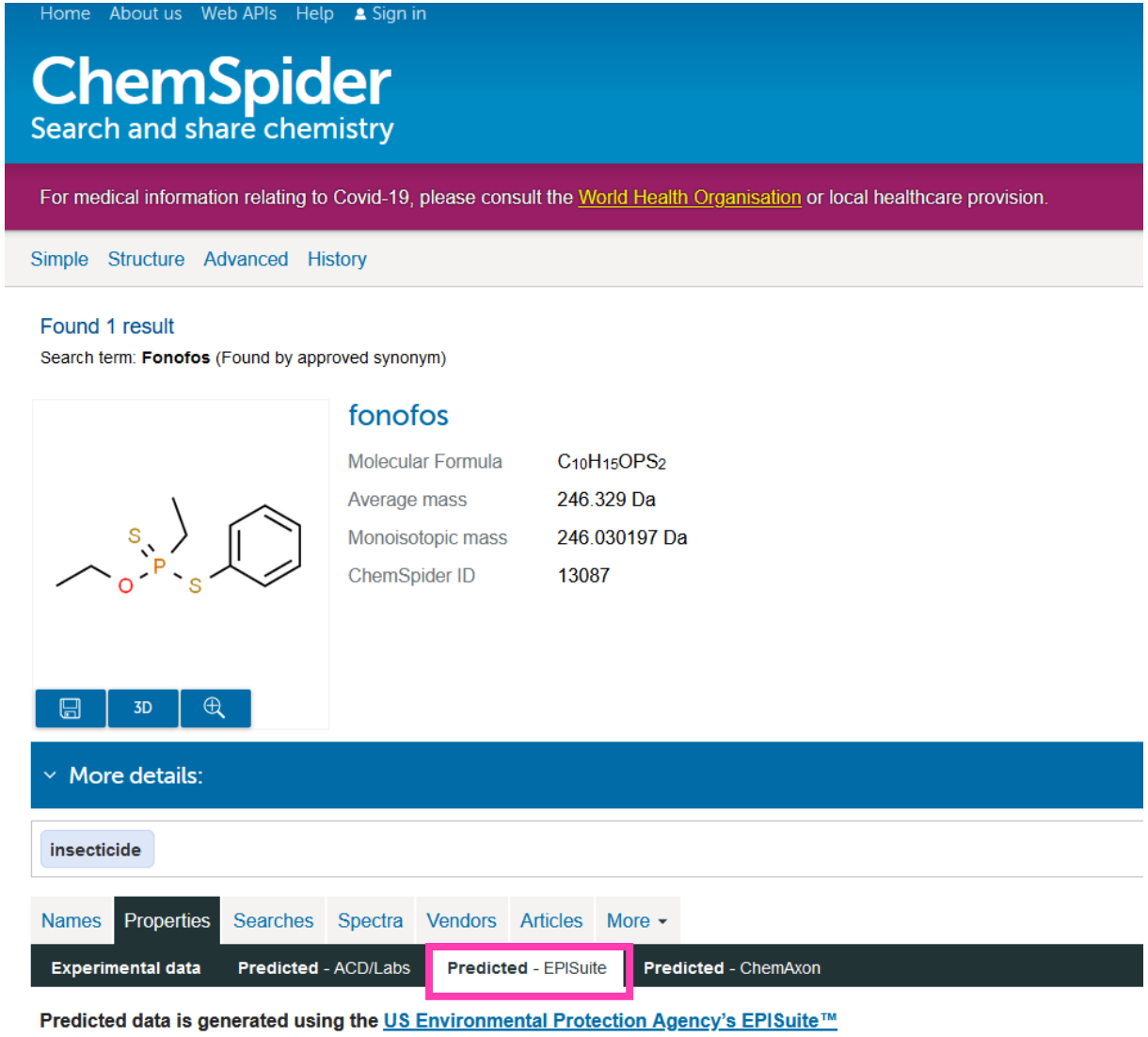
- if experimental value exists: give experimental, if not, give predicted values

http://www.chemspider.com/

Give information for:

- Vapor pressure:
- Log K_{ow}:
- Solubility in water:
- Half life in air:

Properties databases



MUNI | RECETOX

Properties databases

Task 4: Follow the link below and look for information on Fonofos (pesticide) using EpiSuite values

- if experimental value exists: give experimental, if not, give predicted values

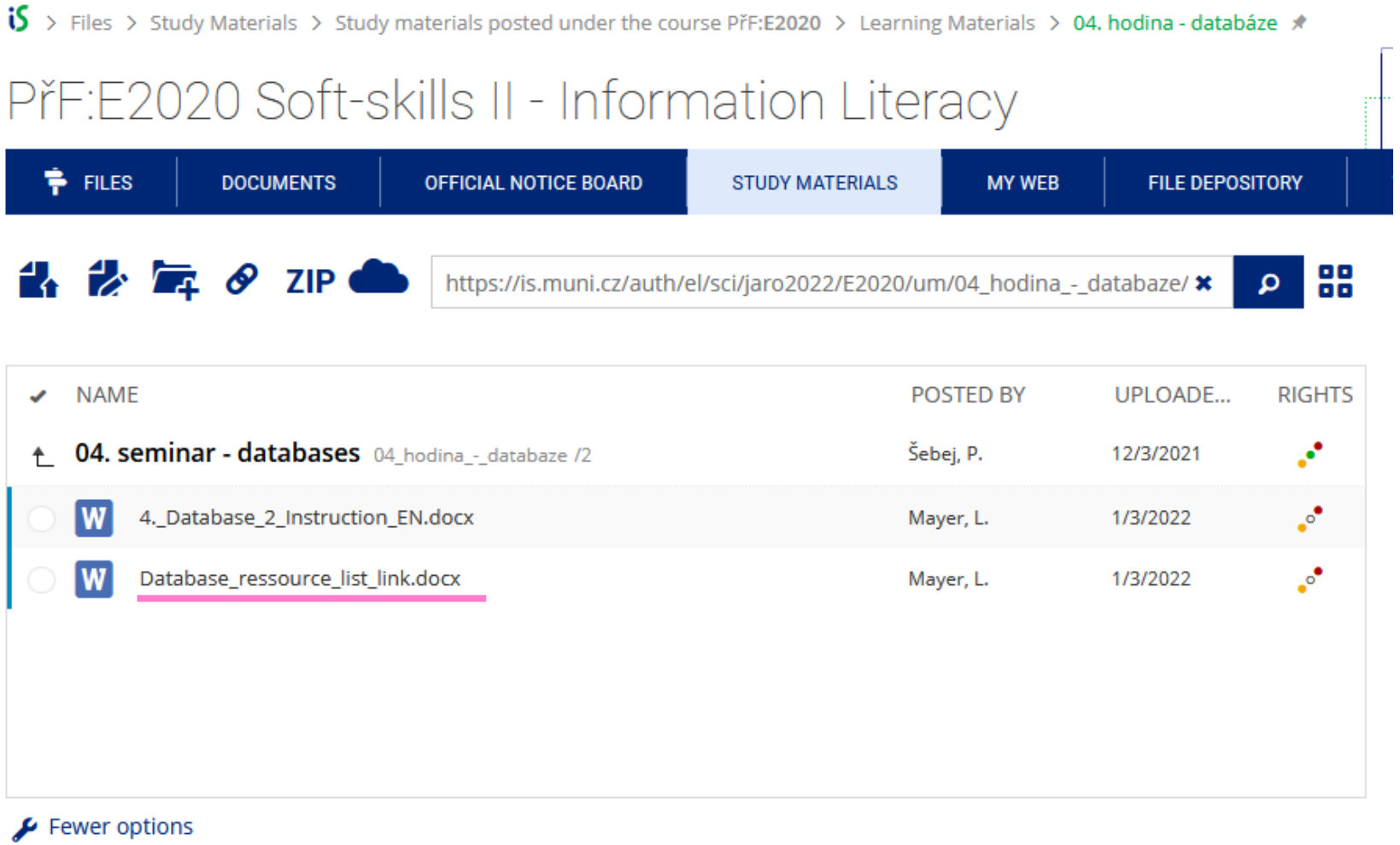
http://www.chemspider.com/

Give information for:

- Vapor pressure: 3.38 e-4 (25°C) mmHg
- Log K_{ow}: 3.94
- Solubility in water: 15.70 mg/L
- Half life in air: 1.493 hours

Also Available on IS

- IS: open word .doc
  - with list of tasks



MUNI | RECETOX

E2020: Soft-Skills 2 : Information Literacy

Ressources, a non-exaustive list of link for database and useful link

Database of information and articles:
scifinder.cas.org
reaxys.com/#/search/quick
chemspider.com
isiknowledge.com
pubchem.ncbi.nlm.nih.gov
ncbi.nlm.nih.gov
http://polysearch.ca/ (biomedical text mining to discover potential associations between various types of biomedical entities)
https://www.storkapp.me/ (alerts for papers based on keywords)

Spectra database
webbook.nist.gov/chemistry
sdbs.db.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi
echa.europa.eu/home (European Chemical agency)
MassBank (jp, eu, na) – mass spectra

Database of chemical manufacturers and sellers (just a few examples):
sigmaaldrich.com
acros.com
tcichemicals.com/en/cz
https://www.emolecules.com/ (Structure searching)
https://zinc.docking.org/ (Commercially available compound database)

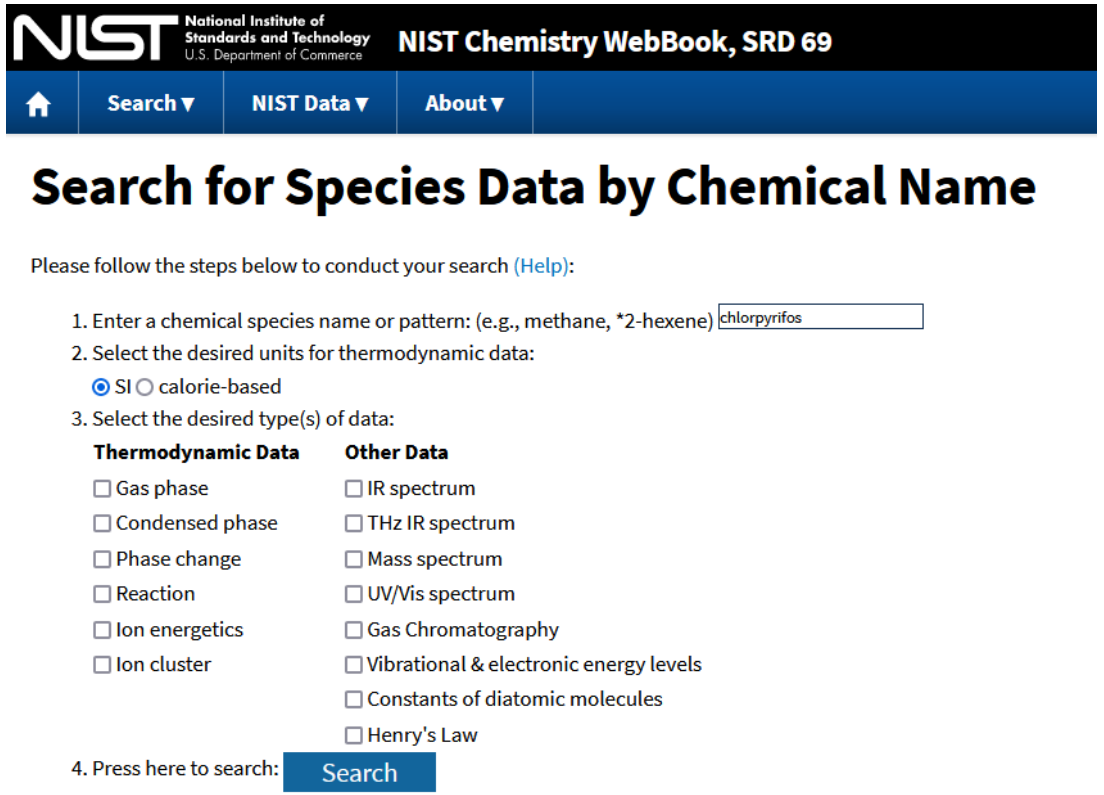
Database of biological structures and biomolecules:
www.uniprot.org
web.expasy.org/protparam
reactome.org
brenda-enzymes.or

Compounds and Spectra databases

https://webbook.nist.gov/chemistry/

- Environmental/human samples
- How to analyse them?
  - **Chromatography: to separate constituents of a sample**
  - **Mass spectrometry: to determine constituent of a sample**

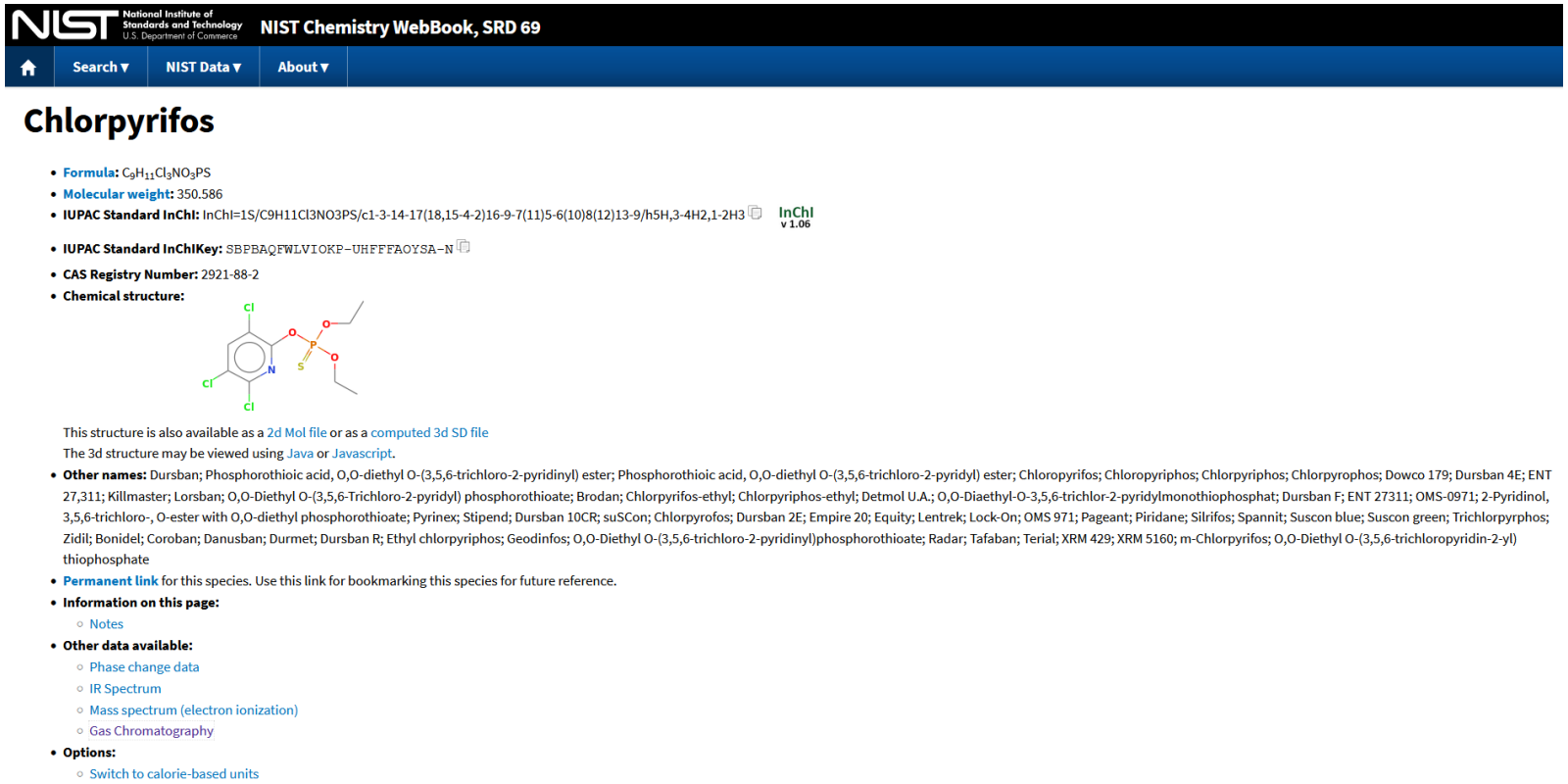
Exist databases of spectra to help identify what is present in samples



MUNI | RECETOX

Compounds and Spectra databases

https://webbook.nist.gov/chemistry/



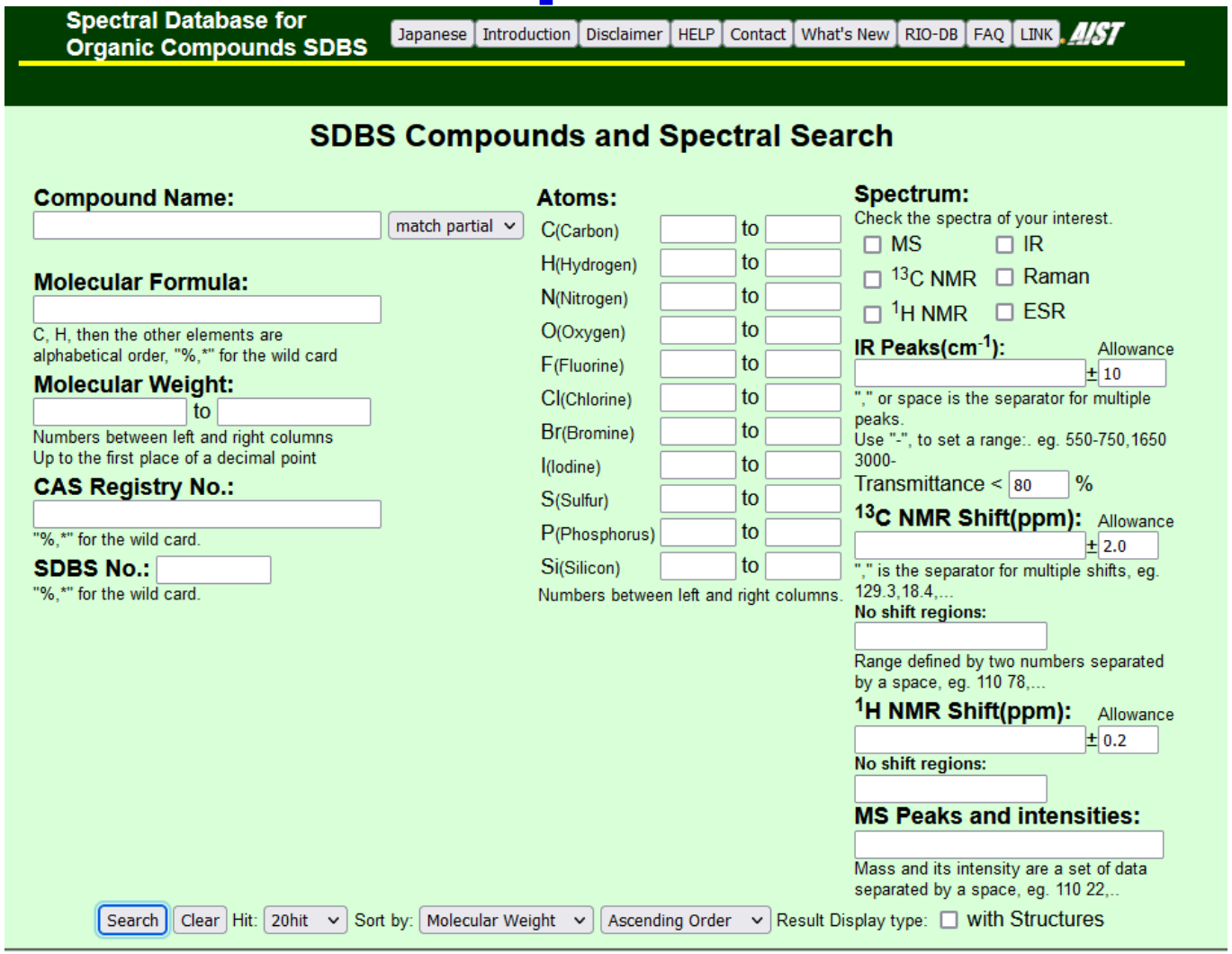
MUNI | RECETOX

Compounds and Spectra databases

Task 4: Look for Spectra for D-glucose, using only the number of atoms: $C_6H_{12}O_6$;
Copy/paste the spectra below and indicate the associated InChI of the molecule

https://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_index.cgi

Compounds and Spectra databases



MUNI | RECETOX

Compounds and Spectra databases

SDBS Search Results: 1 - 20 out of 21 hits Sort by: Molecular Weight | Ascending Order | Search

SDBS No	Molecular Formula	Molecular Weight	MS	CNMR	HNMR	IR	Raman	ESR	Compound Name
1137	C6H12O6	180.2	Y	Y	N	Y	N	N	D-mannopyranose
1139	C6H12O6	180.2	Y	Y	N	Y	N	N	D-fructose
1183	C6H12O6	180.2	N	Y	Y	Y	Y	N	D-galactopyranose
2023	C6H12O6	180.2	N	Y	Y	Y	N	N	D-glucopyranose
2046	C6H12O6	180.2	N	Y	Y	Y	Y	N	myo-inositol
2990	C6H12O6	180.2	Y	Y	Y	Y	N	N	L-(-)-sorbose
4571	C6H12O6	180.2	Y	N	N	Y	N	N	1,3-dihydroxyacetone dimer
5544	C6H12O6	180.2	Y	Y	Y	Y	N	N	D-(+)-talose
10271	C6H12O6	180.2	Y	N	N	Y	N	N	DL-glyceraldehyde dimer
11513	C6H12O6	180.2	N	N	N	Y	N	N	D-mannose
11521	C6H12O6	180.2	Y	N	N	Y	N	N	D-glucose
15696	C6H12O6	180.2	Y	Y	Y	Y	N	N	D-tagatose
16041	C6H12O6	180.2	N	Y	Y	N	N	N	beta-D-fructopyranose
16339	C6H12O6	180.2	N	Y	N	N	N	N	alpha-L-mannopyranose
22009	C6H12O6	180.2	Y	N	Y	Y	N	N	L-(-)-mannose
32639	C6H12O6	180.2	Y	N	N	Y	N	N	alpha-L-sorbopyranose
32672	C6H12O6	180.2	Y	N	N	Y	N	N	beta-D-glucopyranose
50268	C6H12O6	180.2	Y	N	N	Y	N	N	D-allose
35361	C6H12CuN2O6	271.7	N	N	N	Y	N	N	trans-bis(DL-serinato)copper(II)
35377	C6H12CuN2O6	271.7	N	N	N	Y	N	N	trans-bis(L-serinato)copper(II)

1 2

Compounds and Spectra databases

SDBS Information

SDBS No.: 11521

Compound Name:
D-glucose

Molecular Formula: $C_6H_{12}O_6$

Molecular Weight: 180.2

CAS Registry No.:
50-99-7

Spectral Code:
Mass :
IR : nujol mull

Chemical Information:

Return to Search:
Return to Result:

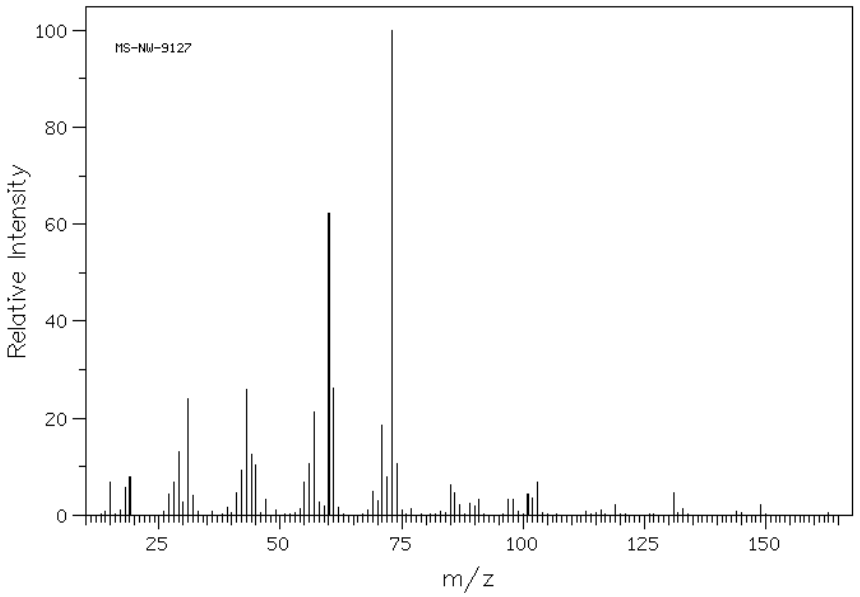
URL for this Compound:
https://sdbs.db.aist.go.jp/sdbs/cgi-bin/landingpage?sdbsno=11521

External Information:
external link displays in a separate page

- Japan Chemical Information Link Center (in Japanese)
  - JST Nikkaji Web (in Japanese)

SDBS-Mass

MS-NW-9127 SDBS NO. 11521
D-glucose
C6H12O6 (Mass of molecular ion: 180)



Source Temperature: 200 °C
Sample Temperature: 140 °C
Direct, 75 eV

peak data

MS-NW-9127 SDBS NO. 11521

DOI:

Sample InChI:
InChI=1S/C6H12O6/c7-1-3(9)5(11)6(12)4(10)2-8/h1,3-6,8-12H,2H2/t3-,4+,5+,6+/m0/s1

Sample InChIKey:
GZCGUPFRVQAUEE-SLPGGIOYSA-N

MUNI | RECETOX

Small molecule mass spectral databases

- Typically mass spectra compound are shared as text files with limited ease of use

- Spectral matching is a key component of a compound identification

NAME: Phosmet
RETENTIONTIME: 25.331
PRECURSORMZ: 316.992126464844
PRECURSORTYPE: [M]+
IONMODE: Positive
COLLISIONENERGY: 70
FORMULA: C10H7NO4
SMILES: COP(=S)(OC)SCN1C(=O)C2=CC=CC=C2C1=O
INCHIKEY: WQINSVOOIJDOLJ-UHFFFAOYSA-N
Num Peaks:15

76.03039	938223
77.03818	1629764
78.91741	417336
78.99396	406285
79.05384	448837
93.00958	2273096
104.02512	1400371
105.03298	1295700
124.94133	446127
124.98142	449841
130.02811	1679535
133.02771	3772589
160.03839	32462268
161.0417	3305389
192.01024	413541

Elliott J. Price and his team (RCX)

Small molecule mass spectral databases

- https://massbank.eu/ : hosts spectra for around 15 000 compounds
  - Pros: diverse instrumentation and ionization modes
  - Cons: intensity of spectra are very different from what we can encounter in biological or environmental samples

- https://metlin.scipps.edu/ : hosts spectra for 1 million compounds
  - Pro: easy search, good indexing
  - Con: prioprietary spectra, not available for download, only for LC-MS/MS

Small molecule mass spectral databases

- https://gnps.ucsd.edu/: hosts spectra for around 255 000 compounds
  - Pros: easy to compare spectra with many tools & links to molecular networking
  - Cons: contain natural products,
    uses MGF text format which loses information such as chromatographic retention

- https://hmdb.ca/: hosts spectra for 100 000 compounds
  - Pro: integrated with many other databases and user-friendly tools
  - Con: merged in silico and experimental spectra
    only cover humans

Content

– 2nd part: Practical – you are at the helm and you find the information!

- Publication databases
- Properties databases
- Spectral databases
- Biological databases
- Registration databases
- Patent database

Last week

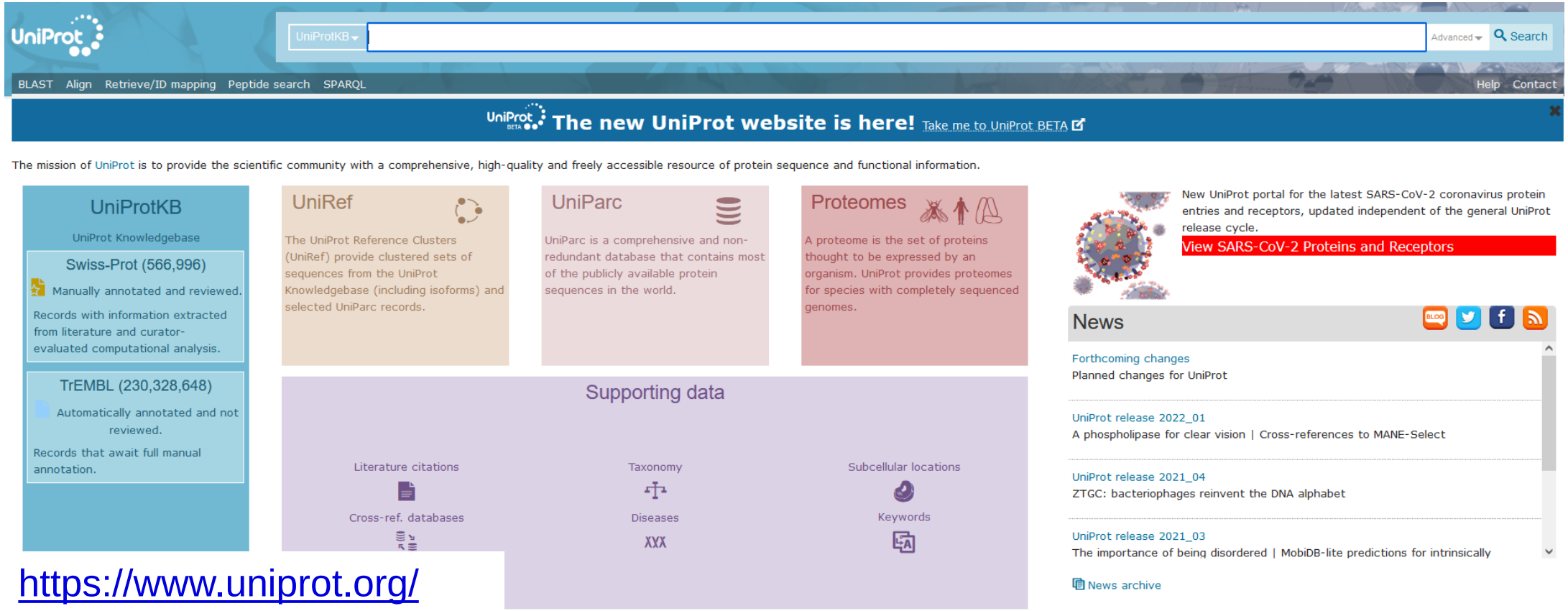
This week

Biological databases

- Protein and DNA databases
- They work similarly to chemical DB
- Focus on biomolecules and biologically important objects

Protein databases

– Sequence, structure, function, domains, localization, interactions, related proteins...



https://www.uniprot.org/

MUNI | RECETOX

Protein databases

UniProtKB - P04554 (PRM2_HUMAN)

Display | Help video

Entry
Publications
Feature viewer
Feature table

None
Function
Names & Taxonomy
Subcellular location
Pathology & Biotech
PTM / Processing
Expression
Interaction
Structure
Family & Domains
Sequences (2)
Similar proteins
Cross-references
Entry information
Miscellaneous

Top

BLAST | Align | Format | Add to basket | History

Protein **Protamine-2**
Gene **PRM2**
Organism *Homo sapiens (Human)*
Status Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level

Function

Protamines substitute for histones in the chromatin of sperm during the haploid phase of spermatogenesis. They compact sperm DNA into a highly condensed, stable and inactive complex.
By similarity

GO - Molecular function

- cadmium ion binding — Source: CAFA
- DNA binding — Source: ProtInc
- zinc ion binding — Source: CAFA

Complete GO annotation on QuickGO ...

GO - Biological process

- chromosome condensation — Source: UniProtKB-KW
- DNA packaging — Source: ProtInc
- nucleus organization — Source: GO_Central
- spermatid development — Source: GO_Central
- spermatogenesis — Source: UniProtKB

Complete GO annotation on QuickGO ...

Keywords

Molecular function	Developmental protein, DNA-binding

MUNI | RECETOX

Protein databases

Sequences (2)

Sequence status: Complete.
Sequence processing: The displayed sequence is further processed into a mature form.
This entry describes **2** isoforms produced by **alternative splicing**. Align Add to basket

Isoform 1 (identifier: **P04554-1**) [UniParc] FASTA Add to basket

This isoform has been chosen as the canonical sequence. All positional information in this entry refers downloadable versions of the entry.
« Hide

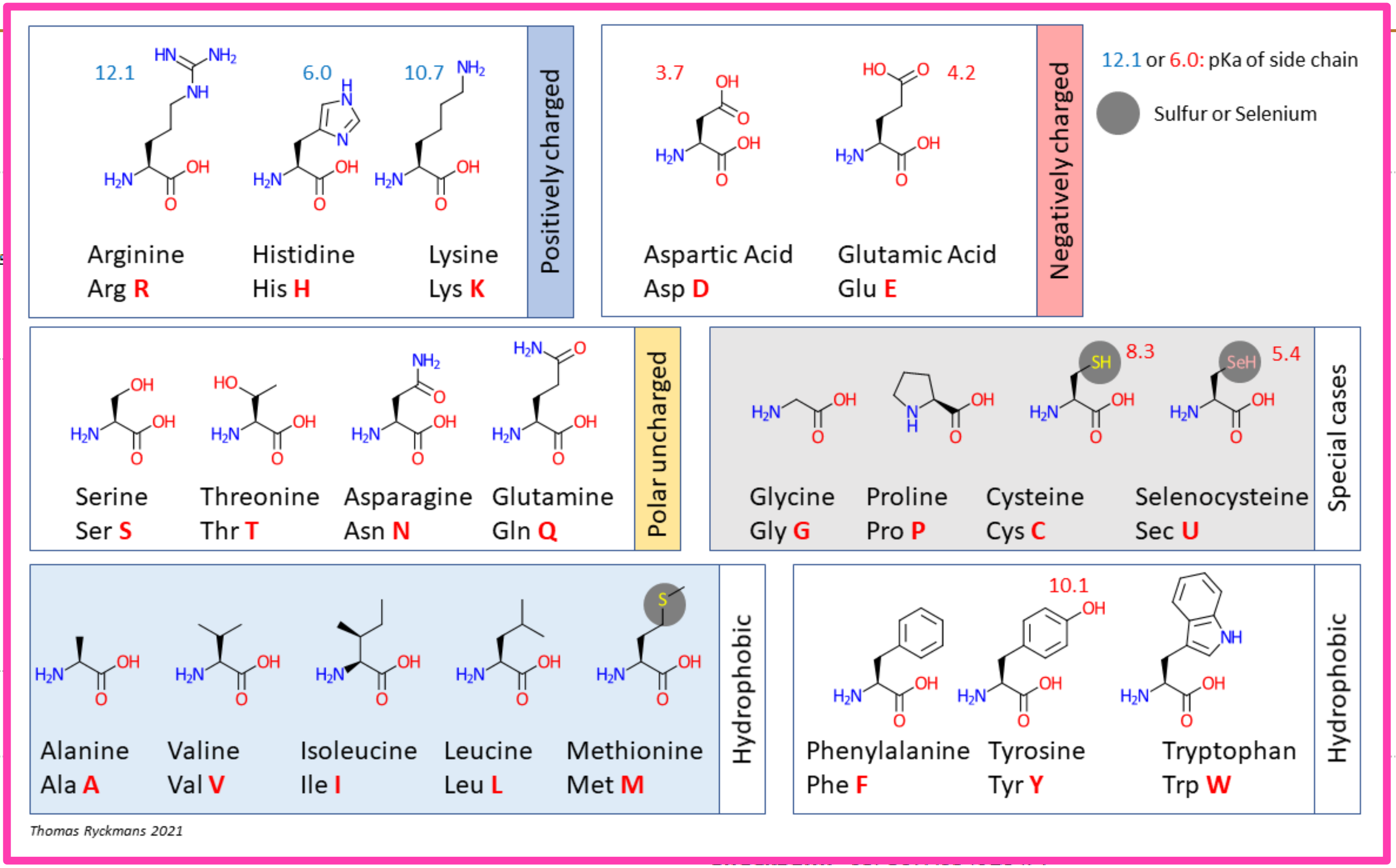
```
        10         20         30         40         50
MVRYRVRSLS ERSHEVYRQQ LHGQEQGHHG QEEQGLSPEH VEVYERTHGQ
        60         70         80         90        100
SHYRRRHCSR RRLHRIHRRQ HRSCRRRKRR SCRHRRRHRR GCRTRKRTCR

RH
```

Isoform 2 (identifier: **P04554-2**) [UniParc] FASTA Add to basket

The sequence of this isoform differs from the canonical sequence as follows:
91-102: GCRTRKRTCRRH → ESLGDPLNQNFLSQKAAEPGREHAEGTKLPGPLTPSWKLRKSRPKHQVRP

Show »



BLAST GO

MUNI | RECETOX

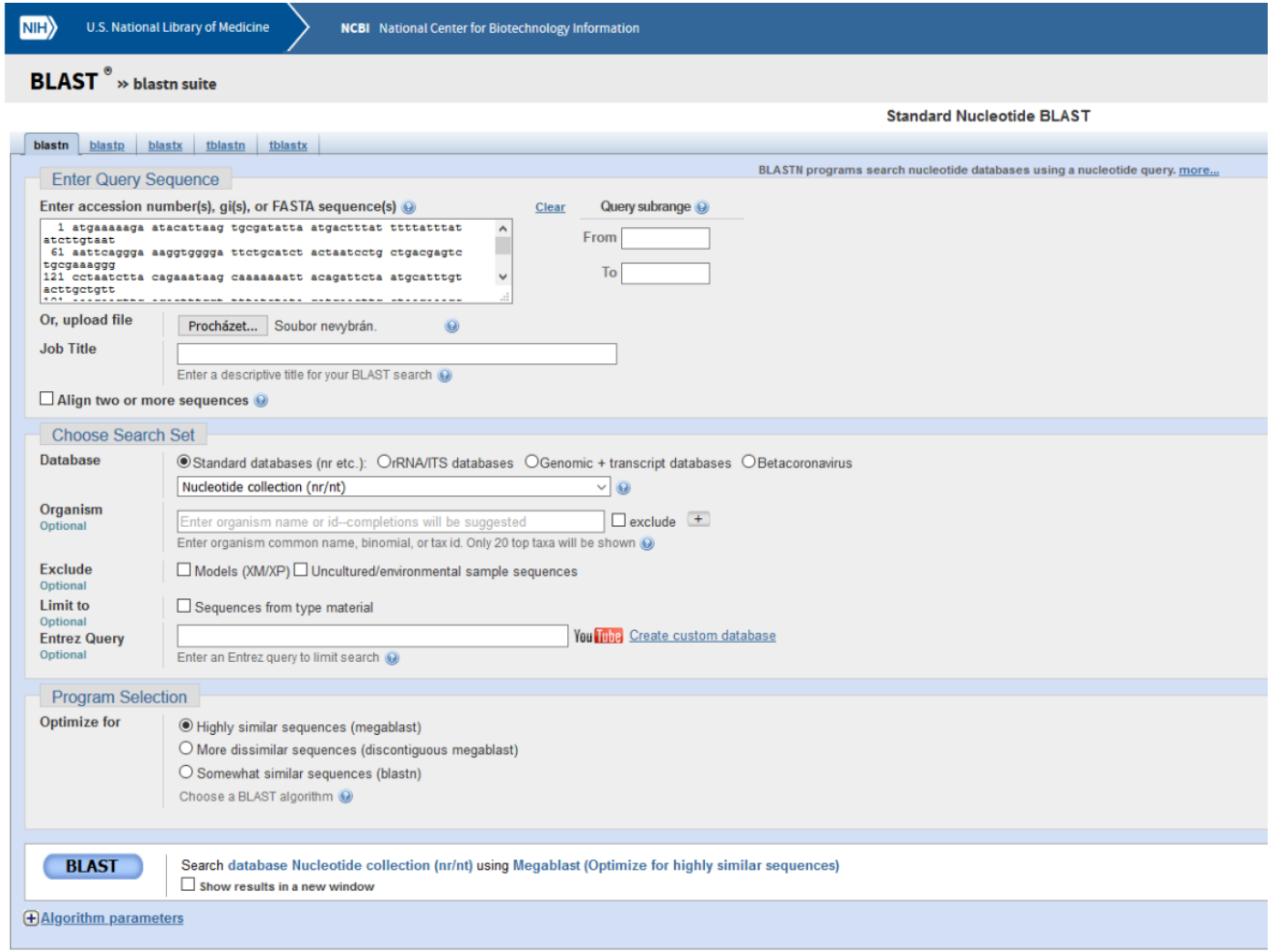
Protein databases

- Sequence, structure, function, domains, localization, interactions, related proteins…
- UniProt KB - aggregation of other databases, Specialized
- Physico-chemical parameters from the sequence - ProtParam
- Biological pathways - Reactome
- Enzymes - Brenda
- Proteomics - ProteomicsDB
- Posttranslational modifications - iPTMnet
- Domains - Pfam
- And many others

MUNI | RECETOX

DNA sequence database

- BLAST =Basic Local Alignment Search Tool
- Search and comparison of sequences
- → GenBank
- Finding protein from DNA sequence
- Also for exists proteins (BLAST protein)



MUNI | RECETOX

BLAST

- BLAST =Basic Local Alignment Search Tool

- Widely used sequence similarity search tool
  - Finds high scoring local alignments between two sequences (Protein or DNA)

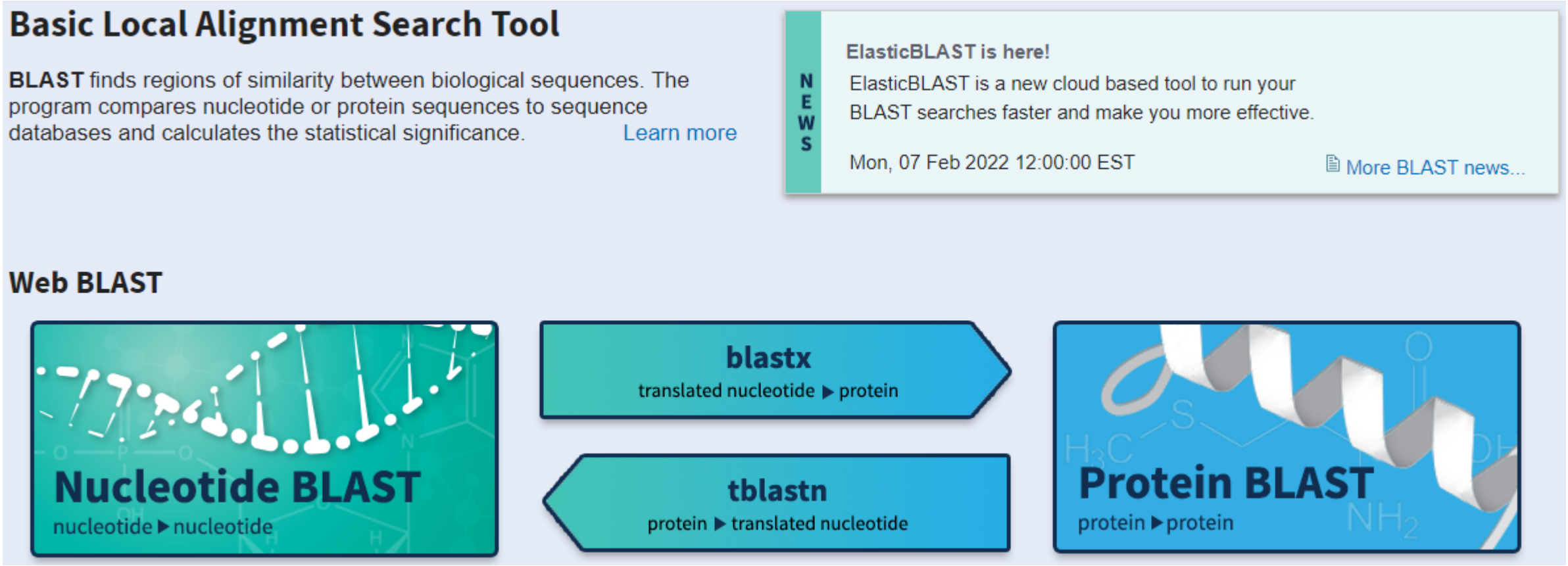
BLAST

I have a DNA or protein sequence

- What is it related to? What does it do? (Homology, conserved domains)
- Is it already in the database (identification)?
  - Find matching sequence in the database or organism of origin
- Where is it located? How is it organized in a genome?

BLAST

https://blast.ncbi.nlm.nih.gov/Blast.cgi



MUNI | RECETOX

BLAST

Task 5: Identification search:

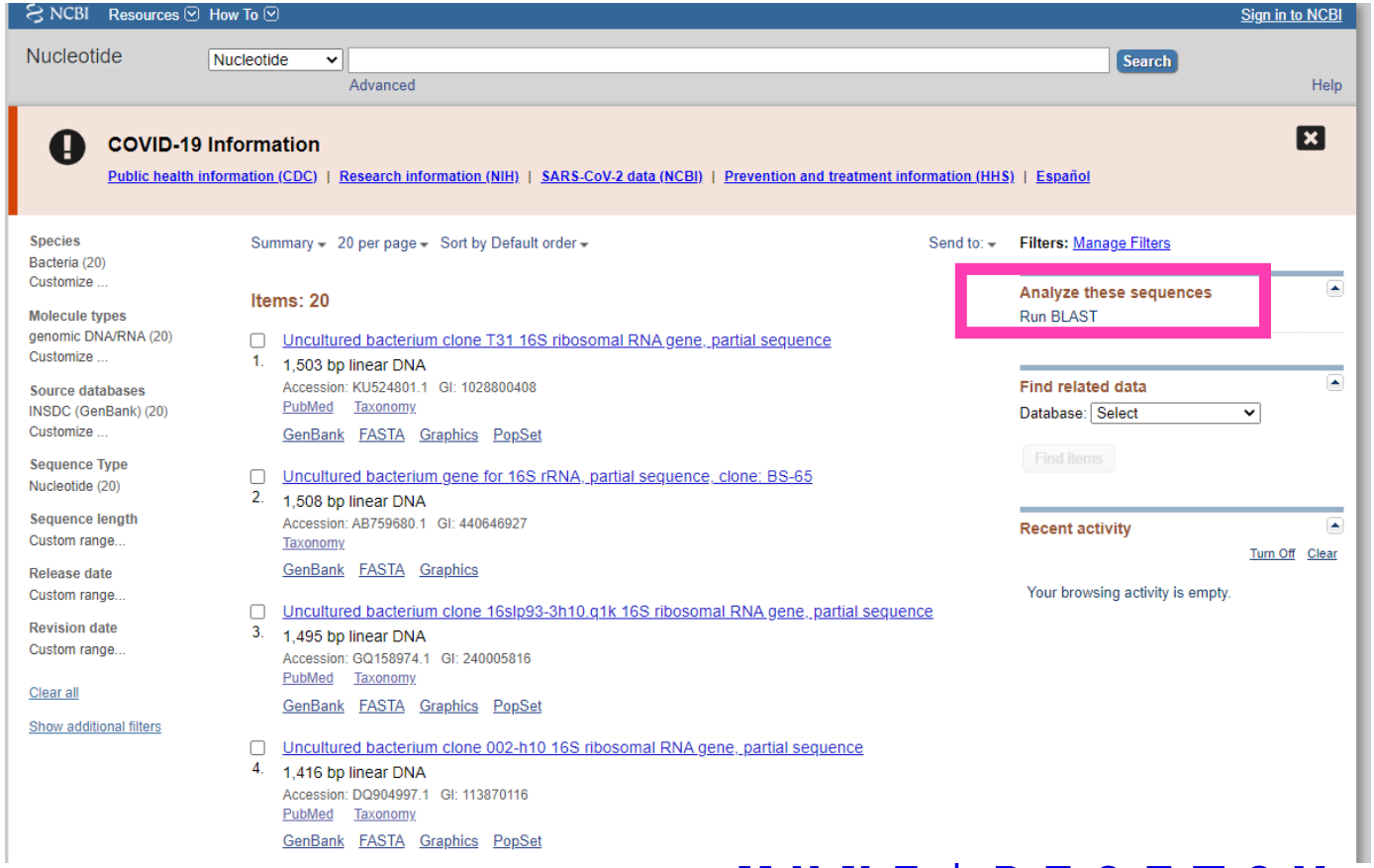
- Look within the 16S rRNA type strains database
- Visualize distance trees of results
- Manipulate tree viewer display
- Use MOLE-BLAST tool

- https://blast.ncbi.nlm.nih.gov/Blast.cgi

BLAST

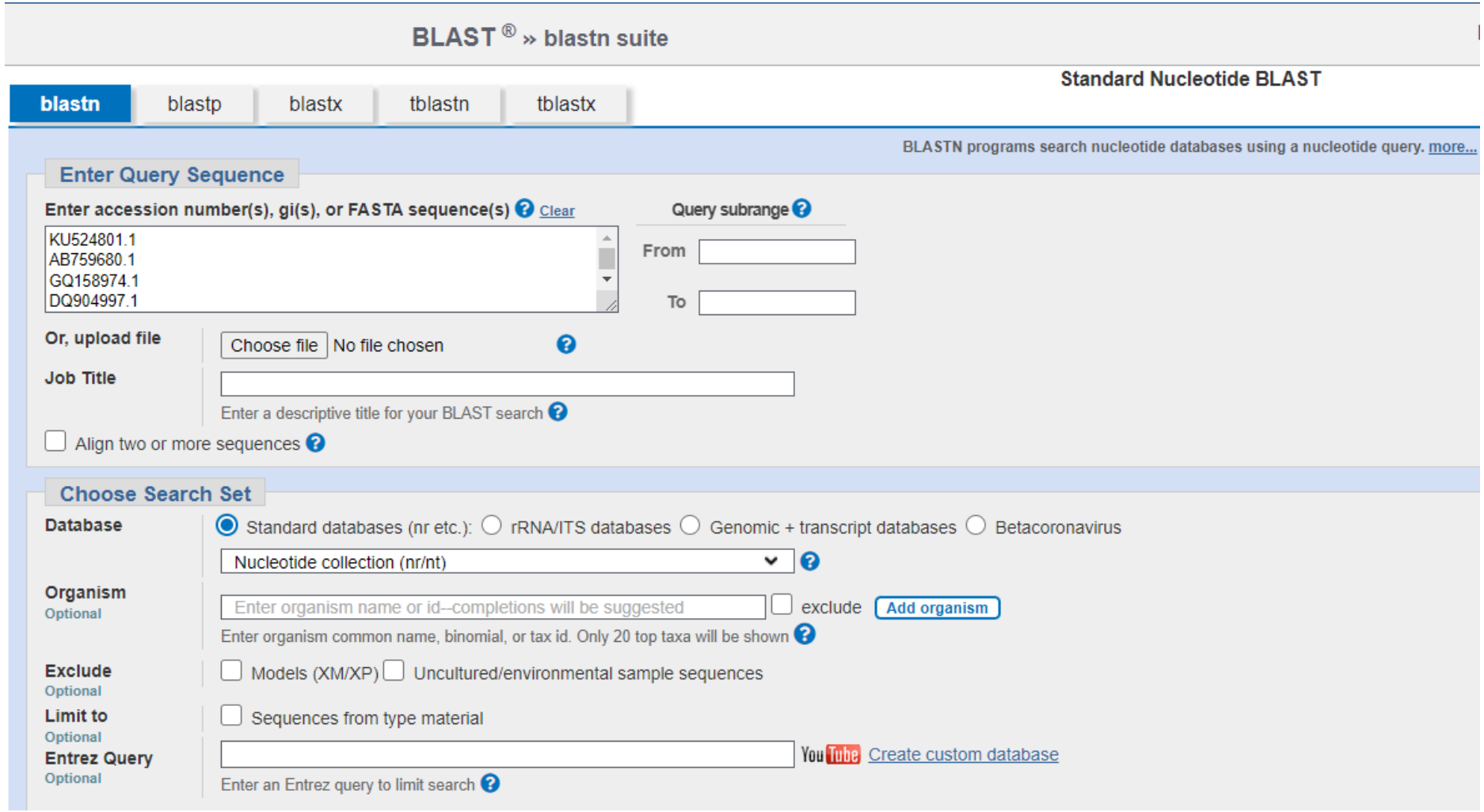
20 uncultured bacterias; Link to retrieve data: https://go.usa.gov/xUEX4

Available on IS:
Word document:



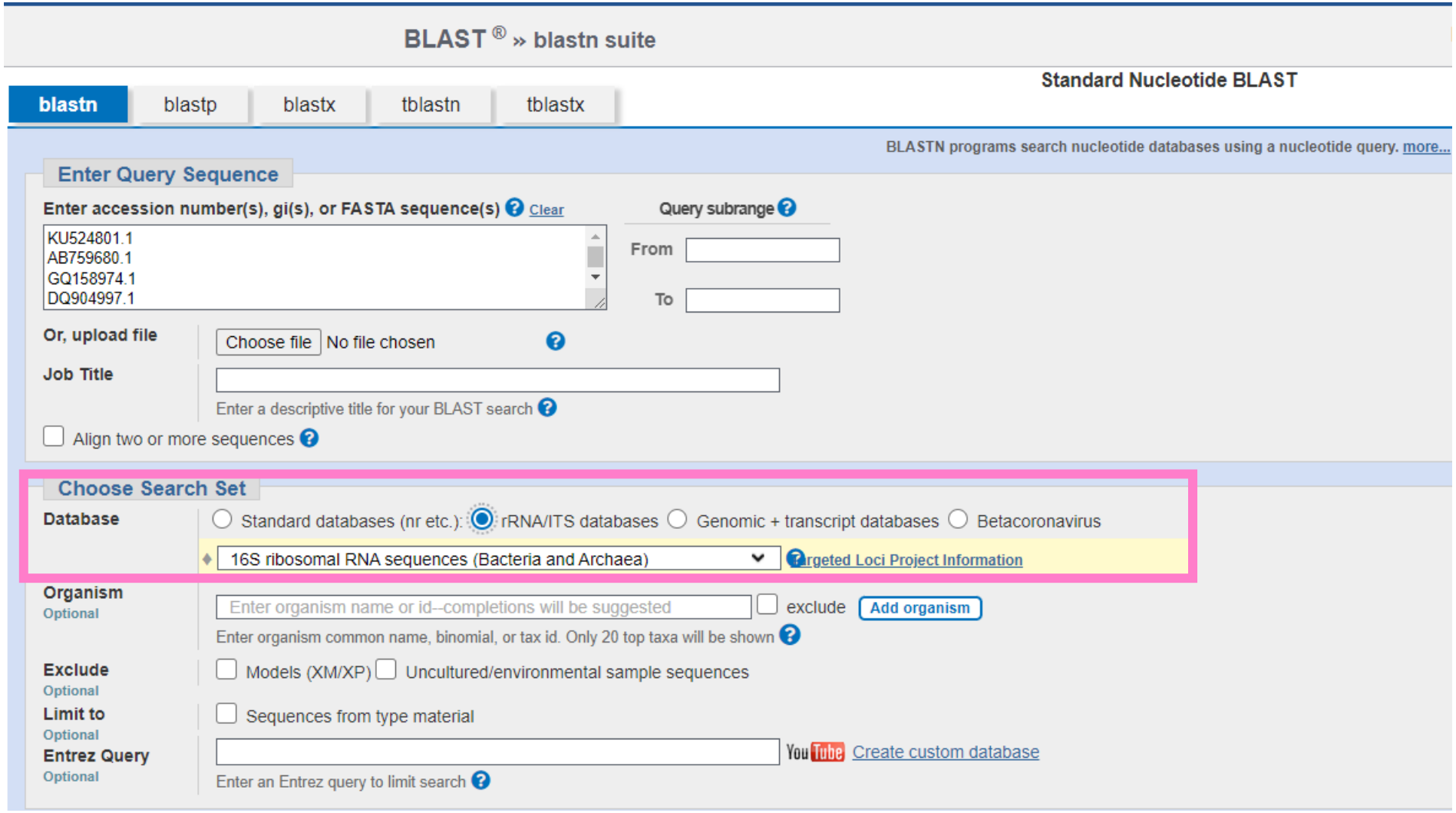
MUNI | RECETOX

BLAST



MUNI | RECETOX

BLAST



MUNI | RECETOX

BLAST

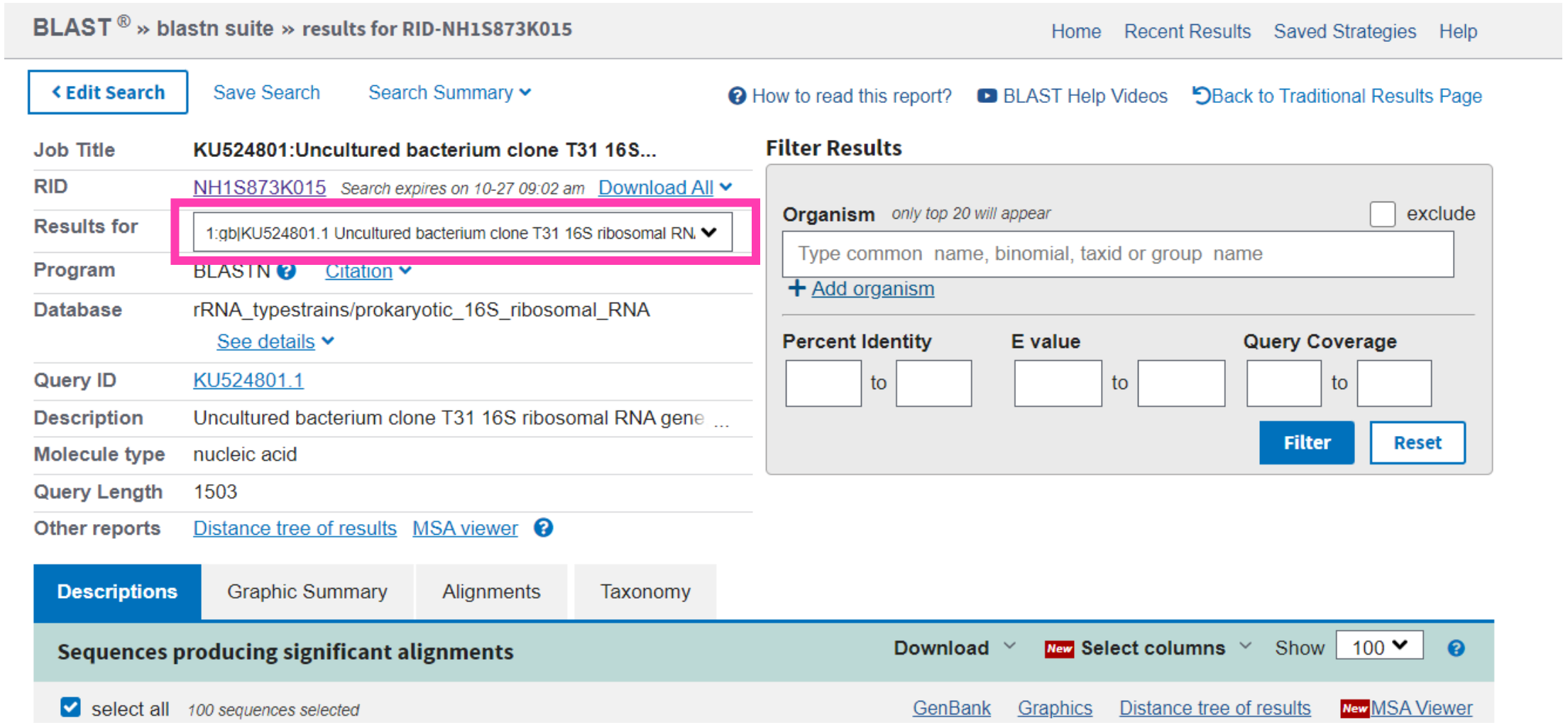
Job Title: gb|KU524801.1|

Request ID	**1YBMVEFY016**
Status	Searching
Submitted at	Wed Mar 2 03:45:15 2022
Current time	Wed Mar 02 03:45:20 2022
Time since submission	00:00:05

This page will be automatically updated in **2** seconds

Link to get results:
https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&RID=NH1S873K015

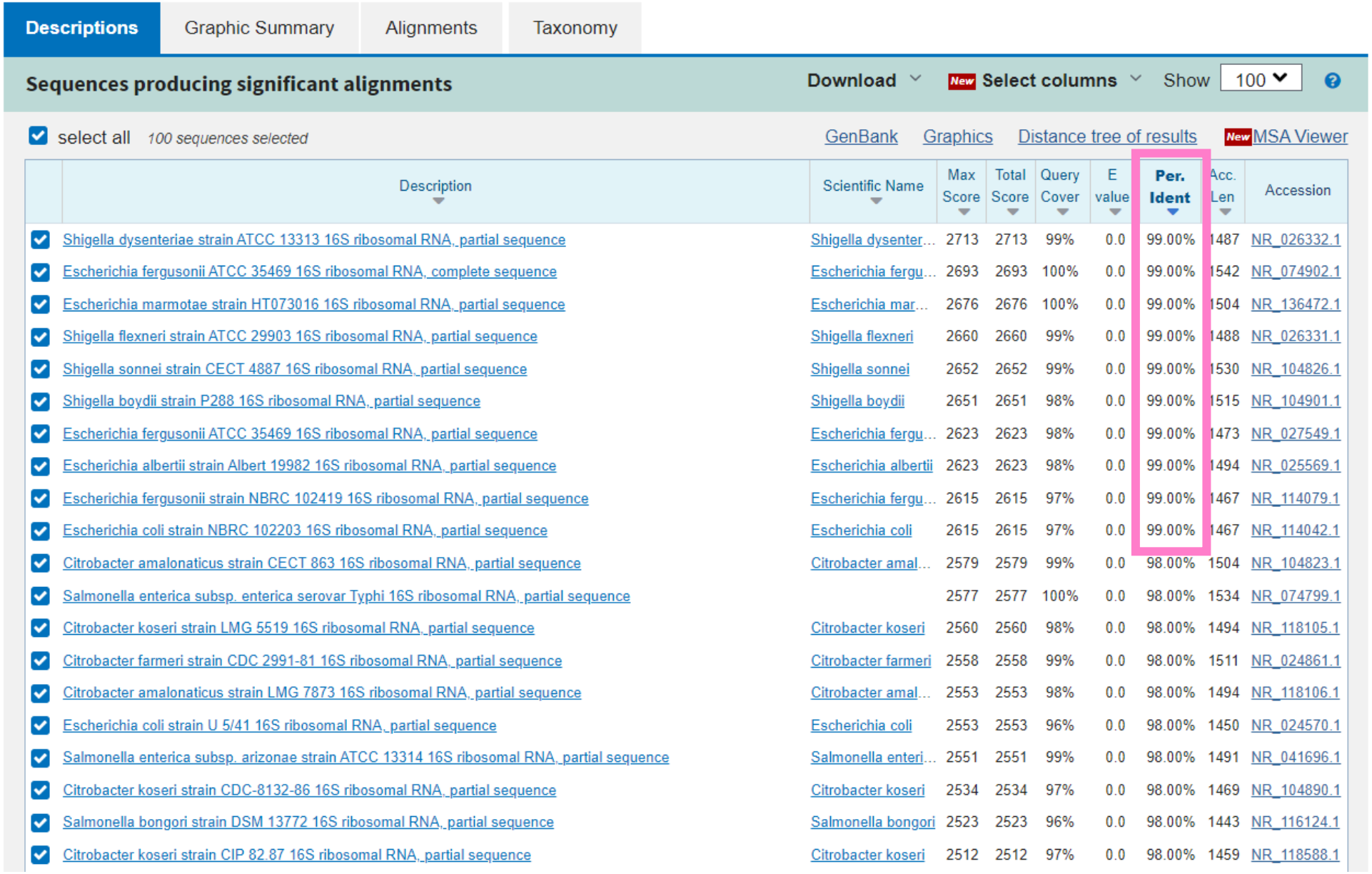
BLAST



MUNI | RECETOX

BLAST

– Identification



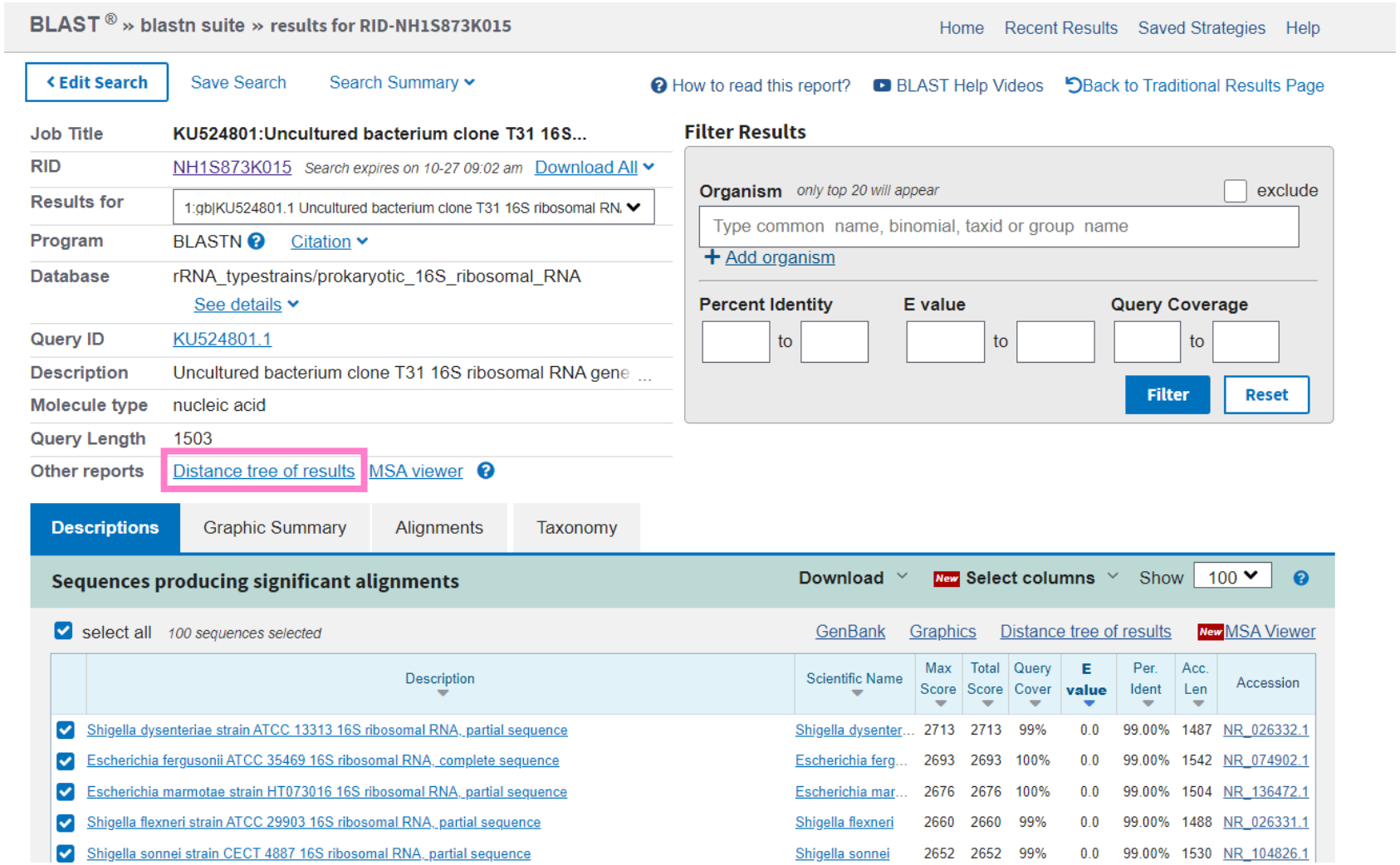
Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download — New Select columns — Show 100

select all *100 sequences selected* — GenBank — Graphics — Distance tree of results — New MSA Viewer

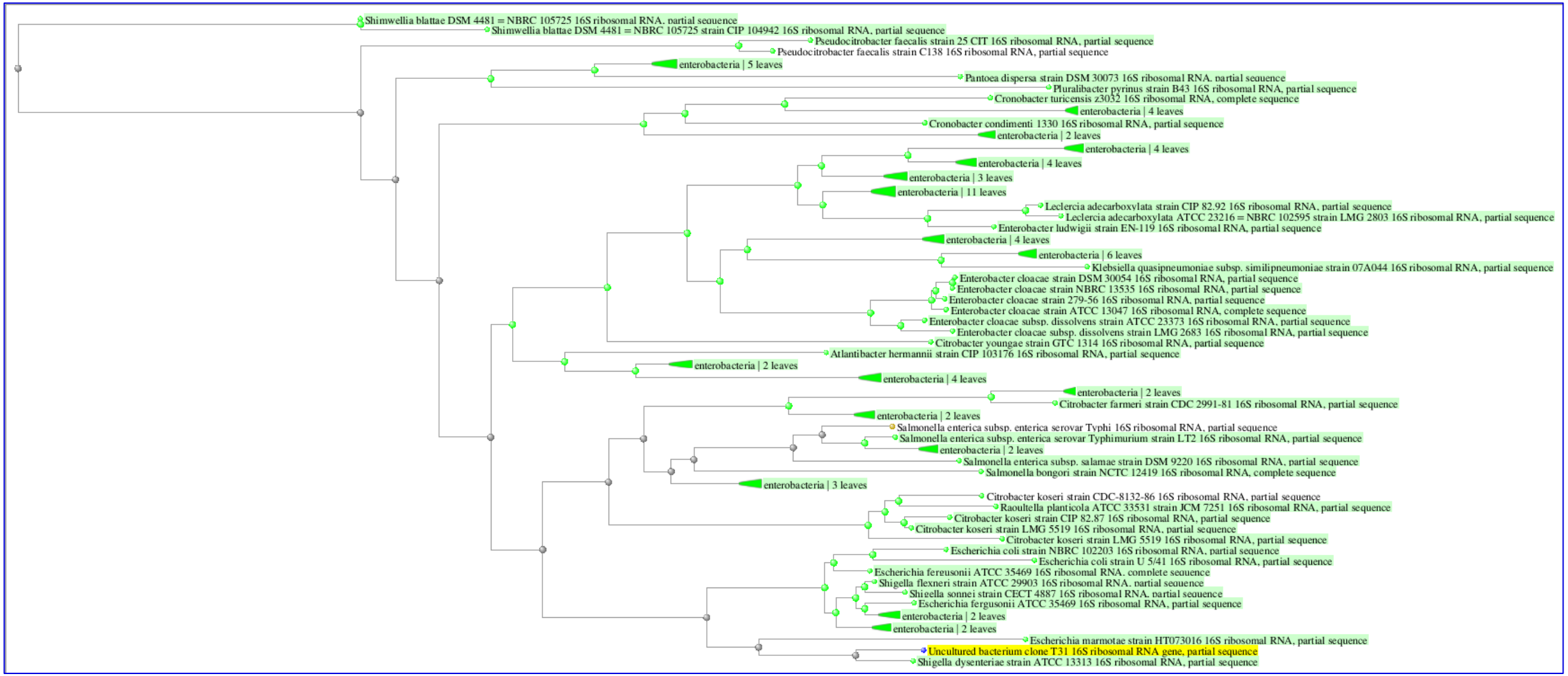
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Shigella dysenteriae strain ATCC 13313 16S ribosomal RNA, partial sequence	Shigella dysenter...	2713	2713	99%	0.0	99.00%	1487	NR_026332.1
Escherichia fergusonii ATCC 35469 16S ribosomal RNA, complete sequence	Escherichia fergu...	2693	2693	100%	0.0	99.00%	1542	NR_074902.1
Escherichia marmotae strain HT073016 16S ribosomal RNA, partial sequence	Escherichia mar...	2676	2676	100%	0.0	99.00%	1504	NR_136472.1
Shigella flexneri strain ATCC 29903 16S ribosomal RNA, partial sequence	Shigella flexneri	2660	2660	99%	0.0	99.00%	1488	NR_026331.1
Shigella sonnei strain CECT 4887 16S ribosomal RNA, partial sequence	Shigella sonnei	2652	2652	99%	0.0	99.00%	1530	NR_104826.1
Shigella boydii strain P288 16S ribosomal RNA, partial sequence	Shigella boydii	2651	2651	98%	0.0	99.00%	1515	NR_104901.1
Escherichia fergusonii ATCC 35469 16S ribosomal RNA, partial sequence	Escherichia fergu...	2623	2623	98%	0.0	99.00%	1473	NR_027549.1
Escherichia albertii strain Albert 19982 16S ribosomal RNA, partial sequence	Escherichia albertii	2623	2623	98%	0.0	99.00%	1494	NR_025569.1
Escherichia fergusonii strain NBRC 102419 16S ribosomal RNA, partial sequence	Escherichia fergu...	2615	2615	97%	0.0	99.00%	1467	NR_114079.1
Escherichia coli strain NBRC 102203 16S ribosomal RNA, partial sequence	Escherichia coli	2615	2615	97%	0.0	99.00%	1467	NR_114042.1
Citrobacter amalonaticus strain CECT 863 16S ribosomal RNA, partial sequence	Citrobacter amal...	2579	2579	99%	0.0	98.00%	1504	NR_104823.1
Salmonella enterica subsp. enterica serovar Typhi 16S ribosomal RNA, partial sequence		2577	2577	100%	0.0	98.00%	1534	NR_074799.1
Citrobacter koseri strain LMG 5519 16S ribosomal RNA, partial sequence	Citrobacter koseri	2560	2560	98%	0.0	98.00%	1494	NR_118105.1
Citrobacter farmeri strain CDC 2991-81 16S ribosomal RNA, partial sequence	Citrobacter farmeri	2558	2558	99%	0.0	98.00%	1511	NR_024861.1
Citrobacter amalonaticus strain LMG 7873 16S ribosomal RNA, partial sequence	Citrobacter amal...	2553	2553	98%	0.0	98.00%	1494	NR_118106.1
Escherichia coli strain U 5/41 16S ribosomal RNA, partial sequence	Escherichia coli	2553	2553	96%	0.0	98.00%	1450	NR_024570.1
Salmonella enterica subsp. arizonae strain ATCC 13314 16S ribosomal RNA, partial sequence	Salmonella enteri...	2551	2551	99%	0.0	98.00%	1491	NR_041696.1
Citrobacter koseri strain CDC-8132-86 16S ribosomal RNA, partial sequence	Citrobacter koseri	2534	2534	97%	0.0	98.00%	1469	NR_104890.1
Salmonella bongori strain DSM 13772 16S ribosomal RNA, partial sequence	Salmonella bongori	2523	2523	96%	0.0	98.00%	1443	NR_116124.1
Citrobacter koseri strain CIP 82.87 16S ribosomal RNA, partial sequence	Citrobacter koseri	2512	2512	97%	0.0	98.00%	1459	NR_118588.1

BLAST



MUNI | RECETOX

BLAST


Shimwellia blattae DSM 4481 = NBRC 105725 16S ribosomal RNA, partial sequence
Shimwellia blattae DSM 4481 = NBRC 105725 strain CIP 104942 16S ribosomal RNA, partial sequence
Pseudocitrobacter faecalis strain 25 CIT 16S ribosomal RNA, partial sequence
Pseudocitrobacter faecalis strain C138 16S ribosomal RNA, partial sequence
enterobacteria | 5 leaves
Pantoea dispersa strain DSM 30073 16S ribosomal RNA, partial sequence
Pluralibacter pyrinus strain B43 16S ribosomal RNA, partial sequence
Cronobacter turicensis z3032 16S ribosomal RNA, complete sequence
enterobacteria | 4 leaves
Cronobacter condimenti 1330 16S ribosomal RNA, partial sequence
enterobacteria | 2 leaves
enterobacteria | 4 leaves
enterobacteria | 4 leaves
enterobacteria | 3 leaves
enterobacteria | 11 leaves
Leclercia adecarboxylata strain CIP 82.92 16S ribosomal RNA, partial sequence
Leclercia adecarboxylata ATCC 23216 = NBRC 102595 strain LMG 2803 16S ribosomal RNA, partial sequence
Enterobacter ludwigii strain EN-119 16S ribosomal RNA, partial sequence
enterobacteria | 4 leaves
enterobacteria | 6 leaves
Klebsiella quasipneumoniae subsp. similipneumoniae strain 07A044 16S ribosomal RNA, partial sequence
Enterobacter cloacae strain DSM 30054 16S ribosomal RNA, partial sequence
Enterobacter cloacae strain NBRC 13535 16S ribosomal RNA, partial sequence
Enterobacter cloacae strain 279-56 16S ribosomal RNA, partial sequence
Enterobacter cloacae strain ATCC 13047 16S ribosomal RNA, complete sequence
Enterobacter cloacae subsp. dissolvens strain ATCC 23373 16S ribosomal RNA, partial sequence
Enterobacter cloacae subsp. dissolvens strain LMG 2683 16S ribosomal RNA, partial sequence
Citrobacter youngae strain GTC 1314 16S ribosomal RNA, partial sequence
Atlantibacter hermannii strain CIP 103176 16S ribosomal RNA, partial sequence
enterobacteria | 2 leaves
enterobacteria | 4 leaves
enterobacteria | 2 leaves
Citrobacter farmeri strain CDC 2991-81 16S ribosomal RNA, partial sequence
enterobacteria | 2 leaves
Salmonella enterica subsp. enterica serovar Typhi 16S ribosomal RNA, partial sequence
Salmonella enterica subsp. enterica serovar Typhimurium strain LT2 16S ribosomal RNA, partial sequence
enterobacteria | 2 leaves
Salmonella enterica subsp. salamae strain DSM 9220 16S ribosomal RNA, partial sequence
Salmonella bongori strain NCTC 12419 16S ribosomal RNA, complete sequence
enterobacteria | 3 leaves
Citrobacter koseri strain CDC-8132-86 16S ribosomal RNA, partial sequence
Raoultella planticola ATCC 33531 strain JCM 7251 16S ribosomal RNA, partial sequence
Citrobacter koseri strain CIP 82.87 16S ribosomal RNA, partial sequence
Citrobacter koseri strain LMG 5519 16S ribosomal RNA, partial sequence
Citrobacter koseri strain LMG 5519 16S ribosomal RNA, partial sequence
Escherichia coli strain NBRC 102203 16S ribosomal RNA, partial sequence
Escherichia coli strain U 5/41 16S ribosomal RNA, partial sequence
Escherichia fergusonii ATCC 35469 16S ribosomal RNA, complete sequence
Shigella flexneri strain ATCC 29903 16S ribosomal RNA, partial sequence
Shigella sonnei strain CECT 4887 16S ribosomal RNA, partial sequence
Escherichia fergusonii ATCC 35469 16S ribosomal RNA, partial sequence
enterobacteria | 2 leaves
enterobacteria | 2 leaves
Escherichia marmotae strain HT073016 16S ribosomal RNA, partial sequence
Uncultured bacterium clone T31 16S ribosomal RNA gene, partial sequence
Shigella dysenteriae strain ATCC 13313 16S ribosomal RNA, partial sequence

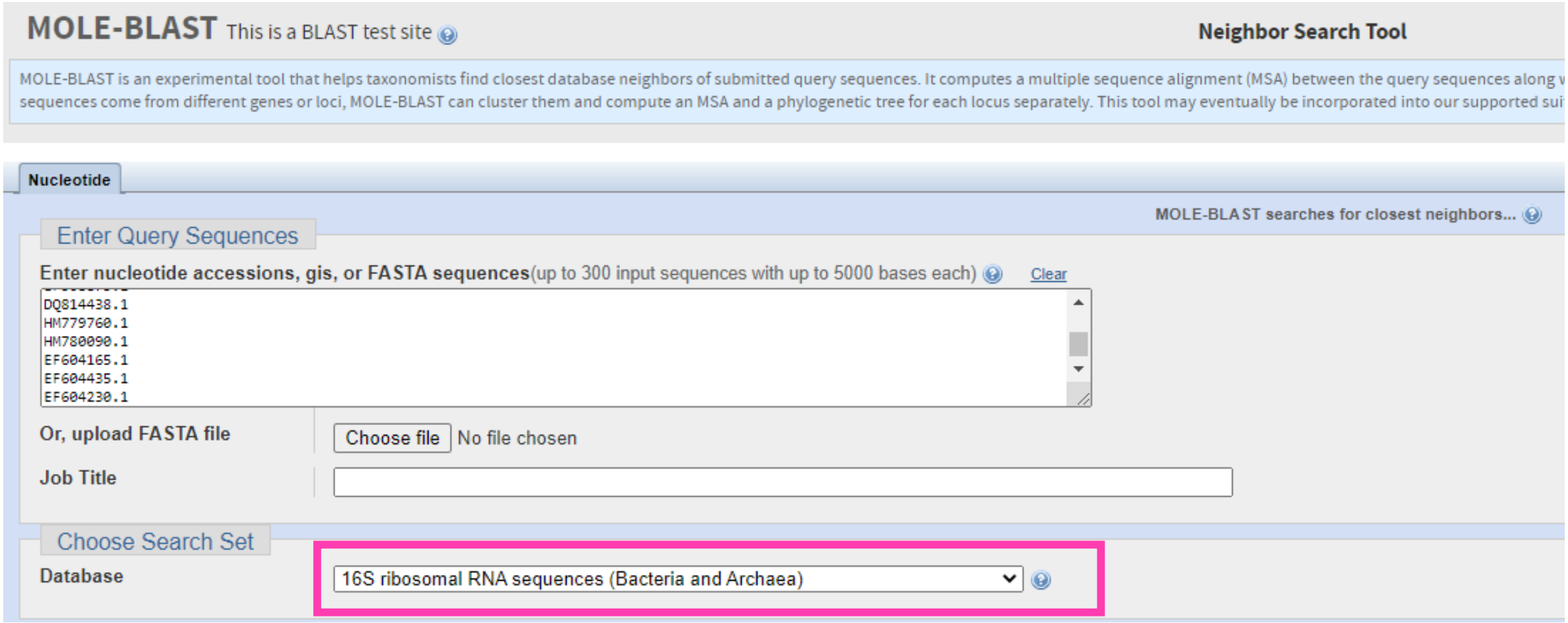
MOLE-BLAST

- Extension tool: cluster things together
- https://blast.ncbi.nlm.nih.gov/moleblast/moleblast.cgi
- Does BLAST search, but also full multiple sequence alignment
- Will give a tree that cluster the 20 sequences
  - With 16s sequences
  - also with each other

- ***For time constraining reason: Copy/paste the first 3 sequences (see word document)***

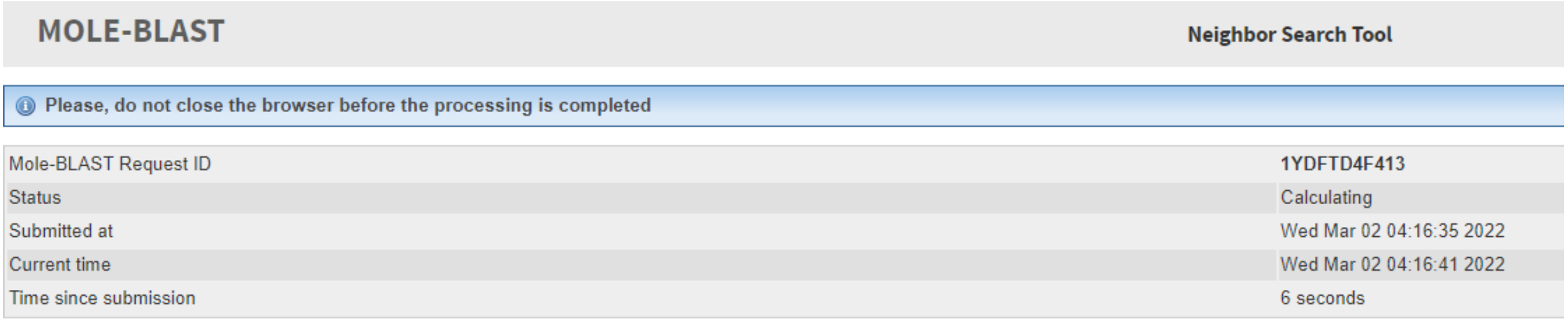
KU524801.1
AB759680.1
GQ158974.1
DQ904997.1
EU488411.1
KX431275.1
EU556993.1
FM873915.1
FM874039.1
HM124388.1
FJ624883.1
FJ625334.1
EU236261.1
EF508875.1
DQ814438.1
HM779760.1
HM780090.1
EF604165.1
EF604435.1
EF604230.1

MOLE-BLAST



MUNI | RECETOX

MOLE-BLAST



- Then we wait!
- *(Let it run in the background while we continue, it will take a few minutes…)*

- https://www.youtube.com/watch?v=JKD5laNtwSc&ab_channel=TheNationalLibraryofMedicine (source for the exercice)

Registration databases

- Look for the status of a compound (authorized for use *vs* banned)
- ECHA: European Chemical Agency
  - Shows substances registered for use in EU
  - Toxicity criteria and labelling
  - Somehow of a mess to find registration status
  - **Chemicals under REACH Legislation**

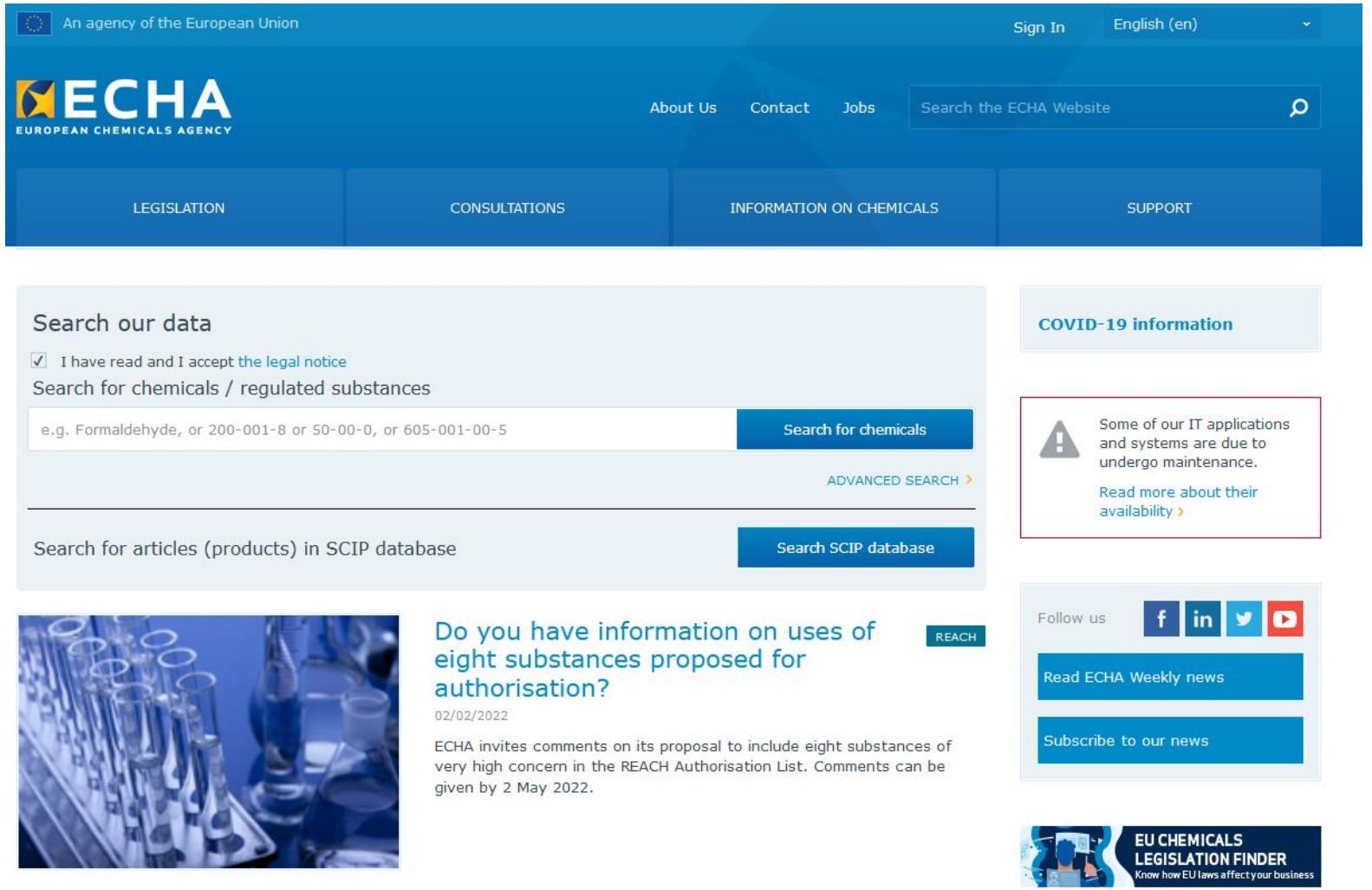
    REACH = Registration, **E**valuation, **A**uthorization, **Ch**emicals

    Applies to all chemicals sold on the European market (more than 1 tonne per year)

    Insure safety of chemicals used for humans and environment

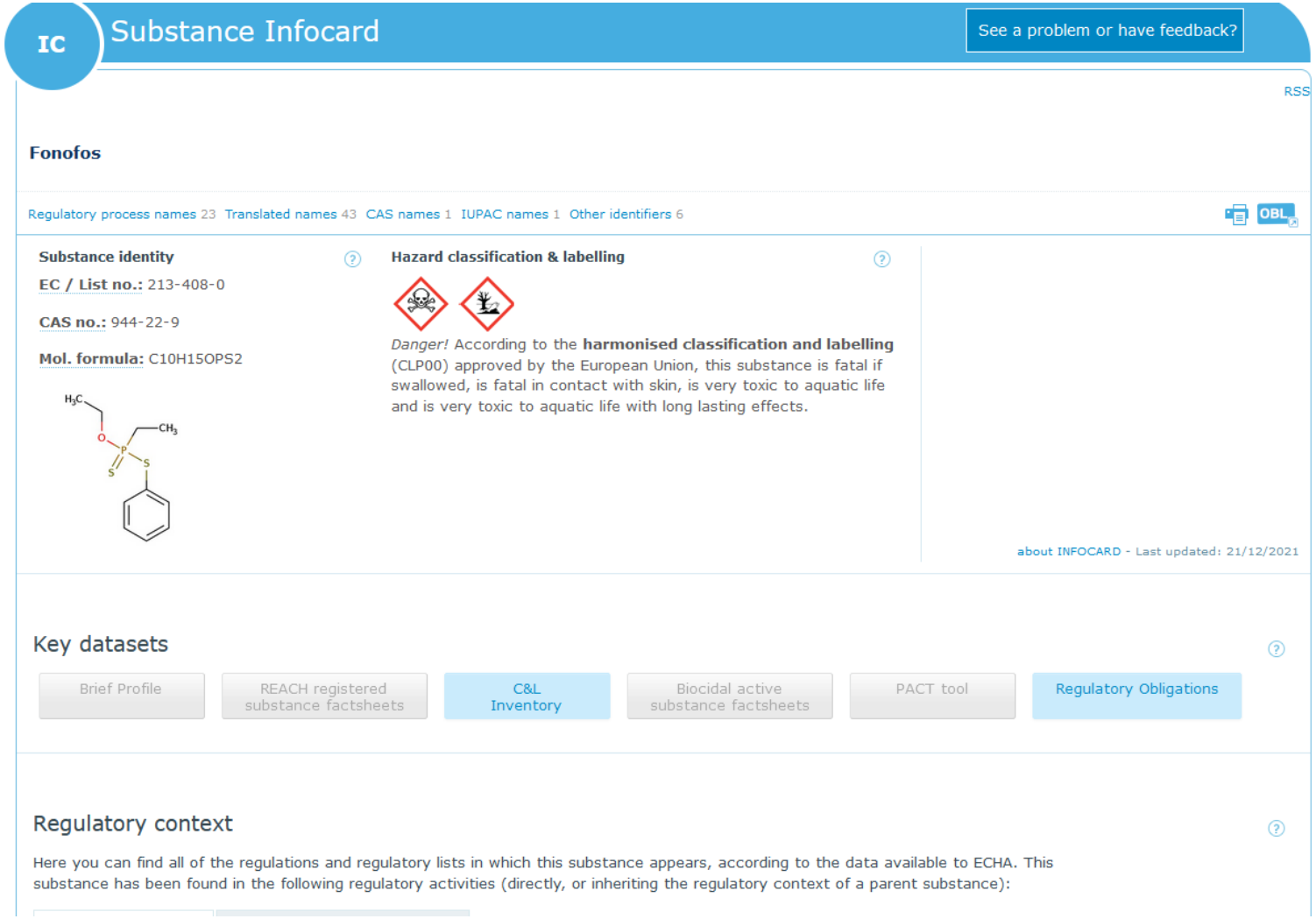
- https://echa.europa.eu/home

Registration databases



MUNI | RECETOX

Registration databases



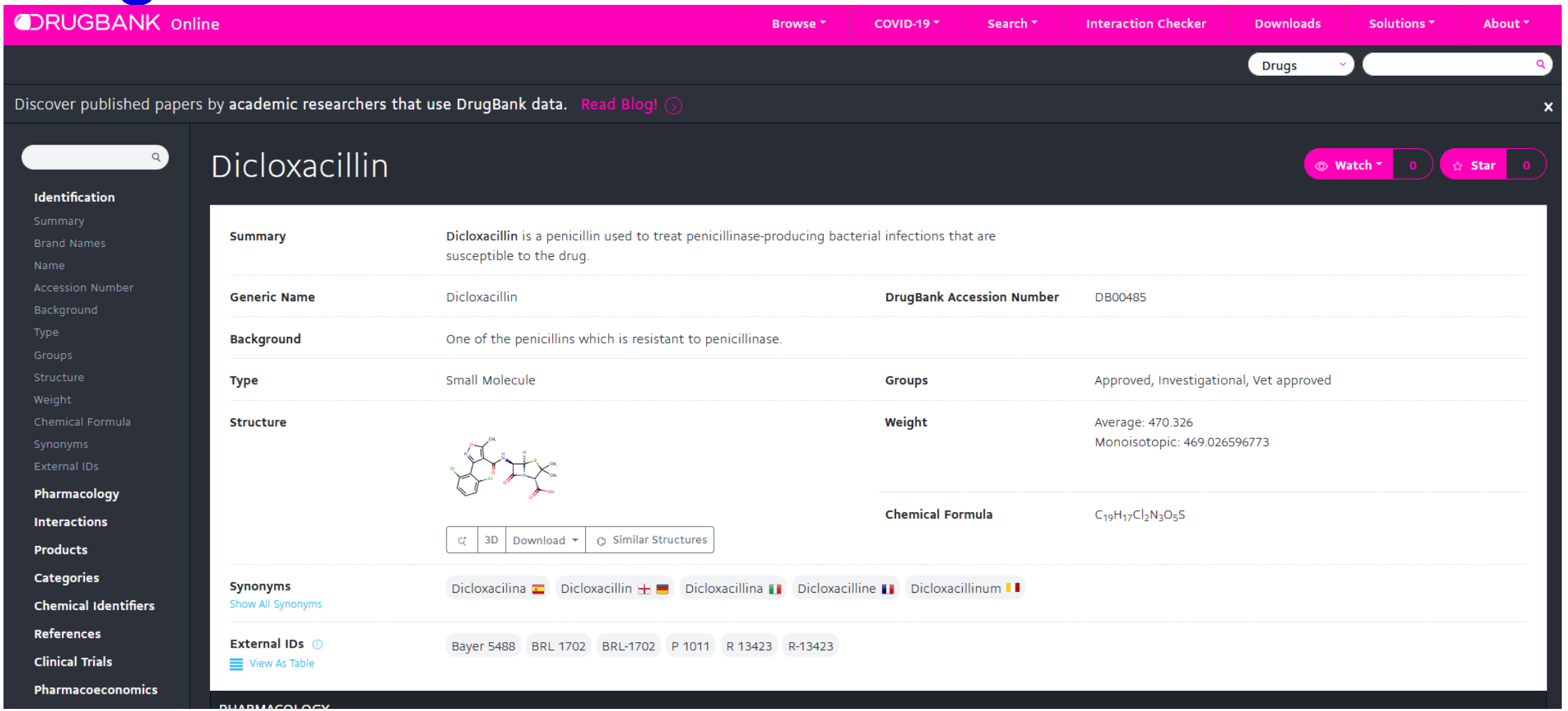
MUNI | RECETOX

Registration databases

- **Task 5: Find EU authorization status of the following compound, using following links**
- https://go.drugbank.com/ (Drug database)
- https://ec.europa.eu/food/plants/pesticides/eu-pesticides-database_en (EU pesticide database)
- https://www.echemportal.org/echemportal/ (EU chemical portal)
- https://risctox.istas.net/en/ (Toxic substance database)
- https://echa.europa.eu/home (ECHA website)

For these compounds:
Dicloxacillin – Epoxiconazole – Decachlorobiphenyl

Registration databases

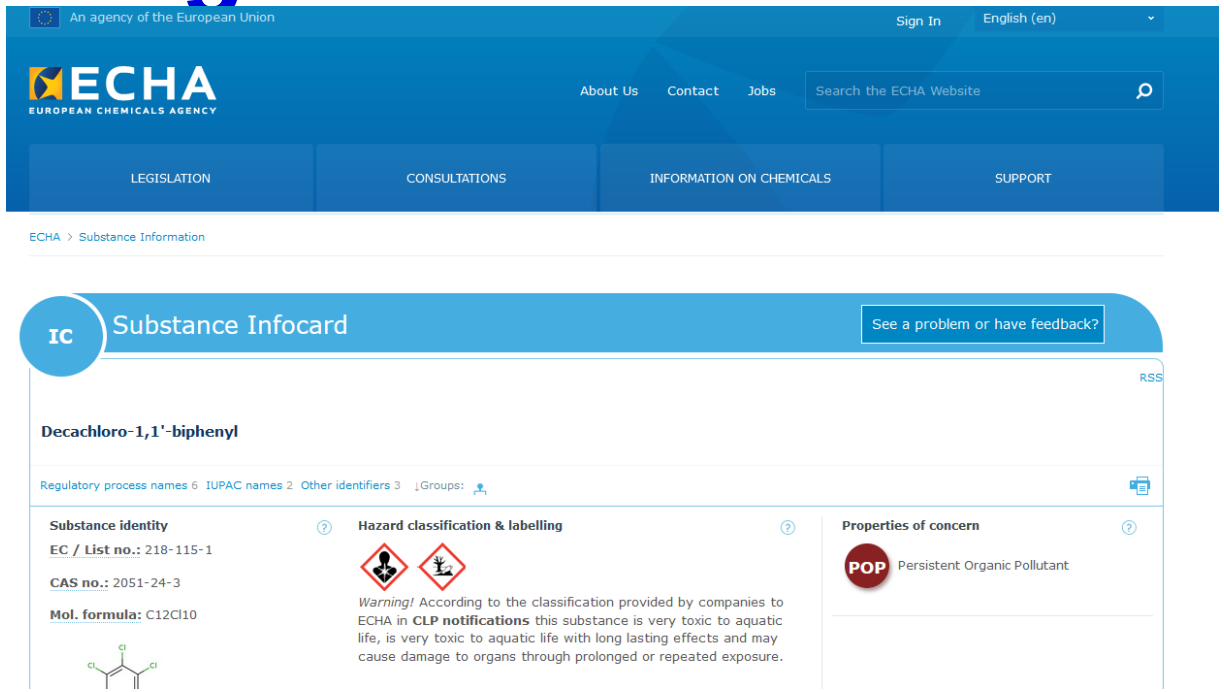


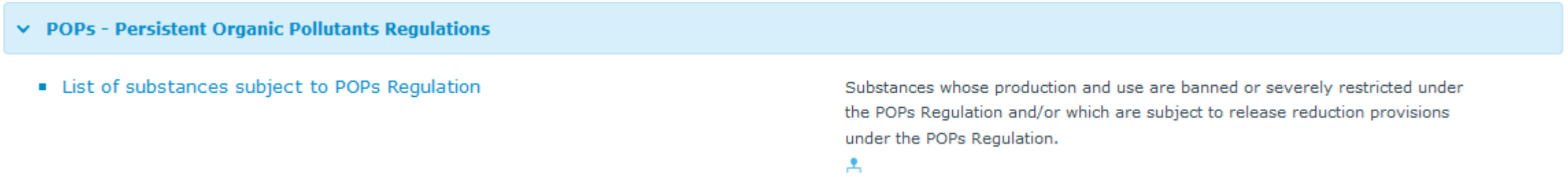
MUNI | RECETOX

Registration databases

MUNI | RECETOX

Registration databases





MUNI | RECETOX

Registration databases

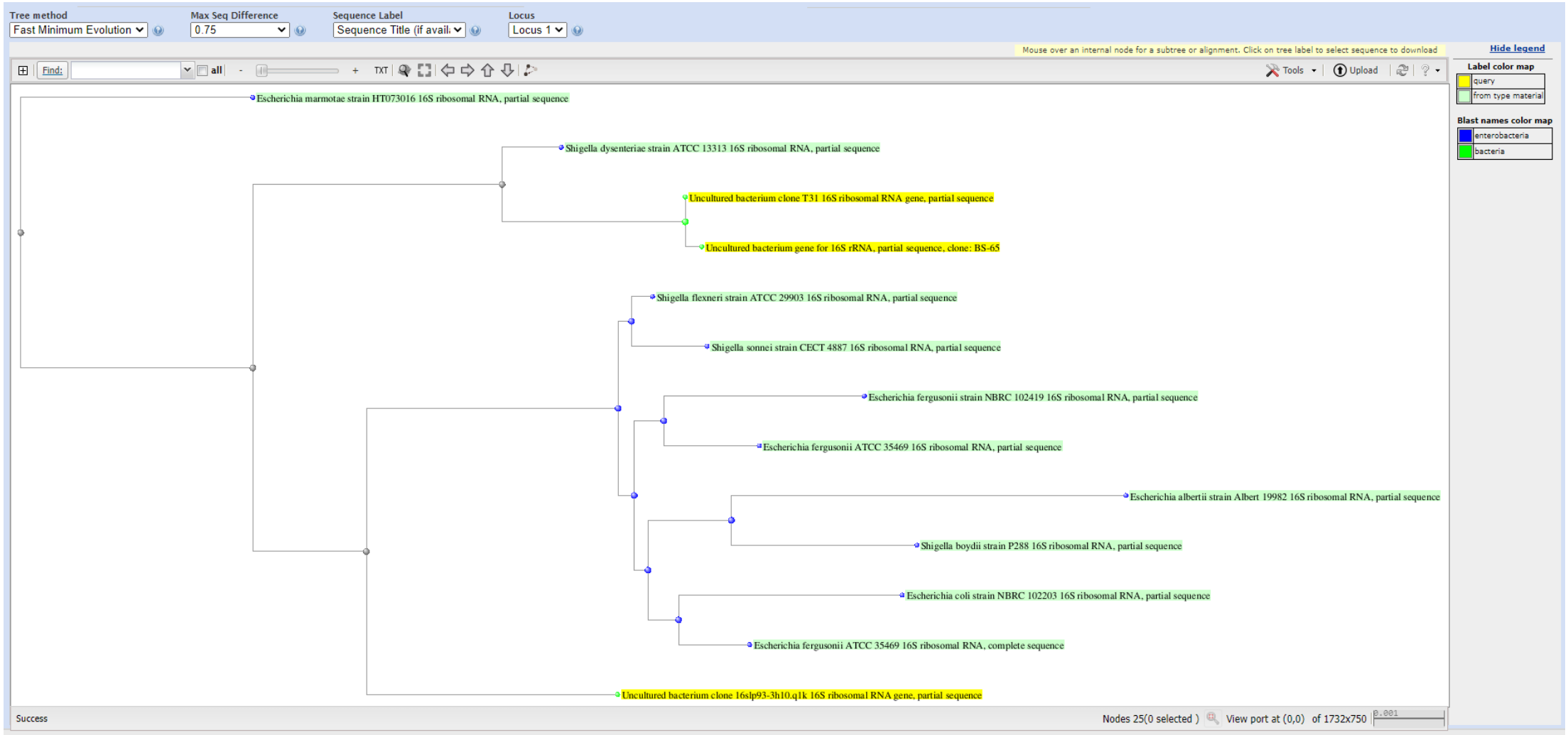
- Very clear DB for some type of compounds:
  - Pesticides
  - Drugs
- But not for other types of compounds

- Alternatives: googling it, news articles can be useful or EU legislation text
- But information can be very complicated to find
- Public information / private company → *Full disclosure legislation doesn't exist*

Registration databases

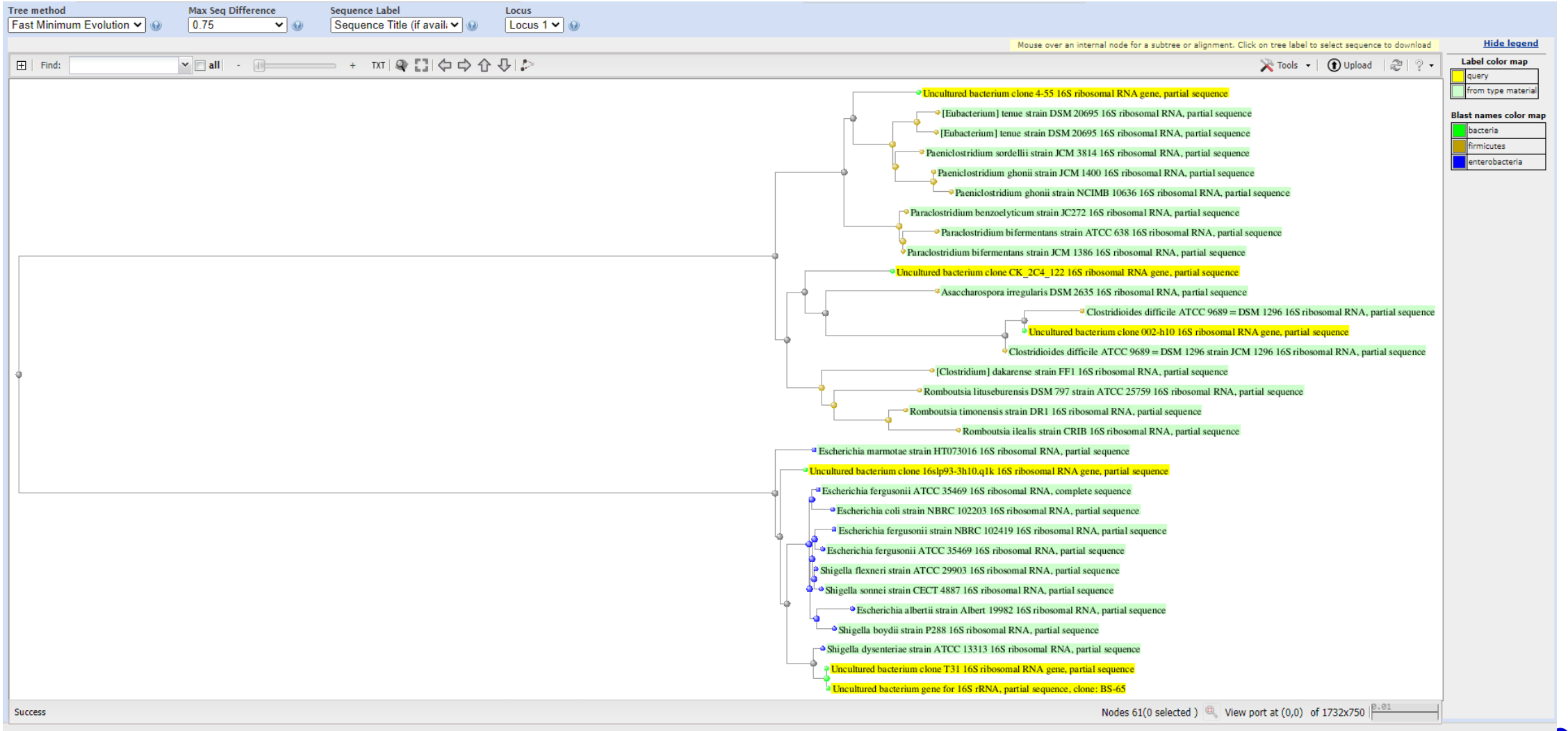
- Other sources of chemical information (incl. Databases)
  - Manufacturers and dealers: Merck, TCI, Acros,…
  - Large projects: NIH library (US -National Institute of Health, etc.)

MOLE-BLAST



MUNI | RECETOX

MOLE-BLAST

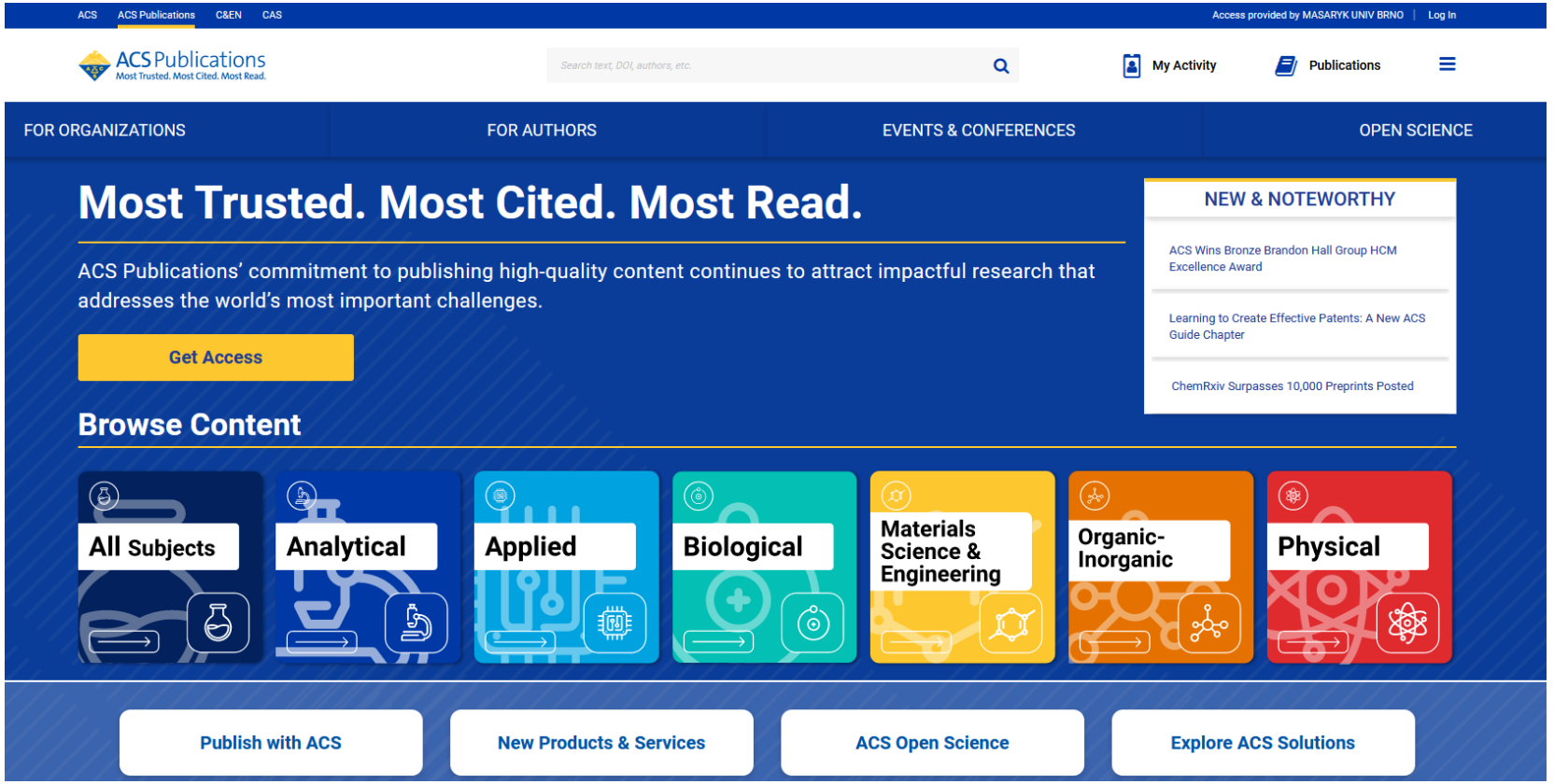

Tree method
Fast Minimum Evolution
Max Seq Difference
0.75
Sequence Label
Sequence Title (if avail
Locus
Locus 1
Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download
Hide legend
Label color map
query
from type material
Blast names color map
bacteria
firmicutes
enterobacteria
Find:
all
TXT
Tools
Upload
Uncultured bacterium clone 4-55 16S ribosomal RNA gene, partial sequence
[Eubacterium] tenue strain DSM 20695 16S ribosomal RNA, partial sequence
[Eubacterium] tenue strain DSM 20695 16S ribosomal RNA, partial sequence
Paeniclostridium sordellii strain JCM 3814 16S ribosomal RNA, partial sequence
Paeniclostridium ghonii strain JCM 1400 16S ribosomal RNA, partial sequence
Paeniclostridium ghonii strain NCIMB 10636 16S ribosomal RNA, partial sequence
Paraclostridium benzoelyticum strain JC272 16S ribosomal RNA, partial sequence
Paraclostridium bifermentans strain ATCC 638 16S ribosomal RNA, partial sequence
Paraclostridium bifermentans strain JCM 1386 16S ribosomal RNA, partial sequence
Uncultured bacterium clone CK_2C4_122 16S ribosomal RNA gene, partial sequence
Asaccharospora irregularis DSM 2635 16S ribosomal RNA, partial sequence
Clostridioides difficile ATCC 9689 = DSM 1296 16S ribosomal RNA, partial sequence
Uncultured bacterium clone 002-h10 16S ribosomal RNA gene, partial sequence
Clostridioides difficile ATCC 9689 = DSM 1296 strain JCM 1296 16S ribosomal RNA, partial sequence
[Clostridium] dakarense strain FF1 16S ribosomal RNA, partial sequence
Romboutsia lituseburensis DSM 797 strain ATCC 25759 16S ribosomal RNA, partial sequence
Romboutsia timonensis strain DR1 16S ribosomal RNA, partial sequence
Romboutsia ilealis strain CRIB 16S ribosomal RNA, partial sequence
Escherichia marmotae strain HT073016 16S ribosomal RNA, partial sequence
Uncultured bacterium clone 16slp93-3h10.q1k 16S ribosomal RNA gene, partial sequence
Escherichia fergusonii ATCC 35469 16S ribosomal RNA, complete sequence
Escherichia coli strain NBRC 102203 16S ribosomal RNA, partial sequence
Escherichia fergusonii strain NBRC 102419 16S ribosomal RNA, partial sequence
Escherichia fergusonii ATCC 35469 16S ribosomal RNA, partial sequence
Shigella flexneri strain ATCC 29903 16S ribosomal RNA, partial sequence
Shigella sonnei strain CECT 4887 16S ribosomal RNA, partial sequence
Escherichia albertii strain Albert 19982 16S ribosomal RNA, partial sequence
Shigella boydii strain P288 16S ribosomal RNA, partial sequence
Shigella dysenteriae strain ATCC 13313 16S ribosomal RNA, partial sequence
Uncultured bacterium clone T31 16S ribosomal RNA gene, partial sequence
Uncultured bacterium gene for 16S rRNA, partial sequence, clone: BS-65
Success
Nodes 61(0 selected)
View port at (0,0) of 1732x750

Magasine database

- Publishers (industry-specific or broad-based)
- Larger consortia or aggregators
- Other organizations: e.g., NIH

Magasine database

– https://pubs.acs.org/ (American Chemical society)



Magasine database

ACS Publications
Most Trusted. Most Cited. Most Read.

dicofol

My Activity | Publications

NARROW RESULTS

CONTENT TYPE

Book Chapter	28
Reference/Standard	2
C&EN Article	32
Journal Article	204

ARTICLE SUBJECT

Article	134
News	26
Review	8
Concentrates	4
Features	4

MORE (15)

PUBLICATION DATE

RESULTS: 1 - 20 **of** 301

Follow results:

REFINE SEARCH | PER PAGE: 20 50 100 | SORT: RELEVANCE

1 2 3 4 5 6 7 >

Article

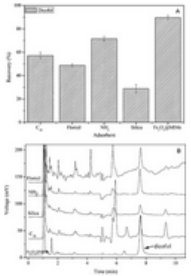
Synthesis of Multi-core–shell Magnetic Molecularly Imprinted Microspheres for Rapid Recognition of Dicofol in Tea

Hongyuan Yan*, Xiaoling Cheng, and Ning Sun

Journal of Agricultural and Food Chemistry 2013, 61, 11, 2896-2901 **(Article)** Subscribed
Publication Date (Web): February 23, 2013
DOI: 10.1021/jf400847b

Abstract | Full text | PDF | LinkSource

ABSTRACT

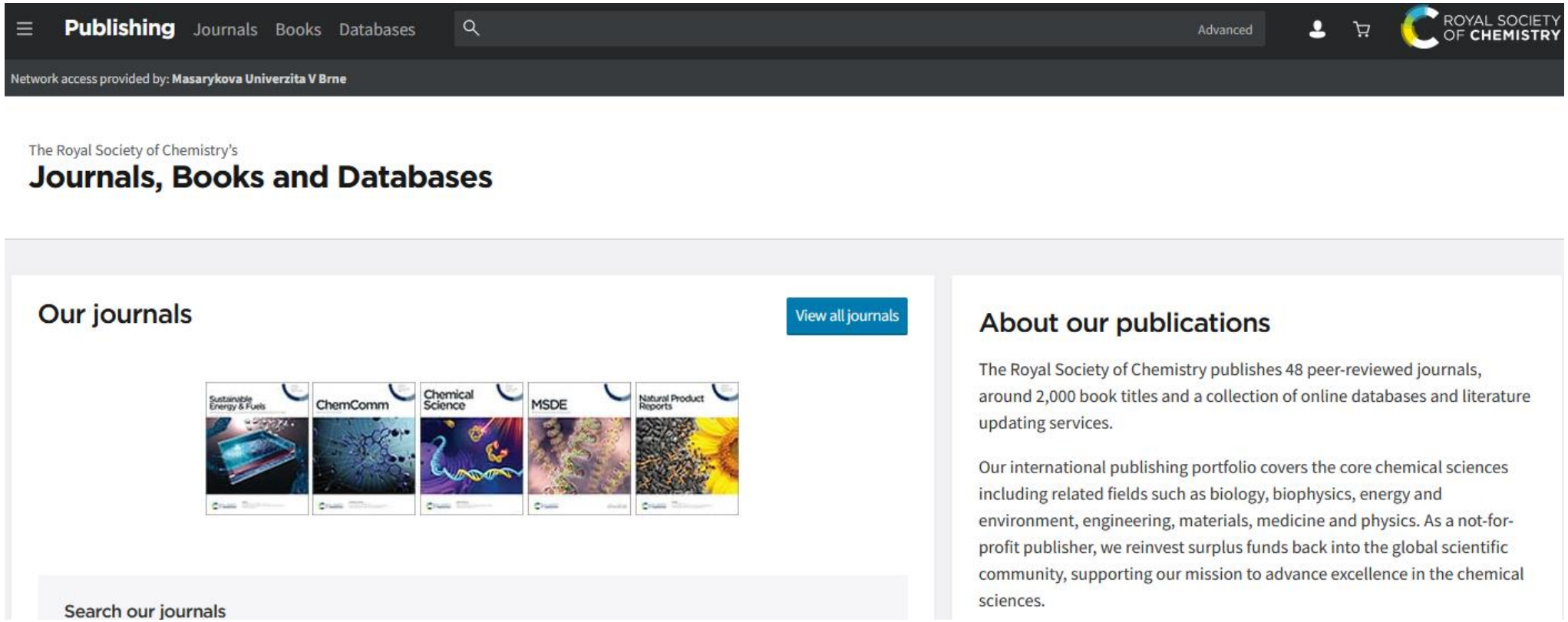


JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

MUNI | RECETOX

Magasine database

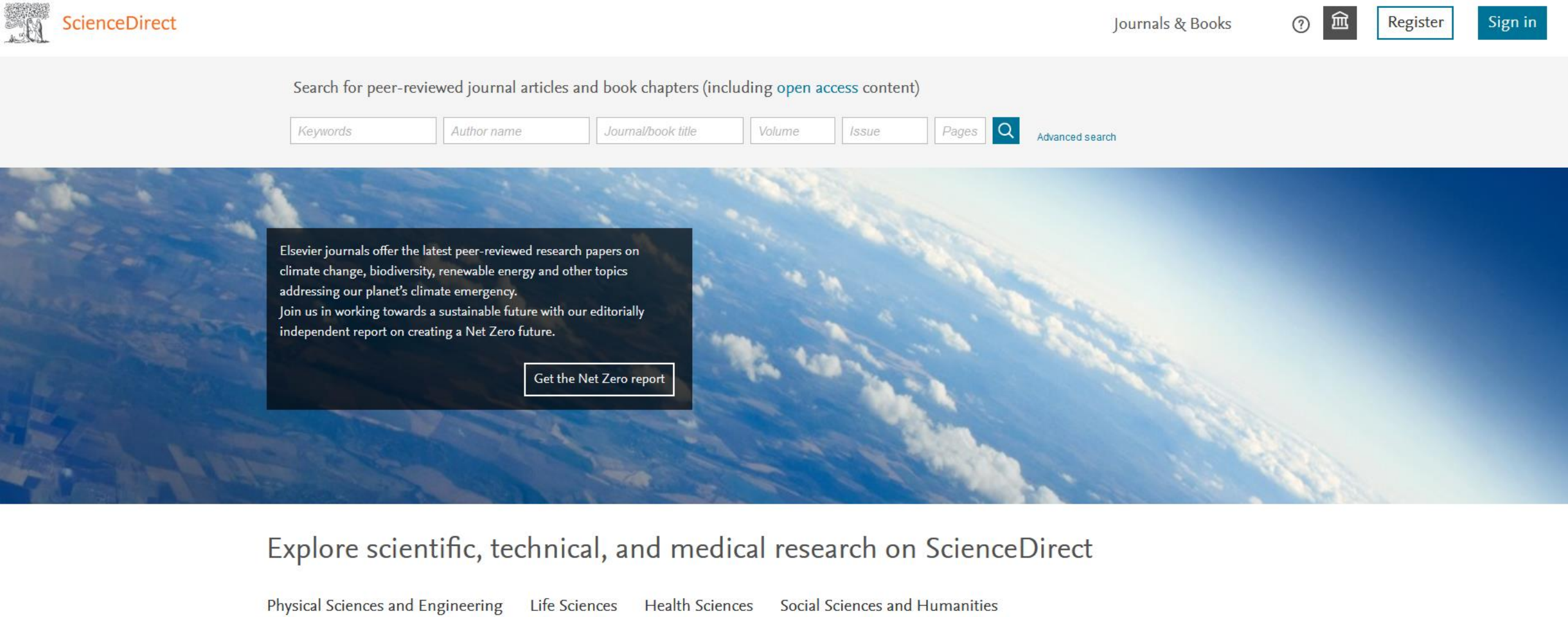
– https://pubs.rsc.org/ (Royal Society of Chemistry)



Other magazine publishers

- Elsevier
- Wiley & Sons.
- Nature publishing group
- AAAS
- Hindawi
- Universities, AV, Research organizations
- Many other…

https://www.sciencedirect.com/

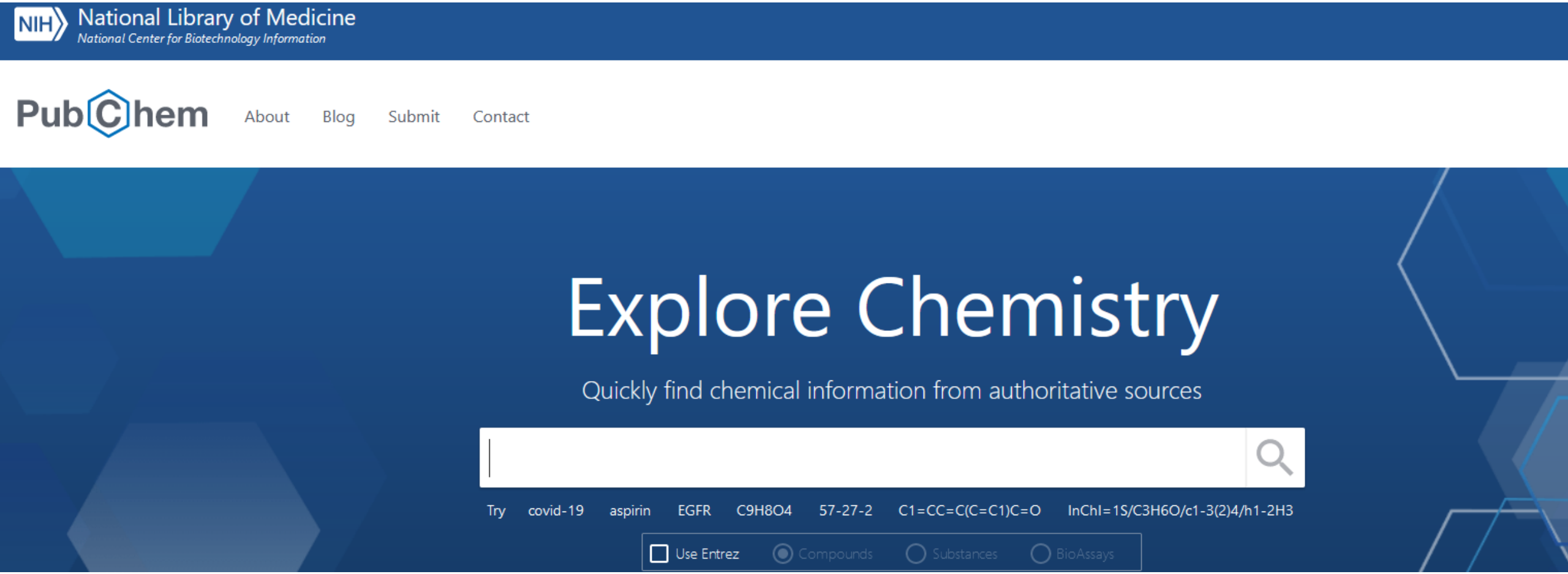


MUNI | RECETOX

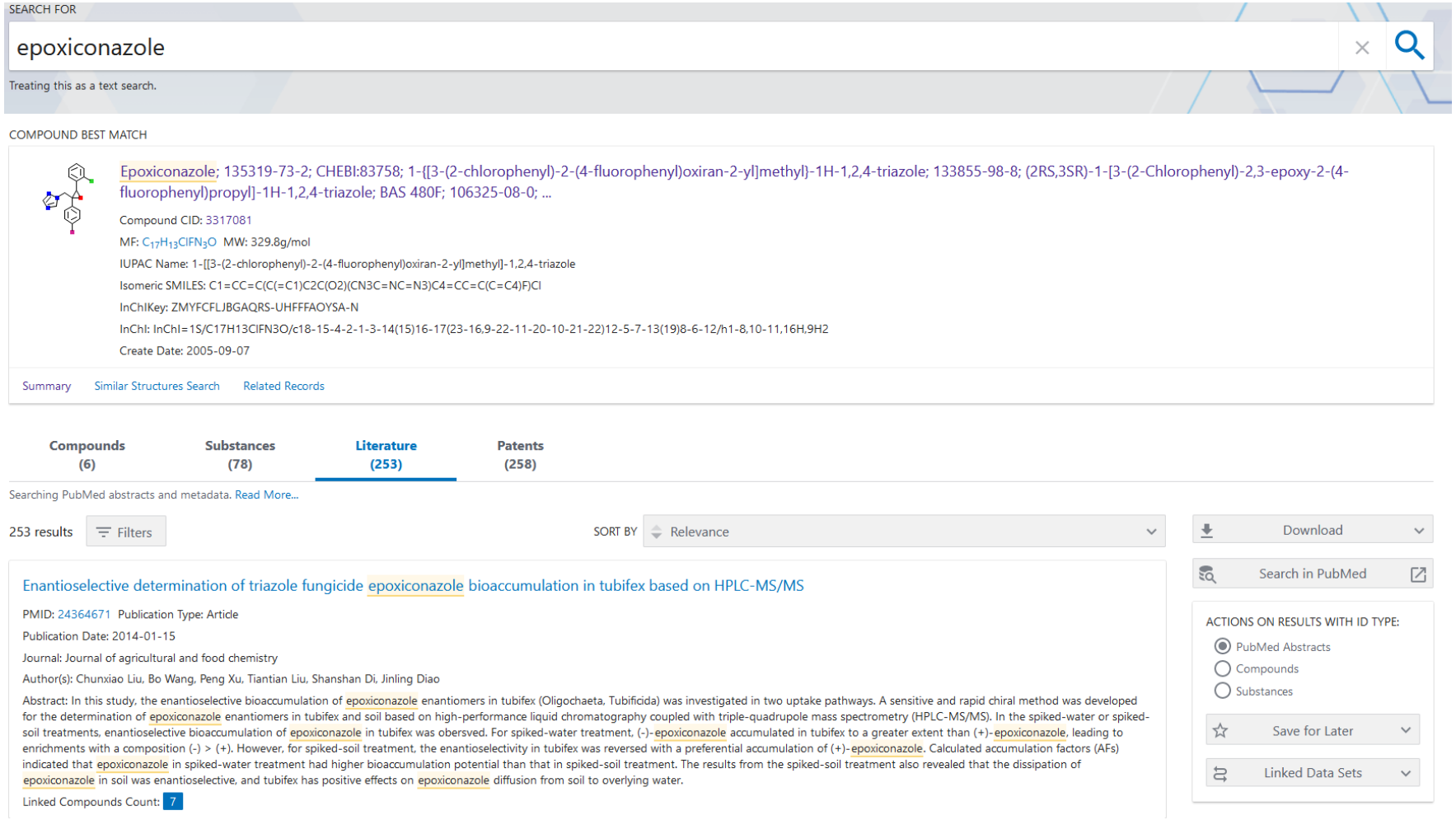
CAS: Chemical Abstracts Services

- Published by ACS as an abstract review journal
  - It included more than 8,000 magazines and conferences, as well as technical reports, dissertations, etc.
  - 1907-2010• Today, its role has been taken over by online databases…
- Abstract databases
  - They no longer make much sense
  - Replaced by general ones
- More wider databases•
  - PubChem - one of the CAS substitutes
  - And one of the databases operated by NIH NLM: ncbi.nlm.nih.gov

PUBCHEM


NIH
National Library of Medicine
National Center for Biotechnology Information
PubChem
About
Blog
Submit
Contact
Explore Chemistry
Quickly find chemical information from authoritative sources
Try
covid-19
aspirin
EGFR
C9H8O4
57-27-2
C1=CC=C(C=C1)C=O
InChI=1S/C3H6O/c1-3(2)4/h1-2H3
Use Entrez
Compounds
Substances
BioAssays

PUBCHEM



MUNI | RECETOX

PUBCHEM

PubChem About Posts Submit Contact Search PubChem

COMPOUND SUMMARY

Epoxiconazole

PubChem CID	3317081
Structure	 2D 3D Find Similar Structures
Chemical Safety	 Health Hazard Environmental Hazard Laboratory Chemical Safety Summary (LCSS) Datasheet
Molecular Formula	$C_{17}H_{13}ClFN_3O$
Synonyms	Epoxiconazole 135319-73-2 1-[[3-(2-chlorophenyl)-2-(4-fluorophenyl)oxiran-2-yl]methyl]-1,2,4-triazole CHEBI:83758 1-{[3-(2-chlorophenyl)-2-(4-fluorophenyl)oxiran-2-yl]methyl}-1H-1,2,4-triazole More...
Molecular Weight	329.8
Dates	Modify 2023-02-18 Create 2005-09-07

1-{[3-(2-chlorophenyl)-2-(4-fluorophenyl)oxiran-2-yl]methyl}-1H-1,2,4-triazole is an epoxide that is oxirane substituted by a 2-chlorophenyl, 4-fluorophenyl and a 1H-1,2,4-triazol-1-ylmethyl groups. It is an epoxide, a member of monochlorobenzenes, a member of monofluorobenzenes and a member of triazoles.

ChEBI

Cite Download

CONTENTS

MUNI | RECETOX

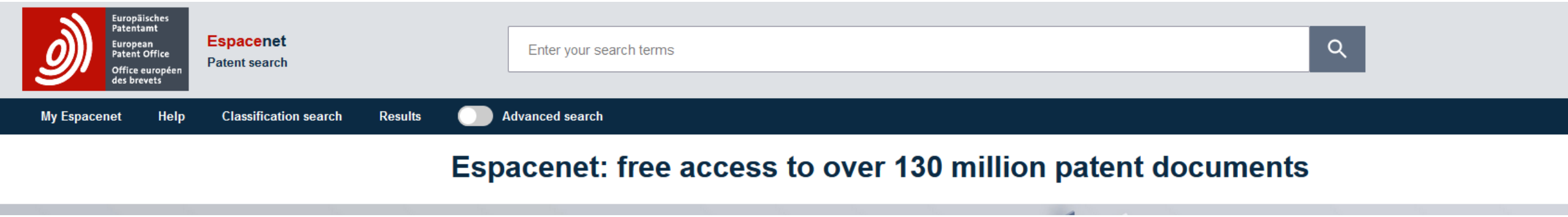
Patent Database

- Only partial overlap with articles and monographs
- Different style of text and content

  - European: ESPACENET
  - US: USPTO
  - Japanese
  - Czech
  - and others, eg:
  - patents.google.com
  - Patentscope (WIPO)
  - freepatentsonline.com

Patent Database

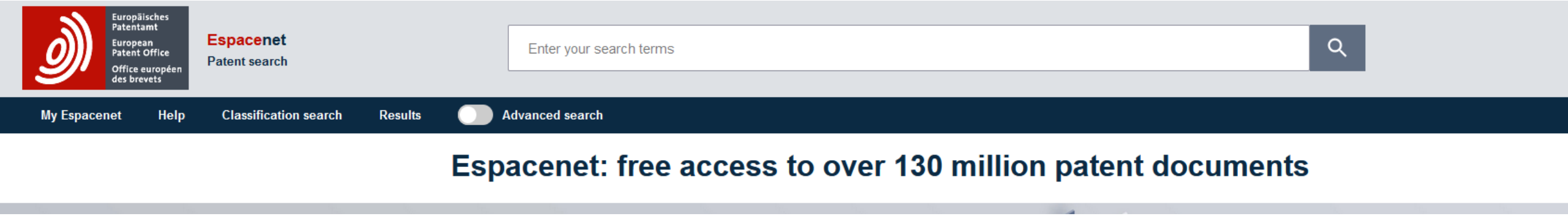
– https://worldwide.espacenet.com/



MUNI | RECETOX

Patent Database

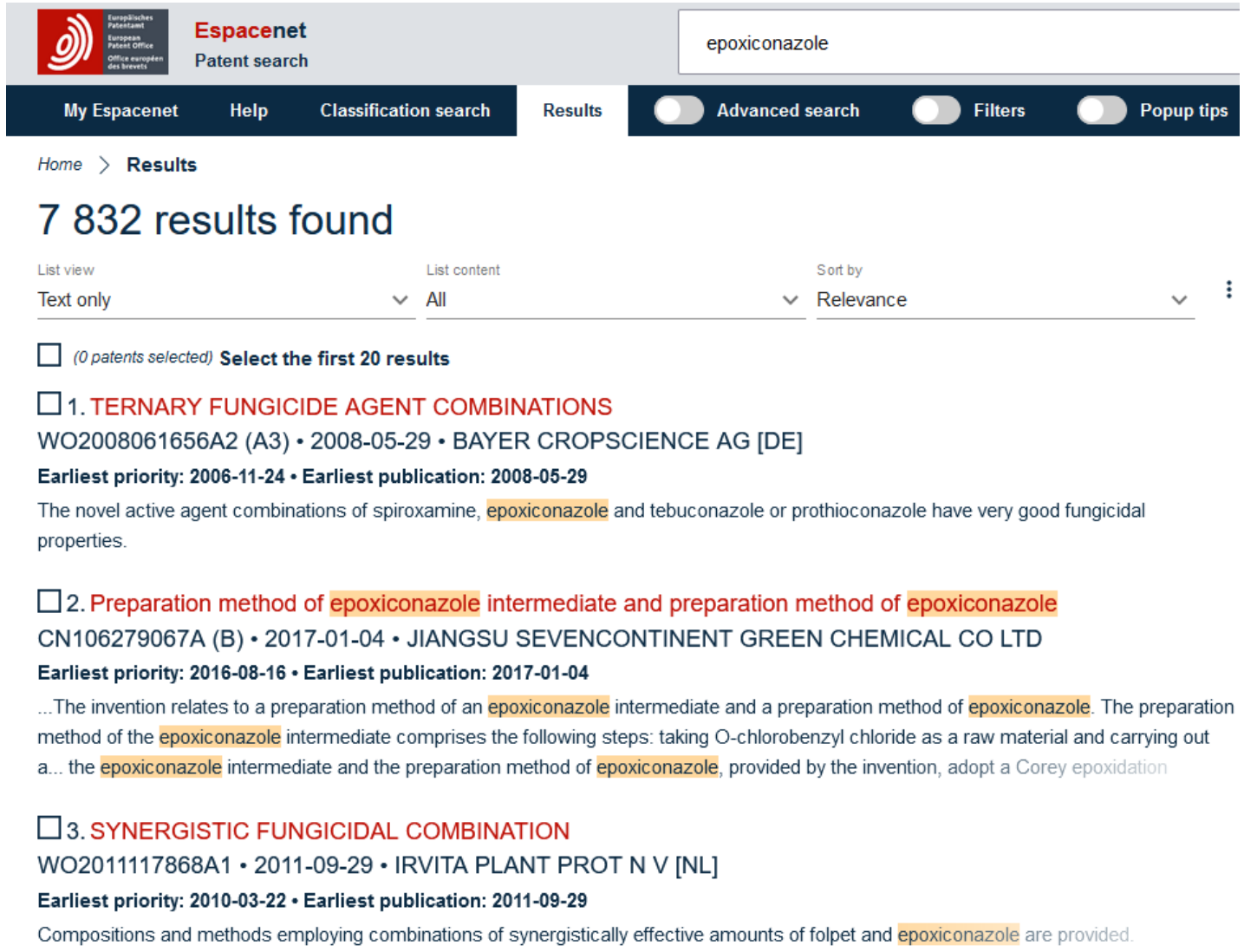
– https://worldwide.espacenet.com/



– Let's look for something
  – Suggestions?

MUNI | RECETOX

Patent Database



MUNI | RECETOX

Conclusions

- Lots of information is available online
- Just need to know where to look for it …
- … and how to formulate your query

- Accessibility may sometimes be an issue…

Exam: Assignment 2

- **For April 5th – as a word document or pdf – submitted on IS**
- Instructions: From the following list of journals, select the one that has the same serial number as the last digit of your UČO number:

Last number of your UČO	Journal
1	**Environmental Science and Technology**
2	**Chemosphere**
3	**ACS Earth and Space Chemistry**
4	**Science**
5	**TrAC Trends in Analytical Chemistry**
6	**Environmental Science & Policy**
7	**Limnologica**
8	**Energy & Environmental Science**
9	**Atmospheric Science Letters**
0	**Risk Analysis**

Exam: Assignment 2

- **For April 5th – as a word document or pdf – submitted on IS**
- Instructions: From the following list of journals, select the one that has the same serial number as the last digit of your UČO number:

- Then find out for the journal you have been given:
  - a) the name of the publishing house
  - b) the name of the chairman of the editorial board
  - c) is there a member of the editorial board who has an affiliation with a Czech institution?
  - d) what types of articles does this journal accept for publication (e.g. reports, full papers, letters, reviews, opinions, etc.

- Write a small report (10 lines maximum)

Questions?

– Contact me anytime via email: ludovic.mayer@recetox.muni.cz