Základy praktické Bioinformatiky

PETRA MATOUŠKOVÁ

2025/2026

Organizace:

Každý student: 1 gen/protein

Zápočet: **domácí úkoly**: vyhledávání informací o konkrétním genu/proteinu

(kontrolováno průběžně nebo odevzdat v celku na konci semestru)

účast: 8/10 lekcí (+ další pouze s omluvenkou)

Zkouška: „**písemná**“ u počítače – soubor úkolů (2x5)

termín pro ZK mezi 13.4.-30.4.

Týden	Datum	Online 1 – DU1	Čas
1	25.2.2026	2	St 17:50 - 20:05
2	2.3.2026	3	Po 17:40 - 19:15
2	4.3.2026	4	St 17:50 - 20:05
3	*11.3.2026*	*X*	*St 17:50 - 20:05*
4	*16.3.2026*	*X*	*Po 17:40 - 19:15*
4	18.3.2026	5	St 17:50 - 20:05
5	25.3.2026	6	St 17:50 - 20:05
6	30.3.2026	7	Po 17:40 - 19:15
6	1.4.2026	8	St 17:50 - 20:05
7	8.4.2026	9	St 17:50 - 20:05
8	13.4.2026	10	Po 17:40 - 19:15
10	27.4.2026	X	Po 17:40 - 19:15

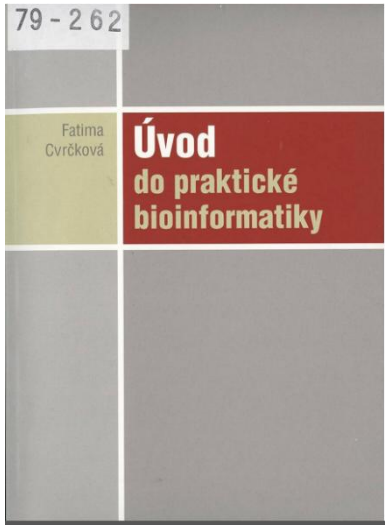
ZK?

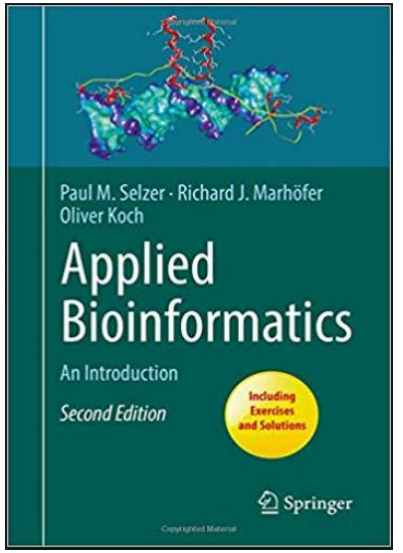
Bioinformatika

= vědní disciplína zabývající se rozsáhlými soubory biologických dat.

Shromažďování, zpracování a analýza dat o sekvencích a struktuře biologických makromolekul

in silico práce či simulace na počítači (*in vitro, in vivo*)





Není třeba nakupovat a instalovat programy, vše běží v rámci internetového prohlížeče (+Java).

Bioinformatika může během pár hodin, ušetřit měsíce práce, dříve strávené v laboratoři.

Vyhledávání a získávání sekvencí z databází (nukleotidových (nt)-**DNA** / aminokyselinových (AA)-**protein**)

Výpočet vlastností: nt/AA složení, molekulová hmotnost, isoelektrický bod

Výpočet hydrofobicity/hydrofilicity proteinů, předpověď antigenních oblastí

Predikce transmembránových úseků

Identifikace restrikčních míst, tvorba restrikčních map

Návrh primerů pro polymerázovou řetězovou reakci (PCR)

„Otevírání“ čtecích rámců - open reading frames (ORFs)

Predikce DNA/proteinových sekundárních struktur

Vyhledání 3-D struktury proteinu a organizace domén

Vyhledání podobných sekvencí, klasifikace do rodin

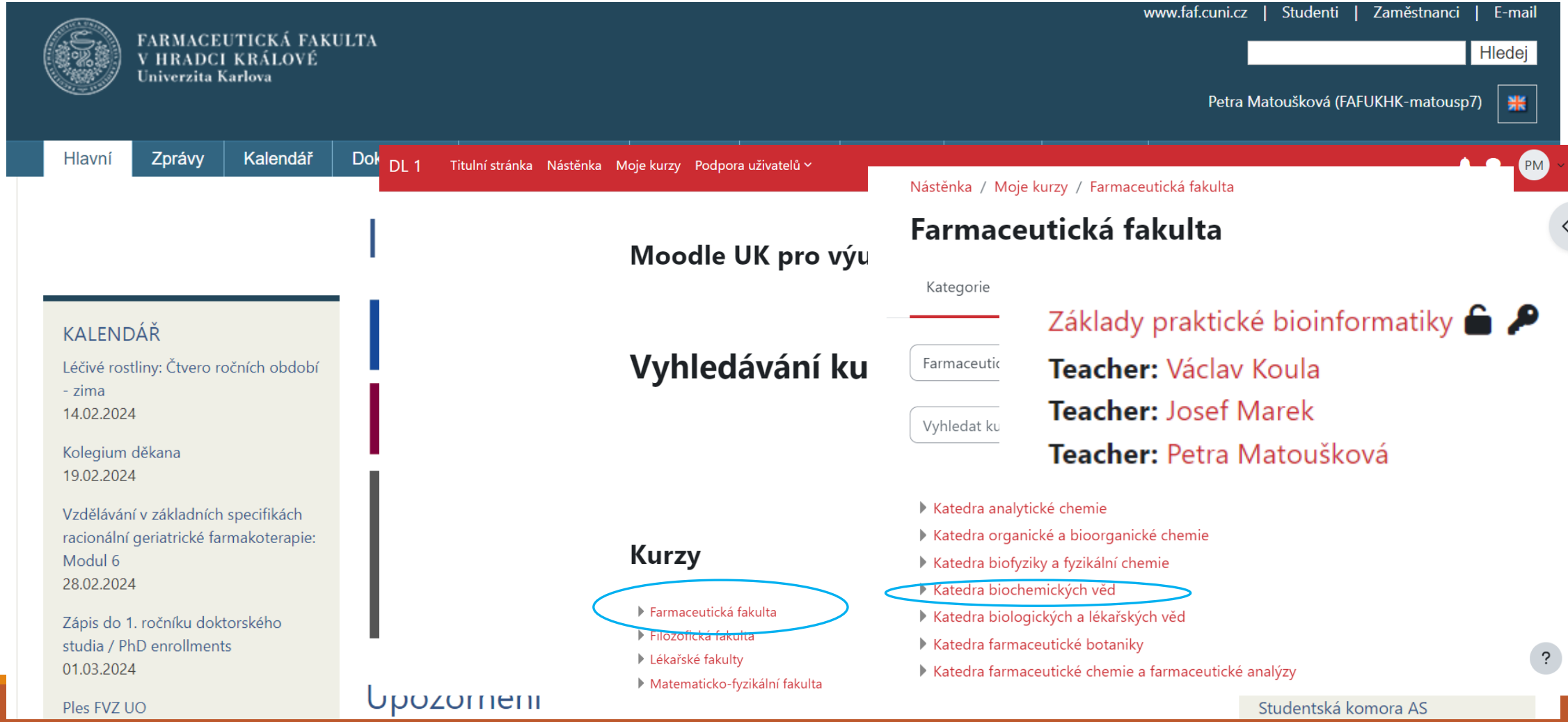
Určení evolučních vztahů mezi proteiny, kreslení evolučních stromů

Výpočet optimálního párového/mnohonásobného porovnání

Vyhledání genových polymorfismů (single nucleotide polymorphisms, SNPs)

Vyhledávání hladin expresí jednotlivých genů

Moodle (1)


FARMACEUTICKÁ FAKULTA
V HRADCI KRÁLOVÉ
Univerzita Karlova
www.faf.cuni.cz | Studenti | Zaměstnanci | E-mail
Hledej
Petra Matoušková (FAFUKHK-matousp7)
Hlavní
Zprávy
Kalendář
DL 1
Titulní stránka
Nástěnka
Moje kurzy
Podpora uživatelů
PM
Nástěnka / Moje kurzy / Farmaceutická fakulta
Farmaceutická fakulta
Moodle UK pro výu
Vyhledávání ku
Kategorie
Farmaceutic
Vyhledat ku
Základy praktické bioinformatiky
Teacher: Václav Koula
Teacher: Josef Marek
Teacher: Petra Matoušková
Kurzy
Farmaceutická fakulta
Filozofická fakulta
Lékařské fakulty
Matematicko-fyzikální fakulta
Katedra analytické chemie
Katedra organické a bioorganické chemie
Katedra biofyziky a fyzikální chemie
Katedra biochemických věd
Katedra biologických a lékařských věd
Katedra farmaceutické botaniky
Katedra farmaceutické chemie a farmaceutické analýzy
KALENDÁŘ
Léčivé rostliny: Čtvero ročních období - zima
14.02.2024
Kolegium děkana
19.02.2024
Vzdělávání v základních specifikách racionální geriatrické farmakoterapie: Modul 6
28.02.2024
Zápis do 1. ročníku doktorského studia / PhD enrollments
01.03.2024
Ples FVZ UO
Upozornění
Studentská komora AS

Základy praktické bioinformatiky

Téma 2/10

Proteinová bioinformatika I

Cíle:

Student bude schopen vyhledat a stáhnout požadovanou sekvenci proteinu. Na základě sekvence bude schopen určit (nalézt nebo spočítat) fyzikálně-biochemické vlastnosti proteinu.

Rešeršní projekt: NQO1

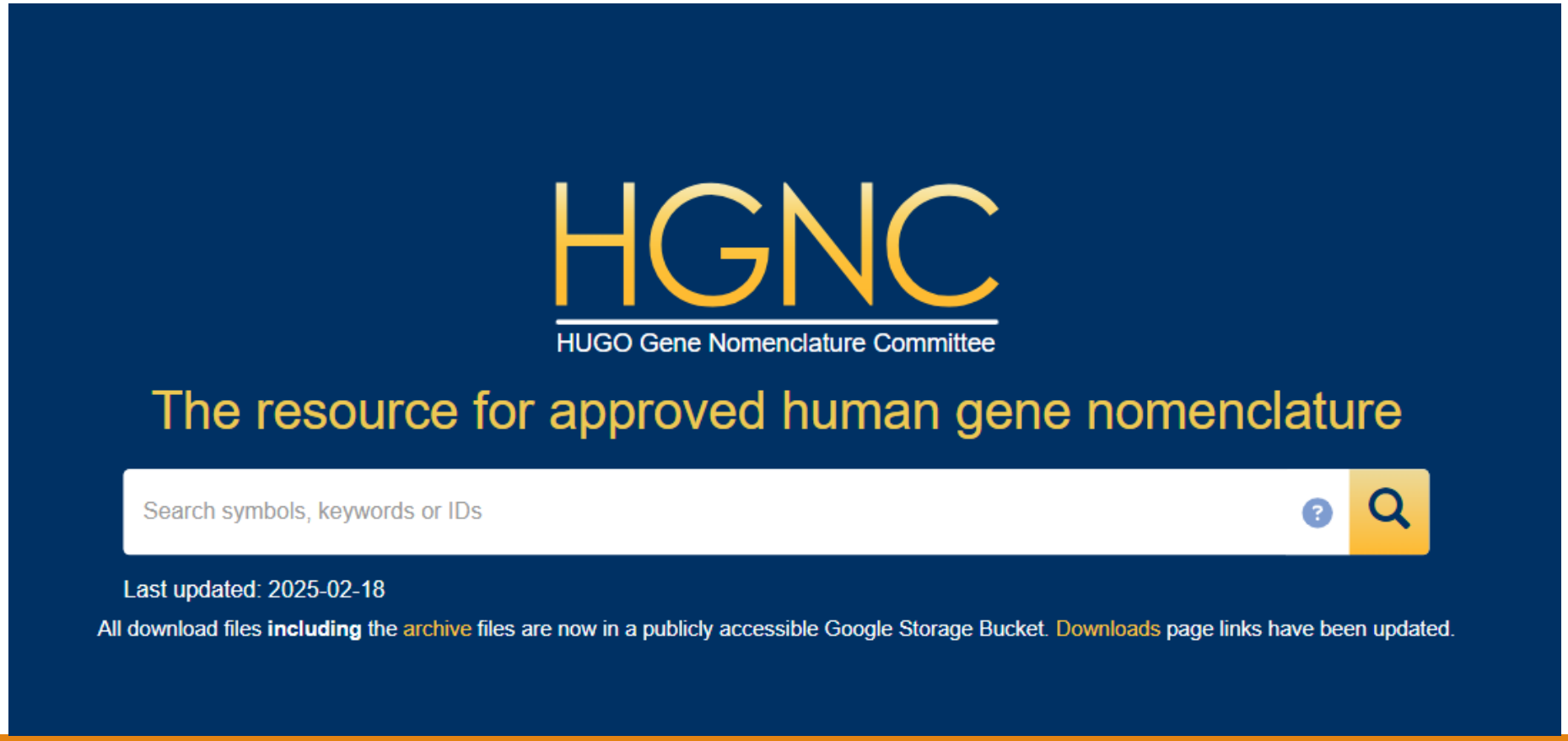
Vyhledejte dostupné informace o NQO1, získejte co nejvíce literárních, sekvenčních, případně i strukturních a dalších údajů o tomto genu/proteinu.

Nalezněte podobné sekvence a porovnejte je na proteinové úrovni.

- **NAD(P)H:chinonoxidoreduktasa / reference**

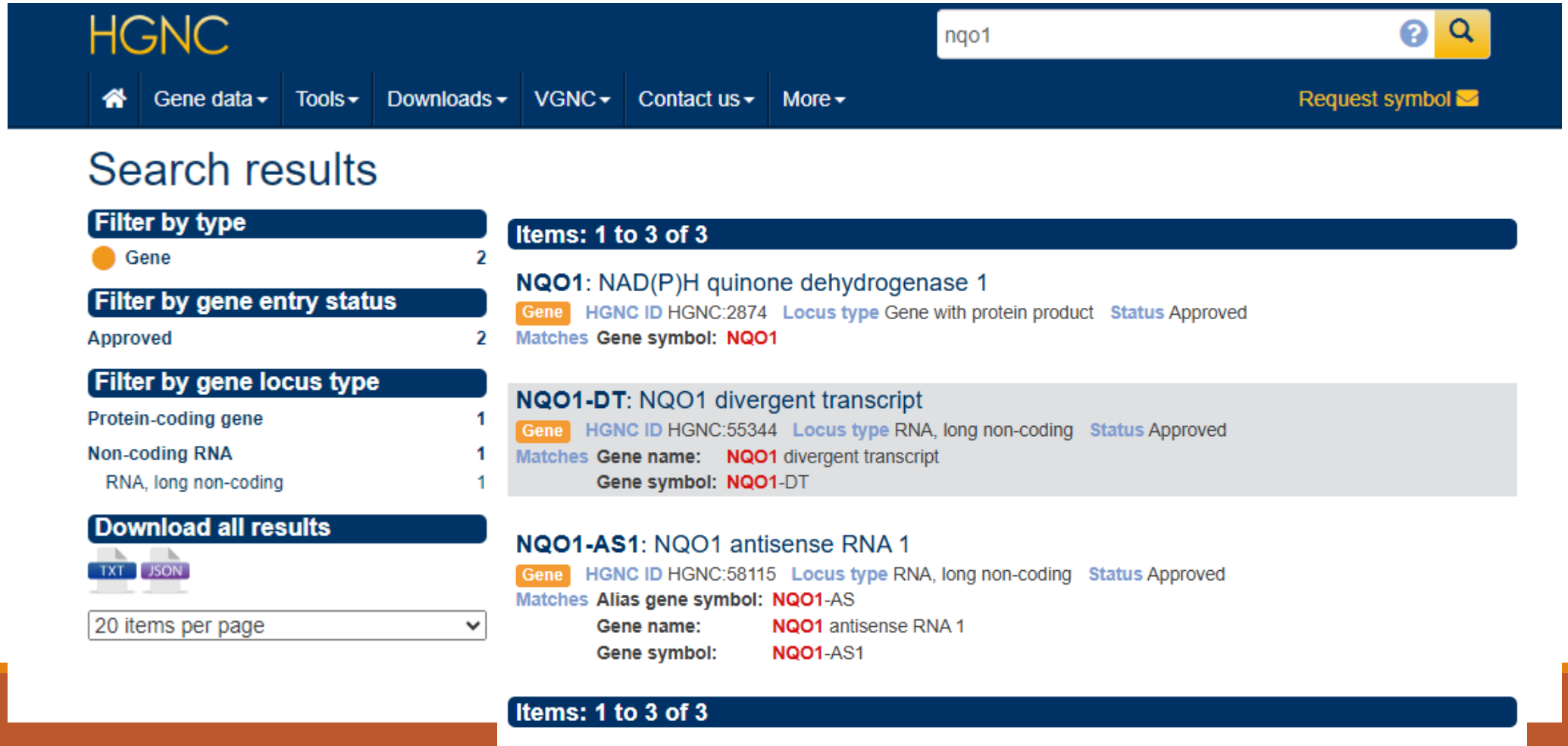
Identifikujte správně váš gen/protein

HUGO database



Identifikujte správně váš gen/protein

HUGO database



„Proteinová bioinformatika I“

Vyhledávání AMK sekvencí

Analýza vlastností sekvencí (aminokyselinové složení, molekulová hmotnost, isoelektrický bod…)

Štěpení proteasami

Analýza hydrofobních segmentů, transmembránových úseků

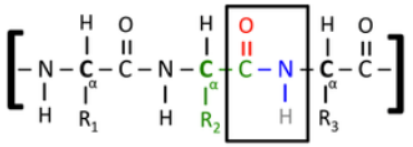
Predikce sekundární struktury

3D-struktura, vizualizace

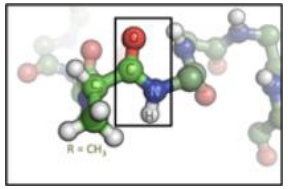
Vyhledání a porovnání podobných sekvencí

Evoluční příbuznost sekvencí

…



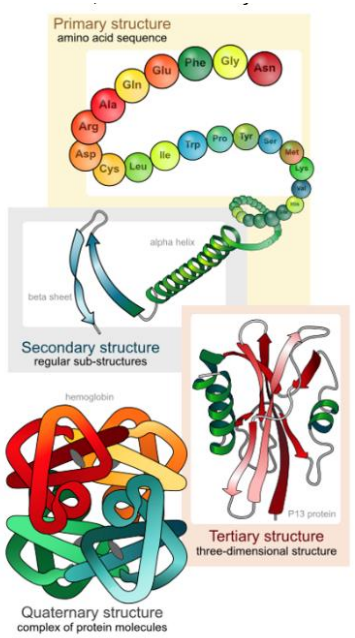
Proteiny



20 Aminokyselin – primární struktura:

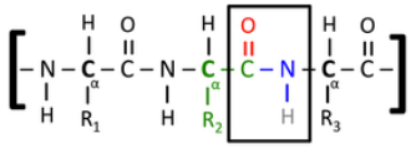
(Frederick Sanger-1958 Nobelova cena za sekvenování insulinu)

Sekundární struktura
Terciární struktura
Kvartérní struktura

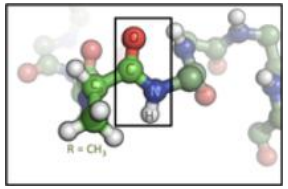


1-letter code	3-letter code	Amino acid	Possible codons
A	Ala	Alanine	GCA, GCC, GCG, GCT
B	Asx	Asparagine or Aspartic acid	AAC, AAT, GAC, GAT
C	Cys	Cysteine	TGC, TGT
D	Asp	Aspartic acid	GAC, GAT
E	Glu	Glutamic acid	GAA, GAG
F	Phe	Phenylalanine	TTC, TTT
G	Gly	Glycine	GGA, GGC, GGG, GGT
H	His	Histidine	CAC, CAT
I	Ile	Isoleucine	ATA, ATC, ATT
K	Lys	Lysine	AAA, AAG
L	Leu	Leucine	CTA, CTC, CTG, CTT, TTA, TTG
M	Met	Methionine	ATG
N	Asn	Asparagine	AAC, AAT
P	Pro	Proline	CCA, CCC, CCG, CCT
Q	Gln	Glutamine	CAA, CAG
R	Arg	Arginine	AGA, AGG, CGA, CGC, CGG, CGT
S	Ser	Serine	AGC, AGT, TCA, TCC, TCG, TCT
T	Thr	Threonine	ACA, ACC, ACG, ACT
V	Val	Valine	GTA, GTC, GTG, GTT
W	Trp	Tryptophan	TGG
X	X	Stop codon	TAA, TAG, TGA
Y	Tyr	Tyrosine	TAC, TAT
Z	Glx	Glutamine or Glutamic acid	CAA, CAG, GAA, GAG

SEKVENCE⇨STRUKTURA⇨FUNKCE

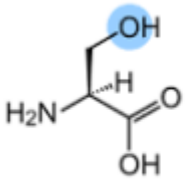


Proteiny

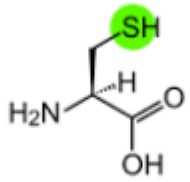


20 Aminokyselin – primární struktura:

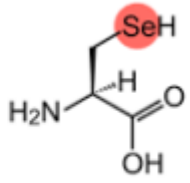
J	Xle	Isoleucine/Leucine
O	Pyl	Pyrrolysine
U	Sec	Selenocysteine
X	Xaa	Any residue



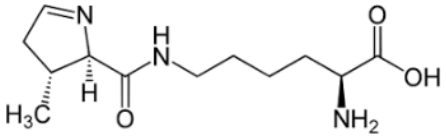
Serine (Ser)



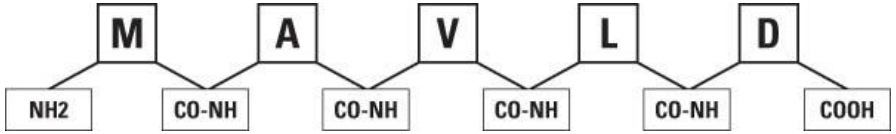
Cysteine (Cys)



Selenocysteine (Sec)



N-konec → C-konec



1-letter code	3-letter code	Amino acid	Possible codons
A	Ala	Alanine	GCA, GCC, GCG, GCT
B	Asx	Asparagine or Aspartic acid	AAC, AAT, GAC, GAT
C	Cys	Cysteine	TGC, TGT
D	Asp	Aspartic acid	GAC, GAT
E	Glu	Glutamic acid	GAA, GAG
F	Phe	Phenylalanine	TTC, TTT
G	Gly	Glycine	GGA, GGC, GGG, GGT
H	His	Histidine	CAC, CAT
I	Ile	Isoleucine	ATA, ATC, ATT
K	Lys	Lysine	AAA, AAG
L	Leu	Leucine	CTA, CTC, CTG, CTT, TTA, TTG
M	Met	Methionine	ATG
N	Asn	Asparagine	AAC, AAT
P	Pro	Proline	CCA, CCC, CCG, CCT
Q	Gln	Glutamine	CAA, CAG
R	Arg	Arginine	AGA, AGG, CGA, CGC, CGG, CGT
S	Ser	Serine	AGC, AGT, TCA, TCC, TCG, TCT
T	Thr	Threonine	ACA, ACC, ACG, ACT
V	Val	Valine	GTA, GTC, GTG, GTT
W	Trp	Tryptophan	TGG
X	X	Stop codon	TAA, TAG, TGA
Y	Tyr	Tyrosine	TAC, TAT
Z	Glx	Glutamine or Glutamic acid	CAA, CAG, GAA, GAG

Databáze

PRIMÁRNÍ

Ukládání primárních dat (sekvence, struktury…)

Přímé vkládání

Genbank, DDBJ, ENA

Protein Data Bank

GEO…

SEKUNDÁRNÍ

Shromažďování, analýza a propojování dat

Primární databáze, literatura (po analýze a interpretaci primárních dat)

SwissProt, InterPro, Ensembl, OMIM, STRING…

THE HUMAN PROTEIN ATLAS

RESOURCES ABOUT NEWS LEARN DATA HELP

The open access resource for human proteins

Search for specific genes/proteins or explore the nine different resources

nqo1 Search Fields »

e.g. ACE2, GFAP, EGFR Search help

Multiplex tissue image of the month - TMEM213 in salivary gland

Using multiplex immunohistochemistry (mIHC/IF)

The 2025 HUPO HPP report on the human proteome

THE HUMAN PROTEIN ATLAS RESOURCES ABOUT NEWS LEARN DATA HELP

nqo1 Search Fields »

Search result (1 genes): **NQO1**

NQO1

SUMMARY | TISSUE | BRAIN | SINGLE CELL | SUBCELL | CANCER | BLOOD | CELL LINE | STRUCTURE | INTERACTION

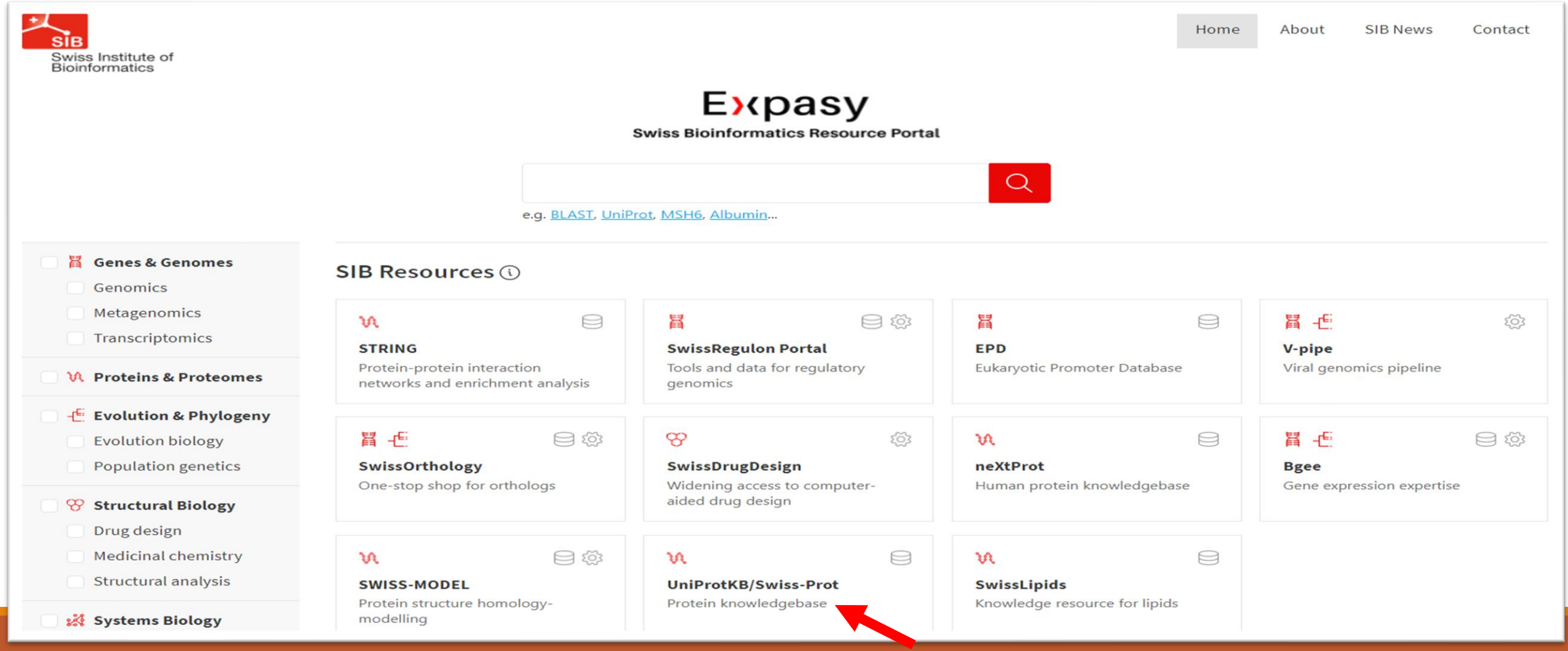
vyzkoušejte

Ve kterých tkáních je protein exprimován?

Jaká je jeho subcelulární lokalizace?

…

Proteinové databáze: UniProt



Možnosti databáze UniProt

http://www.uniprot.org/

Vyhledávání sekvencí, vlastnosti sekvencí

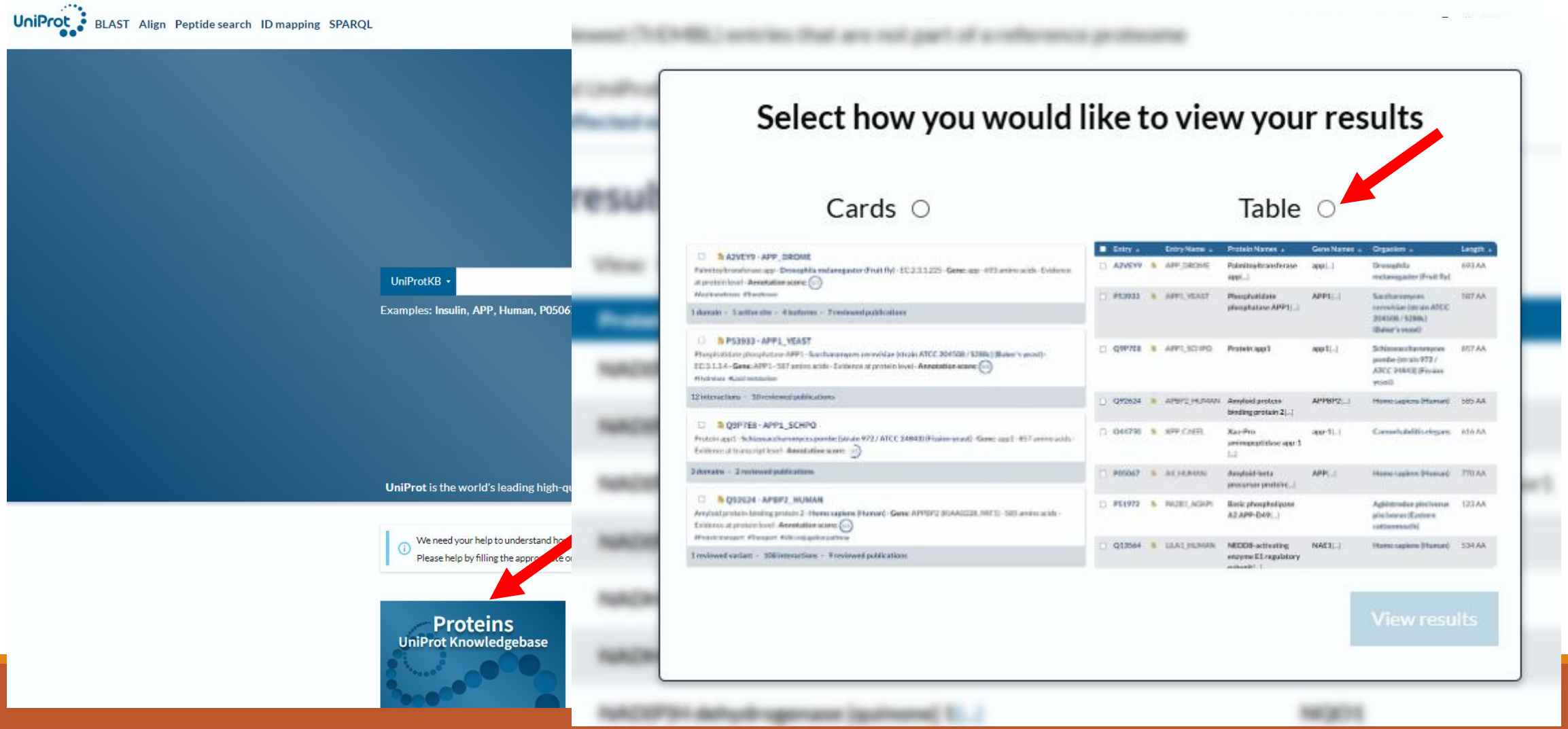
Funkce proteinů, interakce, dráhy

Propojení s nemocemi, mutace

Propojení do dalších databází

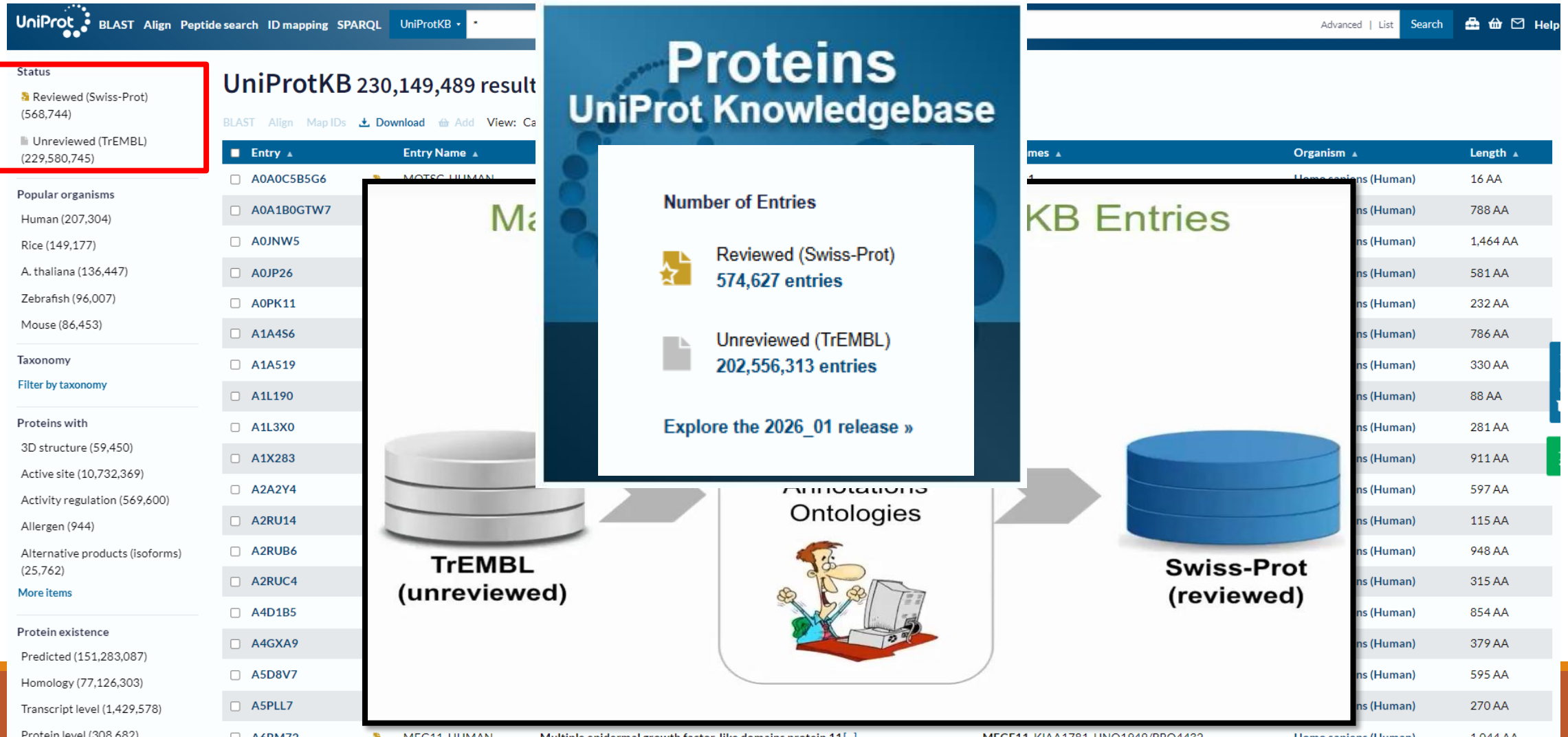
Proteinové databáze: UniProt

http://www.uniprot.org/



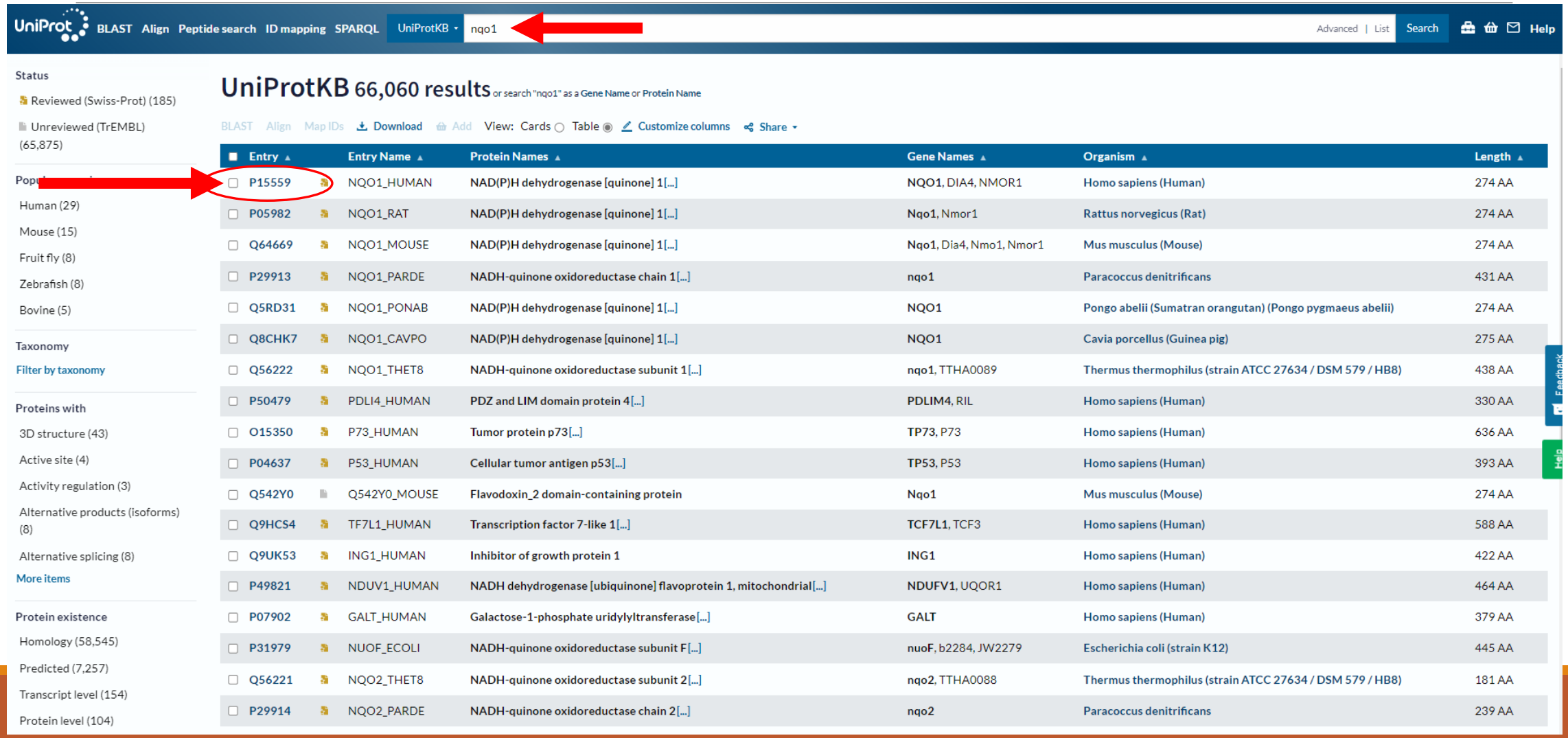
Proteinové databáze: UniProt

http://www.uniprot.org/

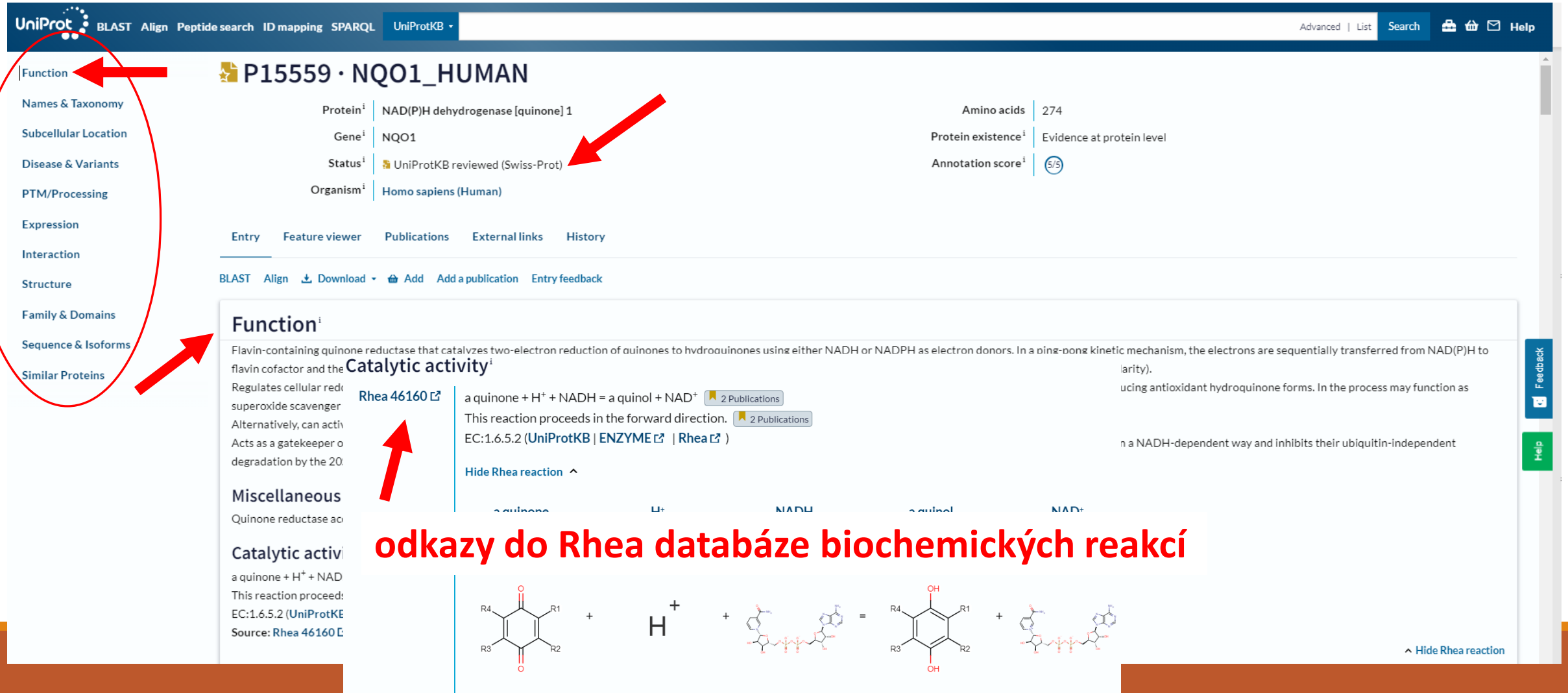


Proteinové databáze: UniProt

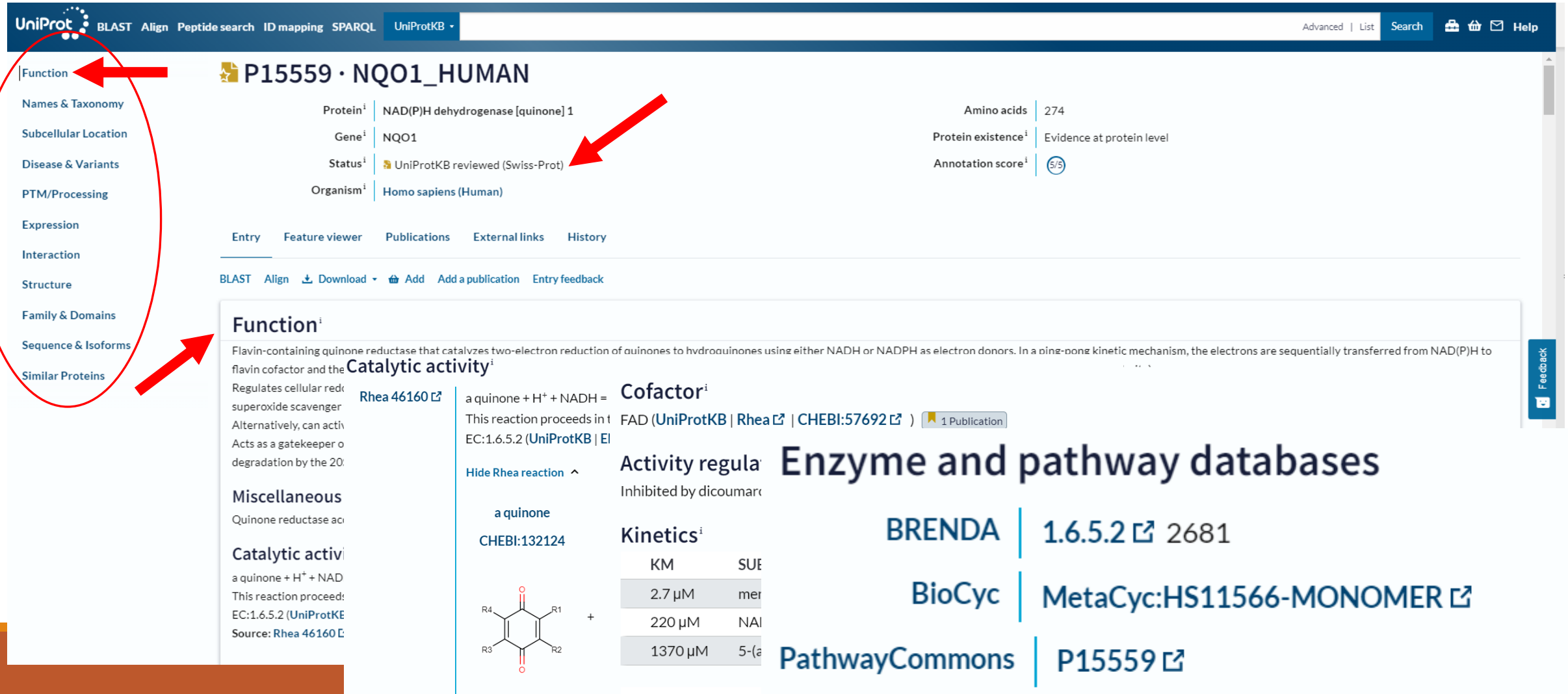
http://www.uniprot.org/



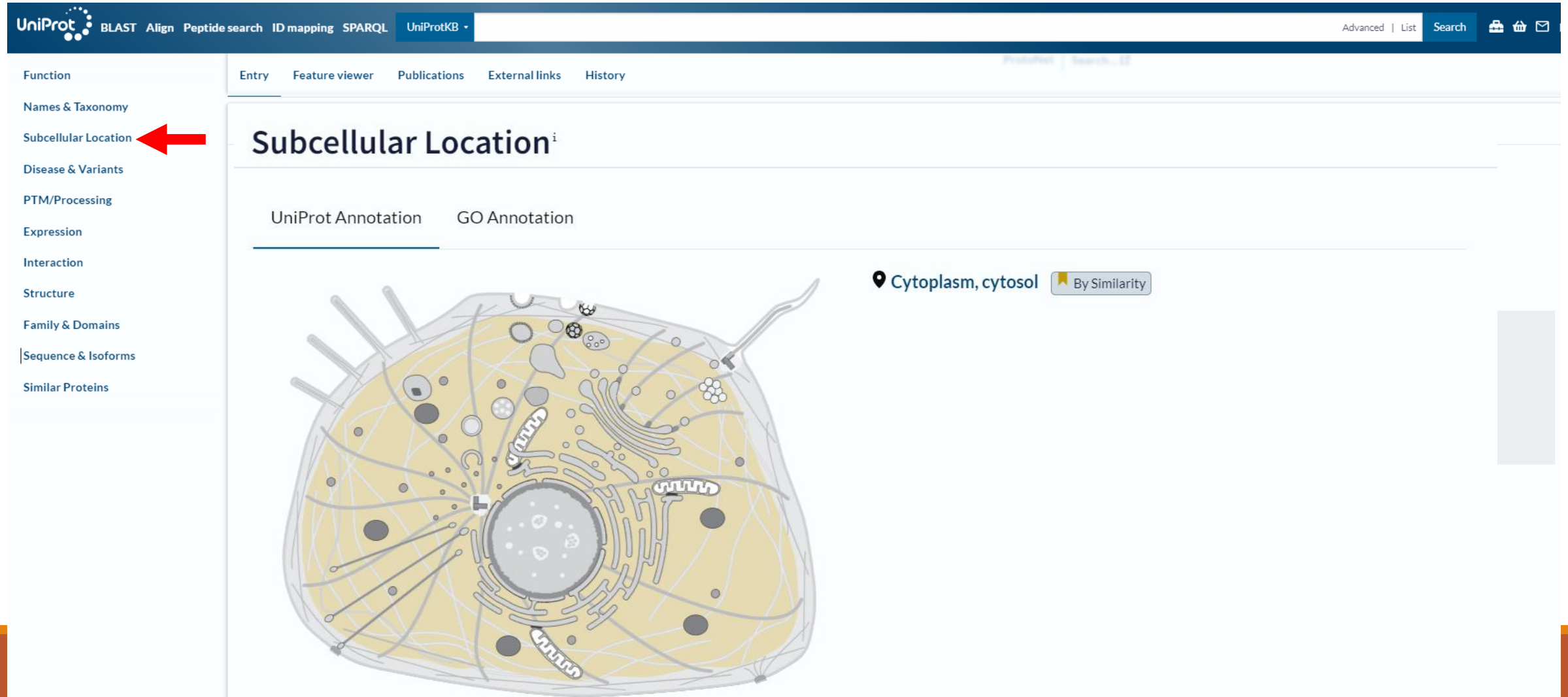
Proteinové databáze: UniProt



Proteinové databáze: UniProt



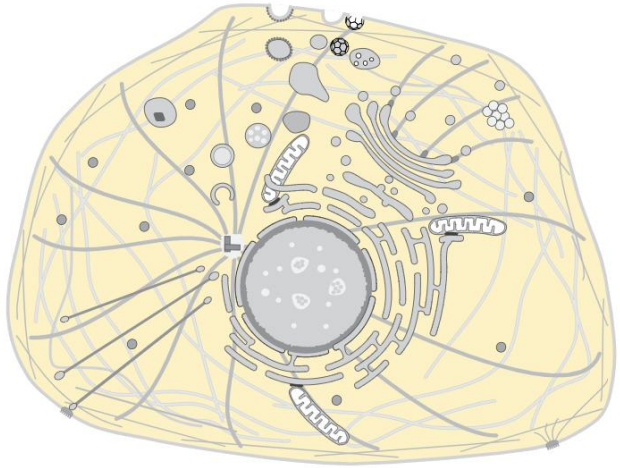
Proteinové databáze: UniProt

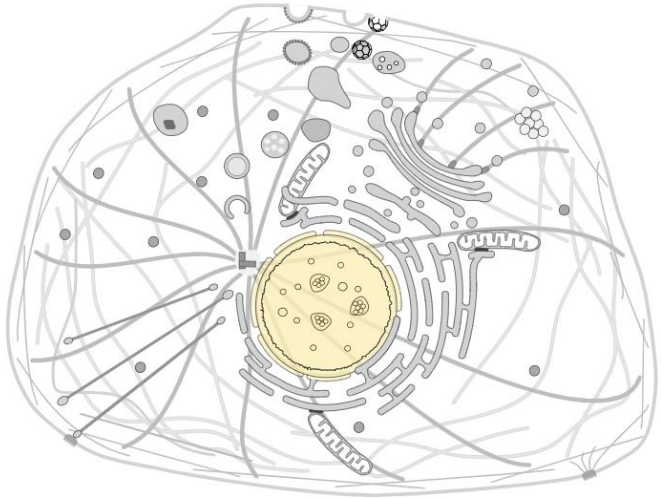


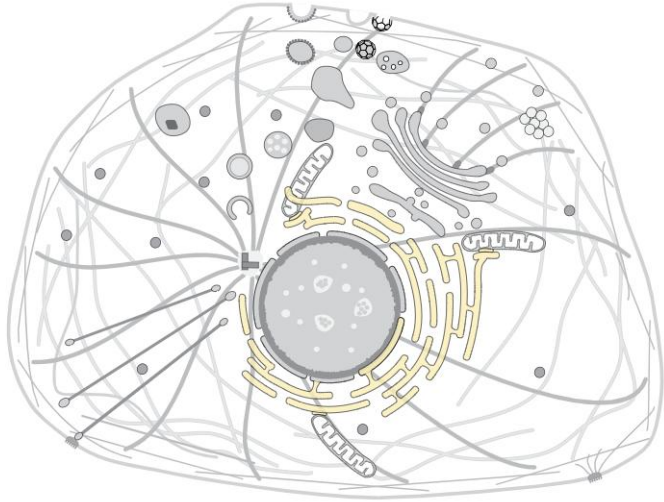
UNIPROT / subcelulární lokalizace

Subcellular location [i]

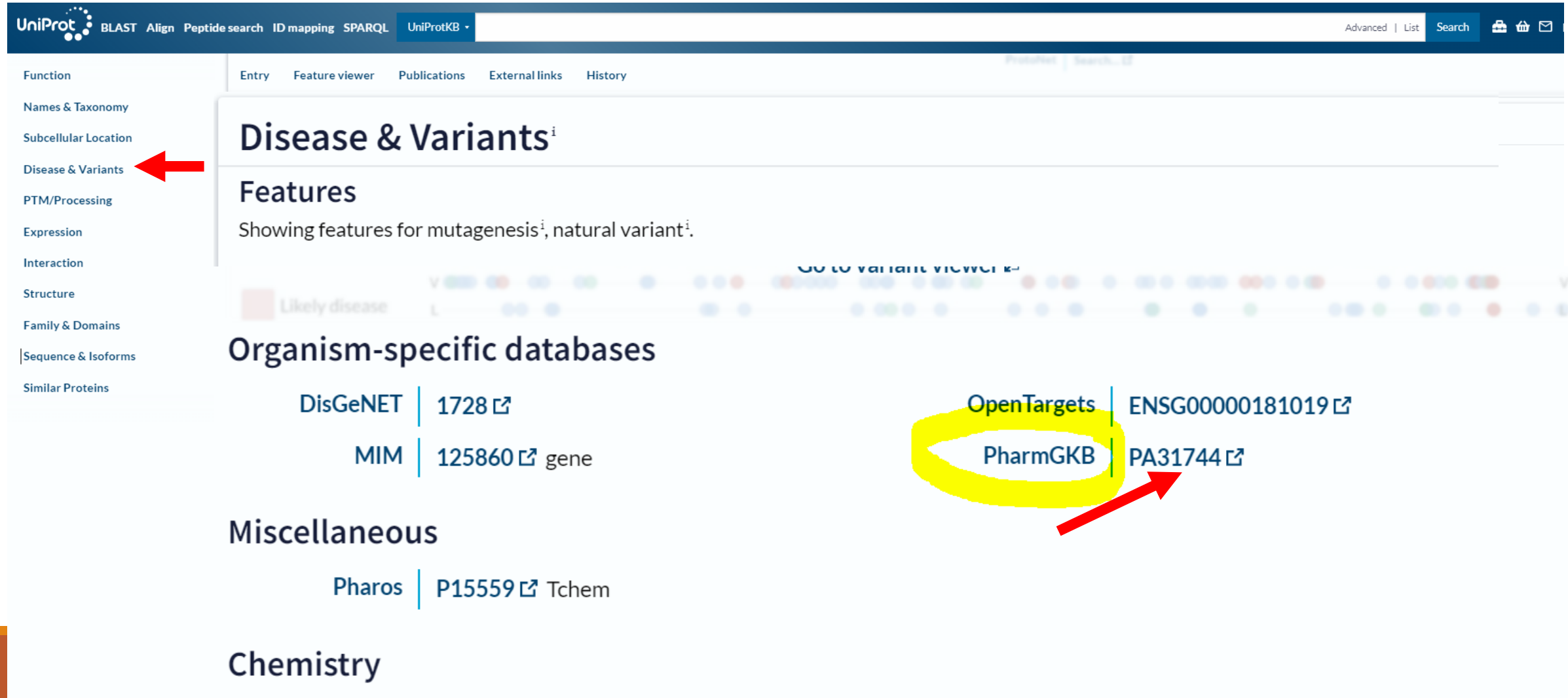
UniProt annotation | GO - Cellular component



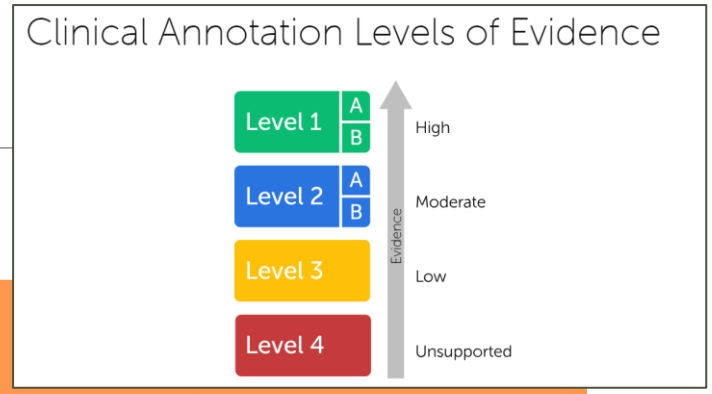




Proteinové databáze: UniProt



PHARMGKB



PHARMGKB

NQO1 × Add a term to make a combination...

NQO1

- Overview
- Prescribing Info
- Drug Label Annotations
- Clinical Annotations
- Variant Annotations
- Named Alleles
- Literature
- Pathways

PRESCRIBING INFO 0

CLINICAL ANNOTATIONS 8

PATHWAYS

Pediatric

Overview

VIP Tier 2

NQO1 catalyzes the reducti
resistance to chemotherape

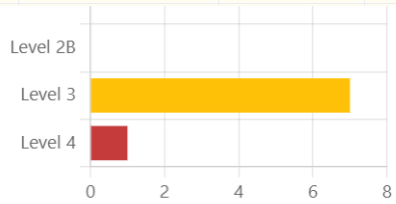
more of the Very Important

Clinical Annotations

PharmGKB clinical annotations provide in
pairs based primarily on variant annotatic
specific prescribing guidance from clinica
drug labels, when available. Curators mar
create genotype-based summaries descri
the variant. Each clinical annotation is assigned a Level of Evidence, which is generally informed by the clinical annotation's score.

Note: Alleles in PharmGKB are mapped to the positive chromosomal strand. Therefore, variants in genes on the "minus" strand (eg. *VKORC1*) are complemented in PharmGKB annotations.

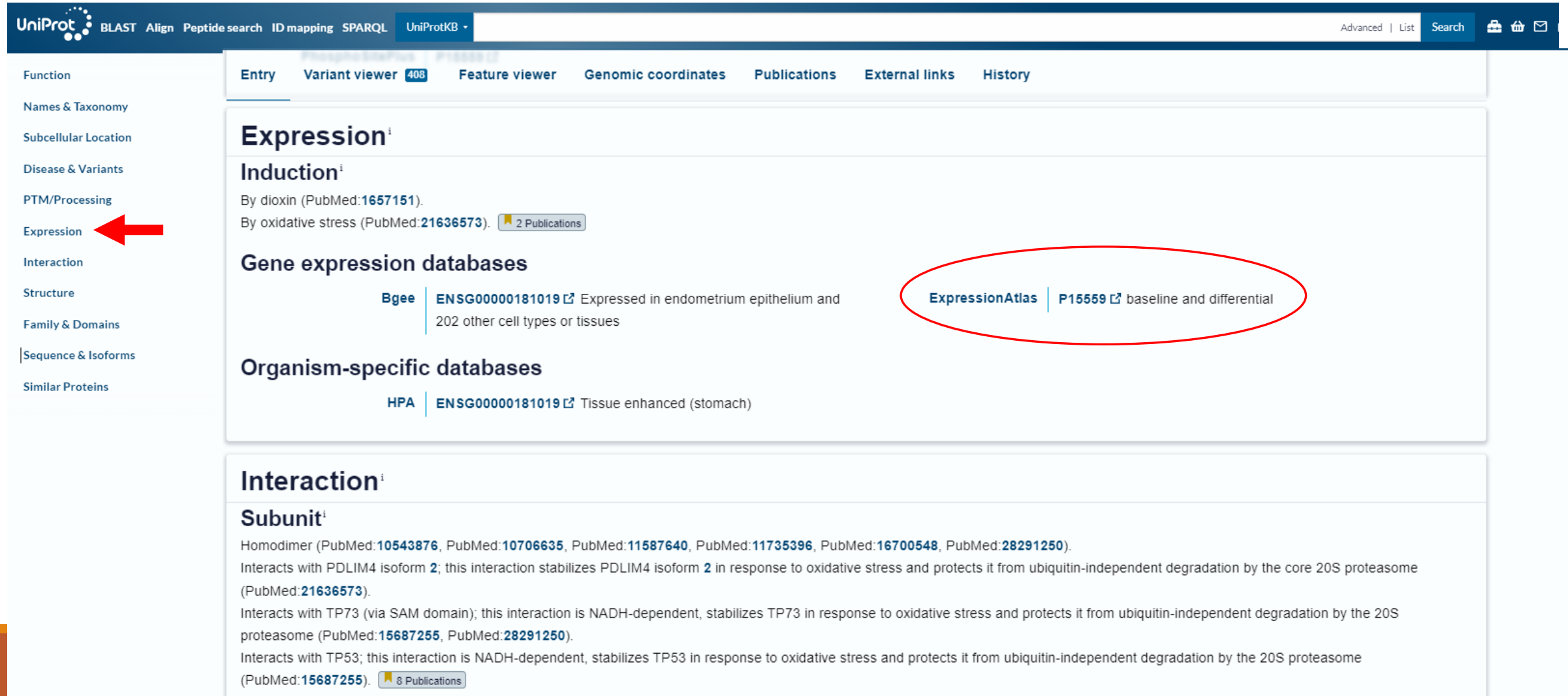
	LEVEL	VARIANT	GENE	DRUGS	PHENOTYPE CATEGORIES
Details	Level 3	rs10517	NQO1	imatinib	Efficacy
Details	Level 3	rs1800566	NQO1	epirubicin, fluorouracil, oxaliplatin	Efficacy
Details	Level 3	rs10517	NQO1	warfarin	Dosage
Details	Level 3	rs1800566	NQO1	warfarin	Efficacy



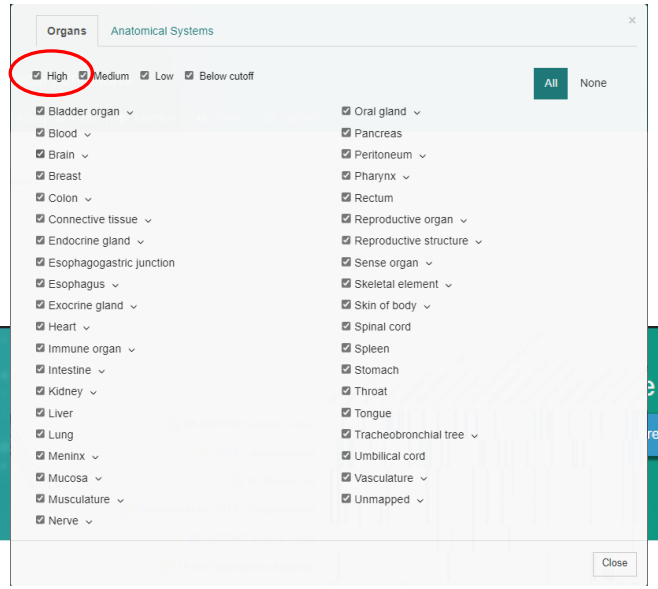
Annotations by level

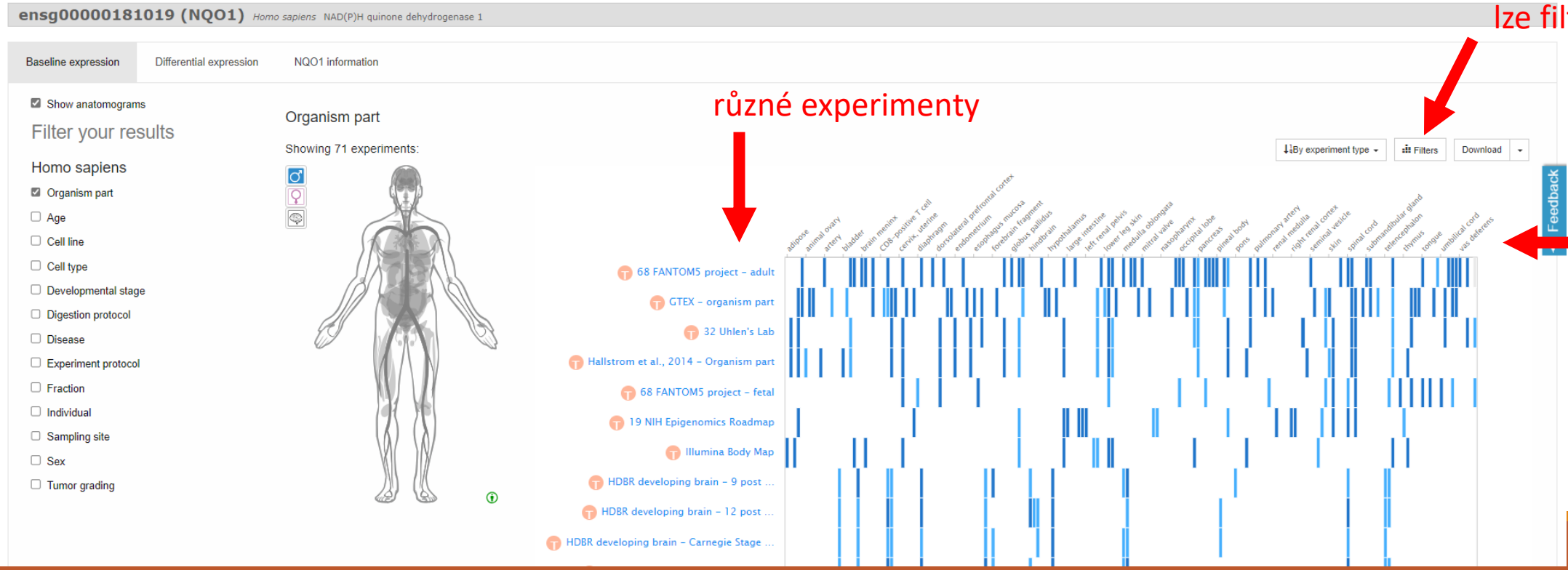
Location

Proteinové databáze: UniProt

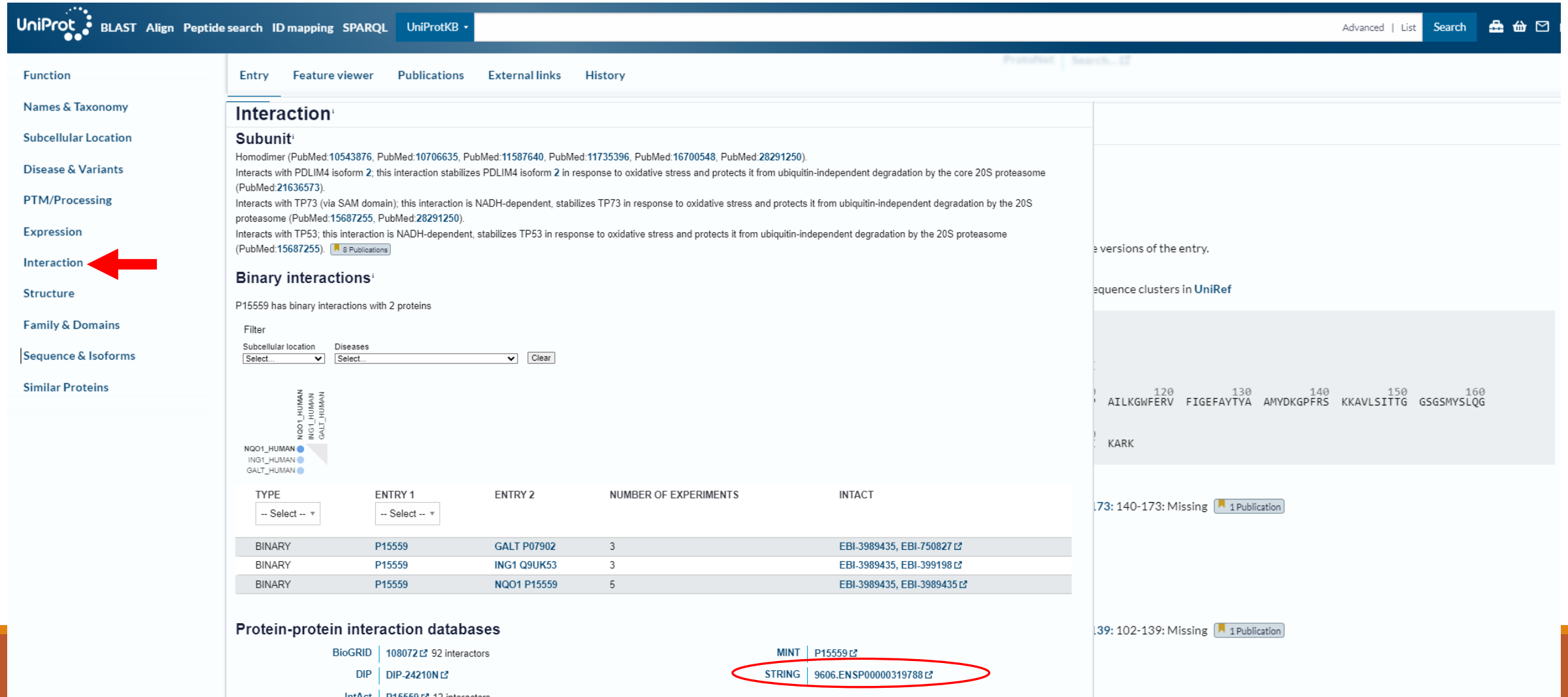


Gene Expression Atlas

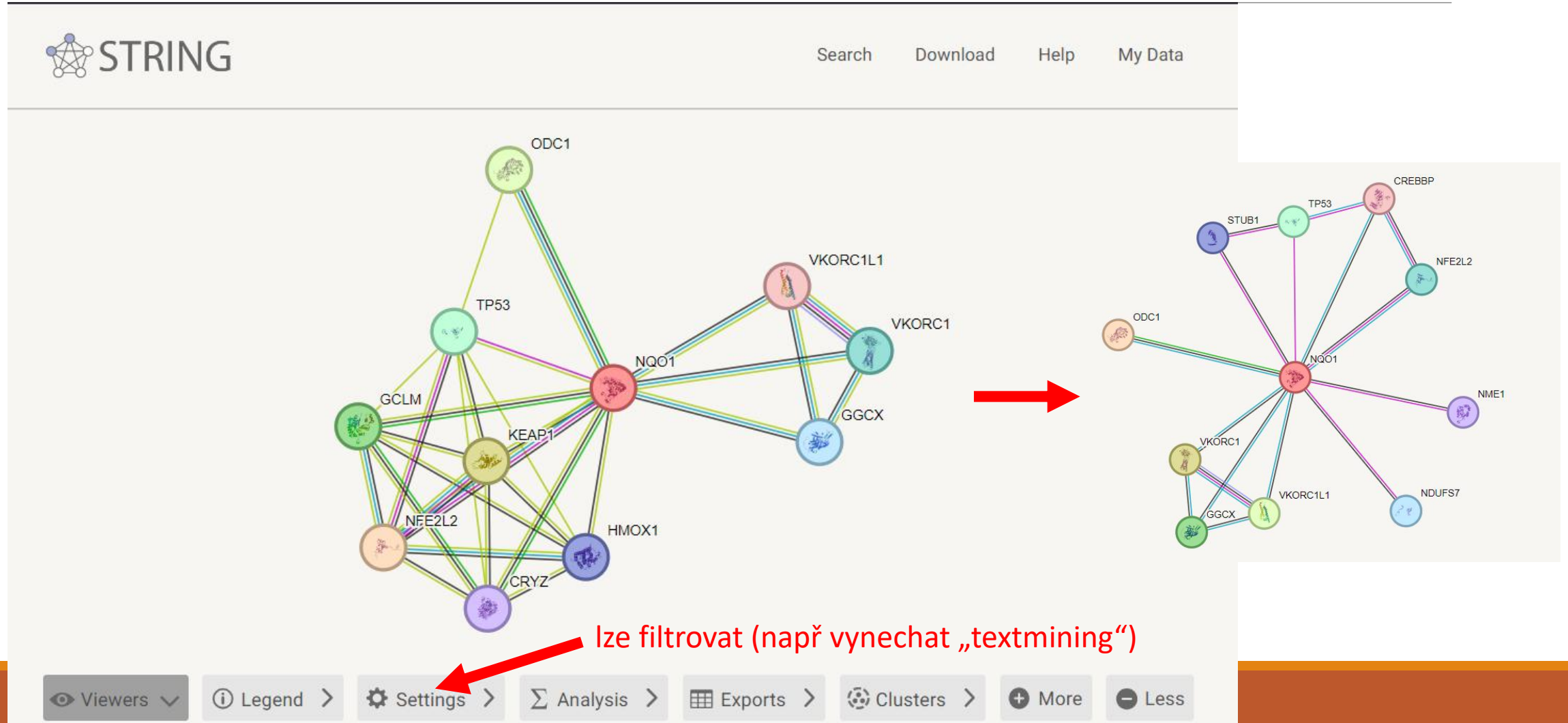




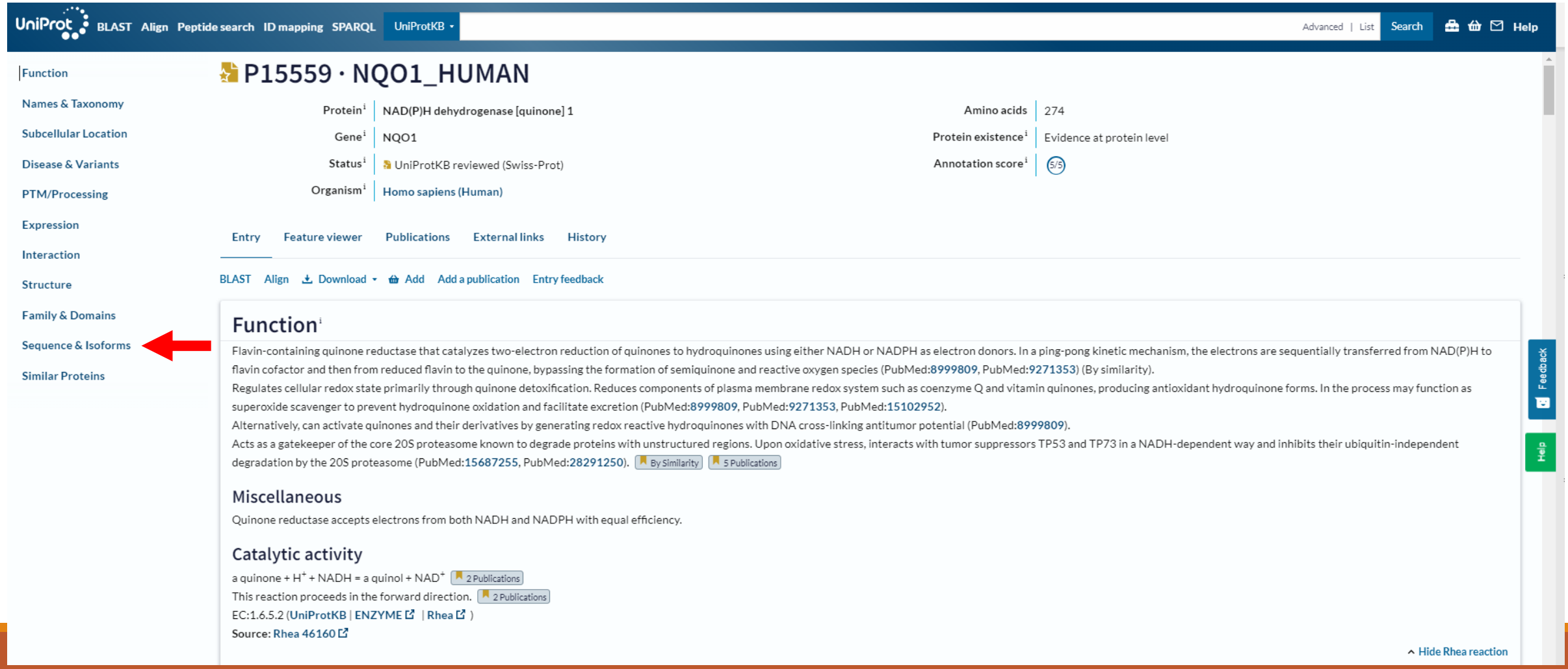
Proteinové databáze: UniProt



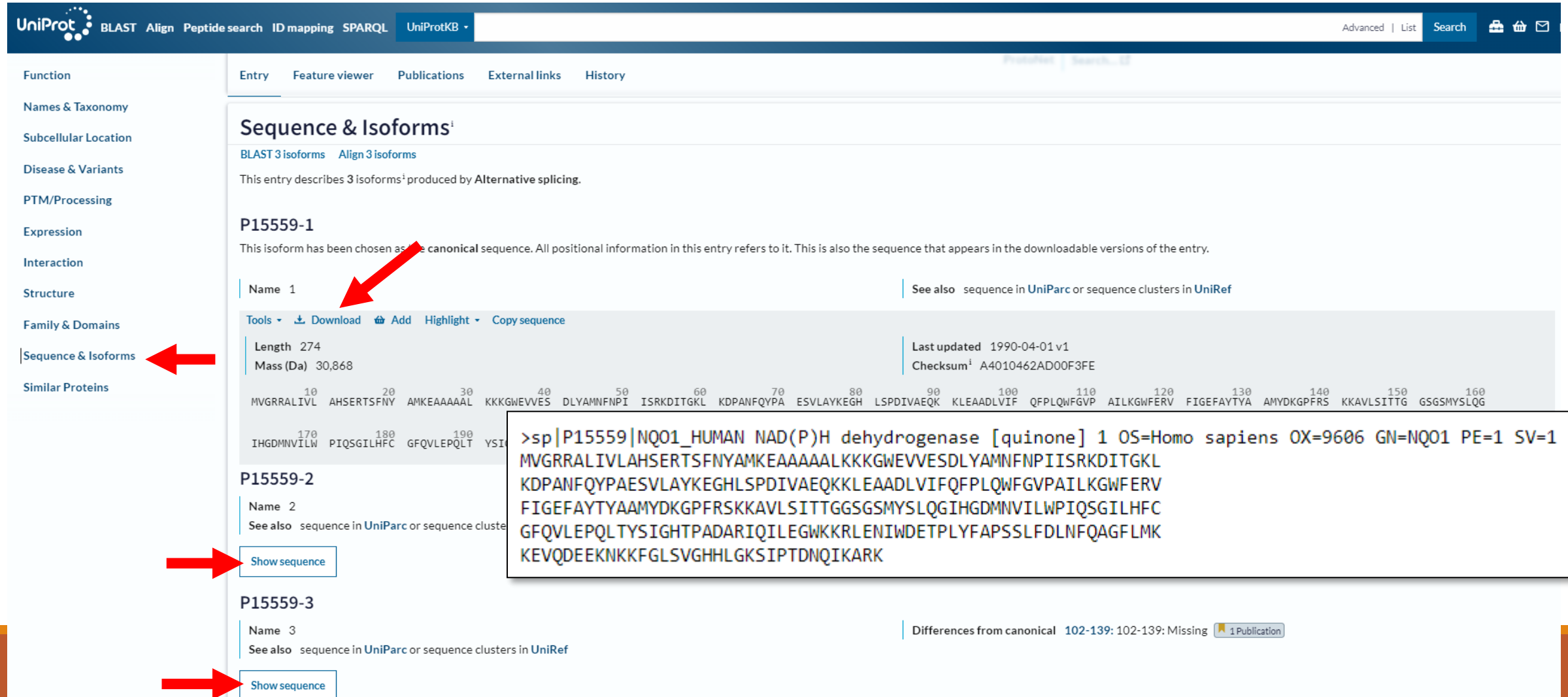
STRING: protein-protein interakce



Proteinové databáze: UniProt



Proteinové databáze: UniProt



FASTA (a RAW) formát

název odvozen:

FASTA = populární program pro porovnávání sekvencí a prohledávání databází

W.R. Pearson a D.J. Lipman 1988

následující formát sekvence:

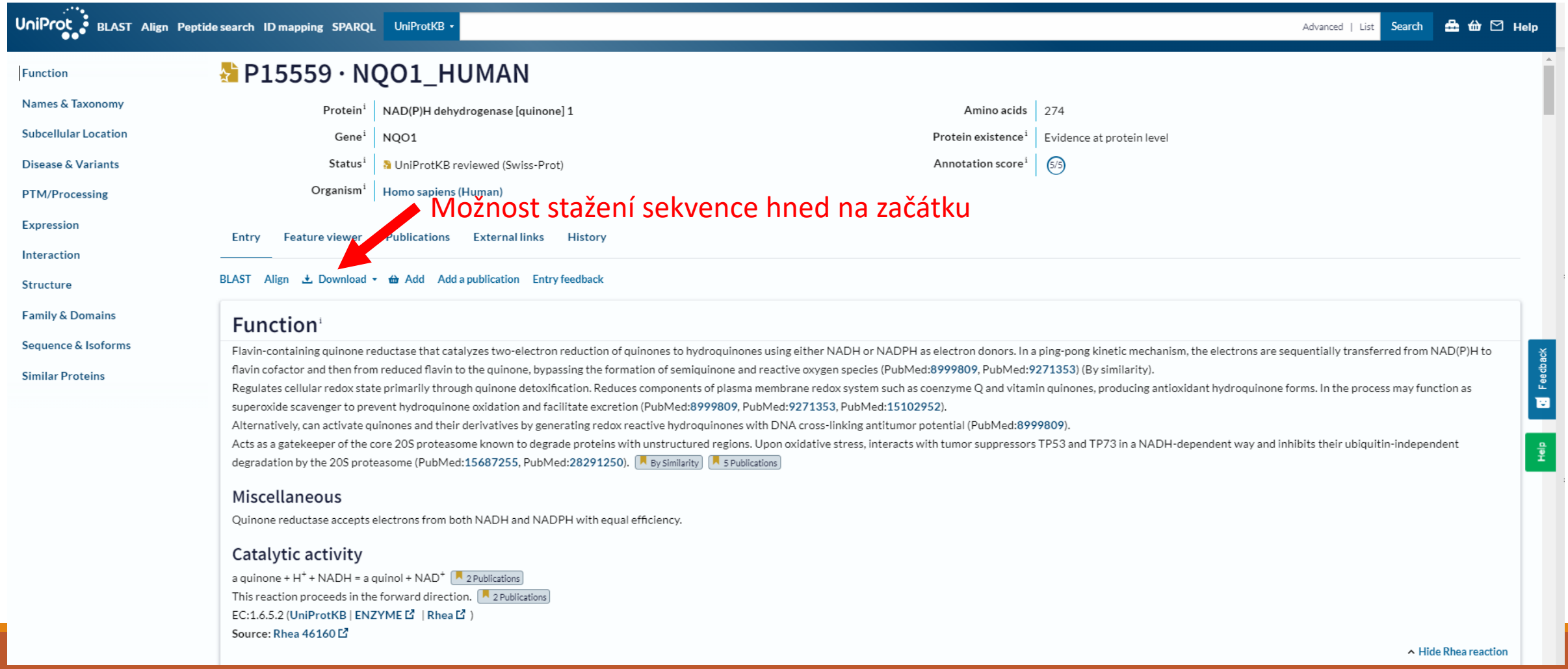
>NQO1_homo
MVGRRALIVLAHSERTSFNYAMKEAAAAA
LKKKGWEVVESDLYAMNFNPIISRKDITG
KLKDPANFQYPA

MVGRRALIVLAHSERTSFNYAMKEAAAAA
LKKKGWEVVESDLYAMNFNPIISRKDITG
KLKDPANFQYPA

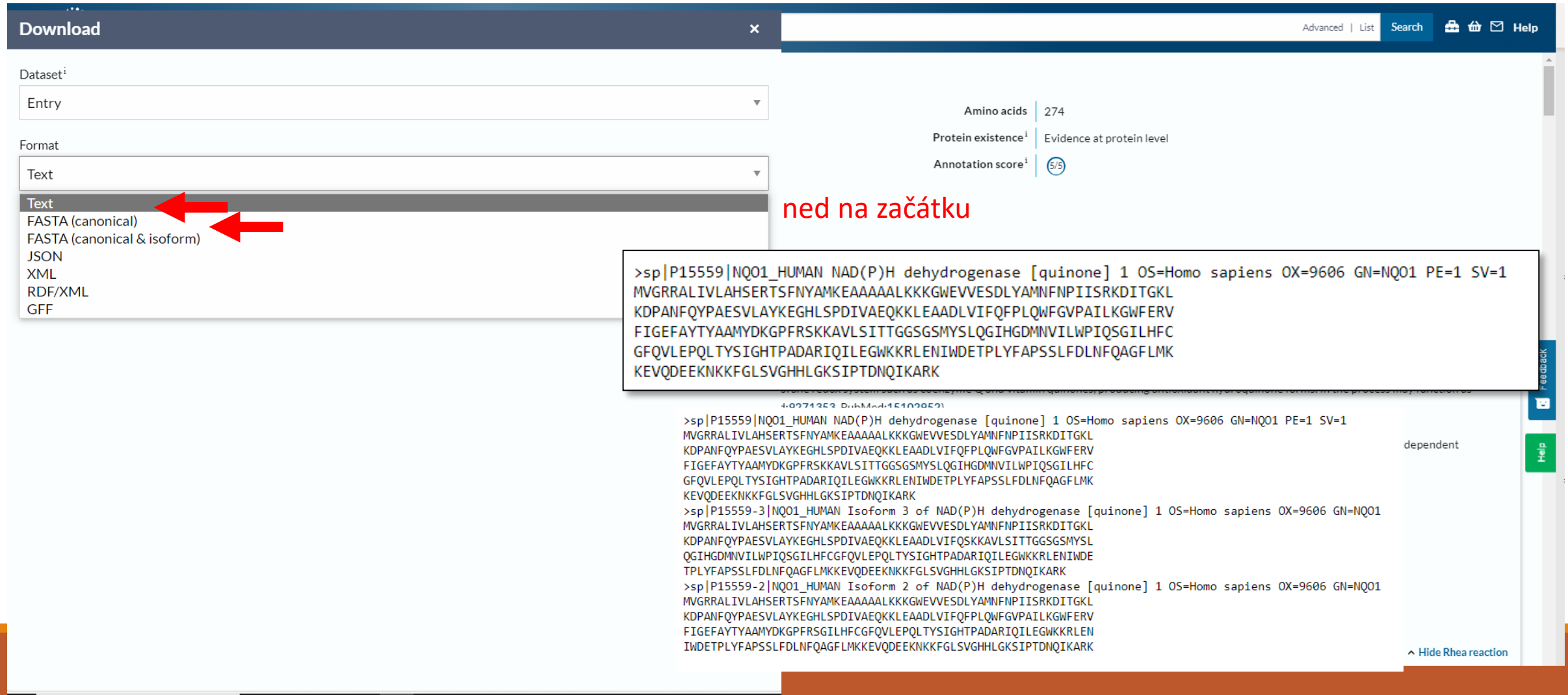
1. „definiční“ řádek (hlavička) začíná **>** a obsahuje unikátní identifikátor, následovaný krátkým popisem (nemusí být).

2. řádek – vlastní sekvence (DNA/protein- v jedno-písmenném kódování) až po další symbol >, určující začátek další sekvence.

Proteinové databáze: UniProt



Proteinové databáze: UniProt



Praktické procvičování

Vyzkoušejte si UniProt proteinovou databázi

Najděte sekvenci lidské GSTP1:

- jaké má přístupové číslo? (entry)
- jak dlouhý protein to je?
- jedná se o enzym?
- kde je lokalizován?
- V jakých tkáních je exprimován?
- Má nějaké isoformy?

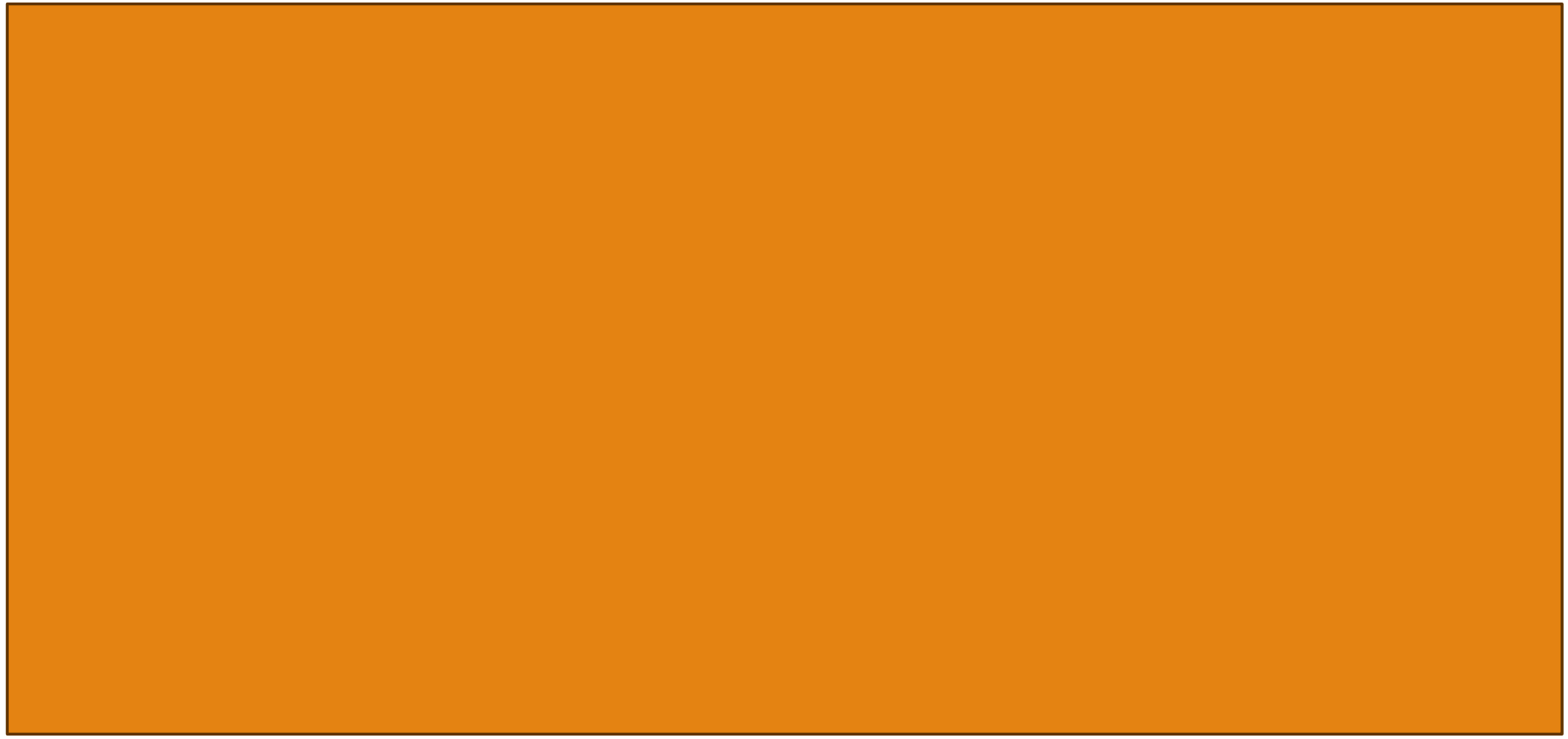
Procvičování v UniProt

Najděte přístupový kód lidského cytochromu P450 isoformy 3A4 (CYP3A4)

Jakou má funkci a kde je lokalizován enzym Q9C0B1?

Jakou má funkci enzym O95251 a kolik má isoforem?

Procvičování v UniProt



Procvičování v UniProt / DU

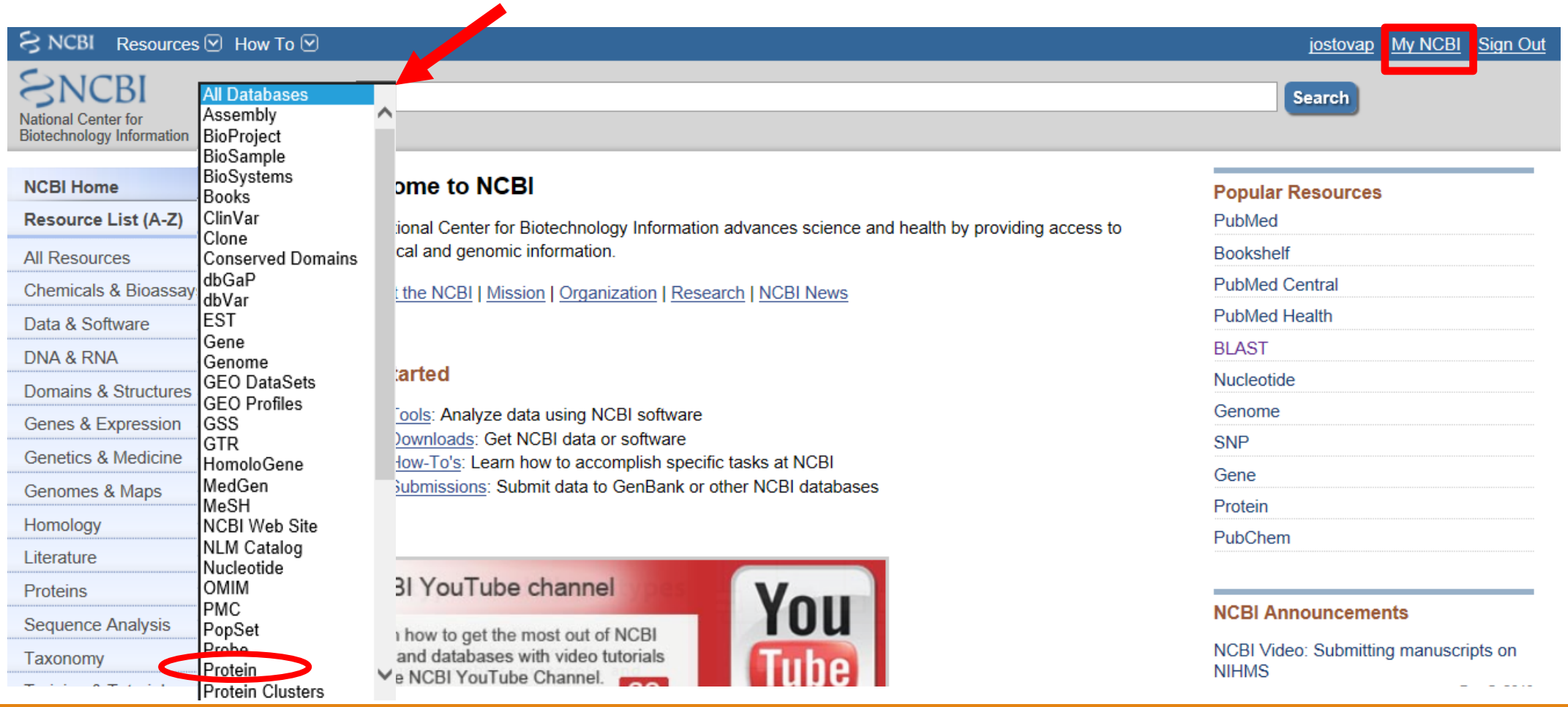
04:14

Najděte si „svou vlastní“ sekvenci

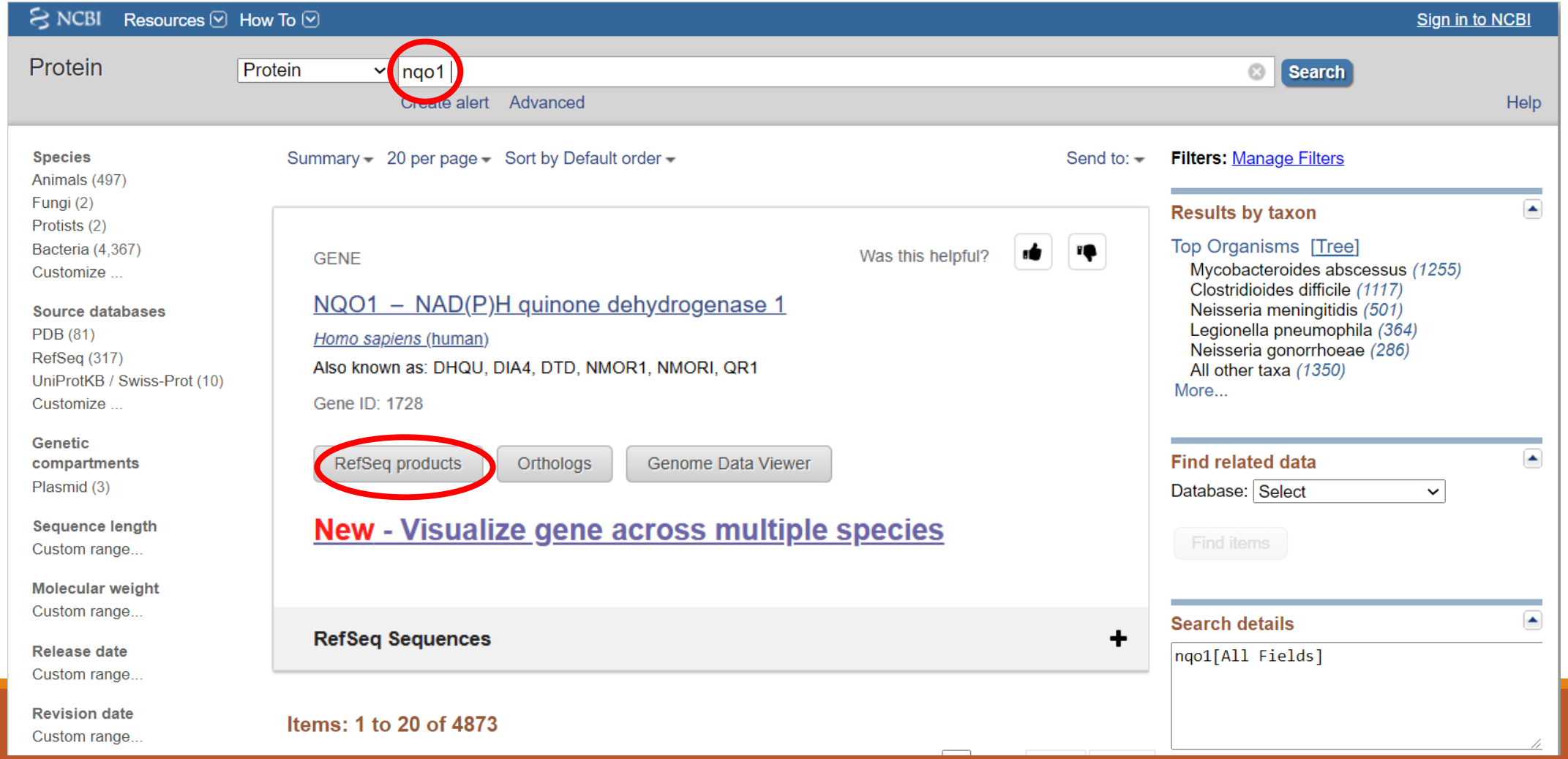
-podívejte se na: lokalizaci, počet isoforem…

- uložte si FASTA formát

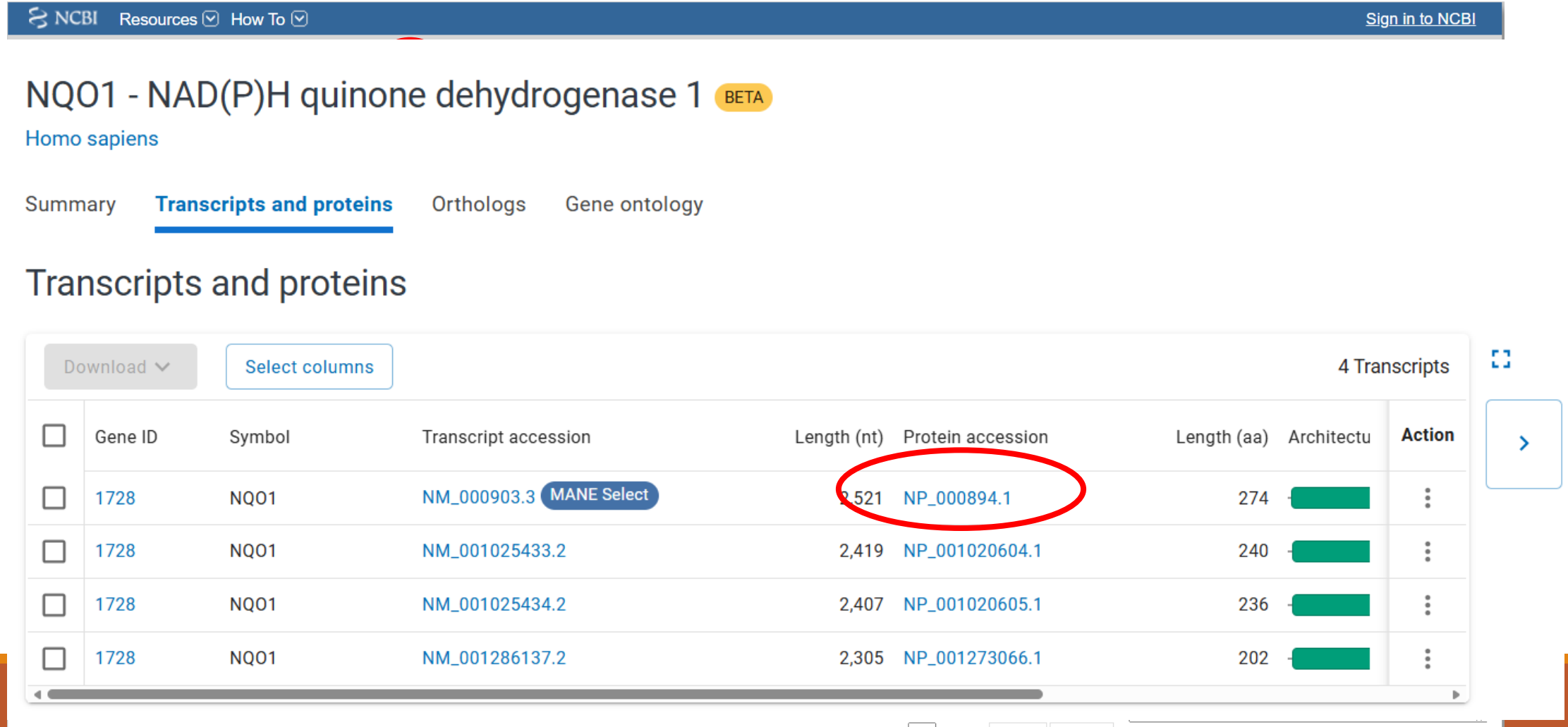
Proteinové databáze: NCBI



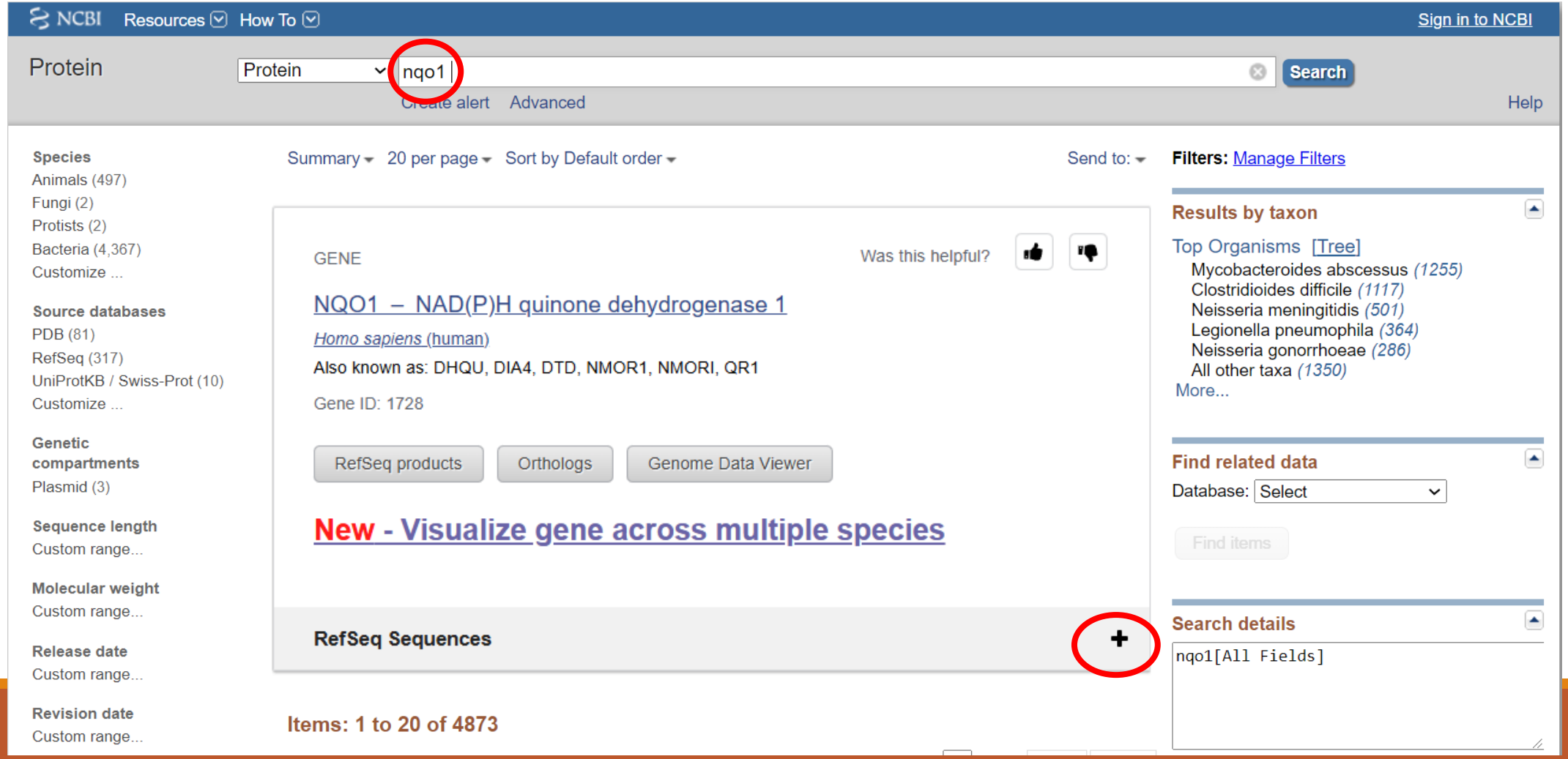
Proteinové databáze: NCBI



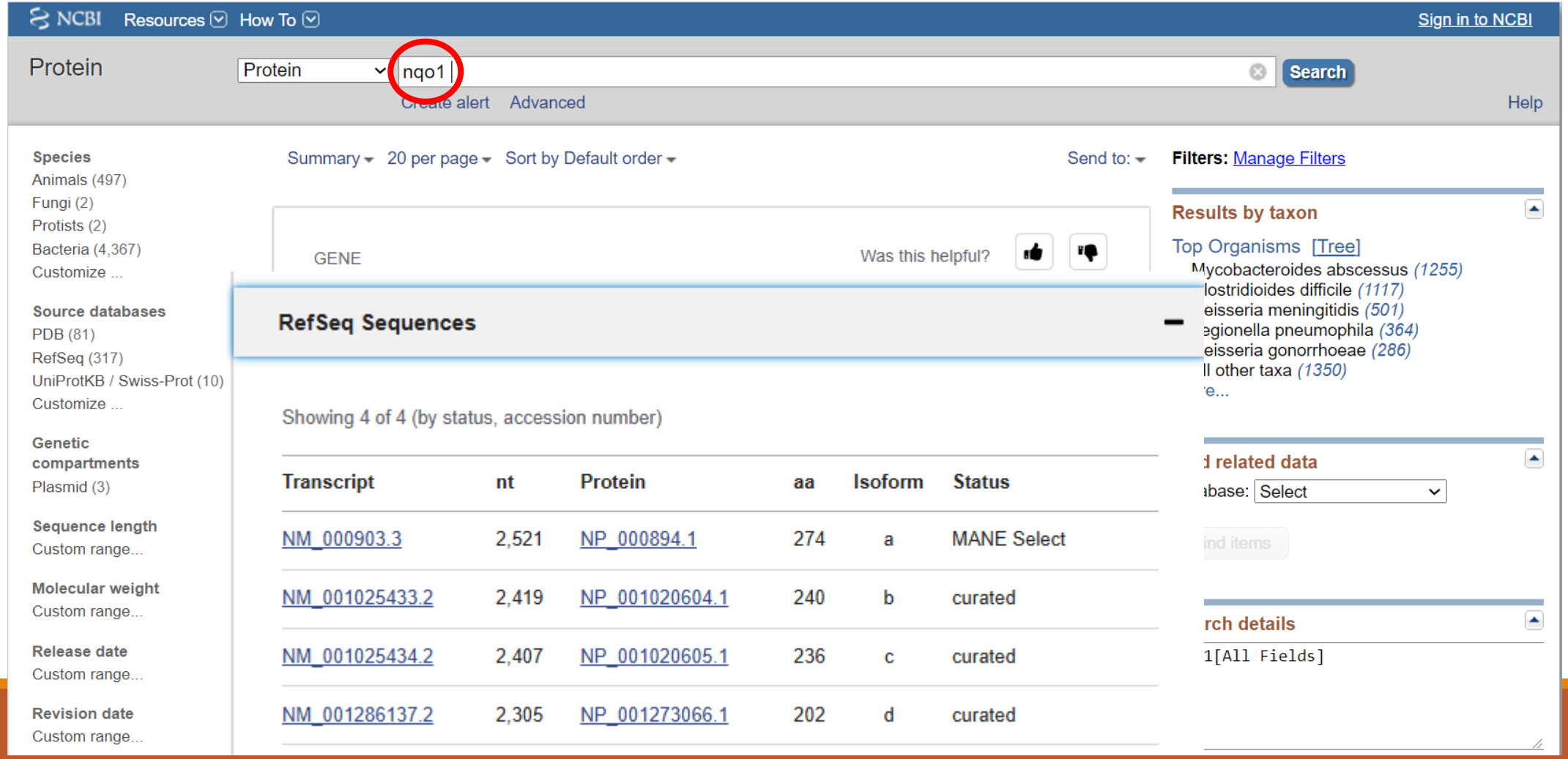
Proteinové databáze: NCBI



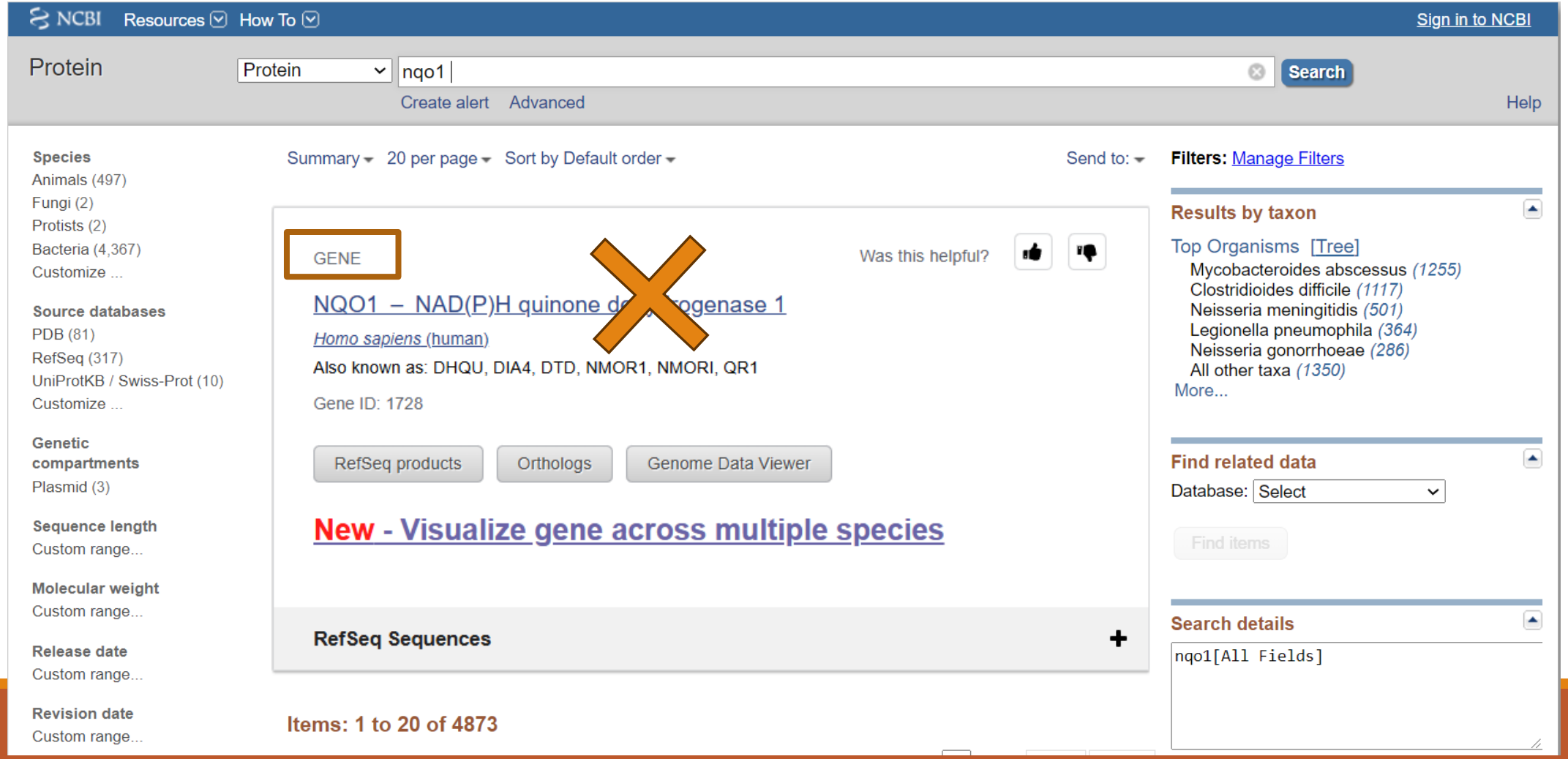
Proteinové databáze: NCBI



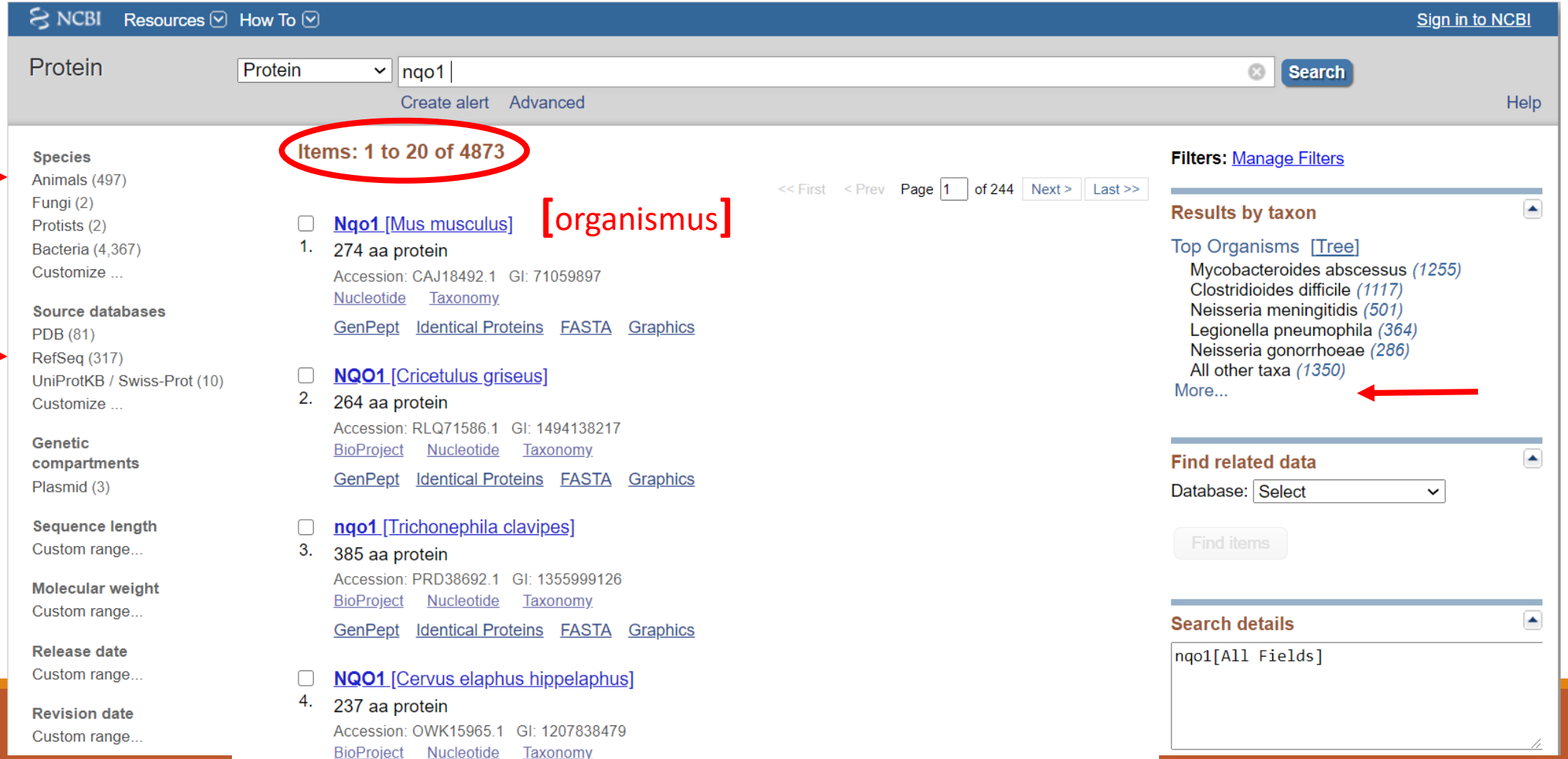
Proteinové databáze: NCBI



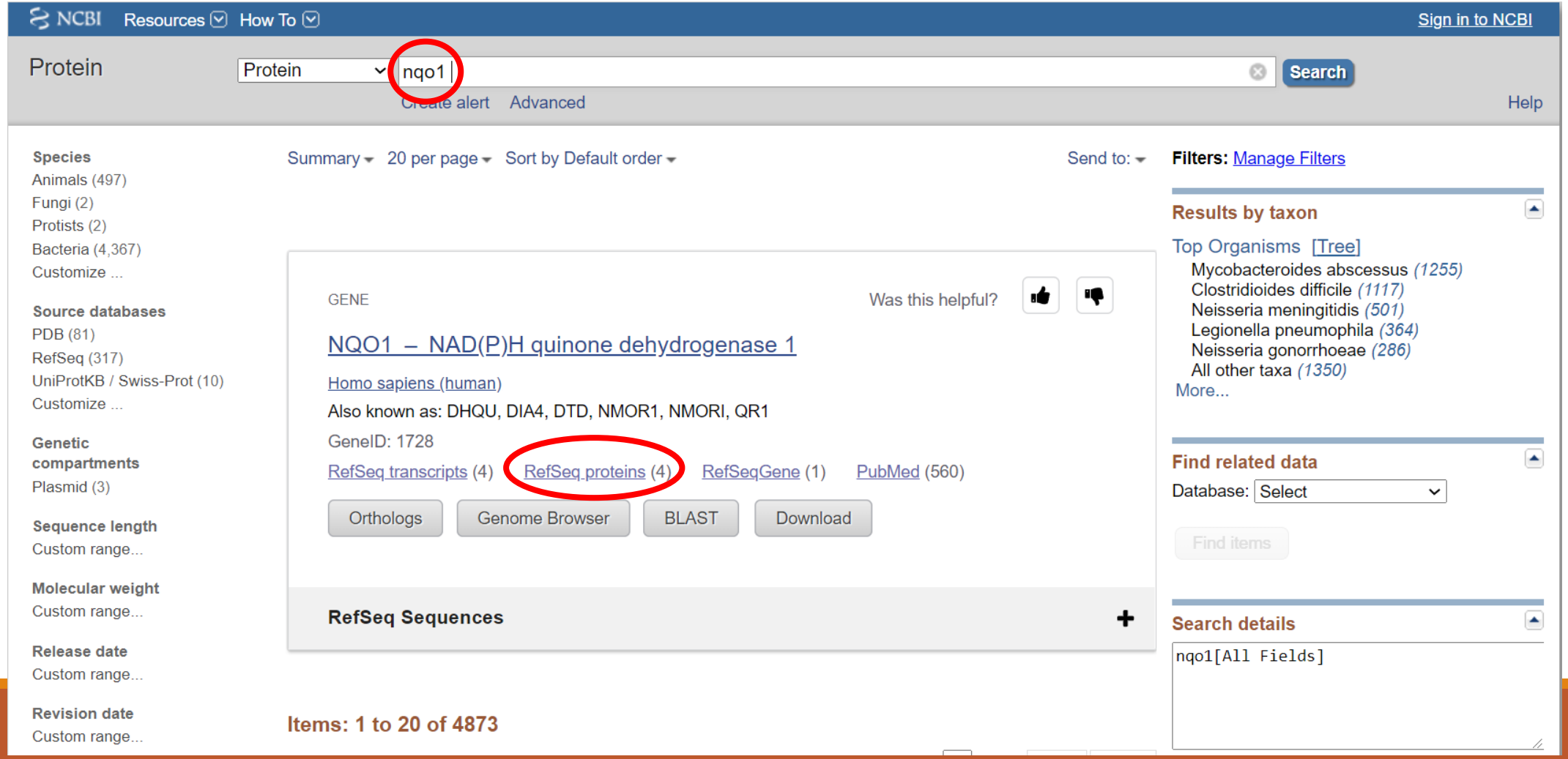
Proteinové databáze: NCBI



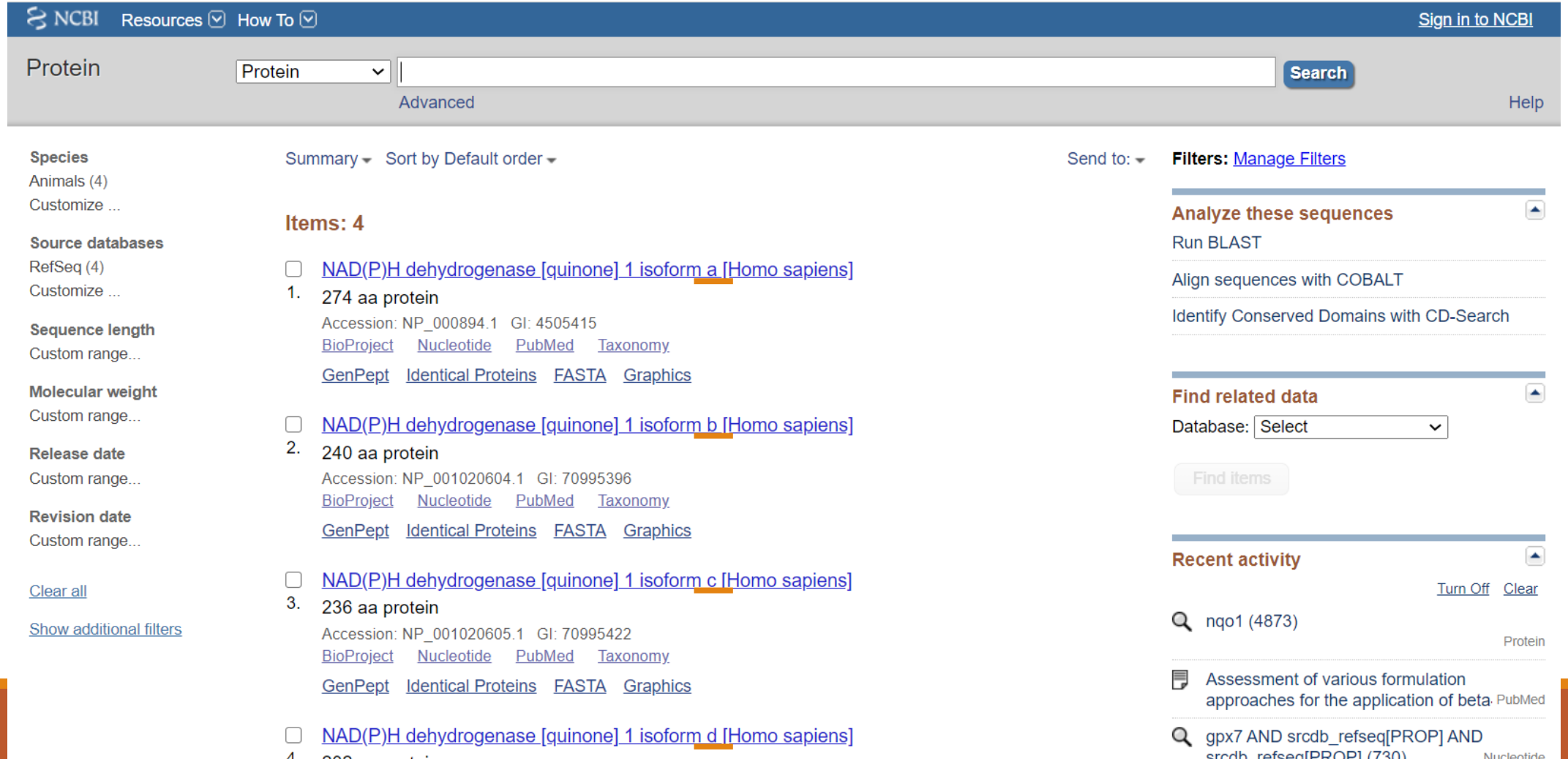
Proteinové databáze: NCBI



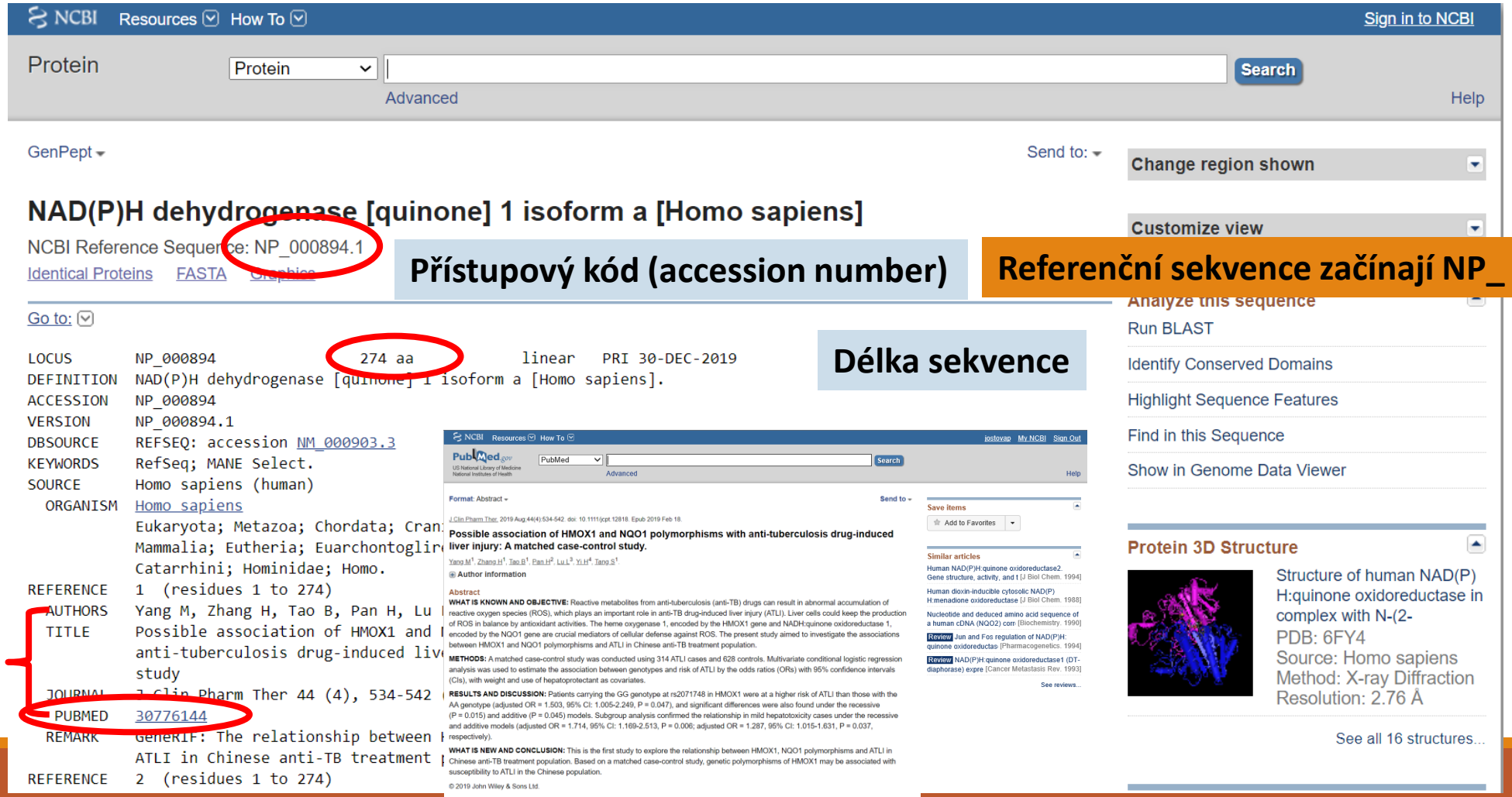
Proteinové databáze: NCBI



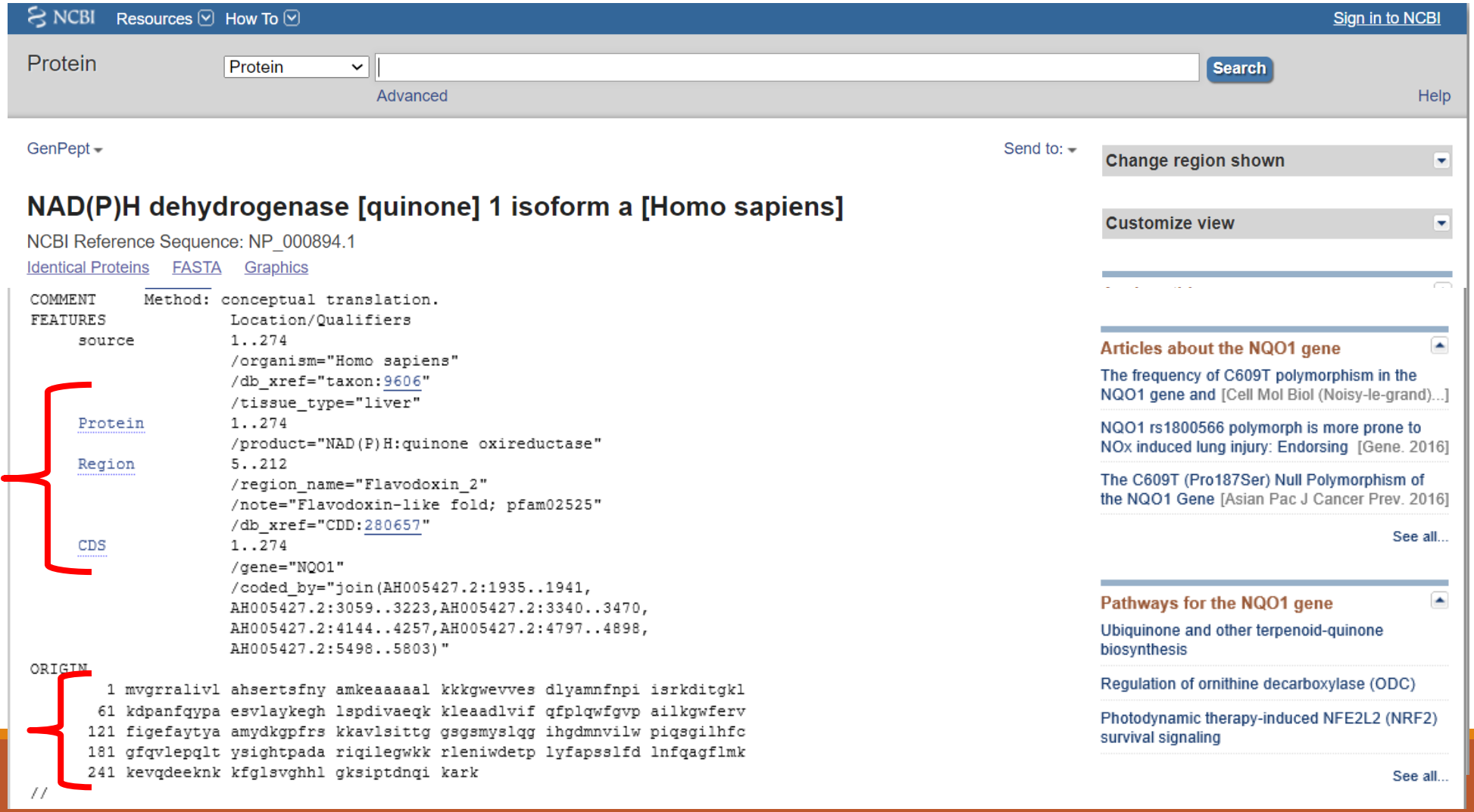
Proteinové databáze: NCBI



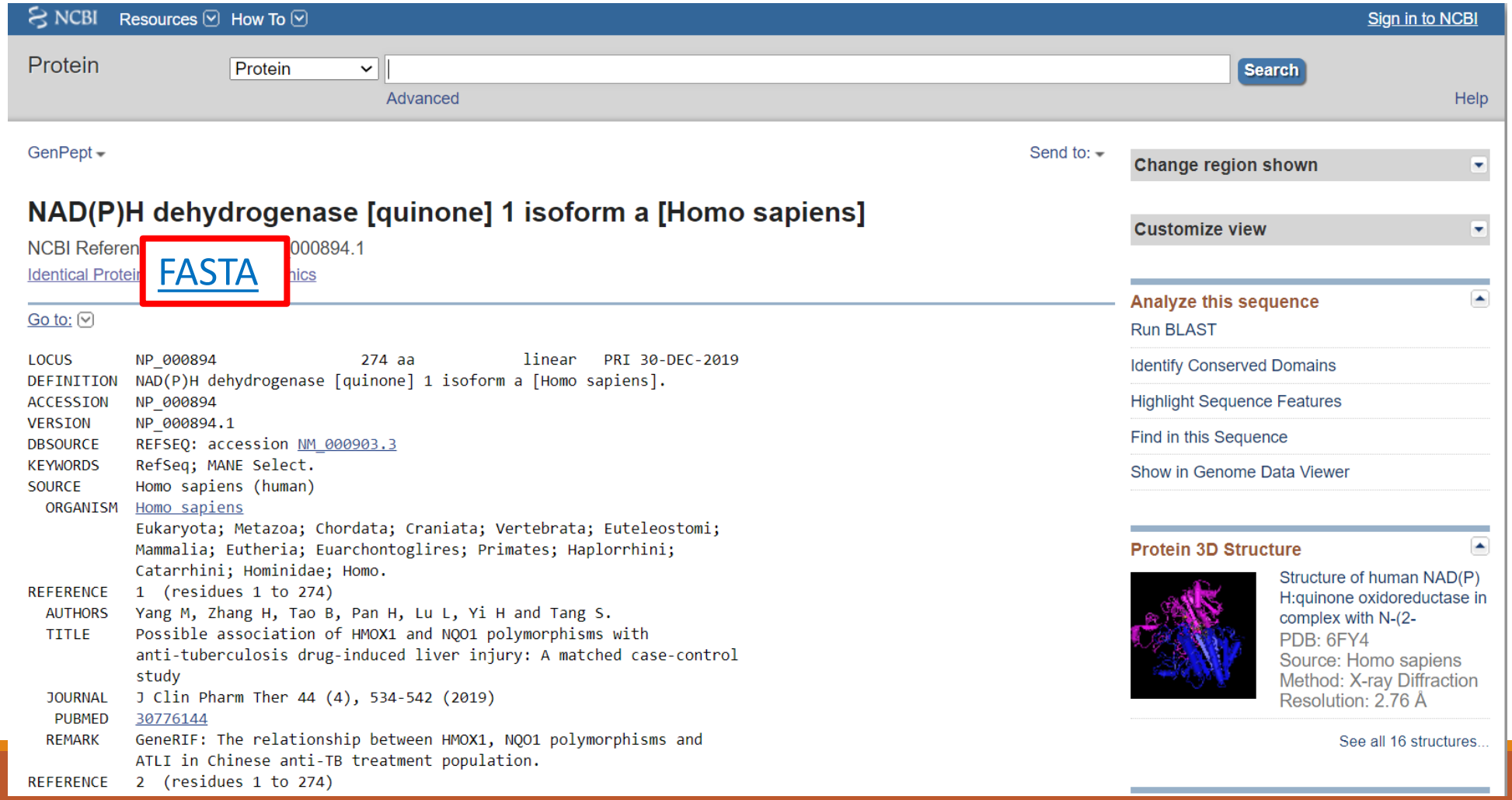
Proteinové databáze: NCBI



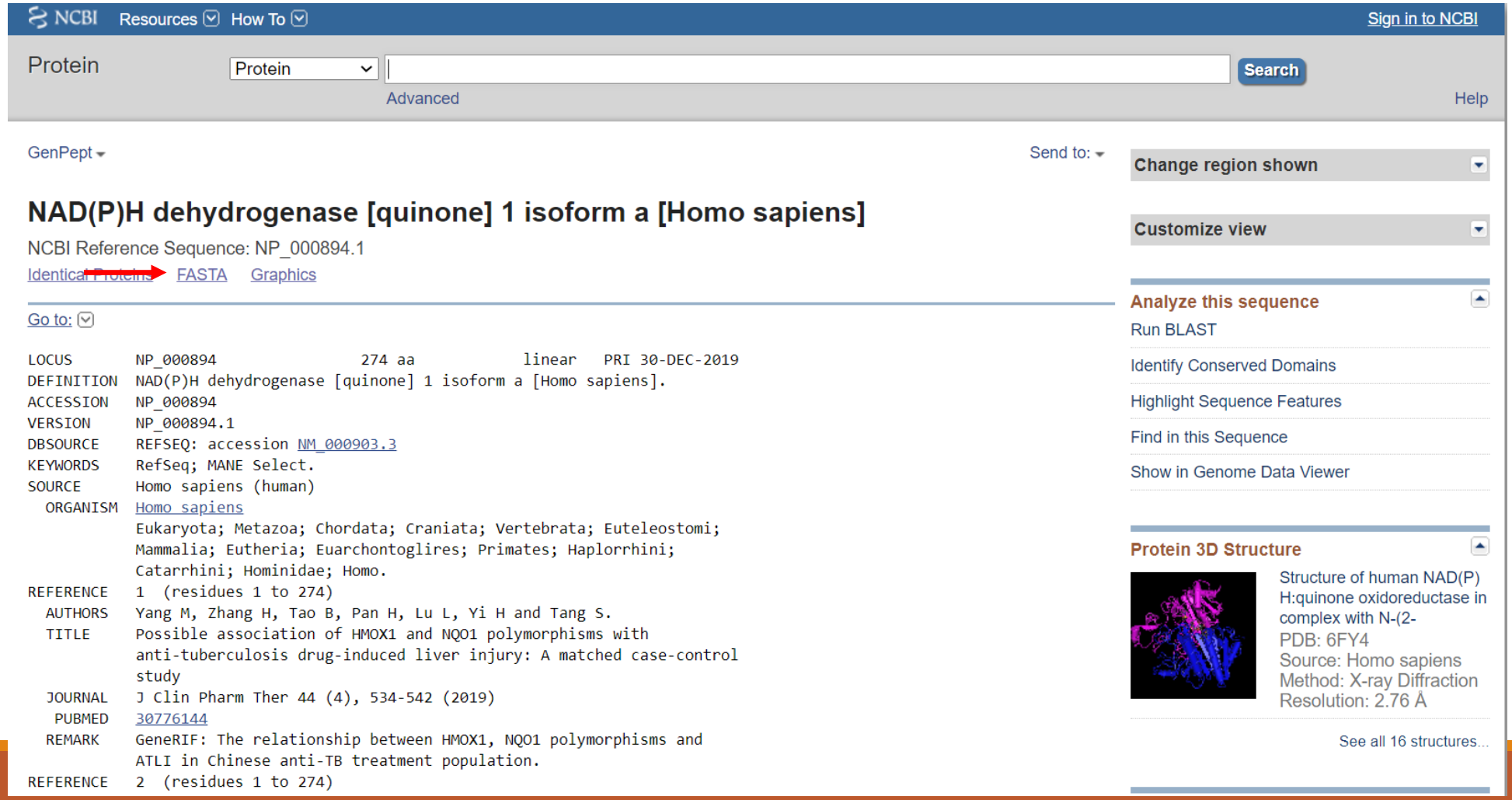
Proteinové databáze: NCBI



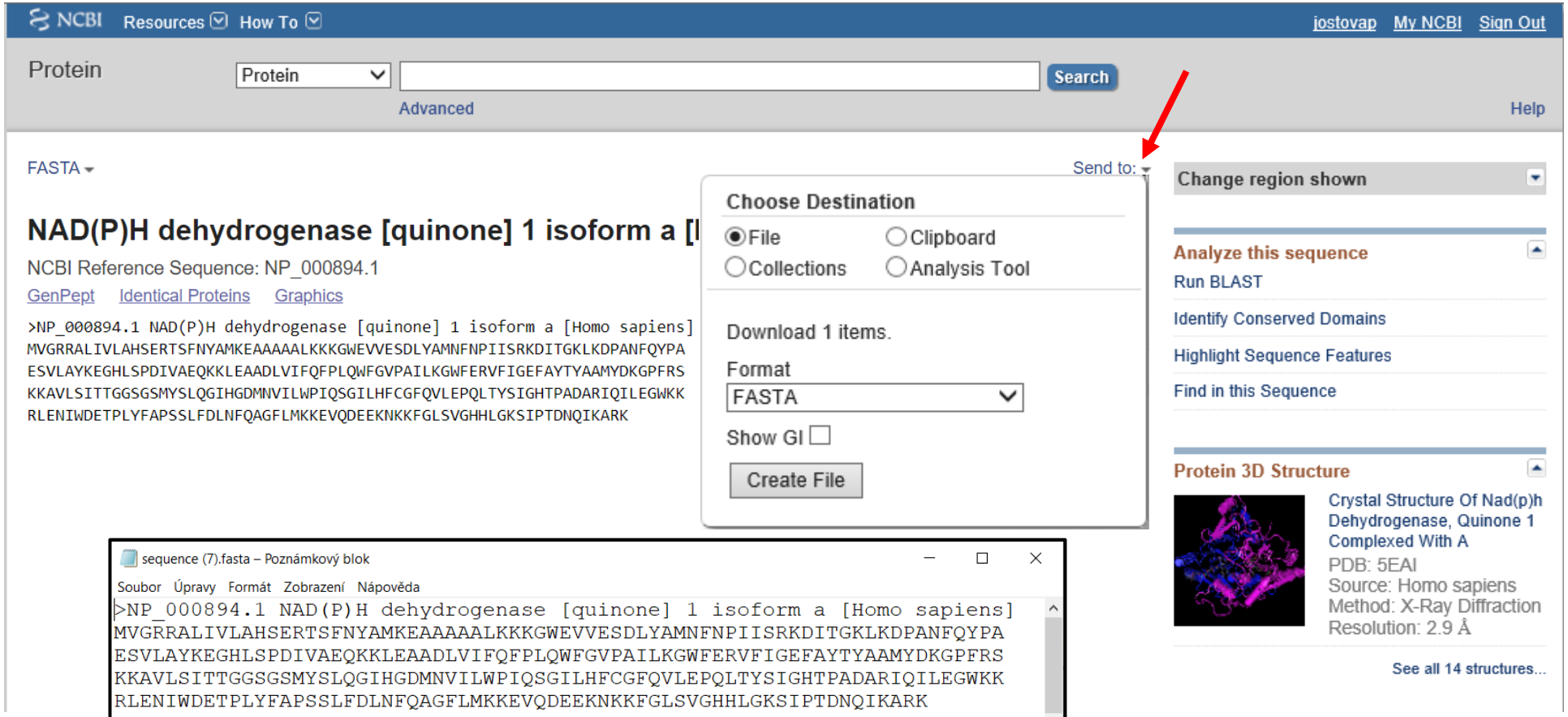
Proteinové databáze: NCBI



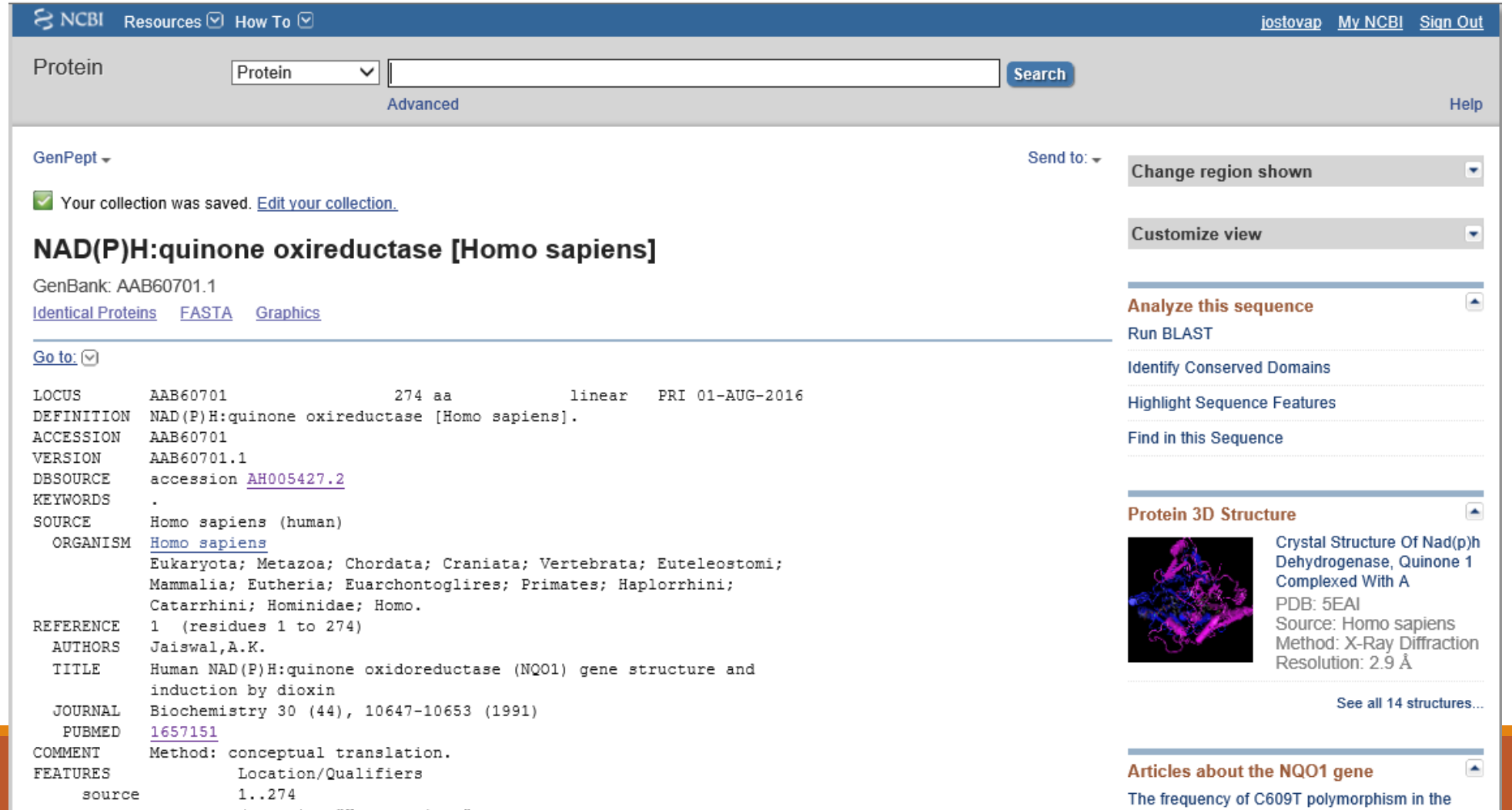
Proteinové databáze: NCBI



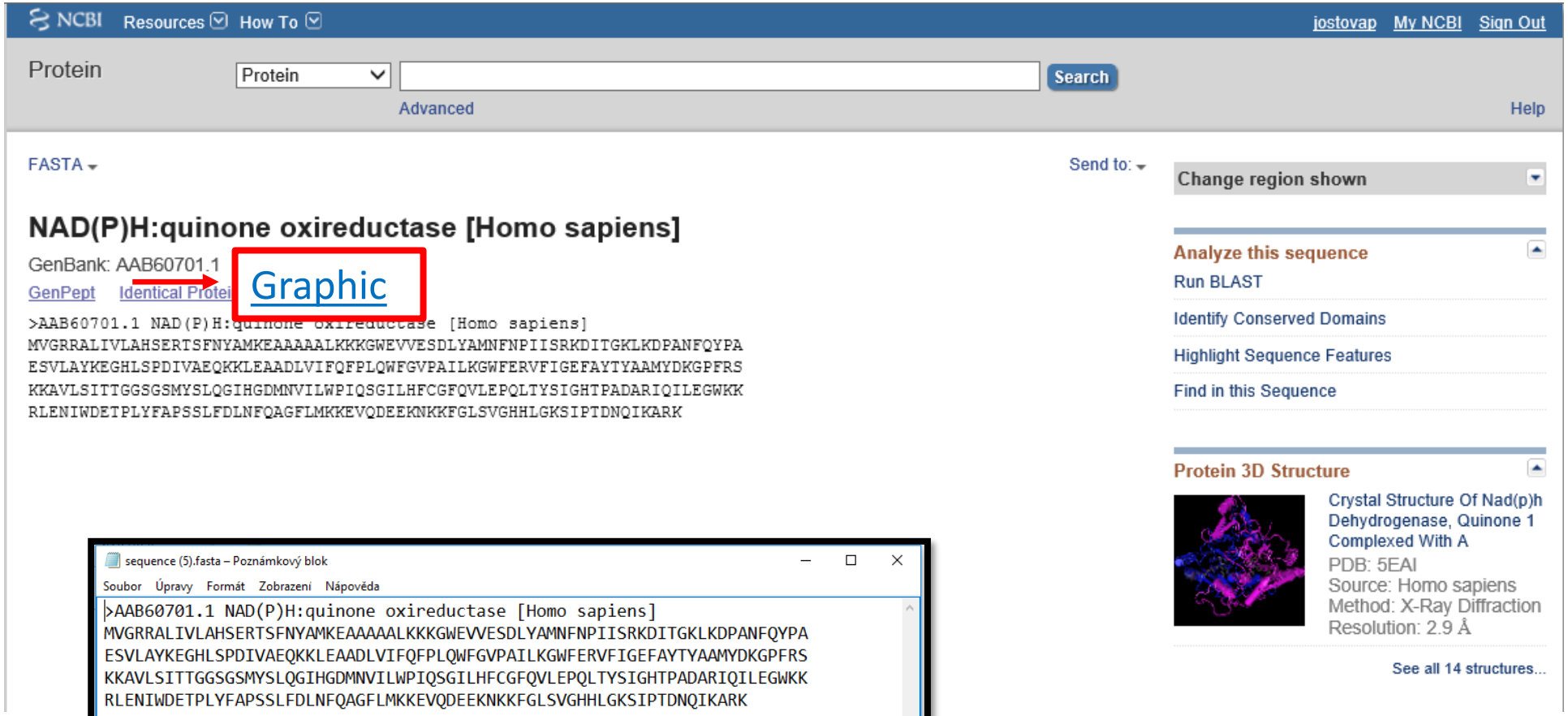
Proteinové databáze: NCBI



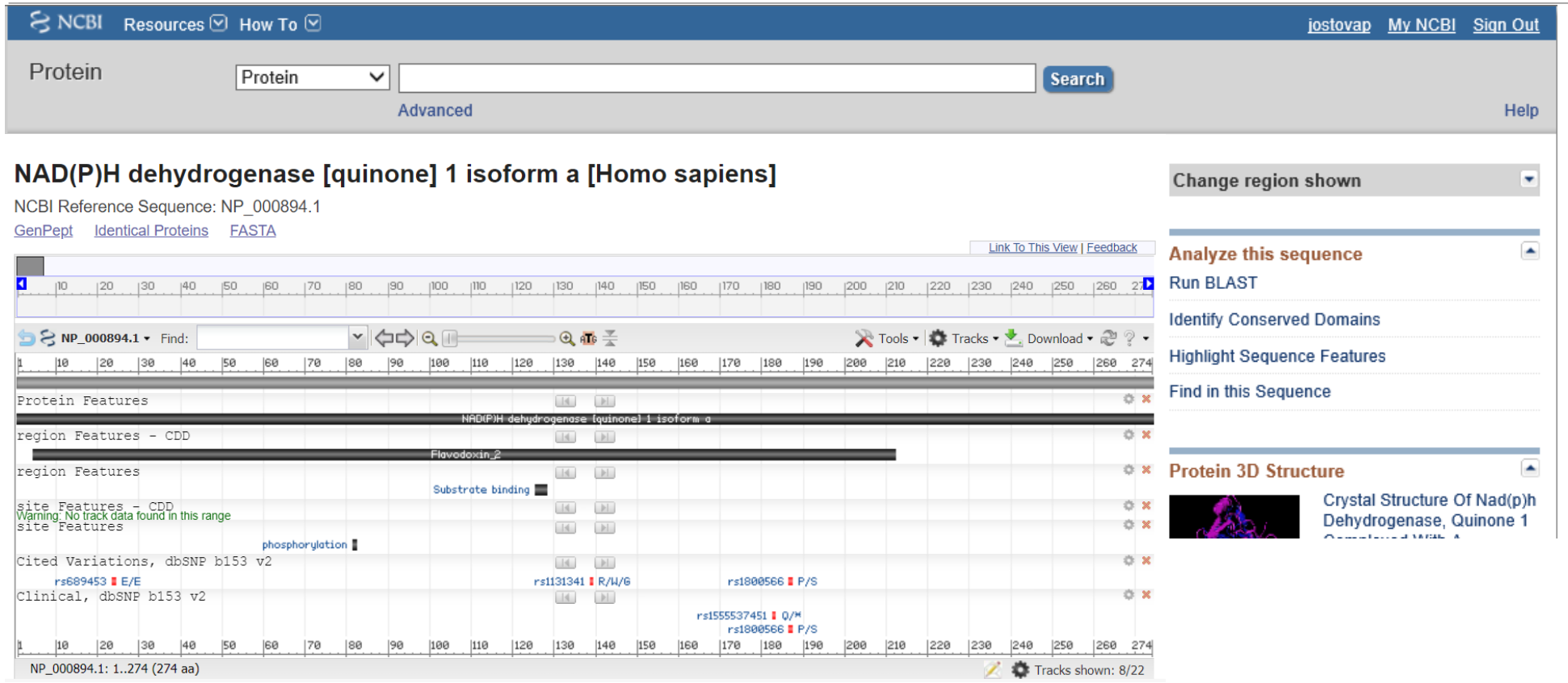
Proteinové databáze: NCBI → **My NCBI**



Proteinové databáze: NCBI



Proteinové databáze: NCBI



Proteinové databáze: UNIPROT →NCBI

Sequence & Isoforms - posunout se níže na:

Sequence databases

CCDS	CCDS10883.1 [P15559-1] CCDS32471.1 [P15559-3] CCDS32472.1 [P15559-2]
PIR	A41135 A30879
RefSeq	NP_000894.1 NM_000903.2 [P15559-1] NP_001020604.1 NM_001025433.1 [P15559-2] NP_001020605.1 NM_001025434.1 [P15559-3]

SEQUENCE	PROTEIN	MOLECULE TYPE	STATUS
J03934 (EMBL | GenBank | DDBJ)	AAA59940.1 (EMBL | GenBank | DDBJ)	mRNA	
M81600 (EMBL | GenBank | DDBJ)	AAB60701.1 (EMBL | GenBank | DDBJ)	Genomic DNA	
M81596 (EMBL | GenBank | DDBJ)	AAB60701.1 (EMBL | GenBank | DDBJ)	Genomic DNA	
M81597 (EMBL | GenBank | DDBJ)	AAB60701.1 (EMBL | GenBank | DDBJ)	Genomic DNA	
M81598 (EMBL | GenBank | DDBJ)	AAB60701.1 (EMBL | GenBank | DDBJ)	Genomic DNA	
M81599 (EMBL | GenBank | DDBJ)	AAB60701.1 (EMBL | GenBank | DDBJ)	Genomic DNA	
AY281093 (EMBL | GenBank | DDBJ)	AAP20940.1 (EMBL | GenBank | DDBJ)	Genomic DNA	Sequence problems.

Vyzkoušejte si NCBI proteinovou databázi

Najděte si „svou vlastní“ sekvenci

-vyzkoušejte hledat „přímo“ v NCBI
-vyzkoušejte „projít“ skrz „Sequence databases“ a „Refseq“ v Uniprot databázi

čas ?

„Proteinová bioinformatika“

Vyhledávání AMK sekvencí

Analýza vlastností sekvencí (aminokyselinové složení, molekulová hmotnost, isoelektrický bod…)

Štěpení proteasami

Analýza hydrofobních segmentů, transmembránových úseků

Predikce sekundární struktury

3D-struktura, vizualizace

Vyhledání a porovnání podobných sekvencí

Evoluční příbuznost sekvencí

…

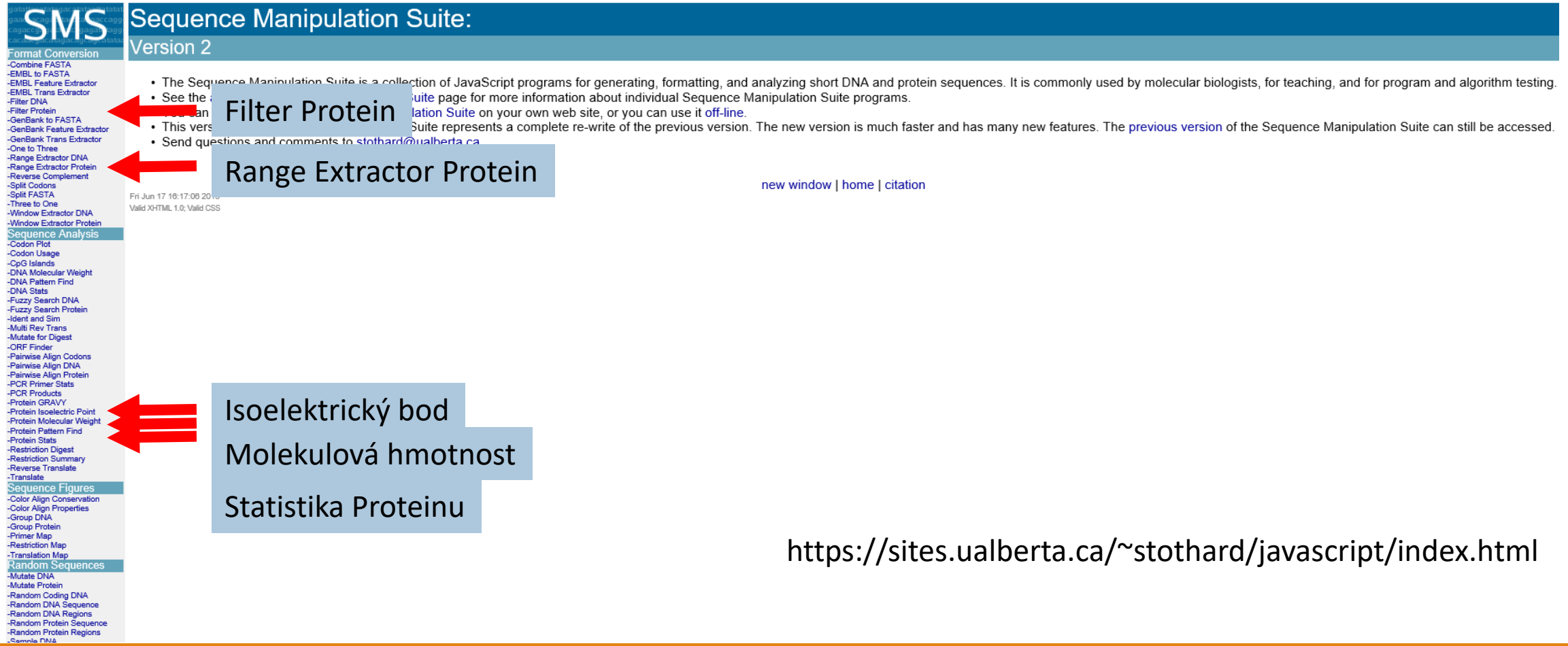
Analýza vlastností proteinu/sekvence

Molekulární hmotnost

Isoelektrický bod

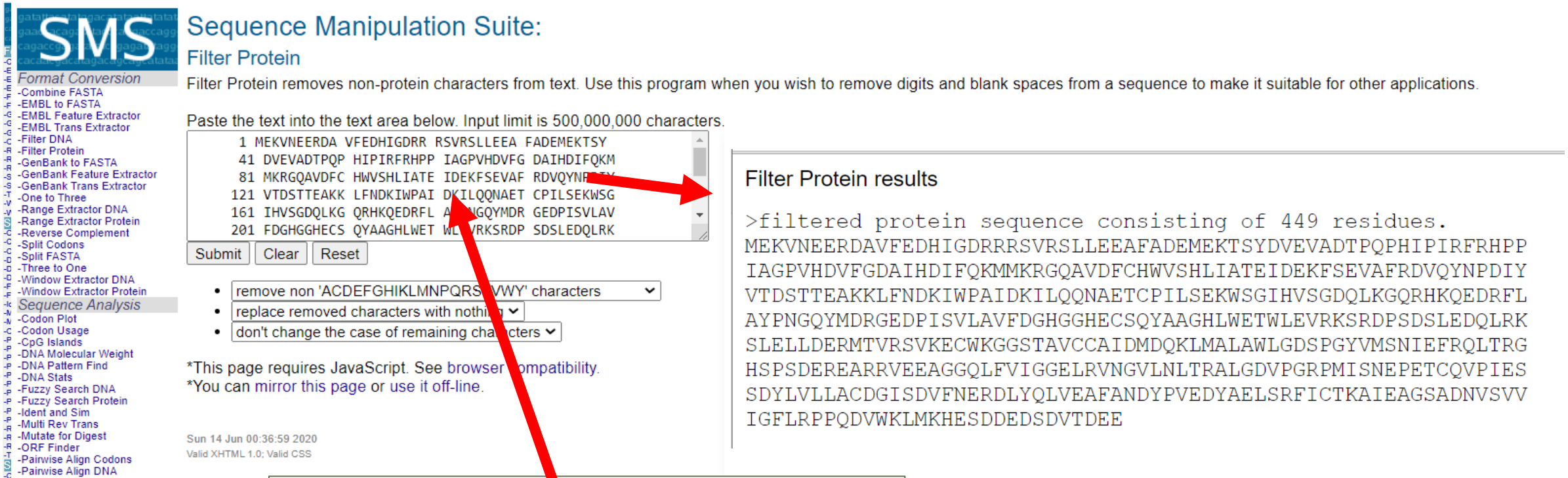
Složení AMK

Analýza vlastností proteinu/sekvence



https://sites.ualberta.ca/~stothard/javascript/index.html

Analýza vlastností proteinu/sekvence



možná testová otázka:

Přepište tento neznámý protein do fasta formátu:

1 MEKVNEERDA VFEDHIGDRR RSVRSLLEEA FADEMEKTSY
41 DVEVADTPQP HIPIRFRHPP IAGPVHDVFG DAIHDIFQKM
81 MKRGQAVDFC HWVSHLIATE IDEKFSEVAF RDVQYNPDIY
121 VTDSTTEAKK LFNDKIWPAI DKILQQNAET CPILSEKWSG
161 IHVSGDQLKG QRHKQEDRFL AYPNGQYMDR GEDPISVLAV

Analýza vlastností proteinu/sekvence

SMS

Format Conversion
-Combine FASTA
-EMBL to FASTA
-EMBL Feature Extractor
-EMBL Trans Extractor
-Filter DNA
-Filter Protein
-GenBank to FASTA
-GenBank Feature Extractor
-GenBank Trans Extractor
-One to Three
-Range Extractor DNA
-Range Extractor Protein
-Reverse Complement
-Split Codons
-Split FASTA
-Three to One
-Window Extractor DNA
-Window Extractor Protein

Sequence Analysis
-Codon Plot
-Codon Usage
-CpG Islands
-DNA Molecular Weight
-DNA Pattern Find
-DNA Stats
-Fuzzy Search DNA
-Fuzzy Search Protein
-Ident and Sim
-Multi Rev Trans
-Mutate for Digest
-ORF Finder
-Pairwise Align Codons
-Pairwise Align DNA
-Pairwise Align Protein
-PCR Primer Stats
-PCR Products
-Protein GRAVY
-Protein Isoelectric Point
-Protein Molecular Weight
-Protein Pattern Find
-Protein Stats
-Restriction Digest
-Restriction Summary
-Reverse Translate
-Translate

Sequence Figures
-Color Align Conservation
-Color Align Properties
-Group DNA
-Group Protein
-Primer Map
-Restriction Map
-Translation Map

Random Sequences
-Mutate DNA
-Mutate Protein
-Random Coding DNA
-Random DNA Sequence
-Random DNA Regions
-Random Protein Sequence
-Random Protein Regions
-Sample DNA
-Sample Protein
-Shuffle DNA
-Shuffle Protein

Miscellaneous

Sequence Manipulation Suite:

Protein Stats

Protein Stats returns the number of occurrences of each residue in the sequence you enter. Percentage totals are also given for each residue, and for certain groups of residues, allowing you to quickly compare the results obtained for different sequences.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 500000 characters.

```
ESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAA
MYDKGPFRS
KKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADAR
IQILEGWKK
RLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTDNQIK
ARK
```

Please check the browser compatibility page before using this program.

Submit Clear Reset

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

Fri Jun 17 16:17:06 2016
Valid XHTML 1.0; Valid CSS

Sequence Manipulation Suite - Internet Expl...
about:blank
Soubor Úpravy Zobrazit Oblíbené položky Nástroje Nápověda

Protein Stats results

Results for 274 residue sequence "AAB60701.1 NAD(P)H:quinone oxireductase [Homo sapiens]" starting "MVGRRALIVL"

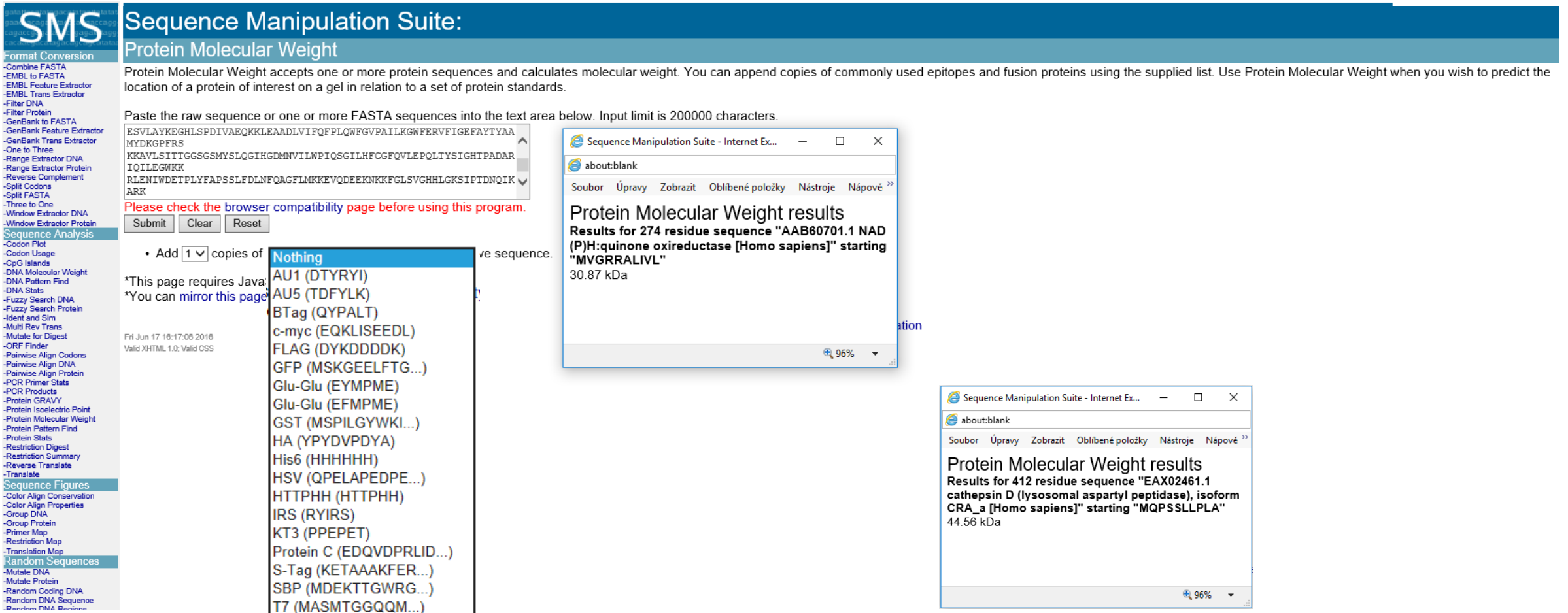
Pattern:	Times found:	Percentage:
A	25	9.12
B	0	0.00
C	1	0.36
D	12	4.38
E	17	6.20
F	17	6.20
G	21	7.66
H	7	2.55
I	19	6.93
K	24	8.76
L	25	9.12
M	7	2.55
N	9	3.28
P	13	4.74
Q	12	4.38
R	9	3.28
S	17	6.20
T	9	3.28

96%

možná testová otázka:

a Identifikujte počet cysteinů

Analýza vlastností proteinu/sekvence



Analýza vlastností proteinu/sekvence

SMS

Format Conversion
-Combine FASTA
-EMBL to FASTA
-EMBL Feature Extractor
-EMBL Trans Extractor
-Filter DNA
-Filter Protein
-GenBank to FASTA
-GenBank Feature Extractor
-GenBank Trans Extractor
-One to Three
-Range Extractor DNA
-Range Extractor Protein
-Reverse Complement
-Split Codons
-Split FASTA
-Three to One
-Window Extractor DNA
-Window Extractor Protein
Sequence Analysis
-Codon Plot
-Codon Usage
-CpG Islands
-DNA Molecular Weight
-DNA Pattern Find
-DNA Stats
-Fuzzy Search DNA
-Fuzzy Search Protein
-Ident and Sim
-Multi Rev Trans
-Mutate for Digest
-ORF Finder
-Pairwise Align Codons
-Pairwise Align DNA
-Pairwise Align Protein
-PCR Primer Stats
-PCR Products
-Protein GRAVY
-Protein Isoelectric Point
-Protein Molecular Weight
-Protein Pattern Find
-Protein Stats
-Restriction Digest
-Restriction Summary
-Reverse Translate
-Translate
Sequence Figures

Sequence Manipulation Suite:

Range Extractor Protein

Range Extractor Protein accepts a protein sequence along with a set of positions or ranges. The residues corresponding to the positions or ranges are returned, either as a single new sequence, a set of FASTA records, as uppercase text, or as lowercase text. Use Range Extractor Protein to obtain subsequences using position information.

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 500,000,000 characters.

```
ESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAY
TYAAMYDKGPFRS
KKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTP
ADARIQILEGWKK
RLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTD
NQIKARK
```

POZOR: Při vkládání sekvence ve FASTA formátu je lepší za názvem vložit „enter“. Někdy nerozpozná program nový řádek.

Enter the residue positions or ranges to be extracted. Use ".." to represent a range, and use a comma to separate entries. The words 'start', 'end', 'center', and 'length' can be used to represent the beginning, end, middle, and length of the sequence. Arithmetic expressions can be included in the ranges. For example, to obtain the last three residues, 'end - 2)..end' can be used. To obtain the 30 bases on either side of the center residue along with the center residue, the ranges '(center - 30)..(center + 30)' can be used.

10..50

Submit | Clear | Reset

- Sequence segments should be returned as a new sequence

Valid XHTML 1.0; Valid CSS

POZOR: Při vkládání sekvence a požadovaného rozsahu sekvence (..) je nutné zmáčknout „**submit**“ ne „enter“.

Sequence Manipulation Suite

about:blank

Range Extractor Protein results

```
>results for 274 residue sequence "NP_000894.1 NAD(P)H dehydrogenase [quinone]
LAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPI
```

Vyzkoušejte si SMS suite

Analyzujte si „svou vlastní“ sekvenci

- kolik má váš protein prolinů?
- identifikujte 100. aminokyselinu
- jaká bude molekulová hmotnost vašeho proteinu s histidinovou kotvou (6xHis)?

„Proteinová bioinformatika“

Vyhledávání AMK sekvencí

Analýza vlastností sekvencí (aminokyselinové složení, molekulová hmotnost, isoelektrický bod…)

Štěpení proteasami

Analýza hydrofobních segmentů, transmembránových úseků

Predikce sekundární struktury

3D-struktura, vizualizace

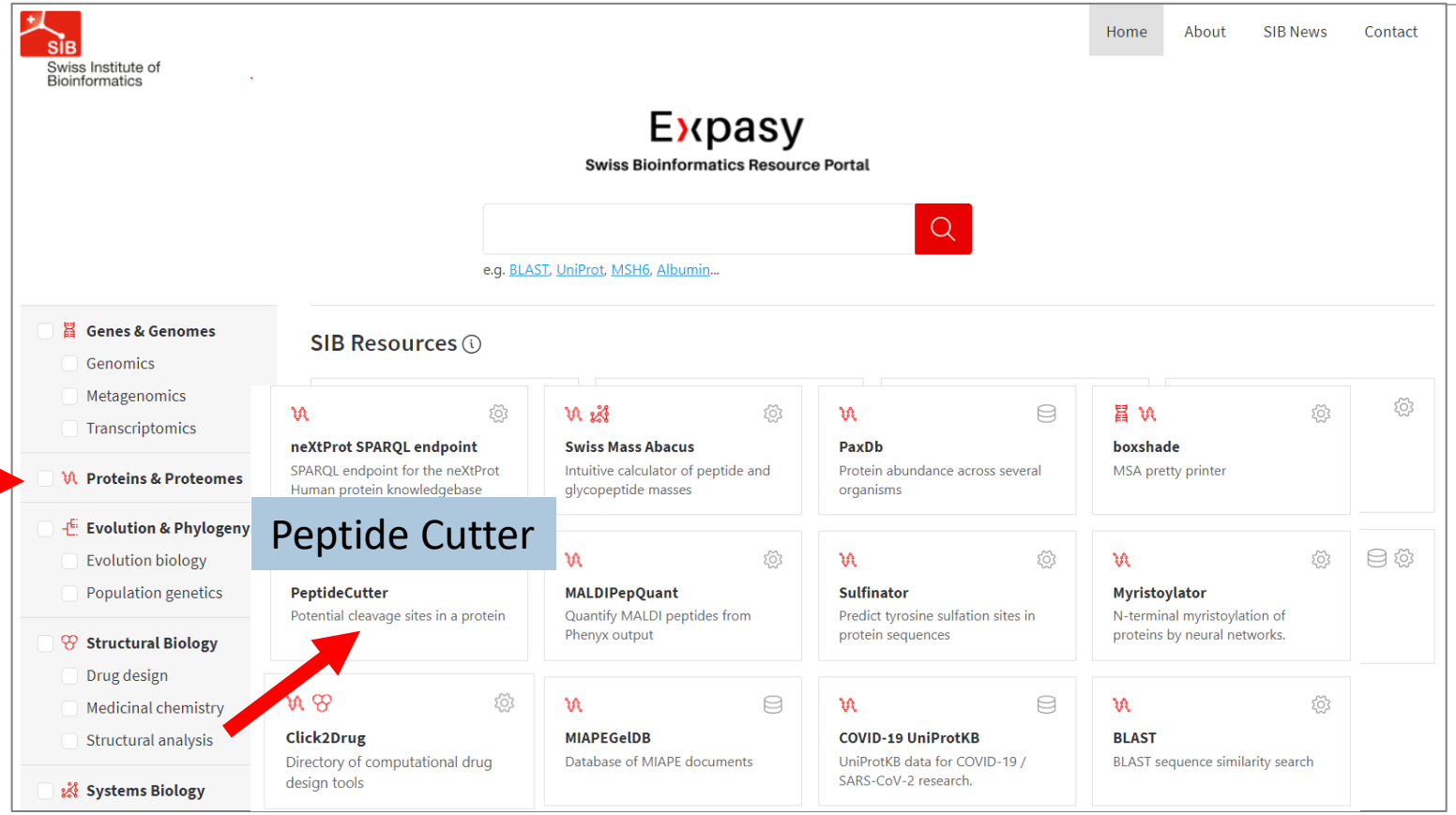
Vyhledání a porovnání podobných sekvencí

Evoluční příbuznost sekvencí

…

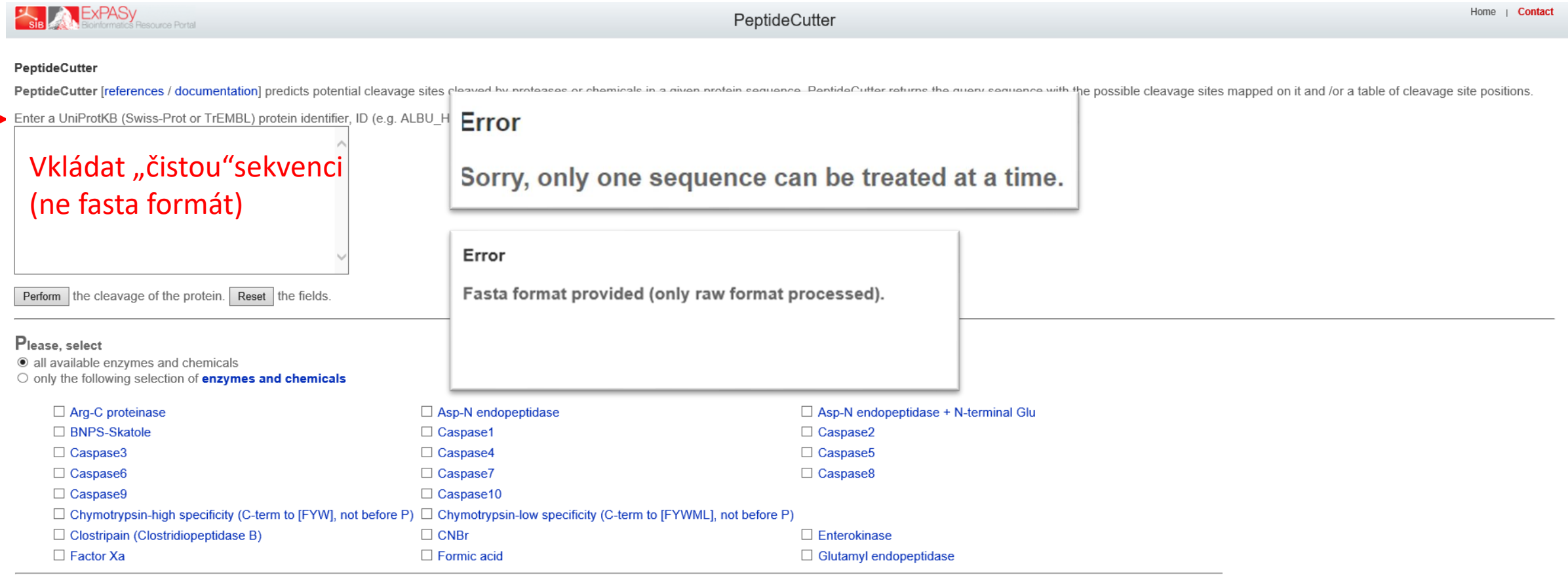
čas ?

Štěpení proteasami

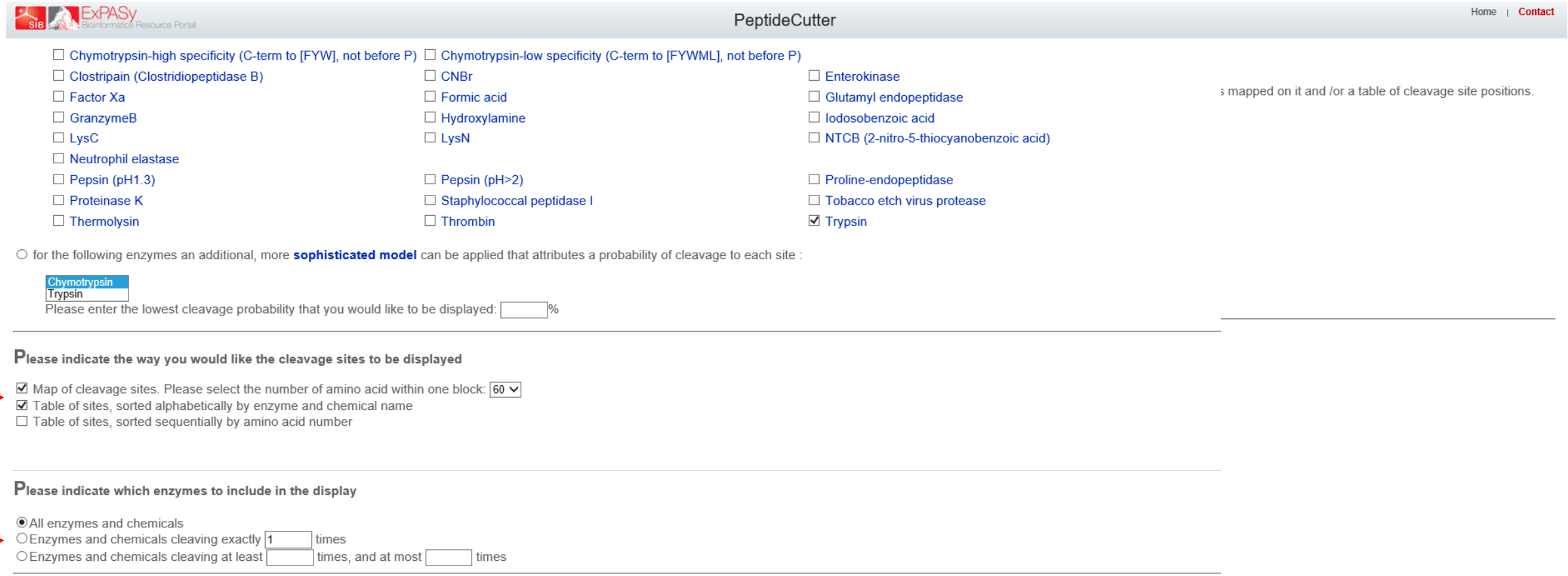


http://web.expasy.org/peptide_cutter/

Štěpení proteasami



Štěpení proteasami



Štěpení proteasami-PeptideCutter

Name of enzyme	No. of cleavages	Positions of cleavage sites
Arg-C proteinase	9	4 5 15 53 119 139 201 211 273
Asp-N endopeptidase	12	40 54 61 83 95 133 163 198 216 229 244 266
Asp-N endopeptidase + N-terminal Glu	29	13 23 35 38 40 54 61 70 77 83 87 92 95 117 123 133 163 185 198 205 212 216 217 229 241 244 245 246 266
BNPS-Skatole	6	35 106 116 170 208 216
CNBr	7	1 22 45 132 155 165 239
Chymotrypsin-high specificity (C-term to [FYW], not before P)	30	18 20 35 43 47 66 76 100 106 107 116 117 121 125 127 129 133 138 156 179 182 191 208 216 222 223 229 233 237 252
Chymotrypsin-low specificity (C-term to [FYWML], not before P)	67	1 7 10 12 18 20 22 30 35 42 43 45 47 60 66 74 76 80 81 92 97 100 104 106 107 113 116 117 121 125 127 129 133 138 145 156 158 162 165 169 177 178 179 182 185 189 191 195 205 208 212 216 221 222 223 228 229 231 233 237 238 239 252 254 258 259 260
Clostripain	9	4 5 15 53 119 139 201 211 273
Enterokinase	1	248
Formic acid	12	41 55 62 84 96 134 164 199 217 230 245 267
Glutamyl endopeptidase	17	14 24 36 39 71 78 88 93 118 124 186 206 213 218 242 246 247
Iodosobenzoic acid	6	35 106 116 170 208 216
LysC	24	23 31 32 33 54 59 61 77 90 91 114 135 141 142 209 210 240 241 248 250 251 262 271 274
LysN	24	22 30 31 32 53 58 60 76 89 90 113 134 140 141 208 209 239 240 247 249 250 261 270 273
NTCB (2-nitro-5-thiocyanobenzoic acid)	1	179
Pepsin (pH1.3)	59	9 10 18 29 30 41 42 46 59 60 65 66 73 74 80 91 96 97 99 100 102 103 106 107 112 113 117 120 124 125 145 157 158 168 176 177 178 179 181 182 184 189 204 205 220 222 227 228 229 230 231 232 233 236 237 238 251 254 259
Pepsin (pH>2)	82	9 10 18 19 20 29 30 41 42 43 46 59 60 65 66 68 73 74 75 76 80 91 96 97 99 100 102 103 105 106 107 112 113 115 117 120 124 125 126 127 128 129 132 133 145 155 156 157 158 168 170 176 177 178 179 181 182 184 189 190 191 204 205 207 208 215 216 220 222 227 228 229 230 231 232 233 236 237 238 251 254 259
Proteinase K	142	2 6 7 8 9 10 11 14 16 18 20 21 24 25 26 27 28 29 30 35 36 37 38 39 42 43 44 47 50 51 56 57 60 64 66 68 70 71 73 74 75 76 78 81 85 86 87 88 92 93 94 95 97 98 99 100 102 104 106 107 109 111 112 113 116 117 118 120 121 122 124 125 126 127 128 129 130 131 133 138 143 144 145 147 148 149 156 158 161 167 168 169 170 172 176 177 179 182 184 185 186 189 190 191 193 196 198 200 202 204 205 206 208 212 213 215 216 218 219 221 222 223 224 228 229 231 233 235 237 238 242 243 246 247 252 254 256 260 264 266 270 272
Staphylococcal peptidase I	16	14 24 36 39 71 78 88 93 118 124 186 206 213 218 242 246
Thermolysin	90	1 5 6 7 8 9 10 17 20 21 25 26 27 28 29 37 43 44 46 49 50 59 63 65 69 72 73 74 80 85 86 91 94 97 98 99 103 106 110 111 112 116 119 120 121 125 129 130 131 137 142 143 144 146 154 157 160 166 167 168 171 175 176 178 181 183 184 188 192 197 201 203 204 211 214 220 222 227 228 232 234 236 237 238 251 253 255 259 269 271
Trypsin	33	4 5 15 23 31 32 33 53 54 59 61 77 90 91 114 119 135 139 141 142 201 209 210 211 240 241 248 250 251 262 271 273 274

These chosen enzymes do not cut:

Caspase1
Caspase10

„defaultní“ nastavení

Štěpení proteasami-PeptideCutter

Name of enzyme	No. of cleavages	Positions of cleavage sites
Trypsin	33	4 5 15 23 31 32 33 53 54 59 61 77 90 91 114 119 135 139 141 142 201 209 210 211 240 241 248 250 251 262 271 273 274

Konkrétní enzym

[*] *NOTE: Proline-endopeptidase was reported to cleave only substrates whose sequences do not exceed 30 amino acids. An unusual beta-propeller domain regulates proteolysis: see Fulop et al., 1998.*
You have chosen to display only those enzymes that cleave exactly 1 times. However, the following enzymes also cleave but not with the selected frequency:
Staphylococcal peptidase I , Pepsin (pH1.3) , Glutamyl endopeptidase , CNBr , Pepsin (pH>2) , Asp-N endopeptidase , Asp-N endopeptidase + N-terminal Glu , Formic acid , Iodosobenzoic acid , Arg-C proteinase , Thermolysin , Trypsin , Clostripain , Proteinase K , Chymotrypsin-high specificity (C-term to [FYW], not before P) , Chymotrypsin-low specificity (C-term to [FYWML], not before P) , LysC , BNPS-Skatole , LysN ,

These enzymes cleave the sequence:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Enterokinase	1	248
NTCB (2-nitro-5-thiocyanobenzoic acid)	1	179

Co štěpí přesně jednou?

Štěpení proteasami-PeptideCutter

Name of enzyme	No. of cleavages	Positions of cleavage sites
Arg-C proteinase	9	4 5 15 53 119 139 201 211 273
Asp-N endopeptidase	12	40 54 61 83 95 133 163 198 216 229 244 266
Asp-N endopeptidase + N-terminal Glu	29	13 23 35 38 40 54 61 70 77 83 87 92 95 117 123 133 163 185 198 205 212 216 217 229 241 244 245 246 266
BNPS-Skatole	6	35 106 116 170 208 216
CNBr	7	1 22 45 132 155 165 239
Chymotrypsin-high specificity (C-term to [FYW], not before P)	30	18 20 35 43 47 66 76 100 106 107 116 117 121 125 127 129 133 138 156 179 182 191 208 216 222 223 229 233 237 252
Chymotrypsin-low specificity (C-term to [FYWML],	67	1 7 10 12 18 20 22 30 35 42 43 45 47 60 66 74 76 80 81 92 97 100 104 106 107 113 116 117 121 125 127 129 133 138 145 156 158 162 165 169 177 178 179 182 185 189 191 195 205 208 212 216 221 222 223 228 229
Clostripain		
Enterokinase		
Formic acid		
Glutamyl endopeptid		
Iodosobenzoic acid		
LysC		
LysN		
NTCB (2-nitro-5-thio acid)		
Pepsin (pH1.3)		
Pepsin (pH>2)		
Proteinase K		
Staphylococcal pepti		
Thermolysin		
Trypsin		

These chosen enzymes do not

Name of enzyme	No. of cleavages	Positions of cleavage sites
Trypsin	33	4 5 15 23 31 32 33 53 54 59 61 77 90 91 114 119 135 139 141 142 201 209 210 211 240 241 248 250 251 262 271 273 274

At these positions the following enzymes cleave:

- Please note that the size of the peptides are calculated as if **all chosen enzymes were present** during digestion. If you want to obtain the size of the peptides resulting from the cleavage of only one enzyme, please, deselect the others
- Please be aware of the fact th of the occurring amino acid re

[*] *NOTE: Proline-endopeptidase was reported to cleave only substrates whose sequences do not exceed 30 amino acids. An unusual beta-propeller domain regulates proteolysis: see Fulop et al., 1998.*

You have chosen to display only those enzymes that cleave exactly 1 times. However, the following enzymes also cleave but not with the selected frequency:
Staphylococcal peptidase I , Pepsin (pH1.3) , Glutamyl endopeptidase , CNBr , Pepsin (pH>2) , Asp-N endopeptidase , Asp-N endopeptidase + N-terminal Glu , Formic acid , Iodosobenzoic acid , Arg-C proteinase , Thermolysin , Try
Chymotrypsin-high specificity (C-term to [FYW], not before P) , Chymotrypsin-low specificity (C-term to [FYWML], not before P) , LysC , BNPS-Skatole , LysN ,

Position of cleavage site	Name of cl
4	Trypsin
5	Trypsin
15	Trypsin
23	Trypsin
31	Trypsin
32	Trypsin
33	Trypsin
53	Trypsin
54	Trypsin
59	Trypsin
61	Trypsin
77	Trypsin
90	Trypsin
91	Trypsin
114	Trypsin
119	Trypsin
135	Trypsin
139	Trypsin

These enzymes cleave the sequence:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Enterokinase	1	248
NTCB (2-nitro-5-thiocyanobenzoic acid)	1	179

At these positions the following enzymes cleave:

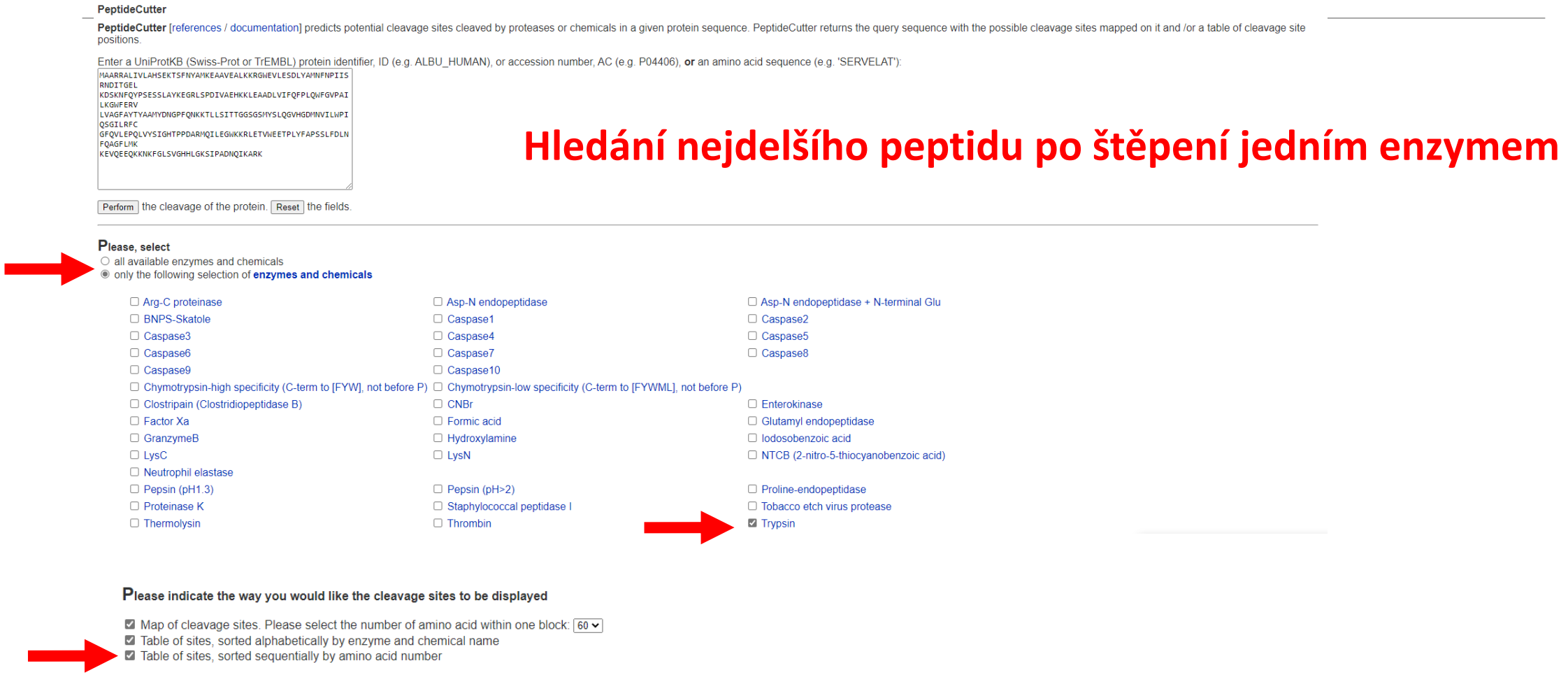
- Please note that the size of the peptides are calculated as if **all chosen enzymes were present** during digestion. If you want to obtain the size of the peptides resulting from the cleavage of only one enzyme, please, deselect th
- Please be aware of the fact that the present version of the PeptideCutter program does not take into consideration any kind of **modification** neither of the protein sequence nor of modifications evoked by the cleavage. Mass co of the occurring amino acid residues, and giving peptide masses as [M]. If you want to select different parameters, we recommend to use PeptideMass.

Position of cleavage site	Name of cleaving enzyme(s)	Resulting peptide sequence (see explanations)
179	NTCB (2-nitro-5-thiocyanobenzoic acid)	MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKLKDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRSKKAVLSITTGGSGSMYSLQGIHGDM
248	Enterokinase	CGFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFL MKKEVQDEEK
274	**end of sequence**	NKKFGLSVGHHLGKSIPTDNQIKARK

These are the cleavage sites of the chosen enzymes and chemicals mapped onto the entered protein sequence:

- You have chosen a block size of **60** for the map

Štěpení proteasami-PeptideCutter



Hledání nejdelšího peptidu po štěpení jedním enzymem

ptideCutter

Name of enzyme	No. of cleavages	Positions of cleavage sites
Trypsin	32	4 5 15 23 31 32 33 53 61 64 77 80 90 91 114 119 141 142 178 201 209 210 211 240 241 248 249 251 262 271 273 274

At these positions the following enzymes cleave:

- Please note that the size of the peptides are calculated as if **all chosen enzymes were present** during digestion. If you want to obtain the size of the peptic
- Please be aware of the fact that the present version of the PeptideCutter program does not take into consideration any kind of **modification** neither of the p
amino acid residues, and giving peptide masses as [M]. If you want to select different parameters, we recommend to use PeptideMass.

y sequence with the possible cleavage sites mapped on it and /or a table of cleavage site

"):

Position of cleavage site	Name of cleaving enzyme(s)	Resulting peptide sequence (see explanations)	Peptide length [aa]	Peptide mass [Da]
4	Trypsin	MAAR	4	447.553
5	Trypsin	R	1	174.203
15	Trypsin	ALIVLAHSEK	10	1080.293
23	Trypsin	TSFNYAMK	8	[illegible]
31	Trypsin	EAAVEALK	8	829.949
32	Trypsin	K	1	146.189
33	Trypsin	R	1	174.203
53	Trypsin	GWEVLESDLYAMNFNPIISR	20	2354.662
61	Trypsin	NDITGELK	8	888.973
64	Trypsin	DSK	3	348.356
77	Trypsin	NFQYPSESSLAYK	13	1533.657
80	Trypsin	EGR	3	360.370
90	Trypsin	LSPDIVAEHK	10	1108.260
91	Trypsin	K	1	146.189
114	Trypsin	LEAADLVIFQFPLQWFGVPAILK	23	2616.141
119	Trypsin	GWFER	5	693.760
141	Trypsin	VLVAGFAYTYAAMYDNGPFQNK	22	2440.755
142	Trypsin	K	1	146.189
178	Trypsin	TLLSITTGGSGSMYSLQGVHGDMNVILWPIQSGILR	36	3803.406
201	Trypsin	FCGFQVLEPQLVYSIGHTPPDAR	23	2574.937
209	Trypsin	MQILEGWK	8	[illegible]
210	Trypsin	K	1	[illegible]
211	Trypsin	R	1	174.203
240	Trypsin	LETVWEETPLYFAPSSLFDLNFQAGFLMK	29	3394.886
241	Trypsin	K	1	146.189
248	Trypsin	EVQEEQK	7	888.930
249	Trypsin	K	1	146.189
251	Trypsin	NK	2	260.293
262	Trypsin	FGLSVGHHLGK	11	1151.333
271	Trypsin	SIPADNQIK	9	985.105
273	Trypsin	AR	2	245.282
274	**end of sequence**	K	1	146.189

rminal Glu

enzoic acid)

Hledání nejdelšího peptidu po štěpení

největší číslo

Vyzkoušejte si PeptideCutter

Analyzujte si „svou vlastní“ sekvenci

- Štěpí nějaký enzym pouze jedenkrát? (-vyzkoušejte štěpení všemi (default nastavení)
  (nebo: Vyberte pouze enzym, který štěpí jedenkrát.)
- Jak dlouhý je nejdelší vzniklý peptid po štěpení trypsinem? (omezte štěpení pouze na Trypsin, zaškrtněte si tabulku s „počtem aminokyselin“)

„Proteinová bioinformatika“

Vyhledávání AMK sekvencí

Analýza vlastností sekvencí (aminokyselinové složení, molekulová hmotnost, isoelektrický bod…)

Štěpení proteasami

Analýza hydrofobních segmentů, transmembránových úseků

Predikce sekundární struktury

3D-struktura, vizualizace

Vyhledání a porovnání podobných sekvencí

Evoluční příbuznost sekvencí

…

shrnutí I - Rešeršní projekt: NQO1

Vyhledejte dostupné informace o NQO1, získejte co nejvíce literárních, sekvenčních, případně i strukturních a dalších údajů o tomto genu/proteinu.

Nalezněte podobné sekvence a porovnejte je na proteinové úrovni.

- **NAD(P)H:chinonoxidoreduktasa / reference**
- **Protein: NP_000894, P15559 / 274 AMK; 30,8 kDa; trypsin štěpí 33x (59AMK nejdelší peptid)…**

DÚ2: Vyhledávání proteinů

Pracujte s „vaším“ genem/proteinem/enzymem (př. NQO1)

1) Vyhledejte „referenční“ NCBI a UniProt přístupové kódy

2) Najděte a uložte si „FASTA“ formát sekvence a uložte si ji (jako text-ne jako výstřižek)

3) Vyberte část proteinu od **10. do 50.** aminokyseliny

4) Spočítejte molekulovou hmotnost tohoto úseku proteinu

5) Zjistěte kolikrát bude vaše **celá** sekvence štěpena trypsinem

- vše zpracujte do bloku OneNote, Word, Pdf

Literární rešerše: DÚ2-příklad řešení

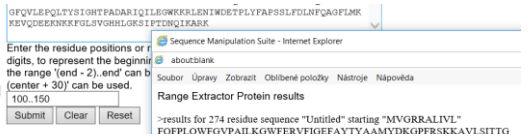
DÚ2

NAD(P)H:quinone oxireductase [Homo sapiens]

GenBank: AAB60701.1

UniProtKB - P15559 (NQO1_HUMAN)

>AAB60701.1 NAD(P)H:quinone oxireductase [Homo sapiens]
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKLKDPANFQYPA
ESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRS
KKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKK
RLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK



Protein Molecular Weight results
Results for 51 residue sequence "Untitled" starting "FQFPLQWFGV"
5.88 kDa

Trypsin 33

Your collection was saved. Edit your collection.

NAD(P)H:quinone oxireductase [Homo sapiens]

GenBank: AAB60701.1

Identical Proteins FASTA Graphics

Nebo stačí takhle:

NQO1:	AAB60701
	P15559
	>AAB60701.1 NAD(P)H:quinone oxireductase [Homo sapiens] MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKLKDPANFQYPA ESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRS KKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKK RLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
	FQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRSKKAVLSITTG 5.88 kDa
	Trypsin:33x
	Your collection was saved. Edit your collection.

NAD(P)H:quinone oxireductase [Homo sapiens]

GenBank: AAB60701.1

Identical Proteins FASTA Graphics