Introduction to applied bioinformatics

PETRA MATOUŠKOVÁ

2023/2024

2/10

„Protein bioinformatics I“

Retrieving protein sequences from databases (Uniprot: FASTA formate)

Computing amino-acids compositions, molecular weight, isoelectric point, and other parameters (SMS)

Prediction of proteases cutting (PeptideCutter)

Predicting elements of protein secondary structure, domains

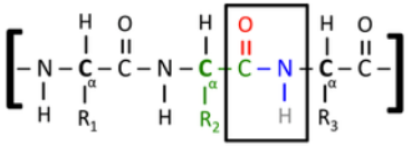
Predicting 3-D structure and the domain organization of proteins

Finding all proteins that share a similar sequence and Classifying proteins into families

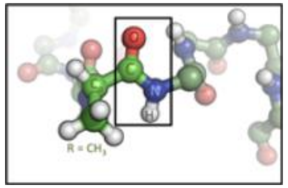
Finding evolutionary relationships between proteins, drawing proteins’ family trees

Computing the optimal alignment between two or more protein sequences

…



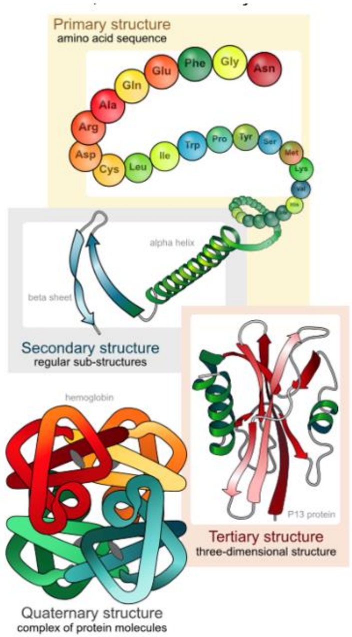
Proteins

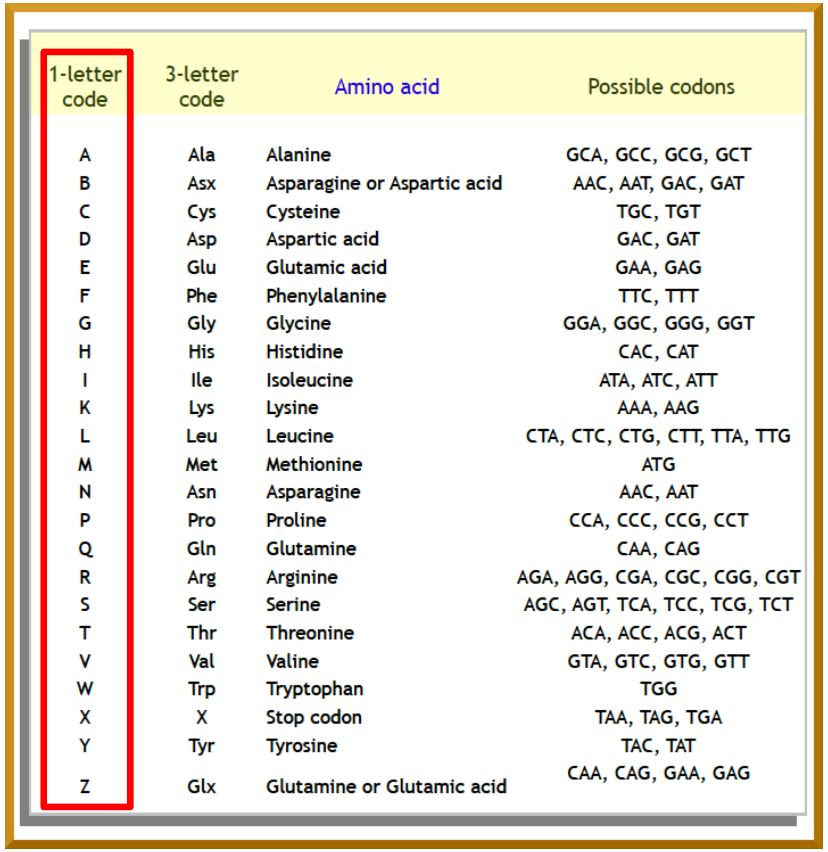


20 Aminoacids – primary structure:

(Frederick Sanger-1958 Nobel prize for insulin sequencing)

Secondary structure
Tertiary structure
Quaternary structure

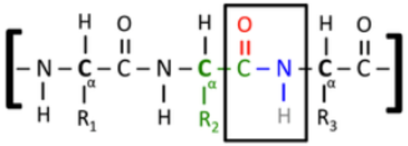




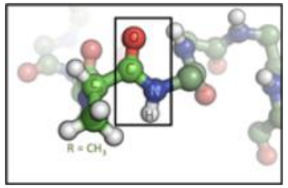
1-letter code	3-letter code	Amino acid	Possible codons
A	Ala	Alanine	GCA, GCC, GCG, GCT
B	Asx	Asparagine or Aspartic acid	AAC, AAT, GAC, GAT
C	Cys	Cysteine	TGC, TGT
D	Asp	Aspartic acid	GAC, GAT
E	Glu	Glutamic acid	GAA, GAG
F	Phe	Phenylalanine	TTC, TTT
G	Gly	Glycine	GGA, GGC, GGG, GGT
H	His	Histidine	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

SEQUENCE ⇨ STRUCTURE ⇨ FUNCTION

Protein sequences are the fundamental determinants of biological structure and function.

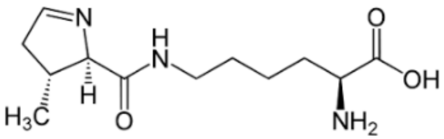


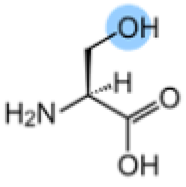
Proteins



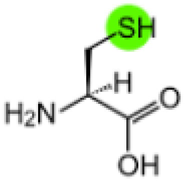
20 Aminoacids – primary structure:

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

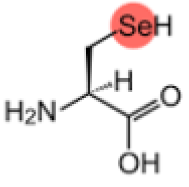




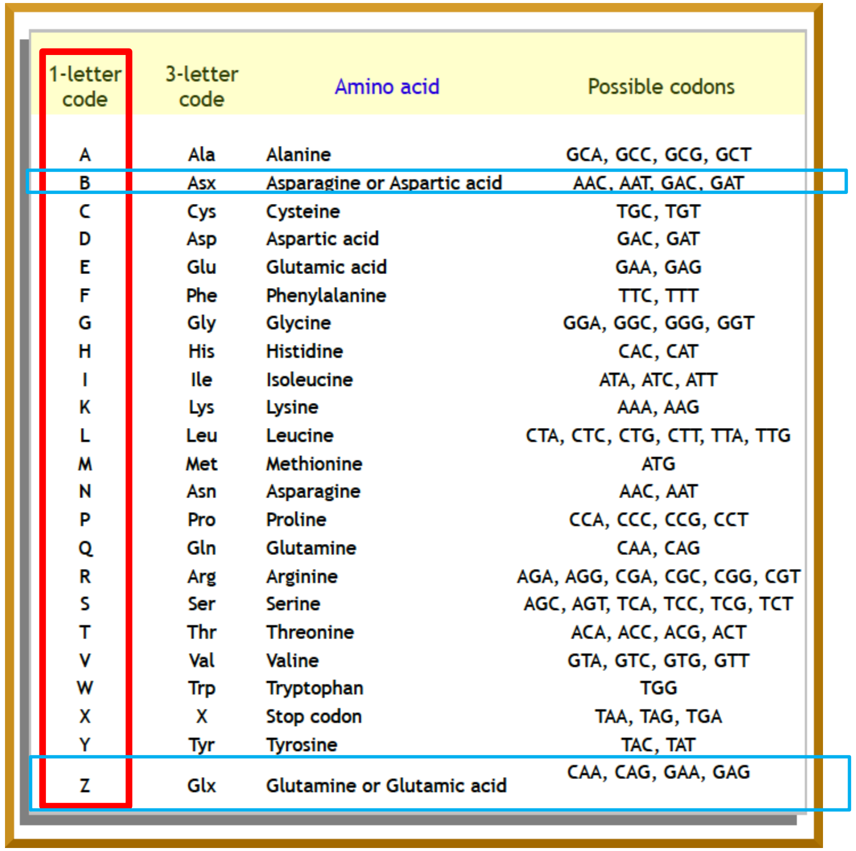
Serine (Ser)



Cysteine (Cys)

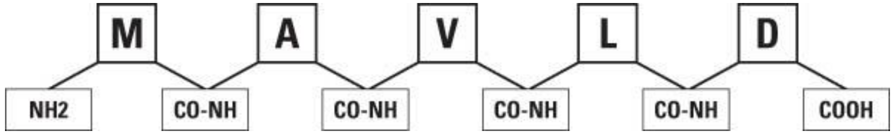


Selenocysteine (Sec)

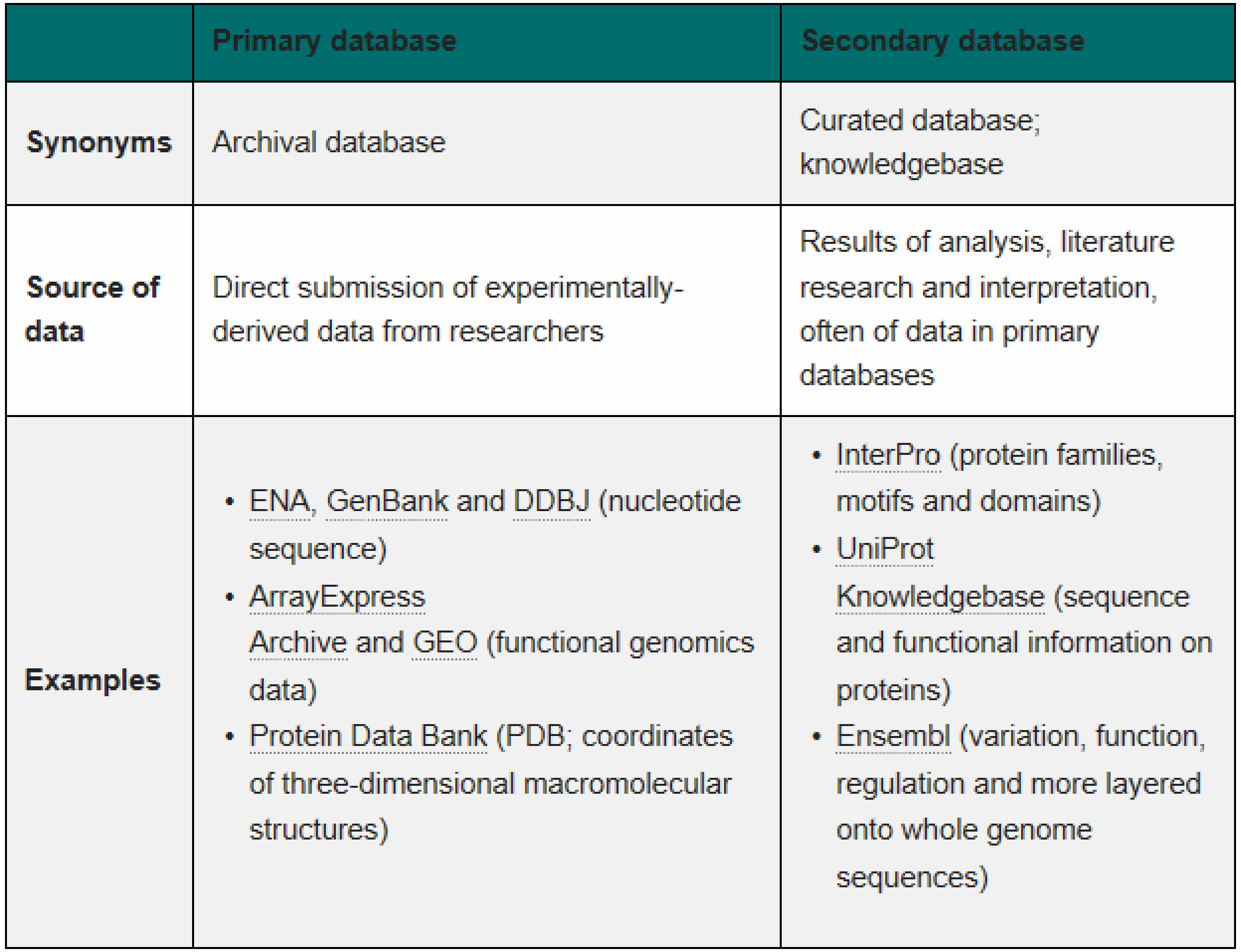


1-letter code	3-letter code	Amino acid	Possible codons
A	Ala	Alanine	GCA, GCC, GCG, GCT
B	Asx	Asparagine or Aspartic acid	AAC, AAT, GAC, GAT
C	Cys	Cysteine	TGC, TGT
D	Asp	Aspartic acid	GAC, GAT
E	Glu	Glutamic acid	GAA, GAG
F	Phe	Phenylalanine	TTC, TTT
G	Gly	Glycine	GGA, GGC, GGG, GGT
H	His	Histidine	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

N-terminus → C-terminus

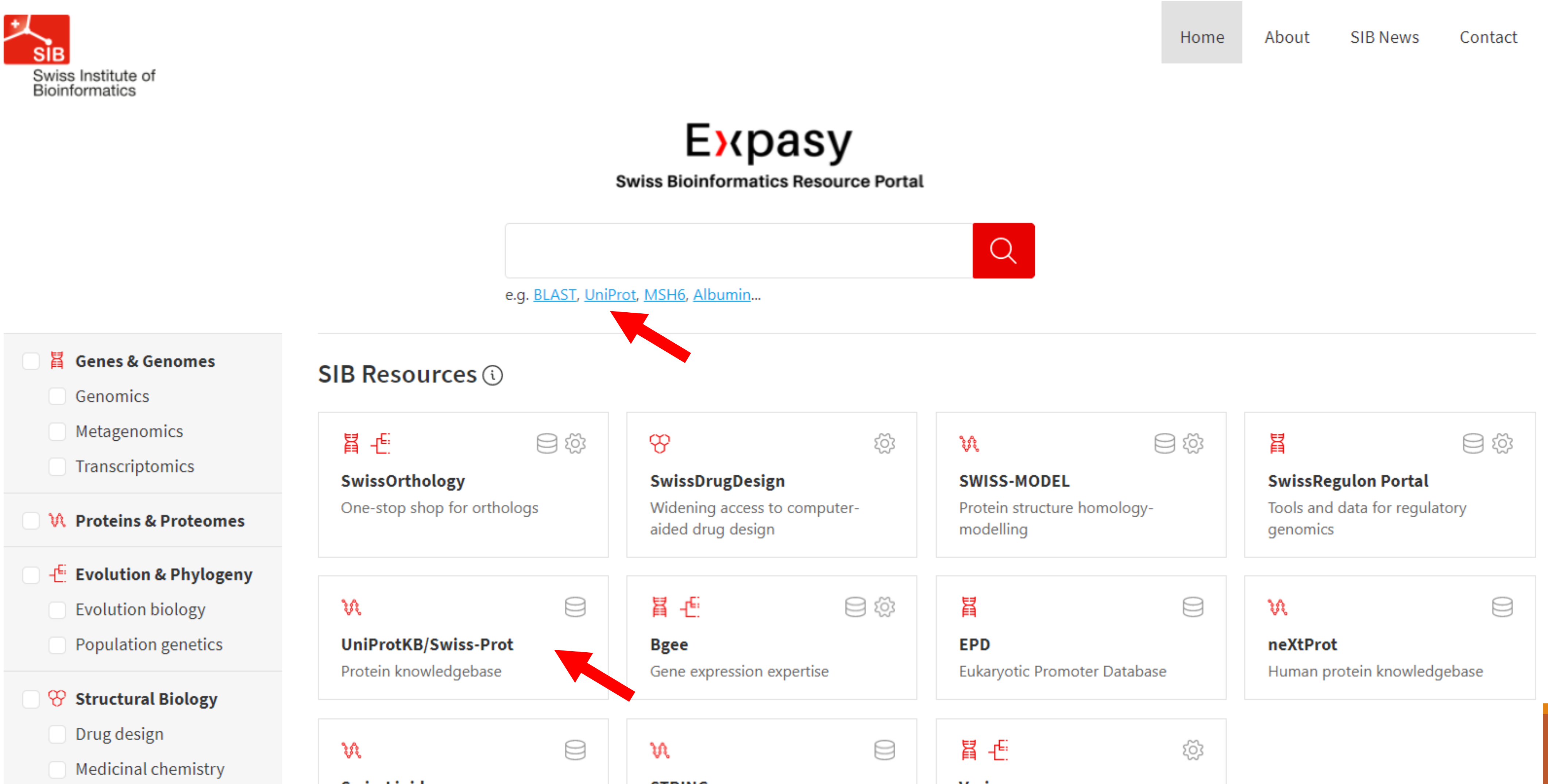


Databases



	Primary database	Secondary database
Synonyms	Archival database	Curated database; knowledgebase
Source of data	Direct submission of experimentally-derived data from researchers	Results of analysis, literature research and interpretation, often of data in primary databases
Examples	• ENA, GenBank and DDBJ (nucleotide sequence) • ArrayExpress Archive and GEO (functional genomics data) • Protein Data Bank (PDB; coordinates of three-dimensional macromolecular structures)	• InterPro (protein families, motifs and domains) • UniProt Knowledgebase (sequence and functional information on proteins) • Ensembl (variation, function, regulation and more layered onto whole genome sequences)

Protein database: Expasy/UniProt



Home | About | SIB News | Contact

SIB Swiss Institute of Bioinformatics

Expasy

Swiss Bioinformatics Resource Portal

e.g. BLAST, UniProt, MSH6, Albumin...

- □ Genes & Genomes
  - □ Genomics
  - □ Metagenomics
  - □ Transcriptomics
- □ Proteins & Proteomes
- □ Evolution & Phylogeny
  - □ Evolution biology
  - □ Population genetics
- □ Structural Biology
  - □ Drug design
  - □ Medicinal chemistry

SIB Resources

SwissOrthology
One-stop shop for orthologs

SwissDrugDesign
Widening access to computer-aided drug design

SWISS-MODEL
Protein structure homology-modelling

SwissRegulon Portal
Tools and data for regulatory genomics

UniProtKB/Swiss-Prot
Protein knowledgebase

Bgee
Gene expression expertise

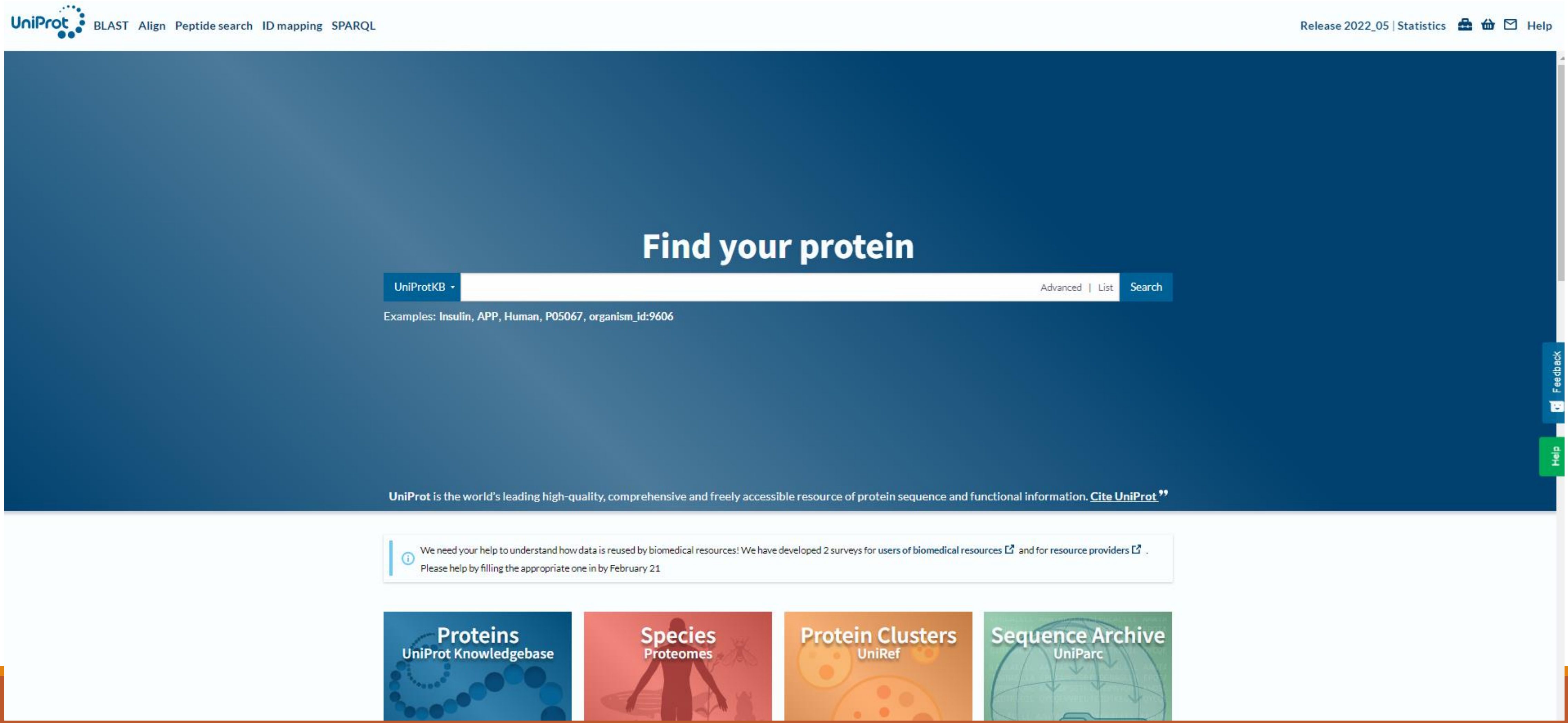
EPD
Eukaryotic Promoter Database

neXtProt
Human protein knowledgebase

http

Protein database: UniProt

http://www.uniprot.org/



Find your protein

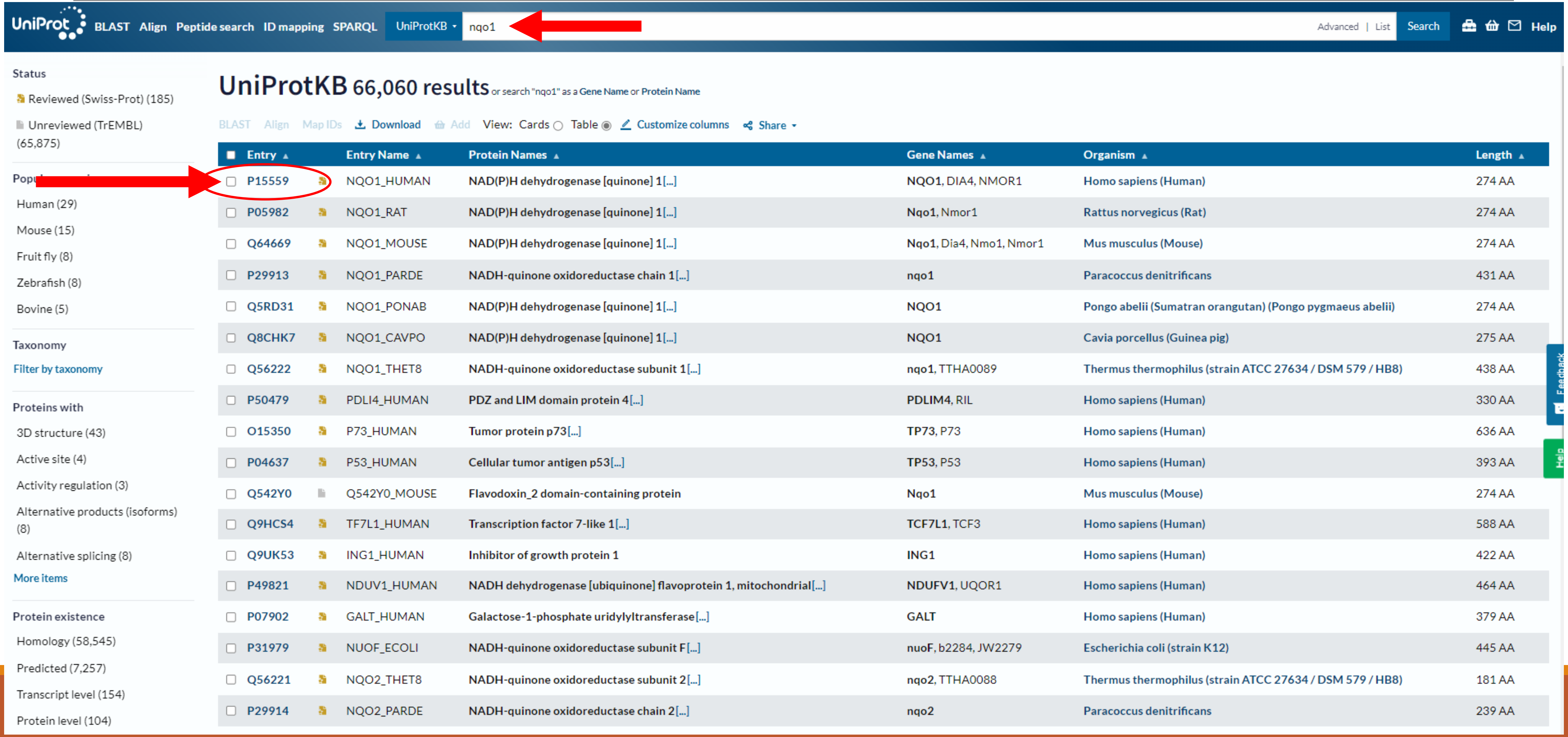
Examples: Insulin, APP, Human, P05067, organism_id:9606

UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. Cite UniProt

We need your help to understand how data is reused by biomedical resources! We have developed 2 surveys for users of biomedical resources and for resource providers. Please help by filling the appropriate one in by February 21

Protein database: UniProt

http://www.uniprot.org/



nqo1

Status

Reviewed (Swiss-Prot) (185)

Unreviewed (TrEMBL) (65,875)

Popu

Human (29)

Mouse (15)

Fruit fly (8)

Zebrafish (8)

Bovine (5)

Taxonomy

Filter by taxonomy

Proteins with

3D structure (43)

Active site (4)

Activity regulation (3)

Alternative products (isoforms) (8)

Alternative splicing (8)

More items

Protein existence

Homology (58,545)

Predicted (7,257)

Transcript level (154)

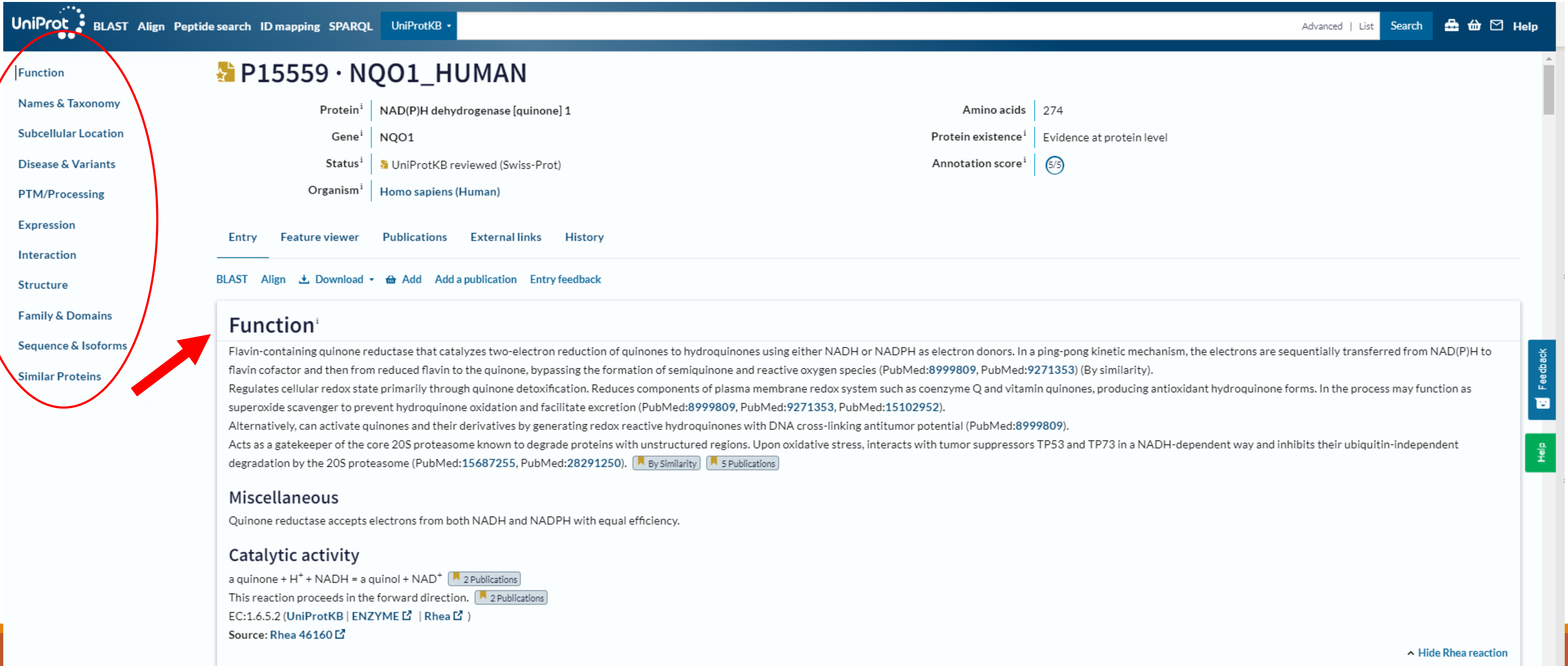
Protein level (104)

UniProtKB 66,060 results or search "nqo1" as a Gene Name or Protein Name

BLAST Align Map IDs Download Add View: Cards ○ Table ● Customize columns Share ▾

☐	Entry ▲	Entry Name ▲	Protein Names ▲	Gene Names ▲	Organism ▲	Length ▲
☐	P15559	NQO1_HUMAN	NAD(P)H dehydrogenase [quinone] 1[...]	NQO1, DIA4, NMOR1	Homo sapiens (Human)	274 AA
☐	P05982	NQO1_RAT	NAD(P)H dehydrogenase [quinone] 1[...]	Nqo1, Nmor1	Rattus norvegicus (Rat)	274 AA
☐	Q64669	NQO1_MOUSE	NAD(P)H dehydrogenase [quinone] 1[...]	Nqo1, Dia4, Nmo1, Nmor1	Mus musculus (Mouse)	274 AA
☐	P29913	NQO1_PARDE	NADH-quinone oxidoreductase chain 1[...]	nqo1	Paracoccus denitrificans	431 AA
☐	Q5RD31	NQO1_PONAB	NAD(P)H dehydrogenase [quinone] 1[...]	NQO1	Pongo abelii (Sumatran orangutan) (Pongo pygmaeus abelii)	274 AA
☐	Q8CHK7	NQO1_CAVPO	NAD(P)H dehydrogenase [quinone] 1[...]	NQO1	Cavia porcellus (Guinea pig)	275 AA
☐	Q56222	NQO1_THET8	NADH-quinone oxidoreductase subunit 1[...]	nqo1, TTHA0089	Thermus thermophilus (strain ATCC 27634 / DSM 579 / HB8)	438 AA
☐	P50479	PDLI4_HUMAN	PDZ and LIM domain protein 4[...]	PDLIM4, RIL	Homo sapiens (Human)	330 AA
☐	O15350	P73_HUMAN	Tumor protein p73[...]	TP73, P73	Homo sapiens (Human)	636 AA
☐	P04637	P53_HUMAN	Cellular tumor antigen p53[...]	TP53, P53	Homo sapiens (Human)	393 AA
☐	Q542Y0	Q542Y0_MOUSE	Flavodoxin_2 domain-containing protein	Nqo1	Mus musculus (Mouse)	274 AA
☐	Q9HCS4	TF7L1_HUMAN	Transcription factor 7-like 1[...]	TCF7L1, TCF3	Homo sapiens (Human)	588 AA
☐	Q9UK53	ING1_HUMAN	Inhibitor of growth protein 1	ING1	Homo sapiens (Human)	422 AA
☐	P49821	NDUV1_HUMAN	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial[...]	NDUFV1, UQOR1	Homo sapiens (Human)	464 AA
☐	P07902	GALT_HUMAN	Galactose-1-phosphate uridylyltransferase[...]	GALT	Homo sapiens (Human)	379 AA
☐	P31979	NUOF_ECOLI	NADH-quinone oxidoreductase subunit F[...]	nuoF, b2284, JW2279	Escherichia coli (strain K12)	445 AA
☐	Q56221	NQO2_THET8	NADH-quinone oxidoreductase subunit 2[...]	nqo2, TTHA0088	Thermus thermophilus (strain ATCC 27634 / DSM 579 / HB8)	181 AA
☐	P29914	NQO2_PARDE	NADH-quinone oxidoreductase chain 2[...]	nqo2	Paracoccus denitrificans	239 AA

Protein database: UniProt



P15559 · NQO1_HUMAN

Protein	NAD(P)H dehydrogenase [quinone] 1
Gene	NQO1
Status	UniProtKB reviewed (Swiss-Prot)
Organism	Homo sapiens (Human)
Amino acids	274
Protein existence	Evidence at protein level
Annotation score	5/5

Function

Flavin-containing quinone reductase that catalyzes two-electron reduction of quinones to hydroquinones using either NADH or NADPH as electron donors. In a ping-pong kinetic mechanism, the electrons are sequentially transferred from NAD(P)H to flavin cofactor and then from reduced flavin to the quinone, bypassing the formation of semiquinone and reactive oxygen species (PubMed:8999809, PubMed:9271353) (By similarity).

Regulates cellular redox state primarily through quinone detoxification. Reduces components of plasma membrane redox system such as coenzyme Q and vitamin quinones, producing antioxidant hydroquinone forms. In the process may function as superoxide scavenger to prevent hydroquinone oxidation and facilitate excretion (PubMed:8999809, PubMed:9271353, PubMed:15102952).

Alternatively, can activate quinones and their derivatives by generating redox reactive hydroquinones with DNA cross-linking antitumor potential (PubMed:8999809).

Acts as a gatekeeper of the core 20S proteasome known to degrade proteins with unstructured regions. Upon oxidative stress, interacts with tumor suppressors TP53 and TP73 in a NADH-dependent way and inhibits their ubiquitin-independent degradation by the 20S proteasome (PubMed:15687255, PubMed:28291250). By Similarity 5 Publications

Miscellaneous

Quinone reductase accepts electrons from both NADH and NADPH with equal efficiency.

Catalytic activity

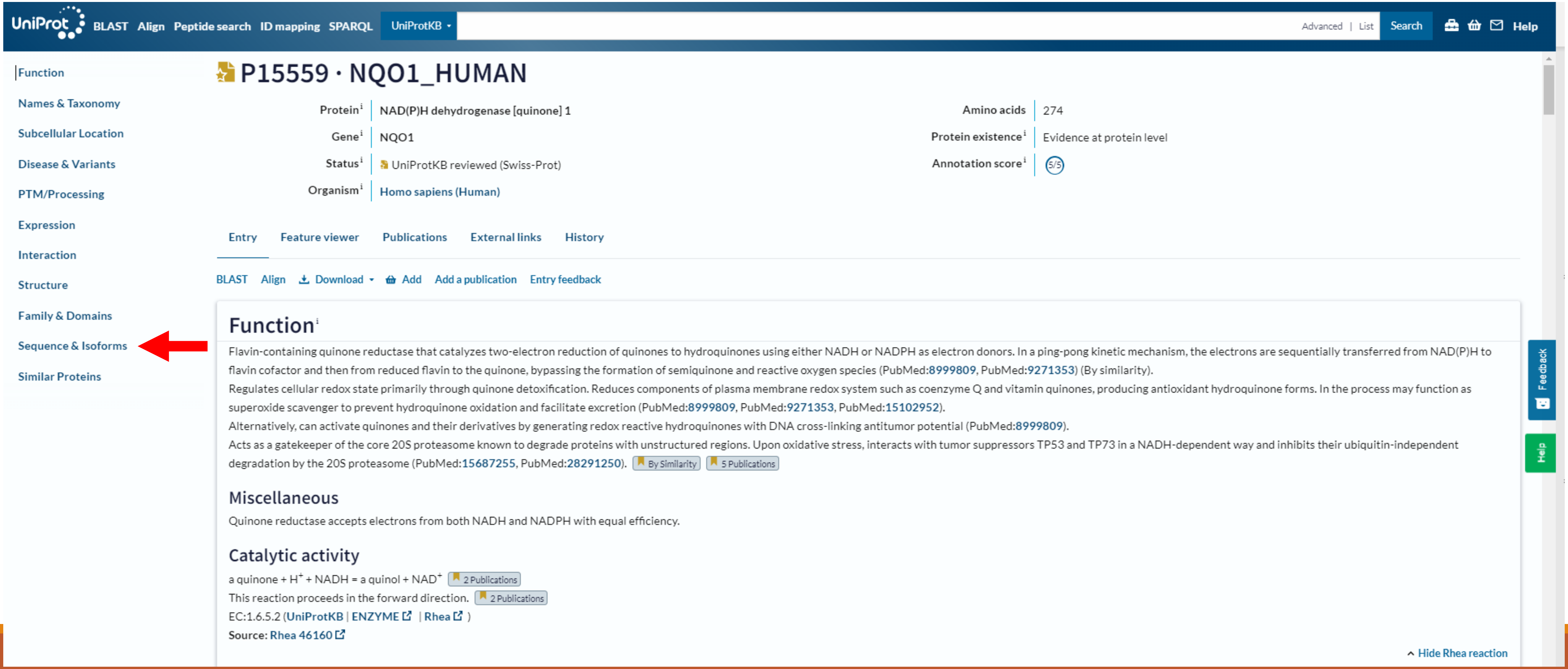
a quinone + H^+ + NADH = a quinol + NAD^+ 2 Publications

This reaction proceeds in the forward direction. 2 Publications

EC:1.6.5.2 (UniProtKB | ENZYME | Rhea)

Source: Rhea 46160

Protein database: UniProt



P15559 · NQO1_HUMAN

Protein	NAD(P)H dehydrogenase [quinone] 1	Amino acids	274
Gene	NQO1	Protein existence	Evidence at protein level
Status	UniProtKB reviewed (Swiss-Prot)	Annotation score	5/5
Organism	Homo sapiens (Human)		

Function

Flavin-containing quinone reductase that catalyzes two-electron reduction of quinones to hydroquinones using either NADH or NADPH as electron donors. In a ping-pong kinetic mechanism, the electrons are sequentially transferred from NAD(P)H to flavin cofactor and then from reduced flavin to the quinone, bypassing the formation of semiquinone and reactive oxygen species (PubMed:8999809, PubMed:9271353) (By similarity).
Regulates cellular redox state primarily through quinone detoxification. Reduces components of plasma membrane redox system such as coenzyme Q and vitamin quinones, producing antioxidant hydroquinone forms. In the process may function as superoxide scavenger to prevent hydroquinone oxidation and facilitate excretion (PubMed:8999809, PubMed:9271353, PubMed:15102952).
Alternatively, can activate quinones and their derivatives by generating redox reactive hydroquinones with DNA cross-linking antitumor potential (PubMed:8999809).
Acts as a gatekeeper of the core 20S proteasome known to degrade proteins with unstructured regions. Upon oxidative stress, interacts with tumor suppressors TP53 and TP73 in a NADH-dependent way and inhibits their ubiquitin-independent degradation by the 20S proteasome (PubMed:15687255, PubMed:28291250). By Similarity 5 Publications

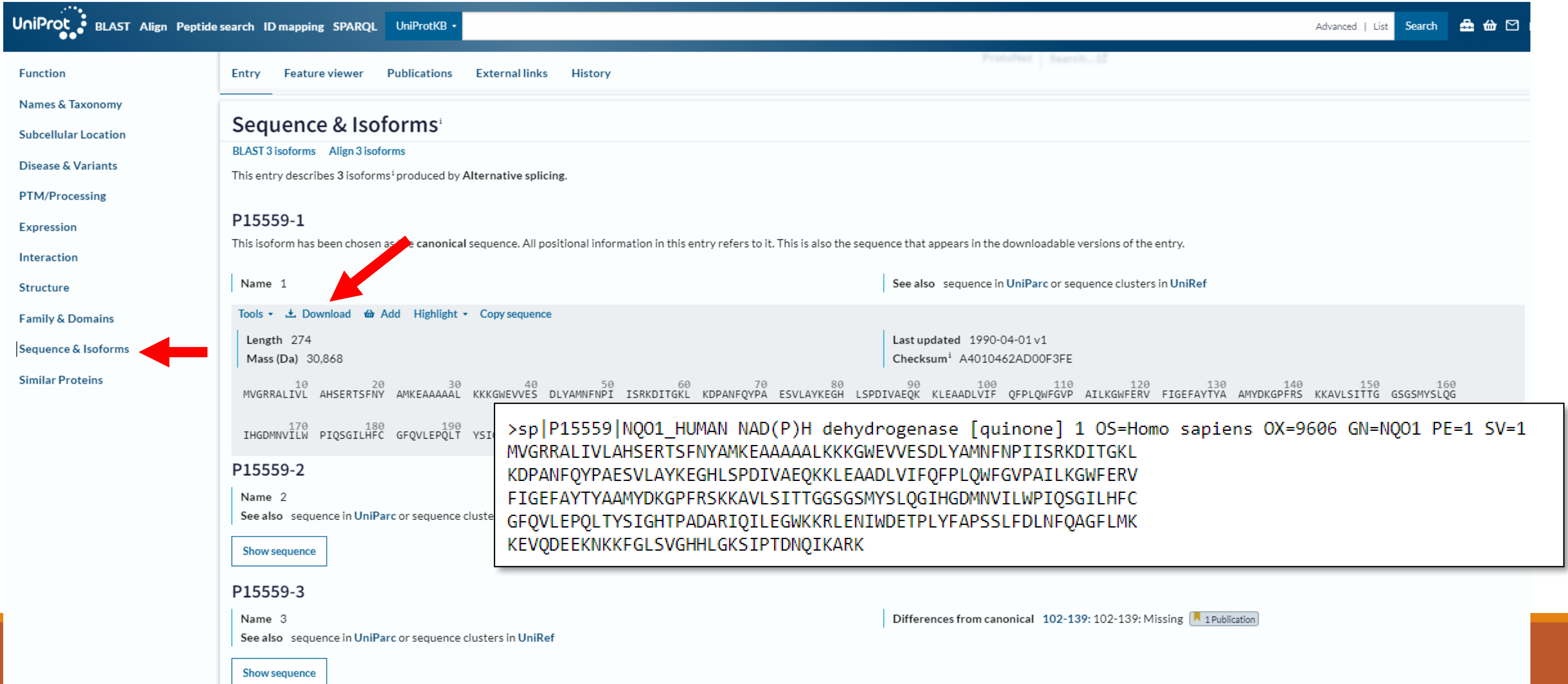
Miscellaneous

Quinone reductase accepts electrons from both NADH and NADPH with equal efficiency.

Catalytic activity

a quinone + H^+ + NADH = a quinol + NAD^+ 2 Publications
This reaction proceeds in the forward direction. 2 Publications
EC:1.6.5.2 (UniProtKB | ENZYME | Rhea)
Source: Rhea 46160

Protein database: UniProt



Sequence & Isoforms

This entry describes **3** isoforms produced by **Alternative splicing**.

P15559-1

This isoform has been chosen as the **canonical** sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

Name 1

See also sequence in UniParc or sequence clusters in UniRef

Length 274
Mass (Da) 30,868

Last updated 1990-04-01 v1
Checksum A4010462AD00F3FE

```
        10         20         30         40         50         60         70         80         90        100        110        120        130        140        150        160
MVGRRALIVL AHSERTSFNY AMKEAAAAAL KKKGWEVVES DLYAMNFNPI ISRKDITGKL KDPANFQYPA ESVLAYKEGH LSPDIVAEQK KLEAADLVIF QFPLQWFGVP AILKGWFERV FIGEFAYTYA AMYDKGPFRS KKAVLSITTG GSGSMYSLQG

       170        180        190
IHGDMNVILW PIQSGILHFC GFQVLEPQLT YSI
```

P15559-2

Name 2

See also sequence in UniParc or sequence cluste

P15559-3

Name 3

See also sequence in UniParc or sequence clusters in UniRef

Differences from canonical 102-139: 102-139: Missing

```
>sp|P15559|NQO1_HUMAN NAD(P)H dehydrogenase [quinone] 1 OS=Homo sapiens OX=9606 GN=NQO1 PE=1 SV=1
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL
KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV
FIGEFAYTYAAMYDKGPFRSKKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFC
GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK
KEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
```

FASTA (and RAW) format

FASTA = popular tool for sequence comparison and database searching

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

fasta format:

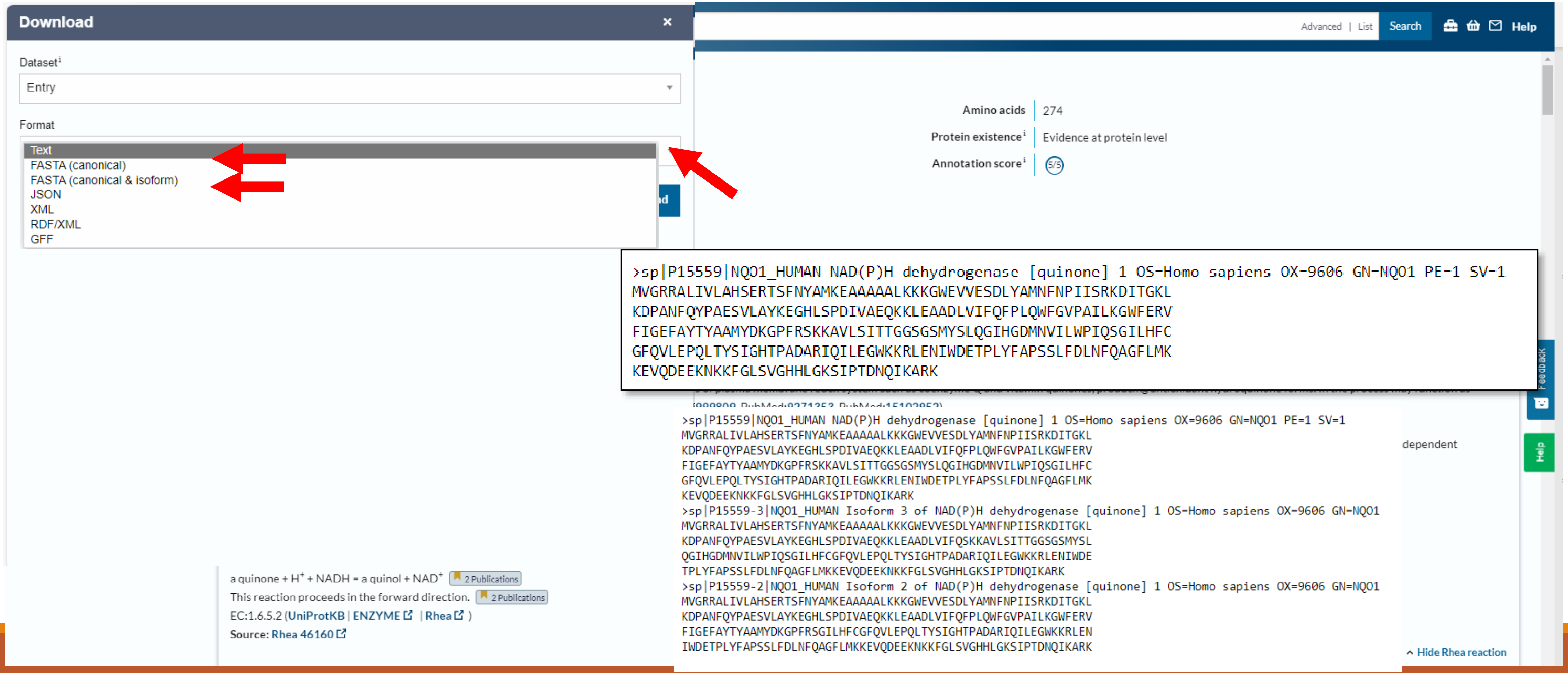
```
>NQO1_homo
MVGRRALIVLAHSERTSFNYAMKEAAAAA
LKKKGWEVVESDLYAMNFNPIISRKDITG
KLKDPANFQYPA
```

```
MVGRRALIVLAHSERTSFNYAMKEAAAAA
LKKKGWEVVESDLYAMNFNPIISRKDITG
KLKDPANFQYPA
```

1. „definition“ line starts with **>** and unique identification follows.

2. line-a sequence (DNA/protein- single letter code).

Protein database: UniProt



Dataset[i]

Entry

Format

Text
FASTA (canonical)
FASTA (canonical & isoform)
JSON
XML
RDF/XML
GFF

Amino acids	274
Protein existence[i]	Evidence at protein level
Annotation score[i]	5/5

```
>sp|P15559|NQO1_HUMAN NAD(P)H dehydrogenase [quinone] 1 OS=Homo sapiens OX=9606 GN=NQO1 PE=1 SV=1
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL
KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV
FIGEFAYTYAAMYDKGPFRSKKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFC
GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK
KEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
```

```
>sp|P15559|NQO1_HUMAN NAD(P)H dehydrogenase [quinone] 1 OS=Homo sapiens OX=9606 GN=NQO1 PE=1 SV=1
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL
KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV
FIGEFAYTYAAMYDKGPFRSKKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFC
GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK
KEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
>sp|P15559-3|NQO1_HUMAN Isoform 3 of NAD(P)H dehydrogenase [quinone] 1 OS=Homo sapiens OX=9606 GN=NQO1
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL
KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQSKKAVLSITTGGSGSMYSL
QGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDE
TPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
>sp|P15559-2|NQO1_HUMAN Isoform 2 of NAD(P)H dehydrogenase [quinone] 1 OS=Homo sapiens OX=9606 GN=NQO1
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL
KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV
FIGEFAYTYAAMYDKGPFRSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKKRLEN
IWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
```

dependent

a quinone + H^+ + NADH = a quinol + NAD^+ 2 Publications

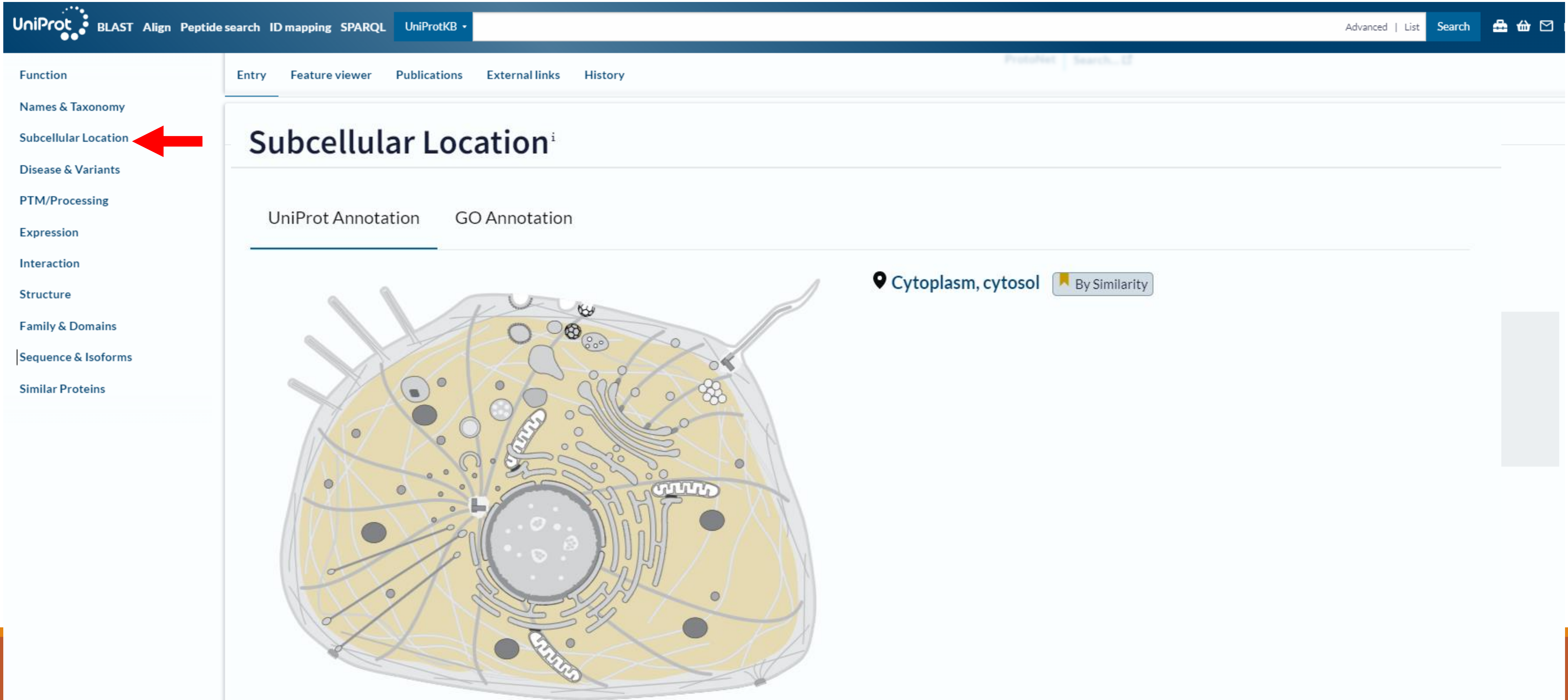
This reaction proceeds in the forward direction. 2 Publications

EC:1.6.5.2 (UniProtKB | ENZYME | Rhea)

Source: Rhea 46160

^ Hide Rhea reaction

Protein database: UniProt



Subcellular Location

Cytoplasm, cytosol By Similarity

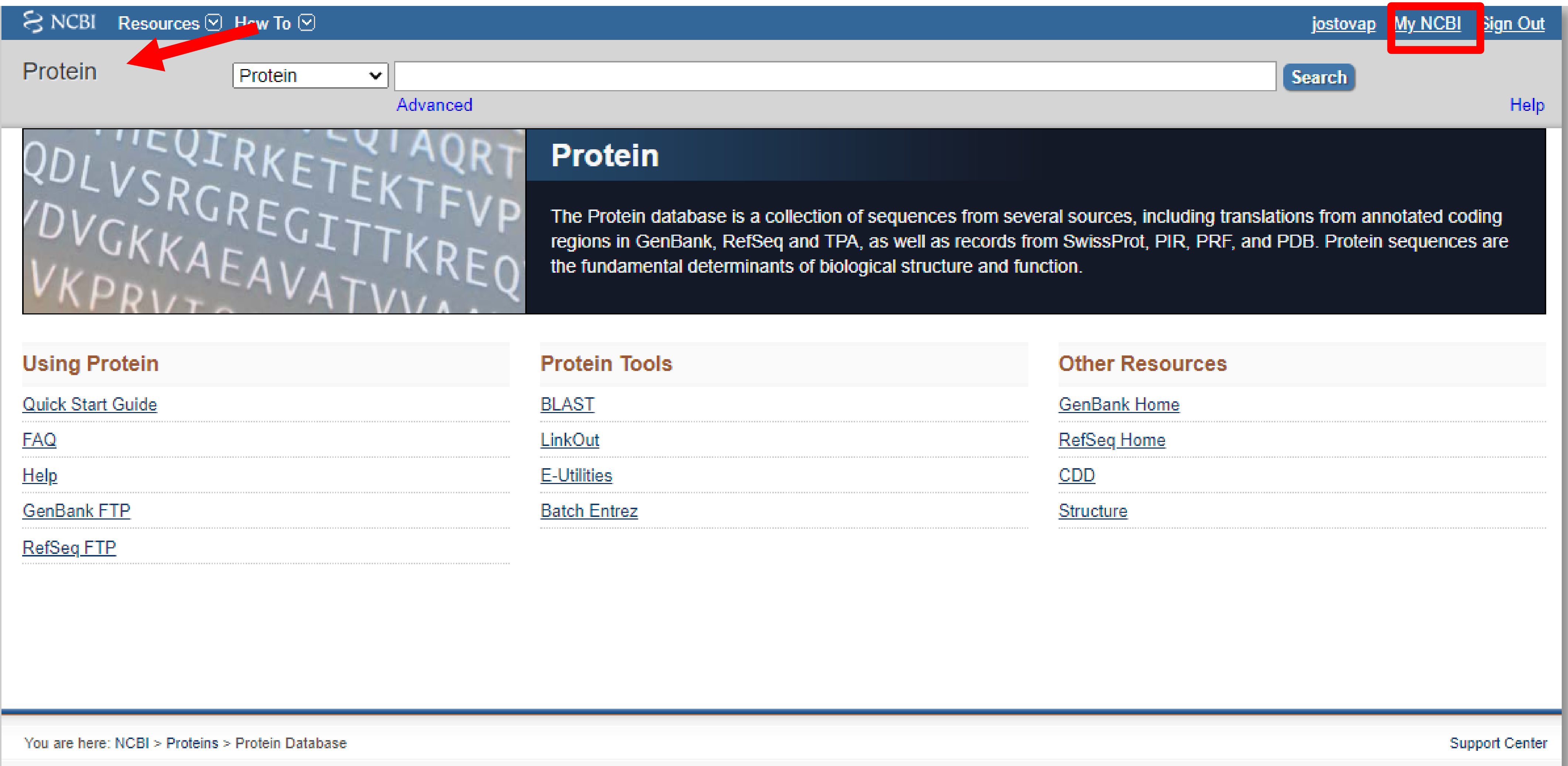
Practical part with NQO1

Try Uniprot!

Find your sequnce

-what is the function of your protein?
-find subcellular location of your protein
-is it involved in any pathology?
-how many amino acids has your sequence?

Protein database: NCBI



Protein

The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.

Using Protein

- Quick Start Guide
- FAQ
- Help
- GenBank FTP
- RefSeq FTP

Protein Tools

- BLAST
- LinkOut
- E-Utilities
- Batch Entrez

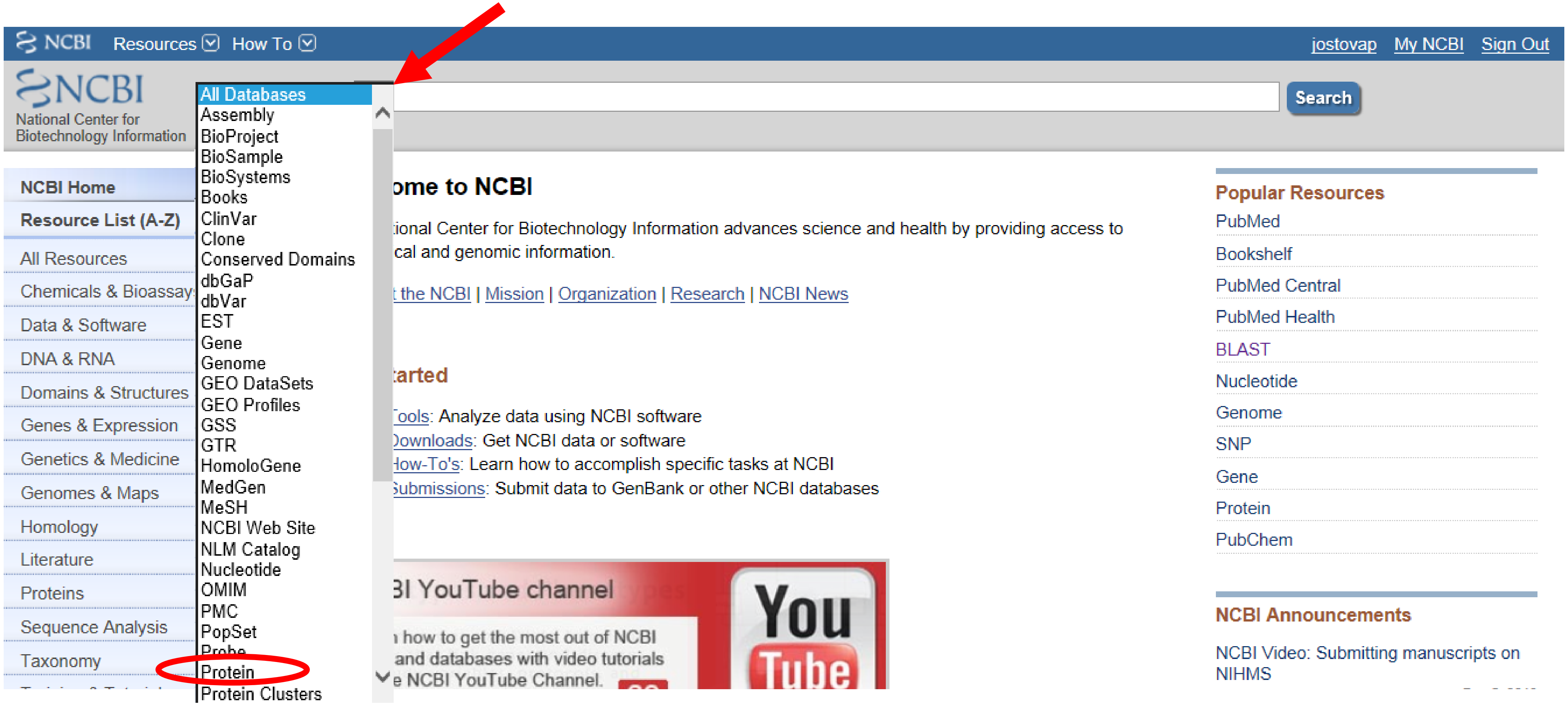
Other Resources

- GenBank Home
- RefSeq Home
- CDD
- Structure

You are here: NCBI > Proteins > Protein Database

Support Center

Protein database: NCBI



ome to NCBI

tional Center for Biotechnology Information advances science and health by providing access to cal and genomic information.

t the NCBI | Mission | Organization | Research | NCBI News

tarted

Tools: Analyze data using NCBI software

Downloads: Get NCBI data or software

How-To's: Learn how to accomplish specific tasks at NCBI

Submissions: Submit data to GenBank or other NCBI databases

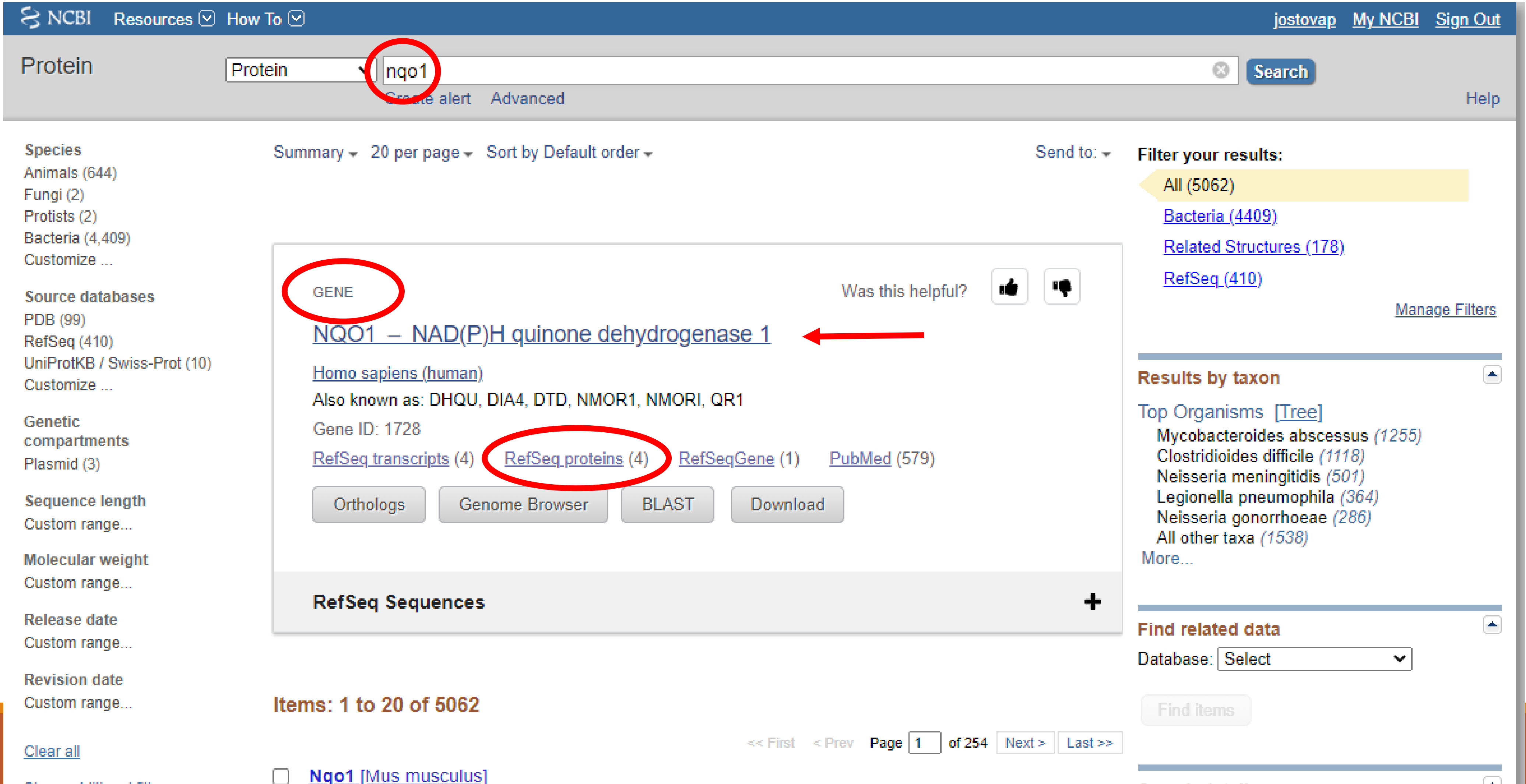
BI YouTube channel

how to get the most out of NCBI and databases with video tutorials e NCBI YouTube Channel.

NCBI Announcements

NCBI Video: Submitting manuscripts on NIHMS

Protein database: NCBI



NCBI Resources How To jostovap My NCBI Sign Out

Protein

Protein nqo1 Search

Create alert Advanced Help

Species
Animals (644)
Fungi (2)
Protists (2)
Bacteria (4,409)
Customize ...

Source databases
PDB (99)
RefSeq (410)
UniProtKB / Swiss-Prot (10)
Customize ...

Genetic compartments
Plasmid (3)

Sequence length
Custom range...

Molecular weight
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all

Summary 20 per page Sort by Default order Send to:

GENE

Was this helpful?

NQO1 – NAD(P)H quinone dehydrogenase 1

Homo sapiens (human)

Also known as: DHQU, DIA4, DTD, NMOR1, NMORI, QR1

Gene ID: 1728

RefSeq transcripts (4) RefSeq proteins (4) RefSeqGene (1) PubMed (579)

Orthologs Genome Browser BLAST Download

RefSeq Sequences +

Items: 1 to 20 of 5062

<< First < Prev Page 1 of 254 Next > Last >>

□ Ngo1 [Mus musculus]

Filter your results:
All (5062)
Bacteria (4409)
Related Structures (178)
RefSeq (410)

Manage Filters

Results by taxon

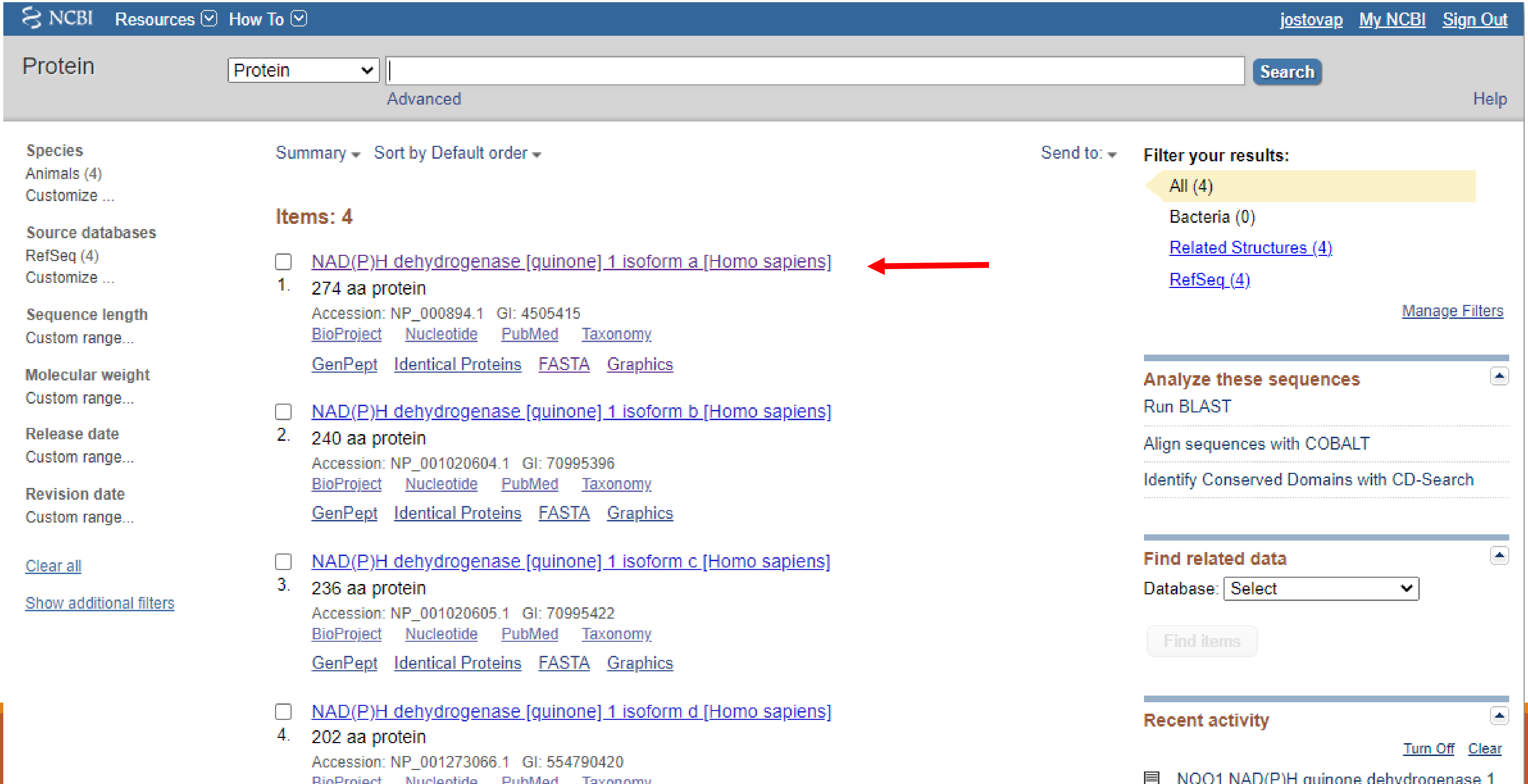
Top Organisms [Tree]
Mycobacteroides abscessus (1255)
Clostridioides difficile (1118)
Neisseria meningitidis (501)
Legionella pneumophila (364)
Neisseria gonorrhoeae (286)
All other taxa (1538)
More...

Find related data

Database: Select

Find items

Protein database: NCBI



NCBI Resources How To jostovap My NCBI Sign Out

Protein

Protein ▾ | Search

Advanced Help

Species
Animals (4)
Customize ...

Source databases
RefSeq (4)
Customize ...

Sequence length
Custom range...

Molecular weight
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all

Show additional filters

Summary ▾ Sort by Default order ▾ Send to: ▾

Items: 4

□ 1. NAD(P)H dehydrogenase [quinone] 1 isoform a [Homo sapiens]
274 aa protein
Accession: NP_000894.1 GI: 4505415
BioProject Nucleotide PubMed Taxonomy
GenPept Identical Proteins FASTA Graphics

□ 2. NAD(P)H dehydrogenase [quinone] 1 isoform b [Homo sapiens]
240 aa protein
Accession: NP_001020604.1 GI: 70995396
BioProject Nucleotide PubMed Taxonomy
GenPept Identical Proteins FASTA Graphics

□ 3. NAD(P)H dehydrogenase [quinone] 1 isoform c [Homo sapiens]
236 aa protein
Accession: NP_001020605.1 GI: 70995422
BioProject Nucleotide PubMed Taxonomy
GenPept Identical Proteins FASTA Graphics

□ 4. NAD(P)H dehydrogenase [quinone] 1 isoform d [Homo sapiens]
202 aa protein
Accession: NP_001273066.1 GI: 554790420
BioProject Nucleotide PubMed Taxonomy

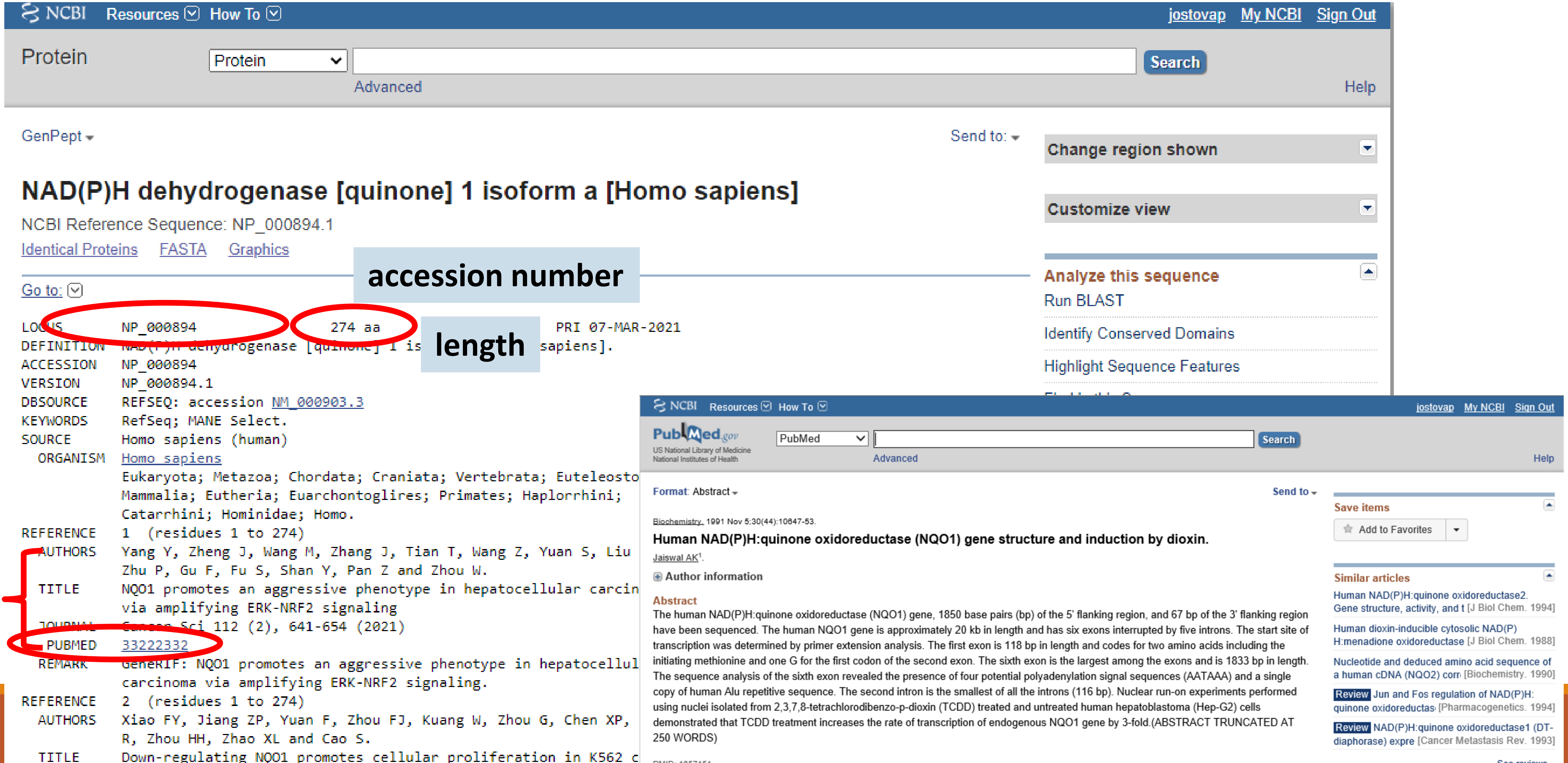
Filter your results:
All (4)
Bacteria (0)
Related Structures (4)
RefSeq (4)
Manage Filters

Analyze these sequences
Run BLAST
Align sequences with COBALT
Identify Conserved Domains with CD-Search

Find related data
Database: Select ▾
Find items

Recent activity
Turn Off Clear
NQO1 NAD(P)H quinone dehydrogenase 1

Protein database: NCBI



NAD(P)H dehydrogenase [quinone] 1 isoform a [Homo sapiens]

NCBI Reference Sequence: NP_000894.1

Identical Proteins FASTA Graphics

accession number

length

```
LOCUS       NP_000894                274 aa            PRI 07-MAR-2021
DEFINITION  NAD(P)H dehydrogenase [quinone] 1 is          sapiens].
ACCESSION   NP_000894
VERSION     NP_000894.1
DBSOURCE    REFSEQ: accession NM_000903.3
KEYWORDS    RefSeq; MANE Select.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleosto
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 274)
  AUTHORS   Yang Y, Zheng J, Wang M, Zhang J, Tian T, Wang Z, Yuan S, Liu
            Zhu P, Gu F, Fu S, Shan Y, Pan Z and Zhou W.
  TITLE     NQO1 promotes an aggressive phenotype in hepatocellular carcin
            via amplifying ERK-NRF2 signaling
  JOURNAL   Cancer Sci 112 (2), 641-654 (2021)
   PUBMED   33222332
  REMARK    GeneRIF: NQO1 promotes an aggressive phenotype in hepatocellul
            carcinoma via amplifying ERK-NRF2 signaling.
REFERENCE   2  (residues 1 to 274)
  AUTHORS   Xiao FY, Jiang ZP, Yuan F, Zhou FJ, Kuang W, Zhou G, Chen XP,
            R, Zhou HH, Zhao XL and Cao S.
  TITLE     Down-regulating NQO1 promotes cellular proliferation in K562 c
```

Format: Abstract

Biochemistry. 1991 Nov 5;30(44):10647-53.

Human NAD(P)H:quinone oxidoreductase (NQO1) gene structure and induction by dioxin.

Jaiswal AK[1].

Author information

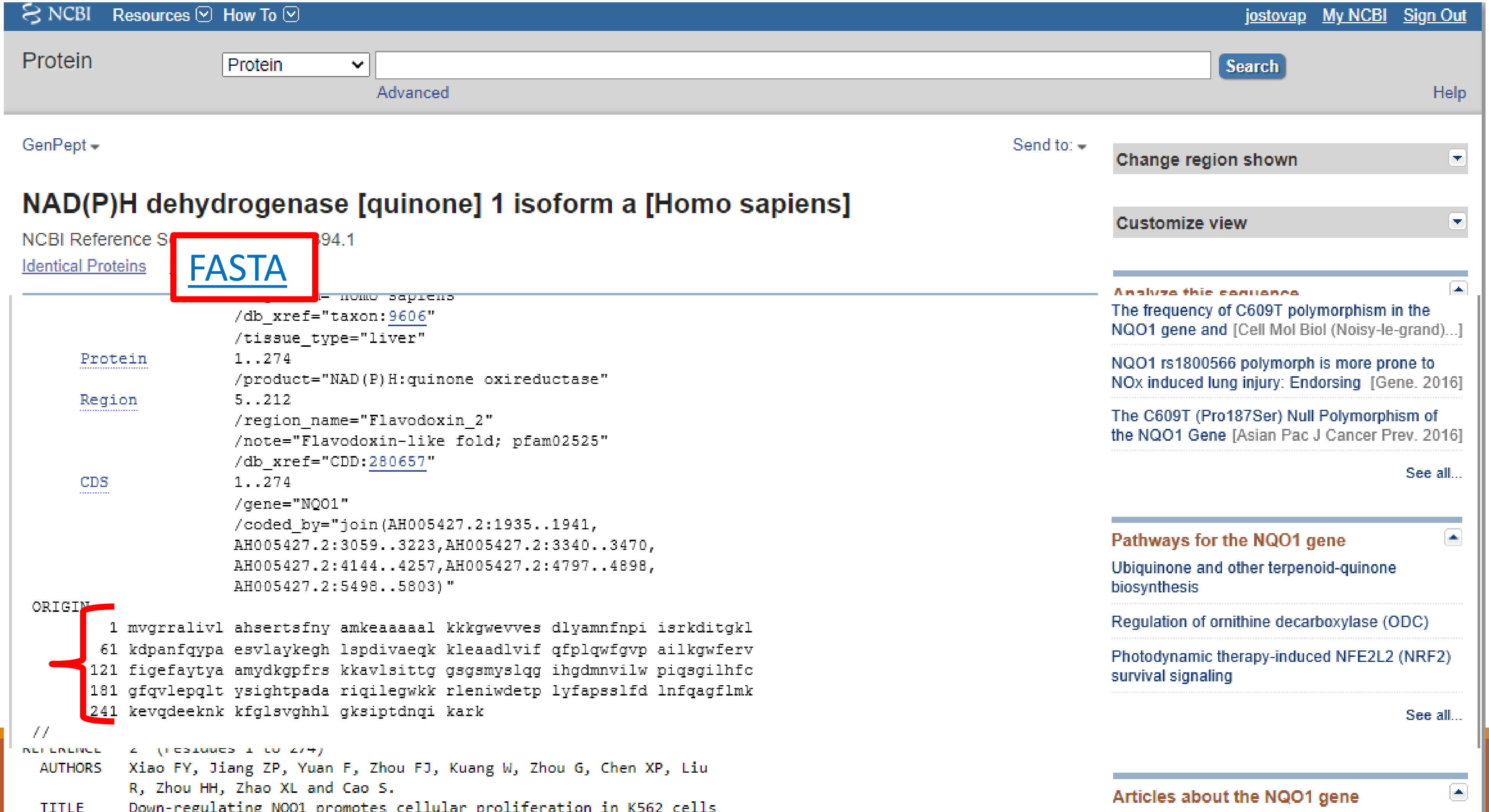
Abstract

The human NAD(P)H:quinone oxidoreductase (NQO1) gene, 1850 base pairs (bp) of the 5' flanking region, and 67 bp of the 3' flanking region have been sequenced. The human NQO1 gene is approximately 20 kb in length and has six exons interrupted by five introns. The start site of transcription was determined by primer extension analysis. The first exon is 118 bp in length and codes for two amino acids including the initiating methionine and one G for the first codon of the second exon. The sixth exon is the largest among the exons and is 1833 bp in length. The sequence analysis of the sixth exon revealed the presence of four potential polyadenylation signal sequences (AATAAA) and a single copy of human Alu repetitive sequence. The second intron is the smallest of all the introns (116 bp). Nuclear run-on experiments performed using nuclei isolated from 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) treated and untreated human hepatoblastoma (Hep-G2) cells demonstrated that TCDD treatment increases the rate of transcription of endogenous NQO1 gene by 3-fold.(ABSTRACT TRUNCATED AT 250 WORDS)

Similar articles

- Human NAD(P)H:quinone oxidoreductase2. Gene structure, activity, and t [J Biol Chem. 1994]
- Human dioxin-inducible cytosolic NAD(P) H:menadione oxidoreductase [J Biol Chem. 1988]
- Nucleotide and deduced amino acid sequence of a human cDNA (NQO2) corr [Biochemistry. 1990]
- Review Jun and Fos regulation of NAD(P)H: quinone oxidoreductas [Pharmacogenetics. 1994]
- Review NAD(P)H:quinone oxidoreductase1 (DT-diaphorase) expre [Cancer Metastasis Rev. 1993]

Protein database: NCBI



NAD(P)H dehydrogenase [quinone] 1 isoform a [Homo sapiens]

NCBI Reference S…94.1

Identical Proteins

FASTA

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     Protein         1..274
                     /product="NAD(P)H:quinone oxireductase"
     Region          5..212
                     /region_name="Flavodoxin_2"
                     /note="Flavodoxin-like fold; pfam02525"
                     /db_xref="CDD:280657"
     CDS             1..274
                     /gene="NQO1"
                     /coded_by="join(AH005427.2:1935..1941,
                     AH005427.2:3059..3223,AH005427.2:3340..3470,
                     AH005427.2:4144..4257,AH005427.2:4797..4898,
                     AH005427.2:5498..5803)"
ORIGIN
        1 mvgrralivl ahsertsfny amkeaaaaal kkkgwevves dlyamnfnpi isrkditgkl
       61 kdpanfqypa esvlaykegh lspdivaeqk kleaadlvif qfplqwfgvp ailkgwferv
      121 figefaytya amydkgpfrs kkavlsittg gsgsmyslqg ihgdmnvilw piqsgilhfc
      181 gfqvlepqlt ysightpada riqilegwkk rleniwdetp lyfapsslfd lnfqagflmk
      241 kevqdeeknk kfglsvghhl gksiptdnqi kark
//
REFERENCE   2  (residues 1 to 274)
  AUTHORS   Xiao FY, Jiang ZP, Yuan F, Zhou FJ, Kuang W, Zhou G, Chen XP, Liu
            R, Zhou HH, Zhao XL and Cao S.
  TITLE     Down-regulating NQO1 promotes cellular proliferation in K562 cells
```

The frequency of C609T polymorphism in the NQO1 gene and [Cell Mol Biol (Noisy-le-grand)...]

NQO1 rs1800566 polymorph is more prone to NOx induced lung injury: Endorsing [Gene. 2016]

The C609T (Pro187Ser) Null Polymorphism of the NQO1 Gene [Asian Pac J Cancer Prev. 2016]

See all...

Pathways for the NQO1 gene

Ubiquinone and other terpenoid-quinone biosynthesis

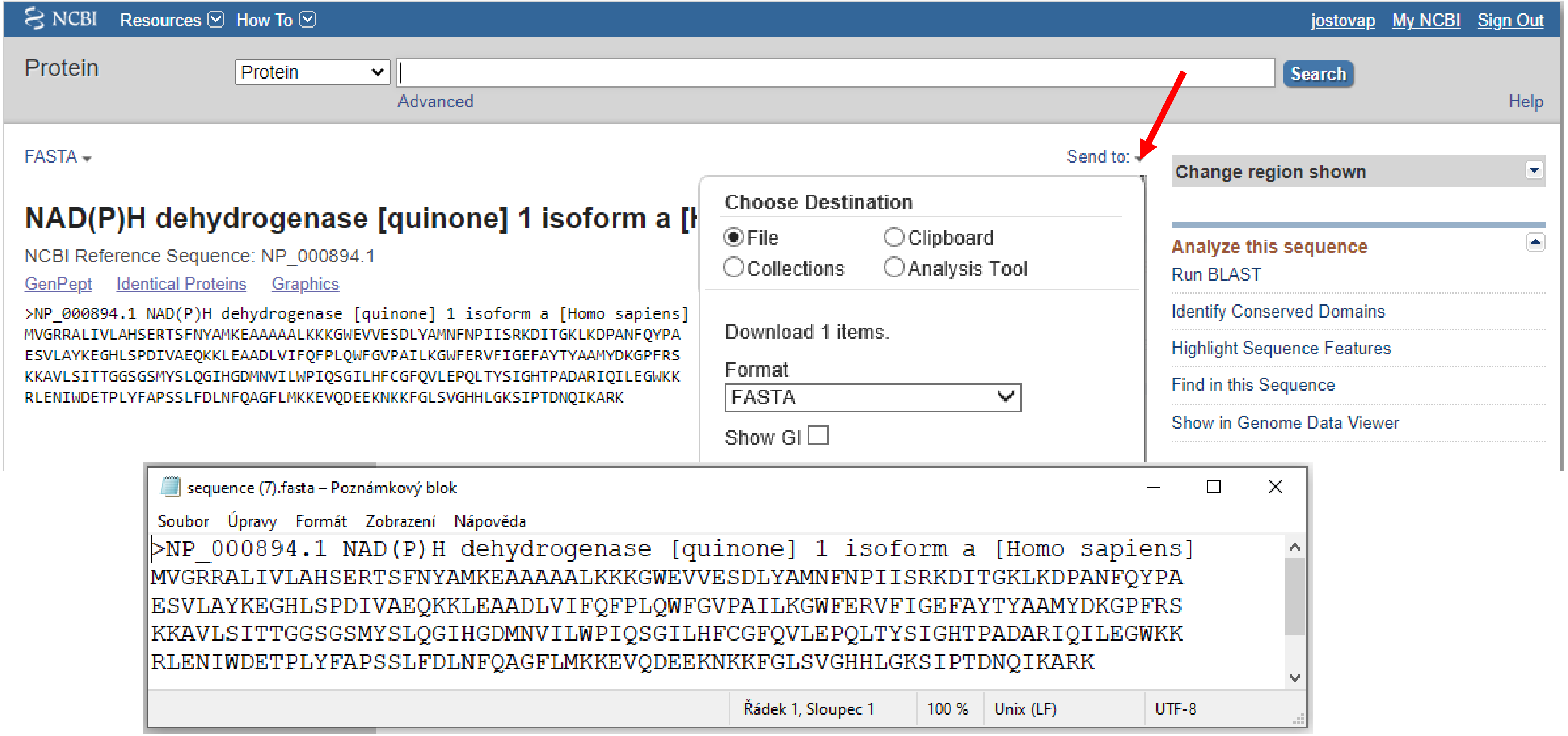
Regulation of ornithine decarboxylase (ODC)

Photodynamic therapy-induced NFE2L2 (NRF2) survival signaling

See all...

Articles about the NQO1 gene

Protein database: NCBI



NAD(P)H dehydrogenase [quinone] 1 isoform a [H

NCBI Reference Sequence: NP_000894.1

GenPept Identical Proteins Graphics

```
>NP_000894.1 NAD(P)H dehydrogenase [quinone] 1 isoform a [Homo sapiens]
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKLKDPANFQYPA
ESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRS
KKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKK
RLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
```

Choose Destination

● File ○ Clipboard
○ Collections ○ Analysis Tool

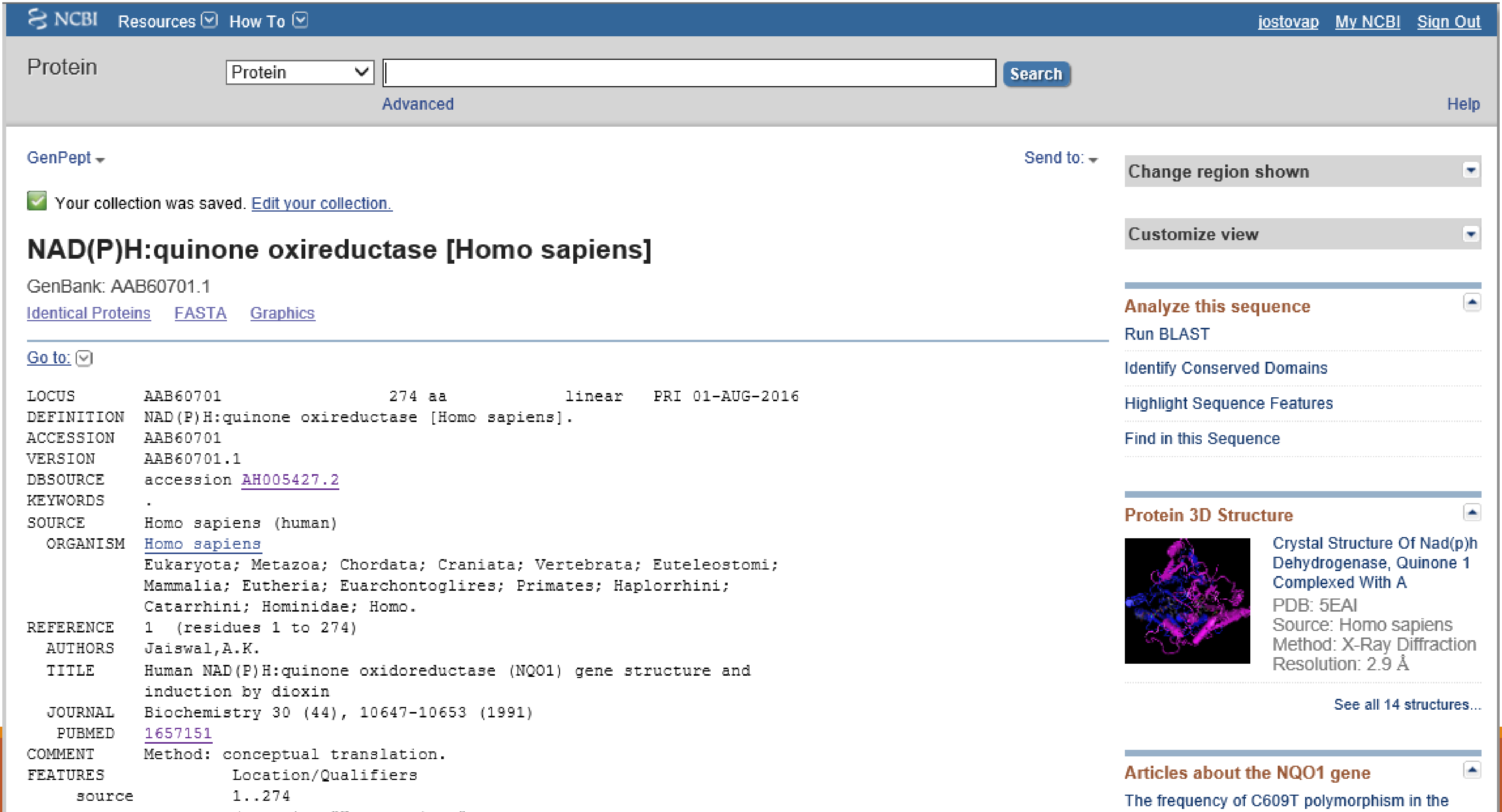
Download 1 items.

Format: FASTA

Show GI □

```
>NP_000894.1 NAD(P)H dehydrogenase [quinone] 1 isoform a [Homo sapiens]
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKLKDPANFQYPA
ESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRS
KKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKK
RLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
```

Protein database: NCBI → MyNCBI



Your collection was saved. Edit your collection.

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

GenBank: AAB60701.1

Identical Proteins FASTA Graphics

```
LOCUS       AAB60701                 274 aa            linear   PRI 01-AUG-2016
DEFINITION  NAD(P)H:quinone oxireductase [Homo sapiens].
ACCESSION   AAB60701
VERSION     AAB60701.1
DBSOURCE    accession AH005427.2
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 274)
  AUTHORS   Jaiswal,A.K.
  TITLE     Human NAD(P)H:quinone oxidoreductase (NQO1) gene structure and
            induction by dioxin
  JOURNAL   Biochemistry 30 (44), 10647-10653 (1991)
   PUBMED   1657151
COMMENT     Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..274
```

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Protein 3D Structure

Crystal Structure Of Nad(p)h Dehydrogenase, Quinone 1 Complexed With A

PDB: 5EAI

Source: Homo sapiens

Method: X-Ray Diffraction

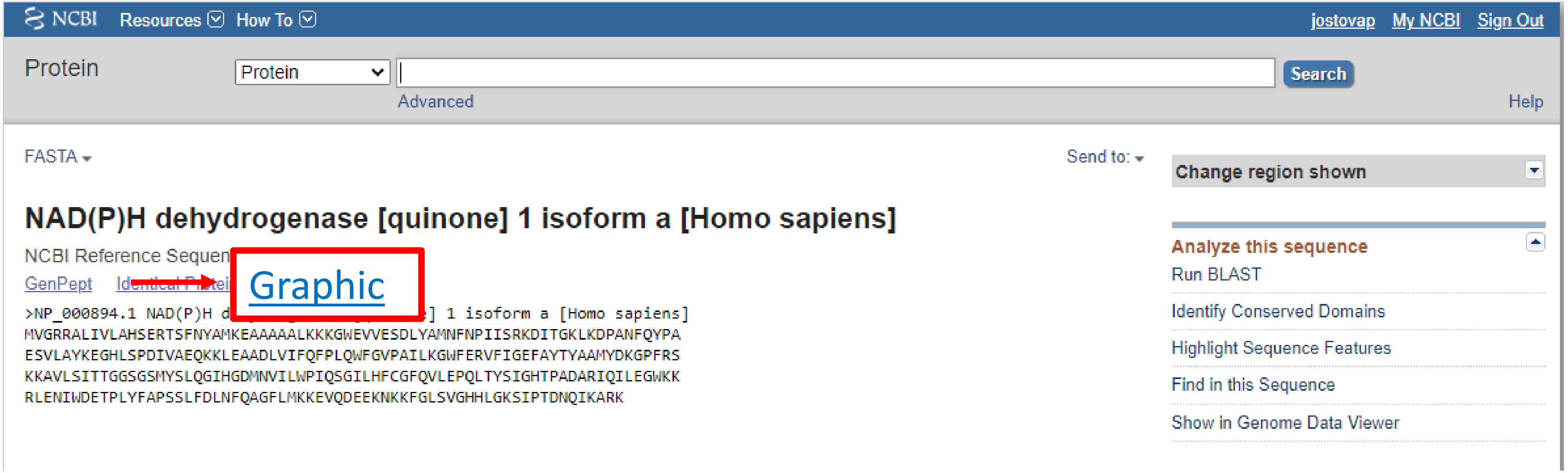
Resolution: 2.9 Å

See all 14 structures...

Articles about the NQO1 gene

The frequency of C609T polymorphism in the

Protein database: NCBI



NAD(P)H dehydrogenase [quinone] 1 isoform a [Homo sapiens]

NCBI Reference Sequen

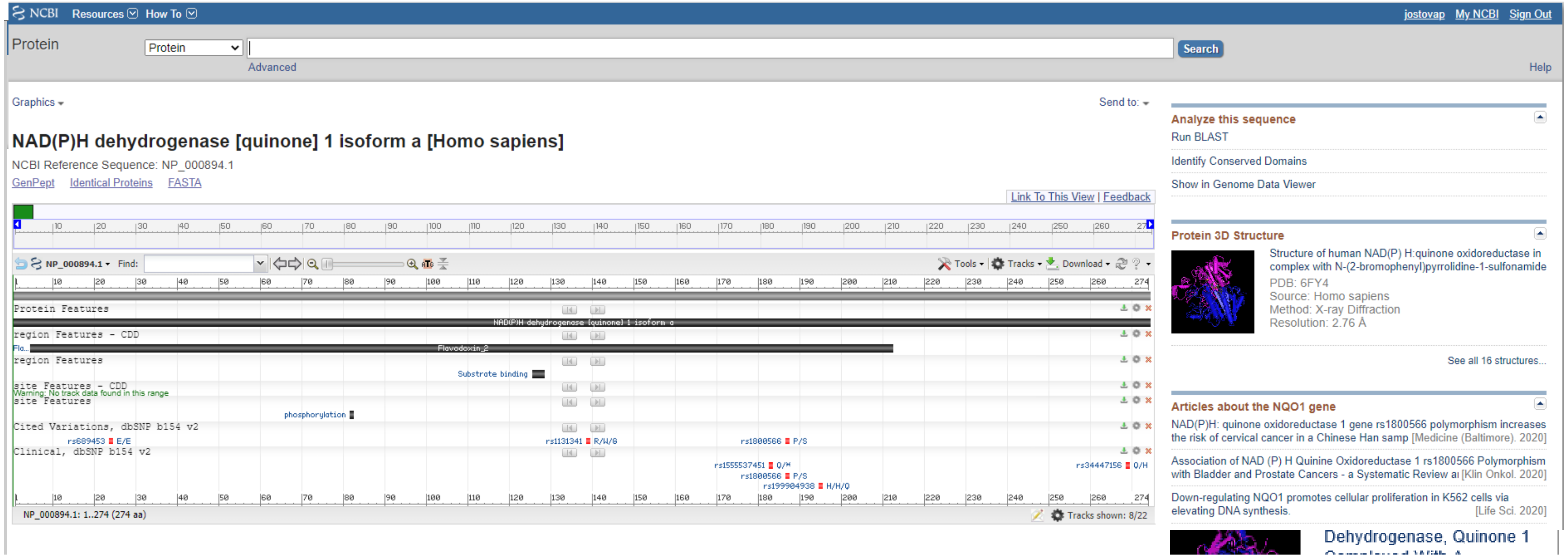
GenPept Identical Protei

Graphic

```
>NP_000894.1 NAD(P)H d                    ] 1 isoform a [Homo sapiens]
MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKLKDPANFQYPA
ESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRS
KKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYSIGHTPADARIQILEGWKK
RLENIWDETPLYFAPSSLFDLNFQAGFLMKKEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
```

Analyze this sequence
Run BLAST
Identify Conserved Domains
Highlight Sequence Features
Find in this Sequence
Show in Genome Data Viewer

Protein database: NCBI



NAD(P)H dehydrogenase [quinone] 1 isoform a [Homo sapiens]

NCBI Reference Sequence: NP_000894.1

GenPept Identical Proteins FASTA

Analyze this sequence

Run BLAST

Identify Conserved Domains

Show in Genome Data Viewer

Protein 3D Structure

Structure of human NAD(P) H:quinone oxidoreductase in complex with N-(2-bromophenyl)pyrrolidine-1-sulfonamide

PDB: 6FY4
Source: Homo sapiens
Method: X-ray Diffraction
Resolution: 2.76 Å

See all 16 structures...

Articles about the NQO1 gene

NAD(P)H: quinone oxidoreductase 1 gene rs1800566 polymorphism increases the risk of cervical cancer in a Chinese Han samp [Medicine (Baltimore). 2020]

Association of NAD (P) H Quinine Oxidoreductase 1 rs1800566 Polymorphism with Bladder and Prostate Cancers - a Systematic Review ar [Klin Onkol. 2020]

Down-regulating NQO1 promotes cellular proliferation in K562 cells via elevating DNA synthesis. [Life Sci. 2020]

Dehydrogenase, Quinone 1

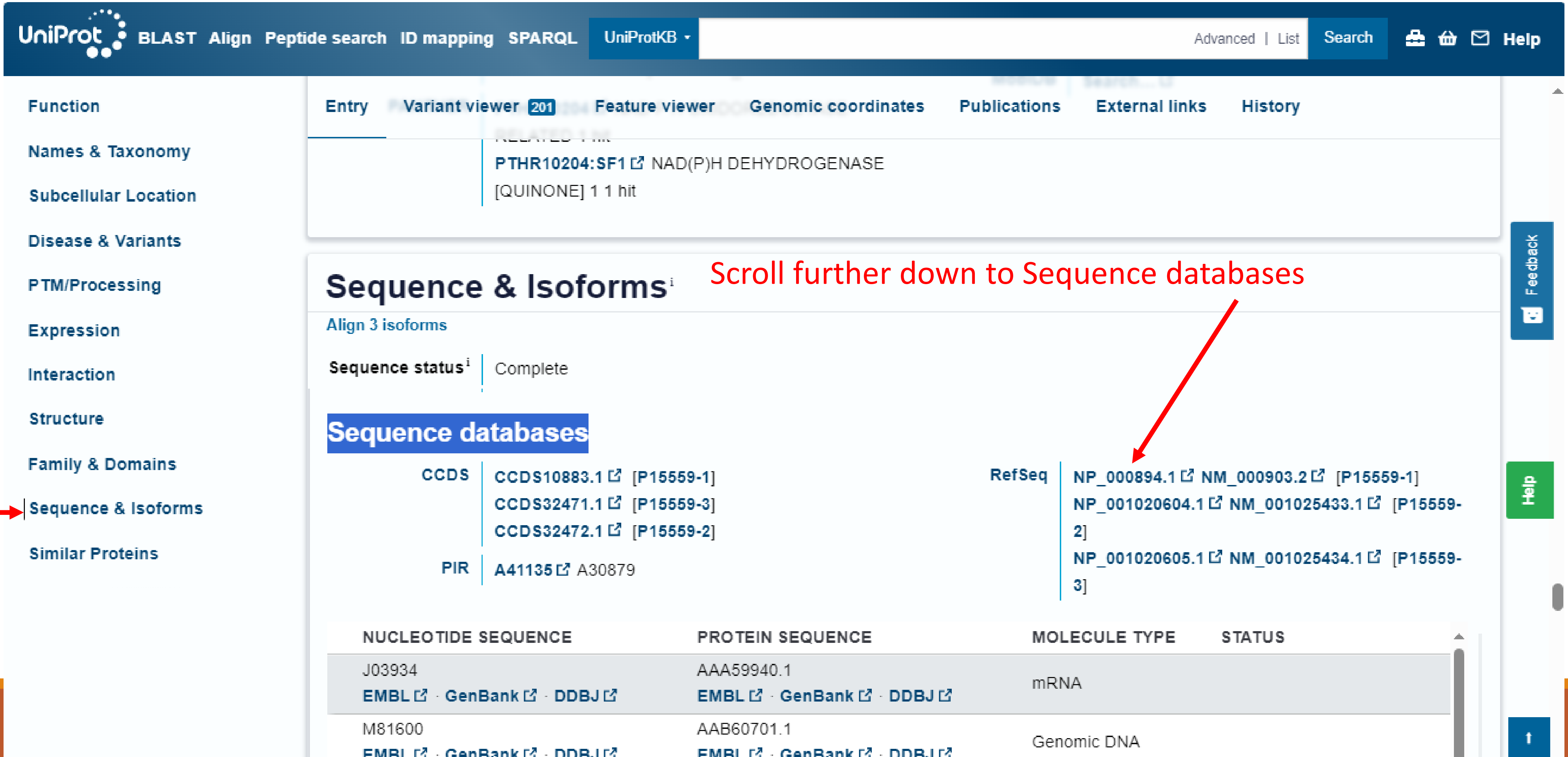
Try Protein database: NCBI

Find your sequence

-try direct search in NCBI

-**try link through Sequence databases → Refseq from Uniprot**

link through Sequence databases → Refseq from Uniprot



PTHR10204:SF1 NAD(P)H DEHYDROGENASE [QUINONE] 1 1 hit

Sequence & Isoforms

Align 3 isoforms

Scroll further down to Sequence databases

Sequence status Complete

Sequence databases

CCDS	CCDS10883.1 [P15559-1] CCDS32471.1 [P15559-3] CCDS32472.1 [P15559-2]	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]
PIR	A41135 A30879		

NUCLEOTIDE SEQUENCE	PROTEIN SEQUENCE	MOLECULE TYPE	STATUS
J03934 EMBL · GenBank · DDBJ	AAA59940.1 EMBL · GenBank · DDBJ	mRNA	
M81600 EMBL · GenBank · DDBJ	AAB60701.1 EMBL · GenBank · DDBJ	Genomic DNA	

„Protein bioinformatics I“

Retrieving protein sequences from databases

→ **Computing amino-acids compositions, molecular weight, isoelectric point, and other parameters**

Prediction of proteases cutting

Predicting elements of protein secondary structure, domains

Predicting 3-D structure and the domain organization of proteins

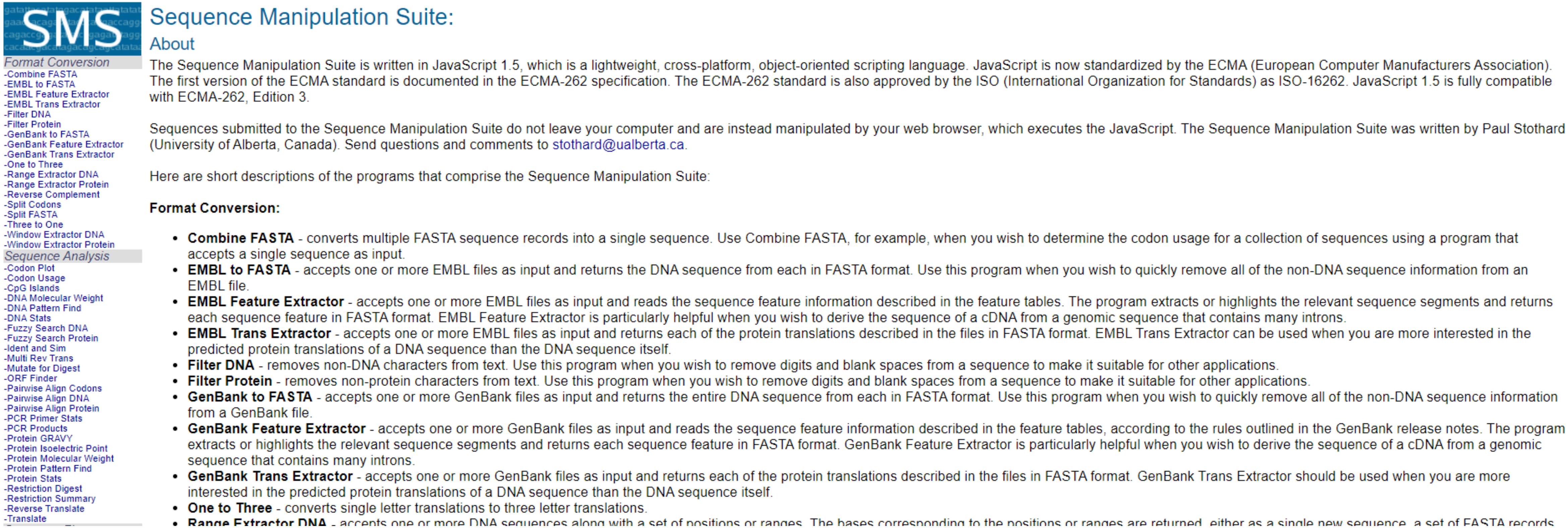
Finding all proteins that share a similar sequence and Classifying proteins into families

Finding evolutionary relationships between proteins, drawing proteins’ family trees

Computing the optimal alignment between two or more protein sequences

…

Protein Sequence Analysis



Sequence Manipulation Suite:

About

The Sequence Manipulation Suite is written in JavaScript 1.5, which is a lightweight, cross-platform, object-oriented scripting language. JavaScript is now standardized by the ECMA (European Computer Manufacturers Association). The first version of the ECMA standard is documented in the ECMA-262 specification. The ECMA-262 standard is also approved by the ISO (International Organization for Standards) as ISO-16262. JavaScript 1.5 is fully compatible with ECMA-262, Edition 3.

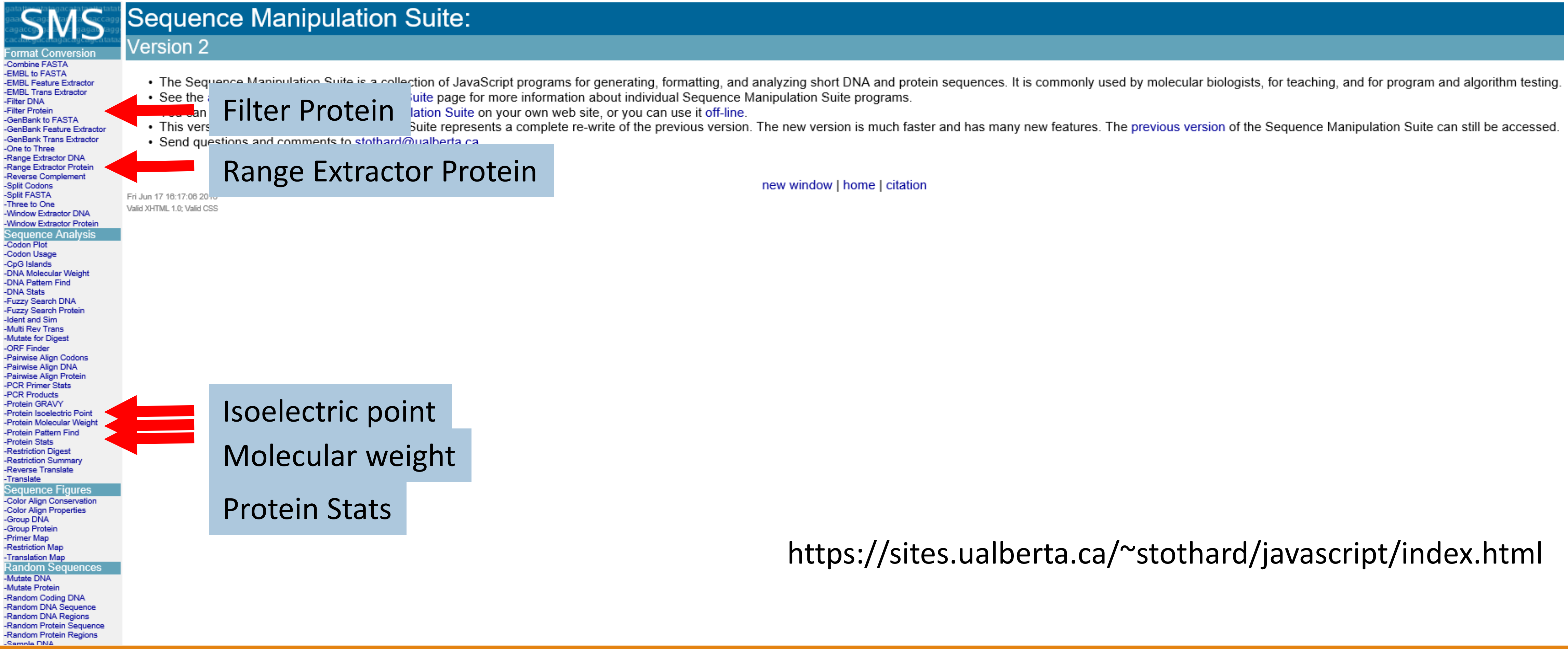
Sequences submitted to the Sequence Manipulation Suite do not leave your computer and are instead manipulated by your web browser, which executes the JavaScript. The Sequence Manipulation Suite was written by Paul Stothard (University of Alberta, Canada). Send questions and comments to stothard@ualberta.ca.

Here are short descriptions of the programs that comprise the Sequence Manipulation Suite:

Format Conversion:

- **Combine FASTA** - converts multiple FASTA sequence records into a single sequence. Use Combine FASTA, for example, when you wish to determine the codon usage for a collection of sequences using a program that accepts a single sequence as input.
- **EMBL to FASTA** - accepts one or more EMBL files as input and returns the DNA sequence from each in FASTA format. Use this program when you wish to quickly remove all of the non-DNA sequence information from an EMBL file.
- **EMBL Feature Extractor** - accepts one or more EMBL files as input and reads the sequence feature information described in the feature tables. The program extracts or highlights the relevant sequence segments and returns each sequence feature in FASTA format. EMBL Feature Extractor is particularly helpful when you wish to derive the sequence of a cDNA from a genomic sequence that contains many introns.
- **EMBL Trans Extractor** - accepts one or more EMBL files as input and returns each of the protein translations described in the files in FASTA format. EMBL Trans Extractor can be used when you are more interested in the predicted protein translations of a DNA sequence than the DNA sequence itself.
- **Filter DNA** - removes non-DNA characters from text. Use this program when you wish to remove digits and blank spaces from a sequence to make it suitable for other applications.
- **Filter Protein** - removes non-protein characters from text. Use this program when you wish to remove digits and blank spaces from a sequence to make it suitable for other applications.
- **GenBank to FASTA** - accepts one or more GenBank files as input and returns the entire DNA sequence from each in FASTA format. Use this program when you wish to quickly remove all of the non-DNA sequence information from a GenBank file.
- **GenBank Feature Extractor** - accepts one or more GenBank files as input and reads the sequence feature information described in the feature tables, according to the rules outlined in the GenBank release notes. The program extracts or highlights the relevant sequence segments and returns each sequence feature in FASTA format. GenBank Feature Extractor is particularly helpful when you wish to derive the sequence of a cDNA from a genomic sequence that contains many introns.
- **GenBank Trans Extractor** - accepts one or more GenBank files as input and returns each of the protein translations described in the files in FASTA format. GenBank Trans Extractor should be used when you are more interested in the predicted protein translations of a DNA sequence than the DNA sequence itself.
- **One to Three** - converts single letter translations to three letter translations.
- **Range Extractor DNA** - accepts one or more DNA sequences along with a set of positions or ranges. The bases corresponding to the positions or ranges are returned, either as a single new sequence, a set of FASTA records

Protein Sequence Analysis



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

Sequence Manipulation Suite:

Version 2

- The Sequence Manipulation Suite is a collection of JavaScript programs for generating, formatting, and analyzing short DNA and protein sequences. It is commonly used by molecular biologists, for teaching, and for program and algorithm testing.
- See the uite page for more information about individual Sequence Manipulation Suite programs.
- You can lation Suite on your own web site, or you can use it off-line.
- This vers Suite represents a complete re-write of the previous version. The new version is much faster and has many new features. The previous version of the Sequence Manipulation Suite can still be accessed.
- Send questions and comments to stothard@ualberta.ca

new window | home | citation

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

Filter Protein

Range Extractor Protein

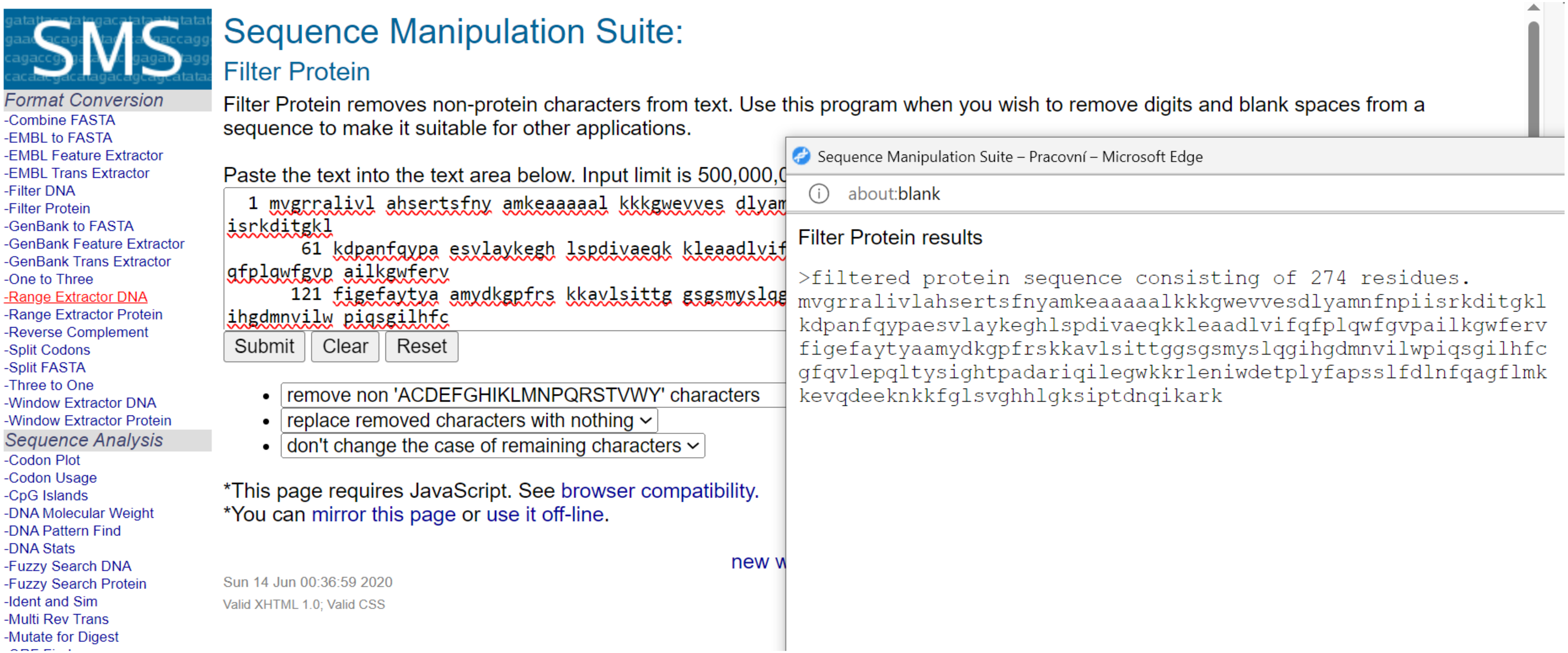
Isoelectric point

Molecular weight

Protein Stats

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

Protein Sequence Analysis



Sequence Manipulation Suite:

Filter Protein

Filter Protein removes non-protein characters from text. Use this program when you wish to remove digits and blank spaces from a sequence to make it suitable for other applications.

Paste the text into the text area below. Input limit is 500,000,(

```
  1 mvgrralivl ahsertsfny amkeaaaaal kkkgwevves dlyam
isrkditgkl
        61 kdpanfqypa esvlaykegh lspdivaeqk kleaadlvif
qfplqwfgvp ailkgwferv
       121 figefaytya amydkgpfrs kkavlsittg gsgsmyslqg
ihgdmnvilw piqsgilhfc
```

- remove non 'ACDEFGHIKLMNPQRSTVWY' characters
- replace removed characters with nothing
- don't change the case of remaining characters

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

Sun 14 Jun 00:36:59 2020
Valid XHTML 1.0; Valid CSS

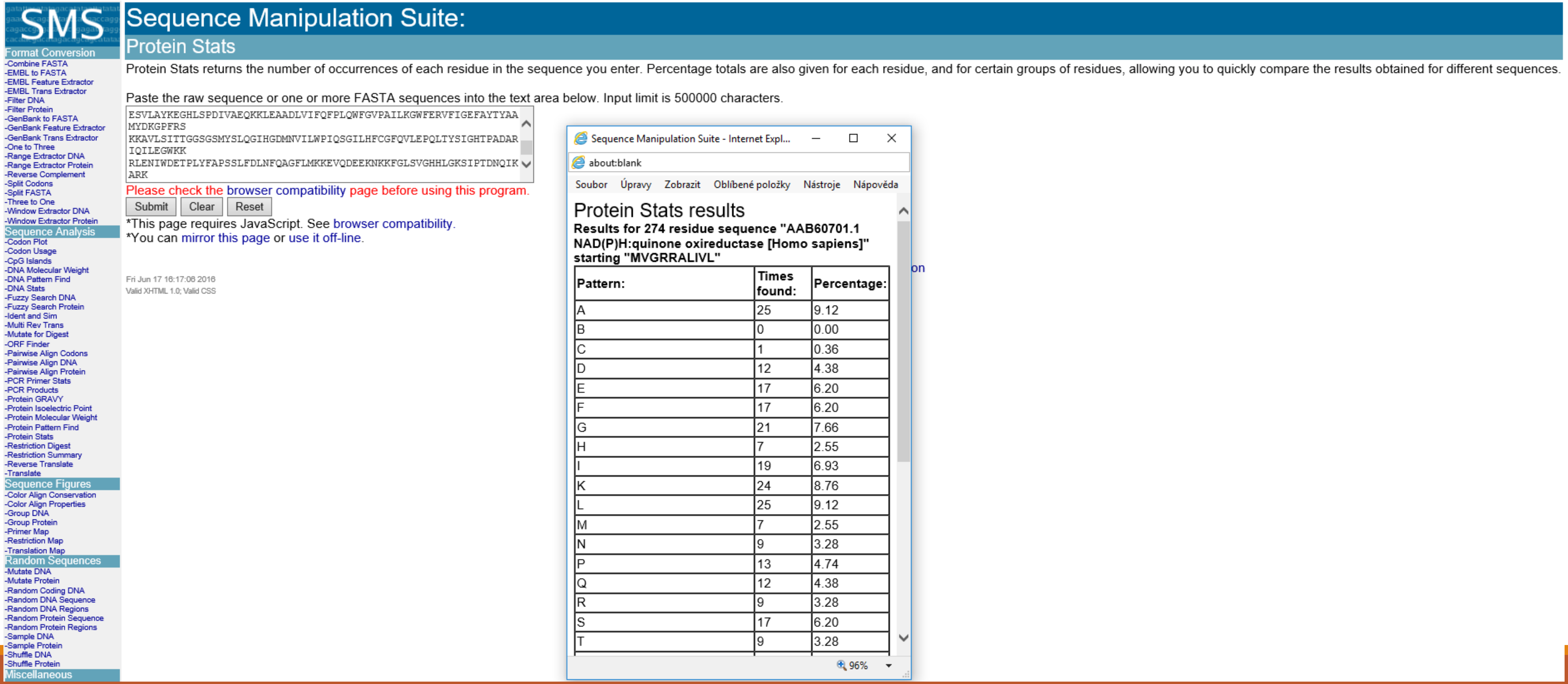
Sequence Manipulation Suite – Pracovní – Microsoft Edge

about:blank

Filter Protein results

```
>filtered protein sequence consisting of 274 residues.
mvgrralivlahsertsfnyamkeaaaaalkkkgwevvesdlyamnfnpiisrkditgkl
kdpanfqypaesvlaykeghlspdivaeqkkleaadlvifqfplqwfgvpailkgwferv
figefaytyaamydkgpfrskkavlsittggsgsmyslqgihgdmnvilwpiqsgilhfc
gfqvlepqltysightpadariqilegwkkrleniwdetplyfapsslfdlnfqagflmk
kevqdeeknkkfglsvghhlgksiptdnqikark
```

Protein Sequence Analysis



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.

*This page requires JavaScript. See browser compatibility.

*You can mirror this page or use it off-line.

Fri Jun 17 16:17:06 2016

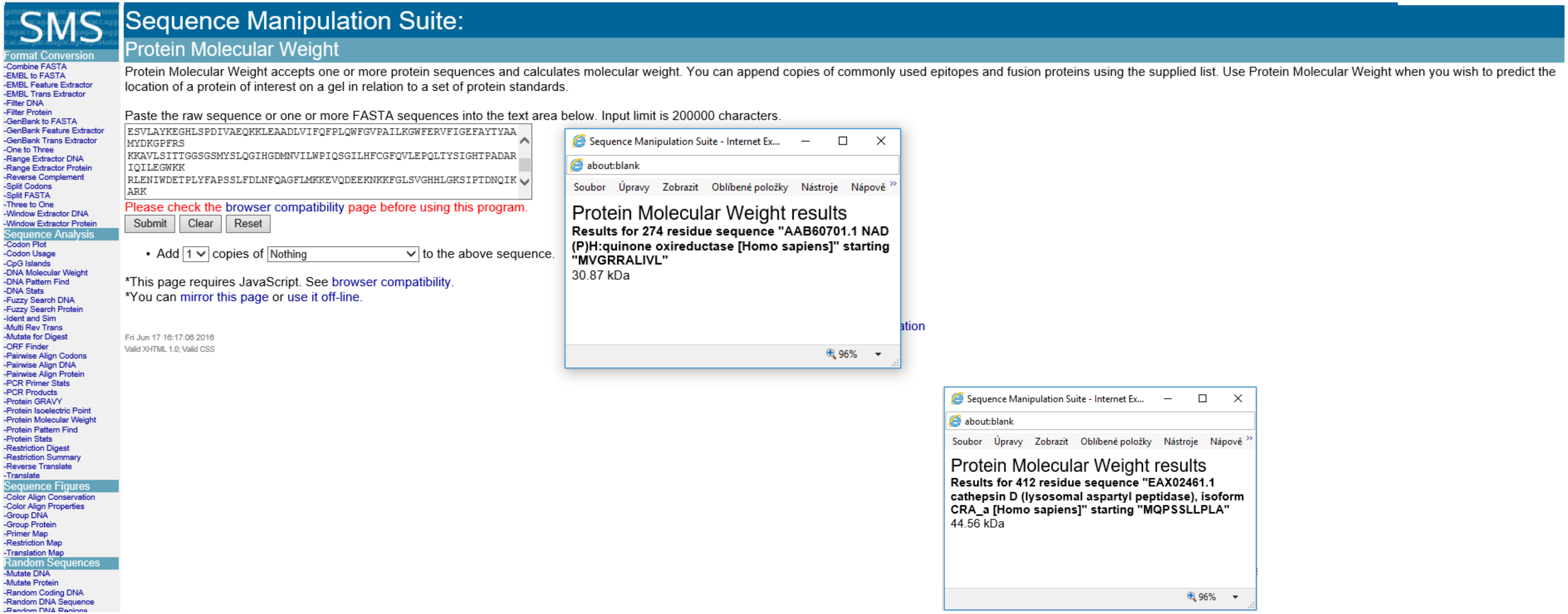
Valid XHTML 1.0; Valid CSS

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

Protein Sequence Analysis



Sequence Manipulation Suite:

Protein Molecular Weight

Protein Molecular Weight accepts one or more protein sequences and calculates molecular weight. You can append copies of commonly used epitopes and fusion proteins using the supplied list. Use Protein Molecular Weight when you wish to predict the location of a protein of interest on a gel in relation to a set of protein standards.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 200000 characters.

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

Please check the browser compatibility page before using this program.

- Add 1 copies of Nothing to the above sequence.

*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

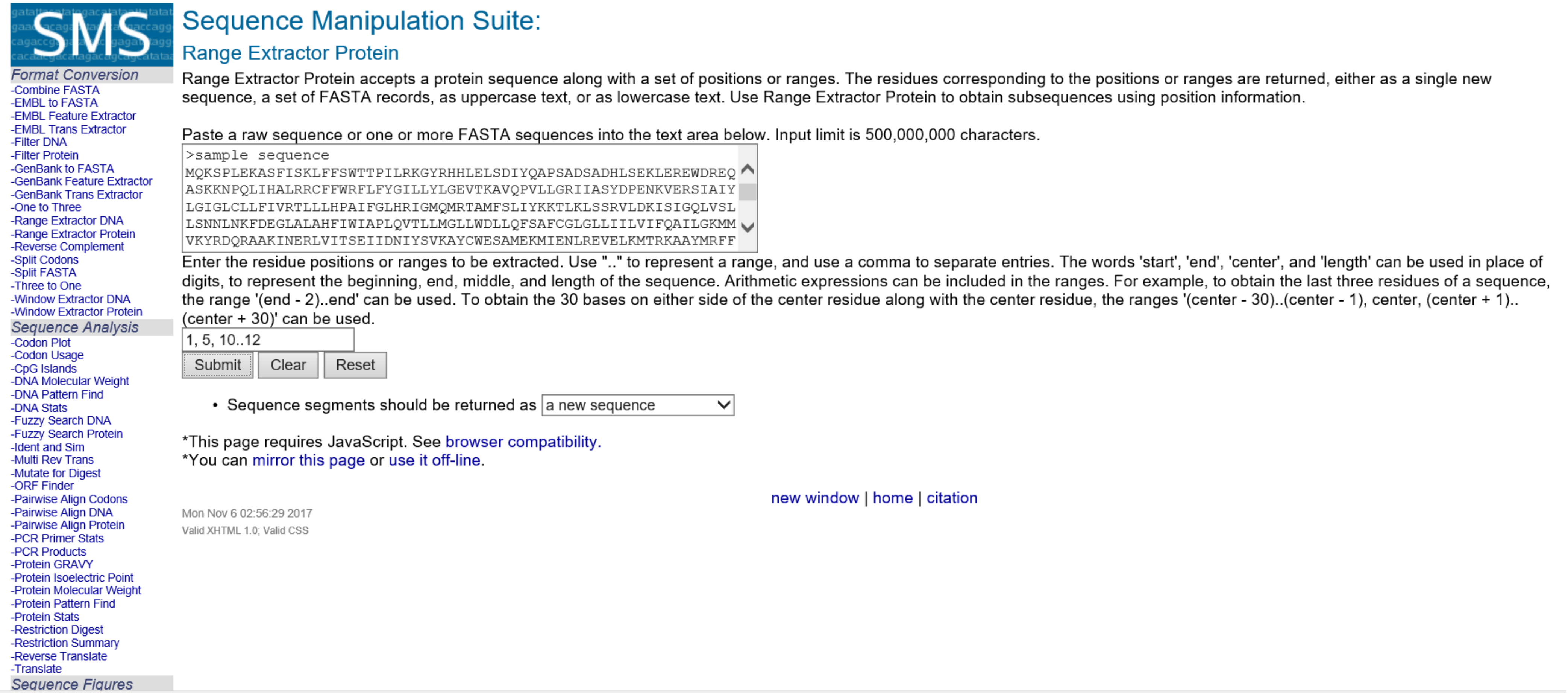
Protein Molecular Weight results

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

Protein Molecular Weight results

Results for 412 residue sequence "EAX02461.1 cathepsin D (lysosomal aspartyl peptidase), isoform CRA_a [Homo sapiens]" starting "MQPSSLLPLA"
44.56 kDa

Protein Sequence Analysis



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.

```
>sample sequence
MQKSPLEKASFISKLFFSWTTPILRKGYRHHLELSDIYQAPSADSADHLSEKLEREWDREQ
ASKKNPQLIHALRRCFFWRFLFYGILLYLGEVTKAVQPVLLGRIIASYDPENKVERSIAIY
LGIGLCLLFIVRTLLLHPAIFGLHRIGMQMRTAMFSLIYKKTLKLSSRVLDKISIGQLVSL
LSNNLNKFDEGLALAHFIWIAPLQVTLLMGLLWDLLQFSAFCGLGLLIILVIFQAILGKMM
VKYRDQRAAKINERLVITSEIIDNIYSVKAYCWESAMEKMIENLREVELKMTRKAAYMRFF
```

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 in place of digits, 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 of a sequence, the range '(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 - 1), center, (center + 1)..(center + 30)' can be used.

1, 5, 10..12

- Sequence segments should be returned as a new sequence

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

Mon Nov 6 02:56:29 2017
Valid XHTML 1.0; Valid CSS

Try SMS suite

Analyze your sequence

- how many cysteins are in your sequence?

HW2

3) Extract peptide from 10. - 50. amino acid.
4) Count the molecular weight of this (short) peptide.
5) How many cysteins are in your (whole) sequence?

„Protein bioinformatics II“

Retrieving protein sequences from databases

Computing amino-acids compositions, molecular weight, isoelectric point, and other parameters

→ **Prediction of proteases cutting**

Predicting elements of protein secondary structure, domains

Predicting 3-D structure and the domain organization of proteins

Finding all proteins that share a similar sequence and Classifying proteins into families

Finding evolutionary relationships between proteins, drawing proteins’ family trees

Computing the optimal alignment between two or more protein sequences

…

Prediction of proteases cutting

protease = enzyme that catalyzes proteolysis (*e.g.* digestion)

Examples: **trypsin -** digestive enzyme, present in duodenum)

- cleaves sequence „behind“ K(lysin) or R (arginin)

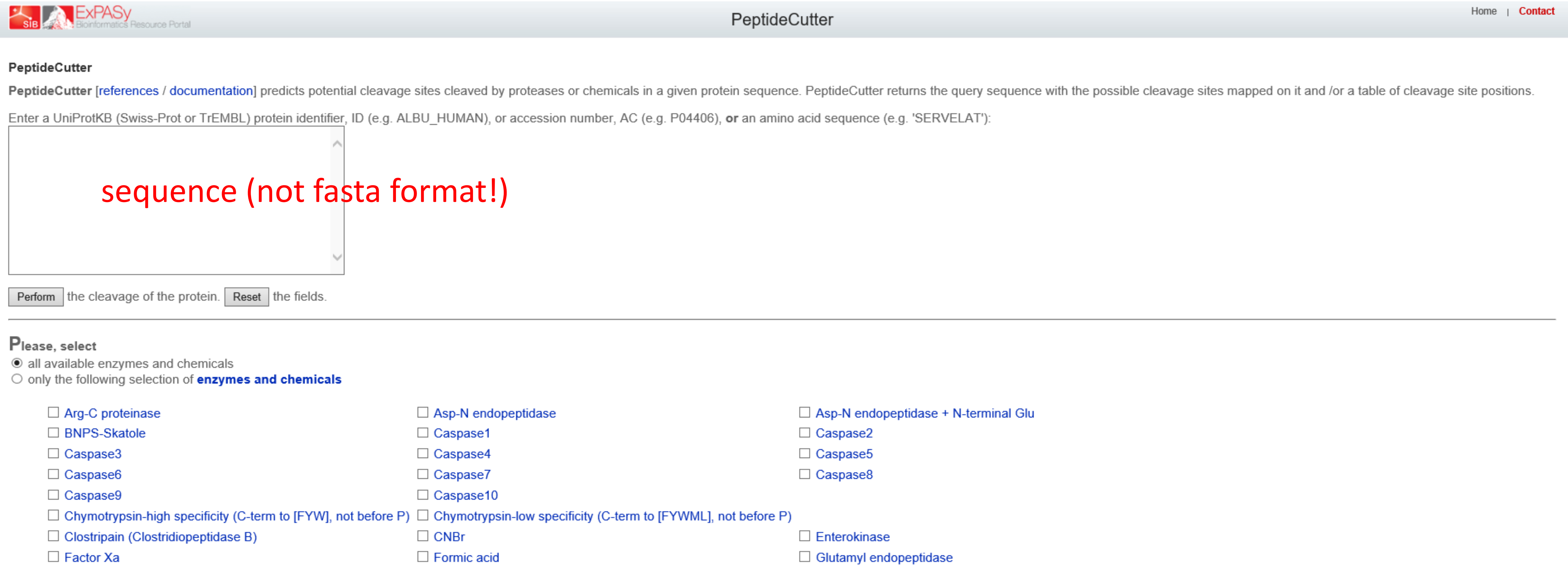
proteinase K - commonly used in molecular biology to digest protein and remove contamination from preparations of nucleic acid.

- cleaves ubiquitously

enterokinase - activation of zymogens (precursors of digestive enzymes like trysinogen)

- specific cleavage site (Asp-Asp-Asp-Asp-Lys)

Prediction of proteases cutting



PeptideCutter

PeptideCutter

PeptideCutter [references / documentation] predicts potential cleavage sites cleaved by proteases or chemicals in a given protein sequence. PeptideCutter returns the query sequence with the possible cleavage sites mapped on it and /or a table of cleavage site positions.

Enter a UniProtKB (Swiss-Prot or TrEMBL) protein identifier, ID (e.g. ALBU_HUMAN), or accession number, AC (e.g. P04406), **or** an amino acid sequence (e.g. 'SERVELAT'):

sequence (not fasta format!)

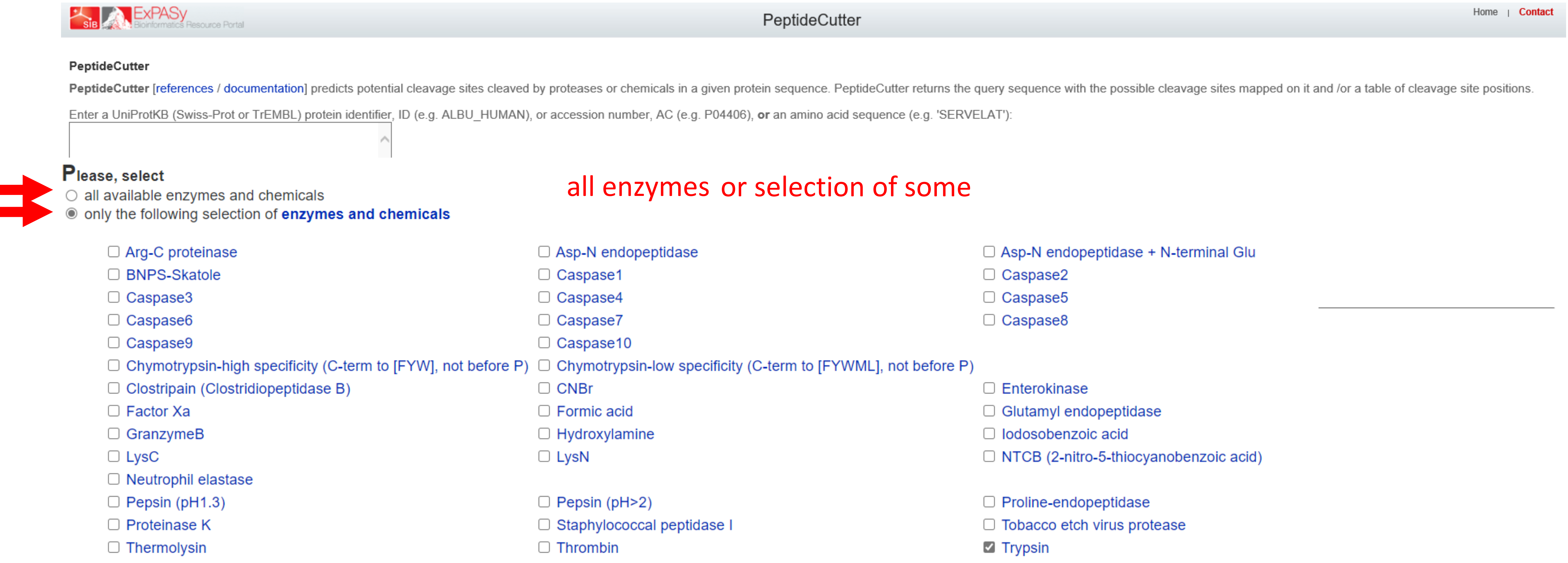
Perform the cleavage of the protein. Reset the fields.

Please, select

● all available enzymes and chemicals
○ only the following selection of **enzymes and chemicals**

- □ Arg-C proteinase
- □ Asp-N endopeptidase
- □ Asp-N endopeptidase + N-terminal Glu
- □ BNPS-Skatole
- □ Caspase1
- □ Caspase2
- □ Caspase3
- □ Caspase4
- □ Caspase5
- □ Caspase6
- □ Caspase7
- □ Caspase8
- □ Caspase9
- □ Caspase10
- □ Chymotrypsin-high specificity (C-term to [FYW], not before P)
- □ Chymotrypsin-low specificity (C-term to [FYWML], not before P)
- □ Clostripain (Clostridiopeptidase B)
- □ CNBr
- □ Enterokinase
- □ Factor Xa
- □ Formic acid
- □ Glutamyl endopeptidase

Prediction of proteases cutting



PeptideCutter

PeptideCutter

PeptideCutter [references / documentation] predicts potential cleavage sites cleaved by proteases or chemicals in a given protein sequence. PeptideCutter returns the query sequence with the possible cleavage sites mapped on it and /or a table of cleavage site positions.

Enter a UniProtKB (Swiss-Prot or TrEMBL) protein identifier, ID (e.g. ALBU_HUMAN), or accession number, AC (e.g. P04406), **or** an amino acid sequence (e.g. 'SERVELAT'):

Please, select

○ all available enzymes and chemicals

● only the following selection of **enzymes and chemicals**

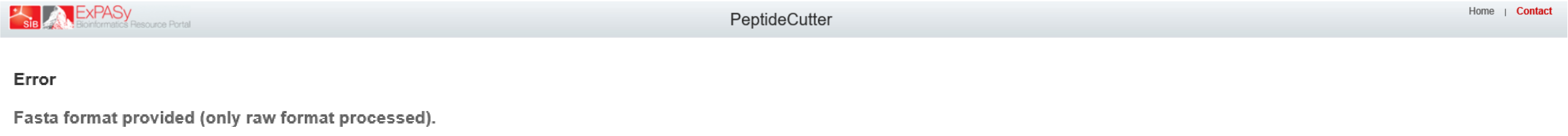
all enzymes or selection of some

□ Arg-C proteinase
□ BNPS-Skatole
□ Caspase3
□ Caspase6
□ Caspase9
□ Chymotrypsin-high specificity (C-term to [FYW], not before P)
□ Clostripain (Clostridiopeptidase B)
□ Factor Xa
□ GranzymeB
□ LysC
□ Neutrophil elastase
□ Pepsin (pH1.3)
□ Proteinase K
□ Thermolysin

□ Asp-N endopeptidase
□ Caspase1
□ Caspase4
□ Caspase7
□ Caspase10
□ Chymotrypsin-low specificity (C-term to [FYWML], not before P)
□ CNBr
□ Formic acid
□ Hydroxylamine
□ LysN
□ Pepsin (pH>2)
□ Staphylococcal peptidase I
□ Thrombin

□ Asp-N endopeptidase + N-terminal Glu
□ Caspase2
□ Caspase5
□ Caspase8
□ Enterokinase
□ Glutamyl endopeptidase
□ Iodosobenzoic acid
□ NTCB (2-nitro-5-thiocyanobenzoic acid)
□ Proline-endopeptidase
□ Tobacco etch virus protease
☑ Trypsin

Prediction of proteases cutting



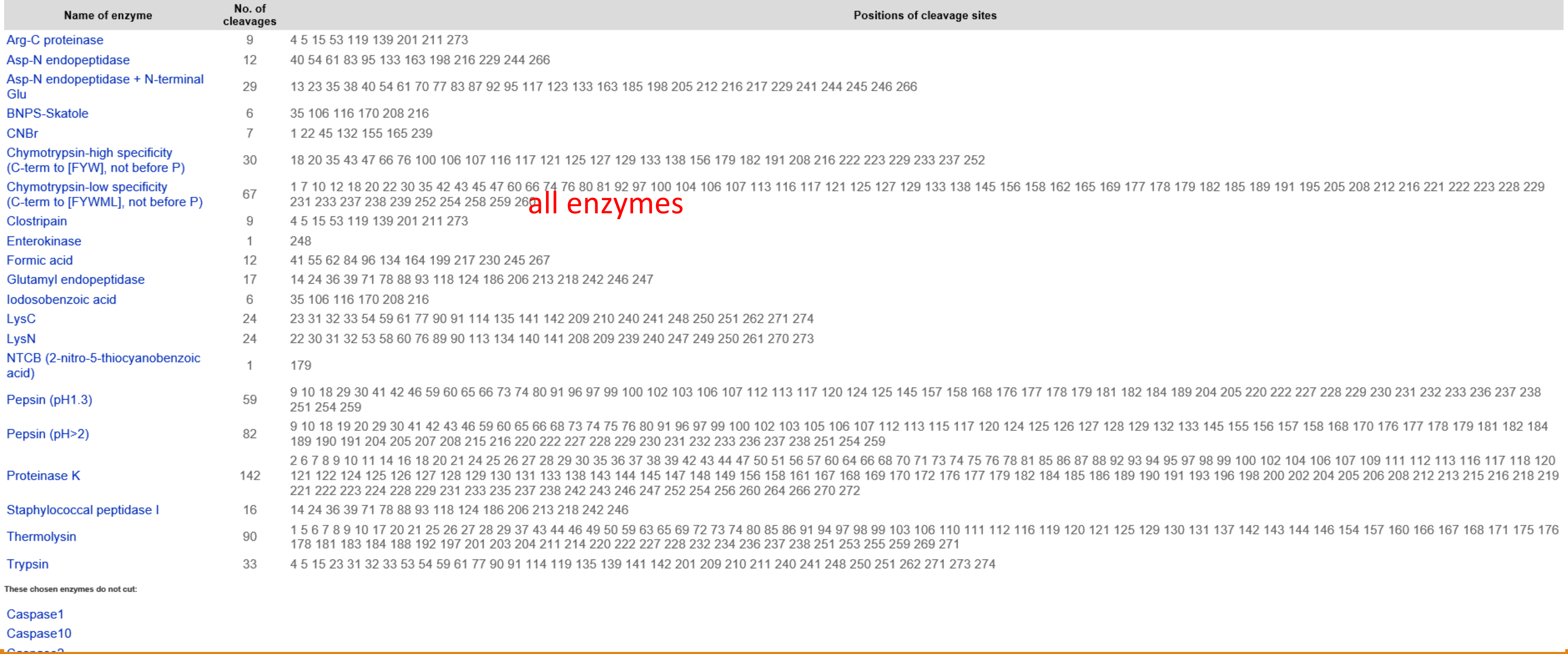
PeptideCutter

Error

Fasta format provided (only raw format processed).

sequence (not fasta format!)

Prediction of proteases cutting



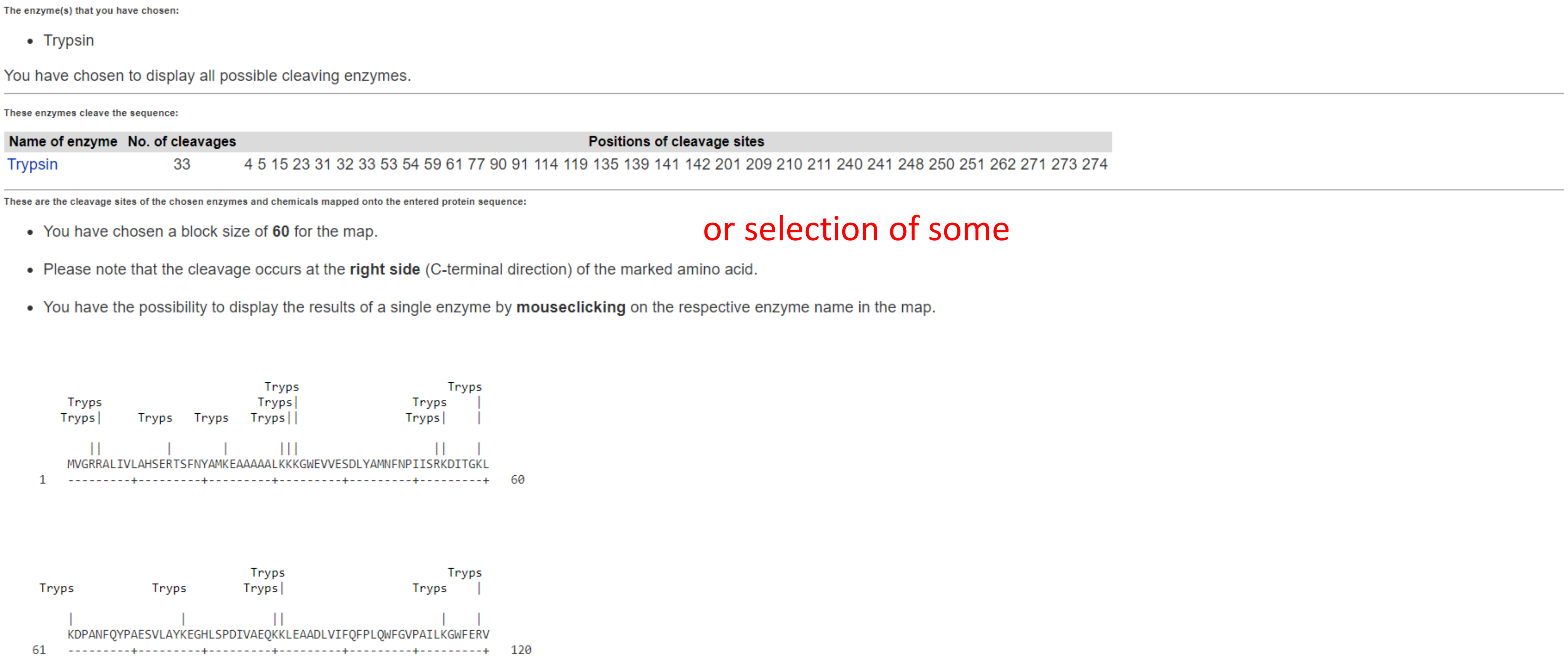
Name of enzyme	No. of cleavages	Positions of cleavage sites
Arg-C proteinase	9	4 5 15 53 119 139 201 211 273
Asp-N endopeptidase	12	40 54 61 83 95 133 163 198 216 229 244 266
Asp-N endopeptidase + N-terminal Glu	29	13 23 35 38 40 54 61 70 77 83 87 92 95 117 123 133 163 185 198 205 212 216 217 229 241 244 245 246 266
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 26
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

all enzymes

These chosen enzymes do not cut:

Caspase1
Caspase10
Caspase3

Prediction of proteases cutting



The enzyme(s) that you have chosen:

- Trypsin

You have chosen to display all possible cleaving enzymes.

These enzymes cleave the sequence:

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

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.
- Please note that the cleavage occurs at the **right side** (C-terminal direction) of the marked amino acid.
- You have the possibility to display the results of a single enzyme by **mouseclicking** on the respective enzyme name in the map.

or selection of some

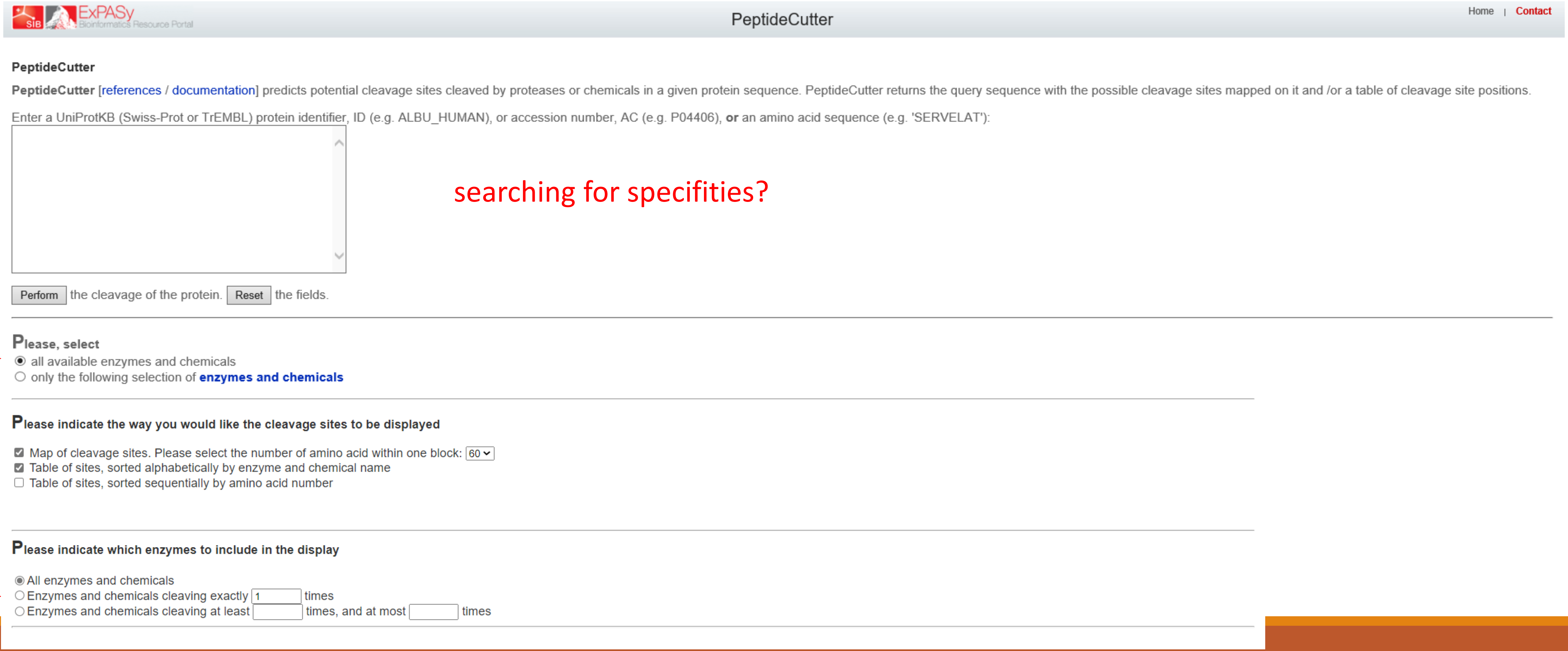
```
                                 Tryps                     Tryps
     Tryps                      Tryps|                Tryps    |
    Tryps|     Tryps   Tryps   Tryps||               Tryps|    |

        ||         |       |       |||                   ||    |
     MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL
 1   ---------+---------+---------+---------+---------+---------+    60


                               Tryps                       Tryps
 Tryps           Tryps        Tryps|                  Tryps    |

     |               |            ||                      |    |
     KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV
61   ---------+---------+---------+---------+---------+---------+    120
```

Prediction of proteases cutting



PeptideCutter

PeptideCutter

PeptideCutter [references / documentation] predicts potential cleavage sites cleaved by proteases or chemicals in a given protein sequence. PeptideCutter returns the query sequence with the possible cleavage sites mapped on it and /or a table of cleavage site positions.

Enter a UniProtKB (Swiss-Prot or TrEMBL) protein identifier, ID (e.g. ALBU_HUMAN), or accession number, AC (e.g. P04406), **or** an amino acid sequence (e.g. 'SERVELAT'):

searching for specifities?

Perform the cleavage of the protein. Reset the fields.

Please, select

● all available enzymes and chemicals
○ only the following selection of **enzymes and chemicals**

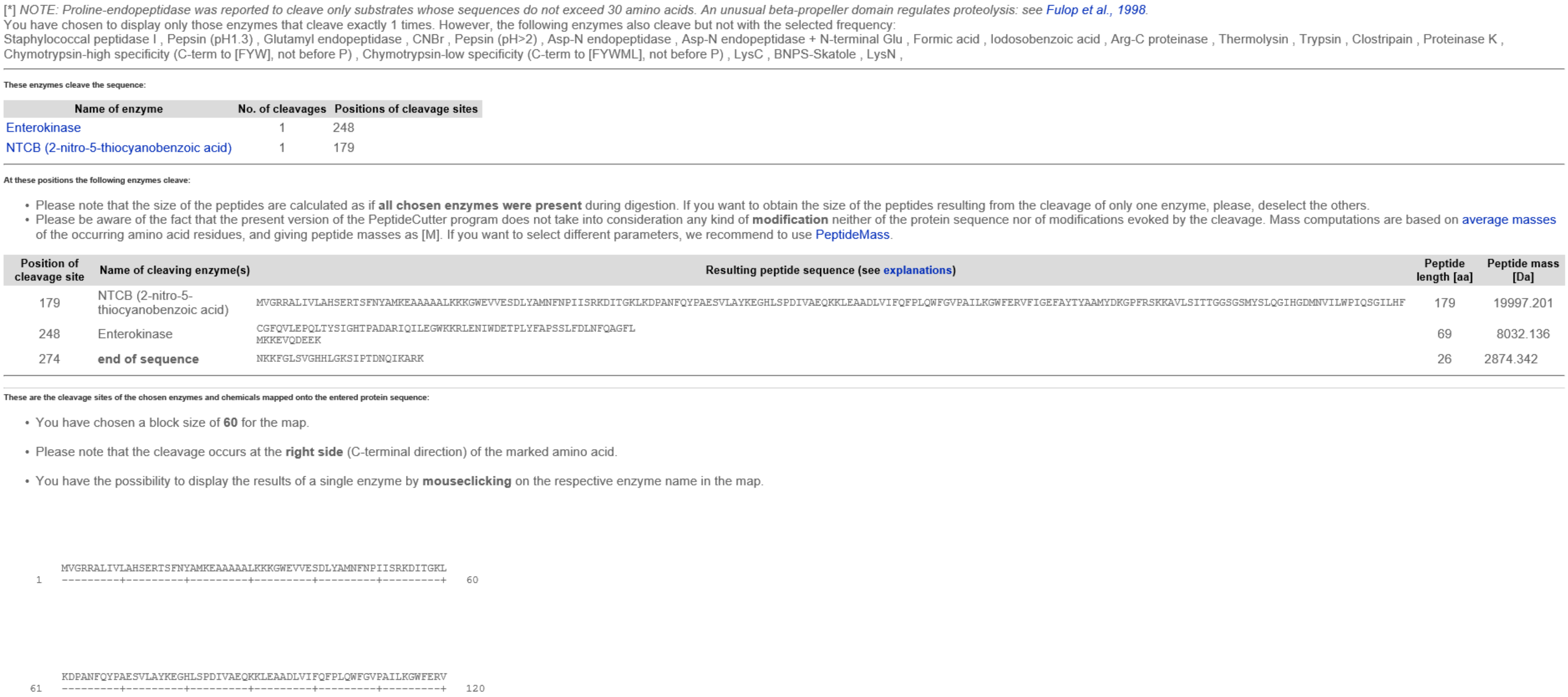
Please indicate the way you would like the cleavage sites to be displayed

☑ Map of cleavage sites. Please select the number of amino acid within one block: 60
☑ Table of sites, sorted alphabetically by enzyme and chemical name
□ Table of sites, sorted sequentially by amino acid number

Please indicate which enzymes to include in the display

● All enzymes and chemicals
○ Enzymes and chemicals cleaving exactly 1 times
○ Enzymes and chemicals cleaving at least ____ times, and at most ____ times

Prediction of proteases cutting



[*] *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

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 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 computations are based on average masses 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)	Peptide length [aa]	Peptide mass [Da]
179	NTCB (2-nitro-5-thiocyanobenzoic acid)	MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKLKDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRSKKAVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHF	179	19997.201
248	Enterokinase	CGFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFL MKKEVQDEEK	69	8032.136
274	**end of sequence**	NKKFGLSVGHHLGKSIPTDNQIKARK	26	2874.342

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

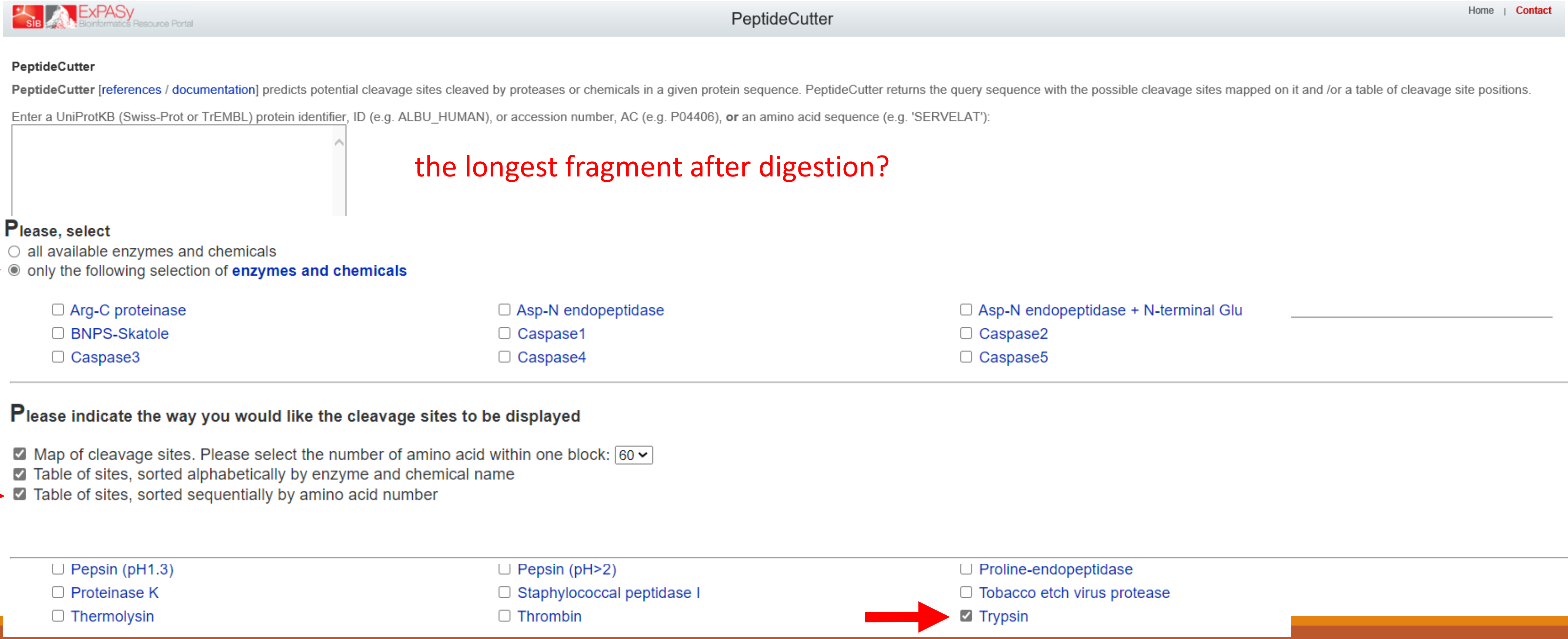
- You have chosen a block size of **60** for the map.
- Please note that the cleavage occurs at the **right side** (C-terminal direction) of the marked amino acid.
- You have the possibility to display the results of a single enzyme by **mouseclicking** on the respective enzyme name in the map.

```
      MVGRRALIVLAHSERTSFNYAMKEAAAAALKKKGWEVVESDLYAMNFNPIISRKDITGKL
 1    ---------+---------+---------+---------+---------+---------+   60



      KDPANFQYPAESVLAYKEGHLSPDIVAEQKKLEAADLVIFQFPLQWFGVPAILKGWFERV
61    ---------+---------+---------+---------+---------+---------+   120
```

Prediction of proteases cutting



PeptideCutter

PeptideCutter

PeptideCutter [references / documentation] predicts potential cleavage sites cleaved by proteases or chemicals in a given protein sequence. PeptideCutter returns the query sequence with the possible cleavage sites mapped on it and /or a table of cleavage site positions.

Enter a UniProtKB (Swiss-Prot or TrEMBL) protein identifier, ID (e.g. ALBU_HUMAN), or accession number, AC (e.g. P04406), **or** an amino acid sequence (e.g. 'SERVELAT'):

the longest fragment after digestion?

Please, select

○ all available enzymes and chemicals
● only the following selection of **enzymes and chemicals**

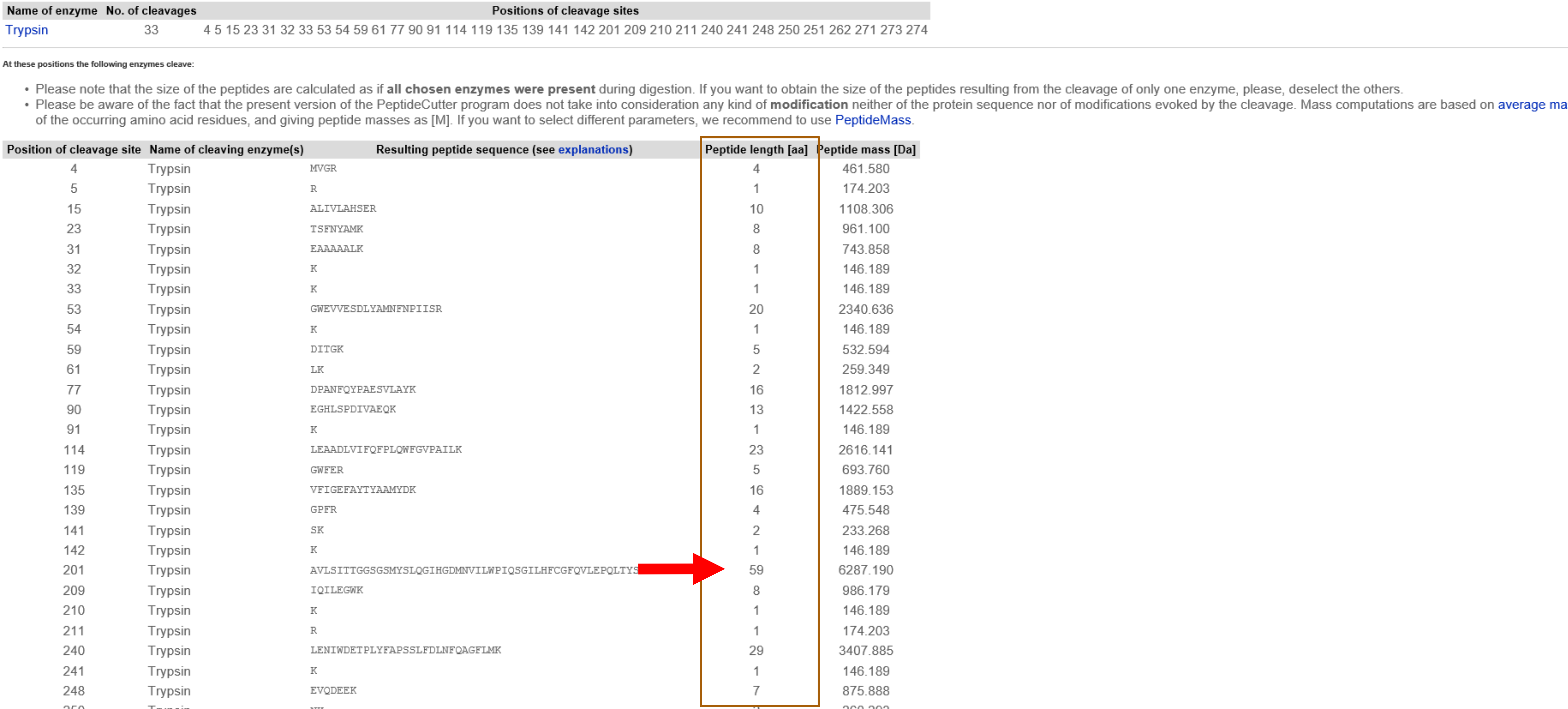
□ Arg-C proteinase
□ BNPS-Skatole
□ Caspase3
□ Asp-N endopeptidase
□ Caspase1
□ Caspase4
□ Asp-N endopeptidase + N-terminal Glu
□ Caspase2
□ Caspase5

Please indicate the way you would like the cleavage sites to be displayed

☑ Map of cleavage sites. Please select the number of amino acid within one block: 60
☑ Table of sites, sorted alphabetically by enzyme and chemical name
☑ Table of sites, sorted sequentially by amino acid number

□ Pepsin (pH1.3)
□ Proteinase K
□ Thermolysin
□ Pepsin (pH>2)
□ Staphylococcal peptidase I
□ Thrombin
□ Proline-endopeptidase
□ Tobacco etch virus protease
☑ Trypsin

Prediction of proteases cutting



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 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 computations are based on average ma 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)	Peptide length [aa]	Peptide mass [Da]
4	Trypsin	MVGR	4	461.580
5	Trypsin	R	1	174.203
15	Trypsin	ALIVLAHSER	10	1108.306
23	Trypsin	TSFNYAMK	8	961.100
31	Trypsin	EAAAAALK	8	743.858
32	Trypsin	K	1	146.189
33	Trypsin	K	1	146.189
53	Trypsin	GWEVVESDLYAMNFNPIISR	20	2340.636
54	Trypsin	K	1	146.189
59	Trypsin	DITGK	5	532.594
61	Trypsin	LK	2	259.349
77	Trypsin	DPANFQYPAESVLAYK	16	1812.997
90	Trypsin	EGHLSPDIVAEQK	13	1422.558
91	Trypsin	K	1	146.189
114	Trypsin	LEAADLVIFQFPLQWFGVPAILK	23	2616.141
119	Trypsin	GWFER	5	693.760
135	Trypsin	VFIGEFAYTYAAMYDK	16	1889.153
139	Trypsin	GPFR	4	475.548
141	Trypsin	SK	2	233.268
142	Trypsin	K	1	146.189
201	Trypsin	AVLSITTGGSGSMYSLQGIHGDMNVILWPIQSGILHFCGFQVLEPQLTYS	59	6287.190
209	Trypsin	IQILEGWK	8	986.179
210	Trypsin	K	1	146.189
211	Trypsin	R	1	174.203
240	Trypsin	LENIWDETPLYFAPSSLFDLNFQAGFLMK	29	3407.885
241	Trypsin	K	1	146.189
248	Trypsin	EVQDEEK	7	875.888

Try PeptideCutter

Analyze your sequence

How many times is your sequence cut by trypsin (HW3)

Is there any enzyme that cuts just once?

How long is the longest product after trypsin digest?

„Protein bioinformatics I“

Retrieving protein sequences from databases

Computing amino-acids compositions, molecular weight, isoelectric point, and other parameters

Prediction of proteases cutting

Predicting elements of protein secondary structure, domains

Predicting 3-D structure and the domain organization of proteins

Finding all proteins that share a similar sequence and Classifying proteins into families

Finding evolutionary relationships between proteins, drawing proteins’ family trees

Computing the optimal alignment between two or more protein sequences

…

Homework 2

1) find both (Uniprot/NCBI) accession numbers for reference sequence (isoform 1)

2) download your sequence in FASTA format

3) Extract peptide form positions 10 to 50

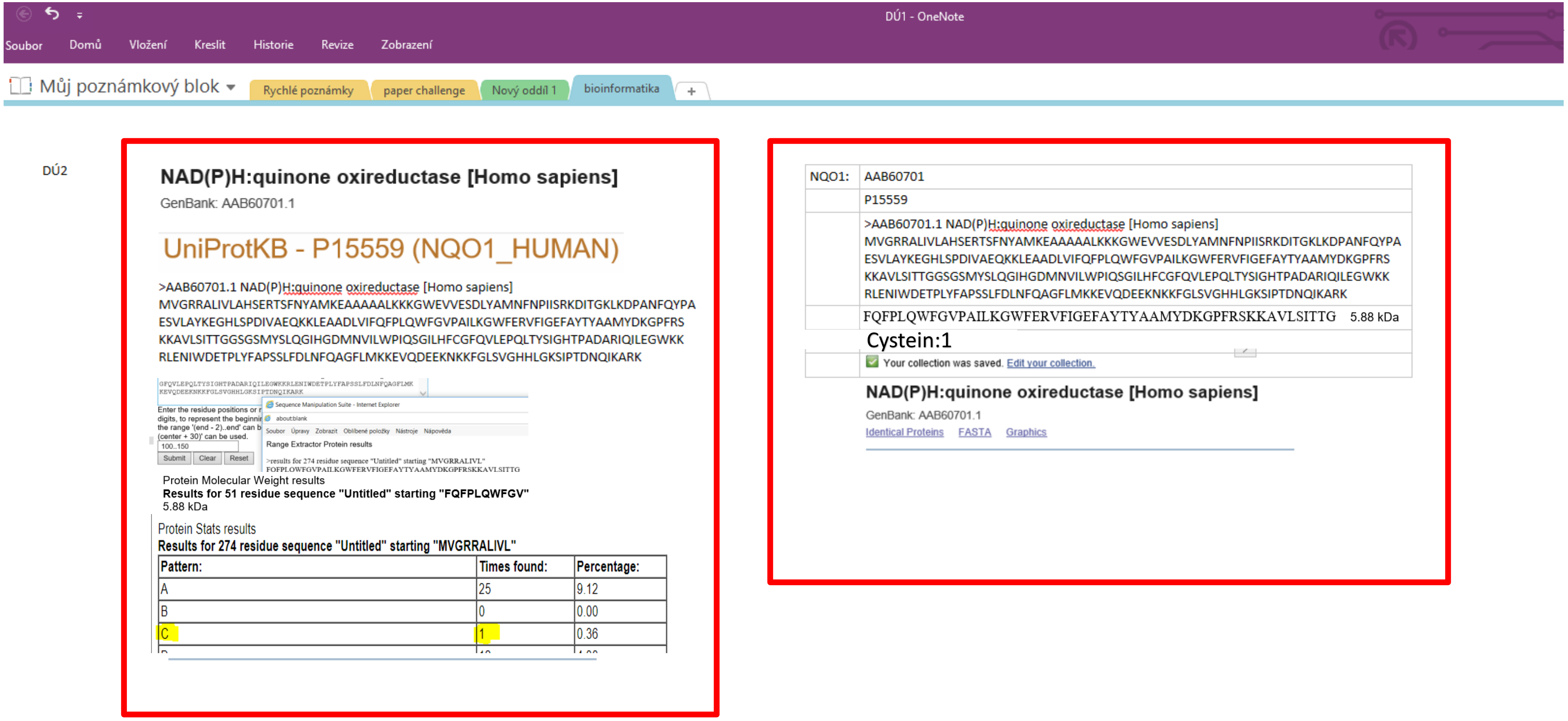
4) Predict molecular weight of this peptide

5) How many cysteins are in your sequence?

E.g use „výstřižky“ „snipping tool“

- Compile in „one note“ (or word, or pdf)

Homework 2: examples



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

```
GFQVLEPQLTYSIGHTPADARIQILEGWKKRLENIWDETPLYFAPSSLFDLNFQAGFLMK
KEVQDEEKNKKFGLSVGHHLGKSIPTDNQIKARK
```

Enter the residue positions or r
digits, to represent the beginnir
the range '(end - 2)..end' can b
(center + 30)' can be used.

100..150

Submit Clear Reset

Sequence Manipulation Suite - Internet Explorer

about:blank

Range Extractor Protein results

>results for 274 residue sequence "Untitled" starting "MVGRRALIVL"
FQFPLQWFGVPAILKGWFERVFIGEFAYTYAAMYDKGPFRSKKAVLSITTG

Protein Molecular Weight results

Results for 51 residue sequence "Untitled" starting "FQFPLQWFGV"

5.88 kDa

Protein Stats results

Results for 274 residue sequence "Untitled" starting "MVGRRALIVL"

Pattern:	Times found:	Percentage:
A	25	9.12
B	0	0.00
C	1	0.36

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

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

GenBank: AAB60701.1

Identical Proteins FASTA Graphics