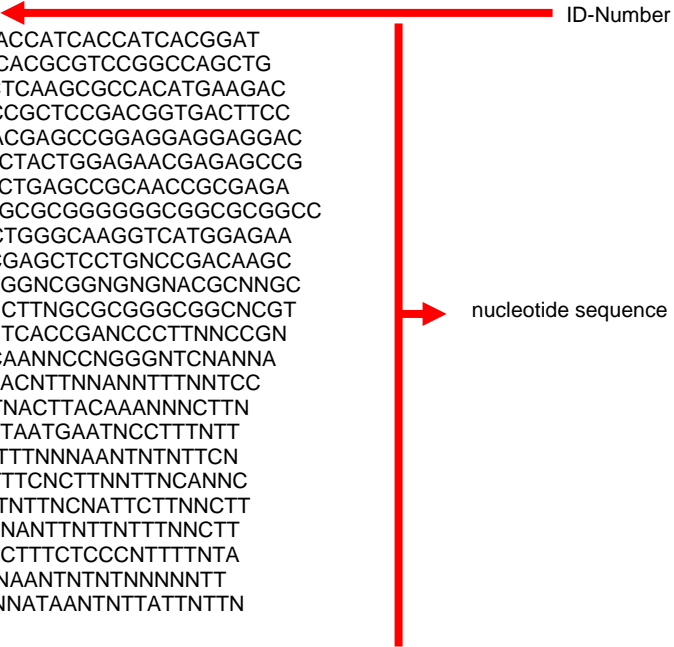


Example for the antigen scenario showing the outline of data and involved steps:

1. Data of the SEREX experiments will be sent to the ACGT database in form of an ASCII File. In Step 1 only the ID numbers and the nucleotide sequences of positive clones are included.

ASCII File:

```
MPMGp800A01579 MPMGp800A01579Q231 ← ID-Number
ANTTAGAGAAATTACTATGAGAGGATCGCATCACCATCACCATCACGGAT
CCTATTTAGGTGACACTATAGAATCGTCGACCCACGGTCCGGCCAGCTG
TGCGACACGCGTCTCGCAGGCCAGCAAGCTCAAGCGCCACATGAAGAC
GCACATGCACAAGGCCGGCTCGCTGGCCGGCCGCTCCGACGGTGACTTCC
GCCACCACGAGAGCGACCCGCTCGCTGGGCCACGAGCCGGAGGAGGAGAC
GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCTGCTACTGGAGAACGAGAGCCG
GCCCGAGTCGAGCTTCAGCATGGACTCGGAGCTGAGCCGCAACCGCGAGA
ACGGCGGTGGTGGGGTCCCGGGGCGCGGGGGGCGGCGCGGCC
AAGGCGCTGGCTGACGAGAAGGCGCTGGTGTGGGCAAGGTCATGGAGAA
CGTGGGCCTAGGCGCACTGCCGCANTACGGCGAGCTCCTGNCCGACAAGC
ANAAGCGCGGCCTTCCGAANCCTGCGGCGGNCGGNGNGNACGCNNGC
NACTACNACTACGCGGGCGGCTGNNGGNACNCTTNGCGCGGGCGGCNCGT
TCAACGGNCGCGNGNGCGGCTTCTCNCCANTACCGANCCCTTNNCCGN
GTTNTTCCNGCGANAATNCTNNNNCNCTNCAANNCCNGGGNTCNANNA
CTCCGCAATNNNCNTCAANGNTGNAAAAANACNTNANNNTTNNNTCC
NGCNCTNNTTANTNCCNATTAATAAANNTATNACTTACAAANNCTTN
GTTGNGCNTANACTCTCNTTTGTNNTTTNATTAATGAATNCCTTTNTT
GGNCTTTNCAAAAAATNNNNNTNNTTTTTTTTTTTNNNAANTNTNTTCN
GNAATATNCTTCTCCAAAAATNNTGAATCNCTTTCNCTTNNNTNCANN
CCCCCGGNATCCTNNCNTTAAAAAATNTTTNTTNCNATTCTTNNCTT
TNTNATNNTCTNTTNNNAANNTATNANCTTNTNANTTNTTNTTNNCTT
TNTNTTNTGNTTGTNNANNTTCNTNNAATNTCTTCTCCCNTTTNTA
ATTTTTNAAATNATTTTCNNTANTTATTTTTANNAANTNTNTNNNNNTT
TTNTNNCCTTNTTNNNNNTTNTNNAACNTNNAANTNTTATNTTN
```



There are few steps that will be done with this nucleotide sequence. The first one is to find a protein (1st protein), the second one is to find a human gene for this nucleotide sequence. For this human gene the corresponding protein (2nd protein) will be looked for. By comparing these two proteins this nucleotide sequence can be distinguished in inframe and not-inframe proteins. Inframe proteins are characterized by the fact that protein 1 and 2 are identical and the protein is expressed in humans. If the sequence of both proteins are not matching this is called not-inframe. Such proteins are unknown in humans. Nevertheless these proteins have shown an antibody reaction. This might be true because of a similar tertiary structure of the not-inframe protein and a really existing human protein, with the same tertiary structure. Those proteins are mimotops. There are no existing software to find the corresponding human proteins. In the further analysis this non-inframe proteins will be used, but not further analysed. Analysis of these not-inframe proteins can only be done in future, when software tools are available, that can generate and compare tertiary structure of proteins according to their sequence.

Here are the following steps of step 1 of the antigen scenario:

2. Nucleotide sequences will be given to the translation tool of ExPasy (<http://www.expasy.org/tools/dna.html> – DNA → Protein) and translated into six possible frames (3 from 3' → 5' and 3 from 5' → 3'). The output format “Verbose” is used.

- The possible frame is characterized by a repeated **Histidin sequence** and a **TISS** sequence as a result of the cloning of the plasmid for the bacterium used in the experiment.
- The nucleotide sequence of the positive clone will be given to the NCBI webtool: (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi?CMD=Web&LAYOUT=TwoWindows&AUTO_FORMAT=Semiauto&ALIGNMENTS=50&ALIGNMENT_VIEW=Pairwise&CLIENT=web&DATABASE=nr&DESCRIPTIONS=100&ENTREZ_QUERY=%28none%29&EXPECT=10&FILTER=L&FORMAT_OBJECT=Alignment&FORMAT_TYPE=HTML&NCBI_GI=on&PAGE=Nucleotides&PROGRAM=blastn&SERVICE=plain&SET_DEFAULTS.x=34&SET_DEFAULTS.y=8&SHOW_OVERVIEW=on&END_OF_HTTPGET=Yes&SHOW_LINKOUT=yes&GET_SEQUENCE=yes)
A request ID will be given and search started with "format"

NCBI Blast - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi?CMD=Web&LAYOUT=TwoWindows&AUTO_FORMAT=Semiauto&ALIGNMENTS=50&ALIGNMENT_VIEW=Pairwise&CLIENT=web&DATABASE=nr&DESCRIPTIONS=100&ENTREZ_QUERY=%28none%29&EXPECT=10&FILTER=L&FORMAT_OBJECT=Alignment&FORMAT_TYPE=HTML&NCBI_GI=on&PAGE=Nucleotides&PROGRAM=blastn&SERVICE=plain&SET_DEFAULTS.x=34&SET_DEFAULTS.y=8&SHOW_OVERVIEW=on&END_OF_HTTPGET=Yes&SHOW_LINKOUT=yes&GET_SEQUENCE=yes

NCBI Nucleotide Protein Translations Retrieve results for an RID

nucleotide-nucleotide **BLAST**

Search: MPMGp800A01579 MPMGp800A01579Q231
 ANTTAGAGAAAATTACTATGAGAGGATCGCATCACCATCACCATCAGGAT
 CCTATTTAGGTGACACTATAGAAATCGTCGACCCACGGTCGGCCAGCTG
 TCGGACCAAGCGTGTCTCGCAGGCCAGCAAGCTCAAGCGCCACATGAAAGAC
 GCACATGCACAAAGCCGGCTCGCTGGCCGGCCGCTCCGACGGTGACTTCC
 GCCACCACGAGAGCGACCCGCTCGCTGGCCACGAGCCGAGGAGGAGGAC

Set subsequence From: To:

Choose database: nr

Now: **BLAST!** or **Reset query** **Reset all**

Options for advanced blasting

Limit by entrez query or select from: All organisms

Choose filter: Low complexity Human repeats Mask for lookup table only Mask lower case

Expect: 10

Word Size: 11

As a result of Now: Blast you will get the ID for search:

NCBI Blast - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi

NCBI Nucleotide Protein Translations Retrieve results for an RID

formatting **BLAST**

WARNING: 5 illegal characters were removed: 4 Ps, 1 Q

Your request has been successfully submitted and put into the Blast Queue.

Query = (1208 letters)

The request ID is 1153932842-17707-84449463301.BLASTQ4

Format! or **Reset all**

The results are estimated to be ready in 10 seconds but may be done sooner.

Please press "FORMAT!" when you wish to check your results. You may change the formatting options for your result via the form below and press "FORMAT!" again. You may also request results of a different search by entering any other valid request ID to see other recent jobs.

Format

Show: Graphical Overview Linkout Sequence Retrieval NCBI-gi Alignment in HTML format

CDS feature

Masking Character: Lower Case Masking Color: Grey

After Format! You will get:

NCBI *results of BLAST*

BLASTN 2.2.14 [May-07-2006]

Reference:
Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

RID: 1153932842-17707-84449463301.BLASTQ4

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, environmental samples or phase 0, 1 or 2 HTGS sequences)
4,265,188 sequences; 17,783,982,231 total letters

If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#)
[Taxonomy reports](#)

Query=
Length=1208

Distribution of 207 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

Color key for alignment scores

Score Range	Color
<40	Black
40-50	Blue
50-80	Green
80-200	Magenta
>=200	Red

Query

0 200 400 600 800 1000 1200

5. The most possible genes will be given by the search and human genes has to be selected by the abbreviation NM or (XM if human). This can be found by scrolling down the page and the first one will be used:

RID=1153932842-17707-84449463301.BLASTQ4, - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi

Distance tree of results **NEW**

Sequences producing significant alignments:

Sequences producing significant alignments:	Score (Bits)	E Value
gi 11558487 emb AJ404614.1 HSA404614 Homo sapiens mRNA for B...	271	5e-69
gi 62822550 gb AC159040.2 Pan troglodytes BAC clone CH251-14...	271	5e-69
gi 62702371 gb AC136807.2 Pan troglodytes BAC clone CH251-35...	271	5e-69
gi 15282111 emb AF262151.5 CNS01RHO Human chromosome 14 DNA s...	271	5e-69
gi 14018354 emb AL359681.5 CNS05TEZ Human chromosome 14 DNA s...	271	5e-69
gi 12597634 ref NM_008841.1 Homo sapiens B cell CLL/lymphoma...	271	5e-69
gi 49574493 ref NM_138576.2 Homo sapiens B-cell CLL/lymphoma...	271	5e-69

- By clicking on this selected gene the information about Gene ID (gene), Gene symbol (given in the definition in brackets) and the protein ID (NP at NCBI) will be given and direct link to the correlating NCBI page.

NCBI Sequence Viewer v2.0 - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Nucleotide&list_uids=12597634&dopt=GenBank

NCBI Nucleotide

Search: Nucleotide for [] Go Clear

Display: GenBank Show 20 Send to []

Range: from begin to end Reverse complemented strand Features: SNP STS + Refresh

1: [NM_022898](#) Reports Homo sapiens B-ce...[gi:12597634] Links

[Comment](#) [Features](#) [Sequence](#)

LOCUS NM_022898 7603 bp mRNA linear PRI 04-DEC-2005

DEFINITION Homo sapiens B-cell CLL/lymphoma 11B (zinc finger protein) (BCL11B), transcript variant 2, mRNA.

ACCESSION NM_022898

VERSION NM_022898.1 GI:12597634

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 7603)
 AUTHORS Cismasiu,V.B., Adamo,K., Gecewicz,J., Duque,J., Lin,Q. and Avram,D.
 TITLE BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter
 JOURNAL Oncogene 24 (45), 6753-6764 (2005)
 PUBMED [16091750](#)
 REMARK GenERIF: transcriptional repression fuction is mediated by NuRD complex

REFERENCE 2 (bases 1 to 7603)
 AUTHORS Przybylski,G.K., Dik,W.A., Wanzeck,J., Grabarczyk,P., Majunke,S., Martin-Subero,J.I., Siebert,R., Dolken,G., Ludwig,W.D., Verhaaf,B., van Dongen,J.J., Schmidt,C.A. and Langerak,A.W.
 TITLE Disruption of the BCL11B gene through inv(14)(q11.2q32.31) results in the expression of BCL11B-TRDC fusion transcripts and is associated with the absence of wild-type BCL11B transcripts in ...

Gene symbol will be found in brackets under Definition on top of the page

By scrolling this page one will find further necessary informations under Features → gene:

Gene ID, gene symbol and gene MIM.

One will also find the protein ID under CDS → /protein_id

These data have to be used for further characterisation.

NCBI Sequence Viewer v2.0 - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Nucleotide&list_uids=12597634&dopt=GenBank

Summary: This gene encodes a C2H2-type zinc finger protein and is closely related to BCL11A, a gene whose translocation may be associated with B-cell malignancies. The specific function of this gene has not yet been determined. Two alternatively spliced transcript variants, which encode distinct isoforms, have been reported.

Transcript Variant: This variant (2) lacks an exon within the coding region, when compared to variant 1. The translation remains in-frame, and thus results in an isoform (2) that lacks an internal segment, as compared to isoform 1.

COMPLETENESS: complete on the 3' end.

FEATURES

source 1..7603
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="14"
 /map="14q32.2"

gene 1..7603
 /gene="BCL11B"
 /note="synonyms: RIT1, CTIP2, CTIP-2"
 /db_xref="GeneID:64919"
 /db_xref="HGNC:13222"
 /db_xref="MIM:606558"

CDS 268..2739
 /gene="BCL11B"
 /go_component="nucleus"
 /go_function="metal ion binding; nucleic acid binding; zinc ion binding"
 /go_process="regulation of transcription, DNA-dependent; transcription"
 /note="isoform 2 is encoded by transcript variant 2; B-cell lymphoma/leukaemia 11B; zinc finger protein hRit1 alpha"
 /codon_start=1
 /product="B-cell CLL/lymphoma 11B isoform 2"
 /protein_id="NP_075049.1"
 /db_xref="GI:12597635"
 /db_xref="CCDS:CCDS9949.1"
 /db_xref="GeneID:64919"
 /db_xref="HGNC:13222"

7. The linked protein page has to be selected and on the connected page the information about the expressed protein will be given.

NCBI Sequence Viewer v2.0 - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NP_075049.1

NCBI Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for [] Go Clear

Display GenPept Show 5 Send to

Range: from begin to end Features: SNP + Refresh

1: NP_075049.1 Report B-cell CLL/lympho.. [gi:12597635] BLink, Links

Comment Features Sequence

LOCUS NP_075049 823 aa linear PRI 04-DEC-2005

DEFINITION B-cell CLL/lymphoma 11B isoform 2 [Homo sapiens].

ACCESSION NP_075049

VERSION NP_075049.1 GI:12597635

DBSOURCE REFSEQ: accession [NM_022898.1](#)

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 823)
 Cismasiu, V.B., Adamo, K., Gecewicz, J., Duque, J., Lin, Q. and Avram, D.
 TITLE BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter
 JOURNAL Oncogene 24 (45), 6753-6764 (2005)
 PUBMED 16091750

REMARK GenEPIF: transcriptional repression function is mediated by NuRD complex

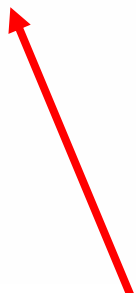
REFERENCE 2 (residues 1 to 823)
 Przybylski, G.K., Dik, W.A., Wanzeck, J., Gramarczyk, P., Majunke, S., Martin-Subero, J.I., Siebert, R., Dolken, G., Ludwig, W.D., Verhaaf, B., van Dongen, J.J., Schmidt, C.A. and Langerak, A.
 TITLE Disruption of the BCL11B gene through inv(14)(q11.2q32.31) results in the expression of BCL11B-TRDC fusion transcripts and is associated with the absence of wild-type BCL11B transcripts in ...

8. The protein sequence has to be written into fasta format (using the display in the task line).

1: NP_075049. Reports B-cell CLL/lympho... [gi:12597635] BLink, Links

```
>gi|12597635|ref|NP_075049.1| B-cell CLL/lymphoma 11B isoform 2 [Homo sapiens]
MSRRKQGNPQHLSQRELITPEADHVEAAILEEDEGLEIEEPSGLGLMVGGPDPDLLTCGQCMMNFFLGD I
LVFIEHKKKQCGGSLGACYDKALDKDSDPPSSRSELKRVSEPEVIGIQVTPDEDDHLLSPTKGIQPKQEN
IAGKDEPSSYICTTCKQPFNSAWFLLQHAQNTGFRILEPGPASSSLTPRLTIPPLGPEAVAQSPLMN
FLGDSNPFNLLRMTGPIIRDHPGFGEGRLPGTFLFSPPPRLHLDPHRLSAEEMGLVAQHPSAFDRVMRL
NPMADSPAMDFSRRLRELAGNSSTPPFVSPGRGNPMHRLNPFQSPKSPFLSTPPLPMPGGTPPPQ
PPAKSKSCEFCGKTFKQSNLIVHRRSHTGKPKYKQCDHACSQASKLKRHMKTHMHKAGSLAGRSDDG
LSAASSPEFOTSELAGEGLKAADGDFRHESDPSLGHEPEEEEEEEEEELLENESPESSFSMDSE
LSRNRENGGGVPGVPGAGGGAAKALADEKALVLGKVENVGLGALPYGELLADKQKRGAFLLKRAAGG
DAGDDDDAGCGGDAGAGGAVNGRGGGAPGTEPFPGLFPRKPAPLSPGLNSAAKRIKVEKDLLEPPAAL
IPSENVYSQVLVGYAASRHMMDPFLGFTDARQSPFATSSSEHSSENGSLRFSTPPGDLDDGGLSGRSGTA
SGGSTPHLGGPGPRPSSKEGRSDTCEYCGKVKNCNSLTVHRRSHTGERPKYKELCNACQSSKLTR
HMKTHGQIGKEVYRDCIQMPFVSVYSTLEKHMKKVHGEHLLTNDVKIEQAERS
```

[Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)



This is the protein sequence of the corresponding gene (2nd protein sequence)

9. The two protein sequences will be compared at the NCBI (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>)

This page is used for comparing proteins, so go to blastp and enter the two protein sequences (1st protein, 2nd protein)

Blast 2 Sequences - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi

SMART: Main page | BBCU - Sequence analysis/Gene ... | KEGG PATHWAY Database | **Blast 2 Sequences** | Translate Tool - Results of transl...

Reward for a match: Penalty for a mismatch:

Use [Mega BLAST](#) Strand option: Not Applicable View option: Standard

Masking character option: X for protein, n for nucleotide Masking color option: Black

Show CDS translation

Open gap: and extension gap: penalties

gap_x_dropoff: expect: word size: Filter

Sequence 1

Enter accession, GI or sequence in FASTA format from: to:

```

A A S X X S P X P X P X X S X X N X X X X X X X X X L P
P X X X X X X X X X X X X X X X X X X L K K X X T Y K X
L X X X X T L X L X F X L Met N X F X X L X K K X X X I F F
F X X X X X X Y X S S K N X E X L X L X X X P R X S X X
Stop K X F X X X S X L X X X X X X Y X L X X X F X F X X
X X V X X X N X F L X F X N F X N X F X X L F F X X X X X
X X X P X X X X X N X X X X Y X X
  
```

or upload FASTA file

Sequence 2

Enter accession, GI or sequence in FASTA format from: to:

```

LSAASSPEPGTSELAGEGLKAADGDFRHESDPSLGHPEEEDDEEEELLENESRI
LSRNRENGGGVPGVPGAGGGAALADEKALVLGKVMENVLGLALPQYGELLADKQKRG.
DAGDDDDAGCCGDAGAGGAVNRRGGGFAPGTEPFGPLFRKPAFLPSPGLNSAAKRIKVEI
IPSENVYSQWLVGVAASRHFMKDPFLGFTDARQSPFATSEHSSSENGSLRFPSTPGDLLD
SGGSTPHLGGPGPRPSSKEGRSDTCEYCGKVFKNCSNLTVHRRSHTGERPYKCELCNY.
HMKTHGQIGKEVYRCDICQMPFSVYSTLEKHKKWHGEHLLTNDVKIEQAERS
  
```

or upload FASTA file

Go to align. If the two proteins match, the following result will be given:

Blast Result - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?0

SMART: Main page | BBCU - Sequence analysis/Gene ... | KEGG PATHWAY Database | **Blast Result** | Translate Tool - Results of transl...

NCBI Blast 2 Sequences results

PubMed Entrez BLAST OMIM Taxonomy Structure

BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.14 [May-07-2006]

Matrix: BLOSUM62 gap open: gap extension:

x_dropoff: expect: wordsize: Filter View option: Standard

Masking character option: X for protein, n for nucleotide Masking color option: Black

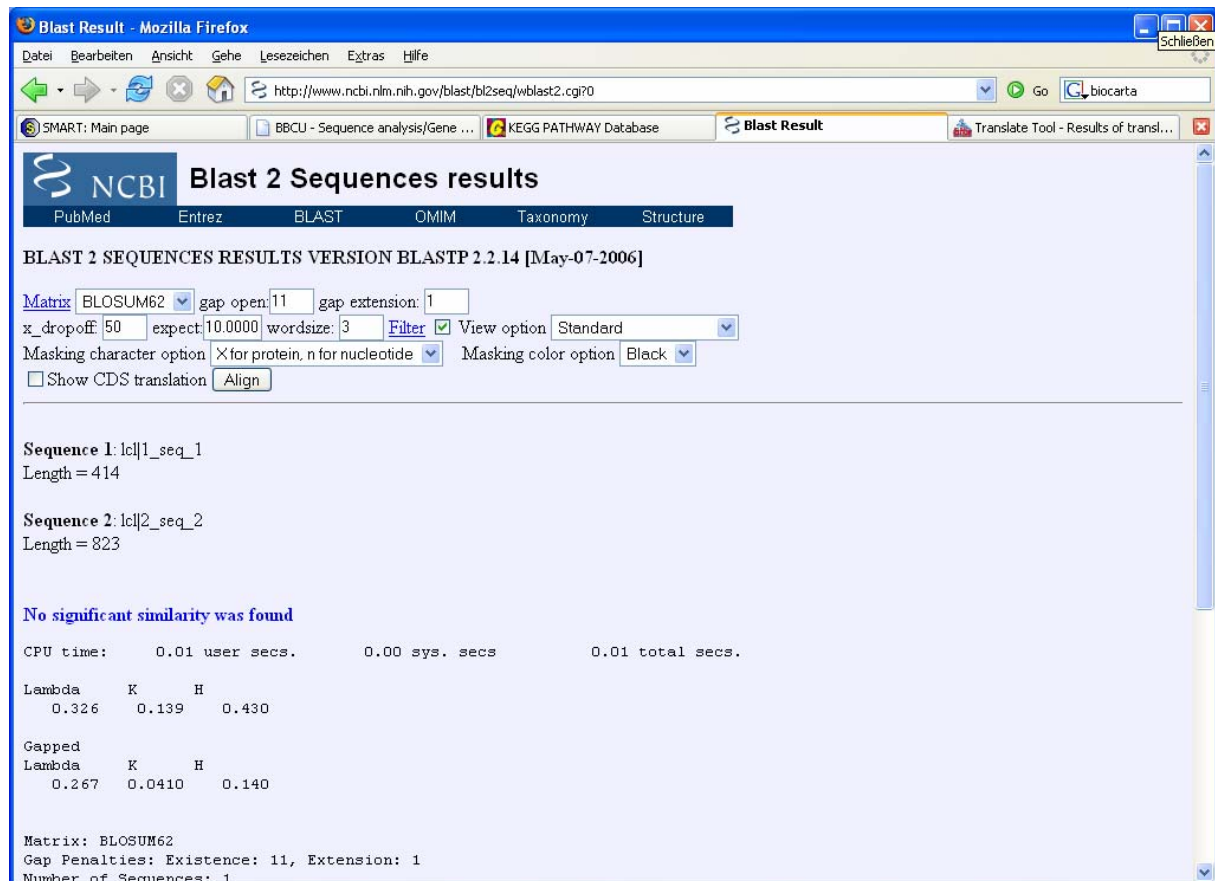
Show CDS translation

Sequence 1: lc|1_seq_1
Length = 414 (1 .. 414)

Sequence 2: lc|2_seq_2
Length = 823 (1 .. 823)

NOTE: Bitscore and expect value are calculated based on the size of the nr database.

If the proteins do not match (not-inframe) the result looks in the following way:



10. If there is an alignment the sequence is in frame and the protein will be expressed by the tumor.

11. Further characterisations are possible by the use of the databases described in the following table

Information on chromosomal localization, protein function, and subcellular localization has to be retrieved from	National Center for Biotechnology Information	www.ncbi.nlm.nih.gov
	GeneCards	http://www.genecards.org/index.shtml
	EBI	http://www.ebi.ac.uk/
	Swiss Prot	http://www.ebi.ac.uk/swissprot/access.html
Information on pathways has to be retrieved from	KEGG PATHWAY database	http://www.genome.jp/kegg/pathway.html
Information on pathways has to be retrieved from	Biocarta database	http://www.biocarta.com/search/index.asp
Information on domains has to be retrieved from	SMART database	http://smart.embl-heidelberg.de
Information about antigens found in other tumours from	SEREX database Cancer Immunome database	http://www2.licr.org/CancerImmunomeDB/
	CAP * (Cancer associated proteins) database	http://www.bioinf.uni-sb.de/CAP/
Translation of DNA to Protein	Swiss Institute of Bioinformatics	http://www.expasy.org/tools/dna.html
Information about autoimmunity of antigens from	the autoimmune database	http://www.wiley-vch.de/contents/jc_2040/2005/25481_s.pdf

12. As an identification the NCBI accession number, gene number or gene symbol will be used. The NCBI database will also provide links to publications of each of the antigens.

- o Antigen function, localisation in the cell and the intracellular processing will be characterized by using the gene ontology homebase from the NCBI homepage (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene>) by entering the gene symbol.

Gene - Mozilla Firefox
 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene

NCBI Entrez Gene

Search Gene for BCL11B

Display Summary Show 20 Send to

All: 11 Current Only: 11 Genes Genomes: 11 SNP GeneView: 9

Items 1 - 11 of 11 One page.

1: **BCL11B**
 Official Symbol: BCL11B and Name: B-cell CLL/lymphoma 11B (zinc finger protein) [*Homo sapiens*]
 Other Aliases: CTIP-2, CTIP2, RIT1
 Other Designations: B-cell CLL/lymphoma 11B; B-cell lymphoma/leukaemia 11B; zinc finger protein hRit1 alpha
 Chromosome: 14; Location: 14q32.2
 MIM: 606558
 GeneID: 64919

2: **Bcl11b**
 Official Symbol: Bcl11b and Name: B-cell leukemia/lymphoma 11B [*Mus musculus*]
 Other Aliases: B630002E05Rik, Ctip2, Rit1
 Other Designations: B-cell lymphoma/leukaemia 11B; COUP-TF interacting protein 2; CTIP2 protein
 Chromosome: 12; Location: 12 52.0 cM
 GeneID: 58208

3: **BCL11B**
 B-cell CLL/lymphoma 11B (zinc finger protein) [*Gallus gallus*]
 Chromosome: 5
 GeneID: 423444

4: **Bcl11b_predicted**
 Official Symbol: Bcl11b_predicted and Name: B-cell leukemia/lymphoma 11B (predicted) [*Rattus norvegicus*]
 Chromosome: 6; Location: 6q32

As a result the following will be available:

Gene - Mozilla Firefox
 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Display&DB=gene

NCBI Entrez Gene

Search Gene for

Display Full Report Show 5 Send to

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

1: **BCL11B B-cell CLL/lymphoma 11B (zinc finger protein)** [*Homo sapiens*]
 GeneID: 64919 Primary source: [HGNC:13222](#) updated 21-Jul-2006

Summary

Official Symbol: BCL11B and Name: B-cell CLL/lymphoma 11B (zinc finger protein) provided by [HUGO Gene Nomenclature Committee](#)

See related: [HPRD:05950](#), [MIM:606558](#)

Gene type: protein coding
 Gene name: BCL11B
 Gene description: B-cell CLL/lymphoma 11B (zinc finger protein)
 RefSeq status: Reviewed
 Organism: [Homo sapiens](#)
 Lineage: *Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominae; Homo*
 Gene aliases: RIT1; CTIP2; CTIP-2
 Summary: This gene encodes a C2H2-type zinc finger protein and is closely related to BCL11A, a gene whose translocation may be associated with B-cell malignancies. The specific function of this gene has not yet been determined. Two alternatively spliced transcript variants, which encode distinct isoforms, have been reported.

Genomic regions, transcripts, and products

(minus strand) [RefSeq below](#)

NC_000014.7

98607575 98795377

5' 3'

NM_022895.1 NP_075049 isoform 2 CDS3949..1
 NM_138576.2 NP_612808 isoform 1 CDS3950..1

- Correlating pathways of the antigen will be provided by the Bicarta database by the gene name or the gene symbol.

GENE SEARCH

Conduct any broad molecular or genetic investigation - by organism, area of research, or keyword search - using multiple online databases. Locate gene-specific information, including sequence data, publications and reviews, disease correlation, and interrelationship with other proteomic pathways. Or, save favorite search parameters to your My BioCarta account.

SEARCH

CATEGORY: All → ORGANISM: Human
 GENE NAME: CONTAINS OXA1L (*required)
 GENE SYMBOL:

SEARCH

▲ TOP

Privacy Policy | Disclaimer | Terms & Conditions | Sponsor Information

The following result will be given:

GENE SEARCH SEARCH RESULTS

Click here to initiate a **NEW SEARCH**

[Results 1 to 1 of 1 found]

H.Sapiens - OXA1L oxidase (cytochrome c) assembly 1-like

COMMENTS SAVE LINK

Gene Results:

Review	DNA/RNA	Protein	Others	Publications	Biocarta Results
Omim	Entrez KEGG Locus UniGene	EntrezProtein SwissProt	Homol Genecard MapView Wormbase SNP	PubMed	

▲ TOP

Privacy Policy | Disclaimer | Terms & Conditions | Sponsor Information

One can choose the KEGG database giving the following result:

http://www.genome.ad.jp - DBGET Search Result: GENES OXA1L - Mozilla Firefox

Database: GENES
Search term: OXA1L (Total 7 hits)

hsa:5018 OXALL; oxidase (cytochrome c) assembly 1-like; K03217 preprotein translocase YidC subunit
 mmu:69089 Oxall; oxidase assembly 1-like; K03217 preprotein translocase YidC subunit
 rno:498512 LOC498512; similar to Oxall protein
 cfa:480239 LOC480239; similar to Inner membrane protein OXAll, mitochondrial precursor (Oxidase assembly 1-like protein) (OXAL-like protein) (0:
 bta:520335 MGC127460; similar to Inner membrane protein OXALL, mitochondrial precursor (Oxidase assembly 1-like protein) (OXAL-like protein) (0:
 gga:425893 LOC425893; similar to Inner membrane protein OXALL, mitochondrial precursor (Oxidase assembly 1-like protein) (OXAL-like protein) (0:
 dre:562963 LOC562963; similar to Inner membrane protein OXAll, mitochondrial precursor (Oxidase assembly 1-like protein) (OXAL-like protein)

DBGET integrated database retrieval system, GenomeNet

By searching the database for human genes (hsa): the following will be displayed:

http://www.genome.ad.jp - DBGET Result: H.sapiens 5018 - Mozilla Firefox

KEGG Homo sapiens (human): 5018 Help

Entry	5018	CDS	H.sapiens
Gene name	OXALL		
Definition	oxidase (cytochrome c) assembly 1-like		
KO	KO: K03217 preprotein translocase YidC subunit		
Pathway	PATH: hsa03060 Protein export		
Class	Gene catalog		
SSDB	Ortholog Paralog Gene cluster		
Motif	Pfam: OKD_IMP Motif		
Other DBs	OMIM: 101066 NCBI-Gene: 4826880 NCBI-GeneID: 5018 UniProt: Q15070		
LinkDB	PDB All DBs		
Position	14q11.2		
AA seq	495 aa AA seq DB search MVTMLYFLPSTNMAAKLRSLPPDLRLQFWLHARLQKCFLSRGCGSYCAGAKASPLPGK MAMGLMGRRELLRLQSGRRVHVSAGPSQWLGKPLTRLLFPVAPCCRPHYLFLAASG PRSLSTSAISFAEVQVQAPFVVAATPSPAVPEVASGETADVQVTAAREQSPAEGLGYSIT PVGLIQLLLEFMHVDLGLPMMGAIAACTVFARCLIFPLIVTGQREARIRHNLPEIQKFS SRIREALAGDHIYYKASSEMALYQKKHGKLYKPLILPVTQAPIFISFFIALREMANL PVPSLQSGLWVFQDLTVSDPIYILPLAVTATMMAVLELGAETGVQSSDLQMMRNVIRMM PLITLPTMHPPTAVFMYLSSMLFSLVQVSLRIPAVRTVLKIPQVVDLKLPPREG FLESFKGWKNAEMTRQLRERQRMRNQLELAARGPLRQTFTHNPLLQPGKDNPPNIPSS SSKPKSYFPHDITLG		
NT seq	1488 nt NT seq +upstream0 nt +downstream0 nt atgggtacggtgctttacagattttaccacttcaaatatggccgccaagctccgtctc ctcttaccgctgatctacggctacaattctggttcctatgccgcctccaaaagtcttc ctctcgggggttggttcttactgcgcaggcgcaaaagcaagtcctcttccgggcaaa atggcgcgggactaatgtcggacccggagcttctgcgcttgctacagctccgggcgt cgggtccacagcgtcgcaggccctcgaatggcttgggaaccgctgaccacacggctc ctattccagtaqccccgtctctctcgcacacactactcttctcttcgqctccgctc		

Clicking to pathways one will get the following result:

http://www.genome.ad.jp - Protein export - Homo sapiens (human) - Mozilla Firefox

Protein export - Homo sapiens (human)

[Pathway menu | Ortholog table]

Homo sapiens (human) Go Current selection Select

PROTEIN EXPORT

Sec dependent pathway

SecB	SecA	SecY	SecE	SecG
SecD	SecF	YajC	YidC	

Signal peptidase

SPase I	SPase II
---------	----------

SRP (signal recognition particle) dependent pathway

SRP9	SRP72	SRP19	SRPR	RNA
SRP14	SRP68	SRP54		4.5S

Tat (twin-arginine translocation) system

TatA	TatB	TatC
TatE		

03060 6/16/03

- o The SMART database has to be used to show the protein domains (enter the protein ID or the complete protein sequence)

SMART: Main page - Mozilla Firefox

http://smart.embl-heidelberg.de/

SMART: Main page BCCU - Sequence... KEGG PATHWAY... NCBI Blast Translate Tool - R... NCBI gene ontolo... gene ontology N... SMART: Main p...

SMART MODE: NORMAL GENOMIC

Simple Modular Architecture Research Tool

Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95: 6857-6864
Letunic et al. (2006) Nucleic Acids Res 34: D257-D260

HOME SETTING FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

Sequence analysis

You may use either a [UniProt](#) / [ensembl](#) sequence identifier (ID) / accession number (ACC) of the protein sequence itself to request the SMART service.

Sequence ID or ACC

Sequence

```
MSRRKQGNPQHLSORELITPEADHVEAAALLEEDEGLEIEE P SGLG:
LVFIEHKRQCGGSLGACYKALDKD S P P S R S E L R K V S E P V E I
I A K D E P S S Y I C T T C K P F N S A W F L Q H A Q N T H G F R I Y L E P G P A S :
F L G D S N P F M L L R M T G F I L R D H P F G E G R L P G T P P L F S P P P R H H L D :
N P M A I D S P A M D F S R R L R E L A G N S S T P P P V S P G R G N P M H R L L N P F Q :
P P A K S K S C E F C G K T F K F Q S N L I V H R R S H T G E K P Y K C L C D H A C S Q :
L S A A S S P E P G T S E L A G E G L K A A D G F R H H E S D P S L G H E P E E D E E I
L S R N R E N G G G V P G V F G A G G G A A K A L A D E K A L V L G K V M E N V L G A :
```

Sequence SMART Reset

HMMER searches of the SMART database occur by default. You may also find:

- Outlier homologues and homologues of known structure
- PFAM domains
- signal peptides
- internal repeats
- intrinsic protein disorder

If you have multiple sequences to analyze, try [batch access](#) to SMART database.

Architecture analysis

You can search for proteins with combinations of [specific domains](#) in different species or taxonomic ranges. You can input the domains directly into "Domain selection" box, or use "GO terms query" to get a list of domains. See [What's New](#) for more info.

Domain selection

Example: TyrKc AND SH3 AND NOT SH2

GO terms query

Example: membrane AND signal transduction

Taxonomic selection

Select a taxonomic range via the selection box or type it into the text box below:

All

Examples: Dictyostelium discoideum, Porifera

Architecture query Zurücksetzen

You can try an [Advanced Query](#) if you're familiar with SQL.

Alert SMART

If you want to be automatically informed each time a new protein with a defined domain composition is deposited in the database, please use 'Alert SMART' (this facility is also available following an architecture analysis query).

Domains detected by SMART

The result will be:

SMART: Sequence analysis results for Q9C0K0-2 - Mozilla Firefox

SMART MODE: **NORMAL GENOMIC**

Simple Modular Architecture Research Tool

Domains within *Homo sapiens* protein Q9C0K0-2

Splice isoform 2 of Q9C0K0

Mouse over domain / undefined region for more info; click on it to go to detailed annotation; right-click to save whole protein as PNG image

Transmembrane segments as predicted by the *TMHMM2* program (■), coiled coil regions determined by the *Coils2* program (■), segments of low compositional complexity determined by the *SEG* program (■). Signal peptides determined by the *SignalP* program (■). Intron positions are indicated with vertical lines showing the intron phase and exact position in AA.

Additional information

- [Display other IDs, orthology and alternative splicing data for this sequence.](#)
- Domain architecture analysis**
- This domain architecture was probably invented with the emergence of *Eukaryota*.
- [Display all proteins with similar domain organisation.](#)
- [Display all proteins with similar domain composition.](#)

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by **SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstructured regions > Low complexity**. In either case, features not shown in the above diagram are marked as 'overlap' in the second table below.

Confidently predicted domains, repeats, motifs and features:

- The SEREX database will be analysed to detect other tumours expressing the antigens found in the actual experiment.

By the input of the nucleotide sequence the following result will be given:

Cancer Immunome Database

Ludwig Institute For Cancer Research
Academy of Cancer Immunology
European Cancer Immunome Program

Welcome, Alexander Hoppe

Genes: 1545 HUGO/RefSeq genes, 1325 loci; SEREX: 2743 sequences, 2316 clones;

Check similarity against Serex sequences/genes

Paste your query, not longer than 1000 bases:

```
AMTTAGAAATTAATGAGGATTGCATCATCATCATCATCGAT  
CTATTAGGTAATATAGAAATGATGATGATGATGATGATGAT  
TGCATGATGATGATGATGATGATGATGATGATGATGATGAT  
GCATGATGATGATGATGATGATGATGATGATGATGATGAT  
GCATGATGATGATGATGATGATGATGATGATGATGATGAT  
GCATGATGATGATGATGATGATGATGATGATGATGATGAT  
GCATGATGATGATGATGATGATGATGATGATGATGATGAT  
GCATGATGATGATGATGATGATGATGATGATGATGATGAT
```

search similarity against SEREX sequences

show top 10 entries

Search (might take several minutes)

This will give the following result:

The screenshot shows the CIDB SEREX website interface. The main heading is "Cancer Immunome Database". Below it, there is a navigation menu with categories like Home, Serex, Peptides, New/edit..., Serology, Microarray, Genetic, Epigenetic, and Antibodies. The central content area displays "Smith-Waterman search results" in a table format.

Serex sequence	score	
NGO-Pr-31	41	show alignment
NGO-Br-64	40	show alignment
NY-Co-9	39	show alignment
NGO-St-2	36	show alignment
TC70	36	show alignment
NY-SAR-33	36	show alignment
NGO-St-61-5'	35	show alignment
NGO-St-142	35	show alignment
NY-SCLC-2	35	show alignment
TE-70	35	show alignment

- The relation between the expressed genes and the chromosomal localisation as well as the expression in different human tissues will be found in the gene cards by entering the external id.

The screenshot shows the GeneCards website homepage. The search bar contains the text "BCL11B" and a "Go" button. Below the search bar, there are radio buttons for different search criteria: "Keywords", "Symbol only", "Symbol/alias", "GC id", and "External id". The "External id" option is selected and highlighted with a yellow starburst icon. A red arrow points from the text in the list above to this "External id" option.

Examples of search results for "BCL11B" are shown below the search bar:

Keywords	Symbol only	Symbol/alias	GC id	External id
tay sachs	BTK	FRAXA	GC17M0384*	7431
dimerization AND diabetes	wnt*			P12004
neurodegenerative OR senile				ENSG00000185515

The following will be displayed:

GeneCard for protein-coding *BCL11B*
GC14M098705

B-cell CLL/lymphoma 11B (zinc finger protein)
Symbol approved by the *HUGO Gene Nomenclature Committee (HGNC) database*

Aliases and Descriptions
(According to ¹HGNC, ²Entrez Gene, ³UniProt/Swiss-Prot, ⁴UniProt/TrEMBL, ⁵GDB, ⁶OMIM, ⁷GeneLoc, and/or ⁸Ensembl)

Aliases
CTIP-2 ^{1,2,5}
CTIP2 ^{2,3,6}
RIT1 ^{2,3}
hRit1 ³

Descriptions
B-cell CLL/lymphoma 11B ³
B-cell CLL/lymphoma 11B (zinc finger protein) ^{1,2,5}
B-cell lymphoma/leukemia 11B ³
COUP-TF interacting protein 2 ³
Radiation- induced tumor suppressor gene 1 protein ³

External Ids
HGNC: 13222¹
Entrez Gene: 64919²
UniProt: Q9CDK0³
Ensembl: ENSG00000127152⁸

Genomic Location
(According to GeneLoc and/or HGNC, and/or Entrez Gene (NCBI build 35), and/or miRBase, Genomic Views According

Chromosome: **14**
Entrez Gene cytogenetic band: [14q32.2](#) Ensembl cytogenetic band: [14q32.2](#)
Gene in genomic location: bands according to Ensembl, locations according to [GeneLoc](#) (and/or Entrez Gene and/or Ensembl if different)

Chr 14
p13 p12 p11.2 p11.1 q11.1 q11.2 q12 q13.1 q21.1 q21.2 q21.3 q22.1 q22.3 q23.1 q23.2 q23.3 q24.1 q24.2 q24.3 q31.1 q31.3 q32.11 q32.12 q32.13 q32.2 q32.32 q32.33

- There will be detailed information from other homepages concerning the isolated gene at gene cards too, like the Human Genome database (<http://www.gdb.org/>).

The GDB Human Genome Database

The Official World-Wide Database for the Annotation of the Human Genome
Hosted by RTI International, North Carolina USA
Available at [mirror sites worldwide](#)

[Information on the New GenePoint Database](#)

[GenePoint Beta Users Log-in Here](#)

Simple Search

Search: Genomic Segments All Biological Data People Citations

By: Name/GDB ID DNA Sequence ID

Note: For Name/ID searches, adding * to the end of the search text may improve results.

Advanced Search

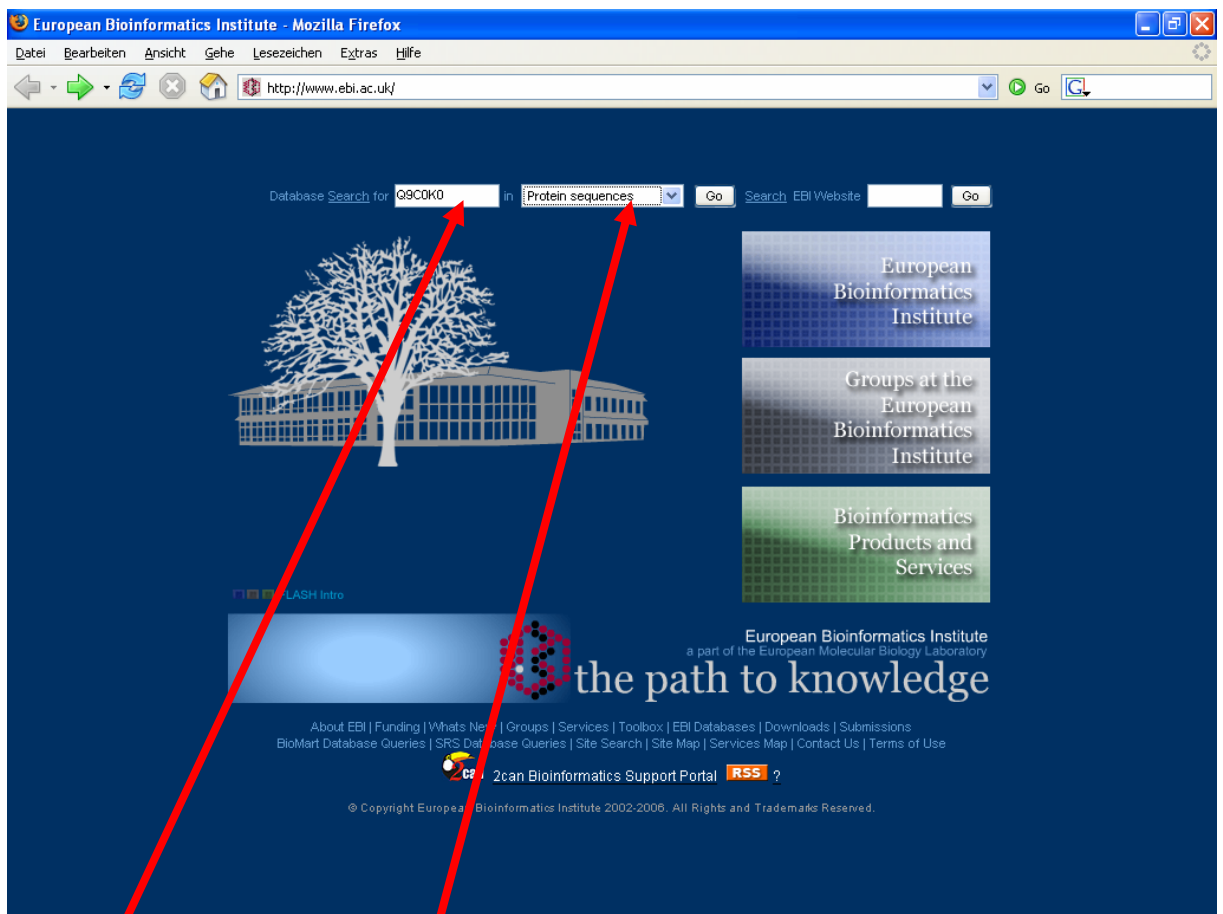
- [Preferences](#)
- [Other Search Options](#)

Submit and Edit Data

- [How to submit data, get D seq #s **NEW!**](#)
- [Insert Object](#)
- [Edit Record](#)
- [Search for Unreleased Records](#)
- [Apply for Account](#)

Navigation Menu: Help, Site Map, Reports, Resources, GDB Operations, CF Mutations, HUGO Chromosomes, GDB Redevelopment Survey, What's New: GDB Host, GDB New Look, GDB Staff

Further databases are: EBI (<http://www.ebi.ac.uk/>)



By entering the name of the protein sequence (Swiss Prot accession No., for example: Q9CoKO for BCL11B) you will get the following result:

Basic UniProtKB Entry Viewer

Your Query Result Sets (Page - 1)[Data Set Manager]

Database	Accession	Count
TrEMBL	Q9C0K0	1 entry
Swiss-Prot	BCL11B	3 entries

Protein BC11B_HUMAN [New Query](#) [Submit Annotation](#) [Download Protein](#) [Bookmark Protein \(Ctrl+D\)](#)

[Q9C0K0](#)

Basic | Extended Viewers: [Fasta](#) | [Flat File](#) | [XML](#) | [Expasy](#) | [SRS](#) | [PIR](#)

General information about the UniProtKB/Swiss-Prot entry

Entry name	BC11B_HUMAN
Primary accession number	Q9C0K0
Secondary accession numbers	Q9H162
Integrated into UniProtKB/Swiss-Prot	01-MAR-2004
Sequence was last modified	01-JUN-2001, version 1
Entry was last modified	30-MAY-2006, version 29

Protein description

Protein name	B-cell lymphoma/leukemia 11B
Synonyms	B-cell CLL/lymphoma 11B Radiation-induced tumor suppressor gene 1 protein hRit1 COUP-TF-interacting protein 2

The Swiss Prot accession No. can be get from : (<http://www.ebi.ac.uk/swissprot/access.html>)

UniProtKB/Swiss-Prot Database Access

The UniProtKB/Swiss-Prot database can be accessed using the following search engines:

(1) [SRS](#) - this search tool can also be used for more complex and/or multiple database queries.

Search: UniProt/Swiss-Prot for: BCL11B

(2) [UniProt Power Search](#) - Provides full text, advanced search, set manipulation and search filtering on the Universal Protein Resource. [[help](#)]

Library: UniProtKB/Swiss-Prot Only
 line Type: Any Field
 Operator: Exact Match
 Query Text:

Other Access Methods

The UniProtKB/Swiss-Prot database can also be accessed using:

- [ExPASy Server](#) - in Geneva offers the choice of full-text search or of individual lines (e.g. ID, AC, DE, OS, OG, GN, RL, RA).
- [SP-ML](#) - the UniProtKB/Swiss-Prot & UniProtKB/TrEMBL in XML format.

[Contact](#)

The result will be:

Query "[swissprot-AllText:BCL11B*]" found 2 entries

UniProtKB/Swiss-Prot	Accession	UniSave	Description	GeneName	Species	Ke
<input type="checkbox"/> UniProtKB/Swiss-Prot: BCL11B_HUMAN	Q9C0K0	Q9C0K0	B-cell lymphoma/leukemia 11B (B-cell CLL/lymphoma 11B) (Radiation- induced tumor suppressor gene 1 protein) (hRit1) (COUP-TF-interacting protein 2).	BCL11B	HOMO SAPIENS (HUMAN).	Alte spli Met Nuc prot Rep Tra Tra reg Znc Znc
<input type="checkbox"/> UniProtKB/Swiss-Prot: BCL11B_MOUSE	Q99PV8	Q99PV8	B-cell lymphoma/leukemia 11B (B-cell CLL/lymphoma 11B) (Radiation- induced tumor suppressor gene 1 protein) (mRit1) (COUP-TF-interacting protein 2).	BCL11B	MUS MUSCULUS (MOUSE).	Alte spli Met Nuc prot Poly Rep Tra Tra reg Znc Znc