Biological databases
an introduction

By Dr. Erik Bongcam-Rudloff

SGBC-SLU

2016

VALIDATION

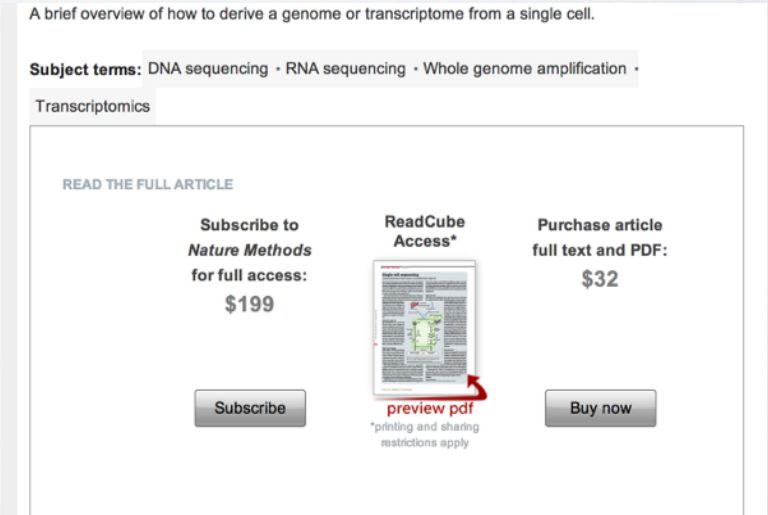
- Experimental
- Literature
- Manual or semi-automatic computational analysis

EXPERIMENTAL

- Costs
- Needs skilled manpower
- Increase in sequencing unparalleled

LITERATURE

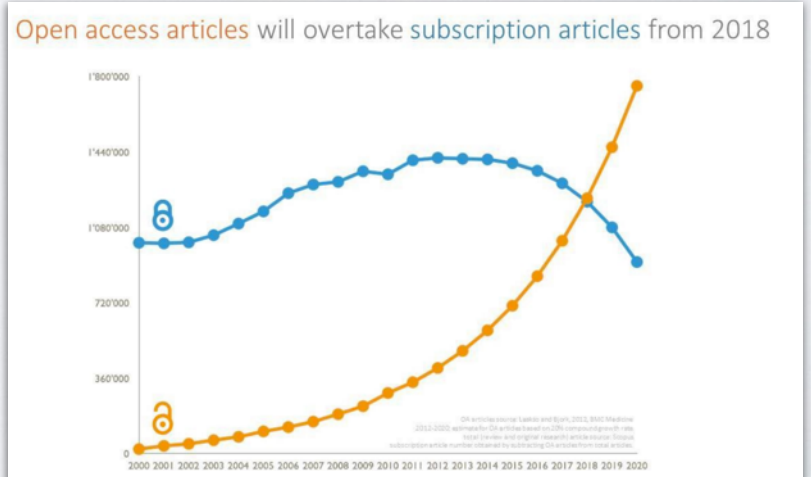
- NOmenclature
- Publishing culture
  Old ways of work and resistance to changes the culture
- PUBMED: human Centric, ONLY abstracts
- No text mining allowed



IF SCIENCE IS GOING TO SAVE THE WORLD, WE NEED TO MAKE IT OPEN

- On Friday 27 May, EU ministers of science, innovation, trade and industry published a progressive commitment calling for full open access to scientific research by 2020.

- US Vice President Joe Biden announced the launch of an open-access cancer database to allow researchers to better understand the disease and develop more effective treatments.



WHY IS THIS NOT HAPPENING?

- We only publish positive results
- No negative results even if they represent 90% of the situations
- Again we need to re-think
- We need to create Open Access repositories with original data
- All actions have a counter action

DATA SHARING

- There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as "research parasites." Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.N Engl J Med 2016; 374:276-277January 21, 2016
- …"or even use the data to try to disprove what the original investigators had posited…"

COMPUTATIONAL METHODS

- Most based on similarity
- Most tools rely on the metadata associated to each sequence

DATA BASES

- Nucleic: ENA, GenBank, DDBJ
- Protein: SwissProt, RefSeq, TREMBL
- Genomic: ENSEMBL
- Structural: PDB

- Biological databases are libraries of life sciences information, collected from scientific experiments, published literature, high-throughput experiment technology, and computational analysis.
- They contain information from research areas including genomics, proteomics, metabolomics, microarray gene expression, and phylogenetics. Information contained in biological databases includes gene function, structure, localisation (both cellular and chromosomal), clinical effects of mutations as well as similarities of biological sequences and structures.

Biological Databases

- Sequence Databases
- Genome Databases
- Structure Databases

Sequence Databases

- The sequence databases are the oldest type of biological databases, and also the most widely used

Sequence Databases

- Nucleotide: ATGC
- Protein: MERITSAPLG

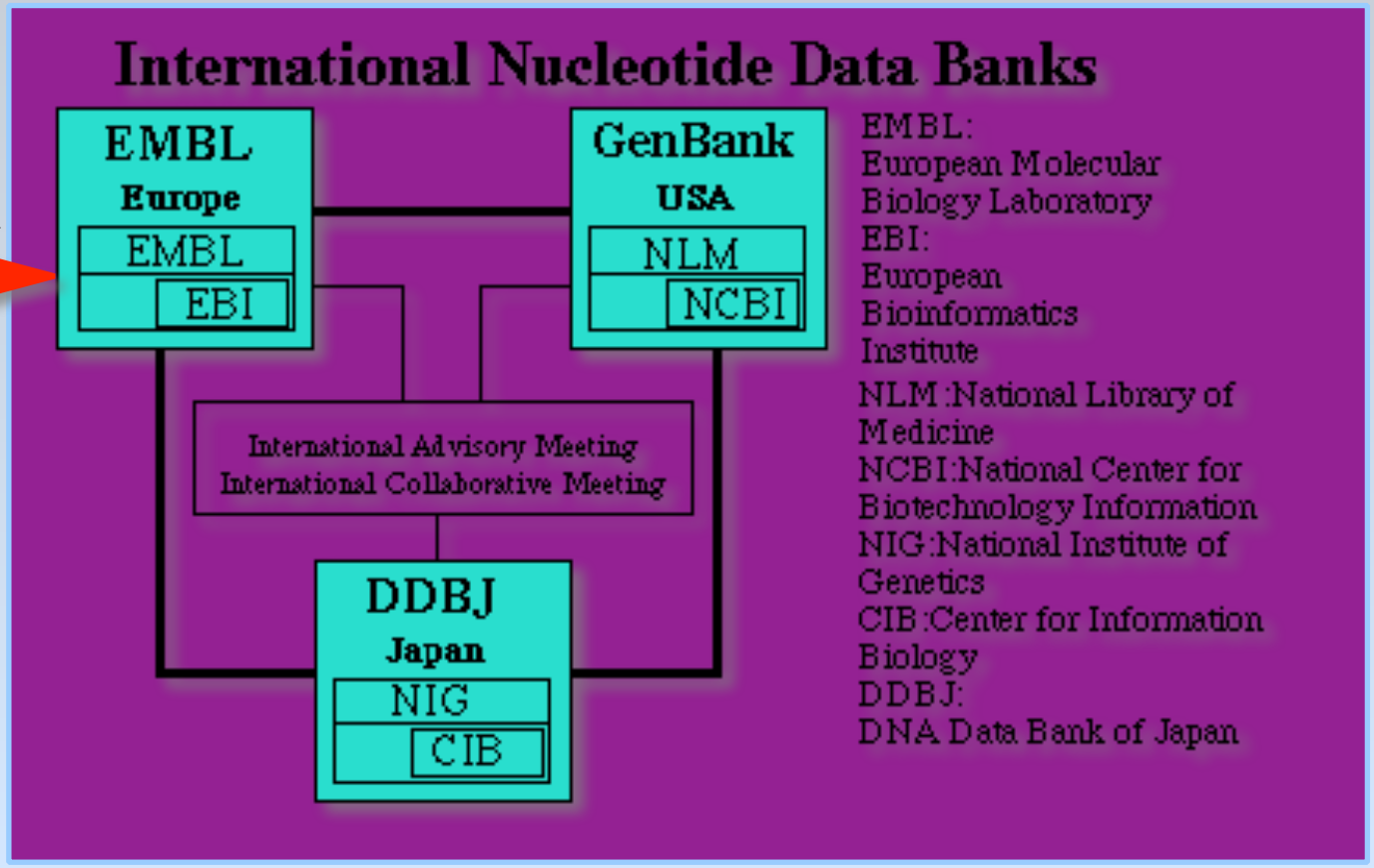
The nucleotide sequence repositories

- There are three main repositories for nucleotide sequences: EMBL, GenBank, and DDBJ.
- All of these should in theory contain "all" known public DNA or RNA sequences
- These repositories have a collaboration so that any data submitted to one of databases will be redistributed to the others.

- The three databases are the only databases that can issue sequence accession numbers.
- Accession numbers are unique identifiers which permanently identify sequences in the databases.
- These accession numbers are required by many biological journals before manuscripts are accepted.

- It should be noted that during the last decade several commercial companies have engaged in sequencing ESTs and genomes that they have not made public.

Today ENA



EST databases

- Expressed sequence tags (ESTs) are short sequences from expressed mRNAs.
- The basic idea is to get a handle on the parts of the genome that is expressed as mRNA (often called the *transcriptome*).
- ESTs are generated by end-sequencing clones from cDNA libraries from different sources.

Ideal minimal content of a « sequence » db

Sequences !!
Accession number (AC)
References
Taxonomic data
ANNOTATION/CURATION
Keywords
Cross-references
Documentation

Example:
Swiss-Prot entry

Entry name
Accession number

SWISS-PROT: O75144

NiceProt - a user-friendly view of this SWISS-PROT entry

```
ID   ICOL_HUMAN      STANDARD;      PRT;   302 AA.
AC   O75144; Q9NRQ4; Q9HD18;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ICOS ligand precursor (B7 homolog 2) (B7-H2) (B7-like protein G150)
DE   (B7-related protein-1) (B7RP-1).
GN   ICOSL OR B7H2 OR B7RP1 OR KIAA0653.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Dendritic cell;
RX   MEDLINE=20477846; PubMed=11023515; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Wang S., Zhu G., Chapoval A.I., Dong H., Tamada K., Ni J., Chen L.;
RT   "Costimulation of T cells by B7-H2, a B7-like molecule that binds
RT   ICOS.";
RL   Blood 96:2808-2813(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
RC   TISSUE=Peripheral blood lymphocytes;
RX   MEDLINE=20465019; PubMed=11007762; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Yoshinaga S.K., Zhang M., Pistillo J., Horan T., Khare S.D., Miner K.,
RA   Sonnenberg M., Boone T., Brankow D., Dai T., Delaney J., Han H.,
RA   Hui A., Kohno T., Manoukian R., Whoriskey J.S., Coccia M.A.;
RT   "Characterization of a new human B7-related protein: B7RP-1 is the
RT   ligand to the co-stimulatory protein ICOS.";
RL   Int. Immunol. 12:1439-1447(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Leukocyte;
RX   MEDLINE=20126021; PubMed=10657606; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ling V., Wu P.W., Finnerty H.F., Bean K.M., Spaulding V., Fouser L.A.,
RA   Leonard J.P., Hunter S.E., Zollner R., Thomas J.L., Miyashiro J.S.,
RA   Jacobs K.A., Collins M.;
RT   "Identification of GL50, a novel B7-like protein that functionally
RT   binds to ICOS receptor.";
RL   J. Immunol. 164:1653-1657(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Ling V., Dunussi-Joannopolulos K.;
RT   "G150 molecules and uses therefor.";
RL   Patent number WO0121796, 29-MAR-2001.
```

```
CC   -!- FUNCTION: LIGAND FOR THE T-CELL-SPECIFIC CELL SURFACE RECEPTOR
CC       ICOS. ACTS AS A COSTIMULATORY SIGNAL FOR T-CELL PROLIFERATION AND
CC       CYTOKINE SECRETION; INDUCES ALSO B-CELL PROLIFERATION AND
CC       DIFFERENTIATION INTO PLASMA CELLS. COULD PLAY AN IMPORTANT ROLE IN
CC       MEDIATING LOCAL TISSUE RESPONSES TO INFLAMMATORY CONDITIONS, AS
CC       WELL AS IN MODULATING THE SECONDARY IMMUNE RESPONSE BY CO-
CC       STIMULATING MEMORY T-CELL FUNCTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ISOFORM 1 IS WIDELY EXPRESSED (BRAIN, HEART,
CC       KIDNEY, LIVER, LUNG, PANCREAS, PLACENTA, SKELETAL MUSCLE, BONE
CC       MARROW, COLON, OVARY, PROSTATE, TESTIS, LYMPH NODES, LEUKOCYTES,
CC       SPLEEN, THYMUS AND TONSIL), WHILE ISOFORM 2 IS DETECTED ONLY IN
CC       LYMPH NODES, LEUKOCYTES AND SPLEEN.
CC   -!- INDUCTION: CONSTITUTIVE EXPRESSION IS FURTHER ENHANCED BY
CC       TREATMENT WITH TNF-ALPHA IN PERIPHERAL BLOOD B-CELLS AND
CC       MONOCYTES, WHILE IT IS DECREASED IN DENDRITIC CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. BTN/MOG
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- CAUTION: Ref.4 sequence differs from that shown in position 300
CC       onward for an unknown reason.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF199028; AAF34739.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF289028; AAG01176.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF216749; AAK16241.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AB014553; BAA31628.1; ALT_SEQ. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AX100595; CAC36465.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   MIM; 605717; -. [NCBI / EBI]
DR   GeneCards; ICOSL.
DR   GeneLynx; ICOSL.
DR   Ensembl; O75144.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig MHC.
DR   InterPro; IPR003600; Ig like.
DR   InterPro; Graphical view of domain structure.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00410; IG like; 1.
DR   ProDom [Domain structure / List of seq. sharing at least 1 domain]
DR   BLOCKS; O75144.
DR   ProtoMap; O75144.
DR   PRESAGE; O75144.
DR   DIP; O75144.
DR   ModBase; O75144.
DR   HUGE; KIAA0653; -.
DR   SWISS-2DPAGE; GET REGION ON 2D PAGE.
KW   B-cell activation; Immune response; Glycoprotein;
KW   Immunoglobulin domain; Signal; Transmembrane; Multigene family;
KW   Alternative splicing.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   302       ICOS LIGAND.
FT   DOMAIN      19   256       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   257   277       POTENTIAL.
FT   DOMAIN     278   302       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      30   120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     151   223       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    37   113       POTENTIAL.
FT   DISULFID   158   216       POTENTIAL.
FT   CARBOHYD    70    70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   137   137       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   173   173       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   186   186       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   225   225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   300   302       GHV -> ESWNLLLLLS (IN ISOFORM 2).
SQ   SEQUENCE   302 AA;  33349 MW;  647934E21B55E34A CRC64;
     MRLGSPGLLF LLFSSLRADT QEKEVRAMVG SDVELSCACP EGSRFDLNDV YVYWQTSESK
     TVVTYHIPQN SSLENVDSRY RNRALMSPAG MLRGDFSLRL FNVTPQDEQK FHCLVLSQSL
     GFQEVLSVEV TLHVAANFSV PVVSAPHSPS QDELTFTCTS INGYPRPNVY WINKTDNSLL
     DQALQNDTVF LNMRGLYDVV SVLRIARTPS VNIGCCIENV LLQQNLTVGS QTGNDIGERD
     KITENPVSTG EKNAATWSIL AVLCLLVVVA VAIGWVCRDR CLQHSYAGAW AVSPETELTG
     HV
//
```

sequence

SWISS-PROT: O75144

NiceProt - a user-friendly view of this SWISS-PROT entry

Protein name

Gene name

Taxonomy

```
ID   ICOL_HUMAN     STANDARD;      PRT;   302 AA.
AC   O75144; Q9NRQ1; Q9HD18;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ICOS ligand precursor (B7 homolog 2) (B7-H2) (B7-like protein G150)
DE   (B7-related protein-1) (B7RP-1).
GN   ICOSL OR B7H2 OR B7RP1 OR KIAA0653.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Dendritic cell;
RX   MEDLINE=20477846; PubMed=11023515; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Wang S., Zhu G., Chapoval A.I., Dong H., Tamada K., Ni J., Chen L.;
RT   "Costimulation of T cells by B7-H2, a B7-like molecule that binds
RT   ICOS.";
RL   Blood 96:2808-2813(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
RC   TISSUE=Peripheral blood lymphocytes;
RX   MEDLINE=20465019; PubMed=11007762; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Yoshinaga S.K., Zhang M., Pistillo J., Horan T., Khare S.D., Miner K.,
RA   Sonnenberg M., Boone T., Brankow D., Dai T., Delaney J., Han H.,
RA   Hui A., Kohno T., Manoukian R., Whoriskey J.S., Coccia M.A.;
RT   "Characterization of a new human B7-related protein: B7RP-1 is the
RT   ligand to the co-stimulatory protein ICOS.";
RL   Int. Immunol. 12:1439-1447(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Leukocyte;
RX   MEDLINE=20126021; PubMed=10657606; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ling V., Wu P.W., Finnerty H.F., Bean K.M., Spaulding V., Fouser L.A.,
RA   Leonard J.P., Hunter S.E., Zollner R., Thomas J.L., Miyashiro J.S.,
RA   Jacobs K.A., Collins M.;
RT   "Identification of GL50, a novel B7-like protein that functionally
RT   binds to ICOS receptor.";
RL   J. Immunol. 164:1653-1657(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Ling V., Dunussi-Joannopolulos K.;
RT   "G150 molecules and uses therefor.";
RL   Patent number WO0121796, 29-MAR-2001.
```

```
CC   -!- FUNCTION: LIGAND FOR THE T-CELL-SPECIFIC CELL SURFACE RECEPTOR
CC       ICOS. ACTS AS A COSTIMULATORY SIGNAL FOR T-CELL PROLIFERATION AND
CC       CYTOKINE SECRETION; INDUCES ALSO B-CELL PROLIFERATION AND
CC       DIFFERENTIATION INTO PLASMA CELLS. COULD PLAY AN IMPORTANT ROLE IN
CC       MEDIATING LOCAL TISSUE RESPONSES TO INFLAMMATORY CONDITIONS, AS
CC       WELL AS IN MODULATING THE SECONDARY IMMUNE RESPONSE BY CO-
CC       STIMULATING MEMORY T-CELL FUNCTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ISOFORM 1 IS WIDELY EXPRESSED (BRAIN, HEART,
CC       KIDNEY, LIVER, LUNG, PANCREAS, PLACENTA, SKELETAL MUSCLE, BONE
CC       MARROW, COLON, OVARY, PROSTATE, TESTIS, LYMPH NODES, LEUKOCYTES,
CC       SPLEEN, THYMUS AND TONSIL), WHILE ISOFORM 2 IS DETECTED ONLY IN
CC       LYMPH NODES, LEUKOCYTES AND SPLEEN.
CC   -!- INDUCTION: CONSTITUTIVE EXPRESSION IS FURTHER ENHANCED BY
CC       TREATMENT WITH TNF-ALPHA IN PERIPHERAL BLOOD B-CELLS AND
CC       MONOCYTES, WHILE IT IS DECREASED IN DENDRITIC CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. BTN/MOG
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- CAUTION: Ref.4 sequence differs from that shown in position 300
CC       onward for an unknown reason.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF199028; AAF34739.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF289028; AAG01176.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF216749; AAK16241.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AB014553; BAA31628.1; ALT_SEQ. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AX100595; CAC36465.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   MIM; 605717; -. [NCBI / EBI]
DR   GeneCards; ICOSL.
DR   GeneLynx; ICOSL.
DR   Ensembl; O75144.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig MHC.
DR   InterPro; IPR003600; Ig like.
DR   InterPro; Graphical view of domain structure.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00410; IG like; 1.
DR   ProDom [Domain structure / List of seq. sharing at least 1 domain]
DR   BLOCKS; O75144.
DR   ProtoMap; O75144.
DR   PRESAGE; O75144.
DR   DIP; O75144.
DR   ModBase; O75144.
DR   HUGE; KIAA0653; -.
DR   SWISS-2DPAGE; GET REGION ON 2D PAGE.
KW   B-cell activation; Immune response; Glycoprotein;
KW   Immunoglobulin domain; Signal; Transmembrane; Multigene family;
KW   Alternative splicing.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   302       ICOS LIGAND.
FT   DOMAIN      19   256       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   257   277       POTENTIAL.
FT   DOMAIN     278   302       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      30   120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     151   223       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    37   113       POTENTIAL.
FT   DISULFID   158   216       POTENTIAL.
FT   CARBOHYD    70    70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   137   137       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   173   173       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   186   186       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   225   225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   300   302       GHV -> ESWNLLLLLS (IN ISOFORM 2).
SQ   SEQUENCE   302 AA;  33349 MW;  647934E21B55E34A CRC64;
     MRLGSPGLLF LLFSSLRADT QEKEVRAMVG SDVELSCACP EGSRFDLNDV YVYWQTSESK
     TVVTYHIPQN SSLENVDSRY RNRALMSPAG MLRGDFSLRL FNVTPQDEQK FHCLVLSQSL
     GFQEVLSVEV TLHVAANFSV PVVSAPHSPS QDELTFTCTS INGYPRPNVY WINKTDNSLL
     DQALQNDTVF LNMRGLYDVV SVLRIARTPS VNIGCCIENV LLQQNLTVGS QTGNDIGERD
     KITENPVSTG EKNAATWSIL AVLCLLVVVA VAIGWVCRDR CLQHSYAGAW AVSPETELTG
     HV
//
```

SWISS-PROT: O75144

NiceProt - a user-friendly view of this SWISS-PROT entry

```
ID   ICOL_HUMAN     STANDARD;      PRT;   302 AA.
AC   O75144; Q9NRQ1; Q9HD18;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ICOS ligand precursor (B7 homolog 2) (B7-H2) (B7-like protein G150)
DE   (B7-related protein-1) (B7RP-1).
GN   ICOSL OR B7H2 OR B7RP1 OR KIAA0653.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Dendritic cell;
RX   MEDLINE=20477846; PubMed=11023515; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Wang S., Zhu G., Chapoval A.I., Dong H., Tamada K., Ni J., Chen L.;
RT   "Costimulation of T cells by B7-H2, a B7-like molecule that binds
RT   ICOS.";
RL   Blood 96:2808-2813(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
RC   TISSUE=Peripheral blood lymphocytes;
RX   MEDLINE=20465019; PubMed=11007762; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Yoshinaga S.K., Zhang M., Pistillo J., Horan T., Khare S.D., Miner K.,
RA   Sonnenberg M., Boone T., Brankow D., Dai T., Delaney J., Han H.,
RA   Hui A., Kohno T., Manoukian R., Whoriskey J.S., Coccia M.A.;
RT   "Characterization of a new human B7-related protein: B7RP-1 is the
RT   ligand to the co-stimulatory protein ICOS.";
RL   Int. Immunol. 12:1439-1447(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Leukocyte;
RX   MEDLINE=20126021; PubMed=10657606; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ling V., Wu P.W., Finnerty H.F., Bean K.M., Spaulding V., Fouser L.A.,
RA   Leonard J.P., Hunter S.E., Zollner R., Thomas J.L., Miyashiro J.S.,
RA   Jacobs K.A., Collins M.;
RT   "Identification of GL50, a novel B7-like protein that functionally
RT   binds to ICOS receptor.";
RL   J. Immunol. 164:1653-1657(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Ling V., Dunussi-Joannopolulos K.;
RT   "G150 molecules and uses therefor.";
RL   Patent number WO0121796, 29-MAR-2001.
```

References

```
CC   -!- FUNCTION: LIGAND FOR THE T-CELL-SPECIFIC CELL SURFACE RECEPTOR
CC       ICOS. ACTS AS A COSTIMULATORY SIGNAL FOR T-CELL PROLIFERATION AND
CC       CYTOKINE SECRETION; INDUCES ALSO B-CELL PROLIFERATION AND
CC       DIFFERENTIATION INTO PLASMA CELLS. COULD PLAY AN IMPORTANT ROLE IN
CC       MEDIATING LOCAL TISSUE RESPONSES TO INFLAMMATORY CONDITIONS, AS
CC       WELL AS IN MODULATING THE SECONDARY IMMUNE RESPONSE BY CO-
CC       STIMULATING MEMORY T-CELL FUNCTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ISOFORM 1 IS WIDELY EXPRESSED (BRAIN, HEART,
CC       KIDNEY, LIVER, LUNG, PANCREAS, PLACENTA, SKELETAL MUSCLE, BONE
CC       MARROW, COLON, OVARY, PROSTATE, TESTIS, LYMPH NODES, LEUKOCYTES,
CC       SPLEEN, THYMUS AND TONSIL), WHILE ISOFORM 2 IS DETECTED ONLY IN
CC       LYMPH NODES, LEUKOCYTES AND SPLEEN.
CC   -!- INDUCTION: CONSTITUTIVE EXPRESSION IS FURTHER ENHANCED BY
CC       TREATMENT WITH TNF-ALPHA IN PERIPHERAL BLOOD B-CELLS AND
CC       MONOCYTES, WHILE IT IS DECREASED IN DENDRITIC CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. BTN/MOG
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- CAUTION: Ref.4 sequence differs from that shown in position 300
CC       onward for an unknown reason.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF199028; AAF34739.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF289028; AAG01176.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF216749; AAK16241.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AB014553; BAA31628.1; ALT_SEQ. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AX100595; CAC36465.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   MIM; 605717; -. [NCBI / EBI]
DR   GeneCards; ICOSL.
DR   GeneLynx; ICOSL.
DR   Ensembl; O75144.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig MHC.
DR   InterPro; IPR003600; Ig like.
DR   InterPro; Graphical view of domain structure.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00410; IG like; 1.
DR   ProDom [Domain structure / List of seq. sharing at least 1 domain]
DR   BLOCKS; O75144.
DR   ProtoMap; O75144.
DR   PRESAGE; O75144.
DR   DIP; O75144.
DR   ModBase; O75144.
DR   HUGE; KIAA0653; -.
DR   SWISS-2DPAGE; GET REGION ON 2D PAGE.
KW   B-cell activation; Immune response; Glycoprotein;
KW   Immunoglobulin domain; Signal; Transmembrane; Multigene family;
KW   Alternative splicing.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   302       ICOS LIGAND.
FT   DOMAIN      19   256       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   257   277       POTENTIAL.
FT   DOMAIN     278   302       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      30   120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     151   223       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    37   113       POTENTIAL.
FT   DISULFID   158   216       POTENTIAL.
FT   CARBOHYD    70    70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   137   137       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   173   173       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   186   186       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   225   225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   300   302       GHV -> ESWNLLLLLS (IN ISOFORM 2).
SQ   SEQUENCE   302 AA;  33349 MW;  647934E21B55E34A CRC64;
     MRLGSPGLLF LLFSSLRADT QEKEVRAMVG SDVELSCACP EGSRFDLNDV YVYWQTSESK
     TVVTYHIPQN SSLENVDSRY RNRALMSPAG MLRGDFSLRL FNVTPQDEQK FHCLVLSQSL
     GFQEVLSVEV TLHVAANFSV PVVSAPHSPS QDELTFTCTS INGYPRPNVY WINKTDNSLL
     DQALQNDTVF LNMRGLYDVV SVLRIARTPS VNIGCCIENV LLQQNLTVGS QTGNDIGERD
     KITENPVSTG EKNAATWSIL AVLCLLVVVA VAIGWVCRDR CLQHSYAGAW AVSPETELTG
     HV
//
```

SWISS-PROT: O75144

NiceProt - a user-friendly view of this SWISS-PROT entry

```
ID   ICOL_HUMAN     STANDARD;      PRT;   302 AA.
AC   O75144; Q9NRQ1; Q9HD18;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ICOS ligand precursor (B7 homolog 2) (B7-H2) (B7-like protein G150)
DE   (B7-related protein-1) (B7RP-1).
GN   ICOSL OR B7H2 OR B7RP1 OR KIAA0653.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Dendritic cell;
RX   MEDLINE=20477846; PubMed=11023515; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Wang S., Zhu G., Chapoval A.I., Dong H., Tamada K., Ni J., Chen L.;
RT   "Costimulation of T cells by B7-H2, a B7-like molecule that binds
RT   ICOS.";
RL   Blood 96:2808-2813(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
RC   TISSUE=Peripheral blood lymphocytes;
RX   MEDLINE=20465019; PubMed=11007762; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Yoshinaga S.K., Zhang M., Pistillo J., Horan T., Khare S.D., Miner K.,
RA   Sonnenberg M., Boone T., Brankow D., Dai T., Delaney J., Han H.,
RA   Hui A., Kohno T., Manoukian R., Whoriskey J.S., Coccia M.A.;
RT   "Characterization of a new human B7-related protein: B7RP-1 is the
RT   ligand to the co-stimulatory protein ICOS.";
RL   Int. Immunol. 12:1439-1447(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Leukocyte;
RX   MEDLINE=20126021; PubMed=10657606; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ling V., Wu P.W., Finnerty H.F., Bean K.M., Spaulding V., Fouser L.A.,
RA   Leonard J.P., Hunter S.E., Zollner R., Thomas J.L., Miyashiro J.S.,
RA   Jacobs K.A., Collins M.;
RT   "Identification of GL50, a novel B7-like protein that functionally
RT   binds to ICOS receptor.";
RL   J. Immunol. 164:1653-1657(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Ling V., Dunussi-Joannopolulos K.;
RT   "G150 molecules and uses therefor.";
RL   Patent number WO0121796, 29-MAR-2001.
```

Comments

```
CC   -!- FUNCTION: LIGAND FOR THE T-CELL-SPECIFIC CELL SURFACE RECEPTOR
CC       ICOS. ACTS AS A COSTIMULATORY SIGNAL FOR T-CELL PROLIFERATION AND
CC       CYTOKINE SECRETION; INDUCES ALSO B-CELL PROLIFERATION AND
CC       DIFFERENTIATION INTO PLASMA CELLS. COULD PLAY AN IMPORTANT ROLE IN
CC       MEDIATING LOCAL TISSUE RESPONSES TO INFLAMMATORY CONDITIONS, AS
CC       WELL AS IN MODULATING THE SECONDARY IMMUNE RESPONSE BY CO-
CC       STIMULATING MEMORY T-CELL FUNCTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ISOFORM 1 IS WIDELY EXPRESSED (BRAIN, HEART,
CC       KIDNEY, LIVER, LUNG, PANCREAS, PLACENTA, SKELETAL MUSCLE, BONE
CC       MARROW, COLON, OVARY, PROSTATE, TESTIS, LYMPH NODES, LEUKOCYTES,
CC       SPLEEN, THYMUS AND TONSIL), WHILE ISOFORM 2 IS DETECTED ONLY IN
CC       LYMPH NODES, LEUKOCYTES AND SPLEEN.
CC   -!- INDUCTION: CONSTITUTIVE EXPRESSION IS FURTHER ENHANCED BY
CC       TREATMENT WITH TNF-ALPHA IN PERIPHERAL BLOOD B-CELLS AND
CC       MONOCYTES, WHILE IT IS DECREASED IN DENDRITIC CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. BTN/MOG
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- CAUTION: Ref.4 sequence differs from that shown in position 300
CC       onward for an unknown reason.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF199028; AAF34739.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF289028; AAG01176.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF216749; AAK16241.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AB014553; BAA31628.1; ALT_SEQ. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AX100595; CAC36465.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   MIM; 605717; -. [NCBI / EBI]
DR   GeneCards; ICOSL.
DR   GeneLynx; ICOSL.
DR   Ensembl; O75144.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig MHC.
DR   InterPro; IPR003600; Ig like.
DR   InterPro; Graphical view of domain structure.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00410; IG like; 1.
DR   ProDom [Domain structure / List of seq. sharing at least 1 domain]
DR   BLOCKS; O75144.
DR   ProtoMap; O75144.
DR   PRESAGE; O75144.
DR   DIP; O75144.
DR   ModBase; O75144.
DR   HUGE; KIAA0653; -.
DR   SWISS-2DPAGE; GET REGION ON 2D PAGE.
KW   B-cell activation; Immune response; Glycoprotein;
KW   Immunoglobulin domain; Signal; Transmembrane; Multigene family;
KW   Alternative splicing.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   302       ICOS LIGAND.
FT   DOMAIN      19   256       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   257   277       POTENTIAL.
FT   DOMAIN     278   302       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      30   120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     151   223       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    37   113       POTENTIAL.
FT   DISULFID   158   216       POTENTIAL.
FT   CARBOHYD    70    70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   137   137       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   173   173       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   186   186       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   225   225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   300   302       GHV -> ESWNLLLLLS (IN ISOFORM 2).
SQ   SEQUENCE   302 AA;  33349 MW;  647934E21B55E34A CRC64;
     MRLGSPGLLF LLFSSLRADT QEKEVRAMVG SDVELSCACP EGSRFDLNDV YVYWQTSESK
     TVVTYHIPQN SSLENVDSRY RNRALMSPAG MLRGDFSLRL FNVTPQDEQK FHCLVLSQSL
     GFQEVLSVEV TLHVAANFSV PVVSAPHSPS QDELTFTCTS INGYPRPNVY WINKTDNSLL
     DQALQNDTVF LNMRGLYDVV SVLRIARTPS VNIGCCIENV LLQQNLTVGS QTGNDIGERD
     KITENPVSTG EKNAATWSIL AVLCLLVVVA VAIGWVCRDR CLQHSYAGAW AVSPETELTG
     HV
//
```

SWISS-PROT: O75144

NiceProt - a user-friendly view of this SWISS-PROT entry

```
ID   ICOL_HUMAN     STANDARD;      PRT;   302 AA.
AC   O75144; Q9NRQ1; Q9HD18;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ICOS ligand precursor (B7 homolog 2) (B7-H2) (B7-like protein G150)
DE   (B7-related protein-1) (B7RP-1).
GN   ICOSL OR B7H2 OR B7RP1 OR KIAA0653.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Dendritic cell;
RX   MEDLINE=20477846; PubMed=11023515; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Wang S., Zhu G., Chapoval A.I., Dong H., Tamada K., Ni J., Chen L.;
RT   "Costimulation of T cells by B7-H2, a B7-like molecule that binds
RT   ICOS.";
RL   Blood 96:2808-2813(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
RC   TISSUE=Peripheral blood lymphocytes;
RX   MEDLINE=20465019; PubMed=11007762; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Yoshinaga S.K., Zhang M., Pistillo J., Horan T., Khare S.D., Miner K.,
RA   Sonnenberg M., Boone T., Brankow D., Dai T., Delaney J., Han H.,
RA   Hui A., Kohno T., Manoukian R., Whoriskey J.S., Coccia M.A.;
RT   "Characterization of a new human B7-related protein: B7RP-1 is the
RT   ligand to the co-stimulatory protein ICOS.";
RL   Int. Immunol. 12:1439-1447(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Leukocyte;
RX   MEDLINE=20126021; PubMed=10657606; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ling V., Wu P.W., Finnerty H.F., Bean K.M., Spaulding V., Fouser L.A.,
RA   Leonard J.P., Hunter S.E., Zollner R., Thomas J.L., Miyashiro J.S.,
RA   Jacobs K.A., Collins M.;
RT   "Identification of GL50, a novel B7-like protein that functionally
RT   binds to ICOS receptor.";
RL   J. Immunol. 164:1653-1657(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Ling V., Dunussi-Joannopolulos K.;
RT   "G150 molecules and uses therefor.";
RL   Patent number WO0121796, 29-MAR-2001.
```

Cross-references

```
CC   -!- FUNCTION: LIGAND FOR THE T-CELL-SPECIFIC CELL SURFACE RECEPTOR
CC       ICOS. ACTS AS A COSTIMULATORY SIGNAL FOR T-CELL PROLIFERATION AND
CC       CYTOKINE SECRETION; INDUCES ALSO B-CELL PROLIFERATION AND
CC       DIFFERENTIATION INTO PLASMA CELLS. COULD PLAY AN IMPORTANT ROLE IN
CC       MEDIATING LOCAL TISSUE RESPONSES TO INFLAMMATORY CONDITIONS, AS
CC       WELL AS IN MODULATING THE SECONDARY IMMUNE RESPONSE BY CO-
CC       STIMULATING MEMORY T-CELL FUNCTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ISOFORM 1 IS WIDELY EXPRESSED (BRAIN, HEART,
CC       KIDNEY, LIVER, LUNG, PANCREAS, PLACENTA, SKELETAL MUSCLE, BONE
CC       MARROW, COLON, OVARY, PROSTATE, TESTIS, LYMPH NODES, LEUKOCYTES,
CC       SPLEEN, THYMUS AND TONSIL), WHILE ISOFORM 2 IS DETECTED ONLY IN
CC       LYMPH NODES, LEUKOCYTES AND SPLEEN.
CC   -!- INDUCTION: CONSTITUTIVE EXPRESSION IS FURTHER ENHANCED BY
CC       TREATMENT WITH TNF-ALPHA IN PERIPHERAL BLOOD B-CELLS AND
CC       MONOCYTES, WHILE IT IS DECREASED IN DENDRITIC CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. BTN/MOG
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- CAUTION: Ref.4 sequence differs from that shown in position 300
CC       onward for an unknown reason.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF199028; AAF34739.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF289028; AAG01176.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF216749; AAK16241.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AB014553; BAA31628.1; ALT_SEQ. [EMBL / GenBank / DDBJ] [CoDingSequence
DR   EMBL; AX100595; CAC36465.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   MIM; 605717; -. [NCBI / EBI]
DR   GeneCards; ICOSL.
DR   GeneLynx; ICOSL.
DR   Ensembl; O75144.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig MHC.
DR   InterPro; IPR003600; Ig like.
DR   InterPro; Graphical view of domain structure.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00410; IG like; 1.
DR   ProDom [Domain structure / List of seq. sharing at least 1 domain]
DR   BLOCKS; O75144.
DR   ProtoMap; O75144.
DR   PRESAGE; O75144.
DR   DIP; O75144.
DR   ModBase; O75144.
DR   HUGE; KIAA0653; -.
DR   SWISS-2DPAGE; GET REGION ON 2D PAGE.
KW   B-cell activation; Immune response; Glycoprotein;
KW   Immunoglobulin domain; Signal; Transmembrane; Multigene family;
KW   Alternative splicing.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   302       ICOS LIGAND.
FT   DOMAIN      19   256       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   257   277       POTENTIAL.
FT   DOMAIN     278   302       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      30   120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     151   223       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    37   113       POTENTIAL.
FT   DISULFID   158   216       POTENTIAL.
FT   CARBOHYD    70    70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   137   137       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   173   173       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   186   186       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   225   225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   300   302       GHV -> ESWNLLLLLS (IN ISOFORM 2).
SQ   SEQUENCE   302 AA;  33349 MW;  647934E21B55E34A CRC64;
     MRLGSPGLLF LLFSSLRADT QEKEVRAMVG SDVELSCACP EGSRFDLNDV YVYWQTSESK
     TVVTYHIPQN SSLENVDSRY RNRALMSPAG MLRGDFSLRL FNVTPQDEQK FHCLVLSQSL
     GFQEVLSVEV TLHVAANFSV PVVSAPHSPS QDELTFTCTS INGYPRPNVY WINKTDNSLL
     DQALQNDTVF LNMRGLYDVV SVLRIARTPS VNIGCCIENV LLQQNLTVGS QTGNDIGERD
     KITENPVSTG EKNAATWSIL AVLCLLVVVA VAIGWVCRDR CLQHSYAGAW AVSPETELTG
     HV
//
```

SWISS-PROT: O75144

NiceProt - a user-friendly view of this SWISS-PROT entry

```
ID   ICOL_HUMAN     STANDARD;      PRT;   302 AA.
AC   O75144; Q9NRQ1; Q9HD18;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ICOS ligand precursor (B7 homolog 2) (B7-H2) (B7-like protein G150)
DE   (B7-related protein-1) (B7RP-1).
GN   ICOSL OR B7H2 OR B7RP1 OR KIAA0653.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Dendritic cell;
RX   MEDLINE=20477846; PubMed=11023515; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Wang S., Zhu G., Chapoval A.I., Dong H., Tamada K., Ni J., Chen L.;
RT   "Costimulation of T cells by B7-H2, a B7-like molecule that binds
RT   ICOS.";
RL   Blood 96:2808-2813(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
RC   TISSUE=Peripheral blood lymphocytes;
RX   MEDLINE=20465019; PubMed=11007762; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Yoshinaga S.K., Zhang M., Pistillo J., Horan T., Khare S.D., Miner K.,
RA   Sonnenberg M., Boone T., Brankow D., Dai T., Delaney J., Han H.,
RA   Hui A., Kohno T., Manoukian R., Whoriskey J.S., Coccia M.A.;
RT   "Characterization of a new human B7-related protein: B7RP-1 is the
RT   ligand to the co-stimulatory protein ICOS.";
RL   Int. Immunol. 12:1439-1447(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Leukocyte;
RX   MEDLINE=20126021; PubMed=10657606; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ling V., Wu P.W., Finnerty H.F., Bean K.M., Spaulding V., Fouser L.A.,
RA   Leonard J.P., Hunter S.E., Zollner R., Thomas J.L., Miyashiro J.S.,
RA   Jacobs K.A., Collins M.;
RT   "Identification of GL50, a novel B7-like protein that functionally
RT   binds to ICOS receptor.";
RL   J. Immunol. 164:1653-1657(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Ling V., Dunussi-Joannopolulos K.;
RT   "G150 molecules and uses therefor.";
RL   Patent number WO0121796, 29-MAR-2001.
```

Keywords

```
CC   -!- FUNCTION: LIGAND FOR THE T-CELL-SPECIFIC CELL SURFACE RECEPTOR
CC       ICOS. ACTS AS A COSTIMULATORY SIGNAL FOR T-CELL PROLIFERATION AND
CC       CYTOKINE SECRETION; INDUCES ALSO B-CELL PROLIFERATION AND
CC       DIFFERENTIATION INTO PLASMA CELLS. COULD PLAY AN IMPORTANT ROLE IN
CC       MEDIATING LOCAL TISSUE RESPONSES TO INFLAMMATORY CONDITIONS, AS
CC       WELL AS IN MODULATING THE SECONDARY IMMUNE RESPONSE BY CO-
CC       STIMULATING MEMORY T-CELL FUNCTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ISOFORM 1 IS WIDELY EXPRESSED (BRAIN, HEART,
CC       KIDNEY, LIVER, LUNG, PANCREAS, PLACENTA, SKELETAL MUSCLE, BONE
CC       MARROW, COLON, OVARY, PROSTATE, TESTIS, LYMPH NODES, LEUKOCYTES,
CC       SPLEEN, THYMUS AND TONSIL), WHILE ISOFORM 2 IS DETECTED ONLY IN
CC       LYMPH NODES, LEUKOCYTES AND SPLEEN.
CC   -!- INDUCTION: CONSTITUTIVE EXPRESSION IS FURTHER ENHANCED BY
CC       TREATMENT WITH TNF-ALPHA IN PERIPHERAL BLOOD B-CELLS AND
CC       MONOCYTES, WHILE IT IS DECREASED IN DENDRITIC CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. BTN/MOG
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- CAUTION: Ref.4 sequence differs from that shown in position 300
CC       onward for an unknown reason.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF199028; AAF34739.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF289028; AAG01176.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF216749; AAK16241.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AB014553; BAA31628.1; ALT_SEQ. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AX100595; CAC36465.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   MIM; 605717; -. [NCBI / EBI]
DR   GeneCards; ICOSL.
DR   GeneLynx; ICOSL.
DR   Ensembl; O75144.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig MHC.
DR   InterPro; IPR003600; Ig like.
DR   InterPro; Graphical view of domain structure.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00410; IG like; 1.
DR   ProDom [Domain structure / List of seq. sharing at least 1 domain]
DR   BLOCKS; O75144.
DR   ProtoMap; O75144.
DR   PRESAGE; O75144.
DR   DIP; O75144.
DR   ModBase; O75144.
DR   HUGE; KIAA0653; -.
DR   SWISS-2DPAGE; GET REGION ON 2D PAGE.
KW   B-cell activation; Immune response; Glycoprotein;
KW   Immunoglobulin domain; Signal; Transmembrane; Multigene family;
KW   Alternative splicing.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   302       ICOS LIGAND.
FT   DOMAIN      19   256       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   257   277       POTENTIAL.
FT   DOMAIN     278   302       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      30   120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     151   223       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    37   113       POTENTIAL.
FT   DISULFID   158   216       POTENTIAL.
FT   CARBOHYD    70    70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   137   137       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   173   173       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   186   186       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   225   225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   300   302       GHV -> ESWNLLLLLS (IN ISOFORM 2).
SQ   SEQUENCE   302 AA;  33349 MW;  647934E21B55E34A CRC64;
     MRLGSPGLLF LLFSSLRADT QEKEVRAMVG SDVELSCACP EGSRFDLNDV YVYWQTSESK
     TVVTYHIPQN SSLENVDSRY RNRALMSPAG MLRGDFSLRL FNVTPQDEQK FHCLVLSQSL
     GFQEVLSVEV TLHVAANFSV PVVSAPHSPS QDELTFTCTS INGYPRPNVY WINKTDNSLL
     DQALQNDTVF LNMRGLYDVV SVLRIARTPS VNIGCCIENV LLQQNLTVGS QTGNDIGERD
     KITENPVSTG EKNAATWSIL AVLCLLVVVA VAIGWVCRDR CLQHSYAGAW AVSPETELTG
     HV
//
```

SWISS-PROT: O75144

NiceProt - a user-friendly view of this SWISS-PROT entry

```
ID   ICOL_HUMAN     STANDARD;      PRT;   302 AA.
AC   O75144; Q9NRQ1; Q9HD18;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   ICOS ligand precursor (B7 homolog 2) (B7-H2) (B7-like protein G150)
DE   (B7-related protein-1) (B7RP-1).
GN   ICOSL OR B7H2 OR B7RP1 OR KIAA0653.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Dendritic cell;
RX   MEDLINE=20477846; PubMed=11023515; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Wang S., Zhu G., Chapoval A.I., Dong H., Tamada K., Ni J., Chen L.;
RT   "Costimulation of T cells by B7-H2, a B7-like molecule that binds
RT   ICOS.";
RL   Blood 96:2808-2813(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
RC   TISSUE=Peripheral blood lymphocytes;
RX   MEDLINE=20465019; PubMed=11007762; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Yoshinaga S.K., Zhang M., Pistillo J., Horan T., Khare S.D., Miner K.,
RA   Sonnenberg M., Boone T., Brankow D., Dai T., Delaney J., Han H.,
RA   Hui A., Kohno T., Manoukian R., Whoriskey J.S., Coccia M.A.;
RT   "Characterization of a new human B7-related protein: B7RP-1 is the
RT   ligand to the co-stimulatory protein ICOS.";
RL   Int. Immunol. 12:1439-1447(2000).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Leukocyte;
RX   MEDLINE=20126021; PubMed=10657606; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ling V., Wu P.W., Finnerty H.F., Bean K.M., Spaulding V., Fouser L.A.,
RA   Leonard J.P., Hunter S.E., Zollner R., Thomas J.L., Miyashiro J.S.,
RA   Jacobs K.A., Collins M.;
RT   "Identification of GL50, a novel B7-like protein that functionally
RT   binds to ICOS receptor.";
RL   J. Immunol. 164:1653-1657(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811; [NCBI, ExPASy, EBI, Israel, Japan]
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Ling V., Dunussi-Joannopolulos K.;
RT   "G150 molecules and uses therefor.";
RL   Patent number WO0121796, 29-MAR-2001.
```

Feature table (sequence description)

```
CC   -!- FUNCTION: LIGAND FOR THE T-CELL-SPECIFIC CELL SURFACE RECEPTOR
CC       ICOS. ACTS AS A COSTIMULATORY SIGNAL FOR T-CELL PROLIFERATION AND
CC       CYTOKINE SECRETION; INDUCES ALSO B-CELL PROLIFERATION AND
CC       DIFFERENTIATION INTO PLASMA CELLS. COULD PLAY AN IMPORTANT ROLE IN
CC       MEDIATING LOCAL TISSUE RESPONSES TO INFLAMMATORY CONDITIONS, AS
CC       WELL AS IN MODULATING THE SECONDARY IMMUNE RESPONSE BY CO-
CC       STIMULATING MEMORY T-CELL FUNCTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ISOFORM 1 IS WIDELY EXPRESSED (BRAIN, HEART,
CC       KIDNEY, LIVER, LUNG, PANCREAS, PLACENTA, SKELETAL MUSCLE, BONE
CC       MARROW, COLON, OVARY, PROSTATE, TESTIS, LYMPH NODES, LEUKOCYTES,
CC       SPLEEN, THYMUS AND TONSIL), WHILE ISOFORM 2 IS DETECTED ONLY IN
CC       LYMPH NODES, LEUKOCYTES AND SPLEEN.
CC   -!- INDUCTION: CONSTITUTIVE EXPRESSION IS FURTHER ENHANCED BY
CC       TREATMENT WITH TNF-ALPHA IN PERIPHERAL BLOOD B-CELLS AND
CC       MONOCYTES, WHILE IT IS DECREASED IN DENDRITIC CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. BTN/MOG
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- CAUTION: Ref.4 sequence differs from that shown in position 300
CC       onward for an unknown reason.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF199028; AAF34739.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF289028; AAG01176.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AF216749; AAK16241.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AB014553; BAA31628.1; ALT_SEQ. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   EMBL; AX100595; CAC36465.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   MIM; 605717; -. [NCBI / EBI]
DR   GeneCards; ICOSL.
DR   GeneLynx; ICOSL.
DR   Ensembl; O75144.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig MHC.
DR   InterPro; IPR003600; Ig like.
DR   InterPro; Graphical view of domain structure.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00410; IG like; 1.
DR   ProDom [Domain structure / List of seq. sharing at least 1 domain]
DR   BLOCKS; O75144.
DR   ProtoMap; O75144.
DR   PRESAGE; O75144.
DR   DIP; O75144.
DR   ModBase; O75144.
DR   HUGE; KIAA0653; -.
DR   SWISS-2DPAGE; GET REGION ON 2D PAGE.
KW   B-cell activation; Immune response; Glycoprotein;
KW   Immunoglobulin domain; Signal; Transmembrane; Multigene family;
KW   Alternative splicing.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   302       ICOS LIGAND.
FT   DOMAIN      19   256       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   257   277       POTENTIAL.
FT   DOMAIN     278   302       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      30   120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     151   223       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    37   113       POTENTIAL.
FT   DISULFID   158   216       POTENTIAL.
FT   CARBOHYD    70    70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   137   137       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   173   173       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   186   186       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   225   225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   300   302       GHV -> ESWNLLLLLS (IN ISOFORM 2).
SQ   SEQUENCE   302 AA;  33349 MW;  647934E21B55E34A CRC64;
     MRLGSPGLLF LLFSSLRADT QEKEVRAMVG SDVELSCACP EGSRFDLNDV YVYWQTSESK
     TVVTYHIPQN SSLENVDSRY RNRALMSPAG MLRGDFSLRL FNVTPQDEQK FHCLVLSQSL
     GFQEVLSVEV TLHVAANFSV PVVSAPHSPS QDELTFTCTS INGYPRPNVY WINKTDNSLL
     DQALQNDTVF LNMRGLYDVV SVLRIARTPS VNIGCCIENV LLQQNLTVGS QTGNDIGERD
     KITENPVSTG EKNAATWSIL AVLCLLVVVA VAIGWVCRDR CLQHSYAGAW AVSPETELTG
     HV
//
```

Sequence database: example

…a SWISS-PROT entry, in fasta format:

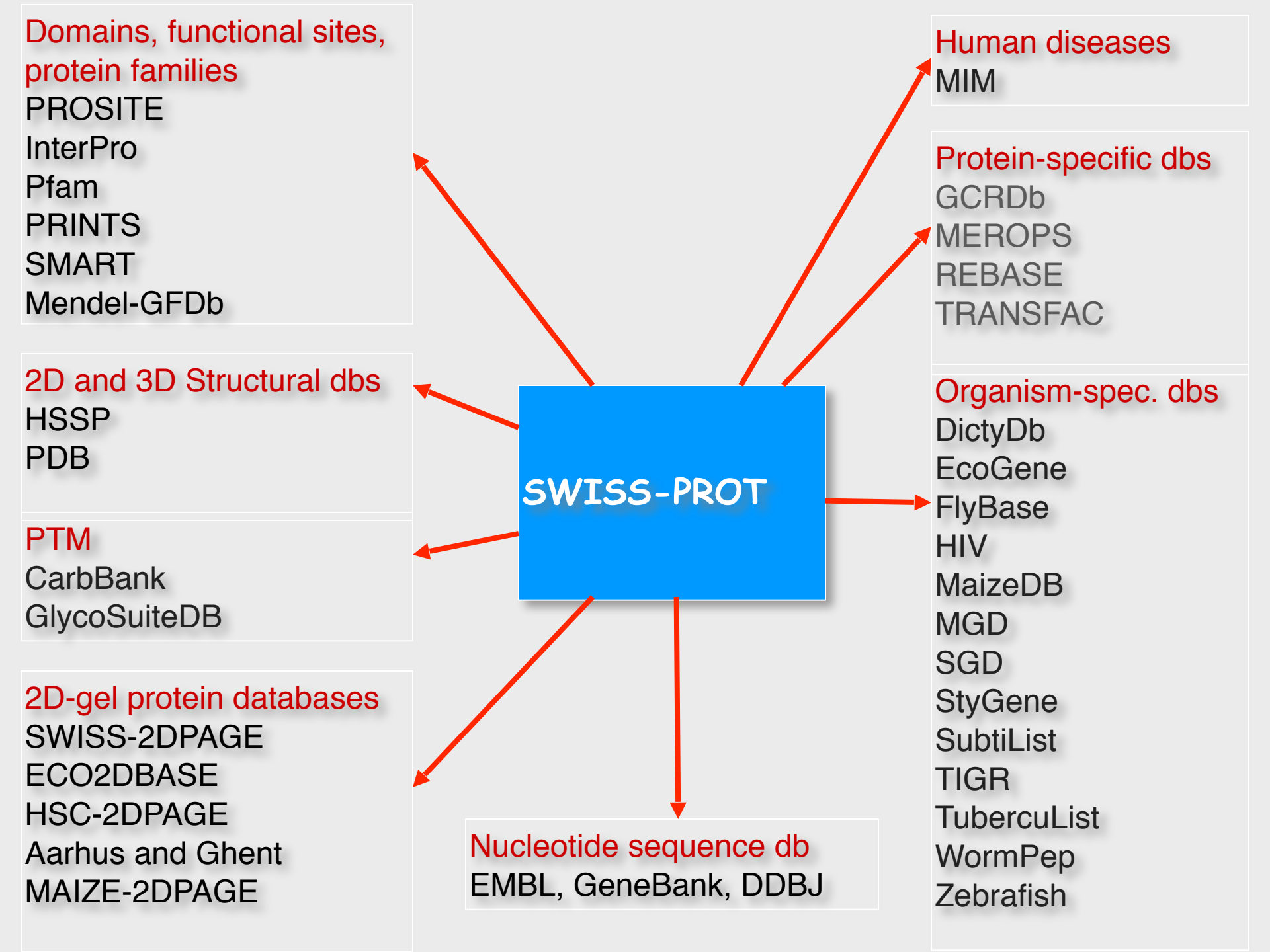
```
>sp|P01588|EPO_HUMAN ERYTHROPOIETIN PRECURSOR - Homo sapiens(Human).
MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE
NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEA
VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPD
AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR
```

SWISS-PROT knowledgebase

- Created by Amos Bairoch in 1986
- Collaboration between the SIB (CH) and EBI (UK)
- Annotated (manually), non-redundant, cross-referenced, documented protein sequence database.
- ~122 ’000 sequences from more than 7’700 different species; 192 ’000 references (publications); 958 ’000 cross-references (databases); ~400 Mb of annotations.
- Weekly releases; available from more than 50 servers across the world, the main source being ExPASy

SWISS-PROT: species

- 7’700 different species
- 20 species represent about 42% of all sequences in the database
- 5’000 species are only represented by one to three sequences. In most cases, these are sequences which were obtained in the context of a phylogenetic study

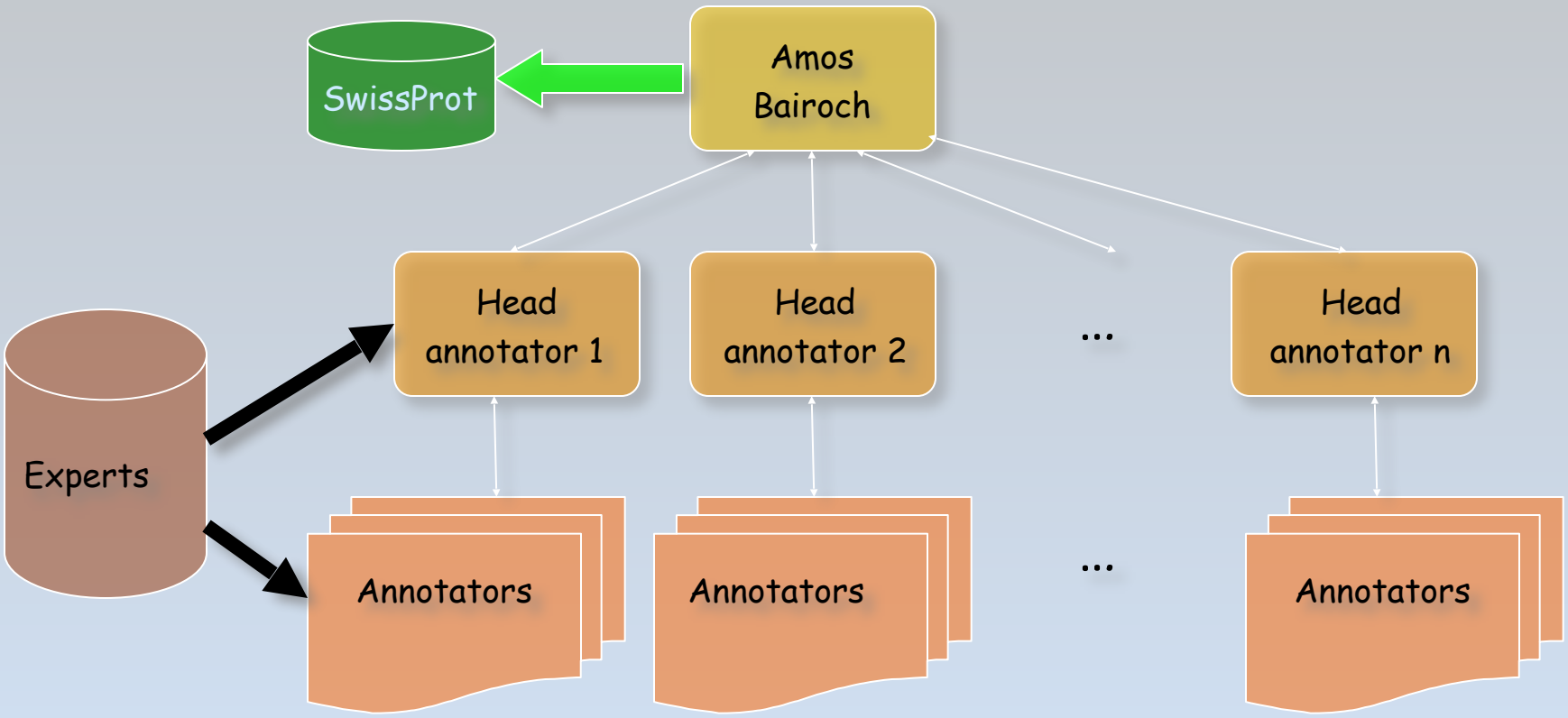

Domains, functional sites, protein families
PROSITE
InterPro
Pfam
PRINTS
SMART
Mendel-GFDb
2D and 3D Structural dbs
HSSP
PDB
PTM
CarbBank
GlycoSuiteDB
2D-gel protein databases
SWISS-2DPAGE
ECO2DBASE
HSC-2DPAGE
Aarhus and Ghent
MAIZE-2DPAGE
SWISS-PROT
Nucleotide sequence db
EMBL, GeneBank, DDBJ
Human diseases
MIM
Protein-specific dbs
GCRDb
MEROPS
REBASE
TRANSFAC
Organism-spec. dbs
DictyDb
EcoGene
FlyBase
HIV
MaizeDB
MGD
SGD
StyGene
SubtiList
TIGR
TubercuList
WormPep
Zebrafish

Annotations

- Function(s)
- Post-translational modifications (PTM)
- Domains
- Quaternary structure
- Similarities
- Diseases, mutagenesis
- Conflicts, variants
- Cross-references

…

Annotation schema



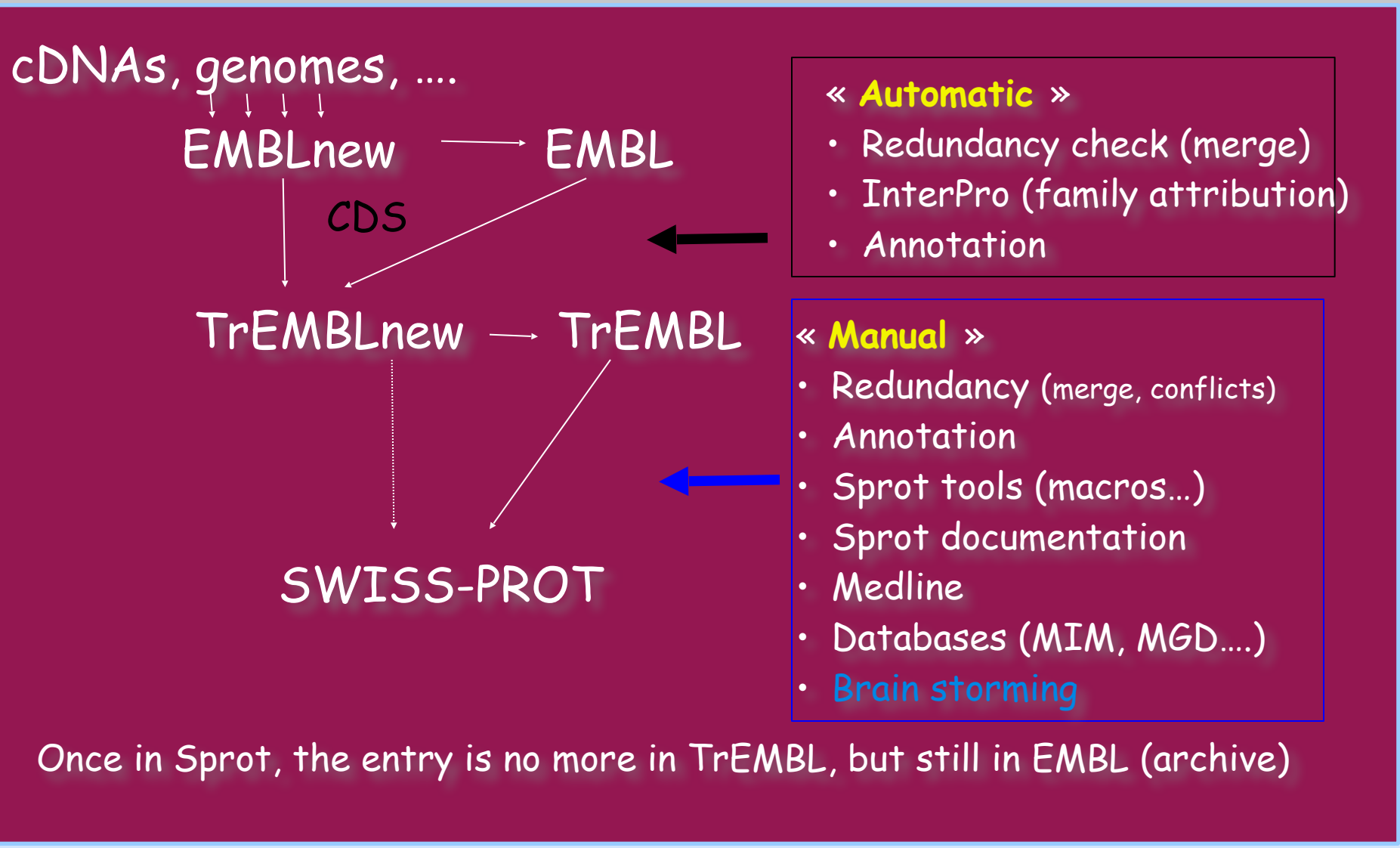
Manual annotation

Code	Content	Occurrence in an entry
ID	Identification	One; starts the entry
AC	Accession number(s)	One or more
DT	Date	Three times
DE	Description	One or more
GN	Gene name(s)	Optional
OS	Organism species	One or more
OG	Organelle	Optional
OC	Organism classification	One or more
OX	Taxonomy cross-references	One or more
RN	Reference number	One or more
RP	Reference position	One or more
RC	Reference comment(s)	Optional
RX	Reference cross-reference(s)	Optional
RA	Reference authors	One or more
RT	Reference title	Optional
RL	Reference location	One or more
CC	Comments or notes	Optional
DR	Database cross-references	Optional
KW	Keywords	Optional
FT	Feature table data	Optional
SQ	Sequence header	One
	Amino Acid Sequence	One or more
//	Termination line	One; ends the entry

TrEMBL (Translated EMBL)

- **TrEMBL**: created in 1996;
- Computer-annotated supplement to SWISS-PROT, as it is impossible to cope with the flow of data…
- Well-structure SWISS-PROT-like resource
- Derived from automated EMBL CDS translation (maintained at the EBI (UK))
- TrEMBL is automatically generated and annotated using software tools (incompatible with the SWISS-PROT in terms of quality)
- TrEMBL contains all what is **not yet** in SWISS-PROT
- Yerk!! But there is no choice and these software tools are becoming quite good !

The simplified story of a Sprot entry



Once in Sprot, the entry is no more in TrEMBL, but still in EMBL (archive)

TrEMBL: example

TrEMBL: Q9UDZ0

```
ID   Q9UDZ0      PRELIMINARY;      PRT;   136 AA.
AC   Q9UDZ0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ERYTHROPOIETIN PROTEIN (FRAGMENT).
GN   ERYTHROPOIETIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93384593. [NCBI, ExPASy, Israel, Japan]
RA   Funakoshi A., Muta H., Baba T., Shimizu S.;
RT   "Gene expression of mutant erythropoietin in hepatocellular
RT   carcinoma.";
RL   Biochem. Biophys. Res. Commun. 195(2):717-722(1993).
DR   EMBL; S65458; AAD13964.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
DR   INTERPRO; IPR001323; -.
DR   INTERPRO; IPR003013; -.
DR   PFAM; PF00758; EPO_TPO; 1.
DR   PRINTS; PR00272; ERYTHROPTN.
DR   PROTOMAP; Q9UDZ0.
DR   PRESAGE; Q9UDZ0.
DR   SWISS-2DPAGE; GET_REGION ON 2D_PAGE.
FT   NON_TER       1      1
SQ   SEQUENCE   136 AA;  15173 MW;  BCB9B1F0D8190AB3 CRC64;
     EHCSLNENIT VPDTKVNFYA WKRMEVGQQA VEVWQGLALL SEAVLRGQAL LVNSSQPWEP
     LQLHVDKAVS GLRNFTTLLR ALGAQKEAIS PQDAASAAPL RTITADTFRK LFRVYSNFLR
     GKLKLYTGEA CRTGDR
//
```

Original TrEMBL entry which has been integrated into the SWISS-PROT EPO_HUMAN entry and thus which is not found in TrEMBL anymore.

Some protein motif databases

- Prosite - Regular expression built from SWISS-PROT
- PRINTS - aligned motif consensus built from OWL
  - (http://bioinf.man.ac.uk/dbbrowser/PRINTS/PRINTS.html)
- BLOCKS - PRINTS-like generated from PROSITE families
  - (http://www.blocks.fhcrc.org/)
- IDENTIFY - Fuzzy regular expressions derived from PROSITE
- pfam - Hidden Markov Model built from SWISS-PROT
  - (http://www.sanger.ac.uk/Software/Pfam)
- Profiles - Weight Matrix profiles built from SWISS-PROT
- Interpro - All of the above (almost)
  - (http://www.ebi.ac.uk/InterPro)

A domain database synchronised with SWISS-PROT

The  database

History

- Founded by Amos Bairoch
- 1988 First release in the PC/Gene software
- 1990 Synchronisation with Swiss-Prot
- 1994 Integration of « profiles »
- 1999 PROSITE joins InterPro
- January 2003 Current release 17.32

Database content

- Official Release (20.128) 2016
  - ~1309 Patterns PSxxxxx PATTERN
  - ~1161 Profiles PSxxxxx MATRIX
  - 1175 Rules PSxxxxx RULE
  - ~1762 Documentations PDOCxxxxx

- Pre-Release
  - ~150 Profiles PSxxxxx MATRIX
  - ~100 Documentations QDOCxxxxx

Prosite (pattern): example

General information about the entry	
Entry name	**EPO_TPO**
Accession number	**PS00817**
Entry type	PATTERN
Date	OCT-1993 (CREATED); NOV-1995 (DATA UPDATE); JUL-1998 (INFO UPDATE).
PROSITE documentation	PDOC00644
Name and characterization of the entry	
Description	Erythropoietin / thrombopoeitin signature.
Pattern	P-x(4)-C-D-x-R-[LIVM](2)-x-[KR]-x(14)-C.

Numerical results

- SWISS-PROT release number: **40.7**, total number of sequence entries in that release: **103373**.
- Total number of hits in SWISS-PROT: **14 hits in 14 different sequences**
- Number of hits on proteins that are known to belong to the set under consideration: **14 hits in 14 different sequences**
- Number of hits on proteins that could potentially belong to the set under consideration: **0 hits in 0 different sequences**
- Number of false hits (on unrelated proteins): **0 hits in 0 different sequences**
- Number of known missed hits: **0**
- Number of partial sequences which belong to the set under consideration, but which are not hit by the pattern or profile because they are partial (fragment) sequences: **1**
- Precision (true hits / (true hits + false positives)): **100.00 %**
- Recall (true hits / (true hits + false negatives)): **100.00 %**

Prosite (pattern): example

Comments

- Taxonomic range: **Eukaryotes**
- Maximum known number of repetitions of the pattern in a single protein: **1**
- `Interesting' site in the pattern: **3,disulfide**
- `Interesting' site in the pattern: **11,disulfide**

Cross-references

SWISS-PROT

True positive hits:

EPO_BOVIN (P48617), EPO_CANFA (P33707), EPO_FELCA (P33708),
EPO_HUMAN (P01588), EPO_MACFA (P07865), EPO_MACMU (Q28513),
EPO_MOUSE (P07321), EPO_PIG (P49157), EPO_RAT (P29676),
EPO_SHEEP (P33709), TPO_CANFA (P42705), TPO_HUMAN (P40225),
TPO_MOUSE (P40226), TPO_RAT (P49745)

`Potential' hits (partial sequences which belong to the set under consideration, but which are not hit by the pattern or profile because they are partial (fragment) sequences):

TPO_PIG (P42706)

Retrieve an alignment of SWISS-PROT true positive hits:

[Clustal format, color, condensed view] [Clustal format, color] [Clustal format, plain text] [Fasta format]

Database content: documentation

Description of pattern(s) and/or profile(s)	
Consensus pattern	G-[LIVMFY]-x(1,3)-[AGC]-[NASM]-x-C-[FYW]-[LIVMFC]-[NST]-[SACV]-x-[LIVMS]-Q [C is the putative active site residue]
Sequences known to belong to this class detected by the pattern	ALL, except for two sequences.
Other sequence(s) detected in SWISS-PROT	NONE.
Consensus pattern	Y-x-L-x-[SAG]-[LIVMFT]-x(2)-H-x-G-x(4,5)-G-H-Y [The two H's are putative active site residues]
Sequences known to belong to this class detected by the pattern	ALL.
Other sequence(s) detected in SWISS-PROT	NONE.
Sequences known to belong to this class detected by the profile	ALL.
Other sequence(s) detected in SWISS-PROT	NONE.
Note	these proteins belong to family C19 in the classification of peptidases [3,E1].
Note	this documentation entry is linked to both a signature pattern and a profile. As the profile is much more sensitive than the pattern, you should use it if you have access to the necessary software tools to do so.
Last update	
December 2001 / Patterns and text revised; profile added.	

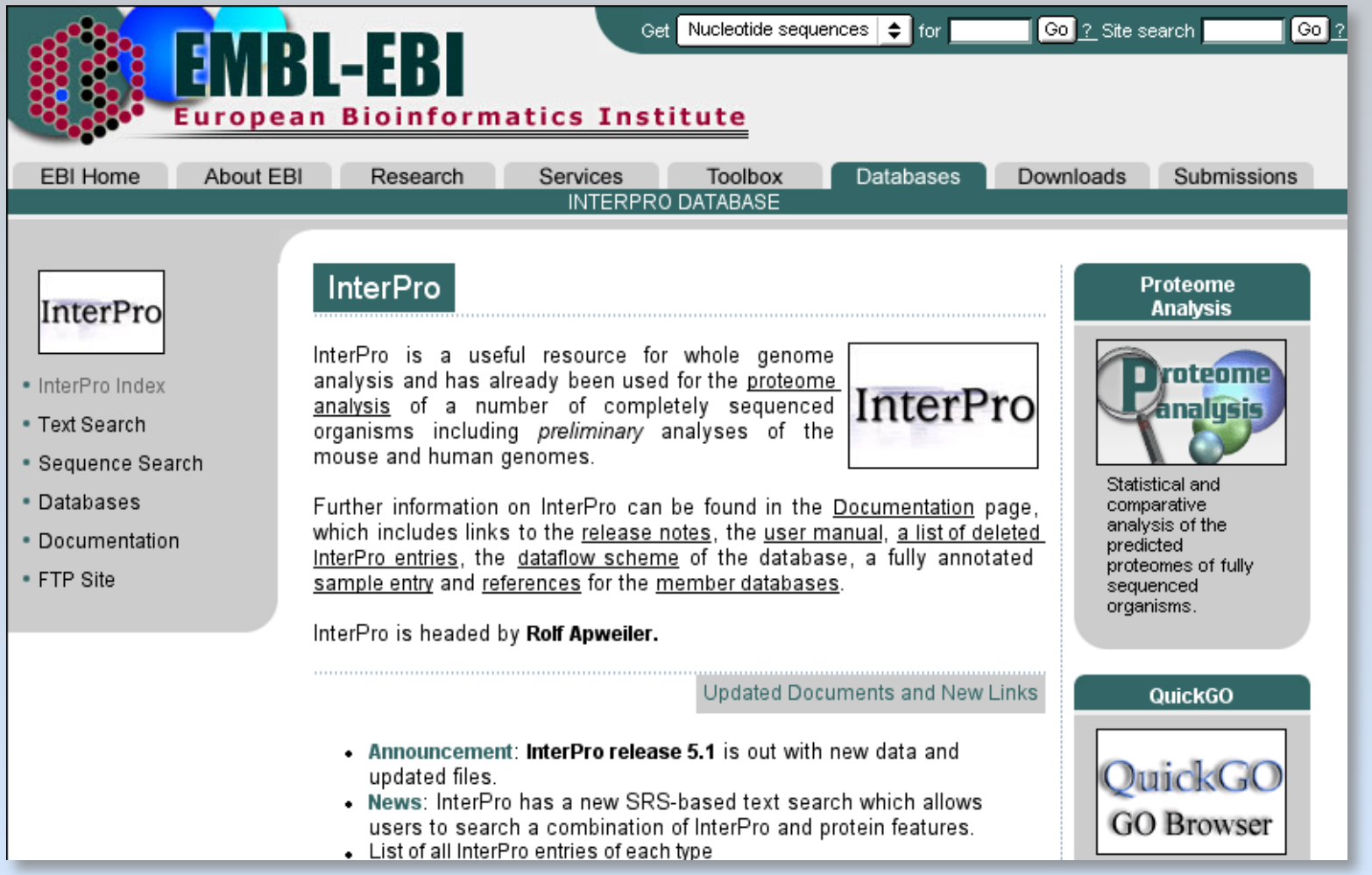
Other protein domain/family db

PROSITE	Patterns / Profiles
ProDom	Aligned motifs (PSI-BLAST) (Pfam B)
PRINTS	Aligned motifs
Pfam	HMM (Hidden Markov Models)
SMART	HMM
TIGRfam	HMM

Interpro

DOMO	Aligned motifs
BLOCKS	Aligned motifs (PSI-BLAST)
CDD(CDART)	PSI-BLAST(PSSM) of Pfam and SMART

InterPro: www.ebi.ac.uk/interpro



InterPro example

InterPro Entry IPR001323

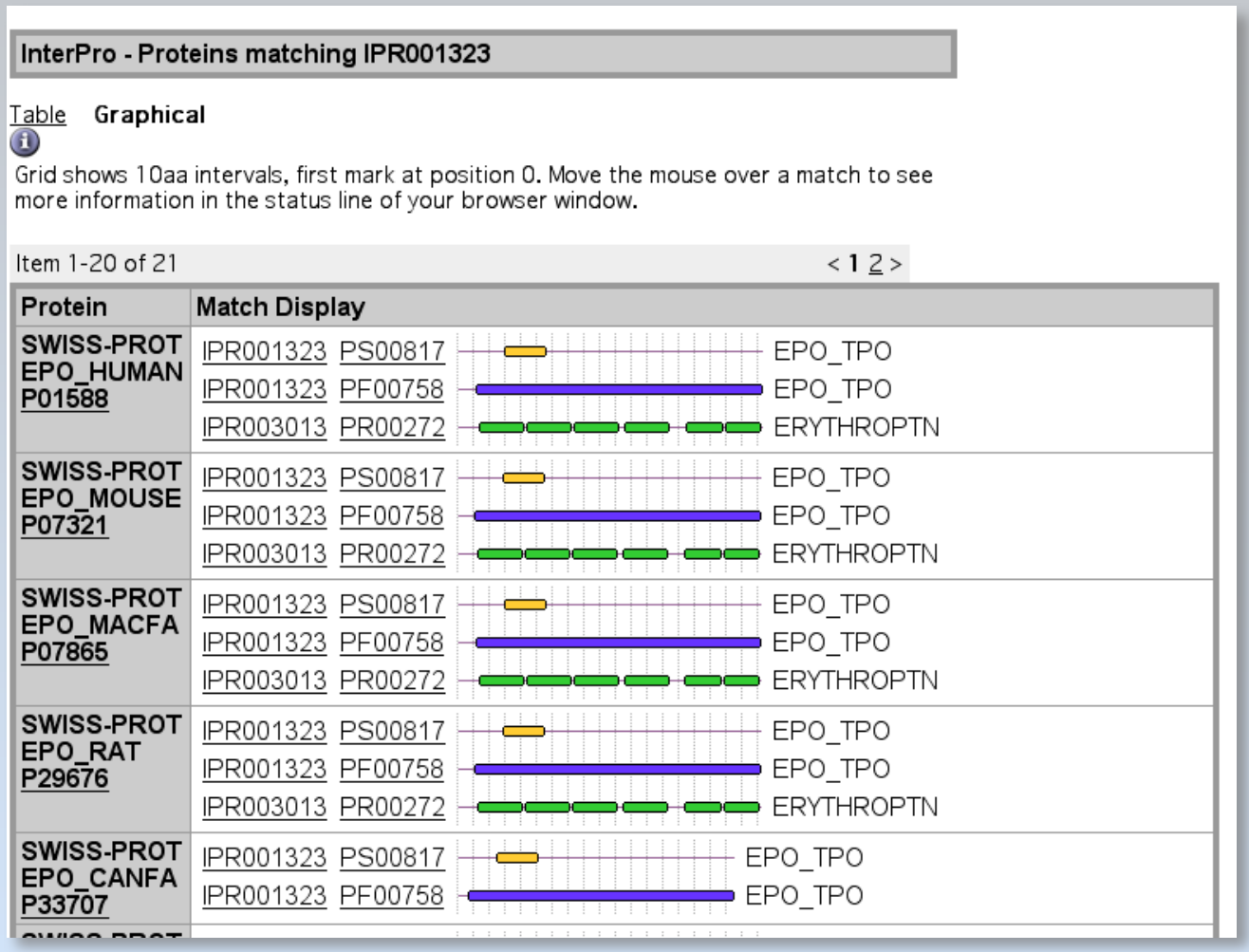
Erythropoietin/thrombopoeitin

Database	InterPro
Accession	IPR001323; EPO_TPO (matches 21 proteins)
Name	Erythropoietin/thrombopoeitin
Type	Family
Dates	08-OCT-1999 (created) 23-NOV-2000 (last modified)
Signatures	PS00817; EPO_TPO (19 proteins) PF00758; EPO_TPO (21 proteins)
Children [tree]	IPR003013; Erythropoietin (12 proteins) IPR003978; Thrombopoeitin (5 proteins)
Function	glycopeptide hormone (GO:0005181)
Component	extracellular (GO:0005576)
Abstract	Erythropoietin, a plasma glycoprotein, is the primary physiological mediator of erythropoiesis [1]. It is involved in the regulation of the level of peripheral erythrocytes by stimulating the differentiation of erythroid progenitor cells, found in the spleen and bone marrow, into mature erythrocytes [2]. It is primarily produced in adult kidneys and foetal liver, acting by attachment to specific binding sites on erythroid progenitor cells, stimulating their differentiation [3]. Severe kidney dysfunction causes reduction in the plasma levels of erythropoietin, resulting in chronic anaemia - injection of purified erythropoietin into the blood stream can help to relieve this type of anaemia. Levels of erythropoietin in plasma fluctuate with varying oxygen tension of the blood, but androgens and prostaglandins also modulate the levels to some extent [3]. Erythropoietin glycoprotein sequences are well conserved, a consequence of which is that the hormones are cross-reactive among mammals, i.e. that from one species, say human, can stimulate erythropoiesis in other species, say mouse or rat [4]. Thrombopoeitin (TPO), a glycoprotein, is the mammalian0 hormone which functions as a megakaryocytic lineage specific growth and differentiation factor affecting the proliferation and maturation from their committed progenitor cells acting at a late stage of megakaryocyte development. It acts as a circulating regulator of platelet numbers.

InterPro example

Examples	• P49745 TPO_RAT • P33709 EPO_SHEEP • P33708 EPO_FELCA View examples
References	1. Shoemaker C.B., Mitsock L.D. *Murine erythropoietin gene - Cloning, expression , and human gene homology.* Mol. Cell. Biol. 6: 849-858(1986). [MEDLINE:87039105] 2. Takeuchi M., Takasaki S., Miyazaki H., Kato T., Hoshi S., Kochibe N., Kobata A. *Comparative study of the asparagine-linked sugar chains of human erythropoietins purified from urine and the culture medium of recombinant chinese hamster ovary cell.* J. Biol. Chem. 263: 3657-3663(1988). [MEDLINE:88153657] 3. Lin F.K., Lin C.H., Lai P.H., Browne J.K., Egrie J.C., Smalling R., Fox G.M., Chen K.K., Castro M., Suggs S. *Monkey erythropoietin gene - Cloning, expression and comparison with the human erythropoietin gene.* Gene 44: 201-209(1986). [MEDLINE:87055236] 4. Nagao M., Suga H., Okano M., Masuda S., Narita H., Ikura K., Sasaki R. *Nucleotide sequence of rat erythropoietin.* Biochim. Biophys. Acta 1171: 99-102(1992). [MEDLINE:93042015]
Database links	PROSITE doc; PDOC00644 Blocks; IPB001323
Matches	Table all Graphical all Condensed graphical view

InterPro graphic example



Genomic Databases

- Genome databases differ from sequence databases in that the data contained in them are much more diverse.
- The idea behind a genome database is to organize all information on an organism (or as much as possible).
- In many cases they stem out of the necessity for a centralized resource for a particular genome project. But of course they are also important resources for the research community.

Genomic Databases

- Ensembl
- Genome Browser
- NCBI

Structure Databases

- PDB
- SCOP

PDB

- The Protein Data Bank (PDB) was established at Brookhaven National Laboratories (BNL) (1) in 1971 as an archive for biological macromolecular crystal structures.
- The three dimensional structures in PDB are primarily derived from experimental data obtained by X-ray crystallography and NMR .

UniProt: United Protein database

- SWISS-PROT + TrEMBL + PIR = UniProt
- Born in October 2002
- NIH pledges cash for global protein database
  - The United States is turning to European bioinformatics facilities to help it meet its researchers' future needs for databases of protein sequences.
  - European institutions are set to be the main recipients of a $15-million, three-year grant from the US National Institutes of Health (NIH), to set up a global database of information on protein sequence and function known as the United Protein Databases, or UniProt (Nature, 419, 101 (2002))

Some examples of integrated biological database resources are:

- SRS (Sequence Retrieval System)
- MRS (Open source SRS)
- Entrez Browser (at NCBI)
- ExPASy (home of SwissProt)
- Ensembl (Open Source based system)
- Human Genome Browser (Jim Kents creation)

Acknowledgments

- Laurent Falquet, SIB and EMBnet-CH for slides and information on SwissProt and Prosite