The annotation of proteins from pathogens in UniProtKB/Swiss-Prot: current status and future plans

Amos Bairoch; University of Geneva and

Swiss Institute of Bioinformatics (SIB)

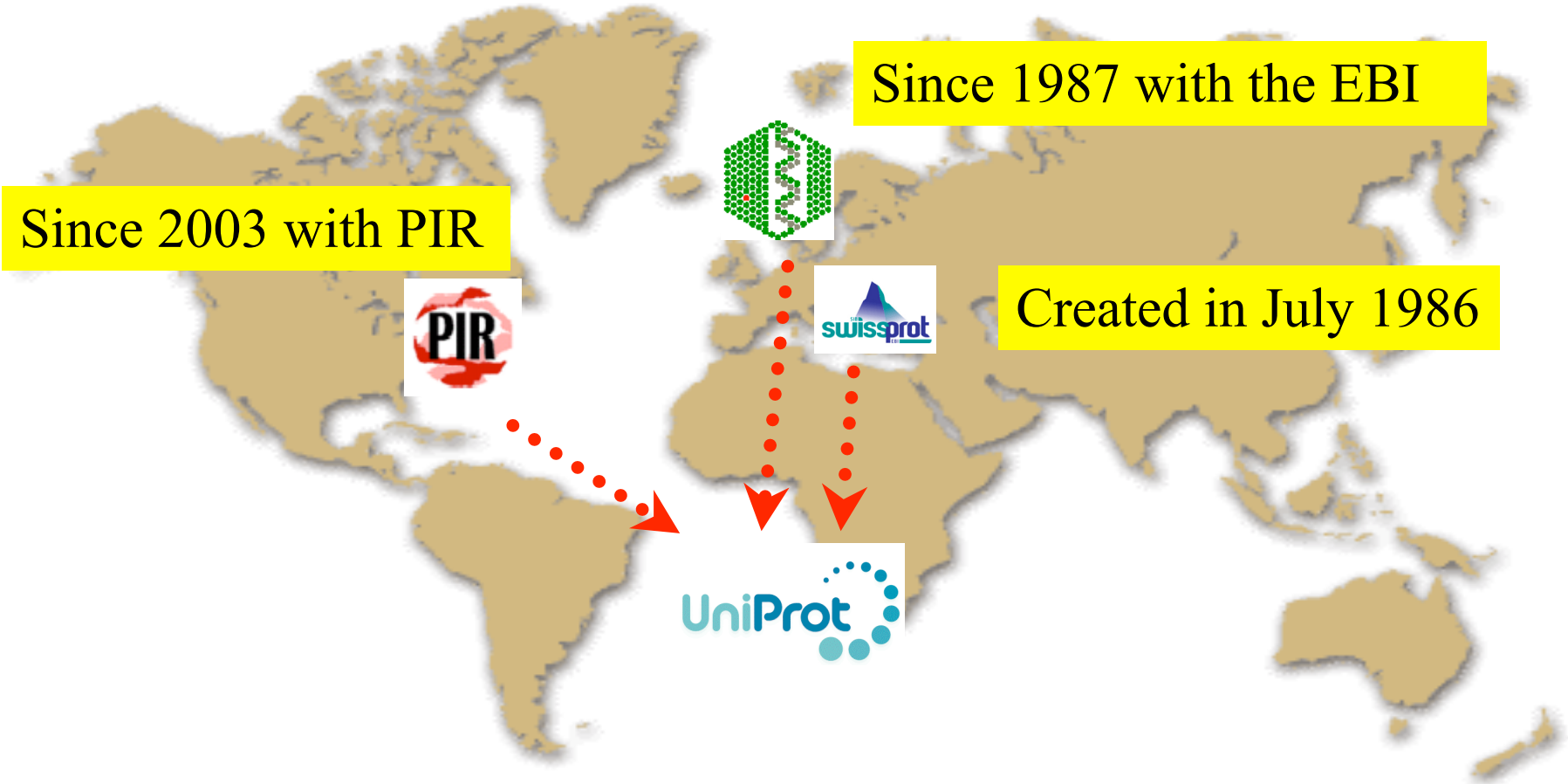
Swiss-Prot group

Nairobi – May 29, 2007

The Swiss-Prot staff at SIB and EBI

- **Group leaders:** Amos Bairoch, Rolf Apweiler, Lydie Bougueleret
- **Annotators/curators:** Yasmin Alam-Faruque, Philippe Aldebert, Ruth Akhtar, Severine Altairac, Nicola Althorpe, Ghislaine Argoud Puy, Andrea Auchincloss, Kristian Axelsen, Kirsty Bates, Marie-Claude Blatter, Emmanuel Boutet, Silvia Braconi Quintaje, Lionel Breuza, Alan Bridge, Paul Browne, Evelyn Camon, Wei mun Chan, Luciane Ciapina, Guy Cochrane, Danielle Coral, Elisabeth Coudert, Isabelle Cusin, Tania de Oliveira Lima, Kirill Degtyarenko, Paula Duek, Ruth Eberhardt, Anne Estreicher, Livia Famiglietti, Nathalie Farriol-Mathis, Nadeem Faruque, Serenella Ferro, Marc Feuermann, Rebecca Foulger, Gill Fraser, Gabriella Frigerio, John Garavelli, Vivienne Gerritsen, Arnaud Gos, Nadine Gruaz-Gumowski, Ursula Hinz, Chantal Hulo, Nicolas Hulo, Julius Jacobsen, Janet James, Silvia Jimenez, Florence Jungo, Vivien Junker, Guillaume Keller, Kati Laiho, Lydie Lane, Petra Langendijk-Genevaux, Duncan Legge, Philippe Lemercier, Virginie Lesaux, Damien Lieberherr, Michele Magrane, Karine Michoud, Madelaine Moinat, Anne Morgat, Nicola Mulder, Marisa Nicolas, Claire O'Donovan, Sandra Orchard, Ivo Pedruzzi, Sandrine Pilbout, Sylvain Poux, Manuela Prüss, Sorogini Reynaud, Catherine Rivoire, Bernd Röchert, Michel Schneider, Christian Sigrist, André Stutz, Shyamala Sundaram, Michael Tognoli, Claudia Vitorello, Eleanor Whitfield, Luiz Fernando Zuleta
- **Programmers and system administrators:** Delphine Baratin, Daniel Barrell, Laurent Bollondi, Lawrence Bower, Matias Castro, Michael Darsow, Edouard deCastro, Paula de Matos, Mike Donnelly, Séverine Duvaud, Alexander Fedetov, Wolfgang Fleischmann, Elisabeth Gasteiger, Alain Gateau, Sebastien Gehant, Andre Hackmann, Henning Hermjakob, Alessandro Innocenti, Eric Jain, Phil Jones, Alexander Kanapin, Paul Kersey, Ernst Kretschmann, Corinne Lachaize, Vincente Lara, Vincent Le Texier, Maria-Jesus Martin, Xavier Martin, John O'Rourke, Salvo Paesano, Sam Patient, Isabelle Phan, Astrid Rakow, Nicole Redaschi, Emilio Salazar, Nataliya Skylar, Karin Sonesson, Peter Sterk, Daniela Wieser, Dan Wu, WeiMin Zhu
- **Research staff:** Valeria Amendolia, Brigitte Boeckmann, Lorenzo Cerutti, Fabrice David, David Perret, Violaine Pillet, Anne-Lise Veuthey, Lina Yip
- **Clerical and secretarial assistance:** Dolnide Dornevil, Claudia Sapsezian, Kerry Smith, Laure Verbregue

The Swiss-Prot group works in collaboration with



And together they form UniProt,

The Universal Protein Knowledgebase

An avalanche of data

- In 1954: publication of the first sequence of a protein: bovine insulin by Frederick Sanger

Date	DNA	Protein	3D
1964	70 bp	65	2
1974	0.1 Mb	500	10
1984	2 Mb	3'000	250
1994	220 Mb	70'000	3'000
2004	78'425 Mb	2'000'000	28'000

- More than 50% of the biomolecular data available today was produced in the last two years;
- In 1986: 4'000 proteins in Swiss-Prot; today: 4'000 new proteins will enter Swiss-Prot+TrEMBL.

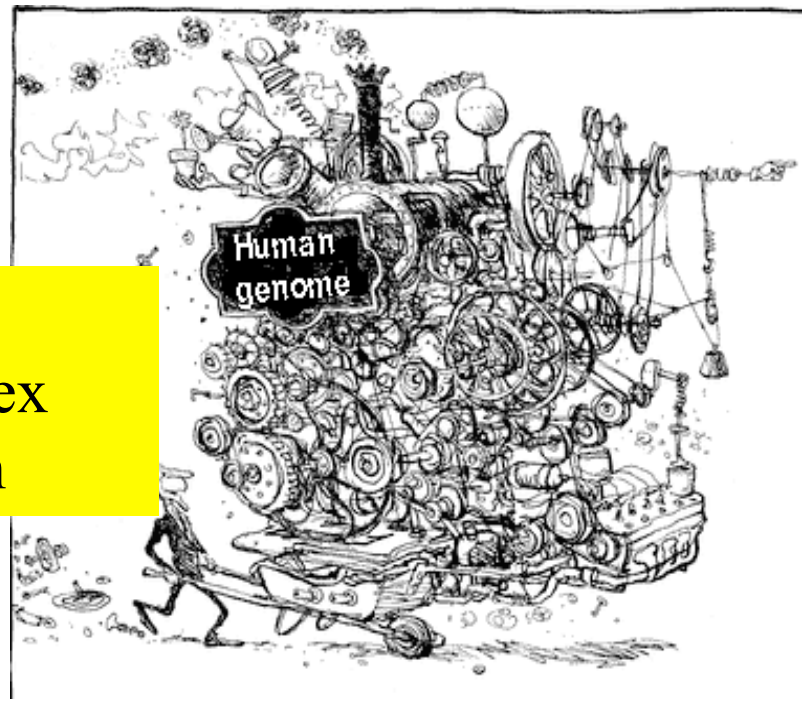
The implications...

- The Life Sciences have undergone a dramatic revolution in the last 20 years:
  - ✓They used to be rich in hypotheses, well-off in knowledge and poor in data;
  - ✓They are now very rich in data, not so well-off in knowledge and very poor in hypotheses.

A list of parts

How do we go from:

To a complex system



The universe in which Swiss-Prot evolves

1953: 1st sequence (bovine insulin)

1986: 4’000 sequences

2007: 5 million sequences

Where will it stop?

179'000'025'042 (179 billion)

179'000'025'042

1st estimate: ~30 million species (1.5 million named)

2nd estimate:

20	million bacteria/archea	x	4'000 genes
5	million protists	x	6'000 genes
3	million insects	x	14'000 genes
1	million fungi	x	6'000 genes
0.6	million plants	x	20'000 genes
0.2	million molluscs, worms, arachnids, etc.	x	20'000 genes
0.2	million vertebrates	x	25'000 genes

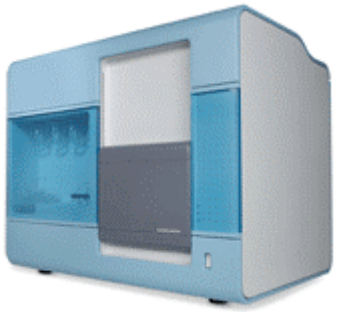
The calculation:

$2x10^7x4000+5x10^6x6000+3x10^6x14000+10^6x6000+6x10^5x20000+2x10^5x20000+2x10^5x25000+25000$(Craig Venter)+42(Douglas Adam)

Caveat: this is an estimate of the number of potential sequence entries, but not that of the number of distinct protein entities in the biosphere.

Will all the different proteins in the biosphere be ever sequenced?

Probably yes!



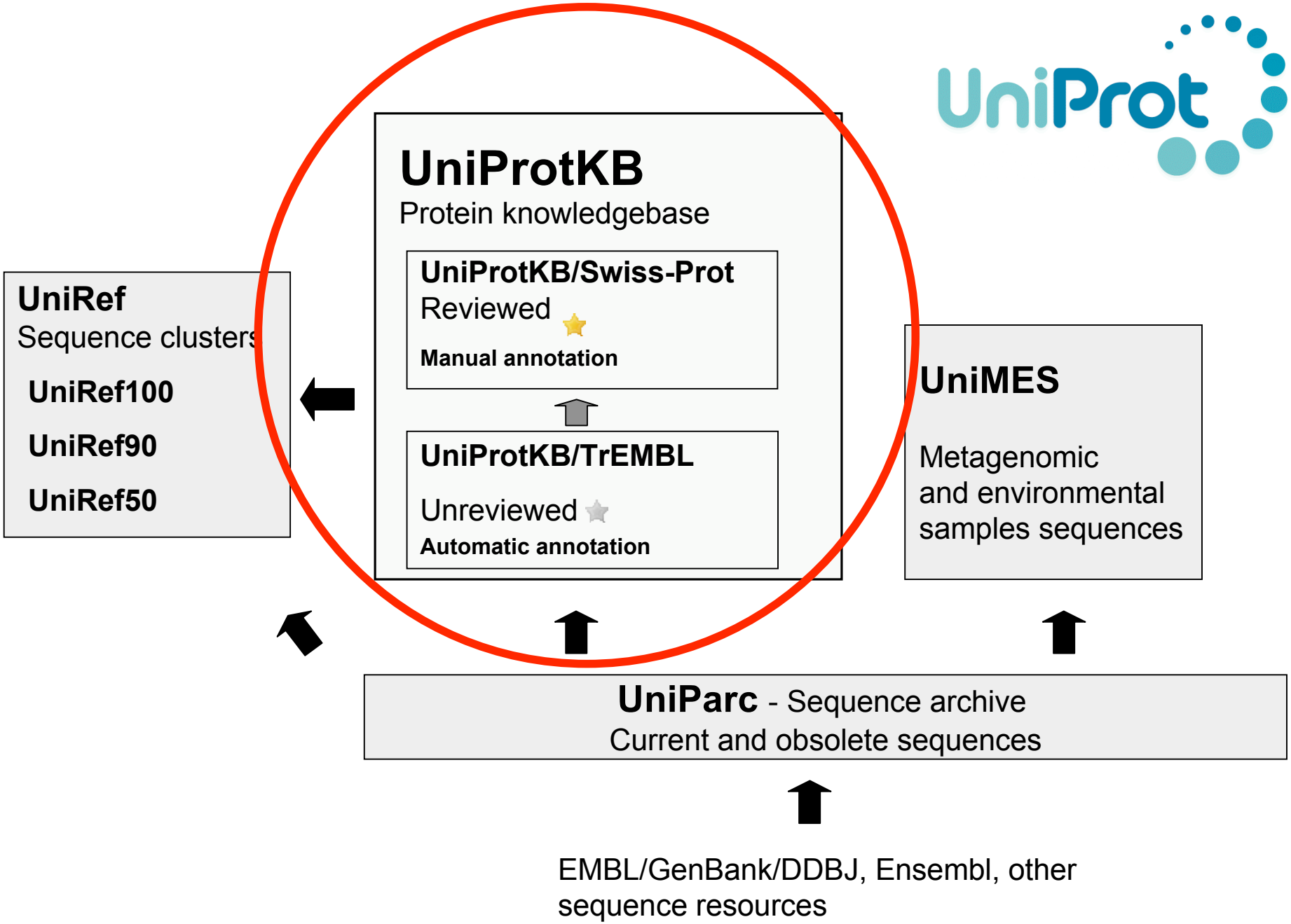
Press Release

FOR IMMEDIATE RELEASE

More than Six Million New Genes, Thousands of New Protein Families, and Incredible Degree of Microbial Diversity Discovered from First Phase of Sorcerer II Global Ocean Sampling Expedition

DNA Sequencing with Solexa® Technology

Generating one billion bases of high quality DNA sequence per run at less than 1% of the cost of capillary-based methods, the Illumina Genome Analyzer is designed to enable researchers to dramatically improve the efficiency and speed of current applications. Now an expanded scale of research that was previously unimaginable with other technology platforms is possible with the Genome Analyzer.


UniProt
UniProtKB
Protein knowledgebase
UniProtKB/Swiss-Prot
Reviewed
Manual annotation
UniProtKB/TrEMBL
Unreviewed
Automatic annotation
UniRef
Sequence clusters
UniRef100
UniRef90
UniRef50
UniMES
Metagenomic
and environmental
samples sequences
UniParc - Sequence archive
Current and obsolete sequences
EMBL/GenBank/DDBJ, Ensembl, other
sequence resources

EMBL DNA db

```
ID   AL713759; SV 1; linear; mRNA; HTC; HUM; 2715 BP.
XX
AC   AL713759;
XX
DT   20-MAR-2002 (Rel. 71, Created)
DT   22-SEP-2004 (Rel. 81, Last updated, Version 2)
XX
DE   Homo sapiens mRNA; cDNA DKFZp564M1682 (from clone DKFZp564M1682)
XX
KW   HTC.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae;
OC   Homo.
XX
RN   [1]
RP   1-2715
RG   The German cDNA Consortium
RA   Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Ossinger A., Fobo G., Han M., Wiemann S.;
RT
RL   Submitted (22-SEP-2004) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
DR   H-InvDB; HIT000026682.
DR   RZPD; DKFZp564M1682.
XX
CC   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research
CC   Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC   sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany)
CC   within the cDNA sequencing consortium of the German Genome Project.
CC   This clone (DKFZp564M1682) is available at the RZPD Deutsches
CC   Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
CC   Please contact RZPD for ordering:
CC   http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp564M1682
CC   Further information about the clone and the sequencing project is
CC   available at http://mips.gsf.de/projects/cdna/
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..2715
FT                   /organism="Homo sapiens"
FT                   /mol_type="mRNA"
FT                   /dev_stage="fetal"
FT                   /clone_lib="564 (synonym: hfbr2). Vector pAMP1; host
FT                   X1-2blue; sites NotI + SalI"
FT                   /clone="DKFZp564M1682"
FT                   /tissue_type="brain"
FT                   /note="hypothetical protein (Mus musculus)"
FT                   /db_xref="taxon:9606"
FT                   /db_xref="RZPD:DKFZp564M1682"
FT   CDS             978..1347
FT                   /codon_start=1
FT                   /gene="DKFZp564M1682"
FT                   /product="hypothetical protein"
FT                   /db_xref="GDB:11514865"
FT                   /db_xref="GDA:Q8TCQ1"
FT                   /db_xref="HGNC:26077"
FT                   /db_xref="InterPro:IPR001841"
FT                   /db_xref="InterPro:IPR011016"
FT                   /db_xref="UniProtKB/TrEMBL:Q8TCQ1"
FT                   /protein_id="CAD28529.1"
FT                   /translation="MLGWCEAIAWPHHIPNDVTRIPEISGDLADASQTSTLNEKSPGRS
FT                   ASRSSWISXASSPFTGIAPRSQSRLSVCPSTQDICRICHCEGDEESPLIPCRCTGTLR
FT                   PVHQSCLHQWIKSSDDRCCELCKYDPIMETIXLKPLRAKWEXLQMTTSERRKIFCSVTPHV
FT                   IAITCVVWSLYVLIDRIAEEIKQGMDNGVLEWPPWTKLVVVAIGFIQGLVPWVQCKVY
FT                   VQLWRRLKAYNRVIPVQMCPDTAKKLEKWFSCNVNTDIKDAVVVPVPQIGANSLPSAEG
FT                   GPPEVVSV"
XX
SQ   Sequence 2715 BP; 721 A; 653 C; 693 G; 648 T; 0 other;
     agaggagggg agaggaetgg ggageeagge eagageeggg ctgeetgeea eccg
     cgtcegctag ctggggagga gegctccacc egcaactgae aaaggatggg aa
     egtcccgggg tgcccgccgc aegcagcctg gcggcccgct gagctactc
     ggtaagtgac tgcaaacatc attcattcaa tcagctctac tgggagcc
     ctggtagtcc tgggcggctt gtcccctgatc ccgagcgggg cttggca
     ggaggccagg cagcaggtgc atttgcctgg tgggtccact ggggag
     gcggcgggtg ctgccaccca acctgcggge gggcgggctcg cccagg
     tgagaggagg cggctccagc acgcatcctg gggtagttgc tactc
     cgtctcgcgc gcgcgccgtt tctggcgtat cccagtgcg cggc
     cgtggtacta gtcaaggggc cgcaccccgt cggaagcgcg gagt
     acaggctcac gcgctcctag tgtgggcttg aaggggcacg ggac
     gcgctgagta cggagaacac agggcagcct ttgtcttggg ttag
     cctgagtcg ggtccatcgc aactgtcgtg tccgatttcg gttcc
     tgggcgagag atgtcattgt gttcctcgcg ccagcggggac tgagag
     gcccaggagg tcctgcgctc acggggcaag tccccccaaa gaactct
     cccctgtcct gctgtgatta aacaagactg ctgtatttta atttcaga
     atagggggaa gggggaaaatg ctgggctggt gtgaagcgat agccctaac
     ttccaaaaca cacgcgaaca cccgagatct caggggattt ggctgagcc
     ccacattgaa tgaaaaatcc ccagggcgat ctgcaagtcg atcaagtaac
     cagcagccc aacaaacggg acgctccca ggagccagtc aaggttgtct
     ccactcaggg catctgcaga atctgtcact gccgaaggggg tgaagagagc
     cacccgtcg ctgcactggg acactgcgct ttgtccacca gtcctgctc
     taaagagctc agatacagc tgctgtgagc tctgcaagta tgactcata atggagacca
     agctcaaacc cctccccaag tcccgcaaac tacagatgac cacaagtcaa aggaccaaaa
```

TrEMBL

```
ID   Q8TCQ1_HUMAN            Unreviewed;        289 AA.
AC   Q8TCQ1;
DT   01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT   01-JUN-2002, sequence version 1.
DT   09-JAN-2007, entry version 23.
DE   Hypothetical protein DKFZp564M1682.
GN   Name=MARCH1; Synonyms=DKFZp564M1682;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
OC   Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Brain;
RG   The German cDNA Consortium;
RA   Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Ossinger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- PATHWAY: Ubiquitin conjugation: third step.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   ---------------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------------
DR   EMBL; AL713759; CAD28529.1; -; mRNA.
DR   SMR; Q8TCQ1; 72-136.
DR   Ensembl; ENSG00000145416; Homo sapiens.
DR   HGNC; HGNC:26077; MARCH1.
DR   ArrayExpress; Q8TCQ1; -.
DR   RZPD-ProtExp; IOH42112; -.
DR   RZPD-ProtExp; W1571; -.
DR   GO; GO:0046872; F:metal ion binding; IEA:UniProtKB-KW.
DR   GO; GO:0005515; F:protein binding; IEA:InterPro.
DR   GO; GO:0008270; F:zinc ion binding; IEA:InterPro.
DR   GO; GO:0006512; P:ubiquitin cycle; IEA:UniProtKB-KW.
DR   InterPro; IPR011016; RING-v.
DR   InterPro; IPR001841; Znf_RING.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00744; RINGv; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Ubl conjugation pathway; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   289 AA;  32308 MW;  923E1809AB3D7087 CRC64;
     MLGWCEAIAR NPHRIPNNTR TPEISGDLAD ASQTSTLNEK SPGRSASRSS NISKASSPTT
     GTAPRSQSRL SVCPSTQDIC RICHCEGDEE SPLITPCRCT GTLRPVHQSC LHQWIKSSDT
     RCCELCKYDP IMETKLKPLR KWEKLQMTTS ERRKIFCSVT FHVIAITCVV WSLYVLIDRT
     LEEIKQGNDN GVLEWPPWTK LVVVAIGFTG GLVPMVVQCK VYVQLWRRLK AYNRVIPVQN
     MCPDTAKKLEK NFSCNVNTDI KDAVVVPVPQ TGANSLPSAE GGPPEVVSV
```

Swiss-Prot

```
ID   UNCX1_HUMAN              Reviewed;          314 AA.
AC   Q17EQ1; Q6NTK1;
DT   16-FEB-2006, integrated into UniProtKB/Swiss-Prot.
DT   11-JUL-2006, sequence version 1.
DT   16-FEB-2011, entry version 31.
DE   E3 ubiquitin-protein ligase UNCX1 (EC 6.3.2.-) (Homeobox-containing
DE   RING finger protein 1) (Homeobox-containing RING-CX protein 1) (UNCX-
DE   1) (RING finger protein 111).
GN   Name=UNCX1; Synonyms=RCF111;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Euteria; Euarchontoglires; Primates; Haplorrhini;
OC   Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [MRNA] AND [XXXXXXXXXXXX] ||XXXXXXXX 1].
RC   TISSUE=Fetal muscle;
RX   PubMed=11113131; DOI=10.1139/g00-[illegible];
RA   Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA   Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA   Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA   Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA   Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA   Sudo H., Hosoiri T., Kaku Y., Kodama H., Kawarabata R., Katai M.,
RA   Takahashi K., Ishikawa S., Yokoyama T., Ishouna Y., Suwa Y.,
RA   Hamadeh L., Takeuchi M., Matsubara T., Nakagawa A., Yoshikawa T.,
RA   Nakamura Y., Yamamoto J., Yoshikawa T., Inoue I., Sugano S.,
RA   Lehrach H., Suzuki Y., Isogai T., Gojobori T., Chenchik A.,
RA   Morishita S., Hirai M., Nomura N., Sugano S.;
RT   "Construction and characterization of a human full-length cDNA
RT   library.";
RL   J. Biomed. Biotechnol. 1:123-133(2001).
CC   -!- FUNCTION: E3 ubiquitin-protein ligase that may mediate
CC       ubiquitination of TYK2, CDK2 and JAK2, and promote their subsequent
CC       proteasomal and lysosomal degradation. Ubiquitinylation of JAK2
CC       ubiquitin ligases accept ubiquitin from an E2 ubiquitin-
CC       conjugating enzyme in the form of a thioester and then directly
CC       transfer the ubiquitin to targeted substrates.
CC   -!- SUBUNIT: Ubiquitin conjugates; Ubl-2 mRNA.
CC   -!- DOMAIN: Contains 1 RING-type zinc finger domain.
CC   -!- MODIFICATION: Ubiquitinylation; multi-gene modifications
CC       protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         Isoid=Q17EQ1-1; Sequence=Displayed;
CC       Name=3;
CC         Isoid=Q17EQ1-3; Sequence=VSP_133131, VSP_133132;
CC   -!- FUNCTION: EC=2.7.11.1; Evidence=ISS; Name=Ser/Thr protein kinase, epsilon and lamp.
CC   -!- DOMAIN: The RING-type zinc finger domain is required for E3
CC       ligase activity [By similarity].
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   ---------------------------------------------------------------------------
CC   Copyright© by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------------
DR   EMBL; AB113126; BAD92126.1; -; mRNA.
DR   EMBL; AC113126; BAA63126.1; -; mRNA.
DR   UniGene; Hs.125811; -.
DR   ZFN; [illegible]; 13-132.
DR   Ensembl; ENSG[illegible]; Homo sapiens.
DR   HGNC; HGNC:2111; UNCX1.
DR   PharmGkb;[illegible]; -.
DR   KEGG-Brief;[illegible]; -.
DR   KEGG-Full;[illegible]; -.
DR   InterPro; IPR[illegible]; RING-.
DR   InterPro; IPR[illegible]; Int_RING.
DR   Pfam; PF[illegible]; rf-C3HC; 1.
DR   SUPFAM; SM[illegible]; RING-; 1.
DR   PROSITE; PS[illegible]; LT_RING_2; 1.
KW   Alternative splicing; Ligase; Homeobox; Ubl-2-binding; Transcriptional
KW   UBL conjugation pathway; Zinc; Zinc-finger.
FT   CHAIN        1   314        E3 ubiquitin-protein ligase UNCX1.
FT                               [illegible]/FRU_[illegible].
FT   TRANSMEM   [illegible]  [illegible]  Helical.
FT   TRANSMEM   [illegible]  [illegible]  Helical.
FT   DOMAIN     [illegible]  [illegible]  RING-type.
FT   VAR_SEQ      [illegible]   [illegible]  Missing (in isoform 3).
FT                               [illegible]/VSP_133131.
FT   VAR_SEQ      [illegible]   [illegible]  [illegible] -> G
FT                               [illegible] (in
FT                               isoform 3).
FT                               [illegible]/VSP_133132.
SQ   SEQUENCE   314 AA;  [illegible] MW;  [illegible] CRC64;
     [illegible]
```

Manual annotation of the sequence and associated biological information

Automated extraction of protein sequence, gene name and references +

Automated annotation

Swiss-Prot

- **Annotated**, **non-redundant**, **cross-referenced**, **documented** protein sequence **knowledge** resource;
- or more simply remember it as an **encyclopedia on proteins!**;
- **268’000** sequences; 134’000 literature references; 4’000’000 cross-references to 100 databases; ~800 Mb of annotations;
- About **4’400’000** sequences in TrEMBL, its computer-annotated supplement.

Contribute
Send feedback
WikiProteins

Reviewed, UniProtKB/Swiss-Prot **P15917** (LEF_BACAN)

Last modified May 1, 2007. Version 85. History...

Clusters with 100%, 90%, 50% identity | Documents (3) | Customize display

TEXT XML RDF/XML FASTA

Names and origin · General annotation (Comments) · Ontologies · Sequence annotation (Features) · Sequences · References · Cross-references · Entry information · Relevant documents

Names and origin

Hide | Top

Protein names	**Lethal factor** [Precursor] *Also known as:* EC 3.4.24.83 LF Anthrax lethal toxin endopeptidase component
Gene names	Name: **lef** Ordered Locus Names: pXO1-107, BXA0172, GBAA_pXO1_0172
Encoded on	Plasmid pXO1
Organism	**Bacillus anthracis** [Complete proteome] [HAMAP]
Taxonomic identifier	1392 [NEWT] [NCBI]
Taxonomic lineage	Bacteria › Firmicutes › Bacillales › Bacillaceae › Bacillus › Bacillus cereus group

Beta web site: beta.uniprot.org; demo on Friday

General annotation (Comments)	Hide | Top
Function	One of the three proteins composing the anthrax toxin, the agent which infects many mammalian species and that may cause death. LF is the lethal factor that, when associated with PA, causes death. LF is not toxic by itself. It is a protease that cleaves the N-terminal of most dual specificity mitogen-activated protein kinase kinases (MAPKKs or MAP2Ks) (except for MAP2K5). Cleavage invariably occurs within the N-terminal proline-rich region preceding the kinase domain, thus disrupting a sequence involved in directing specific protein-protein interactions necessary for the assembly of signaling complexes. There may be other cytosolic targets of LF involved in cytotoxicity. The proteasome may mediate a toxic process initiated by LF in the cell cytosol involving degradation of unidentified molecules that are essential for macrophage homeostasis. This is an early step in LeTx intoxication, but it is downstream of the cleavage by LF of MEK1 or other putative substrates.
Catalytic activity	Preferred amino acids around the cleavage site can be denoted BBBBxHx-|-H, in which B denotes Arg or Lys, H denotes a hydrophobic amino acid, and x is any amino acid. The only known protein substrates are mitogen-activated protein (MAP) kinase kinases.
Cofactor	Binds 1 zinc ion per subunit.
Subunit structure	Anthrax toxins are composed of three distinct proteins, a protective antigen (PA), a lethal factor (LF) and an edema factor (EF). None of these is toxic by itself. PA+LF forms the lethal toxin (LeTx); PA+EF forms the edema toxin (EdTx).
Subcellular location	Secreted protein.
Induction	Positively transcriptionally regulated by AtxA, which, in turn, is induced by bicarbonate and high temperatures (37 degrees Celsius).
Domain	It comprises four domains: domain I binds the membrane-translocating component (PA); domains II, III and IV together create a long deep groove that holds the 16-residue N-terminal tail of MAPKK before cleavage. Domain IV contains the catalytic center. The PA-binding region is found in both B.anthracis EF and LF.
Miscellaneous	LF binds to the heptamer formed by cleaved PA on the host cell membrane. This step is followed by internalization of the hetero-oligomeric complex by receptor-mediated endocytosis. LeTx requires passage through an acidic vesicle for activity; at acidic pH, as the pore is inserted into the membrane, LF is translocated and reaches its cytosolic targets. LF is probably directly involved in its routing, by interacting with the lipid membrane. This interaction could involve a conformational change of LF and/or an oligomerization of the protein. LF may have the capability of partially unfolding in order to cross the membrane.
Sequence similarities	Belongs to the peptidase M34 family.

Ontologies

Keywords

Biological process	Virulence
Domain	Repeat Signal
Ligand	Metal-binding Zinc
Molecular function	Hydrolase Metalloprotease Protease Toxin
Technical term	3D-structure Complete proteome Direct protein sequencing Plasmid

Sequence annotation (Features)

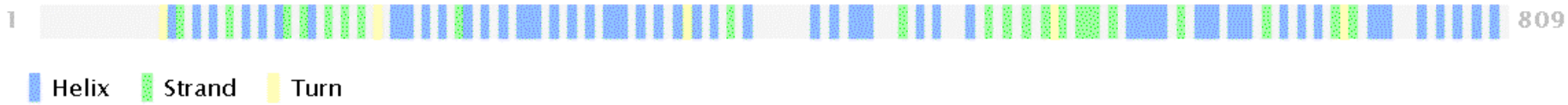
Hide | T

	Feature key	Position(s)	Length	Description	Graphical view
Molecule processing					
	Signal peptide	1 - 33	33		
	Chain	34 - 809	776	Lethal factor	
Regions					
	Repeat	315 - 333	19	1	
	Repeat	342 - 357	16	2	
	Repeat	360 - 378	19	3	
	Repeat	380 - 397	18	4	
	Repeat	399 - 416	18	5	
	Region	34 - 293	260	PA-binding region Potential	
	Region	60 - 295	236	I; PA-binding region Potential	
	Region	296 - 330	35	IIA	
	Region	315 - 416	102	5 X approximate repeats	
	Region	336 - 416	81	III	
	Region	420 - 583	164	IIB	
	Region	585 - 809	225	IV	
Sites					
	Active site	720	1		
	Metal binding	719	1	Zinc (catalytic)	
	Metal binding	723	1	Zinc (catalytic)	
	Metal binding	768	1	Zinc (catalytic)	
Natural variations					
	Natural variant	299	1	A → S in strain: Sterne.	

Experimental info

Mutagenesis	180	1	V → A: No effect on PA-binding ability	
Mutagenesis	181	1	Y → A: Loss of ability to bind to PA	
Mutagenesis	182	1	Y → A: Loss of ability to bind to PA	
Mutagenesis	183	1	E → A: No effect on PA-binding ability	
Mutagenesis	184	1	I → A: Loss of ability to bind to PA	
Mutagenesis	185	1	G → A: No effect on PA-binding ability	
Mutagenesis	186	1	K → A: Loss of ability to bind to PA	
Mutagenesis	220	1	D → A: Loss of ability to bind to PA and loss of toxicity	
Mutagenesis	221	1	L → A: No effect on PA-binding ability and fully toxic	
Mutagenesis	222	1	L → A: No effect on PA-binding ability and fully toxic	
Mutagenesis	223	1	F → A: Loss of ability to bind to PA and non-toxic	
Mutagenesis	719	1	H → A: Loss of activity and zinc binding	
Mutagenesis	720	1	E → C or D: Loss of activity. No effect on zinc binding	
Mutagenesis	723	1	H → A: Loss of activity and zinc binding	

Secondary structure



Details...

Sequences

Sequence	Length	Mass (Da)	
P15917-1 [UniParc].	809	93,770	Blast

Last modified July 5, 2004. Version 2.
Checksum: 2076B4D7277317EE

```
        10         20         30         40         50         60
MNIKKEFIKV ISMSCLVTAI TLSGPVFIPL VQGAGGHGDV GMHVKEKEKN KDENKRKDEE

        70         80         90        100        110        120
RNKTQEEHLK EIMKHIVKIE VKGEEAVKKE AAEKLLEKVP SDVLEMYKAI GGKIYIVDGD

       130        140        150        160        170        180
ITKHISLEAL SEDKKKIKDI YGKDALLHEH YVYAKEGYEP VLVIQSSEDY VENTEKALNV

       190        200        210        220        230        240
YYEIGKILSR DILSKINQPY QKFLDVLNTI KNASDSDGQD LLFTNQLKEH PTDFSVEFLE

       250        260        270        280        290        300
QNSNEVQEVF AKAFAYYIEP QHRDVLQLYA PEAFNYMDKF NEQEINLSLE ELKDQRMLAR

       310        320        330        340        350        360
YEKWEKIKQH YQHWSDSLSE EGRGLLKKLQ IPIEPKKDDI IHSLSQEEKE LLKRIQIDSS

       370        380        390        400        410        420
DFLSTEEKEF LKKLQIDIRD SLSEEEKELL NRIQVDSSNP LSEKEKEFLK KLKLDIQPYD

       430        440        450        460        470        480
INQRLQDTGG LIDSPSINLD VRKQYKRDIQ NIDALLHQSI GSTLYNKIYL YENMNINNLT

       490        500        510        520        530        540
ATLGADLVDS TDNTKINRGI FNEFKKNFKY SISSNYMIVD INERPALDNE RLKWRIQLSP

       550        560        570        580        590        600
DTRAGYLENG KLILQRNIGL EIKDVQIIKQ SEKEYIRIDA KVVPKSKIDT KIQEAQLNIN

       610        620        630        640        650        660
QEWNKALGLP KYTKLITFNV HNRYASNIVE SAYLILNEWK NNIQSDLIKK VTNYLVDGNG

       670        680        690        700        710        720
RFVFTDITLP NIAEQYTHQD EIYEQVHSKG LYVPESRSIL LHGPSKGVEL RNDSEGFIHE

       730        740        750        760        770        780
FGHAVDDYAG YLLDKNQSDL VTNSKKFIDI FKEEGSNLTS YGRTNEAEFF AEAFRLMHST

       790        800
DHAERLKVQK NAPKTFQFIN DQIKFIINS
```

References

Hide | Top

[1] **"Nucleotide sequence and analysis of the lethal factor gene (lef) from Bacillus anthracis."**
Bragg T.S., Robertson D.L.
Gene 81:45-54(1989) [PubMed: 2509294] [Abstract] [Article from publisher]
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA], PROTEIN SEQUENCE OF 34-49.

[2] **"A comparison of Bacillus anthracis sequences."**
Lowe J.
Submitted (APR-1990) to the EMBL/GenBank/DDBJ databases
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA].

[3] **"Sequence and organization of pXO1, the large Bacillus anthracis plasmid harboring the anthrax toxin genes."**
Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K., Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D., Martinez Y., Ricke D., Svensson R., Jackson P.J.
J. Bacteriol. 181:6509-6515(1999) [PubMed: 10515943] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
Strain: Sterne.

[4] **"Comparative genome sequencing for discovery of novel polymorphisms in Bacillus anthracis."**
Read T.D., Salzberg S.L., Pop M., Shumway M.F., Umayam L., Jiang L., Holtzapple E., Busch J.D., Smith K.L., Schupp J.M., Solomon D., Keim P., Fraser C.M.
Science 296:2028-2033(2002) [PubMed: 12004073] [Abstract] [Article from publisher]
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA].
Strain: Ames / isolate Florida / A2012.

[5] **"Bacillus anthracis comparative genomics."**
Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B., Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L., Fraser C.M.
Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases
Cited for: NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
Strain: Ames ancestor.

[6] **"Sequence analysis of the genes encoding for the major virulence factors of Bacillus anthracis vaccine strain 'Carbosap'."**
Adone R., Pasquali P., La Rosa G., Marianelli C., Muscillo M., Fasanella A., Francia M., Ciuchini F.
J. Appl. Microbiol. 93:117-121(2002) [PubMed: 12067380] [Abstract] [Article from publisher]
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 29-809.
Strain: Carbosap and Ferrara.

[7] **"Anthrax lethal factor cleaves the N-terminus of MAPKKs and induces tyrosine/threonine phosphorylation of MAPKs in cultured macrophages."**
Vitale G., Pellizzari R., Recchi C., Napolitani G., Mock M., Montecucco C.
Biochem. Biophys. Res. Commun. 248:706-711(1998) [PubMed: 9703991] [Abstract] [Article from publisher]
Cited for: FUNCTION.

[14] **"Lethal factor active-site mutations affect catalytic activity in vitro."**
Hammond S.E., Hanna P.C.
Infect. Immun. 66:2374-2378(1998) [PubMed: 9573135] [Abstract]
Cited for: MUTAGENESIS OF HIS-719; GLU-720 AND HIS-723.
Strain: Sterne.

[15] **"Involvement of residues 147VYYEIGK153 in binding of lethal factor to protective antigen of Bacillus anthracis."**
Gupta P., Singh A., Chauhan V., Bhatnagar R.
Biochem. Biophys. Res. Commun. 280:158-163(2001) [PubMed: 11162493] [Abstract] [Article from publisher]
Cited for: MUTAGENESIS OF VAL-180; TYR-181; TYR-182; GLU-183; ILE-184; GLY-185 AND LYS-186.
Strain: Sterne.

[16] **"Asp 187 and Phe 190 residues in lethal factor are required for the expression of anthrax lethal toxin activity."**
Singh A., Chauhan V., Sodhi A., Bhatnagar R.
FEMS Microbiol. Lett. 212:183-186(2002) [PubMed: 12113932] [Abstract] [Article from publisher]
Cited for: MUTAGENESIS OF ASP-220; LEU-221; LEU-222 AND PHE-223.
Strain: Sterne.

[17] **"Toxins of Bacillus anthracis."**
Brossier F., Mock M.
Toxicon 39:1747-1755(2001) [PubMed: 11595637] [Abstract] [Article from publisher]
Cited for: REVIEW.

[18] **"Crystal structure of the anthrax lethal factor."**
Pannifer A.D., Wong T.Y., Schwarzenbacher R., Renatus M., Petosa C., Bienkowska J., Lacy D.B., Collier R.J., Park S., Leppla S.H., Hanna P.C., Liddington R.C.
Nature 414:229-233(2001) [PubMed: 11700563] [Abstract] [Article from publisher]
Cited for: X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) IN COMPLEX WITH ZINC IONS AND MAP2K2.

[19] **"The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor."**
Turk B.E., Wong T.Y., Schwarzenbacher R., Jarrell E.T., Leppla S.H., Collier R.J., Liddington R.C., Cantley L.C.
Nat. Struct. Mol. Biol. 11:60-66(2004) [PubMed: 14718924] [Abstract] [Article from publisher]
Cited for: X-RAY CRYSTALLOGRAPHY (3.52 ANGSTROMS) OF 34-809 IN COMPLEX WITH ZINC IONS AND PEPTIDE SUBSTRATE ANALOG.

[20] **"Anthrax lethal factor inhibition."**
Shoop W.L., Xiong Y., Wiltsie J., Woods A., Guo J., Pivnichny J.V., Felcetto T., Michael B.F., Bansal A., Cummings R.T., Cunningham B.R., Friedlander A.M., Douglas C.M., Patel S.B., Wisniewski D., Scapin G., Salowe S.P., Zaller D.M. Hermes J.D.
Proc. Natl. Acad. Sci. U.S.A. 102:7958-7963(2005) [PubMed: 15911756] [Abstract] [Article from publisher]
Cited for: X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 297-809 IN COMPLEX WITH ZINC IONS AND PROTEASE INHIBITOR.

Cross-references

Sequence databases

EMBL	M29081 Genomic DNA. Translation: AAA79216.1. M30210 Genomic DNA. Translation: AAA22569.1. AF065404 Genomic DNA. Translation: AAD32411.1. AE011190 Genomic DNA. Translation: AAM26117.1. AE017336 Genomic DNA. Translation: AAT28913.2. AJ413934 Genomic DNA. Translation: CAC93932.1. AJ413935 Genomic DNA. Translation: CAC93933.1.
PIR	JQ0032.

3D structure databases

PDB	*Structures determined by X-ray crystallography:* 1J7N. Chains A/B map to 34-809. 1JKY. Chain A maps to 34-809. 1PWP. Chains A/B map to 34-809. 1PWQ. Chains A/B map to 34-809. 1PWU. Chains A/B map to 34-809. 1PWV. Chains A/B map to 34-809. 1PWW. Chains A/B map to 34-809. 1YQY. Chain A maps to 297-809.
ModBase	Search...

Protein-protein interaction databases

IntAct	P15917.

Protein family/group databases

MEROPS	M34.001.

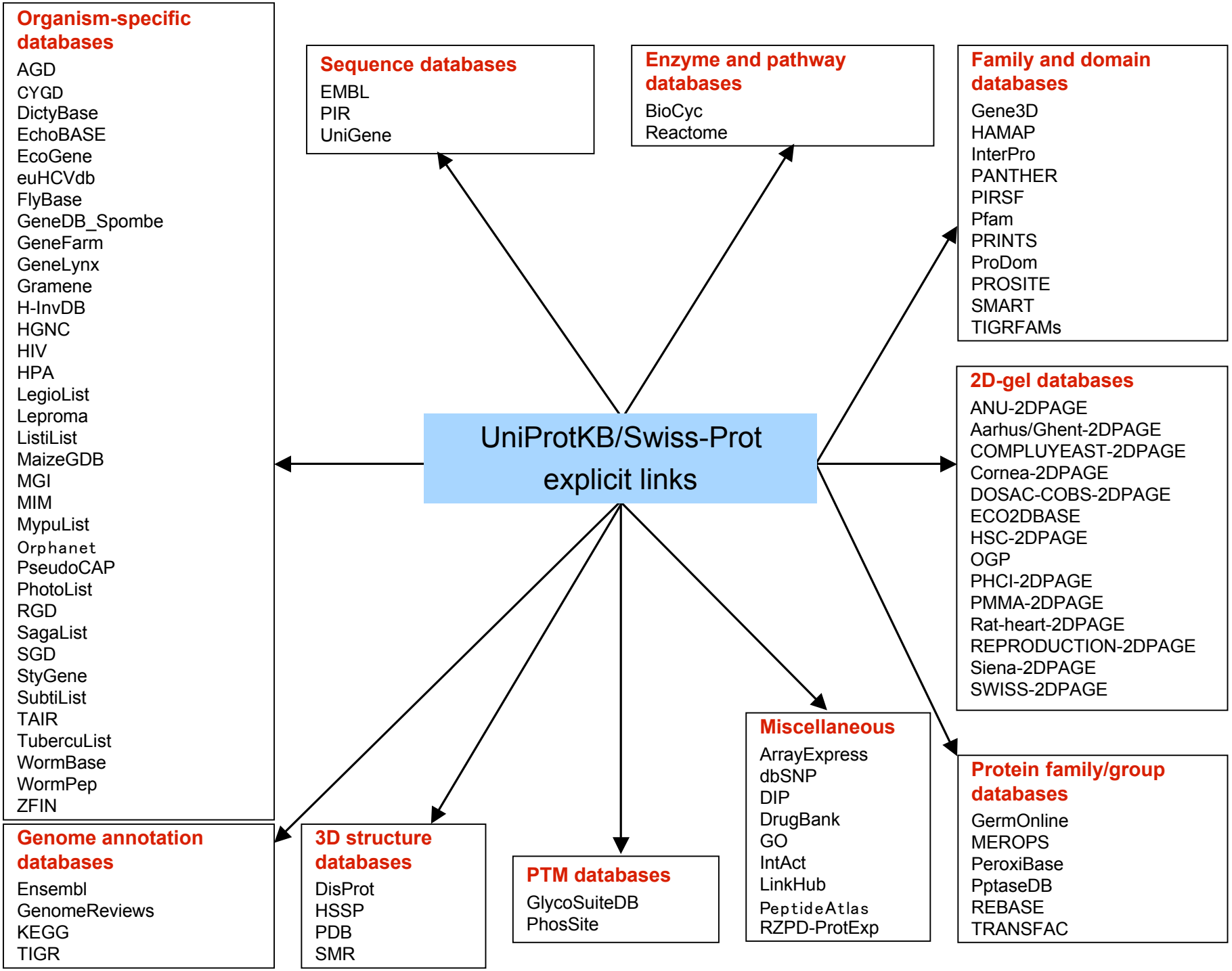
Genome annotation databases

GenomeReviews	Gene locus GBAA_pXO1_0172 in contig AE017336_GR.
KEGG	bar:GBAA_pXO1_0172.

Organism-specific databases	
TIGR	GBAA_pXO1_0172.
HOGENOM	[Family] [Alignment] [Tree]

Family and domain databases	
InterPro	IPR003541. Anthrax_toxinALF_N. IPR006025. Pept_M_Zn_BS. [Graphical view]
Gene3D	G3DSA:3.40.390.10. G3DSA:3.40.390.10. 1 hit. G3DSA:3.90.176.10. G3DSA:3.90.176.10. 2 hits.
Pfam	PF09156. Anthrax-tox_M. 1 hit. PF07737. ATLF. 2 hits. [Graphical view]
PRINTS	PR01392. ANTHRAXTOXNA.
PROSITE	PS00142. ZINC_PROTEASE. 1 hit. [Graphical view]
ProDom	P15917. [Graphical view] [Entries sharing at least one domain]
BLOCKS	Search...

Other Resources	
LinkHub	P15917.
ProtoNet	Search...


UniProtKB/Swiss-Prot explicit links
Organism-specific databases
AGD
CYGD
DictyBase
EchoBASE
EcoGene
euHCVdb
FlyBase
GeneDB_Spombe
GeneFarm
GeneLynx
Gramene
H-InvDB
HGNC
HIV
HPA
LegioList
Leproma
ListiList
MaizeGDB
MGI
MIM
MypuList
Orphanet
PseudoCAP
PhotoList
RGD
SagaList
SGD
StyGene
SubtiList
TAIR
TubercuList
WormBase
WormPep
ZFIN
Sequence databases
EMBL
PIR
UniGene
Enzyme and pathway databases
BioCyc
Reactome
Family and domain databases
Gene3D
HAMAP
InterPro
PANTHER
PIRSF
Pfam
PRINTS
ProDom
PROSITE
SMART
TIGRFAMs
2D-gel databases
ANU-2DPAGE
Aarhus/Ghent-2DPAGE
COMPLUYEAST-2DPAGE
Cornea-2DPAGE
DOSAC-COBS-2DPAGE
ECO2DBASE
HSC-2DPAGE
OGP
PHCI-2DPAGE
PMMA-2DPAGE
Rat-heart-2DPAGE
REPRODUCTION-2DPAGE
Siena-2DPAGE
SWISS-2DPAGE
Genome annotation databases
Ensembl
GenomeReviews
KEGG
TIGR
3D structure databases
DisProt
HSSP
PDB
SMR
PTM databases
GlycoSuiteDB
PhosSite
Miscellaneous
ArrayExpress
dbSNP
DIP
DrugBank
GO
IntAct
LinkHub
PeptideAtlas
RZPD-ProtExp
Protein family/group databases
GermOnline
MEROPS
PeroxiBase
PptaseDB
REBASE
TRANSFAC

Entry information

Entry name	LEF_BACAN
Accession	Primary (citable) accession number: **P15917** Secondary accession number(s): Q8KYJ6, Q933F6
Entry history	Integrated into UniProtKB/Swiss-Prot: April 1, 1990 Last sequence update: July 5, 2004 Last modified: May 1, 2007 This is version 85 of the entry and version 2 of the sequence. [Complete history]
Entry status	Reviewed (UniProtKB/Swiss-Prot)

Relevant documents

PDB cross-references

Index of Protein Data Bank (PDB) cross-references

Peptidase families

Classification of peptidase families and list of entries

SIMILARITY comments

Index of protein domains and families

In a Swiss-Prot entry, you can expect to find:

- All the **names** of a given protein (and of its gene);
- Its biological origin with links to the taxonomic databases;
- A summary of what is known about the protein: function, alternative products, PTM, tissue expression, disease, etc....;
- Selected keywords and ontological descriptions;
- A description of **important sequence features:** domains, PTMs, variations, etc.;
- A selection of **references**;
- Numerous **cross-references;**
- A (often corrected) protein **sequence** and the description of various isoforms/variants.

Annotation projects

- It is not possible to fully annotate all UniProtKB proteins with the current resources;
- It is therefore important to concentrate our efforts in the annotation of proteins that are deemed to be the most important for a majority of users;
- Since 2000 we have initiated a growing number of annotation projects that can be subdivided into 2 distinct subsets:
  - ✓ Horizontal projects that target proteins from specific sets of organisms;
  - ✓ Transversal projects that target aspect of annotations that are common to all horizontal projects (examples: PTMs, 3D-structure, enzymes, etc).

Horizontal annotation projects

The current horizontal projects are targeted towards:

- Mammals (HPI)
- Bacteria and archea (HAMAP)
- Plants (PPAP)
- Fungi (FPAP)
- Viruses
- Insects (mainly Drosophila)
- C.elegans
- Zebrafish
- Xenopus
- Toxins (ToxProt)

Note: the above order reflects the number of annotators involved in the projects. It is not meant to rank their scientific importance/relevance

The UniProt consortium annotators

74 persons are involved in annotation:

49 at SIB, 15 at EBI, 6 at PIR and 4 in Brazil

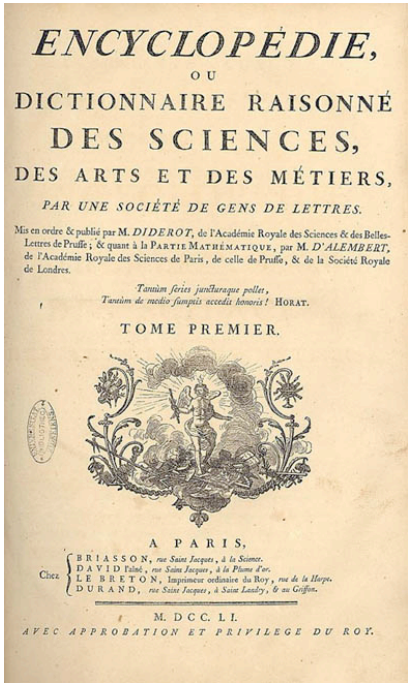
- HPI: Alan, André, Anulka, Bernd, Arnaud, Cecilia, Danielle, Gabriella, Ghislaine, Isabelle, Lionel, Lydie L, Michele, Nadine, Sandra, Serenella, Shyamala, Silvia B, Silvia J, Sorogini, Sylvain, Ursula, Wei Mun, Yasmin
- HAMAP: Andrea, Catherine, Claudia, Elisabeth, Guillaume, Karine, Luciane, Luis, Marisa, Tania, Tatiana, Virginie
- PPAP: Damien, Emmanuel, Michel, Michael
- FPAP: Ivo, Kati, Marc, Vivien
- Viruses: Chantal, Philippe
- ToxProt: Florence, Ruth
- Insects: Eleanor, Sylvain; C.elegans: Duncan
- Zebrafish: Alan, Gill; Xenopus: Alan, Rebecca
- Domains: Anastasia, Christian, Daren, Lai Su, Nicolas, Petra, Virginie
- PTM: Janet, John, Lydie, Nathalie
- 3D: Jules, Sona, Ursula, Vinayaka
- Medical: Arnaud, Livia, Paula
- CVs and taxonomy: Anne, Sandrine, Serenella
- PPI: Bernd; Enzymes: Anne, Kristian; Proteomics: Lydie L.
- Updates/submissions: Claire, Madelaine, Marie-Claude, Michele, Paul, Ruth
- QA: Alan, Amos, Claire, Michele, Sylvain

Note: some people names appears more than once in this list

An important issue…

The process of developing a data resource for the Life Sciences is akin to the work of middle age copists, renaissance encyclopedists or the 19th century OED development….

It is a very tedious, **manually intensive**, **long term** endeavor…



ENCYCLOPÉDIE,
OU
DICTIONNAIRE RAISONNÉ
DES SCIENCES,
DES ARTS ET DES MÉTIERS,
PAR UNE SOCIÉTÉ DE GENS DE LETTRES.

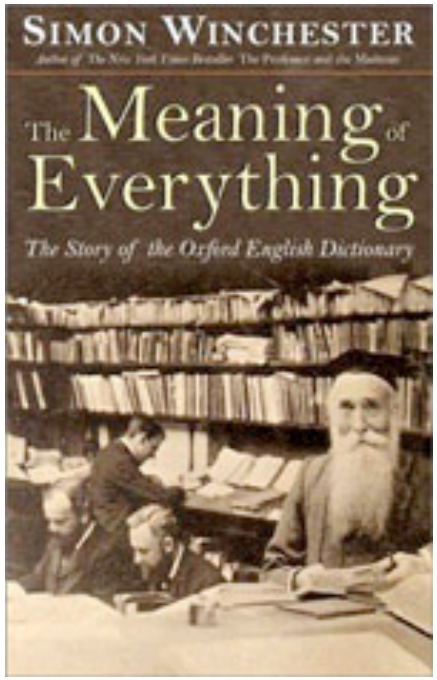
Mis en ordre & publié par M. DIDEROT, de l'Académie Royale des Sciences & des Belles-Lettres de Prusse; & quant à la PARTIE MATHÉMATIQUE, par M. D'ALEMBERT, de l'Académie Royale des Sciences de Paris, de celle de Prusse, & de la Société Royale de Londres.

Tantùm series juncturaque pollet,
Tantùm de medio sumptis accedit honoris! HORAT.

TOME PREMIER.

A PARIS,
Chez BRIASSON, rue Saint Jacques, à la Science.
DAVID l'aîné, rue Saint Jacques, à la Plume d'or.
LE BRETON, Imprimeur ordinaire du Roy, rue de la Harpe.
DURAND, rue Saint Jacques, à Saint Landry, & au Griffon.

M. DCC. LI.
AVEC APPROBATION ET PRIVILEGE DU ROY.



The bacterial «infectome»

In 1995, the first complete sequence of the genome of a microbial organism (*H.influenzae)* became available. Today we have at our disposition the sequence of 500 microbial genomes. This number is currently increasing by about one genome per week.

illustration: Don Smith

Microbial genome and proteomes

UniProtKB/Swiss-Prot Release 52.5 of 15-May-2007: 267354 entries
UniProtKB/TrEMBL Release 35.5 of 15-May-2007: 4361897 entries

Summary statistics				
Type	Proteomes	Total number of entries	Number of UniProtKB/Swiss-Prot entries	Number of UniProtKB/TrEMBL entries
Archaea	36	81782	10292	71490
Bacteria	412	1307918	116281	1191637
Plastids	93	8716	6031	2685
All	541	1398416	132604	1265812

Some human pathogenic bacteria that have been sequenced

• *Bacillus anthracis*	Anthrax
• *Bordetella pertussis*	Whooping cough
• *Borrelia burgdorferi*	Lyme disease
• *Brucella abortus*	Brucellosis
• *Campylobacter jejuni*	Gastroenteritis
• *Chlamydia pneumoniae*	Respiratory tract infections
• *Chlaymida trachomatis*	Trachoma, urogenital infections
• *Escherichia coli O157*	Enterohemorrhagic
• *Haemophilus influenzae*	Respiratory tract infections
• *Helicobacter pylori*	Gastric diseases (ulcers)
• *Mycobacterium leprae*	Leprosy
• *Mycobacterium tuberculosis*	Tuberculosis
• *Mycoplasma genitalium*	Urogenital infections
• *Mycoplasma pneumoniae*	Respiratory tract infections
• *Neisseria gonorhoeae*	Gonorrhea
• *Neisseria meningitidis*	Meningitis
• *Pseudomonas aeruginosa*	Urinary tract infections, burn infections, CF
• *Rickettsia conorii*	Mediterranean spotted fever
• *Rickettsia prowazekii*	Typhus
• *Staphylococcus aureus*	Major hospital acquired infections
• *Streptococcus pneumoniae*	Acute respiratory infections
• *Streptococcus pyogenes*	Scarlet fever, septicemia, etc.
• *Treponema pallidum*	Syphilis
• *Ureaplasma urealyticum*	Urogenital infections
• *Vibrio cholerae*	Cholera

So what does HAMAP means?

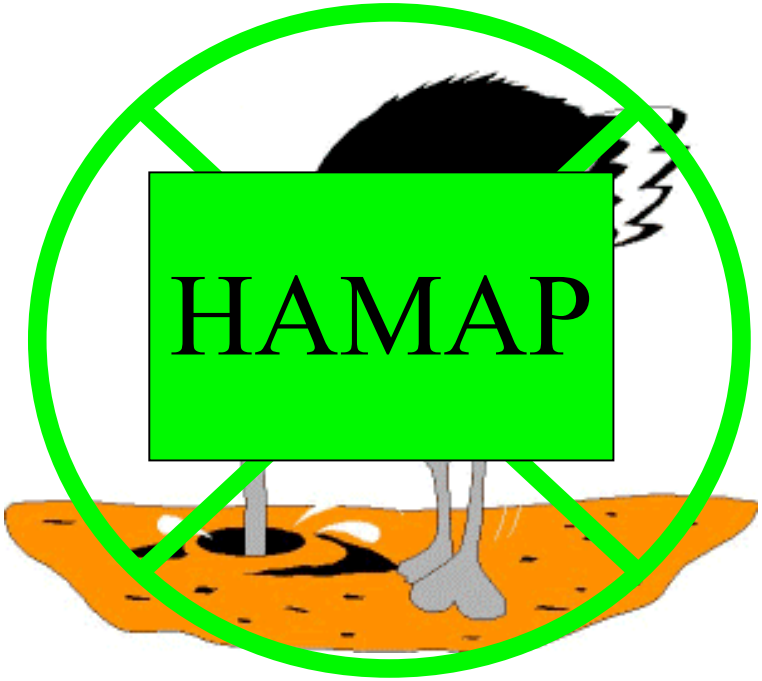
***H*igh quality**

***A*utomated and**

***M*anual**

***A*nnotation of microbial**

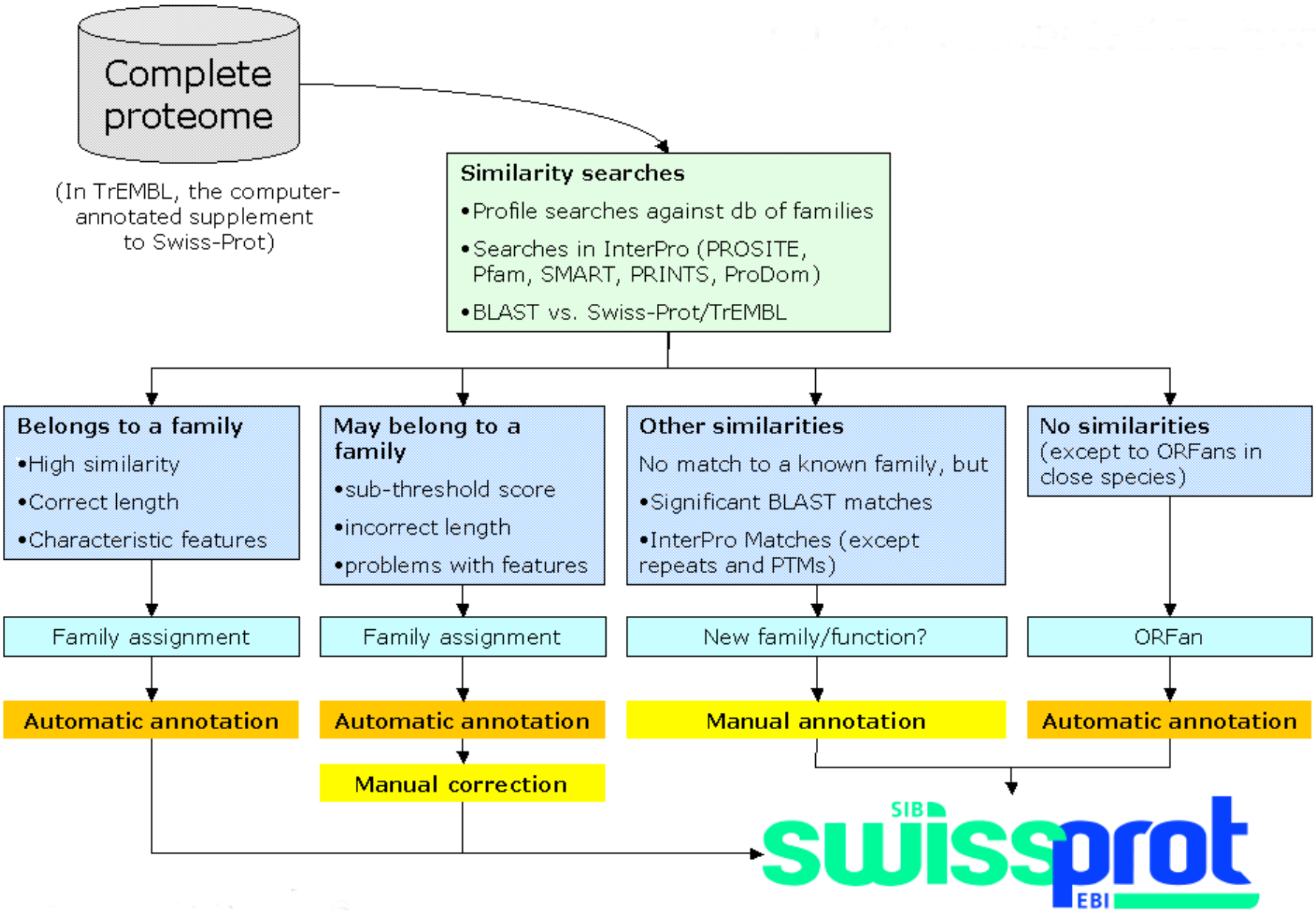
***P*roteomes**



Lots of microbial genomes, lots of proteins. What should we do with them in UniProt?

Automatic annotation of proteins belonging to specified families

- Allows to annotate automatically, yet with a very high level of quality, proteins that belong to well defined protein families;
- Can be applied to both characterized families and to some UPF's (Uncharacterized Protein Family);
- This projects requires the continuous development or adaptation of software tools as well as the development of a database of annotation rules for each type of specified microbial protein (so far about 1'400).


Complete proteome
(In TrEMBL, the computer-annotated supplement to Swiss-Prot)
Similarity searches
•Profile searches against db of families
•Searches in InterPro (PROSITE, Pfam, SMART, PRINTS, ProDom)
•BLAST vs. Swiss-Prot/TrEMBL
Belongs to a family
•High similarity
•Correct length
•Characteristic features
May belong to a family
•sub-threshold score
•incorrect length
•problems with features
Other similarities
No match to a known family, but
•Significant BLAST matches
•InterPro Matches (except repeats and PTMs)
No similarities
(except to ORFans in close species)
Family assignment
Family assignment
New family/function?
ORFan
Automatic annotation
Automatic annotation
Manual annotation
Automatic annotation
Manual correction
SIB
swissprot
EBI

HAMAP annotation rule: MF_00784

[?] General information about the entry

Accession	MF_00784
Dates	18-MAY-2004 (Created) 17-OCT-2006 (Last updated, Version 9)
Data class	Protein; auto
Predictors	HAMAP; MF_00784 ;[distribution of match scores in UniProtKB];[seed alignment for MF_00784]
Identifier	AGRB
	[case <OC:Staphylococcus>]
Description	Accessory gene regulator protein B (EC 3.4.-.-)
Gene name	agrB
	[case not <OC:Staphylococcus>]
Description	Putative agrB-like protein (EC 3.4.-.-)

[?] Comments

- **[case <OC:Staphylococcus>]**
  - ***FUNCTION*:** Essential for the production of a quorum sensing system signal molecule, the autoinducing peptide (AIP). This quorum sensing system is responsible for the regulation of the expression of virulence factor genes. Involved in the proteolytic processing of agrD, the precursor of AIP (By similarity).
- **[case not <OC:Staphylococcus>]**
  - ***FUNCTION*:** May be involved in the proteolytic processing of a quorum sensing system signal molecule precursor (Potential).
- ***SUBCELLULAR LOCATION*:** Cell membrane; multi-pass membrane protein (Potential).
- ***SIMILARITY*:** Belongs to the agrB family.

[?] Cross-references

Pfam	PF04647; AgrB; 1
General	Transmembrane; -; 3-5

[?] Keywords and Gene Ontology

- **Keyword: Quorum sensing**
- **Keyword: Hydrolase**
- **Keyword: Protease**
- **Keyword: Membrane**
- **Keyword: Transmembrane**
- **Keyword: Virulence [case <OC:Staphylococcus>]**
- GO:0008233; Molecular function: peptidase activity.
- GO:0009372; Biological process: quorum sensing.

[?] Characteristics

- Size range: 187-242 amino acids
- Related UniRules: None
- Template: P0C1P7
- Fusion: N-terminal: None; C-terminal: None
- Duplicate: in CLOPE
- Plasmid encoded: None

[?] Sets of member sequences

Bacteria	[25]
All	**[25]**

Listeriaceae

LISIN	Listeria innocua	(AGRB_LISIN)
LISMO	Listeria monocytogenes	(AGRB_LISMO)
LISMF	Listeria monocytogenes serotype 4b (strain F2365)	(AGRB_LISMF)
LISW6	Listeria welshimeri serovar 6b (strain ATCC 35897 / DSM...	(not yet verified: A0AEM5_LISW6)

Staphylococcus

STAAC	Staphylococcus aureus (strain COL)	(AGRB_STAAC)
STAAR	Staphylococcus aureus (strain MRSA252)	(AGRB_STAAR)
STAAS	Staphylococcus aureus (strain MSSA476)	(AGRB_STAAS)
STAAW	Staphylococcus aureus (strain MW2)	(AGRB_STAAW)
STAAM	Staphylococcus aureus (strain Mu50 / ATCC 700699)	(AGRB_STAAM)
STAAN	Staphylococcus aureus (strain N315)	(AGRB_STAAN)
STAA8	Staphylococcus aureus (strain NCTC 8325)	(AGRB_STAA8)
STAA3	Staphylococcus aureus (strain USA300)	(not yet verified: Q2FF88_STAA3)
STAAB	Staphylococcus aureus (strain bovine RF122)	(not yet verified: Q2YUD1_STAAB)
STAES	Staphylococcus epidermidis (strain ATCC 12228)	(AGRB_STAES)
STAEQ	Staphylococcus epidermidis (strain ATCC 35984 / RP62A)	(AGRB_STAEQ)
STAHJ	Staphylococcus haemolyticus (strain JCSC1435)	(AGRB_STAHJ)
STAS1	Staphylococcus saprophyticus subsp. saprophyticus (stra...	(AGRB_STAS1)

Clostridia

Clostridiales

Clostridiaceae

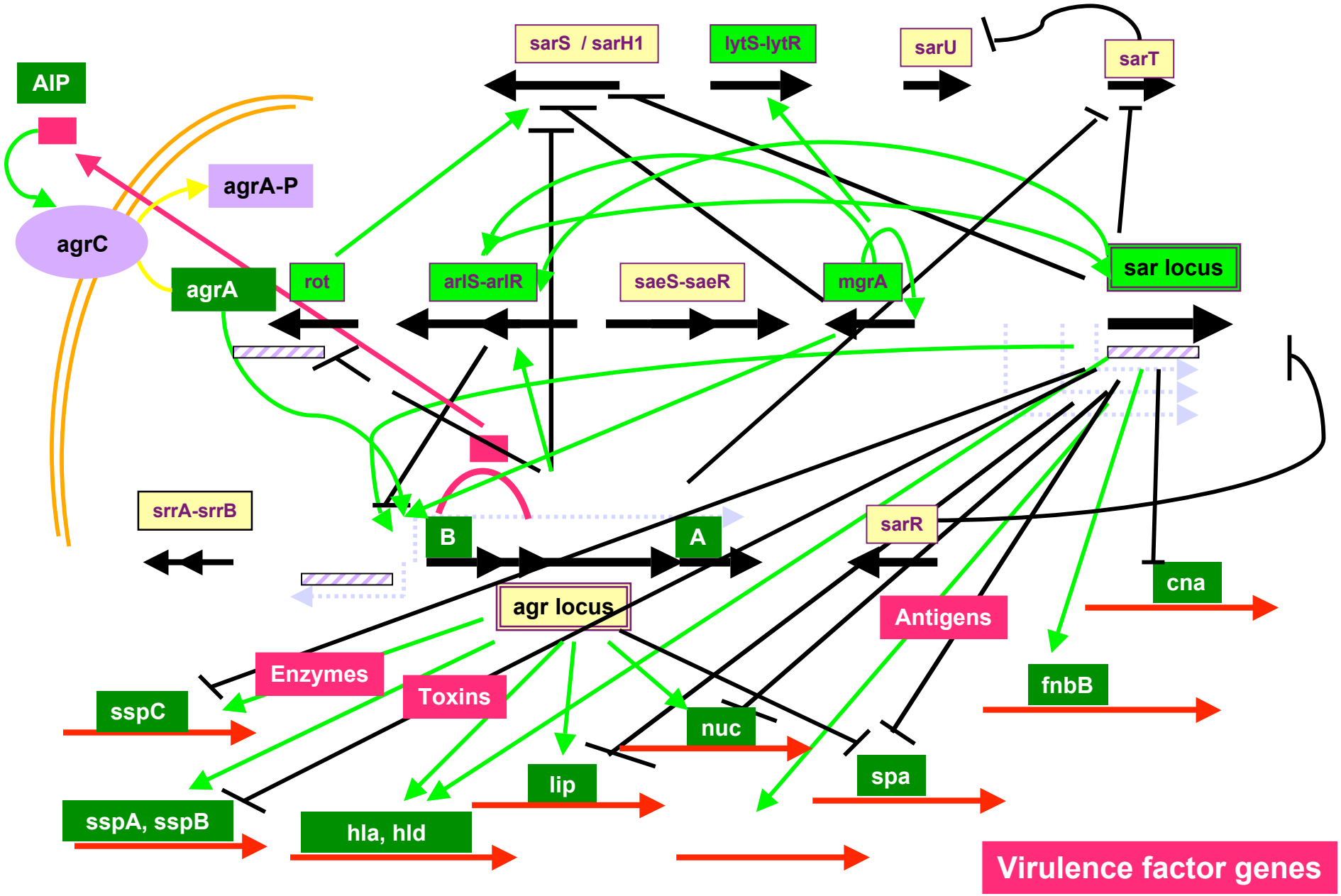
CLOAB	Clostridium acetobutylicum	(AGRB_CLOAB)
CLOD6	Clostridium difficile (strain 630)	(weak match found below threshold: Q183I4_CLOD6)
CLODQ	Clostridium difficile (strain QCD-32g58)	(-)
CLONN	Clostridium novyi (strain NT)	(weak match found below threshold: A0Q0I1_CLONN)
CLOPE	Clostridium perfringens	(AGRB1_CLOPE, AGRB2_CLOPE)
CLOP1	Clostridium perfringens (strain ATCC 13124 / NCTC 8237 ...	(not yet verified: Q0TQ43_CLOP1, Q0TSS5_CLOP1)
CLOPS	Clostridium perfringens (strain SM101 / Type A)	(not yet verified: Q0SSQ8_CLOPS)
CLOTE	Clostridium tetani	(-)
CLOTH	Clostridium thermocellum (strain ATCC 27405 / DSM 1237)	(-)

Summary Statistics	
Number of HAMAP families	1392
Number of alignments	1439
Number of profiles	1461
Coverage of UniProtKB/Swiss-Prot entries	106718
Number of families by taxonomic scopes	• Archaea: 528 • Bacteria: 1216 • Plastid: 132 • Archaea only: 175 • Archaea+Bacteria: 309 • Archaea+Bacteria+Plastid: 44 • Bacteria only: 776 • Bacteria+Plastid: 87 • Plastid only: 1

Using HAMAP, we can currently annotate to Swiss-Prot quality level between 10% to 50% of a complete microbial proteome

But proteins involved in virulence can rarely be annotated in an automated process as there are often species specific or because their implication in virulence is not their 'original' function.

GLOBAL REGULATION OF *Staphylococcus aureus* VIRULENCE FACTORS



Virus annotation program

- Established in 2004; currently 2 persons, but we are currently hiring a 3rd person;
- Goal:
  - Annotate viral proteins with an emphasis on important human, animal and plant pathogens;
  - In collaboration with NCBI and ICTV help to put some order in the taxonomic ‘mess’ that is the hallmark of virus classification and strain naming systems;
  - Create a virus-specific portal to help virologists use the knowledge that is and will be provided in UniProtKB/Swiss-Prot.

What has been already being achieved in term of annotation

- Coronaviruses (including SARS);
- Dengue virus;
- Ebolavirus;
- Hepatitis C virus (in collaboration with IBP – Lyon);
- Human retroviruses (HIV-1, HIV-2, HTLV and spumavirus);
- Influenza types A and B viruses;
- Rhabdoviruses;
- Togaviridae family, including Chikungunya virus, Rubella virus, Semliki forest virus and Sindbis virus;
- Yellow fever virus;
- Spumaviruses;
- Hendra and Nipah viruses (Paramyxoviridae);
- Mimivirus;
- Birnaviruses;
- Porcine circoviruses

Taxonomic issues

- In 2006 we introduced a new line type, OH (Organism Host) in order to indicate the host(s) in viral protein entries;
- Clean up of the classification of viruses in the NCBI taxonomy. Examples: hepatitis C genotypes, dengue isolates, etc.;
- We will soon implement cross-reference to the ICTV taxonomic database.

```
OS   Chandipura virus (strain I653514) (CHPV).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Rhabdoviridae; Dimarhabdovirus supergroup;
OC   Vesiculovirus.
OX   NCBI_TaxID=11273;
OH   NCBI_TaxID=9606; Homo sapiens (Human).
OH   NCBI_TaxID=7198; Phlebotominae (sandflies).
```

Rhabdoviridae [family]

The UniProtKB/Swiss-Prot virus portal (in development)

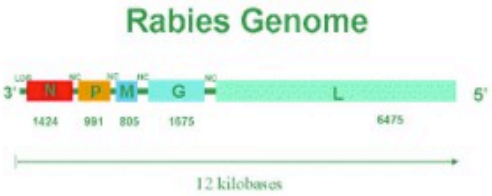
Taxonomy Id:

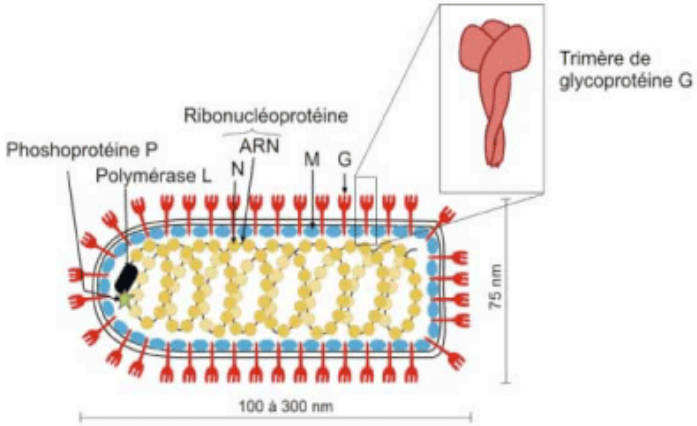
11270

Wiki:

Molecular biology

VIRION





Enveloped, bullet shaped. 180 nm long and 75 nm wide.

Certain plant rhabdoviruses are bacilliform in shape and almost twice the length.

GENOME

Negative-stranded RNA linear genome, about 11-15 kb in size. Encodes for 5 to six proteins.

GENE EXPRESSION

The L protein binds the encapsidated genome at the leader region, then subsequently transcribes each genes by recognizing start and stop signals flanking each gene. mRNAs are capped and polyadenylated by the L protein during synthesis.

REPLICATION

CYTOPLASMIC

1. Virus attaches to host receptors though G glycoprotein and is endocytosed into vesicles in the host cell.
2. Fusion of virus membrane with the vesicle membrane; RNA genome is released into the cytoplasm.
3. Sequential transcription of the genome RNA complexed with N protein yield viral mRNAs.
4. Genomic RNA replication involves synthesis of a full-length positive-sense complementary ssRNA.
5. Encapsidation of newly replicated genomic RNA by N protein, simultaneous condensation of the ribonucleocapsid core by M protein and association with the plasma membrane

TAXONOMY

Group V; ssRNA negative-strand viruses
Order: Mononegavirales *Genus*: Cytorhabdovirus, Ephemerovirus, Lyssavirus, Novirhabdovirus, Nucleorhabdovirus, Vesiculovirus

TYPE SPECIES

Rabies virus

REPRESENTATIVE SPECIES

HOST

Vertebrates; invertebrates; plants.

CELL TROPISM

Rabies virus replicates in neurons

Epidemiology

GEOGRAPHY

Rabies is present in all continents except for Australia and Antarctica.

ASSOCIATED DISEASES

Rabies

TRANSMISSION

Rabies virus: animal bites. VSV: transcutaneous route. Ephemeroviruses, Cytorhabdoviruses, Nucleorhabdoviruses: [arboviruses].

VACCINE

Rabies virus

ANTIVIRAL DRUGS

Filoviridae

Matching UniProtKB/Swiss-Prot entries

Grouped by proteins (reorder by species)

Large structural protein (L protein) (Transcriptase) (Replicase)

Align Retrieve

- L_MABVM Lake Victoria marburgvirus (strain Musoke-80) (MARV) (Marburg virus)
  Large structural protein (L protein) (Transcriptase) (Replicase)
- L_MABVP Lake Victoria marburgvirus (strain Popp-67) (MARV) (Marburg virus)
  Large structural protein (L protein) (Transcriptase) (Replicase)
- L_EBORE Reston ebolavirus (strain Philippines-96) (REBOV) (Reston Ebola virus)
  Large structural protein (L protein) (Transcriptase) (Replicase)
- L_EBORR Reston ebolavirus (strain Reston-89) (REBOV) (Reston Ebola virus)
  Large structural protein (L protein) (Transcriptase) (Replicase)
- L_EBOSM Sudan ebolavirus (strain Maleo-79) (SEBOV) (Sudan Ebola virus)
  Large structural protein (L protein) (Transcriptase) (Replicase)
- L_EBOSU Sudan ebolavirus (strain Uganda-00) (SEBOV) (Sudan Ebola virus)
  Large structural protein (L protein) (Transcriptase) (Replicase)
- L_EBOZ5 Zaire ebolavirus (strain Kikwit-95) (ZEBOV) (Zaire Ebola virus)
  Large structural protein (L protein) (Transcriptase) (Replicase)
- L_EBOZM Zaire ebolavirus (strain Mayinga-76) (ZEBOV) (Zaire Ebola virus)
  Large structural protein (L protein) (Transcriptase) (Replicase)

Nucleoprotein (Nucleocapsid protein)

Align Retrieve

- NCAP_MABVM Lake Victoria marburgvirus (strain Musoke-80) (MARV) (Marburg virus)
- NCAP_MABVP Lake Victoria marburgvirus (strain Popp-67) (MARV) (Marburg virus)
- NCAP_EBORE Reston ebolavirus (strain Philippines-96) (REBOV) (Reston Ebola virus)
- NCAP_EBORR Reston ebolavirus (strain Reston-89) (REBOV) (Reston Ebola virus)
- NCAP_EBOSB Sudan ebolavirus (strain Boniface-76) (SEBOV) (Sudan Ebola virus)
- NCAP_EBOSU Sudan ebolavirus (strain Uganda-00) (SEBOV) (Sudan Ebola virus)
- NCAP_EBOG4 Zaire ebolavirus (strain Gabon-94) (ZEBOV) (Zaire Ebola virus)
- NCAP_EBOZ5 Zaire ebolavirus (strain Kikwit-95) (ZEBOV) (Zaire Ebola virus)
- NCAP_EBOZM Zaire ebolavirus (strain Mayinga-76) (ZEBOV) (Zaire Ebola virus)

Reviewed, UniProtKB/Swiss-Prot **Q05320** (VGP_EBOZM)

Last modified May 1, 2007. Version 53. History...

Contribute
Send feedback
WikiProteins

Clusters with 100%, 90%, 50% identity | Documents (2) | Customize display

TEXT XML RDF/XML FASTA

Names and origin · General annotation (Comments) · Ontologies · Sequence annotation (Features) · Sequences · References · Cross-references · Entry information · Relevant documents

Names and origin

Hide | Top

Protein names	**Envelope glycoprotein** [Precursor] *Also known as:* GP1,2 GP *Cleaved into:* **GP1** **GP2** **GP2-delta**
Gene names	Name: **GP**
Organism	**Zaire ebolavirus (strain Mayinga-76) (ZEBOV) (Zaire Ebola virus)**
Taxonomic identifier	128952 [NEWT] [NCBI]
Taxonomic lineage	Viruses › ssRNA negative-strand viruses › Mononegavirales › Filoviridae › Ebola-like viruses › Zaire ebolavirus
Virus host	Homo sapiens (Human) [TaxID: 9606] Epomops franqueti (Franquet's epauleted bat) [TaxID: 77231] Myonycteris torquata (Little collared fruit bat) [TaxID: 77243]

General annotation (Comments)

Hide | Top

Function	GP1 is responsible for binding to the receptor(s) on target cells. Interacts with CD209/DC-SIGN and CLEC4M/DC-SIGNR which act as cofactors for virus entry into the host cell. Binding to CD209 and CLEC4M, which are respectively found on dendritic cells (DCs), and on endothelial cells of liver sinusoids and lymph node sinuses, facilitate infection of macrophages and endothelial cells. These interactions not only facilitate virus cell entry, but also allow capture of viral particles by DCs and subsequent transmission to susceptible cells without DCs infection (trans infection). Binding to the macrophage specific lectin CLEC10A also seem to enhance virus infectivity. Interaction with FOLR1/folate receptor alpha may be a cofactor for virus entry in some cell types, although results are contradictory. After internalization of the virus into the endosomes of the host cell, proteolysis of GP1 by two cysteine proteases, CTSB/cathepsin B and CTSL/cathepsin L presumably induces a conformational change of GP2, unmasking its fusion peptide and initiating membranes fusion.

	GP2 acts as a class I viral fusion protein. Under the current model, the protein has at least 3 conformational states: pre-fusion native state, pre-hairpin intermediate state, and post-fusion hairpin state. During viral and target cell membrane fusion, the coiled coil regions (heptad repeats) assume a trimer-of-hairpins structure, positioning the fusion peptide in close proximity to the C-terminal region of the ectodomain. The formation of this structure appears to drive apposition and subsequent fusion of viral and target cell membranes. Responsible for penetration of the virus into the cell cytoplasm by mediating the fusion of the membrane of the endocytosed virus particle with the endosomal membrane. Low pH in endosomes induces an irreversible conformational change in GP2, releasing the fusion hydrophobic peptide. GP1,2 mediates endothelial cell activation and decreases endothelial barrier function. Mediates activation of primary macrophages. At terminal stages of the viral infection, when its expression is high, GP1,2 down-modulates the expression of various host cell surface molecules that are essential for immune surveillance and cell adhesion. Down-modulates integrins ITGA1, ITGA2, ITGA3, ITGA4, ITGA5, ITGA6, ITGAV and ITGB1. GP1,2 alters the cellular recycling of the dimer alpha-V/beta-3 via a dynamin-dependent pathway. Decrease in the host cell surface expression of various adhesion molecules may lead to cell detachment, contributing to the disruption of blood vessel integrity and hemorrhages developed during Ebola virus infection (cytotoxicity). This cytotoxicity appears late in the infection, only after the massive release of viral particles by infected cells. Down-modulation of host MHC-I, leading to altered recognition by immune cells, may explain the immune suppression and inflammatory dysfunction linked to Ebola infection. Also down-modulates EGFR surface expression. GP2delta is part of the complex GP1,2delta released by host ADAM17 metalloprotease. This secreted complex may play a role in the pathogenesis of the virus by efficiently blocking the neutralizing antibodies that would otherwise neutralize the virus surface glycoproteins GP1,2. Might therefore contribute to the lack of inflammatory reaction seen during infection in spite the of extensive necrosis and massive virus production. GP1,2delta does not seem to be involved in activation of primary macrophages.
Subunit structure	Homotrimer; each monomer consists of a GP1 and a GP2 subunit linked by disulfide bonds. The resulting peplomers (GP1,2) protrude from the virus surface as spikes. GP1 and GP2delta are part of GP1,2delta soluble complexes released by ectodomain shedding. GP1,2 interacts with host integrin ITGAV/alpha-V and CLEC10A. Also binds human CD209 and CLEC4M (collectively referred to as DC-SIGN(R)), as well as human FOLR1.
Subcellular location	GP2: Virion; virion membrane; single-pass type I membrane protein. Virion; virion membrane; lipid-anchor. Cell membrane; single-pass type I membrane protein. Cell membrane; lipid-anchor. GP1: Virion; virion membrane; peripheral membrane protein. Cell membrane; peripheral membrane protein. GP1,2-delta: Secreted protein. Note=GP1 is not anchored to the viral envelope, but associates with the extravirion surface through its binding to GP2. In the cell, both GP1 and GP2 localize to the plasma membrane lipid rafts, which probably represent the assembly and budding site. GP1 can also be shed after proteolytic processing. GP1,2-delta is shed by the virus after proteolytic cleavage of GP1,2 by host ADAM17.
Domain	The mucin-like region seems to be involved in the cytotoxic function. This region is also involved in binding to human CLEC10A. The coiled coil regions play a role in oligomerization and fusion activity.
Post-translational modification	The signal peptide region modulates GP's high mannose glycosylation, thereby determining the efficiency of the interactions with DC-SIGN(R). N-glycosylated. O-glycosylated in the mucin-like region. Palmitoylation of GP2 is not required for its function.

Ontologies

Keywords

Biological process	Fusion protein Host-virus interaction Viral immunoevasion
Cellular component	Envelope protein Membrane Virion protein
Coding sequence diversity	RNA editing
Domain	Coiled coil Signal Transmembrane
PTM	Cleavage on pair of basic residues Glycoprotein Lipoprotein Palmitate
Technical term	3D-structure

Sequence annotation (Features)

	Feature key	Position(s)	Length	Description
Molecule processing				
☐	Signal peptide	1 – 32	32	Potential
☐	Chain	33 – 676	644	Envelope glycoprotein
☐	Chain	33 – 501	469	GP1
☐	Chain	502 – 676	175	GP2
☐	Chain	502 – 637	136	GP2-delta

[26] **"Effects of Ebola virus glycoproteins on endothelial cell activation and barrier function."**
Wahl-Jensen V.M., Afanasieva T.A., Seebach J., Stroeher U., Feldmann H., Schnittler H.J.
J. Virol. 79:10442-10450(2005) [PubMed: 16051836] [Abstract] [Article from publisher]
Cited for: FUNCTION IN ENDOTHELIAL CELLS ACTIVATION.

[27] **"Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection."**
Chandran K., Sullivan N.J., Felbor U., Whelan S.P., Cunningham J.M.
Science 308:1643-1645(2005) [PubMed: 15831716] [Abstract] [Article from publisher]
Cited for: PROTEOLYSIS OF GP1.

[28] **"Role of Ebola virus secreted glycoproteins and virus-like particles in activation of human macrophages."**
Wahl-Jensen V., Kurz S.K., Hazelton P.R., Schnittler H.J., Stroeher U., Burton D.R., Feldmann H.
J. Virol. 79:2413-2419(2005) [PubMed: 15681442] [Abstract] [Article from publisher]
Cited for: FUNCTION OF GP1,2DELTA.

[29] **"Ebola virus glycoprotein toxicity is mediated by a dynamin-dependent protein-trafficking pathway."**
Sullivan N.J., Peterson M., Yang Z.-Y., Kong W.-P., Duckers H., Nabel E., Nabel G.J.
J. Virol. 79:547-553(2005) [PubMed: 15596847] [Abstract] [Article from publisher]
Cited for: DOWN-MODULATION OF HOST INTEGRIN DIMER ALPHA-V/BETA-3, INTERACTION WITH HUMAN INTEGRIN ITGAV.

[30] **"Role of endosomal cathepsins in entry mediated by the Ebola virus glycoprotein."**
Schornberg K., Matsuyama S., Kabsch K., Delos S., Bouton A., White J.
J. Virol. 80:4174-4178(2006) [PubMed: 16571833] [Abstract] [Article from publisher]
Cited for: PROTEOLYSIS OF GP1.

[31] **"Ebola virus glycoprotein GP is not cytotoxic when expressed constitutively at a moderate level."**
Alazard-Dany N., Volchkova V., Reynard O., Carbonnelle C., Dolnik O., Ottmann M., Khromykh A., Volchkov V.E.
J. Gen. Virol. 87:1247-1257(2006) [PubMed: 16603527] [Abstract] [Article from publisher]
Cited for: FUNCTION.

[32] **"The signal peptide of the ebolavirus glycoprotein influences interaction with the cellular lectins DC-SIGN and DC-SIGNR."**
Marzi A., Akhavan A., Simmons G., Gramberg T., Hofmann H., Bates P., Lingappa V.R., Poehlmann S.
J. Virol. 80:6305-6317(2006) [PubMed: 16775318] [Abstract] [Article from publisher]
Cited for: FUNCTION OF SIGNAL PEPTIDE.

[33] **"Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-A resolution."**
Malashkevich V.N., Schneider B.J., McNally M.L., Milhollen M.A., Pang J.X., Kim P.S.
Proc. Natl. Acad. Sci. U.S.A. 96:2662-2667(1999) [PubMed: 10077567] [Abstract] [Article from publisher]
Cited for: X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 557-630.

Protopap

Protozoan proteomes annotation program

Mission

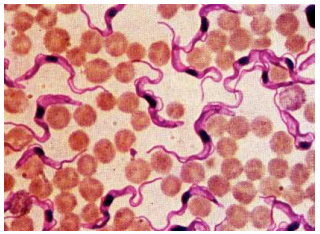
- Annotate proteins orginating from a variety of pathogenic protozoan species;
- The program should concentrate on proteins for which there are published reports;
- It is open-ended (like all other annotaton programs), but we are targeting for a first 3 year funding period.

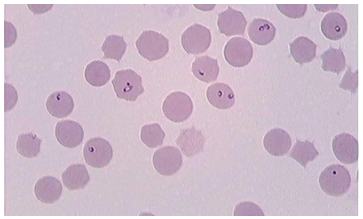
Who and where?

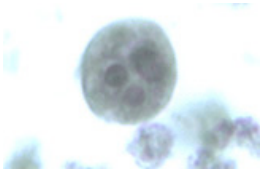
- Have a number of annotators in various countries (Brazil, Cuba?, Mexico?, Kenya, South Africa? and Tunisia?) and at least one in Geneva;
- Scientific collaborations with labs in various tropical countries that work with these pathogenic protozoans;
- Coordination with annotation efforts (at genome level) carried out by the pathogen sequencing unit of the Sanger Center.

Proteins from what species?

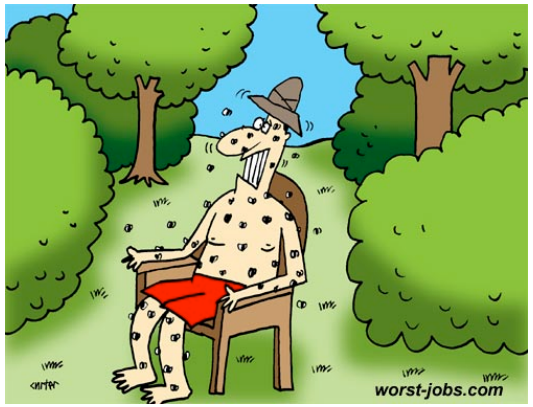
- *Plasmodium falciparum* and related species;
- *Trypanosoma brucei* and *cruzi*;
- *Leishmania major* and related species;
- *Entamoeba histolytica;*
- *Theileria parva* and *annulata;*
- As our efforts will be driven by how we will manage to get funded and by whom is willing to be a long term partner, the above list is going to change.





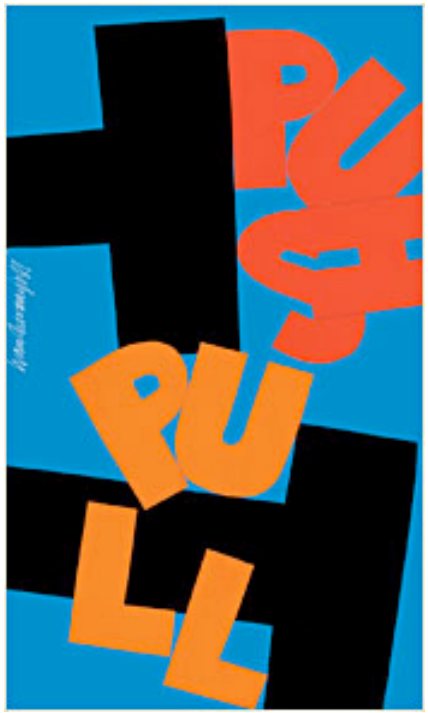


And what about the vectors?

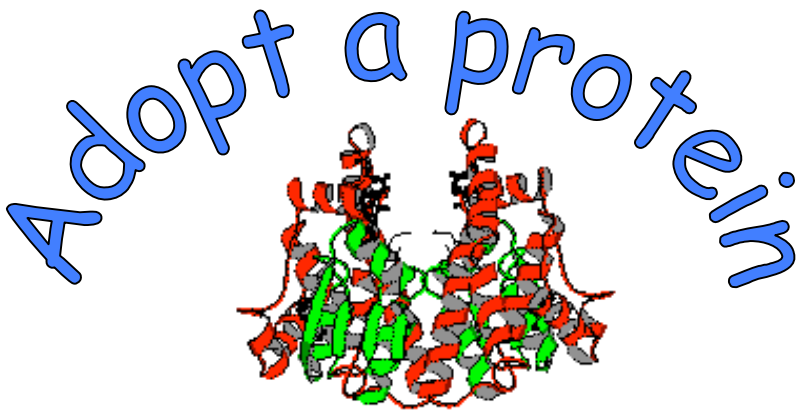


- All those flying and biting insects?;
- Here also we are in a poor shape: we have 2’550 annotated Drosophila melanogaster entries, 640 from other Drosophila, 170 Bombyx, …;
- but only 122 Anopheles, 45 Aedes and 7 Glossina (6 of them added yesterday!);
- So with only a single insect annotator we are not going to make a significant impact in this important taxonomic kingdom;
- May be we also need to build an international effort.

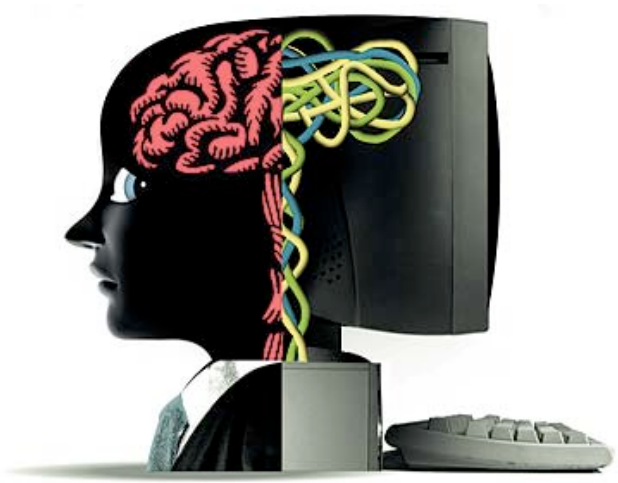
From pull to push..



- For now more than 20 years we have been «pulling» information and knowledge from various sources, but mainly from literature;
- It is now time to make sure that the next 20 years will be defined by the fact that researchers «push» their results and the interpretation of their results in the knowledgebase.



- Attempt to try to get the community to directly submit information on the proteins that they are studying;
- Using a wikepedia-type model/interface;
- Will first be «field-tested» in the yeast community;
- We are hopeful, yet we are realist: only a small percentage of life researchers will take the time and are altruistic enough to fully participate in such a scheme.



Grey grey matter counts!

- Many life scientists with knowledge of the molecular world and that are computer-proficient are reaching retirement age;
- Some want to continue to play a role in the advancement of research, yet they will not be able to do lab work anymore;
- We should offer them the tools necessary for them to contribute to the annotation process.

Education!

- Everyone should feel concerned;
- Awareness of the content and usage of knowledge resources is a pre-requisite to do any type of « serious » research in the field of molecular life sciences;
- Organizations such as EMBNet, EBI, SIB, NCBI, NIG, HUPO, ICGEB, WHO should continue and strenghten their «outreach» efforts;
- We (databases providers) should do more in term of providing tutorials (on-line and on-site).

Issue 82 May 2007

THE POWER BEHIND PAIN

by Vivienne Baillie Gerritsen

[PDF]

We feel pain for a reason. Either to be informed of something that is likely to hurt us more unless we turn our backs on it, or of something that has gone wrong inside us. It is a sensation that has been evolving over millions of years, from yeast to man. Pain is multiple. Understanding its vocabulary and intricate syntax can shed light on what it is, why it is and how it could be countered. Detected by receptors, the sensation of pain can be kick-started from any part of our body. The TRP receptors are a family of such receptors, activated by an array of pain stimuli. They can detect hordes of different noxious chemical compounds but also environmental sensations such as extreme heat and cold. One particular TRP receptor – TRPA1 – comes as a surprise because, unlike many of the other TRP family members, it can detect multiple sensations leading to pain, as opposed to only one.

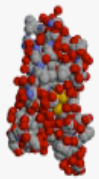
«Over time and as a means of defense, Nature has devised the most diverse ways of hurting. Snakes spit venom. Nettles sting. Bacteria puncture. And dogs bite. »

Over time and as a means of defense, Nature has devised the most diverse ways of hurting. Snakes spit venom. Nettles sting. Bacteria puncture. And dogs bite. However, deprived of the resources to sense pain caused by venom, or a nettle's sting or a dog's fangs, we wouldn't understand the warning that goes with it. Likewise, pain which is caused by something inside us has to be detected so that our attention can be drawn to it. To this end, pain receptors line our body's every nook and cranny, ready to send out a signal which will be relayed to our brain and translated into pain.

The Transient Receptor Potential (TRP) channels – or receptors –

Protein Spotlight (ISSN 1424-4721) is a monthly review written by the Swiss-Prot team of the Swiss Institute of Bioinformatics. Spotlight articles describe a specific protein or family of proteins on an informal tone. This site is hosted on ExPASy.

PROTEIN SNAPSHOT



Tyrannosaurus rex and collagen

Fossils are old because they are made out of stone. Until recently, as far as science was concerned, organic tissues – such as bone matrix – had little chance to survive the passage of time. Organic soft tissues – such as cells and blood vessels for instance – had almost no chance at all. Two years ago though, a 68 million year-old *Tyrannosaurus rex* was unearthed from one thousand cubic metres of sandstone. The mineral from the bone of one of its femurs was removed and, to the scientists' astonishment, they found minute traces of organic soft tissue which had survived millions of years.

[full story]

SEARCH

to all of you in the audience and more specifically to all the organizers who have done an excellent job or organizing this conference

More importantly, I wish good luck to all the efforts to build a solid bioinformatics research AND infrastructure in Africa. I hope the Swiss-Prot group can play a small role in collaborative efforts to annotate proteins from important pathogens

ASBCB